

Pontifícia Universidade Católica de São Paulo  
PUC-SP

Mirella Whiteman

**Os discursos em torno do movimento antivacina no Brasil  
durante a pandemia de Covid-19**

Mestrado em Linguística Aplicada e Estudos da Linguagem

São Paulo  
2024

Pontifícia Universidade Católica de São Paulo  
PUC-SP

Mirella Whiteman

**Os discursos em torno do movimento antivacina no Brasil  
durante a pandemia de Covid-19**

Dissertação apresentada à banca examinadora da Pontifícia Universidade Católica de São Paulo como exigência parcial para obtenção do título de Mestre em Linguística Aplicada e Estudos da Linguagem, sob orientação do Prof. Dr. Antonio Paulo Berber Sardinha.

São Paulo  
2024

Autorizo, exclusivamente para fins acadêmicos e científicos, a reprodução total ou parcial desta dissertação de mestrado por processos fotocopiadores ou eletrônicos.

Assinatura:

Data:

E-mail:

Currículo Lattes:

[Ficha Catalográfica – a ser inserida na versão final pós-defesa]

Mirella Whiteman

Os discursos em torno do movimento antivacina no Brasil durante a pandemia de Covid-19.

Aprovada em: / / .

Dissertação apresentada à Banca Examinadora da Pontifícia Universidade Católica de São Paulo como exigência parcial para obtenção de título de Mestre em Linguística Aplicada e Estudos da Linguagem, sob a orientação do Prof. Dr. Antonio Paulo Berber Sardinha.

Banca Examinadora

---

Profa. Dra. Elisabeth Brait

---

Profa. Dra. Deise Prina Dutra

*A meu marido, Cadu, e meus filhos, Caetano e Matias.*

O presente trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Código de Financiamento 88887.657822/2021-00.

## **Agradecimentos**

É com muita satisfação que expresso aqui meus sinceros agradecimentos a todas as pessoas e instituições que contribuíram de forma significativa para a realização deste trabalho.

Ao meu marido, Cadu, e aos meus queridos filhos, Caetano e Matias, minha fonte inesgotável de amor, apoio e constante encorajamento.

Ao meu estimado orientador, o brilhante Prof. Dr. Tony Berber Sardinha. Sua orientação foi muito além da excelência acadêmica, sendo marcada pela generosidade em compartilhar seu vasto conhecimento em Linguística de Corpus e diversas outras áreas. Foi uma honra poder utilizar uma abordagem de pesquisa pioneira desenvolvida por ele. Expresso minha profunda admiração e respeito por esse professor excepcional, cujos ensinamentos têm sido uma fonte constante de inspiração. Agradeço sinceramente pela oportunidade e pela inestimável mentoria concedida ao longo deste percurso.

À Profa. Dra. Beth Brait, sempre tão elegante e divertida, cuja calorosa acolhida e introdução aos encantos da perspectiva bakhtiniana enriqueceram profundamente meu trabalho. Sua generosidade acadêmica e seu talento extraordinário foram inestimáveis para o desenvolvimento deste estudo.

À Profa. Dra. Deise Dutra, pelo apoio e pelas contribuições valiosas à esta pesquisa. Sua *expertise* e seu comprometimento têm sido fundamentais para o sucesso do nosso projeto sobre discursos infodêmicos.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), por viabilizar minha dedicação integral a este mestrado e pelo seu comprometimento com o avanço da pesquisa científica no Brasil.

À Maria Lúcia dos Reis, por sua dedicação e presteza em ajudar em todas as questões relacionadas ao programa. Agradeço também a todos os funcionários da Pontifícia Universidade Católica de São Paulo por sua atenção e seu profissionalismo.

Ao Grupo de Estudos de Linguística de Corpus (GELC), pelo apoio e pelas colaborações sempre enriquecedoras. Agradeço a todos os membros do grupo por todas as trocas de conhecimento e experiências que tanto contribuíram para o meu desenvolvimento acadêmico.

À minha querida *soul-sister* Arianne Brogini, cuja amizade e alegria contagiante tornaram mais leve e gratificante a jornada desafiadora deste mestrado.

À Aline Zamboni e à Renata Lamberti, pela paciência e pelo carinho demonstrados em todos os momentos deste processo.

À Claudia Delfino, exímia pesquisadora e alma do GELC, cuja liderança e entusiasmo foram fundamentais para a consolidação do grupo. Sem sua incansável dedicação, o GELC não teria alcançado o patamar de excelência e relevância que ostenta hoje.

Ao Carlos Kauffmann, por suas brilhantes reflexões e discussões linguísticas, que expandiram meus horizontes e enriqueceram este trabalho.

Ao inigualável Jeff Buckley, muso e terapeuta de todas as horas.



*The world is multidimensional  
and it gives us headaches.  
We want it to be monochrome  
so it can be clear.  
We want it to be uniform  
so it can be true.  
We want it to be limited  
so we can grasp it, make it stay.  
But its dimensions are endless:  
the world keeps slipping away.*

(Blaga Dimitrova, 1989)

WHITEMAN, Mirella. **Os discursos em torno do movimento antivacina no Brasil durante a pandemia de Covid-19**. 2024. 133 f. Dissertação (Mestrado em Linguística Aplicada e Estudos da Linguagem) – Programa de Linguística Aplicada e Estudos da Linguagem, Pontifícia Universidade Católica de São Paulo, São Paulo, 2024.

### Resumo

A última década testemunhou um aumento significativo dos movimentos antivacina, amplamente impulsionado pela influência das mídias digitais. O ressurgimento de doenças previamente controladas, como sarampo e poliomielite, emergiu como preocupação global, impactando profundamente os profissionais de saúde no Brasil. Desde 2015, observou-se um declínio na cobertura do Programa Nacional de Imunizações (PNI), tendência agravada pela pandemia de Covid-19. Durante o período pandêmico, vacinas e sua eficácia tornaram-se temas centrais nos discursos públicos, intensificados pelo impacto das redes sociais e suas sofisticadas redes de informação, utilizadas para disseminar a hesitação vacinal. Nesse contexto, o objetivo desta pesquisa é investigar o universo discursivo do movimento antivacina brasileiro na plataforma Twitter (atual X), a partir de um referencial teórico-metodológico que compreende a Linguística de Corpus (Berber Sardinha, 2004), a Análise Dialógica do Discurso (Bakhtin, 2016; Volóchinov, 2017) e a Análise Crítica do Discurso (Fairclough, 1989; Van Dijk, 1985; Wodak, 1989). Investigações anteriores em *big data* não direcionaram seu foco para discursos, mas sim para elementos como sentimentos e tópicos (Jang *et al.*, 2021; Wicke; Bolognesi, 2020). Já pesquisas em Análise do Discurso relacionadas à pandemia lidaram com pequenos volumes de dados (Recuero; Stumpf, 2021; Jones, 2021). O uso da Análise Multidimensional Lexical (Berber Sardinha, 2016, 2017, 2020; Berber Sardinha; Fitzsimmons-Doolan, no prelo), abordagem metodológica eleita para este estudo, possibilitou o exame de discursos em corpora de larga escala. Essa abordagem, baseada em cálculos estatísticos, é capaz de detectar grupos de itens lexicais que coocorrem em textos, de modo a revelar discursos recorrentes na linguagem. Para a análise, 16.841 *tweets* foram submetidos à etiquetagem e lematização, seguidas de análise fatorial. A Análise Multidimensional Lexical permitiu a identificação de três dimensões principais, com polos positivos e negativos: alerta sobre riscos cardíacos, males súbitos e falecimentos após doses de reforço, especialmente entre jovens e crianças, considerando a vacina como experimental *versus* questionamento e resistência à

imposição de medidas e decretos compulsórios de combate ao avanço da Covid-19; alegações sobre a origem do vírus em laboratório chinês e interesses de políticos e governos ligados a grandes grupos farmacêuticos *versus* evidências de estudos pseudocientíficos publicados sobre os efeitos adversos das vacinas; e defesa do tratamento precoce frente à letalidade do vírus e à falta de estudos sobre a eficácia e a segurança em longo prazo da vacina emergencial *versus* relatos sobre amigos e terceiros que sofreram com efeitos colaterais graves – dores de cabeça, paralisia, problemas no rim, nas pernas e nos pés – e até fatais após a vacinação. O presente estudo alcançou um grau satisfatório de triangulação teórico-metodológica ao combinar a Linguística de Corpus com duas teorias discursivas complementares. Ademais, proporcionou uma nova perspectiva sobre os discursos antivacina no Brasil, ao identificá-los e analisá-los de forma integrada, compreendendo como eles interagem e se reforçam mutuamente. Por fim, apresentou uma contribuição significativa ao reconhecer o papel preponderante exercido pelas redes sociais, especialmente o Twitter, como plataformas com arquitetura e dinâmicas intrínsecas que tendem a privilegiar a disseminação sistemática de desinformação.

**Palavras-chaves:** antivacina; hesitação vacinal; Linguística de Corpus; Análise Multidimensional Lexical; Análise do Discurso.

WHITEMAN, Mirella. **Discourses emerging from anti-vaccine movements in Brazil during the covid-19 pandemic**. 2024. 133 f. Dissertation (Master's in Applied Linguistics and Language Studies) - Applied Linguistics and Language Studies Program, Pontifical Catholic University of São Paulo, São Paulo, 2024.

### **Abstract**

The last decade witnessed a significant rise in anti-vaccine movements, widely driven by the influence of digital media. The resurgence of previously controlled diseases, such as measles and polio, has emerged as a global concern, profoundly impacting healthcare professionals in Brazil. Since 2015, a decline in coverage of the National Immunization Program (PNI) has been observed, a trend exacerbated by the Covid-19 pandemic. During the pandemic period, vaccines and their effectiveness became central themes in public discourse, intensified by the impact of social networks and their sophisticated information networks, widely used to disseminate vaccine hesitancy. In this context, the objective of this research is to investigate the discursive universe of the anti-vaccine movement in Brazil on the Twitter (now X) platform, based on a theoretical-methodological framework that comprises Corpus Linguistics (Berber Sardinha, 2004), Dialogic Discourse Analysis (Bakhtin, 2016; Volóchinov, 2017) and Critical Discourse Analysis (Fairclough, 1989; Van Dijk, 1985; Wodak, 1989). Previous investigations in big data did not focus on discourses, but rather on elements such as sentiments and topics (Jang *et al.*, 202; Wicke; Bolognesi, 2020). Meanwhile, Discourse Analysis research related to the pandemic dealt with small volumes of data (Recuero; Stumpf, 2021; Jones, 2021). The use of Lexical Multidimensional Analysis (Berber Sardinha, 2016; 2017; 2020; Berber Sardinha; Fitzsimmons-Doolan, in press), the methodological approach chosen for this study, enabled the examination of discourses in large-scale corpora. The approach, based on statistical calculations, is capable of detecting groups of lexical items that co-occur in texts, thus revealing recurrent discourses in language. For the analysis, 16,841 tweets were subjected to tagging and lemmatization, followed by factor analysis. The Lexical Multidimensional Analysis allowed the identification of 3 main dimensions, with positive and negative poles: Warning about cardiac risks, sudden illnesses and deaths after booster doses, especially among youth and children, considering the vaccine as experimental versus Questioning and resistance to the imposition of compulsory measures and decrees to combat the advance of Covid-19; Allegations about the origin of the virus in a Chinese laboratory and interests of politicians and governments linked to large pharmaceutical

groups versus Evidence of pseudoscientific studies published on the adverse effects of vaccines; and Defense of early treatment in the face of the lethality of the virus and the lack of studies on the long-term efficacy and safety of the emergency vaccine versus Reports about friends and third parties who suffered severe side effects – headaches, paralysis, kidney problems, leg and foot problems – and even fatal after vaccination. The present study achieved a satisfactory degree of theoretical-methodological triangulation by combining Corpus Linguistics with two complementary discourse theories. Moreover, it provided a new perspective on anti-vaccine discourses in Brazil by identifying and analyzing them in an integrated manner, understanding how they interact and reinforce each other mutually. Finally, it presented a significant contribution by recognizing the preponderant role played by social networks, with special emphasis on Twitter, as platforms whose architecture and intrinsic dynamics tend to privilege the systematic dissemination of disinformation.

**Keywords:** anti-vaccine; vaccine hesitancy; Corpus Linguistics; Lexical Multidimensional Analysis; Discourse Analysis.

### **Lista de figuras**

<b>Figura 1</b> – Panfleto antivacina norte-americano de 1894 .....	30
<b>Figura 2</b> – Processo cíclico para a compilação de um corpus.....	49
<b>Figura 3</b> – Visão geral do processamento do corpus.....	78
<b>Figura 4</b> – Filtragem de variáveis por palavras-chave .....	79
<b>Figura 5</b> – Procedimentos da Análise Multidimensional Lexical .....	80
<b>Figura 6</b> – Panfleto antivacina de 1885.....	108

## Lista de gráficos

<b>Gráfico 1</b> – <i>Scree plot</i> da extração fatorial.....	81
--	----

## **Lista de quadros**

<b>Quadro 1</b> – Tipologia dos corpora.....	44
<b>Quadro 2</b> – Aspectos das AMDs Funcional e Lexical.....	59
<b>Quadro 3</b> – Composição do CBDIAT.....	80
<b>Quadro 4</b> – Rótulos das dimensões discursivas.....	84
<b>Quadro 5</b> – Variáveis da Dimensão 1.....	85
<b>Quadro 6</b> – Variáveis da Dimensão 2.....	89
<b>Quadro 7</b> – Variáveis da Dimensão 3.....	93
<b>Quadro 8</b> – Discursos das dimensões.....	99



## Sumário

Introdução .....	16
História das pandemias e da infodemia .....	22
A pandemia no Brasil .....	24
História das vacinas .....	28
História dos movimentos antivacina .....	30
História da vacinação no Brasil .....	32
1. Fundamentação teórica .....	36
1.1. Linguística Aplicada .....	36
1.2. Linguística de Corpus .....	38
1.2.1. Linguística de Corpus: histórico .....	40
1.3. Corpus .....	42
1.3.1. Tipologia de corpus .....	44
1.3.2. Design de corpus .....	45
1.3.3. Compilação de corpus .....	49
1.4. Registro .....	50
1.4.1. Tweet como um registro .....	52
1.5. Análise Multidimensional .....	56
1.5.1. Análise Multidimensional Funcional .....	56
1.5.2. Análise Multidimensional Lexical .....	58
1.6. Análise do Discurso .....	63
1.6.1. Análise do Discurso Assistida por Corpus .....	65
1.6.2. Análise Crítica do Discurso .....	66
1.6.3. Análise Dialógica do Discurso .....	68
1.6.4. Enunciado e cadeia discursiva .....	69
1.7. Taxonomia dos argumentos antivacina .....	72
2. Metodologia .....	77
2.1. Coleta e pré-processamento do corpus .....	77
2.2. Visão geral do processamento do corpus .....	77
2.2.1. Filtragem de variáveis por palavras-chave .....	78
2.2.2. Filtragem de variáveis por correlação .....	79
2.3. Composição do corpus .....	80
2.4. Análise Multidimensional Lexical .....	80
3. Apresentação e discussão dos resultados .....	83
3.1. Apresentação dos resultados .....	83
3.1.1. Dimensão 1 – alerta sobre riscos da dose de reforço em jovens versus resistência a medidas compulsórias de combate à Covid-19 .....	85
3.1.2. Dimensão 2 – teorias político-conspiratórias sobre a vacina versus evidências pseudocientíficas contra a vacina .....	89
3.1.3. Dimensão 3 – defesa do tratamento precoce em detrimento da vacina versus relatos pessoais sobre efeitos colaterais da vacina .....	93
3.2. Discussão dos resultados .....	97
Considerações finais .....	103
Referências .....	115

## Introdução

No dia 31 de dezembro de 2019, enquanto o mundo se preparava para celebrar a virada do Ano Novo, autoridades chinesas emitiam um comunicado oficial à Organização Mundial da Saúde (OMS) alertando sobre um surto de pneumonia de origem desconhecida na cidade de Wuhan, situada na província de Hubei, na China. Pouco depois, casos adicionais foram detectados além das fronteiras chinesas, atingindo países como Tailândia e Estados Unidos, o que culminou na declaração de Emergência de Saúde Pública de Importância Internacional (ESPII) em 30 de janeiro de 2020 (Haraki, 2021).

A data marcou o início de uma sequência de eventos que desencadeariam uma crise global, posteriormente conhecida como a pandemia de Covid-19. Esse desafio para a saúde pública mundial suscitou ainda um fenômeno que permeou os anos que se seguiram: a infodemia. De acordo com a OMS (2020), a infodemia pode ser definida como um grande aumento no volume de informações associadas a um determinado assunto, as quais podem se multiplicar exponencialmente em pouco tempo devido a um evento específico. Isso pode dificultar a identificação de fontes e orientações confiáveis, além de incitar a desconfiança em relação a autoridades, prejudicando a capacidade de resposta no âmbito da saúde pública.

A infodemia é impulsionada pela rápida disseminação de informações nas plataformas digitais, que são diariamente abastecidas por uma abundância de narrativas de confiabilidade incerta sobre questões relevantes, passadas de usuário para usuário. Esses rumores e boatos não verificados podem gerar medo excessivo, hostilidade, desconfiança e perturbação social (Taylor, 2020). Nas formas mais extremas dessas especulações, as chamadas teorias da conspiração, eventos importantes são interpretados como resultantes de tramas secretas envolvendo figuras poderosas e grandes corporações frequentemente ocultas, o que pode provocar consequências reais muito graves (Douglas *et al.*, 2019).

Durante os surtos de ebola e cólera, por exemplo, emergiram teorias conspiratórias sugerindo que as doenças eram causadas por equipes médicas, resultando, em alguns casos, em agressões e até mesmo assassinatos de profissionais de saúde (Cohn, 2017). A propensão a acreditar nessas teorias está

associada a diversas variáveis, como a falta de alfabetização midiática<sup>1</sup>, a desconfiança de figuras de poder e a rejeição da ciência convencional em favor da pseudociência (Douglas *et al.*, 2019).

O avanço da Covid-19 desencadeou uma mobilização global sem precedentes para acelerar o desenvolvimento, a produção e a distribuição de imunizantes, com o objetivo de conter a pandemia, que já acumulava milhões de mortes em todo o mundo. O desenvolvimento acelerado de vacinas contra o coronavírus foi resultado de uma combinação de avanços científicos, investimentos financeiros substanciais e colaboração global. No entanto, essa velocidade gerou desconfiança, devido à percepção de que os processos normais de avaliação de segurança e eficácia poderiam ter sido comprometidos.

Embora o SARS-CoV-2 fosse um vírus recém-descoberto, os esforços de pesquisa e desenvolvimento de vacinas puderam se beneficiar de conhecimentos preexistentes de tecnologias em estudo. Apesar da novidade representada pelas vacinas de RNA mensageiro, essa plataforma já vinha sendo investigada no campo da oncologia e no desenvolvimento de outras vacinas. Aliado ao conhecimento científico acumulado, o maciço direcionamento de recursos financeiros, humanos e de infraestrutura para essa finalidade possibilitou que múltiplas vacinas contra a Covid-19 fossem rapidamente desenvolvidas e disponibilizadas em um período de tempo notavelmente curto, dada a magnitude e complexidade do desafio enfrentado.

Além disso, os ensaios clínicos continuaram a seguir protocolos rigorosos estabelecidos por agências reguladoras internacionais, como a Food and Drug Administration (FDA) nos Estados Unidos, a European Medicines Agency (EMA) na União Europeia e a Agência Nacional de Vigilância Sanitária (Anvisa) no Brasil. Os dados dos ensaios também foram revisados por especialistas independentes antes da aprovação das vacinas para uso emergencial (OPAS, 2020).

Assim, as vacinas passaram a ocupar um papel central no debate público, contribuindo para a polarização das discussões nas redes sociais e a formação de “câmaras de eco”, conforme descrito por Boulianne (2020). Nesses ambientes,

---

<sup>1</sup> Segundo a UNESCO (2013), a alfabetização midiática e informacional abrange conhecimentos cruciais sobre as funções da mídia, bibliotecas, arquivos e outros provedores de informação em sociedades democráticas. Isso inclui entender como esses meios podem efetivamente desempenhar seus papéis e como avaliar sua eficácia por meio da análise de conteúdo e serviços oferecidos. Essa compreensão capacita os usuários a se envolverem de maneira significativa com a mídia e os canais de informação. Além disso, as habilidades adquiridas nesse processo promovem o desenvolvimento do raciocínio crítico dos cidadãos.

indivíduos tendem a ser expostos apenas a informações e opiniões que corroboram suas próprias convicções, o que intensifica as divisões. Esse fenômeno ressalta a complexidade dos debates sobre vacinação, que vão além de uma simples avaliação de riscos, custos e benefícios.

Os processos individuais de tomada de decisão são influenciados por uma variedade de fatores, incluindo emoções, cultura, religião e contexto socioeconômico, como discutido por Baumgaertner, Carlisle e Justwan (2018). A essa lista é possível adicionar a crescente influência da internet e das redes sociais na vida dos usuários. Hoje, a internet rivaliza com a medicina tradicional como a principal fonte de conselhos de saúde, configurando-se como um novo paradigma pós-moderno em que o poder passou dos médicos para os pacientes, a legitimidade da ciência é questionada e a especialização é redefinida (Kata, 2012). No Brasil, segundo pesquisa do Instituto de Ciência, Tecnologia e Qualidade – ICTQ (2018), mais de 40% dos brasileiros faz autodiagnóstico médico pela internet. Esse cenário favorece a visibilidade de práticas alternativas que frequentemente entram em conflito com protocolos médicos fundamentados cientificamente.

Nesse contexto, o uso excessivo de plataformas digitais por campanhas antivacina para difundir discursos contrários à imunização, independentemente de sua veracidade, intensificou em todo o mundo um quadro de hesitação vacinal, que pode ser entendido como o atraso na aceitação ou recusa de vacinas, apesar da disponibilidade de serviços de imunização (MacDonald, 2015). Países como a Polônia, onde a confiança nas vacinas diminuiu significativamente, têm sido associados a movimentos antivacina altamente organizados e a uma mobilização *on-line* intensa (Figueiredo *et al.*, 2020). A hesitação vacinal tem sido uma preocupação contínua nos últimos anos e figura entre as dez principais ameaças à saúde global, conforme apontado pela OMS (2019).

A pandemia de Covid-19, a princípio, deveria ter gerado um amplo respaldo às vacinas, considerando sua eficácia comprovada ao longo dos anos e seu papel decisivo na elevação da expectativa de vida e na erradicação de diversas doenças. De acordo com a OMS, a vacinação salva anualmente mais de 5 milhões de pessoas, proporcionando ainda uma economia superior a US\$ 1 bilhão em tratamentos médicos (INSTITUTO BUTANTAN, 2022). No entanto, o cenário pandêmico não seguiu o curso esperado. A resistência às campanhas de vacinação ao redor do mundo e também no Brasil, revelou-se uma profunda contradição, que pode ser atribuída a uma série de

fatores, incluindo desconfiança nas instituições, crenças pessoais, desinformação e influência dos chamados ecossistemas infodêmicos. Esses ecossistemas, caracterizados pelo excesso de informações disseminadas através de várias plataformas de mídia e redes sociais, desempenharam um papel fundamental na moldagem das opiniões e atitudes em relação à vacinação. Isso ocorreu, em grande parte, devido à propagação de desinformação sobre as vacinas e seus efeitos, alimentando teorias da conspiração, incitando temores infundados sobre possíveis efeitos colaterais e semeando dúvidas quanto à eficácia dos imunizantes. Surge, assim, a necessidade de maior compreensão sobre os discursos em torno das vacinas que circulam nas plataformas digitais, considerando seus intrincados sistemas de engajamento. A análise desses discursos torna possível o exame de como informações são configuradas e disseminadas, influenciando percepções e decisões em relação à vacinação.

Desde seu surgimento, no ano de 1996, as redes sociais conquistaram mais da metade dos quase 8 bilhões de habitantes do nosso planeta. A base de usuários das plataformas digitais na última década disparou de 970 milhões em 2010 para uma notável marca que ultrapassou os 4,95 bilhões de usuários em outubro de 2023 (Dean, 2023). O aumento exponencial do uso e da influência das mídias sociais aponta para a necessidade de compreender como as teorias linguísticas interpretam esses ambientes dinâmicos em constante evolução.

Diante desse cenário, a Linguística de Corpus emerge como uma ferramenta indispensável na compreensão e análise dos ecossistemas infodêmicos. Ao contrário das abordagens linguísticas tradicionais, que muitas vezes se concentram na análise de textos individuais ou em pequenos conjuntos de dados, é uma área que se ocupa de grandes volumes de texto. Investigações em Análise do Discurso acerca da pandemia têm se debruçado sobre conjuntos de dados reduzidos (Recuero; Stumpf, 2021; Jones, 2021), uma tendência semelhante à observada em estudos multimodais (Caple; Huan; Bednarek, 2020). Por outro lado, pesquisas que exploraram *big data* relacionadas à pandemia focalizaram aspectos como tópicos, sentimentos e *frames* (Jang *et al.*, 2021; Wicke; Bolognesi, 2020), não se concentrando em discursos.

O presente estudo, parte integrante do Programa Estratégico Emergencial de Prevenção e Combate a Surtos, Endemias, Epidemias e Pandemias, instituído pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), assume uma importância significativa ao visar preencher essa lacuna na literatura. Dentro

desse projeto, no campo da Linguística, o Grupo de Estudos de Linguística de Corpus (GELC) da Pontifícia Universidade Católica de São Paulo (PUC-SP), sob a liderança do Professor Doutor Tony Berber Sardinha, tem desenvolvido abordagens inovadoras para Análise do Discurso em Linguística de Corpus (Berber Sardinha, 2016, 2017, 2020, 2021), *expertise* que se revela imprescindível para a realização desta pesquisa. A relevância científica deste estudo reside no desenvolvimento de novos métodos para identificar discursos subjacentes à infodemia em um conjunto volumoso de textos. Para tanto, será empregado um arcabouço interdisciplinar, integrando contribuições da Linguística de Corpus, Análise Multidimensional Lexical e Análise do Discurso Assistida por Corpus.

A Linguística de Corpus guiará os princípios de criação de corpora representativos da infodemia, bem como as bases de criação de sentido na língua em uso por meio da coocorrência sistemática de itens lexicogramaticais. A Análise Multidimensional emerge como uma ferramenta eficaz na identificação desses ecossistemas como dimensões de variação. Berber Sardinha (2020, 2021) e Fitzsimmons-Doolan (2014) propuseram a integração de dimensões de variação como uma abordagem na Análise do Discurso, evidenciando a viabilidade do uso desse conceito para descrever a variação presente nos discursos, como será visto adiante detalhadamente. A Análise Multidimensional Lexical, vertente discursiva da Análise Multidimensional, será acionada de maneira específica no contexto deste projeto, visando identificar os discursos em torno da vacina de Covid-19. Por fim, será usada a Análise do Discurso Assistida por Corpus, uma área multidisciplinar voltada à identificação de discursos por meio da análise de corpora (Friginal; Hardy, 2020).

A literatura prévia tem focado a infodemia principalmente do ponto de vista do excesso e rápida difusão da comunicação verbal (textual). No entanto, propomos que a infodemia é mais abrangente do que simplesmente a demasia de informação. Na verdade, trata-se de uma profusão de discursos, ou seja, “maneiras de enxergar o mundo, de construir objetos e conceitos de certos modos, de representar a realidade” (Baker; McEnery, 2015, p. 5), que por sua vez “enquadram certos problemas, [distinguindo] certos aspectos da situação em vez de outros” (Hajer, 1993, p. 45). O presente estudo entende a infodemia como um embate de discursos conflitantes, dos quais se valeram governos, empresas e demais atores sociais, para pensar e agir sobre a pandemia.

Configurando-se como um ecossistema discursivo, a infodemia dissemina

visões de mundo, ideologias e representações que indexam valores, intenções, conceitualizações e modos de agir (Foucault; Faubion, 2000; Van Dijk, 2008). Considerando o contexto apresentado, esta pesquisa se desenvolve para responder as seguintes perguntas:

1. Quais são as dimensões lexicais dos discursos em torno do movimento antivacina?
2. Quais os principais discursos associados às dimensões lexicais?

O principal objetivo deste estudo é investigar os discursos emergentes em torno do movimento antivacina no Brasil, no Twitter. Para garantir uma amostra ampla, o período selecionado compreende o início da pandemia e sua etapa final, entre 2020 e 2022, levando em consideração o grande aumento de discursos sobre as vacinas nesse intervalo de tempo.

O Twitter (atualmente conhecido por X<sup>2</sup>) foi escolhido como a mais relevante rede social a ser analisada ao destacar-se das outras plataformas por ter sido criado explicitamente para viabilizar a militância. Lançado em 2006 como uma base ativista, deu continuidade a um programa criado nos Estados Unidos chamado *TXTMob*. Através dele, em 2004, militantes usaram mensagens de texto para compartilhar informações e coordenar protestos descentralizados durante a convenção nacional do Partido Republicano, em Nova York (Barboza, 2021), surgindo, assim, uma nova forma de participação em mobilizações em massa. A rede oferece ainda um ambiente social que recompensa a comunicação curta, agressiva e descontextualizada (Haigh; Haigh, 2020), a arena ideal para a circulação de desinformação.

Outro fator importante para a escolha do Twitter reside no fato de que o então presidente do Brasil, Jair Bolsonaro, o adotou como sua plataforma oficial durante a pandemia, assim como seus apoiadores. Bolsonaro assimilou estratégias semelhantes às do ex-presidente dos EUA, Donald Trump, no enfrentamento da pandemia e transformou a vacinação em uma disputa político-ideológica, utilizando amplamente o Twitter para confrontar perspectivas favoráveis à imunização e promover um ambiente marcado pela desinformação e polarização.

Com o intuito de aprofundar nossa compreensão sobre a construção dos discursos em torno das vacinas, mais especificamente a de Covid-19, e sua

---

<sup>2</sup> Elon Musk, o bilionário da tecnologia que comprou o Twitter em 2022, renomeou a plataforma social para X e substituiu o logotipo do pássaro por uma versão estilizada da 24ª letra do alfabeto latino. Para mais informações, ver Mac e Hsu (2023).

pertinência no cenário sociocultural contemporâneo, as subseções a seguir elucidam a formação dessa trama contextual. Buscamos enriquecer a discussão sobre a circulação desses discursos nas redes sociais, um domínio ainda pouco explorado na Linguística e que carece de respaldo, utilizando um corpus de pesquisa meticulosamente compilado para esse propósito.

### **História das pandemias e da infodemia**

Pandemias são surtos de doenças infecciosas que se espalham globalmente, afetando milhões de pessoas. O termo se refere à distribuição geográfica de uma doença, e não à sua gravidade (OPAS, [s. d.]). Ao longo da história, inúmeras pandemias foram registradas (Snowden, 2019), como a Peste Bubônica, também conhecida como Peste Negra (1346-1356), a Gripe Russa (1889-1890), a Gripe Espanhola (1918-1920), a persistente crise de HIV/AIDS (desde 1981), a Síndrome Respiratória Aguda Grave – SARS (2002-2004), a Gripe Suína (2009-2010), o surto do vírus Zika (2015-2016) e a pandemia de Covid-19 (2020-2023).

A gestão de pandemias por parte dos governos e autoridades públicas busca contemplar duas estratégias fundamentais: sensibilizar as pessoas para sua gravidade e incentivar a adoção de práticas recomendadas de mitigação, como o distanciamento social, sem, no entanto, desencadear pânico em massa. A mitigação de pandemias envolve a comunicação de riscos, pela qual autoridades de saúde e governamentais informam o público sobre o nível de perigo apresentado e as estratégias que serão implementadas. Essa comunicação inclui a orientação à população sobre diversas práticas de higiene, estratégias de distanciamento social, recomendações ou mandatos de permanência em casa, uso de máscaras, vacinação, além do fechamento estratégico de locais de aglomeração, como bares, restaurantes e outros lugares de entretenimento, escolas e igrejas (Taylor, 2020).

A adesão popular a esses métodos de contenção é crucial para sua eficácia. Relatos divulgados na mídia e nas redes sociais desempenham um papel significativo na formação da percepção pública e nas reações a ameaças à saúde (Fung; Namkoong; Brossard, 2011; Kilgo; Yoo; Johnson, 2019). A Covid-19 emergiu na era das mídias sociais e da interconectividade digital global, o que, em muitos países, desencadeou ondas de ansiedade antecipatória que precederam a chegada efetiva do coronavírus (Taylor *et al.*, 2020). Essa dinâmica não é inédita, tendo ocorrido de maneira semelhante em 1889 durante a Gripe Russa, a primeira pandemia a se



desenrolar em uma época de jornais de circulação barata em massa e comunicação global via telégrafo. A Gripe Russa foi, assim, a primeira pandemia a receber ampla cobertura na imprensa diária. Esse fenômeno levou o público a tomar conhecimento da pandemia antes mesmo de atingir suas comunidades, gerando ansiedade antes da chegada efetiva do vírus (Kempinska-Miroslawska; Woźniak-Kosek, 2013).

A mídia é frequentemente criticada por exagerar a periculosidade de surtos anteriores, como o da SARS e a pandemia da Gripe Suína (Taylor, 2019). Análises dos relatos midiáticos durante a pandemia da Gripe Suína sugerem que veículos de comunicação contribuíram para aumentar as percepções de risco principalmente através da intensidade da cobertura, com foco desequilibrado nos perigos e menor atenção às medidas que as pessoas poderiam adotar para se proteger (Klemm; Das; Hartmann, 2016).

O sensacionalismo midiático pode ter efeitos polarizadores, e algumas pessoas podem ficar excessivamente ansiosas em relação à ameaça da doença, enquanto outras podem menosprezar a gravidade da situação, considerando-a exagerada (Taha; Matheson; Anisman, 2013). Aqueles que acreditam que os relatos midiáticos são exagerados têm menor probabilidade de adotar os comportamentos de saúde recomendados pelas autoridades sanitárias (Rubin *et al.*, 2009). Esses tipos de notícias, quando disseminados nas redes sociais, podem intensificar ainda mais o sensacionalismo, pois os usuários dessas plataformas tendem a fazer compartilhamentos seletivos, de modo que histórias sensacionalistas são mais propensas a serem compartilhadas do que informações precisas, porém menos impactantes (Taylor, 2020).

A expressão híbrida *infodemia*, combinando as palavras *informação* e *epidemia*, reflete o impacto desproporcional que as novas tecnologias de informação têm exercido na comunicação contemporânea sobre saúde. O cientista político David Rothkopf cunhou o termo pela primeira vez durante a pandemia de SARS, em 2003: “Alguns fatos, misturados com medo, especulação e boato, amplificados e rapidamente transmitidos mundialmente por meio de tecnologias de informação modernas” (Rothkop, 2003)<sup>3</sup>.

Segundo Rothkopf (2003), a infodemia tem afetado questões econômicas, políticas e até mesmo de segurança de maneiras completamente incompatíveis com

---

<sup>3</sup> Original: “A few facts, mixed with fear, speculation and rumor, amplified and relayed swiftly worldwide by modern information technologies”.

a realidade. Embora o termo seja relativamente recente, a ligação entre epidemias e desinformação remonta a muito tempo. Seja a Peste Bubônica no século XIV ou a epidemia de HIV/AIDS no final do século XX, surtos de doenças frequentemente desencadearam uma avalanche de rumores confusos, teorias conspiratórias e curas miraculosas. Nas últimas duas décadas, o avanço das tecnologias de informação tem amplificado consideravelmente esse processo, tornando-o mais desafiador de ser controlado.

### **A pandemia no Brasil**

O Brasil registrou seu primeiro caso confirmado de Covid-19 em São Paulo, em 26 de fevereiro de 2020. O paciente era um homem de 61 anos residente na capital paulista que havia retornado de uma viagem à Itália, sendo esse considerado o marco inicial da contaminação pelo novo coronavírus no país. Pouco tempo depois, em 12 de março, o Ministério da Saúde confirmou o primeiro óbito decorrente da doença no Brasil. A vítima, Rosana Aparecida Urbano, de 57 anos, faleceu no Hospital Municipal Doutor Carmino Cariccio, localizado na Zona Leste da cidade de São Paulo, um dia após ser internada. A propagação do vírus acelerou rapidamente em território brasileiro, e já em abril de 2020 o país contabilizava cerca de 50 mil casos da doença e aproximadamente 3 mil mortes. No mês subsequente, os registros de óbitos por Covid-19 ultrapassavam 700 por dia, atingindo a marca de mil mortes diárias a partir da segunda quinzena de maio (Primeiro..., 2020).

Durante o avanço da pandemia de Covid-19 no Brasil e no mundo, o então presidente, Jair Bolsonaro, fez declarações controversas em diversos pronunciamentos, classificando a doença como uma “gripezinha” e minimizando sua gravidade. Também afirmou que, devido ao seu “histórico de atleta”, não precisaria se preocupar, caso contraísse o coronavírus. Essas declarações foram proferidas com o objetivo de mitigar o que Bolsonaro percebia como um momento desnecessário de “pânico e histeria” que se alastrava pelo país, a despeito da escalada da crise sanitária causada pela pandemia. Paralelamente, o então ministro da saúde, Luiz Henrique Mandetta, anunciou por meio das redes sociais sua demissão, a qual foi atribuída a divergências nas estratégias de enfrentamento à doença entre ele e o presidente. Sucedendo Mandetta, o oncologista Nelson Teich assumiu o Ministério da Saúde, mas deixou o cargo menos de um mês depois de sua nomeação. Em seu lugar, assumiu o então secretário-executivo da pasta, Eduardo Pazuello, um general sem experiência

prévia na área da Saúde (Linha..., 2021).

Enquanto a rejeição oficial à cloroquina na maior parte do mundo começou ainda no primeiro semestre da pandemia, o Brasil seguiu o caminho oposto. Bolsonaro promoveu medicamentos e outras substâncias que supostamente poderiam tratar ou prevenir a doença viral, desde a hidroxicloroquina até a ivermectina (Casarões; Magalhães, 2021). No período de março de 2020 a janeiro de 2021, ocorreram ao menos quatro medidas federais que direta ou indiretamente promoveram a prescrição desses medicamentos, de acordo com o levantamento conduzido por pesquisadores do Centro de Estudos e Pesquisas de Direito Sanitário da USP (Cepedisa) em colaboração com a Conectas Direitos Humanos<sup>4</sup>. As medidas incluíram um protocolo do Ministério da Saúde datado de 20 de maio de 2020 que recomendava o uso da cloroquina em todos os casos de Covid-19, bem como o aplicativo TrateCov, também desenvolvido pelo referido ministério, que sugeria a médicos a prescrição de medicamentos como hidroxicloroquina, cloroquina, ivermectina e azitromicina. No entanto, o aplicativo foi retirado do ar pouco tempo depois.

Em diversas ocasiões, Bolsonaro foi fotografado e filmado segurando embalagens de cloroquina, o que foi vastamente interpretado como uma campanha de promoção do medicamento. Com o tempo, a divulgação evoluiu para uma defesa mais ampla – tanto por parte do presidente quanto pelo Ministério da Saúde sob a gestão de Pazuello – do que passou a ser conhecido como “tratamento precoce”. O tratamento incluía um coquetel de remédios composto de hidroxicloroquina, azitromicina, ivermectina, nitazoxanida, suplementos de zinco e as vitaminas C e D (Idoeta, 2021).

Observando o tratamento precoce a partir de uma perspectiva econômica, em 2020 o Governo Federal gastou aproximadamente R\$ 90 milhões na aquisição de medicamentos como cloroquina, azitromicina, Tamiflu e hidroxicloroquina, que não possuem comprovação científica para o tratamento da Covid-19. A maior parte dessas compras foi realizada com dispensa de licitação, levantando questões sobre a transparência e a gestão dos recursos públicos. Destaca-se que o exército pagou preços acima do mercado por insumos para a produção de cloroquina, o que resultou em uma investigação pelo Tribunal de Contas da União. Esse cenário suscitou críticas quanto à priorização de tratamentos não comprovados em detrimento da aquisição de

---

<sup>4</sup> Organização não governamental e sem fins lucrativos que propõe soluções e denuncia violações relacionadas a questões de direitos humanos.

vacinas e da eficácia das medidas de combate à pandemia (Shalders, 2021). Como pandemias recentes ilustraram, medicamentos foram cada vez mais incorporados ao repertório performático dos políticos, uma tendência populista de oferecer soluções simples para problemas complexos (Lasco; Yu, 2021).

Paralelamente, durante 2020, o governo liderado por Bolsonaro optou por recusar a oferta de vacinas da Pfizer, que estavam disponíveis pela metade do preço pago por países como Estados Unidos, Reino Unido e União Europeia. Em agosto do mesmo ano, Pazuello considerou as doses caras, apesar da possibilidade de adquirir até 70 milhões delas por US\$ 10 cada a partir de dezembro. A recusa em adquirir essas vacinas mais cedo levantou debates sobre a possibilidade de uma vacinação antecipada ter potencialmente evitado mortes e prejuízos econômicos bilionários causados pelo fechamento da economia. Enquanto isso, os Estados Unidos e o Reino Unido pagaram cerca de US\$ 20 por dose da vacina Pfizer, o dobro do valor recusado pelo Brasil durante vários meses em 2020. Na União Europeia, as doses foram adquiridas por US\$ 18,60 cada. Devido aos atrasos nos contratos, as primeiras doses da Pfizer só chegaram ao Brasil em abril de 2021, oito meses após a oferta inicial (Canzian; Cancian, 2021).

Apesar de recusar inicialmente a compra das vacinas, no começo de 2021 observou-se uma alteração no discurso de Bolsonaro, indicando uma tentativa de reconfigurar sua abordagem sobre o tema. Em meio ao auge da crise pandêmica, com o Brasil ultrapassando a marca de 270 mil óbitos e diante da entrada do ex-presidente Lula como um possível adversário na corrida presidencial de 2022, o então presidente adotou uma postura mais receptiva à vacinação, marcando uma mudança significativa em relação à sua posição anteriormente contrária à imunização, apesar de manter a defesa de medicamentos sem comprovação científica e as críticas às medidas restritivas adotadas por prefeitos e governadores.

Essa mudança de postura foi acompanhada por declarações que enfatizavam a importância crucial da vacina como prioridade na luta contra o coronavírus. Algumas das declarações anteriores mais proeminentes do então presidente sobre o tema demonstram essa mudança radical de atitude frente à vacinação. Bolsonaro afirmou: “Não será comprada. O povo brasileiro não será cobaia de ninguém”, ao desautorizar o ministro da saúde e negar a compra da Coronavac (21 de outubro de 2020). Além disso, em entrevista à Rádio Jovem Pan, ele declarou: “Da China nós não compraremos, é decisão minha. Eu não acredito que ela [a vacina] transmita

segurança suficiente para a população pela sua origem” (21 de outubro de 2020); em outro momento, Bolsonaro expressou: “Ninguém pode obrigar ninguém a tomar vacina. Eu não vou tomar. Eu já tive o vírus. Já tenho anticorpos. Para que tomar vacina de novo?” (17 de dezembro de 2020); e durante um discurso em Porto Seguro, na Bahia, ele disse: “Se tomar e virar um jacaré é problema seu. Se virar um super-homem, se nascer barba em mulher ou homem falar fino, ela [Pfizer] não tem nada com isso” (17 de dezembro de 2020), referindo-se à Pfizer, uma das fabricantes mundiais de vacina contra a doença (Relembre..., 2021).

O avanço da pandemia pelo território nacional levou à superlotação de hospitais e ao colapso do sistema de saúde em muitas regiões do país, o que resultou na falta de leitos de UTI para pacientes graves, bem como na escassez de materiais de segurança (EPIs) para profissionais de saúde e também de elementos básicos para o tratamento dos doentes – como foi o caso da crise de oxigênio em Manaus, no estado do Amazonas, marcada pela falta de cilindros de oxigênio nos hospitais em janeiro de 2021. Mais tarde, naquele mesmo ano, o Brasil viveu a segunda onda da Covid-19, a mais longa e mais letal até então. Em abril de 2021, o país contabilizava mais de 13 milhões de pessoas infectadas, batendo o recorde de 4.211 óbitos em um único dia.

Foi também no ano de 2021 que a vacinação teve início no país. A primeira dose de uma vacina da Covid-19 foi aplicada em 17 de janeiro, e, pouco mais de seis meses depois, cerca de 50% da população já havia tomado ao menos uma dose. Em abril de 2022, o então ministro da saúde, Marcelo Queiroga, assinou uma portaria que declarava o fim da Emergência em Saúde Pública de Importância Nacional (ESPIN) causada pela pandemia da Covid-19 no Brasil. Apesar das dificuldades criadas pelo próprio Governo Federal em relação à aquisição e à produção das vacinas contra a Covid-19, bem como a disseminação de desconfiança sobre sua eficácia, a alta cobertura vacinal alcançada no Brasil emergiu como um dos principais fatores contribuintes para a redução da transmissão do coronavírus. Até dezembro de 2022, mais de 80% da população brasileira estava vacinada com duas doses da vacina ou com a vacina de dose única. No mesmo período, mais de 107 milhões de pessoas já haviam tomado a dose de reforço, o que equivale a quase metade da população do país (Guitarrara, [s. d.]).

Após vencer a disputa presidencial, em janeiro de 2023, o novo presidente eleito, Luiz Inácio Lula da Silva, e a nova ministra da saúde, Nísia Trindade, anunciaram o lançamento do Movimento Nacional pela Vacinação. A iniciativa tem

como objetivo implementar medidas para aumentar a cobertura vacinal de todos os brasileiros com as vacinas disponíveis no Sistema Único de Saúde (SUS). De acordo com o ministério, até maio de 2023, 16 milhões de pessoas já haviam recebido a vacina bivalente contra a Covid-19 (Brasil..., 2023b). A partir de janeiro de 2024, a vacina contra Covid-19 para crianças de 6 meses a menores de 5 anos passou a fazer parte do calendário nacional de vacinação. Também podem se vacinar pessoas com mais de 5 anos, mesmo não fazendo parte dos grupos prioritários, que não foram vacinadas antes ou receberam apenas uma dose (Pessoa, 2024). Dados atualizados pelo Ministério da Saúde mostram que até fevereiro de 2024 o Brasil contabilizava 38.407.327 casos acumulados de Covid-19 e 709.765 mortes confirmadas (Brasil, 2024).

### **História das vacinas**

O avanço nas áreas de Ciência, Microbiologia, Farmacologia e Imunologia tem, ao longo dos anos, integrado-se aos estudos de epidemiologia e sociologia, revelando o significativo impacto que as vacinas têm exercido na sociedade contemporânea (Feijó; Sáfiadi, 2006).

As vacinas se constituem como produtos biológicos que estimulam a resposta imunológica do corpo contra microrganismos, tais como vírus e bactérias, que podem desencadear doenças. Elas podem ser desenvolvidas a partir de microrganismos enfraquecidos, inativados ou de alguns de seus componentes. Ao ser vacinada, a pessoa tem seu sistema imunológico ativado, detectando os elementos presentes na vacina e gerando uma resposta de defesa, conhecida como produção de anticorpos. Esses anticorpos permanecem no organismo, conferindo imunidade e prevenindo a ocorrência da doença no futuro, proporcionando assim uma proteção duradoura (Unicef, [s. d.]).

A vacinação em larga escala começou no início do século XIX, após a apresentação, por Edward Jenner, de um artigo à Royal Society of London em 1796. No documento, Jenner detalhou seu êxito na prevenção da varíola em 13 indivíduos por meio da inoculação com material infeccioso vivo obtido de pessoas infectadas com varíola bovina. Esse processo induzia uma forma leve da doença viral, conferindo imunidade à varíola. Jenner denominou o material da varíola bovina de *vaccine* (do latim *vacca*, significando *vaca*) e o procedimento de *vaccination*, sendo o primeiro a conferir *status* científico a esse procedimento, o que foi fundamental para sua

popularização (Wolfe; Sharp, 2002).

Vacinas rotineiramente recomendadas começaram a ser desenvolvidas no início do século XX, incluindo aquelas contra coqueluche (1914), difteria (1926) e tétano (1938). A combinação dessas três vacinas resultou na vacina DTP, administrada a partir de 1948, marcando o início de esforços significativos no controle de doenças. No início da década de 1950, o médico norte-americano Jonas Salk alcançou sucesso com a primeira vacina contra a poliomielite, testando-a em 1,6 milhão de crianças em 1954. Simultaneamente, o microbiologista Albert Sabin desenvolveu a vacina oral contra a doença. A década de 1960 viu o desenvolvimento da vacina contra o sarampo, seguido pelas vacinas contra caxumba e rubéola, combinadas na vacina tríplice viral por Maurice Hilleman em 1971. A vacina contra o *Haemophilus influenzae* (*Hib*) foi licenciada em 1985 e adicionada ao calendário recomendado em 1989, enquanto a vacina contra hepatite B foi incorporada em 1994.

Conforme mais vacinas se tornavam disponíveis, atualizações anuais no cronograma eram essenciais. Entre 1995 e 2010, varicela (1996), rotavírus (1998-1999, 2006, 2008), hepatite A (2000) e vacina pneumocócica (2001) foram introduzidas. Novas versões, como a vacina acelular contra coqueluche (DTaP, 1997) e influenza intranasal (2004), surgiram. O período entre 1995 e 2010 também trouxe recomendações adicionais para vacinas existentes, como influenza (2002) e hepatite A (2006). Em 2011, a vacinação rotineira contra o HPV foi recomendada para homens, e em 2018 a vacina contra influenza intranasal foi novamente sugerida.

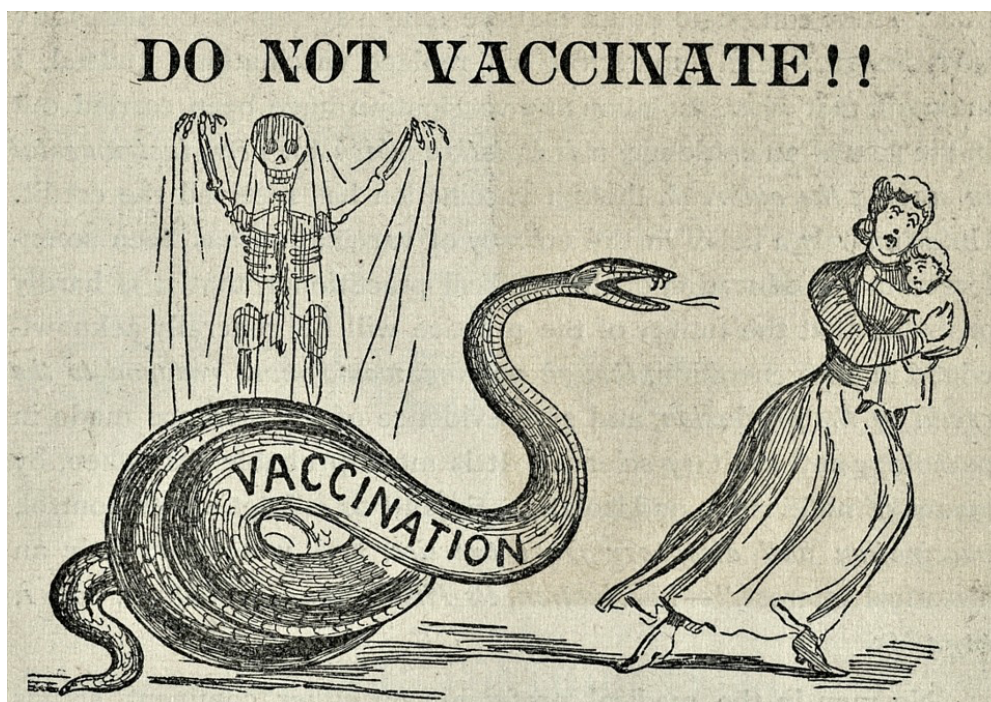
Durante a pandemia de Covid-19, a partir de 2020, recomendações para a vacina contra a doença foram ampliadas para grupos mais jovens. Indicada em 2021 para crianças de 5 a 11 anos, a vacinação foi estendida a menores de 5 anos em 2022. No Brasil, somente em 2024 a vacina contra a Covid-19 foi oficialmente incluída no cronograma de imunização.

O impacto positivo das vacinas na saúde pública global é hoje incontestável. Através de sua capacidade de induzir imunidade protetora contra agentes patogênicos, os imunizantes desempenharam um papel fundamental na erradicação e controle de graves doenças infecciosas. Campanhas de vacinação em massa permitiram a eliminação da varíola e a redução significativa da poliomielite e do sarampo em diversas regiões do planeta, salvando milhões de vidas anualmente. Campanhas contra doenças como difteria, tétano e gripe também se mostraram amplamente eficazes na prevenção de surtos e na proteção da saúde coletiva. Esse

histórico de resultados positivos evidencia a importância das vacinas como uma das mais relevantes ferramentas da medicina preventiva na sociedade contemporânea, constituindo um pilar da promoção do bem-estar e da redução do impacto negativo das doenças infecciosas em escala global<sup>5</sup>.

### História dos movimentos antivacina

**Figura 1** – Panfleto antivacina norte-americano de 1894



Fonte: The Historical Medical Library of The College of Physicians of Philadelphia.

A resistência à imunização não é um fenômeno recente, existindo desde os primórdios da própria prática vacinal (Figura 1). Apesar das vacinas terem desempenhado um papel fundamental na eliminação e na redução de muitas doenças, a propagação de boatos e informações equivocadas representa um desafio significativo na promoção da vacinação (Berman, 2020).

Após Jenner desenvolver a primeira vacina contra a varíola, em 1796, a ideia de imunização começou a ser aceita e considerada um avanço revolucionário na saúde pública. No entanto, à medida que os programas de vacinação foram sendo implementados e, especialmente, tornando-se compulsórios, movimentos de

---

<sup>5</sup> Informações sobre a história das vacinas encontradas no *site* da World Health Organization. Para detalhes, ver WHO ([s. d.]).



oposição começaram a surgir. A grande virada veio com a promulgação de leis de vacinação. Em 1853, o Reino Unido liderou o caminho, tornando a vacinação obrigatória para bebês com até 3 meses de idade. Posteriormente, em 1867, a exigência foi estendida até os 14 anos, com penalidades para os que recusassem a vacinação (Durbach, 2000).

Esses marcos legislativos foram o estopim para o surgimento de movimentos antivacina. As medidas governamentais na área da Saúde Pública geraram preocupações sobre a autonomia do corpo, a liberdade individual e a excessiva intervenção estatal na medicina. Essas apreensões alimentaram a resistência, principalmente entre aqueles que encaravam tais medidas como uma violação de suas liberdades pessoais e convicções religiosas. O contexto de vacinação compulsória alavancou o crescimento de grupos organizados se opondo à imposição da imunização e defendendo abordagens médicas alternativas (Berman, 2020).

Nos Estados Unidos, um evento semelhante ocorreu em 1905, quando a Suprema Corte confirmou uma lei de Massachusetts que exigia a vacinação obrigatória contra a varíola, estabelecendo, assim, a autoridade do estado para impor tal medida (Gostin, 2005). Nesse contexto, começaram a surgir grupos que se opunham à imunização no país. Muitos especialistas apontam que o movimento antivacina norte-americano atual pode ser rastreado até 1982, quando o canal de TV NBC exibiu um documentário intitulado *DPT: Vaccine Roulette*. O programa abordava uma controvérsia crescente na Inglaterra: a suposta ligação entre a vacina contra a coqueluche – uma doença potencialmente fatal que pode causar problemas pulmonares – e convulsões em crianças pequenas.

Embora sensacionalista e baseado em informações falsas, o programa teve um impacto sem precedentes no país. Através de ampla cobertura jornalística, o medo em relação à vacina DPT se disseminou pela população, resultando em processos judiciais contra fabricantes e escassez de suprimentos. Esse pânico, que durou de 1982 a 1986, influenciou tanto a mídia quanto o Congresso dos Estados Unidos. Em 1986, após anos de debate, o Ato Nacional de Lesões por Vacinação Infantil mudou a estrutura do sistema judicial americano, mostrando que o medo provocado pelo sensacionalismo midiático teve um impacto mais significativo do que qualquer evidência científica (Park, 2020).

Um evento decisivo para a disseminação mundial de desinformação sobre vacinas ocorreu na década seguinte, com a publicação de um artigo fraudulento do

cirurgião britânico Andrew Wakefield e colegas na renomada revista médica *Lancet*, em 1998, associando a vacina tríplice viral (que protege contra o sarampo, caxumba e rubéola) ao autismo (Wakefield *et al.*, 1998, retratados). Em 2010, o General Medical Council inglês publicou um relatório detalhando a atitude antiética e irresponsável dos autores. Wakefield enfrentou responsabilização criminal e teve sua licença médica revogada. O artigo em questão foi retirado dos arquivos do periódico (Rao; Andrade, 2011).

Contudo, a constante cobertura midiática acabou por desinformar o público sobre a tríplice viral (Dobson, 2003). Apesar do esforço de cientistas e organizações em todo o mundo para refutar os resultados de Wakefield e seus coautores, expondo a fraude científica que fundamentava o estudo, muitos pais optaram por não vacinar seus filhos devido ao receio de um suposto risco de autismo. Os surtos de sarampo no Reino Unido em 2008 e 2009, bem como os focos da doença nos Estados Unidos e Canadá, foram atribuídos à falta de vacinação de crianças. A fraude perpetrada por Wakefield provavelmente será lembrada como uma das mais graves da história médica (Rao; Andrade, 2011).

Ao longo do tempo, ideias, emoções e crenças profundamente enraizadas – sejam de natureza filosófica, política ou religiosa – que sustentam a oposição à imunização têm permanecido consistentes desde os dias de Edward Jenner, quando a vacinação foi introduzida. Com a busca por uma vacina contra o coronavírus, testemunhamos um renovado impulso de resistência por parte do movimento antivacina. Durante a pandemia, teorias da conspiração intermináveis e campanhas de desinformação *on-line* proliferaram em todo o mundo. Uma revisão histórica dos argumentos antivacinação pode nos ajudar a compreender mais sobre a hesitação vacinal, como será demonstrado mais adiante.

### **História da vacinação no Brasil**

Embora haja alguma divergência entre os historiadores, é bastante provável que a vacina contra a varíola tenha chegado pela primeira vez ao Brasil em 1804. Relatos do serviço de vacinação revelam que inicialmente houve uma aceitação da vacina na Corte, graças ao respaldo e à interferência da política absolutista de Dom João VI. Ele mandou vacinar seus filhos, Dom Pedro e Dom Miguel, em Portugal, e ordenou a tradução e a publicação da obra de Jenner sobre a vacina antivariólica. Em abril de 1811, foi estabelecida no Rio de Janeiro a Junta da Instituição Vacínica,

subordinada ao Intendente Geral da Polícia (Chalhoub, 1996).

Em 1846, a vacinação tornou-se obrigatória em todos os municípios do país por meio do Decreto Imperial n. 464, mas dificuldades técnicas encontradas inviabilizaram sua implementação. Somente em 1887 a técnica da vacina animal foi realmente introduzida no país. A vacina contra a varíola foi tornada compulsória no estado de São Paulo em 1891, como resultado da segunda lei estadual aprovada após a organização do serviço sanitário no estado (Chalhoub, 1996).

Embora a obrigatoriedade não tenha provocado grande resistência em São Paulo, uma década mais tarde ela desencadeou um conflito no estado do Rio de Janeiro (Tetarolli Júnior, 1996). Em 1904, diante de uma epidemia de varíola na cidade, então capital do país, o Governo Federal decreta a vacinação antivariólica obrigatória. A regulamentação da lei, feita por Oswaldo Cruz, diretor de saúde pública, era muito rígida, prevendo multas e demissões para os que não se vacinassem. Isso gerou grande oposição popular, já que a qualidade das vacinas e dos profissionais de saúde era questionada e muitos defendiam o direito de escolher se vacinar ou não. Após uma manifestação contra a vacinação compulsória ser reprimida violentamente pela polícia, a revolta popular cresceu nos dias seguintes, levando o governo a declarar estado de sítio e a mobilizar exército e marinha para conter a rebelião.

Diante da escalada dos protestos, o decreto da vacinação obrigatória foi revogado. A Revolta da Vacina fez parte de um processo mais amplo de remodelação urbana e sanitária do Rio de Janeiro que acabou levando à reestruturação dos serviços de saúde no país nos anos seguintes (Sevcenko, 1993). Vale ressaltar que a referida manifestação, além de ser o único registro significativo de uma ação antivacina com esse perfil no país, foi pontual e sem desdobramentos em longo prazo.

Em 1973, o Ministério da Saúde implementou o Programa Nacional de Imunizações (PNI)<sup>6</sup>, com o intuito de unificar e aprimorar as ações de imunização no país. Até então, essas ações eram marcadas por falta de continuidade, caráter episódico e alcance limitado. O PNI tornou-se um dos programas de imunização mais abrangentes do mundo, desempenhando um papel crucial na redução e na erradicação de doenças preveníveis por vacinação no Brasil. O sucesso do programa ao longo de décadas, com coberturas vacinais atingindo ou superando as metas

---

<sup>6</sup> Mais informações sobre o programa disponíveis em: <https://www.gov.br/saude/pt-br/acao-a-informacao/acoes-e-programas/programa-nacional-de-imunizacoes-vacinacao>. Acesso em: 21 mar. 2024.

estabelecidas, resultou em melhorias significativas para a saúde pública. Esse êxito também é atribuído à participação ativa e receptiva da sociedade brasileira nas estratégias implementadas pelo programa (Frugoli *et al.*, 2021).

No entanto, a partir de 2013, a taxa de cobertura anteriormente alta do PNI começou a declinar (Brown *et al.*, 2018). Uma pesquisa conduzida pela Unicef (2020) revelou uma baixa percepção entre os pais e responsáveis quanto aos reais riscos representados por doenças já erradicadas. A falta de uma exposição direta a seus efeitos resultou na perda de perspectiva sobre a gravidade dessas enfermidades. O êxito anterior das campanhas de vacinação também gerou certa apatia por parte das autoridades públicas, resultando na diminuição de iniciativas destinadas a ressaltar a importância da imunização. Concomitantemente, em 2016 ocorreu a deposição da presidenta Dilma Rousseff, como resultado de uma manobra neoliberal que reduziu drasticamente os recursos estatais destinados a questões sociais como a vacinação.

Observa-se uma lacuna significativa na pesquisa brasileira relacionada à hesitação vacinal (Frugoli *et al.*, 2021) e à aplicação do modelo dos 3Cs da OMS – confiança, complacência e conveniência (MacDonald, 2015). Enquanto países europeus, por exemplo, estão mais avançados nesse campo, o Brasil enfrenta um desafio de desconhecimento do perfil de hesitação vacinal em sua população. Essa falta de entendimento tem sido apontada como uma das principais causas para a diminuição das taxas de cobertura vacinal em outras nações (Succi, 2018).

Durante a pandemia de Covid-19, ocorreu uma notável diminuição nas campanhas de vacinação, acompanhada pelo crescente descrédito em relação à ciência por parte das autoridades, particularmente pelo Governo Federal. Estudos recentes indicam que a aplicação de alguns imunizantes chegou a ter queda de 65% em alguns estados brasileiros em 2020 (Zorzetto, 2022). Dados do PNI no DataSUS mostram uma queda histórica na imunização de crianças e adolescentes em 2021. A vacina tríplice viral (contra sarampo, caxumba e rubéola), que em 2015 chegou a 96% das crianças, caiu para 71%; a pentavalente (difteria, tétano, coqueluche, hepatite B e hemófilo B), caiu de 96% para 68% no mesmo período; e a de poliomielite (ou paralisia infantil), que caiu de 98% a 67%. Bolsonaro deixou a administração do país com um dos índices mais baixos de vacinação desde 2015 (Andrade; Dourado, 2022).

Apesar dos impactos da pandemia, que trouxe as vacinas e sua eficácia para o centro do debate público, com as redes sociais desempenhando um papel significativo na disseminação de movimentos de hesitação vacinal, há indícios de uma

possível reversão na tendência de queda dos índices de vacinação enfrentada pelo Brasil nos últimos anos. Segundo dados do Ministério da Saúde referentes ao período de janeiro a outubro de 2023, constatou-se um aumento na cobertura vacinal de oito imunizantes recomendados pelo calendário infantil, em comparação com os números totais do ano anterior. Destaca-se que, para crianças de 1 ano de idade, as vacinas contra hepatite A, poliomielite, pneumocócica, meningocócica, DTP e tríplice viral, na primeira e na segunda dose, apresentaram um aumento significativo em termos de adesão. Além disso, houve uma expansão na cobertura da vacina contra a febre amarela, indicada para crianças aos 9 meses de idade. Esse crescimento foi observado em todas as regiões do país, e o avanço pode ser em parte atribuído ao planejamento estratégico implementado pelo Ministério da Saúde desde o início do novo mandato do presidente Lula, que inclui iniciativas como o lançamento do já mencionado Movimento Nacional pela Vacinação e do programa Saúde com Ciência (Brasil..., 2023a).

\*

Por meio desta introdução, o presente estudo foi situado no âmbito dos principais temas convergentes às questões de pesquisa propostas. A estrutura do trabalho está dividida em três capítulos distintos. No primeiro, é apresentada uma breve introdução à Linguística Aplicada, seguida dos conceitos teóricos da Linguística de Corpus que fundamentam esta pesquisa, assim como os de Análise do Discurso. No segundo, é abordada a metodologia de pesquisa empregada, incluindo os passos para a compilação do corpus de estudo e os procedimentos metodológicos para a realização da Análise Multidimensional Lexical. Na terceira seção, são apresentadas a análise e discussão dos dados. Por fim, são tecidas considerações finais e explicitadas as referências bibliográficas.

## **1. Fundamentação teórica**

O presente estudo pertence ao domínio da Linguística Aplicada, explorando o emprego da língua em uso e seu impacto nos discursos em torno da vacina de Covid-19 no Brasil durante a pandemia. Concomitantemente, enquadra-se no arcabouço teórico da Linguística de Corpus, a qual se vale da análise de dados textuais em corpora eletrônicos utilizando programas computacionais. O método de pesquisa prevalente compreende a interpretação de dados estatísticos por meio de análises fatoriais e classificação de textos.

Inicialmente, apresentaremos um panorama geral da Linguística Aplicada, abordando os conceitos fundamentais que embasaram este estudo. Em seguida, traçaremos um panorama teórico e histórico da Linguística de Corpus, destacando os princípios de duas metodologias essenciais para o escopo desta pesquisa: a Análise Multidimensional (Berber Sardinha; Veirano Pinto, 2019; Biber, 1988) e a Análise Multidimensional Lexical, central em nossa investigação (Berber Sardinha, 2020, 2021).

Como mencionado anteriormente, por meio da Linguística de Corpus, com base na Análise Multidimensional, o objetivo deste estudo é aprofundar a análise da circulação de discursos nas redes sociais, um campo que ainda necessita de uma investigação mais detalhada no âmbito da Linguística. Para isso, empregaremos o conceito de registro, que determina as escolhas linguísticas específicas de um falante ou escritor em diferentes contextos (Biber; Conrad, 2009), a fim de examinar a linguagem das mídias digitais, com um enfoque particular no Twitter.

Por fim, com os resultados em mãos, investigaremos as dimensões resultantes da Análise Multidimensional Lexical com o suporte da Análise Crítica do Discurso e, adicionalmente, da Análise Dialógica do Discurso, fundamentada na perspectiva teórica de Mikhail Bakhtin, filósofo e pensador russo. Para tanto, mobilizaremos os conceitos de enunciado e cadeia discursiva, conforme preconizado pela metodologia bakhtiniana.

### **1.1. Linguística Aplicada**

A obra de Maria Antonieta Alba Celani apresenta uma reflexão necessária sobre os desafios envolvendo uma definição precisa da Linguística Aplicada (doravante LA) e sua legitimidade como uma “área de conhecimento com autonomia”

(Celani, 1992, p. 15). A autora delinea um panorama histórico dessa disciplina, que tem suas raízes fincadas no final do século XIX e ganha maior relevância no último quarto do século XX, com o estabelecimento de associações e programas de pós-graduação, bem como o surgimento de agências de financiamento à pesquisa. Segundo Celani (1992), a LA é uma disciplina que se entrelaça com outras áreas, como Psicologia, Antropologia e Pedagogia, dependendo da natureza do estudo em questão.

Na pesquisa aqui apresentada, a LA estabelece conexões significativas com diversas disciplinas acadêmicas. Em especial, evidencia-se sua interseção com as Ciências da Computação pelo uso de ferramentas informáticas e princípios de programação para desenvolver *scripts* destinados à limpeza de textos e à identificação de variáveis relevantes. Da mesma forma, sua colaboração com a Matemática e a Estatística se faz presente através da aplicação de algoritmos para discernir as dimensões temáticas presentes nos dados. Além disso, sua relação com a Análise do Discurso é estabelecida por meio da análise qualitativa dos textos investigados.

Essa abordagem multidisciplinar não apenas enriquece a compreensão do fenômeno em estudo, mas também demonstra a amplitude de aplicação e relevância da LA. A interdisciplinaridade é um dos elementos fundamentais para o aprimoramento contínuo da LA contemporânea, visto que se trata de um campo dinâmico que permite um movimento constante de construção e reconstrução de conhecimentos, dada a complexidade inerente às interações humanas (Moita Lopes, 2006). Diante do contínuo fluxo de transformações nas dinâmicas sociais, é improvável que todas as possíveis questões relacionadas aos diversos usos da linguagem tenham sido abordadas em sua totalidade (Veiga, 2020). Assim, o propósito central da LA é abordar os “problemas humanos que surgem dos diversos usos da linguagem” (Celani, 1992, p. 21), adotando uma perspectiva crítica e interdisciplinar que captura os desafios emergentes nas práticas linguísticas.

Conforme destacado por Pennycook (1998), a LA engloba dois aspectos intrinsecamente políticos da existência humana: linguagem e educação. Logo, segundo o autor, é incumbência da LA examinar a base ideológica do conhecimento que ela produz. Pennycook (1998) enfatiza a necessidade premente de professores e pesquisadores adotarem posturas analíticas, visando melhorar e transformar um mundo caracterizado pela desigualdade, além de compreender as implicações políticas de suas práticas.

Uma LA crítica deve reconhecer, de saída, que o conhecimento gerado está inextricavelmente ligado a interesses específicos, de modo a tornar crucial o rompimento com modelos de investigação desprovidos de contexto social, apolíticos e a-históricos (Pennycook, 1998). Nesse sentido, este estudo propõe que, como pesquisadores de LA, reconheçamos a natureza ideológica subjacente ao nosso trabalho e repensemos o papel da linguagem como um componente imprescindível não apenas na transformação da realidade social, mas também na compreensão do mundo contemporâneo.

A herança de uma abordagem apolítica e a-histórica da linguagem, positivista, centrada na concepção de um indivíduo racional, resultou em concepções de linguagem e comunicação que negligenciam questões relacionadas ao poder e à desigualdade, falhando em explicar os conflitos que emergem em torno da construção de significados. Nesse contexto, a presente pesquisa adota uma perspectiva interpretativista, que enriquece o estudo da linguagem ao considerar as diversas subjetividades dos participantes imersos no contexto social em análise (Moita Lopes, 1994), bem como as novas dinâmicas de comunicação apresentadas pelas redes sociais. Utilizaremos a abordagem interpretativista para buscar compreender a experiência humana por meio do discurso, reconhecendo-o como uma construção social em constante evolução e negociação de sentidos, permeada por relações de poder e ideologias.

## **1.2. Linguística de Corpus**

Esta pesquisa está fundamentada na Linguística de Corpus (doravante LC), que pode ser descrita como uma abordagem para o estudo da língua em uso, abrangendo uma variedade de propósitos e métodos de pesquisa, “com ênfase na descrição da linguagem e na observação sistemática do comportamento linguístico, não na intuição de falantes nativos” (Conrad, 2011, p. 49). A LC foi desenvolvida com o intuito de superar as limitações intrínsecas à compreensão linguística impostas pela consciência humana. Assim como as calculadoras foram concebidas para lidar com operações matemáticas complexas, que ultrapassam a capacidade de nosso processamento mental, a LC se apresenta como um instrumental analítico capaz de lidar com a complexidade da linguagem que excede nossas capacidades cognitivas individuais. Essa abordagem, no entanto, demandou tempo para amadurecer completamente, visto que a linguagem é ainda mais intrincada e multifacetada do que



as operações matemáticas. Com o avanço das tecnologias computacionais e o desenvolvimento de métodos e técnicas específicas, a LC se consolidou como uma abordagem comprovadamente eficaz para a análise e a compreensão da linguagem de formas anteriormente inacessíveis, possibilitando a investigação de padrões e variação em grandes volumes de dados linguísticos autênticos.

De acordo com Berber Sardinha (2004, p. 3), a abordagem:

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

Em complemento à afirmação de Berber Sardinha (2004), Baker e Egbert (2016) situam a LC como um método ou conjunto de princípios e procedimentos que se baseia em extensas coleções de textos de linguagem em uso natural, sejam escritos, falados ou mediados por computador, que são amostrados e equilibrados para representar uma variedade específica da língua. A característica central da LC é sua capacidade de investigar a real língua em uso, utilizando corpora de textos e descrevendo padrões linguísticos por meio da análise destes.

Diferentemente de métodos que se baseiam apenas em modelos artificiais da linguagem, a LC emprega exclusivamente exemplos reais de uso da língua. Conforme aponta Baker (2023), a LC possibilita ao pesquisador fazer generalizações sobre as variedades linguísticas analisadas, combinando elementos automatizados e humanizados para a análise da língua. A automação auxilia na identificação de aspectos do corpus que o pesquisador talvez não tenha considerado à primeira vista. Contudo, a pesquisa também requer a intervenção humana para interpretar os resultados encontrados.

Uma definição atualizada de Berber Sardinha e Moreira (2023) estabelece a LC como uma área dos estudos linguísticos que emprega corpora, isto é, coleções de textos de diversos modos semióticos (falados, escritos, sonoros ou visuais) armazenadas em formato de computador, com a finalidade de descrever situações de uso das linguagens e entender a relação entre o uso sistemático de recursos de expressão e os contextos em que esses recursos são produzidos.

A LC adota uma abordagem empirista que parte do pressuposto de que a língua é um sistema probabilístico no qual os traços linguísticos não ocorrem de

maneira aleatória e as palavras não se combinam ao acaso. Assim, é possível identificar e quantificar padrões presentes em uma língua ou em uma variedade linguística, estabelecendo uma correlação entre os traços linguísticos e os contextos situacionais de uso da linguagem (Berber Sardinha, 2004). Essa abordagem, pioneira na LA, revelou uma nova perspectiva para o estudo da linguagem, cujas descobertas têm contribuído significativamente para a pesquisa em diversas áreas, como ensino de língua estrangeira, análise de gênero, descrição linguística, elaboração de dicionários, extração e tratamento de metáforas, linguística forense e tradução (Berber Sardinha, 2009).

### *1.2.1. Linguística de Corpus: histórico*

Estudos que envolvem a compilação manual de grandes conjuntos de textos são abundantes ao longo da história da humanidade. Na Grécia Antiga, Alexandre, o Grande, estabeleceu o Corpus Helenístico, enquanto na Antiguidade e na Idade Média foram produzidos corpora de citações da Bíblia (Berber Sardinha, 2004). Outras compilações estão associadas à elaboração de dicionários, como o abrangente dicionário de inglês compilado pelo Dr. Samuel Johnson em 1755 e o Oxford English Dictionary, que foi manualmente compilado durante a década de 1880. Destaca-se também a análise de Becket das obras de Shakespeare, a qual identificou, no século XVIII, concordâncias de uso (McCarthy; O'Keeffe, 2010).

No século XX, educadores como Thorndike (1927) e linguistas de campo como Boas (1911) se dedicaram à descrição da linguagem através de corpora, lançando as bases para futuros estudos. Charles Carpenter Fries utilizou um corpus de cartas governamentais para desenvolver uma gramática do uso da língua em 1940, além de um corpus de conversas telefônicas para elaborar uma gramática da língua falada em 1952 (Biber, 2010).

Um marco na transição para a análise computacional foi o trabalho de Roberto Busa, que criou um índice lematizado eletrônico de toda a obra de São Tomás de Aquino iniciado na década de 1950 e concluído no final da década de 1970 (McCarthy; O'Keeffe, 2010). A década de 1960 trouxe outro marco importante com o lançamento do primeiro corpus linguístico eletrônico, o *Corpus Brown*, que inaugurou a era dos corpora eletrônicos (Berber Sardinha, 2002). Entre os corpora pioneiros do armazenamento digital estão o LOB (*Lancaster-Oslo-Bergen*, 1978, de 1 milhão de palavras do inglês britânico escrito) e o LLC (*London-Lund Corpus*, 1980, de 500 mil

palavras do inglês britânico falado). Biber (1988) utilizou partes dos dois últimos – com a seleção de alguns subcorpora de gêneros textuais – para seu notório estudo sobre variação linguística entre o inglês falado e escrito.

Destacam-se ainda as investigações conduzidas por Sinclair (1972, 1987, 1991), que analisou instâncias da língua em uso, compiladas em um corpus, visando extrair conceitos gramaticais pertinentes à sua funcionalidade. A aplicação desse conhecimento beneficiou significativamente os domínios lexicográfico e pedagógico, contribuindo para o desenvolvimento de dicionários, gramáticas, abordagens de ensino e materiais didáticos inovadores. No campo teórico, sua pesquisa reforçou a ideia de que as dimensões lexical e gramatical estão interconectadas, em contraste com abordagens tradicionais que as estudavam de forma fragmentada. A noção de léxico-gramática foi aprimorada através das investigações de Sinclair (Stubbs, 1993), que demonstraram a interdependência entre o léxico e a gramática ao examinar o fenômeno da coocorrência lexical e os padrões de uso das formas linguísticas (Sinclair, 1991).

Na década de 1990, a LC consolidou-se como uma abordagem metodológica e teórica inovadora para o estudo da linguagem, reconhecendo-se a importância dos corpora e da análise baseada em dados reais na pesquisa linguística (Biber; Conrad; Reppen, 1998). Uma variedade de corpora acessíveis por computador já estava disponível e o interesse por pesquisas baseadas em corpus começou a aumentar. Taylor, Leech e Fliglestone (1991) examinaram 36 corpora em inglês que podiam ser lidos por máquinas, enquanto Altenberg (1991) compilou uma bibliografia com cerca de 650 estudos baseados em textos extraídos dos corpora mais significativos (Biber, 1995). No século XXI, houve uma expansão notável do uso de corpora e ferramentas computacionais na análise linguística, com aplicações em áreas como tradução automática, processamento de linguagem natural e ensino de línguas. O desenvolvimento teórico e prático da LC continua avançando, oferecendo contribuições significativas para a compreensão da linguagem e suas variações (O’Keeffe; McCarthy; Carter, 2007).

No contexto brasileiro, a LC ainda se encontra em estágio inicial de desenvolvimento. Tipicamente, as pesquisas em corpus são conduzidas em centros especializados em Processamento de Linguagem Natural, Lexicografia e Linguística Computacional (Berber Sardinha, 1999). No entanto, observa-se um crescimento da

presença da LC no setor empresarial, no qual há um avanço no interesse em aplicações comerciais de estudos baseados em corpora (Berber Sardinha, 2002).

Em ambientes acadêmicos, dentre os grupos de pesquisa em LC no Brasil destaca-se a atuação do Grupo de Estudos de Linguística de Corpus (GELC), do qual faço parte como pesquisadora, vinculado ao Programa de Linguística Aplicada e Estudos da Linguagem da Pontifícia Universidade Católica de São Paulo (PUC-SP), sob a liderança do Professor Doutor Tony Berber Sardinha. As pesquisas realizadas pelo GELC abrangem diversas áreas, incluindo Aeronáutica (Zuppardo, 2014), Cinema e Televisão (Araújo, 2017; Veirano Pinto, 2013), Direito (Freire, 2010; Scaramuzzi-Rodrigues, 2016), Educação e Ensino de Línguas (Bissaco, 2010; Contrera, 2010; Ferreira, 2010; Mayer, 2012; Delfino, 2016), Jornalismo (Kauffmann, 2005; Lopes, 2010; Souza, 2012), Linguística Forense (Gonzales, 2019), Literatura (Kauffmann, 2020), Medicina (Silva e Teixeira, 2010), Música (Bertóli-Dutra, 201; Delfino, 2022), Tradução (Resende, 2019), Terminologia (Veiga, 2014) e Sustentabilidade (Brogini, 2023). Reconhecido pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) desde 2008, o GELC mantém uma frequência de reuniões semanais para discutir projetos, métodos e perspectivas futuras na área de LC.

### **1.3. Corpus**

É imprescindível que toda e qualquer exploração na área da LC esteja intrinsecamente ligada à organização, estruturação e montagem de corpora. Conforme destacado por Berber Sardinha (2004, p. 3), corpora são “[...] conjuntos de dados linguísticos textuais, em formato legível por computador e coletados criteriosamente, com o propósito de servirem para a pesquisa de uma língua ou variedade linguística”.

Existem diversas categorias de corpora que podem ser desenvolvidas e aplicadas no âmbito das pesquisas em LC, incluindo corpora gerais, corpora de referência, corpora comparáveis, corpora paralelos, corpora de aprendizes e corpora históricos, dentre outros. No entanto, é fundamental destacar que a LC, como toda e qualquer ciência, segue critérios rigorosos. Assim, embora adote uma abordagem empírica, não considera que todo conjunto de dados pode ser designado como um corpus.

Sanchez (1995, p. 8-9) propõe uma definição que engloba as características essenciais do que constitui um corpus:

Um conjunto de dados linguísticos (seja proveniente do uso oral ou escrito da língua, ou de ambos), organizado segundo critérios específicos, com amplitude e profundidade suficientes para representar abrangentemente o uso linguístico total ou parte dele, e disposto de maneira a ser processado por computador, com o intuito de fornecer resultados variados e úteis para descrição e análise.

Segundo Berber Sardinha (2002), a criação de um corpus deve seguir critérios científicos rigorosos, a saber:

- Autenticidade dos dados: os textos incluídos no corpus devem ser autênticos, ou seja, reais e em linguagem natural. Eles não devem ter sido produzidos com o propósito específico de serem utilizados em pesquisas linguísticas e não devem ter sido criados artificialmente, como por linguagem de programação ou inteligência artificial.
- Finalidade do estudo linguístico: o corpus deve ser criado com o propósito de servir como objeto de estudo linguístico. Os dados contidos no corpus são coletados e organizados para permitir a análise e a investigação de fenômenos linguísticos, padrões de uso da linguagem e outras questões linguísticas relevantes.
- Seleção de conteúdo: a composição do corpus deve ser feita cuidadosamente, levando em consideração os objetivos da pesquisa e a representatividade dos dados linguísticos. A seleção dos textos incluídos no corpus deve garantir a diversidade e a relevância dos dados para a análise linguística.
- Legibilidade por computador: os dados contidos no corpus devem ser formatados de modo que sejam legíveis e processáveis por computador. Isso envolve a organização dos textos em um formato adequado para análise computacional, permitindo a aplicação de ferramentas e técnicas de processamento de linguagem natural e análise de corpus.

Ao conceituar o corpus como um repositório de dados linguísticos abrangendo o uso tanto falado quanto escrito de uma língua, e ao considerar que esses “dados linguísticos” são compostos por unidades autônomas denominadas “textos”, é

essencial definir o que é texto dentro dessa conjuntura. Em termos gerais, podemos entender um texto como uma unidade autônoma e identificável da linguagem natural utilizada para comunicação, seja na forma oral ou escrita (Biber; Conrad, 2009). Os textos compilados para esta investigação – *tweets* extraídos da plataforma Twitter – seguem a definição proposta por Biber e Conrad (2009) e podem ser vistos como unidades independentes e reconhecíveis.

### 1.3.1. Tipologia de corpus

Existe uma vasta diversidade de categorias de corpus que podem ser compiladas, cada uma apresentando características únicas e destinadas a aplicações específicas. Os corpora podem ser classificados de diversas maneiras, podendo ser agrupados de acordo com diferentes critérios, como a natureza dos textos incluídos, o propósito da pesquisa e a variedade linguística estudada, uma vez que a nomenclatura empregada é extensa e diversificada. O Quadro 1 tem como base tipologias propostas por Berber Sardinha (2004) e visa resumir os critérios subjacentes aos diferentes tipos de corpus empregados em estudos linguísticos:

**Quadro 1** – Tipologia dos corpora

<b>MODO</b>	Falado	Composto de porções de fala transcritas.
	Escrito	Composto de textos escritos, impressos ou não.
<b>TEMPO</b>	Sincrônico	Compreende um período de tempo.
	Diacrônico	Compreende vários períodos de tempo.
	Contemporâneo	Representa o período de tempo corrente.
	Histórico	Representa o período de tempo passado.
<b>SELEÇÃO</b>	De amostragem*	Composto de porções de textos ou de variedades textuais e planejado para ser uma amostra finita da linguagem como um todo.
	Dinâmico ou orgânico	Corpus monitor com possibilidade de crescimento ou diminuição.
	Estático	Corpus de amostragem fixo em sua composição.
	Equilibrado**	Corpus com componentes distribuídos em quantidades semelhantes.

<b>CONTEÚDO</b>	Especializado	Composto de textos de um tipo específico.
	Regional ou dialetal	Composto de textos provenientes de uma ou mais variedades sociolinguísticas específicas.
	Multilíngue	Composto de textos escritos em diferentes idiomas.
<b>AUTORIA</b>	De aprendiz	Composto de textos cujos autores não são falantes nativos.
	De língua nativa	Composto de textos cujos autores são falantes nativos.
<b>DISPOSIÇÃO INTERNA</b>	Paralelo	Corpus com textos comparáveis – por exemplo, original e tradução.
	Alinhado	Corpus com traduções abaixo de cada linha original.
<b>FINALIDADE</b>	De estudo	Corpus que se pretende descrever.
	De referência	Corpus usado para fins de contraste com o corpus de estudo.
	De treinamento ou teste	Corpus construído para permitir o desenvolvimento de aplicações e ferramentas de análise.
<p>* Do inglês <i>sample corpus</i>.</p> <p>** Do inglês <i>balanced corpus</i>.</p>		

Fonte: Adaptado de Toledo Dias (2020, p. 15).

Compreender essas definições é fundamental para que o pesquisador possa delinear claramente os parâmetros do seu corpus de estudo e fundamentar suas decisões em relação à inclusão ou exclusão de determinados textos ou variedades linguísticas na amostra de pesquisa. O corpus utilizado neste estudo é uma coleção de textos escritos, sincrônica e especializada, composta exclusivamente por *tweets* da plataforma Twitter.

### 1.3.2. Design de corpus

O desenho ou *design* do corpus refere-se à escolha dos textos que compõem a amostra estudada. A construção de um corpus deve estar fundamentada em bases sólidas, levando em consideração o tamanho da amostra não apenas em termos de número de palavras ou textos, mas também no que diz respeito à inclusão de uma

gama completa de variabilidade em uma população<sup>7</sup>. O *design* do corpus deve ser sistemático, abrangente, legível por computador e útil para análise. Essa abordagem é essencial para que a pesquisa em LC possa fornecer resultados relevantes (Biber, 1993).

Em sua essência, um corpus, independentemente de seu tipo, é considerado representativo da linguagem, de um idioma ou de uma variedade dele (Berber Sardinha, 2002). Garantir a representatividade dos dados coletados é um dos grandes desafios na construção de corpora. Um corpus representativo é uma amostra intencional que reflete um domínio-alvo ou uma população linguística bem-definida. Segundo Biber (1993), a representatividade está relacionada à extensão na qual uma amostra inclui o escopo completo de variação em uma população. Embora o corpus deva ser vasto para ser representativo (Berber Sardinha, 2002), o seu tamanho não pode compensar um design inadequado (Biber, 1993).

Para este estudo, optamos por delimitar o corpus de modo a incluir exclusivamente *tweets* contrários à vacinação provenientes de contas infodêmicas e publicados durante a pandemia de Covid-19 (entre 2020 e 2022). O delineamento do corpus não pode ser dissociado do design da pesquisa (Egbert, 2019), portanto, essa decisão metodológica está alinhada com o objetivo central do estudo, que busca compreender os discursos dos movimentos antivacina dentro deste contexto específico, realizando uma análise detalhada das narrativas<sup>8</sup>, estratégias retóricas<sup>9</sup> e motivações subjacentes a essas mensagens. Embora essa abordagem resulte em um volume reduzido de dados selecionados, em comparação com um corpus mais amplo que incluiria uma variedade de discursos sobre vacinas, ela nos proporciona um corpus altamente infodêmico, delineado para atender às necessidades específicas desta pesquisa.

Ademais, os discursos de oposição à vacinação são minoritários no Twitter, e como tal, requerem uma mineração<sup>10</sup> minuciosa para serem identificados, uma vez que sua detecção não é trivial. Isso foi evidenciado pela pesquisa de Müller, Salathé

---

<sup>7</sup> População é o conjunto completo de todos os elementos que estão sendo estudados, enquanto a amostra é uma porção selecionada dessa população.

<sup>8</sup> Aqui o termo é utilizado para definir o processo ou efeito de narrar, sem ser enquadrado como conceito dentro de uma teoria específica.

<sup>9</sup> Aqui o termo é utilizado para definir a arte do bem dizer, da eloquência, sem ser enquadrado como conceito dentro de uma teoria específica.

<sup>10</sup> Mineração de dados (também conhecida pelo termo inglês *data mining*) é o processo de explorar dados à procura de padrões consistentes para detectar relações sistemáticas entre variáveis, detectando assim novos subconjuntos de dados.



e Kummervold (2023), que analisou uma extensa coleção de *tweets* geolocalizados nos Estados Unidos, abordando os temas sarampo e vacinação. Os autores categorizaram os posts em três grupos distintos: positivo (expressando apoio à vacinação), com uma representação de 51,9%, neutro/outros (irrelevantes ou ambíguos em relação ao tema), com uma frequência de 41,0%, e negativo (oposição à vacinação), com uma incidência de 7,1%. Souza e Loguercio (2024) também afirmam que a despeito de métricas revelarem um nível de atividade mais elevado em termos de compartilhamento de mensagens, os ativistas antivacina são menos numerosos na rede.

Embora os *tweets* antivacina representem uma minoria em relação aos pró-vacina, eles frequentemente recebem maior destaque e repercussão nas redes sociais. Essa disparidade é impulsionada por uma série de fatores complexos, indicando que a disseminação de desinformação ocorre de maneira consideravelmente mais rápida e abrangente no Twitter do que a disseminação de informações precisas.

Um estudo (Vosoughi *et al.*, 2018) conduzido por pesquisadores do Massachusetts Institute of Technology (MIT) examinou aproximadamente 126.000 histórias compartilhadas por cerca de 3 milhões de usuários do Twitter no período de 2006 a 2017, constatando que as postagens contendo desinformação tinham cerca de 70% mais chances de serem *retweetadas* do que informações confiáveis. Contrariando as expectativas, este desequilíbrio é impulsionado principalmente pela ação de usuários, e não por contas automatizadas, conhecidas como *bots*. Os autores observaram que os *bots* desempenharam um papel semelhante na propagação de ambos os tipos de postagens, sugerindo que seres humanos foram mais diretamente responsáveis pelo compartilhamento de desinformação.

O referido estudo adicionalmente investigou elementos estruturais do Twitter, bem como as características individuais dos usuários envolvidos nas conversações, visando encontrar pistas sobre como ocorre a propagação de desinformação. A hipótese era que disseminadores de desinformação deveriam possuir um maior número de seguidores, ser mais ativos na plataforma, ter o selo de usuário verificado ou ter uma conta mais antiga. No entanto, ao compararem os usuários envolvidos em discussões com rumores verídicos e falsos, os pesquisadores descobriram que o oposto era verdadeiro em todos os casos. Fatores como número de seguidores, níveis de atividade, status de verificação e idade da conta não foram determinantes para

explicar a maior disseminação de notícias falsas. De fato, os usuários que compartilhavam informações não confiáveis apresentavam valores significativamente mais baixos nessas métricas de características da rede em comparação com aqueles que compartilhavam notícias verídicas. A desinformação se espalhou mais amplamente apesar dessas diferenças e não devido a elas. Vosoughi e seus colegas sugerem que usuários podem ser mais inclinados a compartilhar desinformação devido ao seu caráter surpreendente, semelhante ao atrativo de manchetes sensacionalistas de *click bait*.

A pesquisa apresentou ainda diversas métricas para quantificar esse fenômeno. Por exemplo, observou-se que informações precisas demandam aproximadamente seis vezes mais tempo para atingir 1.500 pessoas em comparação com postagens de desinformação. No contexto das “cascatas” do Twitter, isto é, cadeias de *retweets* contínuos, *posts* de desinformação atingem uma profundidade de cascata de 10, aproximadamente 20 vezes mais rápido do que fatos verídicos (ou seja, são *retweetados* por 10 níveis de usuários, aproximadamente 20 vezes mais rapidamente).

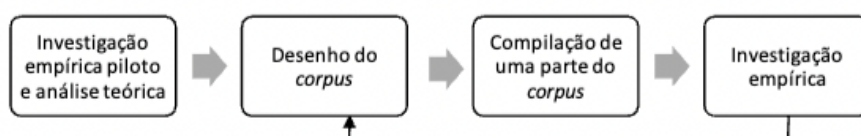
Em consonância com o estudo, Ceylan, Anderson e Wood (2023) defendem que o principal impulsionador da propagação de desinformação nas redes sociais é a estrutura de recompensas nessas plataformas, que incentiva os usuários a adotarem o hábito de compartilhar notícias que geram engajamento e recebem reconhecimento social, independentemente de sua veracidade. Uma vez formados esses hábitos de compartilhamento, os usuários respondem automaticamente aos estímulos presentes nas plataformas, mostrando-se relativamente insensíveis à precisão ou às inclinações políticas das informações compartilhadas, e até mesmo ao fato delas muitas vezes entrarem em conflito com suas crenças pessoais.

Dessa forma, embora o recorte do corpus possa inicialmente parecer restritivo, na realidade, abarca *tweets* que exerceram uma influência ampla e significativa nas discussões no Twitter sobre o tema. Portanto, o corpus coletado criteriosamente para este estudo circunscreve postagens antivacina que influenciaram ativamente as conversas em torno do assunto na plataforma durante o período selecionado.

### 1.3.3. Compilação de corpus

Biber (1993) propõe um processo cíclico de compilação de corpora, conforme ilustrado na Figura 2:

**Figura 2** – Processo cíclico para a compilação de um corpus



Fonte: Adaptada de Biber (1993, p. 256).

De acordo com Biber (1993), antes da coleta do corpus, é preciso realizar uma análise teórica para identificar os parâmetros do estudo e justificar a escolha dos textos. Isso envolve definir os objetivos da pesquisa e determinar as características linguísticas a serem investigadas. Em seguida, aparece o *design* do corpus, que consiste na seleção dos textos para compor a amostra. Nessa fase, são estabelecidos critérios para inclusão ou exclusão, levando em conta aspectos linguísticos e contextuais.

Após definir o desenho, começa a compilação do corpus, na qual os textos são coletados e organizados. É essencial garantir que a amostra seja representativa da população de textos do tema escolhido. Com o corpus compilado, inicia-se a investigação empírica, que envolve a análise dos textos com base nos objetivos da pesquisa. Nesse ponto são realizadas análises linguísticas e interpretação dos resultados obtidos a partir do corpus.

O processo destaca a importância de uma abordagem sistemática e iterativa na construção e análise de corpora, permitindo ajustes e refinamentos ao longo do processo para garantir a validade e a relevância dos resultados da pesquisa linguística. Egbert (2019, p. 36) utilizou o modelo cíclico de construção de corpora proposto por Biber (1993) como base para a elaboração de um corpus de pesquisa, expandindo-o em nove passos distintos:

- Passo 1: estabelecer (e projetar) objetivos e *design* da pesquisa;
- Passo 2: definir o domínio-alvo (a população);
- Passo 3: projetar o corpus;

- Passo 4: coletar a amostra;
- Passo 5: anotar o corpus;
- Passo 6: avaliar a representatividade do domínio-alvo;
- Passo 7: avaliar a representatividade linguística;
- Passo 8: repetir os passos 3-5, se necessário;
- Passo 9: elaborar relatórios.

A etapa final da expansão proposta por Egbert (2019) – elaborar relatórios – merece destaque, pois aponta a importância da documentação detalhada do processo de concepção e compilação do corpus ao longo da pesquisa. Essa documentação serve tanto para orientar o pesquisador como para embasar pesquisas futuras. Isso reflete a disposição dos pesquisadores de LC para o compartilhamento de melhores práticas, essencial para o avanço da área. A documentação cuidadosa e a colaboração entre pesquisadores são fundamentais para o progresso e a credibilidade das pesquisas em LC.

#### **1.4. Registro**

Na LC, em termos gerais, registro se refere a qualquer variedade textual associada a contextos situacionais ou propósitos comunicacionais específicos. Embora as distinções de registro sejam definidas em termos não-linguísticos, geralmente existem diferenças linguísticas importantes entre os múltiplos registros, que se manifestam através de escolhas específicas no léxico, na gramática e na estrutura do texto (Biber, 2015). Portanto, as relações funcionais entre o contexto situacional e as características linguísticas emergem como o principal componente na descrição de um registro.

Um dos principais argumentos sustentados nessa definição é que as características linguísticas são sempre funcionais. Em outras palavras, certas características linguísticas são recorrentes em determinado registro porque se mostram especialmente adequadas aos seus propósitos comunicativos ou ao contexto situacional. Logo, essa análise funcional é um componente essencial em qualquer descrição de registro (Biber; Conrad, 2009).

Em muitos casos, os registros são identificados como variedades dentro de uma cultura – romances, cartas, editoriais, sermões e debates. Podem ser definidos em diferentes níveis de generalidade: por exemplo, a prosa acadêmica representa um

registro bastante amplo, enquanto os artigos metodológicos em Psicologia constituem um registro mais específico (Biber, 1995).

Tradicionalmente, a complexidade da variação de registros em uma língua não era adequadamente abordada pelas descrições linguísticas convencionais. Cada registro possuía sua própria gramática de uso, influenciada por diferentes contextos e objetivos comunicativos (Biber; Conrad, 2009). Para superar essa limitação, Biber (1988) desenvolveu a Análise Multidimensional, uma abordagem multivariada abrangente para comparar vários registros ao longo de dimensões específicas de variação, conforme veremos detalhadamente na próxima seção.

Biber e Egbert (2016) explicam que as diferenças entre textos são mais bem-compreendidas em termos de conjuntos de características gramaticais que coocorrem embasadas funcionalmente. Essa abordagem implica que o propósito de um texto pode ser revelado pela análise de seus padrões de coocorrência<sup>11</sup> gramatical. A Análise Multidimensional desenvolvida por Biber (1988) destaca-se por estar teoricamente fundamentada na noção de coocorrência linguística.

No estudo seminal de 1988, Biber detalhou como características linguísticas coocorrentes em inglês se unem para permitir que os usuários realizem diferentes funções na linguagem escrita e falada, agrupando-as em dimensões de variação. As cinco dimensões encontradas receberam as seguintes denominações:

- Dimensão 1 – Produção interativa *versus* produção informacional.
- Dimensão 2 – Preocupações narrativas *versus* não-narrativas.
- Dimensão 3 – Referências explícitas *versus* referências dependentes de contexto.
- Dimensão 4 – Expressão explícita de persuasão *versus* não explícita.
- Dimensão 5 – Informação abstrata *versus* não-abstrata.

Essas dimensões têm mantido sua relevância, sendo consideradas indicadores confiáveis de variação no inglês (Berber Sardinha; Veirano Pinto, 2019; Friginal, 2013). Ao longo dos anos, a Análise Multidimensional foi aplicada a uma ampla variedade de idiomas e domínios discursivos, fornecendo evidências de que a variação de registro é a norma, e não a exceção, no uso da linguagem (Berber Sardinha, 2018). Uma vez que o registro é um indicador fundamental da variação linguística (Biber, 2012), neste estudo o entendimento do Twitter como um registro

---

<sup>11</sup> Relação entre uma unidade linguística e outra que ocorre com ela no mesmo contexto.

específico, com suas características e coerções singulares, mostra-se essencial para a análise dos resultados que será apresentada mais adiante.

#### 1.4.1. *Tweet como um registro*

Textos gerados em contextos semelhantes têm a tendência de apresentar padrões linguísticos correlatos (Gray; Egbert, 2019). As postagens – *tweets* – produzidas pelos usuários do Twitter exibem um conjunto de características que definem a plataforma como um registro singular, demarcando-a de maneira única em relação a outras formas de expressão. Ao examinarmos as dinâmicas do Twitter, torna-se evidente que diversos recursos-chave exercem uma influência significativa sobre a linguagem. Cada um desses elementos impõe coerções distintas à expressão do usuário, moldando, assim, o estilo comunicativo geral da plataforma.

Primordialmente, as características situacionais do Twitter desempenham um papel essencial no delineamento do uso da linguagem na plataforma. A brevidade das postagens, uma característica distintiva da plataforma, teve sua origem no limite inicial de 140 caracteres, que refletia a adaptação da rede para mensagens de SMS (Zappavigna, 2018). Essa limitação compele os usuários a destilarem suas ideias de forma sucinta, o que resulta no uso de abreviações, siglas, frases condensadas, gramática e pontuação simplificadas e *emojis*, priorizando a brevidade sobre a complexidade intrincada (Zappavigna, 2018).

O aumento do limite de caracteres do Twitter para 280 em 2017 também foi um marco significativo que impactou diretamente a dinâmica da plataforma (Clarke, 2020). Essa alteração não apenas dobrou o espaço disponível para cada *tweet*, mas também desencadeou uma série de transformações na maneira como os usuários se engajam e compartilham informações. Com o novo limite, a natureza das postagens tornou-se mais diversificada, abrindo a possibilidade para maior contextualização.

A presença marcante da *hashtag* exerce grande influência no cenário comunicativo do Twitter. Adicionar *hashtags* implica incluir uma palavra ou frase precedida pelo símbolo # em um *tweet*, viabilizando a categorização e facilitando a busca por outros usuários. Esse uso estratégico de *hashtags* não só expande consideravelmente o alcance de um *tweet*, abarcando novas audiências, mas também desempenha um papel vital no estabelecimento e na manutenção de conexões sociais (Zappavigna, 2017).

O símbolo @ é usado para a menção explícita de um usuário por meio de um *tweet*. Trata-se de uma referência amplificada que pode ser uma ferramenta valiosa de autopromoção, uma vez que outros usuários que seguem o perfil mencionado podem visualizar a interação. Além disso, as menções podem ser consolidadas e pesquisadas por outros usuários. É possível recuperar todas as instâncias de menções para um usuário específico em um intervalo de tempo determinado usando a interface de busca do Twitter, metadados e sua Application Programming Interface (API – Interface de Programação de Aplicação). Contudo, a plataforma está constantemente ajustando a visibilidade das menções, em uma tentativa de antecipar os interesses e a capacidade de atenção de seus usuários. Essa função fomenta trocas direcionadas e conversacionais, esculpindo um espaço comunicativo único dentro do ambiente (Zappavigna, 2017).

Catalisadora para a rápida disseminação de postagens, a função de *retweet* proporciona a replicação de textos originais. *Retweet* é o ato de repostar o *tweet* de outra pessoa, frequentemente com a adição de um comentário. O *retweet* permite que usuários retransmitam ou encaminhem uma postagem através de sua rede, marcando o texto citado e efetivamente recomendando-o aos seus seguidores. Isso pode amplificar significativamente o alcance de um *tweet*, especialmente quando um usuário com um grande número de seguidores, como uma celebridade, opta pelo *retweet* de uma postagem.

Além da retransmissão, o *retweet* contribui para o ecossistema comunicativo da rede, na qual as conversas são compostas por uma interação pública de vozes que dão origem a uma sensação emocional de contexto conversacional compartilhado. Os *retweets* são marcados pela sigla *RT* na maioria dos casos, seguida pelo caractere @ para atribuir o texto ao seu autor original. *Retweets* também podem ser usados de maneira semelhante à função de resposta, bem como trazer vozes externas para um *tweet*. Além disso, o *retweet* é frequentemente utilizado para sinalizar concordância ou endosso, já que os usuários têm maior probabilidade de retweetar conteúdos que consideram especialmente relevantes e interessantes (Zappavigna, 2017).

Para além das características situacionais, os usuários do Twitter perseguem propósitos comunicativos específicos (Goulart; Biber; Reppen, 2022). A qualidade efêmera dos *tweets* exige o compartilhamento em tempo real, privilegiando a expressão imediata de pensamentos, sentimentos e notícias em detrimento de comentários reflexivos (Zappavigna, 2018). Embora propósitos como o

compartilhamento de informações sejam comuns em diversas plataformas de comunicação, a forma como os usuários os realiza linguisticamente no Twitter segue determinadas convenções. A ausência de recursos extensivos de edição, aliada à brevidade dos *tweets*, fomenta um registro linguístico informal. Essa tonalidade alinha-se mais estreitamente com a fala do que com a escrita formal, tornando gírias, palavrões e *emojicons* partes inerentes da linguagem dessa rede social.

Ultrapassando as fronteiras da Linguística e dos propósitos comunicativos, as coerções do registro também exercem um papel fundamental na configuração dos discursos na plataforma. Retomando a importância das *hashtags*, torna-se crucial compreender o seu papel essencial na congregação de comunidades, as quais se unem em torno de valores e sentimentos compartilhados, mesmo na ausência de interação direta entre os usuários (Zappavigna, 2018). Em outras palavras, as *hashtags* posicionam valores no fluxo das mídias sociais.

A ênfase significativa do Twitter nas *hashtags* impõe uma influência coercitiva, instigando os usuários a participarem de conversas virais. Essa inclinação direciona o conteúdo para tópicos e *memes* populares, marginalizando potencialmente interesses mais específicos. Além disso, o emprego generalizado de *hashtags* deu origem ao conceito de “conversa pesquisável” (Zappavigna, 2017), caracterizando o discurso que se baseia em formas de marcação social para criar afinidades com possíveis audiências.

A necessidade de tornar postagens facilmente encontráveis se configurou como uma preocupação central na esfera das plataformas digitais, refletida na prática de marcar conversas. Ao incorporar o uso de *hashtags*, os usuários podem alinhar-se a tópicos específicos, comunidades e audiências, ampliando efetivamente o alcance de suas mensagens para novos públicos, ao mesmo tempo que qualquer enunciador – até marcas e empresas – pode inserir-se em discursos que sejam convenientes para o desenho de seu propósito.

Outro aspecto de destaque na dinâmica dessa rede é sua configuração-padrão de compartilhamento público, que instiga os usuários a adotarem uma postura performática, compelindo-os a se expressar para audiências desconhecidas. Essa visibilidade intrínseca cria espaço para a natureza performática da linguagem e o potencial sensacionalismo. A construção da identidade emerge como um processo complexo de atribuição de significado, no qual indivíduos organizam suas experiências para dar forma a um sentido coerente de “eu”. Na literatura recente sobre



identidade, observamos uma mudança de ênfase dos atributos individuais para os intrincados sistemas de negociação e *performance*, que evoluem através da interação, especialmente ao incorporar nuances contemporâneas das redes sociais. Nessa perspectiva, a identidade se revela como resultado de dinâmicas estruturais e relações de poder (Papacharissi, 2011). Esse enfoque revela uma complexidade subjacente na formação da identidade online do usuário, destacando sua interconexão com as estruturas e relações de poder que moldam as interações na rede.

Ademais, o conceito de *following* não é recíproco, o que significa que um usuário não é obrigado a seguir de volta quem o segue. Esse dinamismo resulta na formação de redes de seguidores complexas, com conexões tanto unidirecionais quanto bidirecionais (Weller *et al.*, 2014). A estrutura assimétrica de seguidores e seguidos induz os usuários a adotarem uma abordagem estratégica em sua autoapresentação, projetando cuidadosamente um ethos discursivo (Fiorin, 2008), adaptado especificamente à plataforma, que representa uma imagem do autor, e não necessariamente reflete a pessoa real. Essa abordagem estratégica se reflete em *tweets* elaborados para angariar seguidores e consolidar uma posição social. Estudos anteriores abordaram a influência das redes de *following* e *followers* nas práticas comunicativas do Twitter, argumentando que a decisão de publicar um *tweet* é fundamentada na percepção que o usuário tem de sua audiência (Schmidt, 2014).

Em 2019, o então CEO do Twitter, Jack Dorsey, salientou que a ênfase no número de seguidores está intrinsecamente ligada à estrutura da plataforma. Os termos *following* e *followers* são proeminentemente destacados, incentivando a busca por um aumento no número de seguidores, independentemente da razão subjacente. Esse desejo tem sido identificado como uma influência nas práticas de promoção, na construção de imagem pessoal e até mesmo na criação de conteúdo mais provocativo e incisivo. Assim, tanto a capacidade tecnológica de seguir quanto o significado atribuído ao número de seguidores, conforme delineado pelo *design* da plataforma, exercem um impacto significativo sobre as práticas sociais e comunicativas (Clarke, 2022).

Portanto, as singularidades distintivas do Twitter não apenas influenciam, mas também impõem coerções específicas às expressões linguísticas e estratégias

discursivas<sup>12</sup>, promovendo uma fusão estilística caracterizada por brevidade, imediatismo, informalidade, performatividade e intertextualidade, entre outros elementos. Essa influência multifacetada molda significativamente a natureza da comunicação e o compartilhamento de informações em seu ecossistema. Os usuários se alinham às normas particulares da plataforma, consolidando-a, assim, como um registro distinto.

## 1.5. Análise Multidimensional

Entre os instrumentos de análise à disposição da LC, a presente pesquisa selecionou os princípios da Análise Multidimensional (doravante, AMD). Existem dois tipos principais de AMD, a Funcional e a Lexical, que serão apresentadas nas próximas seções.

### 1.5.1. Análise Multidimensional Funcional

Segundo Berber Sardinha (2004, p. 300), a AMD Funcional:

É uma abordagem para a análise de corpus que usa procedimentos estatísticos (principalmente análise fatorial), visando o mapeamento das associações entre um conjunto variado de características linguísticas dentro do corpus de estudo. Também usa procedimentos automáticos e semiautomáticos para análise do corpus, tais como etiquetagem morfossintática (part of speech tagging).

Essa técnica de análise fatorial multivariada foi desenvolvida por Douglas Biber nos anos 1980, inicialmente explorada em sua tese de doutorado (Biber, 1984) e posteriormente consolidada no livro *Variation across speech and writing* (Biber, 1988). Até então, havia uma escassez de estudos abrangentes sobre a variação nos registros linguísticos, e muitos deles estavam predominantemente focados na perspectiva gerativa transformacional de Chomsky, a qual não considerava adequadamente a linguagem natural (ou seja, a língua em uso) nem atribuía igual importância aos modos escrito e falado de forma empírica. Segundo Biber (1988), os dados para análise dentro desse paradigma excluem deliberadamente os erros que ocorrem na “fala real” e quaisquer características linguísticas que dependem de um contexto situacional de

---

<sup>12</sup> Nesse contexto, estratégias discursivas constituem-se em “modos sistemáticos de uso da linguagem, visando atingir objetivos sociais, políticos ou psicológicos” (Wodak, 2005, p. 4).

interpretação, estando muito mais próximos da escrita estereotipada do que da fala em si.

A AMD adota o conceito de registro, que, conforme apontado anteriormente, pode ser definido como “uma variedade linguística determinada por aspectos situacionais, que englobam o propósito do falante, a relação entre falante e ouvinte, e o contexto de produção” (Biber; Conrad, 2009, p. 823). Cada falante faz escolhas sistemáticas tanto na pronúncia quanto na morfologia, na seleção de palavras e na gramática, e tais escolhas estão associadas a diferentes registros, refletindo suas características situacionais. Em outras palavras, os traços linguísticos presentes nos textos variam de maneira sistemática de acordo com os contextos e propósitos comunicativos específicos nos quais são produzidos. Essa variação não é aleatória, pois está correlacionada aos contextos de uso.

Assim, a AMD permite uma investigação empírica sobre como determinadas características linguísticas coexistem em textos de um mesmo registro, além da identificação de variação entre diferentes registros em relação aos padrões de coocorrência dessas características. Esses atributos são reunidos em conjuntos conhecidos como dimensões, que podem ser definidas como agrupamentos distintos de características linguísticas que frequentemente coocorrem em textos. Elas são identificadas estatisticamente por meio de uma análise fatorial e posteriormente interpretadas em termos de funções comunicativas compartilhadas pelas características que coocorrem (Biber, 1995).

Essa abordagem envolve uma análise inicial quantitativa, conduzida por meio de programas computacionais de estatística, os quais agrupam as características coocorrentes no registro estudado em dimensões. Conforme destacado por Kauffmann (2020), a análise fatorial representa um elemento central na AMD, uma vez que é empregada para condensar o amplo e variado conjunto inicial de variáveis em um número reduzido de fatores, os quais explicam significativa parcela da variação observada. Além disso, o estudo de múltiplas variáveis em relação à frequência de associação entre elas seria impraticável de forma manual, sendo possível graças ao princípio da análise fatorial multivariada (Mindrila, 2017).

Posteriormente, conduz-se uma análise qualitativa para interpretar o significado de cada dimensão, baseada em suas funções comunicativas. Essa análise consulta múltiplos textos do corpus para orientar a interpretação. Em seguida, são

atribuídos rótulos às dimensões para refletir as principais funções comunicativas indicadas pelas variáveis coocorrentes.

As dimensões ainda são caracterizadas pela presença de “polos”, que se referem às suas duas extremidades. Esses polos são geralmente identificados no rótulo da dimensão pela palavra *versus*. Os polos são designados como “positivo” ou “negativo”, refletindo a pontuação dos textos na dimensão. Textos com pontuações mais altas em uma dimensão correspondem ao polo positivo, enquanto aqueles com pontuações mais baixas correspondem ao polo negativo. Desse modo, os termos positivo e negativo não implicam qualquer julgamento de valor, são apenas denominações técnicas usadas para descrever posições em uma escala dimensional (Delfino, 2021).

A AMD originalmente desenvolvida por Biber evoluiu para um *framework* robusto e adaptável, aplicável a várias línguas, domínios e modos semióticos. O método de extrair grupos de variáveis para expressar dimensões latentes de variação também se aplica a outras formas de AMD desenvolvidas após a obra de Biber (1988). As novas abordagens não se restringem apenas a categorias gramaticais, incluindo categorias do léxico e da semântica (Berber Sardinha; Veirano Pinto, 2019).

Essas diferentes variedades de AMD, como a Lexical e a Multimodal, adotam o mesmo conjunto básico de princípios estabelecidos na formulação original, como primazia do texto, variação sistemática, uso de conjuntos abrangentes de características e análise fatorial. Hoje, a AMD é reconhecida como uma abordagem influente e bem-sucedida para revelar padrões em larga escala de uso da linguagem por meio de modelagem estatística de variação entre textos e modos.

### 1.5.2. Análise Multidimensional Lexical

A partir da AMD apresentada por Biber (1988), Berber Sardinha propôs, em 2014, um novo modelo que não se baseia na interpretação funcional, mas sim na interpretação discursiva dos fatores: a AMD Lexical. Para tanto, o autor investigou o uso dos adjetivos *American* e *Brazilian*, bem como suas colocações – palavras que ocorrem próximas ao núcleo – para identificar os parâmetros de representação de identidade nacional e cultural através dos quais os EUA e o Brasil são retratados nas produções textuais em inglês a partir do século XIX disponíveis no Google Books (Delfino, 2021). Sob a perspectiva lexical, a AMD utilizada no estudo considerou como

variáveis (unidades de análise) apenas as palavras de conteúdo ou combinações de palavras para a identificação das dimensões de variação.

A AMD Lexical, conforme desenvolvida por Berber Sardinha (2016, 2017, 2020, 2021), é uma abordagem que visa detectar os parâmetros de variação lexical de um corpus, os quais podem sinalizar desde temas até discursos nos dados. Assim como a AMD, essa abordagem também emprega análise estatística multivariada, particularmente a análise fatorial, para detectar as variáveis latentes, isto é, aquelas que operam abaixo da percepção imediata do falante. Essas variáveis subjacentes se concretizam em dimensões de variação, que são conjuntos de itens lexicais correlacionados, ou seja, que tendem a ocorrer simultaneamente nos textos, sejam eles escritos, falados, sonoros ou visuais (Berber Sardinha; Moreira, 2023).

Na AMD tradicional – também conhecida como Funcional – as principais categorias gramaticais, estruturas de sentença e classes semânticas são empregadas para rotular, descrever e categorizar as dimensões de variação que correspondem aos parâmetros funcionais da linguagem presentes nos textos analisados. Já a AMD Lexical considera como variáveis, ou unidades de análise, palavras de conteúdo (substantivos, verbos, adjetivos e advérbios), grupos de palavras (colocações e n-gramas) e elementos extralinguísticos (*hashtags* e *emojis*) para identificar as dimensões de variação. Os *emojis* são automaticamente convertidos em rótulos descritivos, permitindo que sejam tratados como itens lexicais. O Quadro 2 exemplifica as semelhanças e distinções entre a AMD Funcional e a AMD Lexical:

**Quadro 2** – Aspectos das AMDs Funcional e Lexical

	AMD Funcional	AMD Lexical
<b>Objetivo</b>	Identificar parâmetros subjacentes de variação nos textos de um corpus.	
<b>Unidade de observação</b>	Textos ou segmentos de texto	Palavras, colocações
<b>Traços linguísticos</b>	Lexicogramaticais	Lexicais
<b>Base da interpretação</b>	Funcional, comunicativa	Campos semânticos <sup>13</sup> , discurso

Fonte: Berber Sardinha (2017).

<sup>13</sup> Toda a área de significação de uma palavra ou de um grupo de palavras.

A abordagem lexical revela nuances de representações culturais que podem escapar à análise de características gramaticais. Associações com valores, crenças e ideologias específicas podem não se manifestar com clareza em uma análise puramente estrutural da gramática. O emprego de determinadas palavras e expressões muitas vezes é moldado por eventos históricos ou tendências sociais, elementos que nem sempre se refletem nas estruturas gramaticais (Berber Sardinha, 2020). Conceitualizações abstratas como representações culturais ou ideologias são, admitidamente, mais difíceis de serem identificadas em textos devido à sua natureza oculta. A aplicação da AMD Lexical viabiliza uma maneira relativamente indutiva de identificar construtos ideológicos (Fitzsimmons-Doolan, 2014).

Embora os princípios sejam semelhantes aos da AMD Funcional, a abordagem lexical adota um critério específico que seleciona palavras lematizadas, ou seja, palavras agrupadas pelo mesmo núcleo significativo, incluindo todas as suas formas flexionadas, tanto verbais quanto nominais. Ademais, conforme apontado anteriormente, a AMD Lexical considera como variáveis apenas as palavras de conteúdo ou multipalavras para a identificação das dimensões de variação. Enquanto a AMD possui uma lista de variáveis limitada pelo número de categorias gramaticais levantadas pelos programas etiquetadores, na AMD Lexical a lista de variáveis é aberta, variando de acordo com os textos componentes do corpus estudado (Araújo; Berber Sardinha; Delfino, 2018).

Essa abordagem *bottom-up*, orientada por dados, permite que as dimensões lexicais surjam dos textos, capturando os principais parâmetros lexicais subjacentes à variação intertextual. As dimensões lexicais resultantes desse enfoque lançam luz sobre uma série de fenômenos linguísticos realizados pelo léxico (Berber Sardinha, 2020). Com base em Berber Sardinha (2016, 2017, 2020, 2021), o passo a passo para a realização da AMD Lexical nesta pesquisa envolve:

1. Identificação e contagem das palavras;
2. Normalização das frequências das variáveis lexicais;
3. Extração fatorial inicial não rotacionada baseada nas frequências normalizadas para identificar os fatores a serem utilizados;
4. *Scree plot*: definição do número de fatores para análise por meio de um gráfico de análise de sedimentação;

5. Eliminação das variáveis lexicais com comunalidades menores que 0,15 (Cf. Biber, 2006, p. 183);
6. Extração fatorial final rotacionada contendo o número de fatores estabelecidos para análise;
7. Cálculo da quantidade de variação compartilhada pelos fatores extraídos;
8. Checagem da variância dos fatores;
9. Cálculo dos escores de fator de cada texto;
10. Interpretação dos fatores em termos de seus discursos subjacentes por meio da observação dos textos, registros e variáveis.

Uma gama de estudos de LC realizou suas análises utilizando a AMD Lexical, abordando uma variedade de registros linguísticos. Muitos são os exemplos dessa diversidade, como trabalho de Mayer (2018), que investigou a variação lexical em comentários de postagens na web, provenientes de diferentes sites e redes sociais em inglês. Para tal, compilou um corpus diversificado, que abrangia 15 registros diferentes, com textos escritos por usuários com propósitos variados. A AMD Lexical revelou quatro dimensões temáticas: avaliação e conjecturas sobre pessoas; interlocução crítica; interlocução e conjecturas com foco informacional; e interlocução descritiva. Romeiro (2020) examinou a obra da fotógrafa Sally Mann através de sua produção textual e das críticas sobre seu trabalho. O corpus reuniu 12 registros diferentes, incluindo livros de fotografia escritos pela artista, textos de parede de suas exposições, artigos da imprensa geral e especializada, entre outros. Esses registros foram extraídos da biblioteca oficial de Mann e de arquivos públicos, abrangendo um período superior a 30 anos. Foram encontradas sete dimensões lexicais relacionadas aos temas subjacentes à obra fotográfica de Mann, centrados no Sul norte-americano, no fascínio pela mortalidade e na questão familiar.

Kauffmann (2020), analisou a variação linguística na prosa ficcional de Machado de Assis, com foco no estilo, visando destacar as principais dimensões estéticas do autor, tanto em termos de comunicação quanto de temática, manifestadas através da língua. Para conduzir a pesquisa, foram coletados dois corpora: o Corpus Literário de Machado de Assis, abrangendo 9 romances e 76 contos do escritor, e o Corpus Literário Congênere, uma coleção de referência composta por 92 obras de 23 escritores do período de 1850 a 1910. O estudo identificou três dimensões estéticas

distintas, nomeadas como: romantismo introspectivo formal; narrativa oralizada sentimental; e representação dramática.

Veiga (2020) realizou um estudo abrangente dos livros sagrados das principais religiões, traduzidos para o inglês, para identificar seus temas mais representativos. O corpus consistiu nos textos dogmáticos de sete religiões: budismo, espiritismo kardecista, hinduísmo, islamismo, judaísmo, mormonismo e protestantismo. Os principais livros de cada uma delas foram coletados e armazenados em formato eletrônico. O estudo identificou seis dimensões lexicais: o mundo dos espíritos e a evolução moral; fluidez, adoração e celebração à força divina; a retidão para esclarecimento espiritual versus a dádiva da terra e o poder do Senhor; crer ou sofrer as consequências versus a casa do Senhor e dos povos; devoção e respeito temente a Deus versus sacrifícios para proteção e espaços celestiais; e ritos sacrificiais de adoração.

Brogini (2023) investigou as práticas discursivas associadas à sustentabilidade, a fim de compreender como o termo é utilizado atualmente no português brasileiro. Para tal, foi compilado um corpus de 93.689 tweets postados por 42.503 usuários diferentes, no período de 2018 a 2022. Como resultado, foram identificadas oito dimensões lexicais, cada uma compreendendo diferentes discursos: cultura corporativa versus recurso escasso/insustentável; esfera de poder político versus modelo de negócio; critério de metas corporativas versus tema da educação; matriz energética limpa versus instrumento de marketing; tópico de conhecimento versus crédito tangível; desenvolvimento local versus desenvolvimento global; proteção ambiental versus oportunidade empresarial inovadora; e agronegócio versus filosofia de vida.

Whiteman (2023) realizou um estudo piloto da pesquisa aqui apresentada, sobre os discursos antivacina emergentes no Twitter durante a pandemia de Covid-19, apresentado na 12ª Conferência Internacional de Linguística de Corpus (CL2023), na Universidade de Lancaster, Reino Unido. Um corpus de cerca de 8 mil tweets foi coletado, contemplando o período de 2020 a 2022. As sete dimensões encontradas foram rotuladas como defesa das redes sociais; alertas sobre efeitos colaterais das vacinas de Covid-19; crítica ao governo, veículos de mídia oficiais e medidas de saúde pública; relatos sobre efeitos colaterais das vacinas de Covid-19; resistência à vacinação obrigatória; defesa da autonomia parental; e defesa da liberdade de escolha.



A AMD Lexical desempenha um papel central na identificação das dimensões discursivas presentes no corpus compilado para este estudo, viabilizando nossa análise dos discursos sobre vacinação no contexto da pandemia de Covid-19 no Brasil.

## 1.6. Análise do Discurso

O presente estudo é parte integrante de um projeto que sustenta a visão de que a infodemia não se configura como uma rede caótica de informação e desinformação, mas sim como um ecossistema complexo de discursos que coesamente moldaram uma perspectiva viável de abordagens frente à pandemia de Covid-19. O fenômeno da infodemia pode ser entendido como um campo de confronto discursivo em que governos, empresas e diversos atores sociais lançaram mão de discursos como instrumentos fundamentais para conceber e implementar suas estratégias diante da pandemia. Esse fenômeno emerge da interação entre sistemas de perspectivas de mundo e ideologias de diversos grupos sociais.

Nesse contexto, as seguintes definições de discurso foram selecionadas pelo projeto como as mais relevantes:

Um conjunto de ideias, conceitos e categorizações que são produzidos, reproduzidos e transformados em um conjunto específico de práticas e por meio dos quais é atribuído significado às realidades físicas e sociais. (Hajer, 1993, p. 45)<sup>14</sup>

Formas de olhar o mundo, de construir objetos e conceitos de certas maneiras, de representar a realidade. (Baker; McEnery, 2015, p. 5)<sup>15</sup>

Palavras, expressões e proposições obtêm seus significados a partir da formação discursiva à qual pertencem. (Pêcheux, 1982, p. 189)<sup>16</sup>

Práticas que sistematicamente formam os objetos dos quais falam. (Foucault, 1972, p. 49)<sup>17</sup>

---

<sup>14</sup> Original: "An ensemble of ideas, concepts, and categorizations that are produced, reproduced and transformed in a particular set of practices and through which meaning is given to physical and social realities".

<sup>15</sup> Original: "Ways of looking at the world, of constructing objects and concepts in certain ways, of representing reality".

<sup>16</sup> Original: "Words, expressions, and propositions get their meanings from the discursive formation to which they belong".

<sup>17</sup> Original: "Practices which systematically form of the objects of which they speak".

Para esta pesquisa, levamos em consideração as intrincadas estruturas das redes sociais e acrescentamos o conceito proposto por Bakhtin (2016), pelo qual o discurso transcende a mera linguagem formalizada, emergindo da complexidade das interações entre os indivíduos. Esse processo não se limita à comunicação verbal, mas se manifesta na interseção dinâmica das ações, na qual o sujeito não tem álibi. Para compreender o discurso em sua totalidade, é necessário analisar uma sequência de ações. Um ato isolado não revela as complexas interações e tensões entre diferentes elementos do discurso, que são essenciais para atribuir valor e significado a cada ação dentro de um contexto mais amplo (Magalhães; Kogawa, 2019).

O discurso, como fenômeno abstrato, reflete os valores e as ideologias historicamente atribuídas a grupos ou setores sociais. Apesar de sua natureza intangível, os discursos adquirem materialidade na língua em uso. A detecção computacional dessa materialidade possibilita a quantificação dos índices de discurso. Contudo, a análise dos padrões levantados a partir dos dados empíricos provenientes da língua (Biber, 1988; McEnery; Brezina, 2022) se vale de um entrelaçamento com uma abordagem qualitativa, pois os discursos não emergem automaticamente de seus índices. A interpretação da materialidade à luz do contexto, incorporando o conhecimento humano, torna-se, assim, essencial. Entretanto, não existe uma abordagem definitiva para a Análise do Discurso (Doravante AD), o que não significa que todos os métodos sejam viáveis. Além de aplicar conceitos, qualquer abordagem em AD requer uma reflexão mais abrangente de natureza histórica e filosófica, que vai além da simples aplicação de técnicas estanques e procedimentos seriados (Magalhães; Kogawa, 2019).

É fundamental destacar que as teorias do discurso, por si só, não têm o alcance da AMD Lexical na escala exigida pela análise de *big data*, especialmente com o nível de detalhamento e amplitude proporcionados pela abordagem, que reforça o caráter multidimensional do discurso. Portanto, neste estudo, a análise dos resultados só se concretiza através da aplicação da metodologia da AMD Lexical, com o suporte das teorias discursivas.

Esta pesquisa, assistida por corpus, se vale de ferramentas oferecidas pela abordagem metodológica da Análise Crítica do Discurso e da Análise Dialógica do Discurso para apoiar a investigação dos resultados obtidos, como delineado nas subseções a seguir.

### 1.6.1. *Análise do Discurso Assistida por Corpus*

Conforme supracitado, ao arcabouço interdisciplinar desta pesquisa integraremos a Análise do Discurso Assistida por Corpus. Os primeiros estudos utilizando corpora para AD surgiram nos anos 1990, com Caldas-Coulthard (1993, 1995) examinando a representação de gênero em notícias e Hardt-Mautner (1995) destacando a relação promissora entre a LC e a Análise Crítica do Discurso.

Ferramentas como o WordSmith Tools<sup>18</sup>, de Mike Scott, impulsionaram o desenvolvimento de mais trabalhos usando corpora, incluindo análises de identidade por Krishnamurthy (1996) e de discursos por Flowerdew (1997). Em paralelo, duas universidades tornaram-se centros de pesquisa nesse campo: Birmingham, onde os estudiosos foram influenciados pelo trabalho pioneiro realizado por Michael Stubbs, e Lancaster, onde Geoffrey Leech supervisionou a criação do British National Corpus, possibilitando trabalhos sobre variação social, como a pesquisa de McEnery, Baker e Hardie (2000) sobre palavrões e categorias demográficas. A década de 2000 viu a ampliação desse trabalho, com pesquisadores como Janet Holmes (2001) e Alan Partington (2003) explorando o uso de corpora para análise de linguagem sexista e estudos de discurso assistidos por corpus, respectivamente. O método ganhou aceitação, resultando na criação de conferências e publicações específicas sobre o tema (Baker, 2023).

A utilização de corpora na análise do discurso oferece diversas vantagens. Primeiramente, corpora fornecem evidência empírica substancial ao disponibilizar uma vasta e estruturada coleção de dados linguísticos autênticos, possibilitando aos pesquisadores uma base sólida para analisar o uso da linguagem em uma variedade de contextos. Além disso, os corpora visam representar uma gama de usos da linguagem, capturando uma variedade de características e padrões linguísticos presentes na comunicação do mundo real, o que permite análises detalhadas e comparativas entre diferentes gêneros, períodos de tempo ou grupos sociais, fornecendo informações sobre a variação e tendências no discurso.

Os corpora ainda facilitam a análise quantitativa de dados linguísticos, permitindo aos pesquisadores identificar padrões, frequências e distribuições de

---

<sup>18</sup> O WordSmith Tools é um pacote de *software* projetado para linguistas, especialmente para aqueles que trabalham no campo da LC. Consistindo em uma coleção de módulos destinados a explorar padrões na linguagem, é capaz de lidar com uma variedade de idiomas, mostrando-se uma ferramenta versátil para análise linguística.

características linguísticas dentro de um conjunto de dados. Por meio deles, é possível testar hipóteses de maneira sistemática e replicável, fornecendo uma base robusta para avaliar reivindicações teóricas na AD. Também oferecem uma visão contextual do uso da linguagem, preservando o contexto original das ocorrências linguísticas e permitindo uma análise mais profunda da linguagem em seu ambiente natural.

Os corpora diacrônicos, por sua vez, permitem aos pesquisadores rastrear mudanças na linguagem ao longo do tempo, revelando a evolução histórica do vocabulário, da gramática e de padrões de discurso. Ademais, os corpora promovem a pesquisa interdisciplinar ao fornecer uma fonte de dados comum para estudiosos de diversos campos, possibilitando colaborações frutíferas e ideias compartilhadas entre disciplinas (Baker, 2023).

Outra vantagem que se destaca ao utilizar corpora na AD é a capacidade de reduzir o viés humano a um nível gerenciável. Ao basear-se na análise sistemática dos dados linguísticos, os pesquisadores podem mitigar interpretações subjetivas e aumentar a confiabilidade e a objetividade de seus resultados. Utilizando um corpus de larga escala, é possível impor várias restrições aos nossos vieses cognitivos. Torna-se mais difícil ser seletivo em relação a um único artigo de jornal quando se está examinando centenas ou milhares de artigos, já que o esperado é que padrões e tendências gerais sejam destacados. Assim, idealmente, partimos de uma posição em que os dados em si não foram selecionados para confirmar inclinações existentes, conscientes ou subconscientes (Baker, 2023).

#### 1.6.2. *Análise Crítica do Discurso*

Para a investigação dos resultados, nos apoiaremos na Análise Crítica do Discurso (Doravante ACD), que surgiu na década de 1970 como uma abordagem para analisar discursos e textos examinando como a linguagem é usada para estruturar as relações de poder na sociedade (Fairclough, 1989; Fowler *et al.*, 1979; Kress; Hodge 1979; Van Dijk, 1985; Wodak, 1989). Trabalhos pioneiros realizados por estudiosos como Gunther Kress, Roger Fowler, Norman Fairclough, Teun van Dijk e Ruth Wodak delinearam a Critical Linguistics<sup>19</sup> e estabeleceram as bases do que viria a ser a ACD. Essa abordagem considera a linguagem como um fenômeno social (Fairclough; Wodak, 1997) e se interessa pelo modo como instituições e grupos sociais usam a

---

<sup>19</sup> Linguística Crítica.

linguagem de maneira sistemática para expressar significados e valores (Fowler, 1996). Os textos são vistos como unidades relevantes de comunicação (De Beaugrande; Dressler, 1981) e leitores/ouvintes são ativos em sua relação com os textos (Kress, 1990).

A análise das relações estruturais opacas e transparentes de domínio, discriminação, poder e controle manifestadas na linguagem (Wodak, 1996) constitui o cerne da ACD. Seu propósito primordial é a investigação crítica da desigualdade social e das dinâmicas de controle (Reisigl; Wodak, 2001), valendo-se dos conceitos-chave de poder, história e ideologia (Fairclough; Kress, 1993). A ACD é fundamentada em teorias de comunicação e linguística, bem como em teorias sociais, reconhecendo a interdisciplinaridade como um elemento essencial dessa abordagem (Chouliaraki; Fairclough, 1999). Diversas perspectivas teóricas e metodológicas são empregadas na ACD (Titscher *et al.*, 2000), algumas enfocando a análise textual detalhada e outras, os contextos mais amplos (Blommaert; Verschueren, 1999).

Há uma relação inegável entre discursos e práticas sociais, demonstrando como o discurso está profundamente enraizado em contextos institucionais e políticos mais abrangentes (Wodak; Meyer, 2001). A ACD baseada em corpus pode atuar como uma lente pela qual é possível examinar questões socioculturais mais amplas, revelando padrões discursivos recorrentes em grandes conjuntos de dados. Esses padrões discursivos oferecem evidências tangíveis de como os significados predominantes são perpetuados através do discurso (Baker; McEnery, 2015), lançando luz sobre comportamentos inconscientes que sustentam fenômenos sociais, como o preconceito racial ou de gênero (Koller; Mautner, 2004).

Uma perspectiva importante na ACD é que é muito raro um texto ser obra de uma única pessoa. Nos textos, as diferenças discursivas são negociadas; elas são governadas por diferenças de poder que, por sua vez, são em parte codificadas e determinadas pelo discurso e pelo registro. Portanto, os textos frequentemente são locais de luta, pois mostram vestígios de diferentes discursos e ideologias contendendo e competindo por predominância.

Uma característica definidora da ACD é sua preocupação com o poder como condição central na vida social, além de seus esforços para desenvolver uma teoria da linguagem que incorpore isso como uma premissa importante. Não apenas a noção de lutas pelo poder e controle, mas também a intertextualidade e a recontextualização de discursos concorrentes são observadas de perto (Wodak; Meyer, 2001).

A ACD vai além das abordagens linguísticas e psicossociais, adotando uma postura decididamente crítica nas análises dos dados. Essas análises não apenas examinam o discurso – a narrativa social em jogo na investigação –, mas também questionam o porquê de uma pessoa específica estar contando uma história específica. Ela investiga os discursos dominantes e subordinados presentes na sociedade e explora as noções de resistência e apropriação de discursos por diferentes atores sociais. Ao fazer isso, a ACD não apenas captura aspectos importantes do mundo social, mas também desempenha um papel ético e político fundamental ao mostrar como os fenômenos sociais são construídos discursivamente (Hammersley, 2003).

### *1.6.3. Análise Dialógica do Discurso*

Com o intuito de ampliar as ferramentas de investigação e análise deste estudo, optamos por também incorporar conceitos da Análise Dialógica do Discurso (Doravante ADD), que apresenta uma notável articulação com a ACD. De acordo com Wodak (1996), acadêmicos provenientes de diferentes áreas disciplinares que se dedicaram a análises linguísticas e semióticas compartilham uma perspectiva comum na qual os conceitos de ideologia e história desempenham um papel central. Ao traçar um panorama do desenvolvimento de uma tradição crítica na AD, a autora destaca a influência da Linguística de Halliday, da Sociolinguística de Bernstein e também dos trabalhos de críticos da alfabetização e filósofos sociais como Pêcheux, Foucault, Habermas, Bakhtin e Volóchinov. Wodak (1996) endossa a sugestão de outros linguistas críticos que argumentam que as relações entre linguagem e sociedade são tão intrincadas e multifacetadas que exigem uma pesquisa interdisciplinar.

O pensamento dialógico, reconhecido hoje por sua contribuição significativa às Ciências Humanas, foi moldado pelo conjunto de conceitos desenvolvidos por Mikhail Mikhailovich Bakhtin (1895-1975) e seus colegas pensadores russos. Através de círculos de reflexão sobre a linguagem, realizados especialmente entre 1920 e 1930, eles deram origem a um pensamento que destaca a natureza dialógica da linguagem e suas implicações profundas nas interações humanas e na construção do conhecimento. Entre seus principais colaboradores se destacam Valentin Nikolaevich Volóchinov (1895-1936) e Pavel Nikolaevich Medviédov (1892-1938).

Uma das contribuições mais importantes de Bakhtin para os estudos linguísticos é o conceito de gêneros do discurso. Segundo o autor, os gêneros do

discurso são formas-padrão relativamente estáveis de enunciados, moldadas pelas circunstâncias sócio-históricas. Bakhtin (2016) enfatiza que o gênero não é uma forma da língua, mas uma expressão típica do enunciado.

O dialogismo coloca a interação verbal no cerne das relações sociais, entendendo que toda expressão verbal, tanto interna quanto externa, não pode ser atribuída exclusivamente a um sujeito individual isolado (Bakhtin, 2012). Essa compreensão ressoa o propósito deste estudo, que busca identificar discursos coletivos organizados por meio de dimensões lexicais. Ademais, alinha-se à LC ao investigar a linguagem em sua integridade concreta e viva, e não como objeto específico da Linguística (Bakhtin, 2013).

No Brasil, observamos um aumento nos estudos do discurso que se fundamentam nas obras do Círculo, bem como um aumento significativo nas traduções e publicações sistemáticas das obras de Bakhtin entre os anos de 2002 e 2011. Conforme observado por Adail Sobral (2019), o termo “Análise Dialógica do Discurso” foi introduzido pela primeira vez pela pesquisadora Beth Brait, e adotado pela comunidade brasileira de estudiosos bakhtinianos como uma forma de destacar a existência de uma abordagem específica de AD de natureza dialógica, distinta de outras. A abordagem centrada na perspectiva bakhtiniana de linguagem e discurso tem se destacado nas últimas duas décadas no Brasil como um dos mais promissores campos de pesquisa em estudos discursivos.

#### *1.6.4. Enunciado e cadeia discursiva*

A AMD Lexical, central nesta investigação, aponta que a variação e as mudanças na língua frequentemente se manifestam de forma mais evidente no léxico. Em consonância, Bakhtin (2016) argumenta que a relação valorativa entre o falante e o objeto de seu discurso exerce influência direta na seleção dos recursos lexicais e composicionais do enunciado.

A linguagem, concebida como uma atividade humana intrinsecamente social, é materializada através de enunciados concretos, nos quais elementos exteriores ao campo da língua se entrelaçam com aspectos linguísticos, moldando sua singularidade. Segundo Brait (2005), metodologicamente ultrapassamos a materialidade linguística, procurando desvendar a articulação constitutiva que há entre o interno e o externo na linguagem. Para a autora, “o enfrentamento bakhtiniano

da linguagem leva em conta, portanto, as particularidades discursivas que apontam para contextos mais amplos, para um extralinguístico aí incluído” (Brait, 2005, p. 13).

Volóchinov (2017) discorre que cada elemento semântico individual dentro de um enunciado, assim como o enunciado como um todo, é interpretado em um contexto ativo e responsivo. Nesse sentido, toda compreensão é inerentemente dialógica. Assim, a dialogia se desenrola na fronteira entre o que é interno e externo ao texto, e a produção de sentido emerge da tensão nesse limiar entre a língua e o enunciado.

Cada enunciado, segundo Bakhtin (2016), é um evento único e irrepetível na interação entre diferentes vozes, uma resposta a outros enunciados que já existem ou são antecipados. Essa composição reflete a posição avaliativa do sujeito falante em relação às outras vozes presentes no contexto (Volóchinov, 2017). Essas vozes, juntamente com o ambiente social da interação, conferem a cada palavra uma densidade ideológica particular. Volóchinov (2017, p. 99) argumenta que “a palavra é neutra em relação a qualquer função ideológica específica”, ou seja, pode absorver ideologias completamente opostas. A seleção de determinadas palavras inevitavelmente implica ocupar uma posição axiológica, portanto, emoção, juízo de valor e expressão emergem exclusivamente no processo de empregar palavras dentro de um enunciado concreto. Ao negociarem valores por meio de enunciados concretos, os falantes atualizam os sentidos atribuídos a objetos, tornando-os portadores de vozes discursivas e juízos de valor (Bakhtin, 2016).

Segundo o pensamento dialógico de Bakhtin (2016, p. 99), “só o enunciado pode ser verdadeiro (ou não ser verdadeiro), correto (ou falso), belo, justo etc.”. Dessa forma, o enunciado é o elemento concreto que estabelece uma relação direta com os sujeitos falantes e a realidade. Cada enunciado, inserido em uma cadeia discursiva contínua, dialoga com seus predecessores e antecipadores, refletindo e gerando novas expressões. Essa interação, permeada pelos participantes da comunicação, situados em contextos sociais específicos, confere ao enunciado sua individualidade e relevância no âmbito discursivo.

Nessa perspectiva, os *tweets* que compõem o corpus deste estudo emergem como enunciados concretos, inseridos na vasta teia discursiva das redes sociais e enquadrados dentro dos propósitos investigativos da pesquisa. Cada *tweet*, produzido em um contexto singular, carrega consigo as marcas históricas, sociais, culturais e individuais que o tornam único. É através desses enunciados que os projetos



discursivos se manifestam, ganhando concretude e expressão, como aponta Bakhtin (2016, p. 74-75):

[...] por trás de cada texto está o sistema da linguagem. [...]. Concomitantemente, porém, cada texto (como enunciado) é algo individual, único e singular, e nisso reside todo o seu sentido [...]. Esse segundo elemento (polo) é inerente ao próprio texto, mas só se revela numa situação e na cadeia dos textos (na comunicação discursiva de dado campo).

Na investigação de qualquer discurso, como os que orbitam em torno das vacinas, é, portanto, essencial reconhecer a continuidade histórica e o desenvolvimento do fenômeno ao longo do tempo. Cada enunciado destacado constitui um elo na cadeia da comunicação discursiva. Os limites desse processo são claramente delineados pela alternância de sujeitos do discurso. Dentro desses limites, o enunciado reflete não apenas o próprio desenvolvimento do discurso, mas também os enunciados dos outros participantes e, acima de tudo, os ecos precedentes da cadeia discursiva. Esses ecos podem variar de imediatos a campos distantes da comunicação cultural (Bakhtin, 2016). Cada *tweet*, ao responder a outros e influenciado por enunciados anteriores, torna-se um elo na cadeia comunicativa, revelando as interações dialógicas que permeiam a esfera digital: “Todo enunciado concreto é um elo na cadeia de comunicação discursiva de um determinado campo” (Bakhtin, 2016, p. 57).

O enunciado possui direção específica, conclusibilidade (que assegura a possibilidade de resposta ou de compreensão responsiva) e destinatário definido. Ademais, está sempre associado a um sujeito, tendo sempre um autor. Assim, se mostra de suma importância examinar os *tweets* por meio de uma perspectiva que busca identificar e interpretar as pistas deixadas por enunciadoreis. Embora não possamos acessar diretamente o sujeito empírico, é possível inferi-lo a partir das marcas enunciativas presentes no texto, que revelam indícios sobre as formas de produção e o autor-criador por trás dele. Nas redes sociais, mesmo quando estão disponíveis informações como o nome de usuário e, em alguns casos, um ethos (Fiorin, 2008) declarado no perfil, frequentemente nos deparamos com um sujeito virtual, protegido pelo anonimato das plataformas ou pela atividade de *bots*. No entanto, é por meio da análise da materialidade textual que conseguimos decifrar o projeto discursivo e identificar os vestígios deixados por diversos sujeitos.

Segundo Bakhtin (2016), dois enunciados independentes, sem conhecimento mútuo, quando desejam estabelecer conexões com um mesmo pensamento, entram inevitavelmente em interações dialógicas. A AMD Lexical identifica e operacionaliza empiricamente essas complexas interações, agrupando os enunciados que compartilham o mesmo discurso em dimensões. Além disso, as dimensões são detentoras de uma natureza essencialmente dialógica. Cada texto do corpus incorpora simultaneamente todas as dimensões. Os textos são atribuídos a uma determinada dimensão com base em um escore de aderência ao discurso identificado, mas todas as outras dimensões permanecem presentes em cada um deles, em menor intensidade, evidenciando as relações dialógicas entre elas, como descrito por Berber Sardinha (2021, p. 298-299):

O objetivo principal de uma AMD Lexical é descrever a variação do discurso; isto é, como o discurso varia sistematicamente de acordo com o contexto. A variação é modelada a partir de uma perspectiva multidimensional, o que significa que cada texto é visto como sendo moldado simultaneamente pela incidência dos diversos discursos representados pelas várias dimensões.<sup>20</sup>

Por meio da abordagem inovadora na análise de *big data* proporcionada pela AMD Lexical, ao agruparmos milhares de enunciados com um projeto discursivo compartilhado, torna-se possível realizar uma análise sofisticada dos discursos predominantes no corpus, revelando como estão entrelaçados com uma variedade de outras vozes discursivas.

### 1.7. Taxonomia dos argumentos antivacina

Fasce, Schmid, Holford, Bates, Gurevych e Lewandowsky desenvolveram em 2023 uma taxonomia de argumentos contrários à vacinação, um quadro hierárquico que categoriza temas comuns. Fundamentado em uma revisão sistemática de 152 artigos acadêmicos, o trabalho busca proporcionar uma compreensão abrangente das diversas razões pelas quais os indivíduos expressam oposição às vacinas. A taxonomia é estruturada em torno de 11 “raízes de atitudes”, que representam atributos subjacentes que influenciam sentimentos antivacina e apontam para

---

<sup>20</sup> Original: “The primary goal of a lexical MD Analysis of discourse is to describe discourse variation; that is, how discourse varies systematically according to context. The variation is modeled from a multidimensional perspective, meaning that each text is seen as being shaped simultaneously by the incidence of the various discourses represented by the various dimensions”.

motivações e preocupações que contribuem para a hesitação em relação à imunização. A análise desses argumentos, listados abaixo, fornece uma base sólida para examinarmos com maior profundidade os discursos identificados nos resultados deste estudo, como veremos nos próximos capítulos.

1. **Ideação conspiratória:** caracteriza-se pela adesão a teorias da conspiração, mesmo quando outras explicações mais plausíveis estão disponíveis (Aaronovitch, 2010). Os mecanismos sociais, o contexto sociopolítico e os aspectos psicológicos que alimentam essas crenças foram extensivamente estudados, solidificando a ideia conspiratória como um fenômeno psicológico amplamente reconhecido (Goreis; Voracek, 2019). Estudos prévios evidenciaram a associação entre ideação conspiratória e resistência à vacinação, destacando sua relevância em contextos de saúde pública.
2. **Desconfiança:** reflete uma falta de confiança generalizada em autoridades e comunicadores envolvidos no processo de vacinação, abrangendo desde empresas farmacêuticas e cientistas até protocolos médicos e métodos de pesquisa científica. Também abarca a percepção de que políticos e profissionais de saúde trabalham por interesses próprios ou não possuem conhecimento suficiente. Embora haja uma sobreposição com a ideação conspiratória, as razões por trás dessa desconfiança podem ser variadas e incluir experiências vividas, como discriminação contra certos grupos (Lewandowsky *et al.*, 2022), sem necessariamente implicar uma cadeia causal complexa de eventos secretos. Pesquisas anteriores têm consistentemente demonstrado que essa desconfiança é um dos principais preditores da hesitação em relação à vacinação (Jolley; Douglas, 2014; Tram *et al.*, 2022).
3. **Crenças sem embasamento científico:** abrangem uma série de crenças que carecem de respaldo científico, distorcem fatos ou se baseiam em concepções pseudocientíficas. A distinção central entre essa atitude e a ideação conspiratória reside no papel predominante das variáveis cognitivas. Enquanto a ideação é influenciada principalmente pela percepção de ameaça e variáveis emocionais, as crenças são moldadas especialmente por fatores cognitivos (Pierre, 2020; Van Prooijen, 2020).

4. Visão de mundo e política: deriva da visão particular de um indivíduo sobre como a sociedade deve ser organizada, incluindo preditores de oposição à vacinação, como populismo (Kennedy, 2019), nacionalismo (Lewandowsky; Oberauer, 2021), conservadorismo (Motta, 2021) e visões de mundo individualistas/hierárquicas (Hornsey; Harris; Fielding, 2018). Variáveis adicionais incluem opiniões sobre questões relacionadas ao contexto político específico de cada país, como animosidade em relação a um governo, partido político ou figura política específica.
5. Preocupações religiosas: abrangem uma gama de crenças e normas, ou outras manifestações de uma noção ampla de espiritualidade, que foram mostradas como motivadoras de atitudes hesitantes em relação às vacinas (Grabenstein, 2013). Essas preocupações podem ser divididas em quatro grupos: o primeiro trata das violações de normas dietéticas, como componentes sanguíneos e excipientes farmacêuticos de origem suína ou bovina; o segundo explicita violações de códigos religiosos de pureza em relação a linhagens celulares de origem fetal e comportamento sexual (por exemplo, a vacina contra o HPV, que protege contra uma doença sexualmente transmissível); o terceiro trata da defesa da ordem natural, o que muitas vezes se reflete na rejeição da interferência na “providência divina”; o quarto menciona alternativas religiosas à vacinação, como a fé ou a oração para combater doenças. Esta raiz apenas considera argumentos contrários baseados em objeções teológicas, deixando de lado argumentos sociopolíticos. Embora a religiosidade não esteja consistentemente associada a uma maior hesitação vacinal (Eriksson; Vartanova, 2022), as taxas de vacinação são particularmente baixas entre algumas comunidades religiosas (Barskey *et al.*, 2012; Hanratty *et al.*, 2000).
6. Preocupações morais: podem surgir quando as vacinações são percebidas como promotoras de comportamento imoral e/ou desenvolvidas usando meios imorais. Distinguem-se das preocupações religiosas, uma vez que, embora a moralidade possa derivar de crenças religiosas, não é necessário ser religioso para manter uma posição moral específica – por exemplo, muitas pessoas se opõem ao aborto por motivos morais, sem recorrer a crenças religiosas. As preocupações morais sobre vacinas frequentemente

se manifestam como preocupações sobre um ambiente permissivo para jovens mulheres se envolverem em atividade sexual (especialmente em relação à vacina contra o HPV), o uso de linhagens celulares fetais, maus-tratos a animais durante o desenvolvimento e produção de vacinas e resistência à ideia de que a vacinação universal sacrifica alguns para beneficiar muitos, ou seja, o antiutilitarismo (Amin *et al.*, 2017; Rossen *et al.*, 2019).

7. Medo e fobias: capturam os diferentes medos que levam os indivíduos a rejeitar vacinas, adotando uma posição na qual não precisem confrontar sua falta de controle ou a autopercepção negativa associada a ter uma fobia (Hornsey; Fielding, 2017). Os medos relacionados à vacinação costumam ser muito desproporcionais aos perigos existentes e envolvem tanto o medo de efeitos colaterais (Karlsson *et al.*, 2021) quanto a tripanofobia<sup>21</sup> (Hornsey; Harris; Fielding, 2018).
8. Percepção distorcida de risco: atitude presente na maioria das teorias gerais de comportamento em saúde e em modelos específicos de hesitação vacinal. Nesse contexto, a percepção distorcida do risco decorre da falta de medo ou consciência da ameaça representada pela doença, seja para si mesmo ou para os outros. Os indivíduos percebem que a doença tem um risco baixo e isso motiva a crença de que a vacinação é desnecessária ou as desvantagens superam os benefícios (Caserotti *et al.*, 2021; González-Block, 2020).
9. Individualismo: provém da priorização das necessidades próprias em relação às dos outros. Está intimamente relacionado a uma visão de mundo individualista e competitiva, na qual os indivíduos devem se preocupar apenas com eles mesmos e sua família, exibindo, assim, baixa propensão à responsabilidade coletiva – um fator motivador para a imunização (Quadri-Sheriff *et al.*, 2012).
10. Relativismo epistêmico: consiste na visão de que o conceito de “verdade” e seus padrões associados de raciocínio (por exemplo, pensamento crítico, métodos científicos e tomada de decisão baseada em evidências) são

---

<sup>21</sup> Medo exagerado de procedimentos médicos que envolvam agulhas e/ou injeções.

produtos de convenções e estruturas de avaliação (como contextos históricos, normas sociais e culturais e padrões individuais), portanto não pode haver um ponto de vista independente de estrutura a partir do qual se pode obter conhecimento objetivo (Boghossian, 2007; Kusch, 2021).

11. Reatividade: uma construção psicológica de longa data que tem sido consistentemente associada à hesitação vacinal (Finkelstein *et al.*, 2020; Soveri *et al.*, 2020; Sprengholz *et al.*, 2022). É definida como a tendência de um indivíduo em defender sua autonomia quando percebe que outros estão tentando impor sua vontade sobre eles – por exemplo, através de mandatos de vacinação. Dessa forma, esta raiz de atitude inclui uma defesa das liberdades civis, na qual o indivíduo é motivado a reivindicar seu direito de agir contrariamente às normas sociais e injunções políticas. Isso está relacionado a um alto senso de autonomia pessoal e empoderamento, pelo qual as pessoas reagem ao conselho de saúde como uma violação de sua capacidade de escolher uma ação para si mesmas. A reatividade se manifesta em argumentos que proclamam que a decisão de vacinar ou não deve ser completamente livre e autônoma.

## 2. Metodologia

Neste capítulo, descrevemos os procedimentos adotados ao longo da pesquisa, desde a coleta dos dados linguísticos que compõem o CBDIAT (Corpus Brasileiro de Discursos Infodêmicos Antivacina no Twitter) até a realização da análise fatorial final. A análise foi realizada com base em lemas e palavras de conteúdo, que englobam as palavras mais recorrentes do corpus relacionadas a termos específicos em torno da vacina de Covid-19 no Twitter. Essa base de dados possibilitou uma análise qualitativa posterior dos resultados, bem como sua interpretação embasada, conforme abordado anteriormente, seguindo a metodologia da AMD Lexical (Berber Sardinha, 2016, 2017, 2020, 2021).

### 2.1. Coleta e pré-processamento do corpus

O CBDIAT é formado por postagens em português coletadas na rede social Twitter entre 2020 e 2022. Para realizar a coleta, o primeiro passo envolveu estabelecer uma busca relevante na plataforma dentro do escopo do estudo. O rastreio foi feito através das palavras-chave como *vacina*, *pfizer*, *astra*, *astrazeneca*, *coronavac*, *janssen*, *jansen*, *covaxin*, *antivax*, *anti-vax*, *anti-vaxx*, *antivaxx*, *antivacina*, *sputnik*, *dose* e *vachina*, entre outras. Para isso, foi utilizado o Twarc, uma ferramenta de linha de comando e uma biblioteca em Python para coletar e arquivar dados JSON do Twitter por meio da API da plataforma.

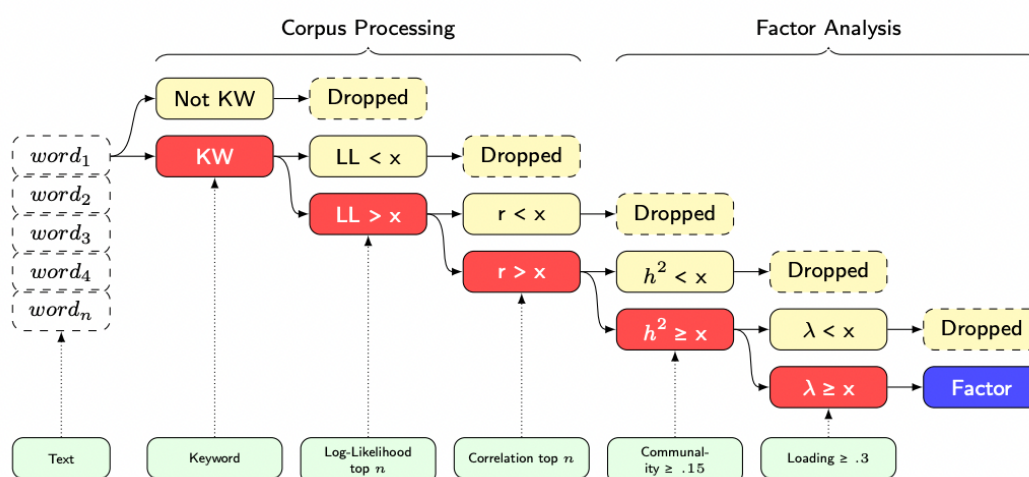
### 2.2. Visão geral do processamento do corpus

Feita a coleta do corpus, o primeiro passo é extrair, de uma lista selecionada, as palavras candidatas à palavra-chave, conforme demonstrado na Figura 3. Esse processo se dá selecionando as palavras de conteúdo, ou seja, no caso da AMD Lexical, substantivos, adjetivos, verbos e advérbios. Essa nova lista de palavras vai compor o conjunto de palavras candidatas à palavra-chave.

Para sabermos se a palavra é chave ou não, deve-se aplicar o método estatístico Log-Likelihood, pelo qual a frequência observada no corpus de estudo deve ser maior do que a esperada no corpus de referência. O Log-Likelihood seleciona as palavras com maior valor de chavicidade (*keyness*). Na etapa seguinte, de correlação, são escolhidas as palavras com alto valor de chavicidade que apresentam maiores valores de correlação. Posteriormente é feita uma filtragem pelo valor de

comunalidade, que é a proporção da variância de uma variável qualquer que pode ser explicada pelos fatores. A comunalidade atribui um peso fatorial às palavras que sobrevivem a esse criterioso processo de filtragem, de modo que tudo abaixo de 0.3 é removido. A aplicação do Princípio da Parcimônia<sup>22</sup> assegura que apenas o mínimo necessário, sem perda de precisão, seja incluído na análise.

**Figura 3** – Visão geral do processamento do corpus



Fonte: Berber Sardinha e Moreira (2023).

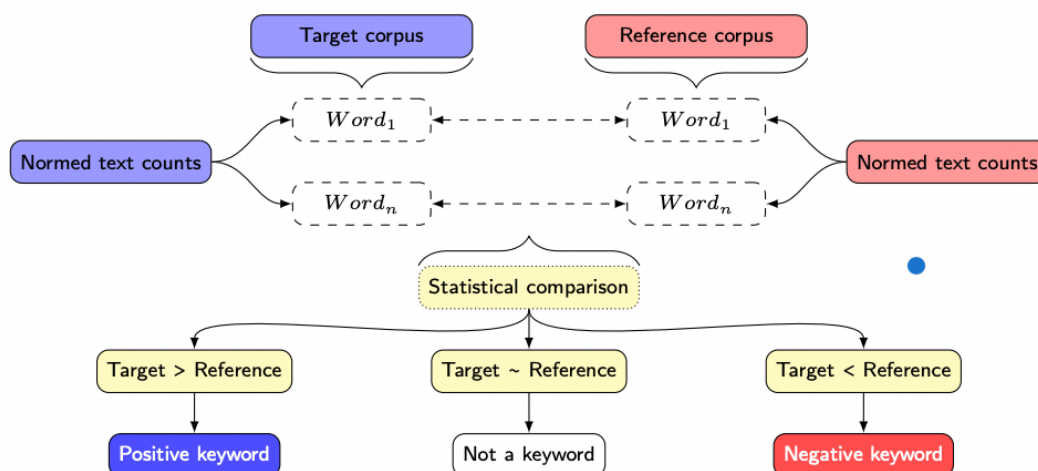
### 2.2.1. Filtragem de variáveis por palavras-chave

As palavras-chave são aquelas que se destacam em termos de frequência em relação a um corpus de referência, indicando, assim, conteúdo relevante – no caso da AMD Lexical, substantivos, adjetivos, verbos e advérbios. Essa distinção estatística entre o corpus em estudo e o de referência, como sugerido por Scott (2000), não apenas revela os tópicos mais proeminentes, mas também destaca diferenças significativas entre os textos ou corpora.

Utilizando um *script* computacional em Shell, identificamos as palavras-chave do corpus e selecionamos aquelas que apresentaram maior chavicidade em comparação com o corpus de referência. Para garantir a precisão dessa seleção, empregamos o teste estatístico Log-Likelihood, que fornece estimativas mais precisas de relevância das palavras. Esse processo de seleção está ilustrado na Figura 4:

<sup>22</sup> A parcimônia é um princípio fundamental na redução da dimensionalidade. Embora a análise comece com uma ampla e diversificada gama de características linguísticas, o objetivo final é identificar o menor número possível de dimensões que possam explicar eficazmente a variação no uso da linguagem.



**Figura 4** – Filtragem de variáveis por palavras-chave

Fonte: Berber Sardinha e Moreira (2023).

### 2.2.2. Filtragem de variáveis por correlação

Após a seleção das variáveis por palavras-chave, o próximo passo consistiu na criação de uma matriz de correlação, a fim de identificar as variáveis com as maiores medidas de correlação. Esse procedimento teve como objetivo principal eliminar correlações estatisticamente baixas, visando otimizar a análise fatorial.

Correlações são estabelecidas com base na frequência em que duas palavras ocorrem juntas em textos, comparada à frequência em que não aparecem juntas. Esse processo gera um número, que representa a correlação, incluindo a sua direção, seja positiva ou negativa. Por exemplo: se a presença da palavra *tratamento* está associada à presença do termo *precoce* em diversos textos, teremos uma correlação positiva; em contrapartida, se a presença de *tratamento* significa a ausência de *precoce* e vice-versa, teremos uma correlação negativa.

É importante notar que, sejam positivas ou negativas, as correlações em foco estão entre as mais proeminentes. Investigamos tanto as associações quanto as dissociações entre palavras, buscando compreender as relações que se formam e aquelas que se afastam. Encerrada essa etapa, tem-se um corpus preparado para análise.

### 2.3. Composição do corpus

A composição final do corpus está disposta no Quadro 3, com um total de 16.841 postagens, totalizando 167.197 de palavras. Nos termos de classificação de corpora, nosso corpus de pesquisa é escrito em português do Brasil (monolíngue) por cidadãos brasileiros, supõe-se, e coletado na rede social Twitter. As postagens selecionadas de março de 2020 a novembro de 2022 foram salvas no formato texto (.txt), em codificação UTF-8.

**Quadro 3** – Composição do CBDIAT

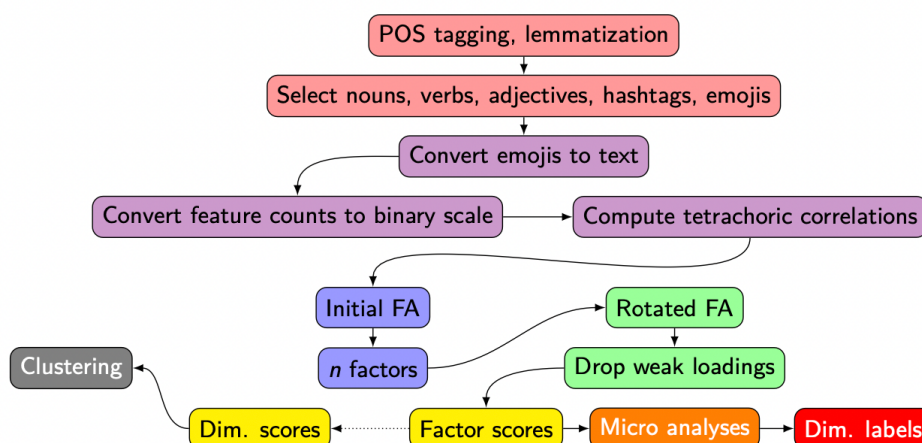
	Número total
<b>Postagens</b>	16.841
<b>Palavras (<i>tokens</i>)</b>	167.197

Fonte: Elaborado pela autora.

### 2.4. Análise Multidimensional Lexical

A metodologia da AMD Lexical, conforme descrito na Figura 5, compreende três etapas fundamentais: 1. O pré-processamento do corpus, abarcando etiquetagem, normalização e seleção das variáveis (traços linguísticos) a serem investigadas; 2. Uma análise fatorial inicial (não rotacionada), na qual o analista determina o número de fatores a serem extraídos por meio da análise do gráfico de escarpa; 3. Uma segunda análise fatorial (rotacionada), que implica extração do número definitivo de fatores, eliminação de variáveis com baixa carga nos fatores, cálculo de escores e, por fim, nomeação das dimensões por meio de microanálises.

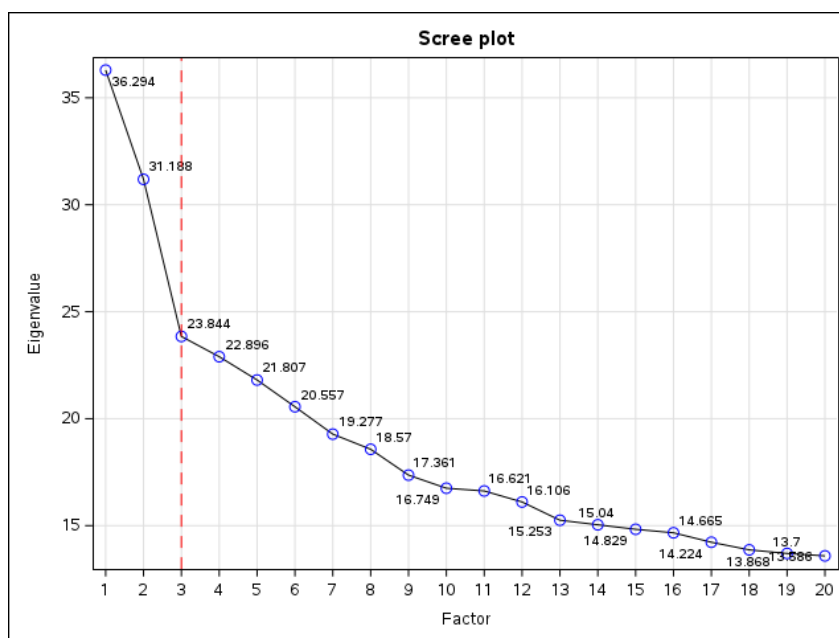
**Figura 5** – Procedimentos da Análise Multidimensional Lexical



Fonte: Berber Sardinha e Moreira (2023).

O processo de etiquetagem utilizou o TreeTagger para a língua portuguesa (Schmid, 2013). Em seguida, foram realizadas verificações manuais no corpus, preservando substantivos, verbos, adjetivos, *emojis* e *hashtags* presentes nas postagens, através de um *script* desenvolvido pelo professor orientador. Os lemas relacionados às categorias gramaticais – substantivos (incluindo os nomes próprios encontrados), verbos e adjetivos –, juntamente com os *emojis* e as *hashtags* presentes no corpus, foram quantificados e sistematizados em planilhas em formato *csv* (*comma-separated values*). A extração fatorial não rotacionada gerou o *scree plot*, ou gráfico de sedimentação, apresentado no Gráfico 1:

**Gráfico 1 – Scree plot da extração fatorial**



Fonte: Elaborado pela autora.

Após uma análise interpretativa baseada nos valores dos *eigenvalues*<sup>23</sup>, que indicam a contribuição para a variação de cada fator, a pesquisa considerou diferentes soluções para a extração final da AMD Lexical, incluindo dois, três e quatro fatores. Entre essas possíveis soluções, a extração de três fatores mostrou-se a mais interpretável, conforme indicado pela seta no Gráfico 1, e, portanto, foi selecionada como a extração definitiva para este estudo. A extração fatorial final foi realizada com o método Promax, que assume a existência de correlações entre os fatores. O

<sup>23</sup> *Autovalores*, em português.

resultado da rotação produziu seis fatores, os quais serão discutidos no capítulo seguinte.

Para avaliar como os textos se relacionam com os fatores em estudo, realiza-se uma Análise de Variância (ANOVA) para identificar os textos que têm maior influência dentro desses fatores. Até esse ponto, os procedimentos descritos são de natureza quantitativa. A partir desses resultados, o pesquisador examinará exemplos de como as variáveis (ou seja, os itens lexicais desta pesquisa) são utilizadas nos textos, especialmente naqueles que demonstram uma representatividade média ou alta conforme avaliado pela ANOVA. Essa etapa da pesquisa assume um caráter qualitativo, enfocando a análise dos padrões de uso das variáveis nas linhas de concordância. Por fim, com base nessa interpretação, foram identificadas as dimensões predominantes presentes no corpus.

Após a identificação das dimensões, atribuímos rótulos a elas a fim de refletir os principais discursos indicados pelas variáveis coocorrentes. Nas próximas seções, detalharemos o processo de rotulação das dimensões e a interpretação dos discursos revelados por meio dessa análise.

### 3. Apresentação e discussão dos resultados

Nas próximas seções, apresentaremos e discutiremos os resultados obtidos a partir do método descrito no capítulo anterior.

#### 3.1. Apresentação dos resultados

Utilizando dados lexicais como variáveis, buscamos identificar a variação discursiva por meio da AMD Lexical (Berber Sardinha, 2016, 2017, 2020, 2021), analisando os arranjos dimensionais do corpus CDBIAT.

Para analisar os resultados obtidos pela AMD Lexical, é necessário examinar os enunciados na forma de *tweets*. Conforme postula a ADD, a palavra é neutra em relação a qualquer ideologia específica. Segundo Volóchinov (2017, p. 99), “ela pode assumir qualquer função ideológica: científica, estética, moral ou religiosa”. O aspecto expressivo das palavras emerge no contexto de seu uso em um enunciado concreto, já que a escolha de determinadas palavras implica assumir um posicionamento valorativo no mundo. Portanto, isoladamente, as variáveis carecem de expressividade: é dentro de um enunciado que elas adquirem sentido e orientação avaliativa. Essa abordagem é relevante, pois nosso interesse não está apenas no significado individual de cada palavra nos *tweets*, mas também em compreender para onde essas variáveis apontam, o que está subjacente ao texto, ou seja, o discurso.

Por meio da análise desses discursos, pretendemos responder às questões centrais do presente estudo: 1. Quais são as dimensões lexicais dos discursos em torno do movimento antivacina? 2. Quais os principais discursos associados às dimensões lexicais?

Cada segmento delinea uma resposta parcial a tais indagações, haja vista que se restringe, em seus respectivos subconjuntos, a elencar as variáveis-chave da dimensão discursiva abordada, assim como a interpretação discursiva do fator, na tessitura dimensional. O Quadro 4 apresenta as três dimensões encontradas e suas respectivas rotulações:

**Quadro 4 – Rótulos das dimensões discursivas**

<b>Dimensão</b>	<b>Rótulo curto</b>	<b>Rótulo longo</b>
1	Alerta sobre riscos da dose de reforço em jovens <i>versus</i> Resistência a medidas compulsórias de combate à Covid-19	Alerta sobre riscos cardíacos, males súbitos e falecimentos após doses de reforço, especialmente entre jovens e crianças, considerando a vacina como experimental <i>versus</i> Questionamento e resistência à imposição de medidas e decretos compulsórios de combate ao avanço da Covid-19
2	Teorias político-conspiratórias sobre a vacina <i>versus</i> Evidências pseudocientíficas contra a vacina	Alegações sobre a origem do vírus em laboratório chinês e interesses de políticos e governos ligados a grandes grupos farmacêuticos <i>versus</i> Evidências de estudos pseudocientíficos publicados sobre os efeitos adversos das vacinas
3	Defesa do tratamento precoce em detrimento da vacina <i>versus</i> Relatos pessoais sobre efeitos colaterais da vacina	Defesa do tratamento precoce frente à letalidade do vírus e à falta de estudos sobre a eficácia e a segurança em longo prazo da vacina emergencial <i>versus</i> Relatos sobre amigos e terceiros que sofreram com efeitos colaterais graves – dores de cabeça, paralisia, problemas no rim, nas pernas e nos pés – e até fatais após a vacinação

Fonte: Elaborado pela autora.

Nesse ponto, mostra-se essencial reiterar que as dimensões discursivas, em sua essência, abarcam uma miríade de significados interligados. Portanto, os rótulos das dimensões devem ser elaborados de forma extensa e intrincada para abarcar essa complexidade. Todavia, para referenciá-las no texto, optamos por uma forma descritiva mais sucinta de cada dimensão, o rótulo curto – como os empregados nos títulos que introduzem as seções subsequentes. Contudo, os rótulos completos das dimensões (mencionados no quadro anterior) são os que capturam o sentido interpretado dos discursos de maneira mais abrangente. De qualquer forma, ambos os rótulos (longo e curto) refletem os mesmos discursos, sendo apenas formulações diferentes em termos de extensão, não de conteúdo.

Por fim, é importante explicitar que todos os *tweets* citados no decorrer deste capítulo estão entre os textos que pontuaram os maiores escores em cada dimensão – tanto do polo positivo quanto do negativo –, ou seja, são os enunciados que carregam grande carga das variáveis lexicais de sua dimensão.

### 3.1.1. Dimensão 1 – alerta sobre riscos da dose de reforço em jovens versus resistência a medidas compulsórias de combate à Covid-19

O Quadro 5 apresenta as variáveis lexicais que compuseram o primeiro fator da extração fatorial em dois polos distintos: alerta sobre riscos cardíacos, males súbitos e falecimentos após doses de reforço, especialmente entre jovens e crianças, considerando a vacina como experimental *versus* questionamento e resistência à imposição de medidas e decretos compulsórios de combate ao avanço da Covid-19.

**Quadro 5 – Variáveis da Dimensão 1**

<b>Polo</b>	<b>Variáveis lexicais</b>
Positivo (65 variáveis)	reforço (0,95635), fonte (0,94996), infarto (0,93656), semana (0,86775), indicar (0,86442), cardíaco (0,86312), falecer (0,83739), publicar (0,83478), pesquisa (0,82977), experimento (0,79311), sofrer (0,76066), jovem (0,74540), pós (0,72900), facebook (0,71919), dia (0,66741), dose (0,65392), súbito (0,65113), alertar (0,63157), criança (0,62101), alerta (0,60826), mês (0,60387), morte (0,58873), fulminante (0,57859), org (0,57492), mal (0,55586), doença (0,55347), morrer (0,54771), estudo (0,54304), saudável (0,52843), miocardite (0,52303), grave (0,49075), global (0,48340), bvsalud (0,47183), risco (0,45471), causar (0,45057), literat (0,44981), óbito (0,44769), primeiro (0,44696), último (0,42515), autoimune (0,42477), ano (0,42149), mrna (0,41604), inocular (0,40904), evento (0,39380), internar (0,38734), articles (0,38307), syringe_e (0,37561), contrair (0,37057), caso (0,36688), dar (0,35343), novo (0,34916), comprovante (0,34781), adverso (0,34257), causa (0,34164), mulher (0,34140), imunidade (0,33854), vida (0,33504), sistema (0,33086), amigo (0,32808), pessoa (0,32791), exame (0,32190), vacinal (0,31053), maioria (0,30991), existir (0,30724), hepatite (0,30286)
Negativo (52 variáveis)	enquete (-0,64958), proposta (-0,60021), imposição (-0,54932), audiência (-0,48156), consciência (-0,47040), termo (-0,46066), gel (-0,44092), Covid19 (-0,44058), zinco (-0,43595), serviço (-0,42600), mercado (-0,42056), tutelar (-0,41945), correiopaulista (-0,41739), dama (-0,41707), piada (-0,41382), álcool (-0,41347), pleno (-0,40453), questionamento (-0,39936), ferro (-0,39698), decreto (-0,39473), combate (-0,39455), restaurante (-0,38842), retirar (-0,38422), lista (-0,38409), conflito (-0,38236), calça (-0,37299), CPI (-0,37061), farmacêutica (-0,36656), distanciamento (-0,36022), comprimido (-0,35825), rebanho (-0,35384), desserviço (-0,35063), jornaldacidadeonline (-0,34456), liberdade (-0,34310), brasil (-0,34096), vice (-0,33947), RNA (-0,33676), senador (-0,33456), conselho (-0,32672), compulsório (-0,32620), suposto (-0,32219), avanço (-0,32124), compra (-0,31953), senso (-0,31852), comprovação (-0,31752), respeitar (-0,31745), confiar (-0,31494), apertar (-0,31342), world (-0,30772), adotar (-0,30588), expressar (-0,30095), vachinas (-0,30063)

Fonte: Elaborado pela autora.

O polo positivo da Dimensão 1 enfoca a disseminação de mensagens que alertam sobre potenciais riscos cardíacos, males súbitos e falecimentos supostamente associados às vacinas contra a Covid-19, com atenção especial aos eventos ocorridos após as doses de reforço. Esses *tweets* enfatizam que tais riscos seriam particularmente preocupantes entre os jovens, apontados como parte significativa das vítimas, conforme os Exemplos 1 e 2:

### Exemplo 1

**Crianças e Adolescentes sofrendo infartos Fulminantes** pouco tempo depois dos **reforços** . . . As **pessoas** não querem admitir a verdade. Poucos **jovens faleceram** da moléstia, muitos **jovens falecendo de infartos fulminantes** após as **doses**. Despertem! ISSO NUNCA FOI NORMAL! <https://t.co/e68plYhO98>

### Exemplo 2

**Mortes nas últimas 48Hs - Pessoas jovens sofrendo infartos.** Ninguém questiona, a despeito da Anvisa ter **alertado** sobre o **risco** de problemas **Cardíacos** e **Pesquisas indicarem o risco d infartos** e AVCs, depois d **meses** após as **doses** . gov. br / anvisa / pt-br / a ... ncbi. nlm. nih. gov / pmc / articles / P ... <https://t.co/W2ANw2mmjO>

Essa narrativa antivacina se alinha com um dos principais temas identificados por Skafle *et al.* (2022) em sua revisão sobre a propagação de desinformação relacionada às vacinas contra a Covid-19 nas mídias sociais: a disseminação de desinformação médica. Conforme observado pelos autores, a apreensão quanto aos supostos efeitos colaterais é uma preocupação central que exerce impacto adverso na disposição e na adesão às vacinas.

De fato, os exemplos adicionais (de 3 a 5) ilustram como esse medo em relação aos potenciais danos é amplificado e transformado em desinformação médica, com alegações de que as vacinas estariam causando infartos, AVCs e mortes súbitas, especialmente entre os jovens. Muitas dessas postagens citam estudos que supostamente corroborariam tais afirmações, descredibilizando as vacinas como um “experimento”:



### Exemplo 3

Vacinado c / 4 **doses**, Eric Gustavo, 32, **contraiu** CV! D várias vezes. Estava de alta hospitalar, qd **sofreu** uma parada **cardíaca** enquanto dormia. **Pesquisas indicam** que a eficácia do **experimento** é negativa após **reforços**. **Fonte:** m. **facebook**. com / story. php? stor ... virologyj. biomedcentral. com / **articles** / 10. 11 ... <https://t.co/LNmjcR8jbF>

### Exemplo 4

Mais **jovens falecendo** por infartos e **males Súbitos**. É importante lembrar que a **pesquisa** científica da @CMAJ, **publicada** nesta **semana**, **indicou** q o **risco** d problemas **cardíacos** é d 100 p / cada 100mil **doses**. É duro saber que estes **morreram** p / q alguns se sentissem seguros. . . <https://t.co/xFSmndVjei>

### Exemplo 5

Lidiane Brugnari era professora em Jacarezinho / PR. Tomou a 2a. **Dose** em agosto / 21 e o **reforço** há pouco mais de 3 **meses**. **Faleceu dia** 26 / 07 / 22 em virtude de um AVC. A OMS ja **alertou** sobre AVC pós **experimento**. **pesquisa**. **bvsalud.org** / global-literat ... m. **facebook**.com / story. php? stor ... m. **facebook**.com

Essa propagação de temores relacionados à vacinação se conecta à taxonomia dos argumentos antivacina apresentada anteriormente, na qual o medo é uma das 11 raízes identificadas. O medo é uma emoção complexa e fundamental, profundamente enraizada na condição humana como um mecanismo de defesa contra ameaças percebidas (Delumeau, 2009). No contexto da vacinação contra a Covid-19, o discurso do medo tende a amplificar de forma desproporcional os riscos envolvidos, provocando preocupação excessiva em relação aos potenciais efeitos adversos.

Já o polo negativo da Dimensão 1 se concentra na resistência à imposição da vacinação obrigatória contra a Covid-19. Os *tweets* refletem uma preocupação com a perda de autonomia e liberdade individual, bem como uma desconfiança em relação às intenções do governo e das autoridades de saúde. Ao rejeitarem a ideia da vacinação obrigatória, os proponentes dessa narrativa buscam proteger sua liberdade individual e resistir à interferência do governo em suas vidas pessoais, como ilustram os exemplos de 6 a 9:

**Exemplo 6**

CONTRA A OBRIGATORIEDADE DA VACINA CHINESA: **ENQUETE** DO PL 4506 / 2020 - BIA KICIS **Proposta retira vacinação compulsória** da lista de medidas de **combate** à Covid-19 <https://t.co/q4oENg6WzK>

**Exemplo 7**

Ao ver alguém ameaçando denunciar pais ao **Conselho Tutelar** por não vacinar seus filhos com uma vacina sem **comprovação** de segurança e eficácia, é impossível não lembrar dos alemães denunciando judeus para o exército nazista! A humanidade não evoluiu muita coisa desde a 2ª guerra!

**Exemplo 8**

A #Pfizer mentiu para todos. CEO da **Farmacêutica** admitiu, em **audiência** no Parlamento Europeu, q o lixo não impede a #infecção e a #transmissão, tornando a **imposição** do passaporte um mecanismo ilegal, pois os √©!n@dos podem ser infectados e transmitir a moléstia. <https://t.co/MXkJRtZLwd>

**Exemplo 9**

Tiranetes montados em retroescavadeira televisiva prometem às empresas farmacêuticas isenção de responsabilidade pelos riscos da falta de segurança e eficácia de **supostas** vacinas contra o #Covid19, transferindo-os para vítimas que pretendem vacinar **compulsoriamente**.

Essa visão é fundamentalmente enraizada na defesa das liberdades individuais contra a intromissão estatal, um embate cujas justificativas modernas encontram elos discursivos com grupos antivacina que se opuseram às políticas de imunização obrigatória desde o século XIX. Conforme analisado por Wolfe e Sharp (2002), os British Vaccination Acts, promulgados entre 1840 e 1867, foram os precursores de uma ampliação dos poderes governamentais sobre as liberdades civis em prol da saúde pública, desencadeando imediata resistência. Grupos organizados como a Liga Antivacina lideraram violentos tumultos, impulsionados por preocupações quanto à violação da liberdade individual e dos direitos decisórios parentais.

Situações semelhantes de resistência à obrigatoriedade da vacinação ocorreram tanto no Brasil como nos Estados Unidos no início do século XX. Em 1904, o Rio de Janeiro foi palco da chamada Revolta da Vacina, um movimento popular que rejeitava a obrigatoriedade da imunização contra a varíola imposta pelo governo. Um ano depois, em 1905, a Suprema Corte dos Estados Unidos confirmou a constitucionalidade de uma lei do estado de Massachusetts que tornou obrigatória a vacinação contra a varíola, suscitando a oposição de grupos contrários à medida.

Embates análogos, caracterizados pelo conflito entre liberdades individuais e ações governamentais em prol da saúde coletiva, pontuaram diversos outros momentos históricos. Essa tensão documentada nos permite traçar um paralelo direto com a raiz de atitude do individualismo, que emerge da primazia conferida às necessidades pessoais em detrimento das alheias. Assim, nesse polo negativo, o discurso é pautado por uma visão de mundo individualista que privilegia o sujeito em relação a qualquer responsabilidade social coletiva. Por trás das justificativas contra a vacinação obrigatória, encontra-se o discurso da defesa da liberdade individual.

### *3.1.2. Dimensão 2 – teorias político-conspiratórias sobre a vacina versus evidências pseudocientíficas contra a vacina*

O Quadro 6 lista as variáveis atuantes dos dois polos existentes no segundo fator extraído para a análise: alegações sobre a origem do vírus em laboratório chinês e interesses de políticos e governos ligados a grandes grupos farmacêuticos *versus* evidências de estudos pseudocientíficos publicados sobre os efeitos adversos das vacinas.

**Quadro 6 – Variáveis da Dimensão 2**

<b>Polo</b>	<b>Variáveis lexicais</b>
Positivo (61 variáveis)	chinês (0,76142), povo (0,74171), comprar (0,71011), governador (0,66846), pagar (0,64473), vender (0,64198), querer (0,63828), laboratório (0,62968), vacina (0,61471), salvar (0,61241), político (0,59485), mundo (0,59394), dar (0,57604), bilhão (0,57162), via (0,56834), cloroquina (0,56221), cobaia (0,56107), garantir (0,55709), ver (0,54319), gente (0,53357), remédio (0,53299), ficar (0,53152), liberdade (0,52182), governo (0,51455), bom (0,50492), esperar (0,49807), verdade (0,49322), perder (0,47940), saber (0,47189), país (0,46664), eficácia (0,45501), colateral (0,44921), dizer (0,44382), aprovar (0,44003), interesse (0,43366), teste (0,43163), pegar (0,43088), efeito (0,42412), tratamento (0,41759), vida (0,41642), obrigatório (0,41486), precoce (0,40288), criar (0,40283), compra (0,40057), lockdown (0,38837), campanha (0,38781), aplicar (0,38636), usar (0,38266), primeiro (0,38253), prefeito (0,37771), pandemia (0,36924), vírus (0,36890), especial (0,36429), ivermectina (0,35185), causa (0,34768), imunização (0,33777), público (0,33559), covid (0,32837), segundo (0,32553), emergencial (0,32380), farmacêutico (0,31582)

Negativo (38 variáveis)	nih (-0,87872), nlm (-0,87872), ncbi (-0,87872), literat (-0,83751), bvsalud (-0,82339), articles (-0,78500), hepatite (-0,59766), pós (-0,57290), fonte (-0,55686), global (-0,51415), célula (-0,50665), facebook (-0,50445), mrna (-0,49270), cdc (-0,48555), publicar (-0,48306), emitir (-0,46946), experimento (-0,44070), mielite (-0,43035), encuesta (-0,40923), atleta (-0,40743), transverso (-0,40650), futebol (-0,40494), alerta (-0,39513), correiopaulista (-0,38708), pesquisa (-0,36311), fulminante (-0,35375), diagnóstico (-0,35348), farmacêutica (-0,34409), rna (-0,33583), infecção (-0,33560), imposição (-0,33243), autoimune (-0,33120), indicar (-0,32435), dama (-0,32138), proposta (-0,31880), world (-0,31081), álcool (-0,30730), postagem (-0,30132)
----------------------------	---

Fonte: Elaborado pela autora.

Na Dimensão 2, os *tweets* do polo positivo buscam semear desconfiança e promover a ideia de que existe uma conspiração para ocultar os reais efeitos dos imunizantes. Além disso, insinuam a existência de um conluio entre autoridades, indústria farmacêutica e mídia com o propósito de manipular a verdade para atender a interesses próprios. Essa abordagem está intimamente ligada à raiz da atitude conspiratória, que se caracteriza pela adesão a teorias da conspiração, mesmo quando outras explicações mais plausíveis estão disponíveis. Os exemplos de 10 a 13 ilustram bem essa conexão:

#### Exemplo 10

**Bom dia Gente!** Um **vírus** mutante **criado** em **laboratório**? Escapa de uma segurança nível 4? Vacina? **Interesses farmacêuticos**? China tentando dominar as criptomoedas? Eu não **digo** é nada **Laboratório chinês** volta a ser cogitado como origem do coronavírus veja. abril. com. br / blog / mundialis ... via @VEJA

#### Exemplo 11

**FICA CLARA A CONSTATAÇÃO QUE POLÍTICOS, GOVERNOS E A MÍDIA COMPRADA, NÃO OUVIRÃO MÉDICOS SÉRIOS E CIÊNTISTAS. POIS OS BILHÕES E BILHÕES DAS VACHINAS FALAM MAIS ALTO.** Se usada no início ou preventivamente, a **ivermectina** pode aniquilar o **vírus chinês** diariodobrasil. org / se-usada-no-in ...

#### Exemplo 12

@BlitzSpitfire Assim que liberarem a vacina pra **Covid**, o governo gastará **bilhões** de reais com anuência do **povo** idiota que acredita na **boa fé** dos nossos **políticos**. Nem desconfiam que os interesses são outros. . .

### Exemplo 13

@carlosjordy - Para a sociedade, **fique** em casa, solde a porta da loja, multa, vacina **obrigatória**, aumento de impostos, e economia a **gente vê** depois; - Para **políticos** e amigos da imprensa, reajuste de 45 a 70 % nas verbas ( pagas com o imposto do pobre ) ; Mas acredite, é tudo por uma **boa causa** !

Como observado por Douglas *et al.* (2019) em sua análise abrangente de teorias conspiratórias, há diversas razões pelas quais uma pessoa pode aderir a elas, que variam desde traços de personalidade até a satisfação de necessidades sociais complexas. Nas redes sociais, essas teorias se disseminam rapidamente devido à natureza interconectada das plataformas e à ação dos algoritmos, criando bolhas de filtro que reforçam as crenças individuais e isolam os usuários de perspectivas alternativas.

Malthouse (2023) ressalta que o viés de confirmação é uma consideração importante na comunicação sobre a imunização, já que há uma propensão a interpretar erroneamente evidências sobre vacinas, de modo a refletir crenças preestabelecidas. Essa dinâmica acaba alimentando um ciclo de desinformação e desconfiança em relação aos envolvidos no processo de vacinação, apontando para o discurso da falta de confiança em figuras de autoridades – desde instituições e governos até empresas e especialistas.

Em contraste, o polo negativo mostra como o movimento antivacina, embora muitas vezes associado a teorias conspiratórias e desinformação, também utiliza, em certa medida, pesquisas científicas para fundamentar suas posições. No entanto, essa apropriação tende a ser enviesada e seletiva, ignorando o consenso científico estabelecido e enfatizando resultados isolados que parecem apoiar sua narrativa. Os exemplos de 14 a 18 ilustram como *links* para importantes centros norte-americanos de pesquisa na área da saúde são listados para respaldar alegações sobre supostos danos causados pelas vacinas:

### Exemplo 14

Existem estudos, **publicados** pela própria OMS, relacionando a doença de #anthonyjohnson à #vacinas de **mRNA** . Confirmam os links: pesquisa. **bvsalud.org** / global-literat ... **ncbi. nlm. nih. gov** / pmc / **articles** / P ...

**Exemplo 15**

Pesquisa científica publicada no @ncbi relacionando a #depressão e a tentativa de suicídio às vacinas. Fonte : **ncbi. nlm. nih. gov / pmc / articles / P ...**  
<https://t.co/q6WyWqwuN1>

**Exemplo 16**

Pesquisa publicada no PubMed sobre a ativação de fenômenos auto-imunes pós-vacina é assustadora! A pesquisa indica que a #vacina de **mRNA** pode provocar problemas hepáticos graves, sequelas irreversíveis no fígado e nefropatias. Fonte :  
 pubmed. **ncbi. nlm. nih. gov / 34957554 / .** <https://t.co/e502VKuxqg>

**Exemplo 17**

A OMS e outras revistas científicas, publicaram estudos relacionando a ideação suicida e tentativa de #suicídio à vacina. Alguns cientistas afirmam q as doses podem suprimir ou interferir na produção da Dopamina e Serotonina, potencializando a depressão. pesquisa. **bvsalud.org / global-literat ...** <https://t.co/90FQcVBTTa>

**Exemplo 18**

Pesquisa publicada no @Pubmed relacionando o aparecimento de Trombose à vacina (estudo multicêntrico) . Fonte : pubmed. **ncbi. nlm. nih. gov / 34370972 /**  
<https://t.co/10cltDPh9U>

Essa prática de utilizar estudos científicos de forma enviesada para contestar a vacinação não é nova. Conforme observado por Guttinger (2019), os avanços recentes na pesquisa do microbioma humano, por exemplo, têm sido alvo de interpretações e manipulações por movimentos antivacina. Esses grupos se apropriam de hipóteses emergentes ou descobertas preliminares e as distorcem ou generalizam para além do escopo original do estudo.

Nesse contexto, as crenças desprovidas de fundamentação científica, outra raiz de atitude antivacina, emergem nesse polo. Embora reivindiquem a ciência, os textos carecem de provas ou plausibilidade e distorcem fatos para se adequarem a concepções preconcebidas, evidenciando um discurso a favor da pseudociência.

### 3.1.3. Dimensão 3 – defesa do tratamento precoce em detrimento da vacina versus relatos pessoais sobre efeitos colaterais da vacina

O Quadro 7 apresenta o conjunto de variáveis coocorrentes do terceiro fator, cujos polos positivos e negativos receberam as seguintes nomeações: defesa do tratamento precoce frente à letalidade do vírus e à falta de estudos sobre a eficácia e a segurança em longo prazo da vacina emergencial *versus* relatos sobre amigos e terceiros que sofreram com efeitos colaterais graves – dores de cabeça, paralisia, problemas no rim, nas pernas e nos pés – e até fatais após a vacinação.

**Quadro 7 – Variáveis da Dimensão 3**

Polo	Variáveis lexicais
Positivo (55 variáveis)	letalidade (0,93945), índice (0,88251), estatística (0,88111), segurança (0,80301), infecção (0,73291), prazo (0,70034), assegurar (0,68924), estudo (0,65232), longo (0,64329), impedir (0,61274), risco (0,61259), transmissão (0,59666), taxa (0,59352), lockdown (0,57503), precoce (0,56670), vírus (0,52658), representar (0,52654), imunidade (0,52510), tratamento (0,51383), máscara (0,50804), saudável (0,49770), sanitário (0,49498), usar (0,49150), eficácia (0,48119), benefício (0,46997), lógico (0,45338), doença (0,45005), médio (0,44215), massa (0,43768), variante (0,42629), pandemia (0,42290), social (0,42220), jovem (0,42181), bom (0,38722), custo (0,38240), científico (0,38085), mentar (0,37752), isolamento (0,37651), criança (0,37113), aprovar (0,36883), emergencial (0,36556), público (0,35527), garantir (0,35513), bilhão (0,35233), natural (0,34869), interesse (0,34267), caráter (0,33719), morte (0,32596), obrigatoriedade (0,31939), maioria (0,31886), forçar (0,31611), perigoso (0,31418), transmitir (0,31083), faixa (0,30887), paciente (0,30236)
Negativo (32 variáveis)	dor (-0,70703), cabeça (-0,70414), perna (-0,58522), ataque (-0,58404), cardíaco (-0,57438), paralisia (-0,57396), amiga (-0,57255), pé (-0,56451), resfriar (-0,54646), trombone (-0,53905), exame (-0,52451), rim (-0,52205), brasil (-0,49335), facial (-0,48978), pele (-0,46964), amigo (-0,46641), falecer (-0,43905), diagnóstico (-0,42023), apertar (-0,40594), sangue (-0,40509), infarto (-0,40205), semana (-0,39123), calça (-0,38275), miocardite (-0,37461), verdade (-0,37454), lista (-0,35126), sofrer (-0,34109), segundo (-0,33326), gripe (-0,32318), pegar (-0,31796), comprimido (-0,31445), news (-0,31229)

Fonte: Elaborado pela autora.

A ênfase na defesa do chamado “tratamento precoce” define o polo positivo da Dimensão 3. Seus apoiadores argumentam que esse tratamento é composto de remédios reconhecidamente seguros e com longa utilização na prática médica tradicional, como indicado nos Exemplos 19 e 20:

### Exemplo 19

**Tratamento precoce** com remédios utilizados há décadas e baixo **risco** de efeitos colaterais - não pode. . . não tem **estudos científicos** suficientes; Vacina com baixo **índice de eficácia** e testes insuficientes - não devemos ser tão **científicos** nesse momento; Tá bom . . . vai lá sabichão!

### Exemplo 20

A urgência justifica o uso de vacinas experimentais sem **estudos científicos** que **assegurem eficácia** e, principalmente, **segurança** quanto aos efeitos colaterais. Mas quando o assunto é o **tratamento precoce** com remédios sabidamente seguros, a urgência deixa de ser urgente!

Essa perspectiva também direciona críticas às vacinas contra a Covid-19, questionando sua eficácia e segurança, bem como as bases científicas que as favorecem em detrimento do tratamento precoce. Além disso, outras estratégias para conter a propagação da pandemia, como o uso de máscaras e as medidas de *lockdown*, são alvo de ataques quanto à sua efetividade, conforme evidenciado nos exemplos de 21 a 24:

### Exemplo 21

Vamos aos fatos: O **vírus** foi feito em laboratório; **Máscara** e **lockdown** são inúteis; Vacinas têm **eficácia** duvidosa e efeitos colaterais **perigosos**; A **letalidade** é baixa em pessoas saudáveis; A imprensa **mente**; Muitos roubaram durante **pandemia**; O **tratamento precoce** salva vidas;

### Exemplo 22

Disseram que o **lockdown** achataria a curva, não achatou; Disseram que **máscaras** impediriam a **transmissão**, não **impediram**; Disseram que a **pandemia** acabaria quando a **maioria** estivesse vacinada, não acabou; Estes mesmos especialistas disseram que o **tratamento precoce** não é eficaz.



### Exemplo 23

@SigaGazetaBR A guerra de informação deixou tudo muito turvo, mas já há **estudos** suficientes para afirmar que o **vírus** veio de laboratório, que o **tratamento precoce** salva vidas, que as vacinas têm **riscos** de efeitos colaterais severos, que **lockdown** e **máscara** são ineficazes, e por aí vai. . .

### Exemplo 24

Defender o **tratamento precoce** é crime pois não há **estudos científicos** comprovando a **eficácia** dos medicamentos, mas questionar a **obrigatoriedade** de **máscaras**, **lockdown**, e vacinas, **usando** o mesmo argumento da ausência de **estudos** comprobatórios, também é crime! Ciência? 😊


Esse fenômeno de promoção de medicamentos ou substâncias não comprovadas como curas milagrosas durante crises de saúde é chamado de “messianismo farmacêutico” por Lasco e Yu (2021). Segundo os autores, medicamentos podem se tornar veículos de ideologias que alteram narrativas sociais de acordo com as circunstâncias políticas. Nesse contexto, identifica-se novamente o discurso favorável à pseudociência, explorado para possibilitar que políticos se apresentem como salvadores, oferecendo soluções mágicas para crises complexas.

Por outro lado, o polo negativo da Dimensão 3 é centrado na disseminação de relatos pessoais descrevendo supostos efeitos colaterais adversos após a vacinação contra a Covid-19, com abordagem minuciosa e íntima, conforme demonstrado nos exemplos de 25 a 28. Essa ênfase na narrativa pessoal amplia a percepção dos riscos potenciais relacionados à vacinação, conferindo uma dimensão mais próxima e concreta à experiência.

### Exemplo 25

Aline tomou a 2ª dose, e 1 mês depois começou a sentir **dores** nos **pés** e **pernas**, até que um dia não tinha forças nas **pernas** nem para ficar em **pé**! Foram muitos **exames** até o **diagnóstico**: Mielite Transversa autoimune, 1 inflamação na medula espinhal. Depoimento da mãe: oscasosraros <https://t.co/e2M4uE4qXu>

### Exemplo 26

A pequena Cleuza é uma grande guerreira, enfrentando vários **diagnósticos** de eventos adversos, depois que recebeu 2 doses da . As reações começaram com **dores** de **cabeça** e pelo corpo, câimbras e formigamentos nas **pernas**, acompanhada de inchaço e vermelhidão, visão embaçada. <https://t.co/pbCTZolgNB>

### Exemplo 27


@AllanGarcs1 Minha sogra tomou a 1º dose da AstraZeneca, 1 tornozelo inchou. Teve covid mesmo após a 1º dose, ficou bem, mas os dois tornelos e **pés** o incharam. Recomendei que fizessem **exames**. Fibrinogênio e D-dímero deram elevados. A **segunda** dose foi suspensa, fazendo uso de anticoagulante

### Exemplo 28

Apenas 5 dias depois de sua 1ª dose, Felipe, de apenas 15 anos, sofre um infarto e vai a óbito. Pela manhã se queixa de **dor** de **cabeça**, e é medicado. À tarde, estava com **amigos**, andando de bicicleta, e desmaiou. <https://t.co/TGIC8WmLxw>

Dentro desse mesmo polo, também são evidentes narrativas extensas e elaboradas, como ilustrado nos exemplos de 29 a 32, *tweets* compartilhados pelo mesmo usuário. Segundo Semino *et al.* (2023), evidências empíricas apontam para uma vantagem persuasiva de relatos pessoais sobre a apresentação de informações (dados e estatísticas) em contextos de comunicação de saúde, pois as narrativas diminuem a contra-argumentação e a resistência à persuasão.

### Exemplo 29

@TonyStarkMeta Uma **amiga** ficou com a **pele** dos **pés** a **cabeça** parecendo uma larva de vulcão. Achei q fosse morrer. Mais de 1 mês internada e os médicos queriam q ela tomasse a **segunda** dose  Esses médicos são assassinos.

**Exemplo 30**

@mjcarvalhop @GFiuzza\_Oficial Temos q orar por esses médicos, o q eles estão enfrentando p revelar a verdade ã é fácil. Uma **amiga** minha, no dia seguinte à vacina, parecia q o **sangue** do corpo dela queria sair pelos poros. Ficou vulcão dos **pés** à **cabeça**. Até a palma da mão. Assustador. + de 1 mês no hospital

**Exemplo 31**

@dramayraoficial Doutora, preciso da sua ajuda Uma **amiga** tomou a vacina da Pfizer e o corpo ficou irreconhecível. Ela está na cor vinho e cheia de bolinhas e inchada dos **pés** à **cabeça** Os médicos dizem q é d um remédio q ela tomava por causa de umas convulsões + é mentira eles ã querem admitir

**Exemplo 32**

A filha de uma **amiga** que se vacinou acabou de morrer. Que tristeza! Deixou uma filhinha. Tomou a 1º dose aí teve problema nos **rins** e na **pele** . . . . ficou roxa e inchada. Os médicos se negaram a dizer que foi da vacina é ã sabiam como tratá -la Essa vacina é veneno. . . .!!!!!!

Dessa forma, observa-se novamente a utilização do discurso do medo, agora mobilizado por meio de relatos pessoais que descrevem experiências extremamente negativas após a imunização, difundindo a ideia de que as vacinas representam riscos significativos e de que seus efeitos colaterais podem causar danos à saúde graves e, em muitos casos, até fatais.

**3.2. Discussão dos resultados**

No estudo piloto sobre os discursos antivacina no Twitter (Whiteman, 2023), foram encontradas as seguintes dimensões:

- Dimensão 1 – defesa das redes sociais.
- Dimensão 2 – alertas sobre efeitos colaterais das vacinas de Covid-19.
- Dimensão 3 – crítica a governo, veículos de mídia oficiais e medidas de saúde pública.
- Dimensão 4 – relatos sobre efeitos colaterais das vacinas de Covid-19.
- Dimensão 5 – resistência à vacinação obrigatória.

- Dimensão 6 – defesa da autonomia parental.
- Dimensão 7 – defesa da liberdade de escolha.

Já a presente pesquisa, como apresentado na seção anterior, encontrou as três dimensões a seguir:

- Dimensão 1 – alerta sobre riscos cardíacos, males súbitos e falecimentos após doses de reforço, especialmente entre jovens e crianças, considerando a vacina como experimental *versus* questionamento e resistência à imposição de medidas e decretos compulsórios de combate ao avanço da Covid-19.
- Dimensão 2 – alegações sobre a origem do vírus em laboratório chinês e interesses de políticos e governos ligados a grandes grupos farmacêuticos *versus* evidências de estudos pseudocientíficos publicados sobre os efeitos adversos das vacinas.
- Dimensão 3 – defesa do tratamento precoce frente à letalidade do vírus e à falta de estudos sobre a eficácia e a segurança em longo prazo da vacina emergencial *versus* relatos sobre amigos e terceiros que sofreram com efeitos colaterais graves – dores de cabeça, paralisia, problemas no rim, nas pernas e nos pés – e até fatais após a vacinação.

Confrontando os dois resultados, observamos sobreposições significativas entre eles. Ambos apontaram dimensões relacionadas a alertas e relatos sobre efeitos colaterais graves das vacinas contra a Covid-19, incluindo eventos adversos fatais. Além disso, as duas análises captaram dimensões que questionam a obrigatoriedade e a imposição de medidas e vacinas relacionadas à pandemia, defendendo a liberdade de escolha individual.

As pesquisas revelaram ainda diferenças discursivas. Enquanto o estudo anterior apresentou dimensões mais genéricas de crítica a governo, veículos de mídia e defesa da autonomia parental, este foi mais específico, identificando dimensões que englobam alegações sobre a suposta origem do vírus em um laboratório chinês e interesses de políticos e governos supostamente ligados a grandes grupos farmacêuticos. Ademais, o presente estudo trouxe à tona uma dimensão inédita sobre o tratamento precoce da Covid-19 e questionamentos sobre a eficácia e a segurança em longo prazo das vacinas emergenciais, aspectos não cobertos pela pesquisa piloto.

Embora exista uma sobreposição conceitual, os dois estudos revelam uma evolução e um refinamento na identificação de dimensões do discurso antivacina. Esta pesquisa é mais detalhada, captando nuances importantes como as alegações sobre a origem do vírus, os questionamentos sobre tratamentos alternativos e o caráter experimental das vacinas. Ao mesmo tempo, mantém o foco em discursos-chave como a resistência à obrigatoriedade de imunização e os alertas e relatos de supostos efeitos adversos graves.

Essa angulação comparativa sugere um aprimoramento na compreensão do fenômeno discursivo antivacina com a ampliação do corpus e o desenvolvimento em longo prazo do trabalho, o que proporcionou a identificação dos quatro principais discursos associados às três dimensões lexicais encontradas neste estudo, como destacado no Quadro 8:

**Quadro 8 – Discursos das dimensões**

<b>DIM 1</b>	Polo positivo	Discurso do medo
	Polo negativo	Discurso da liberdade individual
<b>DIM 2</b>	Polo positivo	Discurso da desconfiança de figuras de autoridade
	Polo negativo	Discurso a favor da pseudociência
<b>DIM 3</b>	Polo positivo	Discurso a favor da pseudociência
	Polo negativo	Discurso do medo

Fonte: Elaborado pela autora.

A investigação dos discursos envolve analisar as práticas sociais que emergem deles. Nesse contexto, as chamadas estratégias discursivas se configuram como “modos sistemáticos” de empregar a linguagem com o propósito de alcançar “objetivos sociais, políticos ou psicológicos” (Wodak, 2005, p. 4). Tais estratégias estão vinculadas a formas de persuasão através do discurso, buscando, assim, convencer indivíduos por meio deles (Vaara, 2010).

Uma estratégia discursiva consiste em um padrão recorrente nas formas de expressão que evidencia um uso específico da linguagem para induzir alguém a determinada crença ou ação. Dessa maneira, o exame dessas estratégias de persuasão representa um elemento fundamental para compreender mais

profundamente os discursos, sua estrutura, propagação e influência na construção (ou desconstrução) do conhecimento sobre variados tópicos.

As estratégias persuasivas empregadas nos discursos antivacina são multifacetadas, abrangendo abordagens emocionais e racionais com o intuito de influenciar atitudes e moldar comportamentos. Seu propósito é modelar não apenas as percepções, mas também as decisões dos indivíduos em relação à vacinação e às medidas de combate à pandemia, valendo-se de múltiplos ângulos e táticas articuladas.

A Dimensão 1 contempla, em seu polo positivo, esforços para incitar receios e resistência frente à dose de reforço vacinal, com especial ênfase nos mais jovens. Nesse âmbito, disseminam-se informações sobre riscos à saúde, como dados de eventos cardiovasculares e óbitos, com o propósito de instilar um sentimento de medo em relação à vacinação. Paralelamente, enfatiza-se o caráter “experimental” da vacina para fomentar desconfianças quanto à sua segurança e eficácia, minando a credibilidade da ciência experimental por trás de seu desenvolvimento. Já no polo negativo, a estratégia fundamental consiste em reivindicar a liberdade e a autonomia pessoal, associando as medidas compulsórias de combate à Covid-19 a uma suposta perda dessas liberdades individuais. Concomitantemente, busca-se promover uma desconfiança das autoridades governamentais, retratando tais medidas como imposições arbitrárias do poder público.

O polo positivo da Dimensão 2 visa espalhar desconfianças nas instituições governamentais e corporações, especialmente empresas farmacêuticas, por meio da disseminação de narrativas conspiratórias acerca da suposta origem proposital do vírus e de relações controversas entre políticos e a indústria. Sugere-se a existência de um plano maior ou uma conspiração para manipular e controlar a população através da vacinação, apelando ao medo das pessoas de serem subjugadas. Nesse contexto, cria-se uma divisão antagônica entre a população comum e as elites políticas e econômicas supostamente por trás dessa alegada trama conspiratória. Por sua vez, no polo negativo busca-se fundamentar as objeções à vacina por meio de estudos e “evidências” pseudocientíficas, desprovidas da devida validação pela comunidade científica, em uma tentativa de conferir credibilidade aos argumentos contrários. Enfatizam-se supostos efeitos adversos da vacina documentados em estudos enviesados ou de metodologia duvidosa, visando dissuadir os indivíduos de se vacinarem. Além disso, destacam-se falhas, inconsistências ou lacunas

propositalmente inseridas em pesquisas sobre a eficácia das vacinas, com o intuito de semear dúvidas quanto à sua real capacidade protetora.

Na Dimensão 3, as estratégias persuasivas empregadas no polo positivo apelam à lógica e à prudência, argumentando ser mais sensato e seguro utilizar tratamentos já conhecidos do que confiar em uma vacina nova cujos efeitos em longo prazo ainda não foram totalmente mapeados. Critica-se a falta de estudos e dados abrangentes sobre a segurança da vacina em um horizonte temporal mais amplo, fomentando hesitações e dúvidas acerca de seus potenciais riscos futuros. Ademais, enfatiza-se a letalidade e os perigos do vírus para justificar o suposto uso de tratamentos precoces alternativos à vacina, contestando a real necessidade da vacinação. No polo negativo, as estratégias persuasivas se baseiam na utilização de relatos e testemunhos pessoais de supostas reações adversas, com o intuito de criar uma conexão emocional com o público e aumentar a credibilidade geral da mensagem antivacina. Espera-se que os indivíduos se identifiquem com as histórias pessoais relatadas, ampliando, assim, o impacto emocional da mensagem e gerando maior empatia com a causa antivacina. Há também uma tendência a enfatizar efeitos colaterais graves supostamente relacionados à vacina, com o objetivo de amplificar o medo e a hesitação em relação a ela entre a população.

Podemos, então, resumir as estratégias persuasivas identificadas neste estudo da seguinte forma:

- Dimensão 1: usa o medo (risco à saúde) e o apelo à liberdade pessoal.
- Dimensão 2: fomenta a desconfiança (teorias conspiratórias) e apela à autoridade científica (evidências).
- Dimensão 3: baseia-se na prudência (tratamento precoce) e na empatia (relatos pessoais).

A análise das estratégias persuasivas empregadas nos discursos antivacina revela uma orquestração cuidadosa de narrativas e táticas discursivas articuladas com o propósito de delinear crenças e incitar ações específicas na população. Portanto, tais estratégias convergem no objetivo central de persuadir indivíduos a adotarem posturas de hesitação, desconfiança e resistência em relação às vacinas e políticas pandêmicas.

No presente capítulo, foram apresentados e discutidos os resultados da AMD Lexical aplicada a um corpus infodêmico de *tweets*, análise que foi conduzida com base nos princípios teórico-metodológicos propostos por Berber Sardinha (2020, 2021). A próxima seção encerra esta dissertação ao sumarizar a pesquisa, fornecendo uma análise crítica, além de examinar algumas possibilidades de aplicação dos resultados obtidos.



## Considerações finais

Esta seção busca concluir o estudo com uma análise dos elementos que nortearam a pesquisa, desde os fatores motivadores até os objetivos traçados e as perguntas de pesquisa formuladas. Serão também discutidos os resultados alcançados, as limitações e as descobertas que podem inspirar novas investigações futuras.

De saída, é fundamental apresentar o Programa Estratégico Emergencial de Prevenção e Combate a Surtos, Endemias, Epidemias e Pandemias, instituído pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Seu objetivo central consiste na ampliação do conhecimento científico e no fomento à pesquisa sobre os impactos sociais, econômicos, culturais e históricos decorrentes da pandemia de Covid-19. Nesse contexto, o presente estudo, como parte integrante do programa, alinha-se a esse propósito ao investigar a disseminação de desinformação sobre vacinas, uma questão urgente de âmbito nacional que se intensificou durante a pandemia.

Durante esse período crítico, observou-se uma queda substancial nas taxas gerais de vacinação no Brasil, evidenciando o impacto tangível da propagação de desinformação. As vacinas e sua eficácia tornaram-se temas centrais no discurso público, especialmente nas redes sociais, destaque intensificado pelo impacto direto da infodemia. Esta pode ser definida como um aumento exponencial no volume de informações associadas a um determinado assunto, as quais se multiplicam rapidamente em pouco tempo (OMS, 2020). Embora a literatura prévia tenha focado a infodemia do ponto de vista do excesso e da rápida difusão da comunicação verbal (textual), este estudo a concebeu como um complexo ecossistema de discursos que se desenvolveu durante a pandemia. Esse ambiente discursivo abrange não apenas informações precisas, mas também diversas formas de desinformação.

Estudos de Análise do Discurso e multimodais sobre a pandemia lidaram com quantidades limitadas de dados. Em contrapartida, pesquisas em *big data* têm se concentrado em aspectos como tópicos, sentimentos e *frames*, não se detendo em discursos. Portanto, no presente trabalho, a Linguística de Corpus (LC) foi uma ferramenta essencial para identificar e analisar discursos em ecossistemas infodêmicos, por sua aptidão e *expertise* em lidar com grandes volumes textuais. Neste estudo, a LC orientou os princípios para a criação do CBDIAT, um conjunto de dados linguísticos representativos da infodemia antivacina no Twitter. Também

forneceu as bases para a compreensão da construção de sentidos na língua em uso, por meio da coocorrência sistemática de itens lexicogramaticais.

As respostas às questões centrais desta pesquisa – 1. Quais são as dimensões lexicais dos discursos em torno do movimento antivacina? 2. Quais os principais discursos associados às dimensões lexicais? – corroboram os objetivos delineados. A metodologia adotada para alcançar esses propósitos foi a Análise Multidimensional Lexical (AMD Lexical) proposta por Berber Sardinha (2016, 2017, 2020, 2021). Foram identificadas as coocorrências mais frequentes de adjetivos, substantivos e verbos nos textos, seguidas de análises fatoriais e de variância e da interpretação dos resultados. A investigação exploratória dos dados revelou três dimensões lexicais e quatro discursos principais.

A análise dos resultados foi concretizada por meio da aplicação da metodologia da AMD Lexical, em conjunto com o respaldo de teorias discursivas, as quais foram fundamentais para a ampliação de nossa investigação. Constituinte desse arcabouço teórico, a Análise Crítica do Discurso (ACD) investiga como o discurso está intrinsecamente ligado a contextos institucionais e políticos mais amplos (Wodak; Meyer, 2001). Portanto, mostra-se indispensável para examinar o contexto ideológico subjacente aos discursos infodêmicos revelados nas dimensões.

Kennedy (2019) investiga a possível correlação entre a ascensão do populismo político e o aumento da hesitação em relação às vacinas. Segundo o autor, tanto a nova onda de idolatria a líderes populistas quanto a disposição antivacina compartilham uma mesma origem: um profundo sentimento de desconfiança e ressentimento em relação a elites, especialistas e instituições estabelecidas, especialmente entre os segmentos economicamente desfavorecidos e marginalizados da sociedade.

As narrativas pós-Segunda Guerra Mundial de otimismo, prosperidade econômica e progresso contínuo não se materializaram para todos os segmentos da sociedade, resultando em um crescente ceticismo e na rejeição dessas concepções abrangentes de avanço civilizacional. O populismo e o movimento antivacina surgem, em parte, como expressões desse descontentamento. Grupos não beneficiados pela globalização e pelo crescimento econômico se sentem cada vez mais excluídos da esfera política e das instituições *mainstream*. Essa erosão geral da confiança pública na ciência e em outras autoridades pode ser interpretada como uma recusa mais

ampla dessa narrativa de progresso que negligenciou esses grupos e se reflete em movimentos como a negação das mudanças climáticas e a hesitação vacinal.

Esse fenômeno também se evidencia na diminuição da confiança nos veículos de mídia tradicionais. Segundo o Relatório de Notícias Digitais do Instituto Reuters (Newman *et al.*, 2019), essa confiança vem declinando tanto no Brasil quanto globalmente. Especialistas apontam que esses meios de comunicação estão desconectados da realidade da população, resultando em um grande abismo entre o que é veiculado pela mídia e o que o público realmente deseja saber. Além disso, o público muitas vezes enfrenta dificuldades para compreender informações complexas apresentadas por especialistas, sobretudo quando se trata de dados de viés negativo. Como resultado dessa falta de conexão, surgem sentimentos de ansiedade, medo e frustração, levando muitas pessoas a buscar fontes de informação alternativas.

No Brasil, nossa breve trajetória democrática tem sido marcada por uma sucessão de líderes políticos com inclinações populistas. Desde Getúlio Vargas, cujo carisma exercia um fator mobilizador das massas, até Fernando Collor de Mello, com sua crítica ferrenha às políticas estatais, a cadeira presidencial foi ocupada diversas vezes por políticos com esse perfil. A vitória eleitoral de Jair Bolsonaro em 2018 acendeu o debate sobre a existência de um novo populismo de direita no país (Codato; Berlatto; Bolognesi, 2018; Tamaki; Fuks, 2020) ou, de forma mais enfática, com aspectos pós-fascistas (Finchelstein, 2019). Esse populismo característico de Bolsonaro se evidenciou de maneira contundente com a crise pandêmica da Covid-19.

Durante esse período, o então presidente transformou a questão da vacinação em uma disputa político-ideológica. Apesar da escalada da crise sanitária no país, com um aumento significativo no número de casos e mortes, Bolsonaro minimizou a gravidade da Covid-19, contrariando as orientações das autoridades sanitárias e da comunidade científica. O ex-mandatário fez diversas declarações públicas contrárias às vacinas aprovadas e em desenvolvimento, semeando dúvidas sobre sua eficácia e segurança. Paralelamente, promoveu o uso de medicamentos sem comprovação científica para o tratamento da doença viral, como a cloroquina e a ivermectina, seguindo o caminho oposto das recomendações internacionais de saúde.

Nesse contexto, ao examinarmos o movimento antivacina e seu ecossistema infodêmico pelo prisma do populismo, podemos descortinar os encadeamentos que compõem a intrincada cadeia discursiva responsável por nutri-los. Para tal, conforme

preconiza a Análise Dialógica do Discurso (ADD), abordagem fundamental para a triangulação interdisciplinar proposta nesta pesquisa, faz-se necessário reconhecer a continuidade histórica e o desenvolvimento gradual desse fenômeno ao longo do tempo, a fim de identificar os elos que entrelaçam sua complexa teia de comunicação discursiva.

O populismo é um movimento político capaz de articular uma série de discursos e narrativas de forma engenhosa, criando uma rede semântica coesa que apela a determinados setores da população. Um dos fios condutores dessa teia discursiva é a exaltação do individualismo e da liberdade de escolha pessoal acima das responsabilidades coletivas e das restrições estatais. Essa mensagem encontra terreno fértil no viés de confirmação, uma tendência humana a privilegiar informações que reforçam crenças e visões de mundo pré-existentes. Os argumentos populistas costumam validar as percepções prévias de seus apoiadores, ignorando ou desacreditando evidências contraditórias. Essa dinâmica é reforçada pelo emprego recorrente de teorias conspiratórias, que criam uma sensação de que há forças ocultas atuando contra os interesses autênticos do “povo”.

Tais narrativas conspiratórias alimentam um sentimento de perseguição e desconfiança generalizada não apenas em relação a grupos específicos, mas também às próprias instituições públicas, à mídia tradicional e até mesmo à ciência estabelecida. Esse ceticismo em relação às autoridades tradicionais abre espaço para a proliferação de fontes de informação alternativas e teorias não comprovadas. Não raro, esses discursos flertam com ideias pseudocientíficas. A adoção dessas crenças infundadas é facilitada pela desconfiança já instalada em relação à *expertise* científica convencional.

Além de desenredar esse emaranhado de discursos que interagem no ecossistema infodêmico antivacina, faz-se necessário identificar ecos anteriores de sua cadeia discursiva. É notório que o movimento não é recente, remontando a Edward Jenner e à introdução da vacina contra a varíola em 1796. Na Inglaterra vitoriana, a oposição à imunização compulsória estava profundamente enraizada nas divisões de classe social e nas desigualdades vigentes à época.

Para as camadas desfavorecidas da população, o Ato de Vacinação de 1853 representou uma imposição das classes abastadas e privilegiadas. Com acesso precário a água potável, cuidados médicos adequados e condições sanitárias, os estratos mais pobres enfrentavam riscos elevados de contrair infecções e sofrer

complicações decorrentes das práticas de vacinação muitas vezes insalubres praticadas naquele período. Registros documentam casos de transmissão acidental de doenças secundárias durante a vacinação “braço a braço”, na qual o mesmo instrumento era usado em vários pacientes sem a devida esterilização. Além disso, o ônus econômico das taxas de vacinação obrigatória recaía de maneira desproporcional e severa sobre as famílias empobrecidas.

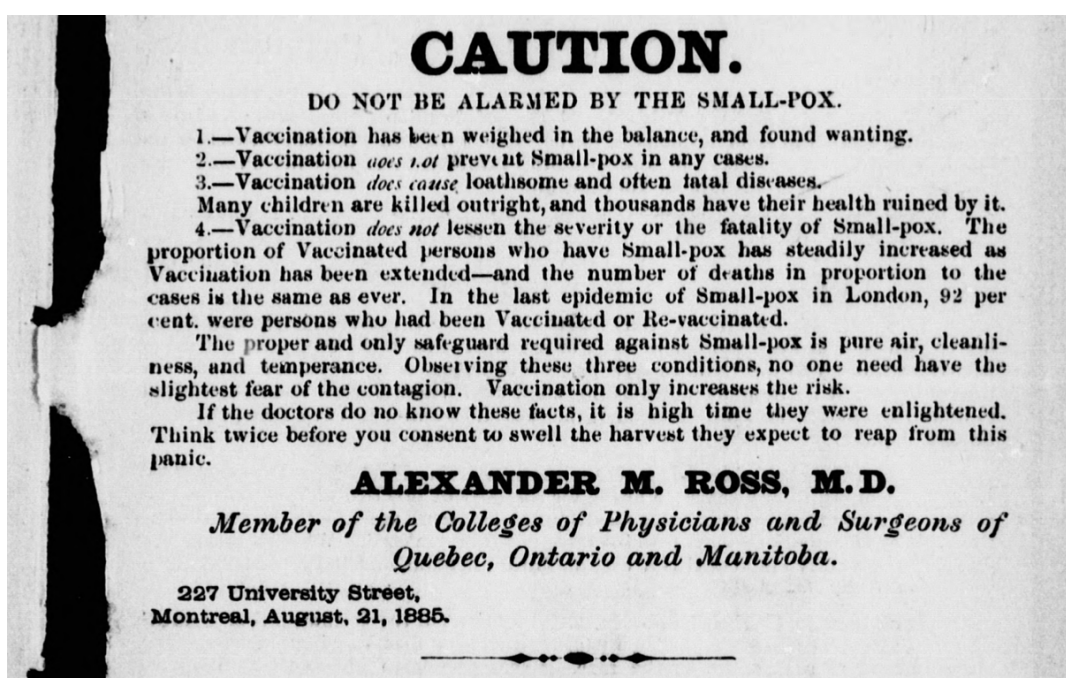
As mulheres, destituídas de muitos direitos sobre seus próprios corpos, enxergavam a vacinação compulsória como mais uma instância de controle estatal sobre suas vidas, à semelhança das Leis de Doenças Contagiosas que permitiam inspeções médicas forçadas em suspeitas de prostituição, independentemente de consentimento (Berman, 2020). Tanto os pobres quanto as mulheres nutriam temores legítimos de que as políticas de vacinação estatais submeteriam ainda mais seus corpos e sua autonomia pessoal aos desígnios de uma estrutura de poder patriarcal dominada pela burguesia.

Políticos oportunistas, grupos organizados e indivíduos com agendas próprias exploraram os receios legítimos das classes trabalhadoras e das mulheres sobre a vacinação compulsória. Utilizando táticas enganosas, esses atores amplificaram medos e desconfianças existentes, distorcendo estatísticas e selecionando dados para corroborar alegações alarmistas. Manipulando preocupações reais, a retórica inflamatória pintava a vacinação como ameaça às liberdades, afronta aos direitos individuais e fonte de doenças graves.

Em contraste com a manipulação de questionamentos legítimos do passado, fundamentados nas práticas médicas rudimentares ou na luta de classes daquela época, o que se propaga atualmente são temores e teorias conspiratórias destituídos de embasamento factual, como narrativas de que as vacinas contra a Covid-19 contêm “chips líquidos” para controle por inteligência artificial ou causam ideação suicida. Discursos antigos, como o da liberdade individual e o do medo, são reutilizados devido à sua comprovada capacidade de aderência ao longo da história. Porém, esses discursos são ressignificados e recontextualizados para instrumentalizar inseguranças coletivas, perpetuar dúvidas e polarizações, mesmo diante de sólidas evidências científicas que reconhecem amplamente o sucesso das vacinas na erradicação de doenças e na prevenção de mortes e a existência de rígidos protocolos sanitários vigentes globalmente, os quais garantem a segurança dos imunizantes.

O movimento antivacina tem ainda uma longa tradição em promover as palavras de supostos “especialistas” que corroboram sua narrativa. Um exemplo histórico é o do Dr. Alexander M. Ross. Durante a epidemia de varíola em Montreal, em 1885, enquanto autoridades de saúde pública buscavam aumentar a cobertura vacinal, o médico canadense teve seus panfletos antivacina amplamente divulgados (Figura 6). Ross aproveitou a oportunidade para ganhar notoriedade e fama pessoal, autodeclarando-se como o “único médico que ousou duvidar do fetiche da vacinação” (Larson, 2020). Posteriormente, no entanto, descobriu-se que ele próprio havia sido vacinado durante aquela epidemia, fato extensamente relatado pelos principais jornais da época (Larson, 2020). Essa contradição evidencia como, mesmo no passado, líderes do movimento exploravam a desconfiança pública para obter projeção pessoal, valendo-se de discursos dúbios e ações controversas.

**Figura 6** – Panfleto antivacina de 1885



Fonte: Biblioteca Digital HathiTrust.<sup>24</sup>

Essa estratégia persiste até os dias atuais, com o movimento contando com porta-vozes carismáticos, a exemplo de Andrew Wakefield, o ex-cirurgião inglês que ficou mundialmente conhecido por seu estudo fraudulento que alegava uma relação

<sup>24</sup> Panfleto completo disponível em: <https://babel.hathitrust.org/cgi/pt?id=aeu.ark:/13960/t7wm29713&seq=1>. Acesso em: 11 jun. 2023.

entre a vacina tríplice viral e o autismo. Conforme relatado por Boseley (2018), após ter sua licença cassada, Wakefield retornou aos holofotes na esteira da presidência populista de Donald Trump, tendo participado da posse presidencial e dirigido o documentário *Vaxxed*, no qual insiste em disseminar os resultados do seu trabalho desacreditado.

Apesar das evidências incontestáveis de que sua pesquisa foi eticamente comprometida e metodologicamente falaciosa, em seu documentário Wakefield alega que os Centros de Controle e Prevenção de Doenças (CDC) – a agência mais importante de proteção à saúde dos Estados Unidos – teriam acobertado o suposto vínculo causal entre a vacina tríplice e o desenvolvimento do transtorno do espectro autista em crianças. O resultado de associar-se à narrativa antiestablishment trumpista foi uma renovada projeção midiática de Wakefield, consolidando-o como uma das principais vozes contemporâneas do movimento antivacina.

Essa reflexão histórica abrangente está intimamente ligada a uma compreensão eficaz dos discursos, pois permite considerar diversas perspectivas ao longo do tempo. Dessa forma, é possível identificar com maior clareza as continuidades e rupturas nas estruturas de conhecimento, poder e significado, assegurando que as evidências atuais não sejam tratadas como verdades absolutas e imutáveis (Magalhães; Kogawa, 2019). Tal abordagem propicia uma visão mais matizada e contextualizada dos fenômenos, reconhecendo a natureza dinâmica e multifacetada dos processos sociais e das construções de sentido que os permeiam.

Partindo dessa ponderação abrangente, é possível observar que os quatro discursos identificados nas dimensões – medo, desconfiança de figuras de autoridade, defesa da pseudociência e liberdade individual – se complementam e reforçam mutuamente, indicando uma narrativa argumentativa coesa por parte do movimento. Essa coesão cria um ambiente com pouco espaço para contradições e debates internos, gerando uma mensagem unificada. Nesse contexto, a natureza dialógica das dimensões fica ainda mais evidente, pois os discursos interagem com facilidade entre si. Embora um discurso seja o principal em uma dimensão, é fácil notar a presença dos outros orbitando o mesmo espaço conceitual. Um exemplo disso é o seguinte *tweet* do polo positivo da Dimensão 1:

Mais **jovens falecendo** por infartos e **males Súbitos**. É importante lembrar que a **pesquisa** científica da @CMAJ, **publicada** nesta **semana**, **indicou** q o **risco** d problemas **cardíacos** é d 100 p / cada 100mil **doses**. É duro saber que estes **morreram** p / q alguns se sentissem seguros. . . <https://t.co/xFSmndVjei>

O discurso em destaque, apontado pelas variáveis lexicais sublinhadas, é o medo dos efeitos colaterais da vacina. No entanto, é possível observar, em menor intensidade, o discurso a favor da pseudociência – quando cita estudos enviesados – e o da desconfiança – quando afirma que existem grupos (não especificados) se beneficiando dessas supostas mortes. Em um único enunciado concreto, pode-se constatar a característica mais relevante da variação multidimensional lexical descrita por Berber Sardinha (2021): cada texto é moldado simultaneamente pela incidência dos diversos discursos representados pelas dimensões.

Outro aspecto relevante da AMD Lexical reside na sua capacidade de apreciar as nuances presentes nos discursos. Tal fato se torna evidente quando duas dimensões convergem em torno de um mesmo fenômeno, a exemplo do que ocorre com o efeito colateral, identificado no polo positivo da Dimensão 1 e no polo negativo da Dimensão 3. Embora os agrupamentos de *tweets* compartilhem o mesmo discurso central – o medo acerca dos efeitos colaterais da vacina –, a AMD Lexical logrou captar e distinguir sutilezas entre eles. Quando os textos são agrupados por meio da coocorrência de variáveis lexicais, aspectos que transcendem aos temas superficiais são revelados. Além de identificar o discurso subjacente a um determinado tema, essa abordagem possibilita compreender as diferentes formas pelas quais ele é articulado e mobilizado em enunciados concretos.

A AMD Lexical mostrou que o receio acerca dos efeitos colaterais constitui um trunfo de tamanha relevância para o movimento antivacina que tem sido articulado e mobilizado de diferentes formas, visando atingir públicos com distintos níveis de adesão. No contexto específico do Twitter, durante a pandemia de Covid-19, o discurso do medo assumiu a forma de alertas sobre supostas reações adversas das vacinas e relatos pessoais descrevendo experiências negativas associadas à vacinação.

Em contraste, a Análise de Modelagem de Tópicos, por exemplo, amplamente empregada em estudos que investigam *big data*, poderia apenas identificar o tópico no corpus, o efeito colateral. Assim, essa habilidade de discernir matizes discursivos



se reveste de grande importância para uma análise aprofundada e refinada de dados em grandes volumes textuais, ultrapassando uma visão superficial dos fenômenos linguísticos.

Outro fato expressivo revelado pela análise dos resultados é a ausência de elementos extralinguísticos entre as variáveis, com exceção de um único *emoji* identificado como variável lexical na Dimensão 1. Embora os *emojis* sejam frequentes na comunicação digital, desempenhando diversas funções junto à linguagem e a outros recursos semióticos (Zappavigna; Logi, 2023), é possível supor que o movimento antivacina opte por não utilizá-los amplamente no Twitter, mantendo um tom de seriedade nas postagens. Conforme observado, as mensagens têm a vocação de apresentar “a verdade” supostamente velada e oculta ao leitor/ouvinte, portanto, o uso excessivo de *emojis* poderia criar uma dissonância cognitiva, tornando o conteúdo menos sério e crível. Assim, a relativa ausência desses recursos pode ser interpretada como uma estratégia discursiva intencional, conferindo um tom mais sóbrio e assertivo às mensagens, alinhado à proposta de revelar verdades negligenciadas.

A ausência de *hashtags* é ainda mais intrigante, visto que elas assumem a função de vinculação social e cultivo de comunidades organizadas em torno de valores e crenças específicas no espaço dinâmico do discurso das mídias sociais (Zappavigna, 2018). Uma possível explicação pode ser encontrada na própria estrutura do Twitter. Em 2015, a plataforma divulgou um relatório centrado em anúncios de resposta direta, os quais visam gerar resultados específicos, como a instalação de um aplicativo ou a visita a um *site*. No entanto, constatou-se que, quando esses anúncios incluíam uma *hashtag*, seu desempenho era insatisfatório.

Embora o estudo tenha sido elaborado para anunciantes, suas descobertas também oferecem um conselho prático para qualquer tipo de *tweet* projetado para alcançar um resultado específico, como a interação com um *link*. Anne Mercogliano, então responsável pelo *marketing* para pequenas e médias empresas no Twitter, sugeriu que outras partes clicáveis de um *tweet* podem distrair os usuários do objetivo principal pretendido (Southern, 2015). Dada a frequência de *links* clicáveis em muitas das postagens analisadas, a ausência de *hashtags* pode ser explicada como uma indicação do profundo entendimento do movimento antivacina sobre o funcionamento da plataforma. Conforme observado por Pfeffer *et al.* (2023), *tweets* desprovidos de *hashtags* podem revelar uma dinâmica distinta no discurso do Twitter. Eles propiciam discussões mais pessoais sobre notícias, política e esportes, ampliando a

compreensão das interações interpessoais na plataforma.

Essa constatação direciona nossa atenção para as questões estruturais inerentes ao Twitter. Ao considerar a plataforma como um registro único neste estudo, levamos em conta todas as suas idiossincrasias e coerções na análise dos dados. A desinformação surge como uma das diversas manifestações de desonestidade *on-line*, cujo aumento pode ser compreendido ao examinarmos as características gerais da interação social na internet e as *affordances* oferecidas por plataformas como o Twitter (Haigh; Haigh, 2020).

As *affordances* das plataformas de mídia social referem-se às possibilidades percebidas e concretizadas para ação, relação e construção de significado que emergem por meio das interações dinâmicas e dos engajamentos entre o ambiente sociotécnico de uma plataforma e seus múltiplos usuários/*stakeholders*. Essa definição abrange as diferentes interfaces e pontos de acesso, como APIs, *sites* e aplicativos, por meio dos quais diversas entidades experimentam e interagem com o ambiente da plataforma. Essencialmente, refletem a natureza dinâmica, adaptativa e personalizada dessas redes, cada vez mais impulsionadas por algoritmos e dados dos usuários. Por fim, destacam o caráter multidirecional e relacional, pelo qual usuários e plataformas moldam e coconstroem mutuamente o que uma rede oferece por meio de suas interações em todo o seu ecossistema.

No Twitter, as *affordances* exercem um papel importante na facilitação e na escalada da disseminação de desinformação. A própria natureza dessa plataforma digital, concebida para otimizar a propagação de conteúdo de maneira viral e em larga escala, é um fator-chave a ser considerado. Os mecanismos de compartilhamento e viralização intrínsecos a essa rede social podem fomentar a difusão exponencial de desinformação. Postagens instigantes, sensacionalistas ou polêmicas tendem a ser amplamente disseminadas, independentemente de sua veracidade, devido à dinâmica de engajamento e às lógicas que governam esse ambiente digital. Desse modo, mesmo sendo minoritárias, as postagens antivacina ganharam destaque e relevância no Twitter durante a pandemia.

Essa dinâmica de viralização de desinformação está intimamente relacionada a um fenômeno mais amplo observado nas redes sociais contemporâneas. Embora não tenham sido explicitamente concebidas como tal, devido às suas próprias características intrínsecas (*affordances*), essas plataformas acabam por criar ecossistemas autoritários não declarados. Esse fenômeno se manifesta na ausência

de um diálogo verdadeiramente democrático, uma vez que as interações se restringem àqueles que compartilham visões e opiniões similares, enquanto vozes discordantes são frequentemente alvo de ataques, cancelamentos e exclusões. Assim como ocorre em dinâmicas autoritárias tradicionais, em que o diálogo é suprimido, prejudicando a construção e a disseminação do conhecimento, essas plataformas dificultam a troca autêntica de ideias e perspectivas distintas.

Sob essa ótica, o Twitter, assim como outras redes sociais, apresenta estruturas nas quais a diversidade de pensamento é suprimida em favor de uma estreita conformidade ideológica. O paradoxo emerge quando tais plataformas são promovidas como espaços de livre expressão sem precedentes, não obstante perpetuarem dinâmicas estruturais que dificultam o diálogo e a negociação de ideias divergentes. Os algoritmos de recomendação empregados, que personalizam os *feeds* baseados nos interesses e comportamentos dos usuários, podem inadvertidamente amplificar a circulação de desinformação, reforçando as câmaras de eco e a polarização.

A análise criteriosa dos dados evidenciou que a relutância em aderir aos programas de vacinação tem sido impulsionada por uma multiplicidade de fatores, que vão desde preocupações legítimas fundamentadas em experiências adversas até teorias conspiratórias destituídas de qualquer embasamento científico. Dentre esses fatores, destacam-se os temores relacionados aos potenciais efeitos colaterais indesejados das vacinas e a objeção à intervenção governamental sobre decisões individuais. Adicionalmente, outras narrativas, como alegações de motivações financeiras por parte da indústria farmacêutica e de governos, também têm sido articuladas como justificativas para a rejeição à imunização.

A identificação desses discursos em corpora de larga escala e sua análise crítica são fundamentais para aprimorar nossa compreensão sobre como a linguagem é instrumentalizada para construir realidades e moldar comportamentos sociais. Essa abordagem pode fornecer dados elucidativos sobre a evolução temporal de discursos específicos, o modo como eles são influenciados por diferentes fatores socioeconômicos e culturais e como podem ser desafiados e desconstruídos por novas ideias e iniciativas.

O presente estudo acredita ter alcançado um grau satisfatório de triangulação teórico-metodológica ao combinar a LC com duas teorias discursivas complementares: a ACD, pela sua ênfase no desvelamento das ideologias implícitas,

e a ADD. Esta última estabeleceu um diálogo profícuo com a AMD Lexical, concebendo os *tweets* analisados como enunciados concretos e, assim, fornecendo uma ferramenta eficaz para a análise dos resultados obtidos. Além disso, a ADD salientou a importância das cadeias discursivas e reforçou a natureza dialógica das dimensões.

Ademais, esta pesquisa proporcionou uma nova perspectiva sobre os discursos antivacina no Brasil ao identificá-los e analisá-los de forma integrada, compreendendo como eles interagem e se reforçam mutuamente. Por fim, apresentou uma contribuição significativa ao reconhecer o papel preponderante exercido pelas redes sociais, com especial ênfase no Twitter, como plataformas cuja arquitetura e dinâmicas intrínsecas tendem a privilegiar a disseminação sistemática de desinformação. Portanto, com este estudo, esperamos oferecer perspectivas embasadas em evidências empíricas que possam subsidiar formuladores de estratégias comunicacionais voltadas ao enfrentamento do fenômeno da hesitação vacinal.

Todavia, é imperativo reconhecer que esta pesquisa apresenta algumas limitações inerentes. Ao centrar nossa análise exclusivamente na plataforma Twitter, reconhecemos o risco de que outros discursos relevantes, porventura existentes em outras redes sociais, não tenham sido contemplados. Exemplos potenciais incluem discussões sobre a autonomia parental no cuidado com os filhos ou questionamentos sobre a intrusão ocidental (ou neocolonial) em comunidades tradicionais. Essa lacuna abre espaço para futuras investigações que explorem outros ambientes do ecossistema digital, a fim de obter uma compreensão mais abrangente do fenômeno em questão.

Além disso, o recorte temporal adotado, focado no período pandêmico, foi uma decisão norteadada pelos objetivos específicos do programa de pesquisa concebido pela CAPES. Contudo, as evoluções dos discursos do movimento antivacina ao longo de diferentes linhas temporais merecem ser exploradas em pesquisas subsequentes.

Ao reconhecer essas limitações, este estudo não só reforça a necessidade de uma investigação contínua e aprofundada sobre o tema, como também abre caminhos para futuras pesquisas complementares que possam preencher as lacunas identificadas e enriquecer ainda mais o entendimento sobre os complexos fenômenos discursivos e ideológicos relacionados à hesitação vacinal.

## Referências

- AARONOVITCH, D. **Voodoo histories**: How conspiracy theory has shaped modern history. New York: Random House, 2010.
- AMIN, A. B. *et al.* Association of moral values with vaccine hesitancy. **Nature Human Behaviour**, v. 1, n. 12, p. 873-880, 2017.
- ANDRADE, T.; DOURADO, I. Vacinação despenca e aponta crise na cobertura vacinal em 2023. **Correio Braziliense**, 5 dez. 2022. Disponível em: <https://www.correiobraziliense.com.br/brasil/2022/12/5056707-vacinacao-despenca-e-aponta-crise-na-cobertura-vacinal-em-2023.html>. Acesso em: 21 mar. 2024.
- ARAÚJO, R. F. de; BERBER SARDINHA, T.; DELFINO, M. C. N. Revista Brasileira de Linguística Aplicada: multidimensões temáticas. *In*: FINATTOO, M. J. B. *et al.* (ed.). **Linguística de corpus**: perspectivas. Porto Alegre: Editora da UFRGS e do Instituto de Letras, 2018. p. 93-125.
- BAKER, P. **Sociolinguistics and Corpus Linguistics**. Edinburgh: Edinburgh University Press, 2010.
- BAKER, P. **Using Corpora in Discourse Analysis**. 2. ed. London: Bloomsbury Academic, 2023.
- BAKER, P.; EGBERT, J. Do all roads lead to Rome? Modeling Register Variation with Factor Analysis and Discriminant Analysis. *In*: BAKER, P.; EGBERT, J. (ed.). **Triangulating methodological approaches in corpus linguistic research**. New York; London: Routledge, 2016. p. 14-32.
- BAKER, P.; MCENERY, T. Introduction. *In*: BAKER, P.; MCENERY, T. (ed.). **Corpora and discourse studies**: Integrating discourse and corpora. Palgrave Macmillan: Basingstoke, 2015. p. 1-20.
- BAKHTIN, M. **O Freudismo**: um esboço crítico. São Paulo: Perspectiva, 2012.
- BAKHTIN, M. O discurso em Dostoiévski. *In*: BAKHTIN, M. **Problemas da poética de Dostoiévski**. 5. ed. Tradução direta do russo, notas e prefácio de Paulo Bezerra. Rio de Janeiro: Forense Universitária, 2013. p. 207-310.
- BAKHTIN, M. **Os gêneros do discurso**. São Paulo: Editora 34, 2016.
- BARBOZA, A. P. V. **“Desce do salto e vai viver”**: uma análise interdiscursiva do movimento #kuToo. 2021. 127 f. Dissertação (Mestrado em Linguística Aplicada e Estudos da Linguagem) – Faculdade de Filosofia, Comunicação, Letras e Artes, Pontifícia Universidade Católica de São Paulo, São Paulo, 2021.
- BARSKEY, A. E. *et al.* Mumps outbreak in Orthodox Jewish communities in the United States. **New England Journal of Medicine**, v. 367, n. 18, p. 1704-1713, 2012.

BAUMGAERTNER, B.; CARLISLE, J. E.; JUSTWAN, F. The influence of political ideology and trust on willingness to vaccinate. **PLoS ONE**, v. 13, n. 1, e0191728, 2018.

BERBER SARDINHA, T. Processamento Computacional do Português. *In*: INTERCÂMBIO DE PESQUISAS EM LINGUÍSTICA APLICADA, 9., 1999, São Paulo. **Anais** [...]. São Paulo: PUC-SP, 1999.

BERBER SARDINHA, T. Linguística de Corpus: histórico e problemática. **DELTA**, v. 16, n. 2, p. 323-367, 2000.

BERBER SARDINHA, T. Linguística de Corpus: histórico e conceitos básicos. **DELTA**, v. 18, n. 2, p. 323-352, 2002.

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

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

BERBER SARDINHA, T. Corpus linguistics and historiography: Finding the major discourses in the first 50 years of TESOL Quarterly. **Journal of Research Design & Statistics in Linguistics & Communication Science**, v. 7, n. 1, p. 69, 2016.

BERBER SARDINHA, T. Lexical priming: Advances and applications. *In*: BERBER SARDINHA, T. **Lexical priming and register variation**. Amsterdam: John Benjamins Publishing Company, 2017. p. 190-230.

BERBER SARDINHA, T. Dimensions of variation across Internet registers. **International Journal of Corpus Linguistics**, v. 23, n. 2, p. 125-157, 2018.

BERBER SARDINHA, T. A historical characterisation of American and Brazilian cultures based on lexical representations. **Corpora**, v. 15, n. 2, 2020.

BERBER SARDINHA, T. Discourse of academia from a multi-dimensional perspective. *In*: FRIGINAL, E.; HARDY, J. (ed.). **The Routledge Handbook of Corpus Approaches to Discourse Analysis**. Abingdon: Routledge. 2021. p. 298-328.

BERBER SARDINHA, T.; FITZSIMMONS-DOOLAN, S. **Lexical Multidimensional Analysis**. Cambridge Elements in Corpus Linguistics. (no prelo).

BERBER SARDINHA, T.; MOREIRA, M. M. F. P. **Deus, Pátria e família: os discursos bolsonaristas na rede social Twitter**. 2023. 23 f. Relatório de Trabalho de Conclusão de Curso (Graduação em Linguística Aplicada e Estudos da Linguagem), Pontifícia Universidade Católica de São Paulo, São Paulo, 2023.

BERBER SARDINHA, T.; VEIRANO PINTO, M. (ed.). **Multi-Dimensional Analysis: Research methods and current issues**. London: Bloomsbury Academic, 2019.

BERMAN, J. M. **Anti-vaxxers: How to challenge a misinformed movement**. Cambridge: MIT Press, 2020.

BIBER, D. **A model of textual relations within the written and spoken modes**. 1984. PhD Thesis – University of Southern California, Los Angeles, 1984.

BIBER, D. **Variation across speech and writing**. Cambridge: Cambridge University Press, 1988.

BIBER, D. Representativeness in Corpus Design. **Literary and Linguistic Computing**, v. 8, n. 4, p. 243-257, 1993.

BIBER, D. **Dimensions of Register Variation – A Cross-Linguistic Comparison**. Cambridge: Cambridge University Press, 1995.

BIBER, D. **University language**: A corpus-based study of spoken and written registers. Amsterdam: John Benjamins, 2006.

BIBER, D. What can a corpus tell us about registers and genres? *In*: O'KEEFFE, A.; MCCARTHY, M. (ed.). **The Routledge Handbook of Corpus Linguistics**. Nova York: Routledge, 2010. p. 241-254.

BIBER, D. Register as a predictor of linguistic variation. **Corpus Linguistics and Linguistic Theory**, v. 8, n. 1, p. 9-37, 2012.

BIBER, D.; CONRAD, S. **Register, genre, and style**. Cambridge: Cambridge University Press, 2009. (Cambridge textbooks in linguistics).

BIBER, D.; CONRAD, S.; REPPEN, R. **Corpus linguistics**: Investigating language structure and use. Cambridge: Cambridge University Press, 1998.

BIBER, D.; EGBERT, J. Register variation on the searchable web: a multidimensional analysis. **Journal of English Linguistics**, v. 44, n. 2, p. 95-137, 2016.

BIBER, D.; EGBERT, J.; KELLER, D. Reconceptualizing register in a continuous situational space. **Corpus Linguistics and Linguistic Theory**, v. 16, n. 3, p. 581-616, 2020.

BLOMMAERT, J.; VERSCHUEREN, J. **The Pragmatics of International and Intercultural Communication**. Amsterdam: John Benjamins Publishing, 1999.

BOAS, F. Introduction. **Handbook of American Indian Languages**, v. 1, p. 1-83, 1911.

BOGHOSSIAN, P. **Fear of knowledge**: Against relativism and constructivism. Oxford: Clarendon Press, 2007.

BOSELEY, S. How disgraced anti-vaxxer Andrew Wakefield was embraced by Trump's America. **The Guardian**, 18 jul. 2018. Disponível em: <https://www.theguardian.com/society/2018/jul/18/how-disgraced-anti-vaxxer-andrew-wakefield-was-embraced-by-trumps-america>. Acesso em: 10 jun. 2024.

BOULIANNE S.; KOC-MICHALSKA, K.; BIMBER, B. Right-wing populism, social media and echo chambers in western democracies. **New Media & Society**, v. 22, n. 4, p. 683-699, 2020.

BRAIT, B. (org.). **Bakhtin: conceitos-chave**. São Paulo: Contexto, 2005.

BRASIL reverte tendência de queda nas coberturas vacinais e oito imunizantes do calendário infantil registram alta em 2023. **Fiocruz**, 20 dez. 2023a. Disponível em: <https://www.canalsaude.fiocruz.br/noticias/noticiaAberta/brasil-reverte-tendencia-de-queda-nas-coberturas-vacinais-e-oito-imunizantes-do-calendario-infantil-registram-alta-em-202320122023>. Acesso em: 21 mar. 2024.

BRASIL tem 16 milhões de vacinados com dose bivalente contra covid-19. **Agência Brasil**, 12 maio 2023b. Disponível em: <https://agenciabrasil.ebc.com.br/saude/noticia/2023-05/brasil-tem-16-milhoes-de-vacinados-com-dose-bivalente-contra-covid-19>. Acesso em: 21 mar. 2024.

BRASIL. MINISTÉRIO DA SAÚDE. Covid-19 no Brasil – dados até 16/03/2024. **Ministério da Saúde**, 2024. Disponível em: [https://infoms.saude.gov.br/extensions/covid-19\\_html/covid-19\\_html.html](https://infoms.saude.gov.br/extensions/covid-19_html/covid-19_html.html). Acesso em: 21 mar. 2024.

BROGINI, A. A. **Representações contemporâneas da sustentabilidade**: uma análise multidimensional lexical discursiva como contribuição para o portal multimodal/multilíngue. 2023. 113 f. Dissertação (Mestrado em Linguística Aplicada e Estudos da Linguagem) – Programa de Pós-Graduação em Linguística Aplicada e Estudos da Linguagem, Pontifícia Universidade Católica de São Paulo, São Paulo, 2023.

BROWN, A. L. *et al.* Vaccine confidence and hesitancy in Brazil. **Cadernos de Saúde Pública**, v. 34, n. 9, e00011618, 2018.

BROWN, P.; FRASER, C. Speech as a marker of situation. *In*: RAINER, K.; GILES, S. (ed.). **Social markers in speech**. Cambridge: Cambridge University Press, 1979. p. 33-62.

CALDAS-COULTHARD, C. From Discourse Analysis to Critical Discourse Analysis: The differential re-representation of women and men speaking in written News. *In*: SINCLAIR, J.; HOEY, M.; FOX, G. (ed.). **Techniques of Description – Spoken and Written Discourse**. London: Routledge, 1993. p. 196-208.

CALDAS-COULTHARD, C. Man in the news: the misrepresentation of women speaking in news as narrative discourse. *In*: MILLS, S. (ed.). **Language and Gender: Interdisciplinary Perspectives**. London: Longman, 1995. p. 226-239.

CANZIAN, F.; CANCIAN, N. Bolsonaro recusou vacina a 50 do valor pago por EUA e União Europeia. **Folha de S.Paulo**, 6 jun. 2021. Disponível em: <https://www1.folha.uol.com.br/equilibrioesaude/2021/06/bolsonaro-recusou-vacina-a-50-do-valor-pago-por-eua-e-uniao-europeia.shtml>. Acesso em: 21 mar. 2024.



CAPLE, H.; HUAN, C.; BEDNAREK, M. **Multimodal News Analysis across Cultures**. Cambridge: Cambridge University Press, 2020.

CASARÕES, G.; MAGALHÃES, D. A aliança da hidroxicloroquina: como líderes de extrema direita e pregadores da ciência alternativa se reuniram para promover uma droga milagrosa. **Revista de Administração Pública**, Rio de Janeiro, v. 55, n. 1, p. 197-214, 2021.

CASEROTTI, M. *et al.* Associations of COVID-19 risk perception with vaccine hesitancy over time for Italian residents. **Social Science & Medicine**, v. 272, 113688, 2021.

CELANI, M. A. A. Afinal, o que é Lingüística Aplicada? *In*: PASCHOAL, M. S. Z.; CELANI, M. A. A. (org.). **Lingüística Aplicada: da Aplicação da Lingüística à Lingüística Transdisciplinar**. São Paulo: Educ, 1992. p. 15-23.

CEYLAN, G.; ANDERSON, I. A.; WOOD, W. Sharing of misinformation is habitual, not just lazy or biased. **PNAS**, v. 120, n. 4, e2216614120, 2022.

CHALHOUB, S. **Cidade febril: cortiços e epidemias na Corte Imperial**. São Paulo: Companhia das Letras, 1996.

CHOULIARAKI, L.; FAIRCLOUGH, N. **Discourse in Late Modernity**. Edinburgh: Edinburgh University Press, 1999.

CLARKE, I. **Linguistic variation across Twitter and Twitter trolling**. 2020. Unpublished doctoral dissertation – University of Birmingham, Birmingham, 2020.

CLARKE, I. A Multi-Dimensional Analysis of English Tweets. **Language and Literature: International Journal of Stylistics**, v. 31, n. 2, p. 124-149, 2022.

CODATO, A.; BERLATTO, F. BOLOGNESI, B. Tipologia dos políticos de direita no Brasil: uma classificação empírica. **Análise Social**, v. 229, n. 4, p. 870-897, 2018.

COHN, S. K. Cholera revolts: a class struggle we may not like. **Social History**, v. 42, n. 2, p. 162-180, 2017.

DE BEAUGRANDE, R.; DRESSLER, W. **Introduction to Text Linguistics**. London: Longman, 1981.

DEAN, B. Social Media Usage & Growth Statistics. **Backlinko**, 2023. Disponível em: <https://backlinko.com/social-media-users>. Acesso em: 21 mar. 2024.

DELFINO, M. C. Análise multidimensional: os números na linguística. **Cadernos de Linguística**, v. 2, n. 4, e474, 2021.

DELUMEAU, J. **História do medo no ocidente**. São Paulo: Companhia das Letras, 2009.

DIMITROVA, Blaga. **Because the sea is black: poems of Blaga Dimitrova**. Middletown: Wesleyan University Press, 1989.

DOBSON, R. Media misled the public over the MMR vaccine, study says. **BMJ**, v. 326, n. 7399, p. 1107, 2003.

DOUGLAS, K. M. *et al.* Understanding conspiracy theories. **Political Psychology**, v. 40, n. 1, p. 3-35, 2019.

DURBACH, N. They might as well brand us: Working class resistance to compulsory vaccination in Victorian England. **The Society for the Social History of Medicine**, v. 13, p. 45-62, 2000.

EGBERT, J. Corpus Design and Representativeness. *In*: BERBER SARDINHA, T.; VEIRANO PINTO, M. (org.). **Multi-Dimensional Analysis: Research Methods and Current Issues**. London: Bloomsbury, 2019. p. 27-42.

EGBERT, J.; BIBER, D.; GRAY, B. **Designing and Evaluating Language Corpora – A Practical Framework for Corpus Representativeness**. Cambridge: Cambridge University Press, 2022.

ERIKSSON, K.; VARTANOVA, I. Vaccine confidence is higher in more religious countries. **Human Vaccines & Immunotherapeutics**, v. 18, n. 1, p. 1-3, 2022.

FAIRCLOUGH, N. **Language and Power**. London: Longman, 1989.

FAIRCLOUGH, N.; KRESS, G. Critical discourse analysis. *In*: VAN DIJK, T. (ed.). **Discourse & Society**. Thousand Oaks: Sage, 1993.

FAIRCLOUGH, N.; WODAK, R. Critical discourse analysis. *In*: VAN DIJK, T. (ed.). **Discourse Studies: A Multidisciplinary Introduction**. Thousand Oaks: Sage, 1997. v. 2. p. 258-284.

FASCE, A. *et al.* A taxonomy of anti-vaccination arguments from a systematic literature review and text modelling. **Nature Human Behaviour**, v. 7, n. 9, p. 1462-1480, 2023.

FEIJÓ, R. B.; SÁFADI, M. A. P. Imunizações: três séculos de uma história de sucessos e constantes desafios. **Jornal de Pediatria**, Rio de Janeiro, v. 82, n. 3, s1-3, 2006.

FIGUEIREDO, A. *et al.* Mapping global trends in vaccine confidence and investigating barriers to vaccine uptake: a large-scale retrospective temporal modelling study. **Lancet**, v. 26, p. 898-908, 2020.

FINCHELSTEIN, F. **From fascism to populism in history**. Berkeley: University of California Press, 2019.

FINKELSTEIN, S. R. *et al.* Psychological reactance impacts ratings of pediatrician vaccine-related communication quality, perceived vaccine safety, and vaccination priority among US parents. **Human Vaccines & Immunotherapeutics**, v. 16, n. 5, p. 1024-1029, 2020.

FIORIN, J. L. **Em busca do sentido: estudos discursivos**. São Paulo: Contexto, 2008.

FITZSIMMONS-DOOLAN, S. Using lexical variables to identify language ideologies in a policy corpus. **Corpora**, v. 9, n. 1, p. 57-82, 2014.

FLOWERDEW, J. The discourse of colonial withdrawal: a case study in the creation of mythic discourse. **Discourse and Society**, v. 8, n. 4, p. 453-477, 1997.

FOUCAULT, M. **The Archaeology of Knowledge**. London: Tavistock, 1972.

FOUCAULT, M.; FAUBION, J. (ed.). **The Essential Works of Michel Foucault 1954–1988: Power**. New York: New York Press, 2000.

FOWLER, R. **Linguistic Criticism**. Oxford: Oxford University Press, 1996.

FOWLER, R. *et al.* **Language and Control**. London: Routledge, 1979.

FRIGINAL, E. Twenty-five years of Biber's Multi-Dimensional Analysis: Introduction to the special issue and an interview with Douglas Biber. **Corpora**, v. 8, n. 2, p. 137-152, 2013.

FRIGINAL, E.; HARDY, J. (ed.). **Routledge Handbook of Corpus Approaches to Discourse Analysis**. London: Routledge, 2020.

FRUGOLI, A. G. *et al.* Fake News sobre vacinas: uma análise sob o modelo dos 3Cs da Organização Mundial da Saúde. **Revista da Escola de Enfermagem da USP**, São Paulo, 55, e03736, 2021.

FUNDO DAS NAÇÕES UNIDAS PARA A INFÂNCIA – UNICEF. **Estudo qualitativo sobre os fatores econômicos, culturais e políticas de saúde relacionados à redução das coberturas vacinais de rotina em crianças menores de cinco anos**. Brasília: Unicef, 2020. Disponível em: <https://www.unicef.org/brazil/media/11001/file/estudo-fatores-relacionados-reducao-coberturas-vaciniais-de-rotina-em-criancas-menores-5-anos.pdf>. Acesso em: 21 mar. 2024.

FUNDO DAS NAÇÕES UNIDAS PARA A INFÂNCIA – UNICEF. Vacinas. **Unicef Brasil**, [s. d.]. Disponível em: <https://www.unicef.org/brazil/vacinas-perguntas-e-respostas>. Acesso em: 21 mar. 2024.

FUNG, T. K. F.; NAMKOONG, K.; BROSSARD, D. Media, social proximity, and risk: a comparative analysis of news- paper coverage of avian flu in Hong Kong and in the United States. **Journal of Health Communication**, v. 16, n. 8, p. 889-907, 2011.

GONZÁLEZ-BLOCK, M. Á. *et al.* Influenza vaccination hesitancy in five countries of South America. Confidence, complacency and convenience as determinants of immunization rates. **PloS ONE**, v. 15, n. 12, e0243833, 2020.

GOREIS, A.; VORACEK, M. A systematic review and meta-analysis of psychological research on conspiracy beliefs: Field characteristics, measurement instruments, and associations with personality traits. **Frontiers in Psychology**, v. 10, p. 205, 2019.

GOSTIN, L. Jacobson vs. Massachusetts at 100 years: Police powers and civil liberties in tension. **American Journal of Public Health**, v. 95, n. 4, p. 576-81, 2005.

GOULART, L.; BIBER, D.; REPPEN, R. In this essay, I will...: Examining variation of communicative purpose in student written genres. **Journal of English for Academic Purposes**, v. 59, p. 101-159, 2022.

GRABENSTEIN, J. D. What the world's religions teach, applied to vaccines and immune globulins. **Vaccine**, v. 31, n. 16, p. 2011-2023, 2013.

GRAY, B.; EGBERT, J. Register and register variation. *In*: GRAY, B.; EGBERT, J. (ed.). **Register Studies**. Amsterdam: John Benjamins, 2019. p. 1-9.

GUIARRARA, P. Pandemia de Covid-19. **Brasil Escola**, [s. d.]. Disponível em: <https://brasilecola.uol.com.br/geografia/pandemia-de-covid-19.htm>. Acesso em: 21 mar. 2024.

GUTTINGER, S. The anti-vaccination debate and the microbiome: How paradigm shifts in the life sciences create new challenges for the vaccination debate. **EMBO Reports**, v. 20, n. 3, e47709, 2019.

HAIGH, M.; HAIGH, T. Fighting and framing fake news. *In*: BAINES, P.; O'SHAUGHNESSY, N.; SNOW, N. (ed.). **The Sage Handbook of Propaganda**. Thousand Oaks: Sage, 2020. p. 303-323.

HAJER, M. Discourse coalitions and the institutionalization of practice. *In*: FISCHER, F.; FORESTER, J. (ed.). **The argumentative turn in policy analysis and planning**. Durham: Duke University Press, 1993. p. 43-76.

HALLIDAY, M. A. K. **Language as Social Semiotic: The Social Interpretation of Language and Meaning**. [S. l.]: Edward Arnold, 1978.

HAMMERSLEY, M. Conversation analysis and discourse analysis: Methods or paradigms? **Discourse & Society**, v. 14, n. 6, p. 751-781, 2003.

HANRATTY, B. *et al.* UK measles outbreak in non-immune anthroposophic communities: The implications for the elimination of measles from Europe. **Epidemiology & Infection**, v. 125, n. 2, p. 377-383, 2000.

HARAKI, C. A. C. Estratégias adotadas na América do Sul para a gestão da infodemia da COVID-19. **Revista panamericana de salud publica**, v. 45, e43, 2021.

HARDT-MAUTNER, G. **"Only Connect": Critical Discourse Analysis and Corpus Linguistics**. Lancaster: University of Lancaster, 1995.

HOLMES, J. A corpus based view of gender in New Zealand English. *In*: HELLINGER, M.; BUSSMAN, H. (ed.). **Gender Across Languages – The Linguistic Representation of Women and Men**. Amsterdam: John Benjamins, 2001. v. 1. p. 115-136.

HORNSEY, M. J.; FIELDING, K. S. Attitude roots and Jiu Jitsu persuasion: Understanding and overcoming the motivated rejection of science. **American Psychologist**, v. 72, n. 5, p. 459-473, 2017.

HORNSEY, M. J.; HARRIS, E. A.; FIELDING, K. S. The psychological roots of anti-vaccination attitudes: A 24-nation investigation. **Health Psychology**, v. 37, n. 4, p. 307-315, 2018.

IDOETA, P. A. A história de Bolsonaro com a hidroxicloroquina em 6 pontos: de tuítes de Trump à CPI da Covid. **BBC News Brasil**, 21 maio 2021. Disponível em: <https://www.bbc.com/portuguese/brasil-57166743>. Acesso em: 21 mar. 2024.

INSTITUTO BUTANTAN. O mundo antes e depois das vacinas: a história comprova que o caminho para a erradicação de doenças é a imunização. **Portal do Butantan**, 14 mar. 2022. Disponível em: <https://butantan.gov.br/noticias/o-mundo-antes-e-depois-das-vacinas-a-historia-comprova-que-o-caminho-para-a-erradicacao-de-doencas-e-a-imunizacao>. Acesso em: 21 mar. 2024.

INSTITUTO DE CIÊNCIA, TECNOLOGIA E QUALIDADE – ICTQ. Pesquisa – Autodiagnóstico Médico no Brasil. **ICTQ**, 2018. Disponível em: <https://ictq.com.br/pesquisa-do-ictq/786-pesquisa-autodiagnostico-medico-no-brasil-2018>. Acesso em: 21 mar. 2024.

JANG, H. *et al.* Tracking COVID-19 Discourse on Twitter in North America: Infodemiology Study Using Topic Modeling and Aspect-Based Sentiment Analysis. **Journal of Medical Internet Research**, v. 23, n. 2, e25431, 2021.

JOLLEY, D.; DOUGLAS, K. M. The Effects of Anti-Vaccine Conspiracy Theories on Vaccination Intentions. **PLoS ONE**, v. 9, n. 2, e89177, 2014.

JONES, R. (ed.). **Viral Discourse**. Cambridge: Cambridge University Press, 2021.

KARLSSON, L. C. *et al.* Fearing the disease or the vaccine: The case of COVID-19. **Personality and individual differences**, v. 172, 110590, 2021.

KATA, A. Anti-vaccine activists, Web 2.0, and the postmodern paradigm: An overview of tactics and tropes used online by the anti-vaccination movement. **Vaccine**, v. 30, n. 25, p. 3778-3789, 2012.

KAUFFMANN, C. H. **Linguística de corpus e estilo**: análises multidimensional e canônica na ficção de Machado de Assis. 2020. 276 f. Tese (Doutorado em Linguística Aplicada e Estudos da Linguagem) – Programa de Pós-Graduação em Linguística Aplicada e Estudos da Linguagem, Pontifícia Universidade Católica de São Paulo, São Paulo, 2020.

KEMPINSKA-MIROŚLAWSKA, B.; WOŹNIAK-KOSEK, A. The influenza epidemic of 1889–90 in selected European cities – a picture based on the reports of two Poznan daily newspapers from the second half of the nineteenth century. **Medical Science Monitor**, v. 19, p. 1131-1141, 2013.

KENNEDY, J. Populist politics and vaccine hesitancy in Western Europe: an analysis of national-level data. **European Journal of Public Health**, v. 29, n. 3, p. 512-516, 2019.

KILGO, D. K.; YOO, J.; JOHNSON, T. J. Spreading Ebola panic: newspaper and social media coverage of the 2014 Ebola health crisis. **Journal of Health Communication**, v. 34, n. 8, p. 811-817, 2019.

KLEMM, C.; DAS, E.; HARTMANN, T. Swine flu and hype: a systematic review of media dramatization of thenH1N1 influenza pandemic. **Journal of Risk Research**, v. 19, n. 1, p. 1-20, 2016.

KOLLER, V.; MAUTNER, G. Computer application in critical discourse analysis. *In*: COFFIN, C.; HEWINGS, A.; O'HALLORAN, K. (ed.). **Applying English Grammar: Functional and Corpus Approaches**. London: Arnold, 2004. p. 216-228.

KRESS, G. Critical discourse analysis. **Annual Review of Applied Linguistics**, v. 11, p. 84-99, 1990.

KRESS, G.; HODGE, R. **Language as Ideology**. London: Routledge & Kegan Paul, 1979.

KRISHNAMURTHY, R. Ethnic, Racial and Tribal: The language of racism? *In*: CALDAS-COULTHARD, C. R.; COULTHARD, M. (ed.). **Texts and Practices: Readings in Critical Discourse Analysis**. London: Routledge, 1996. p. 137-157.

KUSCH, M. **Relativism in the Philosophy of Science**. Cambridge: Cambridge University Press, 2021.

LARSON, P. Covid-19 anti-vaxxers use the same arguments from 135 years ago. **The Conversation**, 4 out. 2020. Disponível em: <https://theconversation.com/covid-19-anti-vaxxers-use-the-same-arguments-from-135-years-ago-145592>. Acesso em: 10 jun. 2024.

LASCO, G.; YU, V. G. Pharmaceutical messianism and the COVID-19 pandemic. **Social Science & Medicine**, v. 292, 114567, 2021.

LEECH, G.; FALLON, R. Computer corpora – what do they tell us about culture? **ICAME Journal**, v. 16, p. 1-22, 1992.

LEWANDOWSKY, S. *et al.* **The COVID-19 Vaccine communication handbook**. Bristol: University of Bristol, 2020.

LEWANDOWSKY, S. *et al.* When science becomes embroiled in conflict: Recognizing the public's need for debate while combating conspiracies and misinformation. **The Annals of the American Academy of Political and Social Science**, v. 700, n. 1, p. 26-40, 2022.

LEWANDOWSKY, S.; OBERAUER, K. Worldview-motivated rejection of science and the norms of science. **Cognition**, v. 215, 104820, 2021.

LINHA do tempo mostra os principais fatos da pandemia no Brasil. **O Globo**, 25 fev. 2021. Disponível em: <https://oglobo.globo.com/brasil/linha-do-tempo-mostra-os-principais-fatos-da-pandemia-no-brasil-24897725>. Acesso em: 21 mar. 2024.

MAC, R; HSU, T. From Twitter to X: Elon Musk Begins Erasing an Iconic Internet Brand. **The New York Times**, 24 jul. 2023. Disponível em: <https://www.nytimes.com/2023/07/24/technology/twitter-x-elon-musk.html>. Acesso em: 21 mar. 2024.

MACDONALD, N. E. SAGE Working Group on Vaccine Hesitancy. Vaccine hesitancy: definition, scope and determinants. **Vaccine**, v. 33, n. 34, 4161-4, 2015.

MAGALHÃES, A. S.; KOGAWA, J. **Pensadores da Análise do Discurso**: uma introdução. Jundiaí: Paco Editorial, 2019.

MALTHOUSE, E. Confirmation bias and vaccine-related beliefs in the time of COVID-19. **Journal of Public Health**, v. 45, n. 2, p. 523-528, 2023.

MAYER, C. **O que e como escrevemos na web**: um estudo multidimensional de variação de registro em língua inglesa. 2018. 126 f. Tese (Doutorado em Linguística Aplicada e Estudos da Linguagem) – Programa de Pós-Graduação em Linguística Aplicada e Estudos da Linguagem, Pontifícia Universidade Católica de São Paulo, São Paulo, 2018.

MCCARTHY, M.; O'KEEFFE, A. Historical Perspective: What are corpora and how have they evolved?. In: O'KEEFFE, A.; MCCARTHY, M. (ed.). **The Routledge Handbook of Corpus Linguistics**. Nova York: Routledge, 2010. p. 3-13.

MCENERY, T.; BAKER, P.; HARDIE, A. Swearing and abuse in modern British English. In: LEWANDOWSKA-TOMASZCZYK, B.; MELIA, J. (ed.). **PALC 99 Practical Applications in Language Corpora**. Hamburg: Peter Lang, 2000. p. 37-48.

MCENERY, T.; BREZINA, V. **Fundamental Principles of Corpus Linguistics**. Cambridge: Cambridge University Press, 2022.

MCENERY, T.; WILSON, A. **Corpus linguistics**. Edinburgh: Edinburgh University Press, 2001.

MCINTYRE, L. **Post-Truth**. Cambridge: MIT Press, 2018.

MINDRILA, D. Exploratory Factor Analysis: An Overview. In: MINDRILA, D. (ed.). **Exploratory Factor Analysis**: Application in School Improvement Research. Nova York: Nova Science Publishers, 2017.

MOITA LOPES, L. P. Pesquisa interpretativista em lingüística aplicada: a linguagem como condição e solução. **DELTA**, v. 10, n. 2, p. 329-338, 1994.

MOITA LOPES, L. P. (org.). **Por uma Linguística Aplicada INdisciplinar**. São Paulo: Parábola Editorial, 2006.



MOTTA, M. Republicans, not democrats, are more likely to endorse anti-vaccine misinformation. **American Politics Research**, v. 49, n. 5, p. 428-438, 2021.

MÜLLER, M.; SALATHÉ, M.; KUMMERVOLD, P. E. COVID-Twitter-BERT: A natural language processing model to analyse COVID-19 content on Twitter. **Frontiers in Artificial Intelligence**, v. 6, 2023.

NEWMAN, N. *et al.* **Reuters Institute Digital News Report 2019**. [S. l.]: Reuters Institute; University of Oxford, 2019. Disponível em: [https://reutersinstitute.politics.ox.ac.uk/sites/default/files/201906/DNR\\_2019\\_FINAL\\_1.pdf](https://reutersinstitute.politics.ox.ac.uk/sites/default/files/201906/DNR_2019_FINAL_1.pdf). Acesso em: 10 jun. 2024.

O'KEEFFE, A.; MCCARTHY, M.; CARTER, R. **From corpus to classroom: Language use and language teaching**. Cambridge: Cambridge University Press, 2007.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS PARA A EDUCAÇÃO, A CIÊNCIA E A CULTURA – UNESCO. **Alfabetização midiática e informacional: currículo para formação de professores**. Brasília: UNESCO; UFTM, 2013. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000220418>. Acesso em: 21 mar. 2024.

ORGANIZAÇÃO MUNDIAL DA SAÚDE – OMS. **Entenda a infodemia e a desinformação na luta contra a COVID-19**. [S. l.]: Organización Panamericana de la Salud, 2020. Disponível em: <https://bit.ly/3sANrxh>. Acesso em: 21 mar. 2024.

ORGANIZAÇÃO PAN-AMERICANA DA SAÚDE – OPAS. Histórico da pandemia de COVID-19. **OPAS**, [s. d.]. Disponível em: <https://www.paho.org/pt/covid19/historico-da-pandemia-covid-19>. Acesso em: 21 mar. 2024.

ORGANIZAÇÃO PAN-AMERICANA DA SAÚDE – OPAS. Trabalho para desenvolver vacina contra COVID-19 é mais rápido do que nunca, mas processos de segurança e eficácia permanecem os mesmos, afirma OPAS. **OPAS**, 21 out. 2020. Disponível em: <https://www.paho.org/pt/noticias/21-10-2020-trabalho-para-desenvolver-vacina-contracovid-19-e-mais-rapido-do-que-nunca-mas>. Acesso em: 21 mar. 2024.

PAPACHARISSI, Z. **A Networked Self: Identity, Community and Culture on Social Network Sites**. New York: Routledge, 2011.

PARK, J. **Anatomy of a Public Health Scare: Fear and Accountability in the Creation of Vaccine Courts**. 2020. 57 f. Senior Thesis – Barnard College, Department of History, 2020

PARTINGTON, A. **The Linguistics of Political Argument: The Spin-doctor and the Wolf-pack at the White House**. London: Routledge, 2003.

PFEFFER, J. *et al.* Just Another Day on Twitter: A Complete 24 Hours of Twitter Data. **ArXiv**, v. 2301, 11429, 2023.

PÊCHEUX, M. (ed.). **Language, semantics and ideology – stating the obvious**. London: Macmillan Press, 1982.



PENNYCOOK, A. A Linguística Aplicada dos anos 90: em defesa de uma abordagem crítica. In: SIGNORINI, I.; CAVALCANTI, M. C. **Linguística Aplicada e Transdisciplinaridade**: questões e perspectivas. Campinas: Mercado das Letras, 1998. p. 23-50.

PESSOA, C. Vacina contra covid-19 para crianças entra no Calendário de Vacinação. **Agência Brasil**, 1º jan. 2024. Disponível em: <https://agenciabrasil.ebc.com.br/radioagencia-nacional/saude/audio/2024-01/vacina-contracovid-19-para-criancas-entra-no-calendario-de-vacinacao>. Acesso em: 21 mar. 2024.

PIERRE, J. M. Mistrust and misinformation: A two-component, socio-epistemic model of belief in conspiracy theories. **Journal of Social and Political Psychology**, v. 8, n. 2, p. 617-641, 2020.

PRIMEIRO caso confirmado de Covid-19 no Brasil ocorreu em SP e completa seis meses nesta quarta. **G1 SP**, 26 ago. 2020. Disponível em: <https://g1.globo.com/sp/sao-paulo/noticia/2020/08/26/primeiro-caso-confirmado-de-covid-19-no-brasil-ocorreu-em-sp-e-completa-seis-meses-nesta-quarta.ghtml>. Acesso em: 21 mar. 2024.

QUADRI-SHERIFF, M. *et al.* The role of herd immunity in parents' decision to vaccinate children: A systematic review. **Pediatrics**, v. 130, n. 3, p. 522-30, 2012.

RAO, T. S.; ANDRADE, C. The MMR vaccine and autism: Sensation, refutation, retraction, and fraud. **Indian Journal of Psychiatry**, v. 53, n. 2, p. 95-96, 2011.

RECUERO, R.; STUMPF, E. M. Características do Discurso Desinformativo no Twitter: Estudo do discurso antivacinas do COVID-19. In: CAIADO, R. V. R.; LEFFA, V. J. (org.). **Linguagem**: tecnologia ensino. Campinas: Pontes Editores, 2021. p. 111-137.

REISIGL, M.; WODAK, R. **Discourse and Discrimination**: Rhetorics of Racism and Antisemitism. London: Routledge, 2001.

RELEMBRE ataques de Bolsonaro contra vacinas e veja como ele agora tenta esconder essas investidas. **Folha de S.Paulo**, 11 mar. 2021. Disponível em: <https://www1.folha.uol.com.br/poder/2021/03/relembre-ataques-de-bolsonaro-contravacinas-e-veja-como-ele-agora-tenta-esconder-essas-investidas.shtml>. Acesso em: 21 mar. 2024.

ROMEIRO, Y. M. de T. D. **A linguagem verbal das artes visuais**: uma análise multidimensional do discurso sobre a fotografia de Sally Mann. 2020. 182 f. Dissertação (Mestrado em Linguística Aplicada e Estudos da Linguagem) – Programa de Pós-Graduação em Linguística Aplicada e Estudos da Linguagem, Pontifícia Universidade Católica de São Paulo, São Paulo, 2020.

ROSSEN, I. *et al.* Accepters, fence sitters, or rejecters: Moral profiles of vaccination attitudes. **Social Science & Medicine**, v. 224, p. 23-27, 2019.

ROTHKOP, D. J. When the Buzz Bites Back. **The Washington Post**, 11 maio 2003. Disponível em: <https://www.washingtonpost.com/archive/opinions/2003/05/11/when-the-buzz-bites-back/bc8cd84f-cab6-4648-bf58-0277261af6cd/>. Acesso em: 21 mar. 2024.

RUBIN, G. J. *et al.* Public perceptions, anxiety, and behaviour change in relation to the Swine flu outbreak: cross sectional telephone survey. **BMJ**, v. 339, b2651, 2009.

SANCHEZ, A. Definicion e historia de los corpus. *In*: SANCHEZ, A. *et al.* (org.). **CUMBRE – Corpus Lingüístico de Español Contemporáneo**. Madrid: SGEL, 1995.

SCHMID, H. Probabilistic Part-of-Speech Tagging Using Decision Trees. **New Methods in Language Processing**, n. 154, p. 1-9, 2013.

SCHMIDT, J. Twitter and the Rise of Personal Publics. *In*: WELLER, K. *et al.* (ed.). **Twitter and Society**. New York: Peter Lang, 2014. p. 3-14.

SCOTT, M. Focusing on the text and its key words. *In*: INTERNATIONAL CONFERENCE ON TEACHING AND LANGUAGE CORPORA, 3., 2000, Frankfurt. **Proceedings** [...]. Frankfurt: Peter Lang, 2000.

SEMINO, E. *et al.* Narratives, Information and Manifestations of Resistance to Persuasion in Online Discussions of HPV Vaccination. **Health Communication**, v. 21, p. 1-12, 2023.

SEVCENKO, N. **A Revolta da Vacina**. São Paulo: Editora Scipione, 1993.

SHALDERS, A. “Tratamento precoce”: governo Bolsonaro gasta quase R\$ 90 milhões em remédios ineficazes, mas ainda não pagou Butantan por vacinas. **BBC News Brasil**, 21 jan. 2021. Disponível em: <https://www.bbc.com/portuguese/brasil-55747043>. Acesso em: 21 mar. 2024.

SINCLAIR, J. **A Course in Spoken English: Grammar**. London: Oxford University Press, 1972.

SINCLAIR, J. **Looking Up: An Account of the COBUILD Project in Lexical Computing**. London: Collins, 1987.

SINCLAIR, J. **Corpus, Concordance, Collocation**. Oxford: Oxford University Press, 1991.

SKAFLE, I. *et al.* Misinformation About COVID-19 Vaccines on Social Media: Rapid Review. **Journal of Medical Internet Research**, v. 24, n. 8, e37367, 2022.

SNOWDEN, F. M. **Epidemics and Society: From the Black Death to the Present**. New Haven: Yale University Press, 2019.

SOUTHERN, M. G. Tweets Without Hashtags or Mentions Get More Click-Throughs [STUDY]. **Search Engine Journal**, 2015.

SOUZA, B. T. de; LOGUERCIO, R. de Q. Plataformização da Verdade: Os Grupos Discursivos Sobre Vacinação Contra COVID-19 no Twitter. **Revista Brasileira de Pesquisa em Educação em Ciências**, e47798, p. 1-21, 2024.

SOVERI, A. *et al.* Trait reactance and trust in doctors as predictors of vaccination behavior, vaccine attitudes, and use of complementary and alternative medicine in parents of young children. **PLoS ONE**, v. 15, n. 7, e0236527, 2020.

SPRENGHOLZ, P. *et al.* Vaccination policy reactance: predictors, consequences, and countermeasures. **Journal of Health Psychology**, v. 27, n. 6, p. 1394-1407, 2022.

STUBBS, M. British traditions in text analysis. *In*: BAKER, M.; FRANCIS, G.; TOGNINI-BONELLI, E. (org.). **Text and Technology – In Honour of John Sinclair**. Philadelphia; Amsterdam: John Benjamins, 1993. p. 1-33.

SUCCI, R.C. M. Vaccine refusal: what we need to know. **The Journal of Pediatrics**, v. 94, n. 6, p. 574-581, 2018.

SVÄRD, R. *et al.* Recent anti-tuberculosis drug resistance levels in Honduras. **Transactions of The Royal Society of Tropical Medicine and Hygiene**, v. 94, n. 5, p. 544-546, 2020.

TAHA, S.; MATHESON, K.; ANISMAN, H. The 2009 H1N1 influenza pandemic: the role of threat, coping, and media trust on vaccination intentions in Canada. **Journal of Health Communication**, v. 18, p. 278-290, 2013.

TAMAKI, E. R.; FUKS, M. Populism in Brazil's 2018 General Elections: an Analysis of Bolsonaro's Campaign Speeches. **Lua Nova**, n. 109, p. 103-127, 2020.

TAYLOR, S. **The Psychology of Pandemics**: Preparing for the Next Global Outbreak of Infectious Disease. Newcastle upon Tyne: Cambridge Sch., 2019.

TAYLOR, S. The Psychology of Pandemics. **Annual Review of Clinical Psychology**, v. 18, p. 581-609, 2020.

TAYLOR, S. *et al.* COVID stress syndrome: concept, structure, and correlates. **Depression and Anxiety**, v. 37, p. 706-714, 2020.

TETAROLLI JÚNIOR, R. **Poder e saúde**: as epidemias e a formação dos serviços de saúde em São Paulo. São Paulo: Universidade Estadual Paulista, 1996.

THORNDIKE, E. L. The law of effect. **The American Journal of Psychology**, v. 39, n. 2/4, p. 212-222, 1927.

TITSCHER, S. *et al.* **Methods of Text and Discourse Analysis**. Thousand Oaks: Sage, 2000.

TOLEDO DIAS, Y. M. **A linguagem verbal das artes visuais**: uma análise multidimensional do discurso sobre a fotografia de Sally Mann. 2020. 182 f. Dissertação (Mestrado em Linguística Aplicada e Estudos da Linguagem) –

Programa de Estudos Pós- Graduação em Linguística Aplicada e Estudos da Linguagem, Pontifícia Universidade Católica de São Paulo, São Paulo, 2020.

TRAM, K. H. *et al.* Deliberation, dissent, and distrust: understanding distinct drivers of coronavirus disease 2019 vaccine hesitancy in the United States. **Clinical Infectious Diseases**, v. 74, n. 8, p. 1429-1441, 2022.

VAARA, E. Critical discourse analysis as methodology in strategy as practice research. *In*: GOLSORKHI, D. *et al.* (ed.). **Cambridge handbook of strategy as practice**. Cambridge: Cambridge University Press, 2010. p. 217-229.

VAN DIJK, T. A. (ed.). **Handbook of discourse analysis**. London; New York: Academic Press, 1985.

VAN DIJK, T. A. **Discourse and power**. Basingstoke: Palgrave Macmillan, 2008.

VAN PROOIJEN, J. W. An existential threat model of conspiracy theories. **European Psychologist**, v. 25, n. 1, p. 16-25, 2020.

VEIGA, A. T. **As dimensões da fé**: sete religiões mundiais em uma análise multidimensional lexical. 2020. 417 f. Tese (Doutorado em Linguística Aplicada e Estudos da Linguagem) – Programa de Pós-Graduação em Linguística Aplicada e Estudos da Linguagem, Pontifícia Universidade Católica de São Paulo, São Paulo, 2020.

VOLÓCHINOV, V. **Marxismo e filosofia da linguagem**: problemas fundamentais do método sociológico na ciência da linguagem. Tradução de Sheila Camargo Grillo e Ekaterina Vólkova Américo. São Paulo: Editora 34, 2017.

VOSOUGHI, S.; ROY, D.; ARAL, S. The spread of true and false news online. **Science**, v. 9, n. 359, p. 1146-1151, 2018.

WAKEFIELD, A. J. *et al.* RETRACTED: Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. **The Lancet**, 1998.

WELLER, K. *et al.* Twitter and Society: An Introduction. *In*: WELLER, K. *et al.* (ed.). **Twitter and Society**. New York: Peter Lang, 2014. p. xxix--xxxviii.

WHITEMAN, M. Discourses emerging from anti-vaccine movements in Brazil on twitter during the Covid-19 pandemic. *In*: INTERNATIONAL CORPUS LINGUISTICS CONFERENCE, 12., 2023, Lancaster. **Anais [...]**. Lancaster: Lancaster University, 2023.

WICKE, P.; BOLOGNESI, M. M. Framing COVID-19: How we conceptualize and discuss the pandemic on Twitter. **PLoS One**, v. 15, n. 9, e0240010, 2020.

WODAK, R. (ed.). **Language, power, and ideology**: Studies in political discourse. Amsterdam; Philadelphia: John Benjamins, 1989.

WODAK, R. **Disorders of Discourse**. London: Longman, 1996.

WODAK, R. **Feminist critical discourse analysis**: new perspectives for interdisciplinary gender studies. Atenas: [S. n.], 2005.

WODAK, R.; MEYER, M. (ed.). **Methods of Critical Discourse Analysis**. Thousand Oaks: Sage, 2001.

WOLFE, R. M.; SHARP, L. K. Anti-vaccinationists past and present. **BMJ**, v. 325, n. 7361, p. 430-432, 2002.

WORLD HEALTH ORGANIZATION – WHO. A brief History of Vaccines. **WHO**, [s. d.]. Disponível em: <https://www.who.int/news-room/spotlight/history-of-vaccination/a-brief-history-of-vaccination>. Acesso em: 21 mar. 2024.

WORLD HEALTH ORGANIZATION – WHO. Ten threats to global health in 2019. **WHO**, 2019. Disponível em: <https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019>. Acesso em: 21 mar. 2024.

ZAPPAVIGNA, M. Twitter. In: HOFFMANN, C. R.; BUBLITZ, W. (ed.). **Pragmatics of Social Media**. Berlin; New York: De Gruyter Mouton, 2017. p. 737-762.

ZAPPAVIGNA, M. **Searchable talk**: Hashtags and social media metadiscourse. London: Bloomsbury, 2018.

ZAPPAVIGNA, M.; LOGI, L. **Emoji and Social Media Paralanguage**. Cambridge: Cambridge University Press, 2023.

ZORZETTO, R. O tombo da vacinação infantil. **Revista Pesquisa Fapesp**, v. 313, 2022. Disponível em: <https://revistapesquisa.fapesp.br/o-tombo-na-vacinacao-infantil/>. Acesso em: 4 abr. 2024.