

BERBER SARDINHA, Tony; ALENCAR, André Luiz Siqueira; DA SILVA, Cicero Soares; GIL, Cristina Borges; LOPES, Maurício José Ferreira; HUGHES, Stephan Arthur Solomon. "#EUNAOVOUTOMARVACINA": Uma abordagem da linguística de corpus e da análise multimodal imagética. *Revista Intercâmbio*, v.LI: 178-209, 2022. São Paulo: LAEL/PUCSP. ISSN 2237-759X

<https://doi.org/10.23925/2237.759X.2022V51.e58515>

#EUNAOVOUTOMARVACINA: UMA ABORDAGEM DA LINGUÍSTICA DE CORPUS E DA ANÁLISE MULTIMODAL IMAGÉTICA

#EUNAOVOUTOMARVACINA¹: A CORPUS LINGUISTICS AND MULTIMODAL IMAGING ANALYSIS APPROACH

Tony BERBER SARDINHA
(Pontifícia Universidade Católica de São Paulo – PUC-SP)
tonyberber@gmail.com

André Luiz Siqueira ALENCAR
(Pontifícia Universidade Católica de São Paulo – PUC/SP)
andrealencar.professor@gmail.com

Cícero Soares DA SILVA
(Pontifícia Universidade Católica de São Paulo – PUC-SP)
pardonmester@gmail.com

Cristina Borges GIL
(Pontifícia Universidade Católica de São Paulo – PUC-SP)
cristinaborgesgil@gmail.com

Maurício José Ferreira LOPES
(Pontifícia Universidade Católica de São Paulo – PUC-SP)
maleferlopes@gmail.com

Stephan Arthur Solomon HUGHES
(Pontifícia Universidade Católica de São Paulo – PUC-SP)
stephan.hughes@gmail.com

RESUMO: O objetivo deste artigo é a análise da *hashtag* #eunaovoutomarvacina, através da compilação de um *corpus* de Twitter. Como método estudo, usou-se a Linguística de *Corpus* (BERBER SARDINHA, 2004), na vertente da Análise Multidimensional Imagética (AMDI) (BERBER SARDINHA, 2022), para analisar 2600 imagens presentes em 17267 tweets, abrangendo o período do início da pandemia

¹ Em inglês, #Nottakingthevaccine

BERBER SARDINHA, Tony; ALENCAR, André Luiz Siqueira; DA SILVA, Cicero Soares; GIL, Cristina Borges; LOPES, Maurício José Ferreira; HUGHES, Stephan Arthur Solomon. "#EUNAOVOUTOMARVACINA": Uma abordagem da linguística de corpus e da análise multimodal imagética. *Revista Intercâmbio*, v.LI: 178-209, 2022. São Paulo: LAEL/PUCSP. ISSN 2237-759X

de COVID-19 (início de 2020) até meados de novembro de 2021. Assim, três dimensões semânticas foram identificadas e interpretadas como (1) Ironia ao fanatismo coletivo vs. Politização/Judicialização da vacina, (2) Sátira ao discurso antivacina, (3) Discurso da relação de poder e exploração da vacina como benefício pessoal.

PALAVRAS-CHAVE: Linguística de Corpus; Análise Multidimensional Imagética; Twitter; Imagens.

ABSTRACT: *This article seeks to analyze the hashtag #eunaovoutomarvacina (#nottakingthevaccine), through a compilation of a Twitter-based corpus. Corpus Linguistics was adopted as the main method of research, more specifically Multidimensional Image Analysis (Berber Sardinha, 2022), to study 2,600 images in 17,267 tweets, spanning the period of the start of the Covid-19 pandemic until mid-November 2021. The data reveal three semantic dimensions that can be identified and interpreted as: (1) Mockery of the collective fanaticism versus Politicization/Judicial activism of the vaccine, (2) Satire of the anti-vaccine discourse, (3) Discourse of the relation between power and exploitation of the vaccine for personal use.*

KEYWORDS: *Corpus Linguistics; Multi-dimensional Image Analysis; Twitter; Images.*

1. Introdução

Este artigo tem como objetivo principal analisar o discurso antivacina da *hashtag* #EuNaoVouTomarVacina por meio da compilação de um *corpus* de *Twitter* com 2600 imagens presentes em 17267 tweets, que abrangem o período do início da pandemia de COVID-19 no início de 2020 até meados de novembro de 2021.

O negacionismo, dentre os quais se insere o discurso antivacina, é um comportamento tendencioso que consiste na recusa a aceitar a existência, a validade ou a verdade de algo, como eventos históricos ou fatos científicos, apesar das evidências ou argumentos que o comprovam. De origem francesa, foi usado pela primeira vez por intelectuais franceses para questionar a validade de uma página sombria de nossa história, o movimento negacionista contemporâneo, denominado os negacionistas do Holocausto (VALIM, SÁ AVELAR, 2020).

Este movimento era formado pelos intelectuais franceses de extrema-direita Maurice Bardèche e Paul Rassinier e o estadunidense Harry Elmer Barnes (CASTRO, 2014, p.7), que escreviam livros e se

BERBER SARDINHA, Tony; ALENCAR, André Luiz Siqueira; DA SILVA, Cicero Soares; GIL, Cristina Borges; LOPES, Maurício José Ferreira; HUGHES, Stephan Arthur Solomon. "#EUNAOVOUTOMARVACINA": Uma abordagem da linguística de corpus e da análise multimodal imagética. *Revista Intercâmbio*, v.LI: 178-209, 2022. São Paulo: LAEL/PUCSP. ISSN 2237-759X

posicionavam através dos meios de comunicação, alegando que o Holocausto foi uma farsa. De acordo com Porfírio (2022), esses intelectuais negacionistas afirmavam que os judeus inflacionaram os relatos e que a maior parte das mortes foi causada por guardas e até pelos próprios prisioneiros. Além do mais, eles alegavam que a câmara de gás nunca existiu. Inescrupulosamente, até justificavam as práticas cruéis contra os judeus.

Atualmente, o discurso negacionista antivacina ganhou bastante notoriedade em meio à pandemia de COVID-19 pondo em risco os preceitos já sedimentados pela ciência. No Brasil, esta motivação disseminou-se muito rapidamente através de repetidas mensagens nas mídias sociais, tais como *Facebook*, *Twitter* e, principalmente em grupos fechados de *WhatsApp*, pondo em xeque a segurança das vacinas.

É sabido que movimentos antivacina estão presentes no mundo desde a criação da vacina. De acordo com a FIOCRUZ e o Instituto Butantan, em 1904, o sanitarista Oswaldo Cruz ordenou a vacinação compulsória da população carioca e isso causou um dos episódios mais bizarros da história do Brasil, a Revolta da Vacina, que foi marcada por mortes, estado de sítio e notícias fraudulentas. Isso se deu pelo fato de que a população desconhecia a eficácia das vacinas, pois não existiu uma campanha, explicação ou uma forma de diálogo entre as autoridades sanitárias e a população.²

Em contraponto, hoje há informações e estudos empíricos sobre a eficácia das vacinas, bem como a erradicação de doenças com a vacinação em massa. No entanto, movimentos antivacina, através da desinformação, teoria da conspiração e, principalmente, de notícias fraudulentas, vêm crescendo incomensuravelmente.

O advento das mídias sociais tornou-se uma ferramenta muito poderosa e um meio de comunicação incrivelmente rápido no mundo contemporâneo. De acordo com Berber Sardinha (2022, no prelo)³:

As mídias sociais referem-se a plataformas que permitem aos usuários enviar mensagens de texto e conteúdo multimídia para um grupo de outros usuários.⁴

²<https://butantan.gov.br/noticias/ha-mais-de-100-anos-revolta-da-vacina-foi-marcada-por-mortes-estado-de-sitio-e-fake-news> (Acesso em 02/02/2022)
<https://portal.fiocruz.br/noticia/cinco-dias-de-furia-revolta-da-vacina-envolveu-muito-mais-do-que-insatisfacao-com-vacinacao> (Acesso em 02/02/2022)

³ Tradução nossa. Doravante, todas as traduções são feitas pelos autores.

⁴ "Social media refers to platforms that enable users to send text messages and multimídia content to a group of users" (BERBER SARDINHA, 2022 no prelo)

BERBER SARDINHA, Tony; ALENCAR, André Luiz Siqueira; DA SILVA, Cicero Soares; GIL, Cristina Borges; LOPES, Maurício José Ferreira; HUGHES, Stephan Arthur Solomon. “#EUNAOVOUTOMARVACINA”: Uma abordagem da linguística de corpus e da análise multimodal imagética. *Revista Intercâmbio*, v.LI: 178-209, 2022. São Paulo: LAEL/PUCSP. ISSN 2237-759X

Há diferentes plataformas sociais como *YouTube, Instagram, Facebook, WhatsApp, Snapchat, Tumblr, TikTok* e *Twitter*, entre outras. Berber Sardinha (2022, no prelo) afirma que “cada uma dessas plataformas foi criada para servir a um propósito peculiar para um público específico”⁵.

Apesar de todos os benefícios proporcionados e oferecidos pelas mídias sociais, indubitavelmente, elas facilitaram o surgimento das câmaras de eco (*echo chambers*)⁶, que são “uma situação em que as pessoas só ouvem opiniões de um tipo, ou opiniões semelhantes às suas”. As câmaras de eco podem criar desinformação e distorcer a perspectiva de uma pessoa, impedindo-a de considerar diferentes pontos de vista e discutir assuntos complexos. Elas são confirmações enviesadas que tendem a reforçar e favorecer crenças existentes.

Neste estudo, o discurso multimodal prioriza os texto-imagens dos tuítes que carregam em sua superfície as marcas da multimodalidade, ou seja, a imagem e o texto. A *hashtag #EuNaoVouTomarVacina* apontou não apenas o discurso antivacina – mas, também, o discurso pró-vacina. Essa divergência de opiniões levantou algumas questões devido ao fato de que as *hashtags*, quando divulgadas ou anunciadas nas mídias sociais, tornam-se domínio público e, conseqüentemente, esse tipo de cruzamento ideológico será inevitável, pressupondo ser pervasivo e ubíquo, resultando na sua viralização.

A metodologia principal consiste na Linguística de *Corpus* (BERBER SARDINHA, 2004), em sua vertente da Análise Multidimensional Imagética (AMDI) (BERBER SARDINHA, 2021). Também chamada de Análise de Constituintes Visuais (ACV), segundo Christiansen, Dance & Wild (2020, p.149), é:

Um método da construção do *corpus* multimodal que permite aos pesquisadores construir e analisar aspectos visuais da mídia online em *corpora* de grande porte através do processamento de imagens utilizando o Google’s Cloud Vision API⁷.

⁵ “Each was introduced to serve a particular purpose for a particular audience.”(BERBER SARDINHA, 2022 no prelo).

⁶ *echo chamber noun* [C] (SITUATION) a situation in which people only hear opinions of one type, or opinions that are similar to their own. Cambridge Dictionary (Acesso em 12/10/2022).

⁷ Visual Constituent Analysis (VCA), a method of multimodal corpus construction that allows researchers to construct and analyse visual aspects of online media in large-scale corpora, using computer ‘Vision’-based image annotation (in our case Google Cloud Vision) (CHRISTIANSEN, DANCE & WILD, 2020, p.149).

BERBER SARDINHA, Tony; ALENCAR, André Luiz Siqueira; DA SILVA, Cicero Soares; GIL, Cristina Borges; LOPES, Maurício José Ferreira; HUGHES, Stephan Arthur Solomon. "#EUNAOVOUTOMARVACINA": Uma abordagem da linguística de corpus e da análise multimodal imagética. *Revista Intercâmbio*, v.LI: 178-209, 2022. São Paulo: LAEL/PUCSP. ISSN 2237-759X

Em termos gerais, esta abordagem se baseia na construção de uma representação simplificada de um *corpus* texto-imagem como uma série de palavras-chaves ou termos genéricos.

2. Fundamentação Teórica

2.1. Linguística de *Corpus*

O artigo aqui proposto tem como fundamentação teórica principal a Linguística de *Corpus* (doravante LC), que se ocupa da coleta e da exploração de *corpora*, ou conjunto de dados linguísticos textuais coletados criteriosamente, com o propósito de servirem para a pesquisa de uma língua ou variedade linguística. Como tal, dedica-se à exploração da linguagem por meio de evidências empíricas, extraídas por computador (BERBER SARDINHA, 2004). Assim, o objeto de estudo da LC é o que pode ser definido como um corpo de linguagem natural (autêntica) que pode ser usado como base para pesquisa linguística (BERBER SARDINHA, 2004; SINCLAIR, 1995).

De acordo com Biber, Conrad & Reppen (1998, p.4), as pesquisas baseadas em LC apresentam como características principais: (I) ser empírica e analisar padrões reais de uso de textos naturais; (II) utilizar uma coleção grande de textos naturais e criteriosamente selecionados, conhecida como *corpus*; (III) fazer análises por meio de técnicas tanto automáticas, como interativas e (IV) empregar técnicas de análise ambas quantitativas e qualitativas (interpretativas).

Segundo Berber Sardinha (2004), o objeto central do estudo da LC é o *corpus*, que Sanchez (1995, p.8-9) define como:

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

Esta definição é mais completa porque menciona vários pontos importantes: **a origem**: os dados devem ser autênticos; **o propósito**: o *corpus* deve ter a finalidade de ser um objetivo de estudo linguístico; **a composição**: o conteúdo do *corpus* deve ser criteriosamente escolhido; **a formatação**: os dados do *corpus* devem ser legíveis por computador; **a representatividade**: o *corpus* deve ser representativo de uma língua ou variedade e **a extensão**: o *corpus* deve ser vasto para ser representativo.

BERBER SARDINHA, Tony; ALENCAR, André Luiz Siqueira; DA SILVA, Cicero Soares; GIL, Cristina Borges; LOPES, Maurício José Ferreira; HUGHES, Stephan Arthur Solomon. "#EUNAOVOUTOMARVACINA": Uma abordagem da linguística de corpus e da análise multimodal imagética. *Revista Intercâmbio*, v.LI: 178-209, 2022. São Paulo: LAEL/PUCSP. ISSN 2237-759X

2.2 Análise Multidimensional

Desenvolvida por Douglas Biber (1988, 2009) (BERBER SARDINHA, 2014; PINTO, 2019), a Análise Multidimensional (doravante AMD) é uma abordagem metodológica, vertente americana da LC. Ela é hoje considerada a principal contribuição "de Biber para o estudo da linguagem baseado em *corpus* eletrônico" (BERBER SARDINHA, 2004, p.299). A AMD busca investigar e comparar os traços linguísticos de textos escritos e orais, identificando padrões de variação em diferentes registros de uma língua (BERBER SARDINHA; VEIRANO PINTO, 2019, p.1).

Os conceitos primordiais dessa abordagem são registros e dimensões. Biber define registro como "termo para qualquer variedade linguística definida pelas suas características situacionais incluindo o propósito do falante, a relação entre o falante e o ouvinte e o contexto de produção" ⁸(BIBER, 2009, p.823). Dimensões são padrões de coocorrência de elementos lexicogramaticais que subjazem aos textos de uma língua. De acordo com Berber Sardinha (2004, p. 300), a AMD é definida como:

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

A AMD (BIBER, 2001) tem como princípio fundamental estudar textos, registros e tipos de textos e não traços linguísticos individuais. Ela denomina-se multidimensional devido ao fato de que os parâmetros múltiplos de variação operam em qualquer domínio do discurso. Sendo assim, a coocorrência de padrões de linguagem define cada dimensão, compreendida como o estatuto em que um fator (produto da análise fatorial) assume assim que ele é interpretado do ponto de vista de sua função comunicativa, levando em consideração não apenas as características linguísticas, mas também as características partilhadas pelos registros que estão apresentados no fator (BERBER SARDINHA, 2004).

⁸ No original. "A cover term for any language variety defined by its situational characteristics, including the speaker's purpose, the relationship between speaker and hearer, and the production circumstances" (BIBER, 2009 p. 823)

BERBER SARDINHA, Tony; ALENCAR, André Luiz Siqueira; DA SILVA, Cicero Soares; GIL, Cristina Borges; LOPES, Maurício José Ferreira; HUGHES, Stephan Arthur Solomon. "#EUNAOVOUTOMARVACINA": Uma abordagem da linguística de corpus e da análise multimodal imagética. *Revista Intercâmbio*, v.LI: 178-209, 2022. São Paulo: LAEL/PUCSP. ISSN 2237-759X

Por intermédio da AMD, torna-se visível que é possível identificar textos e registros relacionados a parâmetros situacionais ou funcionais, tais como: formais/informais, interativos/não-interativos.

Biber (1988) sugere o estudo por meio de dimensões de variação linguística posto que esses parâmetros podem ser considerados dimensões, e dessa forma "definem contínuos de variação ao invés de polos separados. Por exemplo, apesar de ser possível descrever um texto simplesmente como formal ou informal"⁹.

Com a finalidade de exemplificar uma das dimensões de variação que Biber (1988) obteve na pesquisa da língua inglesa, podemos citar 'Produção com interação versus informacional', onde todos os textos analisados possuem características relativas a esses parâmetros. No entanto, existe uma tendência dos textos interativos a serem informativos e vice-versa.

2.3. Multimodalidade

O termo multimodalidade surgiu na década de 1920, ligada à área da Psicologia da Percepção, expressando o efeito que distintas percepções sensoriais têm umas sobre as outras. No entanto, a multimodalidade foi tomada, também, por linguistas e analistas do discurso a fim de incorporar diferentes recursos e meios comunicativos, por exemplo, a linguagem escrita, a imagem, o som e a música em textos multimodais e eventos comunicativos (VAN LEEUWEN, 2011).

Para Kress (1996) e Van Leeuwen (2006), toda forma de comunicação é multimodal, pois tanto no contexto social quanto nas práticas sociais as pessoas se utilizam de meios de comunicação, uma vez que os significados comunicativos são construídos e compartilhados por meio de vários modos, tais como a escrita, a fala, a imagem, o som, o gesto, a tipografia, a imagem em movimento (CALLOW, 2014). Dessa forma, a multimodalidade propõe um entendimento de que as pessoas tenham habilidades que vão além das competências comunicativas da leitura e escrita.

A multimodalidade também abrange a habilidade de ler imagens. Para Kress (1996) e Van Leeuwen (2006), a linguagem visual não deve ser vista simplesmente como um complemento da linguagem verbal. As imagens representam um forte veículo de comunicação e informação,

8 define continuums of variation rather than discrete poles. For example, although it is possible to describe a text as simply formal or informal, it is more accurate to describe it as more or less formal" (BIBER, 1988, p. 9)

BERBER SARDINHA, Tony; ALENCAR, André Luiz Siqueira; DA SILVA, Cicero Soares; GIL, Cristina Borges; LOPES, Maurício José Ferreira; HUGHES, Stephan Arthur Solomon. "#EUNAOVOUTOMARVACINA": Uma abordagem da linguística de corpus e da análise multimodal imagética. *Revista Intercâmbio*, v.LI: 178-209, 2022. São Paulo: LAEL/PUCSP. ISSN 2237-759X

uma vez que a imagem nem sempre transmitirá a mesma mensagem do texto escrito.

2.4. *Twitter*

Fundado em 2006 com o intuito de enviar e receber mensagens via SMS, o micro blog conhecido como *Twitter* se reposicionou como principal fonte de notícias e palco para transformações sociais e políticas. Atualmente disponível em 35 idiomas diferentes, a plataforma conta com 316 milhões de usuários cadastrados. Ao longo dos seus quinze anos de existência, a rede social mostra poder de transformação e inovação sem igual, ressignificando o vocábulo "*twitter*", que está atrelado a "assobios de pássaros" ou "uma pequena explosão de informações". Segundo o ex-diretor criativo da empresa, Doug Bowman, o *Twitter* é "a representação definitiva de liberdade, esperança e possibilidades ilimitadas".

Alinhados com o conceito de liberdade e possibilidades ilimitadas do Bowman foram o aumento de 140 para 280 caracteres, a inclusão de recursos visuais (imagens, gifs, *emojis*) e o encurtamento automático de links compartilhados, as três novidades que potencializaram a disseminação de mais conteúdo, sem perder a dinâmica da interação social do *Twitter* que lhe é peculiar.

Como fruto dessa dinâmica social tem-se uma das características mais relevantes para este artigo: a brincadeira virtual dos memes. O Brasil detém o posto de maior exportador de memes do mundo, tendo vencido a primeira "guerra mundial memeeal" em 2016 (é tricampeão da competição).

Os resultados da pesquisa realizada pelos autores com o *hashtag* #EuNãoVouTomarVacina revela o papel central de memes e imagens em geral, na disseminação de ideias e informações embasadas ou não, e na construção das multidimensões e da multimodalidade da linguagem (BERBER SARDINHA, 2000).

Com o passar dos anos, notou-se a eminência do *Twitter* na disseminação de notícias e transmissão de eventos políticos, culturais, econômicos e sociais. Sua larga utilização por empresas de comunicação se deve em grande parte pela rápida veiculação de fatos, pelo caráter instantâneo, pela possibilidade de estabelecer um diálogo com o público em tempo real, e por último, pela oportunidade de medir o impacto de uma notícia.

Essa vocação jornalística do *Twitter*, capaz de aproximar o jornalista do seu público-alvo, traz benefícios semelhantes para outras empresas, assim como para órgãos oficiais, associações e grupos informais ou de militância política. Todos os citados fazem amplo uso de, talvez, a

BERBER SARDINHA, Tony; ALENCAR, André Luiz Siqueira; DA SILVA, Cicero Soares; GIL, Cristina Borges; LOPES, Maurício José Ferreira; HUGHES, Stephan Arthur Solomon. "#EUNAOVOUTOMARVACINA": Uma abordagem da linguística de corpus e da análise multimodal imagética. *Revista Intercâmbio*, v.LI: 178-209, 2022. São Paulo: LAEL/PUCSP. ISSN 2237-759X

funcionalidade mais impactante do micro blog: o *hashtag*. Com ela, é possível chamar o público usuário das redes sociais e direcioná-lo para uma discussão em torno de assuntos correntes de ordem pública, caso evidenciado no embate com relação à campanha de vacinação contra o vírus da COVID-19.

Assim como notícias em primeira mão são agora mais "primeiras" do que nunca, os negócios podem se armar com o imediatismo do *Twitter* para inovar e construir relacionamentos como nunca (COMM, p.22, 2009). O *hashtag* serve para facilitar a localização de tuítes em torno de determinado assunto, e, assim, auxilia o usuário a encontrar indivíduos com quem compartilha ideias, valores, crenças e objetivos, dando origem a verdadeiras comunidades de prática (WENGER, 1999).

Essa natureza noticiosa do *Twitter* se deve a cinco fundamentos, denominados por Dawson:

1. O "aqui e agora": o que vemos, ouvimos e pensamos pode ser tuitado instantaneamente a milhões de pessoas ao redor do mundo;
2. Apesar de veicular informação duvidosa, é possível checar e complementar dados iniciais;
3. A corrida desenfreada pela divulgação de notícias em primeira mão, ou o furo, faz aumentar a velocidade de transmissão de informações pelo *Twitter*;
4. Cria-se uma hierarquia de credibilidade dos meios de comunicação, e dos próprios usuários, junto ao público.
5. O *Twitter* e os meios de comunicação vivem uma relação simbiótica, uma vez que as pessoas acessam o segundo por meio do micro blog e este se torna fonte de "notícias quentes" (*breaking news*) ao propiciar histórias em potencial para os meios de comunicação.

Em tempos em que se produzem informações cada vez mais rápido e em volumes cada vez maiores, as notícias precisam ser divulgadas de forma quase que instantânea, evidente no fundamento do "aqui e agora" do Dawson. O *Twitter* se apresenta como ferramenta ideal para atender à tal necessidade, como afirma Seixas (2009):

O *Twitter* foi o precursor, aquele que definiu o conceito, as novas possibilidades e a nova forma de irrigar o mundo com conteúdo. Permitiu que uma verdadeira legião de programas, *sites* e *mash-ups*¹⁰ pudessem proporcionar formas diferentes de publicar e interagir com a quantidade colossal de mini conteúdos já disponível nesse pequeno, e crescente, universo. Já se percebe que ao redor dessa plataforma se forma uma economia própria, onde ela própria é o início e o fim, o meio e o objetivo. (p. 45).

¹⁰ Website ou uma aplicação web que usa conteúdo de mais de uma fonte para criar novo serviço completo (WIKIPÉDIA, 2009).

BERBER SARDINHA, Tony; ALENCAR, André Luiz Siqueira; DA SILVA, Cicero Soares; GIL, Cristina Borges; LOPES, Maurício José Ferreira; HUGHES, Stephan Arthur Solomon. "#EUNAOVOUTOMARVACINA": Uma abordagem da linguística de corpus e da análise multimodal imagética. *Revista Intercâmbio*, v.LI: 178-209, 2022. São Paulo: LAEL/PUCSP. ISSN 2237-759X

A grande quantidade de informações disponibilizada através das redes sociais e a velocidade com a qual elas se espalham abrem precedentes para o compartilhamento de notícias falsas ou dados não verificados, as chamadas notícias fraudulentas (*fake news*).

Segundo Geeng, Yee e Roesner (2020), o fato de as redes sociais passarem a ser o principal meio de acesso a notícias por um número cada vez mais crescente de usuários, torna urgente compreender como as pessoas buscam e interagem com postagens não confiáveis. O *Twitter*, portanto, passa a ser arena e laboratório ao mesmo tempo: um espaço onde grupos antagônicos se enfrentam e onde se analisa o comportamento humano como ser social.

3. Metodologia

A fim de analisarmos o *corpus* de estudo, esta pesquisa será subsidiada, conforme descrito no capítulo da fundamentação teórica, pela LC e AMDI.

A AMDI requer uma série de etapas para a sua realização. Tal como em uma Análise Multidimensional tradicional (BIBER, 2009), a AMDI requer, basicamente oito passos: 1) coleta de um *corpus* de estudo; 2) identificação das variáveis linguísticas a serem incluídas no estudo; 3) uso de *softwares* para anotação e etiquetagem do *corpus*; 4) normalização das frequências das variáveis linguísticas em cada texto do *corpus*; 5) identificação da coocorrência entre as variáveis linguísticas, por meio da análise fatorial; 6) interpretação temática dos fatores em dimensões de variação subjacentes; e, por fim, 7) cálculo dos escores das dimensões para cada texto e comparação da média de escore das dimensões para cada registro.

Na próxima seção, serão detalhados todos os passos mencionados acima.

3.1 Coleta do *corpus*

O primeiro passo na coleta do *corpus* foi a identificação de potenciais candidatos para o estudo, a partir da rede social *Twitter*. Após extensas pesquisas, a *hashtag* #EuNaoVouTomarVacina demonstrou-se uma boa candidata, uma vez que não só contempla discursos antivacina – o que era obviamente esperado – mas também discursos pró-vacina, o que surpreendeu pela grande quantidade de usuários a favor do medicamento.

BERBER SARDINHA, Tony; ALENCAR, André Luiz Siqueira; DA SILVA, Cicero Soares; GIL, Cristina Borges; LOPES, Maurício José Ferreira; HUGHES, Stephan Arthur Solomon. “#EUNAOVOUTOMARVACINA”: Uma abordagem da linguística de corpus e da análise multimodal imagética. *Revista Intercâmbio*, v.LI: 178-209, 2022. São Paulo: LAEL/PUCSP. ISSN 2237-759X

No entanto, parece-nos muito natural o fato de discursos contrários digladiarem-se no palco das redes sociais, já que *hashtags*, uma vez lançadas nas redes, estão sob domínio público e, portanto, esse tipo de cruzamento ideológico inevitavelmente sempre estará presente, não importa a rede social utilizada para a sua viralização.

Definido o *corpus*, passou-se à sua compilação, por meio de um *script open-source* denominado *snsrape*¹¹. Para que o *script* funcione corretamente, é preciso certificar-se de que o usuário tenha instalado a versão 3.9 da linguagem de programação *Python*, bem como o sistema operacional Ubuntu¹², cujo núcleo opera sob o sistema operacional Linux.

A partir da tela do Ubuntu, instala-se a versão de desenvolvimento de *snsrape* por meio da seguinte linha de comando *pip3 install git+https://github.com/JustAnotherArchivist/snsrape.git*, conforme a figura abaixo:

Figura 01 – Instalação do *snsrape* pelo repositório *github*

```
cygnusx1@DESKTOP-F8DANIV:~$ pip3 install git+https://github.com/JustAnotherArchivist/snsrape.git
Collecting git+https://github.com/JustAnotherArchivist/snsrape.git
  Cloning https://github.com/JustAnotherArchivist/snsrape.git to /tmp/pip-req-build-w2ablhtg
  Running command git clone -q https://github.com/JustAnotherArchivist/snsrape.git /tmp/pip-req-build-w2ablhtg
Requirement already satisfied: beautifulsoup4 in ./local/lib/python3.9/site-packages (from snsrape==0.4.1.20211214) (4.9.3)
Requirement already satisfied: lxml in ./local/lib/python3.9/site-packages (from snsrape==0.4.1.20211214) (4.6.3)
Requirement already satisfied: requests[socks] in /usr/lib/python3/dist-packages (from snsrape==0.4.1.20211214) (2.22.0)
Requirement already satisfied: soupsieve>1.2; python_version >= "3.0" in ./local/lib/python3.9/site-packages (from beautifulsoup4) (2.2.2)
Requirement already satisfied: PySocks!=1.5.7,>=1.5.6 in ./local/lib/python3.9/site-packages (from requests[socks]->snsrape==0.4.1.20211214) (1.7.1)
Building wheels for collected packages: snsrape
  Building wheel for snsrape (setup.py) ... done
  Created wheel for snsrape: filename=snsrape-0.4.1.20211214-py3-none-any.whl size=53994 sha256=e6ab438d354a23aed3b8283dcd908e41
  Stored in directory: /tmp/pip-ephem-wheel-cache-fv28e9_/wheels/1a/ba/e2/39fa3a11802c4a622f2efc8be3f5ff854481051d0b4c95c1fd
Successfully built snsrape
Installing collected packages: snsrape
```

Fonte: pesquisadores

Em seguida, os tuítes da *hashtag* #EuNaoVouTomarVacina são baixados a partir do comando: *snsrape --jsonl twitter-search "#eunãovoutomarvacina" > tweets_eunãovoutomarvacina.txt*. Ao término do processo, os tuítes são salvos em arquivo único no formato *.json*. (*eunãovoutomarvacina.json*).

Um segundo *script* é usado, mas dessa vez, criado especificamente para este estudo¹³. Sua função é baixar as mídias associadas aos tuítes contidos no arquivo *.json*, o que, neste caso, refere-se a todas as imagens associadas a cada tuíte. Para esse fim, cria-se uma pasta no Ubuntu chamada *media_snsrape* e uma subpasta denominada *inputfolder*.

O *script* é copiado diretamente para a pasta *media_snsrape*, ao passo que o arquivo *.json* é copiado para a pasta *inputfolder*. Roda-se o

¹¹ Disponível em <https://github.com/JustAnotherArchivist/snsrape>

¹² Disponível em <https://ubuntu.com/download>.

¹³ Criado por Berber Sardinha.

BERBER SARDINHA, Tony; ALENCAR, André Luiz Siqueira; DA SILVA, Cicero Soares; GIL, Cristina Borges; LOPES, Maurício José Ferreira; HUGHES, Stephan Arthur Solomon. "#EUNAOVOUTOMARVACINA": Uma abordagem da linguística de corpus e da análise multimodal imagética. *Revista Intercâmbio*, v.LI: 178-209, 2022. São Paulo: LAEL/PUCSP. ISSN 2237-759X

script a partir do comando `bash split_snsrape.sh`, como mostra a figura abaixo:

Figura 02 – Divisão dos tuítes em arquivos individuais.

```
cygnusx1@DESKTOP-F8DANIV:~/Curso_Tony_2021-2/split_snsrape$ bash split_snsrape.sh
Enter project name:
eunaovoutomarvacina
split_snsrape.sh: line 20: /usr/bin/rm: Argument list too long
--- doing outputfolder/tweetfile_1454756389218357251_eunaovoutomarvacina.txt ----
--- doing tweetsfolder/tweet_1454756389218357251_eunaovoutomarvacina.txt ----
--- doing outputfolder/tweetfile_1452612167312756743_eunaovoutomarvacina.txt ----
--- doing tweetsfolder/tweet_1452612167312756743_eunaovoutomarvacina.txt ----
```

Fonte: pesquisadores

O *script* exige que o pesquisador nomeie o projeto. Para esta pesquisa, o projeto foi nomeado *discursoantivacina*. Após o término do processo de extração, o *script* cria automaticamente uma pasta denominada *outputfolder*, onde se encontram todas as imagens associadas a cada um dos tuítes.

3.2 Variáveis linguísticas a serem incluídas no estudo

Uma vez que este estudo lança mão de imagens como objeto principal de análise, as variáveis selecionadas para a pesquisa diferem daquelas normalmente encontradas em estudos tradicionais de AMD.

Assim, a compilação do corpus de imagens dos tuítes deve, primeiramente, ser etiquetada a partir das variáveis encontradas nas imagens e, para isso, usamos a API do *Google Cloud Vision*. Existem diversas categorias de detecção de imagens no site oficial do *Google Cloud Vision*¹⁴, tal como demonstrado na tabela abaixo:

Tabela 01 – Exemplos de etiquetas identificadas por meio do *Google Cloud Vision*

LABELS	WEB ENTITIES	SAFE SEARCH
Ruas	Imagens: correspondência completa	Adulto
Cidades	Imagens: correspondência parcial	Humor
Noite	Páginas com imagens correspondentes	Área médica

¹⁴ Para melhor compreensão das etiquetas aqui usadas, o leitor pode consultar o link <https://cloud.google.com/vision/docs/how-to>

BERBER SARDINHA, Tony; ALENCAR, André Luiz Siqueira; DA SILVA, Cicero Soares; GIL, Cristina Borges; LOPES, Maurício José Ferreira; HUGHES, Stephan Arthur Solomon. "#EUNAOVOUTOMARVACINA": Uma abordagem da linguística de corpus e da análise multimodal imagética. *Revista Intercâmbio*, v.LI: 178-209, 2022. São Paulo: LAEL/PUCSP. ISSN 2237-759X

Plantas	Imagens visualmente semelhantes	Violência
Seres animados	Rótulos de melhores sugestões	Racismo

Fonte: pesquisadores

Assim, para que possamos ter acesso a essas diversas categorias de etiquetas, o próximo passo consiste na criação de uma conta de *e-mail* no *Google* e a autorização do *Google Cloud API*. Todo o processo de autorização – desde a criação do *Google Cloud Storage Bucket* até o upload das imagens em lote no *Google Cloud Vision* encontra-se nesse [link](#).

Ao certificar-se de que as dependências do *Google Cloud* estão operando corretamente no *Ubuntu*, faz-se o *upload* de todas as imagens contidas na pasta *outputfolder* para o *Google Cloud Vision Bucket*, como mostra a figura abaixo:

Figura 03 – Upload das imagens para o *Google Cloud Vision Bucket*

Nome	Tamanho	Tipo	Criado	Classe de armazenamento	Última modificação	Acesso público	Histórico de versões
1250084591706267649_000001_discursoantivacina	222,6 KB	image/jpeg	15 de no...	Standard	15 de nov. de 20...	Não público	–
1320073419929636866_000001_discursoantivacina	67,5 KB	image/jpeg	15 de no...	Standard	15 de nov. de 20...	Não público	–
1325961002064650240_000001_discursoantivacina	97,3 KB	image/jpeg	15 de no...	Standard	15 de nov. de 20...	Não público	–
1325961002064650240_000002_discursoantivacina	36,4 KB	image/jpeg	15 de no...	Standard	15 de nov. de 20...	Não público	–
1326671029540884481_000001_discursoantivacina	577,2 KB	image/jpeg	15 de no...	Standard	15 de nov. de 20...	Não público	–

Fonte: <https://console.cloud.google.com>

3.3 Uso de *softwares* para anotação e etiquetagem do *corpus*

Três *scripts*¹⁵ em linguagem *shell* foram produzidos, a fim de processar as imagens contidas no *bucket*. Cada um dos *scripts* possui uma função específica de anotação do *Google Cloud*. Por exemplo, o *script safe_search.sh* lida com os conteúdos violentos, racistas, adultos etc. e sua probabilidade de estarem presentes nas imagens.

Já o *script labels.sh* lida com a extração de informações a respeito de diversas categorias presentes em cada imagem, tais como objetos, espécies de animais, plantas, atividades, locais etc. Abaixo, um exemplo de extração das anotações pelo *script web_entities*:

¹⁵ Criados por Berber Sardinha: *safe_search.sh*, *labels.sh* e *web_entities.sh*.

BERBER SARDINHA, Tony; ALENCAR, André Luiz Siqueira; DA SILVA, Cicero Soares; GIL, Cristina Borges; LOPES, Maurício José Ferreira; HUGHES, Stephan Arthur Solomon. "#EUNAOVOUTOMARVACINA": Uma abordagem da linguística de corpus e da análise multimodal imagética. *Revista Intercâmbio*, v.LI: 178-209, 2022. São Paulo: LAEL/PUCSP. ISSN 2237-759X

Figura 04 – Extração das anotações de *web_entities*.

```
cygnusx1@DESKTOP-F8DANIV:~/Curso_Tony_2021-2/google_cloud/script/web_entities$ bash web_entities.sh
--- annotating 1250084591706267649_000001_discursoantivacina.jpg ...
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
100 3798    0 3419 100  379  1557   172  0:00:02  0:00:02  --:--:-- 1729
--- annotating 1320073419929636866_000001_discursoantivacina.jpg ...
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
100 4538    0 4159 100  379  2010   183  0:00:02  0:00:02  --:--:-- 2194
--- annotating 1325961002064650240_000001_discursoantivacina.jpg ...
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
100 7185    0 6806 100  379  2424   135  0:00:02  0:00:02  --:--:-- 2558
--- annotating 1325961002064650240_000002_discursoantivacina.jpg ...
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
100 8512    0 8133 100  379  3262   152  0:00:02  0:00:02  --:--:-- 3414
```

Fonte: pesquisadores

Mais um *script*¹⁶, por sua vez, foi criado, no intuito de fazer a contagem de cada anotação específica gerada pelos três *scripts* anteriores. Roda-se o *script* três vezes no terminal com o comando¹⁷ *bash googlevision_tagcount.sh*:

Figura 05 – Contagem de *labels* pelo *script googlevision_tagcount.sh*.

```
cygnusx1@DESKTOP-F8DANIV:~/Curso_Tony_2021-2/google_cloud/tagcount$ bash googlevision_tagcount.sh
--- textcounts : 1250084591706267649_000001_discursoantivacina.jpg ---
--- textcounts : 1320073419929636866_000001_discursoantivacina.jpg ---
--- textcounts : 1325961002064650240_000001_discursoantivacina.jpg ---
--- textcounts : 1325961002064650240_000002_discursoantivacina.jpg ---
--- textcounts : 1326671029540884481_000001_discursoantivacina.jpg ---
--- textcounts : 1326671280150630400_000001_discursoantivacina.jpg ---
```

Fonte: pesquisadores.

Um penúltimo *script*¹⁸ analisa todas as etiquetas identificadas pelo *Google Cloud Vision* e converte-as em um formato mais legível para o SAS – programa de análise estatística que fará as análises das variáveis em questão. Esse *script* lança mão das etiquetas semânticas contidas no etiquetador USAS¹⁹ e contém diversas categorias semânticas, semelhantes ao *Google Cloud Vision*. A tabela abaixo contém o total de classes semânticas reconhecidas pelo USAS:

¹⁶ Criado por Berber Sardinha: *googlevision_tagcount.sh*.

¹⁷ É preciso entrar no script e selecionar uma anotação (*labels*, *web_entities* e *safe_search*) individualmente. Após o término do processamento, seleciona-se outra anotação e repete-se o processo.

¹⁸ Criado por Berber Sardinha: *googlecloud_usas.sh*

¹⁹ Disponível em <http://ucrel.lancs.ac.uk/usas/>.

BERBER SARDINHA, Tony; ALENCAR, André Luiz Siqueira; DA SILVA, Cicero Soares; GIL, Cristina Borges; LOPES, Maurício José Ferreira; HUGHES, Stephan Arthur Solomon. "#EUNAOVOUTOMARVACINA": Uma abordagem da linguística de corpus e da análise multimodal imagética. *Revista Intercâmbio*, v.LI: 178-209, 2022. São Paulo: LAEL/PUCSP. ISSN 2237-759X

Tabela 02 – Categorias semânticas do USAS.

A general and abstract terms	B the body and the individual	C arts and crafts	E emotion
F food and farming	G government and public	H architecture, housing and the home	I money and commerce in industry
K entertainment, sports and games	L life and living things	M movement, location, travel and transport	N numbers and measurement
O substances, materials, objects and equipment	P education	Q language and communication	S social actions, states and processes
T Time	W world and environment	X psychological actions, states and processes	Y science and technology
Z names and grammar			

Fonte: <http://ucrel.lancs.ac.uk/usas/>

Finalmente, um último *script*²⁰ gera um arquivo em formato *.tsv* com as contagens das etiquetas identificadas pelo *script googlecloud_usas.sh*, conforme se observa na figura abaixo:

Figura 06 – Contagem das etiquetas geradas pelo *googlecloud_usas.sh*

```
cygnusx1@DESKTOP-F8DANIV:~/Curso_Tony_2021-2/google_cloud/usas$ bash usas_tagcount.sh
--- 000001 tagsonly/1250084591706267649_000001_discursoantivacina.jpg_000001.txt ---
--- 000002 tagsonly/1320073419929636866_000001_discursoantivacina.jpg_000002.txt ---
--- 000003 tagsonly/1325961002064650240_000001_discursoantivacina.jpg_000003.txt ---
--- 000004 tagsonly/1325961002064650240_000002_discursoantivacina.jpg_000004.txt ---
--- 000005 tagsonly/1326671029540884481_000001_discursoantivacina.jpg_000005.txt ---
--- 000006 tagsonly/1326671280150630400_000001_discursoantivacina.jpg_000006.txt ---
--- 000007 tagsonly/1326671434022785034_000001_discursoantivacina.jpg_000007.txt ---
--- 000008 tagsonly/1326671680010412032_000001_discursoantivacina.jpg_000008.txt ---
--- 000009 tagsonly/1326671891290087424_000001_discursoantivacina.jpg_000009.txt ---
--- 000010 tagsonly/1326672087319257088_000001_discursoantivacina.jpg_000010.txt ---
--- 000011 tagsonly/1326672203732176897_000001_discursoantivacina.jpg_000011.txt ---
--- 000012 tagsonly/1326672373291114496_000001_discursoantivacina.jpg_000012.txt ---
--- 000013 tagsonly/1326672626245398530_000001_discursoantivacina.jpg_000013.txt ---
--- 000014 tagsonly/1330919294381813760_000001_discursoantivacina.jpg_000014.txt ---
--- 000015 tagsonly/1334557097162371075_000001_discursoantivacina.jpg_000015.txt ---
```

Fonte: pesquisadores.

3.4 Identificação da coocorrência entre as variáveis linguísticas, por meio da análise fatorial.

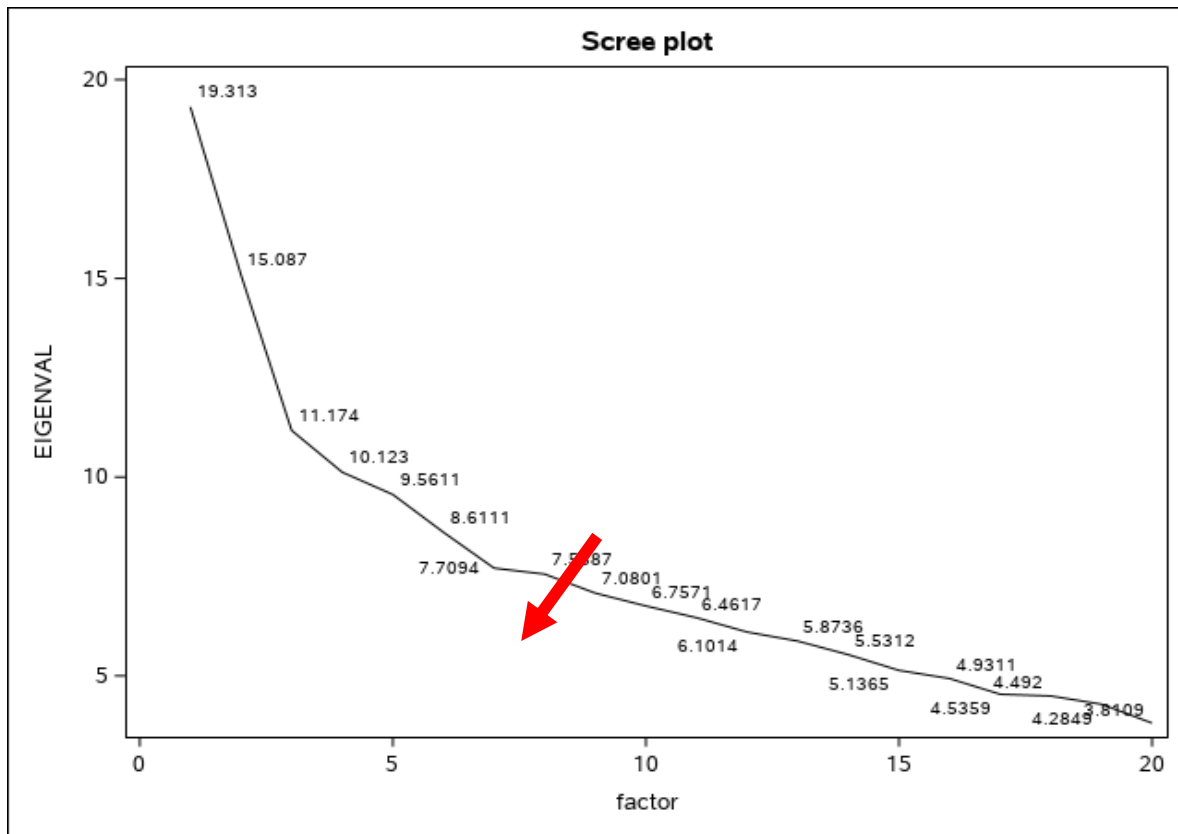
²⁰ Criado por Berber Sardinha: *usas_tagcounts.sh*

BERBER SARDINHA, Tony; ALENCAR, André Luiz Siqueira; DA SILVA, Cicero Soares; GIL, Cristina Borges; LOPES, Maurício José Ferreira; HUGHES, Stephan Arthur Solomon. "#EUNAOVOUTOMARVACINA": Uma abordagem da linguística de corpus e da análise multimodal imagética. *Revista Intercâmbio*, v.LI: 178-209, 2022. São Paulo: LAEL/PUCSP. ISSN 2237-759X

A planilha gerada pelo *script* é importada ao programa SAS²¹, juntamente com um programa criado especificamente para este estudo²², denominado *socialmediaclass.sas*. O programa realiza a análise fatorial, mencionada na seção Fundamentação Teórica.

Segundo Kauffmann (2020), a análise fatorial consiste na redução de uma ampla gama de variáveis linguísticas a um número reduzido de fatores, em que cada fator contém um determinado número de variáveis que coocorrem entre si e permitem sua interpretação em termos funcionais²³. A fim de encontrarmos o número ideal de fatores, uma primeira análise fatorial não rotacionada foi realizada. A solução inicial pode ser vista no gráfico abaixo:

Gráfico 01 – *Scree plot* ou Gráfico de Sedimentação da solução não rotacionada.



Fonte: SAS.

²¹ https://www.sas.com/pt_/software/on-demand-for-academics.html.

²² Programa em linguagem sas, criado por Berber Sardinha.

²³ Neste estudo, como dito anteriormente, as variáveis consistem em variáveis de imagens e, portanto, trata-se de uma interpretação.

BERBER SARDINHA, Tony; ALENCAR, André Luiz Siqueira; DA SILVA, Cicero Soares; GIL, Cristina Borges; LOPES, Maurício José Ferreira; HUGHES, Stephan Arthur Solomon. "#EUNAOVOUTOMARVACINA": Uma abordagem da linguística de corpus e da análise multimodal imagética. *Revista Intercâmbio*, v.LI: 178-209, 2022. São Paulo: LAEL/PUCSP. ISSN 2237-759X

O gráfico acima sugere a extração de sete fatores, uma vez que o cotovelo mais acentuado se situa na sétima iteração, indicada com a seta vermelha. A partir desse ponto, os valores tornam-se estáveis, ou seja, não há captura de diferenças significativas entre os fatores.

Uma análise com sete fatores foi conduzida. No entanto, a partir do terceiro fator, não há captura de diferenças entre os fatores que sejam passíveis de análise. Desse modo, considerou-se uma extração rotacionada de três fatores, por mostrarem adequação ao conteúdo propositivo das imagens.

Por fim, o escore de cada imagem é calculado para cada um dos fatores e, a partir desse ponto, é possível interpretar os fatores em dimensões²⁴ e analisar as imagens com maiores escores médios em cada uma das dimensões. A estrutura final dos fatores pode ser vista na tabela abaixo:

Tabela 03: Estrutura dos fatores.

Variável	Peso	Variável	Peso
Fator 1		<i>Trash can (Z9)</i>	0,58
<i>Other proper names (Z3)</i>	1,06	<i>Moving, coming and going (M1)</i>	0,57
<i>Farming & Horticulture (F4)</i>	0,98	<i>General And Abstract Terms (A1)</i>	0,55
<i>Places (M7)</i>	0,91	<i>Helping/hindering (S8)</i>	0,54
<i>Frequency (N6)</i>	0,86	<i>Arts and crafts (C1)</i>	0,53
<i>Geographical terms (W3)</i>	0,85	<i>Linguistic Actions, States And Processes; Communication (Q1)</i>	0,50
<i>Money generally (I1)</i>	0,80	<i>Sensory (X3)</i>	0,43
<i>Kin (S4)</i>	0,78	<i>Affect (A2)</i>	0,42
<i>Personal names (Z1)</i>	0,72	Fator 3	
<i>Green issues (W5)</i>	0,72	<i>Mental actions and processes (X2)</i>	0,89
<i>Plants (L3)</i>	0,70	<i>Medicines and medical treatment (B3)</i>	0,80
<i>Getting and giving; possession (A9)</i>	0,62	<i>Work and employment (I3)</i>	0,75
<i>Living creatures: animals, birds, etc. (L2)</i>	0,59	<i>Furniture and household fittings (H5)</i>	0,67
<i>Grammatical bin (Z5)</i>	0,54	<i>Recorded sound (K3)</i>	0,67
<i>The universe (W1)</i>	0,50	<i>Speech (Q2)</i>	0,63
<i>Time (T1)</i>	0,43	<i>Information technology and computing (Y2)</i>	0,63
<i>Probability (A7)</i>	-0,59	<i>Science and technology in general (Y1)</i>	0,61
<i>Crime, law and order (G2)</i>	-0,52	<i>Power relationship (S7)</i>	0,59

²⁴ Os conceitos de fator estão detalhados na seção Fundamentação Teórica.

BERBER SARDINHA, Tony; ALENCAR, André Luiz Siqueira; DA SILVA, Cicero Soares; GIL, Cristina Borges; LOPES, Maurício José Ferreira; HUGHES, Stephan Arthur Solomon. "#EUNAOVOUTOMARVACINA": Uma abordagem da linguística de corpus e da análise multimodal imagética. *Revista Intercâmbio*, v.LI: 178-209, 2022. São Paulo: LAEL/PUCSP. ISSN 2237-759X

<i>Government and Politics (G1)</i>	-0,52	<i>Parts of buildings (H2)</i>	0,58
Fator 2		<i>Social Actions, States And Processes (S1)</i>	0,55
<i>Anatomy and physiology (B1)</i>	0,89	<i>Geographical names (Z2)</i>	0,52
<i>Measurement (N3)</i>	0,83	<i>Time: Old, new and young; age (T3)</i>	0,50
<i>Sailing, swimming, etc. (M4)</i>	0,77	<i>Classification (A4)</i>	0,48
<i>Clothes and personal belongings (B5)</i>	0,74	<i>Stationary (M8)</i>	0,47
<i>Clothes and personal belongings (B5)</i>	0,74		
<i>Being (A3)</i>	0,73		
<i>Cleaning and personal care (B4)</i>	0,73		
<i>Physical attributes (O4)</i>	0,68		
<i>Happiness and Contentment (E4)</i>	0,66		
<i>Electricity and electrical equipment (O3)</i>	0,61		
<i>Objects generally (O2)</i>	0,58		

Fonte: pesquisadores.

4. Resultados

A interpretação das três dimensões semânticas resultantes desta análise será apresentada a seguir.

4.1 Dimensão 1: Ironia ao fanatismo coletivo *versus* politização/judicialização da vacina

O polo positivo da Dimensão 1 apresentou quinze variáveis semânticas: *other proper names*, que distingue ou identifica um produto, uma empresa etc.; *farming & horticulture*, que remetem à agricultura e horticultura de um modo geral; *places*, que designam espaços geográficos e conceituais; *frequency*, que se refere à taxa de recorrência de um evento; *geographical terms*, isto é, termos geográficos de um modo geral; *money generally*, termos relacionados a dinheiro; *kin*, que se refere a termos ligados ao parentesco; *personal names*, nomes que identificam um indivíduo; *green issues*, que indicam termos referentes ao meio-ambiente; *plants*, termos referentes a plantas; *getting and giving; possession*, que apontam para termos abstratos relacionados à aquisição, recebimento etc., *living creatures: animals, birds, etc.*, isto é, termos relacionados a criaturas vivas (exceto humanos); *gramatical bin*, preposições, advérbios etc., *the universe*, que indicam termos

BERBER SARDINHA, Tony; ALENCAR, André Luiz Siqueira; DA SILVA, Cicero Soares; GIL, Cristina Borges; LOPES, Maurício José Ferreira; HUGHES, Stephan Arthur Solomon. "#EUNAOVOUTOMARVACINA": Uma abordagem da linguística de corpus e da análise multimodal imagética. *Revista Intercâmbio*, v.LI: 178-209, 2022. São Paulo: LAEL/PUCSP. ISSN 2237-759X

relacionados ao universo/cosmos e *time*, que indica a duração de um determinado evento, o período específico de um determinado evento etc. No geral, houve a prevalência das variáveis semânticas *farming & horticulture, places, green issues, plants* e *living creatures: animals, birds*, que, juntas, ironizam a cegueira coletiva de determinados estratos sociais no que diz respeito à negação da vacina.

Já o polo negativo apresentou três variáveis: *probability*, que se refere a termos abstratos de um modo geral; *Crime, law and order*, isto é, termos relacionados a atividades criminosas, criminais e ao sistema legal e *Government and Politics*, termos ligados à política e ao governo.

De acordo com a prevalência das variáveis semânticas presentes em cada um dos polos, pôde-se rotular a Dimensão 1 como "Ironia ao fanatismo coletivo *versus* Politização/Judicialização da vacina". Exemplos de imagens dos polos positivo e negativo encontram-se nas figuras abaixo:

Figura 07 – Dimensão 1 (Polo Positivo) – Ironia ao fanatismo coletivo



Da esquerda para a direita, de cima para baixo:

BERBER SARDINHA, Tony; ALENCAR, André Luiz Siqueira; DA SILVA, Cicero Soares; GIL, Cristina Borges; LOPES, Maurício José Ferreira; HUGHES, Stephan Arthur Solomon. "#EUNAOVOUTOMARVACINA": Uma abordagem da linguística de corpus e da análise multimodal imagética. *Revista Intercâmbio*, v.LI: 178-209, 2022. São Paulo: LAEL/PUCSP. ISSN 2237-759X

1 - Tuíte número 1339252106813976576: "ado a ado quem não tomar vacina é viado#EuNaoVouTomarVacina <https://t.co/LhwoK9grWd>";

2 - Tuíte número 1339242032343166976: "Eu e a pessoa que tweeta #EuNaoVouTomarVacina sem ser zoas... <https://t.co/BwROyhVEwx>";

3 - Tuíte número 1339256312002211842: "Eu entrando no tt e vendo uma merda dessas #EuNaoVouTomarVacina <https://t.co/hBON6NIpcP>"

4 - Tuíte número 1339364119590387714: "Vai insistir?#EuNaoVouTomarVacina <https://t.co/45Ta20uJYE>".

Fonte: pesquisadores.

Figuras 08 e 09 - Dimensão 1 (Polo Negativo) - Politização/Judicialização da vacina.



BERBER SARDINHA, Tony; ALENCAR, André Luiz Siqueira; DA SILVA, Cicero Soares; GIL, Cristina Borges; LOPES, Maurício José Ferreira; HUGHES, Stephan Arthur Solomon. "#EUNAOVOUTOMARVACINA": Uma abordagem da linguística de corpus e da análise multimodal imagética. *Revista Intercâmbio*, v.LI: 178-209, 2022. São Paulo: LAEL/PUCSP. ISSN 2237-759X



De cima para baixo:

1 - Tuíte número 1339252023368294400: "Zé Gotinha tá de mascara ou foi vedado de fita adesiva para não poder falar...?#CoisasDeBozo#EuNaoVouTomarVacina <https://t.co/p2iHVZx834>";

2 - tuíte 1339687192990498816: "Ao pessoal que está dizendo #EuNaoVouTomarVacina , saiba que vocês vão sim por que o \$TF vai apertar vocês de todos os jeitos, e considerando o quanto brasileiro é passivo e terceiriza suas guerras ao presidente, certamente todo mundo vai entrar na agulha made in China. <https://t.co/qMRXx3KqmJ>"

Fonte: pesquisadores

4.2 Dimensão 2 - Sátira ao discurso antivacina

A Dimensão 2, marcada apenas pelo polo positivo, apresentou dezoito variáveis semânticas: *anatomy and physiology*, indica termos relacionados ao corpo humano e processos corporais; *measurement*, refere-se a medidas de modo em geral; *sailing and swimming*, algo relacionado a meios de transporte e água; *clothes and personal belongings*, termos e objetos relacionados a peça de vestuário e outros acessórios pessoais; *being*, o que significa termos abstratos relacionados

BERBER SARDINHA, Tony; ALENCAR, André Luiz Siqueira; DA SILVA, Cicero Soares; GIL, Cristina Borges; LOPES, Maurício José Ferreira; HUGHES, Stephan Arthur Solomon. "#EUNAOVOUTOMARVACINA": Uma abordagem da linguística de corpus e da análise multimodal imagética. *Revista Intercâmbio*, v.LI: 178-209, 2022. São Paulo: LAEL/PUCSP. ISSN 2237-759X

ao ser e à existência; *cleaning and personal care*, faz-se referência à higiene doméstica e pessoal; *physical attributes*, isto é, termos relacionados à aparência em geral e propriedades físicas; *happiness and contentment*, refere-se a termos que se referem ao nível de felicidade; *electricity and electrical equipment*, representa termos relacionados à eletricidade e equipamentos elétricos; *objects generally* ortográficos ou palavras que não foram ainda incluídas no léxico; *moving, coming and going*, ou seja, termos que, de modo geral, indicam movimento; *general and abstract terms*, termos abstratos de modo geral; *helping/hindering*, que indicam termos relacionados a objetos em geral; *trash can*, representam erros relacionados a termos que descrevem ajuda e obstáculo; *arts and crafts*, termos relacionados a atividades e criatividades artísticas; *linguistics actions, states and processes; communication*, termos relacionados à comunicação de modo geral; *sensory*, termos sensoriais e *affect*, termos gerais e abstratos que denotam propensão para mudança.

As variáveis semânticas que mais pontuaram foram: *anatomy and physiology*, marcadas por características ou traços físicos e expressões faciais, satirizadas por um certo tom de deboche e desprezo; *clothes & personal belongings*, representadas por acessórios de uso pessoal (óculos e roupas) e destacadas, principalmente, por máscaras medievais, no formato de mordança e focinheira e máscaras contemporâneas protetoras do vírus da COVID-19, ambas marcadas por experiências de tempos sombrios da nossa história; *happiness and contentment*, representadas paradoxalmente por multifaces, que expressam um certo incômodo, um tom de deboche e sarcasmo e, por último, *linguistic actions, states and processes, communication*, marcada pela linguagem corporal e, talvez, pela linguagem de sinais.

Pode-se dizer que toda a forma de discurso está sendo representada pela expressão facial de cada pessoa, em diferentes momentos, cenários e situações contextuais, ou seja, um ato de protesto. Parece, até mesmo, que há um mal-estar internamente e as pessoas nem se dão conta de que toda essa emoção está sendo expressada ou exteriorizada através do semblante delas.

BERBER SARDINHA, Tony; ALENCAR, André Luiz Siqueira; DA SILVA, Cicero Soares; GIL, Cristina Borges; LOPES, Maurício José Ferreira; HUGHES, Stephan Arthur Solomon. "#EUNAOVOUTOMARVACINA": Uma abordagem da linguística de corpus e da análise multimodal imagética. *Revista Intercâmbio*, v.LI: 178-209, 2022. São Paulo: LAEL/PUCSP. ISSN 2237-759X

Figura 10 e 11 – Dimensão 2 – Sátira ao discurso antivacina.



BERBER SARDINHA, Tony; ALENCAR, André Luiz Siqueira; DA SILVA, Cicero Soares; GIL, Cristina Borges; LOPES, Maurício José Ferreira; HUGHES, Stephan Arthur Solomon. "#EUNAOVOUTOMARVACINA": Uma abordagem da linguística de corpus e da análise multimodal imagética. *Revista Intercâmbio*, v.LI: 178-209, 2022. São Paulo: LAEL/PUCSP. ISSN 2237-759X



De cima para baixo:

1 - Tuíte número 1339218820716785665" Evolução de um político <https://t.co/RsZOt35goN> por @roxmo";

2 - Tuíte número1339279905670959111:"Ja te obrigaram a usar FOCINHEIRA, você acha que vai escapar da vacina?!#EuNaoVouTomarVacina_ <https://t.co/jlCSiI8rnI>";

Fonte: pesquisadores

BERBER SARDINHA, Tony; ALENCAR, André Luiz Siqueira; DA SILVA, Cicero Soares; GIL, Cristina Borges; LOPES, Maurício José Ferreira; HUGHES, Stephan Arthur Solomon. "#EUNAOVOUTOMARVACINA": Uma abordagem da linguística de corpus e da análise multimodal imagética. *Revista Intercâmbio*, v.LI: 178-209, 2022. São Paulo: LAEL/PUCSP. ISSN 2237-759X

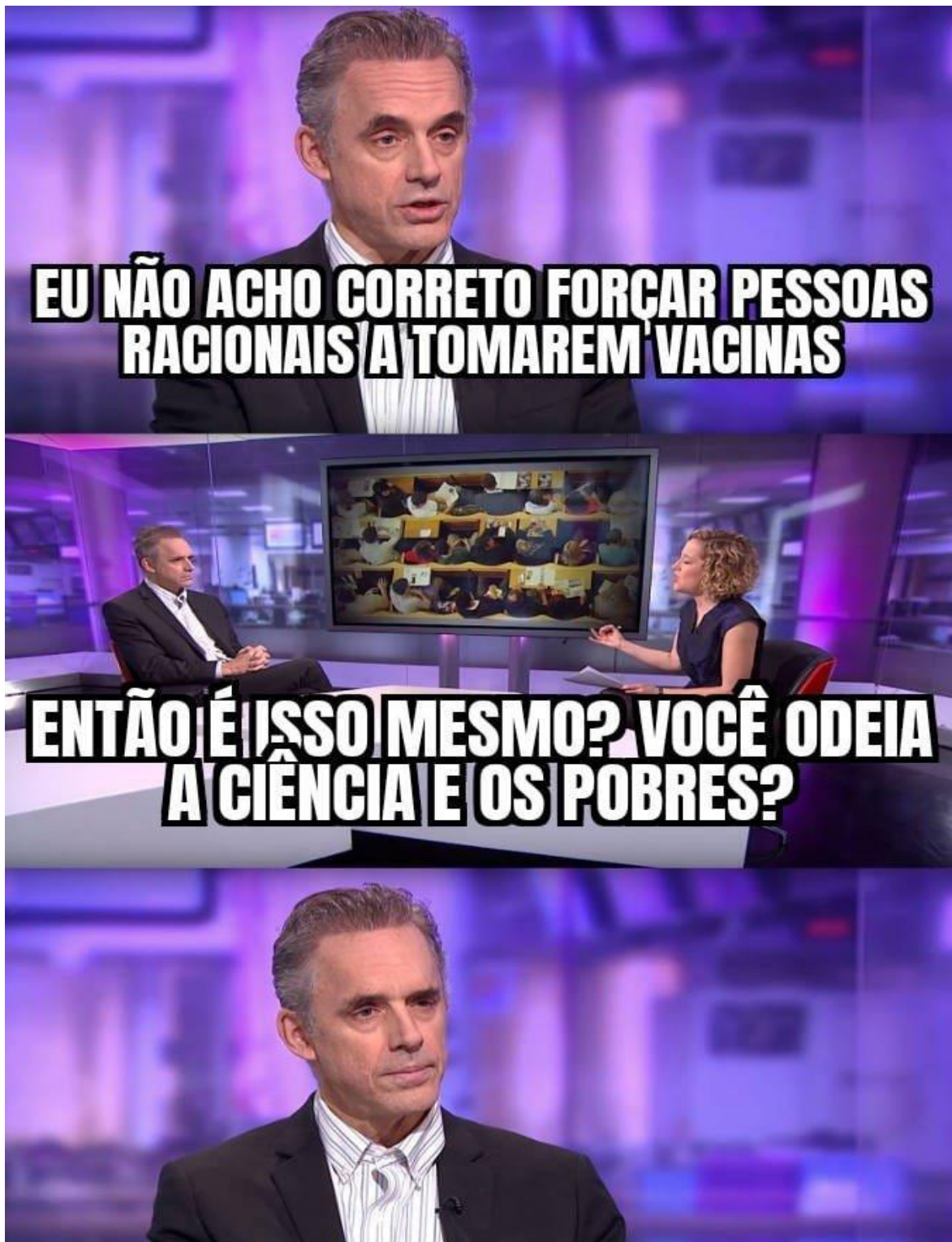
4.3 Dimensão 3 - Discurso da relação de poder e exploração da vacina como benefício pessoal

A dimensão 3, marcada apenas pelo polo positivo, apresentou quatorze variáveis semânticas: *mental actions and processes*, ações e processos mentais; *medicines and medical treatment*, medicamentos e tratamento médico; *Work and employment*, Trabalho e emprego; *Furniture and household fittings*, móveis e acessórios de casa; *Recorded sound*, som gravado; *Speech*, discurso; *Information technology and computing, science and technology in general*, tecnologia da informação e computação; *Science and technology in general*, ciência e tecnologia em geral; *Power relationships*, relações de poder; *parts of building*, componentes prediais; *social actions and processes*, ações, estados e processos sociais; *Geographical names*, nomes geográficos; *Time: Old, new and young*; tempo- velho, novo e jovem.

No geral, houve a prevalência das variáveis semânticas de Discurso (*speech*), representando não somente ações e processos sociais (*social actions and processes*), mas também marcando as relações de poder (*Power relationships*) nos tweets que pontuaram neste polo (Fator 3). Seguem três exemplos:

BERBER SARDINHA, Tony; ALENCAR, André Luiz Siqueira; DA SILVA, Cicero Soares; GIL, Cristina Borges; LOPES, Maurício José Ferreira; HUGHES, Stephan Arthur Solomon. "#EUNAOVOUTOMARVACINA": Uma abordagem da linguística de corpus e da análise multimodal imagética. *Revista Intercâmbio*, v.LI: 178-209, 2022. São Paulo: LAEL/PUCSP. ISSN 2237-759X

Figura 12, 13 e 14 – Dimensão 3 – Discurso da relação de poder e exploração da vacina como benefício pessoal.



BERBER SARDINHA, Tony; ALENCAR, André Luiz Siqueira; DA SILVA, Cicero Soares; GIL, Cristina Borges; LOPES, Maurício José Ferreira; HUGHES, Stephan Arthur Solomon. "#EUNAOVOUTOMARVACINA": Uma abordagem da linguística de corpus e da análise multimodal imagética. *Revista Intercâmbio*, v.LI: 178-209, 2022. São Paulo: LAEL/PUCSP. ISSN 2237-759X

De um lado o Exterminador do Futuro e do outro lado o ator Arnold Schwarzenegger.



BERBER SARDINHA, Tony; ALENCAR, André Luiz Siqueira; DA SILVA, Cicero Soares; GIL, Cristina Borges; LOPES, Maurício José Ferreira; HUGHES, Stephan Arthur Solomon. "#EUNAOVOUTOMARVACINA": Uma abordagem da linguística de corpus e da análise multimodal imagética. *Revista Intercâmbio*, v.LI: 178-209, 2022. São Paulo: LAEL/PUCSP. ISSN 2237-759X



1 - Tuíte número 1339266621580390409: "#EuNaoVouTomarVacina <https://t.co/wSrI6vsYFY>",

2 - Tuíte número 1343359460564209664: "É a pura realidade, não é @jdoriajr ? Acabando com São Paulo, vamos te dar o troco brevemente nas urnas, canalha!#EuNaoVouTomarVacina <https://t.co/vSKaNmJs77>",

3 - Tuíte número 1339375012135006208: "#EuNãovouTomarVacina <https://t.co/7cKshccais>",

Fonte: pesquisadores

5. Considerações finais

A pesquisa apresentada neste artigo buscou revelar quais são as dimensões temáticas de variação subjacentes na *hashtag* #EuNaoVouTomarVacina da rede social *Twitter*. Os resultados mostraram três dimensões temáticas de variação do *corpus hashtag* #EuNaoVouTomarVacina da rede social *Twitter*.

A Dimensão 1, nomeada "Ironia ao fanatismo coletivo versus Politização/Judicialização da vacina" ironiza a cegueira coletiva de determinados estratos sociais no que diz respeito à negação da vacina.

Já a Dimensão 2, nomeada de "Sátira ao discurso antivacina", representada pela expressão facial de cada pessoa, em diferentes momentos, cenários e situações contextuais, um ato de protesto, satiriza um mal-estar contido internamente nestas pessoas, não percebendo que toda essa emoção está sendo expressa ou exteriorizada.

E, por último, a Dimensão 3, "Discurso da relação de poder e exploração da vacina como benefício pessoal", sugerindo ações, estados e processos sociais através de entidades geográficas não apenas o tempo: velho, novo e jovem, mas também ações e processos sociais relacionados ao poder.

Foi com surpresa que observamos, ao pesquisar #EuNaoVouTomarVacina, que na *hashtag* havia tanto o discurso pró-vacina quanto o discurso antivacina.

A Dimensão 1: Ironia ao fanatismo coletivo versus Politização/Judicialização da vacina revelou que no discurso pró-vacina, marcado subjacentemente nas imagens do polo positivo (vide figura 07), havia uma predominância da ironia em relação a quem se opunha à vacinação. Outra revelação foi o discurso da judicialização e a politização, marcado latentemente nas imagens do polo negativo (vide figuras 08 e 09), uma questão que, a princípio, deveria dizer respeito à esfera da saúde pública.

A Dimensão 2: Sátira ao discurso antivacina nos indica o deboche a quem questiona a eficácia da vacina e reforça o que havia sido pontuado na dimensão 1, o escárnio em relação a esse discurso negacionista.

A Dimensão 3: Discurso da relação de poder e exploração da vacina como benefício pessoal ressalta, por sua vez, o pouco comprometimento do poder público com a saúde da população, assim como reforça a politização dessa questão como anteriormente apontado na Dimensão 1.

Apesar da Análise Multidimensional Imagética ainda estar dando os seus primeiros passos, com esta pesquisa esperamos ter demonstrado que a incorporação de diferentes recursos e meios comunicativos, e, principalmente, a multimodalidade, os eventos comunicativos texto-

BERBER SARDINHA, Tony; ALENCAR, André Luiz Siqueira; DA SILVA, Cicero Soares; GIL, Cristina Borges; LOPES, Maurício José Ferreira; HUGHES, Stephan Arthur Solomon. "#EUNAOVOUTOMARVACINA": Uma abordagem da linguística de corpus e da análise multimodal imagética. *Revista Intercâmbio*, v.LI: 178-209, 2022. São Paulo: LAEL/PUCSP. ISSN 2237-759X

imagens, na análise de mídias sociais, torna mais robusta a investigação que até então era feita apenas com o texto.

Referências bibliográficas

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

BERBER SARDINHA, T. *Corpus linguistics and the study of social media: a case study using multi-dimensional analysis*. *The Routledge Handbook of Corpus Linguistics*. 2nd edition. Routledge, 2022.

BERBER SARDINHA, T. *International Perspectives on Corpus Technology for Language Learning*, U. of Queensland, Brisbane, Australia, 11/30/2021.

BERBER SARDINHA, T.; VEIRANO PINTO, M. *Multi-dimensional analysis: research methods and current issues*. Londres: Bloomsbury, 2019.

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

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

BIBER, D. *Multidimensional approaches*. In: LÜDELING, A. KYTÖ, M. (Eds.). *Corpus linguistics: An international handbook v.2*. Berlin/New York: Walter de Gruyter, 2009.

BRAVO, Hebe. *Microblogging, hacia un futuro móvil e inalámbrico*. Disponível em: <<http://www.maestrosdelweb.com/editorial/microblogging-hacia-un-futuro-movil-e-inalambrico/>> Acesso em: 15 janeiro de 2021.

CALLOW, J. *The shape of the text to come - When image and text meet: teaching with visual and multimodal texts*. Disponível em: <<http://wweaching Resources/PPs/PETAA PAPER 181.aspx>>. 2011.

CAMBRIDGE DICTIONARY. Disponível em <https://dictionary.cambridge.org>. Acesso em 12/10/2022.

CARVALHO, J. *Bestializados*; CASTRO, S. *República*; CHALHOUB, S. *Cidade. A trajetória de Oswaldo Cruz e sua luta como médico sanitário no século 19; Fiocruz*. Disponível em:

BERBER SARDINHA, Tony; ALENCAR, André Luiz Siqueira; DA SILVA, Cicero Soares; GIL, Cristina Borges; LOPES, Maurício José Ferreira; HUGHES, Stephan Arthur Solomon. "#EUNAOVOUTOMARVACINA": Uma abordagem da linguística de corpus e da análise multimodal imagética. *Revista Intercâmbio*, v.LI: 178-209, 2022. São Paulo: LAEL/PUCSP. ISSN 2237-759X

<<http://www.fiocruz.br/cgi/cgilua.exe/sys/start.htm?infoid=1084&sid=194%3E>> Acesso em 20 de janeiro de 2022.

CASELLI, Thais; PIMENTA, Francisco. Twitter: a nova ferramenta do jornalismo. In: XVI Congresso Brasileiro Ciências da Comunicação. 2011. p.5

CASTRO, F. R. *O negacionismo do Holocausto: pseudo-história pública*. RESGATE - Vol. XXII, N.28 - Jul./dez. 2014 (pp.5-12).

CHRISTIANSEN, A.; DANCE, W.; WILD, A. *Constructing corpora from images and text: An introduction to visual Constituent Analysis. Chapter 7 (pp.149-174)*. In: RUDIGER, S.; DAYTER, D. *Corpus approach to Social Media*, 2020 – John Benjamins B.V.

COMM, J; BURGE, K. *O poder do Twitter: estratégias para dominar seu mercado e atingir seus objetivos com um tweet por vez*. São Paulo: Gente, 2009.

CONRAD, S.; BIBER, D. *Variation in English. Multidimensional Studies*. Pearson, London. 2001.

ESCOBAR, H. *A Ciência contra o negacionismo*. *Jornal da USP*, São Paulo, 22/01/2021. Acesso em 02 de fevereiro de 2022.

GEENG, C., SAVANNA Y., ROESNER, F. *Fake news on Facebook and Twitter: Investigating how people (don't) investigate"* *Proceedings of the 2020 CHI conference on human factors in computing systems*. 2020.

KAUFFMANN, C. *Linguística de corpus e estilo: análises multidimensional e canônica na ficção de Machado de Assis*. [s.l.] Pontifícia Universidade Católica de São Paulo, 2020.

KRESS, G. R. *Multimodality: A social semiotic approach to contemporary communication*, Taylor & Francis. New York: 2010.

KRESS, G. R. *Multimodal discourse analysis from: The Routledge Handbook of Discourse Analysis* Routledge. 2011. Disponível em: <<https://www.routledgehandbooks.com/doi/10.4324/9780203809068.ch3>> Acesso em 09 de maio de 2022.

BERBER SARDINHA, Tony; ALENCAR, André Luiz Siqueira; DA SILVA, Cicero Soares; GIL, Cristina Borges; LOPES, Maurício José Ferreira; HUGHES, Stephan Arthur Solomon. "#EUNAOVOUTOMARVACINA": Uma abordagem da linguística de corpus e da análise multimodal imagética. *Revista Intercâmbio*, v.LI: 178-209, 2022. São Paulo: LAEL/PUCSP. ISSN 2237-759X

KRESS, G. R.; VAN LEEUWEN, T. *Reading images: the grammar of visual design*. London, New York: Routledge, [1996], 2006.

KRESS, G. R.; VAN LEEUWEN, T. *Multimodal discourse: the modes and media of contemporary communication*. Hodder Education, Oxford University Press Inc., New York, 2001.

PINHEIRO, M. S. *Multimodalidade e Letramento Visual na sala de aula de Língua Espanhola: análise de uma atividade de produção escrita*. V. 16, n. 4, pp.575-593. RBLA, Belo Horizonte, 2016.

PORFÍRIO, F. "O que é negacionismo?." *Brasil Escola*. Disponível em: <<https://brasilecola.uol.com.br/o-que-e/o-que-e-sociologia/o-que-e-negacionismo.htm>>. Acesso em 11 de fevereiro de 2022.

SANCHEZ, A." Definición e historia de los corpus". In: SANCHEZ, A. et al. (orgs.). *CUMBRE: corpus lingüístico de Español contemporaneo*. SGEL, 1995. p.7 -24.

SEIXAS, Fábio. Micro-blogging. In: SPYER, Juliano (Org.). *Para entender a internet: noções, práticas e desafios da comunicação em rede*. Disponível em: <<http://stoa.usp.br/oerworkshop/files/1333/7925/Para+entender+a+Internet.pdf>> Livro publicado em 2009 e distribuído gratuitamente somente pela internet. Acesso em 02 de abril de 2009.

TOLEDO, K. *Negacionismo científico: a produção política e cultural da desinformação*. Agência FAPESP. 02/09/2020.

VALIM, P.; SÁ AVELAR, A. *Negacionismo histórico: entre a governamentalidade e a violação dos direitos fundamentais*. Disponível em: <<https://revistacult.uol.com.br/home/negacionismo-historico/>> Acesso em: 05 de maio de 2022.

WENGER, E. *Communities of practice: Learning, meaning, and identity*. Cambridge University Press, 1999.

Recebido em: 21/06/2022
Aprovado em: 06/10/2022