REVISTA EntreLínguas
Ensino e aprendizagem de Línguas Estrangeiras Modernas

INFODEMIC AND XENOPHOBIC DISCOURSES AGAINST CHINA ON TWITTER DURING THE COVID-19 PANDEMIC: A LEXICAL MULTIDIMENSIONAL ANALYSIS

INFODEMIA E DISCURSOS XENOFÓBICOS CONTRA A CHINA NO TWITTER DURANTE A PANDEMIA DE COVID-19: UMA ANÁLISE MULTIDIMENSIONAL LEXICAL

INFODEMIA Y DISCURSOS XENOFÓBICOS CONTRA CHINA EN TWITTER DURANTE LA PANDEMIA DE COVID-19: UN ANÁLISIS MULTIDIMENSIONAL LEXICAL

Cícero SOARES DA SILVA[1]
e-mail: pardonmester@gmail.com

Eduardo Batista da Silva[2]
e-mail: eduardo.silva@ueg.br

Aline Andrea Zamboni Milanez[3]
e-mail: alinezmilanez@gmail.com

Amanda Chiarelo Boldarine[4]
e-mail: amandaboldarine@gmail.com

Tony Berber Sardinha[5]
e-mail: tonycorpuslg@gmail.com

[1] Pontifical Catholic University of São Paulo (PUCSP), São Paulo – SP – Brazil. Ph.D. candidate (LAEL), Applied Linguistics and Language Studies.
[2] State University of Goiás (UEG), Morrinhos – GO – Brazil. Professor (POSLLI), Stricto Sensu Graduate Program in Language, Literature, and Interculturality.
[3] Pontifical Catholic University of São Paulo (PUCSP), São Paulo – SP – Brazil. Ph.D. (LAEL), Applied Linguistics and Language Studies.
[4] Pontifical Catholic University of São Paulo (PUCSP), São Paulo – SP – Brazil. Ph.D. candidate (LAEL), Applied Linguistics and Language Studies.
[5] Pontifical Catholic University of São Paulo (PUCSP), São Paulo – SP – Brazil. Coordinator and Professor (LAEL), Applied Linguistics and Language Studies.

How to reference this paper:

SOARES DA SILVA, C.; SILVA, E. B.; MILANEZ, A. A. Z.; BOLDARINE, A. C.; BERBER SARDINHA, T. Infodemic and xenophobic discourses against China on Twitter during the covid-19 pandemic: a lexical multidimensional analysis. **Rev. EntreLínguas**, Araraquara, v. 11, n. 00, e025013, 2025. e-ISSN: 2447-3529. DOI: 10.29051/el.v11i00.19844

| **Submitted**: 25/11/2024
| **Revisions required**: 26/03/2025
| **Approved**: 15/10/2025
| **Published**: 20/12/2025

ABSTRACT: This study investigates the dimensions of variation based on the patterns of cooccurrence of lexical elements on Twitter present in the infodemic discourses of anti-Chinese xenophobia during the COVID-19 pandemic in Brazil. The methodology used was Lexical Multidimensional Analysis (Berber Sardinha, 2019; Berber Sardinha; Fitzsimmons-Doolan, 2025). This study used the Corpus of Anti-Chinese Xenophobia (COXAC) in Portuguese, which comprises ca. 100K tweets and 1,697,408 words. We analyzed the hashtags #viruschines, #pragachinesa, #pestechinesa, #pasteldeflango, #boicoteachina, #vachina, #ChainaVirus, and #wuhanvirus. Six discursive dimensions were identified: 1) Pandemic manipulation versus Political mockery; 2) Rejection of political establishment versus Misrepresentation of culture; 3) Criticism of corporate media versus Rejection of Chinese vaccines; 4) Suppression of freedom versus Reallocation of government funds; 5) Overthrow of the political system versus Misrepresenting Chinese eating habits; 6) Anti-China campaign versus Pandemic hoax framing. This research can serve as foundation for intercultural awareness work, providing evidence to combat disinformation.

KEYWORDS: Infodemics. Xenophobia. Lexical Multidimensional Analysis. Artificial Intelligence. Intercultural Awareness.

***RESUMO**: Este estudo investiga as dimensões de variação, examinando os padrões de coocorrência de elementos lexicais no Twitter presentes nos discursos infodêmicos da Xenofobia antichinesa durante a pandemia de covid-19 no Brasil. A metodologia empregada foi a Análise Multidimensional Lexical (Berber Sardinha, 2019; Berber Sardinha; Fitzsimmons-Doolan, 2025). O estudo utilizou o Corpus de Xenofobia Antichinesa (COXAC) em português, composto por 100 mil tuítes e 1.697.408 palavras. Analisamos as hashtags #viruschines, #pragachinesa, #pestechinesa, #pasteldeflango, #boicoteachina, #vachina, #ChainaVirus e #wuhanvirus. Seis dimensões discursivas foram identificadas: 1) Manipulação pandêmica versus Ridicularização política; 2) Crítica ao sistema versus Sentimento antichina; 3) Crítica à mídia versus Rejeição às vacinas chinesas; 4) Supressão da liberdade versus Realocação de fundos governamentais; 5) Derrubada do sistema político versus Distorção alimentar chinesa; 6) Campanha antichina versus Conspiração pandêmica. Esta pesquisa pode servir como um ponto de partida para um trabalho de conscientização intercultural, oferecendo evidências para o combate à desinformação.*

***PALAVRAS-CHAVE**: Infodemia. Xenofobia. Análise Multidimensional Lexical. Inteligência Artificial. Conscientização Intercultural.*

***RESUMEN**: Este estudio investiga las dimensiones de variación examinando los padrones de coocurrencia de elementos léxicos en Twitter en los discursos infodémicos de xenofobia antichina durante la pandemia de COVID-19 en Brasil. La metodología utilizada fue el Análisis Multidimensional Lexical (Berber Sardinha, 2019; Berber Sardinha; Fitzsimmons-Doolan, 2025). Este estudio utilizó el Corpus de Xenofobia Antichina (COXAC) en portugués, compuesto por 100 mil tuits y 1.697.408 palabras. Analizamos los hashtags #viruschines, #pragachinesa, #pestechinesa, #pasteldeflango, #boyicoteachina, #vachina, #ChainaVirus y #wuhanvirus. Fueron identificadas seis dimensiones: 1) Manipulación de la pandemia versus Absurdo político; 2) Crítica al sistema versus Sentimiento antichina; 3) Crítica a los medios versus Rechazo a las vacunas chinas; 4) Supresión de la libertad versus Reasignación de fondos gubernamentales; 5) Destituición del sistema político versus Distorsión alimentaria china; 6) Campaña antichina versus Conspiración pandémica.* Esta investigación puede servir como punto de partida para la sensibilización intercultural, ofreciendo evidencia para combatir la desinformación.

PALABRAS CLAVE: *Infodemia. Xenofobia. Análisis Multidimensional Lexical. Inteligencia artificial. Conciencia intercultural.*

Research Contextualization

On March 17, 2020, former United States President Donald Trump labeled the coronavirus as the "Chinese virus" and posted the following tweet: "The United States will be powerfully supporting those industries, like Airlines and others, that are particularly affected by the Chinese Virus. We will be stronger than ever before" (Dubey, 2020, p. 2). Aligned with Trump's xenophobic stance, Eduardo Bolsonaro, a federal congressman and son of former Brazilian President Jair Bolsonaro, also used the term "Chinese virus" in one of his Twitter posts, which fueled anti-Chinese sentiment in the country and resulted in diplomatic tensions between Brazil and China (Patrick, 2020).

With regard to pandemics or epidemics, it is evident that, throughout history, marks of stigma and intolerance toward the other have consistently been manifested through language. By means of lexical indices—more specifically, "national adjectives" or "demonyms" (Ribeiro, 2022)—to name a disease or virus according to our identities and political and ideological positions, we inevitably create divisions within society. By convention, we establish that the side where the "I" and the "we" presumably reside is associated with positivity, whereas the other side, where "they" and "the others" oddly converge, is framed negatively.

Drawing on the symbolism of the "West versus East" label, the COVID-19 pandemic generated, both in the sphere of international politics and in the political–health arena, feelings of hatred, fear, aversion, and distrust toward migrants and people of Asian descent, especially Chinese individuals. According to Ribeiro (2022), Chinese people were stigmatized by a discourse that accused them of being the first disseminators of COVID-19 worldwide.

In Brazil, it is important to note that racial issues were already embedded in immigration projects in the nineteenth century (Seyferth, 2014). An immigration process took place between 1888 and 1914, driven by the need to replace enslaved labor on coffee plantations (Seyferth, 2014). However, policies grounded in eugenic and assimilationist ideas restricted the entry of groups deemed ethnically undesirable.

According to Kohatsu, Saito, and Andrade (2021), in the 1850s Chinese workers were rejected for temporary employment on coffee plantations. This situation generated debates with the Ministry of Agriculture, Commerce, and Public Works, which justified the exclusion of this group from the country based on prejudiced conceptions of a "mongrel race" and a civilization described as "decadent and corrupted by opium" (Seyferth, 2014, p. 118, our translation).

The origin of the word *xenophobia* derives from Greek, composed of the words *xenos*, meaning "foreigner," and *phóbos*, meaning "fear" (La Garza, 2011, p. 1). Xenophobia refers to feelings of hatred and fear and describes prejudice toward those perceived as a threat, such as foreigners and out-groups. According to She *et al*. (2022, p. 1), "Xenophobia increases during periods of disease threat because outsiders are believed to be carriers of germs and infections".

Introduction

Considering that pandemics and epidemics are historically grounded in a pattern of assigning blame during previous health crises, xenophobic representations against China, its people, and its culture have used social media as a means to disseminate sinophobia. Within this context, Twitter became a genuine arena in which antagonistic groups confronted one another and positioned themselves politically and ideologically, spreading misinformation about COVID-19 through posts. This scenario gave rise to so-called echo chambers, defined as "environments in which a person encounters only information or opinions that reflect and reinforce their own" (Cambridge University Press and Assessment, 2024).

In the literature review, previous studies have addressed xenophobic attitudes and stigmatization in the context of pandemics and epidemics, including the works of Person *et al*. (2004), Lee *et al*. (2005), Hoppe (2018), and Reny and Barreto (2020).

In the study by Person *et al*. (2004) on stigma, it was revealed that during the outbreak of Severe Acute Respiratory Syndrome (SARS)—a disease caused by a coronavirus—in the United States, many Asian Americans were targets of stigmatization.

A similar experience also occurred during the H1N1 outbreak. In this case, Lee *et al*. (2005) observed that xenophobic attitudes were directed toward people from Mexico, the first country to report cases of the H1N1 outbreak. At that time, many global citizens stigmatized most Mexicans, resulting in exclusion, rejection, and marginalization.

Hoppe (2018) highlights a crucial issue that reveals complexities in social and political realities: the absence, in history, of an epidemic episode colloquially known as the "American flu" or the "European flu." By raising this point, Hoppe suggests that it reflects global inequalities that enable powerful countries in the Global North to attribute responsibility for diseases and epidemics to foreigners and foreign nations, rather than assuming responsibility themselves (Hoppe, 2018, p. 1463, our translation).

For Reny and Barreto (2020), in the case of the "Chinese virus," the lexical index used to name the disease as "Chinese virus" also encompasses a xenophobic ideology. When connected to political circumstances—especially those involving the United States and China—the "Chinese virus" discourse in memes can even be used as a means of glorifying anti-Chinese sentiment in the United States or, more broadly, in a global context.

An "An 'infodemic' is defined as "an overabundance of information - some accurate and some not-occurring during an epidemic" (Zielinski, 2021). Although the term applies to other epidemics, it was during the COVID-19 pandemic that the infodemic crisis intensified, due to the volume and speed of information circulation at the global level (Zielinski, 2021), driven by social media platforms (Banerjee; Meena, 2021).

In this study, we argue that the infodemic does not constitute a chaotic set of information and misinformation, but rather a complex ecosystem of discourses that shape a coherent set of ways of thinking and acting. Given that social groups hold systematic worldviews and ideologies, it is reasonable to assume the existence of initiatives aimed at defending and disseminating group-based perspectives, ultimately leading to their acceptance. In the post-pandemic context, despite the urgency of discussions surrounding the infodemic, we observe a lack of recognition of the role of discourse in shaping public opinion and public policy.

In light of the above, we outline a linguistic–computational investigation employing Multidimensional Lexical Analysis to examine xenophobic representations—infodemic discourse—against China on Twitter during the COVID-19 pandemic. To assess infodemic discourse, we manually coded a sample of 3,000 tweets.

According to Berber Sardinha and Moreira (2023), the purpose of infodemic discourses is "to support antidemocratic causes or ideologies through the rapid and large-scale circulation of information on social media" (Berber Sardinha & Moreira, 2023, p. 1). In the context of the COVID-19 pandemic, the infodemic involved an uncontrolled dissemination of information and became a widely perceptible phenomenon, permeating all sectors of society, influencing public opinion, and triggering debates and conflicts regarding strategies to address the health crisis (Banerjee; Meena, 2021; Mesquita *et al*., 2020). According to Zielinski (2021), it is important to emphasize that the infodemic is not limited to the COVID-19 pandemic, suggesting that the experiences and lessons learned from this phenomenon may offer valuable guidelines for the future.

We draw on Multidimensional Lexical Analysis (Berber Sardinha, 2019; Berber Sardinha; Fitzsimmons-Doolan, 2025; Fitzsimmons-Doolan, 2019, 2023; Clarke, 2022), a corpus-based approach designed to detect parameters of lexical variation within a corpus, which may signal themes, topics, and broader discursive patterns, using large datasets (big data).

The present study aims to investigate the dissemination of infodemic discourses that promoted anti-Chinese xenophobia during the COVID-19 pandemic. Through a detailed analysis of messages posted on Twitter, we seek to understand how such discourses were disseminated, particularly against China, its people, and its culture. Our specific objective is, first, to problematize infodemic discourses, their lexical content, and their social impact with regard to the promotion of falsehoods, limiting beliefs, and prejudiced views.

To capture these representations, we compiled a specific corpus and employed linguistic–computational software for data processing. A machine learning model was trained specifically for the purposes of this research (see Section 3).

Considering the scope and urgency of the topic addressed here, this study is intended for a broad target audience, including students, teachers, and researchers in the field of Language and Linguistics, as well as in other areas of knowledge.

Theoretical Framework

Multidimensional Lexical Analysis (hereafter MLA) (Berber Sardinha, 2014; 2019; Berber Sardinha; Fitzsimmons-Doolan, 2025; Fitzsimmons-Doolan, 2019, 2023; Clarke *et al*., 2022) derives from Multidimensional Grammatical–Functional Analysis (Biber, 1988, 2009; Berber Sardinha; Veirano Pinto, 2014, 2019). Developed by Berber Sardinha (2014), MLA can be understood as a branch of Corpus Linguistics aimed at detecting parameters of variation in correlated lexical items, which can be interpreted as indices of the presence of discourses within a corpus.

The MLA also employs multivariate statistical analysis, particularly factor analysis, to detect latent variables—those that operate below the speaker's immediate perception. These latent variables are realized as dimensions of variation, understood as sets of correlated lexical items that tend to co-occur in texts (written, spoken, visual, etc.) (Berber Sardinha, 2023).

MLA has previously been applied to lexical analyses. Crossley and Louwerse (2007) identified dimensions of variation using bigrams in an English-language corpus. Zuppardi and Berber Sardinha (2020) produced collocation lists for academic writing in English, and Berber

Sardinha, Acunzo, and Ferreira (2016) conducted a multidimensional lexical study identifying collocational dimensions in a Brazilian Portuguese corpus.

In 2020, Berber Sardinha explored cultural representations associated with the lexical indices (bigrams) “American” and “Brazilian” in a diachronic corpus composed of English-language texts drawn from the Google Books n-grams corpus, a database comprising documents (primarily books) published between 1800 and 2008. In this study, the lexical indices “American” and “Brazilian” were used to identify parameters of national and cultural identity representation in order to understand what it has meant to be American or Brazilian over time, as represented in the corpus. Five dimensions were identified for “American,” and five dimensions were revealed for “Brazilian,” as shown in Chart 1 below.

Chart 1 – Dimensions for *American* and *Brazilian*

Dimensions for “American”	Dimensions for “Brazilian”
1. Superpower versus regional status	1. Economy and politics
2. People, the flag, and institutions	2. Traditional arts, sciences, people, and land
3. Individuals, community, and culture	3. Raw materials and landscape
4. Armed forces, slavery, and ideals	4. New artistic forms, women and men, religion, and the environment
5. Literate expression versus revolution and the new nation	5. The monarchy, steam transport, and uninhabited areas

Source: Adapted from Berber Sardinha (2020).

The study demonstrated that a methodology based on Multidimensional Analysis proved effective in detecting discursive and cultural representations. The association between MLA and Discourse Analysis makes it possible to delineate discourses based on indices of their materialization in language.

Corpus-Assisted Discourse Analysis (CADA) is a multidisciplinary field focused on identifying discourses through corpus Analysis (Friginal; Hardy, 2020). Although there are multiple definitions of discourse, two are highlighted here: a set of meanings, representations, and statements that, in some way, generate a particular version of events (Burr, 1995, p. 48); and ways of seeing the world, constructing objects and concepts in specific manners, and representing reality (Baker; Mcenery, 2015, p. 5).

Thus, discourse is an abstract phenomenon that encompasses the values and ideologies of historically defined groups or sectors of society. Although abstract, discourses are materialized in language in use. Because such materiality can be detected computationally, it is possible to identify discourse indices quantitatively. At the same time, CADA requires qualitative analysis, since discourses are not automatically derived from their indices. Interpretation of materiality in light of context and human knowledge is therefore essential.

From Pêcheux's perspective (1982, p. 82, our translation), discourse is understood as "the effect of meanings between interlocutors and represents the point of encounter between language and ideology," leading to the understanding that ideology is expressed through language itself. Orlandi (2015, p. 36, our translation) supports this conception by stating that "all discourse is ideologically marked, and it is in language that ideology finds its materialization. Therefore, discourse is the space in which language and ideology interact."

We also draw on Serge Moscovici's theory of social representations. According to Moscovici (2012, p. 39, our translation), "these social representations are composed of a complex interaction of informative, cognitive, ideological, and normative elements, as well as beliefs, values, attitudes, opinions, and images."

Taking a historical retrospective in the context of pandemics, studies by Dionne and Turkmen (2020), Ittefaq *et al*. (2022), White (2020), and Mansouri (2020) argue that xenophobia may manifest through othering, branding, scapegoating, and racialization. These are psychosocial mechanisms that individuals and groups use to create and reinforce collective identity, establish social hierarchies, and justify hostility toward other groups (Dionne; Turkmen, 2020; Ittefaq *et al*., 2022; Mansouri, 2020; White, 2020).

Despite efforts by the World Health Organization (WHO) to discourage the use of national adjectives or stigmatizing terms in disease nomenclature—such as "Ebola virus," "Spanish flu," and "Mexican swine flu," as well as labels like *Spanish Lady* and *Blackman's disease*—xenophobic practices persisted during the COVID-19 pandemic. The term "Chinese virus" produced effects similar to those of earlier disease names with regard to stigmatization.

Materials and Methods

The materials used in this study were: tweets from the social media platform Twitter—currently known as X; the Anti-Chinese Xenophobia Corpus (COXAC), compiled specifically for this research; the TreeTagger tagger; a Unix Shell script; SAS OnDemand software; the Python programming language; and the FastText software.

The methodological procedures comprised the following stages: collection, selection, and analysis of tweets; morphosyntactic tagging and lemmatization of the tweets; machine learning and lexical annotation; selection of the classification model; evaluation by the researchers; and, finally, corpus processing using SAS OnDemand. The following paragraphs describe in detail the steps carried out with the materials listed above.

We compiled the Anti-Chinese Xenophobia Corpus (COXAC), a corpus of tweets consisting of 100,000 posts in Brazilian Portuguese collected from Twitter—illustrated in Figure 1—covering the period from 2020 to 2022 and containing the following hashtags: #viruschines, #víruschines, #víruschinês, #viruschiunês, #pragachinesa, #pestechinesa, #pasteldeflango, #boicoteachina, #vachina, and #ChainaVirus. The *snscraper* command generated a file containing 100,000 tweets, each associated with ten search hashtags. The resulting file was saved in JSON format and structured as delimited fields.

Figure 1 – Samples of Twitter posts

```
--- BR POST: 000751 ---
 Agências de suposto fact-checking correram para gritar “É mentira
 que a China não usa Coronavac!!” naqueles movimentos ultra-espontâneos.
 A tática é só dizer isso no título, mas nos textos, que ninguém
 lê, apenas se diz que preferiram a outra, mas nada de errado,
 veja bem...
```

Source: Prepared by the authors.

The cleaned and standardized corpus presents the characteristics shown in Table 1.

Table 1 – *Corpus* design

Year	Tweets	Lexical forms
2020	73.781	1.525.503
2021	7.080	154.946
2022	795	16.959
Total	81.656	1.697.408

Source: Prepared by the authors.

Using the TreeTagger tagger, it was possible to perform both morphosyntactic tagging and lemmatization of the tweets.

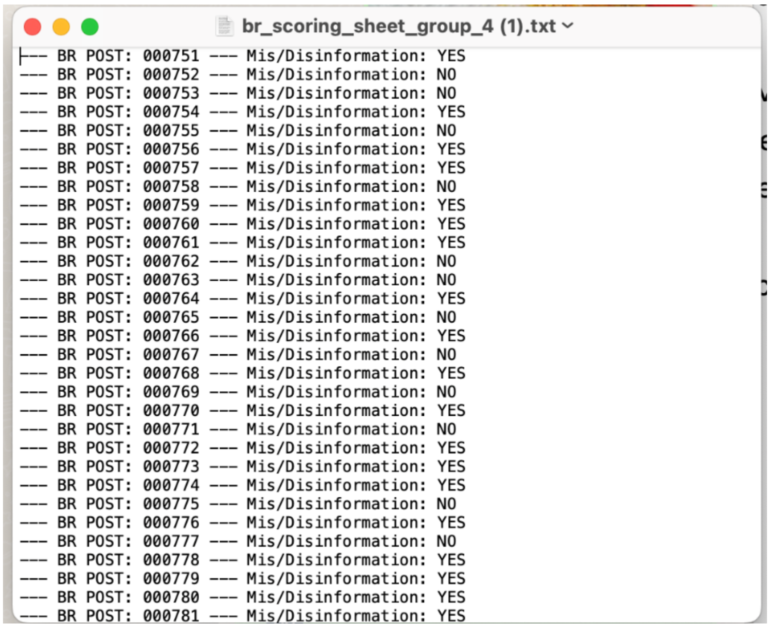
With the FastText tool—an open and free library that enables the creation of text representations and classifiers—we developed a model for identifying xenophobic discourse using artificial intelligence. However, it was first necessary to train the model, that is, within the scope of this study, to teach the machine to classify tweets as having or not having a xenophobic meaning effect.

To this end, we selected hashtags relevant to the scope of our research, namely: #viruschines, #víruschines, #víruschinês, #viruschiunês, #pragachinesa, #pestechinesa,

#pasteldeflango, #boicoteachina, #vachina, #ChainaVirus, and #wuhanvirus. These hashtags were drawn from 500 Brazilian influencer accounts.

In order to train the FastText software, it was necessary to manually annotate 250 tweets in Portuguese with the label "Mis/Disinformation." The annotation categorization—yes or no—was based on the answer to the following question: *Does the meaning effect of the post below contribute to a xenophobic cause or movement?*

Figure 2 – Annotation Sheet



br_scoring_sheet_group_4 (1).txt

```
├── BR POST: 000751 --- Mis/Disinformation: YES
--- BR POST: 000752 --- Mis/Disinformation: NO
--- BR POST: 000753 --- Mis/Disinformation: NO
--- BR POST: 000754 --- Mis/Disinformation: YES
--- BR POST: 000755 --- Mis/Disinformation: NO
--- BR POST: 000756 --- Mis/Disinformation: YES
--- BR POST: 000757 --- Mis/Disinformation: YES
--- BR POST: 000758 --- Mis/Disinformation: NO
--- BR POST: 000759 --- Mis/Disinformation: YES
--- BR POST: 000760 --- Mis/Disinformation: YES
--- BR POST: 000761 --- Mis/Disinformation: YES
--- BR POST: 000762 --- Mis/Disinformation: NO
--- BR POST: 000763 --- Mis/Disinformation: NO
--- BR POST: 000764 --- Mis/Disinformation: YES
--- BR POST: 000765 --- Mis/Disinformation: NO
--- BR POST: 000766 --- Mis/Disinformation: YES
--- BR POST: 000767 --- Mis/Disinformation: NO
--- BR POST: 000768 --- Mis/Disinformation: YES
--- BR POST: 000769 --- Mis/Disinformation: NO
--- BR POST: 000770 --- Mis/Disinformation: YES
--- BR POST: 000771 --- Mis/Disinformation: NO
--- BR POST: 000772 --- Mis/Disinformation: YES
--- BR POST: 000773 --- Mis/Disinformation: YES
--- BR POST: 000774 --- Mis/Disinformation: YES
--- BR POST: 000775 --- Mis/Disinformation: NO
--- BR POST: 000776 --- Mis/Disinformation: YES
--- BR POST: 000777 --- Mis/Disinformation: NO
--- BR POST: 000778 --- Mis/Disinformation: YES
--- BR POST: 000779 --- Mis/Disinformation: YES
--- BR POST: 000780 --- Mis/Disinformation: YES
--- BR POST: 000781 --- Mis/Disinformation: YES
```

Source: Prepared by the authors.

To ensure inter-rater reliability, each researcher independently annotated 25 tweets—as shown in Figure 2—and the annotation data were used to calculate Cohen's Kappa. The calculation was performed using SAS OnDemand. After each round of annotation and calculation, the researchers discussed disagreements in order to align the classification criteria. After three rounds, the annotators reached a Kappa value of 1.0 (almost perfect agreement). Subsequently, 3,000 tweets were manually annotated.

Given that the study required machine-based annotation, we chose to compare the level of accuracy of manual annotation with that performed by artificial intelligence. A sample of 100 tweets was annotated both by the researchers and by the FastText artificial intelligence model.

Table 2 – Machine Learning Model

Manual annotation		Machine annotation	
Yes	No	Yes	No
29	71	11	66

Source: Prepared by the authors.

As shown in Table 3, in the manual annotation, the researchers labeled 29 posts as "yes," indicating a meaning effect related to sinophobia. The artificial intelligence model assigned "yes" to 11 posts. Regarding the "no" label, indicating the absence of a meaning effect related to sinophobia, the researchers identified 71 posts, while the artificial intelligence model identified 66 posts.

To determine whether automatic annotation would be viable, we used a confusion matrix, the results of which are presented in Table 3.

Table 3 – Confusion Matrix

Elements	%
Recall	37,93
Precision	68,75
Correct classification	77
Incorrect classification	23

Source: Prepared by the authors.

Based on the data provided by the confusion matrix, we observed a 77% rate of correct classification between manual and machine annotation, that is, coinciding "yes" or "no" responses. We considered this value acceptable for the purposes of the study, and the machine learning model was therefore trained using the 3,000 tweets previously labeled manually.

Subsequently, the corpus was processed using a Unix Shell script developed by the supervising professor, which performed the operations described in Chart 2.

Chart 2 – Functions of the corpus processing script developed for the study

Function	Description
Cleantweets	Cleaning of JSON files, reducing them to the fields essential for analysis: text, user, date, and tweet ID
Removedupes	Removal of duplicate tweets from the corpus, retaining only one instance of each
Tokenizing	Tokenization of tweets, that is, insertion of spaces around each word to isolate it from other words and from orthographic elements such as emojis, hashtags, and punctuation
Emoji	Conversion of emojis into descriptive textual labels using the Python library *demoji*
Treetagging	Morphosyntactic tagging and lemmatization of each tweet using the TreeTagger tagger, assigning each word a grammatical class and a canonical lexical form (lemma)

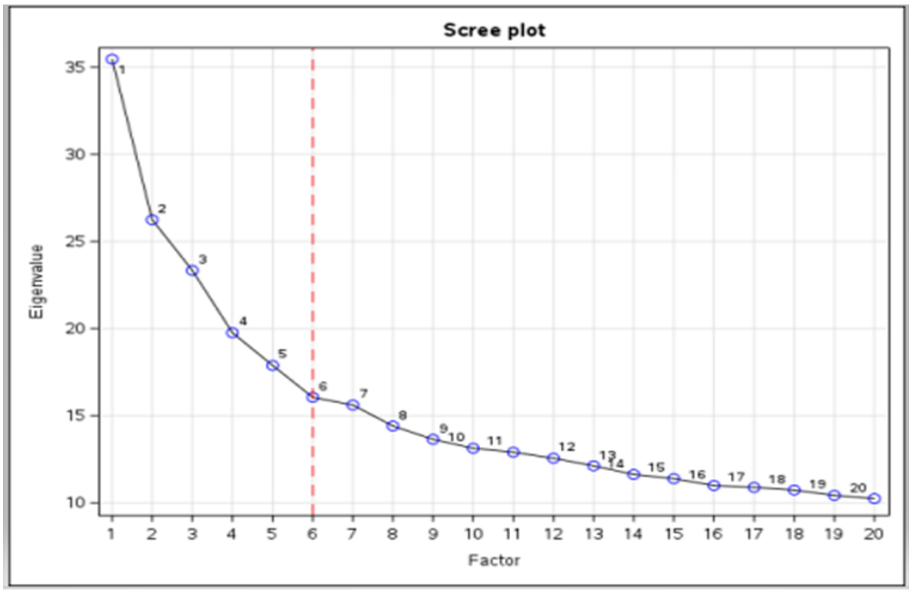
Tokenstypes	Listing of items (tokens) and vocabulary types associated with content word classes, namely nouns, verbs, and adjectives, as well as emojis and hashtags
Removedupes	Removal of partially repeated tweets based on tokens
Toplemmas	Counting of lemmas and listing of the 1,000 most recurrent lemmas, based on the number of tweets in which they occur
Sas	Creation of input files for the SAS OnDemand statistical package
Datamatrix	Calculation of correlations between lemma counts using Python routines
Correlationmatrix	Generation of a correlation matrix formatted for SAS OnDemand
Dates	Listing of tweet dates and creation of a metadata file formatted for SAS OnDemand
Wcount	Listing of tweet length in number of words (tokens) and creation of a metadata file formatted for SAS OnDemand
Formats	Creation of format files for SAS OnDemand
Jqlisting	Generation of a reference file of JSON data to facilitate data extraction in the subsequent routine
Examples	Listing of examples for each dimension

Source: Adapted from Berber Sardinha and Moreira (2023).

The data generated by the script were loaded into the SAS OnDemand statistical analysis package. A SAS Program, that is, a sequence of instructions developed by the supervising professor, was created to perform the statistical analysis required by Multidimensional Lexical Analysis. Essentially, a Factor Analysis was conducted based on the counts of the 1,000 most frequent lemmas, in terms of the number of texts in which they occur. An initial factor analysis produced the eigenvalues associated with the factors, allowing the analyst to determine the optimal number of factors present in the data.

Six factors were identified according to the scree plot generated by SAS, as shown in Figure 3.

Figure 3 – Scree plot produced by Factor Analysis



Source: Prepared by the authors.

A second factor analysis was conducted by extracting exactly six factors, those that presented the greatest variance. Each of these factors comprised two components—or poles—consisting of lemmas distributed across the texts in a complementary manner; that is, when the lemmas of one pole occur in a given text, the lemmas of the opposite pole occur much less frequently or do not occur at all. By convention, these poles are referred to as positive and negative, a designation that merely indicates this complementary distribution and does not imply any evaluative judgment.

The texts were scored on each factor through the sum of the lexical features that loaded on that factor. Accordingly, each tweet received a value corresponding to the incidence of the factor in that text. This procedure made it possible to rank the tweets according to their greater or lesser adherence to each pole of each factor. For the qualitative interpretative analysis, the tweets with the most salient scores in each pole were selected.

Results and data analysis

The factors were qualitatively interpreted as dimensions, such that each factor corresponds to a dimension and each dimension reflects predominant discourses. In this section, we present the findings related to the analysis of the six identified factors:

- Dimension 1 – Pandemic manipulation versus political ridiculing;
- Dimension 2 – Criticism of the system versus anti-China sentiment;
- Dimension 3 – Criticism of the media versus rejection of Chinese vaccines;
- Dimension 4 – Suppression of freedom versus reallocation of government funds;
- Dimension 5 – Overthrow of the political system versus Chinese food distortion; and
- Dimension 6 – Anti-China campaign versus pandemic conspiracy.

For each dimension, a table is presented showing the factorial pattern of both poles, as well as examples of posts that include the most prominent lemmas.

Each table lists the lemma—the canonical form of the word used in the tweet—followed by a numerical value representing the factorial weight, or factor loading, of the lemma in the factor. The lemmas are ordered in descending order by weight, such that the first items have higher weights than subsequent ones. Items marked with the suffix “_e” correspond to emojis represented by their English descriptive labels generated by the *demoji* library, as described in the methodology. Items ending in "_h" are hashtags used in the tweets. Items enclosed in parentheses indicate those with a higher loading on another pole—either within the same factor or in a different factor.

Dimension 1: Political manipulation versus political ridiculing

Chart 3 – Factorial pattern of the positive pole of the first factor

Factorial Pattern
tratamento (.666), isolamento (.656), risco (.618), prefeito (.605), coletivo (.603), medida (.595), eleitoral (.578), campanha (.573), federal (.573), pânico (.564), vida (.561), grupo (.557), público (.553), contágio (.551), óbito (.548), histeria (.546), fome (.546), ação (.541), efeito (.530), abrir (.524), dinheiro (.500), mão (.498), mês (.494), estado (.490), população(.484), salvar(.481), médico (.470), (reduzir(.466)), doença (.463), coronavírus (.462), casa (.458), carnaval (.457), fechar (.455), (fundo (.447)), vazio (.447), cloroquina (.446), hospital (.445), gripe (.440), (município (.436)), cancelar (.436), governador (.432), pegar (.429), social (.428), (salário (.427)), lei (.422), próximo (.422), economia (.417), (destinar (.415)), (Covid (.410)), rua (.408), trabalhar (.405), dia (.396), eleição (.393), conta (.378), soltar (.373), morte (.368), usar (.366), grande (.361), (precoce (.358)), preso (.355), víruschines_h (.353), pessoa (.345), morrer (.343), (hidroxicloroquina (.342)), torcer (.341), (torcida (.337)), olho (.330), guerra (.323), matar (.317), (imprensa (.314)), enfrentar (.311), criar (.306).

Source: Prepared by the authors.

The following are sample tweets that illustrate the positive pole:

Chart 4 – Examples of Dimension 1: Positive pole

Example 1	@govbr Estou vendo ações de **governadores** p **enfrentar** o #viruschines: superfaturar respiradores e **hospitais** de **campanha**, desmoralizar e até esconder estoques de hidróxicloroquina, que é um remédio testado é **usado** em vários países e está **salvando vidas**. Mas é barato e já está disponível (**Metadados**: # 000011, score = 8, url:https://twitter.com/MartaScramin/status/1257617058960998405, id:12 ação). Lemas do polo que ocorrem no tuíte: ação, campanha, enfrentar, governador, hospital, salvar, usar, vida
Example 2	@GeracaodeValor O debate já existe e a solução é bem clara. **Isolamento** no **grupo** de **risco** e algumas **medidas** de restrição **coletiva**. No mais vida que segue. Não temos **mortes** da **população** economicamente ativa e saudável. #BolsonaroTemRazao #VirusChines (**Metadados:** # 000018, score = 8, url:https://twitter.com/Martinera10/status/1243327824259428353, id:1243327824259428400, d:2020-03-27, u:martinera10) Lemas do polo que ocorrem no tuíte: coletivo, grupo, isolamento, medida, morte, população, risco, vida

Source: Prepared by the authors.

Chart 5 – Factorial pattern of the negative pole of the first factor

Factorial Pattern
(foratoffoli_h (-.792)), (rat_e (-.729)), (cnnlixo_h (-.659)), forayangwanming_h (-.655), money_mouth_face_e (-.652), virusxingling_h (-.638), (forastf_h (-.621)), nauseated_face_e (-.611), robot_e (-.6006), bat_e (-.6000), (golpepolitico_h (-.596)), (forawitzel_h (-.595)), (manipulacaopolitica_h (-.585)), (jogopolitico_h (-.565)), (viruscorona_h (-.531)), (japanese_symbol_for_beginner_e (-.526)), virusmaia_h (-.507),

(intervencaomilitarja_h (-.507)), (forarodrigomaia_h (-.483)), (chinesewuhanvirus_h (-.457)), respeitemopresidente_h (-.444), (foraalcolumbre_h (-.436)), respeitem57milhoesdeeleitores_h (-.435), (ptnuncamais_h (-.426)), (euapoiobolsonaro_h (-.385)), (virusdachina_h (-.361)), (chinesecoronavirus_h (-.358)), respeitemopresidente_h (-.444), (foraalcolumbre_h (-.436)), respeitem57milhoesdeeleitores_h (-.435), (ptnuncamais_h (-.426)), (euapoiobolsonaro_h (-.385)), (virusdachina_h (-.361)), (chinesecoronavirus_h (-.358)), (bandlixo_h (-.356)), (bandeira (-.331)), (cristão (-.321)), (gripechinesa_h (-.301)), (parir (-.301).

Source: Prepared by the authors.

The following are sample tweets that illustrate the negative pole.

Chart 6 – Examples of Dimension 1: Negative pole

Example 3	@MarinaSilva Vá pra China, preste solidariedade abraçando todos seus amiguinhos comunistas! #RespeiteOPresidente **#Respeitem57MilhoesDeEleitores** #EduardoBolsonaroTemRazao #VirusChines **#VirusDaChina** #VirusChino **#VirusXingLing** **#VirusMaia** (**Metadados**: # 000001, score = -3, url:https://twitter.com/Manoeljan1/status/1240987822858874880, id:1240987822858875000, d:2020-03-20, u:manoeljan1) Lemas do polo que ocorrem no tuíte: respeitem57milhoesdeeleitores_h, virusdachina_h (secondary), virusmaia_h, virusxingling_h
Example 4	O Brasil vai dá certo! 💪 #RespeiteOPresidente **#Respeitem57MilhoesDeEleitores** #EduardoBolsonaroTemRazao #VirusChines **#VirusDaChina** #VirusChino **#VirusXingLing** **#VirusMaia** #Vera500k #VerbaMagalhaes #verasonegadora #VeraFakeNews #VeraMentirosa (**Metadados**: # 000004, score = -3, url:https://twitter.com/Manoeljan1/status/1241430880608047106, id:1241430880608047000, d:2020-03-21, u:manoeljan1) Lemas do polo que ocorrem no tuíte: respeitem57milhoesdeeleitores_h, virusdachina_h (secondary), virusmaia_h, virusxingling_h

Source: Prepared by the authors.

Based on the qualitative interpretation, we observed that, in the positive pole of the first factor, the tweets—predominantly posted by right-wing supporters—are marked by a discourse of manipulation during the pandemic. The discursive purpose is to denounce the involvement and protection of corrupt public officials and legislators, restrict access to available treatments, inflate death figures, and harm the economy. This style assumes a personal stance, characterized by the use of first-person singular and plural pronouns, and by lexical indices such as nouns and verbs employed to defend the governmental interests of that period.

In the negative pole of this factor, in order to reach a larger number of followers, verbal language predominates through the use of hashtags, revealing a collective discourse that ridicules China, supports the public policies of that moment, and attacks political opponents.

Based on these observations, we formulated the following designation for the first factor, thus defining it as a discursive dimension:

a) Positive Pole: Manipulation during the pandemic involves protecting corrupt public officials and legislators, restricting access to available treatments, inflating death figures, and harming the economy; and
b) Negative Pole: Messages driven by hashtags that ridicule China, support the president's policies, and attack his opponents.

Dimension 2: Criticism of the system versus anti-China sentiment

Chart 7 – Factorial pattern of the positive pole of the second factor

Factorial Pattern
forawitzel_h (1.125), foraalcolumbre_h (1.076), euapoiobolsonaro_h (1.036), forastf_h (.941), foradoria_h (.911), foramaia_h (.860), bandlixo_h (.835), cnnlixo_h (.827), stfvergonhanacional_h (.821), foratoffoli_h (807), remediodobolsonaro_h (.804), foramandetta_h (.761), maiainimigodobrasil_h (.751), somostodosbolsonaro_h (.750), bolsonaroestavacerto_h (.680), impeachmentdodoria_h (.672), japanese_symbol_for_beginner_e (.663), globolixo_h (.644), impeachmentrodrigomaia_h (.630), bolsonarotemrazão_h (.622), forarodrigomaia_h (.606), obrasilnaopodeparar_h (.567), isolamentovertical_h (.519), ptnuncamais_h (.494), bolsonarotemrazaosim_h (.491), bolsonarotemrazao_h (.487), dorialunatico_h (.481), (bolsonarotinharazao_h (.411)), doriavaiquebrarsp_h (.365), (traidor (.306).

Source: Prepared by the authors.

The following are sample tweets that illustrate the positive pole.

Chart 8 – Examples of Dimension 2: Positive pole

Example 5	@CarlaZambelli38 @jairbolsonaro **#BolsonaroEstavaCerto** **#RemedioDoBolsonaro** **#Bolsonarotemrazao** **#SomostodosBolsonaro** **#EuapoioBolsonaro** #Bolsonaro38 #naosourobo **#Globolixo** **#Bandlixo** **#CNNlixo** **#ForaDoria** **#ForaWitzel** **#ForaRodrigoMaia** **#ForaAlcolumbre** #ForaTofolli **#ForaMandetta** #viruschines #PTladrao (**Metadados**: # 000001, score = 13, url:https://twitter.com/kleberalcantara/status/1248026875932352513, id:1248026875932352500, d:2020-04-08, u:kleberalcantara). Lemas do polo que ocorrem no tuíte: bandlixo_h, bolsonaroestavacerto_h, bolsonarotemrazao_h, cnnlixo_h, euapoiobolsonaro_h, foraalcolumbre_h, foradoria_h, foramandetta_h, forarodrigomaia_h, forawitzel_h, globolixo_h, remediodobolsonaro_h, somostodosbolsonaro_h
Example 6	@ConversaAfiada @alefrota77 **#globolixo** **#bandlixo** **#cnnlixo** #foradiastofolli **#foramaia** #forapaulocamara **#foradoria** **#forawitzel** **#forastf** #forageraldojulio #foralula **#bolsonarotemrazao** #foraflaviodino **#euapoiobolsonaro** **#somostodosbolsonaro** #viruschines **#foraalcolumbre** **#remediodobolsonaro** #bolsonaroestacerto (**Metadados**: # 000002, score = 12, url:https://twitter.com/TENORIO1971/status/1248307725303414786, id:1248307725303414800, d:2020-04-09, u:tenorio1971) Lemas do polo que ocorrem no tuíte: bandlixo_h, bolsonarotemrazao_h, cnnlixo_h, euapoiobolsonaro_h, foraalcolumbre_h, foradoria_h, foramaia_h, forastf_h, forawitzel_h, globolixo_h, remediodobolsonaro_h, somostodosbolsonaro_h

Source: Prepared by the authors.

Chart 9 – Factorial pattern of the negative pole of the second factor

Factorial Pattern
cachorro (-.553)), face_with_symbols_on_mouth_e (-.457), (money_bag_e (-.419)), (warning_e (-.417)), (perdão (-.407)), (unamused_face_e (-.367)), (perseguir (-.361), final (-.359), (cancelar (-.353)), parir (-.352), (microbe_e (-.347)), (contágio (-.342)), (segundo (-.328)), face_vomiting_e (-.326), onda (-.315), (partidário (-.311)), (desculpa (-.310)

Source: Prepared by the authors.

The following are sample tweets that illustrate the negative pole:

Chart 10 – Examples of Dimension 2: Negative pole

Example 7	@PSL_Nacional Que vergonha um partido se prestar a isso... ninguém é idiota. Todos sabemos que é #VirusChines. Aguardem a resposta será dada nas urnas. Assim como: PT, PSOL, PCdoB, PSDB vão ter o que merecem. Continuem subestimando o povo. O povo já respondeu em 2018. AGUARDEM... 🤮 🤬 (**Metadados:** # 000001, score = -2, url:https://twitter.com/SandraSolange12/status/1240787246816677890, id:1240787246816678000, d:2020-03-19, u:sandrasolange12) Lemas do polo que ocorrem no tuíte: face_vomiting_e, face_with_symbols_on_mouth_e
Example 8	@TabanduQuebru Puta que **pariu** velho, e **cachorro** frito vivo, sopa de morcego, sapo destroçado e comido vivo e por aí vai ... Aí ainda que achar ruim chamar de #VirusChines a porra do vírus que veio de lá ... Se k vírus veio de lá ele e o que ? #marciano? (**Metadados:** # 000004, score = -2, url:https://twitter.com/SrgioNe63476210/status/1240855963428585479, id:1240855963428585500, d:2020-03-20, u:srgione63476210) Lemas do polo que ocorrem no tuíte: cachorro, parir

Source: Prepared by the authors.

In Dimension 2, we observed that, in the positive pole of the second factor, the tweets are strongly marked by hashtags that primarily reference Brazilian political spheres, such as the legislative and judicial branches. In addition to these branches, the Brazilian press also stands out. The discursive purpose driven by hashtags is the rejection of the political system, corporate media, and their association with China. Ironically, it is important to note that the executive branch—specifically President Bolsonaro—is portrayed as the only actor capable of resolving the COVID-19 pandemic in Brazil.

In the negative pole of this factor, verbal language is marked by the use of emojis and pejorative adjectives aimed at promoting a discourse that normalizes anti-China sentiment, appealing to common sense while distorting Chinese culture.

Based on these observations, the second factor was designated as a discursive dimension as follows:

a) Positive Pole: Hashtag-driven rejection of the political system, corporate media, and their association with China; and

b) Negative Pole: Normalization of anti-China sentiment, appealing to common sense while distorting Chinese culture.

Dimension 3: Criticism of the media versus rejection of Chinese vaccines

Chart 11 – Factorial pattern of the positive pole of the third factor

Factorial Pattern
imprensasordida_h (1.223), noblatnacadeiaja_h (1.223), estadodedefesa_h (1.168), jairmaisfortedoquenunca_h (1.022), vera500k_h (1.020), luladoria_h (.944), pracimadelespresidente_h (.921), jairnaocainemapau_h (.843), viruschinessim_h (.839), boicoteachina_h (.761), (somostodosbolsonaro_h (.723)), impeachmentdodoria_h (.612), perdão (.537), (bolsonarotemrazão_h (.516)), tiro (.467), (chinaliedpeopledie_h (.439)), verbamagalhaes_h (.436), (impeachmentrodrigomaia_h (.409)), (bolsonarotemrazaosim_h (.393)), (respeitem57milhoesdeeleitores_h (.370)), viruschines_h (.368), (respeitemopresidente_h (.349)), (doriavaiquebrarsp_h (.337)), (bolsonaroestavacerto_h (.334)), (obrasilnaopodeparar_h (.321)), (torcer (.316).

Source: Prepared by the authors.

The following are sample tweets that illustrate the positive pole.

Chart 12 – Examples of Dimension 3: Positive pole

Example 9	**#BoicoteAChina #LulaDoria** #AdolfWitzel **#EstadoDeDefesa #impeachmentdoDoria #imprensasordida #NoblatNaCadeiaJa** #BolsonaroTemRazão **#VirusChines #VirusChinesSim** #VirusComunista **#Vera500k #PraCimaDelesPresidente** #JairNaoCaiNemAPau **#JairMaisForteDoQueNunca** #SomosTodosBolsonaro #foraPT https://t.co/xLH5aEDZcO **Metadados:** # 000009, score = 10, url:https://twitter.com/Fabio_Lemos55/status/1247655156809957376, id:1247655156809957400, d:2020-04-07, u:fabio_lemos55) Lemas do polo que ocorrem no tuíte: boicoteachina_h, bolsonarotemrazão_h (secondary), estadodedefesa_h, impeachmentdodoria_h (secondary), imprensasordida_h, jairmaisfortedoquenunca_h, luladoria_h, noblatnacadeiaja_h, pracimadelespresidente_h, vera500k_h, viruschines_h viruschinessim_h
Example 10	@veramagalhaes @SergioLimafoto Hoje é o Dia do Furo de Reportagem! #BoicoteAChina **#LulaDoria** #AdolfWitzel **#EstadoDeDefesa** #impeachmentdoDoria **#imprensasordida #NoblatNaCadeiaJa #BolsonaroTemRazão #VirusChines #VirusChinesSim** #VirusComunista **#Vera500k #PraCimaDelesPresidente #JairNaoCaiNemAPau** #ForaRodrigoMaia (**Metadados**: #000007, score = 10, url:https://twitter.com/Fabio_Lemos55/status/1247656278266183680, id:1247656278266183700, d:2020-04-07, u:fabio_lemos55) Lemas do polo que ocorrem no tuíte: boicoteachina_h, bolsonarotemrazão_h (secondary), estadodedefesa_h, impeachmentdodoria_h (secondary), imprensasordida_h, jairnaocainemapau_h, luladoria_h, noblatnacadeiaja_hpracimadelespresidente_h, vera500k_h, viruschines_h, viruschinessim_h

Source: Prepared by the authors.

Chart 13 – Factorial pattern of the negative pole of the third factor

Factorial Pattern
fraudemia_h (-.642)), vacina (-.593), vacinar (-.587)), precoce (-.571), (doriamentiroso_h (-.532)), Covid (-.518), (clown_face_e (-.485), tratamentoprecocesalvavidas_h (-.448), flag_china_e (-.422), (informação (-.420)), backhand_index_pointing_right_e (-.390), (farsa (-.388)), (omitir (-.388)), víruschinês_h (-.387), (bat_e (-.387)), (Covid19_h (-.381)), positivo (-.377) (stfvergonhanacional_h (-.341)), segundo (-.336)), (víruschines_h (-.334)), peste (-.334), (biológico (-.325)), viruschinês_h (-.324), (coronavírus_h (-.306.))

Source: Prepared by the authors.

The following are sample tweets that illustrate the negative pole.

Chart 14 – Examples of Dimension 3: Negative pole

Example 11	@paulomathias Por mim, acho ótimo **vacinar** os promotores do MP primeiro! Alais , podem ficar com as doses que por ventura seriam destinadas a mim, pois não irei tomar **vacina** 💉 nemhuma contra o **#VirusChinês** 🦠 🇨🇳 (**Metadados**: # 000001, score = -4, url:https://twitter.com/brisola_ricardo/status/1334618094053036038, id:1334618094053036000, d:2020-12-03, u:brisola_ricardo,) Lemas do polo que ocorrem no tuíte: flag_china_e, vacina, vacinar, viruschinês_h
Example 12	@SpeechesBolso @DouglasGarcia @kimpaim Só imbecis apoiam a **vacina** de um país....... pra combater um vírus criado por esse mesmo país ! 🧟 **#vírusChinês** 😈 🇨🇳 CovidNaVeiaDessesImbecis 🇨🇳 (**Metadados**: # 000008 score = -3, url:https://twitter.com/afonsogarbo/status/1338557732912308226, id:1338557732912308200, d:2020-12-14, u:afonsogarbo) Lemas do polo que ocorrem no tuíte: flag_china_e, vacina, víruschinês_h

Source: Prepared by the authors.

In Dimension 3, we observed that, in the positive pole of the third factor, the posted tweets are strongly marked by hashtags that portray the Brazilian press as China's main ally in a conspiracy against Brazil. By placing the Brazilian press in a central role, the discursive purpose conveyed through hashtags is to criticize corporate media, local officials, and China.

Conversely, in the negative pole of this factor, verbal language is marked by the use of emojis and hashtags, as well as verbs and nouns, promoting a discourse that rejects treatment with Chinese vaccines to combat a virus originating in China—despite the fact that CoronaVac was the main COVID-19 vaccine in Brazil, produced with Chinese expertise, something opposed by the right-wing federal administration, which hindered its distribution to vaccination centers Nationwide.

Based on these observations, the third factor was designated as a discursive dimension as follows:

a) Positive Pole: Hashtag-driven criticism of corporate media, local officials, and China; and
b) Negative Pole: Rejection of treatment with Chinese vaccines to combat a virus originating in China.

Dimension 4: Suppression of Freedom versus Reallocation of Government Funds

Chart 15 – Factorial pattern of the positive pole of the fourth factor

Factorial Pattern
destruir (.723), liberdade (.717), expressão (.675), esconder (.654), virusdachina_h (.625), espalhar (.618), desculpa (.606), campanha (.573), omitir (.600), china (.575), humanidade (.540), perseguir (.538), (chinaassumateuvirus_h (.503)), dever (.502), informação (.499), cristão (.480), ditadura (.477), crime (.475), ponto (.446), mundo (.439), puto (.437), (pânico (.435)), (chinaliedpeopledied_h (.430)), mundial (.408), rede (.394), (morcego (.388)), imprensa (.379), (cachorro (.376)), (risco (.374)), pano (.371), correr (.369), afirmar (.362), (histeria (.355)), verdade (.342), verdade (.342), biológico (.337), governo (.336), pedir (.334), virus (.330), (forayangwanming_h (.329), (saco (.320)), (economia (.318)), desligueatv_h (.311), (gripe (.301)).

Source: Prepared by the authors.

Below are samples of tweets illustrating the positive pole:

Chart 16 – Examples of Dimension 4: positive pole

Example 13	**Verdades** sejam **espalhadas**. Mas como o REGIME DO PARTIDO COMUNISTA CHINÊS É DITADOR né @EmbaixadaChina, preferiu **esconder** as **informações** sobre o vírus e ainda **mentir** na OMS que não havia evidências de transmissão entre humano para humano. **PEÇAM DESCULPAS** AO **MUNDO**. #VirusChines (**Metadados:** # 000001, score = 7, url:https://twitter.com/renatohimura/status/1243523372883030023, id:1243523372883030000, d:2020-03-27, u:renatohimura Lemas do polo que ocorrem no tuíte: desculpa, esconder, espalhar, informação, mundo, pedir, verdade
Example 14	@pedrodoria Cara, vai pra **china** então, bom que cancela o Twitter já que lá "jornalista" só fala oq o **governo** quer... Mas criticar não pode, foda-se a **liberdade** de **expressão**, né? Tem que aplaudir **ditadura** e passar **pano** quando **omitem** dados que matam milhares, afinal, é "fobia". #VirusChines (**Metadados:** # 000006, score = 7 url:https://twitter.com/lbdeboa/status/1240799568482177024, id:1240799568482177000, d:2020-03-20,u:lbdeboa) Lemas do polo que ocorrem no tuíte: china, ditadura, expressão, governo, liberdade, omitir, pano

Source: Prepared by the authors.

Chart 17 – Factorial pattern of the negative pole of the fourth factor

Padrão Fatorial
rat_e (-.762), reduzir (-.697), clown_face_e (-.652), doriamentiroso_h (-.623), (noblatnacadeiaja_h (-.617)), partidário (-.614), (money_mouth_face_e (-.577)), (imprensasordida_h (-.533)), warning_e (-.512), (virus_h (-.471)), destinar (-.466), (jairmaisfortedoquenunca_h (-.456)), money_bag_e (-.454), município (-.450), fundo (-.448), salário (-.434), (eleitoral (-.420)), (luladoria_h (-.393)), (maiainimigodobrasil_h (-.391)), (backhand_index_pointing_right_e (-.339)), (hidroxicloroquina_h (-.315)), (óbito (-.311)), (tratamentoprecocesalvavidas_h (-.301)).

Source: Prepared by the authors.

Below are samples of tweets illustrating the negative pole:

Chart 18 – Examples of Dimension 4: negative pole

Example 15	@profcabarros @AzevedoKelvia Leis absurdas não se cumpre. Ainda mais sabendo que tem **salário** do establishment político, Altos **salários** da elite dos servidores públicos que podem ser **reduzidos**, **fundão eleitora**l e **fundo partidário** deverião ser **destinados** para o combate ao #VirusChines (**Metadados:** # 000001, score = -5, url:https://twitter.com/JosMauroAparec1/status/1250012315128410112, id:1250012315128410000, d:2020-04-14, u:josmauroaparec1) Lemas do polo que ocorrem no tuíte: destinar, eleitoral (secondary), fundo, partidário, reduzir, salário
Example 16	@SenadoFederal Esses 🐀 🐀 🐀 🐀 de esgoto estão de olho no 💰 💰 💰 dos chineses para as próximas eleições. VAMOS LIMPAR O NOSSO CONGRESSO. Vai sair 🐀 🐀 🐀 🐀 Por todos os lados. #VirusChines chineses criminosos. (**Metadados:** # 000041, score = -2, url:https://twitter.com/boschi_cleiz/status/1240786065952292864, id:1240786065952292900, d:2020-03-19, u:boschi_cleiz) Lemas do polo que ocorrem no tuíte: money_bag_e, rat_e

Source: Prepared by the authors.

In Dimension 4, the positive pole of the fourth factor is characterized by tweets marked by nouns and verbs accusing China of being a communist dictatorship that concealed and omitted information about the novel coronavirus. Beyond accusations of omission, there is an explicit demand that China apologize to the world. In contrast, the negative pole is marked by the use of emojis, nouns, and verbs aimed at demanding that the political establishment reduce salaries and reallocate party funds to combat COVID-19.

Based on these observations, the fourth factor was labeled as a discursive dimension:

a) Positive Pole: China is a dictatorship that suppresses freedom of expression.

b) Negative Pole: Elected politicians should reduce their salaries and reallocate party budgets to COVID-19 relief efforts.

Dimension 5: Overthrow of the Political System versus Distortion of Chinese Eating Habits

Chart 19 – Factorial pattern of the positive pole of the fifth factor

Factorial Pattern
manipulacaopolitica_h (1.340), jogopolitico_h (1.313), golpepolitico_h (1.288), intervencaomilitarja_h (1.241), quarentena_h (1.072), pandemia_h (1.060), coronavirusnobrasil_h (.991), hidroxicloroquina_h (.978), hidroxicloroquina (.932), clapping_hands_light_skin_tone_e (.904), unamused_face_e (.768), bolsonarotinharazao_h (.721), regra (.442), (face_with_symbols_on_mouth_e (.439)), (rede (.368)), organizar (.345), flag_brazil_e (.317), (isolamento (.302)).

Source: Prepared by the authors.

Below are samples of tweets illustrating the positive pole:

Chart 20 – Examples of Dimension 5: positive pole

Example 17	📷 Toma-te, ALCOLUMBRE CORRUPTO!!! 👊 👏 🤬 🇧🇷 . . . #intervencaomilitarja #DIASTOFFOLIBANDIDO #coronavirusnobrasil #pandemia #quarentena #viruschines **#golpepolitico #manipulacaopolitica #jogopolitico #hidroxicloroquina** #governadorescorruptos... https://t.co/G86jV2d8pM **(Metadados**: # 000001, score = 10, url:https://twitter.com/JrDamasceno/status/1290908535396749313, id:1290908535396749300, d:2020-08-05, u:jrdamasceno) Lemas do polo que ocorrem no tuíte: lapping_hands_light_skin_tone_e, coronavirusnobrasil_h, face_with_symbols_on_mouth_e (secondary), flag_brazil_e, golpepolitico_h, hidroxicloroquina_h, intervencaomilitarja_h, jogopolitico_h, manipulacaopolitica_h, pandemia_h, quarentena_h
Example 18	🤬 👏 🇧🇷 . . . **#intervencaomilitarja** #notade200 #DIASTOFFOLIBANDIDO #coronavirusnobrasil **#pandemia #quarentena** #viruschines **#golpepolitico #manipulacaopolitica #jogopolitico #hidroxicloroquina** #governadorescorruptos… https://t.co/wLnHj6yVT1 (**Metadados:** # 000003, score = 10, url:https://twitter.com/JrDamasceno/status/1289227238936686592, id:1289227238936686600, d:2020-07-31, u:jrdamasceno Lemas do polo que ocorrem no tuíte: clapping_hands_light_skin_tone_e, coronavirusnobrasil_h, flag_brazil_e, golpepolitico_h, hidroxicloroquina_h, intervencaomilitarja_h, jogopolitico_h, manipulacaopolitica_h, pandemia_h, quarentena_h

Source: Prepared by the authors.

Chart 21 – Factorial pattern of the negative pole of the fifth factor

Padrão Fatorial
vermelho (-.460), morcego (-.414), saco (-.364), microbe_e (-.364), pestechinesa_h (-.333), (pracimadelespresidente_h (-.332)), rato (-.320), (flag_china_e (-.318)).

Source: Prepared by the authors.

Below are samples of tweets illustrating the negative pole:

Chart 22 – Examples of Dimension 5: negative pole

Example 19	@oaquinTeixeira Não é só #hotdog eles comem espetinho de gato, sushi de **rato**, sopa de **morcego** de girino e tudo mais que voa, nada, anda e rasteja. Povo estranho! Não é atoa que criaram a **#PesteChinesa** #VirusChines 😷 🦠 🇨🇳 (**Metadados**: # 000001, score = -4, url:https://twitter.com/jbgois/status/1240811601923620864, id:1240811601923620900, d:2020-03-20, u:jbgois) Lemas do polo que ocorrem no tuíte: flag_china_e (secondary), microbe_e, morcego, pestechinesa_h, rato
Example 20	@EmbaixadaChina @BolsonaroSP Chinaredo! Vocês comem **morcego, rato**, barata, etc., lógico que o resultado é #VirusChines (**Metadados:** # 000012, score = -2, url:https://twitter.com/cesarcascos/status/1240798037590622218, id:1240798037590622200, d:2020-03-20, u:cesarcascos) Lemas do polo que ocorrem no tuíte: morcego, rato

Source: Prepared by the authors.

In Dimension 5, the positive pole of the fifth factor is strongly marked by hashtags and emojis promoting outrage against a corrupt political system and calling for the overthrow of the government through military means. In contrast, the negative pole is characterized by emojis expressing disgust and revulsion toward alleged Chinese eating habits.

Based on these observations, the fifth factor was labeled as a discursive dimension:

a) Positive Pole: Emoji- and hashtag-centered outrage against the corrupt political system and calls for the overthrow of the government through military means;

b) Negative Pole: Distortion of Chinese eating habits.

Dimension 6: Anti-China Campaign versus Pandemic Conspiracy

Chart 23 – Factorial pattern of the positive pole of the sixth factor

Factorial Pattern
Covid一19_h (.864), china_h (.852), viruschino_h (.850), Covid_19_h (.816), virus_h (.801), chinamustpay_h (.766), coronavirus_h (.747), chinaliedandpeopledied_h (.746), chinavirus_h (.734), chinesevirus_h (.728), chinaassumateuvirus_h (.722), Covid19_h (.716), viruscorona_h (.712), chinaliedpeopledie_h (.645), gripechinesa_h (.634), chinesecoronavirus_h (.614), chinesewuhanvirus_h (.611), brasil_h (.530), chinaliedpeopledied_h (.516), (quarentena_h (.427)), (viruschinessim_h (.408)), (virusxingling_h (.403)), espanhol (.381), (coronavirusnobrasil_h (.376)), coronavírus_h (.354), (pandemia_h (.339)), (virus (.302)).

Source: Prepared by the authors.

Below are samples of tweets illustrating the positive pole:

Chart 24 – Examples of Dimension 6: positive pole

Example 21	xi jinping control #xijinping **#China** **#ChinaVirus** **#ChinaLiedPeopleDied** #tedrosadhanom #Trump #Huawei **#Covid_19** **#COVID19** #Covid **#COVID□19** **#VirusCorona** #VirusChines **#ChineseVirus** #ChineseVirus19 #COVID19france #Trump2020 #TrumpPence2020 #Truth #C19 #coronavirus #Corona #pandemic https://t.co/38bZrxpd8 (**Metadados:** # 000001, score = 9, url:https://twitter.com/DrawtheTruth/status/1308406464986349568, id:1308406464986349600, d:2020-09-22, u:drawthetruth Lemas do polo que ocorrem no tuíte: china_h, chinaliedpeopledied_h, chinavirus_h, chinesevirus_h, coronavirus_h, Covid19_h, Covid19_h, Covid_19_h, viruscorona_h
Example 22	**#ChinaLiedPeopleDied** इस चीनी महिला की Diary में है Corona का रहस्य, खौफ में चीन **#ChinaVirus** **#China** **#VirusCorona** #VirusChines #VirusDiary **#coronavirus** **#COVID19** **#Covid_19** #GoCoronaGo https://t.co/4FNDmdPSme (**Metadados:** # 000006, score = 7, url:https://twitter.com/VeblrOfficial/status/1254711119333199872, id:1254711119333200000, d:2020-04-27, u:veblrofficial) Lemas do polo que ocorrem no tuíte:china_h, chinaliedpeopledied_h, chinavirus_h, coronavirus_h, Covid19_h, Covid19_h, Covid_19_h, viruscorona_h

Source: Prepared by the authors.

Chart 25 – Factorial pattern of the negative pole of the sixth factor

Padrão Fatorial
farsa (-.555), (robot_e (-.432)), (nauseated_face_e (-.429)), (estadodedefesa_h (-.428)), (fraudemia_h (-.382)), bandeira (-.373), torcida (-.362), pedido (-.346), (vermelho (.334)), (correr (-.322)), traidor (-.320), (organizar (-.301)), (vazio (-.301)).

Fonte: Elaboração dos autores.

Below are samples of tweets illustrating the negative pole:

Chart 26 – Examples of Dimension 6: negative pole

Example 23	Esse #VirusChines logo...logo...vai ter até **torcida organizada**, vai tem **bandeiras** e balões misturados com CUT, UNE e MST (**Metadados:** # 000001, score = -2, url:https://twitter.com/tonny_coxa/status/1243273650016522244, id:1243273650016522200, d:2020-03-26, u:tonny_coxa) Lemas do polo que ocorrem no tuíte: bandeira, organizar (secondary), torcida
Example 24	E alguém da extrema imprensa está interessado em lógica ou razão? Não estamos num debate, estamos numa **torcida** pelo #VirusChines. E quem não entrou em pânico sabe bem o porquê. Mas a fome - e ela virá -, vence o medo. Essa **farsa** tem os dias contados. (**Metadados:** # 000003, score = -2, url:https://twitter.com/diza2012/status/1246794219764813825, id:1246794219764813800, d:2020-04-05, u:diza2012) Lemas do polo que ocorrem no tuíte: farsa, torcida

Source: Prepared by the authors.

In Dimension 6, the positive pole of the sixth factor is primarily marked by hashtags in multiple languages, with the discursive purpose of discrediting China in both national and international arenas. In contrast, the negative pole is characterized by nouns and adjectives that frame the pandemic as a sham financed by China and orchestrated by corporate media and corrupt leaders.

Based on these observations, the sixth factor was labeled as a discursive dimension:

a) Positive Pole: Anti-China messages based on hashtags aimed at national and international audiences; and

b) Negative Pole: Framing of the pandemic as a sham financed by China and orchestrated by corporate media and corrupt leaders.

Final Considerations

Our findings are consistent with previous studies (Person *et al*., 2004; Lee *et al*., 2005; Hoppe, 2018; Reny; Barreto, 2020), which identified xenophobic and stigmatizing attitudes in the context of pandemics and epidemics. We also confirm the proposition put forward by Berber Sardinha and Moreira (2023), as the tweets analyzed support antidemocratic causes or ideologies. Since discourse enables the construction of concepts (Baker; Mcenery, 2015), it is necessary to acknowledge the pernicious impact of xenophobic discourse on public opinion formation and on decision-making processes within public and private institutions during a period of public health crisis. The infodemic negatively affected the social fabric, as it lacked commitment to scientific knowledge and contributed to confusion among segments of society.

The quantitative and qualitative analysis of the tweets broadened our understanding of the discourses present in the infodemic texts that circulated in Brazil during this period. It can be stated that xenophobia did not emerge spontaneously as a result of immediate emotional reactions to the unknown and the uncertainty generated by the virus.

Hashtags such as #viruschines, #pragachinesa, #pestechinesa, #pasteldeflango, #boicoteachina, #vachina, #ChainaVirus, and #wuhanvirus were used by right-wing supporters in Brazil to discredit China and disseminate hatred against its people and culture. As China was considered the initial epicenter of the pandemic, Chinese people were frequently subjected to accusations, insults, persecution, and aggression, reflecting a historical pattern of scapegoating during previous health crises—a phenomenon also identified by Dionne and Turkmen (2020), Ittefaq *et al*. (2022), Mansouri (2020), and White (2020).

Xenophobia may also have been driven by interests linked to geopolitical conflicts and international competition. China's rise as a global power in recent decades, challenging the hegemony of Western countries, has generated broad geopolitical competition with implications for trade, technology, security, and regional influence. In this context, nationalist sentiments and the pursuit of strategic advantages may have contributed to the demonization or dehumanization of China as "the other."

From this perspective, the identified dimensions of variation reveal six dimensions, each composed of a positive and a negative pole. In Dimension 1, the positive pole addresses manipulation during the pandemic, involving the protection of corrupt public officials and legislators, restrictions on access to available treatments, inflation of death tolls, and damage to the economy. In the negative pole, messages driven by hashtags ridicule China, support presidential policies, and attack political opponents. In Dimension 2, the positive pole reveals hashtag-driven rejection of the political system, corporate media, and their association with China, while the negative pole encompasses the normalization of anti-China sentiment, appealing to common sense while distorting Chinese culture. In Dimension 3, the positive pole reflects hashtag-driven criticism of corporate media, local officials, and China, whereas the negative pole expresses rejection of Chinese vaccines as a treatment for a virus that originated in China. In Dimension 4, the positive pole portrays China as a dictatorship that suppresses freedom of expression, while the negative pole argues that elected officials should reduce their salaries and reallocate party budgets to COVID-19 relief efforts. In Dimension 5, the positive pole conveys emoji- and hashtag-centered outrage against a corrupt political system and calls for the overthrow of the government through military means; the negative pole includes references to the distortion of Chinese eating habits. Finally, in Dimension 6, the positive pole consists of anti-China messages based on hashtags aimed at national and international audiences, while the negative pole frames the pandemic as a sham financed by China and orchestrated by corporate media and corrupt leaders.

The discourses identified across the six dimensions indicate an anti-China sentiment associated with intercultural situations of stigmatization related to geographic location—the pandemic epicenter—as well as to perceived characteristics such as race and lifestyle.

In summary, the analyzed posts are embedded in echo chambers—environments that generate misinformation and distort individuals' perspectives—preventing engagement with alternative viewpoints and the discussion of complex issues, while reinforcing and legitimizing limiting and/or prejudiced beliefs. This configuration constitutes a fertile environment for the infodemic itself.

REFERENCES

BAKER, P.; MCENERY, T. **Corpora and discourse studies**: integrating discourse and corpora. Basingstoke: Palgrave Macmillan, 2015.

BANERJEE, D.; MEENA, K. S. COVID-19 as an "infodemic" in public health: critical role of the social media. **Frontiers in public health**, v. 9, n. 231, 2021.

BERBER SARDINHA. 25 years later: Comparing internet and pre-internet registers. *In*: BERBER SARDINHA, T.; VEIRANO PINTO, M. (ed.). **Multi-Dimensional Analysis**: 25 years on: a tribute to Douglas Biber. Amsterdam; Philadelphia, PA: John Benjamins, 2014. p. 81-105.

BERBER SARDINHA, T.; SÃO BENTO F. T. (ed.). **Working with Portuguese corpora**. London; New York: Bloomsbury: Continuum, 2014.

BERBER SARDINHA, T., ACUNZO M, C. ; FERREIRA, S. B. T. Metáforas da economia no dicionário de colocações do português brasileiro: uma análise multidimensional baseada em corpus. **Filologia e Linguística Portuguesa**, v. 18, n. 1, p. 175-198, 2016.

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

BERBER SARDINHA, T. A historical characterisation of American and Brazilian cultures based on lexical representations. **Corpora**, v. 2, n. 15, p. 183-212, 2020.

BERBER SARDINHA, T. Discourse of academia from a multi-dimensional perspective. *In*: FRIGINAL, E.; HARDY, J. **The Routledge handbook of corpus approaches to Discourse Analysis**. Abingdon: Routledge, 2021. p. 298-318.

BERBER SARDINHA, T.; RESENDE,V. S.; SHIMAZUMI, M. Linguística de Corpus. *In*: OTHERO, G. A.; FLORES, V. N. (org.). **A Linguística hoje**: múltiplos domínios. São Paulo: Contexto, 2023. p. 103-117.

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.; FITZSIMMONS-DOOLAN, S. **Lexical Multidimensional Analysis**. Cambridge: Cambridge University Press, 2025.

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

BIBER, D. Multi-dimensional approaches. *In*: LÜDELING, A.; KYTÖ, M. (ed.). **Corpus Linguistics**: an international handbook. Berlin; New York: Walter de Gruyter, 2009. p. 822-855.

BURR, V. **An introduction to social constructionism**. London: Routledge, 1995.

CAMBRIDGE UNIVERSITY PRESS AND ASSESSMENT. Echo-chamber. 2024. Disponível em: https://dictionary.cambridge.org/dictionary/english/echo-chamber. Acesso em: 17 maio 2024.

CLARKE, I.; BROOKES, G.; McENERY, T. Keywords through time: tracking changes in press discourse of Islam. **International journal of Corpus Linguistics**, v. 27, n. 4, p. 399-427, 2022.

CROSSLEY, S.; LOUWERSE, M. M. Multi-dimensional register classification using bigrams. **International journal of Corpus Linguistics**, v. 12, n. 4, p. 453-478, 2007.

DIONNE, K.Y.; TURKMEN, F. F. The politics of pandemic othering: putting COVID-19 in global and historical context. **International organization**, v. 74, n. S1, p. 213-230, 2020.

DUBEY, A. D. The resurgence of cyber racism during the COVID-19 Pandemic and its aftereffects: analysis of sentiments and emotions in tweets. **JMIR Public Health Surveill**, v. 6, n. 4, p. 1-7, 2020.

FITZSIMMONS-DOOLAN, S. Language ideologies of institutional language policy: exploring variability by language policy register. **Language policy**, v. 18, n. 2, p. 169-189, 2019.

FITZSIMMONS-DOOLAN, S. 21st century ideological discourses about us migrant education that transcend registers. **Corpora**, v. 18, n. 2, p. 143-173, 2023.

FRIGINAL, E.; HARDY, J. (ed.). **Routledge handbook of corpus approaches to Discourse Analysis**. London: Routledge, 2020.

HOPPE, T. "Spanish Flu": when infectious disease names blur origins and stigmatize those infected. **American Journal of Public Health**, v. 108, n. 11, p. 1462-1464, 2018.

ITTEFAQ, M. *et al*. A pandemic of hate: social representations of COVID-19 in the media. **Analyses of Social Issues and Public Policy**, n. 22, p. 225-252, 2022.

KOHATSU, L. N.; SAITO, G. K.; ANDRADE, P. F. **Imigração, mídia e xenofobia**: a ameaça imaginária em questão. São Paulo: Blucher, 2021.

LA GARZA, C. Xenofobia. **Laboreal**, v. 7, n. 2, 2011. Disponível em: http://journals.openedition.org/laboreal/7924. Acesso em: 30 jul. 2023.

LEE, S. *et al*. The experience of SARS-related stigma at Amoy Gardens. **Social Science and Medicine**, v. 61, n. 9, p. 2038-2046, 2005.

MANSOURI, F. **The socio-cultural impact of COVID-19**: exploring the role of intercultural dialogue in emerging responses. Paris: UNESCO Briefing Papers, 2020.

MESQUITA, C. T.; OLIVEIRA, A. ; SEIXAS, F. L.; PAES, A. Infodemia, fake news and medicine: science and the quest for truth. **International Journal of Cardiovascular Sciences**, v. 33, n. 3, p. 203-205, 2020.

MOSCOVICI, S. **A psicanálise, sua imagem e seu público**. Petrópolis, RJ: Vozes, 2012.

ORLANDI, E. P. **Análise de discurso**: princípios e procedimentos. 10. ed. Campinas, SP: Pontes, 2015.

PATRICK, I. Latin America used to be positive toward China. Covid-19 might change that. **The Diplomat**. 2020. Disponível em: https://thediplomat.com/2020/04/latin-america-used-to-be-positive-toward-china-Covid-19-might-change-that. Acesso em: 26 nov. 2023.

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

PERSON, B. *et al*. Fear and stigma: The epidemic within the SARS outbreak. **Emerging Infectious Diseases**, v. 10, n. 2, p. 358-363, 2004.

RENY, T. ; BARRETO, M. A. Xenophobia in the time of pandemic: othering, anti-Asian attitudes, and COVID-19. **Politics, Groups, and Identities**, v. 1, n. 24, p. 01-26, 2020.

RIBEIRO, J. **Xenofobia e intolerância linguística**: discurso sobre estrangeiridade e hostilidade brasileira. Campinas: Pontes, 2022.

SEYFERTH, G. O Estado brasileiro e a imigração. *In*: SANTOS, M. O. *et al*. (org.). **Caminhos da migração**: memória, integração e conflitos. Rio de Janeiro: Léo Christiano Editorial, 2014.

SHE, Z.; ZHOU, N.; LI, D. *et al.* Does COVID-19 threat increase xenophobia? The roles of protection efficacy and support seeking. **BMC Public Health**, v. 22, n. 485, p. 01-10, 2022.

WHITE, A. I. Historical linkages: epidemic threat, economic risk, and xenophobia. **The Lancet**, v. 395, n. 10.232, p. 1250-1251, 2020.

ZIELINSKI, C. Infodemics and infodemiology: a short history, a long future. **Revista Panamericana de Salud Publica**, n. 45, p. 1-8, 2021.

ZUPPARDI, M. C.; BERBER SARDINHA, T. A multi-dimensional view of collocations in academic writing. **Studies in Corpus Linguistics**, v. 95, p. 333-354, 2020.

CRediT Author Statement

- **Acknowledgements**: We thank the Coordination for the Improvement of Higher Education Personnel (CAPES); the Corpus Linguistics Study Group (GELC), Pontifical Catholic University of São Paulo (PUCSP); the State University of Goiás (UEG); and The Hong Kong Polytechnic University (PolyU), Hong Kong, China.
- **Funding**: Coordination for the Improvement of Higher Education Personnel (CAPES), grant numbers 88887.704152/2022-00 and 88881.981386/2024-01.
- **Conflicts of interest**: The authors declare no conflicts of interest.
- **Ethical approval**: Not applicable.
- **Data and material availability**: Open Science Framework: 10.17605/OSF.IO/KPYSR.
- **Authors' contributions**: Cícero Soares da Silva conducted the field data collection, participated in the structuring of the manuscript, and contributed to the development of the theoretical framework; Eduardo Batista da Silva participated in the structuring of the manuscript and contributed to the development of the theoretical framework; Aline Andrea Zamboni Milanez and Amanda Chiarelo Boldarine conducted the field data collection; and Tony Berber Sardinha contributed to the field data collection and to the development of the theoretical framework.

Processing and editing: Editora Ibero-Americana de Educação
Proofreading, formatting, standardization and translation