

TRANSLATING THE AGGRESSIONS EXPERIENCED DURING THE COVID-19 PANDEMIC CRISIS: A CORPUS-BASED ANALYSIS OF VIOLENCE IN ENGLISH, FRENCH, PORTUGUESE, AND SPANISH NEWSPAPERS FROM 2020

TRADUZINDO AS AGRESSÕES EXPERIENCIADAS DURANTE A CRISE PANDÉMICA DE COVID-19: UMA ANÁLISE DO CORPUS DE VIOLÊNCIA EM JORNAIS DE LÍNGUAS INGLESA, FRANCESA, PORTUGUESA E ESPANHOLA DE 2020



Letícia Ferreira dos SANTOS
Mestranda

Universidade Estadual Paulista “Júlio de Mesquita Filho”
Instituto de Biociências, Letras e Ciências Exatas
Programa de Pós-Graduação em Estudos Linguísticos
São José do Rio Preto, São Paulo, Brasil
<https://lattes.cnpq.br/4016458573185447>
<https://orcid.org/0000-0002-0440-7452>
leticia.ferreira.santos@unesp.br

Talita SERPA
Pós-doutoranda

Universidade Estadual Paulista “Júlio de Mesquita Filho”
Instituto de Biociências, Letras e Ciências Exatas
Programa de Pós-Graduação em Estudos Linguísticos
São José do Rio Preto, São Paulo, Brasil
<http://lattes.cnpq.br/7935107964292543>
<https://orcid.org/0000-0003-3324-9593>
talitasrp82@gmail.com

1

Abstract: The main purpose of this paper is to analyze how the words *violência*, “violence,” *violencia* and *violence* are used in four comparable *corpora* based on articles from *Folha de São Paulo*, *CNN*, *El País* and *Le Monde* webpages gathered from January to July 2020. We followed Lexicology, Corpus Linguistics and Corpus-Based Translation Studies to extract the most frequently used words in each *corpora* and verify the contexts in which they are used, relating them to the concept of “violence” based on *collocations*, *clusters* and *concordances* extracted with *AntConc* software (2020). Additionally, we employed Discourse Analysis (DA) in Translation to verify how aggressive actions are linguistically expressed and to highlight the importance of understanding context in the translation process. As a result, we noticed that despite being considered translational equivalents, the words are used in contexts with different semantic meanings, depending on the culture to which they belong.

Keywords: Corpus-Based Translation Studies. Corpus Linguistics. Discourse analysis. Journalistic Texts. Violence.

Resumo: O objetivo principal deste trabalho é analisar o uso das palavras “*violência*”, violence, *violencia* e *violence* em quatro corpora comparáveis com base em artigos das páginas da internet dos jornais *Folha de São Paulo*, *CNN*, *El País* e *Le Monde* coletados de janeiro a julho de 2020. Tomamos por teoria as bases da Lexicologia, a Linguística de Corpus e os Estudos de Tradução Baseados em Corpus para extrair as palavras mais frequentes nos corpora mencionados e verificar seus contextos, relacionando-os com o conceito de “*violência*” a partir de colocações, clusters e concordâncias extraídos com o software *AntConc* (2020). Além disso, utilizamos a Análise do Discurso (AD) na Tradução para verificar como as ações agressivas são expressas linguisticamente e destacar a importância de conhecer o contexto no processo tradutório. Como



Este é um artigo em acesso aberto distribuído nos termos da Licença Creative Commons Atribuição que permite o uso irrestrito, a distribuição e reprodução em qualquer meio desde que o artigo original seja devidamente citado.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original article is properly cited.

resultado, observamos que, apesar de serem consideradas equivalentes, as palavras são utilizadas em contextos com significados diferentes, dependendo da cultura à qual pertencem.

Palavras-chave: Estudos de tradução baseados em corpus. Linguística de corpus. Análise de discurso. Textos jornalísticos. Violência.

Worldwide, violence has increased due to restrictions and blockages that governments have imposed during the health crisis of Covid-19. This situation gained enormous proportions during the initial months of 2020 when economic, political, and social structures began to fail. Consequently, the world media started exposing these facts, drawing our attention to the concept of "violence," which can be used differently in discursive acts from different countries and cause some problems when translated to other languages.

Studies such as those carried out by Bradbury-Jones and Isham (2020) and Evans et al. (2020) point to a high incidence of attacks against women and other people in vulnerable situations – for example, children and the elderly – in lower-class homes. They also suggest alcoholism, lack of financial resources and previous abusive relationships as the main factors that triggered this phenomenon.

This investigation turned its gaze to reflect on how the word "violence" is related to brutality and presents different meanings in journalistic texts from four countries (USA, Brazil, Spain, and France). We observed how these newspapers described the processes of aggression faced by populations in the initial year of the pandemic. To this end, we created small, multilingual and comparable *corpora* made up of articles published on *CNN*, *Folha de São Paulo*, *El País* and *Le Monde* websites from January to July 2020,¹ in order to understand how *violência*, "violence," *violencia* and *violence* are used within the discourse of these media and how, despite being considered possible translational equivalents (Tognini-Bonelli, 1996), they can encompass different meanings associated to the reality of each community, as well as represent social phenomena specific to each country. Thus, these apparently similar words may differ among semantic and pragmatic points of view in the complex semiotic systems to which they belong, since each country has an established relationship with the concepts of laws, norms, violation, and damage. Additionally, the role of women differs among these peoples, seeing as the processes of colonization, miscegenation, and invasion were also very different in each country.

We employed Corpus Linguistics (Berber Sardinha, 2004; Sinclair, 1991, 1992, Corpus-Based Translation Studies (Baker, 1995, 1996) and Lexicology (Barbosa, 1990, 1998; Dubois et al., 1973; Sager, 1998) to select the most frequent words in the *corpora* with the help of *AntConc* software (Anthony, 2020). Based on this data, we verified discursive differences between possible equivalents of "violence." The theoretical framework used was Discourse Analysis in Translation (P. Baker, 2006; Beaugrande, 1994; Faiq, 2019; Mason, 1994) and Corpus Linguistics (Bührig & House, 2004; House, 2014; M. Baker, 2006). Our aim was to understand how the concept of "violence" was defined in the selected textual sets in each language and how its discursive representation can distance or bring together interpretive readings and assist in the translation process.

Corpus Linguistics and Corpus-Based Translation Studies: Suitable Concepts to our Investigation

Corpus Linguistics (CL) has been commonly used in several areas of Linguistic Studies and has gained significance in the last decade. Berber Sardinha (2004) emphasizes that the conceptual framework of this area presents an empiricist approach based on the view of language as a probabilistic system.

This area favors the investigation of data gathered from the observation of language in its natural use, allowing for the understanding of semantic networks and lexical fields through the manipulation of continuous texts, and enabling a confrontation between theory and empirical data, which expands scientific knowledge about language structures through the collection and analysis of common language examples.

Data from statistical calculations led us to understand, corroborating Sinclair (1991) and Halliday (1991, 1992), that frequency is an inseparable attribute of the word, as it reveals its occurrence in a text. The frequency (high, low, intermediate etc.) has a defining role in a word's meaning (Berber Sardinha, 2004, pp. 162–163).

Sinclair (1991, 1992) defines a *corpus* as an artefact developed for research:

A corpus is a collection of naturally-occurring language text, chosen to characterize a state or variety of a language: In modern computational linguistics, a corpus typically contains many millions of words: this is because it is recognized that the creativity of natural language leads to such immense variety of expression that it is

difficult to isolate the recurrent patterns that are the dues to the lexical structure of the language. (Sinclair, 1991, p. 171)

In turn, Baker (1995) considers *corpora* analysis a rich source of descriptive-comparative material that can help us perceive differences between the language of a translation and that of a source text. In this study, we used four comparable *corpora* initially written in Portuguese, English, Spanish and French, respectively, as they enable a better identification of the form and function of words.

Regarding the concept of word, in the *Dictionnaire de Linguistique*, Dubois et al. (1973) present the following definition:

(...) a significant linguistic element composed of one or more phonemes; this sequence is susceptible to a written transcription (ideogrammatical, syllabary or alphabetical) comprised between two blank spaces; it retains its form, wholly or partially (in the case of inflexion), in its various syntagmatic uses; the word denotes an object (noun), an action or a state (verb), a quality (adjective), a relation (preposition), etc., our translation.² (Dubois et al., 1973, p. 343, our translation)

4

Therefore, we believe that words are multifunctional, and that the determination of their functionality depends on their insertion in a discursive context (Barbosa, 1990, 1998).

***Corpora* Studies in Translation and the Notion of Equivalence**

Corpus-Based Translation Studies (CBTS) is an area of research that adopts and develops CL methodologies to analyze translation practices for theoretical, descriptive, and applied purposes. More specifically, investigations in this area are seen as central to Translation Studies because (1) they are committed to integrate cultural studies and to explore its interrelatedness, (2) they show awareness of the effects of ideology on translation theory, its practice and pedagogy, and (3) they aim to tailor modern technologies to improve theory, study, and empirical knowledge for the benefit of both translator training and the work of professional translators.

Within the field of CBTS, Tognini-Bonelli (2001) stands out, especially concerning the conceptualization of comparable *corpora* and the development of steps for data analysis.

The first step is based on identifying and classifying the lexical and grammatical pattern (collocational and colligational patterns) within the context of a word. The second step enables us to recognize the first meaning (*prima facie*) of the word in the Source Text (ST) and the Target Text (TT), comparing their form and function. The third step is to observe the form used (colational or coalitional patterns) in Target Language (TL). This method suggested by Tognini-Bonelli (2001) is related to coding and decoding in another language. With the help of comparable *corpora*, researchers have access to concepts as they are used within specific contexts, which allows them to make an adequate choice regarding the *equivalent* for its translation, based on objective evidence of use in both languages.

In this regard, Tognini-Bonelli (1996) promotes the notion of *translational equivalence* using the theories of Firth (1957) to associate it with the constructs of *corpora*. Firth presents a contextual view of meaning, which arises from the relationships between the linguistic item and its use in context. He calls a "context of situation" a group of interrelated verbal and non-verbal categories. This context involves participants in linguistic and extralinguistic interaction.

The translation process, therefore, firstly identifies the linguistic level, seeking to negotiate equivalence on the extralinguistic level, i.e., the relationship between the "relevant participants" (the source writer and his audience), the verbal action (ST) and the "verbal action effect" (TT) – encompassing different languages, as well as different social contexts and contextual environments.

As a result, the search for words that are equivalents in the four languages analyzed in this study demonstrates how these terms (*violência*, "violence," *violencia*, and *violence*) behave in diverse settings, fostering subtle changes in meaning despite their proximity in use.

Discourse Analysis, Translation and Use of *Corpora*

In our study, we understand words as representations of objects and events, and that their meanings are predispositions about which aspects deserve consideration (Beaugrande, 1994, p. 6), allowing us to verify the *corpora* with greater specificity using CL.

Beaugrande (1994, p. 6) points out that since language is a system of codes, translation is more than a linguistic activity because it is not just about changing from one language to another. It requires psychological, intertextual, and narrative knowledge; thus,

it can express "deep" meanings of a text, even violating lexical and referential fidelity (Eco, 2001, pp. 13-14).

M. Baker (2006), in turn, considers discourse to be a way of categorizing reality through narratives (M. Baker, 2006, p. 16). Baker defines a narrative as a social and communicative point of view which describes a dynamic entity that can influence people's actions and beliefs: "They change subtly or radically as people experience and expose themselves to new stories daily" (M. Baker, 2006, p. 3).

According to House (2014), translation, as a continual phenomenon in communicative events, employs Discourse Analysis (DA) as an instrumental contribution to address translation equivalence concerns. The author underlines that, while interpreting a text, readers, listeners, and translators must retrieve its meaning and turn it into a communicative reality; they must interpret it as a conversation. Thus, the text is the linguistic quality in a person's speech or writing. A speech is a meaning that a person wishes to communicate while writing a text, which a receiver (and a translator) must subsequently interpret. Understanding the DA approach allows Translation Studies to identify patterns and systems in the decisions made by translators as well as to verify possible explanations for these decisions based on procedures of discourse analysis. For Translation Studies, DA and CL methods assist in detecting, describing, and evaluating linguistic item choices and item sequences in terms of their suitability to the type of translated text (Bührig & House, 2004). Therefore, the link between DA and CBTS is undeniable since both integrate discursive research approaches based on examples from source texts and translated texts, with verification processes in quantitative and qualitative *corpora*.

Fairclough and Wodak (1997) name eight principles that govern DA and which can be supported by CL analyses: 1. Addressing social problems; 2. Power relations are discursive; 3. Discourse constitutes society and culture; 4. Speech does ideological work; 5. Speech is historical; 6. The link between text and society is mediated; 7. Discourse analysis is interpretive and explanatory; 8. Speech is a form of social action. Between both theories, there is a shared understanding of the importance of data authenticity and language in use. DA, CBTS and CL study actual and often extended instances of social interaction, usually present in texts, making them naturally compatible.

For P. Baker (2006), in *Using Corpora in Discourse Analysis*, the search for usage patterns, placements, frequency, and agreements allows analysts to verify how the linguistic

item, the word, behaves in terms of placement with other items throughout the *corpus*. Therefore, in a qualitative discourse analysis of individual *corpora* texts, there is the option of resorting to concordance *software*, compiling and studying a frequency list, and analyzing the concordances of repeated lexical items. The discovery of patterns used in *corpora* can be seen as a worthwhile pursuit, thus making a legitimate contribution to the critical study of language and society.

Materials and Methods

In this section, we display a description of the Portuguese, Spanish, French and English *corpora*, as well as the methodological steps followed in this research. Below, Table 1 portrays the statistical data of the *corpora*.

Table 1 — Statistical data of comparable corpora

Corpora	Tokens	Types	Type/Token Ratio	Standardized Type/Token Ratio
Portuguese Corpus – Folha de São Paulo	1,016,366	46,791	4,60	48,79
English Corpus – CNN	469,211	23,047	4,91	43.98
Spanish Corpus – El País	372,485	30,518	8.19	45.74
French Corpus – Le Monde	257,747	22,475	8.72	45.48

Source: elaborated by the authors

7

The texts were collected from January 2020 to July 2020, all belonging to the field of journalism and selected from several newspapers' websites based on the search words *violência*, "violence," *violencia* and *violence*. As for the procedures adopted for the treatment of the *corpora*, we cleaned and saved the documents as plain text (.txt) to be processed by the *AntConc* program (Anthony, 2020). Each textual set received a name that identified it within the *corpora* of study, namely: <corpusfolha.>, <corpuscnn>, <corpuselpais> and <corpuslemonde>. Regarding the available tools, this research used *WordList* and *Concordance*, as well as their respective applications (*collocates* and *clusters*).

Data Analysis

As shown in Table 1, although the *corpora* were compiled in the same period, the size of the Portuguese-language *corpus* is more extensive, with a total of 1,016,366 words, which leads us to the hypothesis that the theme *violência* is addressed more frequently in the Brazilian media than in the other countries chosen for this investigation.

The second largest *corpus*, in English, included 469,211 words – 547,155 words less than the *corpus* in Portuguese. However, when we look at the type/token ratio³, which determines the frequency of words in relation to the total number of words, we see that the two largest *corpora* have a similar vocabulary density; that is, the proportions of lexical items are 4.60 and 4.91. The speaker's lexical density is measured by calculating the relationship between the total number of lexical items and the total number of words in the text produced by the individual.

However, when we examine the two smallest *corpora*, we notice that they present a much higher word density – 8.19 in Spanish and 8.72 in French. This can lead us to the conclusion that these texts use more words to compose their speeches despite having a shorter extension, being more creative and diverse.

We understand that both *corpora* retrieved from *El País* and *Le Monde* may present data related to the context of the words *violencia* and *violence* that work on social, cultural and discursive aspects differently from those of the *corpora* retrieved from *Folha de SP* and *CNN*. Additionally, the varying sizes reveal two possible discursive aspects of the texts that make up our analysis, the first being that the Portuguese and English languages may need more linguistic items to express the studied theme, and the second being that there is a higher rate of violent events in the countries where the newspapers in these two languages were produced and, consequently, a more significant number of articles involving the subject. We then used the *WordList* tool to list the most frequent nouns⁴ in each *corpus*. Below, we present Tables 2 to 5 with the first ten words of each list.

Table 2 — *Ten most frequent words in the corpus of Folha de São Paulo*

Words	Freq.
PRESIDENTE	2.014
VIOLÊNCIA	1,960
BOLSONARO	1,887
PESSOAS	1,638
GOVERNO	1,494
POLÍCIA	1.060
ESTADO	1,026
MULHERES	971
PANDEMIA	880
SEGURANÇA	878

Source: elaborated by the authors

Table 3 — *Ten most frequent words in the corpus of CNN*

Words	Freq.
TRUMP	2,028
PRESIDENT	1.299
POLICE	1,240
PEOPLE	1,147
VIOLENCE	737
STATE	662
PROTESTERS	610
RACIAL	539
PROTESTS	478
RECKONING	478

Source: elaborated by the authors

Table 4 — *Ten most frequent words in the corpus of El País*

Words	Freq.
MUJERES	1437
VIOLENCIA	1196
AÑOS	1053
PAÍS	554
GÉNERO	525
PERSONAS	502
MUNDO	413
VIDA	389
NIÑOS	358
GOVERNO	322

Source: elaborated by the authors

Table 5 — *Ten most frequent words in the corpus of Le Monde*

Words	Freq.
VIOLENCE/S	351/369
PRÉSIDENT	328
TEMPS	290
PERSONNES	266
FEMMES	259
POLICE	212
MORT	210
HOMME	207
POLITIQUE	192
MINISTRE	185

Source: Elaborated by the authors

We observed that frequency lists can be combined with DA using *corpora*. Thus, we noticed that the Portuguese texts of *Folha de SP* focused on issues concerning the topic of the presidency of Brazil, in such a way that the word *resident* is mentioned 2014 times – being associated to the name of the previous head of state, *Bolsonaro*, in 1887 occurrences. It is possible to assume that these news articles addressed issues of national security (878), since *Estado* (1026) and the *polícia* (1060) are constantly brought up as recurrent themes. *Violência* (1960), in this *corpus*, therefore, seems to come mainly from the *governo* (1494) against the population (*pessoas* [1638]).

Within the scope of the English language *corpus*, the name of the US president at the time, "Trump," appears as the most frequent word, with a total of 2028 occurrences. The frequency with which it is referred to leads us to believe that its role as an individual permeates its function, considering that the word "president" appears 1299 times. In this corpus, we see that *CNN* uses "violence" in 757 instances within contexts that still address

the notions of "racial" (539), "protests" (478), "police" (1240) and "state" (662). Consequently, it can be assumed that these violent acts are related to issues of racism and xenophobia and that the texts are related to the case of George Floyd.⁵

The *corpus* of *El País* is the first to bring up the concept of "women" among its most frequent words, being the most common noun in the collected texts. With this, we are led to return to the propositions of Bradbury-Jones and Isham (2020) and Evans et al. (2020), among other authors previously mentioned, which point to the increase in aggressive behavior against vulnerable people during the pandemic. We were able to observe that the texts compiled in Spanish still present the word *violencia* (1196) as their second most frequent word. We understand, then, that the newspaper's scope corroborates what has been discussed regarding the social and economic data collected during the pandemic. Additionally, this *corpus* further reinforces such readings by frequently mentioning the words *género* (525) and *niños* (358), which relate to relationships and family.

Finally, the set of texts collected from *Le Monde* presents different data from the others. In addition to displaying the word *violence* as the most recurrent noun, the *corpus* also presents its plural form as the second most frequent word, totaling 720 occurrences. There is also, among the data, the mention of *femmes* (259) and a reference to *homme* (207), which does not occur in the other *corpora*. Therefore, these factors help us understand that the concept of *violence* takes on different proportions in texts from different cultures, leading us to the proposition that there is no way to find translational equivalents among the words analyzed.

Corpora Data in Context Using the Concordance Tool

Folha de São Paulo. Based on the survey of the most frequent words, we started to search the contexts of use for *violência* in the *corpus* of *Folha de SP* using the *Concordance* tool. Below, we present figures 1 to 3, with examples of *concordance*, *clusters/n-grams* and *collocates*:

Figure 1 — Concordance of the word violência in the corpus of Folha de São Paulo

Concordance	Concordance Plot	File View	Clusters/N-Grams	Collocates	Word List	Keyword List
Concordance Hits 1960						
Hit	KWIC					
1		que a paz enfim triunfou sobre a violência, a agressão e a guerra . Por isso				
2		pode voltar a explodir com ainda mais violência, a baleia pode cuspir de uma só				
3		do crime, o que teria motivado a violência . A bike não havia sido encontrada até				
4		registrado um grande aumento nos índices de violência . A cidade tem um número elevado de				
5		mazelas, como a ditadura, a tortura, a violência, a corrupção, a miséria . Mas era como				
6		seguro dessa e de outras formas de violência . A despeito de toda a nossa encravilhada				
7		por expor os imigrantes a situações de violência . A diminuição expressiva do número de pessoas				
8		do novo coronavírus, com o aumento da violência, a discriminação e com a dificuldade de				
9		a céu aberto. A proximidade inibiria a violência, a distância, a estimularia . É uma discussão				
10		toritarismo, o gosto pela hierarquia mantida pela violência, a exploração predatória da natureza, a boçalidad				
11		autoridades para conseguirem sair do ciclo de violência . A forma como a Lei Maria da				
12		é obscura, carregado de sujeiras, vísceras e violência . A forma do popo se torna o				
13		ras promessas masculinas. Afinal, quanta ilusão e violência a frase fique em casa, vou cuidar				
14		algum mal-entendido gera alguma situação de violência . A gente não gosta de ver nossos				
15		[está mais escassa]. Na avaliação dele, a violência, a lei seca e a crise afetaram				
16		ênfase dada a jogos de ação e violência . A lógica de matar ou morrer é				
17		vão trazer o terror, o medo, a violência . A pandemia de Covid-19, segundo ela, não				
18		vulnerabilidade social agravada (e o medo da violência a partir daí) e o aumento da				
19		o tempo todo. Isso é uma grande violência . A partir de agora a cantora e				
20		um tempo de mais bondade e menos violência, a passagem só faz sentido porque diz				

Source: elaborated by the authors

We also present table 6 with contexts of use for *violência* in the *corpus* of *Folha de São Paulo*.

Table 6 — Contexts of use for the word violência in the Portuguese corpus

Entre elas, ameaças (1.844), cárcere privado (1.243), tentativa de feminicídio (2.688), violência doméstica e familiar (35.769) e violência obstétrica (116). E o que isso quer dizer? Que a mulher, cada vez mais, observa as deficiências que a cercam, se une a outras e luta para ter voz e ocupar os seus espaços. Na política não tem sido diferente.

A violência sexual durante o conflito da Síria não é uma forma de tortura e humilhação dirigida apenas às mulheres: homens também são vítimas, especialmente os LGBT.

No gabinete que ocupa desde dezembro, em Buenos Aires, a algumas quadras da Casa Rosada, a ex-deputada Victoria Donda, 42, afirma ser possível trabalhar para acabar com a violência contra negros, pobres, mulheres e outras ditas minorias — “que muitas vezes nem minorias são”.

Source: elaborated by the authors

Contexts of use for the word *violência* in the Portuguese *corpus* show that violence is always a noun followed by prepositions, adjectives, or verbs, such as *aumentar* (to increase), *contra* (against) and *policial* (police). Furthermore, *violência policial* (police violence) against Black people is presented as a public issue related to structural racism, a historical debt left by colonialism and slavery.

According to the context of use, quarantine aimed at controlling Covid-19 infections increased the number of acts of *violência* (violence), especially against women. The explanation for this is that couples were forced to spend more time together, resulting in more time spent with an aggressive person in an abusive relationship. The excerpts suggest

that quarantine has become a risk factor for *violência doméstica* (domestic violence), and the reports urge neighbors to be alert to any suspicion of aggression and call the police whenever necessary.

Figure 2 — Clusters/N-grams for the word *violência* in the corpus of Folha de São Paulo

Concordance			Concordance Plot		File View	Clusters/N-Grams	Collocates	Word List	Keyword List
						Total No. of Cluster Types	4490	Total No. of Cluster Tokens	5880
Rank	Freq	Range	Cluster						
1	80	7	violência contra a						
2	74	7	violência contra a mulher						
3	42	5	violência policial contra						
4	30	3	violência policial contra negros						
5	24	3	violência policial no						
6	23	7	violência contra as						
7	22	6	violência contra as mulheres						
8	21	5	violência ou grave						
9	21	5	violência ou grave ameaça						
10	21	3	violência policial contra negros	voltarfacebookwhatsappwitt					
11	20	7	violência doméstica e						
12	20	6	violência policial e						
13	19	6	violência de gênero						

Source: elaborated by the authors

12

In figure 2, we see that *clusters* determine the most frequent combinations of words, displaying 4490 words that constantly co-occur in the *corpus*. Among them, we highlight *violência* (violence) and *contra* (against) that occur in association 80 times and which bring as the object of aggression the *mulher* (woman) in 74 contexts, and the *negros* (Blacks) in 21 contexts. We can also verify that violent acts are related to police force concerning the racial issue, leading us to understand that, in the Brazilian cultural environment, there is a tendency for the police to act against groups of Blacks and mestizos, generally also related to lower classes.

In turn, when the environment is *doméstico* (domestic), aggressions take on a familiar and misogynistic aspect and are concentrated against *mulheres* (women), being *psicológica* (psychological) and *sexual* (sexual). We understand, therefore, that the word *violência* (violence) in the Portuguese language *corpus* tends to build its meaning linked to issues of gender, sexuality, and vulnerability.

Figure 3 — Collocates for the word violência in the corpus of Folha de São Paulo

Concordance						Concordance Plot	File	View	Clusters/N-Grams	Collocates	Word List	Keyword List
Total No. of Collocate Types: 57						Total No. of Collocate Tokens: 390						
Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate							
9	21	21	0	12.70559	pedem							
10	1	1	0	12.63520	vítimas							
11	1	1	0	12.37217	franca							
12	2	0	2	11.95713	america							
13	3	0	3	11.89824	ananindeua							
14	4	0	4	11.56481	ce							
15	1	0	1	11.37217	salto							
16	4	0	4	11.17577	mata							
17	1	1	0	11.05024	franceses							
18	1	0	1	10.86967	apos							
19	21	0	21	10.84166	negros							
20	1	1	0	10.49770	posts							
21	21	21	0	10.39235	fim							
22	4	0	4	10.35722	meninas							
23	21	21	0	10.07799	eua							
24	26	0	26	9.95540	policial							
25	31	0	31	9.63987	foto							
26	2	0	2	9.49770	latina							
27	26	0	26	9.48015	loading							
28	1	1	0	9.00294	org							

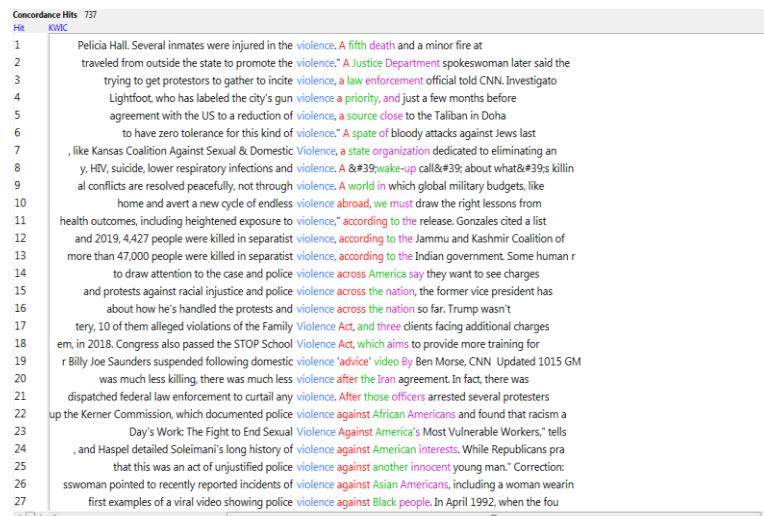
Source: elaborated by the authors

Among collocates, that is, words that occur more frequently in proximity within the same sentence, we find examples such as *vítimas* (victims), *negros* (Blacks), *meninas* (girls). We narrowed the search for up to five words to the right and to the left, considering the search word *violência*, which leads us to believe that all these nouns are related to the group of people who suffer the violent act. This is implied in the meaning of the word *vítima*, which is someone who is dominated, subjugated or harmed in some way, that is, affected by an aggressive action without a chance to defend themselves. Among these, we can include two groups of people: "Blacks" and "girls".

It is interesting to note that, although there are also cases of abuse of male minors, the word that appears most frequently is in the female gender. An explanation can be found in an article by Tolentino (2020), published on the website *CartaCapital*. According to the columnist, the 2018 *Public Security Yearbook* registered 66,000 rapes in Brazil, with 53.8% of the victims being girls 13 years of age or younger. For her, the need for social isolation made girls more prone to abuse.

CNN. Regarding the English *corpus*, we present figure 4, with the most frequent concordances of the word "violence."

Figure 4 — Concordance of the word “violence” in the corpus of CNN



Source: elaborated by the authors

When analyzed in *AntConc* software, "violence" is related to negative words such as "injured" and "separatist;" qualifiers such as "domestic," "police" and "gun;" verbs that denote the idea of an active action, such as "to promote" and "to incite;" and quantifiers such as "any" and "less" (Figure 4). In most cases, the word appears at the end of the sentence and serves as a noun. However, it is possible to find "violence" preceding words related to geographic spaces, such as "abroad" and "across," linked to the preposition "against," or exercising the function of an adjective, as in "violence advice" and "violent act". We see, in Table 7, contexts for the use of the word:

Table 7 — Contexts of use for the word “violence” in the English corpus

This week, in a nation riven by a pandemic and violence in the streets, the uneasy marriage between President and platform came to a spectacular crackup. For the first time, Twitter added a link to two of Trump's tweets, providing a fact check that disputed his condemnation of mail-in ballots.
In a statement previously provided to CNN, Biden's deputy campaign manager and communications director, Kate Bedingfield, denied Reade's allegation, calling it "untrue." "Vice President Biden has dedicated his public life to changing the culture and the laws around violence against women.
We are active participants in all major international organizations that promote universal human rights and work to eliminate discrimination against women and girls. All Nordic countries have ratified the Council of Europe Convention on preventing and combating violence against women and domestic violence (the Istanbul Convention).

Source: elaborated by the authors

Figure 5 — Clusters/N-grams for the word "violence" in the corpus of CNN

Clusters Results 1			
		Total No. of Cluster Types	Total No. of Cluster Tokens
Rank	Freq	Range	Cluster
1	76	7	violence and
2	70	7	violence in
3	65	7	violence against
4	36	6	violence against women
5	19	6	violence that
6	18	6	violence in the
7	14	5	violence is
8	13	6	violence or
9	11	3	violence against women act
10	11	5	violence the
11	10	4	violence prevention
12	9	2	violence in afghanistan
13	9	4	violence to
14	9	5	violence, and
15	8	3	violence as
16	8	3	violence by
17	8	5	violence has
18	8	4	violence on
19	6	4	violence in the region

Source: elaborated by the authors

In Table 7 and Figure 5, the words often related to "violence" are "against women (action)," "in Afghanistan" and "the region." Thus, "violence" is a noun which names a sequence of aggressive acts – in this case, more physical than psychological – against a group, or it is characteristic of some locality, country or region.

The conflict begins with the construction or deconstruction of the enemy, "another so foreign and so distant" that when dehumanized and, consequently, made unworthy of fundamental rights, can be "tortured, maimed, slaughtered" (Nelson, 2002, as cited in M. Baker, 2006, p. 14). Furthermore, unlike ideological and class conflicts, it is not about choosing a side but an identity, posing the question: What are you? (Huntington, 1993, p. 27).

15

Figure 6 — Collocates for the word violence in the corpus of CNN

Collocates					
		Total No. of Collocate Types:	2035	Total No. of Collocate Tokens:	7370
Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate
1	2	1	1	10.29232	endangers
2	1	0	1	9.29232	wrought
3	2	2	0	9.29232	whirlpools
4	1	0	1	9.29232	wedge
5	1	0	1	9.29232	vigorously
6	1	1	0	9.29232	upticks
7	1	0	1	9.29232	troubles
8	3	3	0	9.29232	triumphed
9	1	0	1	9.29232	thrived
10	1	0	1	9.29232	thief
11	1	1	0	9.29232	tamp
12	1	1	0	9.29232	suppresses
13	1	1	0	9.29232	steered
14	1	0	1	9.29232	stds
15	1	1	0	9.29232	stalk
16	1	1	0	9.29232	small
17	1	0	1	9.29232	slayings
18	1	0	1	9.29232	singles
19	1	1	0	9.29232	sikhs
20	*	*	*	0.200222	unknown

Source: elaborated by the authors

When analyzing collocations of words that often accompany violence, the one that appears the most in both L1 and R1 is "endangering," a verb conjugated in the third person singular in the simple present tense. According to *Cambridge Dictionary* (2021), "endanger" means putting someone or something at risk or in danger of being harmed, damaged or destroyed.

Another word is "vigorous," an adjective that frequently appears in R1. In reinforcing the word's meaning, "vigorous" can describe the violent act (very forceful, extreme violence), or a more assertive way to fight the violence itself.

The occurrence of the word "troubles," on the other hand, refers to a conflict in Northern Ireland that ended in 1998 with the *Good Friday Agreement* in Belfast. Since 2020, the pandemic and the departure of the United Kingdom from the European Union (Brexit) have brought severe consequences for the region, which means that after 22 years of the peace agreement, the separatist conflict could return.

16 El País. In the Spanish *corpus*, the word *violencia* is presented in Figure 7 among its most frequent contexts of use.

Figure 7 — Concordance of the word *violencia* in the corpus of *El País*



Source: elaborated by the authors

The word *violencia* is close to verbs, usually in the object position. In some cases, the verbs express an action against the subject (*Tendrá que sufrir más violencia*), or an action of the individual in relation to itself (*porque infinge violencia a su propio cuerpo*). However, its most common use as a noun refers to a phenomenon imposed on other people, groups or entities (*Un segundo brote bajo la violencia/comunidades que huyen de la*

*violencia/castigadas por el machismo y la violencia/víctimas de la violencia etc.). Furthermore, violencia appears as a consequence of corruption (*la herencia de la corrupción*) and lack of awareness, as in many expressions: *Hay que tomar conciencia de la violencia/ Ignorar es una alta forma de violencia* etc.*

Figure 8 — Clusters/N-grams for the word violencia in the corpus of El País

Concordance			Concordance Plot			File View		Clusters/N-Grams	Collocates	Word List	Keyword List
								Total No. of Cluster Types	2534	Total No. of Cluster Tokens	3588
Rank	Freq	Range	Cluster								
1	267	7	violencia de género								
2	45	7	violencia contra las								
3	43	7	violencia contra las mujeres								
4	24	6	violencia contra la								
5	23	7	violencia de género en								
6	23	4	violencia machista en								
7	15	4	violencia contra la mujer								
8	15	2	violencia machista las								
9	13	3	violencia contra los								
10	13	2	violencia machista las 24 horas								
11	12	2	violencia machista las 24 horas y								
12	12	5	violencia sexual y								
13	11	3	violencia de género de								

Source: elaborated by the authors

Analyzing the *clusters* for *violence*, the form of violence that appears with the highest occurrence is that against women, or sexual violence. Both are related to the same paradigms, namely rape, harassment, and death. At the same time, it deals with practices related to different groups, such as child sexual harassment and physical aggression against women (*violencia de género, violencia contra las mujeres, violencia machista*).

Figure 9 — Collocates for the word violencia in the corpus of El País

Concordance			Concordance Plot			File View		Clusters/N-Grams	Collocates	Word List	Keyword List
								Total No. of Collocate Types:	2406	Total No. of Collocate Tokens:	11956
Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate						
1	4	2	2	9.26155	machistacronología						
2	2	1	1	9.26155	incesante						
3	2	0	2	9.26155	hereda						
4	2	1	1	9.26155	corroboran						
5	2	1	1	9.26155	calificación						
6	5	2	3	8.99852	genero						
7	5	4	1	8.58348	suprime						
8	1	0	1	8.26155	zamarreo						
9	1	0	1	8.26155	yihadismo						
10	1	0	1	8.26155	wuhan						
11	1	0	1	8.26155	vulnere						
12	1	0	1	8.26155	vulnera						
13	1	1	0	8.26155	visitadas						
14	2	1	1	8.26155	visibilizó						

Source: elaborated by the authors

Also, in *collocates* (Figure 9), it is possible to notice the wide attention given to domestic violence, including in other places, like *Wuhan*, where the first case of Covid-19 was reported and brought the need for social isolation. Once again, we have a relationship between the increase in violence against women and the pandemic. Thus, we find words that demonstrate a concern to prevent and reduce this type of violence (*corroboran*, *calificación*) and an attempt to understand the causes and consequences of violent acts, as seen in *hereda*, *vulnera*, *machista* and *incesante*.

We can also verify that the most frequent combinations are those between the word *violencia* and the prepositions *de*, *contra* and *a*, in addition to some modifiers of violence: *machista*, *policial*, *racial*, *simbólico*, *postelectoral*.

Violence is a noun followed by another noun. The most frequently used are *precariedad* (precariousness), *guerra* (war) and *biopolítica* (biopolitics). These words are found in texts about immigrants, war refugees and their living conditions, working from home in quarantine, and things the government must do to control the Covid-19 disease.

The fourth most frequent words are verbs, specifically "to generate," which means violence is a consequence of people's actions and something created by reality and the environment. Below, we can verify Table 8 with some examples of contexts of use for the word *violencia* in Spanish:

Table 8 — *Contexts of use for the word violencia in the Spanish corpus*

En el vestíbulo decorado con montones de basura, grafitis caóticos y panteletos rojos del Frelimo en columnas agrietadas, Carlos enumera las tribulaciones del infame laberinto: alcoholismo, violencia doméstica, tráfico de drogas, y los niños que, jugando, caen a los fosos vacíos de los ascensores o de las cornisas sin parapetos.
La actividad fue intensa y, de nuevo, despertó en el país el debate acerca de la desigualdad y de la inseguridad de las mujeres. Los colectivos de defensa de sus derechos no se cansan de denunciar el creciente clima de violencia basada en el género y, sobre todo, las consecuencias para las víctimas y para la sociedad, porque se está transmitiendo una peligrosa sensación de impunidad.
El informe contiene datos, orientaciones, análisis, advertencias. Todos ellos necesarios. Pero para Najat Maala M'jid, representante especial del secretario general de la ONU sobre Violencia contra la Infancia, el mensaje fundamental es uno: "Todas las formas de violencia contra la infancia se pueden prevenir. No lo perdamos de vista".

Source: elaborated by the authors

The samples above show that gender violence and against children are the majority of occurrences in the Spanish *corpus*. This is one of its remarkable differences in relation to

other *corpora*, which does not mean this theme has no appearance or is treated as less necessary by the latter.

Le Monde. In Figure 10, contexts analyzed by using AntConc's Concordance tool in the French corpus show that violence appears several times as an adjective: phase de violence, victime de violence, l'augmentation de la violence, des scènes de violence etc., most of them in the middle of the sentence. When the word's function is a noun, as in violence administrative, it is qualified or specified with very negative words, such as arbitraire, aveugle, au malaise (arbitrary, blinded, and discomfort). It is also possible to find violence associated with certain locations (aux États-Unis; au Brésil) and always related to physical aggression.

In some sentences, the word appears at the end, with a more general sense – in opposition to harmony and peace – and quotation marks, bringing a voice of authority and exempting the publication from any responsibility regarding the displayed opinion.

Figure 10 — Concordance of the word violence in the corpus of Le Monde

Hit	KWC
1	à labdomen, « une scène dune certaine violence », a complété le procureur de la République. De
2	nest pas le moment dencourager la violence », a déclaré Joe Biden. « Nous traversons une cri
3	s sont suspendus pour éviter tout risque de violence », a expliqué Sorbonne Université. A Nanterre, u
4	nt reprendre de façon harmonieuse et sans violence », a précisé Abdourahmane Diallo, un autre respon
5	nvie de les entraîner de nouveau dans cette violence », a-t-il ajouté. Il a expliqué participer
6	sont traduits par endroits par des actes de violence », a-t-il reconnu, en faisant état de
7	avant de menacer [de usage] de la violence », a écrit lartiste de 30 ans sur son
8	partir de juin 2019, une troisième phase de violence a été observée avec des attaques contre la
9	té du territoire national » et « incitation à la violence », accusations qui visent généralement les milita
10	e bananière, ou Papers, Please, qui parle de violence administrative et de contrôle des frontières. Les
11	été victime dune agression « dune extrême violence », alors qu'il voulait contrôler le ticket d
12	été victime dune agression « dune extrême violence », alors qu'il voulait contrôler le ticket d
13	été victime dune agression « dune extrême violence », alors qu'il voulait contrôler le ticket d
14	de prouver quelle a été victime de violence arbitraire au lycée. Voulez-vous parler de cette
15	jours du week-end, selon lorganisation Gun Violence Archives. De quoi assombrir encore un peu plus
16	emain précédente. « Laugmentation de la violence armée survient à un moment où la chaleur
17	28 avril. RICK RYCROFT / AP Des scènes de violence au Brésil et au Liban Le confinement n
18	cadres larges et profondeur de champ) à la violence, au malaise, à la frustration de son héros
19	doute : celle dune inexorable spirale de la violence, au sein dune société où la haine
20	étéée lundi dune « escalade imminente de la violence » autour de Tripoli. « LUE appelle toutes les

19

Source: elaborated by the authors

In clusters of 2 to 5 words, *violence* is more frequently used with prepositions and conjunctions: *de*, *et*, *qui*, *contra*, *entre* (of, and, that, against and between). In this context, *violence* is a noun, and its specification is used by prepositions *de* and *contra*, as shown in figure 11 below. The preposition *entre* is used to express mutual acts of violence between two entities; in this case, the most frequent words are *Iran* and *États-Unis*. Another frequent

word is *policière* (police), which is an adjective and, in this way, qualifies the word *violence*. Most of these sentences are accompanied by the word *racisme* (racism) and are more frequent in July, as mentioned before, due to an international reaction and focus on police and racial violence during the last week of May.

The word violence is related to physical and mutual aggressions between the police and the population, wars and terrorism between countries.

Figure 11 — Clusters/N-grams for the word violence in the corpus of Le Monde

Total No. of Cluster Types 1125			Total No. of Cluster Tokens 1404		
Rank	Freq	Range	Cluster	Concordance	Concordance for the Year
1	26	6	violence et	0.000000	0.000000
2	15	6	violence de	0.000000	0.000000
3	13	5	violence à	0.000000	0.000000
4	12	5	violence des	0.000000	0.000000
5	10	5	violence qui	0.000000	0.000000
6	8	4	violence contre	0.000000	0.000000
7	7	4	violence contre les	0.000000	0.000000
8	7	4	violence policière	0.000000	0.000000
9	7	5	violence », a	0.000000	0.000000
10	6	3	violence du	0.000000	0.000000
11	6	1	violence entre	0.000000	0.000000
12	6	1	violence entre l	0.000000	0.000000
13	6	1	violence entre liran	0.000000	0.000000
14	6	1	violence entre liran et	0.000000	0.000000
15	6	2	violence physique	0.000000	0.000000
16	6	3	violence à l	0.000000	0.000000
17	5	3	violence est	0.000000	0.000000
18	5	4	violence et de	0.000000	0.000000
19	5	4	violence au	0.000000	0.000000

Source: elaborated by the authors

We can see in figure 11 that the word *violence* is often coordinated with negative words from the same semantic field or is presented as a cause/consequence of aggressiveness (attack, authoritarianism, abuse, harassment, insecurity etc.). In figure 11, the preposition *à* refers to a location or an action on something or someone.

In collocates (Figure 12), considering 5L and 5R, the most highly ranked word is *elan* (impulse), followed by *catalysez* (catalyze), *speed* and *momentum*. Violence occurs, then, when there is a lack of emotional control and a lack of self-control.

Figure 12 — Collocates for the word violence in the Le Monde corpus

Collocates for the word violence in the Le Monde corpus					
Concordance		Concordance Plot		File View	Clusters/N-Grams
Total No. of Collocate Types: 1156		Total No. of Collocate Tokens: 3510			
Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate
1	2	0	2	10.49476	elan
2	2	1	1	10.49476	catalysez
3	3	1	2	10.07972	speed
4	3	1	2	10.07972	momentum
5	1	0	1	9.49476	établit
6	1	1	0	9.49476	érotisé
7	1	0	1	9.49476	épanche
8	1	0	1	9.49476	émane
9	1	1	0	9.49476	ébranlés
10	1	1	0	9.49476	vulnérable
11	1	0	1	9.49476	voulez
12	1	0	1	9.49476	viscérale
13	1	0	1	9.49476	veronica
14	1	0	1	9.49476	verbale
15	1	0	1	9.49476	typhons
16	1	1	0	9.49476	transsexualité
17	1	1	0	9.49476	tranchent
18	1	1	0	9.49476	tortueux
19	2	2	0	9.49476	tolérerai
20	1	1	0	9.49476	tison

Source: elaborated by the authors

21

Table 9 — Contexts of use for the word violence in the French corpus

De plus, « j'ai entraîné ma famille dans une épreuve d'une violence inouïe et je n'ai aucune envie de les entraîner de nouveau dans cette violence », a-t-il ajouté. Il a expliqué participer à cette émission « pour défendre l'honneur de [sa] femme et [sa] famille », dont l'« honneur a été déchiqueté dans cette affaire ».

L'Union européenne (UE) s'est inquiétée lundi d'une « escalade imminente de la violence » autour de Tripoli. « L'UE appelle toutes les parties à s'engager dans un processus politique sous l'égide des Nations unies », a plaidé le chef de la diplomatie européenne, Josep Borrell.

Le Mali, où le conflit a éclaté en 2012 avant de se propager, le Burkina Faso et le Niger sont en proie à une alarmante dégradation sécuritaire. Ils sont le théâtre d'une forte poussée djihadiste combattue par les armées nationales et étrangères, de brutalités intercommunautaires et de multiples trafics alimentant la violence. Celle-ci a fait des milliers de morts, combattants et civils, et des centaines de milliers de déplacés.

Source: elaborated by the authors

The word *violence* shown in Table 9 is related to institutions, military power, and international conflicts (warzones). It can describe a scene (*une scène d'une certaine violence*) or an event with no precedent (*une épreuve d'une violence inouïe*), as well as a *diplomatique effort* by the European Union to decrease such acts. Here, *violence* is a noun accompanied by adjectives and pronouns (*certaine, cette*) which refer to already mentioned (or soon-to-be mentioned) semantic meanings linked to the word in question, such as

harming, security degradation, combat by national and international military forces, brutality between communities and traffic (*le blessant à l'abdomen, une alarmante dégradation sécuritaire... combattue par les armées nationales et étrangères, de brutalités intercommunautaires et de multiples trafic*) etc.

Final Remarks

Analyzing these four sets of texts about violence, we observed similarities and differences in the whole *corpora*. Some types of conflicts are presented in all of them, such as violence against women and racial violence exerted by public agents, like the police. However, some of them are very particular to some contexts, like xenophobia, war and terrorism. Some conflicts may stand out more than others because of people's experiences and perspectives about reality.

M. Baker (2006) explains some theories about conflict. For Bachrach and Baratz (1962/2017), for instance, observable conflicts are the presence of power because one party tries to subvert another "to act against its will or what is perceived to be its interests." On the other hand, power is adequate when there is no observable conflict. Therefore, M. Baker (2006, p. 1) defines a conflict as a "situation in which two or more parties seek to undermine each other because they have incompatible goals, competing interests, or fundamentally different values."

Additionally, the word "violence" has different positions and functions depending on the language. In French, Portuguese, and Spanish, it was common to find violence in the middle of sentences, while in English, most occurrences happened at the end. In English, violence can be an adverb of manner, describing phenomena, acts and feelings. Although this last meaning of violence is found in the four languages, we could not observe it in the *corpora*. Moreover, the most common meaning for violence was physical violence, whereas psychological violence appeared in Portuguese and Spanish *corpora*, but not as frequently as its counterpart.

Some verbs presented violence as motivated, created and directed against a group of subjugated people. According to M. Baker (2006, pp. 16–17), narratives fulfill the human need to categorize things, people, and reality. They are not different from stories, and they consist of different versions, which depend on individuals' stories and experiences. They can mold people's lives because they are in some way believable and verifiable by reality.

Moreover, present narratives are made of past narratives repeatedly retold as a means to control and a way to define identity and social roles (M. Baker, 2006, pp. 20–21). They can be used as a way of oppressing and exerting power. Proportionally, when there is more freedom of speech, the more conflicts there are between competing narratives (M. Baker, 2006, pp. 22–23).

Finally, xenophobia is a problem for European countries since many immigrants escaping from war arrive there. In countries where presidents are ultraconservative and populist in their attitudes, decisions, and speeches, violence is motivated between the general population and vulnerable communities. Violence against women is a consequence of sexism and the idea of the superiority of men over women, similar to racial violence, which categorizes black people as criminal and aggressive.

REFERENCES

Anthony, L. (2020). *AntConc* (Version 3.5.9) [Computer Software]. AntLab Solutions.
<https://www.laurenceanthony.net/software/antconc/>

Bachrach, P., & Baratz, M. S. (2017). Two faces of power. In J. R. Champlin (Ed.), *Paradigms of political power* (pp. 118–131). Routledge. (Reprinted from “Two faces of power,” 1962, *American Political Science Review*, 56[4], 947–952, <https://doi.org/10.2307/1952796>).

Baker, M. (1995). Corpora in translation studies: an overview and some suggestions for future research. *Target*, 7(2), 223–243.

Baker, M. (1996). Corpora in translation studies: the challenges that lie ahead. In H. Somers (Ed.), *Terminology, LSP and translation studies in language engineering: in honour of Juan C. Sager* (pp. 177–186). John Benjamins.

Baker, M. (2006). *Translation and conflict: A narrative account*. Routledge.
<https://doi.org/10.4324/9780203099919>

Baker, P. (2006). *Using corpora in discourse analysis*. A&C Black.

Barbosa, M. A. (1990). Considerações sobre a estrutura e funções da obra lexicográfica: metodologia, tecnologia e condições de produção. *Colóquio de Lexicologia e Lexicografia*, Universidade Nova de Lisboa, 229–241.

Barbosa, M. A. (1998). Terminologização, vocabularização, científicidade, banalização: relações. *Acta Semiótica et Lingvistica*, 25–44.

Beaugrande, R. (1994). Cognition, communicatio, translation, instruction: The geopolitics of discourse. In R. Beaugrande, A. Shunnaq, & M. H. Helie (Eds.), *Language*,

discourse and translation in the west and middle east (pp. 1–22). John Benjamins Publishing Company. <https://doi.org/10.1075/btl.7.05bea>

Berber Sardinha, T. (2004). *Linguística de corpus* [Corpus linguistics]. Manole.

Bradbury-Jones, C., & Isham, L. (2020). The pandemic paradox: The consequences of COVID-19 on domestic violence. *Journal of Clinical Nursing*, 29(13–14), 2047–2049. <https://doi.org/10.1111/jocn.15296>

Bührig, K., & House, J. (2004). Connectivity in translation transitions from orality to literacy. *Multilingual communication*, 3, 87.

Cambridge Dictionary. (2021). Endanger. In Cambridge online dictionary. Retrieved February 3, 2023, from <https://dictionary.cambridge.org/us/dictionary/english-portuguese/endanger>

Dubois, J., Giacomo, M., Guespin, L., Marcellesi, C., Marcellesi, J., & Merel, J. (1973). *Dictionnaire de Linguistique*. Larousse.

Eco, U. (2001). *Experiences in translation* (A. McEwen, Trans.). Toronto: University of Toronto Press.

24 Evans, M. L., Lindauer, M., & Farrell, M. E. (2020). A pandemic within a pandemic: Intimate partner violence during Covid-19. *New England Journal of Medicine*, 383(24), 2302–2304. <https://doi.org/10.1056/NEJMp2024046>

Faiq, S. (2019). *Discourse in translation*. Taylor & Francis.

Fairclough, N., & Wodak, R. (1997). Critical discourse analysis. In T. Van Dijk (Ed.), *Discourse as Structure and Process* (pp. 258–284). Sage.

Firth, J. R. (1957). *Man and Culture: an evaluation of the work of Bronislaw Malinowski*. Andesite Press.

Halliday, M. A. K. (1991). Corpus studies and probabilistic grammar. In K. Aijmer, & B. Altenberg (Eds.), *English corpus linguistics*. Longman.

Halliday, M. A. K. (1992). Language as system and language as instance: the corpus as a theoretical construct. In J. Svartvik (Ed.), *Directions on Corpus Linguistics. Proceedings of the Nobel Symposium 82*. Mouton de Gruyter: Stockholm.

House, J. (Ed.). (2014). *Translation: a multidisciplinary approach*. Springer.

Huntington, S. (1993). The clash of civilizations? *Foreign Affairs*, 31.

Mason, I. (1994). Discourse, ideology and translation. In R. Beaugrande, A. Shunnaq, & M. H. Heliel (Eds.), *Language, discourse and translation in the west and middle east* (pp. 23–33). John Benjamins Publishing Company.
<https://doi.org/10.1075/btl.7.06mas>

Sager, J. (1998). In search of a foundation: towards a theory of the term. *Terminology*, 5, 41–57.

Sinclair, J. M. (1991). *Corpus, concordance, collocation*. Oxford.

Sinclair, J. M. (1992). Priorities in discourse analysis. In M. Coulthard (Ed.), *Advances in spoken discourse analysis*. Routledge.

Tognini-Bonelli, E. (1996). Section 2: the Malvern Seminar: Towards Translation Equivalence from a Corpus Linguistics Perspective. *International Journal of Lexicography*, 9(3), 197–217.

Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. John Benjamins.

Tolentino, L. (2020, August 21). A cumplicidade nos casos de abuso sexual contra meninas [Complicity in cases of sexual abuse against girls]. *CartaCapital*. <https://www.cartacapital.com.br/opiniao/a-cumplicidade-nos-casos-de-abuso-sexual-contra-meninas/>

¹ It is important to emphasize that, for this article, we do not focus on studies on journalistic discourse or journalistic translation.

² In the original : "(...) le **mot** est un élément linguistique significatif composé d'un ou de plusieurs phonèmes; cette séquence est susceptible d'une transcription écrite (ideogrammatique, syllabaire ou alphabétique) comprise entre deux blancs; dans ses divers emplois syntaxiques, elle garde sa forme, soit totalement, soit partiellement (dans le cas de la flexion). Sur le plan sémantique, le mot dénote un objet (substantif), une action ou un état (verbe), une qualité (adjectif), une relation (préposition), etc." (Dubois et al., 1973, p. 343)

³ For Baker (1995), the survey of the number of *words* of a text presupposes the count of all words that occurred in the text (*tokens*), whereas each different occurrence (*types*) is counted only once.

⁴ We only use nouns because they have specific meanings.

⁵ We will not address the case specifically. For more information: <https://www.nytimes.com/2020/05/31/us/george-floyd-investigation.html>