Cogent Arts & Humanities

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/oaah20

Tracing the scope of fear in corpus: similarities and differences in cross-domain/genre texts

Ignacio Rodríguez Sánchez, Antonio Reyes Pérez, Manuel Cebral Loureda, Ester Bautista Botello & Nohemí Lugo Rodríguez

To cite this article: Ignacio Rodríguez Sánchez, Antonio Reyes Pérez, Manuel Cebral Loureda, Ester Bautista Botello & Nohemí Lugo Rodríguez (2024) Tracing the scope of fear in corpus: similarities and differences in cross-domain/genre texts, Cogent Arts & Humanities, 11:1, 2416732, DOI: 10.1080/23311983.2024.2416732

To link to this article: https://doi.org/10.1080/23311983.2024.2416732

COGENT ARTS & HUMANITIES
2024, VOL. 11, NO. 1, 2416732
https://doi.org/10.1080/23311983.2024.2416732

LITERATURE, LINGUISTICS & CRITICISM | RESEARCH ARTICLE

OPEN ACCESS Check for updates

Tracing the scope of fear in corpus: similarities and differences in cross-domain/genre texts

Ignacio Rodríguez Sánchez[a], Antonio Reyes Pérez[a], Manuel Cebral Loureda[b], Ester Bautista Botello[a] and Nohemí Lugo Rodríguez[b]

[a]Universidad Autónoma de Querétaro, México; [b]Instituto Tecnológico y de Estudios Superiores de Monterrey, México

ABSTRACT

Even if to mainstream psychologists fear is one of the seven universal emotions, discrete, measurable and with clearly distinct features, in the humanities we consider fear as a widespread concept we associate with more complex prompts than the physiological response to a hazard. This research explores various ways we describe FEAR in Spanish. For this we have made use of digital humanities tools and methods, mainly corpus linguistics and natural language processing, which enable us to explore, present and visualize linguistic elements that define fear in Mexican society. Thus we have explored this emotion (and its family: anxiety, horror, apprehension, dread, panic, terror) by examining the way it is verbalized in an *ad hoc* corpus covering four genres/domains: chronicle, essay, the press, and social media, specifically, tweets. A set of semantic similarity and ranking metrics were applied to the texts to identify each genre's characteristics in association with fear. The results show that fear is an emotion that, even if it differs depending on the genre, responds to the prompts of a modern society in which danger is still being represented by illness, violence, power, or an out-group.

ARTICLE HISTORY
Received 3 May 2024
Revised 8 October 2024
Accepted 10 October 2024

KEYWORDS
Fear; corpus linguistics; semantic similarity; hierarchical classification;·multi-dimensional analysis

SUBJECTS
Language & Linguistics; Corpus Linguistics; Semantics

1. Introduction

1.1. Psychologists' traditional view on emotions

The semantics of emotions is a recent research field in which linguistic material is analyzed in order to infer the way we experience certain emotions. Broadly speaking, work in this field frequently features two general types of research: one is statistical processing of textual information, and the other involves biometric studies of how people experience certain feelings. In a manner similar to color analysis (Berlin & Kay, 1969; Biggam et al., 2011; Cook et al., 2005), when focusing on emotions, researchers tend to conclude there are universals based on their findings, but there is also a significant difference between languages and cultures (Jackson et al., 2019).

The second type of research (based on surveys or experimental settings on what people perceive) focuses mostly on the predominant paradigm of Basic Emotion Theory (Ekman, 2007; Ekman & Friesen, 2003; Tomkins, 1962). Ekman's work contemplates seven basic universal emotions (and their families). Researchers consider that in order for a mood or feeling to acquire the status of emotion, it must comply with a dozen characteristics (for example, that it be universal, possess a distinctive physiology, and be present in other primates, etc.). In recent reformulations of this theory (Ekman & Cordaro, 2011), the characteristics that define the concept of basic emotion have increased slightly; the catalog of emotions has also grown to almost twenty basic emotions (albeit not all universal).

CONTACT Ignacio Rodríguez Sánchez i.rodriguez@uaq.edu.mx Lecturer/researcher at the Facultad de Lenguas y Letras, Universidad Autónoma de Querétaro, México.

In general, the relationship affective psychologists have with linguistics is ambivalent, and although they carry out studies based on verbal information, most tend not to delve in linguistic topics because they believe 'Language is socially constructed; basic emotions are not' (Ekman & Cordaro, 2011, p. 369). To them, language is basically a symbolic and cultural mechanism used for labeling and categorizing things and 'categorization and labeling [are] secondary, optional steps' (Scherer & Moors, 2019, p. 720)

The GRID project (Fontaine et al., 2013) is the emotion-themed equivalent of the World Color Survey Database (Cook et al., 2005) project. Research related to this project (Scherer & Fontaine, 2019) analyzes the semantic structure of words associated to emotions in 24 languages. Participants classify terms referring to emotions according to 142 characteristics (referring to evaluation, bodily symptoms, expressions, action tendencies, subjective feelings and regulation). With this information they carry out a series of hierarchical regression analyses that reduce the characteristics to five.

Research in which participants report how they emotionally experience certain words, through norms of valence, arousal and dominance (Bradley & Lang, 1999; Guasch et al., 2016; Stadthagen-Gonzalez et al., 2017; Stadthagen-González et al., 2018; Warriner et al., 2013), implies even greater simplification of the meaning of words. These lists are useful to behavioral psychologists, computer specialists and marketing experts, who converge in new disciplines such as Sentiment Analysis. The virtue of these approaches is also their main defect: they are used to quickly and roughly compute text valuations.

In sum, psychologists normally hold a view that emotions are constituted by a fixed set of dimensions of meaning motivated by human physiology. This idea is prototypical of psychologists' approach to language and in particular to the lexicon: they solely take into account the referential (denotative) aspects, disregarding what is related to language use, in other words, context, grammar, genre, and the type of circumstances in which it appears. This is of course psychologists' deliberate methodological choice, not an error or an oversight.

1.2. Alternative views on emotions: context

No one argues that language is a window that enables us to get information regarding feelings and emotions, but psychologists don't usually agree with anthropologists and linguists when reminded that 'English words like *fear* or *anger* refer to essentially American ethnopsychological concepts and cannot thus be used as culture-free analytical tools in analyses of any disciplinary orientation' (Ogarkova, 2013, p. 47).

New theories are beginning to take hold and replacing traditional psychological views of emotions. Most of them underline the importance of context in the semantics of emotions. For example, the Theory of Constructed Emotion (Barrett & Russell, 2014) works on the premise that variation in emotions is the norm and that concepts are culture-related, collecting evidence from neuroscience, graph theory and prototype theory. In very much the same way color does not exist in the real world (it is how the brain interprets photons), emotions do not exist without a brain interpreting instances of emotions and categories. In other words, we construct our own emotional experiences based on context (body language, social situation, and cultural expectations).

The emotion theory paradigm in linguistics currently corresponds to the Extended Conceptual Metaphor Theory (Kövecses, 2020), which is a development of the pioneer work of Lakoff (Lakoff & Johnson, 1980) within the field of cognitive linguistics. Kövecses (2015) states that the use of emotion metaphors in discourse is influenced by a variety of contextual factors which can be grouped into four large categories: situational context, discourse context, conceptual-cognitive context, and bodily context.

This multiplicity of perspectives and approaches converge in what is normally described as methodological triangulation. In this case, data obtained by qualitative interviews or responses to tests can be complemented by relevant examples or data retrieved from a corpus using different techniques. The retrieval of these text examples, which are called concordance lines or KWIC (key words in context) or linguistic data, will be reported in this paper. Triangulation (McEnery & Hardie, 2011, p. 209) enables us to confirm findings from different perspectives.

1.3. Purpose

Within this framework underscoring the importance of context, the purpose of our corpus linguistics-based research is to verify the extent to which words associated with the emotion of fear express not solely

what they commonly denote, but other nuances according to co-text, context, text type, as well as to the genre in which they are inserted. We intend to look at how communities of speakers construct meanings and negotiate them from interactions and social behavioral patterns. Concretely, in this paper we are going to see the lexical characteristics (similarities and differences) of four different sections of a corpus collected on the topic of fear.

This paper is structured thus: in Section 2 the theoretical framework that gives this study conceptual and methodological sense is described, and in particular, emphasis is placed on corpus linguistics and natural language processing. Classification work, emphasizing the application of semantic similarity and hierarchical classification metrics, as well as corpus building is explained in Section 3. Section 4 focuses on the experiments undertaken; results are analyzed and discussed. Lastly, Section 5 discusses our findings and future lines of work.

2. Theoretical framework

Similar to the changes mentioned in the previous section, since the end of the last century there has been a paradigm change in linguistics, which is our theoretical starting point. This change basically implies the end of the predominance of a rationalist model, represented by universal grammar, which focused on describing speakers' linguistic competence. Moreover, the new emergentist paradigm is data-driven (empiricist and quantitative) (Sampson, 2013), wherein cognition embraces the computational development corpus linguistics and connectionist models rely on (Beckner et al., 2009; Ellis & Larsen-Freeman, 2006, 2009). In this new paradigm, taking grammar and lexicon as separate entities ceases to make sense and a lexical grammar approach is considered, where each word tends to behave according to the text, context and co-text in which it is confined, establishing semantic, lexical, discursive and genre associations with other texts and elements. Instead of focusing on what words evoke, Firth's original idea is recalled: 'You shall know a word by the company it keeps' (Firth, 1957/1961). Sinclair's idiom principle (Sinclair, 1991, 2004), Hoey's lexical priming theory (Hoey, 2005, 2009), the pattern grammar principle (Hunston & Francis, 2000), Biber's et al. (1998) lexical bundle, collostructional analysis (Gries & Stefanowitsch, 2006), and construction grammar (Goldberg, 1995; 1996; Langacker, 1987), among others, are studies within this paradigm. They are all usage-based (Bybee & Hopper, 2001) and start out inductively.

This paradigm change implies that speakers in their interactions construct and negotiate word meanings and that language is not made up of a set of rules but stems from a complex interaction between speakers. This vision of language based on the negotiation of meanings between agents in a particular context contrasts with the understanding of how Basic Emotion Theory psychologists deal with emotions in language use and comes close to the approach of psychologists who adhere to the Theory of Constructed Emotion.

We will not address all the new paradigm theories in detail, but it is important to note that this is a phenomenon of broad impact in the scientific community. We will now focus on Hoey's and Biber's work as they are especially suitable to understand the scope of this research.

2.1. Lexical priming

The Lexical Priming Theory (Hoey, 2005, 2013) focuses on the various relationships and interactions between words (or word sequences) and other linguistic elements. Lexical priming involves a positive pole (priming) and a negative or excluding one (a restriction) at various levels. There are many types of lexical priming, which may nest together to create networks of complex interactions. These are the types of priming Hoey describes (2005, p. 13): at a lexical level, where there is co-occurrence between two words in a co-text, called *collocation*; the co-occurrence between a word and a semantic field, called *semantic prosody* ('cause HARM'); the relationship between a word and a grammatical category, called *colligation*; each word occurs with specific pragmatic functions (*pragmatic associations*); each word appears with or avoids cohesive relations in discourse (*textual collocations*), semantic relations in discourse (*textual semantic associations*) and certain positions in discourse (*textual colligations*). To properly understand the work presented here, it is important to underscore that these types of priming are

restricted by domain and/or genre. It is equally important to note that words are always primed in social contexts where the speaker, the topic, and the genre are determinant. To illustrate this theory with a simple example, we could say that in the epistolary genre the word *Sincerely* usually occurs at the end of a text, sometimes collocating with the word *yours,* used to formally conclude a letter. All words give and receive priming: some very overtly, others in a less transparent manner. This theory is also echoed by the above-mentioned Extended Conceptual Metaphor Theory of Kovecses, who stated 'Importantly, priming is based on the simulation of some shared experience (the dynamically evolving common ground in a situation) that enables the production and comprehension of metaphors in discourse' (2020, p. 102).

2.2. Biber's multidimensional analysis

We will also highlight Biber's work because this author is a pioneer in using multidimensional linguistic analysis to establish the definition of genres, subgenres and text types. His work is canonical in the discipline. With this statistical technique, Biber (1988, 1995) analyzed 68 characteristics of categories and grammatical structures in a carefully designed representative corpus covering a vast array of communication situations, disciplines and text types to establish seven dimensions that enable grouping and classifying most genres and text types. Among the dimensions multidimensional analysis revealed were: involved vs informational production, narrative vs non-narrative discourse, explicit vs situation-dependent reference, overt expression of persuasion vs argumentation, and abstract vs non-abstract style. Biber's work has inspired a large amount of studies linked to language. Among others, we can cite Pennebaker's work, where forensic linguistics and the diagnosis of disorders such as depression and psychology intersect (Pennebaker, 2011; Pennebaker & Evans, 2014; Rude et al., 2004).

As mentioned previously, in the last fifty years quantitative analysis has reintroduced the empiric approach in linguistics studies.

In this paper we are going to explore the lexical characteristics of four sub-corpora on the topic of fear. To do so, we are basically going to employ three increasingly popular statistical techniques in digital humanities: specificities, hierarchical cluster analysis and multidimensional analysis. In our research, the implementation of these techniques will be carried out on lexical data in order to provide a semantic, discursive and cultural interpretation of the findings.

3. Methods and materials

In this section we provide an explanation of the statistical techniques used in this research from a linguistics point of view, as well as a description of our corpus design.

3.1. On the statistical techniques

The most basic statistical technique used in this piece of research is a specificity analysis. It is related to a procedure known in corpus linguistics as keyword extraction. This is a common procedure that compares the frequency of the same word in two corpora: an experimental corpus (the object of study) and a reference corpus (for comparative purposes) (Bondi & Scott, 2010; Scott & Tribble, 2006). If the relative frequency of one word in two corpora is statistically different, then we have identified a key word. This technique offers very important clues about the most noteworthy characteristics (topic or style) of a text. They can be systematically tracked in a corpus (through lemmatized or unlemmatized words searches with or without tags). In our case, the data come from corpora of disparate provenance and the features we analyze are mostly lexical content words (nouns, verbs, and adjectives) that appear in those texts.

The second technique used in this piece or research is hierarchical cluster analysis. Here we count the frequency with which certain words tend to appear together with other words in a given number of segments (say 40 words) and measure how these co-occurrences form patterns that in turn generate word clusters. These groups of words, which are like crystallized networks, reveal the thematic content of the texts. Thus, for example, with this type of analysis we could establish quantitatively that the vocabulary of sociology and anthropology are more similar to each other than to the vocabulary used

in political science or law; and we could also establish that the texts of these four disciplines are more similar to each other than to texts from neuroscience or genetics. Thus hierarchical cluster analysis shows how many different groups can be established in a group of texts or words and how similar or dissimilar these groups are. This type of statistical analysis is usually represented as a dendrogram.

The third technique we use in this work is multidimensional analysis. It is usually used after the hierarchical cluster analysis to establish the distance between groups according to sets of characteristics or factors. We could say that this technique adds a further level of depth that tells us how and how much the groups identified in the initial cluster analysis differ. The technique is part of multivariate statistics and establishes how two or more variables correlate with each other to varying degrees. As Almela, et al. (2022) point out, 'these statistical procedures help the researcher to summarize the data and reduce the number of variables needed to describe them, exploring the contribution of several factors to the same event' (p. 546). Through the process of factor rotation, multidimensional analysis will enable us to choose an interpretation of the topics from a wide variety of possible options. This factor analysis is viewed through the dispersion of the characteristics on axes of the dimensions or factors found. Unlike techniques based on artificial intelligence, explanatory variables of the dependent variable can also be considered.

In linguistics the main reference for the use of this statistical technique is the work of Douglas Biber. He proposed (1988) a multidimensional analysis of dozens of types of English texts based on 67 features. The statistical analysis enables comparisons and identifies the characteristic structures of texts according to their type or to their genre. Of the scarce studies done with data in Spanish, Parodi (2005) stands out as the first linguist to create an adapted replica of the Biber method.

The use of this technique in the social sciences and humanities is increasingly widespread. It has been used, for example, to establish the similarities (isoglosses) and differences that exist between dialects (see, for example, Kessler et al., 1995; Nerbonne & Heeringa, 1997; Perea & Ueda, 2010) based on the use of certain lexical features, or how closely related certain literary genres and subgenres are (for example, adventure novels with detective or fantasy novels) (Allison et al., 2011). In linguistics, studies by Fang and Cao (2015) and by Berber Sardinha and Veirano Pinto (2019, 2014) summarized the most recent advances in computer-aided text classification and disseminated work using this technique. They reviewed how to analyze varying characteristics, from their etymology to their grammar, including pronoun, adverb or adjective analysis in order to classify genres and registers. Their work combines computational methods and a sophisticated linguistic analysis. The work carried out by Stanford's Literary Lab team, headed by Franco Moretti (Allison et al., 2011) is worthy of note in literature. In this study, the group collected over one hundred 19th century novels, classifying them according to genre. These novels were segmented in ten parts, rendered anonymous and the research group performed the task of creating an algorithm that enabled regrouping (with the utmost success) the novels and assigning them to the corresponding literary genre, which had been hidden. Matthew Jockers, one of Moretti's team members, replicated the experiment a little later using 42 high frequency words (pronouns, prepositions, conjunctions, etc.) and punctuation marks. His work (Jockers, 2014, 2013) is mandatory reference in research of this nature.

In this study statistical techniques are done at a lexical level: word counts. To do multidimensional analyses, the free software used for this work was Iramuteq (Ratinaud, 2009), developed by the LERASS laboratory in the University of Toulouse. The use of this resource is widespread in social sciences for Portuguese and French and is usually used to determine topics and groups in interviews. Statistical techniques developed by Reinert (1985, 1986) are used for hierarchical classification. Its ease of use, the relative simplicity with which the statistical procedure is presented and executed, and the striking visual representations of the results make it one of the favorite tools among researchers whose statistical and computer knowledge is limited, but who value interdisciplinarity and the use of mixed research methods.

3.2. Corpus design

The corpus used in this paper was designed following the three stages suggested by Egbert et al. (2022, p.70): definition of the target domain, definition of the operational domain, and text sampling. Adjustments were made after the first unsuccessful trial.

The target domain was designed taking into consideration a large-scale project (http://cartografias.neurohumanities.info/) from which this research derives. This large-scale project aims at identifying what triggers fear at a social level and what the subsequent human reactions are. In this paper we address both questions by carrying out text analysis with tools normally used in corpus linguistics. We decided to focus on non-fictional texts, to adopt a glocal perspective (mainly Mexican texts but without precluding global concerns).

A definition of the operational domain was decided by the authors according to their field of expertise and comprise chronicle, essay, press and texts from social networks. Other factors considered were availability, time scale, and text relevance.

Table 1, Appendices 1 and 2 provide details of the text samples. As for the sample of chronicles, specialists selected four texts dealing with the topic of violence in Mexico (Bautista & Pérez, 2022). The essay section was collected following criteria related to relevance, coverage of several subdisciplines (history, philosophy, sociology, and anthropology, etc.) and relative recency. This section included translated texts widely available to readers of Spanish as we consider them an essential contribution in terms of concept generation, and they also help to capture collective cultural zeitgeist. Although in linguistics the analysis of translated texts together with texts written originally in a target language is uncommon, in our case we could not find enough reasons to discard essays not originally written in Spanish. We understand that this might pose a problem due to transfer issues (the concept of fear can be translated as *miedo* 'fear', *temor* 'apprehension', or *angustia* 'anguish' according to the context), but this research is not meant to describe linguistic units, metaphors to express fear, or grammatical structures but is rather an investigation on how language reflects variation in cultural and social contexts. The press section comprises news from four Mexican newspapers (El Universal, La Jornada, Excélsior, and El Heraldo) and was selected following criteria of availability, a timescale (2019–2023) and word seeds (*miedo* 'fear', *temor* 'apprehension', *desconfianza* 'mistrust', *fobia* 'phobia', *estrés* 'stress', *ansiedad* 'anxiety', *pavor* 'fright', *horror* 'horror', and *terror* 'terror').

Finally, the social networks category contains tweets collected throughout 2023 using the #miedo 'fear' hashtag.

Table 1. Corpus description by sections

Section	Original no. of words	Dates range	Selection criteria	Source	Final no. of words
Chronicles	296,783	2014–2022	Dates and academic relevance	Mexico	296,783
Essays	1,363,823	1978–2022	Dates and academic relevance	Global	653,713
News	9,731,856	2019–2023	Seeds and dates	Mexico	338,828
Tweets	1,709,471	2023	Seed (#miedo 'fear') and dates	Global	635,170
Total	13,101,933				1,921,457

Table 2. Top 10 most specific words in the corpus.

Chronicle	Essay	Press	Tweets
policía 'police'	siglo 'century'	nacional 'national'	tener 'have'
mujer woman'	modo 'way'	stress 'stress'	mío 'mine'
él 'he'	social 'social'	temor 'fear/aprehension'	yo 'I'
desaparecer "disappear'	forma 'forma'	mexicano 'Mexican'	dar 'give'
cuerpo 'body'	sociedad 'society'	covid 'COVID'	miedo 'fear'
hijo 'son'	riesgo 'risk'	'risk'gobierno 'government'	querer 'want'
casa 'home'	peste 'plague'	presidente 'president'	tú 'you'
joven 'young'	peligro 'danger'	pandemia 'pandemia'	derogar 'abolish'
feminista 'feminist'	ciencia 'science'	ansiedad 'anxiety'	puta 'whore'
violencia 'violence'	judío 'Jew'	salud 'health'	inseguridades 'insecurities'

As for the third stage, the evaluation process, the processing demands of the software and of the original corpus on our hardware could not be initially met. We therefore decided to reduce the corpus size. We segmented the sections of the corpus into sentences using the sent-tokenize method from the NLTK platform (Bird et al., 2009). The first 40,000 sentences from each category were then extracted. As a result, the first ten essays of the list in appendix 1 were included in the final sample (and a sample of the main subdisciplines represented), and a subsample of the press and social networks sub-corpora was selected. The size of the analyzed corpus is in the last column of Table 1.

The corpus is available for research purposes at the following DOI address: https://doi.org/10.6084/m9.figshare.25744122

4. Analysis and results

In order to identify the discursive elements most relevant to fear, three representation stages of the corpus data were created with the aforementioned software: (1) specificity extraction, (2) hierarchical classification, and (3) similarity analysis. These three analyses reveal a gradation that ranges from the most general findings to the most concrete and detailed. What we provide in this section is an interpretation of the meaning and connotations provided by contexts where the words tend to occur. Thus, we read many examples (concordance lines or words in context) of the terms 'flagged' by the statistical techniques (or its visual representations). Then we select one prototypical example to illustrate how words tend to co-occur and point out to lexical, semantic and pragmatic associations. It is important to stress that the interpretations that we provide are always corpus-guided (not corpus-based), and we try not to fit the data results into a preexisting theory. In fact, this is the point of this paper and, in general, of digital humanities as we understand them: we are not trying to provide examples that support preexisting ideas, but to process a huge amount of text that can trigger new and sometimes unforeseen interpretations from a humanities perspective.

4.1. Corpus specificities

Specificities extraction consists of identifying the most relevant words as regards each of the four genres/domains that constitute the corpus (chronicle, essay, press, and social networks). These specificities are like key words. The general process consists of dividing the corpus texts into various groups whose words are compared taking the entire corpus into account. Based on hypergeometric distribution (Boucher, 2011), the software identifies which words best represent each variable. The words were divided into two categories: in one there were words with lexical content (nominal, verbal, pronominal and adjectival forms) and in the other function words (determiners, prepositions, adverbs, numerals, and unrecognized forms). Additionally, we chose to lemmatize the entire corpus. The first ten most specific words considering the four corpus categories are shown in Table 2 (proper names were eliminated and translation provided).

Even if this set of words is small, it still gives an idea of the contents of each sub-corpus with words meaning quite different things. The list of the 100 most specific words of each genre/domain is in Appendix 3. Interestingly, not all words retrieved with this technique are related to the topic in a straightforward way. For example, the word *piraña* 'piranha' (specific in the chronicle subcorpus) refers not to a fish species but to a criminal's nickname.

In the chronicle category words indicate contents related to violence and women. The sub-corpus's theme could be summarized by creating a sentence with those ten words. Words such as *cuerpo* ('body') appear, referring to the human corpse; the words *hijo* ('son/child') and *desaparecer* (to disappear) occur as well, since several of these texts include the testimony of women searching for their missing children. The extraction of these words from the chronicle subcorpus enabled us to analyze around thirty sentences where the word *cuerpo* ('body') cooccurs with either the word *hijo* ('son/daughter') or the verb *desaparecer* ('disappear'). Here is just one example:

(1) ¿cómo iba a ser capaz de escribir sobre la necesidad de ir en pos del cuerpo de un ser querido desaparecido si…) (D. Rea, 2020)

['How was I going to be able to write about the need to search for the body of a loved one who had disappeared if…']

Other important specific words of this sub-corpus are *casa* 'home' and *policía* 'police'. *Casa* is relevant because it represents a safe environment which is violated when someone disappears or has been kidnapped (*levantado*). It is strongly associated with family within a traditional framework. *Policía* is important because of the uncertainties that it triggers in Mexico: they are widely considered incompetent, corrupt and untrustworthy.

As for the essay category, those texts are filled with an analysis of society. The weight of historical texts seems evident as there are examples of how fear has been generated by illnesses (*peste* 'plague') or marginalized communities (*judío* 'Jew').

(2) Se quemaron judíos en 1348 en Stuttgart, donde la peste no apareció sino en 1350 (Delumeau, 1978)
[In 1348, Jews were burnt in Stuttgart, where the plague was declared only in 1350]

Moreover, the word *siglo* 'century' signals the historical dimension of a subsection of this subcorpus. *Peligro* and *riesgo* 'danger' and 'risk' are two words that conceptualize aspects related to 'fear'.

The press category clearly shows that the texts were extracted during the COVID-19 pandemic since their content mention country authorities (*gobierno, presidente mexicano, nacional*; 'government', 'Mexican president', and 'national') and health (*stress*, *covid*, *temor*, *ansiedad*, *pandemia*, and *salud*; 'stress', 'COVID', 'apprehension', 'anxiety', 'pandemic', and 'health').

(3) Ante el miedo a un contagio de covid 19 en transporte público los capitalinos están prefiriendo usar su automóvil particular para movilizarse en la ciudad (Excélsior, 5/jun/2021)
[Fearing COVID-19 contagion on public transport, residents of the capital are preferring to use their private car to get around the city.]

Lastly, in the social network texts the word *miedo* 'fear' logically appears as the texts were obtained through the #fear hashtag. Pronouns such as *mío*, *yo*, *tú* 'mine', 'I', 'you' show that in this domain appealing to interlocutors and expressing personal feelings is common.

(4) Contigo puedo ser yo sin miedo a q me juzgues
[With you I can be me without fear of u judging me]

In this example, pronouns, morphemes, and combined forms provide five different references to the first and second persons.

A close look at just the top ten specific words of each sub-corpus, when compared to each other, points to issues of subjectivity, social relationships and the presence of concepts related to fear. For instance, worthy of note is the contrast between the presence of the first and second person pronouns *yo* 'I' and *tú* 'you' in the Twitter sub-corpus with the third person *él* 'he' personal pronoun of the chronicle sub-corpus. This signals the nature of each text type: the social network serves the purpose of venting personal feelings and opinions whereas the chronicle is intended to describe a social situation in a more objective fashion. Moreover, the chronicle is the only section where words referring to particular individuals are so relevant (*hijo* 'son', *joven* 'youngster', *mujer* woman' or 'wife') as opposed to individuals representing institutions (*policía* 'policeman', *presidente* 'president', or *gobierno* 'government') in the press sub-corpus. As for the concepts associated to fear, some of the top ten key words of the essay sub-corpus seem to refer to causes of fear (*riesgo* 'risk' and *peligro* 'danger') whereas in the News and the Twitter subsections the concepts seem to be associated to the effects of fear (*estrés* 'stress' and *ansiedad* 'anxiety' in the press sub-corpus and *inseguridades* 'insecurities' in the Twitter sub-corpus).

A larger set of specific words can be found in Appendix 3. They are basically the data used to conduct the hierarchical classification described in the next section.

4.2. Hierarchical classification and cluster analysis

This section includes the visual representations derived from a hierarchical classification and a multidimensional analysis: a dendrogram and a word map.

The classification enables the bundling of words to indicate which group is closer to another. From there the distance between groups can be inferred in a general manner. The statistical techniques involved in this analysis are different from those of the specificities seen in the previous section. Here the calculations depend on the concept of word co-occurrence whereas in the previous one the calculations depended on the idea of contrasting frequencies. The robustness of the tests, however, show that the grouping provided by both methods are very similar.

As shown in Figure 1, the classification method (clustering analysis) produced six classes, that is, all words are grouped within one of the six clusters. The size of the box contains the percentage of the content words within each class (e.g. class 1 is made up of 26% of the corpus words, etc.)

We will then read these classes from left to right. In general, each class corresponds to each of the four sections of the corpus: class 1 corresponds to the Twitter sub-corpus, the central sections comprising classes 5, 4, and 6, correspond to the Essay sub-corpus, and the last two sections correspond to the News sub-corpus and the Chronicle sub-corpus.

Class 1, in red, is separate from the other categories according to the dendrogram. The reasons why this might have happened are related to the nature of the communication taking place: electronic, instant messaging, non-face to face interactions, anonymity and dialogic form. It is the class with the most verbs. There are high frequency verbs (*ver* 'see', *dar*, 'give'), volition or appraisal verbs (*importar* 'care', *querer* 'want'), misspelled words (*elijen* 'they choose'), interjections, insults, and some adverbs (*solo* 'only', *nunca* 'never') that may imply quite bold statements. It also includes terms typical of Twitter (*rt* means retweet, *https* for a web source). An example of a very popular retweeted message is:

(5) como no voy a tener inseguridades y miedo al abandono si siempre me ha tocado ver como eligen a otras personas antes que a mí
[How can I not have insecurities and fear of abandonment if I have always seen how other people are chosen over me?]

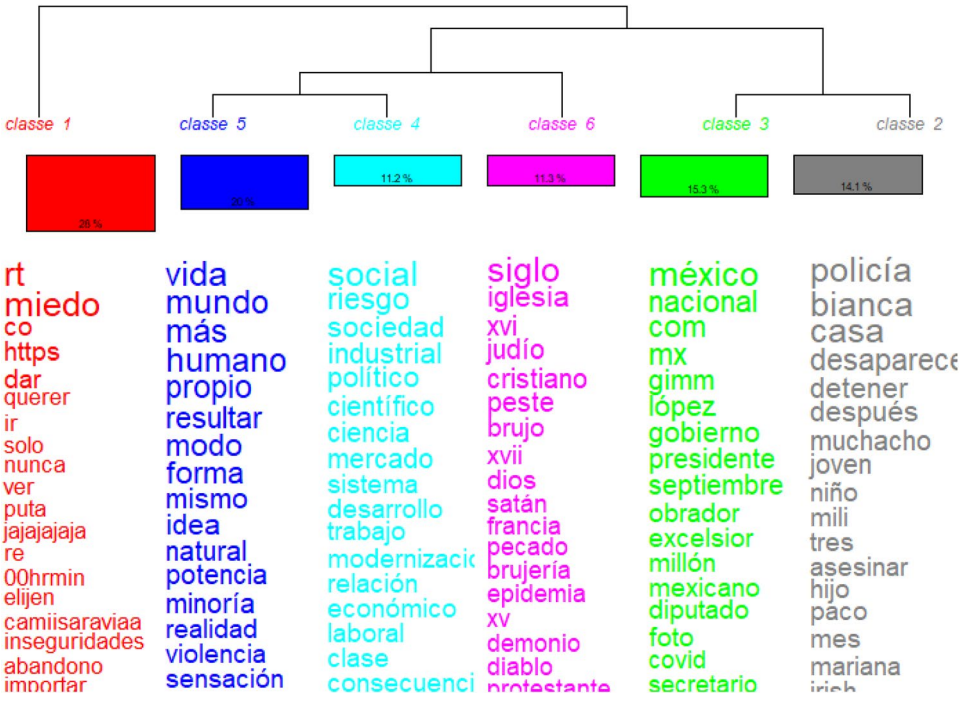

Figure 1. Dendrogram of the classes (groups or clusters) produced from the corpus.

A second cluster is generated with the classes from categories 4, 5 and 6 (presented in light blue, royal blue and pink, respectively). These words seem to come from the essays section of the corpus. The royal blue column seems to correspond to philosophical topics (*idea* 'idea', *potencia* 'potency', *vida* 'life', *realidad* 'reality', and *sensación* 'sensation') as in the following lines.

(1) El miedo nos roba el acto y la potencia al transformar la magnanimidad en apocamiento. El miedo al fracaso, a la impotencia o al ridículo suele disfrazarse de una modestia excesiva. (Castany, 2020)
[Fear robs us of action and potency by transforming magnanimity into timidity. The fear of failure, impotence or ridicule is often disguised as excessive modesty.]

Many words are abstract shell nouns that serve the purpose of encapsulating complex concepts and ideas in order to discuss them (*sensación* 'sensation', *idea*, *modo* 'way', *forma* 'form'). There is only one verb (*resultar* 'to result') which indicates causation, an example of the analytical nature of the genre.

The light blue column seems to identify a socio-economic and political analysis (*social* 'social', *riesgo* 'risk', *ciencia* 'science', *mercado* 'market', *trabajo* 'work', *modernización* 'modernization', *económico* 'economic', and *laboral* 'labor'). The following sentence can be considered prototypical of this:

(2) Las inseguridades generan patologías del vínculo social y, a la inversa, la erosión de la sociabilidad cotidiana acentúa el miedo al otro. No es casual el reciente interés por la confianza. En efecto, ¿qué queda cuando se desvanecen los grandes relatos, las identidades nacionales, las tradiciones consagradas, los paisajes de la infancia? (Villa 2002)
[Insecurities generate social bond pathologies and, conversely, the erosion of everyday sociability accentuates fear of the other. The recent interest in trust is no coincidence. Indeed, what is left when great stories, national identities, consecrated traditions, and childhood landscapes fade?]

No verbs can be found among the dendrogram's first words of class 4.

The pink column seems to correspond to historical essays, of which two are by French authors; that is why the word *Francia* 'France' is present. Most words are linked to what triggered fear from a historical perspective. There are words that are mainly linked to religion (*iglesia* 'church', *cristiano* 'Christian', *satán* 'Satan', *dios* 'God', *pecado* 'sin', *demonio* 'demon', *diablo* 'devil'), diseases, ethnic, and 'heretical' groups (*peste* 'plague', *judíos* 'Jews', *protestantes* 'Protestants', and *brujos* 'witches'), and the centuries analyzed in those texts (XV, XVI, and XVI).

(3) En el ámbito católico, la secesión protestante no hará sino llevar a su paroxismo el miedo de la subversión de la fe, ya muy viva antes, y la tendencia a integrar en el universo de la herejía todas las categorías de sospechosos. En efecto, podemos comprobar que judíos, musulmanes e idólatras domiciliados en territorios que dependen de un príncipe cristiano, se vieron progresivamente asimilados a herejes y, por tanto, punibles como tales. (Delumeau, 1978)
[In the Catholic sphere, Protestant secession would only bring to its paroxysm the fear of the subversion of the faith, already very alive before, and the tendency to integrate all categories of suspects into the universe of heresy. Indeed, we can verify that Jews, Muslims and idolaters domiciled in territories that depended on a Christian prince were progressively assimilated to heretics and, therefore, punishable as such.]

Class 2 is green and seems to correspond to Mexican violence chronicle texts. The following excerpt is an example of the context in which fear is triggered within this genre.

(4) Era la primera vez que hablaban en público; antes tuvieron que tragarse los secuestros, las extorsiones, los asesinatos, por el miedo a los delincuentes y a las autoridades. Todos lloramos esa noche. Hasta los conductores de los autobuses contaron las historias de terror vividas en las carreteras tomadas por criminales. Los decapitados en la Autopista del Sol, los retenes clandestinos, las extorsiones a transportistas, los 'halcones' en las carreteras, el asesinato de un colega. (Turati & Rea 2014)
[It was the first time they spoke in public; before, they had to avoid mentioning the kidnappings, the extortions, the murders, out of fear of criminals and the authorities. We all cried that night. Even the bus drivers

told the horror stories they experienced on the roads taken over by criminals. The beheadings on the Autopista del Sol, the clandestine checkpoints, the extortion of transporters, the 'hawks' on the roads, the murder of a colleague.]

Class 2 is grouped with class 3 (in gray). This class seems to correspond to press texts, marking greater similarity between these two genres. The press sub-corpus refers to names of politicians (López Obrador, Mexican President 2018–2024), authority positions, (secretario 'secretary', diputado 'congressman', and presidente 'president'). The word *millón* 'million' seems to refer to news about economy. While the word *covid* refers to a global and local issue, the words referring to Mexico and Mexicans point to a local dimension.

The results of the clustering technique are in a certain sense to be expected as they are the result of the corpus design. What is interesting to note is that, despite the fact that the word *miedo* 'fear' is a lexical and thematic constant, the texts point to differentiated referents, which may imply the existence of particular semantic fields where the notion of fear is taking place and the aforementioned idea that each word and concept are experience-dependent.

These six classes are represented on a Cartesian plane in Figure 2. Compared to the dendrogram, this scatterplot enables us to see more precisely which words are associated with each class and how distant they are from each other. Thus, taking the dimensions obtained from the multidimensional analysis into consideration, we see classes 4, 5, and 6 are in the upper right quadrant, which correspond to the essays. The words in pink, corresponding to history, are among the more philosophical and more political-economic texts (light blue). The words in gray and green, corresponding to the chronicle and press texts, respectively, are in the lower right quadrant.

The diagram is drawn from the largest dimensions that the factor analysis (out of the total of five dimensions found) allows to identify. The horizontal axis is determined by factor 1 and the vertical axis by factor 2. Our interpretation of dimension 1 points to the 'affectivity-objectivity' opposition. The words on the left of the diagram (for example, *inseguridades* 'insecurities', *infidelidad* 'unfaithfulness', *calmarte* 'relax', *abandon* 'abandonment', *deslealtad* 'disloyalty', *llanto* 'crying', *amar* 'to love, *valiente* 'brave', *odiar* 'hate', *amor* 'love', and *angustia*, 'anguish') represent emotions. This would explain why this area is filled with words from class 1 (in red), which mainly correspond to the tweets. Example (5) below shows how these concepts are used in representative sentences from the corpus. The positive pole of this axis corresponds to the opposite of affectivity, to what is objective, with words that refer to concepts such as *cientificación* 'scientification', *racionalidad* 'rationality', *interpretación* 'interpretation', *consecuencia* 'consequence', *ámbito* 'area' y *técnico* 'technical' (light blue, above), and *independencia* 'independence', *contagios* 'contagions', *encabezar,* 'to head', *ejecutivo*, and 'executive' (in green, below).

Our interpretation of dimension 2 (vertical axis) leads us to indicate the opposition between 'concrete entities—abstract entities.' Thus the lower part has what is concrete: proper names (*Andrés Manuel López Obrador* –México's president–, *Claudia Sheinbaum* –México's then presidential candidate–), institutional positions (*mandatario* 'leader', *entidad* 'entity', *secretario* 'secretary'), places (*Chiapas*, *Oaxaca*, *cdmx*) or names of political parties, newspapers or press agencies (in green). In the upper part (light blue and pink words, corresponding to social science and history texts), there are words that suggest general, global abstract concepts such as *riqueza* 'wealth', *progreso* 'progress', *individualización* 'individualization', *contradicción* 'contradiction', *capitalism* 'capitalism', *actuación* 'acting', *teórico* 'theoretical', *fase* 'stage', *modernización* 'modernization', and *herejía* 'heresy'.

We have underlined seven words associated with moods, emotions, feelings or actions related to fear. To the left, as terms that belong to class 1 (in red, social networks), the statistical analysis highlights the terms *inseguridades* 'uncertainties', *angustia* 'anguish', *abandono* 'abandonment' in addition to *miedo* 'fear'. This finding seems to reinforce those of Cebral-Loureda et al. (2021) regarding the centrality of the word and emotion 'fear' in Tweets regarding the COVID-19 crisis. These words –together with *amor* 'love' and *odio* 'hate'– seem to suggest that fear and failed interpersonal relationships are linked.

Here is another example from the social network sub corpus:

(5) rt @isad_off: me da miedo demostrar mucho amor y que me fallen y quedarme con todo el amor en las manos otra vez

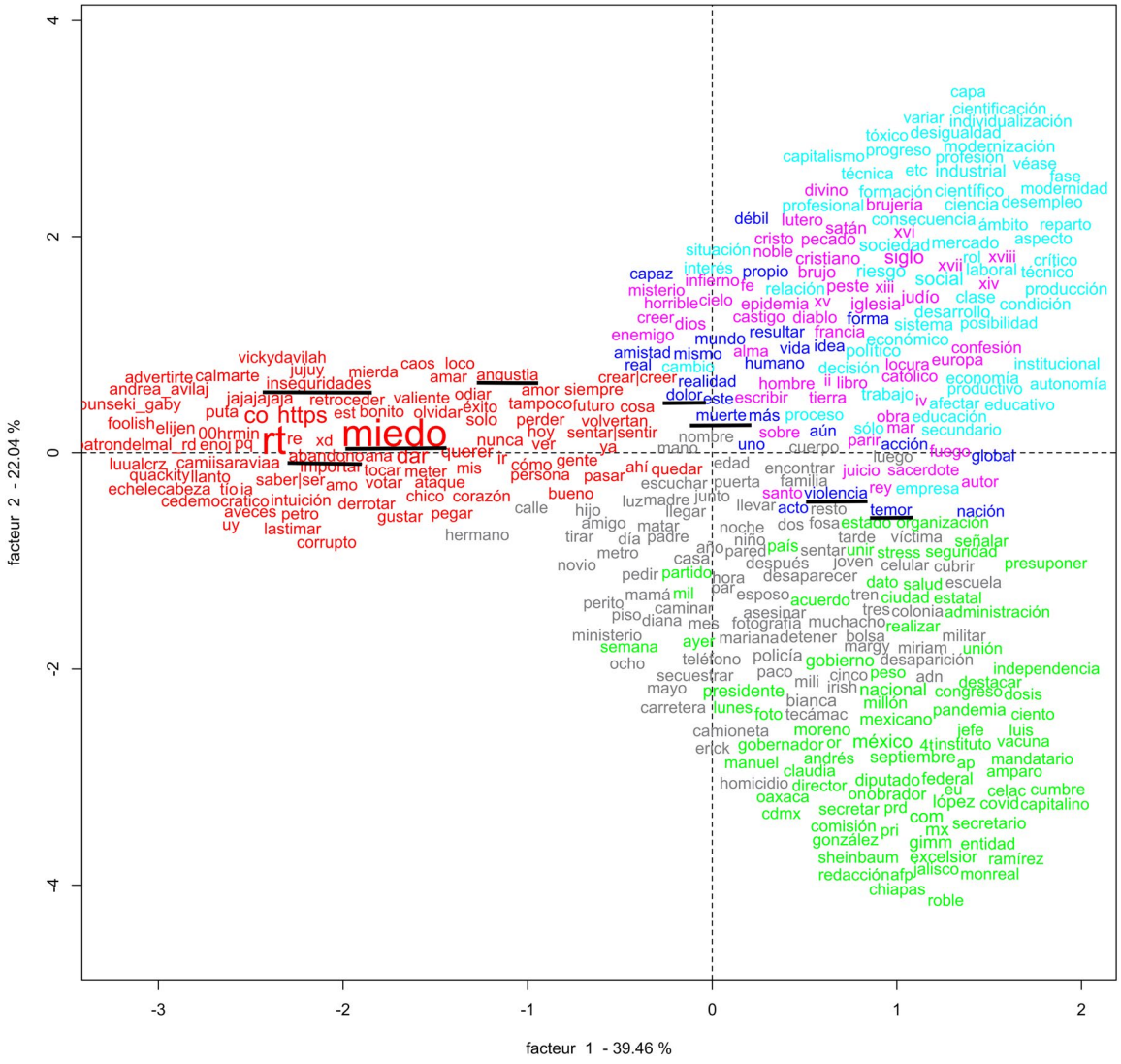

Figure 2. The six classes represented in the diagram according to the first two factors.

[I'm afraid of showing a lot of love and being left with all my love in my hands again]

A qualitative analysis reveals that fear is neither the cause nor the effect of said emotions, but rather that it builds up from these relationships. Most of the time, the content of our social network sub-corpus shows that fear is mostly associated with unrequited love.

The underlined class 5 words (in royal blue) (*dolor* 'pain', *muerte* 'death', violencia 'violence', and *temor* 'apprehension') correspond to philosophical texts. The first three words seem to be situations that trigger fear; two of them are very close to the intersection of the axes, while *temor*, 'apprehension', a word that in Spanish reduces the intensity of fear (called by Kövecses, [1990, p. 80] 'dispositional fear'), is a little further away, to the right.

Several of these words appear in example (11):

(6) podríamos afirmar que, en cierto sentido, el miedo a el hambre y a la sed, o incluso a la ausencia de confort, es un miedo a la muerte. Es muy posible que cierto temor indefinido a la muerte sea una característica innata en nosotros, una ventaja evolutiva. (Nussbaum)

[we could say that, in a way, the fear of hunger and thirst, or even the absence of comfort, is a fear of death. It is very likely that some indefinite fear of death is an innate characteristic in us, an evolutionary advantage.]

4.3. Similarity analysis

After seeing the entire corpus represented in clusters, the last analysis involves a similarity representation in the form of graphs. This type of analysis can identify how fear tends to build up and diversify. It is based on measures from graph theory, which provide evidence of how we construct our communication by focusing on the closeness with which certain words appear in semantic associations, colligations and collocations. A list of seeds representing this emotion at the lexical level was selected for this purpose. The words that make up this list were obtained from the word 'fear' and its expansion onto related emotions (*ansiedad* 'anxiety', *desconfianza* 'mistrust', *fobia*, 'phobia', *horror* 'horror', *pavor* 'fright', *recelo* 'wariness', *estrés* 'stress', *temor* 'apprehension', and *terror* 'terror'), for which information from the FrameNet (Baker et al., 1998) lexical base was used for the purpose of controlling and directing the expansion. Then the relation of these seeds with the 100 most specific words (Appendix 3) obtained for each category (Section 4.1) via a specificities analysis was observed.

In the following subsections, we are presenting five graphs drawn out of the relations established by the seed list. The first one, Figure 3, links this seed list with the entire corpus. The other four show the relations within each sub-corpus independently.

4.3.1. Entire corpus similarity analysis

Nodes have been grouped by color and in halos. The size of the node reflects each word's frequency. In Figure 3 the word *miedo* is represented in the largest (green) node in the center. Some verbal colligations (*dar* 'give', *perder* 'lose', and *eliminar* 'eliminate') appear within this same group. To the right it overlaps with another group that contains the *tener* 'have' and the *ver* 'see'colligations as well as personal pronouns (*yo* 'I', *tú* 'you', and *él* 'he'), and related concepts such as *cuidar* 'take care of', *inseguridades* 'insecurities', and *abandono* 'abandonment'. This is a group related to class 1, the twitter sub-section of the corpus.

Under the *miedo* node there are two groups that seem to have been created from the essays. To one side there is a blue halo, with words such as *sociedad* 'society', *relación* 'relation', *ciencia* 'science', *clase* 'class', and *moderno* 'modern'. From this group stems a second group (the pink halo) referring to history: *época* 'time period', *peste* 'plague', *brujo* 'sorcerer/witch', *ciudad* 'city', *entidad* 'entity', *judío* 'jew', and *cristiano* 'Christian'. To the left of the central *miedo* node there is a beige-colored halo with family relationships (*madre* 'mother' and *hijo* 'child'), and verbs referring to endured violence (*asesinar* 'to kill', *desaparecer* 'to disappear', *buscar* 'to search', *encontrar* 'to find', and *cuerpo* 'body'). From this group, which seems to correspond to the chronicle sub-corpus, there are three others that emerge at a short distance. On one side, in a dark green halo (dark blue nodes) there is the word *joven* 'young/youth' along with other words associated with legality: *policía* 'police', *detener* 'to arrest', *justicia* 'justice', and *procuraduría* 'attorney general's office'; on another side there is a light green halo, under the group just mentioned, where there are terms such as *violencia* 'violence', *víctima* 'victim', *feminicidio* 'femicide', *desaparición* 'disappearance', *terror* 'terror', *horror*, and *poder* 'power'. The third group linked to the beige-colored group has a pink halo and light blue nodes. The words in this group are government related, such as *gobierno* 'government', *diputado* 'congressman', *gobernadora* 'governor', and *presidente* 'president'. Lastly, in the upper section of the graph there are two adjacent groups: the group encased in a violet halo has COVID-19 pandemic-related vocabulary, and just above it there is a light blue halo with terms synonymous to *miedo* 'fear' (*recelo* 'wariness', *desconfianza* 'mistrust', *fobia* 'phobia', *pavor* 'dread', and *temor* 'apprehension').

Thus a global study of this graph would indicate that fear is a historically and socially analyzable phenomenon that is linked to violence, disorders, emotions, and illnesses; it has several facets (almost synonyms), and lastly it is worthy of note that it is also linked to justice and politics. There are some words where clusters seem to pivot. For instance, the words *mujer* 'woman' and *joven* 'youngster' seem to be at the intersection of four clusters on the left of the graph. Stemming from these two nodes we can see verbs that point to the disappearance of individuals (towards the central, beige cluster with light blue nodes) and to violence (light green cluster, yellow nodes). Another pivotal node is *hijo* (light blue node at the center), connecting the words *madre* 'mother' (blue node) within the same brown cluster with words related to the disappearance of individuals; the second word is *dar* 'give' (green node), related to the compound *dar miedo* 'frighten' (yellow cluster), and the third one is *año* 'year' (grey node), part of

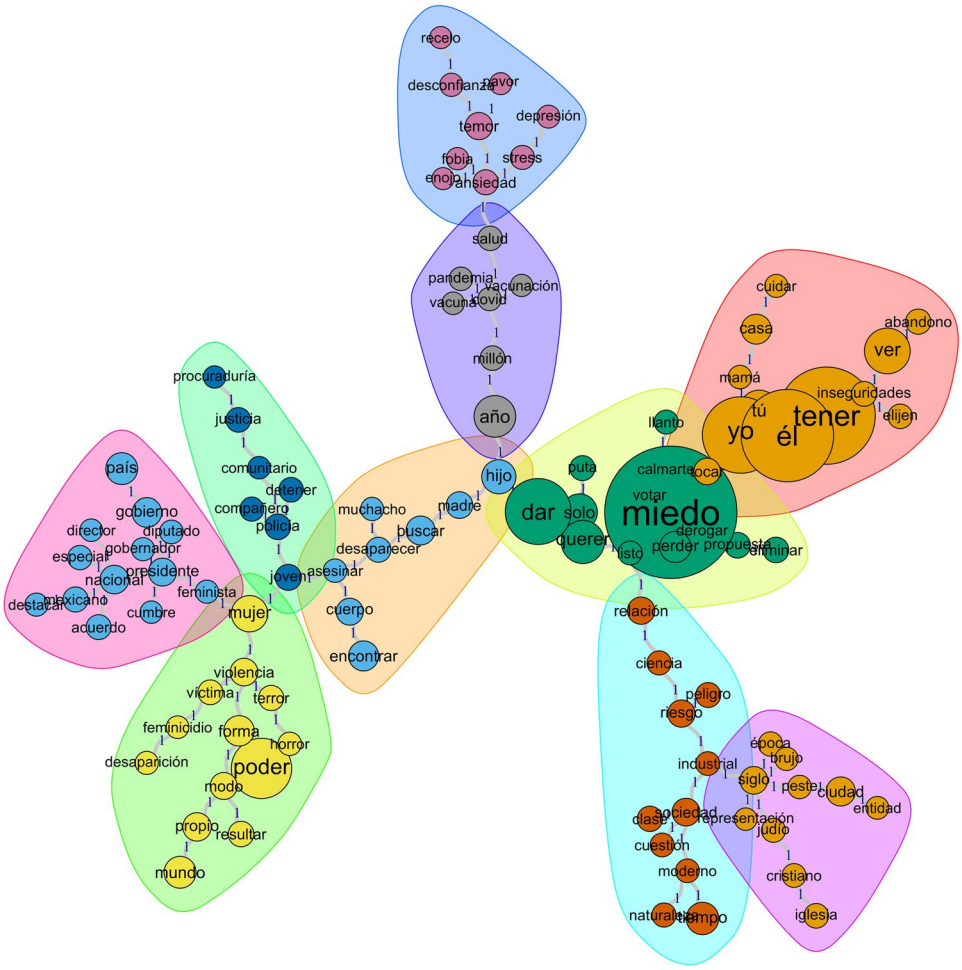

Figure 3. Key word radial representation graph in the fear corpus Similarity analysis considering the seeds for fear and the (a) chronicle, (b) essay, (c) press, and (d) tweets categories.

the purple cluster that introduces a topic related to health (COVID-19 pandemic, vaccine and vaccination). The seed words seem to be isolated at the top of the graph.

4.3.2. Chronicle similarity analysis

Chronicle is a hybrid non-fictional genre, somewhere between journalism and literature. In the five texts that make up the sub-corpus (Appendix 2), the authors report how they investigated femicides, murder cases and how relatives of missing persons set out to search on their own. They recount the testimony of the victims' families and the pain they have endured for their personal loss and the lack of support from the state and the government. These circumstances are contextualized in the so-called 'War against drug cartels' begun during Calderón's government (2009), a process that up to 2021 caused the death of approximately 350,000 people and the disappearance of 72,000 in Mexico (Pardo Veiras & Arredondo, 2021).

The two main clusters in Figure 4 are the red one (with *decir* 'say' as main node) and the purple one (with *mujer* 'woman' as the main node). The centrality of *decir* 'to say' and its size (frequency of use) probably implies the relevance of reported speech in this sub-genre because authors conveyed what the interviewees told them. This red cluster is the largest and it contains very diverse words: feelings (*temor*, 'apprehension', *miedo* 'fear', and *terror*), verbs related to violence (*levantar* 'kidnap' *secuestrar*, 'kidnap', and *matar* 'kill'), nouns referring to state institutions (*procuraduría* 'attorney general's office', *ministerio* 'public prosecution', *perito* 'expert') The blue cluster—around *mujer* 'woman'—comprises words regarding violence: crimen 'crime', asesinato 'murder', and feminicidio(s) 'femicide(s)'. Between the red and blue clusters, the green cluster contains the most words related to family ties: the main node is *hijo* 'son' and we can also read *esposo* 'husband', *padre* 'father', *madre* 'mother', *familia* 'family', and *familiar* 'relative'. Under the main red cluster a small pink cluster is comprised of the noun *casa* 'house/home' and the verbs *salir*

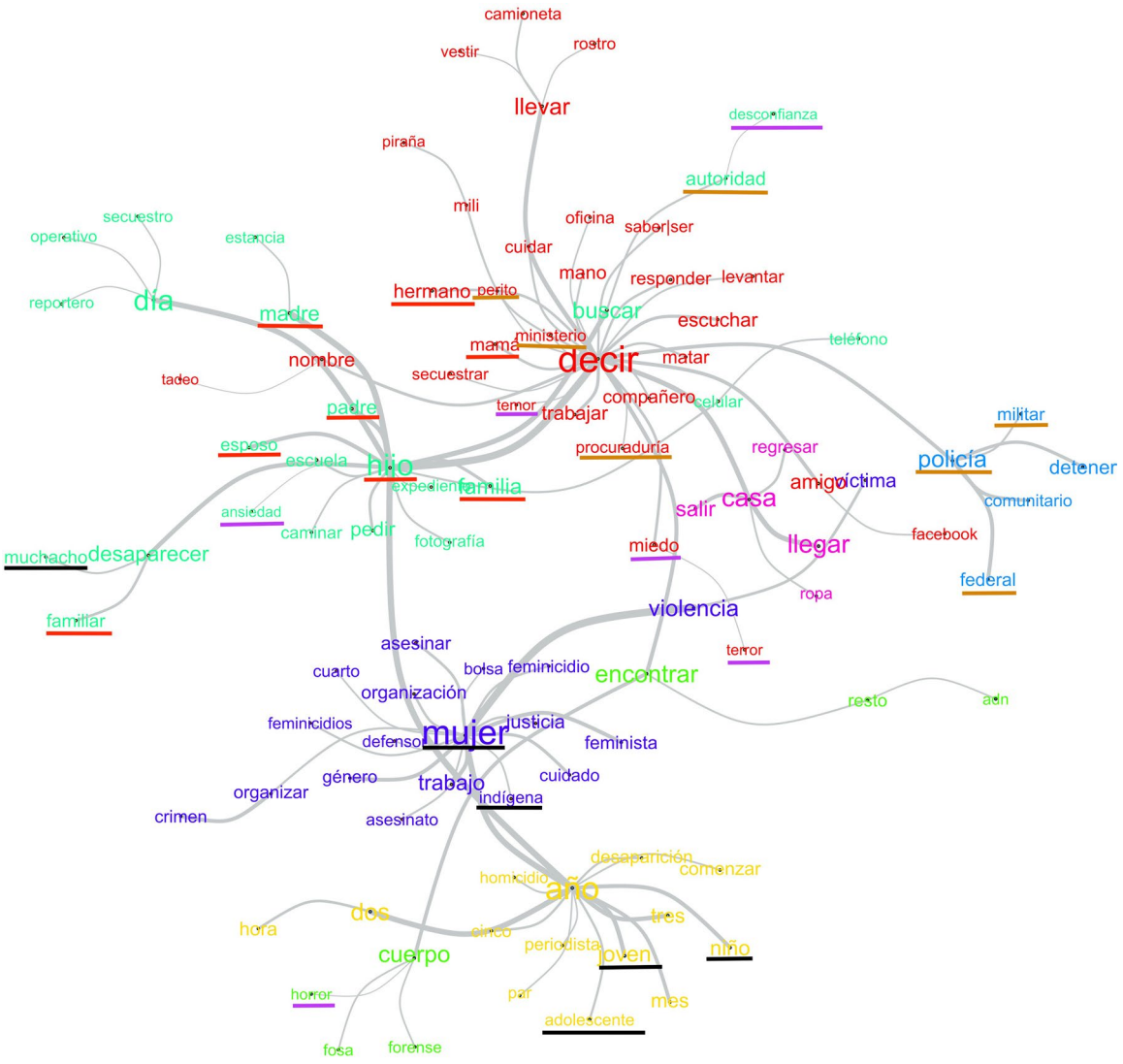

Figure 4. Key word radial representation graph in the chronicle sub-corpus.

'leave', *llegar* 'arrive', and *regresar* 'go back', representing a safety counterpoint to violence. At the bottom of the graph a yellow cluster identifies victims of violence (underlined in black) marked by their young age (*niño* 'child', *joven* 'youngster', and *adolescente* 'teenager'). Close to them, in the blue cluster, word referring to victims according to gender (*mujer* 'woman') or ethnicity (*indígena* 'indigenous').

In this graph there is no perfect match between clusters and categories: there are several cases of words belonging to the same semantic category but placed in different clusters. For instance, horror is a feeling (underlined in purple) that does not appear in the red cluster but is part of a light green cluster at the bottom of the graph, and *ansiedad* 'anxiety' and *desconfianza* 'mistrust' are also feelings but are part of the top green cluster; *hermano* 'brother and *mamá* 'mum' are part of the family category (also underlined in red) but rather than being part of the green cluster, they are in the red one. *Secuestro* 'kidnap' is in the green cluster but *secuestrar* 'to kidnap' and *levantar* 'to kidnap' are in the red one. *Desaparición* 'disappearance' is in the yellow cluster but *desaparecer* 'to disappear' is in the green one. We do not know the reasons behind these apparent inaccuracies: we can speculate that this might be due to the fact that this sub-corpus is not large enough to feed the matrices involved in the calculation.

From the words referring to state institutions (underlined in brown), it is evinced that these institutions (*militar* 'soldier', *policía* 'police, procuraduría 'attorney general's office', and ministerio 'public prosecution') do not mitigate, but rather increase the victims' feelings of fear; that is why words such as *autoridad* 'authority' and *desconfianza* 'mistrust', also occur together (top right green node). The following corpus example shows the association between both concepts:

A lo largo de su vida había aprendido que el Estado nació para protegerla, para darle derechos, para vivir en convivencia. Sin miedo. Pero todo era una farsa. Bajo el cobijo de la bandera fue herida, torturada, violada (Rea, 2016)

[Throughout her life she had learned that the State was created to protect her, to give her rights, to live in coexistence. Without fear. But it was all a farce. Under the shelter of the flag she was wounded, tortured, raped.]

An overall interpretation of Figure 4 suggests that in Mexico fear is strongly associated to violence against women and the loved ones rather than against oneself. In the global south family rather than the state is the institution providing care, support and assistance.

4.3.3. Press similarity analysis

The radial graph of fear-associated words in the press sub-corpus (Figure 5) has one central cluster (royal blue with the word *nacional* 'national' at the center). Within this node we find words related to information regarding events, data, agreements, and matters. Below this cluster, two lines branch out: on the bottom left we have firstly (yellow cluster) words related mainly to Mexico's legislative power commissions, political parties, legislators: *presidente* 'president', *diputado* 'congressman', *senado* 'senate', *candidato* 'candidate' and, interestingly, the word *corrupción* 'corruption'. Further down (green cluster) there are concrete instances of power: governments, authorities, communities, mayors' offices, and municipalities. The

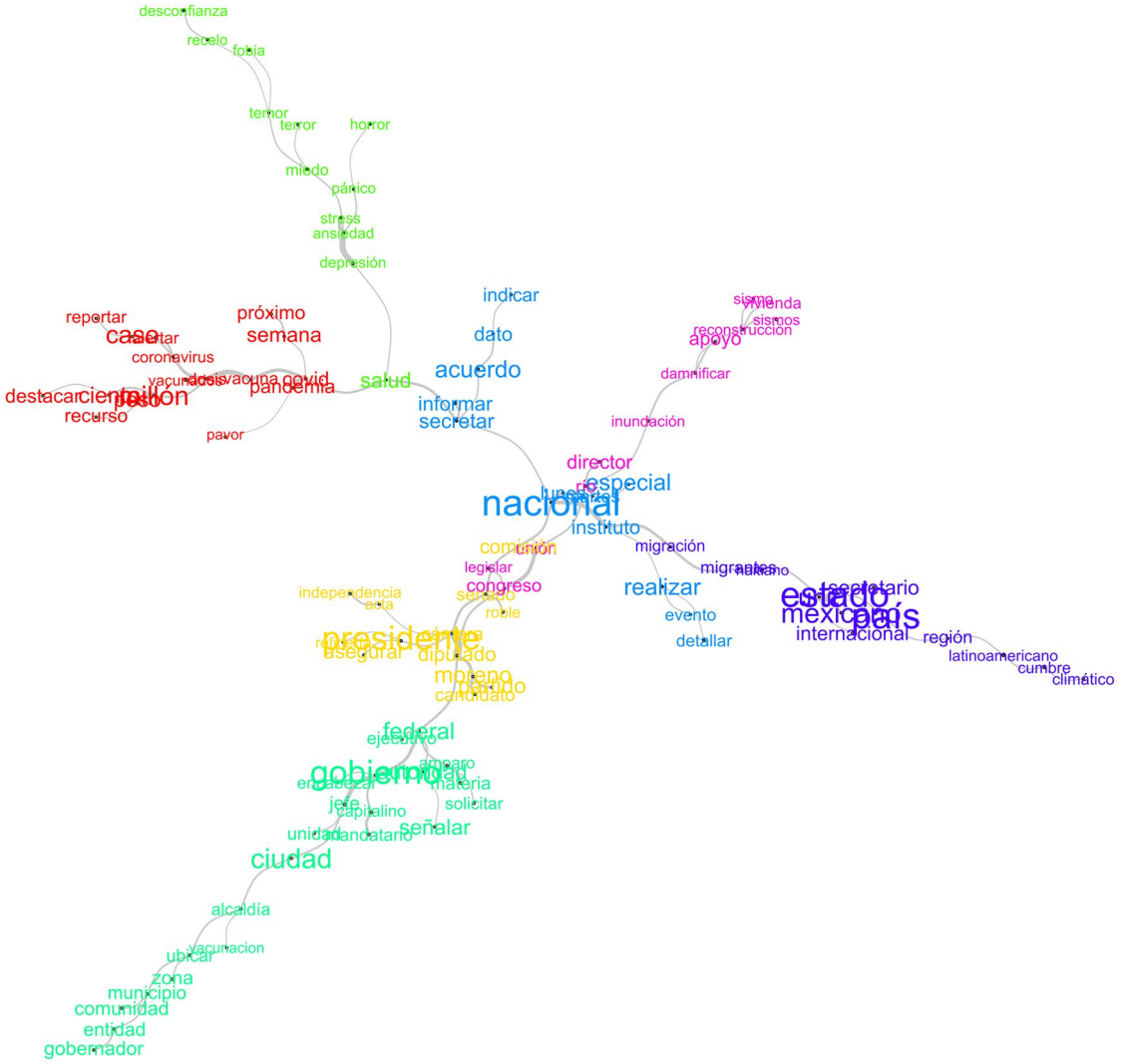

Figure 5. Key word radial representation graph in the press sub-corpus.

appearance of these thematic areas seems to indicate that 'fear itself has been politicized to a point where the debate is no longer about whether or not we should be afraid, but about what or whom we should be afraid of' (Furedi, 2022)

The violet branch towards the bottom right seems to be about Mexican politics regarding migration and climate change.

Shooting off to the upper left side we can find a red cluster where most of the words have to do with the COVID-19 pandemic (the only exception being the word *vacunación* 'vaccination', which is at the bottom of the graph, in the green cluster group); pivotal with the upper green cluster, the word salud 'health' functions as a link to the seed words (top green cluster).

In the pink branch, which shoots upward from the central cluster, words related to natural catastrophes and victims (*sismo* 'earthquake', *inundación* 'flood', and *río* 'river'), seem to be bundled together.

The presence of many terms related to Mexican politics can be explained by the fact that this sub-corpus has a large section (about 60%) which was selected by date, while the remaining 40% were selected via the seeds 'fear' and its synonyms.

4.3.4. Essay similarity analysis

As opposed to the previous graphs, Figure 6 is split into two very large central groups and some small peripheral groups. The red left cluster stemming from the word 'fear' seems more descriptive and

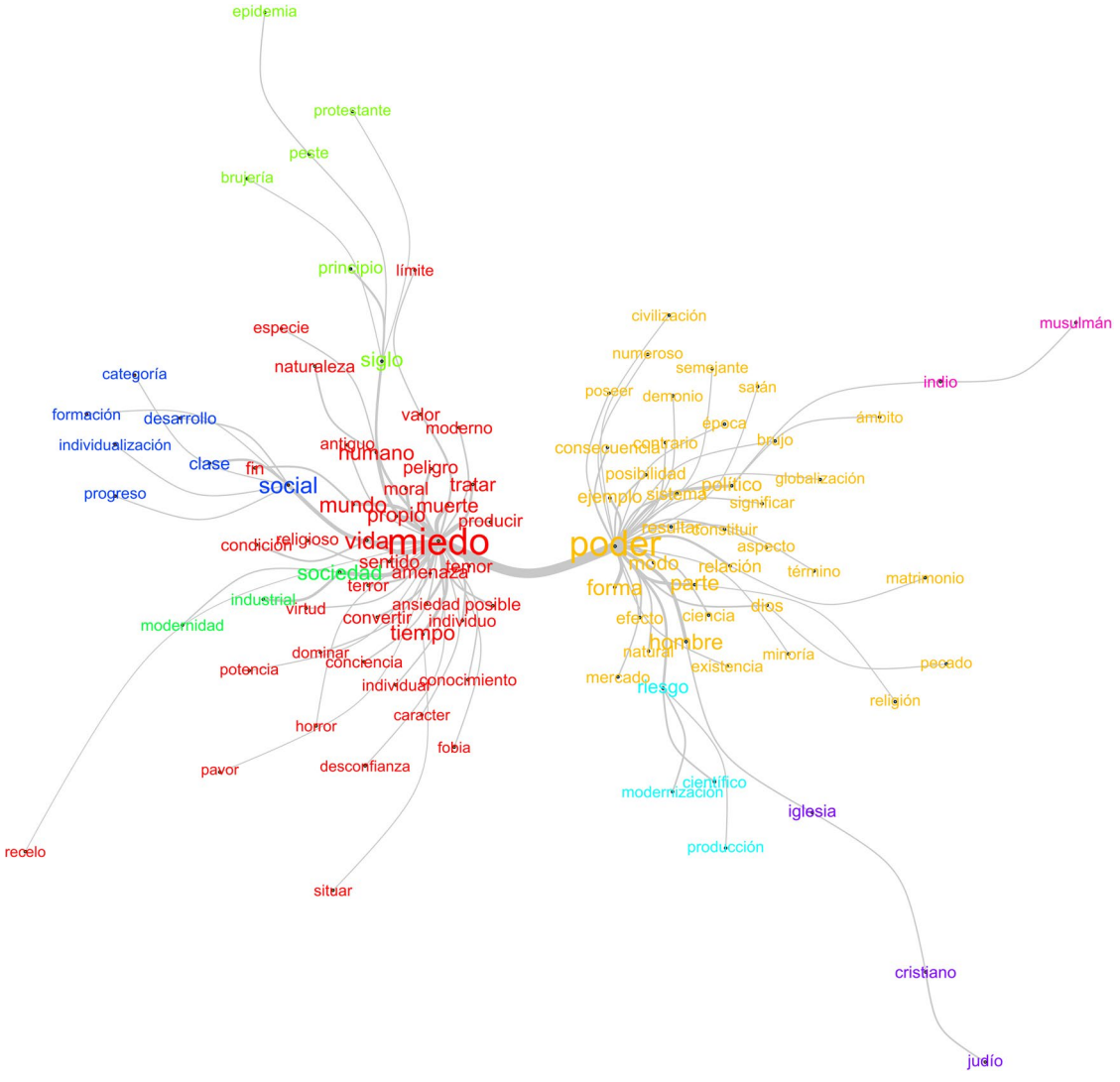

Figure 6. Key word radial representation graph in the essay sub-corpus.

philosophical since it has most of the seeds associated with it (*pavor* 'dread', *terror* 'terror', *horror* 'horror', *fobia* 'phobia', and *temor* 'aprehension'), along with more abstract terms like *vida* 'life', *conciencia* 'conscience', *tiempo* 'time', *individuo* 'individual', *conocimiento* 'knowledge', and *naturaleza* 'nature'. It also includes words that seem to indicate how to exert control: *dominar* 'to dominate', *desconfianza* 'mistrust', *amenaza* 'threat' The yellow right group branching off the word *poder* 'power' seems more analytical, related to political and social sciences: *ciencia* 'science', *consecuencia* 'consequence', *sistema* 'system', and *globalización* 'globalization'.

Three branches that seem to have a historical perspective in common branch out from the red group. In the upper part of the graph (light green) the references seem to point to early history (the words *siglo* 'century', *peste* 'plague', *epidemia* 'epidemic', and *brujería* 'witchcraft'). To the left of the center terms refer to the industrial society and modernity (green) and next to them, in blue, words reminiscent of social relations appear (*clase* 'class', *desarrollo* 'development', *progreso* 'progress', and *individualización* 'individualization').

To the right of the graph there are two small branches, both related to ethnicity and religion, perhaps civilizations (*cristiano* 'Christian', *judío* 'Jew', *musulmán* 'Moslem', and *indio* 'Indian').

This graph's most relevant characteristic is the very strong and direct connection between power and fear. This is a relation reminiscent of what was pointed out in the previous press and chronicle sub-corpora analyses. The chronicle sub-corpus seemed to point to the police force's (and the legal system's) corruption and its collusion with criminals. The press graph also includes the word *corrupción* 'corruption'. However, instead of exposing the ills of a corrupt regime, the graph analyzing the essays seems to imply that threatening behaviors and instilling fear, horror and terror are conscious mechanisms of social control historically used by the political system (including the church).

4.3.5. Tweets similarity analysis

The social network sub-corpus generates a graph (Figure 7) which seems to have the fewest word nodes and probably the largest number of clusters. This seems to imply that in spite of being the largest sub-corpus, it is probably the one with less lexical richness although, paradoxically, it reflects a large thematic variety. Words referring to feelings and emotions are scattered, not bundled around the central light blue cluster containing the word *miedo* 'fear'. Although most of the seed words are in that same light blue bundle branching out towards the upper side of the graph, several other words referring to affects can be seen throughout the graph: *inseguridades* 'insecurities' (royal blue bundle), *desconfianza* 'mistrust' (green), and *angustia* 'anguish' (purple), *deslealtad* 'disloyalty' and infidelidad, (both fuchsia), *ansiedad* 'anxiety', and *abandono* 'abandonment' (both yellow).

Contrary to the other sub-corpus graphs, here a relative large number of adjectives (most of them evaluative) can be found: many of them are pejorative—*cagones* 'cowardly' (orange) *arrogante* 'arrogant', *tonto* 'silly', *loco* 'mad' (all three green)—, and only a few positive: *valiente* 'brave', and *bonito* 'nice' (both also green).

We can also see many other nouns expressing emotions and feelings: *encariñarse* 'be fond of' *calmarte* 'calm down' *llanto* 'crying', *ataque* 'attack' (all four purple) *tranquilizar* 'calm down' (yellow), *mierda* 'shit' (blue), and *aterrador* 'frightening' (blue). The causes of the fear do not seem to be clearly identified and the focus is on the consequences of the experiencer.

Close examination of the semantic relationships of this graph allows us to identify formulaic sequences linked to certain media events. For example, the mention of *éxito,* 'success' (in the orange bundle at the top of the graph) is linked to the cliché 'fear of success', often mentioned regarding a television contest, as revealed by the concordance lines. Close and to the bottom of the central 'fear' node are the words *intuición* 'intuition' and *dar* 'to give', reminiscent of another cliché: *mi intuición me da miedo* 'my intuition frightens me' (as revealed in concordance lines). It is the only graph where pronouns (personal, possesive) appear (*mío* 'mine' [purple], *yo* 'I', *tú* 'you' (both blue], *él* 'he' [orange], *contigo* 'with you' [green bottom], *ustedes* 'you plural' ['pink]).

This is consistent with the nature of the social network domain, one of whose functions is simply to offer the possibility of expressing oneself and sharing and exchanging opinions, feelings and reactions.

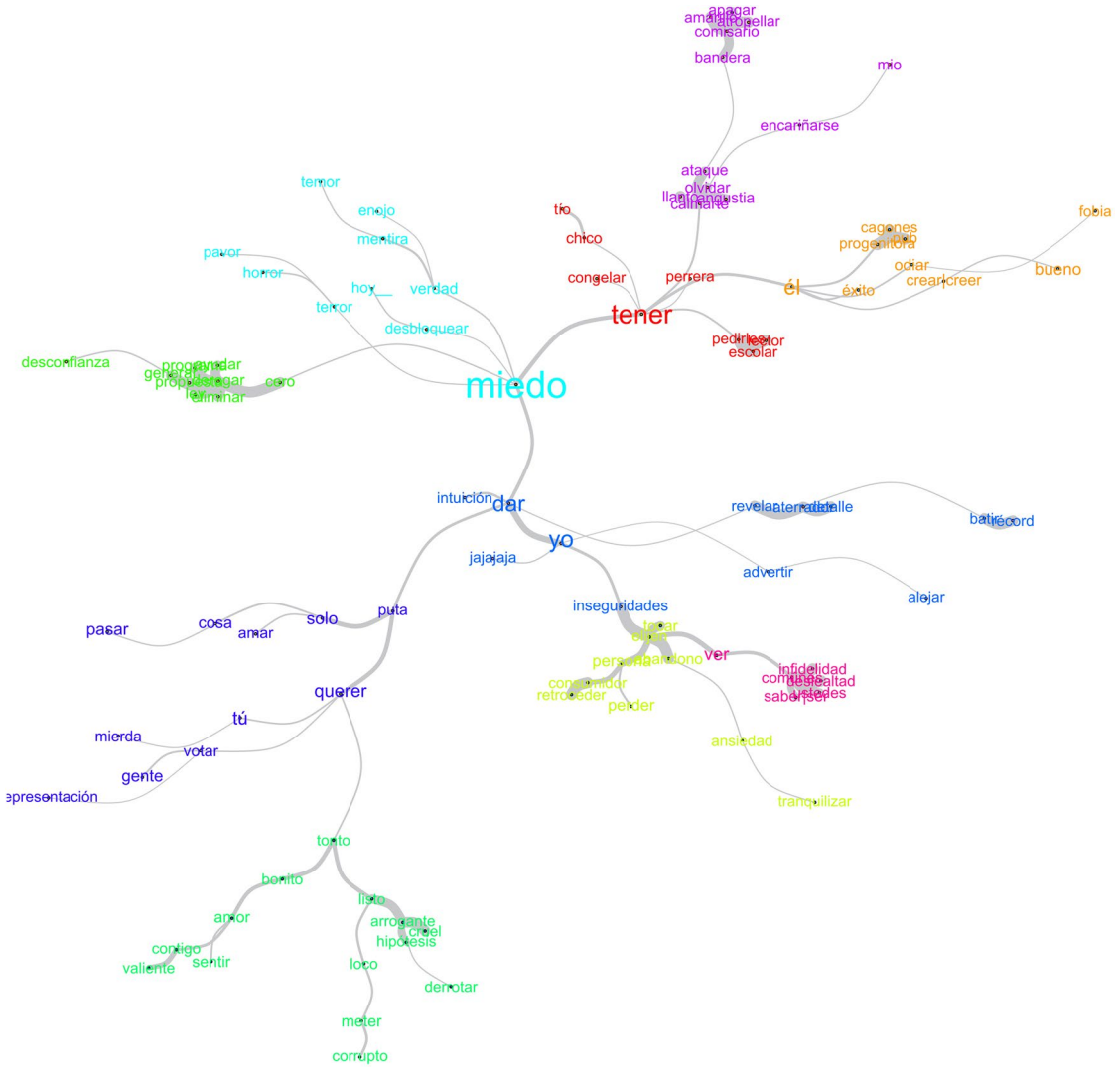

Figure 7. Key word radial representation graph in the social network sub-corpus.

5. Discussion and conclusions

In this section we discuss the implications of our results focusing mainly on theoretical and methodological issues. In this paper we have stated that our position in terms of theories on emotions is close to the theories considering context as a main provider of meaning, be it Hoey's Lexical Priming Theory, Barrett's Theory of Constructed Emotions, or Kövecses's Extended Conceptual Metaphor Theory, since their postulates are reflected in the constructivist approach corpus linguistics is part of. According to our view, meanings are constructed according to the (co[n])text and the genre; the social and historical circumstances in which linguistic productions occur cannot be ignored.

The more traditional approaches to the psychology of emotions tend to reduce the meaning of 'fear' as an emotion to a universal, to some extent essentialist, element. This tendency also seems to predominate in biology, where almost 30 years ago a gene responsible for anxiety was claimed to have been found (Lesch et al., 1996) and very recently the discovery of how a neurotransmitter may be causing the emotion of fear has been announced (Li et al., 2024). Without entering into conflict with other disciplines, from a linguistics standpoint, instead of reducing the scope of the concept of 'fear', our work broadens it, contextualizes it, offers nuances, all this considering a data-driven empirical approach. If applied to the concept of fear, Firth's (1957/1961) famous saying 'you shall know a word by the company it keeps' shows that it is a multifaceted, variable phenomenon depending on social, cultural and historical textual contexts.

The lexical analysis carried out on the basis of an occasional corpus on 'fear' shows the particularities, similarities and differences among the four genres/domains that compose it. In the analysis of the essay sub-corpus we saw that fear is approached differently from different disciplines of the humanities and social sciences. For example, the more philosophical and sociological texts relate it to power and social control, while the historical ones link it to religion, epidemics and stigmatized ethnic groups. Mexican chronicle speaks of fear generated by violence against women and family members, especially young persons, without leaving out condemnation regarding the ineffectiveness of institutions in dealing with this social scourge. In the press, fear is related to how political power deals with certain disasters (earthquakes, floods, COVID-19). In the domain of social networks, fear is linked to a great variety of topics that seem to have the common denominator of referring to personal insecurities that, on an emotional level, are expressed by individuals who participate in them. Essays focus on causality and agency, news and social networks deal with consequences and experiences. Subjectivity (expressive concerns) is present in social media; objectivity (analytical concerns) in the other sub-corpora.

We hope to have shown the usefulness of the statistical techniques presented in this paper in order to identify lexical items specific of each subcorpora/domain regarding the topic of fear. The items identified and their lexical, semantic and pragmatic combinations can be nested and can act as priming for each genre/domain as Hoey stated (see above in Section 2.1).

It should also be noted that the work presented here is not devoid of issues. We see at least three limitations that should be addressed in greater depth in future research of this nature. The first limitation has to do with corpora designed around a topic. Although we took great pains to design a corpus that covered a variety of disciplines, genres and domains, it is inevitable that the findings will reflect any imbalance in the selection of texts. Thus, for example, sub-corpora comparison may provide complementary data, but a comparison of word frequencies between each section would not be appropriate.

The second limitation has to do with specific problems that arise when working with social network corpora. In these cases, a sub-corpus is extracted from certain seeds and this method of extraction inevitably limits the use that can be made of the resulting data. For example, contrary to other subsections of the corpus that enable comparisons between different content units (an essay, an author, a news article), in the case of tweets there is no clear intermediate textual unit between the tweet and the corpus extracted from a seed. It is as if we were to discard sentences that do not contain the target word in the analyses performed on the essay corpus. The corpus selection determines what we are to find in it.

The third limitation lies in the fact that the statistical processing of lexicon in the corpora faces the challenge of overcoming the enormous computational weight of certain calculations. This forced us to work with smaller samples, which may have skewed the scope of some results. Efficient processing will always be related to corpus size, and it is not easy to establish when a sample is large enough to analyze a phenomenon of this nature.

As for the visual representation of statistical analyses, we have shown that a dendrogram, the Cartesian plane and the nodes and clusters graph (corresponding to Figures 1–3, respectively) are three different ways of reporting very similar data. The graph (Figure 3) seems to be the best way to represent clusters and relationships and enables the reader to establish connections that are grammatical, semantic, discursive and textual at a glance. It is colorful, it allows easy identification of nodes, clusters and their sizes (frequencies). It establishes linear connections between words and it is relatively easy to measure distances between them. In sum, it presents information in a simple, didactic, understandable and unproblematic way. On the other hand, Figure 1 has information which looks more like a table, it is the least dynamic and colorful visual representation of the three options; hierarchies are established in it but it is difficult to assess distances and connections among words. Finally, the representation of words in a Cartesian plane (Figure 2) has features present in both Figures 1 and 3. It looks like a word cloud and effectively shows the closeness of some of the clusters, which are occasionally intermingled. However, Figure 2 lacks edges; therefore, reading the data connections and distances is more difficult than in the linear connections of the Figure 3 graph. The apparent simplicity of Figure 3 makes it ideal for summarizing information and it appeals to wide audiences because it is as easy to read and interpret as a map. However, this apparent simplicity could also be considered a hindrance: it is more difficult to uncover subtleties or ambiguities and to identify and understand problematic issues.

Lastly, relevant contribution of this work is that the findings seem to support the postulates of the Theory of Constructed Emotions, namely, that emotions are diffuse, overlap and do not seem to have unique components even within the same genre or domain. Similarly, parallels can be drawn with the ideas of the Extended Conceptual Metaphor Theory, particularly with the assumption that cognition is grounded in the situation in which people act and lead their lives, the discourse they are engaged in at any time, and the conceptual knowledge they have accumulated (Kövecses, 2015, p. 200). Precisely because of all this, and despite the limitations pointed out, we believe that linguistics and digital humanities should play a prominent role in establishing bridges between the humanities and the so-called science of fear.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This study was suported by Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCyT) CF-2023-G-583. Challenge Based Research Funding Program 2022, Instituto Tecnológico y de Estudios Superiores de Monterrey (ITESM) E061 - EHE-GI02 - D-T3 - E.

About the authors

Ignacio Rodríguez Sánchez is a lecturer and researcher at the Facultad de Lenguas y Letras, Universidad Autónoma de Querétaro, México. MA, Universitat de Barcelona (Spain); PhD in Applied Linguistics, Swansea University (United Kingdom). His main interests are corpus linguistics, lexis, and digital humanities.

Antonio Reyes Pérez is a lecturer and researcher at the Facultad de Lenguas y Letras, Universidad Autónoma de Querétaro, Mexico; MA in Linguistics, Universidad Nacional Autónoma de México; PhD in Language Technology, Universidad Politécnica de Valencia (Spain). His main interests are natural language processing, corpus linguistics, and digital humanities.

Manuel Cebral Loureda is an assistant professor and researcher at the Department of Humanistic Studies of the School of Humanities and Education, Instituto Tecnologico y de Estudios Superiores de Monterrey, Campus Monterrey, México. MSC Universidad Nacional a Distancia (Spain); MA, Universidad de Valencia (Spain); PhD in Philosophy, Universidade de Santiago de Compostela (Spain). His main interests are natural language processing, human-computer interaction, and digital humanities.

Ester Bautista Botello is a lecturer and researcher at the Facultad de Lenguas y Letras, Universidad Autónoma de Querétaro, México. MA, University of Athens, Ohio (USA); PhD in Literature, King's College London (United Kingdom). Her main interests are Spanish and Latin American Literature, Women's writing and digital humanities.

Nohemí Lugo Rodríguez is an associate research professor in the Department of Media and Digital Culture at Tecnológico de Monterrey, Querétaro Campus, México. She holds a Master of Arts in Spanish from Western Michigan University (USA) and a PhD in Communication from Pompeu Fabra University (Spain). Her primary research interests include interactive communication, digital methods, digital culture, and neurodiversity.

ORCID

Ignacio Rodríguez Sánchez http://orcid.org/0000-0001-7730-2005
Antonio Reyes Pérez http://orcid.org/0000-0001-9560-1674
Manuel Cebral Loureda http://orcid.org/0000-0001-6359-2427
Ester Bautista Botello http://orcid.org/0000-0002-8429-4856
Nohemí Lugo Rodríguez http://orcid.org/0000-0002-1334-7914

References

Allison, S., Heuser, R., Jockers, M., Moretti, F., & Witmore, M. (2011). *Qualitative formalism: An experiment. Stanford Literary Lab*. https://litlab.stanford.edu/pamphlets/

Almela, Á., Berber Sardinha, T., & Cantos-Gómez, P. (2022). Métodos multidimensionales basados en corpus del español. In G. Parodi, P. Cantos-Gómez, & C. Howe (Eds.), *Lingüística de corpus en español*. Routledge.

Baker, C., Fillmore, C., & Lowe, J. (1998). *The Berkeley FrameNet Project* [Paper presentation]. 86-90-86–90. https://doi.org/10.3115/980451.980860
Barrett, L. F., & Russell, J. A. (2014). *The psychological construction of emotion*. Guilford Publications.
Bautista, E., & Pérez, D. (2022). Cartografías del miedo en la narrativa mexicana contemporánea. In D. Rozotto & M. R. Rojas (Eds.), *Violencia, poder y afectos: Narrativas Del Miedo En Latinoamérica* (pp. 105–118). Boydell & Brewer.
Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., Holland, J., Ke, J., Larsen-Freeman, D., & Schoenemann, T, The "Five Graces Group". (2009). Language is a complex adaptive system: Position paper. *Language Learning*, *59*(s1), 1–26. https://doi.org/10.1111/j.1467-9922.2009.00533.x
Berber Sardinha, T., & Veirano Pinto, M. (Eds.). (2019). Multi-*dimensional analysis. Research methods and current issues*. Bloomsbury Publishing.
Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. University of California Press.
Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press.
Biber, D. (1995). *Dimensions of register variation: A cross-linguistic comparison*. Cambridge University Press.
Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.
Biggam, C. P., Hough, C., Kay, C. J., & Simmons, D. R. (2011). *New Directions in Colour Studies*. John Benjamins Publishing Company.
Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with python: Analyzing text with the natural language toolkit*. O'Reilly.
Bondi, M., & Scott, M. (2010). *Keyness in texts*. John Benjamins Publishing Company.
Boucher, C. (2011). *The hypergeometric distribution*. WOLFRAM Demonstrations Project. http://demonstrations.wolfram.com/TheHypergeometricDistribution/
Bradley, M. M., & Lang, P. J. (1999). Affective norms for English words (ANEW): Instruction manual and affective ratings. In *Technical Report C-1*. The Center for Research in Psychophysiology, University of Florida.
Bybee, J. L., & Hopper, P. (2001). *Frequency and the emergence of linguistic structure*. John Benjamins Publishing Company.
Cebral-Loureda, M., & Sued-Palmeiro, G. E, Tecnológico de Monterrey. (2021). Los inicios de la pandemia de COVID19 en Twitter. Análisis computacional de la conversación pública en lengua española. *Cuadernos.info*, *49*(49), 1–25. https://doi.org/10.7764/cdi.49.27467
Cook, R. S., Kay, P., & Regier, T. (2005). The world color survey database. In H. Cohen & C. Lefebvre (Eds.), *Handbook of categorization in cognitive Science* (pp. 223–241). Elsevier Science Ltd. https://doi.org/10.1016/B978-008044612-7/50064-0
Egbert, J., Biber, D., & Gray, B. (2022). *Designing and evaluating language corpora: A practical framework for corpus representativeness*. Cambridge University Press.
Ekman, P. (2007). *Emotions revealed: Recognizing faces and feelings to improve communication and emotional life*. Henry Holt and Company.
Ekman, P., & Cordaro, D. (2011). What is meant by calling emotions basic. *Emotion Review*, *3*(4), 364–370. https://doi.org/10.1177/1754073911410740
Ekman, P., & Friesen, W. V. (2003). *Unmasking the face: A guide to recognizing emotions from facial clues*. ISHK.
Ellis, N. C., & Larsen-Freeman, D. (2006). Language emergence: Implications for applied linguistics—introduction to the special issue. *Applied Linguistics*, *27*(4), 558–589. https://doi.org/10.1093/applin/aml028
Ellis, N. C., & Larsen-Freeman, D. (2009). *Language as a complex adaptive system*. John Wiley & Sons.
Fang, C. A., & Cao, J. (2015). Text genres and registers. *The computation of linguistic features*. Springer.
Firth, J. R. (1961). A synopsis in linguistic theory. In *Studies in Linguistic Analysis* (Vols. 1952–1959). Blackwell. (Original work published 1957)
Fontaine, J. R. J., Scherer, K. R., & Soriano, C. (Eds.). (2013). *Components of emotional meaning. A sourcebook*. Oxford University Press.
Furedi, F. (2022). *Cómo funciona el miedo: La cultura del miedo en siglo XXI*.
Goldberg, A. (1996). Making one's way through the data. In M. Shibatani & S. Thompson (Eds.), *Grammatical constructions: Their form and meaning.* (pp. 29-53-29–53). Clarendon Press.
Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. University of Chicago Press.
Gries, S. T., & Stefanowitsch, A. (2006). *Corpora in cognitive linguistics: Corpus-based approaches to syntax and lexis*. Walter de Gruyter.
Guasch, M., Ferré, P., & Fraga, I. (2016). Spanish norms for affective and lexico-semantic variables for 1,400 words. *Behavior Research Methods*, *48*(4), 1358–1369. https://doi.org/10.3758/s13428-015-0684-y
Hoey, M. (2005). *Lexical priming: A new theory of words and language*. Routledge.
Hoey, M. (2009). Corpus-driven approaches to grammar. The search for common ground. In U. Römer & R. Schulze (Eds.), *Exploring the lexis-grammar interface*. John Benjamins Publishing Company.
Hoey, M. (2013). *Textual interaction: An introduction to written discourse analysis*. Routledge.
Hunston, S., & Francis, G. (2000). *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. John Benjamins Publishing Company.

Jackson, J. C., Watts, J., Henry, T. R., List, J.-M., Forkel, R., Mucha, P. J., Greenhill, S. J., Gray, R. D., & Lindquist, K. A. (2019). Emotion semantics show both cultural variation and universal structure. *Science*, *366*(6472), 1517–1522. https://doi.org/10.1126/science.aaw8160

Jockers, M. (2014). *Text analysis with R for students of literature*. Springer.

Jockers, M. L. (2013). *Macroanalysis: Digital methods and literary history*. University of Illinois Press.

Kessler, B. (1995). *Computational dialectology in Irish Gaelic* [Paper presentation]. *Computational Linguistics* [Paper presentation]. Seventh Conference of the European Chapter of the Association for, In S. P. Abney & E. W. Hinrichs (Eds.), Association for Computational Linguistics. https://aclanthology.org/E95-1009 https://doi.org/10.3115/976973.976983

Kövecses, Z. (1990). *Emotion concepts*. Springer.

Kövecses, Z. (2015). *Where metaphors come from: Reconsidering context in metaphor*. Oxford University Press.

Kövecses, Z. (2020). *Extended conceptual metaphor theory*. Cambridge University Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.

Langacker, R. (1987). *Foundations of cognitive grammar*. Stanford University Press.

Lesch, K.-P., Bengel, D., Heils, A., Sabol, S. Z., Greenberg, B. D., Petri, S., Benjamin, J., Müller, C. R., Hamer, D. H., & Murphy, D. L. (1996). Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region. *Science (New York, N.Y.)*, *274*(5292), 1527–1531. https://doi.org/10.1126/science.274.5292.1527

Li, H., Jiang, W., Ling, L., Pratelli, M., Chen, C., Gupta, V., Godavarthi, S. K., & Spitzer, N. C. (2024). Generalized fear after acute stress is caused by change in neuronal cotransmitter identity. *Science*, *383*(6688), 1252–1259. https://doi.org/10.1126/science.adj5996

McEnery, T., & Hardie, A. (2011). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.

Nerbonne, J., & Heeringa, W. (1997). Measuring dialect distance phonetically. In *Proceedings of the Third Meeting of the ACL Special Interest Group in Computational*, 11—18. https://www.aclweb.org/anthology/W97-1102.pdf

Ogarkova, O. (2013). Folk emotion concepts: Lexicalization of emotional experiences across languages and cultures. In J. R. J. Fontaine, K. R. Scherer, & C. Soriano (Eds.), *Components of emotional meaning. A sourcebook* (pp. 46–61). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199592746.001.0001

Pardo Veiras, J. L., & Arredondo, I. (2021). *Una guerra inventada y 350,000 muertos en México*. Washington Post. https://www.washingtonpost.com/es/post-opinion/2021/06/14/mexico-guerra-narcotrafico-calderon-homicidios-desaparecidos/

Parodi, G. (2005). Lingüística de corpus y análisis multidimensional: Exploración de la variación en el corpus PUCV-2003. *Revista Española de Lingüística*, *35*(1), 45–76.

Pennebaker, J. W. (2011). *The secret life of pronouns: What our words say about us*. Bloomsbury Publishing USA.

Pennebaker, J. W., & Evans, J. F. (2014). *Expressive writing: Words that Heal*. Idyll Arbor, Incorporated.

Perea, M.-P., & Ueda, H. (2010). Applying Quantitative Analysis Techniques to *La flexió verbal en els dialectes catalans*. *dig*, *18*(1), 99–114. https://doi.org/10.1515/dig.2010.006

Ratinaud, P. (2009). *Iramuteq: Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires [Computer software]*. http://www.iramuteq.org

Reinert, M. (1985). Classification Descendante Hiérarchique: Un algorithme pour le traitement des tableaux logiques de grandes dimensions. In E. Diday, Y. Escouffier, L. Lebart, J. Pages, Y. Schektman, & R. Tomassone (Eds.), *Quatrièmes Journées Internationales" Analyses des Données et Informatique* (pp. 23–28). Elsevier Science publishers BV, Institut national de recherche en sciences et technologies du numérique. https://shs.hal.science/halshs-04066409

Reinert, M. (1986). Un logiciel d'analyse lexicale. *Les Cahiers de l'analyse Des Données*, *11*(4) 471–481

Rude, S., Gortner, E.-M., & Pennebaker, J. (2004). Language use of depressed and depression-vulnerable college students. *Cognition & Emotion*, *18*(8), 1121–1133. https://doi.org/10.1080/02699930441000030

Sampson, G. (2013). The empirical trend: Ten years on. *International Journal of Corpus Linguistics*, *18*(2), 281–289. https://doi.org/10.1075/ijcl.18.2.05sam

Scherer, K. R., & Fontaine, J. R. J. (2019). The semantic structure of emotion words across languages is consistent with componential appraisal models of emotion. *Cognition and Emotion*, *33*(4), 673–682. https://doi.org/10.1080/02699931.2018.1481369

Scherer, K. R., & Moors, A. (2019). The emotion process: Event appraisal and component differentiation. *Annual Review of Psychology*, *70*(1), 719–745. (Volume 70, 2019), https://doi.org/10.1146/annurev-psych-122216-011854

Scott, M., & Tribble, C. (2006). *Textual patterns: Key words and corpus analysis in language education*. John Benjamins Publishing Company.

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

Sinclair, J. M. (2004). Trust the text: Language, corpus and discourse. R. Carter, Ed. Routledge.

Stadthagen-González, H., Ferré, P., Pérez-Sánchez, M. A., Imbault, C., & Hinojosa, J. A. (2018). Norms for 10,491 Spanish words for five discrete emotions: Happiness, disgust, anger, fear, and sadness. *Behavior Research Methods*, *50*(5), 1943–1952. https://doi.org/10.3758/s13428-017-0962-y

Stadthagen-Gonzalez, H., Imbault, C., Pérez Sánchez, M. A., & Brysbaert, M. (2017). Norms of valence and arousal for 14,031 Spanish words. *Behavior Research Methods*, *49*(1), 111–123. https://doi.org/10.3758/s13428-015-0700-2

Tomkins, S. (1962). *Affect imagery consciousness: Volume I: The positive affects*. Springer Publishing Company.

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, *45*(4), 1191–1207. https://doi.org/10.3758/s13428-012-0314-x

Appendixes: Appendix 1. Essay corpus.

References	FileTokens	Discipline
Appadurai, A. (2007). *El rechazo de las minorías. Ensayo sobre la geografía de la furia*. Tusquets México.	39,860	Anthropology
Augé, M. (2014). *Los nuevos miedos*. Paidós.	10,165	Philosophy
Bauman, Z. (2006). *Miedo líquido: La sociedad contemporánea y sus temores*. Paidós.	80,684	Sociology
Beck, U. (1986). *La sociedad del riesgo: Hacia una nueva modernidad*. Group Planeta Spain.	132,385	Sociology
Bude, H. (2017). *La sociedad del miedo*. Herder Editorial.	37,786	Sociology
Butler, J. (2020). *Sin miedo: Formas de resistencia a la violencia de hoy*. Taurus.	25,410	Philosophy
Castany Prado, B. (2022). *Una filosofía del miedo*. Anagrama.	101,611	Philosophy
Delumeau, J. (1978). *El miedo en Occidente*. Taurus.	210,413	History
Duby, G. (1995). *Año 1000, año 2000: La huella de nuestros miedos*. Andres Bello.	17,050	History
Fernández Juárez, G., & Pedrosa, J. M. (Eds.). (2008). *Antropologías del miedo*. Calambur.	82,987	Anthropology
Furedi (2022). *Cómo funciona el miedo: La cultura del miedo en siglo XXI*. Ediciones Rialp, S.A.	103,612	Sociology
Marina, J. A. (2006). *Anatomía del miedo: Un tratado sobre la valentía*. Anagrama.	66,599	Philosophy
Mora Teruel, F. (2016). *¿Es posible una cultura sin miedo?* Alianza Editorial.	43,851	Neuroscience
Nussbaum, M. C. (2019). *La monarquía del miedo: Una mirada filosófica a la crisis política actual*. Ediciones Paidós.	88,215	Philosophy
Reguillo, R. (2000). Los laberintos del miedo. Un recorrido para fin de siglo. Revista de Estudios Sociales, 05, Article 05.	33,291	Sociology
Reguillo, R. (2001). Imaginarios globales, miedos locales: Construcción social del miedo en la ciudad. *Estudios: revista de investigaciones literarias*, 17, 47–64.		
Reguillo, R. (2008). Sociabilidad, inseguridad y miedos: Una trilogía para pensar la ciudad contemporánea. *Alteridades, 18*(36), 63–74.		
Reguillo-Cruz, R. (2002). Miedo al otro. Comunicación, poder y representación en una contemporaneidad sobresaltada. *Anagramas, 1*(1), 51–66.		
Robin, C. (2009). *El Miedo. Historia de Una Idea Politica*. Fondo de Cultura Económcia. (Obra original publicada en 2004)	116,838	Politics
Speckman Guerra, E., Agostoni, C., & Gonzalbo, P. (Eds.). (2009). *Los miedos en la historia*. Colegio de México.	103,582	History
Villa Martínez, M. I. (Ed.). (2002). *El Miedo. Reflexiones sobre su dimension social y cultural*. Corporación Región, Medellín, Colombia.	69,484	Sociology

Appendix 2. Chronicle corpus.

References	Tokens
Turati, M., & Rea, D. (Eds.). (2014). *Entre las cenizas. Historias de vida en tiempo de muerte*. Sumar+	55,553
Rea, D. (2016). *Nadie les pidió perdón: Historias de impunidad y resistencia*. Ediciones Urano.	70,427
Carrión, L. (2018). *La fosa de agua: Desapariciones y feminicidios en el río de los Remedios*. DEBATE.	54,042
Rea, D. (2020). *Ya no somos las mismas: Y aquí sigue la guerra*. PRH Group Editorial.	58,122
Roldán, N. (Ed.). (2022). *Mexicanas en pie de lucha: Pese al gobierno machista, las violencias y el patriarcado*. GRIJALBO.	58,639

Appendix 3. Top 100 specificities per genre.

Chronicle		Essay		Press		Tweets	
bianca	tarde	siglo	principio	nacional	partido	tener	estar
policía	levantar	modo	musulmán	stress	coronavirus	mío	generar
mujer	pedir	social	aspecto	temor	amparo	yo	él
él	ministerio	mundo	construir	mexicano	sismo	dar	siempre
desaparecer	caminar	forma	existencia	covid	realizar	miedo	batir
cuerpo	género	sociedad	brujería	gobierno	martes	querer	éxito
hijo	secuestrar	riesgo	virtud	millón	reportar	tú	récord
casa	matar	peste	modernidad	presidente	haitiano	jajaja	amor
joven	adn	peligro	religión	pandemia	unidad	derogar	mierda
mili	adolescente	ciencia	crítico	ansiedad	latinoamericano	elijen	atropellar
feminista	facebook	judío	individuo	mil	climático	puta	jajaja
violencia	expediente	propio	pecado	salud	solicitar	nunca	bueno
feminicidio	piraña	iglesia	individualización	país	río	solo	pasar
muchacho	regresar	poder	posibilidad	desconfianza	próximo	inseguridades	cruel
detener	escuela	moderno	natural	ayer	roble	cero	hipótesis
asesinar	saber	industrial	demonio	moreno	indicar	abandono	advertir
desaparición	indígena	cristiano	dios	vacuna	cámara	ver	ley
comunitario	llevar	clase	epidemia	entidad	fobia	hoy	tranquilzar
víctima	fosa	brujo	formación	depresión	municipio	calmarte	cosa
madre	teléfono	naturaleza	período	diputado	comisión	enojo	amar
encontrar	reportero	tiempo	antiguo	ciento	congreso	eliminar	comisario
buscar	organizar	relación	amenaza	gobernador	acta	perder	cagones
procuraduría	rostro	resultar	carácter	acuerdo	evento	votar	desbloquear
cuidar	operativo	cuestión	matrimonio	director	región	propuesta	valiente
justicia	homicidio	época	categoría	vacunación	inundación	electoral	derrotar
año	estar	político	moral	destacar	comunidad	llanto	lector
compañero	estancia	científico	producción	especial	asegurar	representación	ataque
mamá	cinco	producir	conciencia	cumbre	legislar	tocar	odiar
camioneta	mano	nuestro	fin	secretar	migrantes	listo	pub
llegar	dos	vida	desarrollo	ciudad	vacunados	arrogante	progenitora
forense	militar	consecuencia	situar	estado	migración	programa	persona
día	asesinato	muerte	protestante	independencia	autoridad	meter	detalle
escuchar	federal	religioso	progreso	lunes	candidato	ayudar	congelar
feminicidios	celular	mercado	límite	contagios	encabezar	pedirles	mentira
fotografía	par	indio	término	señalar	instituto	comunes	angustia
mes	autoridad	condición	numeroso	horror	ejecutivo	ustedes	creer
amigo	perito	efecto	especie	unir	informar	deslealtad	amarillo
familia	padre	sentido	potencia	secretario	capitalino	saber	loco
oficina	ser	modernización	ejemplo	reconstrucción	damnificar	ser	verdad
esposo	comenzar	civilización	semejante	mandatario	internacional	infidelidad	bonito
cuidado	crimen	humano	valor	federal	senado	gente	chico
niño	secuestro	hombre	conocimiento	dosis	alertar	intuición	bandera
nombre	ropa	individual	contrario	vivienda	caso	perrera	escolar
trabajar	responder	minoría	posible	terror	refinería	retroceder	corrupto
organización	periodista	satán	objetivo	dato	alcaldía	contigo	revelar
decir	salir	globalización	dominar	peso	jefe	olvidar	alejar
bolsa	trabajo	tratar	menos	detallar	zona	sentir	apagar
resto	hora	sistema	ámbito	semana	apoyo	encariñarse	tonto
tres	hermano	convertir	poseer	pánico	recurso	aterrador	éxito
defensor	familiar	parte	significar	unión	ubicar	consumidor	tío

Appendix 4. Top 50 words per class.

#	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
1	rt	policía	méxico	social	vida	siglo
2	miedo	bianca	nacional	riesgo	mundo	iglesia
3	co	casa	com	sociedad	más	xvi
4	https	desaparecer	mx	industrial	humano	judío
5	dar	detener	gimm	político	propio	cristiano
6	querer	después	lópez	científico	resultar	peste
7	ir	muchacho	gobierno	ciencia	modo	brujo
8	solo	joven	presidente	mercado	forma	xvii
9	nunca	niño	septiembre	sistema	mismo	dios
10	ver	mili	obrador	desarrollo	idea	satán
11	puta	tres	excelsior	trabajo	natural	francia
12	jajajajaja	asesinar	millón	modernización	potencia	pecado
13	re	hijo	mexicano	relación	minoría	brujería
14	00hrmin	paco	diputado	económico	realidad	epidemia
15	elijen	mes	foto	laboral	violencia	xv
16	camiisaraviaa	mariana	covid	clase	sensación	demonio
17	inseguridades	irish	secretario	consecuencia	moral	diablo
18	abandono	camioneta	federal	condición	temor	protestante
19	importar	noche	moreno	desigualdad	esfuerzo	europa
20	hoy	encontrar	secretar	producción	sentimiento	xiv
21	cómo	dos	gobernador	individualización	considerar	católico
22	patrondelmal_rd	año	entidad	formación	uno	antiguo
23	calmarte	teléfono	estado	profesional	deseo	lutero
24	enoj	tecámac	pandemia	economía	podar|poder	frecuentemente
25	andrea_avilaj	desaparición	comisión	ámbito	filosófico	hombre
26	perder	mamá	manuel	profesión	parecer	turco
27	tocar	esposo	ayer	decisión	tratar	tierra
28	llanto	hora	andrés	etc	tender	divino
29	bunseki_gaby	margy	salud	desempleo	objetivo	francés
30	saber|ser	llegar	ciudad	técnica	acción	infierno
31	tinoobustamante	miriam	peso	técnico	globalización	época
32	echelecabeza	resto	realizar	modernidad	reducir	escribir
33	votar	fotografía	celac	proceso	hecho	religioso
34	pq	tarde	acuerdo	crítico	individuo	xviii
35	cedemocratico	llevar	pri	aspecto	físico	israelita
36	tampoco	matar	país	progreso	muerte	xiii
37	representacion	secuestrar	gonzález	fase	incertidumbre	religión
38	senadogovco	padre	sheinbaum	situación	naturaleza	anticristo
39	olvidar	guaymas	instituto	exigencia	placer	numeroso
40	luualcrz	familia	claudia	institucional	amistad	cristo
41	siempre	cuerpo	jefe	tecnología	experiencia	castigo
42	meter	adolescente	congreso	estructura	profundo	obra
43	rivadavia630	caminar	redacción	contradicción	spinoza	confesión
44	ya	cinco	director	precisamente	virtud	inglaterra
45	retroceder	diana	lunes	posibilidad	moderno	santo
46	comúnes	colonia	mil	cambio	cotidiano	culto
47	deslealtad	procuraduría	unir	competencia	dolor	cielo
48	intuición	araceli	cumbre	modelo	filosofía	renacimiento
49	infidelidad	escuchar	amparo	racionalidad	capacidad	herejes
50	aveces	bolsa	cámara	cientificación	importante	herejía