

# Colocaciones de "Hombre" y "Mujer" en noticias en línea en español: una mirada crítica a la representación de género

*Collocations of "MAN" and "WOMAN" in Spanish online news: A critical look at gender representation*

**Héctor Castro Mosqueda**

*Escuela Normal Superior del Estado de Guanajuato*  
castro32579@hotmail.com

## Resumen

Este estudio utiliza técnicas de la lingüística de corpus para investigar cómo se representan hombres y mujeres en las noticias en línea en español. La investigación explora los patrones de colocaciones verbales de los lemas HOMBRE y MUJER dentro del corpus NOW. Los hallazgos indican que ciertos verbos, como "discutir", "mentir", "amenazar" y "gritar", están más estrechamente asociados con los hombres, mientras que verbos como "protestar", "callar" y "alegar" están más vinculados a las mujeres. Estos patrones reflejan las dinámicas de género en el lenguaje y se alinean con investigaciones en inglés que muestran una asociación más fuerte entre los hombres y acciones violentas. Por otro lado, el verbo "morir" aparece más frecuentemente con HOMBRE, y MUJER con "fallecer", el cual, a diferencia de "morir", lleva consigo matices emocionales y de respeto. Además, el estudio revela la colocación exclusiva de "fracasar" con HOMBRE y "triunfar" con MUJER, arrojando luz sobre las expectativas de logro de ambos géneros. Es notable también que los verbos "lidar" y "enfrentar" ocurren de manera única con MUJER, lo que sugiere una conexión entre las mujeres y el manejo de situaciones difíciles. Este análisis léxico subraya la intrincada interacción entre lenguaje, género y percepciones sociales, enfatizando la necesidad de un análisis e investigación continuos para revelar y desafiar los estereotipos de género y las dinámicas de poder arraigadas en el lenguaje.

**Palabras clave:** Colocaciones, género, corpus, español

## Abstract

This study employs corpus linguistics methods to analyze the representation of men and women in Spanish online news. It examines the collocational patterns of verbs with the lemmas "man" and "woman" within the NOW corpus. The results show that certain verbs, such as "to argue," "to lie," "to threaten," and "to yell," are more strongly associated with men, while verbs like "to protest," "to be silent," and "to argue" are more commonly linked to women. These patterns reflect gender dynamics in language and align with research in English that suggests a stronger connection between men and violent actions. Conversely, the verb "to die" appears more frequently with MAN, while WOMAN is more often paired with "to pass away," which, unlike "to die," carries emotional and respectful connotations. Additionally, the study highlights the exclusive pairing of "to fail" with MAN and "to succeed" with WOMAN, shedding light on achievement expectations for both genders. Notably, verbs such as "to deal with" and "to face" are used only with WOMAN, indicating a link between women and managing difficult situations.

This lexical analysis highlights the intricate interplay between language, gender, and societal perceptions, underscoring the need for ongoing study and research to uncover and challenge the gender stereotypes and power dynamics embedded in language.

**Keywords:** Collocations, gender, corpus, Spanish

## 1. INTRODUCTION

This article aims to achieve two goals. First, it explores applying collocation techniques from corpus linguistics to reveal gender representations in the digital journalism landscape. Second, it contributes to the existing research focused on using corpus linguistics to derive social and cultural insights from Spanish electronic corpora. The study examines how both women and men are portrayed in Spanish online news, an area that has received less scholarly attention in Spanish compared to English.

Gender, as a social construct, encompasses the significance attributed to the societal categories of masculinity and femininity within a specific culture (Eagly & Eagly, 2012). These cultural significances become integral to individuals' identities, shaping their perceptions and self-perceptions, as well as their perceptions of others. Through these portrayals, individuals conceptualize and articulate their identities in relation to the culturally established notions of masculine and feminine traits associated with men and women. Consequently, these gender dimensions may influence their thoughts and behaviors.

Gender expectations reflect prevailing social norms and values, which are reinforced through everyday interactions across different settings. Exploring how gender appears in social situations can provide insights into the supportive framework of social structure and the mechanisms of social control that uphold it (West & Zimmerman, 1987). This analysis is based on the idea of language as a powerful tool, not only for reflecting people's understanding of the world but also for maintaining and reinforcing their perceptions of the phenomena and entities around them (Baker, 2014).

The verbal collocational analysis of the lemmas HOMBRE (MAN) and MUJER (WOMAN) explores a significant social issue: the widespread presence of violence. This issue casts a shadow across many aspects of society, leading to not only physical acts of violence but also affecting psychological and social realms. The pervasive presence of violence appears in the beliefs and ideologies held by certain groups, thereby normalizing abuse and brutality. Consequently, language becomes a tool for justifying violence, a trend driven not only by ordinary people but also by the media and social institutions.

### 1.1 Corpus linguistics and gender

Over the past twenty years, Corpus Linguistics (CL) has significantly influenced the field of language and gender research. Initially, studies in this area relied on small data sets, but CL has revolutionized the approach by utilizing large collections of data consisting of millions of words and sophisticated corpus query tools. The benefits of CL have been recognized across various fields, including translation studies, discourse analysis, applied linguistics, and sociolinguistics, which have integrated these methods into their research practices (Baker, 2010). Furthermore, corpus analysis now finds applications in numerous areas such as forensic linguistics, language education, media studies, gender research, and institutional and workplace discourse (Hyland et al., 2012). Nevertheless, critics have voiced concerns about the potential drawbacks of heavy reliance on quantification through CL. There is a worry that this approach might oversimplify complex language phenomena, reinforce stereotypes, or perpetuate prejudice. It is crucial to understand, however, that CL can complement existing research paradigms rather than replace them entirely (Baker, 2010: 9). Essentially, Corpus Linguistics serves as a valuable addition, enriching the field of linguistic research. McEnery and Baker (2015) support this perspective by noting that researchers often start with quantitative methods, employing techniques such as keyword analysis and collocate examination to guide their investigations. Yet, as their research develops, a natural shift occurs toward a more qualitative and context-sensitive approach. This transition, in turn, decreases dependence on computer software for analysis (p. 2). CL has played a crucial role in shaping the research landscape of language and gender. By utilizing large datasets and advanced tools, CL has created new opportunities for linguistic research across disciplines. While concerns about over-quantification persist, combining CL with traditional methods

can enhance research outcomes, leading to a more thorough understanding of language patterns and behaviors.

Research on gender representation using corpus linguistics remains a fascinating topic. The application of CL methods has expanded into various fields, including politics, video games, workplaces, education, sexuality, and even online journalism (Norberg, 2016; Horton, 2018; Heritage, 2020; Karimullah, 2020; Krendel, 2020; Chuaikun & Wijitsopon, 2021; Norberg & Johansson, 2021; Tsapro & Semeniuk, 2021).

Baker and Brookes (2021) highlight the numerous advantages that corpus linguistics offers for gender analysis. Primarily, the inherent comprehensiveness and representativeness of corpora, compared to smaller text datasets, enable researchers to make more generalizable statements about language patterns. Second, this approach underscores the importance of linguistic nuances that may be subtle when viewed in isolation but become more apparent when examined within a larger discourse or text context. Additionally, using larger corpora allows for the seamless inclusion of a wide range of discourse perspectives on a particular topic, effectively capturing resistant or minority viewpoints.

Furthermore, corpus linguistics adds an essential element of objectivity to discourse analysis through two key principles. The first principle emphasizes the importance of avoiding systematic bias when selecting corpus texts, ensuring that texts are not excluded solely to support pre-existing arguments. The second principle underscores a strong commitment to full accountability for all data collected. Together, these principles help discourse analysts effectively counter accusations of cherry-picking texts or examples that confirm predetermined views, as noted by Widdowson (2004, as cited in Baker & Brookes, 2021).

## 1.2 Previous research

There have been relatively few studies examining how women and men are portrayed linguistically in Spanish, with only a few identified (Alochis, 2016; Castillo, 2019). This is in stark contrast to the extensive research on the linguistic depiction of gender in English. Language and gender studies have a long history, covering topics from differences in language use between women and men and linguistic deficits of women to the idea that gender is a social construct (Weatherall, 2005; Litosseliti, 2014; Coates, 2015; Kendall & Tannen, 2015; Flowerdew & Richardson, 2017). While much linguistic research has employed descriptive and qualitative approaches to analyze how men and women use language, the past decade has seen significant research in Corpus linguistics (CL) focusing on how women and men are represented linguistically in different corpora (Caldas-Coulthard & Moon, 2010; Baker, 2012, 2013; Moon, 2014; McEnery & Baker, 2015). However, most research on gender linguistic representation has been conducted in English, leaving Spanish comparatively understudied. Therefore, this study aims to expand research in this area by applying CL tools and techniques to investigate gender representation in Spanish.

Romaine (2000) utilized the Brown Corpus to examine neutral terms, such as "chairperson." Her results showed that "chairman" appeared 1,142 times, "chairperson" 130 times, and "chairwoman" only 68 times. She suggested that "chairperson" is marked compared to the neutral term "chairman." In a similar study using the British National Corpus (BNC), Romaine found "lady doctor" 125 times, "woman doctor" 20 times, and "female doctor" 10 times, while "gentleman doctor" had no occurrences. "Man doctor" appeared once, and "male doctor" 14 times. "Family man" was found 94 times, but "family woman" only four times. Romaine explained these differences by pointing out societal views that associate careers more with men and mark women who pursue them (2000: 117).

In their comprehensive analysis of various language corpora, including the Brown Corpus, Lancaster-Oslo-Bergen Corpus (LOB), Wellington Corpus of New Zealand English (WWC), Freiburg-Brown Corpus of American English (Frown), and Freiburg-LOB (FLOB) Corpus of British English, Sigley and Holmes (2002) found a decline in the use of sexist terms like "-ess" and polite yet gendered words

like "lady/ladies" since the 1960s. They also observed a decrease in the general use of "man" and terms like "female lawyer." Interestingly, their study highlighted a doubling in the frequency of "woman" and "women," while the use of "man" and "men" significantly decreased. Notably, "woman" was used less frequently than "women."

Using a keyword and collocation approach, Mautner (2007) examined age-related stereotypes. Analyzing the Bank of English Corpus, she found that the term "elderly" appeared 75 times next to "woman," 35 times with "women," 37 times with "lady," and 7 times with "widow." In comparison, "man" was found 51 times with "elderly," while "gentleman" only appeared 13 times. Mautner concluded that "elderly" is a social, not just a chronological, label that connects women to discussions of care, disability, and illness.

Similarly, Pearce (2008) conducted a related study, showing that verbs indicating physical strength and assertion of power, such as "dig," "climb," "jump," "conquer," "dominate," and "lead," are mostly associated with "man." Notably, these collocations did not show similar connections with "woman." When used as objects, actions related to legality, like "apprehend," "arrest," and "convict," often occurred with "man," but not with "woman." Additionally, verbs related to victimhood, such as "kill," "wound," "knife," and "shoot," were linked to "man," while verbs like "assault," "gag," "rape," and "violate" tended to collocate with "woman."

Baker (2013) contributed to the study of gender-marked language by analyzing various corpora created between 1931 and 2006, including the Lancaster corpus (BLOB), LOB, FLOB, and British English 2006 (BE06). His diachronic analysis focused on male and female pronouns, words such as "man" and "woman," professions, and terms like "Mr." and "Ms." The findings showed a decrease in male pronouns, a slight increase in female pronouns, and a persistent gap between them. Inclusive terms ("him/her," "s/he," "he/she," "he or she," and "him or her") grew from 1961 to 1991. However, they dropped by over half in 2006, indicating a decline in usage. The term "spokesman" has been consistently used over the last 15 years (1, 22, 50, and 43 times across different corpora), while "spokeswoman" and "spokesperson" have appeared fewer times. In a more recent study, Moon (2014) analyzed English adjectives used to describe women and men, considering age; querying the Bank of English corpus revealed that the most common collocates of "young man/men" were handsome, nice, bright, tall, and angry. Other collocates appearing six or more times related to relationships (e.g., gay, single, married, lonely, bisexual) and physical traits (e.g., tall, thin, muscular, slender). For middle-aged woman/women, fewer adjectival collocates were found; some included single, lonely, plump, fat, stout, tall, thin, elegant, attractive, beautiful, bored, and healthy. Some of these collocates are positively evaluative, while others imply negative traits for women in midlife.

A recent contribution to this field is the study conducted by Krendel in 2020. This research analyzes data from an anti-feminist Reddit discussion forum. The study highlights widespread negative judgments against women, based on stereotypes related to their gender, including traits like selfishness and manipulation. Women are also dehumanized through comparisons to animals and machines, often being reduced to their physical appearance and how men perceive them. Additionally, the study shows the portrayal of girls exploiting and manipulating men, blaming them for their aggressive behaviors. In contrast, men are depicted as victims, experiencing mistreatment from both female social actors and external institutions. This portrayal fosters feelings of dissatisfaction and insecurity, especially within the Men's Rights sub-corpus.

As mentioned earlier, a common theme across the studies is the consistent linguistic bias favoring men in their portrayals compared to women. Building on this idea, the current research aims to determine whether these similar findings are reflected within the specific context of the Spanish language.

I am running a few minutes late; my previous meeting is running over.

## 2. METHOD

This study explores how women and men are represented linguistically in the Spanish press by analyzing the lemmas HOMBRE (MAN) and MUJER (WOMAN) in the Spanish News on the Web (NOW) corpus, as introduced by Davis (2019). The NOW corpus, which contains 7.6 billion words from online newspapers and magazines across twenty Spanish-speaking countries, serves as the main data source for this research. In linguistic theory, collocations show that a word's full meaning is not only in its form but also in its typical associations with other words or nearby structures (McEnery & Hardie, 2011: 123). By focusing on these word relationships, collocations are key to understanding language and its subtle conveyances, helping to reveal the underlying discourses that shape the portrayal of different groups (Taylor, 2021).

The mutual information (MI) statistic was used to examine how exclusive word pairings are. MI effectively identifies rare collocates that mostly occur with the node, even if these co-occurrences are uncommon throughout the entire corpus (Brezina, 2018: 71).

Figure 1 displays the NOW corpus interface used for parameter input. The analysis was performed with a one-space window span to the right, considering present and past tense verbs alongside each node. The search included web-based news articles and newspapers published between 2012 and 2018 in Latin American countries and Spain. Finally, only verbs with an MI score above three were retained for further analysis.

Figure 1 displays the interface where the lemma WOMAN was entered.

The screenshot shows the 'Corpus del Español: NOW' interface. The top navigation bar includes 'SEARCH', 'FREQUENCY', 'CONTEXT', and 'OVERVIEW'. The 'SEARCH' tab is active. Below the navigation bar, there are input fields for 'Word/Phrase [POS]' and 'Collocates'. The 'Word/Phrase' field contains 'Mujer\_N\*' and the 'Collocates' field contains 'vPRET (hablá)'. There are also buttons for 'Find collocates' and 'Reset'. Below these fields, there are checkboxes for 'Sections', 'Texts/Virtual', 'Sort/Limit', and 'Options'. The 'Sections' checkbox is checked. Below the 'Sections' checkbox, there are two lists: '1 Uruguay' and '2 -IGNORE-'. The 'List' tab is selected. On the right side of the interface, there is a 'SECTIONS' section with a 'SHOW' checkbox and a description of the sections. Below the 'SHOW' checkbox, there are two tables: one for 'ismo in 2016-2018' and another for 'ismo in 2016-18 vs 2012-15'. The first table has two rows: '\*ismo in 2016-2018' and 'NOUN de datos in 2016-2018', and '\*ismo in Venezuela' and 'NOUN DULCE in España'. The second table has two rows: '\*ismo in 2016-18 vs 2012-15' and 'NOUN de datos in 2016-18 vs 2012-2015', and '\*ismo in VE vs CO, MX, ES, AR' and 'NOUN DULCE in ES vs MX'. A note at the bottom states: 'Note: after clicking on a link above, you may need to click on SECTIONS in the search form to see this help file again.'

Figure 1. Corpus NOW interface. Parameters to search the verbal collocations. Source: NOW corpus.

Figure 2 displays the verbs that collocate with the lemma WOMAN. The data shows the specific verbs associated with the lemma and offers insight into their frequency of co-occurrence. A crucial point here is the MI score's role in identifying verbs that are exclusively linked to a particular lemma. This score highlights verbs that occur only in connection with the target lemma, emphasizing the unique nature of these linguistic relationships.



SEE CONTEXT: CLICK ON WORD OR SELECT WORDS + [CONTEXT] [HELP...]

|    | CONTEXT      | FREQ | ALL  | %     | Mi    |
|----|--------------|------|------|-------|-------|
| 1  | BOCINÓ       | 1    | 5    | 20.00 | 10.22 |
| 2  | DESHABITÓ    | 1    | 8    | 12.50 | 9.54  |
| 3  | AMAESTRÓ     | 1    | 13   | 7.69  | 8.84  |
| 4  | NARCOTIZÓ    | 2    | 26   | 7.69  | 8.84  |
| 5  | EMBADURINÓ   | 4    | 63   | 6.35  | 8.57  |
| 6  | CONTRABANDEÓ | 3    | 65   | 4.62  | 8.11  |
| 7  | GRUÑÓ        | 6    | 140  | 4.29  | 8.00  |
| 8  | EMPALIDECIÓ  | 1    | 26   | 3.85  | 7.84  |
| 9  | SENTISTEIS   | 1    | 26   | 3.85  | 7.84  |
| 10 | POLOLEÓ      | 2    | 53   | 3.77  | 7.82  |
| 11 | YACIERON     | 1    | 29   | 3.45  | 7.69  |
| 12 | TACONEÓ      | 1    | 32   | 3.13  | 7.54  |
| 13 | MONETIZÓ     | 1    | 32   | 3.13  | 7.54  |
| 14 | TIMÓ         | 4    | 149  | 2.68  | 7.32  |
| 15 | ESCULCÓ      | 1    | 40   | 2.50  | 7.22  |
| 16 | ERUCTÓ       | 1    | 42   | 2.38  | 7.15  |
| 17 | VOLANTEÓ     | 4    | 176  | 2.27  | 7.08  |
| 18 | CASTRÓ       | 6    | 269  | 2.23  | 7.06  |
| 19 | TAJEÓ        | 1    | 46   | 2.17  | 7.02  |
| 20 | FINGIÓ       | 81   | 4072 | 1.99  | 6.89  |
| 21 | CURVÓ        | 1    | 51   | 1.96  | 6.87  |
| 22 | FORCEJEÓ     | 62   | 3192 | 1.94  | 6.86  |
| 23 | AGONIZÓ      | 18   | 930  | 1.94  | 6.85  |
| 24 | RAPTÓ        | 20   | 1149 | 1.74  | 6.70  |
| 25 | ENSILLÓ      | 1    | 60   | 1.67  | 6.64  |
| 26 | ACOMPRASÓ    | 1    | 61   | 1.64  | 6.61  |
| 27 | LEVITÓ       | 1    | 61   | 1.64  | 6.61  |
| 28 | AMAMANTÓ     | 7    | 441  | 1.59  | 6.57  |

Figure 2. Verbal Collocations of the lemma MUJER (WOMAN). Source: NOW corpus.

A total of 3,200 verbal collocations were extracted from the NOW corpus. The next step involved classifying these verbs, a process facilitated by using the ADESSE taxonomy, as outlined by García-Miguel and Albertuz (2005). This classification not only expanded the scope of the linguistic analysis but also provided a framework for a detailed examination of how men and women are portrayed in discourse.

To visualize and understand the complex data network, this study used Gephi, an open-source software tool developed by Bastian et al. (2009). The resulting graphs, known as constellation networks, served as valuable tools with two main purposes. First, they helped identify key adjectives related to each lemma, providing insights into the dominant semantic contexts. Second, these visual representations enabled detailed comparisons and connections among the verbs. This comparative approach was effective in revealing underlying similarities and differences.

### 3. RESULTS AND DISCUSSION

The main goal of the collocational analysis is to examine how the lemmas HOMBRE (MAN) and MUJER (WOMAN) function in collocations within the NOW Corpus. As noted, 3,200 verbal collocations with a mutual information score above three have been identified. It is important to highlight that due to the large number of collocates, a selection process was used to display only a subset. This was done to enhance the clarity of the network visuals, making the connections between the lemmas and collocates clearer and easier to understand.

These extracted verbal collocations were methodically organized using the ADESSE verb classification system. This ongoing project provides a detailed framework that covers the different semantic categories a verb can belong to. It recognizes that some verbs may fit into multiple categories. After classifying the verbs, further refinement involved placing them into appropriate semantic fields whenever possible. Figure 3 shows the network of verbs assigned to the Communication category.

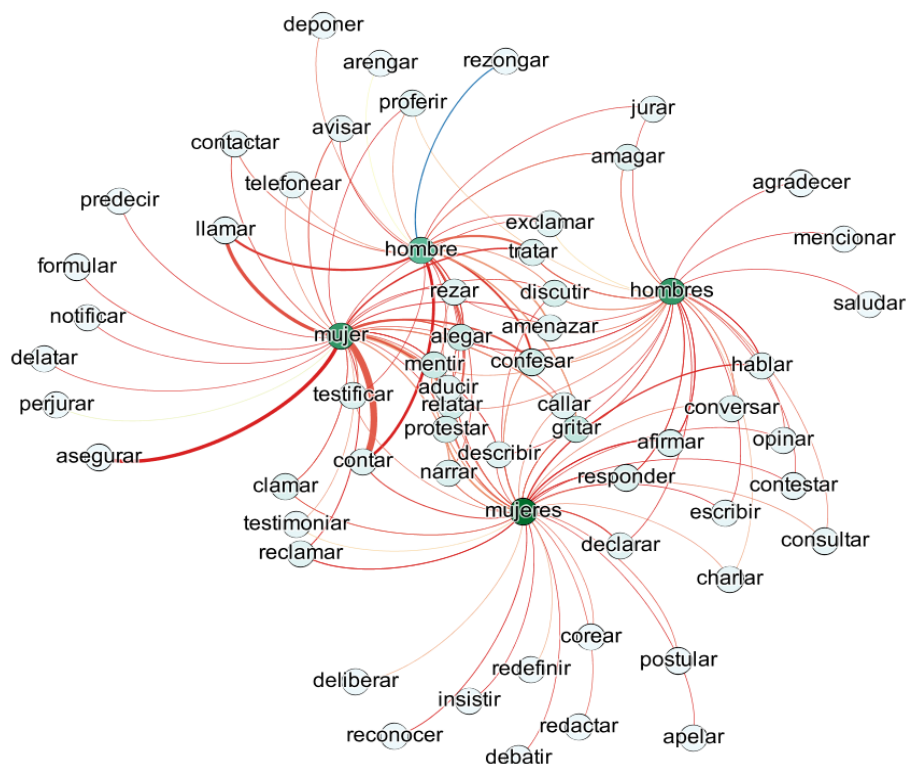


Figure 3. Constellation network of the category of Communication. Source: own elaboration.

The lemma MAN forms exclusive collocations with verbs like "saludar" (to greet), "mencionar" (to mention), "agradecer" (to thank), "amagar" (to threaten), "jurar" (to swear), "rezongar" (to grumble), "deponer" (to testify), and "exclamar" (to exclaim), all of which are associated exclusively with men. Conversely, verbs such as "postular" (to postulate), "apelar" (to appeal), "corear" (to chant), "redefinir" (to redefine), "insistir" (to insist), "debatir" (to debate), "reconocer" (to acknowledge), "deliberar" (to deliberate), "reclamar" (to complain), "testimoniar" (to testify), "asegurar" (to assure), "perjurar" (to perjure), "delatar" (to blow the whistle), "notificar" (to notify), "formular" (to formulate), "predecir" (to predict), and "clamar" (to clamour) are exclusively associated with women.

Among these verbs, a subset belongs to a semantic field called "Emotional Vocal Expression." This includes verbs like "rezongar" (to grumble) and "exclamar" (to exclaim), which are associated with MAN, and verbs like "apelar" (to appeal), "corear" (to chant), "insistir" (to insist), "debatir" (to debate), "reclamar" (to complain), and "clamar" (to call out), which carry the connotation of emotional verbal or vocal expression, at least in Spanish. Most of these are linked with WOMAN.

One verb, "perjurar" (to perjure), has the highest MI score and is exclusively associated with women. Conversely, verbs like "postular" (to postulate), "redefinir" (to redefine), and "reconocer" (to recognize), as well as "deliberar" (to deliberate), "testificar" (to testify), "notificar" (to notify), and "formular" (to formulate), indicate a standard way of communicating information and are also exclusively associated with women. In contrast, verbs like "discutir" (to argue), "mentir" (to lie), "amenazar" (to threaten), and "gritar" (to yell) are more strongly linked to men, while "protestar" (to protest), "callar" (to be silent), and "alegar" (to argue) are more common with women. Notably, the phrase "be silent" ('las mujeres callan') exclusively pairs with women, suggesting that gender roles are often attributed to the whole community rather than individual members, which can lead to deindividuation. This also highlights the concept of "othering," emphasizing the division between "us" and "them." Such linguistic representations of gender reinforce the idea of a binary gender dichotomy that does not account for the diversity within the LGBTQ+ community.

This research aims to show how Corpus Linguistics techniques, especially collocations, can improve qualitative studies. Although this thesis does not focus on gender studies or discourse analysis, I will include a few paragraphs to explore the idea of "othering" and its link to gender representation. According to Jensen (2011), "othering" assumes that people without power are often cast as "the other" in discourse, reinforcing the authority and superiority of the powerful and shaping the identity of subordinates (p. 65). Seeing men as the norm and women as the "other" leads men to define and describe women, which continues their objectification and marginalization. A collocational analysis can help reveal how women are otherized and objectified. Additionally, the process of differentiation and demarcation shown by the collocational analysis in this study helps visualize how language stereotypes women and men.

Below is the constellation network showing the collocations within the Life category. Unlike other categories, the Life category has a relatively small number of collocations, with most verbs tending to collocate with both lemmas.

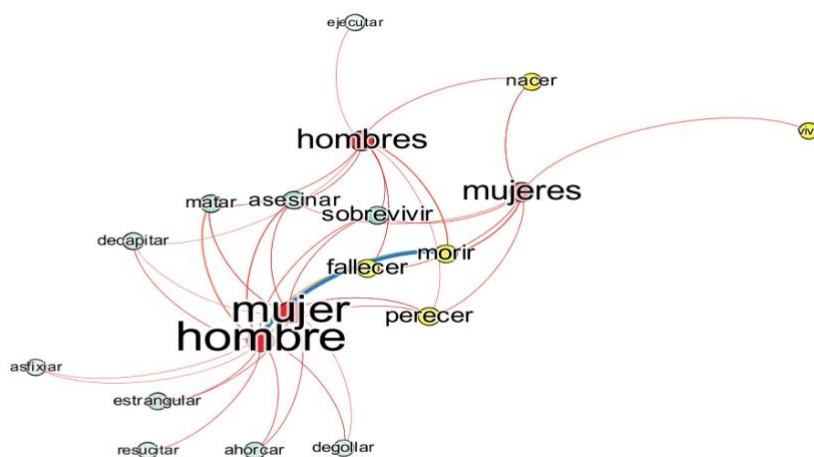


Figure 4. Constellation network of the category of Life. Source: own elaboration.

The verb "vivir" (to live) aligns more strongly with the lemma WOMAN, while "ejecutar" (to execute) and "resucitar" (to resurrect) align more closely with the lemma MAN. Notably, only "to execute" carries a negative connotation. Other verbs like "matar" (to kill), "decapitar" (to behead), "asesinar" (to murder), "estrangular" (to strangle), "degollar" (to behead), "asfixiar" (to suffocate), and "ahorcar" (to hang) also have negative connotations and are associated with both genders; however, these verbs are more strongly linked to MAN. For example, "matar" (to kill) shows a stronger association with MAN (MI score of 7.21) than WOMAN (MI score of 5.22). Similarly, "degollar" has an MI score of 8.33 with MAN and 5.98 with WOMAN. These findings are consistent with prior research in other languages like English, which highlights a stronger connection between men and acts of violence.

Additionally, regarding non-violent verbs, MAN is closely associated with "morir" (to die), while WOMAN is more strongly linked to "fallecer" (to pass away). This is interesting because "to die" is considered a neutral verb, simply indicating the end of life without emotional nuance. Conversely,

"to pass away" is more commonly used in obituaries and situations where a more empathetic tone is preferred. This verb tends to relate more to religious concepts and suggests a gentle transition from life to death. It's used to show respect and avoid sounding blunt, often carrying a more emotional tone than "to die." However, this can vary depending on individual and cultural differences. Although both verbs refer to the same concept, they carry different connotations and are used differently depending on the context, personal preference, and situation. This lexical analysis offers an opportunity to expand the investigation by examining these two lemmas not only as subjects, as shown in this study, but also as objects. By analyzing the recipients of actions described by subjects in online news press, we can gather insights into both perpetrators of violence and victims. This expanded approach has the potential to deepen our understanding of gender representation and language use in the context of violence. The analysis of verbs within the "Life" category demonstrates how collocational analysis can be extended to explore gender portrayal.

The following constellation network examines the results within the Competition category. This network includes verbs that indicate competition, extending beyond sports to various life situations. Figure 5 provides a visual representation of the verbs related to competition. A few verbs are used exclusively with either MAN or "WOMAN." However, the presence of antonyms among these verbs adds interesting significance to the findings.

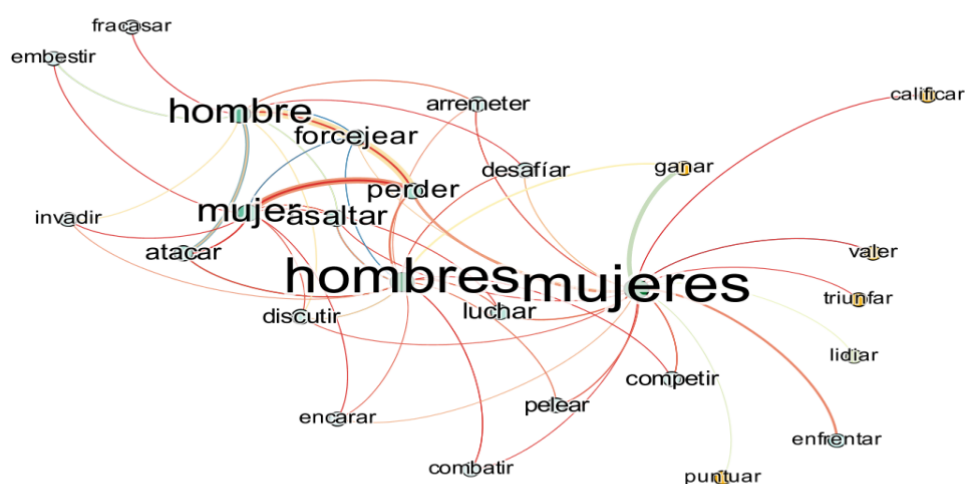


Figure 5. Constellation network of the competition category. Source: own elaboration.

The verb "fracasar" (to fail) only collocates with "man," while "triunfar" (to succeed), "calificar" (to qualify), "valer" (to be worth), "lidar" (to deal with), and "enfrentar" (to cope with) only collocate with "woman."

Interestingly, the verb "to fail" is only associated with MAN, while "to succeed" is linked exclusively to WOMAN. In this context, it is important to revisit the concept of "markedness" to understand the difference between marked and unmarked terms. Marked terms have a less common or more specific meaning, whereas unmarked terms are more widely understood. Adding a linguistic particle, such as

a verb, to a subject (man, woman) alters the meaning of a marked term. Regarding the verbs in the Competition category, the verb "to fail" only appeared with MAN.

Analyzing collocations, like in this study, helps us identify both common and less common expressions, which reveal gender representations embedded in language.

It's worth noting that the verbs "to deal with" and "to cope with," which describe facing challenges, only collocate with WOMAN. Interestingly, no similar verb pattern was found with MAN. This unique collocational pattern suggests that pairing these verbs with WOMAN might imply a link between women and handling difficulties. Conversely, the lack of comparable verb collocations with "man" could suggest a connection between men and problem-solving skills.

Several verbs, such as "luchar" (to struggle), "combatir" (to combat), "desafiar" (to challenge), "competir" (to compete), and "enfrentar" (to face), show collocational ties with both genders. Notably, all these verbs except "to combat" have a much stronger connection to women. This interesting pattern suggests women are more often depicted in situations involving struggle or challenge.

The analysis of verb collocations with the lemmas MAN and WOMAN shows that certain verbs are more closely linked to one gender, emphasizing how language use and gender are connected. This type of analysis can strengthen qualitative studies and help illustrate how women and men are stereotyped through language. The concept of markedness has been used in linguistic analysis to explore gender representation and other social meanings. Pairing specific verbs with gendered subjects reveals underlying assumptions about success, failure, and gender roles. The exclusive association of "to fail" with MAN questions the idea of male success and challenges the norm.

Conversely, linking "to succeed" with WOMAN highlights notable achievements in this context. This analysis also emphasizes the importance of recognizing key expressions to reveal subtle gender representations embedded in language. The exclusive pairing of "to deal with" and "to cope with" with "woman," along with the absence of a similar pattern with "man," raises interesting implications. It suggests a possible association between women and their ability to handle challenges. Similarly, the distinctive pattern of verbs like "luchar" (to struggle), "desafiar" (to challenge), and "enfrentar" (to face), which favor women, underscores their portrayal in stories of resilience and perseverance. These findings show how language both influences and reflects societal attitudes toward gender roles, enriching our understanding of gender dynamics from a linguistic perspective.

#### 4. CONCLUSION

This collocational study helps us identify potential discourses and social practices observed in press studies, expanding the discussion on gender to include not just men and women but also discourses involving the LGBT community. This type of research allows us to examine different aspects, which are briefly summarized below. The corpus for this lexical analysis was created from news articles collected from various online media outlets across Latin American countries and Spain. This enables us to explore how the media influences specific linguistic associations that shape, reinforce, and sustain gender stereotypes through language choices in news reports. It raises questions about the media's responsibility and role in shaping societal perceptions. An important aspect—and a limitation of this study—is that this approach gathers information from diverse contexts, allowing cross-cultural comparisons. Essentially, studies like this can help determine whether similar gender associations exist in other parts of the Spanish-speaking world, including regions with different Spanish dialects and cultural backgrounds. This helps establish whether certain linguistic patterns are universal within the language or more specific to particular cultural environments. By analyzing the collocations of the lemmas "MAN" and "WOMAN," we can identify linguistic gender bias, as some verbs (actions) are associated with a specific gender, reinforcing societal stereotypes and roles. The results may lead

to discussions about promoting gender-neutral language or raising awareness of biases embedded in language. This aligns with broader conversations on language reform and how it might influence societal attitudes and roles. For instance, verbs related to communication and emotional expression could reflect societal expectations about how men and women communicate or how emotions are expressed and perceived based on gender. It also encourages dialogue on how men, women, and other community members are socialized to express their feelings. The lexical analysis in this study identified associations of lemmas with verbs that signify authority and power, such as "demandar" (to demand), "postular" (to apply), and "debatir" (to debate). This prompts reflection on how language reveals power dynamics aimed at establishing or maintaining dominance through linguistic choices.

Additionally, this kind of analysis promotes discussions about how linguistic choices connect to factors like race and sexuality. Finally, as this article emphasizes, using corpus linguistics techniques offers a way to examine societal discourses. Analyzing the collocational patterns within these discourses helps us deepen our understanding of the complex relationship between language, gender, and society.

## References

- Alochis, I. 2016. La representación de la violencia sexual contra las mujeres y las niñas en el léxico y en la construcción discursiva de las noticias en un estudio de caso en la prensa gráfica de Córdoba. Coloquio interdisciplinario internacional "Educación sexualidades y relaciones de género", Universidad Nacional de Córdoba, Argentina. Retrieved from [https://rdu.unc.edu.ar/bitstream/handle/11086/20777/EJE\\_2.pdf](https://rdu.unc.edu.ar/bitstream/handle/11086/20777/EJE_2.pdf)
- Baker, P. 2010. *Sociolinguistics and corpus linguistics*. Edinburgh University Press.
- Baker, P. 2012. Corpora and gender studies. In K. Hyland, C. Meng Huat, & M. Handford (eds.), *Corpus Applications in Applied Linguistics*, 100-116. London: Bloomsbury Publishing. [https://doi.org/10.1007/978-3-319-41733-2\\_14](https://doi.org/10.1007/978-3-319-41733-2_14)
- Baker, P. 2010. 'Will Ms ever be as frequent as Mr? A corpus-based comparison of gendered terms across four diachronic corpora of British English. *Gender and Language* 4.1. 125-149. <https://doi.org/10.1558/genl.v1.i1.17188>
- Baker, P. 201. *Using corpora to analyze gender*. London: Bloomsbury Academic. <https://doi.org/10.1017/s0047404515000093>
- Baker, P., & Brookes, G. 2021. Lovely nurses, rude receptionists, and patronising doctors. In J. Angouri & J. Baxter (eds.), *The Routledge Handbook of Language, Gender, and Sexuality* 559-571. New York: Routledge. <https://doi.org/10.4324/9781315514857-45>
- Bastian, M., Heymann, S., & Jacomy, M. 2009. Gephi: An open source software for exploring and manipulating networks. *Proc. International AAAI Conference on Weblogs and Social Media*, 361-362. <https://doi.org/10.1609/icwsm.v3i1.13937>
- Brezina, V. 2018. *Statistics in corpus linguistics: A practical guide*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781316410899>
- Caldas-Coulthard, C. R., & Moon, R. 2010. 'Curvy, hunky, kinky': Using corpora as tools for critical analysis. *Discourse & Society*, 21.2. 99-133. <https://doi.org/10.1177/0957926509353843>
- Castillo, M. N. 2019. ¿Qué se dice de la mujer y el hombre en el español de Chile?: estudio exploratorio de las combinaciones frecuentes de los vocablos mujer y hombre en un corpus de referencia estratificado. *Boletín de filología*, 54.1. 95-117. Retrieved from <https://boletinfilologia.uchile.cl/index.php/BDF/article/view/53530>

- Chuaikun, D., & Wijitsopon, R. 2021. A corpus-based study of LGBT-related news discourse in Thailand's and international English-language newspapers. *Applied Linguistics Review* .0. 000010151520210036. <https://doi.org/10.1515/applirev-2021-0036>
- Coates, J. 2015. *Women, Men and Language: A sociolinguistic account of gender differences in language*. London: Routledge. <https://doi.org/10.4324/9781315645612>
- Davies, M. 2019. The corpus del Español NOW (news on the web). <http://www.corpusdelespanol.org/now>.
- Eagly, A. H., & Wood, W. 2012. Social role theory. *Handbook of theories of social psychology* 2. 458-476.
- Flowerdew, J., & Richardson, J. E. (eds.). 2017. *The Routledge Handbook of Critical Discourse Analysis*. New York: Routledge. <https://doi.org/10.4324/9781315739342>
- García-Miguel, J., & Albertuz, F. 2005. Verbs, semantic classes, and semantic roles in the ADESSE project. In *Proceedings of the Interdisciplinary Workshop on the Identification and Representation of Verb Features and Verb Classes* 50-55. University of Vigo.
- Heritage, F. 2020. Applying corpus linguistics to videogame data: Exploring the representation of gender in videogames at a lexical level. *Game studies* 20.3. 1-48.
- Horton, R. H. 2018. A corpus analysis of girls and boys in spoken academic English and teaching activities to raise awareness about gendered discourse. *TESOL Working Paper Series* 16. 1-18.
- Hyland, K., Chau, M. H., & Handford, M. 2012. *Corpus applications in applied linguistics*. London: Bloomsbury Publishing.
- Jensen, S. Q. 2011. Othering, identity formation, and agency. *Qualitative Studies* 2.2. 63-78. <https://doi.org/10.7146/qs.v2i2.5510>
- Karimullah, K. 2020. Sketching women: a corpus-based approach to representations of women's agency in political Internet corpora in Arabic and English. *Corpora* 15.1. 21-53. <https://doi.org/10.3366/cor.2020.0184>
- Kendall, S., & Tannen, D. 2015. Discourse and Gender. In D. Schiffrin, D. Tannen, & H. Hamilton (eds.), *The Handbook of Discourse Analysis*, 639-660. Malden, MA: Wiley Blackwell. <https://doi.org/10.1002/9781118584194.ch30>
- Krendel, A. 2020. The men and women, guys and girls of the 'manosphere': A corpus assisted discourse approach. *Discourse & Society* 31.6. 607-630. <https://doi.org/10.1177/0957926520939690>
- Litosseliti, L. 2014. *Gender and language theory and practice*. London: Routledge.
- Mautner, Gerlinde. 2007. Mining large corpora for social information: The case of elderly. *Language in Society* 36.1. 51-72. <https://doi.org/10.1017/S0047404507070030>
- McEnery, A., & Baker, P. (Eds.). 2015. *Corpora and discourse studies: Integrating discourse and corpora*. Lancaster: Palgrave.
- McEnery, T., & Hardie, A. 2011. *Corpus linguistics: Method, theory and practice*. Cambridge University Press.
- Moon, R., .2014. From gorgeous to grumpy: Adjectives, age, and gender. *Gender & Language* 8 .1. 5 -41. <https://doi.org/10.1558/genl.v8i1.5>

- Norberg, C., & Johansson, M. 2021. "Women and 'ideal' women": The representation of women in the construction industry. *Gender Issues* 38.1. 1-24. <https://doi.org/10.1007/s12147-020-09257-0>
- Norberg, C. 2016. Naughty boys and sexy girls: The representation of young individuals in a web-based corpus of English. *Journal of English Linguistics* 44.4. 291-317. <https://doi.org/10.1177/0075424216665672>
- Pearce, M. 2008. Investigating the collocational behavior of man and woman in the BNC using Sketch Engine. *Corpora* 3.1. 1-29. <https://doi.org/10.3366/E174950320800004X>
- Romaine, S. 2000. *Language in society: An introduction to sociolinguistics*. OUP.
- Sigley, R., & Holmes, J. 2002. Looking at girls in corpora of English. *Journal of English Linguistics* 30.2. 138-157. <https://doi.org/10.1177/007242030002004>
- Taylor, C. 2021. Investigating gendered language through collocation. In J. Angouri & J. Baxter (eds.), *The Routledge Handbook of Language, Gender, and Sexuality*, 572-586. London: Routledge. <https://doi.org/10.4324/9781315514857-46>
- Tsapro, G., & Semeniuk, A. 2021. Discursive presentation of women in the 2001-2002 and 2020-2021 media texts: A corpus-based analysis. *Studia Philologica* 1. <https://doi.org/10.28925/2311-2425.2021.168>
- Weatherall, A. 2005. *Gender, language, and discourse*. New York: Routledge.
- West, C., & Zimmerman, D. H. 1987. Doing gender. *Gender & Society* 1.2. 125-151. <https://doi.org/10.1177/0891243287001002002>
- Widdowson, H. G. 2004. *Text, context, pretext: Critical issues in discourse analysis*. Oxford: Blackwell.