

Análisis de las colocaciones verbales de los lemas HOMBRE y MUJER en español en el corpus NOW

*An analysis of the verbal collocations of the lemmas MAN and WOMAN
in Spanish in the Now corpus*

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 “lidiar” 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 uses corpus linguistics techniques to investigate how men and women are represented in online news in Spanish. The research explores the collocational patterns of verbs with the lemmas MAN and WOMAN within the NOW corpus. The findings indicate that certain verbs, like “to argue,” “to lie,” “to threaten,” and “to yell,” are more closely associated with men, while verbs like “to protest,” “to be silent,” and “to argue” are more linked to women. These patterns reflect gender dynamics in language and align with research in English that shows a stronger association between men and violent actions. On the other hand, the verb “to die” appears more frequently with MAN, and WOMAN is more commonly paired with “to pass away,” which, unlike “to die,” carries emotional and respectful nuances.

Additionally, the study reveals the exclusive collocation of “to fail” with MAN and “to succeed” with WOMAN, shedding light on achievement expectations for both genders.

Notably, verbs like “to deal with” and “to face” occur uniquely with WOMAN, suggesting a connection between women and handling difficult situations. This lexical analysis underscores the intricate interplay between language, gender, and societal perceptions, emphasizing the need for ongoing study and research to uncover and challenge gender stereotypes and power dynamics embedded in language.

Keywords: collocations, gender, corpus, Spanish

1. INTRODUCTION

This article aims to achieve two objectives. Firstly, it delves into applying collocation techniques from corpus linguistics to unveil gender representations within the digital realm of journalism. Secondly, it adds to the existing work centered around using corpus linguistics to extract social and cultural insights from Spanish electronic corpora. The investigation scrutinizes the portrayal of both women and men in Spanish online news, which has received comparatively less scholarly attention in Spanish when contrasted with English.

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

Gender expectations reflect prevailing social norms and values upheld and reinforced through everyday interactions across various contexts. Exploring how gender manifests in social situations can provide insights into the supportive framework of social structure and the mechanisms of social control that sustain it (West & Zimmerman, 1987). This inquiry draws upon language as a potent instrument, reflecting people's understanding of the world and perpetuating and strengthening their perceptions of the phenomena and entities surrounding them (Baker, 2014).

The verbal collocational analysis of the lemmas HOMBRE (MAN) and MUJER (WOMAN) delves into a pertinent societal issue: the prevalence of violence. This issue casts its shadow across various facets of society, giving rise not only to physical acts of violence but also permeating the psychological and social realms. This insidious presence of violence finds expression in the beliefs and ideologies upheld by specific segments of society, thereby normalizing mistreatment and brutality. As a consequence, language becomes a conduit for perpetuating justifications for violence; a phenomenon propagated not just by ordinary individuals but also by the media and societal institutions.

1.1 Corpus linguistics and gender

Over the past two decades, Corpus Linguistics (CL) has significantly influenced the field of language and gender research. Initially, studies in this area relied on limited data sets. Still, CL has revolutionized the approach by utilizing vast amounts of data comprising millions of words and sophisticated corpus query tools. The benefits of CL have been recognized by various disciplines, including translation studies, discourse analysis, applied linguistics, and sociolinguistics, which have integrated them into their research methodologies (Baker, 2010). Moreover, corpus analysis has applications in diverse domains such as forensic linguistics, language teaching, media studies, gender studies, and academic and workplace discourses (Hyland et al., 2012). However, critics have raised concerns about the potential drawbacks of relying heavily on quantification through CL. There is a fear that this approach may oversimplify complex language phenomena, perpetuate stereotypes, or reinforce prejudice. It is important to note, though, that CL can coexist with existing research paradigms rather than replace them entirely (Baker, 2010: 9). In its essence, Corpus Linguistics stands as a valuable complement, enhancing the landscape of linguistic inquiry. McEnery and Baker (2015) support this idea by pointing out that researchers typically commence their investigations with quantitative methods, utilizing techniques like keyword analysis and collocate examination to steer their course. However, as their research unfolds, a natural progression emerges, leading to a qualitative and context-driven approach. This transition, in turn, results in a decreased reliance on computer software for analysis (p. 2). Corpus Linguistics has played a pivotal role in shaping the language and gender research landscape. By leveraging vast amounts of data and powerful tools, CL has opened up new possibilities in linguistic inquiry across various fields. While concerns about over-quantification persist, integrating CL alongside traditional methods can enrich research outcomes, offering a more comprehensive understanding of language patterns and behaviors.

Research into gender representation using corpus linguistics continues to be an intriguing subject. The utilization of CL techniques has extended to various domains, ranging from politics and video games to professional settings, education, sexuality, and even the sphere of 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) accentuate the multifaceted advantages of corpus linguistics on gender analysis. Primarily, corpora's inherent comprehensiveness and representativeness, when juxtaposed with smaller text datasets, provide researchers with a heightened ability to assert broader generalizability in their findings concerning language patterns. Secondly, this approach unveils the significance of subtle linguistic nuances that may be considered in isolation but gain prominence when scrutinized within a broader discourse or textual context. Furthermore, larger corpora facilitate the seamless integration of a diverse spectrum of discourse perspectives on a specific subject, thus adeptly capturing resistant or minority viewpoints.

Additionally, corpus linguistics introduces a crucial element of objectivity into discourse analysis through two pivotal principles. The first principle underscores avoiding systematic bias in selecting corpus texts, refraining from excluding texts solely to conform to pre-existing arguments. The second principle underscores an unwavering commitment to absolute accountability for all collected data. These principles collectively empower discourse analysts in effectively addressing allegations of cherry-picking texts or instances

that bolster predetermined viewpoints, as posited by Widdowson (2004, as cited in Baker & Brookes, 2021).

1.2 Previous research

Collocational analyses exploring the linguistic portrayal of women and men in Spanish have been relatively underexplored, with only a handful of related studies identified (Alochis, 2016; Castillo, 2019). This starkly contrasts the extensive body of research dedicated to the linguistic depiction of gender in English. Language and gender studies have a relatively long history, with research in this area ranging from attributing differences in language use between women and men to linguistic deficits on the part of women to differences attributed to the idea that gender is a social construct (Weatherall, 2005; Litosseliti, 2014; Coates, 2015; Kendall & Tannen, 2015; Flowerdew & Richardson, 2017;). While a significant amount of linguistic research has adopted a descriptive and qualitative approach to analyze how men and women use language, in the last decade, there has been extensive research in CL addressing how women and men are linguistically represented in different corpora (Caldas-Coulthard & Moon, 2010; Baker, 2012, 2013; Moon, 2014; McEnery & Baker, 2015). However, gender linguistic representation has been conducted mainly in English, relatively unexplored in Spanish. Therefore, this study aims to expand research in this area by using tools and techniques from CL to explore gender representation in Spanish.

Romaine (2000) used the Brown Corpus to analyze neutral terms like "chairperson." Her findings showed that "chairman" appeared 1,142 times, "chairperson" 130 times, and "chairwoman" just 68 times. She suggested that "chairperson" was marked as compared to the neutral term "chairman." In a similar study with 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 instances. "Man doctor" appeared once, and "male doctor" appeared 14 times. "Family man" occurred 94 times, but "family woman" appeared only four times. Romaine explained these differences by highlighting societal views associating careers more with men and marking women who pursued them (2000: 117).

In their extensive 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."

Mautner (2007) investigated age-related stereotypes through a keyword and collocation methodology. Analyzing the Bank of English Corpus, she discovered that the term "elderly" co-occurred 75 times with "woman," 35 times with "women," 37 times with "lady," and seven times with "widow." In contrast, "man" co-occurred 51 times with "elderly," while "gentleman" had just 13 instances. Mautner concluded that "elderly" is a societal, not just a chronological label, linking women to discourses of care, disability, and illness.

Similarly, Pearce (2008) conducted a related study focusing primarily on three grammatical relations: MAN/WOMAN as subject, MAN/WOMAN as object, and attributive adjectives

associated with MAN/WOMAN. His investigation unveiled that verbs denoting physical strength and assertion of power, such as "dig," "climb," "jump," "conquer," "dominate," and "lead," were predominantly affiliated with "man," with the notable absence of similar associations with "woman." Furthermore, when these verbs were employed as objects, legal-related actions like "apprehend," "arrest," and "convict" were observed to co-occur with "man," but not with "woman." Moreover, verbs related to victimhood, such as "kill," "wound," "knife," and "shoot," were linked predominantly to "man," while verbs like "assault," "gag," "rape," and "violate" were found to collocate more frequently with "woman."

Baker (2013) contributed by studying gender-marked language using various corpora constructed between 1931 and 2006: Lancaster corpus (BLOB), LOB, FLOB, and British English 2006 (BE06). His diachronic analysis focused on male and female pronouns, words like "man," "woman," professions, and terms like "Mr." and "Ms." Findings indicated a decrease in male pronouns, a slight rise in female pronouns, and a persistent gap between them. Inclusive terms ("him/her," "s/he," "he/she," "he or she," and "him or her") increased from 1961 to 1991. Still, they dropped by over half in 2006, suggesting fading use. The term "spokesman" was consistently used over the last 15 years (1, 22, 50, 43 times in different corpora), while "spokeswoman" and "spokesperson" appeared fewer times. In a more recent study, Moon (2014) analyzed English adjectives used to describe women and men considering the age factor; by querying the Bank of English corpus, the results showed that the most common collocates of young man/men were handsome, nice, bright, tall and angry. Other collocates that occurred six or more times were related to the relationship domain (gay, single, married, lonely, bisexual, etc.) and physical (tall, thin, muscular, slender, etc.). Concerning the collocations of middle-aged woman/women, there were fewer adjectival collocates; some of these were single, lonely, plump, fat, stout, tall, thin, elegant, attractive, beautiful, bored, and healthy, among others; some of these collocates are positive in evaluative orientation, but there are some that seem to suggest negative traits for women in midlife.

A recent contribution to this field is the study conducted by Krendel in 2020. This research draws its data from an anti-feminist Reddit discussion forum. The study accentuates the prevalent negative judgments directed towards women, rooted in stereotypes attributed to their gender, encompassing traits like selfishness and manipulation. Women also endure dehumanization through comparisons to animals and machinery, often being reduced to their physical appearance and how men perceive them. Furthermore, the study uncovers the portrayal of girls exploiting and manipulating men, placing blame on them for men's aggressive actions. In contrast, men are depicted as victims, subjected to mistreatment by both female social actors and external institutions. This portrayal fosters discontent and insecurity, particularly in the Men's Rights sub-corpus.

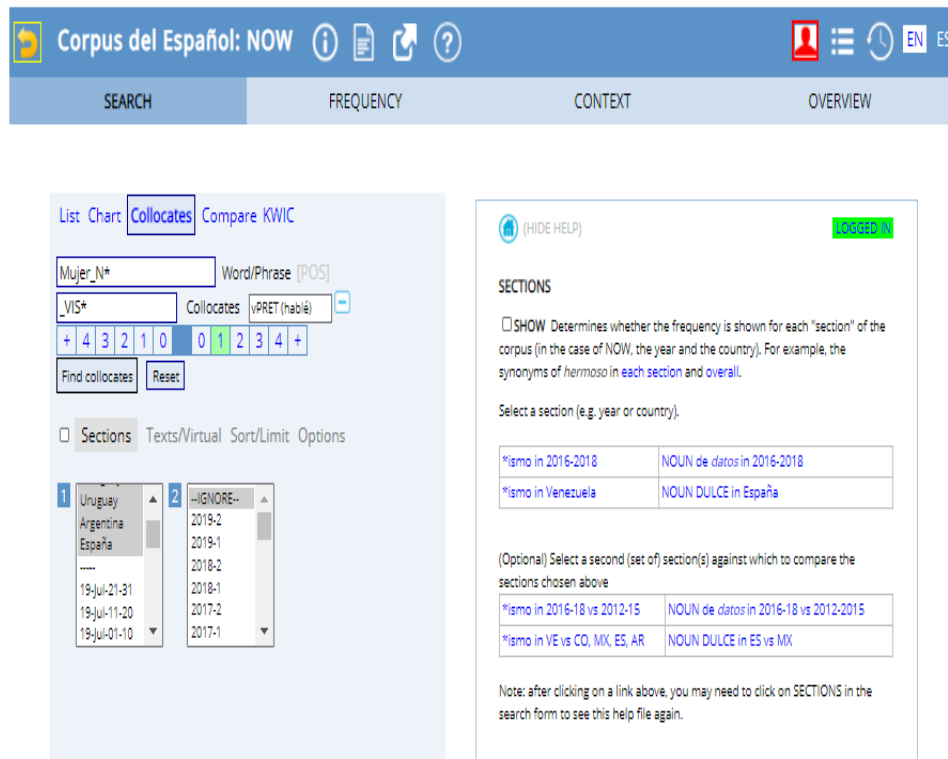
A prevailing theme among the studies mentioned above is the consistent linguistic bias favoring men in their portrayals relative to women. Expanding upon this premise, the present research seeks to delve into whether these analogous findings will be mirrored within the specific context of the Spanish language.

2. METHOD

This study examines linguistic representations of women and men 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, comprising 7.6 billion words from web-based newspapers and magazines across twenty Spanish-speaking nations, serves as the primary data source for this investigation. In linguistic theory, collocations signify that a word's complete meaning does not reside solely within its isolated form but rather emerges from its inherent associations with other words or structures frequently found in proximity (McEnery & Hardie, 2011: 123). By shedding light on these word associations, collocations play a pivotal role in understanding language and its nuanced expressions, thereby aiding in uncovering the underlying discourses that shape the representation of distinct groups (Taylor, 2021).

The statistical measure of mutual information (MI) was employed to investigate the degree of exclusivity between word pairings. MI effectively highlights the rare collocates appearing almost exclusively in conjunction with the node, even if this co-occurrence is infrequent within the entire corpus (Brezina, 2018: 71).

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



The screenshot shows the 'Corpus del Español: NOW' interface. The search parameters are set to 'Mujer_N*' and 'vPRET(habíe)'. The search results are displayed in a table with columns for sections and collocates. The interface includes a navigation bar with 'SEARCH', 'FREQUENCY', 'CONTEXT', and 'OVERVIEW'. The search results are displayed in a table with columns for sections and collocates.

SECTION	COLLOCATE
*ismo in 2016-2018	NOUN de datos in 2016-2018
*ismo in Venezuela	NOUN DULCE in España
*ismo in 2016-18 vs 2012-15	NOUN de datos in 2016-18 vs 2012-2015
*ismo in VE vs CO, MX, ES, AR	NOUN DULCE in ES vs MX

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

Figure 2 shows the verbs that collocate with the lemma WOMAN. The displayed data unveils the specific verbs linked with the lemma and offers insight into the frequency of their co-occurrences. An essential consideration here is the MI score's role in pinpointing verbs exclusively associated with the given lemma. This score highlights those verbs that occur solely in conjunction with the target lemma, emphasizing the distinctive nature of these linguistic partnerships.

	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	EMBADURNÓ	4	63	6.35	8.57
6	CONTRABANDEÓ	3	65	4.62	8.11
7	GRUÑÓ	6	140	4.29	8.00
8	EMPALIDECÓ	1	26	3.85	7.84
9	SENTISTEIS	1	26	3.85	7.84
10	POLOLEÓ	2	53	3.77	7.82
11	YACERON	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	ACOMPASÓ	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 subsequent step involved classifying these verbs, a process facilitated by the utilization of the ADESSE taxonomy, as outlined by García-Miguel and Albertuz (2005). This classification procedure not only widened the scope of the linguistic investigation but also provided a framework for a comprehensive examination of the portrayal of men and women in discourse.

This study harnessed Gephi, an open-source software tool developed by Bastian et al. (2009), to visualize and comprehend the intricate data network. The generated graphs (constellation networks) proved invaluable assets, serving a dual role. Notably, they facilitated the identification of paramount adjectives associated with each lemma, thereby offering insights into the prevailing semantic contexts. Moreover, these graphical representations enabled the intricate comparison and establishment of relationships among the verbs. Such a comparative approach was a potent lens to unveil underlying similarities and differences.

3. RESULTS AND DISCUSSION

The main objective of the collocational analysis is to explore how the lemmas HOMBRE (MAN) and MUJER (WOMAN) behave in collocation within the NOW Corpus. As mentioned, 3,200 verbal collocations with a mutual information score exceeding three have been extracted. It's important to note that due to the volume of collocates, a selection process was carried out to show only a subset of them. This decision was made to enhance the clarity of the constellation networks, ensuring that the associations between the lemmas and collocates are presented more coherently and understandably. Figure 3 illustrates the constellation network comprising verbs allocated to the Communication category.

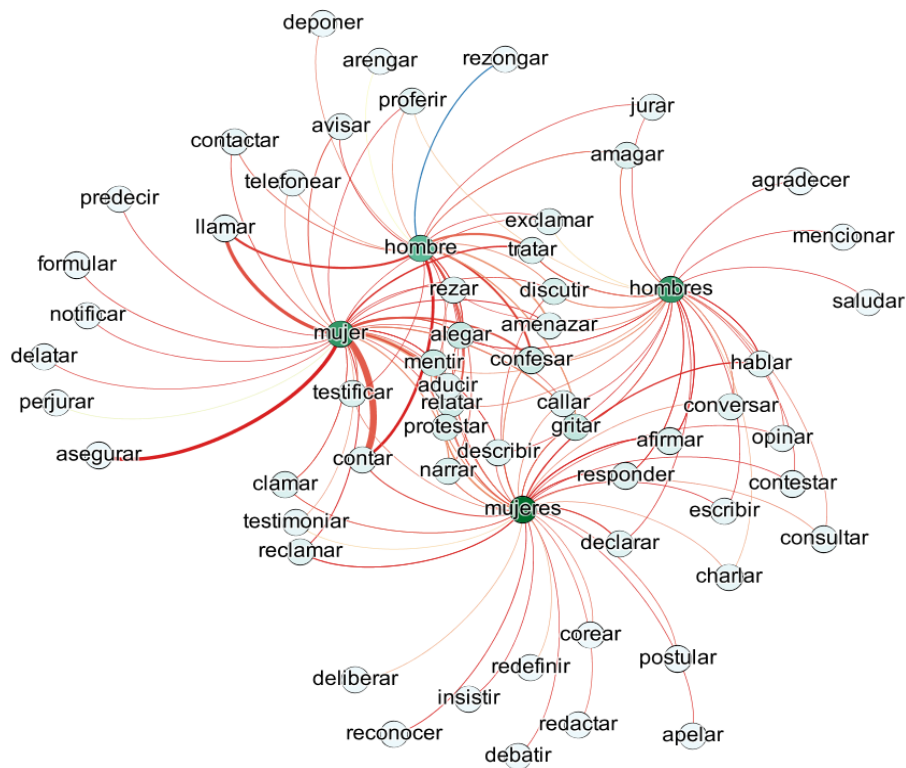


Figure 3. Constellation network of the category of Communication. Source: own elaboration.

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 exclusively associated with men. On the other side, verbs like "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 clamor) exclusively collocate with WOMAN.

Among these verbs, a subset falls within a semantic field named "Emotional Vocal Expression." This includes verbs like "rezongar" (to grumble) and "exclamar" (to exclaim) which are tied to 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, and most of these collate with WOMAN.

One verb, "perjurar" (to perjure), holds the highest MI score and is found exclusively collocated with women. Conversely, verbs such as "postular" (to postulate), "redefinir" (to redefine), "reconocer" (to acknowledge), "deliberar" (to deliberate), "testificar" (to testify), "notificar" (to notify), "formular" (to formulate) suggest a standard mode of communicating information and also exclusively collocate with women. Contrastingly, verbs like "discutir" (to argue), "mentir" (to lie), "amenazar" (to threaten), and "gritar" (to yell) are more strongly associated with men, while "protestar" (to protest), "callar" (to be silent), and "alegar" (to argue) are more prevalent with women. It is interesting to note that the phrase "be silent" ('las mujeres callan') exclusively collocates with women, suggesting that gender representations are often attributed to the entire community rather than individual members, potentially leading to deindividuation. This observation also raises the concept of "othering," emphasizing the division between "us" and "them," and such linguistic representations of gender reinforce the idea of a binary gender dichotomy that does not consider the different members of the LGBTQ+ community.

This research aims to show how Corpus Linguistics techniques, particularly collocations, can enrich qualitative studies. While this thesis doesn't delve into gender studies or discourse analysis, I will dedicate a few paragraphs to explore the concept of "othering" and its relation to gender representation. As per Jensen (2011), "othering" assumes that individuals lacking power are often relegated to being "the other" in discourse, reinforcing the authority and superiority of the powerful and shaping the identity of subordinates (p. 65). Viewing men as the norm and women as the "other" leads men to define and describe women, perpetuating their objectification and otherization. A collocational analysis can help uncover and analyze how women are otherized and objectified. Furthermore, the differentiation and demarcation process highlighted by the collocational analysis in this study helps to visualize how women and men are stereotyped through language.

Below is the constellation network portraying the collocations within the Life category. In contrast to other categories, the Life category presented a relatively modest number of collocations, with most verbs exhibiting collocation tendencies with both lemmas.

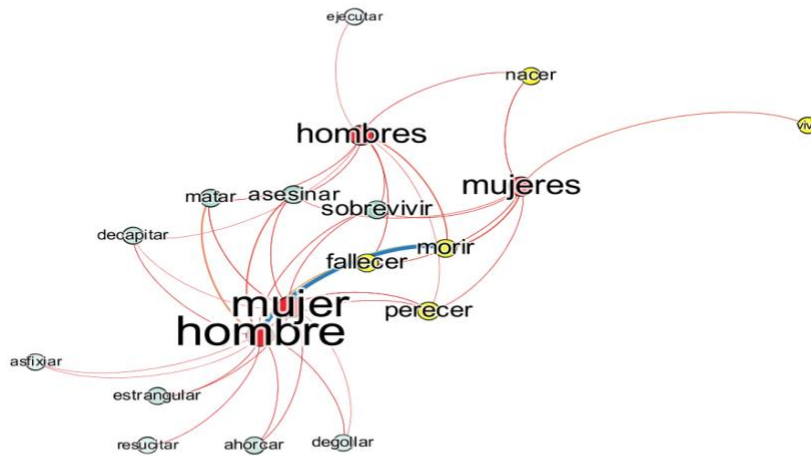


Figure 4. Constellation network of the category of Life. Source: own elaboration.

The verb "vivir" (to live) patterned more strongly with the lemma WOMAN while "ejecutar" (to execute) and "resucitar" (to resurrect) patterned with the lemma MAN. Notably, only the verb "to execute" carries a negative connotation. Other verbs such as "matar" (to kill), "decapitar" (to decapitate), "asesinar" (to murder), "estrangular" (to strangle), "degollar" (to behead), "asfixiar" (to asphyxiate), and "ahorcar" (to choke) also carry negative connotations and are associated with both genders; however, these verbs are more strongly associated with MAN. For example, the verb "matar" (to kill) exhibits a stronger association with MAN (MI score of 7.21) compared to WOMAN (MI score of 5.22). Similarly, "degollar" (to behead) has an MI score of 8.33 with MAN and 5.98 with WOMAN. These findings align with previous research in other languages, such as English, that highlights a more vital link between men and acts of violence.

Additionally, in terms of non-violent verbs, MAN is closely related to "morir" (to die), while WOMAN is more strongly connected to "fallecer" (to pass away). This case is interesting because "to die" is considered a neutral verb since it simply describes the cessation of life and has an aspect of emotional detachment. On the other hand, the verb "to pass away" is more common in obituaries and situations where a more empathetic approach is required or desired. To some extent, this verb is more related to something religious and describes a peaceful transition from life to death. This verb is used to show respect and not sound blunt, and it also has a more emotional tone than the verb "to die." It is essential to mention that this can vary due to individual and cultural differences. While both verbs refer to the same concept of death, they have different connotations and can be used in other contexts depending on the writer's preference and the situation.

This lexical analysis presents an opportunity for expansion by investigating the behavior of these two lemmas not only as subjects, as demonstrated in this study, but also as objects. By examining the recipients of the actions described by the subjects in the online news press, we can gain insights into both the perpetrators of violence and those subjected to such acts. This expanded approach has the potential to provide a deeper understanding of gender representation and violence in language use. The analysis of verbs within the "Life" category exemplifies how collocational analysis results can be extended to analyze gender representation.

The following constellation network delves into the outcomes within the Competition category. This network encapsulates verbs that signify competition, extending beyond sports to encompass various life contexts. Figure 5 serves as a visual representation of the verbs linked to Competition. A limited number of verbs exclusively collocate either MAN or "WOMAN." Nevertheless, the presence of antonyms among these verbs imbues the results with intriguing significance.

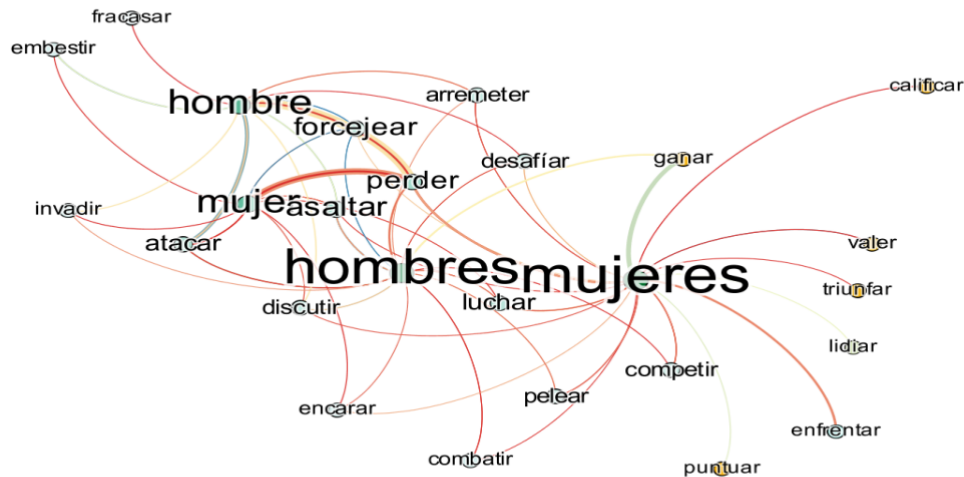


Figure 5. Constellation network of the category of Competition. Source: own elaboration.

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

Interestingly, the verb "to fail" only patterns with MAN while "to succeed" patterns with WOMAN. In this context, it is pertinent to revisit the concept of "markedness" to understand the meaning behind marked and unmarked terms. Marked terms convey a meaning that is not commonly understood, while unmarked terms convey a widely understood meaning. 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. This suggests that men could be marked with certain verbs, indicating that it should not be taken for granted that they succeed most of the time. Conversely, the absence of certain verbs could imply that specific actions are considered default for men. This argument also applies to the fact that the verb "to succeed" only patterns with "woman," which implies that when both "woman" and the verb "to succeed" pattern, it may be something unusual.

An analysis of collocations, such as the one attempted in this study, allows us to identify marked and unmarked expressions, which leads us to determine gender representations embedded in language.

It's worth noting that the verbs "to deal with" and "to cope with," which convey the notion of confronting challenges, exclusively collocate with WOMAN. Interestingly, no parallel verb pattern was observed with MAN. This unique collocational tendency raises thought-provoking implications. Pairing these verbs with WOMAN might suggest a connection between women and handling difficulties. Conversely, the absence of comparable verb collocations with "man" could imply a link between men and problem-solving skills.

Several verbs, such as "luchar" (to struggle), "combatir" (to combat), "desafiar" (to challenge), "competir" (to compete), and "encarar" (to face), display collocational ties with both genders. All these verbs, except "to combat," exhibit a notably stronger association with women. This intriguing pattern suggests a higher likelihood of women being depicted in situations that involve struggle or challenge.

The analysis of verb collocations with the lemmas MAN and WOMAN reveals that certain verbs are more strongly associated with one gender, highlighting how language use and gender are intertwined. This analysis can enrich qualitative studies and help visualize how women and men are stereotyped through language. The concept of markedness has been applied to linguistic analysis to explore gender representation and other social meanings. The pairing of specific verbs with gendered subjects reveals underlying assumptions about success, failure, and gender roles. The exclusive alignment of "to fail" with MAN questions the assumption of male success, challenging the norm.

On the other hand, the association of "to succeed" with WOMAN draws attention to exceptional achievements in this context. This analysis also highlights the significance of uncovering marked expressions to discern subtle gender representations embedded within language. The exclusive collocation of "to deal with" and "to cope with" with "woman," coupled with the absence of a parallel pattern with "man," raises intriguing implications. It suggests a potential connection between women and their ability to manage challenges. Similarly, the distinctive pattern of verbs like "luchar" (to struggle), "desafiar" (to challenge), and "encarar" (to face) leaning towards women underscores their portrayal in contexts of resilience and determination. These findings underscore how language molds and mirrors societal viewpoints on gender roles, thereby enhancing our comprehension of gender dynamics from a linguistic perspective.

CONCLUSION

This collocational study illuminates the discourses and social practices evident in press coverage, enhancing discussions on gender representation for both women and men. Additionally, it provides a thorough investigation covering multiple aspects, as outlined briefly below.

The corpus for this lexical analysis was compiled from news on the web collected from various online media outlets from Latin American countries and Spain. This allows scrutinizing how media language choices contribute to, reinforce, and perpetuate gender stereotypes, prompting reflection on the media's responsibility and influence in shaping societal perceptions.

Another significant facet arising from this study, intertwined with its limitations, is its capacity to extract information from diverse contexts, enabling cross-cultural comparisons. Thus, studies of this nature can ascertain whether similar gender associations prevail across Spanish-speaking regions, varying in linguistic varieties and socio-cultural contexts. This exploration extends to probing whether specific linguistic patterns are universal within the same language and indicative of specific cultural norms.

Analysis of verbal collocations involving the lemmas MAN and WOMAN unveils linguistic gender biases, as certain verbs become associated predominantly with one gender, perpetuating societal stereotypes and roles. These findings catalyze discussions on initiatives to promote gender-neutral language and raise awareness about embedded linguistic biases. Such discussions intersect with broader discourse on language reform and its potential to influence societal attitudes and roles. For instance, examining verbs linked to communication and emotional expression reflects societal expectations regarding gendered modes of communication and emotional display. This prompts exploration into how individuals across genders and communities are socialized to express emotions, thus emphasizing the role of language in shaping social norms.

Moreover, the lexical analysis conducted provides a platform for discussing how language enables the identification of power dynamics perpetuated or established through linguistic choices. Furthermore, this type of analysis facilitates exploration into how our linguistic choices intersect with factors such as race and sexuality.

Finally, this article's acknowledgment of corpus linguistics techniques underscores their utility in capturing societal discourses. Delving into the collocational analysis of such discourse enables a nuanced exploration of the intricate interplay 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
- 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
- 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. 2014. *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.