



## OPEN ACCESS

## EDITED BY

Kostas Maronitis,  
Leeds Trinity University, United Kingdom

## REVIEWED BY

Zuocheng Zhang,  
University of New England, Australia  
Ahlam Alharbi,  
Imam Abdulrahman Bin Faisal University,  
Saudi Arabia

## \*CORRESPONDENCE

Jiaqi Xue  
✉ jqxue00@126.com

RECEIVED 28 November 2024

ACCEPTED 29 January 2025

PUBLISHED 19 February 2025

## CITATION

Xue J (2025) Men as offenders while women as victims: a corpus-based study of men and women in the United Nations.  
*Front. Commun.* 10:1535312.  
doi: 10.3389/fcomm.2025.1535312

## COPYRIGHT

© 2025 Xue. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Men as offenders while women as victims: a corpus-based study of men and women in the United Nations

Jiaqi Xue\*

Hong Kong Polytechnic University, Kowloon, Hong Kong SAR, China

This study examines the representations of men and women in the United Nations Parallel Corpus-English (UNPC-E) by using the corpus linguistics tool, Sketch Engine. Three types of grammatical collocations were explored, including verbs frequently collocating with MAN and WOMAN when they serve as subjects and objects, and adjectives used as predicates in relation to them. The results reveal that within the UN discourse, men are often portrayed as offenders while women tend to be depicted as victims. This pattern of representations may stem from the UN's agenda focused on addressing gender-based inequality, as well as from the conventional reporting bias that might have influenced data availability. In general, this study broadens the scope of research in the field of gender and corpus linguistics to some extent, offering a more comprehensive and international perspective on the representations of men and women through the analysis of the UN corpus.

## KEYWORDS

men and women, representation, the United Nations, corpus linguistics, Sketch Engine

## 1 Introduction

The United Nations (UN) plays a crucial and indispensable role in today's world (Moore and Pubantz, 2022). Endeavoring to “reshape the world” (Mingst et al., 2022, p. 1), its mission encompasses maintaining international peace and security (Badache et al., 2022), promoting sustainable development (Montiel et al., 2021), protecting human rights (Baça and Anxhaku, 2023), and fostering global business collaboration (de Villiers et al., 2021). Through its specialized agencies and a multitude of programs and initiatives, the UN addresses a wide array of challenges and issues that transcend national borders, from climate change and health pandemics to armed conflicts and economic disparities. Its ability to convene nations and provide a platform for dialogue and negotiation is unparalleled, often serving as the last bastion of hope for peaceful resolution in times of international crisis. As such, the considerable contributions of the UN in promoting international stability, advancing sustainable development, ensuring the protection of human rights, and enhancing global cooperation merit recognition and appreciation.

In the realm of human rights, the issue of gender equality is paramount (Reilly, 2009). The UN is committed to addressing and ameliorating the disparities between men and women, with a particular focus on enhancing women's rights (Brannon, 2022; Rothermel, 2020). To be more specific, in light of the acknowledgment that “universally recognized human rights are still not enjoyed equally by women and men” (Pietilä, 2007, p. 27), the UN has facilitated significant advancements. A series of landmark frameworks have been established to protect women's rights, including the Convention on the Elimination of All Forms of Discrimination Against Women, the Beijing Declaration and Platform for Action,

and Resolution 1,325 on “Women, Peace and Security,” among others (ibid.). Despite these initiatives, outcomes regarding women’s public status and their involvement in the UN affairs have been less than satisfactory (Haack, 2014; Skard, 2008; Pietilä and Vickers, 1990). Skard (2008) observes that “while women’s descriptive representation in UN agencies has increased, its unevenness across portfolios and different types of agencies (appointed vs. elected) remains a challenge” (p. 51). This highlights the ongoing need for concerted efforts to ensure equitable participation and representation of women within the UN’s structure and activities.

As previously mentioned, the analysis of women’s roles within the UN frequently emphasizes their actual participation and involvement in related institutions and activities (e.g., Brannon, 2022; Haack, 2014). However, it is essential to acknowledge that the representation of men and women can also be examined from a linguistic perspective. According to the United Nation (2020) guidelines, owing to the key role that language plays in shaping cultural and social attitudes, “using gender-inclusive language is a powerful way to promote gender equality and eradicate gender bias. Being inclusive from a gender language perspective means speaking and writing in a way that does not discriminate against a particular sex, social gender or gender identity, and does not perpetuate gender stereotypes.” Such a statement, to some extent, highlights the significance of investigating the interplay between gender and language within the context of the UN’s framework.

However, relatively few studies have explored men’s and women’s representations from a linguistic perspective within the UN context. On this account, the current research aims to explore the gender representations in the UN domain by utilizing a corpus linguistics tool, Sketch Engine. It is hoped that the present research can not only contribute to broadening the research scope of inquiry into the relationship between gender and linguistics but also reexamine the portrayal of men and women from a global prism through the lens of a large corpus consisting of the official records and documents in the UN.

## 2 Gender and corpus linguistics

Corpora, as Baker (2006) observes, enjoy great adaptability in discourse analysis (DA). Particularly, their introduction “can benefit gender research, addressing questions concerned with gendered *usage of language*” (Baker, 2012, p. 100, italic in the original). Using corpus tools, numerous scholars have extensively examined the language usage of various gender pairs, including *male* and *female*, *men* and *women*, *boys* and *girls*, among others. To be more specific, Pearce (2008) examined the representations of MAN and WOMAN in the British National Corpus (BNC), finding that MAN and WOMAN are presented “in stereotypical ways” (p. 19) with men being characterized by masculine traits, while women are portrayed as weak. Taylor (2013) employed multiple corpora to investigate the image of *boys* and *girls* in the UK national broadsheet newspapers and found that “the ways in which the lemma GIRL, over a very recent 17 year period, is consistently associated with sexual contexts” (p. 108). In the New Model Corpus (NMC), Norberg (2016) similarly investigated the representations of *boys* and *girls*, arguing that “*boys* in the NMC appear as energetic, playful, curious and mischievous, whereas *girls*

to a large extent are represented as objects of desire with a focus on bodily appearance and agreeable behavior” (p. 311).

In addition, the expansion of social identity theory (Augustinoux and Walker, 1995) has broadened the exploration of different gender pairs, including *bachelors* and *spinsters*, *fathers* and *mothers*, *husbands* and *wives*, *sons* and *daughters*, etc. Baker (2006) delved into the image of *bachelors* and *spinsters* in the BNC and pointed out that the seemingly universal term actually enjoys certain undertones in different contexts. By using the corpus linguistics tool, Sketch Engine, Fang and Han (2015) explored the image of *husbands* and *wives* in the BNC. They found that gender inequalities are prevalent, with *husbands* often holding primary decision-making power within the family, while *wives* are more likely to prioritize family harmony. Likewise, this gender inequality can also be reflected through various social roles, e.g., *fathers* and *mothers*, *sons* and *daughters*, etc. According to the enTenTen corpus, Wang and Chen (2016) observed that in Britain, *fathers* are typically associated with higher social status and greater social acceptance, as evidenced by their frequent collocation with terms such as “lawyer” and “doctor”; in contrast, the role of *mothers* is often limited to domestic responsibilities and family life, as indicated by associations with phrases like “at home” and “home-schooling” (p. 62). Similarly, Lu and Wang (2021), also utilizing the enTenTen corpus, argued that *daughters* within a family tend to be marginalized, often linked with terms such as “burden” and “overweight”; on the contrary, *sons* are perceived as integral to the family, being described with nouns like “successor” or “heir” (p. 6).

Within the UN, research on the intersection of gender and language is notably limited. Nevertheless, Brun-Mercer’s (2021) investigation into the collocates of *woman(s)*, *women(s)*, *man(s)*, and *men(s)* within 193 addresses from the UN General Assembly General Debate in the autumn of 2015 sheds light on gender representations by international policymakers. This study highlights the persistent disparities in the frequency and context of references to *men* and *women* within the UN discourse in which women are frequently portrayed as a group in need of protection and support, whereas men are often depicted as political figures and other distinguished individuals. Despite this, there remains a scarcity of scholarly work that probes the nexus of language and gender specifically within the UN setting. In this context, based on Brun-Mercer’s analysis, the current study stands as an endeavor to expand this kind of research, hoping to offer more valuable insights into the linguistic dimensions of gender issues at the international level.

Generally speaking, different as research methods and research perspectives would be, the major findings of these studies (e.g., Brun-Mercer, 2021; Taylor, 2013; Pearce, 2008) tend to converge to some degree. To be more specific, a gender bias in language expressions is consistently observed. Men are prone to remain the focal point across different corpora and the gender pair is exhibited in a stereotypical way (Golombok and Fivush, 1994), wherein men are often depicted as superior to women. Granted, some studies have shown that more neutral language expressions are gradually replacing the existing biased language (Macalister, 2011; Sigley and Holmes, 2002). Nevertheless, this shift in linguistic representation may be limited to specific domains, such as education and appearance. Furthermore, previous research tends to explore gender and language from a relatively narrow point of view. For example, in studies related to the BNC (e.g., Pearce, 2008), men’s and women’s images are characterized

from the perspective of the UK. In other words, a broad and general perspective concerning gender and language may still be lacking.

In this context, the present study aims to explore the representations of men and women in the United Nations Parallel Corpus–English (UNPC-E) at an international level. Furthermore, from the perspective of critical discourse analysis (CDA), linguistic expressions reveal the interplay between “language and other social elements (power relations, ideologies, institutions, social identities, and so forth)” (Fairclough, 2013, p. 9). In this sense, discourses derived from large corpora may still be shaped by cultural, historical and social contexts (ibid.). In this light, as a kind of “corpus-based discourse analysis study” (Norberg, 2016, p. 445), the current research also seeks to discuss potential explanations for the observed linguistic representation of men and women in the corpus in question, which may not have been covered in previous studies. As such, the following three research questions are addressed:

- 1 Which verbs frequently collocate with MAN and WOMAN when they serve as subjects and objects, and which adjectives are commonly used as predicates in relation to MAN and WOMAN within the UNPC-E?
- 2 Based on the collocations identified in the UNPC-E, what are the overarching representations of men and women?
- 3 Which social factors may have shaped the observed representations of men and women in the UNPC-E?

### 3 Material and methodology

The corpus employed for this study is the United Nations Parallel Corpus–English (UNPC-E) which mainly includes official records, documents as well as parliamentary materials with 664,924,245 words in total. In fact, it is a sub-corpus that belongs to one of the parallel corpora of the United Nations Parallel Corpus (UNPC). The texts included are sourced from a diverse array of official UN documents, including speeches by diplomats, resolutions, reports, and documents produced by various UN agencies and bodies (Ziemski et al., 2016). Upon examination of the texts, it is evident that they encompass a multitude of globally significant topics, such as discussions on terrorist incidents in CODESA (Convention for a Democratic South Africa), European immigration issues, global healthcare, and international civic education, among others.

In essence, as Gatea and Lateef (2023) notes, “the UN has identified multilingualism as a strategic goal because it is crucial to peaceful inter-human dialogue” (p. 32). As such, there are six parallel corpora associated with the UNPC, including Arabic, Chinese (Simplified script), English, French, Russian, and Spanish. The rationale for focusing on the English parallel corpus is informed by the longstanding recognition of English as a global *lingua franca* (Crystal, 2003; Jenkins, 2017; Smokotin et al., 2014). According to Choi (2021), “in the context of neoliberal globalization, English has largely been considered a means of global communication” (p. 144).

In this study, Sketch Engine was used as the main research tool. This tool was developed by Kilgarriff et al. (2008) and serves as a network assistance system, which can combine grammatical and lexical knowledge in different kinds of corpora and provide over a hundred million collocations of lemma (Liang et al., 2010). Since its inception, Sketch Engine has undergone refinement and

expansion, establishing itself as a crucial and robust research tool in the field of corpus linguistics (Kilgarriff et al., 2014). Notably, it features a function called Word Sketch Difference, which enables users to compare the collocational behavior of two words. This function aids in identifying differences and similarities in their usage patterns based on large text corpora. Specifically, it provides all possible grammatical roles of the target lemma, facilitating a comprehensive and systematic description of their collocational behavior.

In the present study, the analysis will focus on the collocations with the lemmas of MAN and WOMAN serving three distinct grammatical roles: MAN and WOMAN as subjects (e.g., The men beat the inhabitants), MAN and WOMAN as objects (e.g., The population-related policies...affect men and women in notably different ways.), and adjective predicates related to MAN and WOMAN (e.g., Most men were *reluctant* to accompany their wives). According to Pearce (2008), this kind of categorization “allows to say something about what men and women are represented as doing and experiencing as agents, and undergoing as patients and beneficiaries, and also how they are described and categorized” (p. 7). It is also noteworthy that, for the data analysis, some manual filters and revisions are indispensable to maintain the validity of the study.

## 4 Major findings

### 4.1 MAN and WOMAN as subjects

Through the function of Word Sketch Difference, verbs collocating with MAN and WOMAN in this corpus can be identified. Specifically, the verbs that occur exclusively with MAN are extracted and presented in Table 1.

As shown in Table 1, the verbs collocating with MAN are all action verbs (Macalister, 2011). Except terms of *force* and *assault*, the remaining verbs collocating with MAN demonstrate a relatively high logDice score (above 7.0), indicating a strong collocational relationship where “the theoretical maximum for a collocation is 14, suggesting that every occurrence of X is found with Y” (Norberg, 2016, p. 295). Notably, these verbs are predominantly associated with crimes and violent acts. Essentially, they denote actions that infringe upon another individual’s physical integrity, personal freedom, or dignity, mirroring violent and coercive behavior that severely violates both human rights and legal standards. The following examples further elucidate this observation:

TABLE 1 Verbs collocating with MAN as subjects exclusively.

Verb	Frequency	logDice
beat	111	7.7
abduct	106	8.5
shoot	79	7.4
perpetrate	76	7.4
force	59	2.2
assault	50	6.2
steal	40	7.8

## Extract 1

(1a) the men *beat* the inhabitants, including a three-year-old child, with the butts of their revolvers.

(1b) ...four armed men had *forced* three indigenous passengers to board a car...

(1c) Unidentified men *stole* items from a girls' secondary school in the village of Tayyibah.

(UNPC-E, <https://app.sketchengine.eu>).

As demonstrated in Extract 1, in the UN records and documents, men are the direct sources of all sorts of offences. It is evident that these men engage in acts of violence against women and innocent individuals, extinguishing precious lives and leaving behind shattered families in their wake. Their criminal tendencies extend beyond the realm of physical harm, as they also indulge in acts of theft, causing immense suffering and economic loss to countless people and communities.

On the contrary, within the UNPC-E, only one word *encounter* is found to exclusively collocate with WOMAN, as shown in Table 2.

The verb *encounter* appears with a frequency of 274 and a notably high logDice score of 8.6, which is much higher than *beat* (111) for MAN. It decodes that an especially unexpected adversary or enemy is met. A detailed content analysis of the *encounter* within this corpus reveals that women face a multitude of challenges, including legal barriers, political obstacles, discrimination, and low familial status, as illustrated in the following extract:

## Extract 2

(2a) Women had *encountered* many obstacles.

(2b) Women still *encounter* legal barriers.

(2c) In practice women often *encountered* discrimination in respect of both remuneration and opportunities for promotion.

(UNPC-E, <https://app.sketchengine.eu>).

In this context, it is posited that when MAN and WOMAN are positioned as subjects, men are often depicted as aggressive offenders who are responsible for violence; conversely, women are faced with several problems and difficulties, assuming the role of victims. In essence, these verbs encapsulate the gender power dynamics, portraying men as dominant or superior to women, with power predominantly concentrated in male-dominated institutions such as government, business, and religion (Agarwal, 2015; Diekman et al., 2004).

In addition, in the UNPC-E, it is discovered that with regard to this grammatical role, apart from a few specific verbs that more frequently collocate with MAN, the majority of verbs tend to exhibit a bias toward WOMAN. In this vein, it is necessary to discuss these certain verbs, as presented in Table 3.

As shown in Table 3, when MAN and WOMAN are used as subjects, there is a significantly higher frequency of occurrence when these verbs collocate with MAN compared to WOMAN. Consistent with the earlier discussion, these verbs predominantly fall under the category of "activity verbs" (Norberg, 2016, p. 312) suggesting that within the UN records, men are more commonly depicted as being the offenders and perpetrators of such activities, whereas women tend to

TABLE 2 Verbs collocating with WOMAN as subjects exclusively.

Verb	Frequency	logDice
encounter	274	8.6

be portrayed as victims of these actions. To exemplify this trend, the verb *attack*, which exhibits the largest disparity (189 occurrences with MAN vs. 10 occurrences with WOMAN) is taken as a particular example:

## Extract 3

(3a) ...men *attacked* the UNTAC office in Thpong district in Kompong Speu Province...

(3b) ...men *attacked* a train in Kampot Province...

(3c) ...men *attacked* the Port-au-Prince police station...

(3d) ...men randomly *attacked* residents of Boiphatong in the Johannesburg area.

(3e) ...the woman who *attacked* her did not utter any threats relating to her victim...

(3f) The woman who *attacked* Ms. Uzunova did not utter any threats...

(UNPC-E, <https://app.sketchengine.eu>).

As demonstrated in (3a), (3b), and (3c), men can attack numerous places, including offices, trains, police stations, etc. In fact, according to the data in the UNPC-E, almost every place can be their attacking target. In addition, men's attacks may not be confined to concrete places, but individuals can also become their goal, which further indicates and reinforces their role as offenders. Nevertheless, from (3e) and (3f), attacks by women may not result in any consequences for the target. Thus, it is suggested that in the UN documents, women are faced with physical and psychological limitations, while men are framed as agents of control, violence, and crime.

## 4.2 MAN and WOMAN as objects

As another important grammatical category, MAN and WOMAN as objects deserves particular attention. Verbs that exclusively

TABLE 3 Verbs favoring MAN over WOMAN.

Verb	Frequency	
	MAN	WOMAN
attack	189	10
rape	200	23
threaten	47	9
open	76	16
kill	135	38
abuse	33	11
break	54	19
cross	41	15
drive	53	33
dominate	337	234
disappear	40	30
approach	55	39
wear	174	141
stop	35	33
control	66	65
commit	183	167



collocate with MAN and WOMAN are mainly discussed in this section.

As shown in Table 4, three verbs, i.e., *arm*, *uniform*, and *mask*, exclusively collocate with MAN. All these three verbs portray men in a state of violence, as the term *arm* typically refers to the act of furnishing or equipping with weapons. Additionally, the verbs *uniform* and *mask* describe their attire during violent behavior. Generally, similar to the findings in MAN as subjects, when serving as objects, men are portrayed as violent offenders who carry guns and wear masks and uniforms to carry out attacks. The following extract could further illustrate this representation:

#### Extract 4

(4a) ...armed man with T-56 rifles, and destroyed all the green lights in the mosque...

(4b) ...uniformed men and beat him with a rubber club on his heels and pushed him to the floor...

(4c) The terrorist masked men with military uniforms and abducted a male relative of the Georgian Deputy Chairman of Parliament...

(UNPC-E, <https://app.sketchengine.eu>).

As shown in these three examples, according to the UN records and documents, men have even become synonymous with violence and criminals. They were responsible for the destruction of infrastructure, carried out attacks against strangers, and even went as far as seizing relatives of government officials. These findings solidify men's representation of offenders and wrongdoers. Nevertheless, when examining the verbs that exclusively collocate with WOMAN in this corpus, there is a stark contrast in the representation of men and women, as indicated in Table 5.

In the first place, with regard to the frequency of these verbs, a huge gap is observed between the verbs collocating with WOMAN exclusively and those collocating with MAN in this corpus. It is found that the frequency of verbs associated with WOMAN is considerably higher than that with MAN in the UN domain. For instance, taking the verb *restrict* as an example, it is the word that enjoys the lowest occurrence among the verbs exclusively for WOMAN; however, its frequency (417) is still twice as high as the word *arm* (298) which shares the highest occurrence among verbs for MAN. In this vein, women are much more mentioned than their counterparts in this corpus.

In addition, according to Table 5, these four verbs can be categorized into two types according to their semantic information. On the one hand, the verbs *enhance* and *advance* belong to the first type which denotes the improvement and growth of value, quality, desirability, or attractiveness; on the other hand, terms *underrepresent* and *restrict* fall under the second type, indicating the adversity that is confined with bounds. In fact, these four verbs share a common semantic aspect in that they all

involve an effect or change on a situation, state, or behavior. Given the context of the UN documents as well as the description of women, it is suggested that there might be a causal relationship between these two sets of verbs. It is precisely owing to the limitations and restrictions women are confronted with that the necessity for their enhancement as well as advance arises. However, in other words, women are still disadvantaged in the UN accounts, facing many difficulties and receiving unnecessary restrictions compared to men, which could be exemplified by the following extract:

#### Extract 5

(5a) ...to restrict women's employment, professional opportunities and mobility in the formal sector...

(5b) ...tends to underrepresent women in councils and committees that approved research programmes...

(5c) ...enhance women's employment and conditions of work...

(5d) ...to advance women's rights, to ensure their greater involvement in public life...

(UNPC-E, <https://app.sketchengine.eu>).

As depicted in (5a) and (5b), a multitude of restrictions are imposed on women. To be specific, they are lacking in employment resources and encounter obstacles when attempting to participate in political affairs. In fact, it is important to note that the restrictions faced by women in this corpus are by no means confined to these two examples. They also have challenges regarding education, healthcare, and so on. On this account, the delegates of the UN are compelled to take action in order to improve the situation for women, as evident in (5c) and (5d); however, women's image of victims can still be apparently observed.

Additionally, as a supplement to this grammatical role, there is a significant disparity in the frequency of verbs that collocate with both MAN and WOMAN. Particularly, when these verbs are used in association with WOMAN, they occur 50 times more frequently than when used in conjunction with MAN, as depicted in the following table.

These four words listed in Table 6, to a certain extent, share similar meanings to the words *enhance* and *advance* in that they refer to the improvement and positive changes for women, as shown in the following extract:

#### Extract 6

(6a) It had improved women's health services and now ran annual vaccination campaigns for both women and children.

(6b) ...in protecting women's human rights because...

(6c) ...on promoting women's full participation in economic planning...

(6d) The need to empower women and to improve their status is acknowledged by most Governments.

(UNPC-E, <https://app.sketchengine.eu>).

Similar to (5c) and (5d), on one hand, the use of these words attests to the UN's concern for women's issues and they want to change the current situation of women. On the other hand, these words also highlight the narrative of women as victims and their marginalized status within the world community.

TABLE 4 Verbs collocating with MAN as objects exclusively.

Verb	Frequency	logDice
arm	198	5.8
uniform	68	6.3
mask	26	8.7

TABLE 5 Verbs collocating with WOMAN as objects exclusively.

Verb	Frequency	logDice
enhance	1,590	8.1
underrepresent	617	12.1
advance	614	7.8
restrict	417	8.6

TABLE 6 Verbs favoring WOMAN than MAN (over 50 times).

Verb	Frequency	
	MAN	WOMAN
improve	19	2,192
protect	67	5,377
promote	97	4,872
empower	49	4,690

### 4.3 Adjective predicates of MAN and WOMAN

In previous research, adjectives as modifiers of different gender pairs were explored extensively (Norberg, 2016; Pearce, 2008). In fact, adjective predicates are also of importance in English grammar system as “predicate adjectives predicate a property of the clausal subject” (Bybee and Thompson, 2022, p. 2). In the UNPC-E, the adjective predicates collocating exclusively with MAN are identified as follows:

As shown in Table 7, there are only three adjectives that exclusively function as predicates for MAN in this corpus, namely, *superior*, *guilty*, and *violent*. These three adjective predicates, to some extent, reflect the differences between men and women in the UN records, as demonstrated in the following extract:

#### Extract 7

- (7a) ...that men are *superior* to women...
- (7b) ...men who are *violent* to women are likely to have more sexual partners...
- (7c) ...men who were *guilty* of crimes such as rape and defilement...

(UNPC-E, <https://app.sketchengine.eu>).

According to these three examples, a direct comparison is drawn between men and women and it is suggested that men are superior to women with higher rank, quality, or importance; besides, they are deemed guilty solely because of the violence they made on women, e.g., rape, defilement and so on. The image of men as offenders while women as victims is again reinforced.

However, it is worth noting the occurrence is extremely low in this grammatical category. Furthermore, apart from these three adjectives, the remaining adjectives as predicates enjoy much more frequency when collocating with WOMAN compared to MAN. This observation indicates that men are not the central focus of this grammatical category and that their image as offenders is somewhat diluted. On the contrary, it further confirms that women consistently take the spotlight in the UN records. In this vein, some adjective predicates of

TABLE 7 Adjective predicates of MAN exclusively.

Adjective predicates	Frequency	logDice
superior	20	6.9
guilty	20	5.4
violent	9	3.9

WOMAN exclusively with extremely high frequency (over 100 times) are identified.

According to Table 8, the word, *vulnerable*, is used most frequently (708) among all adjectives. The term *vulnerable* refers to the susceptibility to physical or emotional attack or harm. It can be seen that women in the UN domain are easy to be hurt by something negative or harmful, which further strengthens women's representation as victims. Furthermore, it is found that adjective predicates *pregnant* and *migrant* enjoy relatively high occurrence. After detailed text analysis, both of them reflect the international community's concern for the protection of women, especially the protection of *pregnant* women and *migrant* women workers, as shown in the following extract:

#### Extract 8

- (8a) ... if the woman is *pregnant*, various rights come into effect that are comparable to those of a married couple...
- (8b) Article 83 stipulates that imprisoned women who are *pregnant* and who have babies under two years of age may have the right to live outside the territory of the prison...
- (8c) Women who are *pregnant* are entitled to 90 days' leave...
- (8d) To raise awareness of labor rights among women *migrant* and domestic workers...
- (8e) ...to design and implement projects to promote and protect the rights of women *migrant* domestic workers...
- (UNPC-E, <https://app.sketchengine.eu>).

Apparently, these examples come from some Articles proposed by members of the UN. According to (8a), (8b) and (8c), *pregnant* women are granted equal rights with other people. They are entitled to extended maternity leave and are even allowed temporary release from prison to raise a child. For women who are migrant to make a living, several projects and programs are developed to safeguard and advance their rights, as demonstrated in (8d) as well as (8e). Thus, it is suggested that women's issues have attracted a great deal of attention from the international community, and some relevant programs have been put in place to address these issues.

It is also noted that the adjective predicate, *important*, exhibits a relatively high frequency (103) among those adjectives exclusively for women, which indicates that the status and ability of women are gradually recognized by the world community, as depicted in the following extract:

#### Extract 9

- (9a) Women are *important* economic agents...
- (9b) women were particularly *important* for promoting sustainable development...
- (9c) ...women are *important* when deciding upon transport infrastructure investments...
- (UNPC-E, <https://app.sketchengine.eu>).

TABLE 8 Adjective predicates of WOMAN exclusively (over 100 times).

Adjective predicates	Frequency	logDice
vulnerable	708	8.0
pregnant	349	9.1
migrant	116	7.1
important	103	3.4

From Extract 9, women can play their part in the progress of the whole world. They can contribute to economic advancement and sustainable development as well as the investment of city facilities. Actually, by no means is their importance confined to these areas. According to the recorded data in this corpus, they also play an essential role in other aspects, including education, employment, family building, etc. Admittedly, the frequency of this word is not comparable with those utilized to depict their vulnerable images, especially in terms of the logDice (3.4 vs. 8.0, 9.1, and 7.1). Nevertheless, it is at least a good signal that the international community has recognized the importance of women.

## 5 Discussion and conclusion

The UN prioritizes women's issues and has established numerous agreements and programs to address them (Pietilä, 2007). A general representation of men and women can be obtained within the UNPC-E domain through the current study. It is revealed that in the UN documents, men tend to be represented as offenders while women as victims. Critical discourse analysis, as an effective “explanatory critique” (Fairclough, 2013, p. 9), situates linguistic patterns within a broader social context to elucidate the relationship between text and context (ibid.). In this regard, these linguistic representations are not merely reflections of objective reality; they also contribute to shaping our understanding of that reality. This approach thus enables us to comprehend “the underlying reasons and biases driving specific representations” (Breazu and Katsos, 2024, p. 693). Therefore, it is crucial to explore possible explanations for the observed patterns to deepen our understanding of this subject through the lens of CDA.

First, these representations often align with the broader international agenda focused on addressing gender-based violence and inequality. Historically, women have been disproportionately affected by various forms of violence and discrimination (Burnette, 2015), ranging from domestic violence and sexual assault to systemic discrimination in education, employment, and political participation. This “staggering consequence” of disparity in treatment even starts at an early age, affecting young girls as well (Rafferty, 2013, p. 8). In response, international frameworks such as the UN have underscored the urgent need for states to take proactive measures to protect women's rights and ensure their full participation in society with the advance of a series of documents like Convention on the Elimination of All Forms of Discrimination Against Women, the Beijing Declaration and Platform for Action, etc. By focusing on women's victimization, global initiatives seek to challenge patriarchal structures that perpetuate violence and limit women's opportunities. This approach is integral to fostering a more just and equitable world where all individuals, regardless of gender, can live free from violence and

discrimination. In fact, efforts to elevate women's voices and rights are seen not just as essential for achieving gender equality, but also as fundamental to advancing overall societal progress and stability. As such, the UN serves as a crucial tool for advocating systemic change and promoting a world where gender-based violence is eradicated, and everyone can thrive equally.

Second, this portrayal may also stem from reporting biases and data availability. Incidents of violence against women are often more visible and tend to be “universally reported” (Watts and Zimmerman, 2002, p. 1252). In contrast, men's experiences of victimization might be less frequently disclosed and openly articulated so that the “negative consequences they [men] may suffer have received considerably less attention” (Depraetere et al., 2020, p. 992). The official records within the UN, which comprise the corpus, are likely to consist predominantly of mainstream documents addressing international issues. This could lead to a skewed representation, where women are predominantly portrayed as victims and men as offenders. In essence, this bias in reporting and the resulting availability of data, to a large degree, mirror the deeply entrenched gender stereotype that frames men as perpetrators while women as victims within the discourse on gender issues (Spiegel, 2013).

With regard to the data availability, a straightforward comparison of the frequency of the lemmas MAN and WOMAN in the UNPC-E and the BNC can further elucidate this point.

The BNC, arguably the most representative and widely utilized corpus for examining the representations of men and women, as well as other gender pairs, reveals a total frequency count for MAN (94996) that surpasses that for WOMAN (58205) (Pearce, 2008). In contrast, the UNPC-E exhibits a substantial disparity between the frequency of MAN (150912) and WOMAN (942975). When examining the calculated percentage difference rates in these two corpora, it becomes evident that the disparity between MAN and WOMAN in the UNPC-E is significantly larger than that observed in the BNC. Specifically, in the UNPC-E, the frequency of WOMAN exceeds that of MAN by 525.83%. Conversely, in the BNC, this rate is −38.44%, indicating that the frequency of WOMAN is 38.44% lower than that of MAN. This pattern, in fact, is not unique to the BNC. In the British Lancaster-Oslo-Bergen Corpus (LOB), there are also considerably more occurrences of MAN than WOMAN (Persson, 1990). Moreover, similar asymmetries are observable in other corpora across different languages, including the French Web 2020 and the German Web 2020, among others.

Admittedly, the sheer frequency of references to both MAN and WOMAN alone does not determine how they are represented. Merely increasing the mentions of WOMAN does not effectively eradicate the existing inequalities between men and women. However, when juxtaposed with other corpora, it becomes relatively apparent that within the UN context, women tend to be the focal point of attention in comparison to men. This, at the very least, signifies that the problems and situations faced by women are garnering greater consideration and attention within the international community.

Inclusive language serves as a priority when it comes to language usage in the UN (Kosnick, 2021). With the utilization of Sketch Engine, this study describes the representations of men and women in the UNPC-E. Generally speaking, within the context of the UN, men are often portrayed as offenders while women are depicted as victims. This kind of representations may derive from the UN's agenda focused on addressing gender-based inequality as well as the

conventional reporting bias leading to the data availability. In addition, compared with other mainstream corpora, women tend to receive more attention within the context of the UN although the gender inequality is not eliminated and can still be observed from the language expression in the UNPC-E. This study expands the research scope in the investigation of gender and corpus linguistics to some extent, providing more insights into the representations of men and women by examining the UN corpus.

Granted, this study is confined to examining only three grammatical categories: MAN and WOMAN as subjects, MAN and WOMAN as objects, and adjectives serving as predicates for MAN and WOMAN. As a result, the generalizability of the findings might be somewhat constrained when considering the extensive nature of the corpus. Moreover, by focusing on these two lemmas as the objects of research, the study may capture only the broad representations of men and women within the UN context, potentially overlooking the nuanced portrayals of specific individuals. For instance, if a document in the UN references political figures such as Angela Merkel, it would mention a woman without using the lemma WOMAN. Despite these limitations, this research may serve as a good attempt to explore the representations of men and women in the UN setting. As more content is indexed into the corpus under consideration, the shifts in their representations warrant further examination, thereby enriching our understanding of gender portrayal in this international context.

## Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

## References

- Agarwal, B. (2015). Power of numbers in gender dynamics: illustrations from community forestry groups. *J. Peasant Stud.* 42, 1–20. doi: 10.1080/03066150.2014.936007
- Augustinuous, H., and Walker, I. (1995). *Social cognition: an integrated approach*. London: Sage.
- Baça, F., and Anxhaku, A. (2023). The role of the United Nations as a defender of human rights: a view from Albania. *Access Justice Eastern Eur.* 6, 1–9. doi: 10.33327/AJEE-18-6.1-n000108
- Badache, F., Hellmüller, S., and Salaymeh, B. (2022). Conflict management or conflict resolution: how do major powers conceive the role of the United Nations in peacebuilding? *Contemp. Sec. Policy* 43, 547–571. doi: 10.1080/13523260.2022.2147334
- Baker, P. (2006). *Using corpora in discourse analysis*. London: Continuum.
- Baker, P. (2012). “Corpora and gender studies” in *Corpus applications in applied linguistics*. eds. K. Hyland, C. M. Huat and M. Handford, London: Bloomsbury Publishing, 100–116.
- Brannon, E. L. (2022). The appointment of men as representatives to the United Nations commission on the status of women. *Polit. Res. Q.* 75, 1360–1373. doi: 10.1177/10659129211066124
- Breazu, P., and Katsos, N. (2024). ChatGPT-4 as a journalist: whose perspectives is it reproducing? *Discourse Soc.* 35, 687–707. doi: 10.1177/09579265241251479
- Brun-Mercer, N. (2021). Women and men in the United Nations: a corpus analysis of general debate addresses. *Discourse Soc.* 32, 443–462. doi: 10.1177/09579265211992145
- Burnette, C. (2015). Historical oppression and intimate partner violence experienced by indigenous women in the United States: understanding connections. *Soc. Serv. Rev.* 89, 531–563. doi: 10.1086/683336
- Bybee, J., and Thompson, S. A. (2022). Interaction and grammar: predicative adjective constructions in English conversation. *Languages* 7, 2–17. doi: 10.3390/languages7010002
- Choi, L. J. (2021). English as an important but unfair resource: university students' perception of English and English language education in South Korea. *Teach. Higher Educ.* 29, 1–15. doi: 10.1080/13562517.2021.1965572
- Crystal, D. (2003). *English as a global language*. Cambridge: Cambridge University Press.
- de Villiers, C., Kuruppu, S., and Dissanayake, D. (2021). A (new) role for business – promoting the United Nations' sustainable development goals through the internet-of-things and blockchain technology. *J. Bus. Res.* 131, 598–609. doi: 10.1016/j.jbusres.2020.11.066
- Depraetere, J., Vandeviver, C., Beken, T. V., and Keygnaert, I. (2020). Big boys don't cry: a critical interpretive synthesis of male sexual victimization. *Trauma Violence Abuse* 21, 991–1010. doi: 10.1177/1524838018816979
- Diekmann, A. B., Goodfriend, W., and Goodwin, S. (2004). Dynamic stereotypes of power: perceived change and stability in gender hierarchies. *Sex Roles* 50, 201–215. doi: 10.1023/B:SERS.0000015552.22775.44
- Fairclough, N. (2013). “Critical discourse analysis” in *The Routledge handbook of discourse analysis*. eds. J. P. Gee and M. Handford (London: Routledge), 9–20.
- Fang, W. J., and Han, J. H. (2015). Sketch engine-based contrastive studies of the collocational behavior of husband and wife in the BNC. *J. Huainan Normal Univ.* 17, 38–42.
- Gatea, H. H., and Lateef, B. A. (2023). Translation of archaic expressions in United Nations' documents: corpus-based study. *Alustath J. Hum. Soci. Sci.* 62, 29–38.
- Golombok, S., and Fivush, R. (1994). *Gender development*. Cambridge: Cambridge University Press.
- Haack, K. (2014). Breaking barriers? Women's representation and leadership at the United Nations. *Glob. Gov.* 20, 37–54. doi: 10.1163/19426720-02001004
- Jenkins, J. (2017). “The future of English as a lingua franca?” in *The Routledge handbook of English as a lingua franca*. eds. J. Jenkins, W. Baker and M. Dewey (New York: Routledge), 594–605.
- Kilgarrieff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., et al. (2014). The sketch engine: ten years on. *Lexicography* 1, 7–36. doi: 10.1007/s40607-014-0009-9
- Kilgarrieff, A., Rychly, P., Smrz, P., and Tugwell, D. (2008). *The sketch engine*. *Pract. Lexicography*. ed. Fontenelle, T. Oxford: Oxford University Press, 297–306.

## Author contributions

JX: Writing – original draft, Writing – review & editing.

## Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

## Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.



- Kosnick, K. (2021). "Inclusive language pedagogy for (un) teaching gender in French" in *Teaching diversity and inclusion*. eds. E. N. Meyer and E. Hoft-March (New York: Routledge), 57–67.
- Liang, M. C., Li, W. Z., and Xu, J. J. (2010). *Using corpora: a practical coursebook*. Beijing: Foreign Language Teaching and Research Press.
- Lu, T. R., and Wang, L. J. (2021). A corpus-based noun SON/DAUGHTER collocation pattern analysis. *Overseas English* 6, 5–7.
- Macalister, J. (2011). Flower-girl and bugler-boy no more: changing gender representation in writing for children. *Corpora* 6, 25–44. doi: 10.3366/cor.2011.0003
- Mingst, K. A., Karns, M. P., and Lyon, A. J. (2022). *The United Nations in the 21st century*. New York: Routledge.
- Montiel, I., Cuervo-Cazurra, A., Park, J., Antolín-López, R., and Husted, B. W. (2021). Implementing the United Nations' sustainable development goals in international business. *J. Int. Bus. Stud.* 52, 999–1030. doi: 10.1057/s41267-021-00445-y
- Moore, J. A. Jr., and Pubantz, J. (2022). *The new United Nations: international organization in the twenty-first century*. New York: Routledge.
- Norberg, C. (2016). Naughty boys and sexy girls: the representation of young individuals in a web-based corpus of English. *J. English Linguist.* 44, 291–317. doi: 10.1177/0075424216665672
- Pearce, M. (2008). Investigating the collocational behaviour of man and woman in the BNC using sketch Engine1. *Corpora* 3, 1–29. doi: 10.3366/E174950320800004X
- Persson, G. (1990). *Meanings, models and metaphors: a study in lexical semantics in English*. Stockholm: Almqvist and Wiksell International.
- Pietilä, H. (2007). *The unfinished story of women and the United Nations*. New York: UN, Non-Governmental Liaison Service.
- Pietilä, H., and Vickers, J. (1990). *Making women matter: the role of the United Nations*. London: Zed Books.
- Rafferty, Y. (2013). International dimensions of discrimination and violence against girls: a human rights perspective. *J. Int. Women's Stud.* 14, 1–23.
- Reilly, N. (2009). *Women's human rights*. Cambridge: Polity Press.
- Rothermel, A.-K. (2020). Gender in the United Nations' agenda on preventing and countering violent extremism. *Int. Fem. J. Polit.* 22, 720–741. doi: 10.1080/14616742.2020.1827967
- Sigley, R., and Holmes, J. (2002). Looking at girls in corpora of English. *J. English Linguist.* 30, 138–157. doi: 10.1177/007242030002004
- Skard, T. (2008). Getting our history right: how were the equal rights of women and men included in the charter of the United Nations? *Forum Dev. Stud.* 35, 37–60. doi: 10.1080/08039410.2008.9666394
- Smokotin, V. M., Alekseyenko, A. S., and Petrova, G. I. (2014). The phenomenon of linguistic globalization: English as the global lingua franca (EGLF). *Proc. Soc. Behav. Sci.* 154, 509–513. doi: 10.1016/j.sbspro.2014.10.177
- Spiegel, J. (2013). *Sexual abuse of males: the SAM model of theory and practice*. New York: Routledge.
- Taylor, C. (2013). Searching for similarity using corpus-assisted discourse studies. *CORPORA* 8, 81–113. doi: 10.3366/cor.2013.0035
- United Nation. (2020). Guidelines for gender-inclusive language in English. Available at: <https://www.un.org/en/gender-inclusive-language/guidelines.shtml>
- Wang, L. J., and Chen, Y. E. (2016). A corpus-drive study: discourse construction of family members "father" and "mother". *J. Chuzhou Univ.* 18, 59–64.
- Watts, C., and Zimmerman, C. (2002). Violence against women: global scope and magnitude. *Lancet* 359, 1232–1237. doi: 10.1016/S0140-6736(02)08221-1
- Ziemski, M., Junczys-Dowmunt, M., and Pouliquen, B. (2016). The United Nations parallel corpus v1. 0. Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16) (pp. 3530–3534), Portorož, Slovenia. European Language Resources Association (ELRA).