Article

Computational Techniques for Analyzing Women's Social Change in Saudi Newspapers

Shrouq Almaghlouth [1,*] and Sultan Almujaiwel [2,*]

1   Department of English, College of Arts, King Faial University, Hofuf 31982, Saudi Arabia
2   Department of Arabic, College of Humanities and Social Sciences, King Saud University, Riyadh 11362, Saudi Arabia
*   Correspondence: salmaghlouth@kfu.edu.sa (S.A.); salmujaiwel@ksu.edu.sa (S.A.)

Abstract: This study utilized computational techniques for a reliable analysis of discourse. These techniques were adopted to analyze the progress of Saudi social change in terms of women's empowerment within the Saudi transformation program. The data from open source 2021–2022 Saudi newspaper archives were automatically crawled using cutting-edge computational techniques and structured according to the sections of the Saudi newspapers: front page, economy, international, sports, society, culture and religion. The analysis was based on computing the minimally uneven distribution of the relative frequencies of the occurrence of the central word (the Arabic forms of *woman* and *women*) from the years 2021 to 2022. This produced two samples of text data, each of which represented the respective years. Calculating the normalized and adjusted frequencies of the central word from each section in the data from each year was important to avoid unbalanced absolute frequencies in the qualitative analysis stage. In addition, dispersion measures showed that the amount of variance in terms of the lexical dispersion of the central word was not high. The observable facts from the quantitative analysis produced a more accurate observational sample of citations, which we qualitatively analyzed. The results of the latter showed a considerable ascending change in favor of empowering women as a consequence of Saudi Vision 2030.

Keywords: women's empowerment; social change; Saudi newspaper archives; statistical corpus linguistics; computational techniques

1. Introduction

The media has a tremendous impact on creating and circulating social constructs within any context; thus, analyzing women's empowerment representations as portrayed by newspaper sections may provide useful insight into deconstructing gender symmetry and asymmetry. In recent years, Saudi Arabia has carried out many transformative measures aimed at women's empowerment in the country as a result of its Saudi Vision 2030. Consequently, these measures resulted in a noticeable increase in women in previously male-dominated professions and within socio–cultural and eco–political domains. The media, whether in the traditional sense or its more modernized digital forms, has a tremendous and reciprocal impact on regulating discursive practices in any social and eco–political context. Consequently, the media plays a major role in producing and maintaining hegemonic ideologies in any given society (Sriwimon and Zilli 2017). Such ideologies encompass many social constructs, including identity constructions and gender representations. Considering the key role of the media in negotiating these, there is a constantly growing need in social science research to explore how media discourses interact within diverse contexts and which constructions are created, reproduced and circulated based upon such interactions.

The current paper adds to this rich body of research by highlighting how gender constructions are portrayed in Saudi newspapers from socio–cultural and eco–political perspectives. Keeping in mind the transformative measures taken in Saudi Arabia toward

women's empowerment in the last decade and the progressive nature of the sustainable 2030 vision, which was launched by the Saudi government in 2016. Originally, Saudi Vision 2020 had been proposed, but later it was expanded and incorporated within a broader nationwide scheme aimed at a national transformation program and has been officially referred to as Saudi Vision 2030 (Saudi Vision 2030 2022). As a comprehensive transformation program, Saudi Vision 2030 operates on a sustainable agenda that targets the strategic reduction of the Kingdom's dependence on oil, diversification of economic sources and improving the quality of life across diverse sectors. Due to such motivation, women's empowerment has emerged as a prominent theme within this vision, and in particular, women's contribution to the Saudi 2030 vision has been tremendously encouraged. Consequently, this has stirred up many changes, not only on an economic level but also on political, social and cultural grounds.

With this in mind, a discourse study in such a direction was needed, and this has led us to investigate Saudi societal interactions toward women-related themes. Our focus was on the Saudi newspaper articles published electronically in the two years after Saudi Vision 2020 as an initial stage of the wider 2030 Vision. This study explored woman-related themes by using corpus linguistic methods, which are an understudied area of research in Arabic discourse analysis. The sample text data from 2021 and 2022 Saudi newspapers were automatically crawled and structured from the Saudi newspapers archived on www.sauress.com (accessed on 1 December 2022). The main purpose of this study was twofold: to contribute to the critical discourse analysis (CDA) research by elucidating computational techniques that help to gather a homoscedastic variance of large-scale linguistic data and to utilize these techniques in investigating the progress of women's social change in Saudi Arabia.

2. Related Work

Representation is a recurrent theme in the discourse literature, with many works examining the presence and absence of diverse social actors in different discourses. In media discourse, for instance, the diversity and plurality of representations have often been popular points of examination (Brown et al. 1987). This is primarily because the diversity of media representations can be viewed as a political aim (Sjøvaag and Pedersen 2019) despite many recent works having argued that such diversity is rarely sufficiently translated into media representations, such as Kleemans et al. (2017) and Ross et al. (2013). Identity construction is a key issue in representation. Identity, in particular, has been repeatedly examined in discourse studies and is primarily based on the premise that 'identity is performatively constituted by the very expressions that are said to be its results' (Butler 1999, p. 25). This take on identity should highlight the socially constituted and constituting nature of discourse (Van Dijk 2015), as it simultaneously shapes context while being shaped by it. It also links the current study theoretically to elements of social constructionism since the texts under examination shape the representation of social constructs toward a particular socio–cultural or eco–political issue, which are then transported to the public as they are mediated through the media (Popa and Gavriliu 2015).

Research on the social construction of gender in discourse studies not only highlights the dynamicity of gender construction but also serves to problematize such a construction. Femininity, in particular, is a rather vague construct with overlapping components, which increase the risk of negative connotations in certain cultural contexts. For example, in a Turkish study, femininity was shown to clash with patriarchal masculinities and blend with other alternative narratives, such as modernization, independence and career success, to achieve respectability and self-value (Alemdaroglu 2015). This should be understood in light of the inherent connections created locally and regionally between femininity and feminism. Some examinations of feminism, both locally (Al Maghlouth 2017) and globally (Brooke 2020), reveal its negative construction, thus problematizing notions such as femininity and feminism to their audience. The problem has been worsened by the varying waves of feminism (Romano 2021), each of which constructs diverse and loaded ideologies

with opposing constructs that potentially contradict local cultural norms. Relevant to this, the recurrent issues of colonized feminism (Karimullah 2020) and first-world feminism (Alemdaroglu 2015)—in which stereotypical perceptions of Arab/Muslim women are forcibly cast by Western media discourses to serve colonial agendas—appear too often in the relevant literature so as not to create prejudice against feminism. As a result of such ideological grounding, this study can also be linked to works within critical discourse analysis (CDA). In CDA, the analyst is concerned with the power distribution in discourse (Van Dijk 2015), and by highlighting areas of unbalance, more awareness is created toward them. While gender asymmetry persists in language, many studies report diachronic changes toward leveling it up. For instance, Baker and Freebody (1989) report a negative representation of women in UK textbooks, whereas Wharton (2005) reports that while women are still less visible in the reading books in UK schools, they are represented as more capable due to such awareness. All this serves to highlight the potential of CDA in promoting more gender equality.

A recurrent theme in CDA, feminism and feminist discourse research is women's empowerment, which makes sense since all these enterprises center around power and its unbalanced distribution within any context. Journalism is also drawn to power (Wolfsfeld and Sheafer 2006), thus making exploring different reconstructions of power and empowerment in media discourse quite tempting for research. A preliminary step is understanding what empowerment is in this particular context. A common thread in the relevant literature (Elliott 2008; Kabeer 2005) is that women's empowerment is fundamentally based on a woman's ability to exercise choice over her life, especially concerning significant decisions such as marriage, education and work. Choice, in such cases, represents the tangible translation of power possessed by women. However, to validate women's empowerment, there should be some mechanisms to improve women's decision-making processes, as well as allow them more access to income, self-confidence and solidarity with other women (Kabeer 1999; Mayoux 1998). Thus, on its own, choice is never enough, and it should be paired with awareness, in particular, awareness of possibilities, which is promoted in most pro-women empowerment discourses (Khumalo et al. 2015). An examination of the research on women's empowerment reveals strong connections between women's empowerment on the one hand and access to employment on the other (i.e., feminized labor). Continuing the same line of thought, access to financial freedom and self-independence were prioritized in many works; for instance, in Termine and Percic (2015), as critical factors in women's empowerment and, consequently, in constructing a country's eco–political discourse (Sarfo-Kantankah 2021). This explains why, in many corpus studies that examined women's empowerment within eco–political discourses, terms such as 'employment' and 'work' were always prioritized as keywords; see, for instance, Grunenfelder (2013).

The plethora of research on discourse studies offers a multiplicity of analysis frameworks within gender discourse, most of which center around sexism and gender asymmetry. That being said, the current study adopted a corpus-based approach paired with diverse inspirations from discourse analysis or CDA. This combination is very common, as it allows for an empirical examination of discursive data (Biber et al. 2012), which, in turn, reduces the chances of subjective interpretations, which are occasionally attached to content-based qualitative works (Lee 2018). Approaching analysis from such a perspective allows an analyst to uncover hidden ideologies (Baker 2006) and attempt to impact the values and behaviors of other people (Partington et al. 2013) while exploring the analyst's hypotheses in practice (Baker 2014).

The vast majority of the relevant literature reveals a number of negative constructions of women and gendered discourse across the globe. Some of these corpus studies problematized such negative constructions from a macro lens using corpus methods, for instance, Almujaiwel (2017) and Coimbra-Gomes and Motschenbacher (2019), while others approached it in a more bottom-up fashion. By utilizing gender-based keywords, such as 'man', 'men', 'woman', 'women', 'girl' and 'boy', these studies examined diverse contexts. Earlier studies from nearly three decades ago (Kjellmer 1986) documented such nega-

tive constructions. Despite some improvement, subsequent works documented relatively parallel portrayals (Caldas-Coulthard and Moon 2010; Pearce 2008; Romaine 1999), be they in academic research discourse (Brooke 2020; Grunenfelder 2013), political discourse (Bakar 2014; Sarfo-Kantankah 2021), literature discourse (Eberhardt 2017) or even within discourse attempting to defend and promote women's empowerment, such as United Nations addresses (Brun-Mercer 2021).

Saudi Arabia has implemented plenty of transformative measures in the last two decades, which have escalated enormously since the launch of the Saudi 2020 and 2030 visions in 2016. Among its actively progressive missions, it has materialized women's empowerment plans tremendously in the hope that it can aid in reducing the gender gap from an eco–political perspective. This was translated into the noticeable increase in employment rates of women in male-dominated professions, which began subtly and gradually a little over a decade ago (Al Maghlouth 2017), yet had perceptibly flourished by the end of the same decade with the rapid increase in female leaders in the country (Alkhammash and Al-Nofaie 2020). For instance, the Human Resources Development Fund (HRDF 2023) was founded to publicize and promote Saudi women's employment in the private sector, along with its other primary motives. By the same token, Al-Munajjed (2010) reported in detail how this materialized in reality in recent times.

According to the official annual report issued by the national transformation program (SV2030), a noticeable increase has been detected in the number of working women from 21% in 2017 to 33.5% in 2021. Similarly, the report demonstrates an increase in the economic contributions made by Saudi women from 17% in 2017 to 34.1% in 2021. Along the same line, a parallel increase can also be detected when it comes to women in leading positions from 28.6% in 2017 to 39% in 2021 as a result of some initiatives specifically addressed to equip women with efficient leadership training. The transformation program also released among its plan two initiatives in support of women's empowerment at work, *Qurrah* and *Wusul.* The *Qurrah* initiative offered subsidized childcare to working mothers, reaching a total of more than 6645 female beneficiaries by the end of 2021. *Wusul*, on the other hand, offered subsidized transportation to more than 112,160 female beneficiaries. Due to such reforms, the change in women's representation within socio–cultural and eco–political domains was inevitable. Consequently, all of this attracted media attention globally (Elyas and Aljabri 2020), thus further highlighting the need to explore Saudi women's constructions on media platforms.

Deconstructing the portrayals of Saudi women should acknowledge the multiplicity of their identity to maximize the potential of the analysis. Ethnically, ideationally and regionally, Saudi women can be simultaneously constructed within broader definitions as Arab and Muslim women in addition to their national identity. Unfortunately, most of the relevant literature portrays negative constructions when it comes to the construction of Arab and Muslim women. To name but a few, Al-Hejin (2015), Karimullah (2020), Ruby (2013) and Saleh (2016) all reported negative constructions in which these women were constructed as submissive, passive, oppressed, subordinate and in need of help, often from Western agents. Such portrayals were consistent with a neo-orientalist construction of women (Saleh 2016), victimizing Arab/Muslim women and casting them as third-world women who should be urgently saved (Kahf 1999) to serve post-colonial interventionist agendas (Saleh 2016).

Comparably, constructions of Saudi women as a nationality reveal similar patterns, especially with analyses of English language media sources published in the internationally available literature. For instance, in his corpus-based analysis of Saudi women in BBC coverage, Al-Hejin (2015) documented a negative construction of the hijab (head veil in Islam) as an obstacle to the progress of Saudi women, citing examples in which refusing to wear the hijab was associated with successful businesswomen and female leaders. Similarly, in other media discourse analyses, Bashatah (2017) and Elyas and Aljabri (2020) revealed framing patterns within Western newspapers through which Saudi women were negatively constructed. In addition to this, Karimullah (2020) reported contrasting findings within a

self-built corpus between Saudi women on the one hand and Kurdish or Tunisian women on the other. In this corpus, Saudi women were cast as oppressed and non-agents, along with women in conflict regions, such as Yemen and Afghanistan, while Kurdish and Tunisian women were rather portrayed as active and idealized with empowered constructions of feminine agency. Interestingly, Mishra (2007) conducted a comparative study of American and Saudi newspapers, which revealed that Saudi women were portrayed along the same negative lines of passivation and oppression identified earlier in American newspapers, while American women were, in return, portrayed as superficial and immoral in the Saudi press. However, these same Saudi platforms reported rather positive and active constructions of Saudi women who were in charge of rejecting Westernization and maintaining their moral purity. The same positive and active thread of construction of Saudi women was also reported in another discourse analysis of Saudi newspapers (Elyas et al. 2021).

As is evident in this concise review, the vast majority of gender studies in media discourses were based on English data and/or data from the Western media, especially those with a corpus-based design (Sarfo-Kantankah 2021). This demonstrates the gap in the discourse literature in terms of targeting gender constructions in other languages, such as Arabic, in its investigations. Keeping in mind the scarcity of relevant research from Arabic corpora, the current study attempted to shed light on this gap in the hope that it adds to the growing research and helps to even out gender asymmetry in media discourse. It also attempts to promote awareness of this issue and explore whether the recent social and eco–political transformations in Saudi Arabia were mirrored in this discourse.

Such an urgent need stems from an understanding of the theoretical underpinning of the intricate relationship between discourse on the one hand and context on the other. This emerges as a fundamental theme in discourse studies under the premise that a mutually constituting relationship between these two exists (Fairclough and Wodak 1997) too often not to be missed. In particular, this could be linked to the theory of social constructionism and its emphasis on the role of communication in reality construction. To illustrate, such theorization highlights the discursive potential of linguistic construction in shaping social and cognitive constructs as well as their reciprocal status as a dynamic product of the same constructs (Van Dijk 2015). This, for instance, has been pointed out in the aforementioned comparison of women's representation in UK textbooks (in Baker and Freebody 1989 versus Wharton 2005). If it was not for the potential of such discourse analyses, these improvements could not have been made and consequently exposed and distributed among students.

Examining linguistic manifestations within discourse in search of evidence of social change utilizing a variety of semiotic parameters is a common practice in support of what is often classified as positive discourse analysis (Al Maghlouth 2017). According to Martin (2004), positive discourse analysis does not diverge from works within CDA but rather complements them. In particular, it seeks to embody evidence of resistance to the biased status quo through diverse linguistic tools (more on this in Section 3). In light of this, the main research question in the current study was as follows: to what extent do co-occurrences (collocate + central word) and their broad contexts indicate the discursive and social practices of the women-related themes that were introduced, promoted, or modified over the two years and the multiple sections? The answer to this question is provided in the results section. We arrived at the answer using prominent computational techniques that tested the study data retrieved from the Saudi newspaper archive for the years 2021 and 2022 and provided the distributions and dispersions of the central word and their collocations in a 2n-gram span from the 2021 and 2022 Saudi newspapers and across the seven sections of the newspapers' structure (front page, economy, international, sports, society, culture and religion). Further computational techniques were also adopted to show the amount of variance between the sample data from the two years using dispersion measures.

3. Data and Computational Techniques for Corpus Analysis

The data we collected represented two months from each of the years 2021 and 2022. The data used in this paper were crawled from the Saudi archived newspapers website www.sauress.com (accessed on 1 December 2022), and we assigned two months from each year, due largely to these being the two months in which the intended central word as the nodes المرأة(woman) and النساء(women) appeared more frequently. This was detected using an advanced search we conducted on the website itself. Thus, our sample was from February and March for the year 2021, and from January and February for the year 2022. The Saudi archived newspapers included 43 local newspapers. This archived platform publishes daily news and articles, which helped with the task of simultaneous crawling and structuring.

The relevant literature proposes many linguistic manifestations to underpin the intricate interplay between discourse on the one hand and gender representations on the other; to name a few: semantic macro structures (Al-Hejin 2015), metaphor analysis (Al Maghlouth 2021), titles, (Alkhammash and Al-Nofaie 2020) social actor representation (Almaghlouth 2022), multimodality (Alkhammash 2022) and process type analysis (Koller 2012). The current study was designed based on minute examination of collocations. Collocations are pairs or groups of words that often come together within the same near-linguistic context (Baker 2006). Due to such linguistic proximity, collocations are often examined as evidence of Moscovici's (2000) mental representations since their co-occurrence often suggests that they tend to correlate in cognition as well. Against this backdrop, it is hypothesized in the current study that the aforementioned measures taken by the Saudi government in support of women's empowerment can be traced back to the newspapers' corpus at hand in the form of more empowered/positive women's representations. In that sense, utilizing collocations as an inbuilt tool in corpus processing has been quite informative within many gender studies (see, for instance, Almujaiwel (2017), within a gender-based Saudi context). The reviewed literature, especially from non-Saudi newspapers, has confirmed the negative representations attached to Saudi women within such discourse. By investigating the collocational behaviors of the intended central words, this study might be able to detect a corresponding linguistic change of such representations, especially keeping in mind the reformative changes made in support of more women's empowerment.

As the focus of our analysis was on the collocational behaviors of the intended central word: *woman* (and its plural form: *women*), in the sections of the newspaper structures, information regarding the number of texts in each local newspaper was unnecessary. According to our linguistic raw data, Tables 1 and 2 show the number of texts (files), each of which contained a specific article, and the number of types (unique word forms) and tokens (all running words) across the respective years. The tables show the numbers of texts, types and tokens in the seven sections.

Table 1. Basic statistical information about the 2021 Saudi newspapers.

No. of	Front	Economy	Intl.	Sports	Society	Culture	Religion	Total
Articles	207	1705	1827	1119	1421	681	308	7268
Types	10,101	41,137	45,343	31,113	27,334	45,621	16,754	129,678
Tokens	33,270	379,147	336,037	206,511	225,369	220,740	79,510	1,480,584

Table 2. Basic statistical information about the 2022 Saudi newspapers.

No. of	Front	Economy	Intl.	Sports	Society	Culture	Religion	Total
Articles	1674	1162	1119	757	344	413	78	5547
Types	59,477	40,447	37,779	25,749	12,827	27,859	6689	118,760
Tokens	367,429	301,206	209,319	132,825	63,188	104,977	21,389	1,200,333

The computational techniques adopted for the corpus/data analysis were as follows: first, the raw and normalized frequency analysis of the central word between the two years; second, the raw and normalized frequency analysis of the intended lexical bundles (5n-grams collocation window) between the two years; third, the raw and normalized frequency analysis of the intended lexical bundles between the two years and the newspaper section data; and fourth, the raw and normalized frequency analysis of the central word between the multiple newspaper sections data. The terms and their concepts needed to be well-defined for the sake of clarity. Such clarity paved the way for explaining the computational techniques used for our data analysis. These statistical corpus linguistics terms were analyzed in terms of the normalized frequency—corpus size *n*, relative frequency *rf* and normalization base *nf* (McEnery and Hardie 2012, pp. 50–51)—and dispersion—the standard deviation SD, coefficient of variance *CV* and Juilland's *D* (Brezina 2018, pp. 46–53).

The corpus size *n* is the total number of all running tokens/words in a given corpus. The raw frequency *f* of a word is the absolute number of *f* in a given corpus. The relative frequency *rf* is the number of times a word occurs divided by *n*. The normalized frequency *nf* is simply the result of *f* /*n* × *normalization base*. The normalization base *nf* is a number to be set on average, and they always follow a numerical pattern that starts with 1 followed by zeros. For example, if *n* equals any number in a format of tens of thousands (27,000, 76,938, 89,002 and so on), the *nf* will be set to the base 10,000 or less (1000 or 100) according to the preference of the human analyst. The corpus-to-corpus ratio (*nf* 1/*nf* 2) is computed to provide the number of times the word occurs in a corpus compared with another corpus. It gives the difference ratio between the *nf* over the multiple corpora or sub-corpora. The benefit of using the *nf* measure is to avoid the unevenly distributed number of *f* for a given word while its *nf* is low. For example, in our data, the *f* values for the central words *woman* and *women* between the culture and religion sections for the sample of the year 2021 were 220,740 and 79,510, respectively, but the *nf* (×10,000) values were 6.025 and 8.427, which means that the *nf* in the *religion* section was more than that in the *culture* section.

Dispersion is different and more accurate than *distribution* when it comes to the whole corpus. The parts of a given corpus are simply the *nf* (normalization base). Distribution is usually lacking in denseness, while dispersion tells us how the relative frequencies of a word's per normalization base (×100, ×1000, ×10,000 and so on) are agglomerated between the parts of the intended corpus and on the whole. The merit of dispersion is that it is a set of measures that output the variation within different parts of the corpus. The measures utilized are the coefficient of variance *CV* and Juilland's D. The coefficient of variation *CV* is simply calculated as follows: $CV\ (word) = \frac{standard\ deviation}{mean}$. The result of *CV* is then divided by the square root of the number of parts minus 1: $CV = \frac{CV\ (word)}{\sqrt{number\ of\ the\ corpus\ parts - 1}}$. The final result comes out between 0 and 1. When the *CV* is closer to 0, the given word is more evenly distributed throughout the parts of the corpus. As for Juilland's D, it is a measure that depends on *CV* and is based on the following formula: $Juilland's\ D = 1 - \frac{CV}{\sqrt{number\ of\ the\ corpus\ parts - 1}}$. The result of Juilland's *D* is also between 0 and 1. When it is close to 1, the distribution of the given word is perfect over the parts of the corpus. This means that *CV* and Juilland's *D* are opposite in terms of reporting a perfectly or imperfectly even distribution.

4. Results

In this section, the observable facts are described. The data were analyzed using two main computational techniques. First, the normalized frequency *nf* was used to adjust the frequencies of the central word across two years and among the newspaper sections over the two years. Second, the dispersion measures were used to demonstrate the fairly even distributions of the relative frequencies of the central word across the two years and over the multiple newspaper sections to increase the confidence of how good our sample was in terms of homoscedasticity (the equality of variances in two groups) for qualitatively analyzing the discursive practices of women's social change within the contexts of the

collocations in the 2021 and 2022 Saudi newspapers. Utilizing these techniques allowed us to recognize the scales of the women-related themes and perform qualitative analyses of the intended examples from the data.

4.1. Normalized Frequency and Dispersion Measures

As explained earlier regarding *nf*, comparing the data sizes of the years 2021 and 2022 showed that the text data of the former were larger than the latter (Table 3), even in terms of the *f* and *nf* of the central word. However, comparing the *nf* of the sections of each year was the touchstone.

Table 3. Raw vs. normalized frequency of the central word for the two years.

	N	*f*	*nf* ($\times$**100**)
2021	1,480,584	669	0.045
2022	1,200,333	246	0.020

Tables 4 and 5 include the *n*, *f* and *nf* (*rf* $\times$ 100) of the central word across the seven sections. The notation (*rf* $\times$ 100) denotes the relative frequency of the word occurring per 100 words in all seven sections of the data from each year. This was applied in each section to show the more realistic proportional occurrences of lexical items from each part of the corpus. The *nf* of the central word for the sampled data of the year 2021 showed that it was higher in the *religion* section than in the remaining sections. As for the year 2022, the *nf* was found to be higher in the same section than in the remaining sections, except for the *front page* section. The *nf*, therefore, was the pivot around which the adjusted *nf* evenly redistributed the central word over the multiple sections, and it allowed us to avoid becoming confused by the *n* and *f*. Table 6 reports the section-to-section ratios and Table 7 shows the descending order of the adjusted frequencies of the central word for the sections according to size. Regarding the *section-to-section ratios* of the two sets of sampled data for the respective years, the difference in *nf* sizes varied, but the overall section-to-section ratio of the sampled data for the years 2021 (*N* = 1,480,584) and 2022 (*N* = 1,200,333) was 0.175. This meant that we are confident regarding the qualitative analysis of the contexts of the central word (women-related themes) for the two years.

Table 4. Raw vs. normalized frequencies of the central word in multiple newspaper sections for the year 2021.

Sections (A-0)	*N*	*f*	*nf* ($\times$**100**)
Front page (F-1)	33,270	5	0.015
Economy (E-2)	379,147	277	0.073
International (I-3)	336,037	149	0.050
Sports (S-4)	206,511	19	0.009
Society (O-5)	225,369	19	0.008
Culture (C-6)	220,740	133	0.060
Religion (R-7)	79,510	67	0.084

Table 5. Raw vs. normalized frequencies of the central word in multiple newspaper sections for the year 2022.

Sections (A-0)	*N*	*f*	*nf* ($\times$**100**)
Front page (F-1)	367,429	131	0.036
Economy (E-2)	301,206	42	0.014
International (I-3)	209,319	38	0.018
Sports (S-4)	132,825	10	0.008
Society (O-5)	63,188	2	0.003
Culture (C-6)	104,977	16	0.015
Religion (R-7)	21,389	7	0.033

Table 6. Corpus-to-corpus (section-to-section) ratio for the 2021 and 2022 samples.

Sections (A-0)	2021	2022	Section-to-Section Ratio
	nf (×100)	*nf* (×100)	
Front page (F-1)	0.015	0.036	−0.021
Economy (E-2)	0.073	0.014	0.059
International (I-3)	0.050	0.018	0.032
Sports (S-4)	0.009	0.008	0.002
Society (O-5)	0.008	0.003	0.005
Culture (C-6)	0.060	0.015	0.045
Religion (R-7)	0.084	0.033	0.052

**Table 7.** Descending order of *nf* (×100) of multiple sections for the 2021 and 2022 samples.

Section	2021		Sections	2022	
	f	*nf* (×100)		*f*	*nf* (×100)
Religion (R-7)	67	0.084	Front page (F-1)	131	0.036
Economy (E-2)	277	0.075	Religion (R-7)	7	0.033
Culture (C-6)	133	0.06	International (I-3)	38	0.018
International (I-3)	149	0.050	Economy (E-2)	42	0.014
Front page (F-1)	5	0.015	Culture (C-6)	16	0.015
Sports (S-4)	19	0.009	Sports (S-4)	10	0.008
Society (O-5)	19	0.008	Society (O-5)	2	0.003

Regardless of the *n* and *f* of the central word for the sampled data of the two years, the *nf* (per 100 words) provided the relative distributions of the central word for the sections of the two years, as visualized in Figure 1, where the boxplot shows the *nf* of the central word for the sections, and the red line inside the box is the median of the *nf* (per 100 words). The *nf* values of the central word in sections E-2, C-6 and R-7 were larger. Smaller values can be seen for the *international*, *front page*, *society* and *sports* sections. These adjusted frequencies were essential when analyzing the discursive practices and constructs of the women-related themes in the sections for the two years, and they helped to avoid the bias that occurs when relying solely on the raw frequencies. By looking at Table 4 (the raw frequencies F of the central word for the seven sections in the 2021 data), we can state that if we relied solely on *F*, we would conclude that the central word occurred from higher to lower frequencies in the *economy*, *international*, *culture*, *religion*, *sports*, *society* and *front page* sections. This made it difficult to judge the examples extracted for the qualitative analysis when we looked at the co-occurrences of the central word in terms of the social practices in the discourses from the data, even if the examples representing the raw frequencies were small in number. The order of sections in terms of the adjusted frequencies *nf* was different, as it shows that the central word occurred from higher to lower adjusted frequencies in the *religion*, *economy*, *culture*, *international*, *front page*, *sports* and *society* sections. The same operation was conducted for the seven sections in the 2022 data (Table 5). The *nf* order of sections from higher to lower for the two years is given in Table 7.

Dispersion measures demonstrate the degree to which the data are scattered. This shows whether the data are homogenous or heterogeneous. The dispersion measures used herein were the relative measures of dispersion: the mean of the relative frequency *rf* × 100 (the *nf* of the central word for the seven sections in the data from each year), the standard deviation SD, the coefficient of variance and the powerful measure of dispersion known as Juilland's D. The outputs of these measures are given in Tables 8 and 9.

There were seven data parts, which corresponded to the number of sections in the data from each year. The interpretation of the sampled data was based on the values (between 0 and 1) of the coefficient of variance *CV* and Juilland's D. In Tables 8 and 9, the *CV* values were both closer to 0 than to 1, meaning that the amount of variance was small. This can be

further demonstrated by calculating Juilland's D, the value of which was close to 1 for both years, meaning that the central word was distributed fairly evenly.
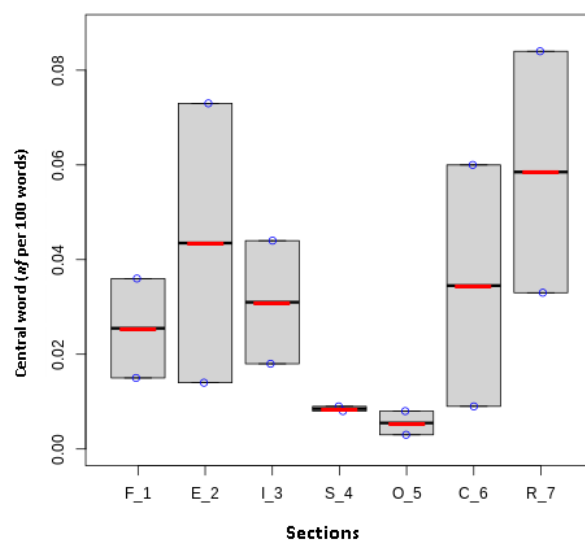
Figure 1. The adjusted frequencies *nf* (*rf* × 100) for the different sections.

Table 8. Dispersion analysis of the central word among the sections for the year 2021.

Mean *rf*	SD	CV	Juilland's D
0.04	0.03	0.28	0.72

Table 9. Dispersion analysis of the central word among the sections for the year 2022.

Mean *rf*	SD	CV%	Juilland's D
0.02	0.01	0.25	0.75

4.2. Normalized Frequencies of Co-Occurrences

Newspaper articles (opinions and editorials) and news are an instance of discourse, where discursive events are affected by social practice, and the latter has an impact on shaping the discursive practice (Fairclough 1992). We obtained the complete excerpts/examples of the co-occurrences concerning women from our quantitative data, and the social practices were excavated from all those examples to unveil the progress of women's empowerment.

The application of the 2n-grams of the central word, excluding the grammatical items to ensure that all the co-occurrences appeared correctly, was the first step for the qualitative observations of the co-occurrences and their broad contexts. Not excluding the grammatical items produced a long list of co-occurrences and removing them reduced the size of the list. This is feasible as long as such items will appear when extracting the citations of the co-occurrences for further contextual interpretations. What we processed reproduced the instantiations of social practices toward women-related themes and rhemes. What we arrived at afterward are examples of the co-occurrences. Table 10 presents the highest absolute frequency collocates associated with the central word, in addition to the sections in which they appeared. It is noticeable that the absolute frequency of the co-occurrences was higher in 2021 than in 2022. However, as we were following the perspective of the *nf* base, and *rf* in particular, some co-occurrences were more frequent in 2022 than in 2021. That is, the normalized frequencies of تمكين, إنجاز, حقوق, عمل and دور were higher in 2022.

Table 10. Absolute frequencies (*f*) of the co-occurrences for 2021 and 2022.

Collocate	2021 *nf*	2021 Sections	2022 *nf*	2022 Sections
Empowering تمكين	0.0015	A-0	0.0052	F-1, E-2, S-4, C-6
Position مكانة	0.0005	E-2	0.0004	E-2
Support دعم	0.0006	E-2, C-6, I-3	0.0001	E-2
Participation مشاركة	0.0012	E-2, C-6, R-7	0.0013	F-1, E-2, S-4
Work عمل	0.0002	E-2, I-3	0.0005	F-1
Against ضد	0.0007	E-2, I-3, C-6	0	
Rights حقوق	0.0005	E-2, I-3, C-6, R-7	0.0012	F-1, I-3
Role دور	0.0006	E-2, I-3, O-5, C-6, R-7	0.0012	F-1, E-2, S-4
Issue قضية	0.0005	E-2, I-3, R-7	0	
Accomplishment إنجاز	0.0010	E-2, S-4	0	
Status وضع	0.0006	I-3	0.0004	F-1
Capabilities إمكانات	0.0002	I-3	0.0000	
Protection حماية	0.0010	I-3, R-7	0.0001	F-1, E-2, I-3
Strengthening تعزيز	0.0006	I-3, R-7	0.0003	E-2, S-4
Body جسد	0.0003	O-5, R-7	0	

Noticeably, the practicality of the *nf* base of each co-occurrence reflected in *rf* (×100) is an orthodox method for undertaking a reasonable analysis of the co-occurrences that reflect the progress of the social change toward women-related themes. Some negative co-occurrences from opponents terminated in the year 2022, namely, ضد (against) and جسد (body), after they occurred in the year 2021 in the *economy* (E-2), *international* (I-3), *culture* (C-6), *religion* (R-7) and *society* (O-5) sections. These two collocates were the only ones found to be negative in the year 2021. The remaining 13 collocates in their co-occurrences were positive. This indicated swift progress in social practices constructed toward women-related themes, keeping in mind the theoretical and methodological grounds highlighted earlier in this study.

The same method was applied to the verbs (Table 11) associated with the central word and the genitive/adjective construction of the co-occurrences for the residuals of the examples (Figures 2 and 3). The verbs found to be negative from their co-occurrences posed by opponents in the sampled data of the years 2021 and 2022 were تستهلك (she consumed), تغار (she is jealous) and أغرى (he tempted her to), which were found in the *front page* (F-1), *international* (I-3), *culture* (C-4) and religion (R-7) sections.

Table 11. Verbal co-occurrences *rf* = (>0.00001 × 100) for 2021 and 2022.

Collocate	2021 *rf* (×100)	Sections	2022 *rf* (×100)	Sections
To have received حظيت	0.00028	C-6, F-1	0.00017	E-2, F-1
To have practiced مارست	0.00021	C-6	0	-
To have enabled مكنت	0.00021	E-2	0	-
To receive تحظى	0.00021	C-6	0	-
To be given منحت	0.00007	S-4	0	-

Table 11. *Cont*.

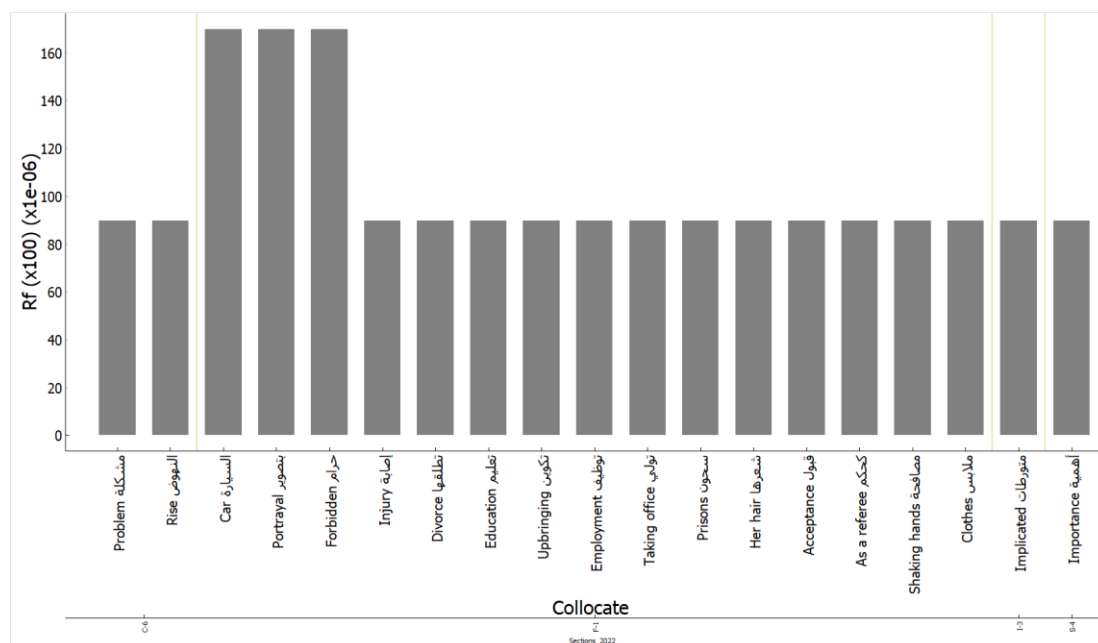| Collocate	2021 *rf* (×100)	Sections	2022 *rf* (×100)	Sections
To have gained نالت	0.00007	C-6	0	-
To enjoy تتمتع	0	-	0.00009	E-2
To find تجد	0	-	0.00009	F-1
She consumes more تستهلك	0	-	0.00034	F-1
To be forced نقحم	0	-	0.00017	I-3
To be weakened تستضعف	0.00014	I-3	0	-
She was forced قهرت	0.00014	C-6	0	-
To be exploited تستغل	0.00014	R-7	0	-
Requesting (her) يُطالبون	0.00014	R-7	0	-
To be jealous تغار	0	-	0.00009	F-1
Tempted (her) أغرى	0.00007	R-7	0	-
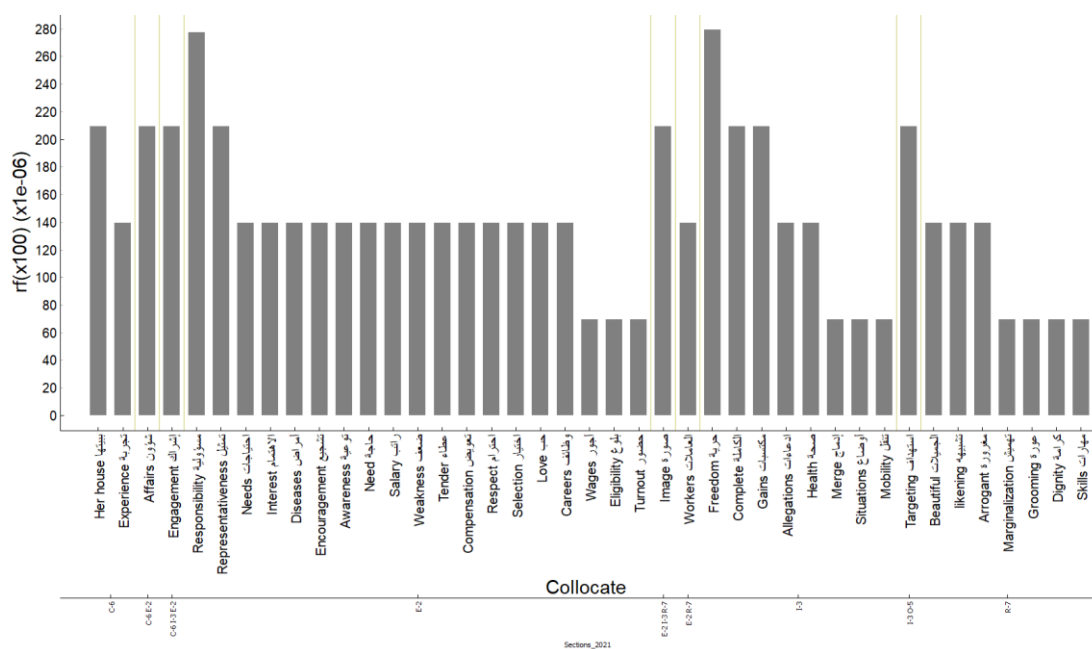
Figure 2. Bar plot of the genitive/adjective construction co-occurrences *rf* = (>0.000001 × 100) for 2021.

The results of the remaining co-occurrences are shown in Figures 2 and 3 and are given according to the relative frequencies in scientific notation due to them being too small. For example, number (200) means that its *rf* is based on 1×10^{-6} (0.000001), and as a result, the number 200 corresponds to a value of 0.000200. The *nf* (*rf* × 100) for the collocate مسؤولية, for instance, was exactly 0.000278 in the output file after the automatic calculation. The collocates that were found to be negative in the remaining co-occurrences were conveyed by مغرورة (image/likening), تشبيه (targeting), استهداف (allegations), ادعاءات بتصوير (forbidden), حرام (car), السيارة (grooming), عورة (marginalization), تهميش (arrogant), المتورطات (who are implicated) and مصافحة (shaking hands), شعرها (her hair), مصافحة (portrayal).

Figure 3. Bar plot of the genitive/adjective construction co-occurrences *rf* = (>0.000001 × 100) for 2022.

4.3. Instantiations of Social Practices

In this section, we present the results in a broader context, meaning that the examples extracted herein were analyzed in their contexts by considering the sections they were mentioned in. The women-related themes and their progress after Saudi Vision 2020 and the constructivist social changes that took place in favor of or against women-related themes were manually extracted. We grouped the co-occurrences and their broad contexts into two facets of social practice constructions, namely, proponents (progressive supporters) and opponents. The former denoted further support and real positive achievements in empowering women, while the latter represented opponents of women's empowerment. In Tables 12–14, the number of positive (supported by proponents) and negative (raised by opponents) examples of co-occurrences is 74 positive and 17 negative co-occurrences according to their broad contexts.

Table 12 contains the examples and the broad contexts of the genitive co-occurrences whose *nf* values were more than 0.0001 per 100 words, and Table 13 contains the examples and the broad contexts of the verbal co-occurrences whose *nf* values were more than 0.00001 per 100 words. Tables 12 and 13 show that a few negative examples were categorized under *opponents*. At a glance, the examples of *positive* genitive and verbal co-occurrences (proponents) were much more frequent than *negative* ones (opponents). Table 14 shows the examples and the broad contexts of the co-occurrences whose *rf* values were less than 0.000001 per 100 words. As for the co-occurrences exemplified therein, negativity increased.

A few examples of negativity from opponents were found in the contexts of the *international* (I-3), *culture* (C-6), *society* (O-5) and *religion* (R-7) sections in Table 12, and in the contexts of the *culture* (C-6), *front page* (F-1), *international* (I-3) and *religion* (R-7) sections in Table 13. The reaming co-occurrences did not match between the 2021 and 2022 data, which resulted in them being presented separately in Table 14. Therein, a few examples of negativity can be seen in the contexts of the *international* (I-3), *society* (O-5), *economy* (E-2), *international* (I-3) and *religion* (R-7) sections in the 2021 data and in the contexts of the *front page* (F-1) and *international* (I-3) sections in the 2022 data. No negativity was detected in the context of the *sports* (S-4) section in either year or in the context of the *economy* (E-2) section in the year 2022.

Table 12. Broad contexts of the genitive construction co-occurrences (*rf* = >0.00001 × 100).

Genitive Construction Co-Occurrences	
Proponents	(1) تمكين: ممارسة الرياضة, من العمل, لرؤية السعودية, من المناصب, في الوزارة, الكفؤة, في الريادة. (2) مكانة: تحسين, إظهار, إصلاح. (3) دعم: برنامج متكامل, ومساعدة, قدراتها, ومساعدتها في صقل الشاعرة, استقرار المنطقة, في مجال العلوم. (4) مشاركة: في كل القطاعات, تعزيز, الانتخابات البلدية, سوق العمل, في منظومة التغيير, في القوى العاملة, العمل بجدية, رفع سقف, ازدادت. (5) عمل: عدم استمراريتها من نفسها, عدم جواز العمل لساعات طويلة, الحق في العمل, التمكن من الابتكار. (6) حقوق: العمل, في تعزيز دورها في التنمية, في ممارسة حقوقها, تهميش دورها. (8) قضية: العنف ضد المرأة, حماية المرأة الشاكية, المرأة الشائكة. (9) إنجاز: في الطب, في علاج السرطان, في الفنون, براءات الاختراع. (10) الكرامة, المجال السياسي. (11) تبيين, إظهار: وضع. (7) دور: في تعزيز العمل, إمكانات: كشريك في المجتمع, الاستفادة القصوى من إمكاناتها. (12) حماية: حقوقها, من العنف. (13) تعزيز: مكاتها, دورها في علم البيانات, دورها في التنمية.
	(1) Empowering: in sports, work, for Saudi's vision, in leaderships, in ministerial positions, in efficient, entrepreneurships (A-0) (F-1) (E-2) (S-4) (C-6). (2) Position: improving, supporting, rectifying (E-2). (3) Support: integrated program, assistance, help in refining, abilities, the poetess, the stability of region, in the field of sciences (E-2) (C-6) (I-3). (4) Participation: in all sectors, enhancing, municipal elections, labor market, in the system of change, in work forces, working hard, raise the ceiling, increased (E-2) (C-6) (R-7) (F-1) (S-4). (5) Work: not continuing due to her own decision, not allowed to work for long hours, the right of work, be able to innovate (E-2) (I-3) (F-1). (6) Rights: work, participate in politics, dignity (E-2) (I-3) (C-6) (R-7) (F-1). (7) Role: in enhancing work, support her role, in development, in exercising her rights, marginalizing her role (E-2) (I-3) (O-5) (C-6) (R-7) (F-1) (S-4). (8) Issue: violence against, protecting women, the complaining woman, the thorny woman (E-2) (I-3) (R-7). (9) Accomplishment: in medicine, in cancer treatment, in arts, patents (E-2) (S-4). (10) Socio–economic status: disclosure, clarifying (I-3) (F-1). (11) Capabilities: as a partner in society, making the most of her potential (I-3). (12) Protection: her rights, against violence (I-3) (R-7) (F-1) (E-2). (13) Strengthening: her position, her role in data science, her role in development (I-3) (R-7) (E-2) (S-4).
Opponents	(1) ضد: المرأة, الزوجة. (2) جسد: مثار إعجاب للرجل.
	(1) Against: woman, wife (I-3) (C-6). (2) Body: seductive for man (O-5) (R-7).

Table 13. Broad contexts of the verbal construction co-occurrences (*rf* = >0.00001 × 100).

Verbal Construction Co-Occurrences	
Proponents	(1) حظيت: بمكانة عالية في عهد محمد بن سلمان, برعاية ورعاية, باهتمام. (2) مارست: التجارة. (3) مكنت: المجال الهندسي, التحول الرقمي. (4) تحظى: بعصرها الذهبي, بالدعم الكامل. (5) منحت: المزيد من الحقوق. (6) نالت: حظها. (7) تتمتع: بكامل حقوقها, بتجارب مميزة. (8) تجد: دورها الفعال. (9) نقحم: لم نقحمها في الجيش فالتجنيد اختياري. (10) تستضعف: في العمل. (11) قهرت: بقوانين صارمة كبلت قدراتها. (12) تُستغل: في العمل, لحاجتها إلى العمل. (13) يُطالبون: بأن تستحي.
	(1) To have received: a high position in the age of Muhammed bin Salman, with care and maintenance (C-4) (F-1) (E-2). (2) To have practiced: trading (C-6). (3) To be enabled: engineering field, digital change (E-2). (4) To receive/has or have: her golden age, the whole support (C-6). (5) To be given: more rights (S-4). (6) To have gained: her chance (C-6). (7) To enjoy: with whole her rights, with distinct experiences (E-2). (8) To find: her efficient role (F-1). (9) Not to be forced: not to be forced by the arm, for enlistment is optional (I-3). (10) To be weakened: in work (I-3). (11) She was forced by strict laws that undermined her capabilities (C-6). (12) To be exploited: in work, for her need to work (R-7). (13) Requesting her to be ashamed (R-7).
Opponents	(1) تستهلك: أكثر. (2) تغار: لدوافع ماديّة. (3) أغرى: على الهرب.
	(1) She consumes more (F-1). (2) To be jealous for material motives (F-1). (3) Tempted (her) to escape (R-7).

Table 14. Broad contexts of the genitive/adjectival construction co-occurrences (nf = <0.00001 × 100).

Genitive/Adjectival Construction Co-Occurrences	
Proponents	(1) اعتنت: ببيتها (2) تجربة: بالقيادة (3) شؤون :تحفيز ,تولي كامل والأسرة. (4) إشراك :في الأنشطة السياسية في صنع القرار (5) مسؤولية: عائلتها. (6) تمثيل :في المحافل الدولية في السلك القضائي (7) احتياجات :المجتمعية الاهتمام (8) بقضايها (9) أمراض :وأورام والولادة (10) تشجيع: على تطوير مهارتهن ,مهارتها في البرمجة الشعر النسائي (11) توعية :بحمايتها المجتمع حاجة (12) :للعمل إلى الوظيفة عطاء (13). المرأة أكبر من الرجل (14) تعويض :المرأة ماليا (15) احترام :وتقدير جهودها ,حريات النساء وحبها المرأة اختيار (16) :لوظيفتها في التجنيد (17). وظائف :تختارها ,يتركنها عند الحمل (18) واحترامها :حب (19) أجور :مساواتها مع أجور الرجل بلوغ (20) الأحقية كبلوغ الرجل حضور (21): يساوي حضور الرجل في قطاعات الأعمال (22) صورة :انتقاصها وعدم تصحيحها العاملات (23) عددهن في مجالات الأعمال حرية (24): الكاملة :والفعالة في المشاركات (25) في اللباس وارتداء النقاب مكتسبات (26): تعظيمها ,حمايتها ,تعزيزها صحة (27): ادعاءاتها تعزيز صحتها وصحة الطفل إدماج (28) وتمكينها في عمليات صنع القرار أوضاع (29): تحقيق المساواة تنقل (30): في مجالات العمل وريادة الأعمال الجميلات (31): يعتقدن بأنهن يسحرن كرامة (32): عدم النيل من كرامتها وإنسانيتها وعقلها وحقوقها مهارات: تطويرها وإتاحة الفرص الاستثمارية (34) راتب :أقل من الرجل ضعف (35): الرواتب ,المادي والمعنوي مشكلة (36) جذب الانتخابات النهوض (37): بأوضاعها لتحقيق المساواة إصابة (38) في الملعب تطليقها (39): المحكمة من زوجها تعليم (40) تمكينها في التعليم تكوين (41): الخَلقي وما يتناسب معه توظيف :تعظيم فرص التوظيف لها في عهد الملك سلمان (43) تولي :المناصب سجون (44): أهمية توظيف النساء في سجون النساء قبول (45): عدم قبول الحامل في الجيش بحكم (46) في البطولات الرياضية ملابس (47): أهمية: دورها في التنمية ,في السجون اختصاص بيعها من قبل النساء (48) في المجتمع وفي السجون.

| | (1) Her house: take care of (C-6). (2) Experience: in leadership (C-6). (3) Affairs: motivation, be responsible for all, the family (C-6) (E-2). (4) Engagement: in political activities, in decision making (C-6) (I-3) (E-2). (5) Responsibility: her family (E-2). (6) Representativeness: in the international forums, in the judiciary (E-2). (7) Needs: social needs (E-2). (8) Interests: in her issues (E-2). (9) Diseases: tumors and childbirth (E-2). (10) Encouragement: to develop their skills, her skills in programming, women poetry (E-2). (11) Awareness: to be aware of society to protect her (E-2). (12) Need: need for a job, for work (E-2). (13) Tender: woman is more tender than the man (E-2). (14). Compensation: compensate woman financially (E-2). (15) Respect: to appreciate her efforts, women freedoms, woman and her love (E-2). (16) Selection: for her job, for enlistment (E-2). (17) Love: to respect her (E-2). (18) Careers: her selection, leave careers at pregnancy (E-2). (19) Wages: to be equal to man wages (E-2). (20) Attainment of eligibility: eligibility as of a man when reaching maturity (E-2). (21) Turnout: in business sectors, her turnout is equal to man's (E-2). (22) Image: to be corrected and not degraded (E-2) (I-3) (R-7). (23) Female workers: their numbers in business fields (E-2) (R-7). (24) Freedom: in dressing and wearing the veil (niqab)(I-3). (25) Complete: efficient participations (I-3). (26) Gains: greatening, protecting, enhancing (I-3). (27) Health: her allegations, enhancing her health and her child's (I-3). (28) Merge: empowering her in the processes of making decisions (I-3). (29) Situations: achieving equality (I-3). (30) Mobility: in business field and entrepreneurships (I-3). (31) Beautiful women: they think that they can enchant (R-7). (32) Dignity: not undermining her dignity, her humanity, her mind and her rights (R-7). (33) Skills: develop skills and provide investment opportunities (R-7). (34) Salary: less than man's wage (E-2). (35) Weakness: salaries, material and physical (E-2). (36) Problem: attract elections (E-6). (37) Rise: improving her conditions to achieve equality (E-6). (38) Injury: in the playground (F-1). (39) Divorce: court divorces her from her husband (F-1). (40) Education: empowering her in education (F-1). (41) Upbringing: her character and what corresponds with it (F-1). (42) Employment: valuing employment opportunities for her in the age of the king, Salman (F-1). (43) Taking office: positions (F-1). (44) Prisons: the importance of employing women in prisons (F-1). (45) Acceptance: accepting in the army, not accepting pregnant women in the army (F-1). (46) As a referee: in the tournaments (F-1). (47) Clothes: to be sold only by women (F-1). (48) Importance: her role in society, development and prisons (S-4). |

| Opponents | (1) ادعاءات :صحتها ضد الزوج (2) استهداف :النساء (3) تشبيه :من لا تشبه النساء (4) التنمر على مغرورة (5) بجمالها تهميش: دورها وتقليصه. (6) عورة :وجها (7) حرام :السيارة ,إسقاط ولايتها ,قيادتها حرام الخلو مع الرجل حرام (8): التعليم ,المصافحة ,صبغ الشعر ,الزنا ,قيادة السيارة بتصوير (9): النساء في الشارع حرام :شعرها (10) مصافحة :حرام (11) المتورطات (12): في قضايا سياسية ,دخولها للمقار الحكومية ,عرض الملابس الداخلية. |

| | (1) Allegations: her health, against husband (I-3). (2) Targeting: women (I-3) (O-5). (3) Image/likening: bullying those who are not like women (E-2) (I-3) (R-7). (4) Arrogant: about her beauty (R-7). (5) Marginalization: her role and shrink it (R-7). (6) Grooming: her face (R-7). (7) Car: driving car by her is forbidden (haram), to topple her rule is forbidden (haram) (F-1). (8) Haram (forbidden): be in seclusion (khalwa) with a man, education, shaking hands, dying of hair, adultery, driving a car, entering government places, show underwear clothing (F-1). (9) Portrayal of women: women in the street (F-1). (10) Her hair: forbidden (haram) (F-1). (11) Shaking hands: forbidden (haram) (F-1). (12) Who are implicated: in socio–cultural issues (I-3). |

5. Discussion

All the observable facts about the normalized frequencies *nf* (relative frequencies *rf* × 100) of the co-occurrences of the central word from the data parts of the 2021 and 2022 samples and about the dispersion measures gave us grounds to look through the citations/examples of the co-occurrences. The next step was to capture the broad contexts of the

collocations in their given citations to categorize the discursive practices. In Tables 12–14, we numbered the collocates associated with the central word and provided the broad context constructs and the contextualized co-occurrences. The quantitative method relating to the measuring of the normalized frequency base of the co-occurrences was used to express the frequencies of the co-occurrences relative to all the study data. This method is powerful as it is based on the average. In addition, such computational techniques utilized for text data are of high importance in discourse analyses, especially critical discourse analysis (CDA), since the media and newspapers have become digitized. The technique of scraping textual data and organizing it in terms of metadata is feasible using different techniques and the library *request* in Python. Moreover, the *nf* and dispersion measures that tackle the unbalanced linguistic data parts provided us with the confidence to analyze and report the progress of women's empowerment in Saudi newspapers after Saudi Vision 2020.

Recall that the study question was the following: to what extent do co-occurrences (collocate + central word) and their broad contexts indicate the discursive and social practices of the women-related themes that were introduced, promoted or modified over the two years and the multiple sections? The findings given in Tables 12–14 included all the examples and their broad contexts with reference to their sections. The number of cases in which the progress of women's empowerment was supported was 74. On the other hand, progress was impeded in only 17 cases. Therefore, opponents were far fewer in number than supporters. Opponents were found in the *international*, *culture*, *society* and *religion* (Table 12), *front page* and *religion* (Table 13) and *economy* sections in the year 2021, and the *international*, *society*, *religion* and *front page* sections in both years (Table 14).

In light of this, the quantitative findings could be taken as empirical evidence of the reciprocal and constructionist link between discourse and context that was established earlier in this paper and is often highlighted in the relevant literature. In particular, what can be inferred from the above discussion is that the reformative measures taken by the Saudi government in accordance with its 2020 and, prospectively, 2030 visions are starting to materialize within discursive practices at various levels. This has not only taken place as these changes have been gradually implemented but has also manifested at a linguistic level through the national media. Women's empowerment emerges within this context as a dominant theme that extends across various fields, and opposition to this theme, which used to be quite strong in previous decades, is starting to fade.

In that sense, it is possible to see the link between our results and what Fairclough (1992, p. 201) refers to as the 'democratization' and 'technologisation' of discourse. Democratization of discourse can be defined as 'the removal of inequalities and asymmetries in the discursive and linguistic rights, obligations and prestige of groups of people', while technologization refers to the opposite in which there is a deliberate or subconscious intervention in discourse to maintain status quo by ensuring that a given discursive hegemony is discursively introduced and distributed. As many of the examined collocations established a shift toward more positive and women-empowering representations—be it through the positive polarity of the noun collocates or the grammatically active verb collocates—some of the linguistic asymmetries disfavoring women are also starting to decline.

Taking into consideration the aforementioned discussion of positive discourse analysis and the socially constituting inherent feature of discourse, one might be able to highlight the potential of such linguistic transformation in pushing forward more gender equality. Since it has been established in the analysis that such reformative measures were correspondingly translated in linguistic circles, it is only fair to predict further social change to be initiated by discourse on its linguistic end. In particular, this should be seen as operating in a circular fashion rather than a linear one. As change is repeatedly introduced and distributed discursively vis-à-vis the status quo within a particular context, it begins to transform gradually into the status quo per se (Al Maghlouth 2017). This usually takes place through several means, one of which is the normalization of such change linguistically and, in consequence, cognitively and socially.

This could be also linked to the role of governing and policymaking as primarily societal factors within a given discourse, as proposed by Bracher (1993, p. 53). By the same token and drawing on Gramsci's structure of power, Fairclough (2013) highlighted the significant role of political power in the domination of certain ideologies, which would gradually inform and be informed by discourse Fairclough (2013) highlighted the significant role of political power in the domination of certain ideologies, which would gradually inform and be informed by discourse. This is consistent with what was highlighted in another local study (Al Maghlouth 2017), in which social change was advocated for and pushed forward by decision makers and the associated policy in the country. Interestingly, in her study, policies in support of women's empowerment faced opposition more than a decade ago; nevertheless, they persisted and paved the way for far more reformative measures to materialize. However, this documented far less opposition, which is a finding that highlights the role of awareness in promoting, accepting and maintaining the desired change alongside governing and policymaking.

Moreover, our study data add to the rich research on the *us–them* representation spectrum, which is a recurrent theme in social psychology. In particular, a detailed examination of the relevant literature clearly documented a distinction between how Saudi women are constructed by the Saudi media locally and how they are portrayed on a more international level by the foreign media. For instance, othering and negative representations in the Western media often seem to be reinforced (Al-Hejin 2015; Elyas and Aljabri 2020; Karimullah 2020; Ruby 2013; Saleh 2016) and sincere changes or reforms in support of women are often overlooked. However, media sources produced, distributed and examined locally appear to be more consistent at reporting such changes and even portraying a more progressive construction of Saudi women. This should be approached from a perspective that validates such findings while acknowledging that this might not always be the case. To elaborate on this, Ndambuki and Janks (2010), for instance, reported a linguistic clash between how Kenyan women are portrayed discursively as lacking in agency by Kenyan political leaders, while in reality, these women were quite agentive despite being surrounded by dominating discourses of patriarchy and rurality in Kenya. What this signifies is that the constructionist link between discourse and context in this Kenyan study was not analyzed based on broad discursive patterns of varying representative data, and thus, such a link should not be taken for granted across different discourses and contexts.

6. Conclusions

In brief, this study analyzed women's social change in Saudi newspapers after the implementation of Saudi Vision 2020 as an initial stage of the national transformation program using computational techniques. One of the techniques used to assess the data from the years 2021 to 2022 was based on the adjusted frequencies of the central word (woman, pl. women) across seven data parts from the datasets from each year. This is a method that is used when the data parts are unbalanced and the raw frequencies of the lexical unit are highly skewed.

The discursive social practices of the broad contexts of the co-occurrences concerning the central word were extracted from the study data. The data was amassed from archived Saudi newspaper articles. As such, we utilized quantitative data obtained from statistical tools from our self-built corpus from Saudi newspapers published in 2021 and 2022. In doing so, the rationale and procedures for data collection and analysis were articulated in detail. The findings clearly indicated a more progressive tone in terms of women's empowerment in the Kingdom, which was consistent with the transformative measures that have been implemented by the Saudi government over the last decade. However, despite the considerable ascending change in favor of empowering women detected in the examples of discursive practices in the 2022 data, some examples conveying negativity were found in the *religion* and *society* sections.

As such, the study offers another insight into the constructionist perspective of discourse, highlighting various linguistic manifestations of social change that are embedded

within discursive practices. The study also demonstrated the need to investigate representations of Saudi women from within the Saudi context rather than from representations stemming from foreign, mostly Western, media. In addition, a number of theoretical and methodological implications can be drawn from the approach taken in this study, especially if one considers the very scarce literature on Arabic corpora in international journals.

References

Al Maghlouth, Shrouq. 2017. A Critical Discourse Analysis of Social Change in Women-Related Posts on Saudi English-Language Blogs Posted between 2009 and 2012. Ph.D. thesis, Lancaster University, Lancaster, UK.

Al Maghlouth, Shrouq. 2021. Metaphorical Analysis of Discourse on Early Saudi Attempts to Include Women in Unconventional Work Environments. *GATR Global Journal of Business and Social Science Review* 9: 1–9. [CrossRef]

Al-Hejin, Bandar. 2015. Covering Muslim women: Semantic macrostructures in BBC News. *Discourse & Communication* 9: 19–46. [CrossRef]

Al-Munajjed, Muna. 2010. Women's employment in Saudi Arabia. A Major Challenge. In *Middle East and North Africa Business*. Report. Riyadh: Booz & Company's Ideation Center.

Alemdaroglu, Ayça. 2015. Escaping femininity, claiming respectability: Culture, class and young women in Turkey. *Women's Studies International Forum* 53: 53–62. [CrossRef]

Alkhammash, Reem. 2022. Multimodal metaphors and sexism in Arabic cartoons depicting gender and gender relations during COVID-19. *Multimodal Communication* 11: 235–46. [CrossRef]

Alkhammash, Reem, and Haifa Al-Nofaie. 2020. Do Saudi academic women use more feminised speech to describe their professional titles? An evidence from corpus. *Training, Language and Culture* 4: 9–20. [CrossRef]

Almaghlouth, Shrouq. 2022. Mourning the lost: A social actor analysis of gender representation in the@ FacesofCovid's tweets. *Frontiers in Psychology* 13: 7614. [CrossRef]

Almujaiwel, Sultan. 2017. Discursive patterns of anti-feminism and pro-feminism in Arabic newspapers of the KACST corpus. *Discourse & Communication* 11: 441–66. [CrossRef]

Bakar, Kesumawati. 2014. Attitude and identity categorizations: A corpus-based study of gender representation. *Procedia, Social and Behavioral Sciences* 112: 747–56. [CrossRef]

Baker, Paul. 2006. *Using Corpora in Discourse Analysis*. London: A&C Black.

Baker, Paul. 2014. *Using Corpora to Analyze Gender*. London: Bloomsbury Publishing.

Baker, Carolyn D., and Peter Freebody. 1989. *Children's First School Books: Introduction to the Culture of Literacy*. Cambridge: Cambridge University Press.

Bashatah, Nahid. 2017. Framing Analysis of British Newspaper Representation of Saudi Women from 2005–2013. Ph.D. thesis, University of Salford, Manchester, UK.

Biber, Doug, Randi Reppen, and Eric Friginal. 2012. Research in Corpus Linguistics. *In The Oxford Handbook of Applied Linguistics, 2nd ed. Edited by Robert Kaplan. Oxford: Oxford University Press, pp. 548–70*. [CrossRef]

Bracher, Mark. 1993. *Lacan, Discourse and Social Change: A Psychoanalytic Cultural Criticism*. Ithaca: Cornell University. [CrossRef]

Brezina, Vaclav. 2018. *Statistics in Corpus Linguistics*. Cambridge: Cambridge University Press. [CrossRef]

Brooke, Mark. 2020. "Feminist" in the sociology of sport: An analysis using legitimation code theory and corpus linguistics. *Ampersand* 7: 1–8. [CrossRef]

Brown, Jane Delano, Carl Bybee, Stanley Wearden, and Dulcie Murdock Straughan. 1987. Invisible power: Newspaper news sources and the limits of diversity. *Journalism & Mass Communication Quarterly* 64: 45–54. [CrossRef]

Brun-Mercer, Nicole. 2021. Women and men in the United Nations: A corpus analysis of general debate addresses. *Discourse & Society* 32: 443–62. [CrossRef]

Butler, Judith. 1999. *Gender Trouble: Feminism and the Subversion of Identity*. New York: Routledge.

Caldas-Coulthard, Carmen Rosa, and Rosamund Moon. 2010. "Curvy, hunky, kinky": Using corpora as tools for critical analysis. *Discourse & Society* 21: 99–133. [CrossRef]

Coimbra-Gomes, Elvis, and Heiko Motschenbacher. 2019. Language, normativity, and sexual orientation obsessive-compulsive disorder (SO-OCD): A corpus-assisted discourse analysis. *Language in Society* 48: 565–84. [CrossRef]

Eberhardt, Maeve. 2017. Gendered representations through speech: The case of the Harry Potter series. *Language and Literature* 26: 227–46. [CrossRef]

Elliott, Carolyn. 2008. Introduction: Markets, communities and empowerment. In *Global Empowerment of Women: Responses to Globalization and Politicized Religions.* Edited by Carolyn Elliot. New York: Routledge.

Elyas, Tariq, Kholoud Ali Al-Zhrani, Abrar Mujaddadi, and Alaa Almohammadi. 2021. The representation(s) of Saudi women pre-driving era in local newspapers and magazines: A critical discourse analysis. *British Journal of Middle Eastern Studies* 48: 1033–52. [CrossRef]

Elyas, Tariq, and Abdulrahman Aljabri. 2020. Representations of Saudi male's guardianship system and women's freedom to travel in Western newspapers: A critical discourse analysis. *Contemporary Review of the Middle East (Online)* 7: 339–57. [CrossRef]

Fairclough, Norman. 1992. *Discourse and Social Change.* Cambridge: Polity Press.

Fairclough, Norman. 2013. *Critical Discourse Analysis: The Critical Study of Language*, 2nd ed. London: Pearson.

Fairclough, Norman, and Ruth Wodak. 1997. Critical discourse analysis. In *Discourse as Social Interaction. Discourse Studies: A Multidisciplinary Introduction*, 2nd ed. Edited by Teun A. Van Dijk. London: Sage, pp. 258–84.

Grunenfelder, Julia. 2013. Discourses of gender identities and gender roles in Pakistan: Women and non-domestic work in political representations. *Women's Studies International Forum* 40: 68–77. [CrossRef]

HRDF. 2023. Human Resources Development Fund. Available online: www.hrdf.org.sa (accessed on 1 December 2022).

Kabeer, Naila. 1999. Resources, agency, achievements: Reflections on the measurement of women's empowerment. *Development and Change* 30: 435–64. [CrossRef]

Kabeer, Naila. 2005. Gender equality and women's empowerment: A critical analysis of the third millennium development goal. *Gender and Development* 13: 13–24. [CrossRef]

Kahf, Mohja. 1999. *Western Representations of the Muslim Woman from Termagant to Odalisque*, 1st ed. Austin: University of Texas Press.

Karimullah, Kamran. 2020. Sketching women: A corpus-based approach to representations of women's agency in political Internet corpora in Arabic and English. *Corpora* 15: 21–53. [CrossRef]

Khumalo, Kathryn, Kimber Haddix McKay, and Wayne Freimund. 2015. Who is a "real woman"? Empowerment and the discourse of respectability in Namibia's Zambezi region. *Women's Studies International Forum* 48: 47–56. [CrossRef]

Kjellmer, Göran. 1986. 'The lesser man': Observations on the role of women in modern English writings. In *Corpus Linguistics II*. Edited by Jan Aarts and Willem Meijs. Leiden: Brill, pp. 163–76. [CrossRef]

Kleemans, Mariska, Gabi Schaap, and Liesbeth Hermans. 2017. Citizen sources in the news: Above and beyond the vox pop? *Journalism* 18: 464–81. [CrossRef]

Lee, Jackie F. K. 2018. Gender representation in Japanese EFL textbooks: A corpus study. *Gender and Education* 30: 379–95. [CrossRef]

Martin, James. 2004. Positive discourse analysis: Solidarity and change. *Revista Canaria de Estudios Ingleses* 49: 179–200.

Mayoux, Linda. 1998. Participatory learning for women's empowerment in micro-finance programmes: Negotiating complexity, conflict and change. *IDS Bulletin* 29: 39–50. [CrossRef]

McEnery, Tony, and Andrew Hardie. 2012. *Corpus Linguistics*. Cambridge: Cambridge University Press. [CrossRef]

Mishra, Smeeta. 2007. "Liberation" vs. "purity": Representations of Saudi women in the American press and American women in the Saudi press. *The Howard Journal of Communications* 18: 259–76. [CrossRef]

Moscovici, Serge. 2000. *Social Representations: Explorations in Social Psychology*. Oxford: Blackwell.

Ndambuki, Jacinta, and Hilary Janks. 2010. Political discourses, women's voices: Mismatches in representation. *CADAAD Journal* 4: 73–92.

Partington, Alan, Alison Duguid, and Charlotte Taylor. 2013. *Patterns and Meanings in Discourse: Theory and Practice in Corpus-assisted Discourse Studies (CADS)*. Amsterdam: John Benjamins.

Pearce, Michael. 2008. Investigating the collocational behaviour of man and woman in the BNC using Sketch Engine. *Corpora* 3: 1–29. [CrossRef]

Popa, Dorin, and Delia Gavriliu. 2015. Gender representations and digital media. *Procedia, Social and Behavioral Sciences* 180: 1199–206. [CrossRef]

Koller, Veronika. 2012. How to analyse collective identity in discourse–textual and contextual parameters. *Critical Approaches to Discourse Analysis across Disciplines* 5: 19–38.

Romaine, Suzanne. 1999. *Communicating Gender*. Mahwah: L. Erlbaum Associates.

Romano, Manuela. 2021. Creating new discourses for new feminisms: A critical socio-cognitive approach. *Language & Communication* 78: 88–99. [CrossRef]

Ross, Karen, Elizabeth Evans, Lisa Harrison, Mary Shears, and Wadia Khursheed. 2013. The gender of news and news of gender: A study of sex, politics, and press coverage of the 2010 British general election. *The International Journal of Press/Politics* 18: 3–20. [CrossRef]

Ruby, Tabassum Fahim. 2013. Muslim women and the Ontario Shari'ah tribunals: Discourses of race and imperial hegemony in the name of gender equality. *Women's Studies International Forum* 38: 32–42. [CrossRef]

Saleh, Layla. 2016. (Muslim) woman in need of empowerment: US foreign policy dicourses in the Arab spring. *International Feminist Journal of Politics* 18: 80–98. [CrossRef]

Sarfo-Kantankah, Kwabena Sarfo. 2021. The discursive construction of men and women in Ghanaian parliamentary discourse: A corpus-based study. *Ampersand* 8: 1–10. [CrossRef]

Sjøvaag, Helle, and Truls Pedersen. 2019. Female voices in the news: Structural conditions of gender representations in Norwegian newspapers. *Journalism & Mass Communication Quarterly* 96: 215–38. [CrossRef]

Sriwimon, Lanchukorn, and Pattamawan Jimarkon Zilli. 2017. Applying critical discourse analysis as a conceptual framework for investigating gender stereotypes in political media discourse. *Kasetsart Journal of Social Sciences* 38: 136–42. [CrossRef]

Saudi Vision 2030. 2022. Kingdom of Saudi Arabia: SV2030. Available online: https://www.vision2030.gov.sa/ (accessed on 14 December 2022).

Termine, Paola, and Monika Percic. 2015. Rural women's empowerment through employment from the Beijing Platform for Action Onwards. *IDS Bulletin* 46: 33–40. [CrossRef]

Van Dijk, Teun A. 2015. Critical Discourse Analysis. In *The Handbook of Discourse Analysis*, 2nd ed. Edited by Deborah Tanned, Heidi Hamilton and Deborah Schiffrin. Hoboken: John Wiley & Sons, pp. 466–85. [CrossRef]

Wharton, Sue. 2005. Invisible females, incapable males: Gender construction in a children's reading scheme. *Language and Education* 19: 238–51. [CrossRef]

Wolfsfeld, Gadi, and Tamir Sheafer. 2006. Competing actors and the construction of political news: The contest over waves in Israel. *Political Communication* 23: 333–54. [CrossRef]