

A CORPUS-BASED STUDY OF ARABIC PREPOSITIONS IN AL-JAZEERA NET NEWS

SHAFRI, M. H.¹ – SHAHARUDDIN, H. N.^{1*} – WAN OTHMAN, W. M. L.¹ – ABU BAKAR, H. R.² – MOHAMAD ISA, A. A.¹

¹ *Akademi Pengajian Bahasa, Universiti Teknologi MARA (UiTM), Selangor, Malaysia.*

² *Fakulti Sains Sosial, Universiti Islam Antarabangsa, Selangor, Malaysia.*

**Corresponding author
e-mail: hanisnajwa[at]uitm.edu.my*

(Received 09th May 2025; revised 13th August 2025; accepted 22nd August 2025)

Abstract. This study investigates the use of ḥarf al-jar in Al-Jazeera Net news articles through a corpus-based approach, focusing on their frequency, distribution and collocational patterns. The research identifies *fī*, *li*, *bi*, *min*, and *‘alā* as the five most frequently occurring prepositions and examines their usage across five news genres: general, economic, technology, culture and freedom. Employing a mixed-methods design, the study extracts data using AntConc 4.2.4 to analyze patterns of occurrence and collocation, offering insights into the syntactic and semantic roles of these prepositions in Arabic news discourse. Based on a corpus of 50 news articles containing 21,686 tokens, *fī* is the most frequently used ḥarf al-jar, appearing 699 times (3.22%), followed by *li* (686 occurrences, 3.16%), *bi* (603 occurrences, 2.78%), *min* (578 occurrences, 2.67%) and *‘alā* (364 occurrences, 1.68%). These findings underscore the critical role of ḥarf al-jar in structuring meaning and enhancing textual coherence in Arabic news discourse. From a pedagogical perspective, this study provides valuable insights for Arabic language learners and educators, particularly in understanding prepositional usage in authentic media contexts. The emphasis on corpus-based analysis further highlights the importance of data-driven approaches in Arabic language teaching and linguistic research.

Keywords: *ḥarf al-jar, prepositions, corpus linguistics, Al-Jazeera Net*

Introduction

Ḥarf al-jar refers to Arabic prepositions that generally function as connectors of meaning between the preceding word, which may be a verb or a noun, and the following noun (Al-Qahtani, 2023). Since only nouns can follow ḥarf al-jar without any restrictions, Arabic grammar scholars classify them as ḥarf al-muḥtaṣ bi al-asmā’, a category of prepositions that specifically precede nouns in Arabic syntactic structures (Aqil, 2005). In addition to their role in forming nouns in the genitive case (*majrūr*), the presence of ḥarf al-jar also determines the overall meaning of a sentence (Isnaini and Hidayat, 2023). Moreover, ḥarf al-jar plays a crucial role in establishing logical relationships between phrases and clauses within a discourse. Their presence assists readers in comprehending the connections between elements in a sentence such as cause-and-effect relationships, as well as spatial and temporal aspects expressed in syntactic structures (Badawi et al., 2015). The widespread use of ḥarf al-jar can be observed in various discourses including news media. In the context of language learning, news is considered a relevant source for enhancing students' language proficiency. This is because lessons, knowledge and various pieces of information are conveyed through multiple communication mediums, including mass communication, particularly in civilised societies (Chew and Mohamed, 2013). Additionally, news serves as an effective resource for language learning since it employs standard language

that reflects contemporary usage and linguistic changes. The precise selection of words in news articles plays a significant role in influencing readers' comprehension (Razak and Mohamed, 2024). According to Nasir et al. (2023), news remains relevant as reading material and an effective medium for language acquisition because the articles used are carefully edited by professional editors. In line with technological advancements, news continues to play a vital role in disseminating authentic information, even though much of it has transitioned to digital formats to reduce costs and expand reader accessibility (Razak and Mohamed, 2024).

The digitisation of news enables researchers to gain easier access to large-scale authentic textual data, thereby reflecting language use in real-world contexts (O'keeffe et al., 2007). This feature serves as the foundation for corpus studies, allowing for a more systematic and efficient analysis of linguistic patterns and language changes (McEnergy and Hardie, 2011). Furthermore, news digitisation facilitates the statistical computation required for corpus analysis (Stubbs, 2001). Through a corpus-based approach, large-scale data analysis can be conducted with greater accuracy and objectivity compared to traditional manual methods, which are less efficient (Biber et al., 1998). This approach also provides a more comprehensive depiction of language usage than studies relying solely on selected samples (Baker et al., 2013). Al-Jazeera Net is one of the most prominent online news portals at the international level. This news portal is easily accessible and available for free. Al-Jazeera is a globally influential news organisation with extensive operations across Arab countries. Established on November 1, 1996, in Qatar during the reign of Sheikh Emir Hamad bin Khalifa Al-Thani, the network is headquartered in Doha but maintains correspondents in various countries, enabling broader news dissemination beyond a single region (Shafri et al., 2023). On January 2, 2001, Al-Jazeera expanded its reach by launching Al-Jazeera Net, an Arabic-language online news platform. The success of Al-Jazeera television enhanced the influence of Al-Jazeera Net, making it a leading news source in the Arab world and internationally. Compared to the television channel, Al-Jazeera Net not only offers online broadcast content but also provides dedicated news sections and exclusive reports (Abdelrahim, 2007). Therefore, this study applies a corpus approach to identify the five most frequently occurring *ḥarf al-jar* in Al-Jazeera Net, followed by a collocation analysis of the two most frequent *ḥarf al-jar*. Referring to two main objectives, this study aims to provide a deeper understanding of the use of *ḥarf al-jar* in news articles, ultimately benefiting researchers and students, particularly in the field of Arabic linguistics.

Problem statement

The use of *ḥarf al-jar* in Arabic plays an essential role in determining sentence structure and intended meaning. *Ḥarf al-jar* is a significant linguistic element in shaping both written and spoken structures. These elements function as fundamental components in sentence construction and serve a key role in ensuring text cohesion (Husni and Zaher, 2020). However, students learning Arabic, especially as a second language, often struggle to accurately comprehend the use of *ḥarf al-jar* (Saleh et al., 2024; Alhawary, 2019). This difficulty arises due to structural differences between Arabic and the students' native languages, as well as a lack of exposure to the appropriate contextual usage (Dajani et al., 2014). Consequently, students tend to make errors in using *ḥarf al-jar*, such as selecting an inappropriate *ḥarf al-jar* for a given sentence context or overlooking the semantic differences between similar *ḥarf al-jar*

(Husni and Zaher, 2020). These mistakes not only affect linguistic accuracy but also reduce clarity and the effectiveness of communication (Mansour et al., 2024). Therefore, it is essential to conduct further investigation into the challenges faced by learners, with the aim of improving their proficiency in using ḥarf al-jar effectively.

While Classical Arabic is predominantly used in literature and formal writing, Modern Standard Arabic (MSA) is the primary variant for spoken communication, especially in professional, academic, and media contexts (Seyidov, 2024). Currently, MSA is used by 22 UNESCO member states, 422 million Arabs, and 1.5 billion Muslims worldwide (Thohir et al., 2020). Therefore, studying the use of ḥarf al-jar in modern texts such as news articles is highly important, as it reflects the evolution of contemporary Arabic within the context of daily communication (Holes, 2004). However, a search on Google Scholar regarding ḥarf al-jar reveals that in the past five years, most studies have focused on analyzing their usage in classical texts, such as the Quran. Studies that discuss this topic include Bahaaldin and Hat (2024), Isnaini and Hidayat (2023) as well as Asdaliah et al. (2022). This focus has resulted in a lack of research examining the use of ḥarf al-jar in Modern Standard Arabic contexts such as news media and academic writing. Therefore, further research is needed to bridge this gap and provide a more comprehensive understanding of the use of ḥarf al-jar across various discourse contexts.

Materials and Methods

This study is a syntactic analysis of Arabic based on a corpus approach, employing a mixed-methods design that integrates both qualitative and quantitative techniques to examine and descriptively explain the use of ḥarf al-jar in Arabic news (*Table 1*). By incorporating both quantitative and qualitative corpus analysis techniques, more accurate and reliable research findings can be obtained (Biber and Conrad, 2019). A total of 50 articles from the online news portal Al-Jazeera Net were selected for this study, covering five genres: general news, economy, technology, culture and freedom, spanning the period from January to December 2023. Each category is represented by 10 articles, amounting to a total of 50 articles. The length of each article is estimated to range between 300 and 500 tokens (words), with a total corpus size of 21,686 tokens. For analytical purposes, the software AntConc 4.2.4 was used to identify the five most frequently occurring ḥarf al-jar in Al-Jazeera Net news articles and to analyze the collocations of the two most frequent ḥarf al-jar. This software was chosen due to its ability to process Arabic script while providing a right-to-left text display option (Anthony, 2023). Each selected article was stored in a separate Microsoft Word file and each file was named using a simple coding system that provided an indication of the information contained within the file. This coding system included the news genre and the article number within the genre, as illustrated in *Table 2*. Data cleaning was performed by removing sections of the news articles that were not relevant to the analysis, such as images, titles, timestamps, dates and author names. The files were annotated with additional information placed at the beginning of the text, including the author's name, genre and publication date of the article. These details were enclosed within angle brackets < > to ensure they were excluded from the generated data. The corpus software settings were adjusted so that any content within angle brackets would be ignored during data processing. Since this study does not involve part-of-speech analysis, words were not tagged with linguistic annotations known as part-of-speech

(POS) tagging. Additionally, all ḥarf al-jar that were attached to other words were separated to ensure that the generated data accurately captured all instances of ḥarf al-jar.

Table 1. *Distribution of tokens by news genre.*

No.	News genre	Number of articles	Number of tokens (Words)
1	General	10	4,562
2	Economy	10	3,917
3	Technology	10	3,912
4	Culture	10	4,636
5	Freedom	10	4,659
	Total	50	21,686

Table 2. *Example of simple coding for Microsoft Word files.*

No.	Simple Coding	Genre	Article Number
1	A1	A-General	1
2	B3	B-Economy	3
3	C5	C-Technology	5
4	D7	D-Cultural	7
5	E9	E-Freedom	9

To achieve the first objective, a word list is generated based on genre. The five most frequent ḥarf al-jar in each genre are selected, and their frequencies are converted into percentages. Meanwhile, the second objective is achieved through the use of the collocation tool available in the AntConc software to generate a list of words that collocate with the two most frequent ḥarf al-jar in Arabic news based on the selected genre. These words are selected with a minimum overall frequency of five occurrences across genres, using a window span setting of one to five words to the right and left of the ḥarf al-jar. These words are then analyzed based on their position and context within sentences to determine their function and meaning.

Results and Discussion

The five most frequent Ḥarf al-Jar in each genre

Analysis of the corpus data across different genres reveals notable variations in the frequency of ḥarf al-jar usage. As presented in *Table 3*, the general news genre, comprising a total of 4,562 tokens, exhibits the five most frequently occurring ḥarf al-jar. The results indicate that *fi* is the most prevalent, accounting for 3.20% of occurrences followed by *li* (3.15%), *bi* (2.89%), *min* (2.65%) and *‘alā* (1.99%). The five most frequently occurring ḥarf al-jar in the economic news genre, which comprises 3,917 tokens. In this genre, *fi* remains the most dominant ḥarf al-jar with a frequency percentage of 3.90% followed by *li* at 3.24%. Compared to the general news genre, *min* and *bi* have switched rankings, with frequencies of 2.70% and 2.47% respectively. Meanwhile, *‘alā* retains its position as the least frequent with an occurrence rate of 1.48%. In contrast, as illustrated in *Table 3*, which presents findings for the technology genre comprising 3,912 tokens, *li* exhibits the highest frequency at 3.39%, followed by *min* (3.11%), *bi* (2.78%), and *fi* (2.45%). Meanwhile, *‘alā* remains the least frequent ḥarf al-jar in this genre, with a frequency of 2.17%. The distinct distribution of ḥarf al-jar in the technology genre compared to other genres suggests differences in syntactic or

semantic tendencies within texts related to technological discourse. For the cultural genre, which consists of 4,636 tokens, *Table 3* indicates that *li* continues to be the most frequently used *ḥarf al-jar*, with a frequency of 3.42%. It is followed by *fī* (2.95%), *bi* (2.58%), *min* (2.09%) and *‘alā* (1.29%). Finally, the findings for the freedom genre, which contains 4,659 tokens. In this genre, *fī* emerges as the most frequently used *ḥarf al-jar*, with a frequency of 3.58%. It is followed by *bi* (3.11%), *min* (2.83%), *li* (2.64%) and *‘alā* (1.50%).

Table 3. *The five most frequently occurring Ḥarf al-Jar.*

No.	Harf al-Jar	Frequency	Percentage (%)
The five most frequently occurring Ḥarf al-Jar in the general news genre			
1	Fī	146	3.20
2	Li	144	3.15
3	Bi	132	2.89
4	Min	121	2.65
5	‘Alā	91	1.99
The five most frequently occurring Ḥarf al-Jar in the economic news genre			
6	Fī	153	3.90
7	Li	127	3.24
8	Min	106	2.70
9	Bi	97	2.47
10	‘Alā	58	1.48
The five most frequently occurring Ḥarf al-Jar in the technology genre			
11	Li	133	3.39
12	Min	122	3.11
13	Bi	109	2.78
14	Fī	96	2.45
15	‘Alā	85	2.17
The five most frequently occurring Ḥarf al-Jar in the cultural genre			
16	Li	159	3.42
17	Fī	137	2.95
18	Bi	120	2.58
19	Min	97	2.09
20	‘Alā	60	1.29
The five most frequently occurring Ḥarf al-Jar in the freedom genre			
21	Fī	167	3.58
22	Bi	145	3.11
23	Min	132	2.83
24	Li	123	2.64
25	‘Alā	70	1.50

Figure 1 summarizes the five most frequently used *ḥarf al-jar* in news articles from the Al-Jazeera Net portal across the selected genres: general, economic, technological, cultural and freedom. The findings indicate that the five most commonly occurring *ḥarf al-jar* across all categories are *fī*, *li*, *bi*, *min* and *‘alā*. However, their frequency rankings vary depending on the news genre. In the general news genre, *fī* records the highest frequency at 3.20%, maintaining its dominance in the economic (3.90%) and freedom (3.58%) genres. In the cultural genre, *fī* ranks second with a frequency of 2.95%, while its frequency declines to 2.45% in the technological genre. *Li* demonstrates consistent usage across all genres, registering the highest frequency in the technological (3.39%) and cultural (3.42%) genres. In the general, economic, and freedom genres, *li* occurs at

3.15%, 3.24%, and 2.64%, respectively. The distribution of *bi* exhibits notable variation across genres. In the general genre, *bi* appears with a frequency of 2.89%, whereas its frequency decreases to 2.47% in the economic genre. In the technological and cultural genres, *bi* occurs at 2.78% and 2.58%, respectively, while its frequency increases to 3.11% in the freedom genre. *Min* maintains relatively stable frequencies across most genres, with the highest occurrences in the technological (3.11%) and freedom (2.83%) genres. In the general, economic, and cultural genres, *min* registers frequencies of 2.65%, 2.70%, and 2.09%, respectively. *‘Alā* consistently records the lowest frequency across all genres, with the highest occurrence in the technological genre (2.17%) and the lowest in the cultural genre (1.29%). In the general, economic, and freedom genres, *‘alā* appears with frequencies of 1.99%, 1.48%, and 1.50%, respectively. Based on the overall corpus data spanning all genres, comprising 21,686 tokens, *fī* emerges as the most frequently used *ḥarf al-jar*, appearing 699 times, accounting for 3.22% of the total. It is followed by *li* (686 occurrences, 3.16%), *bi* (603 occurrences, 2.78%), *min* (578 occurrences, 2.67%), and *‘alā* (364 occurrences, 1.68%). *Fī* remains the most dominant *ḥarf al-jar* across most genres, except in the technological genre.

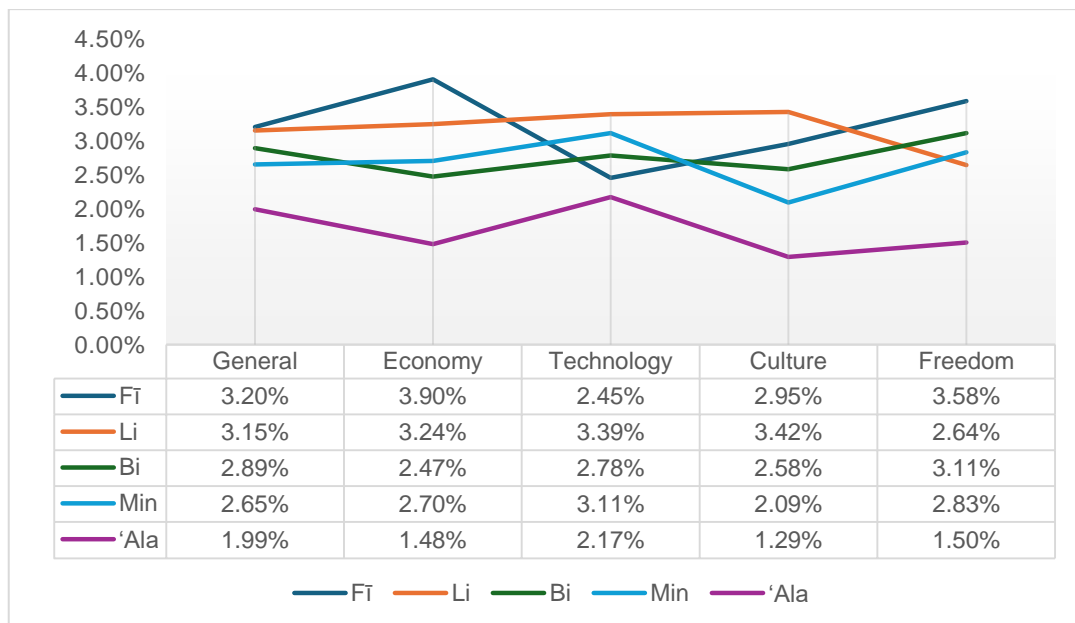


Figure 1. The five most frequently used *Ḥarf al-Jar* across all genres.

Collocational patterns of *Ḥarf al-Jar*

A corpus-based analysis across multiple genres indicates that the two most frequently occurring *ḥarf al-jar* in Al Jazeera Net news articles are *fī* and *li*. Words that co-occur with these *ḥarf al-jar* with a minimum frequency of five occurrences were analyzed based on their syntactic position and contextual usage to determine their function and meaning. This approach is essential, as the interpretation of *ḥarf al-jar* must align with the sentence context to ensure accurate and precise usage (Wahab et al., 2014).

Collocations of words appearing before *fī*

Table 4 presents four lexical items that frequently precede and collocate with *fī* across different genres. The past-tense verb *jā'a* exhibits the highest frequency of co-

occurrence with *fi*, appearing 12 times. This collocation is most prevalent in the economic genre (C) with a frequency of five occurrences, followed by the freedom genre (E) with four occurrences, the general genre (A) with two instances, and the cultural genre (D) with one occurrence. Meanwhile, the verb *tasabbaba* or *tasabbabat*, also in the past tense, collocates with *fi* only four times, each instance appearing in a different genre, general, economic, cultural and freedom, with a frequency of one occurrence per genre. Additionally, two nouns, *‘āmilīn* and *taḥqīq*, also exhibit frequent collocation with *fi*, each occurring four times. The collocation *‘āmilīn fi* is recorded twice in the freedom genre and once each in the general and economic genres. Similarly, *taḥqīq fi* appears in two genres, general and freedom, with a frequency of two occurrences per genre. Overall, these four words exhibit the highest frequency of collocation with *fi* in the freedom genre, with a total frequency of 11 occurrences. However, in the technology genre, only *jā’a* was found to collocate with *fi*, while the other lexical items did not appear in this genre.

Table 4. Collocations of words appearing.

No.	Word	Frequency by genre					Total
		A	B	C	D	E	
Collocations of words appearing before <i>fi</i>							
1	<i>jā’a/ jā’at</i>	2	-	5	1	4	12
2	<i>tasabbaba/tasabbabat</i>	1	1	-	1	1	4
3	<i>‘āmilīn</i>	1	1	-	-	2	4
4	<i>taḥqīq</i>	2	-	-	-	2	4
	Total	6	2	5	2	11	24
Collocations of words appearing after <i>fi</i>							
1	<i>hīn</i>	3	4	2	1	4	14
2	<i>hi</i>	6	2	3	2	1	14
3	<i>bayān</i>	3	3	1	1	4	13
4	<i>‘ām</i>	2	3	-	4	4	13
5	<i>‘ālam</i>	2	1	-	3	4	10
6	<i>waqt</i>	2	4	3	-	-	9
7	<i>jamī‘</i>	2	1	1	-	5	9
8	<i>majāl</i>	1	-	3	1	2	7
9	<i>miṣr</i>	4	2	-	-	1	7
10	<i>sibtambar</i>	-	2	1	2	2	7
11	<i>mantiqatī</i>	-	1	-	1	5	7
12	<i>tārīḥ</i>	2	1	-	2	1	6
13	<i>mamlakatī</i>	1	-	-	3	2	6
14	<i>ḥāl</i>	1	4	-	-	-	5
	Total	29	28	14	20	35	127
Collocations of words appearing before <i>lī</i>							
1	<i>mudīr</i>	1	2	1	1	3	8
2	<i>wafqan</i>	1	2	1	-	3	7
3	<i>nisbatī</i>	2	1	2	1	-	6
4	<i>himāyatī</i>	-	2	-	3	1	6
5	<i>ta’arruḍ</i>	-	-	-	-	5	5
	Total	4	7	4	5	12	32
Collocations of words appearing after <i>lī</i>							
1	<i>hu</i>	4	5	-	2	3	14
2	<i>ṣarikatī</i>	-	1	10	-	1	12
3	<i>hā</i>	2	1	2	3	4	12

4	awwal	2	-	1	5	-	8
5	muddat	1	2	2	1	1	7
6	asaf	5	-	-	-	-	5
7	mukāfaha	-	-	1	2	2	5
	Total	14	9	16	13	11	63

Corpus data indicate that *fī* most frequently collocates with the verb *jā'a* in news articles related to official reports, documents, and statements. This collocation frequently appears alongside expressions such as *bayān ṣaḥafī* (press statement), *risāla* (letter), *taqrīr* (report) and *taḥqīq* (investigation). While *jā'a* literally means "came" in this context, it is more accurately interpreted as "was stated", "was mentioned" or "was reported." For instance, the phrase *jā'a fī al-bayān aṣ-ṣaḥafī...* can be translated as "it was stated in the press statement..." illustrating how the verb signifies the formal transmission of information in written discourse. This semantic shift demonstrates that *ḥarf al-jar* plays a critical role in determining sentence meaning (Assiri, 2021). The collocation *tasabbaba* or *tasabbabat fī* is typically used in news reports to indicate causality, meaning "to cause", "to lead to" or "to result in." For example, the phrase *tasabbaba fī taṣādum maqṣūratayn* can be translated as "resulted in the collision of two gondolas". Meanwhile, the collocation *āmilīn fī* is used in news discourse to denote workers within a specific place, agency, company, or sector. For instance, *āl-āmilīn fī hādā āl-qīṭā'* translates to "workers in this sector". Similarly, *taḥqīq fī* found across multiple news genres, conveys the meaning of "investigation into something," particularly within legal and human rights contexts. For example, the phrase *āl-taḥqīq fī āṣ-ṣakāwā āl-muqaddama* translates to "the investigation into the submitted complaints". However, depending on the context, it may also denote "examination" or "in-depth review".

Collocations of words appearing after fī

Table 4 presents a list of 14 words that frequently appear after *fī* and collocate with it across various genres, with a total occurrence of 127 instances. These words play a crucial role in determining the function of *fī* within a sentence. Among them, the words *hīn* (when) and *hi* (his/its) exhibit the highest frequency, each occurring 14 times. The word *hīn*, which denotes a temporal reference, is predominantly found in the economic and freedom-related genres, whereas *hi*, a personal pronoun, appears more frequently in general and technological genres. Additionally, several words collocating with *fī* are related to temporal meanings, such as *ām* (year) and *waqt* (time), which occur 13 and 9 times respectively. The word *ām* appears more frequently in cultural and economic genres (four times each) and three times in the freedom-related genre, indicating its usage in annual reports or significant events. Meanwhile, *waqt*, also referring to time, is commonly found in economic texts, reflecting its role in financial and economic news contexts. Furthermore, *sibtambar* (September), another temporal collocate of *fī*, appears seven times, particularly in the freedom-related, economic and cultural genres. Its presence highlights its role in referencing specific months within news and historical discourse. Overall, the occurrence of these words demonstrates that *fī* is frequently employed to denote time or specific periods across different discourses.

The word *tārīḥ* (date/history) appears six times in total, predominantly collocating with *fī* in general and cultural genres. This suggests that the collocation *fī tārīḥ* is commonly used in discussions related to historical events or official documents. In the

context of location and geography, the words Miṣr (Egypt) and mantiqāʾ (region) each occur seven times. The word Miṣr appears most frequently in the general genre (four times), while mantiqāʾ is more dominant in the freedom-related genre (five times), illustrating their usage in geopolitical and social discussions. Additionally, the word bayān (statement), which appears 13 times in contexts related to documents and information, indicates that the collocation fī bayān is frequently used in official reports or press releases, particularly in general, economic, and freedom-related genres. Moreover, majāl (field/discipline), occurring seven times, reflects the role of fī in referring to areas of expertise or academic disciplines, particularly within the technological genre. In the context of entities and institutions, words such as mamlakaʾ (kingdom) and ʿālam (world) occur six and ten times, respectively, and are most frequently found in freedom-related and cultural genres. This suggests that fī is commonly used when referring to geographical entities or institutions in news and social analysis discourse. Regarding conditions and situations, the word ḥāl (situation) appears five times, particularly in the economic genre, signifying its use in discussions on current economic and political conditions. From a genre distribution perspective, these words collocate most frequently with fī in the freedom-related genre (35 occurrences), followed by the general genre (29 occurrences), economic genre (28 occurrences), cultural genre (20 occurrences) and technological genre (14 occurrences). In conclusion, this analysis highlights that fī in Arabic news discourse exhibits a strong tendency to collocate with words that follow it, particularly those related to time, location, official documents, areas of expertise, as well as entities and institutions. This demonstrates its crucial role in structuring news reports by providing temporal and spatial references, reinforcing credibility through official records, and framing discussions across various fields.

Collocations of words appearing before li

Table 4 provides a detailed analysis of five words that frequently appear before li and collocate with it across various genres. Among these, mudīr (director) is identified as the most frequent collocate, occurring eight times. This collocation is most prevalent in the freedom-related genre (three occurrences) followed by the economic news genre (two occurrences) and once each in the general, technological and cultural genres. Meanwhile, wafqan (according to/based on), another highly frequent collocate, appears seven times across four different genres. This collocation is most commonly found in the freedom-related genre (three occurrences), followed by the economic genre (two occurrences) and once each in the general and technological genres, but it is absent in the cultural genre. Additionally, four other nouns collocate with li, namely nisbaʾ (in terms of/with respect to), ḥimāyaʾ (protection) and taʿarruḍ or taʿarraḍ (exposure). Nisbaʾ appears six times across four genres, occurring twice each in the general and technological genres and once each in the economic news and cultural genres. Similarly, ḥimāyaʾ collocates with li six times with the highest frequency in the cultural genre (three occurrences), followed by two instances in the economic news genre and one in the freedom-related genre. Taʿarruḍ or taʿarraḍ, however, differs from the others as it exclusively appears in the freedom-related genre, with a total of five occurrences. The genre-based distribution of these li-collocated words reveals that they are most frequently found in the freedom-related genre (12 occurrences), followed by the economic genre (7 occurrences), and the cultural genre (5 occurrences), with fewer occurrences in the general and technological genres.

Corpus data indicates that *mudīr* collocates with *li* in contexts related to official positions within organizations, institutions or companies. The use of *mudīr* in this context suggests that it frequently appears in news discourse to denote an individual's role within an institution, particularly in official reports, announcements or press statements. The presence of *mudīr* across multiple genres further underscores its importance in discussions on administration, policy and institutional decisions affecting society. In this construction, *li* functions as *ḥarf al-jar* linking *mudīr* to the institution or organization it represents. Its primary function is to indicate possession or an official position within an organization, as exemplified in the phrase *āl-mudīr āl-tanfīdī li markaz iḥtijāz muqāta'aī fāyīt* (Executive Director of the Fayette County Detention Center). Additionally, *li* is employed to specify an individual's jurisdiction or scope of responsibility, as seen in *āl-mudīr al-iqlīmī li barnāmij al-aḡziyāī al-'ālamī* (Regional Director of the World Food Programme). The collocation of *wafqan* with *li* in the corpus serves as an expression that attributes information to a specific source, making it a key linguistic tool in news texts, official reports, and academic discourse. This phrase typically precedes the names of organizations, reports or individuals who provide information, thereby enhancing the credibility and legitimacy of statements in journalistic and official contexts. For instance, in *wafqan li idāraī aī-ṭayarān al-fīdrāliyyāī* (according to the Federal Aviation Administration), the collocation signifies that the statement originates from this authoritative body. Similarly, in *wafqan li al-umām al-muttaḥidaī* (according to the United Nations), the phrase highlights that the information is based on an official report or statement from the organization.

The word *nisbaī*, which collocates with *li*, functions as a phrase indicating a relationship or comparison with a particular subject or entity. This expression is frequently used in economic news articles, policy analysis and social reports to establish connections between a given aspect and the entity mentioned afterward. For example, in *al-nisbaī li al-iqtisād al-bākistānī* (with respect to the Pakistani economy), the phrase signals that the subsequent discussion pertains to Pakistan's economic context. Similarly, in *al-nisbaī li wāḍi'ī al-siyāsāt* (with respect to policymakers), the expression indicates that the information or perspective provided is relevant to policymakers. The preposition *li* in this collocation serves to denote perspective, comparison or reference to a specific subject within a discourse. The collocation of *ḥimāyāī* with *li* is commonly used in the context of protection for specific entities, such as humanitarian aid, cultural heritage, children, investments and legal frameworks. In the phrase *yajib an takful quwāt al-da'm al-sarī' al-ḥimāyāī li al-musa'adāt al-insāniyyāī* (The Rapid Support Forces must guarantee protection for humanitarian aid), the phrase illustrates *ḥimāyāī* as a guarantee of safety for a particular object or group. Similarly, in *al-ḥimāyāī al-duwaliyyāī li al-turāth al-thaqāfi* (international protection for cultural heritage), it denotes the role of legal frameworks and agreements in safeguarding cultural assets. Furthermore, this collocation appears in social contexts, such as *tawfir al-mazīd min al-ḥimāyāī li al-aṭfāl* (providing greater protection for children). Within this structure, *li* functions as a marker that connects the concept of *ḥimāyāī* (protection) to the entity being safeguarded. It plays a semantic role similar to "for" or "on behalf of" in English, clarifying the recipient of the protective action. Thus, *li* in this collocation acts as a semantic connector denoting the recipient of protection across various domains.

The phrase *ta'arruḍ* or *ta'arraḍ li* carries the meaning of "experiencing" or "being subjected to" often referring to negative experiences such as violence, abuse or harm. It is used in contexts where an individual or entity is the target of an adverse action.

Examples include *ta'arruḍuhu li al-ḍarb aš-šadīd* (experiencing severe beating) and *ta'arruḍuhu li al-taḥarruṣ al-jinsī* (experiencing sexual harassment). In these cases, *li* functions as a preposition that connects an action to the person who undergoes it. It conveys the meaning of "towards" or "against" establishing a cause-and-effect relationship where the subject becomes the target of an external force or event. In summary, *li* in Arabic frequently collocates in news discourse with words related to authority, attribution, comparison, protection, and exposure to external actions, highlighting its crucial role in structuring information, indicating sources, and establishing relationships within journalistic reporting.

Collocations of words appearing after li

Table 4 presents a list of seven words that frequently appear after *li* and collocate with it across various genres, with a total occurrence of 63 times. These words play a crucial role in determining the syntactic and semantic functions of *li* within a sentence. The corpus data indicates that the most frequent word following *li* is *hu* (his/its), a third-person pronoun that appears 14 times. This pronoun occurs most frequently in the economic genre (5 times), followed by the general genre (4 times), the freedom genre (3 times), and the cultural genre (2 times). In addition to *hu*, the words *šarikaġ* (company/corporation) and *hā* (her/its, feminine form) each occur 12 times. The word *šarikaġ* exhibits the highest frequency in the technology genre (10 times), with the remaining occurrences distributed across the economic and freedom genres (once each). Meanwhile, *hā* is most prevalent in the freedom genre (4 times) followed by the cultural genre (3 times), while the general and technology genres each record two instances and the economic genre records one. Furthermore, *awwal* (first) appears 8 times, primarily in the cultural genre (5 times), followed by the general genre (twice) and the technology genre (once). The word *muddaġ* (duration/time period) is found across multiple genres with two occurrences each in the economic and technology genres, while the general, cultural and freedom genres each contain one instance. In contrast, *asaf* (regret) is found exclusively in the general genre (5 times). Lastly, *mukāfaḥaġ* (combating) appears twice in the cultural and freedom genres and once in the technology genre.

A closer examination of these collocations reveals distinct functional patterns. The word *hu* functions as a pronoun referring to a previously mentioned entity, serving three primary roles: as a personal pronoun, a possessive marker, and an indicator of relational association. In its role as a personal pronoun, *hu* frequently appears in phrases expressing actions or experiences, such as *ta'arraḍ lahu* (experienced something). In this context, *hu* replaces an entity that undergoes a particular event, as observed in *at-ta'nīf aš-šadīd alladī ta'arraḍ lahu aṭ-ṭifl* (the severe violence experienced by the child) and *mā ta'arraḍ lahu al-sajīn fī ġuwāntānāmū* (what the prisoner in Guantanamo experienced). Beyond its role as a personal pronoun, *hu* also functions as a possessive marker, indicating a relationship between an entity and its owner. For instance, in *aṭwal silsilā ħasā'ir lahu* (the longest series of losses for him), *hu* refers to *aš-šikl* (the Shekel currency), thereby signifying that the losses pertain to this currency. A similar possessive function is evident in *a'lā mustawā lahu* (his highest level), where *hu* refers to *mu'aššir ad-dūlār* (the dollar index), indicating the highest value achieved by the index. Additionally, *hu* appears in constructions such as *al-kalimaġ lahu* (his speech), *taġrīdaġ lahu* (his tweet), and *šūraġ lahu* (his picture), all of which indicate direct ownership, often referring to prominent figures such as the Iraqi Prime Minister or an activist mentioned in the text. Moreover, *hu* serves as an indicator of relational

association, signifying a connection between an entity and a broader concept or phenomenon. In the phrase *intihāk lā dā'iyā lahu* (an unnecessary violation), *hu* refers to *intihāk* (violation), clarifying that the violation lacks a clear justification. The presence of *li* in collocation with *hu* is thus integral to structuring meaning within these contexts.

The collocation of *li* with *šarikaġ* in news discourse primarily serves to denote affiliations between individuals and corporate entities. *Li* functions as a possessive marker, linking individuals to the companies they represent, as illustrated in *ar-ra'īs at-tanfīdī li šarikaġ OpenAI* (Chief Executive Officer of OpenAI) and *aš-šarġk al-mu'assis li šarikaġ Apple* (Co-founder of Apple). This usage underscores the individual's role within the organization. Similarly, the collocation of *li* with *hā* highlights possession and attribution in Arabic syntax. In this structure, *li* acts as a preposition that links an object or action to an entity, while *hā* serves as a third-person feminine pronoun referring to a specific entity. For example, in *u'lina fī as-su'ūdiyyaġ 'an ta'sis awwal jam'iyyaġ miħniyyaġ li al-mūsīqā wa tattahīd min al-'āsimaġ ar-riyāq maqarran lahā* (The establishment of the first professional music association, headquartered in Riyadh, was announced in Saudi Arabia), the phrase *lahā* signifies that the headquarters belong to the mentioned association. The collocation of *li* with *awwal* functions as a temporal marker indicating the first occurrence of an event. *Li* introduces *awwal* as an indicator of precedence within a given context. For instance, in *muniħat jā'izaġ būkar li awwal marraġ 'am 1969* (The Booker Prize was awarded for the first time in 1969), *li awwal marraġ* emphasizes the novelty of the event. Additionally, the collocation of *li* with *muddaġ* specifies the duration of an event or condition. Here, *li* functions as a preposition conveying the meaning of "for" or "over a period of" making it a common feature in news reports concerning legal, economic and social contexts. It is frequently used to denote periods of imprisonment, suspension or the effectiveness of particular measures. An example of this usage appears in *qad yuwājih as-siħn li muddaġ 'ām idā tamma iqrār at-tuhmaġ* (He may face imprisonment for a year if the charge is confirmed).

Furthermore, the phrase *li al-asaf* serves as a discourse marker indicating regret or disappointment. This phrase is commonly employed to express negative sentiments regarding an undesirable situation. For example, in *li al-asaf, al-'aqliyyaġ at-taħrībiyyaġ muntaširaġ fī kulli makān wa zamān* (Unfortunately, the destructive mentality is widespread everywhere and at all times), *li al-asaf* underscores concern over a worsening phenomenon. In certain contexts, this phrase may carry a sarcastic tone, as exemplified in *li al-asaf malmūš ġayr 19 milyūn min al-hawā* (Unfortunately, they only managed to raise 19 million out of thin air), which criticizes an unsuccessful fundraising effort. Finally, *li mukāfaħaġ* serves as a prepositional phrase denoting purpose, often translated as "for the purpose of" or "to combat". This construction frequently appears in news reports to indicate initiatives aimed at addressing specific issues. A clear example is found in *ħilāl ħamlaġ li mukāfaħaġ al-fasād* (as part of a campaign to combat corruption), where *li mukāfaħaġ al-fasād* specifies the objective of the campaign. In overall, the collocational patterns of words that appear after *li* in Arabic news discourse reveal its diverse grammatical and semantic roles. Whether indicating possession, affiliation, duration or purpose, *li* plays an essential role in structuring meaning within Arabic texts. Its frequent co-occurrence with pronouns, nouns and descriptive terms underscores its significance in establishing relationships, specifying temporal contexts and marking evaluative perspectives, making it a critical component of Arabic linguistic structures.

Conclusion

This study provides a corpus-based analysis of ḥarf al-jar in Al-Jazeera Net news articles, identifying fī, li, bi, min and ‘alā as the five most frequently occurring prepositions. These elements play a crucial role in structuring sentences and conveying nuanced meanings across various genres. From a pedagogical perspective, this study provides valuable insights for Arabic language learners and educators. By identifying the most frequently used ḥarf al-jar and their collocational patterns, the findings offer a practical framework for understanding their usage in real-world contexts. This is particularly relevant for non-native speakers, who often struggle with the nuanced meanings and grammatical functions of ḥarf al-jar due to structural differences between Arabic and their native languages. The study’s emphasis on corpus-based analysis further underscores the importance of authentic language data in language teaching and research, enabling learners to engage with contemporary Arabic as it is used in modern media. In conclusion, this study contributes to the broader field of Arabic linguistics by providing a systematic analysis of the syntactic and semantic functions of ḥarf al-jar in news discourse. Through a corpus-based approach, it offers an objective and data-driven examination of their usage, enhancing our understanding of their role in structuring meaning and ensuring textual cohesion. Future research could expand on these findings by exploring the use of ḥarf al-jar in other discourse types, such as academic writing or social media to further investigate their linguistic and communicative functions in diverse contexts.

Acknowledgement

This research is self-funded research.

Conflict of interest

The authors confirm that there is no conflict of interest involved with any parties in this research study.

REFERENCES

- [1] Abdelrahim, Y.A.E. (2007): Arab Media and Neo-objectivity: Aljazeera Online News. – *Global Media Journal, Mediterranean Edition* 2(2): 45-58.
- [2] Alhawary, M.T. (2019): Arabic second language learning and effects of input, transfer, and typology. – Georgetown University Press 256p.
- [3] Al-Qahtani, M.M. (2023): Prepositions (Huruf Al-Jar): Addition, Deletion and Attachment To The Infinitives (Al-Afaal Al-Mujarradah) In The Arabic Sentence. – *Ijaz Arabi: Journal of Arabic Learning* 6(2): 319-335.
- [4] Anthony, L. (2023): AntConc (Version 4.2.4) [Computer Software]. – Waseda University, Japan 21p.
- [5] Aqil, I. (2005): Sharh Ibn ‘Āqil ‘alā Alfiyyah ibn Mālik. – Resalah Publishers, Lubnan 1545p.
- [6] Asdaliah, N., Mukhtar, M., Hamzah, H., Mahmud, B., Mujahid, M. (2022): Huruf Jar Ba dan Kandungan Maknanya dalam QS Al-Maidah (The Letter Jar Ba and Its Meaning in QS Al-Maidah). – *Loghat Arabi: Jurnal Bahasa Arab dan Pendidikan Bahasa Arab* 3(1): 69-86.

- [7] Assiri, A.I. (2021): Prepositional Phrases in Modern Standard Arabic: An Agree-Based Analysis. – *Journal of Language Teaching and Research* 12(6): 884-891.
- [8] Badawi, E., Carter, M.G., Gully, A., Awad, M. (2015): *Modern Written Arabic: A Comprehensive Grammar*. – Routledge 828p.
- [9] Bahaaldin, M.S., Hat, N.C. (2024): [AR] Dirasah Nahwiyyah li Ma'ani Huruf al-Jar fi Siyaq Surat al-Naml: A Grammatical Study of the Meanings of Prepositions in the Context of Surat al-Naml. – *Afaq Lughawiyah* 2(2): 197-209.
- [10] Baker, P., Gabrielatos, C., McEnery, T. (2013): *Discourse Analysis and Media Attitudes: The Representation of Islam in the British Press*. – Cambridge University Press 294p.
- [11] Biber, D., Conrad, S. (2019): *Register, genre, and style*. – Cambridge University Press 420p.
- [12] Biber, D., Conrad, S., Reppen, R. (1998): *Corpus Linguistics: Investigating Language Structure and Use*. – Cambridge University Press 300p.
- [13] Chew, F.P., Mohamed, R. (2013): Keberkesanan Penggunaan Akhbar dalam Pengajaran Karangan dalam Kalangan Pelajar Tingkatan Enam. – *Jurnal Pendidikan Bahasa Melayu-JPBM* 3(2): 18-30.
- [14] Dajani, B.A.S., Mubaideen, S., Omari, F.M.A. (2014): Difficulties of learning Arabic for non-native speakers. – *Procedia-Social and Behavioral Sciences* 114: 919-926.
- [15] Holes, C. (2004): *Modern Arabic: Structures, functions, and varieties*. – Georgetown University Press 419p.
- [16] Husni, R., Zaher, A. (2020): *Working with Arabic prepositions: Structures and functions*. – Routledge 238p.
- [17] Isnaini, R.L., Hidayat, M.R. (2023): Analysis of the Lexical Meaning of Harf Jar in Surah Al-Ahqaf. – *Studi Arab* 14(1): 10-28.
- [18] Mansour, M.S., Wan Muhammad, W.S., Abd Rauf, H., Pabiyah, H. (2024): Impact of Linguistic Problems on Learning Arabic Speaking Skills among Non-Native Arabic-Speaking Students. – *International Journal of Academic Research in Progressive Education and Development* 13(4): 3269-3285.
- [19] McEnery, T., Hardie, A. (2011): *Corpus linguistics: Method, theory and practice*. – Cambridge University Press 292p.
- [20] Nasir, S.N.A., Halim, H.A., Yahaya, A.S. (2023): Media Cetak (Akhbar) sebagai Medium Pemerolehan Bahasa Pertama dan Kedua Murid Lamban. – *Journal of Communication in Scientific Inquiry (JCSI)* 5(1): 59-73.
- [21] O'keeffe, A., McCarthy, M., Carter, R. (2007): *From corpus to classroom: Language use and language teaching*. – Cambridge University Press 315p.
- [22] Razak, Z.R.A., Mohamed, Y. (2024): Analisis Konkordan Kombinasi Kata Kerja Bahasa Arab dan Kata Sendi dalam Akhbar Mesir Al-Ahram: A Concordance Analysis on the Use of Verb and Proposition Combination in Al-Ahram Egyptian Newspaper. – *International Journal of Language Education and Applied Linguistics* 14(1): 51-59.
- [23] Saleh, Z., Murtadlo, N., Rosyidi, A.W., Hidayah, E.Z. (2024): The Analysis of Arabic Learning Error in Student of Madrasah Tsanawiyah. – *AL-ISHLAH: Jurnal Pendidikan* 16(3): 4096-4107.
- [24] Seyidov, R. (2024): Difference Between Modern Arabic And Classical Arabic. – *The Scientific Heritage* 135(135): 30-33.
- [25] Shafri, M.H., Bakar, H.R.A., Abd Raup, F.S., Yusof, M.A.M. (2023): Kekerapan Penggunaan Istilah Berkaitan COVID-19 dalam Akhbar Arab: Kajian Korpus: Frequency of COVID-19 Related Terms in Arabic Newspapers: A Corpus Study. – *International Journal of Language Education and Applied Linguistics* 13(1): 25-35.
- [26] Stubbs, M. (2001): *Words and phrases: Corpus studies of lexical semantics*. – John Wiley & Sons 288p.
- [27] Thohir, M., Kurjum, M., Muhid, A. (2020): Design and discourse of modern standard Arabic e-textbook. – *LITERA-Jurnal Penelitian Bahasa, Sastra, Dan Pengajarannya* 19(1): 1-20.

- [28] Wahab, R.A., Mokhtar, M., Shiddiq, M.A., Malla, A.S.B., Syarifuddin, S. (2014): Preposition In Arabic And Bahasa Indonesia: A Comparative Study. – Researchers World 5(4): 67-70.