

CONFLUENCE
Journal of Multidisciplinary Studies

ISSN: 3085-5411
Journal Homepage: https://journals.imist.ma/index.php/confluence/index

The *hijra* in Arabic Media Context: A Corpus-assisted Discourse Collocation Analysis

Azouz Belkharraf[a] and Zakariae Azennoud[b]
[a]Sidi Mohamed Ben Abdellah University, [b]Sultan Moulay Slimane University

https://doi.org/10.34874/PRSM.cjms-vol2iss3.9442

Azouz Belkharraf is a doctoral student at Sidi Mohammed Ben Abdellah University in Fez, affiliated with the Laboratory of *Discourse, Creativity, Society and Religions*. His research focuses on analyzing social phenomena, including migration, through critical discourse analysis.

belkharafazooz@gmail.com https://orcid.org/0009-0000-2011-8178

Zakariae Azennoud is a Lecturer in English at the Polydisciplinary Faculty of Khouribga, Sultan Moulay Slimane University. His research interests include corpus linguistics, discourse analysis, semiotics, sociolinguistics, literature, and foreign language acquisition.

zakariae.azennoud@gmail.com https://orcid.org/0000-0003-2342-8766

Received: 16-10-2025 Accepted: 25-01-2026 Published: 22-07-2026

The *hijra* in Arabic Media Context: A Corpus-assisted Discourse Collocation Analysis

Azouz Belkharraf[a] and Zakariae Azennoud[b]
[a]Sidi Mohamed Ben Abdellah University, [b]Sultan Moulay Slimane University

ABSTRACT

The Arabic word *hijra*, the equivalent of (im)migration, has played a significant part in modern media discourse. Its occurrence is determined by a variety of systematic collocational patterns that convey a range of socio-cultural implications. While the debate has been developing over recent years about the reasons and objectives of *hijra*, as a social problem, youth-centric hope, or geo-political crisis, recent empirical research remains correspondingly limited. The present study adopts a corpus-assisted discourse analysis (CADA) approach to explore the word in context, by operationalizing the collocational networks of relationships among lexical items co-occurring with *hijra*. By running three statistical association measures (AMs), namely Log Dice, MI, and Delta P in GraphColl, the study analyzes four dimensions of collocability related to the node *hijra*: frequency, exclusivity, connectivity, and dispersion, in a specially designed corpus of written media, drawn from various Arabic newspapers. The analysis identifies significant patterns of word combinations that allow reasoned judgments of some underlying socio-cultural factors influencing the *hijra* as a significant discursive unit and as a contingent, prospective act. Founded on corpus-based evidence, this exploratory work seeks to expand research interest in the increasingly influential English-dominant CADA approaches and to encourage their broader application in Arabic

KEYWORDS

Hijra

Migration

Corpus-assisted Discourse Analysis

Collocational Networks

Association Measures

Introduction

Frequency is a key element in corpus linguistics. Abstracted from it are empirical generalizations about the occurrence and co-occurrence patterns of language in natural context. Frequent and exclusive linguistic relationships can be demonstrated by examining chunks of language in use. In particular, John R. Firth identified collocations in his *Papers in Linguistics* as a way of pursuing this empirical demonstration. *Collocation* refers to the systematic co-occurrence of words in texts and discourse that can be highlighted statistically. As corpus linguistics, the empirical approach to the description of language use (Tognini-Bonelli 2), looks at language from a social perspective (Teubert 8), the study of collocations can be extended to the layers of discourse, offering deeper insights into the actual performance of its users.

In this context, the present research introduces a discursive social phenomenon, *hijra*[1] هجرة (immigration), using an analytical corpus linguistic method. Particularly, collocation analysis is adopted in order to investigate the lexico-semantic associations related to the *hijra* in the written Arabic media discourse. The collocational networks of relationships among lexical

Contact ✉ belkharafazooz@gmail.com ✉ zakariae.azennoud@gmail.com
©2026

items co-occurring with *hijra* help interpret the context of use in which the term *hijra* operates. The primary objective of the study is, therefore, to arrive at the underlying socio-cultural factors influencing the *hijra* as a primed discursive concept and act. Far beyond the empirical results, there is the potential to explore how the phenomenon is visualized from regionally specific perspectives, mainly the Maghreb and the Middle East. To achieve this objective, this research addresses three main questions: (i) The core items that co-occur with *hijra* in the written media context, (ii) how they are distributed across the texts produced by different Arabic media sources, and (iii) how the overall collocational networks can be functionally interpreted. To answer the questions above, it is necessary to examine specific pieces of evidence drawn from collocation analysis. The statistical collocation measures produce results along several relevant dimensions. In the present work, four dimensions of collocability related to the node[2] *hijra* are selected: frequency, exclusivity, connectivity, and dispersion. Some preliminary results of the study are presented below to illustrate these dimensions in turn.

Frequency of use is the basis of collocation analysis. It designates the raw and normalized occurrence of a word within the corpus, which can be presented either as specific or as aggregated results. For example, the word *fī* في occurs 1,589 times in our 40,140-token corpus, and its cases as a collocate *al-hijra fī* الهجرة في (immigration in) have a frequency of 43. Hypothetically, if we intended to compare this frequency result for the chunk *al-hijra fī* with one from another corpus which does not exceed 10,000 tokens, we should *normalize* the collocate frequency by a simple calculation (see Friginal and Hardy 302; Biber 76; McEnery and Hardie 49) as follows: (43/40,140)*10,000, which equals about 10. With this in mind, we can say that the collocate *al-hijra fī* occurs 10 times per 10,000 words. The second dimension relates to exclusivity. Let us consider the collocation *al-hijra fī* again; although this collocate occurs 43 times, the word *fī* is found to occur also after *ar-raghba* الرغبة (desire), *al-Maghrib* المغرب (Morocco), *al-filāḥa* الفلاحة (agriculture), etc., which indicates that the relationship between *hijra* and *fī* is not *exclusive*. On the other hand, the relationship between *hijra* and *ghayr* غير (not/non-) is strong, by at least one of the association measures (AMs), suggesting high exclusivity. In other words, when the word *hijra* is used in the Maghreb media (see Section 3), there is a high probability that the following word is *ghayr*. This can be linked, in turn, to the third dimension, connectivity, that is, how positionally and directionally far from or close to the node (in the collocation window) the collocate is. For instance, the word *min* من (from/in/at/etc.) has a high frequency in the corpus (1,016 as a separate item and 28 as a collocate), it is weakly connected to *hijra*, in that its position in the span is distant from the node. Consider the line below:

> في حين كان في المدينة ميزان **هجرة** سلبي في العشر سنوات **من** بين العشرين.
> *fī ḥīnin kāna fī al-madīnati mīzānu **hijratin** silbīyin fī al-ʿashri sanawātin **min** bayni al-ʿishrīn*
> (While there was a negative migration scale in the city during the ten years out of twenty)

The node *hijra* precedes the collocate *min* by five syntactic positions, which renders its semantic influence questionable. Vaclav Brezina and colleagues note that "collocates ... are part of a complex network of semantic relationships which ultimately reveals their meaning, and the semantic structure of a text or corpus should not be considered in isolation but rather as part of larger collocation networks" (141). Given this, the second-category nodes (see collocation graphs in Section 3), such as *ghayr*, are worth examining. Yet not all the collocates can be evenly distributed across the corpus. The distribution of the node and its collocates can be highlighted in terms of the fourth dimension, dispersion. In collocation analysis, dispersion is

identified by comparing the co-occurrences across the texts of the corpus. For example, the word *min* collocates with *hijra* in approximately 68 out of 85 texts, whereas *ghayr* collocates with it only in 22 texts (though its cases are highly frequent as has just been seen). Therefore, the former collocate is considered to be far more evenly distributed than the latter. In sum, dispersion results enable comparison of the broad position and interests of the producer of language regarding a particular phenomenon.

The paper follows a conventional set of procedures adopted in corpus analysis in general and corpus-assisted collocation analysis in particular. Section 2 synthesizes a concise review of the literature on the *hijra*, with major focus on the linguistic research. Section 3 presents the corpus-assisted discourse analysis (CADA) methodology in two main parts, corpus construction and corpus analysis. Section 4 displays and comments on the statistical findings of the study, in accordance with the four dimensions of collocability defined earlier. Section 5 presents a functional interpretation of these findings by establishing reasoned judgments of underlying socio-cultural factors influencing the use of *hijra*. The paper closes with conclusions and directions for future research.

The *hijra* in Literature

The word hijra in Arabic is derived from the root h-j-r ه ج ر, which means 'to leave,' 'to abandon,' or 'to migrate' (Schacht et al. 828). In most of its uses, it refers to migration or emigration, especially the act of leaving one place to settle in another, be it for social, economic, or religious purposes. Historically, the word carries profound significance within Islamic tradition and refers to the Prophet Muhammad's journey from Mecca to Medina in 622 CE, an event which marked the beginning of the Islamic calendar ("Hijrah"). In its contemporary context, hijra is often used in diversified collocations, therefore serving different meanings at the social, cultural, and political levels.

The intersection of migration and media is worth researching and is, in fact, crucial, as it contributes to a better understanding of how the media shape public discourse, attitudes, and policies surrounding migration (Eberl et al. 207). Research has revealed that media accounts depict migrants through prisms of opportunity, threat, and humanitarian need, which influences societal perceptions and political debates (McCann et al. 2). In essence, investigating this relationship from a corpus-based perspective is pivotal as it allows for the systematic analysis of language patterns, collocations, and discourse structures in large datasets of media texts (Rheindorf and Vollmer 8). This approach reveals frequent and recurring themes, biases, and framing strategies, which results in both quantitative and qualitative insights into how migration is shaped in the media (O'Regan and Riordan 749). By using corpus linguistics, researchers can critically assess the power of media in influencing migration narratives and contribute to more informed, balanced, and ethical reporting (Gemi et al. 266).

The discourse on migration within Arabic media reflects a complex blend of local and global influences. Marco Ammar and Pamela Murgia emphasize how Mediterranean migration narratives in Arabic media are influenced by European political and institutional discourses, while maintaining unique, localized characteristics (Ammar and Murgia 486). This dual nature highlights the tensions between global perceptions and region-specific realities, demonstrating how traditional categories of racism are both reinforced and reinterpreted in an Arab context. Expanding on the theme of changing narratives, Ahmad S. Haider and Saleh S. Olimy examined the shifting depiction of refugees and migrants in Jordan News Agency (PETRA). Their research complements Ammar and Murgia's findings by demonstrating how media narratives shift over time from supporting refugees to highlighting resource scarcity and Europe's

management of the crisis (Haider and Olimy 178). These studies underscore the dynamic role of media in shaping migration discourses and public perceptions, both within and beyond Arab societies.

While Ammar and Murgia focus on media representations, Hussein Elasrag redirects the discussion to the socioeconomic implications of migration, particularly the phenomenon of brain drain in the Arab world (Elasrag 7). By investigating how the migration of skilled professionals influences regional development, Elasrag offers a crucial perspective on the long-term effects of migration that go beyond immediate media narratives. This focus on the relationship between talent migration and regional growth enriches the discussion and connects the representation of migrants to broader structural and economic challenges.

Broadening the discussion of migration's impact on identity and integration, the research on media disengagement among Arab immigrants in Sweden investigates how disengagement from local media echoes profound cultural and social rifts (Al-Musawi and Al-Amiri 173). This research emphasizes the challenges immigrants encounter in maintaining a connection with their host society, while also highlighting the risk of alienation when media narratives do not align with their lived experiences (Al-Musawi and Al-Amiri 173). This viewpoint validates the broader themes addressed by Ammar and Murgia, stressing the media's pivotal role in shaping integration and identity (Ammar and Murgia 503). Finally, an ethnographic study by Alice Elliot provides a deep perspective on migration in Morocco, which challenges the traditional distinctions between migrants and non-migrants. By emphasizing the interconnectedness of those who emigrate and those who prefer to stay, this research widens the migration discourse to include its extensive social and cultural dimensions (Elliot 50). This approach complements Elasrag's macro-level analysis by highlighting the micro-level dynamics that push migration's broader effects, which offers a comprehensive understanding of its complexities. Overall, while all the studies reviewed above contributed to the investigation of the theme of migration in relation to media via different frameworks and using different methods, this research provides a quantitative, data-driven analysis of migration-related language in Arabic media using a corpus-assisted approach, which complements the largely qualitative focus of the existing studies. Unlike prior work, much of it centering on Western media, we explore linguistic patterns in Arabic discourse, which, in turn, will guide our socio-political interpretations.

Methodology

The CADA method, adopted in the present study, comprises two main phases: Corpus construction and corpus analysis. The corpus design followed a *sample corpus approach*, a construction method that tends to represent a genre of language over a specific time span, capturing a particular snapshot of slow-moving language changes (Hunston 60; McEnery and Hardie 8). The sample of the study is a specialized corpus of written Arabic media. Texts of the corpus are dated between 12 January and 12 November 2024, spanning 10 months.[3] The corpus consists of 40,140 words (Table 1), drawn from six Arabic newspapers located in two main regions, the Maghreb and the Middle East. The Maghreb includes Hespress and Al-Shorouq newspapers, while the Middle East comprises four newspapers from four countries, hence a difference in the size of each component, with the latter exceeding the former by 11,904 words.

Table 1

Sources and size of the written Arabic media corpus

Region	Newspaper	Headquarters	No of texts	No of words
The Maghreb	Hespress	Morocco	20	11,528
	Al-Shorouq	Algeria	15	5,590
The Middle East	Al-Ahram	Egypt	15	3,435
	Al-Jazeera	Qatar	10	8,983
	Akhbar Al-Khaleej	Bahrain	15	4,227
	Al-Quds Al-Arabi	United Kingdom	10	12,377
	Total number		85	40,140

Corpus investigation through collocation analysis can be carried out using various statistical measures. Given our focus on four dimensions of collocability, namely frequency, exclusivity, connectivity, and dispersion (see Section 1), it was appropriate to use three relevant AMs that have proved effective in a number of studies (cf. Alonso et al.; Brezina et al.; Gablasova et al.; Jhang and Lee; McEnery; Williams): Log Dice, MI, and Delta P.

Log Dice measures the tendency of two words to co-occur relative to the frequency of these words in the corpus (Evert 30; Smadja 155; Smadja et al. 8). It is a standardized measure operating on a scale with a fixed maximum value of 14; in the case of the present study, the maximum score reaches 11.7 in the whole dataset, and 12.5 on a regional scale. Log Dice can mark an entirely exclusive combination between words in the collocation and enable reliable contrast of dispersive collocates. Moreover, it measures the strength of connectivity, that is, how closely the collocate is associated with the node. MI (Mutual Information) is another AM that assesses the tightness and appropriateness of word combinations (González Fernández and Schmitt 96; Hunston 97; Siyanova and Schmitt 435). Theoretically, the MI score, as a normalized value that is comparable across language corpora, does not operate on a specific min-max scale. Most importantly, the more exclusive and rarer the combination between two words, the larger the score is. In our case, the MI score for all collocates ranges from 3.2 to 9.5. *Delta P*, the last metric to be presented, also quantifies the strength of association between words. Furthermore, it has been suggested as a means of assessing the dispersion and directionality of the collocate (Gries, "Phraseology" 20; "50-something Years" 141); that is, it measures whether a collocate has a tendency to precede or follow the node, hence taking into consideration the symmetries and asymmetries of node-collocate relationships. For appropriate visualization of the statistical findings, *GraphColl*, the Graphical Collocations Tool developed at Lancaster University (Brezina et al.), was employed. The tool presents graphs that illustrate the dimensions of collocability in a clear range of symbols and colors.

Results

Frequencies and Association Scores

The term *hijra* occurs 87 times in the corpus, which is the equivalent of 216 per 100,000 words ((87/40,140)*100,000). There are 682 words that co-occur with the node *hijra* in the entire corpus. As seven words were selected in both directions (R7-L7) within the collocation window, 25 out of the 682 qualify as strong collocates given the overall AM scores at hand, with function words included (*min*, *fī*, etc.). They are shown in the figure below.

مع	ميزان	أن	على	نحو	غير	إلى	من	في
ma'a	*mīzān*	*'an*	*ʿalā*	*naḥwa*	*ghayr*	*'ilā*	*min*	*fī*
(with)	(scale)	(that)	(on)	(towards)	(not, non-)	(to)	(from)	(in,..)
	إسرائيل	التي	حيث	عبر	سرية	محاولة	الشباب	أو
	'isrāʾīl	*allatī*	*ḥaythu*	*ʿabra*	*sirriya*	*muḥāwala*	*ash-shabāb*	*'aw*
	(Israel)	(that)	(where)	(across)	(clandestine)	(an attempt)	(the youth)	(or)
	عن	الخارج	سلبي	الأدمغة	لا	جماعية	عملية	الاجتماعي
	ʿan	*al-khārij*	*silbī*	*al-ʾadmigha*	*lā*	*jamāʿiya*	*'amaliya*	*al-ijtimāʿī*
	(on,..)	(abroad)	(negative)	(the brains)	(not)	(collective)	(process)	(social)

Fig. 1. Collocates of *hijra*

When the AMs are run, the majority of the 25 collocates of *hijra* prove to be meaningful items, while structural words have low values.[4] For example, *mīzān*, *silbī*, and *jamāʿiya* are at the top of the three AMs, respectively (Table 2). In Table 2, Log Dice is adopted as the main classification AM.

Table 2

The three AMs scores of the 25 collocates of *hijra*

Collocate	Log Dice	MI	Delta P
mīzān	11.7	9.0	0.1
silbī	11.6	9.0	0.1
jamāʿiya	11.2	9.0	0.08
naḥwa	11.2	6.8	0.1
muḥāwala	11.1	7.7	0.08
ash-shabāb	10.9	6.5	0.09
al-ʾadmigha	10.8	9.2	0.06
al-ijtimāʿī	10.8	7.1	0.07
sirriya	10.8	8.5	0.06
'isrāʾīl	10.6	5.9	0.09
ghayr	10.6	5.5	0.1
al-khārij	10.6	7.3	0.06
ʿabra	10.6	6.4	0.07
'amaliya	10.5	6.6	0.06
'ilā	10.3	4.7	0.3
ḥaythu	10.1	5.4	0.06

ma'a	9.8	4.5	0.09
fī	9.7	4.0	0.5
min	9.7	4.0	0.3
lā	9.3	4.1	0.05
'an	9.3	3.8	0.1
ʿan	9.2	3.9	0.05
'aw	9.1	3.8	0.05
ʿalā	8.9	3.2	0.1
allatī	8.8	3.4	0.05

Yet comparison among the collocates across the three measures can be easily performed, taking into account that the cut-off points range between 8.8 and 11.7 in Log Dice, 3.2 and 9.0 in MI, and 0.05 and 0.5 in Delta P. From a raw-frequency perspective, the minimum occurrence of collocates among the 25 co-occurring hits is five. In the table above, a few collocates are found close to the minimum threshold (i.e., five co-occurrences), including *al-ʾadmigha* and *jamāʿiya*, but their AM scores are high, given their strong values on the dimensions of exclusivity, connectivity, and dispersion. In this regard, the strength of association determines the magnitude of scores. The significance of quantified collocability can be clearly illustrated in graphs, another aspect of the findings to be displayed later.

The dispersion of the collocates is considered across two regional domains of coverage, the Maghreb and the Middle East. Notably, the collocational difference between the Maghrebi (MA.) and Middle Eastern (ME) discursive representation of the *hijra* is statistically fairly apparent. Consider the table below:

Table 3
Contrast between the Maghreb and Middle East use of *hijra*

Collocate	Freq (Corpus)		Log Dice		MI		Delta P	
	MA.	ME	MA.	ME	MA.	ME	MA.	ME
mīzān	-	11	-	12.5	-	9.4	-	0.2
silbī	-	10	-	12.4	-	9.3	-	0.2
al-ʾadmigha	-	5	-	11.6	-	9.5	-	0.1
'isrāʾīl	-	77	-	11.1	-	6.2	-	0.2
ḥaythu	-	38	-	10.9	-	6.6	-	0.1
'ilā	213	300	10.3	10.3	4.3	5.0	0.2	0.3
min	447	569	8.7	10.2	2.5	4.8	0.1	0.5
ma'a	-	121	-	9.9	-	4.9	-	0.1
fī	575	1,014	9.8	9.7	3.6	4.2	0.4	0.5
ʿalā	-	373	-	9.3	-	3.9	-	0.2
jamāʿiya	7	-	12.2	-	8.7	-	0.2	-
muḥāwala	14	-	12.0	-	7.7	-	0.2	-
naḥwa	35	-	11.9	-	6.7	-	0.2	-
sirriya	8	-	11.7	-	8.0	-	0.1	-
al-ijtimāʿī	22	-	11.6	-	6.8	-	0.1	-
'amaliya	16	-	11.5	-	7.0	-	0.1	-

ʿabra	34	-	11.4	-	6.2	-	0.1	-
ash-shabāb	36	-	11.3	-	6.1	-	0.1	-
ghayr	72	-	11.0	-	5.3	-	0.2	-
ʻan	197	-	9.7	-	3.7	-	0.1	-

Since the node is studied within two separate sub-corpora (MA. and ME), the scores of the AMs rise above the cut-off points mentioned earlier; the Log Dice maximum value reaches 12.5, 9.5 in MI, and 0.5 in Delta P, while the minimum values of some items fall below the predictable moderate association scores (hyphenated in Table 3). This indicates that some collocates, such as *mīzān* and *jamāʿiya*, are more significantly associated with the node *hijra* when they are examined from a regional angle.

Collocation Networks

GraphColl uses four symbols to show the significance of the collocate: edge length, the size of the collocate circle, its color, and positionality (see Figure 2). There are 25 collocates attached to the node (in the middle black circle). Their distance from the node varies; the closer the collocate is to the node, the stronger the association is. This edge length also illustrates the connectivity of the collocates, according to Log Dice and MI scores, in appropriate visualization. It is clear, for example, how strong the relationships between *hijra* and *mīzān*, *naḥwa*, *ash-shabāb*, *al-ʾadmigha*, *sirriya*, and *al-khārij* are. On the other hand, function words are distant from the node. These words, albeit distant, are graphed in large *circles*, which indicates their high frequency in the entire corpus as collocates of the node. The distance and connectivity between the node and the collocates not only determine syntactic roles but also have an important impact on the semantic interpretation of their contextual structures as a whole (see Section 5).

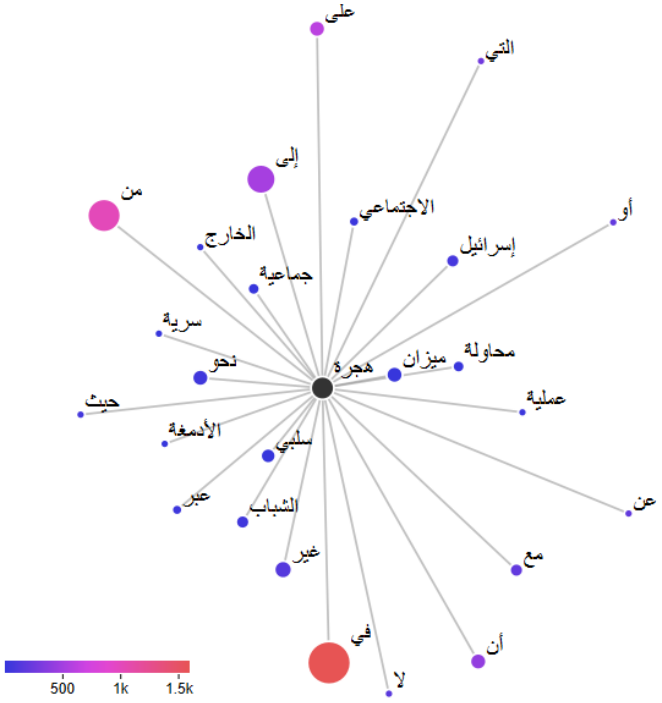


Fig. 2. Graphed 25 collocates of *hijra*

Another variable is the position of the collocate in the graph relative to the node in the center. Items on the right side of the graph are words that occur before the node, while others on the left side occur after the node. How significantly the meanings change due to such positionality is remarkable. For example, the collocate *'amaliya* mostly precedes the node *hijra*, referring to 'an organized *process'*, which is different from '*a practical* immigration' if the collocate followed the node (*hijra 'amaliya* هجرة عملية). In essence, the number of collocates in both directions is highly symmetrical (R12-L13).

Positionality can be extended to confirm our understanding of the context of a collocate. A second-category node analysis is one way of doing that. This technique was employed specifically to elaborate on the collocational network of the two Arabic domains of discourse, the Maghreb and the Middle East; in this vein, it is worth noting that the same principle of analysis adopted for retrieving basic collocates, regarding the collocation window and span, was applied to the second-category node *ghayr*. *ghayr* is a significant collocate of *hijra* in the Maghreb media, having a Log Dice score of 11.0, 5.3 in MI, and 0.2 in Delta P. Consider the elaboration of this collocate in the following graph:

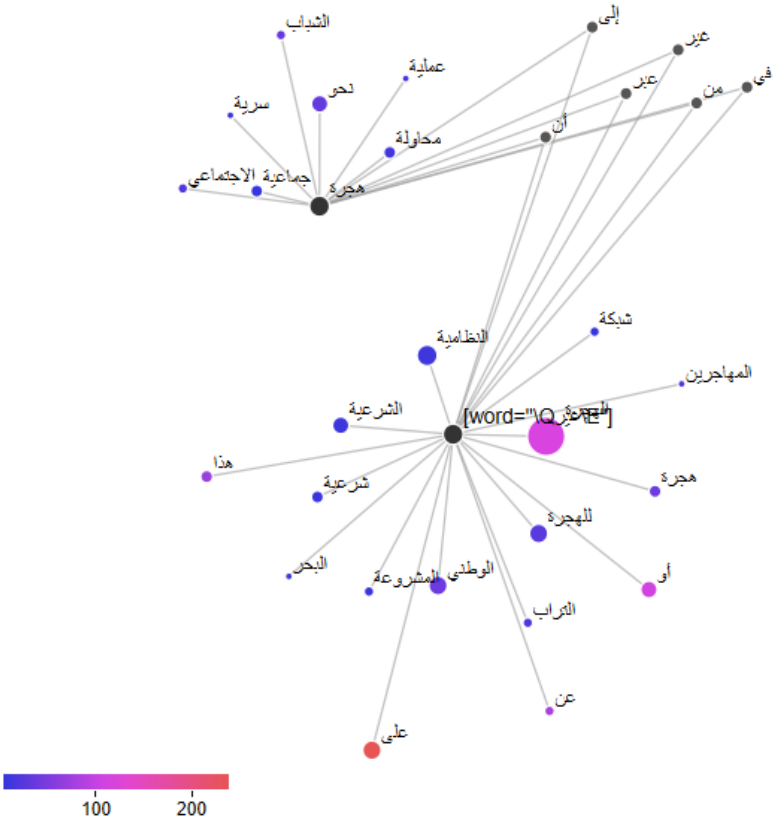


Fig. 3. *ghayr* as a second-category node of hijra (MA)

For the Middle East media, the collocate *min* has a score of 10.2 in Log Dice and 0.5 in Delta P, which are high values. Graphed by GraphColl, *min* is found to be attached to 268 second-category.

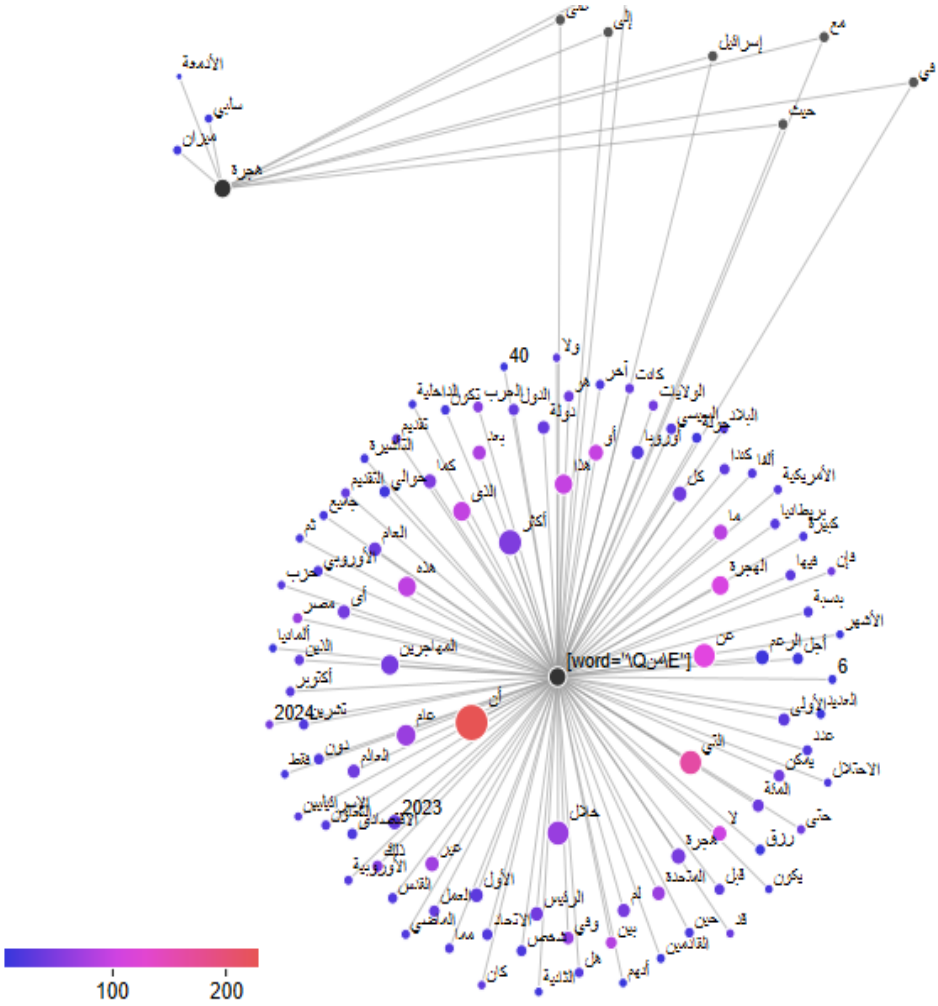


Fig. 4. *min* as a second-category node of hijra (ME)

A number of these collocates are closely related to the *hijra*. Making sense of such relationships, along with interpreting the overall data of collocation analysis, forms the focus of the next section.

Discussion

The 25 collocates found by the AMs in this study establish an interconnected context of migration in which the process, class, reasons, and destination are addressed. The process of migration, frequently associated with *muḥāwala*, implies a continual set of attempts carried out by immigrants to reach their destination *al-khārij*. *muḥāwala* also implies that migration is an intense, laborious struggle against the barriers to the outside world. In this vein, *ash-shabāb* is the major demographic group targeted in the Arab media coverage. Young people are perceived to be highly sensitive to the social (*ijtimāʿī*) driving factors. In the same way, *ash-shabāb* may constitute a large proportion of *al-ʾadmigha* heading to the north, in the Maghreb case, and to the west, in the general Middle Eastern case. It is further noticeable that the collocates referring to route or arrival points outnumber the collocates referring to departures. *naḥwa*, *ʿabra*, *ḥaythu*, *maʾa*, *ʿalā*, and *ʿilā* are six cases in point; they all reveal that the scale (*mīzān*) of migration is

negative (*silbī*), that is, the number of people who migrate from a particular country is higher than the number who migrate to it. In fact, the overall positive/negative semantic associations are regionally oriented.

Technically speaking, *hijra* in Arabic can be used to refer to irregular migration, but also to regular migration or brain drain. Significantly, the corpus data reveal a surprising dominance of the 'irregular migration' theme in the Maghreb countries. This linguistic focus stands in contrast to the socio-economic reality of brain drain reported in external literature. In Morocco, for instance, reports by De Bel-Air reveal that over 10,000 Moroccan professionals, including doctors, engineers, and researchers, emigrate annually, primarily to Europe and North America (De Bel-Air 3-4). In his article, "The Emigration of Algerian Doctors", Ahcene Zehnati asserts that an estimated 30,000 highly skilled Algerian professionals leave annually (Zehnati 5). A significant proportion are medical practitioners, with around 20% of Algeria's trained doctors working abroad as of 2024 (Zehnati 5). Egypt is not an exception to this trend. Research reveals that a vast majority of junior Egyptian physicians voice their strong desire to emigrate due to systemic job dissatisfaction (Kabbash et al. 1104). In addition to this, research has revealed that tens of thousands of Egyptian doctors are permanently employed across the Gulf countries, especially given the competitive reward packages granted by countries such as Saudi Arabia (Almansour et al. 4). Despite all this, the discourse in the media is mainly centered on irregular migration and is, in essence, predominantly descriptive, lacking a deeper evaluative or analytical layer. This descriptive emphasis is marked by a focus on reporting incidents, 'selected' statistics, and immediate consequences, such as the number of migrants apprehended or tragic accounts of failed migration attempts, without exploring the root causes or broader societal implications. For example, coverage often emphasizes irregular migration routes, detentions, or rescue operations in the Mediterranean, thereby providing a factual report rather than interrogating the socio-economic and political factors behind these movements. Such a surface-level narrative constrains public discourse, offering audiences limited perspectives and failing to encourage critical thinking, systemic reform, or innovative policy solutions.

The media discourse on migration has also been shown to be shaped by the prevailing power structures and institutional priorities. Terms like *muḥāwala*, which implies a 'difficult act,' and *sirriya*, which implies something illegal, mirror a narrative that highlights control, security, and legality, and, most importantly, resonates with governmental perspectives that prioritize national interests over humanitarian concerns. This framing portrays migration as a clandestine or risky endeavor, ignores the socio-economic factors such as poverty, unemployment, or conflict, and sustains stereotypes that stigmatize migrant communities. By intentionally focusing on sensationalized aspects like irregular routes, detentions, or rescue operations, media discourse neglects the human stories, dreams, and push factors behind migration. In fact, a shift from a security-focused discourse to one of empathy and understanding can truly change public perceptions and policy responses and foster more inclusive and equitable approaches to migration governance.

Essentially, the Maghreb media discourse on migration plays a crucial role in transforming revolutionary social energy into narratives that maintain the societal status quo. The tendency of Maghreb media to emphasize "descriptive" and "security-focused" collocates reflects a narrative that resonates with broader institutional priorities rather than an analytical investigation of the phenomenon. Rather than an isolated linguistic pattern, this echoes what Ammar and Murgia describe as the "strong presence" (Ammar and Murgia 486) of European discourse paradigms within Arabic media, which results in the "unconscious replication of Western perspectives" (Ammar and Murgia 6). This convergence with external paradigms obscures local socio-economic complexities and pushes media to embrace a "context-bound

system of representation" (Ammar and Murgia 1) that prioritizes institutional narratives over the lived experiences of migrants. In addition, Gemi et al. observe that such reporting is frequently restricted by "news making routines" and a "fixed repertoire" that prioritizes sensational or negative events, such as "illegal" crossings, given that they possess higher "news value" (Gemi et al. 268). By drawing on these standardized routines, in fact, the Maghreb media discourse inadvertently preserves the societal status quo and shapes *hijra* as a series of manageable security incidents instead of a complex social process that is stimulated by deep-grounded structural issues.

On the other hand, the Gulf media discourse on migration has been shown to be evaluative rather than descriptive, as demonstrated by the analyzed corpus. For instance, the frequent use of terms like *mīzān*, *silbī*, *al-'admigha*, and *'isrā'īl* reflects this evaluative stance, as they reveal a narrative that assesses migration in terms of its consequences on the region's socio-economic and geopolitical contexts. By way of illustration, the word *mīzān* is used to examine migration in terms of equilibrium or imbalance and echoes anxieties about preserving social stability and economic equilibrium in Gulf states heavily dependent on expatriate labor. Likewise, *silbī* carries a judgmental tone and portrays migration or its effects as negative or detrimental, especially when discussing irregular migration or cultural integration difficulties. Also, the term *al-'admigha* underscores a concern with brain drain, particularly when addressing the emigration of highly skilled professionals. The migration of this group is depicted as a setback to national progress. Meanwhile, the references to *isrā'īl* introduce a geopolitical dimension that is used to link migration issues to wider regional tensions or to criticize policies as impacted by foreign powers. This evaluative approach not only represents the region's priorities and concerns but also illustrates how media discourse in the Gulf engages actively with the implications of migration, and, therefore, shapes public perception through a critical and analytical lens rather than straightforward reportage.

Conclusion

The CADA method applied to the analysis of *hijra* in the Arabic media context arrives at several conclusions. First, the *hijra* designates a complex process, beset by a combination of hardships and oppositional connotations, positive in a particular coverage and negative in another. Second, the oppositional meanings, derived from collocation analysis, challenge the historical Islamic picture associated with *hijra* as a religiously inspired pursuit. In a wider sense, the *hijra* reflects a forced economic-specific alternative to the homeland where possibilities of well-being and happiness are limited. Third, despite apparent cultural homogeneity, the Maghreb and Middle East discourses seem to vary strongly with respect to social phenomena such as the *hijra*. This coverage disparity has been clearly defined in the discussion. The present study provides significant research value, lying in its application of an empirical approach to the study of a discursively debated concept such as *hijra*. This does not negate several limitations of the present work, notably the relatively small size of the corpus and the constrained scope of interpretation it offers. Indeed, the discussion has been developed within the framework of the statistical significance of the node under investigation relative to the overall size of the textual data. In this vein, there is a need for large-scale application of the approach in terms of an expanded dataset, more in-depth discourse analysis, and a broader diachronic corpus, which would enable the monitoring of lexico-semantic shifts within *hijra* discourse over time.

Notes

1. The transliteration system used throughout this paper is Hans Wehr, with slight modifications related to the following characters: ث *th*, ج *j*, خ *kh*, ذ *dh*, ش *sh*, ظ *ḍh*, and غ *gh*. *t marbūṭa* is chosen as an *–a*, not as *–ah*. Transliterated words are italicized and followed by a gloss.
2. The node is the word we are interested in (i.e., *hijra*). The words that co-occur with it are the *collocates* that exist in a span specified according to researchers' interests, for example five words right (R5) and five words left (L5). This span is called the *collocation window*. How far or close the word is to the node is addressed in the concept of *distance* (see Sinclair et al. 95).
3. Texts were extracted using a simple text-from-file option provided by Microsoft Word. For an effective use of the textual data, the input was refined by deleting supplementary materials such as cover pictures, icons, and hypertexts. It is also worth noting that the corpus was compiled only to answer the research questions, rather than to share it for public or commercial use. In this context, taking content from open-access domains, like newspaper websites, for corpus analysis is deemed an acceptable convention in corpus linguistics (McEnery and Hardie 59).
4. Function words are distinguished from content words based on their role in the collocation window. They are lexico-grammatical units whose semantic interpretation relies on the context in which they exist.

Works Cited

Almansour, Husam, et al. "Migration of Nurses and Doctors: Pull Factors to Work in Saudi Arabia." *Human Resources for Health*, vol. 21, no. 1, 2023. https://doi.org/10.1186/s12960-023-00809-5.

Al-Musawi, Muhammed J., and Aseel Mahdi Rasheed Al-Amiri. "Media Separation: The Relationship of Arab Immigrants with the Media of the Countries of Diaspora (Sweden as a Model)." *ALBAHITH ALALAMI*, vol. 5, no. 22, 2013, pp. 173-96. https://doi.org/10.33282/abaa.v5i22.312.

Alonso, Araceli, et al. "Collocational Networks and their Application to an E-Advanced Learner's Dictionary of Verbs in Science." *Electronic Lexicography in the 21st Century: New Applications for New Users,* edited by Iztok Kosem et al., Bled, 2011, pp. 12-22.

Ammar, Marco, and Pamela Murgia. "Shaping the Migrant: Semantic Strategies to Portray Inward and Outward Migrants as Social Actors in the Arab Press." *Discourse & Communication*, vol. 16, no. 5, 2022, pp. 485-503, https://doi.org/10.1177/17504813221099192.

Biber, Douglas. *Variation across Speech and Writing*. Cambridge University Press, 1988.

Brezina, Vaclav, et al. "Collocations in Context: A New Perspective on Collocation Networks." *International Journal of Corpus Linguistics*, vol. 20, no. 2, 2015, pp. 139-73. https://doi.org/10.1075/IJCL.20.2.01BRE.

De Bel-Air, Françoise. *Migration Profile: Morocco*. Migration Policy Center, Policy Briefs, no. 2016/05, Robert Schuman Center for Advanced Studies, European University Institute, 2016, pp. 1-16. *Cadmus EUI*, hdl.handle.net/1814/41124.

Eberl, Jakob-Moritz, et al. "The European Media Discourse on Immigration and Its Effects: A Literature Review." *Annals of the International Communication Association*, vol. 42, no. 3, 2018, pp. 207-23. https://doi.org/10.1080/23808985.2018.1497452.

Elasrag, Hussein. "Migration of Arab Talent." *Munich Personal RePEc Archive*, Sept. 2016. https://doi.org/10.2139/ssrn.2846685.

Elliot, Alice. *The Outside: Migration as Life in Morocco*. Indiana University Press, 2021.

Evert, Stefan. "Corpora and Collocations." *Corpus Linguistics: An International Handbook*, edited by Anke Lüdeling and Merja Kytö, de Gruyter, 2008, pp. 223-33. https://doi.org/10.1515/9783110213881.2.1212.

Firth, John Rupert. *Papers in Linguistics 1934–1951*. Oxford University Press, 1957.

Friginal, Eric, and Jack A. Hardy. "Conducting Multi-dimensional Analysis Using SPSS." *Multi-dimensional Analysis, 25 Years on: A Tribute to Douglas Biber*, edited by Tony B. Sardinha and Marcia V. Pinto, John Benjamins, 2014, pp. 297-321.

Gablasova, Dana, et al. "Collocations in Corpus-based Language Learning Research: Identifying, Comparing, and Interpreting the Evidence." *Language Learning*, vol. 67, no. 1, 2017, pp. 155-79. https://doi.org/10.1111/lang.12225.

Gemi, Eda, et al. "Migrants and Media Newsmaking Practices." *Journalism Practice*, vol. 7, no. 3, 2013, pp. 266-81. https://doi.org/10.1080/17512786.2012.740248.

González Fernández, Beatriz, and Norbert Schmitt. "How Much Collocation Knowledge Do L2 Learners Have? The Effects of Frequency and Amount of Exposure." *International Journal of Applied Linguistics*, vol. 166, 2015, pp. 94-126. https://doi.org/10.1075/itl.166.1.03fer.

Gries, Stefan Th. "Phraseology and Linguistic Theory: A Brief Survey." *Phraseology: An Interdisciplinary Perspective*, edited by Sylviane Granger and Fanny Meunier, John Benjamins, 2008, pp. 3-25.

Gries, Stefan Th. "50-something Years of Work on Collocations: What Is or Should Be Next." *International Journal of Corpus Linguistics*, vol. 18, 2013, pp. 137-66. https://doi.org/10.1075/ijcl.18.1.09gri.

Haider, Ahmad S., and Saleh S. Olimy. "The Representation of Laji'een (Refugees) and Muhajireen (Migrants) in the Headlines of Jordan News Agency (PETRA)." *International Journal for the Semiotics of Law - Revue Internationale de Sémiotique Juridique*, vol. 32, no. 1, 2019, pp. 155-86. https://doi.org/10.1007/s11196-018-9550-4.

Hunston, Susan. *Corpora in Applied Linguistics*. Cambridge University Press, 2002.

Jhang, Se-Eun, and Sung-Min Lee. "Clusters and Key Clusters in the Maritime English Corpus." *언어과 학*, vol. 20, no. 4, 2013, pp. 199-219. earticle.net/Article/A206411.

Kabbash, Ibrahim, et al. "The Brain Drain: Why Medical Students and Young Physicians Want to Leave Egypt." *Eastern Mediterranean Health Journal*, vol. 27, no. 12, 2021, pp. 1102-08. https://doi.org/10.26719/emhj.21.050.

McCann, Katherine, et al. "The Role of Media Narratives in Shaping Public Opinion toward Refugees: A Comparative Analysis." *Migration Research Series*, no. 72, 2023.

McEnery, Tony. *Swearing in English: Bad Language, Purity and Power from 1586 to the Present*. Routledge, 2006.

McEnery, Tony, and Andrew Hardie. *Corpus Linguistics: Method, Theory, and Practice*. Cambridge University Press, 2012.

Middle East Eye. "Arabic Press Review: Half of Egypt's Doctors Left Country in Unprecedented Brain Drain." 3 June 2022, middleeasteye.net. Accessed 20 Dec. 2025.

O'Regan, Veronica, and Elaine Riordan. "Comparing the Representation of Refugees, Asylum Seekers and Migrants in the Irish and UK Press: A Corpus-Based Critical Discourse

Analysis." *Journal of Language and Politics*, vol. 17, no. 6, 2018, pp. 744-68. https://doi.org/10.1075/jlp.17043.ore.

Rheindorf, Markus, and Bastian Vollmer. "Corpus-Based Discourse Analysis of Migration-Related Discourses in Media, Politics and Civil Society." *I-CLAIM Project*, Feb. 2025, i-claim.eu. Accessed 22 June 2026.

Schacht, Joseph, et al., editors. "Hidschra." *The Encyclopedia of Islam*, vol. 3, 1971, pp. 828-29.

Sinclair, John, et al. *English Collocation Studies: The OSTI Report*. Continuum, 2004.

Siyanova, Anna, and Norbert Schmitt. "L2 Learner Production and Processing of Collocation: A Multi-study Perspective." *Canadian Modern Language Review*, vol. 64, no. 3, 2008, pp. 429-58. https://doi.org/10.3138/cmlr.64.3.429.

Smadja, Frank. "Retrieving Collocations from Text: Xtract." *Computational Linguistics*, vol. 19, no. 1, 1993, pp. 143-77.

Smadja, Frank, et al. "Translating Collocations for Bilingual Lexicons: A Statistical Approach." *Computational Linguistics*, vol. 22, 1996, pp. 1-38. aclanthology.org/J96-1001/.

Teubert, Wolfgang. "My Version of Corpus Linguistics." *International Journal of Corpus Linguistics*, vol. 10, no. 1, 2005, pp. 1-13. https://doi.org/10.1075/ijcl.10.1.01teu.

Tognini-Bonelli, Elena. *Corpus Linguistics at Work*. John Benjamins, 2001.

Wehr, Hans. *A Dictionary of Modern Written Arabic*. Harrassowitz, 1979.

Williams, Geoffrey. "Collocational Networks: Interlocking Patterns of Lexis in a Corpus of Plant Biology Research Articles." *International Journal of Corpus Linguistics*, vol. 3, no. 1, 1998, pp. 151-71. https://doi.org/10.1075/ijcl.3.1.07wil.

Zehnati, Ahcene. "The Emigration of Algerian Doctors: A Normal Phenomenon or a Real Exodus?" *International Development Policy: Revue Internationale de Politique de Développement*, vol. 13, no. 1, 2021. https://doi.org/10.4000/poldev.4443.