

Investigating the Functions of Meta-Discourse Markers in Arabic Academic Writing: A Comparative Study of Iranian and Arab Writers

Received 2025-06-16

Accepted 2026-03-19

Published 2026-07-04

Mahdi Moqadasi Nia^{*1}, Rasool Dehghanzad²^{1,2}Department of Arabic Language and Literature, Faculty of

Literature and Humanities, University of Qom, Qom, Iran

m.moqadasi@qom.ac.ir^{*1}, r.dehghanzad@qom.ac.ir²

To cite this article: Nia, Mahdi Moqadasi., Dehghanzad, Rasool. (2026). Investigating the Functions of Meta-Discourse Markers in Arabic Academic Writing: A Comparative Study of Iranian and Arab Writers. *Ijaz Arabi: Journal of Arabic Learning*, 9 (3), 1143-1160, DOI: <https://doi.org/10.18860/ijazarabi.V9i3.41372>

Abstract

The present study aims to investigate the differences between Iranian (Persian-speaking) and Arab writers in the use of meta/discourse markers in Arabic academic writing. Emphasizing analysis at both the functional and individual marker levels, this study sought to answer the main research question of whether a significant difference exists between the two groups of writers in their use of meta/discourse markers. The research methodology was quantitative and corpus-based. The data comprised 200 academic article abstracts in the humanities and social sciences, selected through purposive sampling (100 abstracts by Iranian writers and 100 abstracts by Arab writers). meta/discourse markers were classified into 12 functions and 11 individual markers based on an integrated framework drawing on models by Fraser (2006) and Hyland (2005). Data were analyzed using Generalized Linear Models with a negative binomial distribution in SPSS software. The findings revealed that at the functional level, the pattern of discourse marker use did not differ significantly between the two groups ($p = 0.063$), with both groups employing additive, referential, explanatory, and introductory functions most frequently. However, at the level of individual markers, a significant difference was observed between the two groups ($p < 0.001$). The marker *wa* ('and') was the most frequent marker overall, accounting for 38.1% of all uses, and was used significantly more often by Arab writers. Arab writers also made significantly greater use of the markers *bal* (contrastive), *min khilāl* (referential), *qad* (hedging), and *mithl* (explanatory). In contrast, Iranian writers employed introductory verbal markers such as *yatanāwal*, *yuhāwil*, and *yas'ā*, as well as inferential markers like *azharat natā'ij...* and *tawuṣṣala*, significantly more frequently.

Keywords: Meta; Discourse Markers; Arabic Academic Writing; Contrastive Analysis; Arabic for Specific Purpose.

INTRODUCTION

Studies on writing become crucial because, when students make the transition from secondary school to a tertiary institution, they encounter many challenges—among them the difficulty of producing a coherent B.A. project. Most undergraduate students, whether they write in L1 (English as a first language) or L2 (English as a second language), still find it hard to argue, discuss or evaluate ideas competently and persuasively (Haruna et al., 2018). In the context of Arabic as a second language, Ismail and Mat (2024) documented a significant gap between students' self-perceived writing abilities and their actual proficiency, revealing notable weaknesses in complex sentence

structure, idea continuity, and paragraph organization among Malaysian learners of Arabic. Closer to the present study's context, Zare (2020) examined writing errors among Persian-speaking learners of Arabic at Yazd University, revealing that Iranian students struggle particularly with grammatical and lexical accuracy in their Arabic writing. Given that many of these observed weaknesses—particularly those related to idea continuity and paragraph organization—are directly tied to how writers establish coherence in their texts, the role of linguistic elements that create textual connectivity becomes particularly relevant.

One such set of resources is meta/discourse markers. These are linguistic elements that establish connections between different parts of a text and contribute to its coherence and cohesion (Fraser, 1999). They play a crucial role in organizing information, expressing semantic relationships, and guiding the audience in understanding the text (Halliday & Hasan, 1976). As Hyland (2005: 157) notes, such resources allow writers to "project themselves into their discourse to signal their attitude towards both the content and the audience of the text," thereby transforming a text into a site of social engagement between writer and reader. As Pearson and Abdollahzadeh (2023) highlight in their systematic review, metadiscourse has become a central concept in applied linguistics, with a growing body of research examining its role across different genres, disciplines, and languages. Following the functional orientation advocated in metadiscourse research (Hyland, 2004), this study adopts a framework that distinguishes between the textual functions of meta/discourse markers—those that organize the discourse by pointing to topic shifts, signaling sequences, and connecting ideas—and their interpersonal functions, which convey the writer's stance through hedges, boosters, self-reference, and evaluative features.

Despite considerable interest in meta/discourse markers among teachers and applied linguists, research in this area has faced persistent challenges. Hyland (2004: 156) observes that the concept has "failed to achieve its explanatory potential due to a lack of theoretical rigour and empirical confusion." A key difficulty lies in the fact that meta/discourse markers do not lend themselves to simple linguistic identification; the same item may function as a discourse marker in some contexts but not in others, and the category remains open to expansion based on the writer's communicative needs (Hyland, 2004: 158). Consequently, analysts have often struggled to operationalize the concept confidently in real texts. These challenges underscore the importance of adopting a clear functional framework when examining meta/discourse markers across different writer groups (Haruna et al., 2018). The complexity of metadiscourse analysis is further compounded by cross-cultural and cross-linguistic variations, which require researchers to carefully consider the specific rhetorical traditions of each language community (Pearson & Abdollahzadeh, 2023).

Existing research has established several foundational patterns in metadiscourse use. Basheer (2019) and Livingstone (2019) both reported that additive markers are the most frequent type of metadiscourse device in academic writing, with contrastive markers also appearing among the three most frequent types (Basheer, 2019). A higher frequency of hedging markers compared to boosters further replicates Livingstone's (2019) findings. More specifically, Hayisama et al. (2019) demonstrated that Thai and Malaysian master's students rely heavily on hedging constructions, particularly the *qad* + present tense verb pattern, to convey uncertainty.

The identification of connective explanatory markers among the most frequent devices aligns with the findings of Alenizy et al. (2024), who reported that in doctoral dissertations of Arab students, temporal, contrastive, and inferential markers follow additive markers in frequency. In a recent corpus-based study, Sanosi and Mohammed (2024) examined the use of interactional metadiscourse markers by Arab scholars in the fields of applied linguistics, translation, and literature. Their analysis revealed an imbalanced use of these markers, with Arab scholars effectively utilizing them to serve various pragmatic functions despite their limited frequency.

Numerous studies have conducted comparative analyses of metadiscourse markers across various languages and writer groups. Leńko-Szymańska (2008) documented systematic differences between native and non-native writers in their use of discoursal features. A comparable pattern emerges in Lee (2009), whose corpus of English postgraduate writing revealed that non-native students relied more heavily on textual metadiscourse (TMD), whereas native writers favoured interpersonal metadiscourse (IMD). Leńko-Szymańska (2008) also reported that foreign language learners and native speaker students employed a higher frequency of connectors than expert native writers—an effect she attributed to novice writing.

In the context of Persian and English academic writing, Mirshamsi and Allami (2013) investigated the use of metadiscourse markers in the discussion and conclusion sections of master's theses written by Iranian EFL learners. Their findings revealed significant cross-cultural differences in the deployment of both interactive and interactional metadiscourse markers. Similarly, Khedri and Basirat (2022) examined interactive metadiscourse in dentistry research articles written by Iranian and non-Iranian academic writers, finding that disciplinary conventions and cultural backgrounds shape the realization of interactive resources.

Pahlavannejad and Alinejad conducted a contrastive study of metadiscourse in 80 descriptive compositions written by Persian native speakers and Arab learners of Persian, based on Hyland's (2004) model. Their results showed that Persian learners used transitions, attitude markers, code glosses, boosters, and self-mentions more frequently than Arab learners, indicating that Persian writers employ a wider variety of markers to make their writings more cohesive and attractive. Boginskaya (2024) conducted a contrastive analysis of metadiscourse patterns in academic texts by non-native authors from Latin American, Asian, and East European backgrounds. The study revealed that East European academic prose contained considerably more hedges and attitude markers than those written by Asian and Latin American authors, confirming that metadiscourse use varies significantly across cultural and linguistic backgrounds.

In the context of Arabic academic writing, Nursa'adah, Syihabuddin, and Supriadi (2026) compared interactive metadiscourse use between native Arab writers and non-native Indonesian writers in Arabic research article abstracts, revealing that non-native writers employed metadiscourse more explicitly while native writers relied more heavily on implicit cohesion. In a broader cross-linguistic context, Assassi and Merghmi (2023) examined metadiscourse use in applied linguistics abstracts written by Algerian, Saudi, and native English-speaking researchers, finding that Algerian writers approximated the native norm in six types of markers, while Saudi writers did so in only two types. Their findings further indicated that proximity to native-like metadiscourse use does not necessarily correlate with publication success in high-indexed journals. Khedri, Chan, and Tan (2013) argued that both disciplinary conventions and language context shape the

manifestation of metadiscourse markers in academic writing. This perspective has important implications for comparative studies, suggesting that differences between writer groups may emerge at different levels of analysis (functional vs. lexical).

Despite these valuable contributions, the examination of differences between native Arab writers and Persian-speaking learners of Arabic in their use of the various functions of meta/discourse markers in Arabic academic writing has received far less attention. While Zare (2020) focused on grammatical and lexical errors in Persian-speaking learners' Arabic writing, and Ismail and Mat (2024) documented a gap between perceived and actual writing proficiency among Malaysian learners, neither study examined metadiscourse markers as an organizational and interactional resource. Furthermore, most previous studies have examined metadiscourse at either the functional level or the individual marker level, but rarely both.

The present study extends this line of inquiry by shifting the focus to Persian-speaking learners of Arabic, thereby examining whether similar or different rhetorical patterns emerge across distinct L1 backgrounds. This research seeks to address this gap in the literature by investigating the differences between Iranian and Arab writers in the use of meta/discourse markers in Arabic academic writing, with an emphasis on analysis at both the functional and individual marker levels. Accordingly, this study aims to answer the questions how do the various functions of meta/discourse markers and individual markers differ in terms of frequency of occurrence, and which are most frequent? Does the pattern of discourse marker use (at both the functional and individual marker levels) differ significantly between the two groups of Iranian and Arab writers? At which level of analysis (functional or individual markers) is the difference between the two language groups more evident? Which individual markers show the greatest differences between the two groups, and within which functions do these differences fall? By addressing these questions, the study seeks to provide a detailed comparative account of discourse marker usage in Arabic academic writing by native and non-native writers. In doing so, it aims to contribute to a better understanding of interlanguage features and academic writing development among Persian-speaking learners of Arabic, while also responding to the need for more theoretically grounded and empirically rigorous analyses of metadiscourse markers in under-researched language contexts.

METHOD

To ensure comparability, two specialized written corpora of Arabic research article abstracts were compiled following established methodological guidelines for corpus design (McEnery & Hardie, 2012). Both corpora are strictly comparable in terms of discipline (Humanities and Social Sciences), genre (abstracts), and text length.

1. Iranian Corpus (Non-native): This corpus comprises 100 Arabic abstracts written by Iranian (Persian-speaking) scholars. The texts, totaling 17,831 words, were extracted from issues 58 to 77 of the *Iranian Association of Arabic Language and Literature Scientific Research Journal*. The selected abstracts cover various subfields, including literary criticism, linguistics, discourse analysis, and Arabic language teaching.
2. Arab Corpus (Native): This corpus comprises 100 Arabic abstracts written by native Arabic-speaking authors. The texts, totaling 14,500 words, were collected from Volumes 17 and 18 (spanning four issues) of the *Lark Journal of Philosophy, Linguistics, and Social Sciences*.

A purposive sampling strategy was employed to ensure the representativeness of the academic discourse communities under investigation (Palinkas et al., 2015). Accordingly, the two corpora are purposefully matched to represent comparable levels of academic discourse, allowing for a systematic analysis of stylistic and lexical differences between native and non-native academic writing in Arabic.

Prior to analysis, several preprocessing steps were undertaken:

1. **Formatting:** All abstracts were converted to plain text (.txt) and encoded in UTF-8 to ensure accurate processing of Arabic script.
2. **Cleaning:** The texts were manually cleaned to maintain orthographic consistency. Extraneous elements such as article titles, author names, affiliations, and keywords were meticulously excluded to retain only the main body of each abstract.

In this study, marker frequency, language group, marker function, individual marker, and log-transformed text length were used as variables. The table below presents the symbol for each variable used in the statistical software package, its type, and its levels/range.

Table 1. Description of Research Variables

Variable	Levels / Range	Variable Type	Symbol
Number of Discourse Marker Occurrences	0 – 58	Dependent (Count)	COUNT
Nationality/Language of the Writer	Arab, Iranian	Independent (Factor)	CORPUS
Marker Function	12 levels	Independent (Factor)	FUNCTION
Meta/discourse markers	87 levels	Independent (Factor)	DM
Log-Transformed	3.53 – 6.03	Offset	ln words

For a systematic analysis of meta/discourse markers in the two corpora, it was first necessary to classify these markers based on their connective, attitudinal, and structural functions. This classification was achieved by integrating prominent theoretical frameworks (particularly Fraser, 2006; Hyland, 2005) and conducting an exploratory analysis of the research data.

The process of classifying meta/discourse markers in this study involved the following steps:

1. **Marker Extraction:** An initial list of common meta/discourse markers in Arabic academic texts was compiled based on previous studies and existing corpora.
2. **Exploratory Analysis:** By examining a sample of data from both corpora, frequent markers, as well as markers that had received less attention in prior research, were identified.
3. **Initial Classification:** Markers were placed into preliminary functional groups based on their discursual role within the context of the sentence and text.
4. **Validation:** To ensure the validity of the classification, 20% of the data were randomly selected and recoded by two additional independent evaluators (a linguistics specialist and an Arabic language specialist). The inter-rater agreement (Cohen's Kappa coefficient) was 0.86, indicating satisfactory reliability of the classification.

Ultimately, meta/discourse markers were classified into three main categories and twelve functions:

1. **Connective Markers:** Markers that establish semantic and logical relationships between different parts of discourse. This group includes six functions:
 - a. **Additive:** Markers that add new information to the text (e.g., *wa* 'and', *fa* 'so/and', *aydan* 'also')

- b. Cause and Effect: Markers that express causal relationships (e.g., *li-dhālika* ‘therefore’, *li-anna* ‘because’)
 - c. Contrast: Markers that indicate opposition or contradiction (e.g., *lākin* ‘but’, *bal* ‘rather/but’)
 - d. Order: Markers that specify sequence or order (e.g., *awwalan* ‘firstly’, *thāniyan* ‘secondly’)
 - e. Explanatory: Markers that provide clarification or elaboration (e.g., *haythu* ‘whereas/in that’)
 - f. Reference and Perspective: Markers that refer to a source or viewpoint (e.g., *bi-ḥasabi* ‘according to’)
2. Attitude Markers: Markers that express the writer's attitude, evaluation, or stance towards the propositional content. This group includes two functions:
 - a. Emphasis: Markers that stress the importance of a point (e.g., *ḥaqqan* ‘truly’, *bi-l-ta’kīd* ‘certainly’)
 - b. Uncertainty (Hedging): Markers that express lack of certainty (e.g., *rubbamā* ‘perhaps’, *mina l-mumkin* ‘it is possible’)
3. Structuring Markers: Markers that help organize the text and guide the reader through the discourse structure. This group includes four functions:
 - a. Introduction: Markers that introduce a new topic or the study's aim (e.g., *yuḥāwil ḥādhā l-baḥṡ* ‘this study attempts’, *yas’ā ḥādhā l-baḥṡ* ‘this study seeks’)
 - b. Inference and Conclusion: Markers that introduce a result or conclusion (e.g., *ladhā* ‘thus/therefore’)
 - c. Structural Reference: Markers that refer to other parts of the text (e.g., *min khilāl* ‘through’, *abra* ‘via’)
 - d. Topic Shift: Markers that signal a change in topic (e.g., *ammā* ‘as for’, *fī-mā yata’allaqu bi-* ‘regarding’)

Examined Markers and Considerations

Based on the above process, a total of 87 individual meta/discourse markers were classified into 12 functions. Several methodological considerations were taken into account in this classification:

1. Polysemy of Markers: Some meta/discourse markers can have different functions in different contexts. In such cases, the function was determined based on the sentence context and its role in the discourse.
2. Compound Markers: Some markers appear in composite forms (e.g., *min khilāl al-i’timād* ‘alā ‘through reliance on’). These were considered as single independent markers.
3. Functional Overlap: In cases where a marker exhibited more than one function, the dominant function observed in the research data was used as the basis for classification.
4. Coding Consistency: To ensure uniformity in coding, a detailed coding manual was developed, in which each marker was presented with examples from different contexts.

This classification enables a systematic comparison of discourse marker usage between the Persian-speaker and Arab-speaker corpora and serves as the basis for the quantitative analyses in the following section.

Statistical Model

Generalized Linear Models (GLM) were used for data analysis. Following established methodological frameworks for count data analysis (Hardin & Hilbe, 2012), these models were selected due to two characteristics of our data: first, the dependent variable is a count variable, meaning the data are in the form of frequencies; second, the data exhibited overdispersion, meaning the variance was greater than expected under a standard Poisson distribution. To address this issue, a Negative Binomial distribution with a Log Link function was employed (Hardin & Hilbe, 2012).

The model formula was defined as follows:

1. $\log(\text{COUNT}) = \beta_0 + \beta_1(\text{CORPUS}) + \beta_2(\text{FUNCTION}) + \beta_3(\text{CORPUS} \times \text{FUNCTION}) + \text{offset}(\ln_words)$. For the complementary analysis at the level of individual markers, the following model was also fitted:
2. $\log(\text{COUNT}) = \beta_0 + \beta_1(\text{CORPUS}) + \beta_2(\text{DM}) + \beta_3(\text{CORPUS} \times \text{DM}) + \text{offset}(\ln_words)$

Procedure and Software Used

After extracting the meta/discourse markers from the corpus and coding their functions, the data were entered into SPSS software (version 26). Analyses were conducted using the Generalized Linear Models procedure. The significance level for all tests was set at 0.05, and Bonferroni correction was applied to control for Type I error in multiple comparisons (Hsu & Nelson, 1998).

RESULTS AND DISCUSSION

The Functions Of Meta/Discourse Markers And Individual Markers

A total of 17,400 observations (200 abstracts $\times$ 87 markers) were analyzed. Table 2 presents the descriptive statistics for the research variables.

Table 2. Descriptive Statistics of Research Variables

Variable	N	Minimum	Maximum	Mean	Std. Deviation
COUNT	17400	0	57	0.28	2.07
TOTAL WORDS	200	34	416	174.32	60.24
ln_words	17400	3.53	6.03	5.09	0.34

As observed, the mean frequency of meta/discourse markers is 0.28 with a standard deviation of 2.07, indicating relatively high dispersion and the presence of overdispersion in the data. This justifies the use of the negative binomial distribution. The range from zero to a maximum of 57 markers in a single abstract shows a wide variation.

Table 3. Goodness of Fit of the First Model (Functional Level)

Index	Value	Interpretation
Deviance/df	0.682	Very good fit (< 1)
Pearson Chi-Square/df	2.213	Acceptable
Log Likelihood	-8420.636	-
AIC	16889.272	-

The Deviance/df ratio is 0.682, and the Pearson Chi-Square/df ratio is 2.213. These values indicate a satisfactory fit of the model to the data.

Table 4. Omnibus Test for Comparison with Baseline Model

Index	Value	df	Sig.
Likelihood Ratio Chi-Square	6318.659	23	0.000

This finding indicates that the present model performs significantly better than the baseline model (without predictors), as the significance level is 0.000. In other words, the

variables included in the model significantly explain the variation in discourse marker frequency.

Table 5. Tests of Model Effects (Type III)

Source	Wald Chi-Square	df	Sig.
CORPUS	0.099	1	0.753
FUNCTION	4621.464	11	0.000
CORPUS × FUNCTION	18.869	11	0.063

The tests of model effects show that the main effect of FUNCTION is significant ($p < 0.001$). This finding indicates that different functions of meta/discourse markers differ significantly in their frequency of occurrence. In contrast, the main effect of CORPUS is not significant ($p = 0.753$), suggesting no overall difference between the two language groups. Furthermore, the interaction effect CORPUS×FUNCTION is not significant ($p = 0.063$), as this value is higher than 0.05, indicating that the pattern of using different discourse marker functions does not differ significantly between the Iranian and Arab writers.

Parameter Estimates for Functions in Both Groups

Table 5 shows the model coefficients for each function across the two language groups. Positive coefficients indicate a higher frequency of use compared to the reference group. Given the significance of the main effect of FUNCTION, Table 6 displays the model coefficients for different discourse marker functions, using Uncertainty markers as the reference category.

Table 6. Parameter Estimates for Discourse Marker Functions in Both Groups

Function	Coefficient (β)	Std. Error	Sig.
Additive	3.993	0.414	0.000
Contrast	0.856	0.437	0.050
Explanatory	1.553	0.441	0.000
Cause and Effect	1.209	0.430	0.005
Order	1.081	0.443	0.015
Reference and Perspective (Connector)	1.811	0.431	0.000
Topic Shift	1.386	0.465	0.003
Introduction	1.574	0.432	0.000
Inference and Conclusion	1.280	0.442	0.004
Structural Reference	0.881	0.463	0.057
Emphasis	0.134	0.484	0.782
Uncertainty (Reference)	0		

As shown, all functions except “Structural Reference” ($p = 0.057$) and “Emphasis” ($p = 0.782$) differ significantly from the reference category (Uncertainty). Additive markers have the highest frequency with a coefficient of 3.993, followed by “Introduction” (1.574) and “Topic Shift” (1.386), and then “Reference and Perspective” (1.811).

Bonferroni Correction for Function Comparisons

To control for Type I error resulting from comparing 11 function pairs (12 functions with Uncertainty as the reference category), Bonferroni correction was applied. The adjusted significance level was calculated as 0.0045 ($0.05 \div 11$).

Table 7. Comparison of Functions with Bonferroni Correction

Function	p-value (Unadjusted)	p-value (Bonferroni Adjusted)	Final Result
Additive	0.000	0.000	Significant
Explanatory	0.000	0.000	Significant

Reference and Perspective (connector)	0.000	0.000	Significant
Introduction	0.000	0.000	Significant
Topic Shift	0.003	0.033	Significant
Inference and Conclusion	0.004	0.044	Significant
Cause and Effect	0.005	0.055	Non-significant
Order	0.015	0.165	Non-significant
Contrast	0.050	0.550	Non-significant
Structural Reference	0.057	0.0627	Non-significant
Emphasis	0.782	1.000	Non-significant
Uncertainty (Reference)	-	-	Reference

After applying the Bonferroni correction (adjusted alpha = 0.0045), the Additive, Explanatory, Reference and Perspective (Connective), Introduction, Topic Shift, and Inference and Conclusion functions showed a statistically significant difference from the reference group (Uncertainty). By contrast, the Cause and Effect, Order, Contrast, Structural Reference, and Emphasis functions did not show a significant difference.

Effect Size Analysis

To determine the magnitude of the observed differences, Incidence Rate Ratios (IRR = $\exp(\beta)$) were calculated. IRR represents the factor by which the expected count of meta/discourse markers changes compared to the reference category (Uncertainty markers). An IRR greater than 1 indicates a higher frequency.

Table 8. Effect Sizes for Significant Functions (IRR)

Function	β	IRR	95% CI for IRR	p (Bonferroni)
Additive	3.993	54.20	[24.07, 122.04]	0.000
Explanatory	1.553	4.73	[1.99, 11.20]	0.000
Reference and Perspective (connector)	1.811	6.12	[2.63, 14.24]	0.000
Introduction	1.574	4.83	[2.07, 11.27]	0.000
Topic Shift	1.386	4.00	[1.61, 9.95]	0.033
Inference and Conclusion	1.280	3.60	[1.51, 8.56]	0.044

The large effect size for Additive markers (IRR = 54.20) indicates that these markers are used approximately 54 times more frequently than Uncertainty markers. This considerable difference reflects the fundamental role of additive markers in establishing coherence and continuity in academic writing. The wide confidence intervals are consistent with the overdispersion observed in the data.

Analysis of Individual Markers (Model 2 – Main Model)

Given that the interaction effect CORPUS $\times$ FUNCTION was not significant at the functional level, but exploratory analyses suggested subtle differences at the individual marker level, the second model was fitted at the level of individual markers. The results of this model are presented below.

Table 9. Fit Indices for Model 2 (Individual Marker Level)

Index	Value	Interpretation
Deviance/df	0.268	Very good fit (< 1)
Pearson Chi-Square/df	0.868	Very good fit (< 1)
Log Likelihood	-4809.878	-
AIC	9967.756	-

The Deviance/df ratio is 0.268, and the Pearson Chi-Square/df ratio is 0.868, indicating a very good fit of the model to the data. A comparison of AIC values between the two models shows that the second model (9967.756) has a better fit than the first (16889.272).

Table 10. Omnibus Test for Comparison with Baseline Model

Index	Value	df	Sig.
Likelihood Ratio Chi-Square	13540.175	173	0.000

Given the significance level of 0.000, this finding indicates that the second model performs significantly better than the baseline model.

Table 11. Tests of Model Effects (Type III)

Source	Wald Chi-Square	df	Sig.
CORPUS	0.076	1	0.783
DM	4872.425	86	0.000
CORPUS × DM	127.362	62	0.000

The tests of model effects indicate that the main effect of DM (individual markers) is significant (Wald $\chi^2 = 4872.425$, df = 87, $p < 0.001$), demonstrating a significant difference among individual markers in terms of frequency of occurrence. In contrast, the main effect of CORPUS is not significant (Wald $\chi^2 = 0.076$, df = 1, $p = 0.783$). However, the interaction effect (CORPUS × DM) is significant (Wald $\chi^2 = 127.362$, df = 62, $p < 0.001$). This finding shows that the pattern of using individual meta/discourse markers differs significantly between the Iranian and Arab writers.

Most Frequent Individual Markers in the Entire Corpus

Table 12. Ten Most Frequent Markers in the Entire Corpus

Rank	Marker	Function	Total Frequency	Percentage
1	wa ('and')	Additive	1852	38.1
2	fa- ('so/and')	Additive	214	4.4
3	min khilāl ('through')	Reference	103	2.1
4	haythu ('whereas/in that')	Explanatory	87	1.8
5	mithl ('like/such as')	Explanatory	76	1.6
6	yatanāwal... (al-baḥṭh/al-dirāsah) ('the study addresses')	Introduction	65	1.3
7	bal ('rather/but')	Contrast	58	1.2
8	yuhāwil... (al-baḥṭh/al-dirāsah) ('the study attempts')	Introduction	52	1.1
9	yas'ā... (al-baḥṭh/al-dirāsah) ('the study seeks')	Introduction	48	1.0
10	wa-lākin ('but')	Contrast	45	0.9

As observed, the marker wa ('and') is by far the most frequent marker, accounting for over 38% of all uses. This finding is consistent with the nature of academic writing, which requires coherence and connectivity.

Markers with the Greatest Differences Between the Two Groups

Analysis of the CORPUS × DM interaction effects revealed that some markers differ significantly between the two groups.

Table 13. Markers Used Significantly More by Arab Writers

Marker	Function	Arab-to-Iranian Ratio	Sig.
wa ('and')	Additive	1.45	0.000
bal ('rather/but')	Contrast	1.38	0.000
min khilāl ('through')	Reference	1.32	0.000
qad ('may')	Uncertainty	1.28	0.000
mithl ('like/such as')	Explanatory	1.25	0.000

Table 14. Markers Used Significantly More by Persian-Speaking (Iranian) Writers

Marker	Function	Arab-to-Iranian Ratio	Sig.
yatanāwal... (al-baḥṡh/al-dirāsah) ('the study addresses')	Introduction	1.52	0.000
yuḥāwil... (al-baḥṡh/al-dirāsah) ('the study attempts')	Introduction	1.44	0.000
aẓharat natā'ij... ('the results showed...')	Inference and Conclusion	1.35	0.000
yas'ā... (al-baḥṡh/al-dirāsah) ('the study seeks')	Introduction	1.31	0.000
tawuṣṣala ('arrived at/found')	Inference and Conclusion	1.27	0.000

1. Similarities in the Most Frequent Functions

The findings of the study showed that in both groups, additive markers (including wa, *fa-*, ayḍan) were the most frequent function. This finding aligns with the nature of academic writing, which requires coherence, continuity, and the development of ideas. Additive markers allow the writer to add new information to the text and form a chain of arguments. This characteristic appears to be common in both Arabic and Persian writing traditions. Furthermore, the tests of model effects at the functional level (Table 5) indicated that the main effect of FUNCTION was significant ($p < 0.001$), meaning that different functions differ significantly in their frequency of occurrence.

2. No Significant Difference at the Functional Level

A key finding of this research is that the pattern of using discourse marker functions does not differ significantly between the Iranian and Arab writer groups. The tests of model effects at the functional level (Table 5) showed that the CORPUS $\times$ FUNCTION interaction effect was not significant ($p = 0.063$). This finding suggests that although the frequency of different functions varies, the distribution pattern of these functions is similar across the two language groups. In other words, Iranian and Arab writers exhibit similar linguistic behavior at the overarching functional level.

3. Significant Difference at the Level of Individual Markers

Despite the lack of difference at the functional level, the analysis at the individual marker level revealed considerable differences. The tests of model effects at the individual marker level (Table 11) demonstrated that the CORPUS $\times$ DM interaction effect was significant ($p < 0.001$). This finding indicates that the pattern of using individual meta/discourse markers differs significantly between the Iranian and Arab writers. In other words, although the two groups use similar functions, their choice of specific markers within each function varies.

4. Markers Used More Frequently by Arab Writers

The results of the interaction effect analysis showed that Arab writers used five individual markers significantly more often than Iranian writers. The additive marker wa ('and') showed the greatest difference, with a ratio of 1.45 and a significance level

of 0.000; this marker, which is also the most frequent marker in the entire corpus, has a higher frequency in the writing of Arab writers. Following this is the contrastive marker *bal* ('rather/but') with a ratio of 1.38 ($p = 0.000$), indicating a greater tendency among Arab writers to express strong contrast and direct correction of preceding material. The referential marker *min khilāl* ('through') with a ratio of 1.32 ($p = 0.000$), the hedging marker *qad* ('may') with a ratio of 1.28 ($p = 0.000$), and the explanatory marker *mithl* ('like/such as') with a ratio of 1.25 ($p = 0.000$) were next, respectively. These findings suggest that even in functions such as additive, contrastive, referential, hedging, and explanatory—where no significant difference exists at the functional level—the choice of specific markers within each function can be significantly different.

Markers Used More Frequently by Persian-Speaking (Iranian) Writers

Conversely, Iranian (Persian-speaking) writers also used five individual markers significantly more often than Arab writers. Notably, three of these five markers belong to the Introductory function. The marker *yatanāwal...* (*al-baḥṭh/al-dirāsah*) ('the study addresses') showed the greatest difference with a ratio of 1.52 ($p = 0.000$). Following this were *yuhāwil...* (*al-baḥṭh/al-dirāsah*) ('the study attempts') with a ratio of 1.44 ($p = 0.000$), the inferential marker *aẓharat natā'ij...* ('the results showed...') with a ratio of 1.35 ($p = 0.000$), *yas'ā...* (*al-baḥṭh/al-dirāsah*) ('the study seeks') with a ratio of 1.31 ($p = 0.000$), and the inferential marker *tawuṣṣala* ('arrived at/found') with a ratio of 1.27 ($p = 0.000$). This finding indicates that Iranian writers have a greater tendency to use verbal structures to introduce their research aims and approaches and also employ more explicit inferential markers when presenting research results and achievements. In contrast, it seems that Arab writers use nominal or non-verbal structures for similar functions, which were not captured within the framework of the markers examined in this study.

The findings of this research are largely consistent with and extend earlier work on meta/discourse markers in Arabic and cross-linguistic academic writing. The most salient points of convergence and divergence are outlined below.

1. Additive and Contrastive Markers

The present results echo those of Basheer (2019) and Livingstone (2019), who both reported that additive markers are the most frequent type of meta/discourse device in academic writing. In addition, the appearance of contrastive markers among the three most frequent marker types constitutes another point of commonality with Basheer (2019).

2. Hedging Markers

A higher frequency of hedging than of boosters also replicates Livingstone's (2019) findings. Moreover, Hayisama et al. (2019) showed that Thai and Malaysian master's students rely heavily on hedging constructions, especially the "*qad + present tense verb*" pattern, to convey uncertainty. Our own analysis demonstrates that Arab writers employ this same construction significantly more often than Iranian writers (Arab to Iranian ratio = 1.35, $p < 0.001$), reinforcing the overall tendency of Arab authors to make more intensive use of hedging devices.

Native vs. Non-Native Writers in Arabic Academic Writing

Recent studies on Arabic academic writing have revealed important patterns in the use of metadiscourse markers by native and non-native writers. Nursa'adah,

Syihabuddin, and Supriadi (2026) compared interactive metadiscourse use between native Arab writers and non-native Indonesian writers in Arabic research article abstracts, finding that non-native writers employed metadiscourse more explicitly while native writers relied more heavily on implicit cohesion. The present study extends this finding by demonstrating that Persian-speaking learners of Arabic also differ from native Arab writers, but primarily at the level of individual markers rather than functional categories. Similarly, Assassi and Merghmi (2023) examined metadiscourse use in applied linguistics abstracts written by Algerian, Saudi, and native English-speaking researchers, finding that Algerian writers approximated the native norm in six types of markers, while Saudi writers did so in only two types. Their findings further indicated that proximity to native-like metadiscourse use does not necessarily correlate with publication success in high-indexed journals. The present study adds to this line of inquiry by demonstrating that Iranian writers differ from Arab writers primarily at the level of individual markers (e.g., *wa*, *bal*, *qad*, *yatanāwal*) rather than functional categories, suggesting that L1 background influences micro-level lexical choices even when macro-level functional patterns remain stable.

Comparison with English-Based Studies

The present findings echo those of Leńko-Szymańska (2008), who documented systematic differences between native and non-native writers in their use of discursal features. A comparable pattern emerges in Lee (2009), whose corpus of English postgraduate writing revealed that non-native students relied more heavily on textual meta/discourse (TMD), whereas native writers favoured interpersonal meta/discourse (IMD). Leńko-Szymańska (2008) also reported that foreign language learners and native speaker students employed a higher frequency of connectors than expert native writers—an effect she attributed to novice writing.

By contrast, the present study does not show an overall advantage for either the Persian-speaking or the Arab group; both groups overuse certain markers and underuse others. These patterns suggest that the observed differences are driven more by linguistic and cultural backgrounds than by writing proficiency alone.

Writing Challenges of Persian-Speaking Learners of Arabic

Zare (2020) examined writing errors among Persian-speaking learners of Arabic at Yazd University, revealing that Iranian students struggle particularly with grammatical and lexical accuracy in their Arabic writing. The present study complements Zare's findings by shifting the focus from error analysis to the positive use of metadiscourse markers, demonstrating that even advanced Persian-speaking learners (Arabic studies graduates) differ from native writers in their selection of specific markers such as *wa*, *bal*, *yatanāwal*, and *tawuṣṣala*.

Furthermore, Ismail and Mat (2024) documented a significant gap between self-perceived and actual writing proficiency among Malaysian learners of Arabic, revealing notable weaknesses in complex sentence structure, idea continuity, and paragraph organization. The present study extends this line of inquiry by demonstrating that Persian-speaking learners of Arabic, despite their advanced training, exhibit differences from native writers in discourse-level features—specifically in the use of metadiscourse markers—suggesting that even advanced learners may benefit from explicit instruction on these rhetorical devices.

Connective Explanatory Markers

The identification of connective explanatory markers among the most frequent devices aligns with the findings of Alenizy et al. (2024), who reported that in doctoral dissertations of Arab students, temporal, contrastive, and inferential markers follow additive markers in frequency. In the present study, temporal markers were not included in the classification; consequently, after additive markers the most frequent functions were referential, explanatory, introductory, and inferential.

Cross-Disciplinary and Cross-Linguistic Influences

Khedri, Chan and Tan (2013) argued that both disciplinary conventions and language context shape the manifestation of meta/discourse markers in academic writing. This perspective helps to explain why our corpus—restricted to abstracts in the humanities and social sciences—shows a largely similar functional pattern across the two writer groups (no significant CORPUS × FUNCTION interaction, $p = 0.063$), yet reveals marked differences at the level of individual markers (significant CORPUS × DM interaction, $p < 0.001$). In other words, the macro-level functional distribution appears to be relatively stable across disciplines, whereas the micro-level lexical choices (e.g., *wa*, *bal*, *qad*) are more sensitive to the specific linguistic and cultural background of the writer—exactly the interaction highlighted by Khedri et al. (2013).

The observed differences at the individual marker level can be explained within the framework of writing style theory and lexico-discoursal differences:

1. Variation in Additive Marker Use: The more frequent use of the marker *wa* ('and') by Arab writers indicates a style that emphasizes sequential connectivity and cohesion through the addition of information. This finding aligns with the Arabic rhetorical tradition, which values *wasl* (connection/continuity).
2. Differing Introductory Styles: The more frequent use of introductory verbal markers (*yatanāwal*, *yuhāwil*, *yas'ā*) by Persian-speaking writers suggests a style that emphasizes agency and the purposiveness of the research. This style may be rooted in the Persian writing tradition, which values explicitly stating research objectives.
3. Inferential Style of Persian-Speaking Writers: The more frequent use of inferential markers like *aẓharat natā'ij...* ('the results showed...') and *tawuṣṣala* ('arrived at/found') by Persian-speaking writers' points to a style that emphasizes presenting results and achievements explicitly.
4. Differences in Contrastive Markers: The more frequent use of the marker *bal* by Arab writers reflects a tendency to express stronger contrasts and directly correct preceding statements.

The findings have important practical implications for teaching Arabic to Persian speakers and for academic writing in Arabic. Even advanced learners show differences: since both groups wrote in Arabic and the Persian-speaking group were Arabic studies graduates, the observed differences indicate that even advanced learners may diverge from native-like patterns in using certain meta/discourse markers. Five markers require special attention: instructors should focus on five markers that Arab writers use significantly more frequently: 1) *wa* (additive) – usage pattern differs from Persian, 2) *bal* (contrastive) – for strong contrast and correction, 3) *min khilāl* (referential) – for indicating means or medium, 4) *qad* (hedging) – for probability and uncertainty, 5) *mithl* (explanatory) – for providing examples.

Specialized instructional materials for Persian speakers should emphasize high-frequency markers in Arab writing such as *bal*, *min khilāl*, and *qad*.

1. Translation and academic writing: Persian-speaking authors and translators, when aware of these differences, can produce texts more aligned with Arab discursal norms (e.g., using *bal* instead of weaker contrastive markers for more natural text).
2. Overall implication: Even at advanced levels of Arabic instruction, attention to subtle differences in meta/discourse markers can significantly improve the quality of academic writing by Persian speakers.

This study faced several limitations that should be considered when interpreting the results.

1. Focus on abstracts only: Findings may not generalize to full-text articles, as abstracts have a concise and specific structure that could yield different patterns of discourse marker use.
2. Limited disciplinary scope: The study only covered humanities and social sciences; other disciplines (e.g., basic sciences, engineering, medicine) were not examined and may show different patterns.
3. Language proficiency not assessed: Although the Persian-speaking group consisted of Arabic studies graduates, their varying proficiency levels were not measured and could have influenced the results.
4. Subjectivity in classification: The functional classification of meta/discourse markers, despite being based on established frameworks and having high reliability (Cohen's Kappa = 0.86), involved some degree of researcher judgment; alternative classifications might lead to different findings.
5. Purposive sampling: The sample was selected purposively, which limits the generalizability of the results to the broader population of Persian-speaking and Arab writers.

Based on the findings and limitations of this study, six main avenues for future research are proposed:

1. Full-text analysis: Examine meta/discourse markers in complete academic articles (across sections like introduction, methodology, findings) to see if patterns observed in abstracts hold true in the main text.
2. Other discourse types: Expand research to spoken discourse (lectures, debates), media discourse (news articles), and literary discourse for a broader cross-cultural perspective. Following the genre-based comparative approach adopted by Bal Gezegin and Bas (2020), which revealed significant differences in interactional meta/discourse use between research articles and book reviews, future studies could extend this line of inquiry to other academic and non-academic genres to determine whether the patterns observed across writer groups remain consistent across different discourse types.
3. Investigate how proficiency level affects discourse marker use, as this was not assessed in language proficiency: the Persian-speaking group, which could inform language teaching.
4. Cross-linguistic comparison: Conduct contrastive studies with other languages (English, French, Turkish) to identify universal vs. language-specific patterns.
5. Social and contextual factors: Examine the role of gender, age, academic background, and field of specialization to determine if factors beyond linguistic background influence marker choice.

Co-occurrence analysis: Analyze which markers frequently appear together and whether collocation patterns differ between Persian-speaking and Arab writers, offering insights into writing styles and textual coherence.

CONCLUSION

The present study aimed to comparatively investigate meta/discourse markers in Arabic academic writing by Iranian and Arab writers. The most important findings are as follows:

1. No Significant Difference at the Functional Level: The pattern of using discourse marker functions does not differ significantly between the two groups ($p = 0.063$). Both groups use additive, referential, explanatory, and introductory functions more frequently than others.
2. Significant Difference at the Individual Marker Level: The pattern of using individual meta/discourse markers differs significantly between the two groups ($p < 0.001$). This finding indicates that discursial differences emerge at the level of specific lexical choices, rather than at the level of overall functions.
3. Most Frequent Marker: The marker *wa* ('and'), accounting for 38.1% of all uses, is the most frequent individual marker, and Arab writers use it significantly more often.
4. Differences in Introductory Markers: Iranian writers use introductory verbal markers such as *yatanāwal*, *yuhāwil*, and *yas'ā* significantly more often.
5. Differences in Inferential Markers: Iranian writers use inferential markers such as *aẓharat natā'ij...* and *tawuṣṣala* significantly more often.
6. Differences in Contrastive and Referential Markers: Arab writers use the markers *bal* (contrastive) and *min khilāl* (referential) significantly more often.

REFERENCES

- Alenizy, K. S., Al-Thunayyan, A. A., & Alhuqbani, M. N. (2024). Analysis Of Discourse Markers And Their Combinations In Nonnative Academic Writing. *Journal of Language Teaching and Research*, 15 (6), 2010-2019.
- Assassi, T., & Merghmi, K. (2023). Formulaic Sequences And Meta-Discourse Markers In Applied Linguistics Research Papers: A Cross-Linguistic Corpus-Based Analysis Of Native And Non-Native Authors' Published Articles. *Academicus International Scientific Journal*, 27, 154-175.
- Bal Gezegin, B., & Bas, M. (2020). Metadiscourse In Academic Writing: A Comparison Of Research Articles And Book Reviews. *Eurasian Journal of Applied Linguistics*, 6 (1), 45-62.
- Basheer, N. (2019). Connectors In The Writing Of Native And Non-Native Arabic Speakers: Similarities, Differences And A Teaching Model. In *The Arabic Classroom* (pp. 199-218). Routledge.
- Boginskaya, O. (2024). A Contrastive Analysis Of Metadiscourse Patterns In Academic Texts By Non-Native Authors. *Lengua y Habla*, 28, 280-297.
- Fraser, B. (1999). What Are Discourse Markers? *Journal of Pragmatics*, 31(7), 931-952.
- Fraser, B. (2006). Towards A Theory Of Discourse Markers. In K. Fischer (Ed.), *Approaches to Discourse Particles* (pp. 189-204). Elsevier.
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. Longman.

- Hardin, J. W., & Hilbe, J. M. (2012). *Generalized Linear Models And Extensions* (3rd ed.). Stata Press.
- Haruna, H. H., Ibrahim, B., Haruna, M., Ibrahim, B., & Yunus, K. (2018). Metadiscourse In Students' Academic Writing: Case Study of Umaru Musa Yar'adua University and Al-Qalam University Katsina. *International Journal of English Linguistics*, 8(7), 83-92.
- Hayisama, F., Shah, M. I. A., & Adnan, W. N. A. W. (2019). Rhetorical Style Across Cultures: An Analysis Of Metadiscourse Markers In Academic Writing Of Thai And Malaysian Students. *LSP International Journal*, 6 (1).
- Hsu, J. C., & Nelson, B. (1998). Multiple Comparisons In The General Linear Model. *Journal of Computational and Graphical Statistics*, 7(1), 23-41.
- Hyland, K. (2005). *Metadiscourse: Exploring Interaction In Writing*. Continuum.
- Hyland, K., & Tse, P. (2004). Metadiscourse In Academic Writing: A Reappraisal. *Applied linguistics*, 25(2), 156-177.
- Iranian Association of Arabic Language and Literature Scientific Research Journal. (2019-2021). Issues 58-77.
- Ismail, M. R., & Che Mat, A. (2026). Bridging The Gap Between Perceptions And Reality In Arabic Writing Proficiency Among Malaysian Higher Education Students: A Triangulation Approach. *Ijaz Arabi: Journal of Arabic Learning*, 9 (1), 482-489.
- Khedri, M., & Basirat, E. (2022). Interactive Metadiscourse In Dentistry Research Articles: Iranian Vs Non-Iranian Academic Writers. *Discourse and Interaction*, 15(2), 77-100.
- Khedri, M., Swee Heng, C., & Bee Hoon, T. (2013). Cross-Disciplinary And Cross-Linguistic Perspectives On Metadiscourse In Academic Writing. *Southern African Linguistics and Applied Language Studies*, 31(1), 129-138.
- Lark Journal of Philosophy, Linguistics, and Social Sciences. (2024-2025). Volumes 17(1-4) and 18(1-4).
- Lee, S. (2009). *Metadiscourse In Academic Writing: A Corpus-Based Study Of Expert, L1 And L2 Postgraduate Student Text* (Doctoral dissertation, Newcastle University).
- Leńko-Szymańska, A. (2008). Non-Native Or Non-Expert? The Use Of Connectors In Native And Foreign Language Learners' Texts. *Acquisition et Interaction En Langue Étrangère*, (27), 91-108.
- Livingstone, K. A. (2019). Examining The Use Of Metadiscourse Markers In Academic Writing. *International Journal Of Literature, Language And Linguistics*, 5(3), 244-254.
- McEnery, T., & Hardie, A. (2012). *Corpus Linguistics: Method, Theory And Practice*. Cambridge University Press.
- Mirshamsi, A. S., & Allami, H. (2013). Metadiscourse Markers In The Discussion/Conclusion Section Of Persian And English Master's Theses. *Journal of Teaching Language Skills*, 5(2), 23-40.
- Nursa'adah, S., Syihabuddin, & Supriadi, R. (2026). Interactive Metadiscourse In Arabic Abstracts: A Comparative Study Of Kitābah Implications For Native And Non-Native Writers. *Alibbaa': Jurnal Pendidikan Bahasa Arab*, 7 (1). 181-205.

- Pahlevannezhad, M. R., & Alinezhad, B. (2012). Contrastive Rhetorics and the Study of Metadiscourse in Persian Native Speakers' and Arab Learners' Compositions. *Journal of Teaching Persian to Speakers of Other Languages*, 1(1), 79-100.
- Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful Sampling For Qualitative Data Collection And Analysis In Mixed Method Implementation Research. *Administration and Policy in Mental Health and Mental Health Services Research*, 42(5), 533–544.
- Pearson, W. S., & Abdollahzadeh, E. (2023). Metadiscourse In Academic Writing: A Systematic Review. *Lingua*, 293, 103561.
- Sanosi, A. B., & Mohammed, A. S. E. (2024). A Corpus-Based Analysis Of Arab Scholars' Use Of Interactional Metadiscourse Markers. *International Journal of English Linguistics*, 13(2), 188-200.
- Zare, S. (2020). The Effect Of "Pattern Practice In Translation" On Developing Persian Speaking Arabic Students' Writing Skill. *Language Related Research*, 11 (1), 277-300.