ISSN: 1994-4217 (Print) 2518-5586(online)

Journal of College of Education

Available online at: https://eduj.uowasit.edu.iq

Dr. Ali Yasir Alsahlanee

Al-Iraqia University, Baghdad, Iraq

Email: ali.fahad@aliraqia.edu.iq

Keywords:

Arts and Humanities, Iraqi academics, Lexical Bundles, Research Article Abstracts

A r t i c l e i n f o

Article history:

Received 20. Jun.2026

Revised 22. Jul.2026

Accepted 29. Jul.2026

Published 25. Nov.2026

An Investigation of Lexical Bundles in the Abstracts of Iraqi and Saudi Arts and Humanities Research Articles

A B S T R A C T

Lexical bundles are building blocks that are used to structure arguments and meet genre-specific expectations. They contribute to textual cohesion and rhetorical organisation, as lexical bundles allow writers to achieve various communicative functions in the genre of research article (RA) abstracts. This study investigates the use of lexical bundles in the texts of Iraqi and Saudi RA abstracts in the Arts and Humanities field. These RAs are published in journals indexed in the Scopus database. A total of 617 four-word lexical bundles were extracted from the corpus of the study. All bundles were classified according to their structure and function, following the taxonomies of Biber et al. (1999) and Hyland (2008b). The Mann–Whitney U test results revealed a significant difference in the overall use of lexical bundles between the abstracts of the two study corpora. The structural distribution of lexical bundles differed significantly between Iraqi and Saudi abstracts, while the functional distribution revealed no significant difference. Iraqi abstracts rely more heavily on noun-phrase-based and phrasal bundles, suggesting a preference for condensed informational packaging. These findings highlight the importance of examining the linguistic features of Iraqi RAs' texts to meet publishers' expectations and secure publication in reputable, indexed journals.

© 2026 EDUJ, College of Education for Humans Sciences, Wasit University

DOI: https://doi.org/10.31185/eduj.Vol65.Iss2.5277

دراسة استقصائية للحزم المعجمية في مستخلصات البحوث العراقية والسعودية في مجالات الآداب والعلوم الإنسانية

م.د. علي ياسر السهلاني

الجامعة العراقية، العراق، بغداد

الملخص:

تُعدّ الحزم اللفظية (Lexical Bundles) وحداتٍ بنائية تُستخدم لتنظيم الحجج وتحقيق المتطلبات الخاصة بالنوع الخطابي. كما تسهم في تحقيق التماسك النصي والتنظيم البلاغي، إذ تُمكّن الكُتّاب من أداء وظائف تواصلية متعددة في نوع مستخلصات المقالات البحثية. تهدف هذه الدراسة إلى استقصاء استخدام الحزم اللفظية في نصوص مستخلصات المقالات البحثية العراقية والسعودية في مجال الآداب والعلوم الإنسانية. وقد نُشرت هذه المقالات في مجلات مفهرسة ضمن قاعدة بيانات سكوبس. تم استخراج ما مجموعه ٦١٧ حزمة لفظية رباعية الكلمات من مدونة الدراسة. وصُنِّفت جميع الحزم وفقًا لبنيتها ووظيفتها اعتمادًا على تصنيفات بيبر وآخرين (١٩٩٩) وهايلاند (٢٠٠٨ب). وكشفت نتائج اختبار مان– ويتني (Mann–Whitney U) عن وجود فرق دال إحصائيًا في الاستخدام الكلي للحزم اللفظية بين مستخلصات مدونتي الدراسة. كما أظهر التوزيع البنيوي للحزم اللفظية فروقًا دالة بين المستخلصات العراقية والسعودية، في حين لم يكشف التوزيع الوظيفي عن أي فروق دالة إحصائيًا. وتعتمد المستخلصات العراقية بدرجة أكبر على الحزم القائمة على العبارات الأسمية والحزم العبارية، مما يشير إلى تفضيل أسلوب التكثيف المعلوماتي. وتبرز هذه النتائج أهمية دراسة الخصائص اللغوية لنصوص المقالات البحثية العراقية من أجل تلبية توقعات الناشرين وضمان النشر في المجلات المرموقة والمفهرسة.

الكلمات المفتاحية: ××××××××××××××××××××××××××××××××

1. Introduction

In the context of academic writing, previous studies have estimated the amount of formulaic language in academic discourse as 20% to 50% (Biber et al., 1999; Erman & Warren, 2000). Formulaic language refers to a ready chunk of language that writers use as a single unit in composing text. The use of formulaic language enables writers to structure arguments, convey disciplinary meanings, and meet genre-specific expectations. Formulaic language units, such as lexical bundles, contribute to textual cohesion and rhetorical organisation (Biber et al., 1999). In research article abstracts, formulaic sequences enable writers to serve multiple communicative functions within a limited textual space (Hyland, 2008b). As a result of this widespread use of formulaic language, corpus studies have paid considerable attention to linguistic patterns such as lexical bundles, collocations, and n-grams in academic discourse. These recurring sequences are classified according to their frequency and function in a discourse.

They serve as building blocks for presenting and communicating findings fluently across different academic sub-fields (Hyland, 2012). A lexical bundle is defined as a sequence of words (usually three words or more) that co-occur frequently in discourse (Biber et al., 2004;

Durrant, 2017; Hyland, 2008b). These lexical sequences play an important role in organising text and meaning (Hyland, 2008a; Ren, 2021). Therefore, academics need to use these building blocks carefully in their RAs, especially in the abstract, since it is the first part of a research article and readers rely on the information presented there to decide whether to read the full text. Most of the previous studies have focused on analysing and comparing lexical choices in texts written by English L1 and L2 writers (Appel, 2022; Cortes, 2006; Esfandiari & Barbary, 2017; Grabowski, 2015; Pan et al., 2016). Despite extensive research on lexical bundles, academic texts written by Arab academics, especially Iraqis, remain relatively underexplored (Sanosi, 2022). Investigating the use of lexical bundles in texts written by Iraqi academics helps reveal the linguistic features of such texts and identify ways to improve academic texts intended for publication.

Iraqi academics, as well as their peers around the globe, are under pressure to publish their findings in international journals, especially those indexed in databases like Scopus and Web of Science, to gain academic promotion. Iraqi Arts and Humanities RAs receive the lowest per-document citation rate among Middle Eastern countries, with 3.25 citations per document. On the other hand, Saudi RAs receive the highest citation rate in the region (*SJR - International Science Ranking*, n.d.). These facts motivate the current study to investigate the writing of Iraqi academics in the Arts and Humanities fields by comparing RAs' abstracts written by Iraqi and Saudi academics to reveal the linguistic features of texts intended for publication purposes and to provide insights into how Iraqi academics construct their academic discourse. Thus, this study aims to answer the following research questions:

1. Do Iraqi and Saudi research article abstracts differ significantly in the overall frequency of lexical bundles?
2. Are there any significant structural differences in the distribution of lexical bundles between Iraqi and Saudi abstracts?
3. Are there any significant functional differences in the distribution of lexical bundles between Iraqi and Saudi abstracts?

2. Literature Review

2.1 Lexical Bundles

Lexical bundles are traditionally defined as recurrent sequences of words or expressions, regardless of their idiomaticity or structure (Biber et al., 1999). They usually consist of three or more words that commonly occur together (Chen & Baker, 2010). In terms of frequency of occurrence, lexical bundles are described as building blocks (sequences of words) that are used together frequently in a particular register (Ädel & Erman, 2012; Cortes, 2004). More attention has been paid to the function of lexical bundles in a specific discourse, Hyland (2008b) states that lexical bundles are sequences of words commonly used together to serve a particular function in that discourse.

To illustrate, the most frequent four-word bundles in the abstracts examined in this study include *the aim of this*, *the results showed that*, and *on the other hand*. Each recurs across dozens of texts and does a recognisable job: the first announces the purpose of a study, the second reports its findings, and the third signals a contrast between propositions. Other attested bundles mark methodological procedures (*data were collected from*), frame the

research context (*in the context of*), or express writer stance (*it was found that*). These examples show how a small stock of prefabricated sequences carries much of the routine communicative work of an abstract.

Previous research has endeavoured to categorise lexical bundles; as a result, they have been classified by structure and function. At the structural level, Biber et al. (1999) identified several common grammatical structures of lexical bundles (Noun phrase + of, e.g., *the results of the*; Prepositional phrase, e.g., *in the context of*; Anticipatory it + VP, e.g., *it was found that*). On the other hand, Hyland (2008b) proposed another classification of lexical bundles based on their function in academic writing discourse; he grouped them into three types (Research-Oriented, Text-Oriented, Participant-Oriented). Research-oriented bundles help writers describe the data, context, and procedures of their research (e.g., *this study aims to*, *data were collected from*). The second group, text-oriented bundles, are usually used to organise, clarify, and guide discourse (e.g., *in the context of*, *on the other hand*). Lastly, participant-oriented bundles are used by writers to express stance, degree of certainty, and engagement (e.g., *it was found that*, *it is recommended that*). Participant-oriented bundles are commonly used in the abstract section of research articles to express the writer's attitude or evaluation.

In corpus linguistics, the term "lexical bundles" has played a significant role in the study of academic writing. At the first stage, studies tried to identify features of these multiword sequences (Biber et al., 1999; Cortes, 2004, 2013; Hyland, 2008b). Then, studies moved to identify the connection between lexical bundles and rhetorical moves and steps of a specific genre of writing (Cortes, 2013; Gray et al., 2020; Mizumoto et al., 2017; Omidian et al., 2018; Qi & Pan, 2020).

The findings of these studies revealed that certain lexical bundles are used more frequently in a specific move. The frequency of lexical bundle use also varies across disciplines, as shown in Omidian et al.'s (2018) study, which investigates the distribution of lexical bundles in the abstracts of six disciplines.

Several studies have compared lexical bundle use among writers with different first-language backgrounds (Appel, 2022; Cortes, 2006; Esfandiari & Barbary, 2017; Grabowski, 2015; Pan et al., 2016). Persian authors were found to use fewer lexical bundles in their research articles than English-native authors (Esfandiari & Barbary, 2017). On the contrary, Lu & Deng (2019) found that Chinese PhD dissertation abstracts contain more lexical bundles than those written by native English-speaking peers. Contrary to both previous studies, Shin (2019) concluded that non-native and native English writers behave similarly when writing argumentative essays.

2.2 Research Articles Abstracts

The abstract of a research article is considered a distinct genre of academic writing; it contains a lot of information in a limited number of words. It is considered the summary of a research article (Corbett, 2020). Salager-Meyer et al. (2014, p. 222) described abstracts as "short, concise, complete, and accurate source of information". Abstracts are the most-read section of research articles, since they are freely accessible and readers rely on them to decide

whether to read the rest of the article (Shiely et al., 2024). Writers are required to present, in their abstracts, a summary of the background, aims, methods, results, and conclusions (Behzadi & Gajdács, 2021). Therefore, they rely on using lexical bundles, especially participant-oriented bundles, to achieve different communicative goals in their abstracts (Cortes, 2004).

Abstracts of research articles are classified generally into three types: informative, indicative, and combinatory (Bhatti et al., 2019; Lorés, 2004). The first and most common type is the informative abstract, which outlines the structure of the article, presenting the information in the introduction, methods, results, and discussion (IMRD). On the other hand, the indicative abstract summarises information on the background of the study, the research gap, the research questions, and the findings. It does not include results, data, or detailed conclusions. This type of abstract focuses on presenting how the study's gap has been addressed and how the research questions have been answered. Lastly, the combinatory abstract is the mix of both informative and indicative abstracts (Bhatti et al., 2019).

Previous studies on the use of lexical bundles in research articles have revealed notable findings and significant differences in their use between abstracts from different disciplines and between abstracts written by English L1 and non-native English writers. Bao & Liu (2022) compared abstracts of linguistics dissertations written by American and Chinese students; they found a significant difference in the use of lexical bundles between the two groups. Shahmoradi et al. (2021) conclude in their study that the use of lexical bundles in abstracts is sensitive to disciplinary norms and rhetorical functions. It has also been noted that scholars in soft disciplines pay considerable attention to highlighting the importance of objectives and implications due to the difficulty of defining theoretical structures (Omidian et al., 2018).

2.3 Iraqi Academics Writing for Publication

Iraqi academics, especially in the Arts and Humanities, are publishing few research articles in international journals indexed in the Web of Science and Scopus databases. This situation adversely affects the rankings and reputations of both academics and institutions, as publication productivity is a key indicator of their performance (Lee et al., 2025). Institutions around the globe encourage their academics to publish their research in international journals, especially those indexed in reputable databases, to achieve high rankings and a prestigious reputation. In addition to publication productivity, the citation rate is another indicator of quality and performance, as it reflects academics' contributions to the community of practice in a specific field (Swales, 1990). It also contributes to the H-index; therefore, academics are eager to earn more citations for their work to increase it. It has been found that an Iraqi Arts and Humanities research article receives around 3.25 citations, the lowest per-document citation rate among regional countries (*SJR - International Science Ranking*, n.d.), as shown in Table 1.

Table 1 Citations Per-Document of Middle East Countries

Rank	Country	Documents	Citations per document
1	Saudi Arabia	4465	21.78
2	Syrian Arab Republic	196	18.93
3	Kuwait	696	17.12
4	Lebanon	1476	15.91
5	United Arab Emirates	2962	13.7
6	Qatar	1188	12.05
7	Yemen	201	11.75
8	Egypt	3597	9.79
9	Oman	735	8.72
10	Turkey	21389	8.22
11	Bahrain	276	7.35
12	Iran	9916	6.77
13	Palestine	479	6.54
14	Jordan	4483	3.43
15	Iraq	1634	3.25

Previous studies on the writings of Iraqi academics have shown weaknesses in English-language skills. In a study of Iraqi postgraduate students, Mohammed et al. (2015) concluded that they commit grammatical mistakes in the use of tenses, prepositions, articles, active and passive voice, and verbs, as well as morphological errors. This low proficiency may impact their academic writing as well as their citation rate, since the citation rate of academic research articles is affected by writing proficiency (Ma & Qin, 2017; Vasconcelos et al., 2008). Iraqi academics were also found to use fewer complex lexical items and syntactic structures in their linguistics research articles than their peers, English L1 and L2 academics (Alsahlanee & Jaganathan, 2022, 2023). Also, Iraqi M.A. students were found to engage in plagiarism in their writing (Hreiz & Al-Attabi, 2024). In general, the limited use or misdeployment of conventionalised lexical bundles may affect Iraqi academics' writing in establishing the rhetorical norms followed in reputable, indexed journals (Alsahlanee & Jaganathan, 2022, 2023). Mastering the use of prefabricated chunks is essential for adhering to disciplines' conventions and realising rhetorical structures of effective abstracts (Omidian et al., 2018).

3. Methodology

3.1 Data of the Study

The data for this study include abstracts of published RAs in the Arts and Humanities. The electronic corpus of this study is divided into two corpora based on country: Iraq and Saudi Arabia. The Iraqi corpus comprises 1,085 RA abstracts, which were all indexed under Iraqi affiliations and published in journals indexed in the Scopus database. On the other hand, the Saudi corpus consists of 2,846 RA abstracts that are indexed and published under Saudi affiliations. The criteria for selecting RAs were to include all RAs in the subject area (Arts and Humanities), written by academics affiliated with Iraqi and Saudi institutions, and published in journals indexed in the Scopus database for the period between 2020 and 2025. The selection of Arts and Humanities RAs was motivated by the fact that Iraqi RAs had the lowest citation rate compared with RAs from other countries in the region. Saudi RAs, on the other hand, were found to have the highest citation rate among Middle Eastern countries. Due to the low productivity of Iraqi academics in international journals indexed by Scopus, the two corpora are not size-balanced. To mitigate the effect of corpus size imbalance, analyses were conducted using distributional percentages and non-parametric statistical tests. In the Iraqi corpus, the number of running words is 203,966, and the average number of words in each abstract is 188, while in the Saudi corpus, the number of running words is 568,819, and the average number of words in each abstract is 199. Prior to corpus analysis, the selected RAs' abstracts were downloaded and saved as plain text files.

3.2 Lexical Bundles Analysis

The criteria used to identify lexical bundles are based on length, frequency of occurrence, and range. This study focused on analysing 4-word lexical bundles, as this type of bundle is considered more common than 5- and 6-word bundles (Hyland, 2008b). Also, the structures of 4-word lexical bundles include many 3-word lexical bundles (Cortes, 2004). This study set a lexical bundle frequency cut-off of 10, following previous studies of lexical bundles (Altenberg, 1998; Shahmoradi et al., 2021; Varghaei & Khodadadi, 2022). Lastly, as a range criterion, only lexical bundles that occurred in at least 10 different texts were included in the analysis.

This study utilised the automatic toolkit AntConc 4.3.1 to analyse and compare lexical bundles in the texts of RAs' abstracts. AntConc 4.3.1 is a concordance program that is freely available for extracting and counting lexical bundles according to preferred criteria. The total number of lexical bundles extracted from Iraqi and Saudi RAs' abstracts is 617. To reveal the structures and functions of the identified lexical bundles, the current study followed Biber et al.'s (1999) classification of lexical bundle structures and Hyland's (2008b) taxonomy of lexical bundle functions. The two frameworks and their sub-categories are summarised in Figure 1, together with corpus examples of each category.

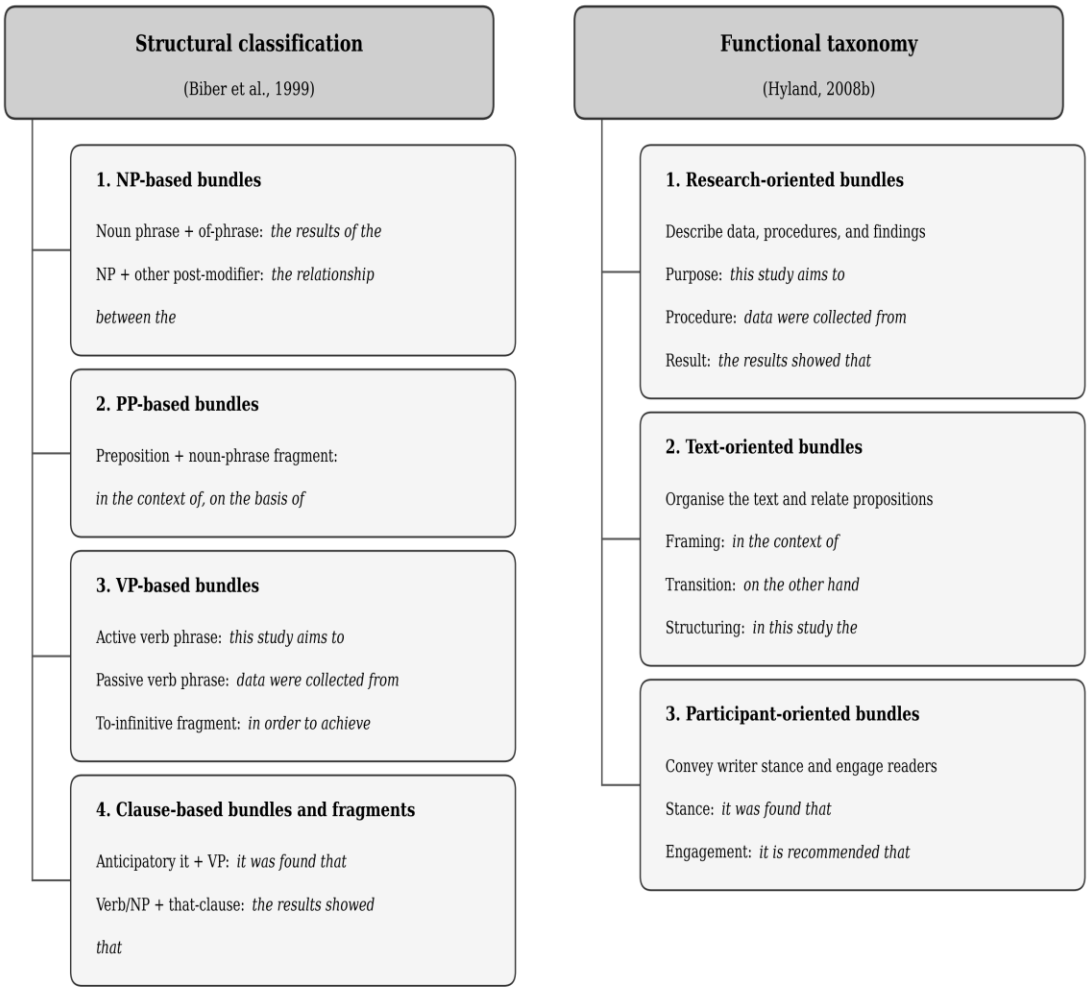


Figure 1 Biber et al.'s (1999) Structural Classification and Hyland's (2008b) Functional Taxonomy of Lexical Bundles, with Examples from the Study Corpus

3.3 Statistical Analysis

This study utilised descriptive and inferential statistics to answer its research questions. IBM SPSS Statistics version 30 was used to analyse the data of the study. The descriptive statistics are presented as frequencies and percentages of lexical bundle distributions across structural types and functional categories in the abstracts of Iraqi and Saudi RAs. In order to answer the first research question, inferential statistics were utilised by conducting a Mann–Whitney U test to compare the use of lexical bundles in the abstracts of Iraqi and Saudi RAs. The Mann–Whitney U test was used because of the non-normal frequency distribution of lexical bundles and differences in corpus sizes. The use of nonparametric tests for skewed linguistic frequency data is well established in statistics and corpus linguistics research (Field, 2018; Gries, 2021). For the second and third research questions, chi-square tests were conducted to determine whether there are significant differences in the structure and function of lexical bundles used in abstracts of Iraqi and Saudi RAs.

4. Results

RQ. 1 Do Iraqi and Saudi research article abstracts differ significantly in the overall frequency of lexical bundles?

To compare the overall use of lexical bundles in the writing of Iraqi and Saudi RAs' abstracts, a non-parametric Mann–Whitney U test was utilised due to the unequal sizes of the two corpora and the non-normal distribution of lexical bundle counts. Examples (1) and (2) illustrate this contrast. Both are typical realisations of the purpose and method moves, assembled from the most frequent bundles attested in each corpus, with raw frequencies given in parentheses. (1) Iraqi corpus: abstracts typically announce the aim through a small set of noun-anchored sequences, as in "The aim of this study is to shed light on …", built from *the aim of this* (25) and *of this study is* (31). Method reporting, by contrast, is rarely formulaic: apart from *is based on the* (13), no passive data-collection bundle reaches the frequency threshold in the Iraqi corpus. (2) Saudi corpus: abstracts deploy the full conventional repertoire across moves, as in "This study aims to examine … Data were collected from … The results showed that …", where each sequence ranks among the most frequent Saudi bundles: *this study aims to* (146), *data were collected from* (50), and *the results showed that* (84).

The results showed a significant difference in the frequency of lexical bundles used in Iraqi and Saudi RAs' abstracts, $U = 3{,}764.00$, $z = -13.41$, $p < .001$, $r = .54$. Saudi abstracts included a significantly higher frequency of lexical bundles than Iraqi abstracts. Since the two corpora are unequal in size, this difference could affect the calculation of lexical bundle density. The density of lexical bundles was calculated per 1,000 words to control for this variable. The results showed that the Saudi corpus exhibited a higher bundle density (0.91 per 1,000 words) than the Iraqi corpus (0.48 per 1,000 words), indicating greater formulaic language use in Saudi abstracts. The normalised density confirms that the observed difference is not due to corpus size.

Table 2 Lexical bundle density per 1,000 words

Corpus	Tokens	Bundles	Density per 1,000 words
Iraqi	203,966	98	0.48
Saudi	568,819	519	0.91

These findings reflect the high density and diversity of lexical bundles in the abstracts of Saudi RAs. Iraqi RAs exhibit lower lexical bundle frequency and diversity than Saudi RAs. Iraqi writers in engineering fields have also been reported to use a small variety of lexical bundles in their academic texts (Alkhafaji et al., 2022). Previous research across various corpora has shown that non-native English writers use a narrower range of lexical bundles, resulting in more repetitive and less diverse discourse (Appel et al., 2024; Bao & Liu, 2023; Li & Lei, 2025). It has been reported that higher bundle density is often associated with greater familiarity with academic phraseological conventions, as lexical bundles represent recurrent patterns that facilitate fluency and cohesion in academic discourse (Esfandiari & Barbary, 2017; Hyland, 2008b). A greater concentration of such sequences may therefore

reflect stronger alignment with conventional academic phraseology. Based on the previous findings, we can conclude that Iraqi academics rely on a limited set of lexical bundles when writing RA abstracts. This could be one factor affecting the citation rate of Iraqi academics, since the abstract section is the first part of RAs, and readers rely on the information presented there to decide whether to read the full text.

RQ. 2 Are there any significant structural differences in the distribution of lexical bundles between Iraqi and Saudi abstracts?

The second research question aimed to compare the structural distribution of lexical bundles in the abstracts of Iraqi and Saudi RAs. To examine the structure of lexical bundles in abstracts from Iraqi and Saudi RAs, a chi-square test of independence was conducted. Examples (3) and (4) illustrate the structural contrast between the two corpora. (3) Iraqi corpus: high-frequency bundles are mostly noun- and preposition-anchored, as in "One of the most important … the relationship between the …": *one of the most* (55), *of the most important* (29), and *the relationship between the* (14) all package information in condensed nominal form and leave the research action implicit. (4) Saudi corpus: verb-based and frequently passive structures are prominent, as in "Data were analyzed using … It was found that …": *data were analyzed using* (23) and *it was found that* (37) foreground research processes in clausal form.

The findings showed a significant relation between variables, χ^2 (26, N = 617) = 79.91, $p < .001$, with a moderate-to-large effect size (Cramer's V = .36). These findings reveal that there are significant differences in terms of the structures of lexical bundles used in Iraqi and Saudi RAs' abstracts. This shows that writers of the two corpora utilise different structural realisations of lexical bundles.

In both corpora, phrasal bundles occurred more frequently than clausal bundles. This finding aligns with previous studies showing that academic texts tend to include more phrasal than clausal lexical bundles (Esfandiari & Barbary, 2017; Li & Lei, 2025; Pan et al., 2016). In this study, most of the extracted bundles comprised nominal, prepositional, and verbal structures. However, significant differences were observed in the distribution of lexical bundles by structure. Nominal and prepositional phrase structures are the most common lexical bundle structures in the abstracts of Iraqi RAs (51.1%). This reflects a tendency to focus on content and framing rather than on elaboration through clausal structures. The use of phrasal bundles is considered one main characteristic of advanced academic writing (Esfandiari & Barbary, 2017; Hyland, 2008b). The extensive use of phrasal bundles signifies the shift from conversational style to a more formal, dense, and grammatically compressed discourse (Biber et al., 1999; Pan et al., 2016). Saudi abstracts, on the other hand, include a high proportion of clausal and verb-based lexical bundles. They utilised a wide diversity of lexical bundles: NP-based bundles (18.7%), VP-based bundles (17.5%), and NP+of bundles (10.8%). The results also showed that Saudi abstracts include more VP-based lexical bundles (7.3%) than Iraqi abstracts (1.0%). These findings revealed that Saudi RAs' abstracts include a wide range of lexical bundles to serve different communicative purposes.

Table 3 Structural distribution of lexical bundles across Iraqi and Saudi abstracts

Structural Category	Iraqi n (%)	Saudi n (%)
NP-based (of-phrase)	23 (23.5)	56 (10.8)
NP-based (general)	8 (8.2)	97 (18.7)
PP-based structures	19 (19.4)	71 (13.7)
VP-based (active)	6 (6.1)	91 (17.5)
VP-based (passive)	1 (1.0)	38 (7.3)
Other structures	41 (41.8)	166 (32.0)
Total	**98 (100)**	**519 (100)**

Note: Minor categories are collapsed for clarity.

The observed differences in structural bundle use can be interpreted in the context of broader patterns of academic discourse. Iraqi abstracts rely more heavily on noun-phrase-based and phrasal bundles, suggesting a preference for condensed informational packaging. In contrast, Saudi abstracts demonstrate greater use of clausal and verb phrase-based bundles, indicating a more dynamic and process-oriented style of expression. This pattern may be explained by the possibility that Saudi writers are more closely aligned with international academic writing conventions, in which procedural clarity and explicit reporting of research processes are emphasised.

RQ. 3 Are there any significant functional differences in the distribution of lexical bundles between Iraqi and Saudi abstracts?

The last research question aimed to determine whether there are significant differences in the functions of lexical bundles used in the abstracts of Iraqi and Saudi RAs. Therefore, the chi-square test was utilised to answer this research question. Examples (5) and (6) illustrate the shared functional profile of the two corpora. (5) Iraqi corpus: the most frequent bundles serve research-oriented purposes, e.g., *this study aims to* (48), followed by text-oriented organisation, e.g., *as well as the* (39); participant-oriented bundles are nearly absent, limited to *that there is a* (27) and *the importance of the* (12). (6) Saudi corpus: the same ordering holds, with research-oriented *the purpose of this* (92), text-oriented *in the context of* (90), and a modest participant-oriented layer, e.g., *it is recommended that* (19).

The results of the chi-square test of independence showed no significant difference: χ^2 (2, N = 617) = 3.27, p = .195, Cramer's V = .073. According to Cohen (2013), this effect size indicates a very small effect, suggesting that no significant difference was found in the distribution of lexical bundles in abstracts from Iraqi and Saudi RAs. The most frequent functional categories used in both Iraqi and Saudi abstracts are research-oriented lexical bundles, followed by text-oriented lexical bundles, and lastly, participant-oriented lexical bundles. Research-oriented bundles were the most frequent category in both corpora, confirming their central role in academic abstracts.

Within this category, bundles associated with procedures (e.g., data were collected from, used to analyse the) and descriptions of results (e.g., the results showed that, findings indicate that) were particularly prominent. These bundles serve to present methodological steps and summarise findings, which are essential communicative functions in abstracts. Text-oriented

bundles, which help organise discourse and guide readers through the text, constituted the second most frequent category in both corpora.

These include transition signals and framing bundles. The similarity in the proportions of text-oriented bundles suggests that both Iraqi and Saudi writers demonstrate awareness of the need to structure information coherently in abstracts. Lastly, participant-oriented bundles were the least frequent category in both corpora, which is consistent with previous findings that academic abstracts tend to minimise explicit author stance and reader engagement. The relatively low frequency of these bundles reflects the impersonal, informational nature of abstracts, where the emphasis is on reporting research rather than engaging with readers.

Table 4 Functional distribution of lexical bundles

Functional Category	Iraqi n (%)	Saudi n (%)
Research-oriented	71 (72.4)	362 (69.7)
Text-oriented	25 (25.5)	122 (23.5)
Participant-oriented	2 (2.0)	35 (6.7)
Total	**98 (100)**	**519 (100)**

These findings show that Iraqi and Saudi abstracts include similar percentages of lexical bundles for achieving goals such as describing methods, reporting results, and organising abstract information. Both groups of writers use lexical bundles from the same functional class; this similarity in functional distribution between the two corpora suggests that, despite structural differences, academics in both contexts follow comparable rhetorical conventions of academic writing. The current findings align with previous studies on non-native academics, which found that they rely on utilising more research-oriented than text-oriented bundles (Li & Lei, 2025; Shirazizadeh & Amirfazlian, 2021).

5. Conclusion

The present study systematically analysed and compared the use of lexical bundles in Iraqi and Saudi RA abstracts. Biber et al.'s (1999) taxonomy was employed for lexical bundle structure analysis, while Hyland's (2008b) taxonomy was applied to determine their functions. Taking into account the size difference between the two corpora of this study, the results revealed significant differences in the use of lexical bundles in the abstracts of Iraqi and Saudi RAs. Iraqi abstracts contained significantly fewer lexical bundles than Saudi abstracts. These results showed that Iraqi academics rely on a limited set of lexical bundles in composing their academic texts. This is supported by findings of other works done on non-native English academics; they conclude that texts written by non-native English academics include a narrower range of lexical bundles and tend to rely on a limited set of bundles in composing their academic texts (Alkhafaji et al., 2022; Appel et al., 2024; Bao & Liu, 2022). Since the use of lexical bundles contributes to textual cohesion and rhetorical organisation (Biber et al., 1999), we can conclude that Iraqi academics, specifically in the field of Arts and Humanities, are not utilising formulaic language in composing their texts intended for publication to the same extent as their peers in the community of practice (Swales, 1990).

Formulaic language helps academics to structure arguments, convey disciplinary meanings, and meet genre-specific expectations.

At the structural level of classification, the findings revealed a significant difference in the use of lexical bundles between the abstracts of Iraqi and Saudi RAs. This difference reflects broader patterns in academic discourse: Iraqi abstracts rely more heavily on noun-phrase-based constructions and phrasal bundles, suggesting a preference for condensed information packaging. In contrast, Saudi abstracts employ clausal and verb-phrase-based bundles more frequently, indicating a dynamic, process-oriented style of expression.

Unlike the structural level of classification, the functional classification shows no significant differences between the two groups of RAs' abstracts. This finding supports Hyland's (2008b) argument that genre conventions regulate functional choices. As reported in previous studies, the data from this study revealed that the research-oriented bundles were the most frequent category in both corpora, confirming their central role in academic abstracts (Li & Lei, 2025; Shirazizadeh & Amirfazlian, 2021).

Despite its findings, this study has several limitations. The two corpora are not fully balanced in size, and the analysis is limited to research article abstracts, which represent a highly condensed academic genre. Future research could extend this work by examining lexical bundle use across sections of research articles, disciplines, or longitudinal learner corpora to explore developmental patterns. Several factors help explain why Saudi academics outperform their Iraqi counterparts in the quantity and diversity of lexical bundles. The first is exposure. Saudi academics publish far more in Scopus-indexed Arts and Humanities venues (4,465 documents against 1,634 Iraqi documents in the SJR ranking), and sustained participation in international publishing familiarises writers with the phraseology that editors, reviewers, and readers expect (Hyland, 2008b). The second is institutional context. Saudi universities have invested heavily in research support, publication incentives, and English-medium instruction over the past decade, whereas Iraqi higher education has operated for years under conflict, sanctions, and relative isolation from the international research community, which limits access to training in English for research publication purposes. The third is language proficiency. Previous work has documented grammatical and lexical weaknesses in the academic English of Iraqi writers (Alsahlanee & Jaganathan, 2022; Mohammed et al., 2015), and command of conventionalised multiword sequences is a late-developing component of second-language writing ability. Together, these conditions give Saudi academics more opportunities to acquire, and stronger incentives to deploy, the formulaic repertoire of international academic English.

6. Implications and Recommendations

The findings carry practical lessons for academics preparing abstracts for international journals, and for Iraqi academics in particular. Lexical bundles are learnable. Academics can build a personal inventory of high-frequency bundles organised by rhetorical move: purpose (*this study aims to*), method (*data were collected from*), results (*the results showed that*), and implications (*it is recommended that*). Reading recently published abstracts in the target journal and noting the recurrent sequences used at each move is a simple way to construct such an inventory. Freely available concordance tools such as AntConc also allow writers to

check their own drafts against a reference corpus and identify moves in which conventional phraseology is missing.

The Iraqi corpus points to two specific gaps that deserve attention. Passive method-reporting frames (*data were collected from*, *data were analyzed using*), which are among the most frequent Saudi bundles, barely occur in Iraqi abstracts; adopting these frames would make the reporting of procedures more explicit and better aligned with international conventions. Participant-oriented bundles are similarly underused, with only two such bundles reaching the threshold in the Iraqi corpus. Sequences that express stance and engage readers (*it is recommended that*, *there is a need*) help authors state the value of their work, which matters in an abstract that must persuade readers to continue to the full text.

For writing instructors and Iraqi institutions, the results support explicit instruction in lexical bundles within courses on English for research publication purposes. Cortes (2006) showed that bundles can be taught productively within disciplinary writing courses and move-based instruction that links specific bundles to specific rhetorical functions (Cortes, 2013) is well suited to abstract writing. A note of caution applies: bundles should serve communicative goals rather than be inserted mechanically, since overreliance on a small set of sequences produces the repetitive discourse observed in the Iraqi corpus. The aim is conventional phrasing deployed with variety, not formula for its own sake.

7. Further Research

The present findings open several lines of research. A natural extension is to compare Iraqi abstracts with abstracts written by English L1 academics, which would separate the effects of writing in a second language from the effects of the national publishing environment; the present Iraqi–Saudi comparison holds first language constant but not academic context. A second direction is disciplinary: parallel analyses of Iraqi science, medical, and engineering abstracts would show whether the patterns observed here are specific to the Arts and Humanities or general to Iraqi academic writing. Third, the analysis can move beyond the abstract to other sections of the research article, particularly introductions and discussions, where bundle repertoires are known to differ from those of abstracts (Cortes, 2013).

Other designs would deepen the picture. Move-based analyses could link each bundle to the rhetorical move it realises, revealing whether Iraqi writers underuse bundles across all moves or only in specific ones. Longitudinal corpora could track whether bundle use in Iraqi research articles changes over time or after targeted instruction. Finally, the relationship suggested here between formulaic language and citation rates remains a hypothesis; studies correlating bundle density with citation counts at the level of individual articles would test whether phraseological conventionality contributes to the visibility of published research.

References

Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes*, *31*(2), 81–92. https://doi.org/10.1016/j.esp.2011.08.004

Alkhafaji, H., Abdulzahra, A., & Alabedi, A. (2022). A Comparative Study of the Use of Lexical Bundles in Academic Writing by Iraqi and Expert Authors. *The Islamic University College Journal*, *2*, 119–147.

Alsahlanee, A., & Jaganathan, P. (2022). Lexical Complexity in The Writings of Iraqi, English L2, and English L1 Writers. *Asian Journal of University Education*, *18*(4), 1105–1118. https://doi.org/10.24191/ajue.v18i4.20044

Alsahlanee, A., & Jaganathan, P. (2023). Phrasal Syntactic Complexity Measures in Linguistics Research Articles Written by Iraqi and English L1 Writers. *3L The Southeast Asian Journal of English Language Studies*, *29*(1), 154–168. https://doi.org/10.17576/3L-2023-2901-11

Altenberg, B. (1998). On the Phraseology of Spoken English: The Evidence of Recurrent Word-Combinations. In *Phraseology* (pp. 101–123). Oxford University PressOxford. https://doi.org/10.1093/oso/9780198294252.003.0005

Appel, R. (2022). Lexical bundles in L2 English academic texts: Relationships with holistic assessments of writing quality. *System*, *110*, 102899. https://doi.org/10.1016/j.system.2022.102899

Appel, R., Geluso, J., & Feng, H.-H. (2024). An examination of phrase-frames in L2 english academic writing: Exploring relationships with writing quality. *System*, *123*, 103349. https://doi.org/10.1016/j.system.2024.103349

Bao, K., & Liu, M. (2022). A Corpus Study of Lexical Bundles Used Differently in Dissertations Abstracts Produced by Chinese and American PhD Students of Linguistics. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.893773

Bao, K., & Liu, M. (2023). Comparative Analysis of Move-Specific Lexical Bundles in Linguistics Dissertation Abstracts: A Study of Students from China and the United States. *Sage Open*, *13*(4). https://doi.org/10.1177/21582440231210055

Behzadi, P., & Gajdács, M. (2021). Writing a strong scientific paper in medicine and the biomedical sciences: a checklist and recommendations for early career researchers. *Biologia Futura*, *72*(4), 395–407. https://doi.org/10.1007/s42977-021-00095-z

Bhatti, I. A., Mustafa, S., & Azher, M. (2019). Genre Analysis of Research Article Abstracts in Linguistics and Literature: A Cross Disciplinary Study. *International Journal of English Linguistics*, *9*(4), 42. https://doi.org/10.5539/ijel.v9n4p42

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at ...: Lexical Bundles in University Teaching and Textbooks. *Applied Linguistics*, *25*(3), 371–405. https://doi.org/10.1093/applin/25.3.371

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). Longman grammar of spoken and written. *English, London: Longman*.

Chen, Y.-H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology*, *14*.

Cohen, J. (2013). *Statistical Power Analysis for the Behavioral Sciences*. Routledge. https://doi.org/10.4324/9780203771587

Corbett, J. B. (2020). Developing Academic Literacies in Medicine and Agriculture: The Design, Delivery and Evaluation of Blended Learning Support for Research Writing in English. *The ESPecialist*, *41*(5). https://doi.org/10.23925/2318-7115.2020v41i5a3

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, *23*(4), 397–423.

Cortes, V. (2006). Teaching lexical bundles in the disciplines: An example from a writing intensive history class. *Linguistics and Education*, *17*(4), 391–406. https://doi.org/10.1016/j.linged.2007.02.001

Cortes, V. (2013). The purpose of this study is to: Connecting lexical bundles and moves in research article introductions. *Journal of English for Academic Purposes*, *12*(1), 33–43. https://doi.org/10.1016/J.JEAP.2012.11.002

Durrant, P. (2017). Lexical Bundles and Disciplinary Variation in University Students' Writing: Mapping the Territories. *Applied Linguistics*, *38*(2), 165–193. https://doi.org/10.1093/applin/amv011

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text & Talk*, *20*(1), 29–62.

Esfandiari, R., & Barbary, F. (2017). A contrastive corpus-driven study of lexical bundles between English writers and Persian writers in psychology research articles. *Journal of English for Academic Purposes*, *29*, 21–42. https://doi.org/10.1016/j.jeap.2017.09.002

Field, A. (2018). *Discovering Statistics Using IBM SPSS Statistics*. SAGE Publications. https://books.google.iq/books?id=JIrutAEACAAJ

Grabowski, Ł. (2015). Keywords and lexical bundles within English pharmaceutical discourse: A corpus-driven description. *English for Specific Purposes*, *38*, 23–33. https://doi.org/10.1016/j.esp.2014.10.004

Gray, B., Cotos, E., & Smith, J. (2020). *Combining rhetorical move analysis with multi-dimensional analysis* (pp. 137–168). https://doi.org/10.1075/scl.95.06gra

Gries, S. Th. (2021). *Statistics for Linguistics with R*. De Gruyter. https://doi.org/10.1515/9783110718256

Hreiz, Z. J., & Al-Attabi, F. K. T. (2024). A Forensic Linguistic Study of Plagiarism in Iraqi M.A. Theses Writing in English. *Journal of Education College Wasit University*, *56*(2), 519–536. https://doi.org/10.31185/eduj.Vol56.Iss2.3996

Hyland, K. (2008a). Academic clusters: text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, *18*(1), 41–62. https://doi.org/10.1111/j.1473-4192.2008.00178.x

Hyland, K. (2008b). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, *27*(1), 4–21. https://doi.org/10.1016/j.esp.2007.06.001

Hyland, K. (2012). Bundles in Academic Discourse. *Annual Review of Applied Linguistics*, *32*, 150–169. https://doi.org/10.1017/S0267190512000037

Lee, S., Kim, S. J., Park, S., & Shin, J. C. (2025). Unequal metrics in research publications: the impact of bibliometric databases and faculty size across academic disciplines on university rankings in South Korea. *Scientometrics*, *130*(2), 641–663. https://doi.org/10.1007/s11192-024-05223-9

Li, Y., & Lei, H. (2025). Lexical Bundles in L1 and L2 English Academic Writing: Convergent and Divergent Usage. *Sage Open*, *15*(2). https://doi.org/10.1177/21582440251333850

Lorés, R. (2004). On RA abstracts: from rhetorical structure to thematic organisation. *English for Specific Purposes*, *23*(3), 280–302. https://doi.org/10.1016/j.esp.2003.06.001

Lu, X., & Deng, J. (2019). With the rapid development: A contrastive analysis of lexical bundles in dissertation abstracts by Chinese and L1 English doctoral students. *Journal of English for Academic Purposes*, *39*, 21–36. https://doi.org/10.1016/j.jeap.2019.03.008

Ma, R., & Qin, X. (2017). Individual Factors Influencing Citation Competence in L2 Academic Writing. *Journal of Quantitative Linguistics*, *24*(2–3), 213–240. https://doi.org/10.1080/09296174.2016.1265793

Mizumoto, A., Hamatani, S., & Imao, Y. (2017). Applying the Bundle–Move Connection Approach to the Development of an Online Writing Support Tool for Research Articles. *Language Learning*, *67*(4), 885–921. https://doi.org/10.1111/lang.12250

Omidian, T., Shahriari, H., & Siyanova-Chanturia, A. (2018). A cross-disciplinary investigation of multi-word expressions in the moves of research article abstracts. *Journal of English for Academic Purposes*, *36*, 1–14. https://doi.org/10.1016/j.jeap.2018.08.002

Pan, F., Reppen, R., & Biber, D. (2016). Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in Telecommunications research journals. *Journal of English for Academic Purposes*, *21*, 60–71. https://doi.org/10.1016/j.jeap.2015.11.003

Qi, H., & Pan, F. (2020). Lexical bundle variation across moves in abstracts of medical research articles. *Southern African Linguistics and Applied Language Studies*, *38*(2), 109–128. https://doi.org/10.2989/16073614.2020.1763814

Ren, J. (2021). Variability and functions of lexical bundles in research articles of applied linguistics and pharmaceutical sciences. *Journal of English for Academic Purposes*, *50*, 100968. https://doi.org/10.1016/j.jeap.2021.100968

Salager-Meyer, F., Ariza, M. Á. A., & Lewin, B. (2014). *Abstract quality in complementary and alternative medicine papers: A structural and cross-generic analysis* (Vol. 187, pp. 221–240).

Sanosi, A. B. (2022). The Use and Development of Lexical Bundles in Arab EFL Writing: A Corpus-driven Study. *Journal of Language and Education*, *8*(2), 106–121. https://doi.org/10.17323/jle.2022.10826

Shahmoradi, N., Jalali, H., & Ghadiri, M. (2021). Lexical Bundles in the Abstract and Conclusion Sections: The Case of Applied Linguistics and Information Technology. *Applied Research on English Language*, *10*(3), 47–76. https://doi.org/10.22108/are.2021.128024.1703

Shiely, F., Gallagher, K., & Millar, S. R. (2024). How, and why, science and health researchers read scientific (IMRAD) papers. *PLOS ONE*, *19*(1), e0297034. https://doi.org/10.1371/journal.pone.0297034

Shirazizadeh, M., & Amirfazlian, R. (2021). Lexical bundles in theses, articles and textbooks of applied linguistics: Investigating intradisciplinary uniformity and variation. *Journal of English for Academic Purposes*, *49*, 100946. https://doi.org/10.1016/j.jeap.2020.100946

SJR - International Science Ranking. (n.d.). Retrieved August 17, 2025, from https://www.scimagojr.com/countryrank.php?area=1200&region=Middle%20East&order=cd&ord=desc.

Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.

Varghaei, E., & Khodadadi, G. (2022). Comparing Lexical Bundles in Medical Research Article Abstracts of Iranian and Foreign Journals. *GEMA Online® Journal of Language Studies*, *22*(3), 86–102. https://doi.org/10.17576/gema-2022-2203-05

Vasconcelos, S. M. R., Sorenson, M. M., Leta, J., Sant'Ana, M. C., & Batista, P. D. (2008). Researchers' writing competence: a bottleneck in the publication of Latin-American science? *EMBO Reports*, *9*(8), 700–702. https://doi.org/10.1038/embor.2008.143