Unravelling Medical Results: A Corpus-Driven Exploration of Lexical Bundles, Lexical Frames, and Rhetorical Moves

Panupan Panthong[1], & Anchalee Wannaruk[1]

[1] School of Foreign Languages, Institute of Social Technology, Suranaree University of Technology, Thailand

Correspondence: Anchalee Wannaruk, School of Foreign Languages, Institute of Social Technology, Suranaree University of Technology, Thailand.

Received: January 6, 2026 Accepted: March 25, 2026 Online Published: May 11, 2026

doi:10.5430/wjel.v16n4p533 URL: https://doi.org/10.5430/wjel.v16n4p533

Abstract

Publishing in international medical journals poses a significant challenge for non-native English-speaking researchers, particularly in Results sections that conform to a discourse community. However, previous studies have examined lexical bundles (LBs) and lexical frames (LFs) separately and did not explore the connection to rhetorical move structures. This corpus-driven study explored the lexical bundles and frames in results sections of medical research articles, analysed their structural types and mapped them to move structures. Python scripts extracted 5-word spans of LBs and LFs from 3,000 research articles in a four-million-word corpus named RESMED with a frequency threshold of 80 occurrences, and a range criterion of 80% across five high-impact medical journals. Concordance line analysis was used to determine their structural patterns and move structure mapping. Results revealed 67 types of lexical bundles and 225 types of lexical frames predominantly related to the medical trial subjects and statistical data. Structurally, most lexical bundles were noun phrase-based, and a new type *explicative with copula be* emerged from the analysis. Lexical frames frequently occurred with non-verb content words, followed by verb-based frames and frames with function words. Move structure mapping showed that most lexical bundles and frames appeared in *Move 3: Describe statistical test and state findings* and *Move 1: Report the findings*, with fewer occurrences in *Move 2: Review the experiment* and *Move 4: Summary report of the overarching findings.* These findings reveal how medical writers construct results sections through FSs in specific moves and steps, offering pedagogical implications.

Keywords: lexical bundles, lexical frames, corpus-driven analysis, rhetorical analysis, medical research articles

1. Introduction

Medical Research Articles (MRAs) serve as the primary channel for exchanging and announcing medical knowledge from research settings (Iyengar et al., 2022; Waaijer et al., 2019). They play an essential role in improving patient outcomes and advancing medical knowledge through evidence-based practice (Bhuiya & Makaryus, 2023). As MRAs are mostly published in English (Davis, 2020; Salager-Meyer, 2014), non-native speaker medical writers experience considerable hardship in producing linguistically and rhetorically competent texts. These writers face language-related problems such as discourse-related errors, lexical issues, grammatical mistakes, mechanical errors, and unsuitable translation (Gholami & Zeinolabedini, 2015; Luo & Hyland, 2019; Rezaeian, 2015). Among these challenges, a good understanding of the discipline-specific rhetorical convention is important as novice writers struggle to engage with the conventional rhetorical move structures in high-impact journals (Raffing et al., 2021).

One of the most rhetorically significant sections of the MRA is the Results section (Habtu & Deressa, 2025). Presenting the raw data processed through statistical analyses in the form of tables, graphs, and illustrations (Williams, 1999), it serves as the data presentation, explanation, and interpretation that substantiates the research findings (Gao & Pramoolsook, 2022). Given these rhetorical functions, the clear, complete, and coherent results sections are strongly associated with acceptance or rejection of manuscripts by journals (Fazilatfar & Naseri, 2016). That is to say, well-written results sections with clear rhetorical structures tend to succeed in making a convincing case for their contribution to the field. Amnuai & Wannaruk (2012) postulated that the rhetorical move structures vary from one discipline to another, suggesting that discipline-specific linguistic knowledge is indispensable for producing a credible and convincing Results section.

Yet, the knowledge of rhetorical move structures without an understanding of the linguistic features in particular moves or steps could not encourage the novice writers to engage the discourse community (Johns, 2008; Omidian et al., 2018; Simpson-Vlach & Ellis, 2010). For this reason, Formulaic Sequences (FS) play an important role since they function as the building blocks of the discourse (Biber et al., 2004; Gray & Biber, 2013), shape the thoughts of learners, guide them as they navigate the linguistic patterns of discourse, and increase the fluency of their writing (Hu, 2024). According to Gray and Biber (2013), FSs are broadly distinguished into two distinct categories: Lexical Bundles (LB) and Lexical Frames (LF). While LBs such as *"the objective was to", "was included in the"* are fixed multi-word sequences that frequently occurred in texts (Biber et al., 1999), LFs including *"the * of the", "it is * accepted"* are variable patterns an open slot that allows lexical items to vary within a frame structure (Cortes, 2023). These two types of FSs are regarded as the discoursal units of rhetorical moves, shaping the ideas in each move and step (Cortes, 2013). Therefore, it is essential to explore the connection

between FSs and move structures, as this connection constitutes the foundation for novice writers to acquire disciplinary linguistic conventions (Le et al., 2021). Also, it could assist language teachers in developing materials that address the discourse community (Mbodj & Cortes, 2024, 2025).

Although a growing body of literature addresses FSs in medical discourse, few studies systematically map these sequences onto rhetorical move structures, thereby limiting findings that could otherwise inform and enhance disciplinary writing pedagogy. Early works by Mbodj-Diop (2016) and Panthong and Poonpon (2020) examined LBs in MRAs, while Mbodj-Diop and Cortes (2024) further compared different types of multiword collocations, including LBs and LFs, among other types within MRAs. Subsequently, attention shifted to the connection between LBs and rhetorical move structures only within different sections of MRAs, including abstracts (Asano et al., 2024), methods (C. Liu & Pan, 2024) and the discussion section (Al-Shujairi et al., 2020). However, two gaps exist in previous works. First, these studies only examined LBs while ignoring LFs. Biber (2012) suggested that researching only one type of FSs might insufficiently capture or represent the language in a particular discourse. Second, relatively few studies have explored how both LBs and LFs function within the rhetorical moves of the Results section to supplement medical discourse. To address these gaps, this present study aims to investigate five-word LBs and LFs across five MRA journals and map them onto rhetorical move structures. The expected results could be maximally useful for medical writers ranking from novice to medical staff, language or content teachers in English for Medical Purposes (EMP) course, material designers, and knowledge gap in corpus linguistics.

1.1 Research Questions

To bridge these gaps, this research addresses the medical discourse community's needs and proposes the following research questions:

1) What lexical bundles and lexical frames emerge in the results sections of medical research articles?

2) What structural types characterise the lexical bundles and lexical frames in medical research article results sections?

3) How do lexical bundles and lexical frames function as linguistic realisations of rhetorical moves and steps in the results sections of medical research articles?

2. Literature Review

2.1 Formulaic Sequences in Academic Writing

Formulaic sequences (FS) are word combinations that act as structural elements within discourse (Schneider, 2023; Wulff, 2018). Stored in long-term memory, they facilitate fluency in linguistic production (Hu, 2024; Yu, 2022) and play a pivotal role in discourse functions (Biber et al., 1999; Biber et al., 2004). In terms of language acquisition, incorporating them as the linguistic resources in the classroom could increase fluency and natural language use for learners (Uçar & Zarfsaz, 2023). While Wray (2013) acknowledged that FSs include a wide range of types such as lexical bundles, lexical frames, collocations, idiomatic expression, and n-grams, Gray and Biber (2013) adopted a broader categorisation between two primary types: Lexical Bundles (LBs) and Lexical Frames (LFs).

2.1.1 Lexical Bundles

Lexical Bundles (LBs) are fixed phrases serving as grammatical structures within a given discourse. They vary from genre to genre and discipline to discipline (Biber et al., 1999). For example, in conversation, LBs include phrases such as *"I don't know if,"* whereas in academic writing, common LBs are *"as shown in Figure"* and *"the fact that the"* (Biber & Barbieri, 2007). Because of these variations, earlier works (Biber et al., 1999; Biber et al., 2004; Biber & Barbieri, 2007) also investigated the structural types of LBs further, proposing a conceptual framework including verb-phrase fragments, prepositional phrase and noun phrase fragments, and dependent clause fragments. Yet, these structural types are seen as too general to identify LBs in the specific genre because they were invented to cover both spoken and written registers. Therefore, Lu and Deng (2019) revised and introduced more academically oriented structural types to use with the written academic discourse. They revised the previous versions of early conceptual framework to cover the academic LBs.

2.1.2 Lexical Frames

In contrast, Lexical Frames (LFs) are discontinuous phrases with a slot that can be filled with lexical items to co-construct semantics in context (Gray & Biber, 2013). Corpus experts (Biber, 2012; Cortes, 2023, 2024) argued that only exploring LBs may overlook the linguistic pattern and a lack of accuracy. LFs also exhibited variation across genres, such as *'I * it was'* to in conversation, which can be filled with *'see'* or *'think'*, and *'it is * that'* in academic writing which can be added with words such as *'possible' 'likely' 'acknowledged'* (Gray & Biber, 2013). LFs with slots exhibit different patterns, such as, i.e., 1*34, 12*4, *234, 123*, and the absence allows word fillers to construct the meaning of phrases. Fuster-Márquez (2022) postulated that studying both types of FSs could supplement and advance our understanding the specific linguistic patterns of a given discourse community.

2.2 Rhetorical Move Structures of the Result section

One of the main sections of a research article is the Results section (Nwogu, 1997). According to Williams (1999), this section presents the data from the Methods section in the form of tables, graphs, and visual aids. Rhetorical move structure analyses have continuously sharpened our insights into the organisation of this section. Williams (1999) adopted the classic framework to explore the Results section of MRAs, finding unique discipline-specific characteristics. Later, Nwogu (1997) analysed the results sections of 30 medical research articles and developed conceptual frameworks that identified move structures unique to the medical discipline. Recently,

Davis (2020) revised Nwogu's work, and he included more recent research papers and a larger corpus. This, as a result, provides more up-to-date data on move structures in MRAs. Despite the advanced knowledge of rhetorical moves, Johns (2008) emphasised that rhetorical moves without linguistic features are too broad for novice learners to write successfully, so she suggested integrating moves with such features to benefit learners maximally.

2.3 The Connection between Formulaic Sequences and Rhetorical Move Structures

Focusing only on the discourse functions of FSs is not sufficient (Simpson-Vlach & Ellis, 2010). Cortes (2013) argued that FSs should be bridged to rhetorical moves because they serve as the fundamental discourse units of specific communicative purposes. Extensive research also showed that FSs are prevalent and essential for rhetorical move structures (Casal & Yoon, 2023; Cortes, 2023; Liu et al., 2023). Pedagogically, this connection could guide novice writers in mastering linguistic patterns (Lu et al., 2021) and facilitate language instructors in designing materials and instruction that match the learners' discourse community (Mbodj & Cortes, 2024).

Methodologically, two traditions explore these connections: the top-down approach, which analyses rhetorical move structures to identify associated linguistic features, and the bottom-up approach, which explores linguistic patterns and maps them onto rhetorical moves (Biber et al., 2007; Cortes, 2023). The top-down approach has been criticised for its subjectivity, small corpus size, and labour intensity (Cortes, 2013; Li et al., 2020; Lu et al., 2021; Omidian et al., 2018). In contrast, the bottom-up approach produces less biased, more generalisable results by relying on high-frequency patterns from large corpora (Li et al., 2020), and it can reveal specialised unexplored linguistic patterns (Biber, 2012; Cortes, 2023).

Both LBs and LFs have been mapped to rhetorical moves across disciplines, but studies remain limited in medical discourse, particularly in the results section. Studies mapping LBs to rhetorical moves have been conducted across various sections including introductions (Cortes, 2013), discussions (Le & Harrington, 2015), and abstracts (Li et al., 2020). LF mapping, however, remains limited and most conducted in top-down approach namely, introduction sections (Lu et al., 2021; Chanh & Pramoolsook, 2025), figure explanations (Liu et al., 2023), and discussion sections (Chen & Zhang, 2025). In medical discourse specifically, bottom-up studies have examined abstracts (Asano et al., 2024), discussions (Al-Shujairi et al., 2020), and methods (Liu & Pan, 2024), with each IMRD section showing distinct linguistic patterns (Liu & Pan, 2024b). To date, few studies have simultaneously mapped both LBs and LFs to rhetorical move structures in the results sections of MRAs by a bottom-up approach. This study, therefore, occupied these gaps to explore to respond the call for such research in the medical discourse (Panthong & Poonpon, 2025; Mbodj & Cortes, 2024, 2025).

3. Method

3.1 Corpus Compilation

This study introduces RESMED, which contains 3,000 Result sections from five leading medical journals (2018-2022), including The Lancet, New England Journal of Medicine (NEJM), Journal of the American Medical Association (JAMA), British Medical Journal (BMJ), and Journal of Clinical Investigation (JCI). The corpus comprises 600 articles per journal with a total of 4,034,851 tokens. All sections were manually extracted, standardised to plain text format, and pre-processed to remove metadata and tables for analysis. The compilation is guided by three main criteria: (1) the selection of reputable and accessible journals, especially those ranked among the highest impact factor journals in the category of Medicine, General and Internal according to Journal Citation Reports (JCR), indexed in Web of Science, and representative of general medicine discourse (Nwogu, 1997), (2) a five-year timespan to capture language variability (Guinda, 2015), and (3) the compilation of a large corpus size to provide sufficient data for pedagogical purposes (Meyer, 2023).

3.2 Data Analysis

This study adopted the bottom-up approach as mentioned in the previous section to guide the data analysis. Therefore, the analysis includes three stages including identifying formulaic sequence, characterising structural types and mapping them to rhetorical moves.

3.2.1 The Identification of Formulaic Sequences

The first stage involved four steps to identify, extract, classify, and filter five-word LBs and LFs from the result sections. First, a Python script (version 3.13.2) was coded to merge results sections from MRAs in each journal into one text file. After that, the Python script was written to extract two separated lists: the first one contains raw lexical bundles, and the second one comprises raw lexical frames. This approach was adopted to address misrepresentation errors produced by corpus software, for example, embedding "*a total of of patients*" within longer sequences like "*a total of 128 of 212 patients*" (Asano et al., 2024; Cortes & Lake, 2023). Second, five-word bundles and frames were selected because they tend to be more discipline-specific (Cunningham, 2017; Golparvar & Barabadi, 2020) and highly associated with rhetorical move structures (Cortes, 2013). Three- and four-word bundles and frames are too general, and six-word bundles are relatively rare in the corpus (Gray & Biber, 2013). In addition, each five-word frame contained an internal slot such as "*in the * group the*" which commonly occurred in academic texts (Gray & Biber, 2013). The frequency threshold was set at 80 times (normalised from Hyland's 20 times per 1 million words). The range required the occurrence of four out of five journals or 80% to avoid the overrepresentation of specific journals (Cunningham, 2017). Third, Type Token Ratio (TTR) and Entropy (Ent) scores of LFs were automatically calculated by the Python script during the extraction of LFs in the first stage. According to Gray and Biber (2013), TTR scores measured variability of frames while Ent scores measured the predictability. Based on Gray and Biber's range, LFs in this study were classified by TTR as highly variable (TTR > 0.70), variable ($0.30 \leq$ TTR ≤ 0.70), or relatively fixed (TTR < 0.30), and by H~norm~ as highly predictable (H~norm~ < 0.30), moderately predictable ($0.30 \leq$ H~norm~ ≤ 0.70), or lowly predictable (H~norm~ > 0.70).

Finally, bundles and frames tied to socio-cultural contexts, e.g., "*the hospital in Thailand*," (Panthong & Poonpon, 2020) or spanning sentence boundaries, e.g., *"in Table below. The,"* (Mbodj & Cortes, 2024) were excluded. Still, numerical bundles and frames such as "*shown in Table 2*" were retained for pedagogical value.

3.2.2 The Analysis of Structural Types of Formulaic Sequences

The second stage was to analyse the structural types of LBs and LFs, involving two steps. The lists of LBs and LFs identified in the previous section were searched using concordance lines in the AntConc software (Anthony, 2024). Two conceptual frameworks guided each analysis.

For the structural analysis of LBs, this study adopted Lu & Deng (2019)'s framework, which revised the classic taxonomies of Biber et al. (1999) and Biber et al. (2004). Lu and Deng introduced five structural types, including noun phrase-based, prepositional phrase-based, verb phrase-based, clause-based and conjunction. As they claimed, this framework was specifically designed for academic genres.

For LFs, this study adopted a taxonomy of structural types proposed by Gray and Biber (2013) that includes three broad categories: verb-based frames, frames with non-verb content words, frames with function words.

3.2.3 Mapping Formulaic Sequences to Moves

The last stage was to map the LBs and LFs to the rhetorical move structures in the results sections. This present work adopted a corpus-driven approach. To map both lexical bundles and frames to the rhetorical move structure, this study used the conceptual framework proposed by Davis (2020). As discussed in the literature review, this framework provides an up-to-date move structure specifically developed for the results sections of MRAs, and it consisted of 4 moves with eight steps in the results sections (see Figure 1). The process begins with manual analysis and coding of salient rhetorical functions within the extended co-text concordance lines, based on the suggestions by Omidian et al. (2018) and Le and Harrington (2015). That is, each type of LBs and LFs was used to search in the concordance line functions in Antconc, and each concordance line was coded at move and step based on the examination of the surrounding contexts. To ensure reliability, 30% of the texts were given to a medical doctor who was trained as an inter-rater to code the; as a result, the agreement for this analysis reached 80.52%. This level of agreement is considered acceptable for discourse analysis research (Kanoksilapatham, 2005).

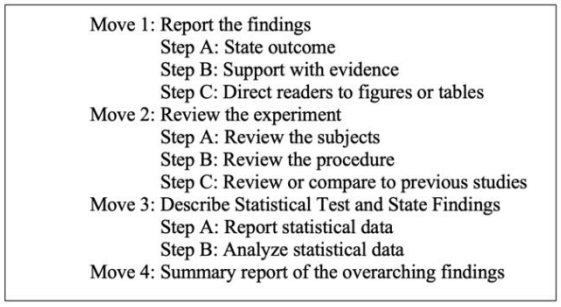

Figure 1. Rhetorical move structures in the Result section of medical research articles (Davis, 2020)

4. Results

4.1 Lexical Bundles in the Result sections of Medical Research Articles

4.1.1 Lexical Bundles Identified and Their Frequency

In response to research question 1, the analysis found 67 types over 80 times with a minimum of four out of five journals, and Table 1 showed the top twenty of lexical bundles in the Results section of RESMED corpus. These LBs mostly characterise two themes: trial participants and the statistical comparison, reflecting the nature and communicative purposes of the results section. First, trial-participant-related LBs such as *"patients in the control group"* and *"in the usual care group"* were frequently used to identify different patient or participant groups. In addition, LBs with the statistical comparison for example, *"there was no significant difference"* and *"did not differ significantly between"* were often written to express statistical comparison fundamental to the randomised control trial.

Table 1. Top twenty lexical bundles in the Result section of RESMED corpus

No.	Lexical Bundles	Frequency/Range	No.	Lexical Bundles	Frequency/Range
1	of the patients in the	376/5	11	were lost to follow up	183/4
2	in the usual care group	337/4	12	patients in the control group	171/5
3	in the intention to treat	305/4	13	participants in the placebo group	165/4
4	there was no significant difference	292/5	14	at the end of the	155/5
5	group than in the placebo	249/4	15	were more likely to be	151/5
6	in the intervention group and	242/4	16	there was no evidence of	141/5
7	there were no significant differences	240/5	17	in the placebo group had	135/4
8	patients in the placebo group	232/4	18	the modified intention to treat	135/4
9	was no significant difference in	205/5	19	in the per protocol population	133/4
10	did not differ significantly between	204/5	20	there was no statistically significant	126/5

4.1.2 Structural Types of Lexical Bundles

Based on research question 2, Table 2 demonstrates that noun-based bundles are the most frequent type (32.84%), followed by VP-based bundles (28.36%), and clause-based bundles (13.24%). The prevalence of noun phrases indicates medical authors' preference for explaining their findings about research participants, procedural elements, and clinical outcomes. However, no LBs connected with conjunctions were found, so it suggests that such bundles are not conventionalised in this genre. It is likely that cohesion in medical results sections is expressed through referential and statistical language.

Table 2. Structural types of lexical bundles in the Result section of RESMED corpus

Categories	**Number**	**Percentage**	**Examples**
1.NP-based	**22**	**32.84**	
1.1 NP with of-phrase fragments	9	13.43	the baseline characteristics of the, a median follow up of
1.2 NP with other post-modifier fragments	9	13.43	participants in the placebo group, patients in the intervention group
1.3 Other NP fragments	4	5.978	no significant between group differences, the most common adverse events
2.PP-based fragments	**18**	**26.87**	
2.1 PP with embedded *of*-phrase	7	10.44	of the patients in the, with a lower risk of
2.2 Other PP fragments	11	16.41	in the usual care group, in the placebo group had
3.VP-based	**19**	**28.36**	
3.1 Copula be + NP/adjective phrase	8	11.94	were more likely to be, were more likely to have
3.2 VP with active verb	2	2.98	did not differ between the, did not differ significantly between
3.3 VP with infinitive verb	1	1.49	to be related to the
3.4 VP with passive verb	5	7.46	were lost to follow up, are shown in table 1
3.5 LBs beginning with past participle	3	4.47	compared with the control group, associated with an increased risk
4.Clause-based	**8**	**11.94**	
4.1 PP + copula be	0	0	
4.2 NP + copula be	0	0	
4.3 Anticipatory it + copula be + adjective phrase	0	0	
4.4 Anticipatory it + passive verb + that	0	0	
4.5 NP/complementizer + passive verb	3	4.47	patients were randomly assigned to, serious adverse events were reported
4.6 NP + active verb	0	0	
4.7 NP + active verb + that	0	0	
4.8 Explicative + copula be	5	7.46	there was no significant difference, there was no significant difference
5. Conjunctions	**0**	**0**	
Total	67	100	

It is worth noting that the analysis identified a structural pattern which was not explicitly seen in Lu and Deng's (2019) taxonomy: explicative constructions with copula be (e.g., *"there was no significant difference", "there were no statistically significant"*), accounting for five types within the clause-based category. As illustrated in Figure 2, this structure serves crucial discourse functions in reporting non-significant statistical findings.

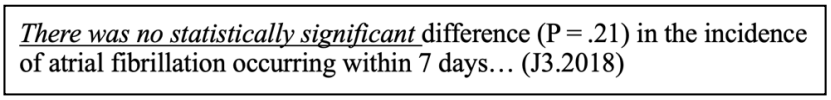

Figure 2. Explicative constructions with copula be

4.1.3 Lexical Bundles in Moves and Steps

To address research question 3, the distribution of LBs across rhetorical moves shows distinct patterns (Table 3). Move 3 contains the highest proportion of LBs (50%), followed by Move 1 (32.65%), Move 2 (16.32%), and Move 4 (1.02%). The number of bundle-moves (98 types) is higher than that of LBs because some LBs are multifunctional across moves and steps. The distribution reflects the key communicative purposes of the result section as this section aims to present the results from the method and the statistical analysis. However, the lower proportion of Move 2 and Move 4 suggests that these moves are less formulaically conventionalised because they are optional and occur less frequently in medical results sections (Davis, 2020).

Table 3. The distribution of Lexical Bundles in each move and step

Moves & Steps	**Number of LBs**	**Percentage**
Move 1: Report the findings	32	32.65
Step A: State outcome	27	27.55
Step B: Support with evidence	2	2.04
Step C: Direct Readers to figure or table	3	3.06
Move 2: Review the experiment	16	16.32
Step A: Review the subjects	6	6.12
Step B: Review the procedure	10	10.20
Step C: Review and compare to previous studies	0	0
Move 3: Describe statistical test and state findings	49	50
Step A: Report statistical data	34	34.69
Step B: Analyse statistical data	15	15.30
Move 4: Summary report of the overarching findings	1	1.02
Total	98	100

Figure 3 illustrates how LBs operate functionally across the rhetorical moves and steps of the Results section, with certain bundles serving multiple discourse purposes.

Move 1 distributes LBs across three steps. In Step A: State outcomes, bundles such as “patients in the placebo group” communicate treatment group distinctions when the authors reported primary and secondary outcomes. Step B: Support with evidence employs bundles such as “health-related quality of life” in patient-reported measures although this step remains comparatively less formulaic than others. Completing this move, Step C: Direct readers to figures or tables employs navigational bundles such as “are shown in table 2” to link textual claims to visual evidence.

Move 2 shifts the rhetorical focus toward methodological review. In Step A: Review the subjects “patients were included in the” reestablished the participant inclusion of the study. Step B: Review the procedure reinforces containing lexical bundles such as “patients were randomly assigned” and “patients were assigned to”. Notably, no LBs were identified in Step C: Review and compare to previous studies, so it points to the absence of conventionalised cross-study comparison at the lexical bundle level in this corpus.

Accounting for the highest proportion of LBs overall, Move 3 reflects the quantitative core of results reporting. In Step A: Report statistical data, bundles such as “there was no statistically significant” deliver outcomes with analytical directness. Step B: Analyse statistical data then extends this with interpretive bundles such as “associated with an increased risk” to bridge raw data and its clinical significance.

By contrast, Move 4 (Summary report of the overarching findings) is the least lexically dense since it was represented by a single bundle type. To close the section, “there was no evidence of” frames an evaluative statement and signals both temporal and rhetorical closure.

Beyond move-specific functions, certain LBs exhibit multifunctionality. The bundle “health-related quality of life”, for instance, recurs across Move 1 Step B and the multifunction category. This bundle demonstrates that some bundles transcend individual rhetorical steps.

Move 1: Report the findings

Step A: State outcome

Depth of response improved during maintenance therapy in 139 (46%) of 302 patients in the ixazomib group and 60 (32%) of 187 *patients in the placebo group* with very good partial response or partial response post trans plantation … (J1.2019)

Step B: Support with evidence

Patients reported an improved *health-related quality of life a*s indicated by improvements in SF-36 physical component score for upadacitinib 15 mg … (J1.2019)

Step C: Direct Readers to figure or table

Additional echocardiographic findings *are shown in Table 2* and eTable 8 in Supplement 1. (J3.2019)

Move 2: Review the experiment

Step A: Review the subjects

A total of 210 *patients were included in the* first randomisation; 71 were assigned to exercise therapy and 139 to surgery. (J4.20218)

Step B: Review the procedure

Thus, 336 *patients were randomly assigned* to the ticagrelor/aspirin group and 339 *patients were assigned to* the clopidogrel/aspirin group (table 1). (J3.2019)

Move 3: Describe statistical test and state findings

Step A: Report statistical data

When analyzed separately, *there was no statistically significant* association between PCI and death at 5 years (15.0% vs 13.9%; HR, 0.94 [95% CI, 0.68-1.29]). (J3.2020)

Step B: Analyse statistical data

Adjusted for age, sex, and comorbidity, the following factors were *associated with an increased risk* of mortality post diagnosis

Move 4: Summary report of the overarching findings

At 16 weeks after infection, *there was no evidence of* active intralymphatic adult parasitism or circulating microfilariae, indicating that lymphatic pathology persists long after initial induction by filarial infection. (J5.2021)

Multifunction Lexical Bundles across Moves

On the *health-related quality of life* total score portion of the questionnaire, scores range from 0 to 100, with higher scores indicating a better quality of life. (J2.2020)

Figure 3. Examples of Lexical Bundles in Each Move and Step

4.2 Lexical Frames in Result Section of Medical Research Articles

4.2.1 Lexical Frames Identified and Their Frequency

To answer research question 1, a total of 225 types of LFs were identified, which is substantially higher than that of LBs (67 types). Table 4 presents the top 20 most frequent LFs with frequency, range, and filler examples. For example, the most common LFs "*in the * group and"* (4,302 times across 5 journals) accommodated discipline-specific fillers such as intervention, placebo, olaparib and pembrolizumab. The presence of cancer treatment drugs as LF fillers, including pembrolizumab, daratumumab and olaparib in the RESMED corpus, demonstrates greater flexibility and discipline specificity than with LBs.

Moreover, LFs in this study accommodate both lexical and numerical fillers. For example, *"of * patients in the"* was filled with lexical items (the, treated, female) and numerical items (nine, seven, 355). This hybrid function of LFs reflects the quantitative nature of the results reporting of MRAs as the medical authors integrate both lexical and numerical data in certain linguistic phrases.

Further analysis of LF patterns found ABC*E (85 tokens, 37.8%), AB*CD (83 tokens, 36.9%), and A*CDE (57 tokens, 25.3%). That is, medical writers use the fourth and the third slots to report various aspects of their research. TTR scores indicated low predictability (110 tokens, 48.9%), moderate predictability (84 tokens, 37.3%), and high predictability (31 tokens, 13.8%). Entropy scores showed relatively fixed frames (150 tokens, 66.7%), variable frames (64 tokens, 28.4%), and highly variable frames (11 tokens, 4.9%). These two statistical measures reveal that many different words can fill the frame slots (low predictability in TTR), but medical writers conventionally rely on a small set of frequently repeated discipline-specific terms (fixed frames in entropy). This is attributed to terminological precision in the clinical report of medical discourse.

Table 4. Top twenty lexical frames in RESMED corpus

No.	Lexical Frames	Frequency/Range	Examples of fillers
1	in the * group and	4302/5	intervention, placebo, vaccine, pembrolizumab, daratumumab, control
2	patients in the * group	1949/5	placebo, control, intervention, pembrolizumab, cannabidiol
3	of * patients in the	1312/5	the, treated, all, nine, female, seven, 355, including
4	in the * group than	845/5	intervention, placebo, remdesivir, daratumumab, closed-loop
5	a total of * patients	765/5	10, 14, 13
6	of the * in the	633/5	patients, participants, women, time, events, study
7	in the * group had	611/5	placebo, control, intervention, sunitinib, chemotherapy
8	participants in the * group	605/5	placebo, intervention, control, vaccine, semaglutide
9	the * group and in	586/4	vedolizumab, intervention, ibrutinib, rivaroxaban, 6-week
10	the * of patients with	453/5	percentage, proportion, number, percentages, subgroup, proportions
11	the * of patients who	396/4	percentage, proportion, number, percentages, subgroup, proportions, numbers, group
12	in the * of the	389/5	analysis, presence, context, expression, rest, case, effect, absence, regulation, majority, distribution, magnitude
13	at the * of the	372/5	end, time, start, beginning, discretion, recommendation, base, request, level
14	in the * group were	358/5	placebo, control, intervention, sunitinib, chemotherapy, surgery
15	of * participants in the	341/5	the, 306, 3421, all,
16	there was no * difference	330/5	significant, between-group, statistical, meaningful, observed, notable
17	was significantly * in the	307/5	higher, lower, longer, greater, reduced, increased, decreased
18	there were no * differences	306/5	significant, between-group, substantial, notable
19	no significant * in the	287/5	difference, differences, change, increase, changes, heterogeneity
20	compared with * in the	286/5	those, patients, that, participants, levels, none, women, physicians

4.2.2 Structural Types of Lexical Frames

With regard to research question 2, across 225 LF types, the analysis identified three different structural types (see Figure 4 for examples). The majority of frames (51.55%) contained non-verb content words, usually embedding nouns associated with measurement, quantification, or patient characteristics. In 44.88%, verb-based frames were used, primarily to convey reporting verbs or research actions that establish an evidential stance. Only 3.5% of instances included function word frames, indicating that lexical rather than grammatical patterns are the main means by which medical discourse creates meaning. This distribution shows that explicit reporting of procedures and concrete referential content are given precedence over abstract relational expressions in research articles.

1) Frames with Non-Verb content words
the baseline *[characteristics, features, demographics, distribution]* of the
the *[percentage, proportion, number, subgroup, majority]* of patients with
the effect of *[aspirin, edoxaban, parkin, education, dapagliflozin]* on
a significantly higher *[proportion, risk, rate, percentage]* of
2) Verb-based frames
results were *[observed, obtained, seen, found]* in the
were not *[included, enrolled, present, involved, observed]* in the
are *[shown, presented, summarized, provided, reported, listed]* in table 2
these *[data, results, findings, studies, outcomes]* suggest that the
3) Frames with function words
of the *[components, results, characteristics, analysis, effects]* of the
with the *[results, use, exception, findings, addition]* of the
in the *[analysis, presence, supplement, expression]* of the
for the *[comparison, analysis, duration, association, purpose]* of the

Figure 4. Examples of structural types of Lexical Frames

4.2.3 Lexical Frames in Moves and Steps

Addressing research question 3, the distribution of LFs across rhetorical moves reveals functional specialisation (Table 5). Move 3 contained the highest concentration (40.37%), followed by Move 1 (38.20%), Move 2 (20.8%), and Move 4 (0.62%). The number of frame-moves (322) is higher than the 225 LF types because certain frames are multifunctional across moves and steps. The higher proportion of LFs in Move 2 and Move 4 compared to LBs reflects the greater flexibility of LFs, thereby allowing them to appear across a wider range of moves including optional ones. Despite this difference, both LBs and LFs were predominantly found in Move 1 and Move 3, which constitute the rhetorical core of the Results section in medical research articles.

Table 5. Distribution of lexical frames across moves and steps

Moves & Steps	Number of LFs	percentage
Move 1: Report the findings	**123**	**38.20**
Step A: State outcome	86	26.7
Step B: Support with evidence	11	3.4
Step C: Direct Readers to figure or table	26	8
Move 2: Review the experiment	**67**	**20.8**
Step A: Review the subjects	33	10.24
Step B: Review the procedure	28	8.7
Step C: Review and compare to previous studies	6	1.89
Move 3: Describe statistical test and state findings	**130**	**40.37**
Step A: Report statistical data	87	27.01
Step B: Analyse statistical data	43	13.35
Move 4: Summary report of the overarching findings	**2**	**0.62**
Total	**322**	**100**

Move 1: Report the findings

Step A: State outcome

The current report presents *the results of the [primary] analysis* involving 1103 patients with severe confirmed Covid-19 …. (J2.2021)

Step B: Support with evidence

These results [suggest] that the incrementally better patient outcomes recorded in wealthier countries were partly explained by patient casemix. (J1.2018)

Step C: Direct Readers to figure or table

As *shown in Figure [3A]*, the liraglutide group had few cases of any cardiovascular event over the first year ….. (J2.2022)

Move 2: Review the experiment

Step A: Review the subjects

The *baseline [demographic] and clinical characteristics* of all the Sisonke vaccine recipients and of unvaccinated individuals included …. (J1.2022)

Step B: Review the procedure

To investigate the [potential] of convalescent plasma therapy, we conducted a nonrandomized, open-label, phase I clinical trial, "Convalescent Plasma Therapy – Zurich Protocol (CPT-ZHP) (J5.2022)

Step C: Review and compare to previous studies

Among children of migrants in China (all of whom were internal migrants) risks for all health outcomes were *consistent with the [results] of* the main analyses. (J1.2018)

Move 3: Describe statistical test and state findings

Step A: Report statistical data

….the reduction in HAM-D total scores from baseline at day 30 *was significantly [higher] in the* BRX60 group (p=0·0044) and BRX90 group (p=0·0481). (J1.2018)

Step B: Analyse statistical data

These data [demonstrate] that the film has a protective role, slowing or preventing the proliferation of bacteria and reducing the movement of bacteria into wounds and skin in the first 12 hours after injury. 3(J5.2018)

Move 4: Summary report of the overarching findings

At the [end] of the study period, 758,118 participants (90%) had received the booster. (J2.2021)

Multifunction Lexical Frames across Moves

Among 781 patients eligible *for the [analysis] of the* primary end point, all-cause death or adjudicated AHF rehospitalization through day 180 occurred in 117 patients (30.6%) (J3.2019)

…. the voclosporin group that returned to near baseline levels by week 4 and remained stable *for the [duration] of the* study (appendix p 17). (J1.2021)

After adjustment for baseline risk factors, the hazard ratio *for the [comparison] of the* surgery group with the control group was 0.70 (95% CI, 0.61 to 0.81; P<0.001) …(J2.2020)

Figure 5. Examples of lexical frames in each move and step

This section describes the LFs identified in these moves and steps and also discusses how some LFs are multifunctional across different moves (see Figure 5).

Move 1 presents the findings objectively, falling into three different steps (Davis, 2020). In *Step A State Outcome*, the LF "*the results of the [primary, secondary, main] analysis"* signals the relative importance of different findings based on the sequences of research. In *Step B: supporting with evidence*, the medical authors use phrases such as "*these results [suggest, indicate, demonstrate] that the"* to interpret their results. In *Step C: Direct readers to figures or tables*, "*shown in figure [the figure number] the"* invites the reader to see the visual information.

Move 2 provides methodological context by reminding the readers in order to remove ambiguity about the experiments. In *Step A: Review the subjects*, medical authors use the LF "*baseline [demographic, patients] and clinical characteristics"* to describe demographic reference to a comparability between study groups. In *Step B: Review the procedure*, the medical writers rely on the phrase "*to investigate the [effect, potential, relationship, significance, effectiveness] of* "to reexplain the method in the result section. In *Step C: Review or compare to previous studies*, only six frames were found in this step. For example, "*consistent with the (results, findings, analysis, absence, effect)"* was used when researchers compared their findings with previous research.

Move 3 is considered the cornerstone of the results section of research articles. In *Step A: Report statistical data*, the LF "*was significantly [higher, lower, different] in the"* was used to report the data. In *Step B: Analyse statistical data*, after the statistical presentation, researchers also interpreted their statistical findings. The LF "*these data [show, demonstrate, indicate, reveal] that" the* appeared when researchers interpreted their findings and explained the implications.

For Move 4, this optional move signals closure and synthesis of the information. The LF "*at the [end] of the"* was used to summarize the study.

Additionally, analysis revealed the multiple functions of LFs. The LF "*for the * of the"* which is related to function words, can function in the following moves: Move 1Step A, Move 2 Step B, and Move 3 Step B.

5. Discussion

5.1 Variations of FSs in Disciplines and Sections

The results found that both LBs and LFs in the results sections of MRAs are specific and distinct from other sections of the IMRD structure. This section discusses three main points.

First, LBs and LFs are embedded with medical terms such as patient, placebo, follow, pharmaceutical compounds, and medical intervention. This suggests that both LBs and LFs are discipline-specific for medical discourse. This finding confirms previous studies, which agreed that FSs are more discipline-specific than general academic English (Asano et al., 2024; Mbodj & Cortes, 2024; Panthong & Poonpon, 2020). It can be tentatively suggested that the linguistic patterns of each research article section should be explored for discipline-specific knowledge and coinciding pedagogical implications.

Second, both LBs and LFs demonstrate particular linguistic patterns that shape and facilitate the presentation of statistical findings and demographics, as evidenced by LBs such as there was no significant difference and patients in the control group. This could be explained by the fact that the results section has its own communicative purposes (Davis, 2020; Nwogu, 1997). The sectional variation observed aligns with Shahriari (2017), who found that LBs vary across different sections of research articles. This study further reflects the existence of linguistic variation among genres.

Third, the inclusion of numerical and lexical items in both LBs and LFs challenges traditional conceptualisations. Biber et al. (1999) considered LBs to be a string of words, and Gray and Biber characterised LFs as structures filled with lexical items. In the medical discipline, Mbodj and Cortes (2024) eliminated LFs filled with numbers in MRAs because these frames did not match characteristics. However, this study found numerical elements to be integral components of formulaic patterns in results sections, where presentation of numerical data is fundamental (Brett, 1994; Williams, 1999). The inclusion of numbers serves as bridge between linguistic structure and statistical content. Therefore, these LFs have potential pedagogical usefulness, despite the different numerical reports.

5.2 Differential Flexibility

This study found 67 lexical bundles and 225 lexical frames (a ratio of nearly 1:4). This difference exists because of their characteristics. LFs allow word variations, so they reach frequency thresholds more easily; however, LBs must appear in exact form to be counted (Cortes, 2023; Gray & Biber, 2013). This confirms earlier research (Biber, 2012; Cortes, 2023; Cunningham, 2017; He et al., 2021), which found that studying both types provides a fuller picture of language use. Biber (2012) warned that studying only lexical bundles misses important patterns; this study confirms that concern—focusing only on lexical bundles would miss about 78% of formulaic patterns.

5.3 Structural Types

The results of structural types introduce the following discussions: noun phrase, explicative construction and structural types of LFs in medical discourse.

The most prevalent type belongs to noun phrase-based bundles, while conjunction-based bundles were absent. In the result sections of MRAs as an academic discourse, noun phrase-based bundles function to identify participants, interventions, and outcomes in medical

experiments. Previous research confirms that noun structures form the foundation of written academic discourse (Biber et al., 1999, 2004; Biber & Barbieri, 2007).

Verb phrase-based bundles emerged as the second most frequent type. Most used copula be constructions and passive voice (e.g., were lost to follow up, are shown in Table 1). This finding agrees with previous studies showing that passive voice is common in academic writing (Biber et al., 1999, 2004; Biber & Barbieri, 2007) and medical discourse (Al-Shujairi et al., 2020; Asano et al., 2024; Mbodj-Diop, 2016; Panthong & Poonpon, 2020). Passive voice emphasizes research findings instead of researchers, creating objectivity across all medical research articles.

The identification of explicative constructions as an additional structural pattern extends Lu and Deng's (2019) taxonomy. This suggests that discipline-specific modifications are necessary to capture formulaic expressions in medical contexts (Panthong & Poonpon, 2020; Mbodj-Diop, 2016). This pattern shows that medical writing has unique features for presenting statistics, and general taxonomies may miss these special patterns.

Lexical frames required more detailed structural categories than Gray and Biber's (2013) taxonomy. Our analysis found subtypes similar to LB structures, such as noun phrases with of-phrase fragments and post-modifiers. Previous studies support dividing structural taxonomies into subtypes (Cunningham, 2017; Golparvar & Barabadi, 2020; Mbodj & Cortes, 2024). These subcategories show that medical writing uses varied formulaic expressions, which can help develop teaching materials and analysis tools for medical writing.

5.4 Rhetorical Distribution Patterns

The mapping between formulaic sequences and rhetorical moves reveals concentration in Move 3: Describe statistical test and state findings, with 50% for LBs and 40.37% for LFs, followed by Move 1 (Report findings), with 32.65% for LBs and 38.20% for LFs. This distribution pattern demonstrates that medical writers frequently employ formulaic sequences to present statistical analyses and report primary outcomes, which constitute the core communicative purposes of results sections. The high frequency in these moves can be explained by the nature of medical reporting, which demands precise and conventional language to convey statistical findings (Davis, 2020; Nwogu, 1997). This finding shares similarities with previous research (Cortes, 2004; Panthong & Poonpon, 2020; Mbodj-Diop, 2016) showing that scientific research articles, especially in medical discourse, primarily use lexical bundles for reporting statistical data.

The relatively lower occurrence of LBs and LFs in Move 2 (Review the experiment) and Move 4 (Summary report) suggests that these moves are optional and do not occur frequently in medical research articles (Davis, 2020). Previous studies using top-down approaches may report lower frequencies due to predetermined move categories that miss these distribution patterns.

Interestingly, the identification of multifunctional bundles and frames across different moves, such as *health-related quality of life* and *for the * of the*, indicates flexibility in formulaic sequence usage. This multifunctionality also accounts for why the move-level totals exceed the actual number of LB and LF types identified. From a pedagogical perspective, this connection between formulaic sequences and moves provides novice writers with concrete linguistic resources tied to specific communicative purposes, addressing the limitation that move structures alone may be too abstract for successful writing (Casal & Kessler, 2020). This mapping thus offers a bridge between macro-level discourse organization and micro-level linguistic realization in medical writing.

5.5 Methodological Reflections & Limitations

The number of LBs and LFs occurrences is relatively low, however, this ensures cross-journal validity and increases the chance that learners might notice these patterns. The high frequency threshold (80 occurrences) and text range (80% across journals) impact the occurrences of LBs and LFs. Experts note that using a high frequency threshold avoids individual writer preferences for writing styles (Biber, 2012; Biber & Barbieri, 2007; Cunningham, 2017; Golparvar & Barabadi, 2020). However, when a lower cut-off point was set, more specific LBs such as a mean of follow up of and LFs like results of the [sensitivity, subgroup, main, secondary, regression, exploratory] analyses were found. This is a limitation of this study, as it aims to investigate medical discourse through manageable analysis. The results are similar to Panthong and Poonpon (2020) and Mbodj (2016), who suggested that setting a low frequency might allow researchers to see more lexical bundles and frames in medical discourse. Therefore, this result could complement discourse knowledge and provide resources for writers to engage more effectively in discourse communities (Mbodj & Cortes, 2024).

To the best of our knowledge, this study is one of a few studies in the field to map LBs and LFs to rhetorical moves using a bottom-up approach. Previous studies typically used top-down approaches, which start with predetermined move categories and may miss naturally occurring patterns in the text (Chanh & Pramoolsook, 2025; Lu et al., 2021; Liu, 2023). The bottom-up approach allows patterns to emerge from the data itself, revealing how formulaic sequences actually function in medical discourse. This methodological choice provides a more accurate picture of the relationship between linguistic features and rhetorical functions (Biber et al., 2007; Le & Harrington, 2015; Omidian et al., 2018).

6. Conclusion

This study explored LBs and LFs in the results sections of MRAs, analysed their structural types, and mapped them to rhetorical move structures. The results found 67 types of LBs and 225 types of LFs, most of which were used to report medical trial subjects and statistical data. These findings suggest that FSs vary across sections and disciplines and contribute to the growing body of knowledge on section-specific and discipline-specific formulaic language. Furthermore, the dual analysis of LBs and LFs provides a more comprehensive account of medical language instead of focusing on a single analysis which lacks accuracy (Biber, 2012). Structurally,

noun phrase fragments were common in LBs while non-verb content words characterised LFs in the Result section of MRAs. A new structural type of LBs *Explicative with Copula Be* was found from the analysis, demonstrating that the structural types of LBs in discipline-specific and section-specific contexts may be different from those found in general academic discourse. Move structure mapping demonstrated that most LBs and LFs concentrated in *Move 3 and Move 1* while fewer occurred in *Move 2 and Move 4*. Additionally, certain lexical bundles and linguistic features are not confined to a single rhetorical move or step but can function across multiple stages of the discourse. These findings indicate that functional segments (FSs), which are discipline-specific, operate as discoursal units that realize specific functions within rhetorical move structures (Cortes, 2013).

Important implications emerge from the research findings. Pedagogically, the identified lexical bundles and frames provide both language and content teachers with solid evidence for corpus-informed teaching. These bundles and frames could be used to guide learners, especially novice and non-native writers, to increase their writing fluency and enhance the linguistic realisation (Simpson-Vlach & Ellis, 2010; Panthong & Poonpon, 2020; Mbodj, 2016). Practically, these findings also offer medical authors references to the authentic expressions and support the development of writing conventions of medical research articles. For material developers, this corpus-driven data could facilitate the creation of medical writing exercises with discipline-specific patterns rather than relying on general academic writing counterparts (Mbodj & Cortes, 2024, 2025). Methodologically, this work, to the best of our knowledge, is one of the earliest studies to map LFs to move structures via a bottom-up approach, allowing authentic patterns to emerge from corpus data rather than imposing predetermined categories. This approach could provide expressions that occur with high frequency and help avoid researcher bias (Biber et al., 2007; Omidian et al., 2018). For the existing literature on discourse analysis, this study contributes novel insights into the variation of different types of formulaic sequences in the results sections of MRAs.

However, this study has some limitations. First, this study only focused on the results sections of medical research, giving an incomplete picture of MRAs. Future research should examine other sections (i.e., introduction, methods, and discussion). Secondly, the high-frequency threshold may have overlooked meaningful low-frequency lexical bundles and frames that serve specialised functions in disciplines. Setting lower frequency thresholds could capture a fuller range of formulaic sequences. Third, treating MRAs as homogeneous discourse overlooks potential sub-disciplinary variation across cardiology, oncology, or neurology. Future comparative analyses should investigate these variations to advance the knowledge of medical discourse.

Acknowledgments

Not applicable

Author contributions

Both authors contributed equally to this work. Panupan Panthong and Anchalee Wannaruk were involved in the study conception, data analysis, manuscript writing, and revision. Both authors read and approved the final manuscript.

Funding

This work was supported by (i) Suranaree University of Technology (SUT), (ii) Thailand Science Research and Innovation (TSRI), and (iii) the National Science, Research and Innovation Fund (NSRF), Project No. 179345.

Competing interests

Not applicable

Informed consent

Obtained.

Ethics approval

The Publication Ethics Committee of the Sciedu Press.

The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Data sharing statement

No additional data are available.

Artificial Intelligence (AI) Usage Disclosure

The authors used Claude AI (Sonnet 4.6, Anthropic) for language revision and polishing only. All research content, analysis, and references were independently developed and curated by the author, who takes full responsibility for the manuscript's accuracy and integrity.

References

Al-Shujairi, Y. B. J., Tan, H., Abdullah, A. N., Nimehchisalem, V., & Imm, L. G. (2020). Lexical bundles in the discussion section moves of high impact medical research articles. *Pertanika Journal of Social Sciences and Humanities*, *28*(3).

Amnuai, W., & Wannaruk, A. (2012). Investigating move structure of English applied linguistics research article discussions published in international and Thai journals. *English Language Teaching*. https://doi.org/10.5539/elt.v6n2p1

Asano, M., Hirosuna, K., & Fujieda, M. (2024). Exploring Lexical Bundles in the Move Structure of English Medical Research Abstracts: A Focus on Vocabulary Levels. *Languages*, *9*(9), 281. https://doi.org/10.3390/languages9090281

Bhuiya, T., & Makaryus, A. N. (2023). The Importance of Engaging in Scientific Research during Medical Training. In *International Journal of Angiology* (Vol. 32, Issue 3). https://doi.org/10.1055/s-0042-1759542

Biber, D. (2012). Corpus-Based and Corpus-driven Analyses of Language Variation and Use. In *The Oxford Handbook of Linguistic Analysis*. https://doi.org/10.1093/oxfordhb/9780199544004.013.0008

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, *26*(3), 263-286. https://doi.org/10.1016/j.esp.2006.08.003

Biber, D., Connor, U., & Upton, T. A. (2007). *Discourse on the move: Using corpus analysis to describe discourse structure*. John Benjamins Publishing. https://doi.org/10.1075/scl.28

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*. Longman.

Casal, J. E., & Kessler, M. (2020). Form and rhetorical function of phrase-frames in promotional writing: A corpus- and genre-based analysis. *System*, *95*, 102370. https://doi.org/10.1016/j.system.2020.102370

Casal, J. E., & Yoon, J. (2023). Frame-based formulaic features in L2 writing pedagogy: Variants, functions, and student writer perceptions in academic writing. *English for Specific Purposes*, *71*, 102-114. https://doi.org/10.1016/j.esp.2023.03.004

Chanh, N. H., & Pramoolsook, I. (2025). English oral and maxillofacial Surgery research articles: Move and phrase frames in the introduction sections. *World Journal of English Language*, *15*(7), 340. https://doi.org/10.5430/wjel.v15n7p340

Chen, J., & Zhang, P. (2025). Phrase frames in rhetorical moves of economic research article discussions. *Southern African Linguistics and Applied Language Studies*, *43*(2), 168-182. https://doi.org/10.2989/16073614.2024.2353720

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

Cortes, V. (2023). Lexical Bundles and Phrase Frames. In *Conducting Genre-Based Research In Applied Linguistics: A Methodological Guide* (pp. 105–126). Taylor and Francis. https://doi.org/10.4324/9781003300847-8

Cortes, V., & Lake, W. (2023). LBIAP. *International Journal of Corpus Linguistics*, *28*(2), 263–277. https://doi.org/10.1075/ijcl.21100.cor

Davis, R. H. (2020). Moves and Steps in the IMRD Sections of Medical Research Articles. *Taiwan International ESP Journal*, *11*(1), 23-39. https://doi.org/10.6706/TIESPJ.202008_11(1).0002

Fazilatfar, A. M., & Naseri, Z. S. (2016). The Study of Rhetorical Moves in Applied Linguistics Research-Based Articles Written by Iranian Researchers. *International Journal of Linguistics*, *8*(6). https://doi.org/10.5296/ijl.v8i6.10468

Fuster-Márquez, M. (2022). Applying a corpus-driven approach in linguistic analyses: The case of lexical bundles and phrase frames. In *Bloomsbury Academic eBooks*. https://doi.org/10.5040/9781350275256.ch-4

Gao, S., & Pramoolsook, I. (2022). Move-Step Structure of The Results and Discussion Section of Electronic Engineering Research Articles Written by Chinese and Thai Researchers. *Journal of Teaching English for Specific and Academic Purposes*, 725. https://doi.org/10.22190/JTESAP2104725G

Gholami, J., & Zeinolabedini, M. (2015). A diagnostic analysis of erroneous language in Iranian medical specialists' research papers. *Journal of Tehran University Heart Center*, *10*(1).

Golparvar, S. E., & Barabadi, E. (2020). Key phrase frames in the discussion section of research articles of higher education. *Lingua*, *236*, 102804. https://doi.org/10.1016/j.lingua.2020.102804

Gray, B., & Biber, D. (2013). Lexical frames in academic prose and conversation. *International Journal of Corpus Linguistics*, *18*(1).

https://doi.org/10.1075/ijcl.18.1.08gra

Guinda, C. S. (2015). Genres on the move: Currency and erosion of the genre moves construct. *Journal of English for Academic Purposes*, *19*, 73-87. https://doi.org/10.1016/j.jeap.2015.07.001

Habtu, Y., & Deressa, W. (2025). How to write a "results section" in biomedical scientific research papers?: Critical review. *Research Methods in Medicine & Health Sciences*, *6*(3), 86-95. https://doi.org/10.1177/26320843241237444

Hu, F. (2024). Related Studies on Formulaic Sequences. In *An MT-Oriented Study of Corresponding Lexical Chunks in Business Correspondences from English to Chinese* (pp. 1-24). Springer Nature Singapore. https://doi.org/10.1007/978-981-99-4977-9_1

Iyengar, K. P., Jain, V. K., & Vaishya, R. (2022). What do editors expect from authors in medical research? *Journal of Orthopaedics*, *32*, 146-150. Reed Elsevier India Pvt. Ltd. https://doi.org/10.1016/j.jor.2022.06.004

Johns, A. M. (2008). Genre awareness for the novice academic student: An ongoing quest. *Language Teaching*, *41*(2). https://doi.org/10.1017/S0261444807004892

Kanoksilapatham, B. (2005). Rhetorical structure of biochemistry research articles. *English for Specific Purposes*, *24*(3), 269-292. https://doi.org/10.1016/j.esp.2004.08.003

Li, L., Franken, M., & Wu, S. (2020). Bundle-driven move analysis: Sentence initial lexical bundles in PhD abstracts. *English for Specific Purposes*, *60*. https://doi.org/10.1016/j.esp.2020.04.006

Liu, C., & Pan, F. (2024). Connecting lexical bundles and moves in medical research articles' Methods section. *Southern African Linguistics and Applied Language Studies*, *42*(1). https://doi.org/10.2989/16073614.2023.2226171

Liu, L., Jiang, F. (Kevin), & Du, Z. (2023). Figure legends of scientific research articles: Rhetorical moves and phrase frames. *English for Specific Purposes*, *70*, 86-100. https://doi.org/10.1016/j.esp.2022.11.005

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

Lu, X., Yoon, J., & Kisselev, O. (2021). Matching phrase-frames to rhetorical moves in social science research article introductions. *English for Specific Purposes*, *61*, 63-83. https://doi.org/10.1016/j.esp.2020.10.001

Luo, N., & Hyland, K. (2019). "I won't publish in Chinese now": Publishing, translation and the non-English speaking academic. *Journal of English for Academic Purposes*, *39*, 37-47. https://doi.org/10.1016/j.jeap.2019.03.003

Mbodj, N. B., & Cortes, V. (2024). Using multiword collocations as a tool to address the demands of conventionalized medical discourse for international publication. *English for Specific Purposes*, *75*, 119–135. https://doi.org/10.1016/j.esp.2023.12.004

Mbodj, N. B., & Cortes, V. (2025). Exploring Within-Discipline Variation of Lexical Bundle Use: An Example from Medical Research Articles and Medical Case Reports. *Revista Brasileira de Linguística Aplicada*, *25*(1). https://doi.org/10.1590/1984-6398202550012

Mbodj-Diop, N. B. (2016). Lexical bundles in medical research articles: Structures and functions (Unpublished master thesis). Michigan State University, East Lansing, MI.

Meyer, C. F. (2023). *English Corpus Linguistics: An Introduction*. Cambridge University Press. https://doi.org/10.1017/9781107298026

Nwogu, K. N. (1997). The medical research paper: Structure and functions. *English for Specific Purposes*, *16*(2), 119-138. https://doi.org/10.1016/S0889-4906(97)85388-4

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

Panthong, P., & Poonpon, K. (2020). Lexical bundles in Thai Medical Research Articles. *Journal of Studies in the English Language*, *15*(1), 59-106.

Raffing, R., Jensen, T. B., Larsen, S., Konge, L., Møller, C., & Tønnesen, H. (2021). Facilitators and barriers for young medical doctors writing their first manuscript for publication. *International Journal of Environmental Research and Public Health*, *18*(16). https://doi.org/10.3390/ijerph18168571

Rezaeian, M. (2015). Disadvantages of publishing biomedical research articles in English for non-native speakers of English. *Epidemiology and Health*, *37*. https://doi.org/10.4178/epih/e2015021

Salager-Meyer, F. (2014). Origin and development of English for Medical Purposes. Part I: Research on written medical discourse. *Medical Writing*, *23*(1). https://doi.org/10.1179/2047480613z.000000000187

Schneider, G. (2023). Do non-native speakers read differently? Predicting reading times with surprisal and language models of native and non-native eye tracking data. In *Language and Linguistics in a Complex World*. https://doi.org/10.1515/9783111017433-008

Shahriari, H. (2017). Comparing lexical bundles across the introduction, method and results sections of the research article. *Corpora*, *12*(1). https://doi.org/10.3366/cor.2017.0107

Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*.

https://doi.org/10.1093/applin/amp058

U çar, S., & Zarfsaz, E. (2023). A corpus-based teaching of lexical bundles to enhance writing skills of prospective Turkish EFL teachers. *International Journal of Learning and Teaching*, *15*(1). https://doi.org/10.18844/ijlt.v15i1.8577

Waaijer, C. J. F., Ommering, B. W. C., van der Wurff, L. J., van Leeuwen, T. N., & Dekker, F. W. (2019). Scientific activity by medical students: the relationship between academic publishing during medical school and publication careers after graduation. *Perspectives on Medical Education*, *8*(4). https://doi.org/10.1007/s40037-019-0524-3

Williams, I. A. (1999). Results Sections of Medical Research Articles: Analysis of Rhetorical Categories for Pedagogical Purposes. *English for Specific Purposes*, *18*(4). https://doi.org/10.1016/S0889-4906(98)00003-9

Wray, A. (2013). Formulaic language. *Language teaching*, *46*(3), 316-334. https://doi.org/10.1017/s0261444813000013

Wulff, S. (2018). Acquisition of formulaic language from a usage based perspective. In *Understanding Formulaic Language: A Second Language Acquisition Perspective* (pp. 19-37). https://doi.org/10.4324/9781315206615

Yu, Y. (2022). The role of psycholinguistics for language learning in teaching based on formulaic sequence use and oral fluency. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.1012225

Appendix A

Examples of Lexical Bundles in each move and step

Move 1: Report the finding

Step A: State outcome

in the usual care group, in the per protocol analysis, the proportion of patients who, health related quality of life, of the participants in the, group and the placebo group

Step B: Support with evidence

the most common adverse events, health related quality of life

Step C: Direct Readers to figure or table

are shown in table 1, are shown in table 2

Move 2: Review the experiment

Step A: Review the subjects

of the patients in the, were included in the analysis, patients were included in the, the baseline characteristics of the, participants in the intervention group

Step B: Review the procedure

patients were randomly assigned to, were lost to follow up, were randomly assigned to the, the presence or absence of, the modified intention to treat, on the basis of the

Step C: Review and Compare to previous studies

N/A

Move 3: Describe statistical test and state findings

Step A: Report statistical data

there was no significant difference, there were no significant differences, was no significant difference in

did not differ significantly between, there was no statistically significant, was not significantly different between not significantly different between the, there were no statistically significant

Step B: Analyse statistical data

participants in the intervention group, the percentage of patients with, the proportion of patients with, the proportion of patients who, group than in the placebo, patients in the control group, participants in the placebo group, serious adverse events were reported

Move 4: Summary report of the overarching findings

there was no evidence of

Appendix B

Examples of Lexical Frames in each move and step

Move 1: Report the finding

Step A: State outcome

results of the [primary, per-protocol, sensitivity, main, final] analysis, showed that the [majority, incidence, risk, expression, proportion, effect, level(s), rate] of, the [result(s), finding(s), analysis, component(s)] of the primary, found that the [expression, level, majority, proportion, percentage, number, presence] of, these results [show, suggested, demonstrate, revealed] that the, the proportion of [patients, participants, women, individual, children] who, in the [subgroup, proportion, percentage, number, group, population] of patients, in the usual, standard care group

Step B: Support with evidence

by the [end, square, time, value,] of the, these results [indicated, suggested] that the, was also [observed, detected, reflect, evident, noted] in the, in the [intervention, guideline, meta-analysis,] compared with

Step C: Direct Readers to figure or table

are shown in [table, supplement, supplement, figure, eTable, appendix] 2, are [shown, provided, presented, summarised, listed, reported, scribed, detailed, given, depicted] in table 1. in table [number of table] in the, shown in figure [number of figure] and, shown in table [number of table] and, and table [number of table] in the

Move 2: Review the experiment

Step A: Review the subjects

the baseline characteristics [of, for, between, in, and] the, and [clinical, disease, baseline, demographic, sociodemographic, hospital-level] characteristics of the, of the [patients, participants, contacts, infants, women, outcome] at baseline, age of [the, these, included, enrolled, all, deceased, matched, study] patients was, of the [patients, participants, contracts, contract, infants, women, outcome] at baseline, with a median [follow-up, age, duration, number] of, the [mean, median, average] age of the, the median [age, duration, date, follow-up, value, dose] of the, received at least [number] dose, a total of [number] patients, a total of [number] participants

Step B: Review the procedure

were randomly [assigned, allocated, selected] to receive, were well balanced [between, in, across, among, and, by] the, assigned to the [placebo, intervention, control] group, the presence [or, and, versus] absence of, patients were [included, enrolled, treated, grouped] in the, we [examined, investigated, assessed, evaluated, tested, analyzed] the effect of, the [end, characteristics, course, duration] of the study

Step C: Review and compare to previous studies

consistent with the [results, findings, pattern, result, effect] of, consistent with those [of, in, form, for] the, were [similar, identical, comparable, opposite], similar to those [of, in, from, for, among, with]

Move 3: Describe statistical test and state findings

Step A: Report statistical data

did not differ [significantly, substantially, statistically, meaningfully, notably] between, was [no, a] significant difference in, there [was, were, is] no statistically significant, were no significant [significant, changes, findings, reductions] in, was significantly [higher, lower, longer, greater, reduced, increased, decreased, improved, attenuated] in the, a [significant, small, substantial] decrease in the, lower in the [intervention, closed-loop, remdesivir, booster, intensive-treatment, high-dose] group, in the [intervention, placebo, surgery, control] group than

Step B: Analyse statistical data

were more [common, frequent, prevalent, durable, poly functional, prominent] in the, these [data, studies] suggest that the, the [incidence, rate(s), majority, type(s)] of adverse events, associated with a [lower, higher, reduced] risk, patients [in, from, of] the placebo group, in the [placebo, intervention, control] group reported, the most [common, frequent] adverse events, differences in the [number, incidence, rates, percentage, prevalence, proportion] of, in the (intervention, induction, treatment) group compared

Move 4: Summary report of the overarching findings

at the [end, recommendation] of the, the effect of [treatment, dulaglutide, vaccination] on