

Exploring Within-Discipline Variation of Lexical Bundle Use: An Example from Medical Research Articles and Medical Case Reports

Explorando a variação intradisciplinar do uso dos grupos lexicais: um exemplo de artigos de pesquisa médica e relatos de casos médicos

Ndeye Bineta Mbodj*

*Georgia State University

Atlanta | GA | USA

ndeye.mbodj@fulbrightmail.org

<https://orcid.org/0000-0002-0604-6101>

Viviana Cortes*

*Georgia State University

Atlanta | GA | USA

vcortes@gsu.edu

<https://orcid.org/0000-0003-4704-779X>

Abstract: Studies that focus on the use of formulaic language in academic writing primarily investigated the use of different types of formulaic sequences in academic research articles in various disciplines. These studies revealed dramatic variations in the use of formulaic language across academic disciplines (e.g., Cortes, 2004; Hyland, 2008; Jalali; Moini, 2014; Shahriari, 2017). However, there is evidence that discipline alone may not be enough to determine linguistic variation (Gray, 2015), as different registers within one discipline may reflect different linguistic characteristics depending on specific communicative purposes (Biber; Conrad, 2019). In the case of medical written discourse, different medical registers may call for the use of different formulas to convey different communicative intents. This is the case of the medical research article (MRA) and the medical case report (MCR). In this study, we will focus on the identification and analysis of lexical bundles in medical discourse, particularly in MRAs and MCRs. Lexical bundles (Biber *et al.*, 1999) are groups of three or more words that frequently occur in a given register. They are identified empirically, using specially designed computer programs that run through the texts in a language corpus and flag expressions of three or more words that repeat frequently, employing a strictly corpus-driven approach (Cortes, 2015). Using specially designed software, lexical bundles of 3+ words were identified in each corpus. These expressions were later

analyzed according to their structure and the functions they performed to conduct a comparison across the two registers using taxonomies adapted from the literature (Biber; Conrad; Cortes, 2004; Hyland, 2008).

Keywords: lexical bundles; medical discourse; register variation.

Resumo: Os estudos que focam no uso da linguagem formulaica na escrita acadêmica investigaram principalmente o uso de diferentes tipos de expressões formulaicas em artigos de pesquisa acadêmica em diversas disciplinas. Esses estudos revelaram variações marcantes no uso da linguagem formulaica entre disciplinas acadêmicas (por exemplo, Cortes, 2004; Hyland, 2008; Jalali; Moini, 2014; Shahriari, 2017). No entanto, há evidências de que a disciplina por si só pode não ser suficiente para descrever a variação linguística (Gray, 2015), uma vez que, dentro de uma disciplina, diferentes registros podem apresentar diferentes características linguísticas que refletem propósitos comunicativos específicos (Biber; Conrad, 2019). No caso do discurso médico escrito, diferentes registros médicos podem exigir a utilização de diferentes expressões formulaicas para transmitir diferentes intenções comunicativas. É o caso do artigo de pesquisa médica (APM) e do relato de caso clínico (RCC). Neste estudo, focaremos na identificação e análise de grupos lexicais no discurso médico, particularmente em APMs e RCCs. Os grupos lexicais (Biber *et al.*, 1999) são grupos de três ou mais palavras que ocorrem frequentemente em um determinado registro. Os grupos lexicais foram identificados empiricamente graças a uma abordagem estritamente orientada pelo *corpus*, na qual foram utilizados programas de computador especialmente projetados para buscar e sinalizar nos textos as expressões de três ou mais palavras que se repetem frequentemente em um *corpus* linguístico (Cortes, 2015). Estas expressões foram posteriormente analisadas, conforme a sua estrutura e as funções que desempenham, para efetuar uma comparação entre os dois registros médicos, utilizando taxonomias adaptadas da literatura (Biber; Conrad; Cortes, 2004; Hyland, 2008).

Palavras-chave: grupos lexicais; discurso médico; variação de registro.

1 Introduction

Lexical bundles have become an important construct in the wide spectrum of formulaic language expressions. The reason for their popularity derives perhaps from the empirical way in which these expressions are identified and from the specific functions they perform in discourse, particularly in specialized texts. Ever since Biber, Johansson, Leech, Conrad, and Finegan (1999) coined the term *lexical bundles* to label any group of three or more continuous words that frequently recurred in a register, studies that analyzed lexical bundle use in academic and professional genres have been continuously published in top journals in the applied linguistics field.

Lexical bundles help language users convey their communicative intentions and for this reason they have been considered building blocks at the sentence level (Cortes, 2013; Biber; Conrad, 1999). They are often not complete structural units, but they can be grouped according to their structural characteristics. In written registers, particularly in academic written registers, lexical bundles tend to be phrasal, made up of noun phrase or prepositional phrase fragments with fragments of other embedded phrases. In spoken registers, bundles tend to be clausal, mostly fragments of verb phrases or dependent clauses with other embedded fragmented structures. Some examples of lexical bundles frequently found in written academic registers are expressions such as *at the end of*, *as a result of the*, *in the context of the*, *it is important to*, *on the other hand*, and *the purpose of this study is to*, to mention only a few examples (Biber *et al.*, 1999).

As discourse building blocks, lexical bundles have been often employed to help describe particular registers because they can help identify the specific language purposes that speakers or writers are trying to communicate (Cortes, 2013, 2023). Frequently, studies that focus on the use of lexical bundles in academic writing have primarily investigated the use of these formulaic sequences in academic research articles in a particular discipline or compared their use across different disciplines. These studies revealed dramatic variation in the use of bundles across academic disciplines (e.g., Cortes, 2004; Hyland, 2008; Jalali; Moini, 2014; Shahriari, 2017). There is ample evidence, however, that discipline alone may not be enough to determine linguistic variation (Gray, 2015), as different registers within one discipline may reflect different linguistic characteristics depending on specific communicative purposes (Biber; Conrad, 2019). In specialized areas, different registers within the same field may make use of different expressions because they need to convey different language functions for different communicative purposes. These differences, however, have not been extensively explored in terms of the use of frequent formulaic expressions compared across different registers within one disciplinary context.

In the medical field, there are several registers that medical professionals produce to satisfy the communication needs of their specialized community. Dorgeloh (2016) explains that there are general discourse goals and purposes common to several medical registers, but there are patterns specific to a given register that may present features that do not occur in texts from other medical registers. Some examples of written registers in this field that could help explore differences in communicative purposes are the medical research article (MRA) and the medical case report (MCR). While both registers are produced by medical professionals, the MRA has been described as a highly technical report of experimental

research, typically organized in the Introduction, Method, Results, and Discussion (IMRD) format (Nwogu, 1997, p. 117). On the other hand, the MCR has been defined as “a detailed narrative that describes, for medical, scientific, or educational purposes, a medical problem experienced by one or several patients” (Gagnier *et al.*, 2013, p. 1). These differences in communicative intentions could reveal variations in the use of specific expressions to convey particular functions.

The purpose of this study is to determine the most frequent lexical bundles identified in MRAs and MCRs to further explore within-discipline variation, focusing on the divergent and convergent structures and functions of these recurrent word combinations. The analysis will combine a quantitative, data-driven approach to the identification of bundles and the qualitative interpretations of the functions these expressions perform in each of the registers under study. It is important to explain, however, that this study of lexical bundles is a small section of a much larger study of MRAs and MCRs that tried to identify the formulaic profile of the texts in these registers by analyzing different types of recurrent word combinations: collocations, multiword collocations, lexical bundles, and phrase frames (Mbodj, 2021).

The rest of the paper will be organized as follows. Section two will present literature related to the two major areas of inquiry covered in this study: register variation in specialized discourse and lexical bundles. In section three, we will introduce our corpora, and the methodological procedures followed for the identification and analysis of bundles, as well as a brief recount of the results of the situational characteristics analysis we conducted to guide the corpus collection and help explain register variation. Section four will present the results of the study and a discussion of the most salient findings. Finally, in section five we conclude by discussing suggestions for pedagogical applications and further studies.

2 Review of Related Literature

In this section we want to focus on two areas of inquiry that are essential to our study: a thorough description of the main construct in our study, lexical bundles, focusing on studies that compared the use of these expressions across registers, and a brief introduction to register variation. We want to provide a solid framework of past studies that may help us define our construct, help us highlight the gap in the literature we are trying to fill in, and could contribute to provide accountability for some of our methodological choices.

2.1 Lexical Bundles

Lexical bundles are frequent expressions of three or more words that recur in a given register. They are identified empirically by a specially designed computer program that runs through a corpus and flags the repeated use of these expressions as well as the number of texts in which the expressions occur to compute their frequency and range.

2.1.1 Identifying Lexical Bundles

The advancement of computer processing and the compilation of large collections of texts or language corpora brought about new perspectives for the study of formulaic language (Cortes, 2015). In the frequency-based tradition to the study of recurrent word combinations, which is based on frequency and statistics, there are two distinct approaches to the identification of these expressions: the corpus-based approach and the corpus-driven approach (Tognini-Bonelli, 2001). The corpus-based approach relies on expressions that have been considered formulaic in linguistic theory, working with a group of pre-selected formulaic sequences and how these sequences are used in a given corpus (Nattinger; DeCarrico, 1992). The corpus-driven approach is an inductive procedure with few theoretical assumptions and little preconception, which uses frequency distribution for the identification of recurrent expressions. Among the formulaic constructs that emerged from the corpus-driven approach, lexical bundles have become a popular feature in studies that tried to describe particular registers from a formulaic perspective (Cortes, 2004; Hyland, 2008).

This procedure is the same used to identify *n*-grams, that is, groups of *n* number of words in a corpus. The difference between these two constructs can be summarized in that all lexical bundles are *n*-grams but not all *n*-grams are lexical bundles because to be considered a bundle, an *n*-gram has to meet conventionally pre-established thresholds for total frequency and number of texts. Biber *et al.* (1999) established a frequency threshold of 10 times per million words (pmw) in a corpus for three, four, five, and six-word expressions, and they used a range of 5 texts to avoid any idiosyncratic use of a given expression. Later studies used more conservative thresholds for frequency (20 or 40 times pmw) and a percentage of the texts in the corpus for range (5% or 10%). The reason for these higher, more conservative, cut-off points is to ensure the expressions are really frequent, as frequency is the only quality of these word sequences. As the number of lexical bundles tends to decrease as their size increases, more recent studies have used stratified frequency thresholds. Cortes (2013, 2023) used 60 times pmw for 3-word bundles, 20 times pmw for 4-word bundles, 15 times pmw for 5-word bundles, 10 times pmw for 6-word bundles, and 6 times pmw for any bundle larger than 6 words that could be identified in her corpora.

2.1.2 Lexical Bundles as Functional Building Blocks

Perhaps one of the reasons why lexical bundles have become popular in the formulaic language spectrum is the fact that they help convey specific functions in discourse (Biber; Conrad; Cortes, 2004). Bundles can help express stance, as in *are more likely to*, *the fact that the*, *it is important to*, *it is necessary to*, *can be used to*, and *it is possible to*. Expressions like *in this chapter we*, *on the other hand* or *as well as the* are frequently used to organize discourse. Furthermore, the most frequent function that lexical bundles perform in written academic discourse is to indicate particular types of reference. Some examples of referential bundles are *is one of the* and *one of the most* (identification/focus reference), *the rest of the* (quantity), *the size of the* and *in the form of* (tangible framing attributes), *in the case of* and *in the context of* (intangible framing attributes), *in the United States* and *at the university* (place reference), *at the same time* and *at the time of* (time reference), *as shown in figure* and *in the next section* (text deictic), and *at the beginning of* and *the top of the* (multifunctional reference), to mention just a few examples

in each subcategory. Furthermore, Cortes (2013) tried to connect lexical bundles frequently present in published research article introductions to the communicative purposes they help convey, as realized by rhetorical moves and steps (Swales, 2004). She found that some lexical bundles are inherently bound to a single move or step in these sections of research articles while others perform functions across different moves and steps. For example, the 4-word lexical bundle *little is known about* was frequently found in move 2 step 1 (indicating a gap), but the bundles *in the context of* and *is based on the* occurred in different moves and steps in the introductions under study. Another interesting finding in Cortes (2013) was the connection between the bundles and the move in terms of initiating or triggering the move/step or commenting on or complementing other discourse that was used to trigger the move/step.

2.1.3 Comparing Lexical Bundles Across Registers

Lexical bundles can be used to identify variation across different registers, particularly across academic disciplines. Reppen and Olson (2020) studied lexical bundles across nine different academic disciplines and found cross-disciplinary bundles, which often function as discourse frames providing guidance to readers, and discipline specific bundles, which were usually related to the specific content presented in the texts under study. In our study, we will be focusing on the identification, analysis, and classification of lexical bundles, and we will conduct a comparison of their use in the two medical registers selected.

Many studies that identified and analyzed lexical bundles centered in the use of these formulaic expressions in published research articles and research article sections from numerous disciplines (Cortes, 2004, 2013; Lake; Cortes, 2020; Wright, 2019). A few studies have focused on the use of lexical bundles in other academic registers. For example, Chen (2018) looked at lexical bundles frequently used in university textbooks in electrical engineering. Other examples of studies of lexical bundles in diverse academic writing registers are the study by Shin (2019), who investigated the use of bundles in the argumentative writing of undergraduate students, and the study by Lu and Deng (2019), who analyzed bundles in doctoral dissertation abstracts. Both of these studies compared the use of bundles in the production of L1 and L2 writers.

The study of lexical bundles by Hyland (2008) compared lexical bundles used in research articles, MA theses, and doctoral dissertations from various disciplines. He discovered that writers used different types of expressions in each of those registers because they needed to perform different communicative purposes required by the situation in which each text type was produced. This has been considered a very important study in the literature because the investigation of lexical bundle use across very different registers in the same discipline or across disciplines has not been extensively explored.

2.2 Specialized Discourse and Register Variation

The notion of variation in language according to context has been of interest in linguistics for several decades and has been explored from different linguistic traditions (Lin; Mwinlaaru; Tay, 2021). The communicative contexts in which languages are produced bring about many situational variables that accompany the communicative act, which are fundamental aspects

of human language. Biber and Conrad (2019, p. 23) explain that, in the case of registers, which are situationally defined varieties of a language, this phenomenon “focuses on the pervasive patterns of linguistic variation across such situations, in association with the functions served by linguistic features”. These authors explain that in order to describe languages or language varieties and to develop cross-linguistic language use theories, it is essential to study variation to understand how linguistic features are used across different types of texts.

Studies in search of outlining the linguistic characteristics of a register sometimes focus on only one register. The majority of studies in this area of inquiry, however, focus on variation. Biber and Conrad (2005, p. 179) present examples of register variation in lexical descriptions, analyzing the use of downtoners in American conversation and academic prose with the purpose of describing the meaning and use of related words. They also studied register variation in grammatical descriptions, investigating the use of *that* clauses with the relativizer *that* retained and without the relativizer *that* (omitted) in fiction, academic prose, conversation, and news reportage. They reported that register variation in the use of these grammatical features depended on “the extent to which the typical functions of the feature fit the typical communicative characteristics of the register” (Biber; Conrad, 2005, p. 183).

In the analysis of specialized discourse, as in the case of academic registers in general and written academic registers in particular, the majority of studies conducted for this purpose take a comparative approach to try to identify and analyze linguistic characteristics across different registers (Gray, 2015). For example, many studies that analyze language use in published research articles often focus on disciplinary variation. Bazerman (1994) maintains that understanding the language used in the disciplines may help know these disciplines better, help writers use language more effectively, guide editors in their evaluation tasks, and help readers identify the discourse of these disciplines. Gray (2015, p. 3) states that “Describing and understanding patterns of language use in academic prose allows us to understand the disciplinary cultures and practices that they embody for a variety of purposes”.

Linguistic variation is a complex phenomenon that needs to be explored comprehensively taking into account many important variables. For example, Gray (2015) explains that linguistic variation is likely to result from the research paradigm favored in specific disciplines and that the choice of a particular research methodology may bring about similar ways to describe procedures that could be used across several disciplines, resulting in different or similar linguistic features. Gray’s study contributes to the discussion that argues that register variation in written academic prose may not be limited to disciplinary variation. Many disciplines make use of a set of established registers to communicate within their academic communities. These registers may have shared traits that come from the intertextuality established by the community that produces and consumes those texts (Porter, 1986).

Studies that analyzed linguistic features across different registers produced by a single discipline, however, have been rare. In the study reported in this article, we will be focusing on a lexicogrammatical perspective on register variation through the study of lexical bundles and their use in two registers produced by the medical community: the research article and the case report.

3 Corpora and Methodology

In this section, we will present a brief report of the situational characteristics analysis that guided the corpus compilation and was later used to interpret some of the most salient findings, describe the corpora used for our study, and discuss the corpus collection procedure. We will also describe the process of bundle identification as well as the structural and functional analysis conducted to reveal the potential converging and diverging lexical bundle use across registers.

3.1 Brief Recount of the Situational Characteristics Analysis

For the analysis of MRAs, we used a framework adapted from Gray (2015) that included eight categories. We conducted a pilot study on 50 MRAs and 50 MCRs to test out the framework categories and to familiarize ourselves with the texts. Table 1 contains the categories used and some examples for each category. The results of this preliminary situational characteristics analysis were used as a guide to collect the rest of the texts that comprised the corpora used for this study.

Table 1 – Situational Characteristics Analysis Framework

Categories	Examples
Participants	Authors, actual writers, professional training/title, and primary audience
Textual layout and organization	Word count, headings, use of abstract, visual elements, and sections/organization
Setting	Type of journal
Subject/topic	Main themes in texts
Purpose	Introduce a case to the medical community
Nature of data or evidence	Presence of observe data, use of numerical evidence, object of study, and primary presentation of evidence
Methodology	Observational, experimental, and use of statistical techniques
Explicitness of research design	Explicitness of purpose, of research questions, of citations, in-text reference to cited authors, explanation of evidence and explanation of procedures

Source: Adapted from Gray (2015).

MCRs were analyzed on a very similar framework, but categories 6, 7, and 8 in the MRA framework were merged into the methodology category because the differences across these categories seemed to be rather blurred in MCR texts. Most of the subcategories were retained with the exception of some subcategories in explicitness of the research design, explicitness of the research questions, and explanation of evidence and procedures as these elements are not found in MCRs.

As expected, some of the similarities derived from the fact that the two registers belonged to the same field, which resulted in the discussion of common topics. MRAs and

MCRs focus on medical pathologies and treatments and the purpose of both types of texts is to advance research and clinical practice in the medical field. The comparison of situational characteristics showed marked differences because, even though the general purpose of these texts may show some overlap, they have divergent approaches to researching these shared topics. While MRAs report generalizable results on medical pathologies and treatments, MCRs describe a single or a few cases that are not generalizable but can inform medical practice and education (Mbodj, 2021, p. 95). MRAs primarily focus on research describing methods and procedures and MCRs focus on the patient, the correct diagnosis of the medical pathology, and the patient's response to the provided treatment.

The organization of these texts also showed differences, which may show differences in writing practices. MRAs do not present problem-solution patterns, which are often found when presenting the case in MCRs. These organizational differences may bring about linguistic differences as different linguistic features may be necessary when describing experimental procedures in contrast to those used to report results of clinical procedures, for example. The comparison of the methodologies in these two registers showed two fundamental differences. In the first case, MRAs present an experimental design and MCRs report observations. This difference also resulted in the description of statistical techniques used in MRAs, which are inexistent in MCRs and extends to differences in the way results are reported in each of these registers. MCRs present many verbs of communication like *confirm*, *indicate*, or *show* to report the outcome of medical procedures or tests. MRAs also use communication verbs, but these texts include reports of the statistical significance of the study results that call for different linguistic features.

Finally, the audiences these texts are produced for are also different. MRAs are primarily directed to medical professionals who are doing research in the medical field to share methodologies, approaches, and results. MCRs, on the other hand, are written for medical students, practitioners, and patients. Thus, while both registers involve the use of specialized language, MCR writers tend to convey their message in less complex language, often resorting to short and straightforward sentences that can be easily understood by medical and non-medical readers.

3.2 Corpora Used in the Study

As previously mentioned, the corpora collection for the Medical Research Article Corpus (MRAC) and the Medical Case Report Corpus (MCRC) was completed using the results of the preliminary situational characteristics analysis as a guide. Table 2 presents some descriptive information about the two corpora. The MRAC was an updated version of a corpus of MRAs used for a preliminary study of lexical bundles in this register (Mbodj-Diop, 2016) and covered articles published from 2005 till 2015. All the papers adhere to the Introduction-Methods-Results-Discussion (IMRD) organization. The corpus was updated in 2021 with the addition of 50 articles published since 2016. The articles were published in five top medical journals, namely, *Science Translational Medicine* (a subsection of the well-known journal *Science*), *The Lancet*, *The New England Journal of Medicine* (NEJM), *The Journal of American Medical Association* (JAMA), and *The Journal of Clinical Investigations* (JCI).

Table 2 – Descriptive information on the MRAC and MCRC

Corpus	Journals	Number of texts	Mean word count per text	SD	Total word count
MRAC	JAMA JCI <i>The Lancet</i> NEJM <i>Science TM</i>	250	4,846	1,527.35	1,209,367
MCRC	JAMA BMJCR JMCR OMCR IMCRJ	704	1,428.8	624	1.005.907

Source: the authors.

The MCRC compilation was guided by the situational characteristics analysis, related literature, information on the journals' websites, and suggestions from expert informants to ensure that all texts included in the corpus comply with the general characteristics of this register as discovered in the situational analysis, and that the texts were published in reputable journals. This quality was essential as most MCRs are published in open-access journals. The texts in the MCRC come from two not open access journals (*JAMA* and *BMJ Case Reports*) and three open-access journals (the *Journal of Medical Case Reports (JMCR)*, *Oxford Medical Case Report (OMCR)*, and *International Medical Case Report Journal (IMCRJ)*), which were recommended by the expert informant.

3.3 Identification and Classification of Lexical Bundles

For this study, we used the Lexical Bundles Identification and Analysis Program (LBiaP), a program that we believe has many advantages over other freeware often used to identify this type of expressions. Cortes and Lake (2023) explain that LBiaP is to this date the only program that eliminates all types of overlapping bundles (simple overlapping, subsuming, and interlocking), which when working with lexical bundles of different sizes is an essential step in the identification (Chen; Baker, 2010). We chose to work with lexical bundles of all possible sizes, from 3-word expressions to the longest expression occurring frequently in the corpus that was 11 words long. As LBiaP only identifies bundles from 3 to 9 words long, we continued exploring the corpus for longer bundles using *AntConc* (Anthony, 2017) until the search yielded no more results. Given that longer bundles are much less frequent than shorter bundles, we followed Cortes (2013) and Cortes and Lake (2023) and discriminated frequency thresholds and ranges for different bundle sizes, as shown in Table 3.

Table 3 – Frequency Thresholds and Ranges for the Identification of Bundles

Bundle Length	Frequency pmw	Range
3 words	20	10
4 words	20	10
5 words	10	10
6 words	8	8
7 + words	6	6

Source: the authors.

As this study is part of a larger investigation that looked at different types of formulaic language, an important consideration had to be established in the bundle identification phase to avoid any overlap between lexical bundles and multiword collocations (Mbodj, 2021). In a previous trial identifying recurrent word combinations, we noticed LBiaP was yielding many expressions that were very frequent in our corpora but did not share the structural characteristics of lexical bundles. These were expressions that could be considered multiword collocations, even when they are identified following a corpus-driven approach and they meet the pre-established frequency and range thresholds. It is undeniable that expressions such as *acid phase reactants*, *shortness of breath*, *intensive care unit*, or *chronic obstructive pulmonary disease*, which recurred frequently and met the frequency and range thresholds established for lexical bundle identification in this study, are very different from lexical bundles, as most of them are dense noun phrases with heavy nominal or adjectival modification. A decision was made that any bundle with more than one lexical word had to be re-examined. We concluded that 3-word and 4-word bundles would only have one lexical word and 5-word and longer bundles could have two or three lexical words but never more than four. There were a few exceptions for expressions in which two or more lexical words were functioning as a unit, as in the case of *a white blood cell count of*, where white blood cell can be considered a unit modifying count and the whole expression has a functional role in discourse, working like many other lexical bundles that are made up mostly of function words.

These processes yielded a list of expressions that were true lexical bundles of a given size (3-word, 4-words, and so on) and allowed us to continue with the structural and functional classifications. For the structural classification, we based our analysis on the frameworks designed by Biber *et al.* (1999), Hyland and Jiang (2018), and Cortes (2023). It was necessary to add some categories to our classification that were not present in those frameworks, as our corpora presented bundles that had not been identified before. Table 4 shows the framework we used for the structural classification of the lexical bundles in the MRAC and MCRC corpora (Mbodj, 2021, p. 48).

Table 4 – Structural classification framework for MRAC and MCRC bundles

Main Categories	Subcategories	Examples from the MRAC and MCRC
Verb phrase-related	Passive verb	<i>were included in the, the patient was started on</i>
	Copular be	<i>were eligible for, is the most common</i>
	Imperative	<i>see the supplementary appendix</i>
	Modals	<i>can be a, may lead to, can also be</i>
Clause-related	Anticipatory it	<i>it is possible that, it is important to</i>
	Abstract subject	<i>his blood pressure was</i>
	Human subject	<i>our patient presented with</i>
	External subject	<i>confirmed the diagnosis of</i>
	As-fragments	<i>as measured by, as described previously</i>
	There fragments	<i>there was no family history of</i>
	(NP) + wh-fragments	<i>when compared with, patients who were</i>
	(VP) + that-fragments	<i>these data suggest that, we found that the</i>
Noun/ Preposition-related	(VP) + to-fragments	<i>to assess the effect of, to confirm the diagnosis</i>
	NP with of-phrase fragment	<i>the onset of, a single dose of, a wide range of</i>
	NP with other post modifier fragments	<i>a man in his, death from any cause</i>
	PP with embedded of-fragment	<i>at the end of the, at the time of diagnosis</i>
	Other prepositional fragments	<i>in individuals with, in addition to the</i>
Adj & Adv-related	Comparative expressions	<i>as in this case</i>
	AdjP /AdvP + PP fragment/ to-inf	<i>consistent with the, more likely to</i>
Function-word only (3-words only)	Superlatives	<i>the most common</i>
	Function words only	<i>as in our, and in the, but not in</i>

Source: the authors.

For the functional classification, we used two different taxonomies adapted from the literature. In the past, functional taxonomies from the literature were adopted and used in a top-down approach to classify lexical bundles identified in different types of texts (Biber; Conrad; Cortes, 2004; Hyland, 2008). Nowadays, it is important to adapt those taxonomies or build new taxonomies in an attempt to use a bottom-up analysis, looking at each lexical bundle in context and identifying its function in discourse (Cortes, 2023). For the lexical bundles identified in the MRAC, we used a taxonomy adapted from Hyland (2008) and Cortes (2023). These taxonomies have been designed based on corpora of research articles and research article sections, thus we believed it would be appropriate to the classification of bundles in this corpus. Table 5 shows the categories and subcategories in this adapted functional taxonomy (Mbodj, 2021, p. 49).

Table 5 – Functional Classification Framework for MRAC Bundles

Categories	Subcategories	Examples from the MRAC
Research-oriented (Structure writers' experiences / activities)	Location (in time/place)	<i>during the study period, at the university of</i>
	Procedure	<i>the use of, by western blotting</i>
	Quantification	<i>the number of, a total of, the proportion of</i>
	Description	<i>the effect of, the risk of, was defined as the</i>
Text-oriented (Organize text and its meaning)	Transition signals	<i>as well as, on the other hand</i>
	Resultative signals	<i>these findings suggest that, as a result of</i>
	Structuring signals ◆ Text reference ◆ Study subject/elements reference	<i>in the present study, are shown in table of patients with, in patients with</i>
	Framing signals	<i>on the basis of, with respect to the</i>
Participant-oriented (Focus on writer or reader)	Stance features	<i>it is possible that, may not be</i>
	Engagement features	<i>it should be noted that</i>
Grammatical only		<i>to that of, any of the, that of the</i>

Source: the authors.

For the analysis of the lexical bundles in the MCRC, we used a taxonomy adapted from the guidelines presented by Gagnier *et al.* (2013) and the frameworks designed by Biber, Conrad and Cortes (2004), Hyland (2008), and Cortes (2023), as shown in Table 6. A close look at the subcategories in Tables 5 and 6 does not reveal dramatic differences in the functions, as the subcategories show a certain level of overlap. The main categories need to be different, however, because, for example, the research-oriented category could only be used for bundles in MRAs and the diagnosis/intervention-oriented category could only be used for bundles in MCRs.

Table 6 – Functional Classification Framework for MCRC Bundles

Categories	Subcategories	Examples from the MCRC
Case-related (presentation and discussion of case)	Description	
	◆ Medical condition-related	<i>we present the case of, has been shown to be</i>
	◆ Subject-related	<i>a woman in her, man with a history of, she had a</i>
	Location (time/place)	<i>at the time of admission, to the emergency department</i>
	Quantification	<i>a high index of suspicion, at a dose of</i>
Diagnosis & Intervention-related (writers' activities and decisions)	Procedure	<i>she was treated with, the patient underwent a</i>
	Result Reporting	<i>with a blood pressure of, was positive for</i>
	Decisions/ Outcome	<i>the decision was made to, complete resolution of</i>

Discourse Organizers	Transition signals	<i>in addition to the, on the other hand, as well as</i>
	Resultative signals	<i>as a result of, due to the presence of</i>
	Structuring signals	
	◆ Text reference	<i>are shown in, reported in the, in this report</i>
	◆ Study subject reference	<i>of these patients, of our patient</i>
	Framing signals	<i>based on the, depending on the, in the absence of other</i>
Grammatical only		<i>and did not, in which the, he did not, it has been</i>
Stance & Engagement	Stance features	<i>may lead to, it is possible that, may be associated with</i>
	Engagement features	<i>should be considered in, physicians should be aware of</i>

Source: the authors.

Once all the lexical bundles identified in both corpora were classified structurally and functionally, we focused on converging bundles, that is, those that were used in the two corpora under study, to later focus on diverging bundles, expressions that were tied to only one of the medical registers represented in our study.

4 Results and Discussion

The use of LBiaP and *AntConc*, the software used to identify bundles, yielded a total of 533 bundles of 3 to 11 words in the MRAC and 505 bundles of 3 to 8 words in the MCRC. Table 7 shows the number of bundle types identified in each corpus by bundle length.

Table 7 – Lexical bundle type totals by bundle length

MRAC						
3-word	4-word	5-word	6-words	7-word	8-words	11-words
362	116	40	11	1	1	1
MCRC						
3-word	4-word	5-word	6-words	7-word	8-words	
270	142	61	30	1	1	

Source: the authors.

The next sections will present the structural and functional classification of the bundles found in the two medical registers under study, highlighting their convergent and divergent use.

4.1 Grammatical Classification of the Lexical Bundles Found in MRAs and MCRs

We have already mentioned that lexical bundles are usually not complete structures, but they can be grouped according to their grammatical correlates. The structural classification of bundles in our corpora showed a high level of overlap in the most frequent structural

groups of the bundles in each corpus. The majority of the bundles belonged to the noun phrase/prepositional phrase-related group, followed by verb-phrase related, and clause-related bundles. Table 8 shows the major structural correlates and their percentage of use in each of the two corpora. The three major structural groups of bundles identified in this study have been frequently found in many written academic registers (Biber; Conrad; Cortes, 2004; Cortes, 2004; Hyland, 2008).

Table 8 – Structural classification of lexical bundles identified in the MRAC and the MCRC

MRAC			MCRC		
Structure	Percentage	Example	Structure	Percentage	Example
Noun-Prep related	53%	<i>of the right eye</i>	Noun-Prep related	50%	<i>an increase in the</i>
Verb-Phrase related	22%	<i>showed no evidence</i>	Verb-Phrase related	26%	<i>had no effect on</i>
Clause-related	17%	<i>we did not observe</i>	Clause-related	20%	<i>we report a case</i>
Adj/Adv related	4%	<i>likely to be</i>	Adj/Adv related	3%	<i>consistent with a</i>
Function-word only	4%	<i>to that of</i>	Function-word only	1%	<i>as in our</i>

Source: the authors.

We need to explain that even though medical research articles and medical case reports are two different registers, and they have different communicative purposes, they share several situational characteristics. The texts may sometimes be directed to different audiences, and they report different types of medical phenomena, but they are written by academics and to some extent consumed by academics. Both registers are heavily informational, which many times results in an extensive use of nominalization and noun phrases. The fragments of noun and prepositional phrases that comprise the major structural group of the bundles identified in these two registers could be proof of this tendency.

Among the noun (NP) and prepositional phrase (PP) related bundles found in both registers, there were some differences according to the bundle length. In the MRAC, the predominant structure was noun phrase related, but while 3- and 6-word bundles tended to be majorly NP related, 4- and 5-word bundles were more often prepositional phrase related. Examples of this structural group in the MRAC are lexical bundles such as *in patients with*, *at a dose of*, *with the use of a*, *the regulation of*, *the number of patients*, and *the baseline characteristics of the*. In contrast, in the MCRC, the predominant structure in the NP and PP-related group was the PP-related group of bundles, which represented 57% of the bundles in this category. Examples of NP and PP-related bundles in the MCRC are *of the skin*, *on the left side*, *at the time of admission*, and *the first case of*.

In the verb-related category, both registers showed a high number of passive constructions. In spite of a recently found decline in the use of passive constructions in scientific writing (Biber; Gray, 2016; Hyland; Jiang, 2018), writers of MRAs and MCRs still favor this structure and very often use bundles that show fragments of passive verb phrases. Examples of these bundles in MRAs are expressions like *are shown in table*, *was performed using the*, *was approved by the*, and *were randomly assigned to receive*. In MCRs, some examples of verb-

related bundles are *was refer to our*, *was found to have*, *should be considered in patients with* and *has been reported to be*, to mention just a few expressions.

The last of the three major grammatical groups of lexical bundles in these registers was clause-related bundles. In the MRAC, these bundles showed an even distribution across the subcategories in the framework used for their classification. These expressions could have human subjects, abstract subjects, or external subjects. Some examples of clause-related bundles in MRAs are expressions like *previous studies suggested*, *we did not observe*, and *all participants provided written informed consent*, among many others. The three types of subjects were also used in MCRs, but the human subjects were the most frequent of the three in these texts, as shown in these examples: *we report a case of*, *this case highlights the importance of*, and *the patient had a history of*.

4.2 Functional Classification of the Lexical Bundles Identified in MRAs and MCRs

We will now describe the major functions performed by lexical bundles in MRAs and MCRs. First, we will discuss the functions performed by those bundles that occurred in both corpora, and we will continue with a discussion of the major functional categories of those bundles that were used exclusively in one register.

4.2.1 Lexical Bundles Shared Across the Two Registers and Their Functions

There were a number of lexical bundle types that occurred in both MRAs and MCRs. Table 9 shows the raw number of shared bundles arranged by bundle length and the percentage of shared bundles for each corpus in 3-, 4-, and 5-word bundles. There were no shared bundles longer than 5 words.

Table 9 – Proportions of Shared Bundles in the MRAC and MCRC by Length

	3-words	4-words	5-words
Raw #_Shared	88	29	5
%_MRAC	24.31%	25%	15.50%
%_MCRC	32.59%	20.42%	8.20%

Source: the authors.

As with the structures of the bundles identified in these medical registers, the functions performed by these shared bundles do not seem to be specific to any of these two registers or the medical field. Most of these expressions have been identified in past studies of lexical bundles as frequently occurring in other academic registers and disciplines (e.g., Biber; Conrad; Cortes, 2004; Cortes, 2004; Hyland, 2008). These bundles did not relate to medical practices or the medical field but performed more general discourse functions. One example of this type of expression is the lexical bundle *at the time of*, which had been functionally classified in the literature as a time marker (Cortes, 2004) and a location time/place marker

(Hyland, 2008). In the MRAC and the MCRC, this bundle fulfilled the same function: specifying time. In fact, 17 out of the 29 4-word bundles shared by the two corpora in our study occurred in the history and biology corpora used by Cortes (2004) and in the biology and engineering corpora used by Hyland (2008). Other examples of these shared bundles were *as a result of*, *as well as the*, *in the context of*, *in the presence of*, *on the basis of*, and *on the other hand*.

The bundles shared by the two medical registers in our study performed varied functions. Table 10 shows the functional categories and the classification of shared bundles by bundle length.

Table 10 – Functions Served by Shared Bundles in the MRAC and MCRC (expressed in bundle types)

	3-words	4-words	5-words
Description	37	8	2
Quantification	10	2	
Location (in time)	5	4	2
Procedure	6	3	
Resultative signals	12	1	
Framing signals	8	6	
Transition signals	1	4	1
Structuring (text reference)	1	0	
Grammatical only	7	0	
Stance	1	1	
<i>Totals</i>	88	29	5

Source: the authors.

Our analysis also showed a number of bundles that include field-related terms, which may be shared across the medical field. Their use across the two registers did not reveal any specific differences, as shown in the following examples.

- (1) *These results question the view that implementing SDM tools for anticoagulant treatment can improve care for patients with AF.* (MRAC_20JAMA10)
- (2) *NPV is suitable for patients with abnormal facial morphologies, excessive oropharyngeal secretions as well as patients who experience anxiety [...].* (MCRC_20OMCR48)

In these examples, the lexical bundle *for patients with* was used as a framing bundle. In example 1, from the MRAC, the bundle is providing a frame regarding who the improved care would benefit. In example 2, from the MCRC, the bundle is used to indicate whom a specific test would be best suited for.

It is necessary to clarify that the number of shared bundle types (122) represented less than a quarter of the total bundles identified in each corpus. Thus, it is important to examine what was not shared, as about 75% of the bundles identified in each corpus belonged to only one of the two registers under study and, even when their structure types overlapped, they

were mostly different expressions. In the following sections, we will focus on the most salient functions performed by bundles exclusive to the MRAC or the MCRC.

4.2.2 Functional Analysis of Lexical Bundles in the MRAC

As explained in the methodology section, we used a framework adapted from Hyland (2008) for the functional classification of bundles in the MRAC. Table 11 shows the distribution of bundles across the main functional categories. Most lexical bundles in this register across all bundle lengths were research-oriented, accounting for 60% of all 3-word expressions, 56% of all 4-word, and 62% of 5-word bundles. Most of the 6-word bundles and the only 8- and 11-word bundles identified in the MRAC were also research-oriented. The rest of the bundles were text-oriented, with a higher number of 4-word bundles in this category, followed by 3-word, 5-word, 6-word, and the only 7-word bundle identified. Participant-oriented and grammatical only bundles were rare in this corpus.

Table 11 – Distribution of MRAC bundles across the major functional categories

Main Functions	MRAC Bundles: Raw Number (%)						
	3-words	4-words	5-words	6-words	7-words	8-words	11-words
Research-Oriented	218 (60.22%)	65 (56.03%)	25 (62.50%)	7 (58.33%)	0%	1 (100%)	1 (100%)
Text-Oriented	143 (39.50%)	51 (43.97%)	14 (35%)	5 (41.67%)	1 (100%)	0%	0%
Participant-Oriented	8 (2.21%)	2 (1.72%)	2 (5%)	0%	0%	0%	0%
Grammatical only (3-words only)	3 (0.83%)						
Total # of bundles	362	116	40	12	1	1	1

Source: the authors.

We further analyzed research- and text-oriented bundles to identify specific functions in the most frequent subcategories. Table 12 shows that the most frequent Research-oriented bundles were mostly descriptive expressions like *the most common*, *the effect of*, *the death from any cause*, *has been associated with*, and *the baseline characteristics of*. The only 11-word bundle (*calculated as weight in kilograms divided by height in meters squared*) was also functioning as a descriptor.

Table 12 – Research-oriented Functions performed by MRAC Bundles

Bundle size	Description	Quantification	Procedure	Location Time/Place
3-word	38.07%	22.02%	29.36%	10.55%
4-word	43.08%	24.62%	21.54%	10.77%
5-word	50%	4%	29%	17%
6-word	43%	--	43%	14%

Source: the authors.

Structurally, research-related bundles were mostly NP-related, PP-related, copular, be, and passive related. Four-word, 5-word, and 6-word bundles were predominantly passive constructions (e.g., *were included in the*, *was defined as a*, *has been associated with*). Even though Millar, Budgell, and Fuller (2013) highlighted the importance of passive constructions to express medical concepts and procedures, passive sequences in the MRAC were also frequently used to explain the use of previous study results when describing the object of the new study or elements related to the new study, as illustrated in examples 3) and 4) below. This way of referring to previous research may be explained by the fact that, following the American Medical Association (AMA) style, cited authors are rarely mentioned in-text in MRAs. Such practice tends to shift the focus from the authors to the topics and elements under investigation, which become the grammatical subjects of the passive constructions.

- (3) *Left ventricular hypertrophy, greater LVMI, or both have been shown to predict CVD outcomes in both observational studies and clinical trials.* (MRAC8JAMA5)
- (4) *Similarly, mania is considered a rare complication of subthalamic nucleus (STN) DBS for Parkinson's disease (46), and stimulation sites that have been associated with mania (47-49) are more connected to our identified mania lesion network than the standard location of stimulation.* (MRAC20JCI7)

The next most frequent function performed by research-related bundles in the MRAC was Procedure. Some examples of bundles functioning in this category are *all participants provided written informed consent*, *were used for*, *were included in*, *was performed using*, *were assigned to*, *were considered statistically significant*, *were excluded from the*, *were enrolled in the*, *was added to the*, *were randomly assigned to receive*, and *were randomly assigned in a*. As shown by these examples, the bundles in this category were mostly passive constructions used to describe medical concepts and methodology procedures, thus lending support to Millar, Budgell, and Fuller's claim regarding the use of the passive constructions in medical writing. Example 5 illustrates such use of bundles with passive constructions.

- (5) *Of these patients, 98% in each molgramostim group and 94% in the placebo group completed the blinded intervention period, and 131 were enrolled in the open-label treatment-extension period.* (MRAC20NEJM3)

The remaining subfunctions, Quantification and Location (in time and place), were performed less frequently and the use of these bundles was consistent with that of bundles in previous studies from the literature (e.g., Biber *et al.*, 1999; Cortes, 2004; Hyland, 2008; Mbodj-Diop, 2016). Some examples of Quantification bundles are expressions like *the number of*, *a total of*, and *the proportion of*, while examples of Location (time/place) bundles are *during the study period*, *at the university of*, and *at the time of*.

In the text-related category, bundles were primarily 3-words and 4-words with very few text-related expressions of 5 and 6 words. Table 13 shows the distribution of 3- and 4-word text-related bundles across the subcategories in the framework. These bundles were majorly Resultative signals, followed by Framing signals.

Table 13 – Text-oriented functions performed by MRAC Bundles

Bundle length	Framing signals	Structuring signals (study subject/elements reference)	Structuring signals (text reference)	Resultative signals	Transition signals
3-words	23.78%	14.69%	16.08%	32.16%	11.19%
4-words	23.53%	9.80%	9.80%	35.2%	21.57%

Source: the authors.

Resultative signals were predominantly clause-related structures consisting of (VP) + *that*-fragment, as in *we observed that, these results demonstrate that, and our data indicate that*). We also found some NP and PP-related structures (e.g., *effect on the, as a result, in response to*) and passive constructions (e.g., *significantly associated with, related to the*). These Resultative signals were used to indicate causative relations between elements being discussed or presented in the studies reported as well as authors’ interpretations of their studies, as illustrated in examples 6 and 7 below.

- (6) *The G6PD c.202T allele is associated with such a mild phenotype that even hemizygous boys and homozygous girls retain 12% of normal G6PD activity and, as a result, are rarely affected by the more severe manifestations of G6PD deficiency [...].* (MRAC15LANCET5)
- (7) *Together, these results indicate that PVcre Syt2fl mice are a reliable genetic animal model for action tremor and a promising candidate to model human essential tremor disorder (Table 1).* (MRAC20JC12).

Framing signals were the second most frequent subfunction among text-related bundles, representing 23.78% of 3-word text-oriented bundles (e.g., *patients in the, t cells in, patients with a*) and 23.57% of 4-words (e.g., *in terms of, in patients who, with respect to*). More than half of the 5-words bundles (e.g., *on the basis of the, on the basis of these, in the context of the, at the discretion of the*), and two of the five 6-word text-oriented bundles (e.g., *in the presence or absence of, according to the manufacturer’s protocol*) were also framing signals.

Structuring signal and Transition bundles were used less frequently than the previous two subcategories. Some examples of Structuring signals were *in the present study, in the supplementary appendix, are shown in table, in our study, in the appendix, are provided in, are shown in, as previously described, and as described above*, among other expressions. The use of these Structuring signals does not differ from what has been reported in previous studies that explain they “provide readers with a cognitive roadmap” (Hyland; Jiang, 2018, p. 18).

Bundles in the Transition signals subcategory were primarily preposition-related sequences, but specific bundle lengths, such as 3-word transition bundles, showed more varied structures. Examples of more varied 3-word bundles included passive-related sequences (e.g., *compared to the, compared with those*) and function words only (e.g., *but not in, than in the*). As shown in previous studies, bundles serving as Transition signals were used to add information or to compare/contrast elements of the study.

We can conclude that many of the functions performed by the bundles in the MRAC overlapped with those reported in previous studies of lexical bundles in published research articles in medicine and in other disciplines. There were bundles specific to the medical field,

but in general, many lexical bundles had discourse functions common to the communicative purposes conveyed in published research articles.

4.2.3 Functional Analysis of Lexical Bundles in MCRC

The functional classification of the lexical bundles identified in the MCRC was conducted using the framework presented in Table 6 in the Methodology section. Table 14 presents the distribution of bundles in this register classified into the major functional categories in the framework.

Table 14 – Distribution of MCRC bundles across the major functional categories

Main Functions	MCRC Bundles: Raw # (%)					
	3-words	4-words	5-words	6-words	7-words	8-words
Case-related	160 (59.26%)	78 (54.93%)	43 (70.49%)	25 (80.65%)	1 (100%)	1 (100%)
Diagnosis & Intervention-related	74 (27.41%)	44 (30.99%)	12 (19.67%)	5 (16.13%)	0%	0%
Discourse Organizers	39 (14.44%)	19 (13.38%)	7 (11.48%)	1 (3.23%)	0%	0%
Grammatical only	10 (3.70%)					
Stance & Engagement	12 (4.44%)	5 (3.52%)	3 (4.92%)	0%	0%	0%

Source: the authors.

As these texts report on medical cases, it is logical that the most frequent function of their lexical bundles of all lengths is Case-related. Some examples of these expressions are *we report a*, *we describe a*, *the nature of*, *is known to* (3 words); *the case of a*, *this case highlights the*, *has been reported in* (4-words); *with a medical history of*, *is the most common site* (5-words); *the most common cause of*, *his medical history was significant for* (6-words); and *this is the first reported case of* (7-words). The second most frequent functional category was Diagnosis and Intervention-related with expressions of different bundle lengths like *differential diagnosis of*, *on physical examination*, *for the diagnosis of*, *she was treated with*, *computer tomography of the*, *the patient was started on*, *with a blood pressure of*, and *a white cell count of*.

Discourse Organizers occurred with less frequency in the MCRC than the previous two categories discussed, but they were still frequent, as they accounted for around 15% of the total 3-word bundles, 14% of the total 4-word bundles and 11% of the 5-words. Bundles expressing stance and engagement as well as 3-word bundles in the grammatical only subcategory, however, were not frequent in the MCRC.

We further analyzed the two most frequently served functional categories (Case-related and Diagnosis/Intervention-related) to better describe the subfunctions served by the bundles identified in the MCRC. We also analyzed Discourse Organizers to compare them with their functional realization with that of similar bundles from the literature.

Table 15 – Case-related bundles in the MCRC across subcategories

Bundle length	Description (medical condition)	Description (subject)	Location (time/ place)	Quantification
3-word	48.13%	16.88%	23.75%	11.25%
4-word	46.15%	21.79%	19.23%	12.82%
5-word	34.88%	44.19%	18.28%	4.65%
6-word	64%	28%	4%	4%

Source: the authors.

Table 15 shows the distribution of Case-related bundles in the MCRC across subcategories. Those Description bundles that were used to describe medical conditions were mostly NPs with embedded *of*-phrases and passive constructions in 3-word bundles (e.g., *the cause of, shortness of breath, is defined as, is caused by*) and passives and copular be expressions in 4-word bundles (e.g., *have been associated with, is thought to be, is more common in, are the most common, is a very rare*). In the case of 5-word bundles, these descriptors were also passive constructions (e.g., *has been reported to be, been shown in the literature*) and clause-related structures with human subjects (e.g., *we report a case of, we present a case of*), while 6-word bundles were generally clause-related bundles with first person plural subjects (e.g., *we report a rare case of, we present a rare case of*). The only 7- and 8-word bundles identified in the MCRC were also clause-related but with abstract subjects (*this is the first reported case of* and *the key to the correct diagnosis is*).

Description bundles used for subjects were mostly clausal sequences with human subjects across bundle lengths (e.g., *the patient had, our patient had a, she had no history of*). The subjects of these expressions often referred to the patient and the verb of these clauses was exclusively *have*. In this functional subcategory, there were also some postmodified NPs that also described the case subjects (e.g., *family history of, woman with a history of*). All these bundles provided details of the case subjects, describing their past medical histories, their current health status, or baseline characteristics like age, sex, or race, as shown in examples 8 and 9 below.

- (8) A 61-year-old **man with a history of hypertension** presented to the emergency department with a 1-day history of fever, dyspnea, and generalized weakness. (MCRC19JAMA149)
- (9) A **woman in her 50s** presented to the emergency department in a comatose condition. (MCRC20JAMA97)

The other two case-related subfunctions in the MCRC, Location (time/place) and Quantification, presented bundle uses very similar to those in the MRAC and to bundles in academic prose in general, as reported in the related literature. The Quantification bundles in MCRC were mostly NPs with embedded *of*-phrase fragments (e.g., *a wide range of, at a dose of, a total of, the majority of*). Location bundles, on the other hand, were mostly PPs with or without embedded *of*-phrases. The bundles in these two categories were used to quantify elements related to a case or to specify time or place, as in examples 10 and 11 respectively.

- (10) *Atrial fibrillation is the most common sustained cardiac arrhythmia and has a wide range of complications including stroke, heart failure, myocardial infarction, sudden cardiac death, chronic kidney disease, cognitive dysfunction, and mortality.* (MCRC19JAMA141)
- (11) *The most common feature of RP at the time of presentation is unilateral or bilateral inflammation of the ear, which is observed in approximately 43% of RP patients.* (MCRC19OMCR149)

We will now turn our attention to the second major functional category, Diagnosis/Intervention-related. Across all bundle lengths, bundles used to report results of diagnosis and intervention procedures were the most frequent ones. The next most frequently served subfunction was Procedure, closely followed in third position by the Decisions/Outcome subcategory. We analyzed all three subcategories because they represent new functions that had previously not been identified in academic writing in general (Table 16).

Table 16 – Distribution of Diagnosis/Intervention Subcategories Served by MCRC Bundles

Bundle lengths	Results Reporting	Procedure	Decisions/Outcome
3-word	40.54%	33.78%	25.68%
4-word	45.45%	29.55%	25%
5-word	41.67%	25%	33.33%

Source: the authors.

Results reporting bundles were mostly clause-related with abstract subjects, external subjects, and *existential there* fragments (e.g., *his blood pressure was*, *vital signs were*, *did not reveal any*, *confirmed the diagnosis of*, *there was no evidence*, and *there were no signs of*). There were also frequent passive constructions in this function, as in *was diagnosed as having*, *was found to have*, *the patient was diagnosed with*. These expressions were used to report physical examination, mostly through clause-related bundles with abstract subjects, as illustrated in example 12 below. Example 13 shows the use of these expressions to report results of tests and other diagnostic procedures, also in clause-related bundles with external fragments or *existential there* fragments.

- (12) *On admission her blood pressure was 140/80 mmHg, breathing 18 breaths per minute, pulse 80 beats per minute, and temperature 36.5 °C.* (MCRC20JMC40)
- (13) *Eye examination revealed signs of conjunctivitis, scleromalacia and mild blepharitis; but there were no signs of corneal injury.* (MCRC20OMCR60)

Bundles in the Procedure subcategory were primarily NPs with embedded *of*-phrases (e.g., *removal of the*, *biopsy of the*, *computed tomography of the*, *MRI of the brain*), PPs with or without embedded *of*-phrases (e.g., *in the treatment of*, *in combination with*, *in association with*, *with the use of*), and clause-related bundles with human subject (e.g., *the patient was started on*, *she was treated with*, *the patient underwent*). NPs were found to be used more frequently to refer to medical procedures during diagnosis (example 14), while PPs and clause-related bundles with human subjects were mostly used for intervention procedures (example 15).

- (14) *Blood tests showed a normal ANA, antiphospholipid screen, full blood count, an ESR of 28 mm/hour and the presence of an atypical ANCA. The ANCA prompted a referral to the rheumatology department. An MRI of the brain showed lesions that involved the deep white matter (Fig.1).* (MCRC19OMCR165)
- (15) *The first line therapy consists of the use of steroids (prednisone 40 mg/daily tapered slowly) in combination with tamoxifen 10 mg twice daily.* (MCR21IMCRJ31)

Decision/Outcome-related bundles were majorly clause-related bundles with human subjects, with the subject almost always being *the patient* (e.g., *the patient was discharged, the patient was referred to*), and clause-related with (VP) + to-fragments (e.g., *we decided to, to confirm the diagnosis, to rule out*).

Finally, we will discuss the use of Discourse Organizers in the MCRC. The subfunctions in this category overlap with those in the Text-oriented category in the classification of bundles in MRAs. In the MCRC, Discourse Organizers were primarily 3-word and 4-word bundles. There were fewer bundles functioning as Discourse Organizers in the MCRC than in the MRAC. The reason for this low number of expressions in this category may result from the short length of case reports, which may force writers to use less discourse organizing markers, resorting to other devices to organize their texts (e.g., report subsections, bullet points). It may also be the case that the use of short and less complex sentences in MCRs requires less discourse organizers.

The most frequent subfunctions of Discourse Organizers in the MCRC were Framing bundles and Resultative signals, illustrated in examples 16 and 17 respectively.

- (16) *High clinical suspicion for Lemierre syndrome in the setting of head and neck infections is important as surgical drainage of collection wherever possible and prolonged antibiotics are necessary.* (MCR19OMCR160)
- (17) *Due to the lack of a proper regimen, our patient is currently taking this injectable-free regimen that is recommended by the WHO for the treatment of pulmonary disease.* (MCR21IMCRJ35)

All in all, lexical bundles in the MCRC performed functions that had not been identified before for lexical bundles in academic registers. The medical case report is a very specialized and not extensively analyzed register, and the communicative purposes conveyed in these texts often call for the use of very specific linguistic exponents.

5 Conclusions

The results of our study showed that the structural characteristics of lexical bundles in MRAs and MCRs are in line with previous descriptions of these expressions in many written academic registers. Structurally, bundles in these registers are mostly NP-related and PP-related. It is important to highlight, however, that passive constructions were found to be very frequent in both the MRAC and the MCRC, in spite of medical journal guidelines recommending using the active voice. The functional analysis showed that lexical bundles in the MRAC were often not different from bundles used in other academic registers, particularly in published research

articles in other disciplines. Bundles in the MCRC, however, performed functions different from those in the MRAC and also from bundles previously identified in studies of academic prose.

While bundles in MRAs were predominantly research-oriented, which is consistent with previous research of academic writing in the hard sciences, given the focus on the description of research procedures and elements, many MCR bundles were used to describe medical pathologies and diagnosis and intervention procedures, which are already slightly different from the description of experimental procedures. The most marked difference between bundles in these two registers, however, was the noticeable proportion of VP-related and clause-related bundles in MCRs, which were used for MCR-specific functions such as reporting results of diagnoses and intervention, describing the patient, and reporting the case outcome and decisions made during case management.

There are some pedagogical implications of the findings of our study that can be directly translated to writing instruction contexts. It is usually understood that disciplinary variation has to be taken into consideration when designing curricula and materials for the teaching of writing in a specific field. We can now suggest that studying variation across specialized registers in medicine may bring about results that can help medical students become effective writers of these specialized texts, as in the case of medical case reports. As we could see, students who may be exposed to the teaching of research articles may lack the specialized formulaic expressions that are necessary to convey the functions and communicative purposes in medical case reports. We can only speculate that this may be the case with other medical registers and perhaps many registers that are often produced and consumed by the academic and professional communities of other fields.

Several pedagogical applications can result from the findings of the study of lexical bundle use across MRAs and MCRs. For example, the use of passive-related bundles. First, it is important to raise students' awareness that using passive voice is common in both of these registers even when the use of the active may be encouraged by journal publishers. Using corpus-based materials such as concordancing lines from real data that show examples of these passive-related bundles and other clausal bundles in the active voice would provide a good discussion point for any discipline specific and register specific writing class.

It is still necessary to explore the description of formulaic language in a register to include more information on the context and co-text in which these sequences often occur. The study of semantic prosodies and preferences for lexical bundles and the connection between lexical bundles and communicative intentions as expressed in moves and steps in specialized discourse across the disciplines could provide more evidence to help lexical bundles become a powerful pedagogical tool.

References

- ANTHONY, L. *AntConc* (Version 3.5.2) [Computer Software]. Tokyo, JP: Waseda University, 2017.
- BAZERMAN, C. *Constructing Experience*. Carbondale, IL: Southern Illinois University Press, 1994.
- BIBER, D.; CONRAD, S. Lexical Bundles in Conversation and Academic Prose. In: HASSELGARD, H.; OKSEFJELL, S. (ed.). *Out of Corpora: Studies in Honor of Stig Johansson*. Amsterdam: Rodopi, 1999. p. 181-189.

- BIBER, D.; CONRAD, S. Register Variation: A Corpus Approach. In: SCHIFFRIN, D.; TANNEN, D.; HAMILTON, H. (ed.). *The Handbook of Discourse Analysis*. London: Blackwell, 2005. p. 175-196.
- BIBER, D.; CONRAD, S. *Register, Genre, and Style*. Cambridge: Cambridge University Press, 2019.
- BIBER, D.; CONRAD, S.; CORTES, V. 'If You Look at ...': Lexical Bundles in University Teaching and Textbooks. *Applied Linguistics*, v. 25, p. 371-405, 2004.
- BIBER, D.; GRAY, B. *Grammatical Complexity in Academic English: Linguistic Change in Writing*. Cambridge: Cambridge University Press, 2016.
- BIBER, D.; JOHANSSON, S.; LEECH, G.; CONRAD, S.; FINEGAN, E. *Longman Grammar of Spoken and Written English*. London: Longman, 1999.
- CHEN, L. *Lexical Bundles in Vocabulary-Based Discourse Units: A Corpus-Based Study of First Year Core Engineering Textbooks*. 2018. Dissertation (Doctorate in Philosophy) – Carleton University, Ottawa, 2018. Unpublished.
- CHEN, Y. H.; BAKER, P. Lexical Bundles in L1 and L2 Academic Writing. *Language Learning & Technology*, v. 14, n. 2, p. 30-49, 2010.
- CORTES, V. Lexical Bundles in Academic Writing. In: JABLONKAI, R.; CSOMAY, E. (ed.). *The Routledge Handbook of Corpora in Language Teaching and Learning*. London: Routledge, 2023. p. 220-233.
- CORTES, V. Lexical Bundles in Published and Student Disciplinary Writing: Examples from History and Biology. *English for Specific Purposes*, v. 23, p. 397-423, 2004.
- CORTES, V. Situating Lexical Bundles in the Formulaic Language Spectrum: Origins and Functional Analysis Developments. In: CORTES, V.; CSOMAY, E. (ed.). *Corpus Linguistics in Applied Linguistics: Studies in Honor of Doug Biber*. Amsterdam; Philadelphia: John Benjamins, 2015. p. 197-216.
- CORTES, V. The Purpose of This Study is to: Connecting Lexical Bundles and Moves in Research Article Introductions. *Journal of English for Academic Purposes*, v. 12, p. 33-43, 2013.
- CORTES, V.; LAKE, W. LBiaP: A Solution to the Problem of Attaining Observation Independence in Lexical Bundle Studies. *International Journal of Corpus Linguistics*, v. 28, n. 2, p. 263-277, 2023.
- DORGELOH, H. The Interrelationship of Register and Genre in Medical Discourse. In: SCHUBERT, C.; SANCHEZ-STOCKHAMMER, C. (ed.). *Variational Text Linguistics: Revising Register in English*. Berlin: De Gruyter, 2016. p. 43-65.
- GAGNIER, J. J.; KIENLE, G.; ALTMAN, D. G.; MOHER, D.; SOX, H.; RILEY, D.; CARE GROUP. The CARE Guidelines: Consensus-Based Clinical Case Reporting Guideline Development. *BMJ Case Reports*, v. 2, n. 5, p. 38-43, 2013.
- GRAY, B. *Exploring Disciplinary Writing Through Corpus Linguistics: When Discipline Tells Only Part of the Story*. 2014. Dissertation (Doctorate in Applied Linguistics) – Northern Arizona University, Flagstaff, 2014. Unpublished.
- GRAY, B. *Linguistic Variation in Research Articles: When Discipline Tells Only Part of the Story*. Amsterdam; Philadelphia: John Benjamins, 2015.
- HYLAND, K. As Can Be Seen: Lexical Bundles and Disciplinary Variation. *English for Specific Purposes*, v. 27, p. 4-21, 2008.

HYLAND, K.; JIANG, F. Academic Lexical Bundles: How Are They Changing? *International Journal of Corpus Linguistics*, v. 23, n. 4, p. 383-407, 2018.

JALALI, Z.; MOINI, M. Structure of Lexical Bundles in Introduction Section of Medical Research Articles. *Procedia – Social and Behavioral Sciences*, v. 98, p. 719-726, 2014.

LAKE, W.; CORTES, V. Lexical Bundles as Reflections of Disciplinary Norms in Spanish and English Literary Criticism, History and Psychology. In: ROMER, U.; CORTES, V.; FRIGINAL, E. (ed.). *Advances in Corpus-Based Research on Academic Writing*. Amsterdam: John Benjamins, 2020. p. 184-203.

LIN, K.; MWINLAARU, I.; TAY, D. Introduction: Genre Research Into the 21st Century – Goals, Approaches, and Applications. In: LIN, K.; MWINLAARU, I.; TAY, D. (ed.). *Approaches to Specialized Genres*. Abingdon; New York: Routledge, 2021. p. 1.

LU, X.; DENG, J. 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*, v. 39, p. 21-36, 2019.

MBODJ, N. B. *Writing in the Disciplines and Within-Discipline Variations: A Comparison of the Formulaic Profiles of the Medical Research Article and the Medical Case Report*. 2021. Dissertation (Doctorate in Applied Linguistics) – Georgia State University, Atlanta, 2021. Unpublished. DOI: <https://doi.org/10.57709/26822844>.

MBODJ-DIOP, N. B. *Lexical Bundles in Medical Research Articles: Structures and Functions*. 2016. Thesis (Master of Arts) – Michigan State University, Michigan, 2016.

MILLAR, N.; BUDGEELL, B.; FULLER, K. “Use the Active Voice Whenever Possible”: The Impact of Style Guidelines in Medical Journals. *Applied Linguistics*, v. 34, n. 4, p. 393-414, 2013. DOI: <https://doi.org/10.1093/applin/ams059>.

NATTINGER, J.; DECARRICO, J. *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press, 1993.

NWOGU, K. N. The Medical Research Paper: Structure and Functions. *English for Specific Purposes*, v. 16, n. 2, p. 119-138, 1997.

PORTER, J. Intertextuality and the Discourse Community. *Rhetoric Review*, v. 5, p. 34-47, 1986.

REPPEN, R.; OLSON, S. Lexical Bundles Across Disciplines: A Look at Consistency and Variability. In: RÖMER, U.; CORTES, V.; FRIGINAL, E. (ed.). *Advancements in Corpus-Based Research on Academic Writing*. Amsterdam; New York: John Benjamins, 2020. p. 169-182.

SHAHRIARI, H. Comparing Lexical Bundles Across the Introduction, Method and Results Sections of the Research Article. *Corpora*, v. 12, p. 1-22, 2017.

SHIN, Y. Evaluative Prosody and Semantic Preference: Extending the Analysis of Recurrent Multiword Sequences. *English for Specific Purposes*, v. 59, p. 48-58, 2019.

SWALES, J. *Research Genres*. Cambridge: Cambridge University Press, 2004.

TOGNINI-BONELLI, E. *Corpus Linguistics at Work*. Amsterdam: John Benjamins, 2001.

WRIGHT, H. Lexical Bundles in Stand-Alone Literature Reviews: Sections, Frequencies, and Functions. *English for Specific Purposes*, v. 54, p. 1-14, 2019.

History

Date of submission: 16/11/2024.

Date of approval: 31/03/2025.

Responsible Editor

Andréa Machado de Almeida Mattos, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Minas Gerais/MG, Brasil. Lattes: <http://lattes.cnpq.br/7749222257907067>, ORCID: <https://orcid.org/0000-0003-3190-7329>, e-mail: andreamattos@ufmg.br.

Authors' Contribution

Ndeye Bineta Mbodj: conceptualization, design, data acquisition, analysis, interpretation.

Viviana Cortes: conceptualization, design, analysis, writing.

Declaration of Conflict of Interest

The authors declare that there is no conflict of interest.

Data Availability Statement

The data is not public but it may be available upon request. Please contact the authors.

Use of AI

AI was not used in any section or process in this article.

Reviews

As part of the commitment made by the Brazilian Journal of Applied Linguistics to Open Science, the journal publishes the reviews issued regarding its published works, when authorized by all parties involved.