12-13-2021

Writing in the Disciplines and Within-discipline Variations: A Comparison of the Formulaic Profiles of the Medical Research Article and the Medical Case Report

Ndeye Bineta Mbodj
Georgia State University

Writing in the Disciplines and Within-discipline Variations: A Comparison of the Formulaic Profiles of the Medical Research Article and the Medical Case Report

by

Ndeye Bineta Mbodj

Under the Direction of Viviana Cortes, PhD

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2021

ABSTRACT

Research of formulaic language in academic writing has primarily investigated the use of single types of formulaic sequences in academic research articles in various disciplines. Studies in this line of research have revealed dramatic variations in the use of formulaic language across academic disciplines (e.g., Cortes, 2004; Hyland, 2008a; Jalali & Moini, 2014; Shahriari, 2017). However, there is evidence that discipline alone does not tell the whole story about linguistic variation (Gray, 2015). Different varieties of texts within one discipline may reflect different linguistic characteristics depending on specific communicative purposes (Biber & Conrad, 2009). It follows that the almost exclusive focus on the academic research article may "limit our knowledge of the discourse practices within discipline" (Gray, 2015, p. 19). Moreover, formulaic language encompasses different types of sequences (e.g., collocations, lexical bundles, frames, etc.) each of which only reveals a partial picture of formulaicity in discourse (Wray, 2005). Thus, studies that investigate the use of single types of formulaic sequences may provide only partial descriptions of the registers they investigate. Therefore, to better serve disciplinary writing instruction, there is a need for studies that provide more comprehensive descriptions of formulaic language in various registers within one discipline.

The present dissertation takes a step in that direction by investigating within-discipline linguistic variation through the comparison of the formulaic profiles of two registers in the field of medicine: the medical research article (MRA) and the medical case report (MCR). These two registers that have both been reported in the medical literature to contribute to advancing research, clinical practice, and education in the field (e.g., Man et al., 2004; Rison et al., 2017). The study proposes a more comprehensive approach to the description of formulaic language and investigates the use of various formulaic sequences that have been described as accounting for the formulaicity of discourse. Such sequences include: (a) *collocations*, pairs of words that tend to co-occur, (b) *multiword collocations*, sequences of three or more words with strong mutual attraction (such sequences consist primarily of lexical words, most of which are technical terms), (c) *lexical bundles,* most frequent sequences of three or more words in a register, described as the building blocks of academic writing (Cortes, 2013), and (d) *frames,* sequences of three or more items with one variable slot. Frames have been described as allowing writers to make more creative use of formulaic language (e.g., Biber, 2009; Gray & Biber, 2013).

The analyses of the formulaic sequences in the two registers often revealed structural similarities but noticeable variations in terms of the discourse functions of the sequences. Such variations reflect the differences in the situational characteristics of the two registers such as communicative purposes, nature of data and evidence, textual organization, to name but a few. The findings of the present study portray MRAs and MCRs as two distinct registers, thus highlighting the importance of investing within-discipline variations to better serve disciplinary writing instruction.

INDEX WORDS: Disciplinary writing, Formulaic language, Linguistic variation, Collocations, Lexical bundles, Frames

Writing in the Disciplines and Within-discipline Variations: A Comparison of the Formulaic Profiles of the Medical Research Article and the Medical Case Report

by

Ndeye Bineta Mbodj

Committee Chair: Viviana Cortes

Committee: Diane Belcher

Youjin Kim

Bethany Gray

Electronic Version Approved: December 13, 2021

Office of Graduate Services

College of Arts and Sciences

Georgia State University

December 2021

DEDICATION

To Mum, Dad, and Papa Aly Ndiaye,

> *Ceux qui sont morts ne sont jamais partis :*
>
> *Ils sont dans l'Ombre qui s'éclaire*
>
> *Et dans l'ombre qui s'épaissit* (Birago Diop, 1960)

Yes, these verses by Birago Diop, 61 years ago, will always hold true. I know from where you are, you are still watching over me, and I hope you are proud of me. You are forever in my heart.

ACKNOWLEDGEMENTS

I would like to express my gratitude to all my professors at the Department of Applied Linguistics and ESL at Georgia State University. Thank you for the knowledge imparted and for the always valuable advice. This truly has been a rewarding experience.

I am forever grateful to Dr. Viviana Cortes who has been much more than an academic advisor to me. I have learned so much both as your student and research assistant. I could never thank you enough not only for your time, guidance, and patience, but also for your compassion and everlasting support in difficult moments. My thanks go also to all members of my dissertation committee, Dr. Diane Belcher, Dr. Youjin Kim, and Dr. Bethany Gray. Thank you for being always available when I needed your help and advice. I feel fortunate to have had the opportunity to work each of you.

I would also like to thank my field-expert informants for the valuable insights and thorough answers to all my questions. My heartfelt thanks go also to my cohort for the mutual support and to all my fellow PhD students. I am thankful for getting to know each of you.

Last but not least, my loving husband! Thanks for believing in me and for being by my side every step of the way. This journey became a million times easier when you entered my life. I love you so much.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AdjP:	Adjective Phrase
AdvP	Adverb Phrase
BMJ:	*British Medical Journal*
ESP:	English for Specific Purposes
EAP:	English for Academic Purpose
IMCRJ:	*International Medical Case Reports Journal*
JAMA:	*Journal of the American Medical Association*
JCI:	*Journal of Clinical Investigation*
JMCR:	*Journal of Medical Case Reports*
L2:	Second Language
MCRs:	Medical Case Reports
MCRC:	Medical Case Report Corpus
MRAs:	Medical Research Articles
MRAC:	Medical Research Article Corpus
NEJM:	*New England Medical Journal*
N:	Noun
NP:	Noun Phrase
OMCR:	*Oxford Medical Case Repors*
PP:	Prepositional Phrase
STM:	*Science Translational Medicine*
TTR:	Type/Token Ratio
VP:	Verb Phrase

1 CHAPTER 1: INTRODUCTION

Formulaic language has been of much interest in the research of academic writing for decades. Such interest can be explained by evidence in the literature that academic writing is highly formulaic (e.g., Biber et al., 1999; Biber et al., 2004) and therefore, formulaic sequences are considered useful tools in the comprehension and production of academic texts (Biber & Barbieri, 2007). Research of formulaic language in academic writing can thus be considered instrumental to the description of written texts in academia and ultimately, can serve as a basis for corpus-based instruction of academic writing (Gray, 2015).

Studies of formulaic language in academic writing have investigated various types of formulaic sequences in various academic disciplines and have indeed contributed to our understanding of how discourse is constructed in those disciplines (e.g., Cortes, 2004; Cunningham, 2017; Grabowski, 2013; 2015; Hyland, 2008a; 2008b; Kanosksilapatham, 2015; Lu et al., 2017; Nekrasova-Beker, 2019; Shahriari, 2017). Some of these studies compared the use of formulaic sequences in different disciplines and revealed some dramatic variations between disciplines (e.g., Cortes, 2004, Hyland, 2008a; 2008b; Kanosksilapatham, 2015; Shahriari, 2017). Such studies have underscored the importance of considering each discipline in its own terms if learners are to be familiarized with writing practices in their disciplines.

There is evidence, however, that there is more to linguistic variation than discipline alone can account for (Gray, 2015). Different varieties of texts within one discipline may reflect different linguistic characteristics depending on specific communicative purposes and situational contexts (Biber & Conrad, 2009). This warrants research that investigates and describes the use of formulaic language in different academic texts that serve as channels for sharing knowledge within academic disciplines. Findings of such research would certainly be beneficial to disciplinary writing instruction. The present study takes a step in that direction by providing a description of two varieties of text in the field of medicine: the medical research article (MRA) and the medical case report (MCR) from the perspective of formulaic language use. The purpose of MRAs, according to Phillips et al. (1991), is to convey knowledge useful to physicians in their practice, the scientific community working to advance the field, and the public who need information about their medical conditions. Nwogu (1997, p.119) defines the medical research article as a highly technical report of experimental research, typically presented in the well-known IMRD (Introduction, Methods, Results, and Conclusion) format. On the other hand, the medical case report is defined by Gagnier et al. (2013) as “a detailed narrative that describes, for medical, scientific, or educational purposes, a medical problem experienced by one or several patients” (p. 1). Despite the apparent similarities in purpose, these definitions foretell potential linguistic variations that warrant thorough description of each of the two registers. But before going further into the aims of the present study, it is necessary to define the term *register* and the perspective I take on the analysis of formulaic language in MRAs and MCRs.

1.1 Definition of Register

Both *register* and *genre* have been used in the literature to refer to text varieties and have sometimes been used interchangeably. Biber & Conrad (2019) distinguish between *register* and *genre* perspectives on text varieties is these terms:

> The *register* perspective combines an analysis of linguistic characteristics that are common in a text variety with analysis of the situation of use of the variety. […] In contrast, the *genre* perspective focuses on conventional structures used to construct a complete text within the variety (for example, the conventional way in which a letter begins and ends) (p. 2).

The authors further explain that linguistic features described from a register perspective are always functional as they are adapted to the communication purposes and the situational context of the register. This is the perspective I take on the analysis of the formulaic language in MRAs and MCRs. The analysis of the formulaic sequences in the present study is informed by the communicative purposes and the situational contexts of each of the two registers. For example, based on the definition of MRAs just provided, it can be expected that formulaic language in this register will include sequences that are specifically related to reporting experimental results in the medical field such as statistical measures, statistical significance, experimental procedures, intervention groups, etc. On the other hand, it can be assumed that such sequences are less likely to occur in a "detailed narrative" of a medical pathology experienced by a single patient.

1.2 Purpose of the Study and Research Questions

The purpose of the present study is two-fold. First, by investigating the formulaic profiles of MRAs and MCRs, the study aims to add to disciplinary writing research by describing potential linguistic variations that can occur within a same discipline. Taking such variations into account can highly contribute to effective disciplinary writing instruction. The choice of MCRs in addition to MRAs is not random. Research of formulaic language in medical writing, has primarily focused on the IMRD research article. Such focus on MRAs can partly be explained by the importance of this register in the medical field. According to Man et al. (2004), the medical research article is not only the most commonly used vehicle to convey new knowledge but also the "principal currency for academic recognition and promotion" (p. 811). However, next to medical research articles, case reports now "account for a growing number of articles in medical journals" (Gagnier et al., 2013; p. 1). A number of articles in the medical field have been published to expressly encourage the publication of MCRs given the contribution of such reports in the field (e.g., Florek & Dellavalle, 2016; Gagnier et al., 2013; 2014; Green & Johnson, 2006; Rison et al., 2017). Key contributions of MCRs listed by Florek & Dellavalle (2016) include, but are not limited to, providing first line of evidence that often leads to new randomized clinical trials, serving as major sources of detecting rare adverse events of treatments and side effects due to drug interactions in clinical practice, and preparing authors and medical students for a scientific career by adding publications to their resumes. Similar highlights of the importance of MCRs can also be found on websites of journals that publish case reports.

Yet, despite this consensus on the importance of MCRs in the medical field, there is a dearth of research in academic writing that describes this register. To my knowledge, only two studies, Goudier (2008) and Helan (2012), looked at MCRs, and both studies used a genre approach describing the rhetorical moves in case reports. Therefore, as mentioned above, with the description of the linguistic and situational characteristics of both MRAs and MCRs, two of the most important registers in the medical field, the present study can add to our understanding of within-discipline linguistic variations and make a valuable contribution to medical writing instruction.

The second purpose of the present study is to propose a more comprehensive approach to the description of formulaic language in a register. As already mentioned, research of formulaic language in academic writing has greatly contributed to our understanding of how discourse is constructed in various academic disciplines by describing key sequences that have been shown to account for the formulaicity of discourse. Such studies have primarily focused on three types of sequences: *collocations* (e.g., Flowerdew & Forest, 2009; Gledhill, 2000; Krummes & Ensslin, 2015; Marco, 2000), *lexical bundles* (e.g., Biber et al. 2004; Cortes, 2004; 2006; 2008; 2013;

Grabowski, 2015; Hyland, 2008a; 2008b), and *lexical frames* (e.g., Biber 2009; Garner, 2016; Gray & Biber, 2013; Römer, 2010).

Collocations have been defined differently in the literature depending on research focus, but whatever the focus, all definitions concur to the notion of frequent co-occurrence of words with strong mutual attraction (Conzett, 2000). *Lexical bundles,* defined by Biber et al., (2004, p. 376) as “the most frequent recurring lexical sequences in a register” have also been described by Cortes (2013) as the building blocks of academic writing. Finally, *frames* are multiword sequences with one variable slot (e.g., *the * of the, to * the effect of; can be * in*). These discontinuous sequences have been reported to allow writers to make more creative use of formulaic language (e.g., Biber, 2009; Cunningham, 2017; Gray & Biber, 2013; Lu et al., 2018).

Taken individually, each of these sequences provides only a partial description of formulaic language in a given register. As rightly noted by Wray (2005), formulaic language encompasses various types of sequences, each of which “has something useful to say” but does not fully “capture the essence of the wider whole” (p.8). The present study differs from most previous studies of formulaic language in that it sets out to investigate the formulaic profiles of MRAs and MCRs, in an aim to provide a closer description of the “wider whole” (i.e., formulaic language) in each of the two registers. In addition to key formulaic sequences mentioned above, I also investigate a less frequently studied type of sequence: *multiword collocations*: sequences of three or more words primarily composed of lexical words with strong collocational strength. Such sequences often consist of multiword technical terms (Biber, 2009). This makes multiword collocations of particular relevance to the description of medical texts as, according to Rezeaian (2015), one of the main challenges in medical writing lies in the appropriate use of conventionalized terms that exist for a range of concepts in the field.

In sum, in an aim to better serve disciplinary writing instruction, the two main purposes of the present study are (a) to investigate within-discipline linguistic variation and (b) to propose a more comprehensive approach to the investigation of formulaic language in a register. Therefore, the present study sets out to answer the following questions.

- Research Question 1: What are the situational characteristics that distinguish MCRs from MRAs?
- Research Question 2a: What collocations are used in MRAs and MCRs? Are they register-specific or rather shared within the same discipline?
- Research Question 2b: What multiword collocations are used in the two registers? How do they compare in terms of structures and functions? Are they register-specific or rather shared within the same discipline?
- Research Questions 2c: How are Lexical Bundles used in both registers? How do they compare to bundles previously identified in academic prose? Are there any similarities and/or differences in terms of length, structure, and function?
- Research Question 3a: How does the use of phrase frames compare in the two registers? Are there any variations in terms of predictability, variability, and structures?
- Research Question 3b How can a grouping by semantic domains of fillers inform the functional analysis of phrase frames? What are the main functions served by some of the most salient frames and their fillers in MRAs and MCRs? Are there any variations between the two registers?

With the pedagogical implications and applications of the findings of the present study in mind, it is important to mention that the focus of the comparison of the use of formulaic language in the two registers is on the structures and functions of formulaic sequences rather than on frequency or on the difference in number of types or tokens of formulaic language across the two registers. I answered each question above by providing a thorough description of the use of the formulaic sequences in each register before discussing the observed similarities and differences.

1.3 Organization of the Dissertation

After this introduction, the rest of the paper is organized as follows. In chapter 2, I provide a review of the literature relevant to the objects of the present study. Then in chapter 3, I present the two corpora collected to represent the two registers under study and explain the methodology I used to answer the research questions just posed. Chapter 4 answers RQ 1 with a description of the situational characteristics of MRAs and MCRs, followed by a discussion of the similarities and differences between the two registers. Then chapters 5, 6, and 7 answer the rest of the research questions, all related to the use of formulaic sequences in the two registers. Chapter 5 presents the findings of the analyses of collocations and multiword collocations in MRAs and MCRs as well as observed similarities and differences in the structures and functions of these sequences in the two registers. Chapter 6 describes the use of lexical bundles in each of the two registers and discusses observed variations. Then chapter 7 answers RQ 3a and RQ 3b, related to the use of frames in the two registers. Finally, chapter 8 is the conclusion of the present study. In this chapter, I provide a summary of the findings of and some relevant methodological and pedagogical implications and applications.

2 CHAPTER 2: LITERATURE REVIEW

In this chapter, I provide an overview of previous research relevant to the description of formulaic sequences investigated and the methodologies used in the present study. As the study focuses on the analysis of formulaic sequences in two potentially distinct medical written registers, I also provide a brief review of previous studies of medical writing. To that end, the rest of the chapter is organized as follows. Section 2.1. reviews the guidelines for conducting a situational analysis of a register, the logical starting point of register analysis. Section 2.2 provides an overview of studies that have described the use of formulaic sequences investigated in the present study (i.e., collocations, multiword collocations, lexical bundles, and lexical frames) in academic writing. Finally, section 2.3. provides a brief review of studies that have contributed to the description of the use of formulaic sequences in the medical writing.

2.1 Register Situational Analysis

Situational analysis, described by Biber & Conrad (2009) as the identification and description of the characteristics of use (or situational characteristics) of registers, is a key step in any register approach on text varieties. Indeed, the analysis of the situational characteristics of a register is essential for both the collection of a representative corpus and the interpretation of the linguistic characteristics of the register, as already explained in section 1.1. of the introductory chapter. According to Biber & Conrad (2009), the situational analysis of any register is inherently comparative as it is "virtually impossible" to provide a thorough description of the characteristics of a register without comparing it to other registers (p. 36). In other words, it is the comparison between two or more registers that highlights the distinctive characteristics of each register.

In the process of describing register situational characteristics, there are several sources the researcher can draw from. Biber & Conrad (2009) suggest the following four sources: (1) the researcher's personal experience and observation, (2) expert informants who can provide invaluable insider information, (3) previous research on the register being investigated, and (4) analysis of sample texts from the register under study. The authors also suggest a framework that includes seven major characteristic categories to consider in any register situational analysis. Those categories of situational characteristics are shown in Figure 2.1. While the framework shown in Figure 2.1 may be too broad for the comparison of more specialized registers, it can serve as a starting point for the development of other more specific frameworks. Indeed, Biber & Conrad (2009) acknowledge that "particular situational characteristics will be more or less important depending on the registers that are being compared" (p. 37).

Drawing from Biber & Conrad (2009) and other previous frameworks developed for the situational analyses of various registers (e.g., Biber, 1994; Conrad, 1996), Gray (2015) developed a framework that is of relevance to the study of within-discipline linguistic variations. Her framework was designed to compare academic research articles in various disciplines, including different subregisters within the same disciplines (e.g., quantitative and qualitative research articles in political science). That framework, presented in the next chapter on Table 3.3., includes eight categories of situational characteristics, namely, *Participants*, *Layout and Organization, Setting, Subject/topic, Purpose, Nature of Data or Evidence, Methodology,* and *Explicitness of Research Design*. Using that framework, Gray established the situational characteristics of qualitative, quantitative, and theoretical research articles in six disciplines including Philosophy, History, Political Science, Applied Linguistics, Biology, and Physics. Her analysis revealed variations in the situational characteristics of research articles, both between and within disciplines, which led to the conclusion that the academic research article is not a monolithic register but rather comprises a variety of subregisters. The frameworks used for the situational analyses of MRAs and MCRs in the present study are primarily based on Gray 's (2015) and Biber & Conrad's (2009) frameworks.

Table 2.1 *Situational characteristics of registers and genres*

I. Participants
- A. Addressor(s) (i.e. speaker or author)
  - 1. single / plural / institutional / unidentified
  - 2. social characteristics: age, education, profession, etc.
- B. Addressees
  - 1. single / plural / un-enumerated
  - 2. self / other
- C. Are there on-lookers?

II. Relations among participants
- A. Interactiveness
- B. Social roles: relative status or power
- C. Personal relationship: e.g., friends, colleagues, strangers
- D. Shared knowledge: personal and specialist

III. Channel
- A. Mode: speech / writing / signing
- B. Specific Medium:
  Permanent: taped / transcribed / printed / handwritten / e-mail / etc.
  Transient speech: face-to-face / telephone / radio / TV / etc.

IV. Production circumstances: real time / planned / scripted / revised and edited

V. Setting
- A. Is the time and place of communication shared by participants?
- B. Place of communication
  - 1. Private / public
  - 2. Specific setting
- C. Time: contemporary, historical time period

VI. Communicative purposes
- A. General purposes: narrate / report, describe, exposit / inform / explain, persuade, how-to / procedural, entertain, edify, reveal self
- B. Specific purposes: e.g., summarize information from numerous sources, describe methods, present new research findings, teach moral through personal story
- C. Factuality: factual, opinion, speculative, imaginative
- D. Expression of stance: epistemic, attitudinal, no overt stance

VII. Topic
- A. General topical "domain": e.g., domestic, daily activities, business / workplace, science, education / academic, government / legal / politics, religion, sports, art / entertainment, etc.
- B. Specific topic
- C. Social status of person being referred to

Figure 2.1. Categories of Characteristics to Consider in Register Situational Analysis (as Proposed by Biber & Conrad, 2009, p.40)

2.2 Formulaic Language in Academic Writing

That language is highly formulaic is now well-established in the field of Applied Linguistics. Several studies have demonstrated that language users follow multiple patterns of word co-selection in producing written and/or spoken texts (e.g., Biber et al., 1999; Erman & Warren, 2000; Sinclair, 1991). Research of formulaic language in academic writing has shown that the use of formulas varies from one discipline to the other and even between registers within the same discipline (e.g., Cortes, 2004; Grabowski, 2013; 2015; Hyland 2008a, 2008b). This section provides an overview of studies that have investigated the use of various formulaic sequences that play important roles in the construction of academic written texts. It focuses on four types of sequences that are investigated in the present study, namely, collocations, multiword collocations, lexical bundles, and lexical frames.

2.2.1 Collocations

The notion of collocation can be traced back to Firth (1957) and his now famous statement "you shall know a word by the company it keeps" (p.11). Sinclair (1991) explained that this tendency of a given word to prefer the company of certain words is guided by two principles: *the open-choice principle* and *the idiom principle.* The open-choice principle suggests that in language production, language users have a wide variety of words to choose from and that choice is constrained only by basic syntactic restrictions, more like the "open slot and fillers" view of language. However, *open-choice* alone cannot account for how meaning arises in a text. Sinclair (1991) contends that the co-occurrence of words in a text is not random and for the construction of meaning, there is the *idiom principle* that puts restrictions to the *open-choice principle* by limiting the extent of the company that a word can keep (p.112). The *idiom-principle* is less concerned with individual words that fill each slot in the syntactic structure and is more about the syntagmatic attraction between one slot filler and the subsequent ones. Sinclair explains that, as a result of that attraction, there are numerous semi-preconstructed phrases that constitute single choices that are available to language users. This is what accounts for formulaicity in language and explains why *powerful coffee* or *heavy coffee,* for instance, would sound less natural than *strong coffee.*

Collocations can be identified empirically and automatically in a collection of texts and there are certain parameters to consider in their identification. Such parameters include, but are not limited to, *frequency of co-occurrence*, the *span* or *window of collocation*; that is the number of relevant lexical items (the *collocates*) on each side of the central word under study (the *node*), and *mutual expectancy*; that is, how glued together the node and the collocates are. There exist corpus tools and software programs like *AntConc* (Anthony, 2017, 2020) or *Collocate* (Barlow, 2004) that automatically identify most frequent collocates with various spans and compute the mutual expectancy between each collocate and the node. Sinclair (2004) suggested a span of up to four words on either side of the node for the study of collocations; but some studies have used a span of ± 5 words, and current corpus tools like *AntConc* allow searches for up to 20 words on each side of the node.

Indeed, in the study of collocation, the span is primarily determined by the object of the investigation (Brezina, McEnery, & Wattam, 2015). An investigation of adjectives that frequently premodify the noun *moon*, for example, might require no more than a span of *– 3L,* that is, only words to the left of the node will be relevant, since the focus is on premodification. As rightly noted by Nattinger and DeCarrico (1992, p.22), smaller windows of collocation will reveal fixed expressions (e.g., ± 1 on either side) and other “short-range relations” like the example of noun premodification above. This is one consideration I took into account for the search for “specialized terms” in the medical field.

Regarding mutual expectancy, it indicates to what extent a collocation is fixed and idiomatic (Nattinger & DeCarrico, 1992). Mutual expectancy can be assessed by means of a statistical measure of cohesiveness such as Mutual Information (MI), for example. This measure is incorporated in most corpus tools that compute MI scores to indicate the probability of the words in the collocation to co-occur. The higher the MI score, the greater consistency there is between the constituents of the collocation. However, researchers have cautioned about the use of MI as a measure of mutual expectancy because of its tendency to favor low frequency words (Biber, 2009; Gablasova et al., 2017). Gries (2013) added that MI and many other measures of collocational strengths are symmetric measures that provide scores that do not reflect directional attraction of collocates. That is, these measures do not reflect whether $word_1$ and $word_2$ have a mutually attraction, or whether $word_1$ is a stronger predictor of $word_2$ or vice versa. This author proposed the use of *delta P*, a directional measure that uses conditional probability to indicate both directional and bidirectional associations of elements of a collocation. In other words, *delta P* can provide information that others cannot, that is, “it can tease apart which collocates in a collocation exhibit the strongest or weakest amounts of attraction or repulsion to the other collocate(s)” (p. 152). According to Gries, the combination of *delta P* with other measures such as dispersion and/or frequency can allow a better description of formulaic sequences like collocations.

With so many characteristics, collocations have been operationalized differently depending on which aspects researchers have focused on. The definition proposed by Nattinger & DeCarrico (1992) is of relevance to the present study. They posit that collocations are “strings of specific lexical items, such as *rancid butter* and *curry flavor*, that co-occur with mutual expectancy greater than chance" (p. 36). One other characteristic of collocations highlighted in their definition is the word class of the constituents that form the collocation. Their definition limits it to lexical words (i.e., nouns, verbs, adverbs, and adjectives) but as rightly suggested by Howarth (1998), while the node is typically a lexical word (e.g., *weather*), its collocates can be either lexical words as in *nice weather* or function/grammatical words, as in *under the weather*. Nevertheless, the core, in both cases remains the lexical word. Nattinger's & DeCarrico's (1992) definition, as well as Howarth's (1998) observation have informed the working definition of collocation in the present study.

While collocational studies have mostly been corpus driven (Biber, 2009), most studies of collocations in academic writing have been corpus-based or rather hybrid. That is, the researcher starts out with a predetermined set of lexical items and/or linguistic features that have been shown to be related and/or relevant to a given register and then investigates their collocational patterns in a collection of texts representing that register (e.g., Gledhill, 2000; Marco, 2000; Cunningham, 2017; Peacock, 2012). Gledhill (2000), for example looked at what he termed as "the phraseology" of Introduction sections of *Cancer* articles and did so by investigating the discourse functions of collocations comprising the verb forms *has, have, been,* and *is,* and the prepositions *of* and *to*. His analysis demonstrated that some patterns such as *X* ***is*** *Y* (e.g., *resistance to therapy* ***is*** *a critical parameter*) were specific to the introduction as they were linked to the explanatory function of this section of the research article. Peacock (2012) investigated the high frequency collocates of abstract nouns in research articles published in eight disciplines: Chemistry, Computer Science, Materials Science, Neuroscience, Economics, Language and Linguistics, Management, and Psychology. His analysis revealed that the same nodes varied across disciplines, thus producing discipline-specific terminologies such as:

- *thermal analysis*, *reaction analysis* (Materials Science), *discourse analysis*, *genre analysis* (Language and Linguistics)
- *user model* (Computer Science) *parallel model* (Psychology), *memory model* (Neuroscience), *profit model* (Economics)
- *corrosion process* (Materials Science), *cognitive process* (Neuroscience)
- etc.

Other studies used a more exploratory approach to the investigation of collocations in academic writing. Flowerdew & Forest (2009), for example, used a corpus driven approach to identify key words and key keywords (i.e., most widely distributed keywords) in their corpus of PhD Literature Review chapters in Applied Linguistics. They then analyzed the collocational patterns of the identified key keywords in their corpus.

Each of these studies revealed salient patterns in the registers they investigated. What should be noted here is that multiple approaches to the study of collocations can be used depending mostly on the research focus. The proposed study is no exception to the rule and the approach described later in chapter 3, section 3.3. is primarily guided by the goal of investigating the "accepted terms" (Rezaeian, 2015, p. 2) that contribute to the construction of the discourse of medical research articles and medical case reports..

2.2.2 Multiword Collocations

Multiword collocations (Biber, 2009) are sequences of three or more words that have the peculiarity of being primarily composed of content words with strong collocational strength. Such sequences may be of relevance to the study of formulaic language in specialized texts in that they "tend to be technical referring expressions" (Biber, 2009, p. 289). They belong to that category of multiword sequences that have been described as useful low-frequency sequences with a mutual expectancy higher than could be expected by chance and that may fail to meet the frequency cutoffs set for the identification of frequently recurring expressions (Simpson-Vlach & Ellis, 2010). Some researchers have advocated the combination of frequency metric with some statistical measure of cohesiveness for the identification of such sequences (e.g., Ellis, 2012; Simpson-Vlach & Ellis, 2010; Paquot, 2017). As already mentioned, the most frequently used statistical measure of cohesiveness in studies of formulaic language has been MI. Such frequent use of MI may be partly explained by convenience, given that there already exist corpus tools like *Antconc* (Anthony, 2020) or *Collocate* (Barlow, 2004) that automatically compute MI scores of sequences of various lengths.

Like 2-word collocations, multiword collocations can thus be identified empirically on the basis of both frequency and some measure of collocational strength. However, the frequency thresholds need not to be too high to allow for the identification of the potentially useful low frequency sequences. In his brief comparison of multiword collocations and lexical bundles, a type of formulaic sequence presented in the next section, Biber (2009) used a lower frequency cutoff of 10 times per million words for the identification of 4-word collocations in lieu of the more conservative cutoff of 20 times per million words for 4-*wordlexical* bundles.

While corpus tools like *Antconc* (Anthony, 2020) or *Collocate* (Barlow, 2004) automatically compute MI scores for sequences of various lengths, it should be mentioned that there is no agreement as to how well MI can measure strength of association for sequences of three or more words as (1) it was initially meant to measure collocational strength between two words, and (2) it is biased towards low frequency words (Biber, 2009; Gablasova et al., 2017). Gries (2013) suggests that *delta P* can also be used for sequences longer than two words. The present study has explored that possibility.

To my knowledge, there is no published study that has focused on multiword collocations and their functions in specific disciplines and/or registers. Biber's (2009) comparison of multiword collocations and lexical bundles was for the purpose of addressing a methodological issue and providing evidence that the frequency approach and the MI approach to the identification of multiword expressions yield two distinct categories of formulaic sequences (i.e., lexical bundles and multiword collocations). However, there is indication of their potentially important role in the construction of discourse in specific disciplines and/or registers.

2.2.3 Lexical Bundles

Lexical bundles are simply the "most frequent recurrent sequences [of three or more] words" in a selection of texts representing a register (Biber et al., 2004, p. 373). They are identified empirically and automatically on the sole basis of their frequency in a corpus and the range of texts in which they occur. Biber et al. (1999) set the frequency cut-off for the identification of lexical bundle at 10 times per million words and across at least five texts. That cutoff has been the baseline in studies of lexical bundles, albeit with some variation, most likely depending on the selected bundle length. Biber et al. (1999) reported that as bundles get longer, they become less frequent, and subsequent studies that investigated bundles of less than five words have used higher frequency cutoffs from 20 times per million words (e.g. Hyland, 2008a; 2008b; Cortes, 2004) to 40 times per million words (e.g. Biber et al., 2004; Biber & Barbieri, 2007). The range of at least five texts, is used to control for potential idiosyncratic uses of certain sequences by individual authors.

2.2.3.1 Structures and functions of lexical bundles

Lexical bundles are analyzed structurally and functionally, mostly following analysis frameworks developed by Biber et al. (1999), Biber et al. (2004), and Hyland (2008a).

Structurally, lexical bundles fall into one of the three major structural categories identified by Biber and colleagues: bundles including verb phrase fragments (e.g. *is going to be, can be used to, are shown in table*), bundles including dependent clause fragments (e.g. *I want you to, if you want to, it's going to be*), and bundles including noun phrase or prepositional phrase fragments (e.g. *on the basis of, in the absence of, a reduction in the, the end of the)*.

For the functional analysis of lexical bundles Biber et al. (2004) also proposed three main categories: stance expressions (e.g. *are more likely to, it is important to, it is possible to*), discourse organizers (e.g., *as well as the, on the other hand*), and referential bundles (e.g. *is one of the, on the basis of, as shown in figure).* This functional analysis framework has been adopted and/or adapted in several subsequent studies (e.g. Biber, 2006; Biber & Barbieri, 2007; Chen & Baker, 2010; Cortes, 2004; Cortes & Csomay, 2007).

Another functional analysis framework was developed by Hyland (2008a) specifically for the analysis of lexical bundles in academic research papers, namely published research articles, doctoral dissertations, and master's theses. Within that framework, lexical bundles fall into one of three categories: research-oriented bundles, used to describe the writers' activities and experiences (e.g., *the baseline characteristics of, in the present study*); text-oriented bundles, serving the function of text and meaning organization (e.g., *in addition to the, with respect to the*); and participant-oriented bundles which include stance and engagement features (e.g.; *it is possible that, it is important to*). Like Biber et al.'s framework, this framework has been either adapted or adopted in a number of studies of lexical bundles. The present study also uses structural and functional classification frameworks based on these taxonomies.

2.2.3.2 Lexical bundles in academic writing

Lexical bundles have been shown to be ubiquitous in academic written texts. Biber et al., (1999) found that these formulaic sequences occurred at the high frequency per million words of over 60,000 times for 3-word bundles and more than 5,000 times for 4-word bundles in their corpus of academic prose. Such pervasiveness of lexical bundles in academic texts has been explained by the fact that they function as building blocks by providing frames for new information that is being expressed in the text. In other words, they function as “a kind of pragmatic ‘head’ for larger phrases and clauses” that are used as a vehicle for new information (Biber, 2009; p. 284). Given their importance in academic discourse, the use of lexical bundles in various academic disciplines and registers has been extensively investigated (e.g., Biber et al., 2004; Cortes, 2004; Grabowski, 2013; 2015; Jalali et al., 2008; Jalali & Moini, 2014; Nesi & Basturkmen, 2006; Hyland, 2008a; Hyland, 2008b). Some of these studies have focused on the use of lexical bundles in single disciplines and registers (e.g., Bal, 2010; Jalali & Moini, 2014; Mbodj-Diop, 2016; Sahriari, 2017), and have thus contributed to the description of language use in those registers and disciplines. Jalali & Moini (2014), for example, investigated the use of 4-word lexical bundles in the introduction sections of medical research articles and found that medical writers have a strong preference for noun phrases and phrasal bundles.

Other studies, on the other hand, have compared the use of lexical bundles in specific registers both across and within disciplines (e.g., Cortes, 2004; Grabowski, 2013; 2015; Hyland 2008a, 2008b). Such studies have been instrumental in the description of disciplinary variations in academic writing. Cortes (2004) compared the use of 4-word lexical bundles in two disciplines, history and biology and found variations in both the structures and functions of lexical bundles in history and biology research articles. Similarly, Hyland (2008a) compared the use of lexical bundles in published research articles in two distinct fields: pure sciences, represented by electrical engineering and microbiology; and social sciences, represented by applied linguistics and business studies. He found that bundles varied not only between the two fields but also between disciplines within the same field, with less than half of the top 50 bundles in each list occurring in any other list. His functional analysis revealed a predominance of research-oriented bundles in biology and engineering texts, text-oriented bundles dominating the applied linguistics and business science corpora, reflecting the empirical and interpretative natures of texts in hard sciences and social sciences, respectively. Grabowski (2015), on the other hand, compared two different registers within a same discipline (pharmacy). Among other features, he compared the use of lexical bundles in Patient Information Leaflets and Summaries of Product Characteristics. He found that none of the top 20 bundles occurring in his two corpora appeared in the other lists.

These studies of disciplinary variations in the use of lexical bundles, both between and within disciplines, have underscored the need for research in lexical bundles – and by extension, formulaic language – to consider each discipline and each register within it in their own rights. As studies of the building blocks of academic writing, these studies have highly contributed to our understanding of writing in the disciplines and have been of importance for writing instruction in EAP and ESP contexts. However, as valuable and important as they are, they do not provide a full account of the use of formulaic language in academic writing. Lexical bundles are continuous fixed expressions, but language users employ both continuous and discontinuous formulaic sequences (Biber, 2009). Moreover, as stated above, other categories of formulaic sequences have been shown to play important roles in the construction of academic discourse. Therefore, for a more thorough description of how discourse is constructed in specific registers and disciplines, it is necessary to supplement the study of the building blocks (i.e., lexical bundles) with an analysis of the use of these other formulaic sequences.

2.2.4 Lexical Frames

Lexical frames (Biber, 2009; Gray & Biber, 2013) or phrase-frames (Stubbs, 2007; Römer, 2010) or are "sets of n-grams which are identical except for one word" (Römer, 2010, p. 98). They are discontinuous sequences of three or more items with one variable slot that can be filled by different *fillers* (Biber, 2009; Gray & Biber, 2013). For instance, the slot of the 4-frame *the * of the* may be filled by fillers such as *sum, square, addition* yielding sequences like *the sum of the, the square of the,* and *the addition of the.* The form of a frame varies from one study to another and is dependent to what is considered as a slot. Some studies consider only inner slots (e.g., *A*C, A*CD, AB*D, A*CDE, AB*DE, ABC*E*) while others consider both inner and outer slot (e.g., **BC, A*C, AB**). Biber (2009) found that frames with inner slots were more characteristic of academic prose while frames with outer slots were more frequently found in his corpus representing conversation.

Like collocations, lexical bundles, and multiword collocations frames are identified empirically and automatically, using programs such as *Antgram* (Anthony, 2020) or *kfNgram* (Fletcher 2002-2007). Earlier studies of frames (e.g., Biber, 2009; Römer, 2010) used a *bundle-to-frame* approach for the identification of frames. That is, a conservative frequency cutoff (e.g., 10 or 20 times per million words) is set for the identification of *n*-grams or bundles that would serve to generate the list of frames. However, there since has been evidence that this approach excludes some high frequency frames that do not involve frequently occurring formulaic sequences like lexical bundles (Gray & Biber, 2013). Gray and Biber then proposed a *fully inductive approach* that consists of identifying frames based on all *n*-grams in the submitted corpus, with the ground (i.e., the frequency threshold) set at 1.

2.2.4.1 Variability and predictability of frames

Variability and predictability of frames are analyzed in studies of frames with the rationale that the occurrence of fillers is not random but rather driven by Sinclair's (1991) *idiom principle* (Römer, 2010). Thus, the predictability and internal variation scores of a frame are indications of the extent to which the *idiom principle* operates in the frames; that is, the extent to which the frame is fixed or variable. Internal variation is typically measured in studies of frames by computing the type-token ratio (TTR). The TTR is computed by dividing the number of fillers of a frame by the total number of occurrences of the frame in the corpus under study. The closer the TTR is to 1, the more variable the frame is, and a score closer to 0 indicates that the frame is relatively fixed, allowing little variation. Predictability, on the other hand, has been measured differently.

The predictability of a frame provides information about the distribution of the fillers across the frame tokens; that is, is there a comparatively even distribution of the fillers or does a small number of high frequency fillers account for the total number of fillers of a frame. Gray and Biber (2013) rightly noted that TTR alone may not reflect the actual predictability of highly frequent frames as it is sensitive to token frequency. The authors suggested combining TTR with other metrics reflecting the distribution of fillers in the frames. As mentioned above, different approaches have been used to compute frame predictability. Biber (2009) and Gray & Biber (2013), divided the number of occurrences of the most frequent filler of a frame by the total number of occurrences of the frame to obtain percentages indicating the degree of predictability of the frames they identified. For instance, the frame *of the * of* occurred 3,434 times in their corpus of academic prose and the most frequent filler *use* occurred 55 times. The predictability score of the frame *of the * of* is then 2% (55 ÷ 3,434), which is quite low. Römer (2010) assessed the predictability of the frames by plugging the number of occurrences in a graph to visually evaluate how predictable each frame was. Another measure of frame predictability is *normalized entropy* which is a statistical measure that provides scores for each frame based on the percentage of each variant within the frame. Entropy scores are expressed on a scale of 0 to 1. The higher the entropy score, the less predictable a frame is. *Antgram* now automatically computes TTR and entropy scores for all frames identified in a corpus. The present study uses TTR and normalized entropy scores to assess the variability and predictability of p-frames.

2.2.4.2 Structures of frames

Gray & Biber (2013) proposed a structural analysis framework that allowed them to identify the structural correlates of frames in academic prose and conversation. Their framework comprised three broadly defined categories:

- Verb based frames: frame contains one or more modal, auxiliary or main verb (e.g. *must be * to, was * in the, I * going to, what did * do*).
- Frames with other content words: frame contains one or more nouns, adjectives, or adverbs but no verbs (e.g., *on the * hand, it * necessary to, I * no idea*).
- Function word frames: frame consists of only function words such as prepositions, determiners, conjunctions, pronouns, complementizers, etc. (e.g., *the * of this, in the * that, as * as you, a * in the*) (p. 122)

Their analysis revealed that frames in academic writing predominantly fall in the third category (i.e., function word frames) and even those falling in the first category (verb-based frames) typically involved auxiliaries rather than lexical verbs.

2.2.4.3 Functional analysis of frames

The functional analysis of frames is done in context with the different fillers in the slots. This probably explains why most studies that have actually looked at the functions of frames (not all of them have) have primarily used Biber et al.'s (2004) and / or Hyland's (2008a) frameworks for the functional analysis of lexical bundles. To my knowledge, no framework has yet been developed specifically for a systematic analysis of the functions of frames. Garner (2016) used the functional analysis framework developed by Biber et al. (2004) to determine the functions of frames with their different fillers in learner corpora representing different levels of proficiency. Fuster-Márquez and Pennock-Speck (2015) drew from both Biber et al.'s (2004) and Hyland's (2008a) frameworks to analyze the functions of frames in their corpus of British hotel websites. Another study by Li, Yoon, and Kisselev (2018) used Biber et al.'s framework to study the functions of frames in the introductions of social science articles.

On the other hand, Cunningham (2017) developed a functional taxonomy specific to the writing of mathematics that comprised three main categories: *Signaling completion of the proof, identifying the location of a proof in the larger discourse of the text,* and *stating the manner of the proof* or *remaining portion of the proof*. Similarly, Römer (2010) identified four main functions of frames and their fillers in her corpus of book reviews: *expressing evaluation, referring to a book's structure*, *referring to the content of a book*, and *organizing the discourse*.

Geluso (2019) used an approach that is of relevance to the present study. He investigated the semantic characteristics of 30 frames in a learner corpus of argumentative and literary essays. His approached included only function word frames (e.g., *in the * of, the * of the*), previously identified by Gray & Biber (2013) to be more frequent in academic prose. This approach is based on claims that (1) language patterning occurs as a result of the interaction between the *syntagmatic* and the *paradigmatic* axes of language (Sinclair, 2004) and (2) words that share the same meaning or class tend to occur in similar grammatical environments (Renouf & Sinclair, 1991). Syntagmatic and paradigmatic axes, as explained by Sinclair (2004), refer to the horizontal and vertical axes of language, respectively, with the syntagmatic axis controlling the grammatical context, or the structure being observed, and the paradigmatic axis specifying the possible lexical choices within a specific position in the structure. This interaction between the two axes of language is what underlies the relationship between frames and their fillers. Geluso (2019) nicely summarizes this observation in these terms: “Because frames […] essentially constitute recurring syntactic patterns of words, they exist on the syntagmatic axis of language. Words that fill the variable slot of these frames then constitute sets of words on the paradigmatic axis of language.” (p. 13).

Geluso’s (2019) study is of relevance to the present study as it lends support to the widely evidenced claim that words with similar meanings tend to occur in the same grammatical environments (e.g., Gledhill, 2000; Hoey, 2005; Hunston & Francis, 2000; Marco, 2000; Renouf & Sinclair, 1991). The present study uses an approach similar to Geluso’s but takes a more exploratory approach that goes beyond one structural category of frames and proposes a classification framework for all three structural groups of frames.

One final point to note regarding research of frames is that most studies have looked at three-word and four-word combinations or just 4-word combinations, except Römer (2010) who explored up to 6-word combinations. This probably can be explained by the large amount of data at hand after the automatic identification of frames. To make the data manageable, most studies have focused on subsets of frames to analyze functionally. Some have focused on the top (and therefore most frequent) items in their frame lists (e.g., Römer, 2010; Garner, 2016). Others like Grabowsky (2015) have suggested a sampling from the top, the middle, and the bottom of the frame list. Lu et al. (2018) decided to exclude all frames that were “not meaningful or pedagogically relevant” (p. 80). Gray and Biber (2013) set a frequency threshold of 200 times per million words, but this was for comparison purposes with frames previously identified by Biber (2009). Whatever the approach, it remains that the analysis of frames requires some "realistic strategies" to make the data manageable (Stubbs, 2007, p. 92).

2.3 Research of Formulaic Language in Medical Writing

Researchers in the medical field have expressed the need for courses that would help familiarize both novice and L2-English writers with medical writing and conventionalized expressions in the field (e.g., Rezaeian, 2015; Yanoff & Burg, 1988). Such call highlights the potential importance of research in and instruction of formulaic sequences in medical writing. Yet, there is a dearth of research in formulaic language in written medical texts. Research in medical writing has mostly focused on the study of (a) selected linguistic features and their rhetorical functions such as modals in expressing epistemic modality (e.g., Yang, Zheng, & Ge, 2015), the use of reporting verbs in medical research articles (e.g. Jirapanakorn, 2012), or conditionals in medical research articles (Ferguson, 2001), (b) medical vocabulary in research articles (e.g., Chen & Ge, 2007; Mungra & Canziani, 2013; Wang, Liang & Ge, 2008), and (c) the textual organization of the medical research article (e.g., Fryer, 2012; Li & Ge, 2009; Nwogu, 1997).

The handful of studies that have looked at formulaic language in medical writing have mostly focused on single types of formulaic sequences (e.g., Gledhill, 2000; Marco, 2000; Abdollahpour & Gholami; 2018; Jalali & Moini, 2014; Jalali, Moini, & Arani, 2015; Mbodj-Diop, 2016). As already mentioned, Gledhill (2000) looked at the phraseology of Introduction sections of *Cancer* articles and investigated the discourse functions of collocations comprising the verb forms *has, have, been,* and *is;* and the prepositions *of* and *to*. Marco (2000) explored the collocational frameworks (pairs of function words that form discontinuous sequences with one variable slot) in medical research articles focusing on intermediate words that fill in the following three frameworks: *the … of, a … of,* and *be … to.* Abdollahpour & Gholami (2018) investigated the use of lexical bundles in the abstracts of MRAs. Jalali & Moini (2014) looked at the use of lexical bundles in the introductions of MRAs. Both Jalali, et al. (2014) and Mbodj-Diop (2016) looked at the structures and functions of lexical bundles in MRAs.

These studies, like other studies of formulaic language in disciplinary writing, have added to our understanding of the use of formulaic language in the medical field. Yet, they only investigated single types of formulaic sequences. Additionally, all these studies have exclusively focused on the medical research article. To my knowledge, only two studies, previously mentioned in the introductory chapter of the present study have investigated the medical case report (Goudier, 2008; Helan 2012). As mentioned in section 1.2., both studies used a genre approach to the study of this register. Goodier compared the moves in case reports written by students and professionals in radiology and Helan described the rhetorical moves of medical case reports. It appears then that there still remains a need to provide a more comprehensive description of other types of formulaic sequences in registers other than the medical research article. The present study takes a step in that direction.

3 CHAPTER 3: CORPORA AND METHODOLOGY

This chapter presents the corpora and methodology used for the analysis of formulaic sequences in the present study. First, I provide a description of the two corpora collected to represent the two registers under study, i.e., medical research articles (MRAs) and medical case reports (MCRs), in section 3.1. Then in section 3.2, I describe the methodology used to identify and analyze the formulaic sequences investigated in the present study.

3.1 Corpora

As very little is known about MCRs, a preliminary situational analysis of the two registers was conducted based on 50 MRAs and 50 MCRs. Academic research articles have already been well-described in the literature (e.g., Gray, 2015), and therefore the objectives of the preliminary situational analysis were to inform the collection of the MCR corpus and also confirm that MCRs and MRAs are indeed two distinct registers. The situational analysis frameworks for both registers will be introduced in the next section and the results of the situational characteristics analyses of each register are presented in Chapter 4. Thus, in this section, I refer to findings of the preliminary situational analysis only when they are relevant to decisions made during the corpus collection.

3.1.1 The Medical Research Article (MRAC)

The MRA corpus is an updated version of a corpus previously collected for the investigation of the use of lexical bundles in MRAs (Mbodj-Diop, 2016). That corpus was built based on previous research on corpus collection (e.g., Biber, 1993; Loi, 2010), previous studies of MRAs (e.g., Nwogu, 1997; Wang et al., 2008), journal websites, sample MRAs, and insight from expert informants. Based on those preliminary investigations, I collected a 1-million-word corpus that included only quantitative MRAs written in the conventional Introduction – Results –

Methods – Discussion (IMRD) format, published in five renowned journals in the medical field, 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).* From each of these journals, I collected 50 articles published in the past ten years at that time (between 2006 and 2015), for a total of 250 texts.

For the purpose of present study, the working definition of MRAs after the preliminary situational analysis was as follows: a research article written in the IMRD format, not necessarily is the same order, and published in renowned peer-reviewed journals in the field of medicine. I updated the initial corpus by 20%. I added 10 articles published in 2020, from each of the journals above, to replace articles published in 2006 and 2007. In other words, I added 50 newly published articles and removed the 50 oldest articles of the initial corpus. The thus updated corpus includes 250 texts for a total over one million words. Table 3.1 summarizes information about word count and articles in the MRAC.

Table 3.1. The Medical Research Article Corpus

Journals & Corpus	Number of texts	Mean word count per text	*SD*	Total word count per journal
JAMA	50	4,197.32	851.79	209,866
JCI	50	5,785.10	1,199.48	289,255
The Lancet	50	4,773.74	1,744.44	238,689
NEJM	50	3,842.18	547.00	192,109
Science TM	50	5,589.00	1,790.57	279,420
MRAC	**250**	**4,846**	**1,527.35**	**1,209,367**

3.1.2 The Medical Case Report Corpus (MCRC)

As mentioned above, very little is known about MCRs. Therefore, the collection of the MCRC was informed by the findings of the preliminary situational analysis, literature from the medical field, journal websites, sample MCRs, and insight from field-expert informants. It was particularly important to ensure two main things: (1) that the medical case report does not consist of multiple subregisters, mainly given Gray's (2015) findings of different subregisters of academic research articles; and (2) that MCRs included in the MCRC were published in reputable journals.

The initial situational analysis revealed the following main three organizational formats of MCRs determined by journals' guidelines for authors: (1) Introduction, Report of a Case, Discussion (IRD); (2) Background, Case Presentation, (Investigation), Treatment, Outcome and Follow-up, Discussion, Learning point/take-home messages (BC(I)TODL); and (3) Report of Case, Diagnosis, What to Do Next, Discussion, (Patient Outcome) (RWD(P)). The bracketed sections may not be included in the MCR, depending on the case being presented. That multiplicity of formats raised the legitimate question of whether there exist multiple subregisters of MCRs and whether to include all three formats in the MCR corpus.

The review of the literature on MCRs in the medical field and insight from field-expert informants led to the conclusion that the MCR is a single register (see Gagnier et al., 2013, 2014; Rison et al., 2017). The differences in section headings stem from a lack of consensus among journals as to whether the main section of MCRs, the Case Presentation (discussed in depth later in section 4.3. of the Situational Analysis chapter), should be presented under a single or multiple headings. I found that regardless of headings, all 50 case reports included the same types of information. That information is listed in the CARE (CAse REport) guidelines proposed by

Gagnier et al. (2013) and revised by Gagnier et al. (2014). The authors describe the medical case report as follows:

> A case report tells a story in a narrative format that includes the *presenting concerns, clinical findings, diagnoses, interventions, outcomes* (including adverse events), and *follow-up*. The narrative should include a discussion of the rationale for any conclusions and any *take-away messages* [emphasis added]. (Gagnier et al., 2013, p. 3)

I therefore decided to include all three formats, based on the Gagnier et al.'s (2013) definition and confirmation from one field-expert informant that MCRs in these three formats "are exactly the same in terms of purpose, audience, content and aims" (M. M. Ka, personal communication, October 22, 2020).

Another parameter I considered for the collection of the MCRC was the reputation of journals as the majority of MCRs are published in open-access journals (Rison et al., 2017). I was able to find two renowned journals that publish case reports and are not open access (*JAMA* and *BMJ Case Report*) and received three other open-access journal suggestions from one field-expert informant: the *Journal of Medical Case Reports (JMCR), Oxford Medical Case Report (OMCR),* and *International Medical Case Report Journal (IMCRJ)*. According to Rison et al. (2017), the most reputable open-access journals are indexed in *PubMed*, a search engine described by Williamson & Minter (2019) as "one of the most widely accessible biomedical resources globally" (p. 16). I personally searched *PubMed* to ensure that all five journals were indexed in *PubMed*, which they were.

In light of the information gathered, the working definition of MCRs for corpus collection was as follows: any medical case report published in a journal indexed in *PubMed* and reflecting, through its headings, the information listed in Gagnier et al.'s (2014) CARE guidelines. Given the scope of general medicine, and the generalist nature of the five journals selected (as revealed by the preliminary situational analysis), I did not make any attempt to control for topic. The MCRC includes 704 case reports on various topics in general medicine. To build a corpus of at least one million words, I collected between 106 and 196 MCRs per journal for a total of 704 texts. Table 3.2 provides a summary of the information about the MCRC.

Table 3.2. The Medical Case Report Corpus

Journals & Corpus	Number of texts	Mean word count per text	Standard Deviation	Total word count per journal
JAMA	196	1,022	444.01	200,212
BMJCR	114	1,758	675.38	200,413
JMCR	106	1,896	547.59	201,032
OMCR	173	1,159	227.45	200,446
IMCRJ	115	1,772	243.14	203,804
MCRC	**704**	**1,428.8**	**624**	**1,005,907**

3.2 Analysis of the Situational Characteristics of MRAs and MCRs

To answer RQ1 that asked about the situational characteristics that distinguish MCRs from MRAs, I conducted a situational analysis of each register and then discussed the similarities and differences between the two registers. To analyze the situational characteristics of MRAs, I used Gray's (2015) framework as a starting point, as that framework was designed to analyze the situational characteristics of academic research articles in various disciplines. As explained in 2.1, in the Methodology chapter, the framework, summarized in Table 3.3, included eight categories of characteristics: *Participants, Layout and Organization, Setting, Subject/topic, Purpose, Nature of Data or Evidence, Methodology*, and *Explicitness of Research Design*. I maintained the eight major categories but adapted some of the subcategories. I also added some factors that appeared to be relevant to both the description of the situational characteristics of

MRAs and the comparison of the two registers under scrutiny. The changes I made and the final situational analysis framework for MRAs are presented in chapter 4, section 4.1.

Table 3.3. Summary of Gray's (2015) Situational Analysis Framework

Major Categories of Characteristics	**Subcategories**
Participants	- *Writer* (single, small group, large groups)
Layout and Organization	- *Length* - *Headings* (none, unnumbered, numbered) - *Use of Abstracts* (yes, no) - *Visual Elements* (none, tables, tables & figures, equations) - *Sections* (IMRD, IMRD with varied order, other, standardized section headings, variable section headings/names)
Setting	- *Nature of Journal* (generalist, specialized)
Subject/topic	- *General Topic(s) of the Discipline*
Purpose	- *General Academic Purpose*
Nature of Data or Evidence	- *Presence of Observed Data* (yes, no) - *Use of Numerical Data* (yes, no) - *Primary Presentation of Evidence* (extensive prose, quantitative displays, mathematical formula) - *Object of Study*
Methodology	- *General Method Type* (observational, experimental, n/a [for theoretical]) - *Statistical Techniques* (n/a [for theoretical and qualitative]), descriptive statistics, statistical difference testing, other advanced statistics)
Explicitness of Research Design	- *Explicitness of Purpose* (Direct statement, Indirect / No discernible statement) - *Explicitness of Research Questions* (Direct statement, Indirect / No discernible statement) - *Explicitness of Citations* (Within the text, In footnotes/endnotes) - *Explanation of Evidence* (Extensive, Mention / No discernible statement) - *Explanation of Procedures* (Extensive, Mention / No discernible statement)

For the situational analysis of MCRs, I developed a second framework, presented in detail in the next chapter in section 4.2. The framework is primarily based on the seven main categories of characteristics proposed by Biber & Conrad (2009) and presented in Figure 2.1. in the Literature Review chapter, Chapter 2. Those seven characteristics include: *Participants, Relations among participants, Channel, Production Circumstances, Setting, Communication purposes,* and *Topic.* Additional sources I resorted to develop the framework include the adapted situational analysis framework for MRAs, expert-informants, journal websites and instructions to authors, sample case reports, and Gagnier et al.'s (2014) CARE guidelines (see chapter 4, section 4.2 for the CARE guidelines).

3.3 Identification and Analysis of Collocations

In order to answer RQ 2 on the use of collocations in MRAs and MCRs, and the potential variations between the two registers, I identified and analyzed frequently cooccurring 2-word sequences in the MRAC and the MCRC. As mentioned in section 2.2.1 of the Literature Review chapter, collocations have been operationalized differently in the literature depending on research focus. Given that one of the aims of the present study is to describe the conventionalized expressions described by Rezaein (2015) as being central to medical discourse, the key characteristics of collocations I considered were mutual expectancy – to capture specific expressions – and word class of the constituents of the collocation. To decide on the word class of collocation elements, I conducted a pilot study that revealed that in addition to pairs of lexical words, some bigrams with strong mutual expectancy included only one lexical word (e.g., *insight into*, *predominance of, determine whether*) or were "borrowed" expressions (e.g., *ex vivo, in vitro, de novo*). Therefore, pairs with only one lexical word were considered as valid candidates for analysis.

As regards mutual expectation, I decided to use *delta P* (ΔP), as suggested by Gries (2013). As explained in 2.2.1, ΔP offers the possibility to identify bigrams with elements that are mutual predictors and other pairs where only one element of the collocation selects the other one. I decided to include both types of bigrams to ensure that important bigrams that include high frequency function words (and therefore with lower ΔP values) are not excluded from the analysis. In light of these considerations, I operationalized collocations as frequently occurring bigrams including at least one lexical word, and with strong mutual or unidirectional expectancy as measured by *delta P*.

3.3.1 Identification of Collocations

Collocations were identified in *Antgram* (Anthony, 2020) with the minimum frequency set at 5 times pmw and a range of 5 texts, in an aim to include low frequency technical expressions used across at least five texts. As already mentioned, I used *delta P* as a measure of collocational strength. *Delta P* values are computed using the following formula proposed by Ellis (2007, p. 11): $\Delta P = P(O|C) – P(O| – C)$. Gries (2013) provides the different steps (see figure 3.1) involved in computing ΔP values, using the example of the bigram *of course* in the BNC corpus. $\Delta P_{2|1}$ is the probability of having word$_2$ if word$_1$ is present, and $\Delta P_{1|2}$ is the probability of having word$_1$ if word$_2$ is present. As I do not have programing skills, I collaborated with a statistician to emulate in Excel the formula shown in figure 3.1 and compute ΔP values for all bigrams identified in *Antgram.*

$$(6) \quad \Delta P_{2|1} = p\,(course\,|\,word_2 = of) - p\,(course\,|\,word_2 \neq of) = \frac{5610}{174548} - \frac{2257}{10235320} \approx 0.032$$

$$(7) \quad \Delta P_{1|2} = p\,(of\,|\,word_2 = course) - p\,(of\,|\,word_2 \neq course) = \frac{5610}{7867} - \frac{168938}{10402001} \approx 0.697$$

Figure 3.1. Formula Used to Compute ΔP Values of Bigrams

We also computed the log ratio (log10) of ΔP values to later facilitate the classification of bigrams. As suggested by Gries (2013), log10 makes "margins more legible" (p.162). Log10 scores are automatically computed in Excel and vary from – 4.0 to + 4.0. Negative values indicate that $word_1$ is a stronger predictor than $word_2$, and positive values indicate $word_2$ as the strongest predictor. Log10 scores between – 0.5 and + 0.5 are indication of mutual attraction between $word_1$ and $word_2$ as there is not much difference between $\Delta P_{1|2}$ and $\Delta P_{2|1}$.

Once all ΔP values and log10 scores were computed in Excel for all identified bigrams, I identified three initial lists of candidates labelled *Bidirectionals*, *Unidirectionals 1*, and *Unidirectionals 2*. *Bidirectionals* refer to bigrams with relatively strong mutual attraction; *Unidirectionals 1* are pairs in which $word_1$ was found to be the strongest predictor; and *Unidirectionals 2* refers to bigrams with $word_2$ as the strongest predictor. Then I used the following thresholds for the classification of bigrams identified in each of the three lists of candidates:

- *Bidirectionals:* $-0.50 \leq \log 10 \leq +0.50$
- *Unidirectionals1:* $-4 \leq \log 10 \leq -0.51$
- *Unidirectionals 2:* $0.51 \leq \log 10 \leq 4$

The next step was to identify pairs with high collocational strength. The thresholds for bigrams in the three lists were as follows:

- *Bidirectionals:* $\Delta P_{1|2} \geq .30$ and $\Delta P_{2|1} \geq .30$
- *Unidirectionals1:* $\Delta P_{2|1} \geq .70$
- *Unidirectionals 2:* $\Delta P_{1|2} \geq .70$

The thresholds for the *Unidirectionals* categories were set higher than those of *Bidirectionals* to ensure that the strong predictor occurs primarily with the second element of the collocation. The three lists of candidates were then cleaned of any bigram that did not meet the ΔP thresholds above, bigrams that did not include at least one lexical word, and proper nouns (e.g., *Melinda Gates, John Hopkins, Santa Cruz*).

The next step was to check overlaps. Each of the remaining bigrams in the three lists was viewed in context using the *Concordance* tool in *Antconc* (Anthony, 2020) to identify pairs that were part of longer sequences. For example, the bigram *care unit* occurred 113 times in the MCRC. It co-occurred 106 times with *intensive*, 5 times with *coronary* and only two times with other words, which did not meet the frequency threshold of 5 times per million words. The bigram was therefore deleted from the list of candidates. Once all overlaps had been taken care of, the three list were ready for analysis.

3.3.2 Analysis of Collocations

Collocations were analyzed both structurally and functionally. For the structural classification, I adapted Ackermann & Chen's (2013) framework for lexical collocations to include grammatical and 'borrowed' collocations. Table 3.4. shows the framework used for the structural classification of collocations from both corpora.

For the functional analysis of MRAC collocations, I used Hyland's (2008a) framework for the analysis of discourse functions of lexical bundles in academic research articles. This framework, presented on Table 3.7., in section 3.4.2, comprises three major functional categories (Research-oriented, Text-oriented, and Participant-oriented) and a number of subcategories. I conducted the qualitative analysis in context using the *Concordance* tool in *Antconc* (Anthony, 2020) to view collocations in context. For the functional analysis of MCRC collocations, I designed a new framework, given that there exists, to my knowledge, no functional analysis framework for formulaic language in MCRs.

Table 3.4. Structural Classification Framework for Collocations, adapted from Ackermann & Chen (2013)

	Structures	Examples from the MRAC and MCRC
1. Lexical Collocations	Noun Combinations	
	• *n + n* • *adj + n*	*viral load, risk factors, chest radiograph, liquid nitrogen, cytoplasmic inclusions, western blot*
	Verb combinations	
	• *v + n*	*count fingers*
	• *v + adj*	*is conceivable*
	Verb-Adv combinations	
	• *adv + v*	
	• *v + adv*	
	• *adv + vpp*	*normally distributed, virally suppressed*
	Adj-Adv combinations	
	• *adv + adj*	*commercially available, critically ill, mutually exclusive*
2. Grammatical Collocations	N & function word	*insight into, predominance of, and colleagues, for example*
	V & function word	*determine whether, confined to, we undertook, counterstained with*
	Adj & function word	*irrespective of, amenable to, compatible with, analogous to*
3. "Borrowed" Collocations		*de novo, vice versa, bona fide, in vivo*

The framework was based primarily on observations from the analyses in context in *Antconc* and the required information that MCRs must include based on Gagnier et al.'s (2013) definition of case reports. I also drew from Biber et al.'s (2004) and Hyland's (2008a) frameworks for the labeling of some of the functions. The functional analysis framework for collocations in MCRs comprises the following five major categories:

- Case-related collocations used to specify case subjects (patients), describe the medical pathology, and refer to other individuals related to the case;
- Diagnosis/Intervention-related collocations used to refer to biomedical elements and/or processes, decisions made during diagnosis, and intervention procedures;

- Outcome/Follow-up-related collocations used to report outcomes of cases being reported and describe follow-ups;
- Discourse organizers, to organize the text into a coherent whole; and
- Stance features, mostly to indicate authors' evaluation of the importance of the case being presented and/or therapeutic approaches they took.

These major categories, as well as their respective subcategories and examples of collocations from the MCRC are shown in Table 3.5.

Table 3.5. Functional Classification Framework for Collocations in MCRs
Based on Gagnier et al.'s (2014) guidelines and Hyland's (2008a) framework

Categories	**Sub-categories**	**Examples from the MCRC**
Case-related	Subject specification	*transplant recipient*
	Description of pathology	*iliac fossa, sleep apnea, heart failure*
	Other Individuals	*general practitioner, family members*
Diagnosis/ Intervention-related	Biological processes/elements	*excisional biopsy, right eye, ejection fraction*
	Decisions	*ruled out, referred to, elected to*
	Procedures	*thrombolytic therapy, treated with*
Outcome/Follow-up – related		*adverse events, transitioned to*
Discourse organizers	Resultative signals	*owing to, leading to, resulted in*
	Framing signals	*depending on, according to, based on*
	Elaboration/Clarifications	*consisting of, characterized by*
Stance Features		*the best, the highest, the rarest*

3.3.3 Identification of Shared Collocations

As part of the comparison of the use of collocations in MRAs and MCRs, I identified and analyzed pairs shared by the two registers. To that end, I copied and pasted the MRAC and MCRC collocations on a same Excel spreadsheet, and then used the *Conditional Formatting* tool in Excel the to highlight overlaps. I highlighted overlaps following the three steps below in Excel

(Steps 2 and 3 are illustrated in Figure 3.2.):

- Step 1: Copy and paste the two lists side by side on an Excel spreadsheet.
- Step 2: Select the two columns containing the two lists. Then in the home menu, click on the *Conditional Formatting* tab (A), then click on *Highlight Cell Rules* (B), and then on *Duplicate Values…* (C).
- Step 3: In the pop-up box that appears, click "OK" (D). All cells containing the shares collocations are then automatically highlighted, as shown in Figure 3.3.

The next step was to delete all non-highlighted bigrams to obtain the list of shared collocations. I used the same procedure for the identification of all shared formulaic sequences investigated in the present study.

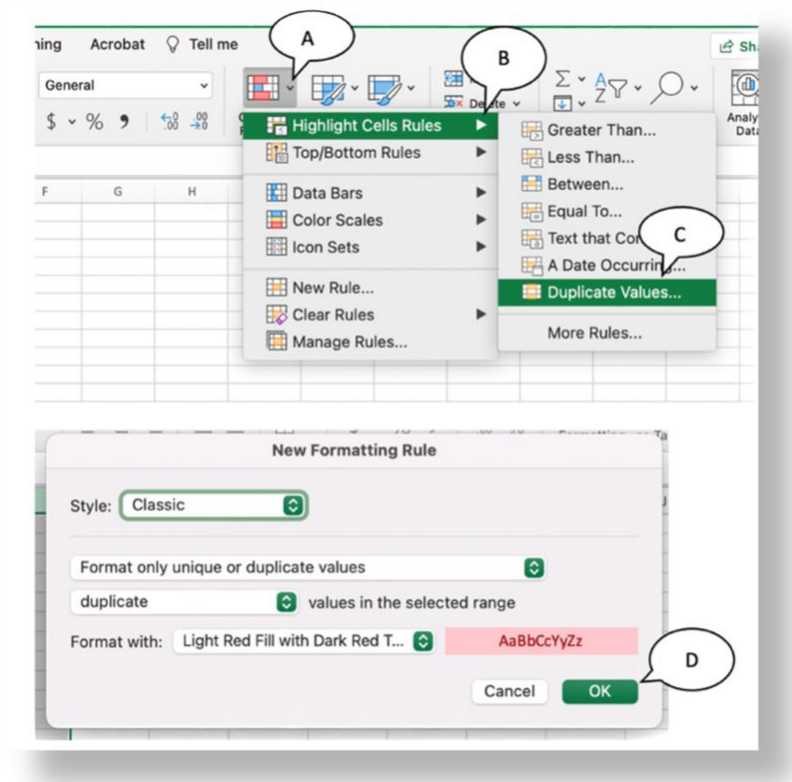


Figure 3.2. Steps for Highlighting Shared Sequences in Excel

	A	B
1	**MCRC**	**MRAC**
2	the rarest	the remainder
3	the highest	to verify
4	the largest	in situ
5	to address	in utero
6	to clarify	the latter
7	to proceed	to deliver
8	to seek	were instructed
9	in vitro	a priori
10	the latter	is conceivable
11	to relieve	and colleagues
12	the entire	the entire
13	the exact	to calculate
14	was extubated	the original
15	was readmitted	as evidenced
16	to remove	we reasoned
17	to investigate	to ensure
18	in situ	to obtain
19	to manage	cell strainer
20	to maintain	figure legends
21	to assess	not necessarily
22	to convert	we undertook
23	to detect	to examine
24	to achieve	to address
25	to determine	to generate
26	to diagnose	to achieve
27	after completing	per kilogram

Figure 3.3. Screenshot of Highlighted Shared Collocations in Excel

3.4 Identification and Analysis of Multiword Collocations

To answer RQ2b (What multiword collocations are used in the two registers? How do they compare in terms of structures and functions? Are they register-specific or rather shared within the same discipline?), I first determined a working definition for multiword collocations based on two main observations. First, multiword collocations consist primarily of lexical words and second, these sequences have strong mutual expectancy (Biber, 2009). To ensure that all sequences considered for analysis were predominantly composed of lexical words, I set the following criteria:

- 3-word, 4-word and 5-word sequences must contain no more than one non lexical word,
- 6-word to 8-word sequences must contain no more than 2 non lexical words, and
- 9-word and more sequences must contain no more than 3 non lexical words.

To determine the collocational strength of multiword collocations, I tested Gries's (2013) suggestion of using *delta P* for sequences longer than two words. That suggestion worked for 3-word multiword collocations but proved to be problematic for longer sequences. Therefore, I decided to use *delta P* for 3-word expressions – partly as some evidence of the feasibility of Gries's suggestion – and MI scores for all other sequences.

Based on the two observations mentioned above, I operationalized multiword collocations as frequently co-occurring sequences of three or more words consisting primarily of lexical words and with strong mutual attraction as determined by ΔP for 3-word sequences and MI for longer sequences.

3.4.1 Identification of Multiword Collocations

Multiword collocations in both corpora were identified following Cortes's (2013) exploratory approach that consists in progressively searching for longer expressions until the search yields no more result. I used the frequency thresholds of 10 times per million words (pmw) for 3-word and 4-word sequences, eight times pmw for 5-word sequences, six times pmw for 6-word and 7-word sequences, and five times pmw for longer multiword collocations. These relatively low frequency thresholds are necessary to identify sequences involving technical words as the latter often occur at a much lower frequency than other non-technical words (Gablasova et al., 2017). Additionally, for sequences longer than three words, the frequency thresholds served to mitigate the tendency of MI to favor low frequency expressions (Biber, 2009; Gablasova et al., 2017).

The range was set at five texts for sequences in both corpora. Given the big difference in the number of texts in each corpus, I conducted a mini pilot study using various ranges, mainly for the MCRC corpus. That pilot study revealed that at only 5% of texts in the MCRC, sequences had to occur across about 35 texts to be identified, which automatically excluded several potentially important sequences. I then decided to use a range of 5 texts in both corpora, given that the same range has been used in multiple studies of formulaic sequences, sometimes for corpora much larger than the MCRC (e.g., Biber et al., 1999).

Sequences of three words in both corpora were identified in *Antgram* (Anthony, 2020) using the thresholds described above. Then ΔP values were computed in Excel following Stefan Gries's suggestion in a personal communication to use the simple approach of calculating the probability of $word_3$ when ($word_1$+$word_2$) is present. In that way, only two ΔP values are computed in Excel: ΔP (A+B), and ΔP (C) with A, B, and C representing $word_1$, $word_2$, and $word_3$, respectively. The log 10 score of these two values was also computed to facilitate the selection of candidates for analysis. Only sequences with a log10 score between -0.50 and +0.50 (i.e., sequences with relatively similar mutual attraction) were considered for the list of candidates. Additionally, where the log10 score was between
-0.50 and + 0.00, the ΔP value of (A+B) had to be equal to or greater than .30; and where the log10 score is between +0.01 and +0.50, the ΔP value of C had to be equal to or greater than .30. In sum 3-word collocations in the list of candidates had to meet the following criteria:

- $-0.50 \leq \log10 \leq +0.00$ and $\Delta P_{C|(A+B)} \geq .30$
- $+0.01 \leq \log10 \leq +0.50$ and $\Delta P_{(A+B)|C} \geq .30$

All other multiword collocations were identified in *Collocate 2.1* (Barlow, 2004), using the explorative approach described above. Initial lists of candidates ranked by MI were generated by *Collocate* and saved to be cleaned later. The cleaning process for all lists including the 3-word list was done in three steps. First, I deleted all sequences that did not meet the criteria for the number of non-lexical words described above. The next step was to delete all sequences that had already been identified as lexical bundles, another type of sequences identified in this study and described in section 3.5. To that end, multiword collocations in each list were copied and pasted side-by-side on an Excel spreadsheet with the list of bundles of the same length. Next, overlaps where highlighted using the *Conditional Formatting* tool in Excel and deleted from the multiword collocation lists. The last step was to attend to sequences embedded in longer ones. Here also, I used concordance lines to manually identify embedded sequences. As was the case with collocations, I deleted shorter sequences if they did not meet the frequency thresholds outside of the longer expressions. After this final cleaning step, all lists were ready for analysis.

3.4.2 Structural and Functional Analyses of Multiword Collocations

For the structural analysis of multiword collocations, I developed a structural classification framework, in the absence, to my knowledge, of any framework to draw from. With the multiword collocations as observations, I identified the following three major structural categories:

- Phrasal Sequences, with three subcategories: Complex noun phrases, Prepositional phrases in the form of *prep +(complex) NP;* and Other phrases;
- Clausal Sequences, with two categories: Declarative clauses/fragments and Dependent clauses/Fragments; and
- Coordinated Binominals, defined by Biber et al. (1999, p. 1030) as “two words from the same grammatical category, coordinated by *and* or *or”*.

These categories are further described in Chapter 5, Section 5.2. Table 3.6 shows the structural classification framework for multiword collocations and some examples from the two corpora.

Table 3.6. Structural Classification Framework for Multiword Collocations

Structural Categories	**Subcategories**	**Examples from MRAC & MCRC**
Phrasal Sequences	Complex Noun Phrases	*acid phase reactants*
		common terminology criteria for adverse events
	Prepositional Phrases	*in nonhuman primates*
		with no significant past medical history
	Other Phrases	*available on request*
Clausal Sequences	Declarative Clauses/ Fragments	*vital signs were stable*
		showed a blood cell count of
	Dependent Clauses/ Fragments	*what we believe*
		to better understand
Coordinated Binominals		*chest pain and shortness of breath*
		palms and soles

The functional analysis of MRAC multiword collocations was based on Hyland's (2008a) framework for lexical bundles described in section 2.2.3 of the Literature Review chapter. I slightly adapted the framework by further dividing the *Topic* subsection into two subgroups: *Institutions* and *Research Objects and Related Elements* to ovoid compounding visibly distinct sequences, like *The World Health Organization* and *posterior circulation territory infarctlike lesions* on a same list. Table 3.7 shows the functional classification framework I used for MRAC multiword collocations. The ***bolded entries in italics*** indicate additions made based on the observed functions of MRAC multiword collocations.

For MCRC multiword sequences, I adapted the framework presented above and used for the functional classification of 2-word collocations. All sequences were functionally analyzed in context in *Antconc* (2020), using the *Concordance* tool and also the *File View* tool in cases where more context was needed.

Table 3.7. Functional Classification Framework for MRAC Multiword Collocations (Adapted from Hyland, 2008a)

Categories	**Subcategories**	**Examples of MWC from the MRAC**
Research-oriented	Location (in time/place)	*low-income and middle-income countries,*
	Procedure (***experiments & interventions***)	*oral glucose tolerance test, induction of mixed chimerism*
	Quantification	*estimated glomerular filtration rate, minutes at room temperature,*
	Description / identification focus	*the primary outcome measure*
	Topic (related to field of study)	
	- ***Institutions***	*the institutional review board, the World Health Organization*
	- ***Research object and related elements***	*seafood omega-3 fatty acids, de novo cholesterol synthesis*
Text-oriented	Transition signals	
	Resultative signals	
	Structuring signals	*available on request*
	Framing signals	*death from cardiovascular causes*
Participant-oriented	Stance features	
	Engagement features	*is worth noting*

3.5 Identification and Analysis of Lexical Bundles

In order to answer RQ 2c that asked about the use of lexical bundles in the two registers, and how they compare to bundles previously identified in academic prose, lexical bundles were identified using a computer program called *Lexical Bundles Identification & Analysis Program* (LBiaP) (Cortes & Lake, forthcoming). The software identifies bundles of up to 9 words and automatically controls for both overlapping and interlocked sequences. Bundles of 3 to 9 words were identified in LBiaP. Then, following Cortes's (2013) explorative approach I continued the search for longer bundles in *Antconc* until the search yielded no more results. Table 3.8 shows the frequency thresholds and ranges for the identification of bundles in both corpora. As overlaps had already been taken care of by LBiaP for bundles of 3 to 8 words, I manually checked for overlaps in lists of bundles of 9 or more words, using the same process described above for collocations and multiword collocations. After that cleaning process, all lists of bundles were ready for analysis.

Table 3.8. 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

The structural classification used for lexical bundles in this study is based primarily on frameworks proposed by Biber et al. (1999), Hyland & Jiang (2018), and Cortes (in press). With the identified bundles as observations, I also added a few structures that did not appear in the three frameworks mentioned above. Table 3.9 shows the structural classification framework for bundles identified in the present study. The **bolded** categories and subcategories were added based on the observed structures of bundles identified in the two corpora.

I used the *Concordance* and *File View* tools in *Antconc* for the qualitative analysis in context of bundles identified in each corpus. The functional classification of MRAC bundles was based on Hyland's (2008a) framework and Cortes (in press) for 3-word bundles. Additionally, I subdivided the Structuring Signals subcategory (in the Participant-oriented category) into Text Reference, for sequences like *table s1 in the supplementary appendix, shown in figure*; *as described above* and Study Subject/Element Reference for bundles like *t cells in, participants in* *the, of the general population.* The final framework used for the functional classification of MRAC bundles is summarized in Table 3.10. The **bolded** category and subcategories represent the additions to Hyland's original framework. For MCRC bundles, I revised the frameworks used for the functional analysis of collocations based on the discourse functions of bundles in MCRs. Table 3.11 shows the final version of the framework I used for the functional classification of MCRC bundles.

Table 3.9. Structural Classification Framework for MRAC and MCRC Bundles. Adapted from Biber (1999), Cortes (in press), and Hyland & Jiang (2018)

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

Table 3.10. Functional Classification Framework for MRAC Bundles. Adapted from Hyland (2008a) and Cortes (in Press)

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

Table 3.11. Functional Classification Framework for MCRC Bundles.
Based on Gagnier et al.'s (2014) guidelines and frameworks by Biber et al. (2004), Cortes (in Press), and Hyland (2008a)

Categories	**Subcategories**	**Examples from the MCRC**
Case-related (presentation and discussion of case)	Description	
	- medical condition-related	*we present the case of, has been shown to be*
	- Subject-related	*a woman in her, man with a history of, she had a*
	Location (time/place)	*at the time of admission, to the emergency department*
	quantification	*a high index of suspicion, at a dose of*
Diagnosis & Intervention-related (writers' activities and decisions)	Procedure	*she was treated with, the patient underwent a*
	Result Reporting	*with a blood pressure of, was positive for*
	Decisions/ Outcome	*the decision was made to, complete resolution of*
Discourse Organizers	Transition signals	*in addition to the, on the other hand, as well as*
	Resultative signals	*as a result of, due to the presence of*
	Structuring signals	
	- Text reference	*are shown in, reported in the, in this report*
	- Study subject reference	*of these patients, of our patient*
	Framing signals	*based on the, depending on the, in the absence of other*
Grammatical only		*and did not, in which the, he did not, it has been*
Stance & Engagement	Stance features	*may lead to, it is possible that, may be associated with*
	Engagement features	*should be considered in, physicians should be aware of*

3.6 Identification and Analysis of Frames

As motioned in chapter 2, section 2.2.4., various approaches have been used for the identifications of frames. For the purpose of the present study, I used a fully inductive approach (with the floor set at 1) as opposed to the bundle-to-frame approach to ensure that high frequency frames that do not involve bundles are included in the analyses. I also decided to include only frames with internal variable slots as these sequences have been shown to be more characteristic of academic writing (Biber, 2009). I thus operationalized frames as sequences of three of more words with internal variable slots occurring at least one time in each corpus.

3.6.1 Identification of Frames

I used both *Antgram* (Anthony, 2020) and *KfNgram* (Fletcher, 2007) for the identification of frames. *Antgram* offers the advantage of directly selecting "inner slots" and it automatically computes the type/token ratio (TTR) and entropy scores, both needed for the analysis of frame variability and predictability, respectively. *KfNgram* does not have information on TTR or entropy, but it provides information on the total number of fillers for each frame and lists all fillers and their frequency of occurrence in the frames, which is essential for the classification of fillers into semantic categories and the functional analysis of frames. I primarily worked with the *Antgram* lists and referred to *KfNgram* lists for the functional analysis of frames and their fillers. Given the amount of data at hand, I decided to set the following frequency thresholds for frames to consider for analysis:

- 40 times pmw for 3-gram frames,
- 20 times pmw for 4-gram and 5-gram frames, and
- 10 times pmw for 6-gram frames.

I decided to include only 3-gram – 6-gram frames as longer frames were found to be less interesting for analysis. The majority of these sequences were relatively fixed with very limited numbers of fillers as shown in figure 3.4, with the top 3 10-gram frames. The numbers in red squares indicate the number of fillers for each frame. Note the Zipfian distribution of fillers for the third frame, indicating an almost exclusive use of the filler *study* in the frame.

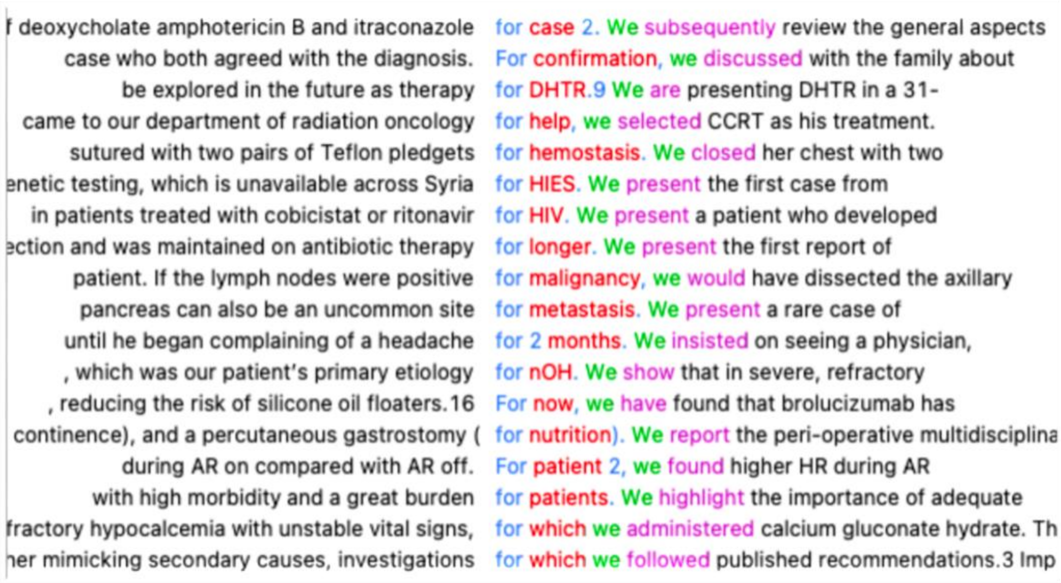
```
supplemental figure * supplemental material available online with this article      23     2
supplemental figure 1 supplemental material available online with this article      14
supplemental figure 1a supplemental material available online with this article     9

supplemental * 1 supplemental material available online with this article   23   2
supplemental figure 1 supplemental material available online with this article       14
supplemental table 1 supplemental material available online with this article        9

the * had no role in study design data collection  22    4
the study had no role in study design data collection     17
the funder had no role in study design data collection     3
the sponsor had no role in study design data collection    1
the funders had no role in study design data collection    1
```

Figure 3.4. Examples of Relatively Fixed Long Frames

The next step after identifying frames that met the above thresholds was to clean the initial lists of candidates. Two types of frames were removed from the lists: frames that primarily included numbers (e.g., *95 ci * 1, 0 * p 0, 0 * 95 ci*) and non-unit frames (e.g., *or * and, the * and, of * and the, for * we*). I refer to the latter as non-unit as even when associated with their fillers, they do not convey much meaning (if at all). I identified such sequences in context and many of these sequences bridged punctuations, as shown in Figure 3.5., with the frame *for * we.* Non-unit frames may be problematic to analyze both functionally and in terms of variability and predictability. They tend to be highly variable but show no real patterns in the fillers they select. For example, the frame *of * and the* occurred 104 times in the MRAC and had 80 fillers. The first six fillers all occurred 2 times and all other fillers occurred only once in the frame. The list of fillers included words such as *delivery, enrollment, presentation, participants, life, data, cancer, Singapore, research*, etc. The frame was also found to bridge punctuations in many instances. Once these two types of frames were removed from the lists, the next step was to attend to overlaps to obtain the final lists for analysis.



f deoxycholate amphotericin B and itraconazole	for case 2. We subsequently review the general aspects
case who both agreed with the diagnosis.	For confirmation, we discussed with the family about
be explored in the future as therapy	for DHTR.9 We are presenting DHTR in a 31-
came to our department of radiation oncology	for help, we selected CCRT as his treatment.
sutured with two pairs of Teflon pledgets	for hemostasis. We closed her chest with two
enetic testing, which is unavailable across Syria	for HIES. We present the first case from
in patients treated with cobicistat or ritonavir	for HIV. We present a patient who developed
ection and was maintained on antibiotic therapy	for longer. We present the first report of
patient. If the lymph nodes were positive	for malignancy, we would have dissected the axillary
pancreas can also be an uncommon site	for metastasis. We present a rare case of
until he began complaining of a headache	for 2 months. We insisted on seeing a physician,
, which was our patient's primary etiology	for nOH. We show that in severe, refractory
, reducing the risk of silicone oil floaters.16	For now, we have found that brolucizumab has
continence), and a percutaneous gastrostomy (	for nutrition). We report the peri-operative multidisciplina
during AR on compared with AR off.	For patient 2, we found higher HR during AR
with high morbidity and a great burden	for patients. We highlight the importance of adequate
fractory hypocalcemia with unstable vital signs,	for which we administered calcium gluconate hydrate. Th
her mimicking secondary causes, investigations	for which we followed published recommendations.3 Imp

Figure 3.5. Example of Frame Bridging Punctuation

3.6.2 Structures, Variability, and Predictability of Frames

In order to answer RQ 3a (How does the use of phrase frames compare in the two registers? Are there any variations in terms of predictability, variability, and structures?), the following procedures were implemented. For the structural analysis, frames of 3 – 6 elements from both corpora were classified into the following three major structural categories proposed by Gray & Biber (2013, p.122). The examples provided for each structural group are from the MRAC and MCRC.

- **Verb based frames:** frame contains one or more modal, auxiliary, or main verb (e.g., *should be * in, were * overnight at, presented to the * department*).
- **Frames with other content words:** frame contains one or more nouns, adjectives, or adverbs but no verbs (e.g., *a * year old female, in * patients, these findings * that*).

- **Function word frames:** frame consists of only function words such as prepositions, determiners, conjunctions, pronouns, complementizers, etc. (e.g., *the * of the, we * that, at the * of the*).

The variability and predictability of frames from both corpora were determined based their TTR and entropy scores, respectively. To determine the degree of variability of frames, I used the same thresholds as Gray & Biber (2013) to classify the analyzed frames into the following three groups:

- Highly variable: TTR >. 70,
- Variable: .30 ≤ TTR ≥ .70, and
- Relatively fixed: TTR < .30

For the predictability of frames, the analyzed sequences were classified into the following three groups based on their entropy scores:

- High Predictability: $H_{\text{norm}} < .30$,
- Moderate Predictability: $.30 \leq H_{\text{norm}} \geq .70$, and
- low Predictability: $H_{\text{norm}} > .70$.

3.6.3 Functional Analysis of Frames

To analyze the discourse functions of frames and their fillers, answering RQ3b (How can the grouping of fillers by semantic domains inform the functional analysis of frames? What are the main functions served by some of the most salient frames and their fillers in MRAs and MCRs? Are there any variations between the two registers?), I developed a functional classification framework for each structural category of frames. I created the frameworks with the pedagogical application of the functional analysis of frames in mind. Each framework is designed to allow a fully exploratory approach. With the frames as observations, I first identified subgroups in each structural category. Then the next step was to create a "skeleton" framework for each structural category that the researcher, instructor, or independent learner can populate as they discover the functions of frames and their semantic groups of fillers. Subgroups can be added or deleted based on lists of frames under investigation. The functional classification frameworks for verb-based, lexical word-based, and content word-based frames are shown together in Table 3.12. The thus designed frameworks were used for the functional classification of some selected frames and their fillers in both corpora.

For the MRAC frames, I decided to analyze only frames in the 4-gram list. This decision was motivated by two main reasons. First, with the amount of data at hand and the time imparted for the completion of the present study, it did not appear realistic to analyze all lists of frames and their fillers. The second reason was that the 4-gram list was found to include a higher proportion of variable frames and all structures were well-represented in that list. That made the list a representative sample for analysis.

The next step was to classify the fillers of 4-gram frames into semantic categories. To that end, I first excluded frames that involved function words other than modal auxiliaries, like the frame *in * placebo groups* (fillers: *the* and *both*), for example. Then for the each of the remaining frames, I identified all fillers that occurred at least five times in the frame and listed them in an Excel spreadsheet (see Figure 3.6). Then I manually classified the fillers of each frame into semantic categories, using concordance lines in case of ambiguity. I also used concordance lines to determine the discourse functions of analyzed frames and their semantic groups of fillers (see Appendices D for example of list of frames and their most frequent fillers).

Table 3.12. Functional Classification Frameworks for the Main Structural Groups of Frames

VERB-BASED FRAMES				
Subgroups	**Filler word class**	**Filler sematic category**	**Function**	**Examples**
Auxiliary-initial				*are * in the, did not * any, is * seen in, has been * to*
copular be-initial				*is a * condition, were * in the*
Determiner/Noun/ Pronoun-initial				*the * was a, it is * for, patient was * to, cells were * for*
Modal-initial				*may be * for, can be * as, should be * in*

CONTENT WORD-BASED FRAMES				
Subgroups	**Filler word class**	**Filler sematic category**	**Function**	**Examples**
Determiner/Noun/ pronoun-initial				*a * increase in, the * effects of, completeness and * of the data, we * the effect of*
Preposition-initial				*at the * level, for the * group, in the * term, on the basis of * results*
to-infinitive initial				*to * the effect of*
Adj/Adv-initial				*significantly * in the, due to * rarity*

FUNCTION WORD-BASED FRAMES				
Subgroups	**Filler word class**	**Filler sematic category**	**Function**	**Examples**
Determiner-initial				*the * of the, a * of,*
preposition-initial				*at a * of,* for * in the, in the * of the
to-infinitive initial				*to * whether the*
pronoun-initial				*we * that the, we * a*
wh-word initial				*which * the, who * with*
that-initial				*that* in the, that* with*

frame	nber of fillers	
the * of the	507	end, use, basis, results, course, effect, time, duration, start, magnitude, design, analysis, presence, characteristics, ability, fide sensitivity, capacity, integrity, writing, sum, distribution, strength, findings, quality, proportion, effectiveness, control, effects, function, robustness, significance, efficacy, sponsor, funders, level, incidence, day, remainder, regulation, bottom, publication, surface, ratio, rate, development, activity, assessment, immunogenicity, spread, details, calculation, head, validity, adequacy, diversity, implementation, conclusion, percentage, severity, association, goal, principles, value, activation, standards, exceptio combination, peak, coefficients, intensity, genotype, base, release, region, cells, kinetics, vicinity, progression, mechanism, de safety
in the * of	325	presence, absence, context, number, pathogenesis, setting, incidence, risk, lungs, development, rate, case, regulation, treatme range, course, use, rates, analysis, management, face, form, feces, blood, induction, subgroup, control, distribution, levels, reg country, concentration, formation, middle, majority, csf, livers, duration, amount, brains, production, cytoplasm, fibroblasts, ac circulation, study, extent, microenvironment, percent, assessment, progression, tissues, quality, effect, numbers, transport, ra center, cbn, shape, spleens, degree, hippocampus, modulation, pathophysiology
in the * group	141	placebo, control, intervention, iron, screening, early-therapy, delayed-therapy, simvastatin-ezetimibe, avelumab, aspirin, telmi varenicline, migraine, preterm, calcineurin-inhibitor, 20-mg, atrasentan, dexamethasone, vtd, amg-157, standard-treatment, r lifestyle, malaria-vaccine, detox, supine, nicotine-replacement, prone, ivabradine, standard, nonintervention, non-invasive, cas ak002, invasive, enhanced-lifestyle-counseling, treatment-failure, treatment, ili, dse, tdf-ftc, tight-control, cbt, lifestyle-couns high-risk, placebo-followed-by-cbt, vaccine, usual-care
on the * of	129	basis, surface, day, incidence, risk, number, progression, presence, results, importance, role, effect, use, development, ability, expression, proportion, utilization, strength, percentage
of the * of	228	number, effect, percentage, use, university, effects, prevalence, lack, nature, proportion, results, burden, declaration, absence, intensity, area, regulation, mechanism, initiation, duration, paucity, concentration, quality, development, ministry, frequencies biology
to the * of	226	development, use, end, number, pathogenesis, risk, level, lack , treatment, intensity, date, expression, effect, initiation, induct start, criteria, sera, release, activation, site, left, activity, addition, amount
with the * of	141	use, exception, addition, results, declaration, number, risk, presence, rest, combination, expression, support, findings, lack, eff
for the * of	161	treatment, development, presence, detection, duration, comparison, prevention, analysis, fidelity, management, lack, purpose production, percentage, design, control, identification, initiation, regulation, remainder, ability, loss, measurement, pathogene
of * in the	278	patients , participants, children, cells, decline, death, variation, screening, mice, differences, bas, stay, treatment, individuals,
the * of a/an	49	use, absence, results, presence, development, effect, measurement

Figure 3.6. Screenshot of Spreadsheet Used to List Frames and their Frequently Occurring Fillers

I used Biber's (2006) taxonomy for the semantic categories of major word classes (nouns, verbs, adjectives, adverbs, and modals). I slightly adapted a couple of categories to allow a more accurate description of the patterns I observed. For nouns, the semantic category of ABSTRACT/ PROCESS appeared to be too broad. From my observations, different types of abstract noun fillers served different functions when associated with certain frames. Therefore, I subdivided abstract nouns into three subgroups: (a) ABSTRACT/ PROCESS, for nouns expressing only process (e.g., *mechanism*, *implementation*, *activation*, *neutralization*); (b) ABSTRACT/ATTRIBUTE, for nouns referring to the characteristics or attributes of something or someone (e.g., *characteristics*, *effectiveness*, *complexity*, *density*); and (c) ABSTRACT /OTHER for other less specific abstract nouns (e.g., *findings, conclusions, base, basis, presence*). Still in the semantic categories of nouns, I changed the category TECHNICAL/ CONCRETE to just TECHNICAL as I noted that technical nouns occurring in a same frame could be concrete or abstract without affecting the function of the frame. Finally, in the semantic categories of verbs, I added INTENTION for verbs like *plan, decide,* or *intend* as some frames were found to have specific discourse functions when associated with this type of verbs. Table 3.13 shows the final semantic category taxonomy I used to classify the fillers of the analyzed frames. The shaded cells indicate semantic categories of fillers I did not identify in the analyzed frames.

For MCRC frames, I decided to analyze only 4-gram verb-based frames. These frames appeared particularly salient as they included sequences related to functions specific to MCRs. In fact, the presence of such sequences made the MCRC 4-gram list of frames more than twice longer than its counterpart in the MRAC. Therefore, I considered that the analysis of 4-gram verb-based frames could bring a non-negligible contribution to the description of the formulaic profile of MCRs and the discussion of differences between MRAs and MCRs. To analyze the discourse functions of MCRC 4-gram verb-based frames, I used the same approach described above for MRAC 4-gram frames. After the analysis of selected frames in each register, I discussed the observed similarities and differences. That discussion was supplemented by the functional analysis of the shared frames initially identified.

3.6.4 Bundles in Frames and Variability of Semantic Groups within Frames

I answered the first sub-question of RQ3c that asked about the variability of semantic domains within frames based on the findings of the functional analysis of selected MRAC and MCRC, described above. To answer the second subquestion of RQ 3c that asked whether the presence of a bundle in a frame affected the variability of the semantic categories of the other fillers, I used the *KfNgram* lists of 4-gram frames identified in the MRAC and MCRC, as they already included the fillers of the frames.

Table 3.13. Semantic Category Taxonomy Used for the Grouping of Analyzed Frame Fillers (Adapted from Biber, 2006)

Word Class	Semantic Category	Examples
Noun	animate	*members, sponsors, sponsor, funders, funder,*
	cognitive	
	concrete	
	technical	*pathogenesis, phenotype, cells, blood, pancreas*
	quantity	*course, magnitude, size, rest, mean, date, sum, proportion,*
	place	*bottom, surface, head, area, center*
	group/institution	*committee, board*
	abstract/process	*use, design, analysis, development, calculation, combination*
	abstract/attribute	*nature, characteristics, importance, specificity, ability*
	abstract other	*basis, presence, control, absence, conclusion*
Verbs	be as main verb	*be*
	activity	*conduct, perform, create*
	communication	*reveal, show, demonstrate*
	mental	*hypothesize, know, think, believe*
	causative	*readmit, admit, transfer*
	occurrence	*increase, decrease, become*
	existence	*possess, have,*
	aspectual	*keep, remain*
Modals	possibility/permission/ability	*may, can, might,*
	necessity/obligation	*should*
	predictive/volition	*will, would*
Adjectives	size attributive	*population-wide, greater, low, high*
	time attributive	*long-term, young, recent*
	color attributive	*green, red, blue*
	evaluative attributive	*significant, substantial, striking, modest, marked, clear*
	relational attributive	*individual, general, secondary, primary*
	topical attributive	*immunomodulatory, antiproliferative, nontransgenic*
Adverbs	certainty	*obviously, certainly*
	likelihood	*typically, commonly*
	style	*strongly, significantly, mainly*
	attitude	

I first copied and pasted side by side the lists of 4-gram frames and 4-word bundles from each corpus on an Excel spreadsheet. Then, using the *Conditional Formatting* tool in Excel, I highlighted all bundles in the list of frames. Figure 3.7 shows some highlighted bundles in the most frequent 4-gram frame in the MRAC (*the * of the*). I next deleted all non-highlighted sequences to create a list of frames and the bundles they involve. Then I compared the semantic categories of other recurrent fillers (already identified during the functional analysis) and the fillers forming the lexical bundles.

	A	B	C
1		**Frames**	**Bundles**
2	1	the * of the	in the placebo group
3		the end of the	on the basis of
4		the use of the	in the control group
5		the basis of the	in the presence of
6		the results of the	in the absence of
7		the effect of the	at the time of
8		the course of the	with the use of
9		the time of the	in the supplementary appendix
10		the design of the	in the intervention group
11		the presence of the	these data suggest that
12		the magnitude of the	in the context of
13		the duration of the	in the united states

Figure 3.7. Example of Bundles in Frame Highlighted in Excel

3.7 Conclusion

I have described in this chapter the corpora I collected to represent the two registers under scrutiny in the present study, and the methodology and procedures I used to answer each research question. In the next chapters, I also provide additional information on some aspects of the methodology for more convenient reading and interpretation of the findings being presented.

4 CHAPTER 4. SITUATIONAL CHARACTERISTICS OF MRAs AND MCRs

As one of the main goals of the present study is to analyze different types of formulaic sequences in MRAs and MCRs and to compare these expressions across the two registers, it is essential to describe these types of texts in terms of their external characteristics that define them in the contexts in which they are used (Biber & Conrad, 2009; Gray, 2015). As mentioned in section 3.1. of the Methodology chapter, I adapted Gray's (2015) framework for the situational analysis of MRAs and drew from both Biber & Conrad (2009) and Gray's adapted framework to analyze the situational characteristics of MCRs. In the next sections, I provide a detailed description of the situational characteristics of MRAs and MCRs as well as the key characteristics that define them as two distinct registers. The findings reported in this chapter will inform the interpretation of the use of formulaic sequences investigated in the present study.

4.1 Situational Characteristics of MRAs

In an aim to have a representative sample from all journals used for the corpus collection, The analyses of the situational characteristics of both MRAs and MCRs were based on subsets of 100 research articles from the journals described in chapter 3, section 3.1. As mentioned in section 3.1.1, only quantitative MRAs written in the IMRD format (not necessarily in this order) were included in the present study. To analyze the situational characteristics of MRAs, I adapted Gray's (2015) framework, based on the observations and information gathered from journal websites, expert-informants, and the selected articles. The few additions I made to Gray's (2015) framework primarily pertain to *Participants*, *Nature of Evidence,* and *Explicitness of Research Design.*

For the category of participants characteristics, I considered that the difference in professional training could be determinant of some linguistic variations in the two registers. Therefore, I included *Professional Training/Title* as a second factor in this category, with the following three distinguishing elements: *PhD* for researchers, *MD/MBBS* for practitioners, and *Both*. MD stands for Medical Doctor, and MBBS is an international equivalent of MD in the United States. This information is apposed to authors' names, under the title of each article. Some authors may have both titles, which explains the presence of the third factor (Both) under *Professional Training/ Title*. Additionally, given that authors' contributions are provided at the end of most MRAs, it is possible to know exactly how many authors participated in the actual writing of the articles. Therefore, I also added *Actual writers* as another element to consider under *Participants.* The final factor I added in this category of situational characteristics is *Audience*. Both MRAs and MCRs are addressed to the medical community but given that the medical community includes researchers, practitioners, medical students, and patients (Chapman et al., 2014), this factor can be determinant in the comparison of the two registers.

For the nature of evidence, the option *mathematical formula* under *Primary Presentation of Evidence* was removed from the framework. Gray (2015) included this factor as her corpus included theoretical physics research articles, but here, the inspection of the results sections of the MRAs (and MCRs) indicated that this factor was not relevant to the registers under study. There were some doubts regarding the use of some formulae like those bolded in example (4.1) and (4.2) below, but insights from expert consultants led to the decision to refer to these formulae as medical formula/shorthand rather than mathematical formula. Therefore, *Prose with medical formula/shorthand* was added as a way of presenting evidence in MRAs. I also added *Prose and figures with extensive description* as MRA authors make extensive use of figures to present evidence. The factor *Prose discussion* was removed from the framework as a quick perusal of the result sections of the MRAs and MCRs did not reveal any use of extensive prose without the figures or medical formulae or shorthand mentioned above.

> *(4.1)* Laboratory findings showed a white cell count (WCC) count of **28.4×109/L with 90.7%** neutrophils, haemoglobin (Hb) level **of 6.7 g/dL,** and a platelet (Plt) count of **11.0×109/L** with schistocytes and dacrocytes (figure 1). His serum AST level **was 93 U/L, ALT 13 U/L**, lactate dehydrogenase (**LDH) 1572 U/L,** creatinine **67.5 µmol/L,** ferritin **15 500 ng/mL** and **CRP 9.40 mg/dL**. (MCRC_20*BMJ*8)

> *(4.2)* The activated DCs then produce **IFN-α, IFN-β, IL-12, IL-23, IL-6**, and **TNF-α,** which activate and polarize autoaggressive Th cells toward **Th1, Th17, and Th22** cell subsets, as well as **γδ T cells** toward the **γδ T17** (**IL-17A–producing γδ T cells)** subset, resulting in an immune imbalance of T cells (**4–6).** Interestingly, **Tcrb−/−** mice treated with **IMQ** had significantly decreased percentages of dermal **γδ T17** cells, whereas **IMQ** treatment of **Tcrd −/−** mice had no effect on dermal **Th17** cells (Supplemental Figure 9G), but both showed comparatively decreased neutrophil infiltration (Supplemental Figure 9G). However, splenic **Th17 or γδ T17** cells were largely unchanged in **Tcrd−/− and Tcrb−/−** mice treated with **IMQ**, respectively. *(MRAC_20JCI5)*

Finally, under *Explicitness of Research Design,* I added *in-text reference to cited authors* with the options *yes* and *no*, as this also could be determinant of some linguistic choices. In most MRAs, references to previous studies are easily noticed in the texts as they are followed by superscript numbers – as shown in example (4.3) below – that correspond to numbered studies in the reference list. It was assumed that whether writers mention cited authors in the text or not could be determinant of linguistic features used to refer to previous research. After these slight changes, the thus adapted framework was used to analyze the situational characteristics of MRAs, summarized in Table 3.1.

> *(4.3)* Atrial fibrillation (AF) is the most common cardiac arrhythmia observed in clinical practice, with more than 5 million people experiencing AF in the US alone.[1,2] Atrial fibrillation is associated with increased stroke and systemic embolism rates and increased

morbidity and mortality.[1] Anticoagulant treatment reduces the risk of stroke by approximately 65% in patients with nonvalvular AF.[3] Almost one-half patients at risk of experiencing stroke do not start, and a similar proportion do not continue, to receive anticoagulant treatment and experience preventable strokes.[4-7] *(20JAMA10)*

Table 4.1. Situational Characteristics of MRAs

Characteristics	**MRAs**
Participants	
Authors	
1	0
2-4	4
5+	96
Actual Writers	
1	6
2	26
3-4	24
5+	23
not disclosed	20
Professional Training/Title	
PhD (researcher)	84
MD/MBBS (Practitioner)	0
Both	16
Primary Audience	Researchers, medical practitioners
Textual Layout & Organization	
Length	
Mean word count	5,451.08
Standard deviation	1,800.33
Headings	
None	0
Un-numbered	100
Numbered	0
Use of Abstracts	
yes	100
no	0
Visual Elements	
None	0
Tables	1
Figures	10
Tables & Figures	89
Sections/Organization	
IMRD	60
IMRD in other order	40
Other	0
Standardized section heading	100
Variable section heading/name	0

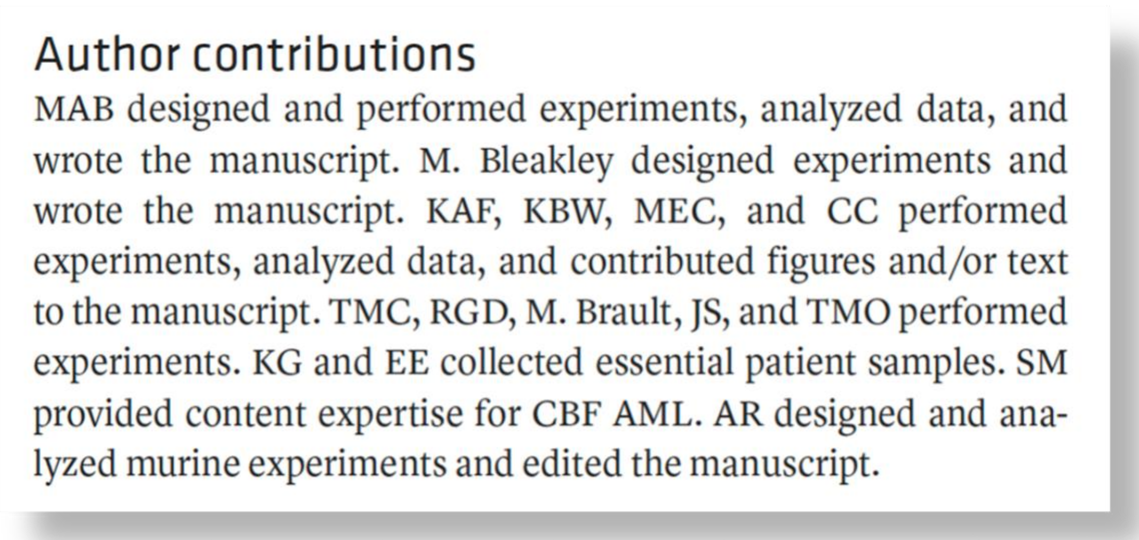
Table 4.1 (Cont'd)

Characteristics	**MRAs**
Setting	
Nature of Journal	
generalist	100
specialized	0
Subject/Topic	
General Topic	medical pathologies and treatments, medication experimentation
Purpose	
General Academic Purpose	To report on the analysis of observed data to advance the field of medicine
Nature of Data or Evidence	
Presence of Observed Data	
yes	100
no	0
Use of Numerical Evidence	
yes	100
no	0
Object of Study	medical pathologies & treatments
Primary Presentation of Evidence	
Prose with medical formula/shorthand	36
Quantitative displays	19
Prose & Figures with extensive description	45
Methodology	
observational	14
experimental	86
Statistical Techniques	
None	0
Descriptive Statistics	100
Statistical Difference Testing	100
Other advanced statistics	86
Explicitness of Research Design	
Explicitness of Purpose	
Direct statement	97
Minimal / No statement	3
Explicitness of RQs	.
Direct statement	20
Minimal / No statement	80
Hypothesis / Hypotheses	17
Explicitness of Citations	
Within text	100
In notes	0
in-text reference to cited authors	
yes	0
No	100
Explanation of Evidence	
extensive	100
mention / none	0
Explanation of Procedures	
extensive	100
mention / none	0

After this summary of the additions to Gray's (2015) framework, I present a description of some key situational characteristics of MRAs in the next section. These characteristics will be later compared to the situational characteristics of MCRs described in section 4.3.

4.1.1 Participants

Medical research articles are typically multi-authored, with 96 out of the 100 analyzed articles having more than five authors, 47 of which had 10 or more authors. However, most of the analyzed MRAs are actually written by two to four authors, as can be noted on Table 4.1. Gray (2015) suggests that this multiplicity of authors is inherent to research in the hard sciences that involve numerous experiments and a variety of equipment to manipulate. This appears to be the case in MRAs as illustrated by the *Author contributions* sections provided at the end of each article. Figure 4.1 show an example of this section describing the roles of each author in the study. Most of these authors are researchers (86%).



Author contributions

MAB designed and performed experiments, analyzed data, and wrote the manuscript. M. Bleakley designed experiments and wrote the manuscript. KAF, KBW, MEC, and CC performed experiments, analyzed data, and contributed figures and/or text to the manuscript. TMC, RGD, M. Brault, JS, and TMO performed experiments. KG and EE collected essential patient samples. SM provided content expertise for CBF AML. AR designed and analyzed murine experiments and edited the manuscript.

Figure 4.1. Example of 'Author contribution' Section in an MRA Describing the Roles of Authors (20JCI10)

The audience of MRAs appears to primarily be researchers and medical practitioners, based on information gathered from journal websites and field-expert informants. *Science Translational Medicine,* for example, states in its guidelines to authors that the journal welcomes articles that "represent significant advance in research" and "report successful progress toward improvements in clinical medicine" (*STM*, 2021, Information for Authors). Given the audience of MRAs, it can be assumed that there is a certain level of shared knowledge, which probably explains why MRAs are perceived by non-specialists as "highly technical" (Nwogu, 1997, p. 119).

4.1.2 Textual Organization and Layout

In terms of textual organization, MRAs follow a very consistent format. All analyzed articles include an abstract and the only noted difference regards the ordering of the sections, which are determined by journal guidelines. Articles published in *JAMA, The Lancet,* and *NEJM* follow the more common IMRD-format, while the sections of MRAs in *Science* and *JCI* are typically presented in the following order: Introduction, Results, Discussion, and (Materials and) Methods. The vast majority of abstracts in MRAs are very structured with distinct subheadings, defined by journal guidelines. The few exceptions were articles published in *JCI* where authors are explicitly required to write abstracts in "one single paragraph". Nevertheless, the guidelines also require authors to include the "rationale, objectives, findings, and conclusions" (*JCI*, 2021, Author Information Center). Figure 4.2 shows an example of the predominant type of abstracts found in MRAs.

IMPORTANCE Shared decision-making (SDM) about anticoagulant treatment in patients with atrial fibrillation (AF) is widely recommended but its effectiveness is unclear.

OBJECTIVE To assess the extent to which the use of an SDM tool affects the quality of SDM and anticoagulant treatment decisions in at-risk patients with AF.

DESIGN, SETTING, AND PARTICIPANTS This encounter-randomized trial recruited patients with nonvalvular AF who were considering starting or reviewing anticoagulant treatment and their clinicians at academic, community, and safety-net medical centers between January 30, 2017 and June 27, 2019. Encounters were randomized to either the standard care arm or care that included the use of an SDM tool (intervention arm). Data were analyzed from August 1 to November 30, 2019.

INTERVENTIONS Standard care or care using the Anticoagulation Choice Shared Decision Making tool (which presents individualized risk estimates and compares anticoagulant treatment options across issues of importance to patients) during the clinical encounter.

MAIN OUTCOMES AND MEASURES Quality of SDM (which included quality of communication, patient knowledge about AF and anticoagulant treatment, accuracy of patient estimates of their own stroke risk [within 30% of their estimate], decisional conflict, and satisfaction), decisions made during the encounter, duration of the encounter, and clinician involvement of patients in the SDM process.

RESULTS The clinical trial enrolled 922 patients (559 men [60.6%]; mean [SD] age, 71 [11] years) and 244 clinicians. A total of 463 patients were randomized to the intervention arm and 459 patients to the standard care arm. Participants in both arms reported high communication quality, high knowledge, and low decisional conflict, demonstrated low accuracy in their risk perception, and would similarly recommend the approach used in their encounter. Clinicians were significantly more satisfied after intervention encounters (400 of 453 encounters [88.3%] vs 277 of 448 encounters [61.8%]; adjusted relative risk, 1.49; 95% CI, 1.42-1.53). A total of 747 of 873 patients (85.6%) chose to start or continue receiving an anticoagulant medication. Patient involvement in decision-making (as assessed through video recordings of the encounters using the Observing Patient Involvement in Decision Making 12-item scale) scores were significantly higher in the intervention arm (mean [SD] score, 33.0 [10.8] points vs 29.1 [13.1] points, respectively; adjusted mean difference, 4.2 points; 95% CI, 2.8-5.6 points). No significant between-arm difference was found in encounter duration (mean [SD] duration, 32 [16] minutes in the intervention arm vs 31 [17] minutes in the standard care arm; adjusted mean between-arm difference, 1.1; 95% CI, −0.3 to 2.5 minutes).

CONCLUSION AND RELEVANCE The use of an SDM encounter tool improved several measures of SDM quality and clinician satisfaction, with no significant effect on treatment decisions or encounter duration. These results help to calibrate expectations about the value of implementing SDM tools in the care of patients with AF.

Figure 4.2. Example of Structured MRA Abstract (20JAMA10)

4.1.3 Topic and Purpose

MRA topics may differ, given the multiple areas covered by general medicine, but all subjects relate to medical pathologies and treatments, and the main purpose of MRAs is to report observed data pertaining to these subjects that are generalizable and can advance the field of medicine, as previously noted in *STM'*s guidelines for authors. This probably explains the extensive description of methods and results noted in all 100 analyzed articles. It should be noted however that despite this extensive description of evidence, MRAs have a relatively limited word count (*M* = 5,451.08, *SD* = 1,800.33), which suggests a certain level of concise writing. This limited word count is determined by journals in the instructions to authors. *The Lancet* for example, allows between 3,500 and 4,500 words.

4.1.4 Explicitness of Research Design

In terms of explicitness of the research design, the purpose of the study is always clearly stated in the abstract, regardless of format, as previously shown in the "OBJECTIVE" section in Figure 4.2 and in example (4.4) below, showing an abstract written in a single paragraph.

> *(4.4)* Psoriasis is a severe disease associated with the disturbance of metabolism and inflammation, but the molecular mechanisms underlying these aspects of psoriasis pathology are poorly understood. **Here, we report** that glutaminase 1–mediated (GLS1-mediated) glutaminolysis was aberrantly activated in patients with psoriasis and in psoriasis-like mouse models, which promoted Th17 and γδ T17 (IL-17A–producing γδ T) cell differentiation through enhancement of histone H3 acetylation of the Il17a promoter, thereby contributing to the immune imbalance and development of psoriasis. **We further demonstrate that** mucosa-associated lymphoid tissue lymphoma translocation protein 1 (MALT1) protease was constitutively active in psoriatic CD4+ and γδ T cells, thereby supporting GLS1 expression by stabilizing c-Jun, which directly binds to the GLS1 promoter region. (20*JCI*7).

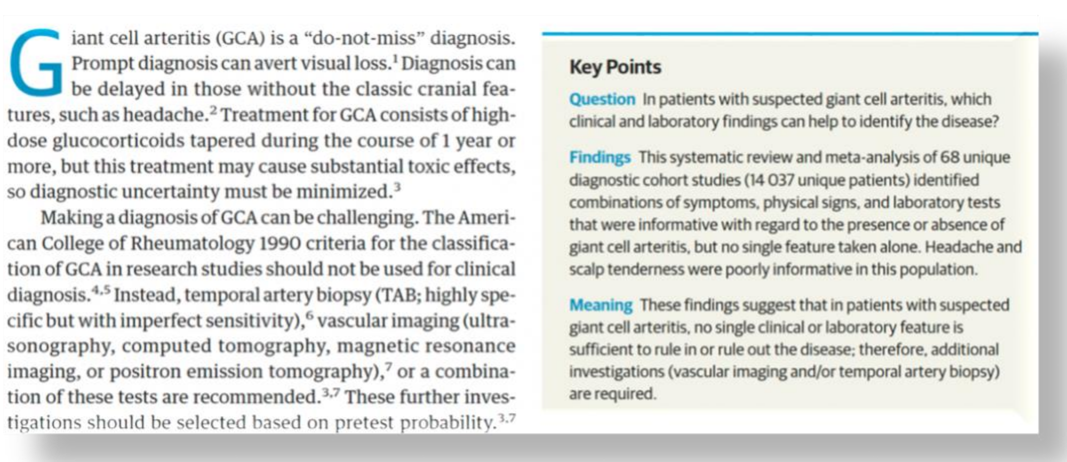
The purpose of the study is usually iterated throughout the paper: in the last paragraph of the introduction (example 4.5), in the methods section (example 4.6), and sometimes in the result section, before the presentation of each set of results, or in the first paragraph of the Discussion section. This iteration of the purpose can involve frequent use of signal phrases (**bolded** in examples 4.4-4.6) announcing the aim of the study.

> *(4.5)* **We set out to address** uncertainties about the early host immune response to ZIKV, once our longitudinal studies in pregnant macaques were completed and passive maternal immunity had waned in offspring. **Here, we report** the longitudinal analysis of macaque offspring born to ZIKV-infected rhesus macaque dams from gestation through an extended

postnatal period. **We evaluated whether** the offspring born to ZIKV-infected mothers had acquired immunological memory to ZIKV that was sufficient to protect against ZIKV reexposure. (20*Science*4)

(4.6) Participants discontinued their current NSAID and took meloxicam daily during a 2-week run-in period. **To examine whether** placebo is noninferior to continued NSAID use, participants who remained eligible after the run-in period were randomized to receive meloxicam or placebo for 4 weeks (double-blinded phase 1). After 4 weeks, participants in the NSAIDs group continued meloxicam. Those in the placebo group stopped taking the placebo and participated in a 10-week telephone-based CBT program. T**he objective of the second phase was to determine whether** CBT (after placebo) is noninferior to continued NSAIDs. Placebo was not continued during phase 2 because it may potentiate the effects of CBT. (20*JAMA*8)

While the purpose of the study is always clearly stated, explicit research questions are very rare in MRAs. In that respect, MRAs are relatively similar to Biology research articles described by Gray (2015). The few articles with explicit statements of research questions were those published in *JAMA.* Even in those articles, the questions are not directly stated in the text. They are presented in a separate box in the introduction, labelled "Key Points", that features the question(s), the summary of the findings, and the importance of the study, as shown in figure 4.3.



Giant cell arteritis (GCA) is a "do-not-miss" diagnosis. Prompt diagnosis can avert visual loss.[1] Diagnosis can be delayed in those without the classic cranial features, such as headache.[2] Treatment for GCA consists of high-dose glucocorticoids tapered during the course of 1 year or more, but this treatment may cause substantial toxic effects, so diagnostic uncertainty must be minimized.[3]

Making a diagnosis of GCA can be challenging. The American College of Rheumatology 1990 criteria for the classification of GCA in research studies should not be used for clinical diagnosis.[4,5] Instead, temporal artery biopsy (TAB; highly specific but with imperfect sensitivity),[6] vascular imaging (ultrasonography, computed tomography, magnetic resonance imaging, or positron emission tomography),[7] or a combination of these tests are recommended.[3,7] These further investigations should be selected based on pretest probability.[3,7]

Key Points

Question In patients with suspected giant cell arteritis, which clinical and laboratory findings can help to identify the disease?

Findings This systematic review and meta-analysis of 68 unique diagnostic cohort studies (14 037 unique patients) identified combinations of symptoms, physical signs, and laboratory tests that were informative with regard to the presence or absence of giant cell arteritis, but no single feature taken alone. Headache and scalp tenderness were poorly informative in this population.

Meaning These findings suggest that in patients with suspected giant cell arteritis, no single clinical or laboratory feature is sufficient to rule in or rule out the disease; therefore, additional investigations (vascular imaging and/or temporal artery biopsy) are required.

Figure 4.3. Example of Presentation of the Research Question in an MRA (20JAMA2)

Finally, one last characteristic that can be determinant of linguistic choices is the citation practices in medical research articles. As previously explained and shown in example (4.3), cited authors are rarely mentioned in the MRAs, if at all. In fact, none of the 100 analyzed MRAs included direct mention of cited authors. Authors are numbered in the order they are cited, as shown in example (4.3) above and as a result, the reference list is numbered. This appears to be the recognized American Medical Association (AMA) style and is explicitly required in journals' guidelines for authors. For example, the guidelines on the *JAMA* website state: "Number references in the order they appear in the text; do not alphabetize" (*JAMA*, 2021, Instructions for Authors). As already mentioned, since the in-text mention of cited authors often entails the use of signal phrases (e.g., *x et al. suggested / have shown / reported/ claimed that …*), it is very likely that the citation practices in MRAs will determine some linguistic features used to report findings of previous research.

4.1.5 Nature of Data and Methodology

The vast majority of analyzed MRAs are experimental studies (86%). The remaining 14% were cohort studies. For example, one study investigated the 1-year mortality related to the COVID-19 pandemic. Data are analyzed using various statistical techniques (even in cohort studies) that are described in the Statistical Analysis subsection of the Methods section. This description is almost always supplemented by additional material online. The presentation of results involves multiple tables and figures, and those figures are accompanied with long legends that entail extensive use of descriptive language, including a lot of passive constructions, as shown in Figure 4.4 (yellow-shaded). Given that the number of tables and figures are limited to 5 – 8, depending on journals, additional figures and/or tables are often provided online, and readers are directed to those supplementary materials when necessary. Often, once the first reference to supplemental materials has been made, the other references are done with minimal text as shown in example (4.7).

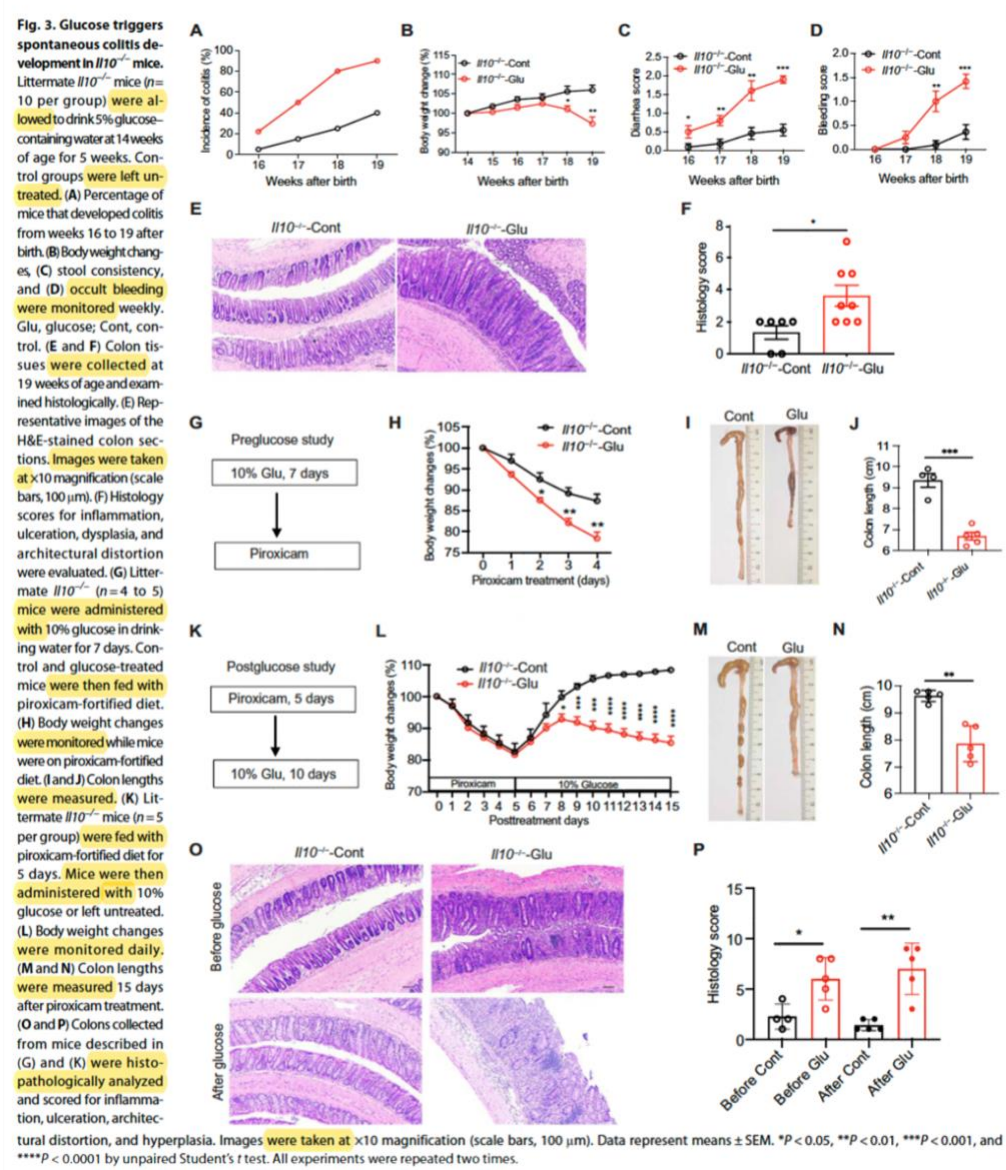


Figure 4.4. Example of Frequent Presentation of Figures and Legends in MRAs

(4.7) Consistent with previous reports, patients with psoriasis showed elevated IL-17A production in serum **(Supplemental Figure 1A; supplemental material available online with this article; https://doi.org/10.1172/JCI129269DS1),** blood CD4+ T cells (**Supplemental Figure 1B**), and skin tissues (**Supplemental Figure 1C**), and IL-17A levels were positively correlated with disease severity (**Supplemental Figure 1D**). As speculated,

> glutaminolysis in CD4+ T cells was aberrantly activated in patients with psoriasis, as indicated by elevated mRNA and protein levels of GLS1 (Figure 1, A and B, **and Supplemental Figure 1E**) and increased production of glutamate (Figure 1C). (20*JCI*5)

4.2 Situational Analysis Framework of MCRs

To my knowledge, there are not situational characteristics frameworks of MCRs that could be used to analyze this type of texts. Thus, I designed a framework based on the seven main register characteristics proposed by Biber & Conrad (2009), namely, *Participants, Relations among participants, Channel, Production Circumstances, Setting, Communication purposes,* and *Topic.* I also drew from the framework used above for MRAs. The effort was made to maintain the same, or at least similar, situational characteristics as for the MRAs whenever possible, with the rationale that doing so would facilitate the comparison between the two registers. I adapted characteristics that appeared similar but were not exactly the same in the two registers and added other elements that appeared to be specific to medical case reports.

In addition to key sources, including but not limited to expert-informants, journal websites and instructions to authors, and sample case reports, both the design of the framework and the subsequent analysis of the situational characteristics of MCRs were informed by the CARE (Case Report) guidelines proposed by Gagnier et al. (2014). The authors contend that "[c]ase reports written without guidance from reporting standards are insufficiently rigorous to guide clinical practice or to inform clinical study design" (p. 46), and to address this issue, they developed a checklist of 13 items shown in Figure 4.5. This checklist includes all information required in a case report, summarized by Gagnier et al. (2013) in these terms:

> A case report tells a story in a narrative format that includes the *presenting concerns, clinical findings, diagnoses, interventions, outcomes* (including *adverse events*), and *follow-up*. The narrative should include a *discussion* of the rationale for any conclusions and any take-away messages [emphasis added]. (p. 3)

Table 1. The CARE guidelines checklist

Item name	Item no.	Brief description
Title	1	The words "case report" (or "case study") should appear in the title along with phenomenon of greatest interest (e.g., symptom, diagnosis, test, intervention)
Keywords	2	The key elements of this case in 2–5 words
Abstract	3	a) Introduction—What does this case add? b) Case Presentation: • The main symptoms of the patient • The main clinical findings • The main diagnoses and interventions • The main outcomes c) Conclusion—What were the main "take-away" lessons from this case?
Introduction	4	Brief background summary of this case referencing the relevant medical literature
Patient information	5	a) Demographic information (e.g., age, gender, ethnicity, occupation) b) Main symptoms of the patient (his or her chief complaints) c) Medical, family, and psychosocial history—including diet, lifestyle, and genetic information whenever possible, and details about relevant comorbidities including past interventions and their outcomes
Clinical findings	6	Describe the relevant physical examination (PE) findings
Timeline	7	Depict important dates and times in this case (table or figure)
Diagnostic assessment	8	a) Diagnostic methods (e.g., PE, laboratory testing, imaging, questionnaires) b) Diagnostic challenges (e.g., financial, language/cultural) c) Diagnostic reasoning including other diagnoses considered d) Prognostic characteristics (e.g., staging) where applicable
Therapeutic intervention	9	a) Types of intervention (e.g., pharmacologic, surgical, preventive, self-care) • Administration of intervention (e.g., dosage, strength, duration) • Changes in intervention (with rationale)
Follow-up and outcomes	10	a) Summarize the clinical course of all follow-up visits including • Clinician and patient-assessed outcomes • Important follow-up test results (positive or negative) • Intervention adherence and tolerability (and how this was assessed) • Adverse and unanticipated events
Discussion	11	a) The strengths and limitations of the management of this case b) The relevant medical literature c) The rationale for conclusions (including assessments of cause and effect) d) The main "take-away" lessons of this case report
Patient perspective	12	The patient should share his or her perspective or experience whenever possible
Informed consent	13	Did the patient give informed consent? Please provide if requested

Figure 4.5. The CARE Guideline Checklist as Presented by Gagnier et al. (2014, p. 49)

Of the seven main characteristics of registers proposed by Biber & Conrad (2009), only *Relations among Participants* was not included in the new framework. Given that the case reports examined in the present study are all written, it would be difficult to speculate on the relationship between authors and readers of MCRs. However, both the “addressors” (i.e., authors) and the addressees (i.e., audience) are included in the situational analysis. The choices of elements to consider under each of the remaining six main characteristic categories (Participants, Channel, Production circumstances, Setting, Communicative purposes, and Topic) are explained in the next sections and summarized in Table 4.2.

4.2.1 Participants

For the first characteristic, participants, I kept the same elements used in the analysis of MRAs, namely, number of authors, number of actual writers, professional training/title, and primary audience. The information on authors involved in the actual writing is available at the end of MCRs, under the heading ‘Contributors’. The identification of authors’ positions (i.e., researcher or practitioner) was less straightforward as the titles PhD, MD, or MBBS are not always added next to authors’ names as was the case with MRAs. In some cases, the positions had to be inferred from the information on authors’ affiliations. Depending on whether authors were affiliated to medical institutions or universities, they were assumed to be practitioners or researchers, respectively.

4.2.2 Channel

Since all MCRs are written, it can be assumed that the organization and layout of the text will have an effect on the language of MCRs. I decided to keep the same elements as those under *Textual layout and organization* in the MRAs situational analysis framework. The *word count* and *visual elements* criteria remain unchanged. I changed *use of abstracts* to *use of abstract and summary* as case reports use the two terms interchangeably, depending on the journal.

Table 4.2. Situational Analysis Framework for Medical Case Reports

1. Participants	
1-1. Authors	1 2-4 5+
1-2. Actual Writers	1 2 3-4 5+ not disclosed
1-3. Professional Training/Title	PhD (researcher) MD/MBBS (Practitioner)
1-4. Audience	will vary
2. Channel	
2-1. Textual layout and organization	
2-1-a. Word count	Mean word count Standard deviation
2-1-b. Use of Abstract / Summary	yes / no
2-1-c. Visual Elements	None Tables Figures Tables & Figures
2-1-d. Sections/Organization	I/BCD BC(I)TODL RWDWD(P) Standardized section heading Variable section heading/name
3. Production circumstances	
3-1. Nature of data or evidence	
3-1-a. Presence of Observed Data	yes /no
3-1-b. Use of Numerical Evidence	yes / no
3-1-c. Primary Presentation of Evidence	Prose with Medical formula/shorthand Prose and Figures with extensive description
3-1-d Object of study	
3-2. Methodology	
3I-2-a. Observational	yes / no
3-2-b. Statistical Techniques	none / Descriptive Statistics
3-3. Explicitness of Research Design	
3-3-a. Explicitness of Purpose	Direct statement Minimal / No statement
3-2-b. Explicitness of Citations	Within text in-notes
3-2-c. in-text reference to cited authors	yes / no
4. Setting	
4-1 Nature of Journal	generalist / specialized
4-2. Open access peer-reviewed journal	yes / no
5. Communicative purposes	
5-1. General purpose	
5-2. Specific purposes	
6. Topic	
6-1. General topical domain	
6-2. Specific topic	

The major change pertains to the sections. There exist guidelines for writing case reports (e.g., Gagnier et al., 2014) but none for section headings and organization. As a result, sections headings and the number of sections vary from one journal to the other. I identified the following five formats of MCRs:

a. Introduction, Report of a case, and Discussion (IRD);
b. Introduction, Case presentation/Case report, and Discission (ICD);
c. Background, Case presentation, and Discussion (BCD);
d. Background, Case Presentation, (Investigation), Treatment, Outcome and Follow-up, Discussion, Learning point/take-home messages (BC(I)TODL), with the bracketed section not always present; and
e. Report of a case, What would you do next?, Diagnosis, What to do next? Discussion, (Patient Outcome) (RWDWD[P]).

The terms *introduction* and *background* are used interchangeably, and so are *report of a case*, *case report,* and *case presentation.* Therefore, I combined a, b, and c into one format I/BCD (Introduction/Background, Case presentation/report, and Discussion). Format d and e are quite distinct from a, b, and c, and are therefore listed individually.

4.2.3 Production Circumstances

For this category of characteristics, I decided to keep the same criteria used for the analysis of MRAs, albeit with some slight changes. *Presence of observed data, Use of numerical evidence,* and *Object of study* remain the same under *nature of data or evidence,* but the primary presentation of data is limited to just *Prose with medical formula/shorthand* and *Prose & figures with extensive description.* Quantitative display was removed from the list of criteria as this was noticed in none of the 100 analyzed case reports. In lieu of the two factors *Experimental* and

Observational, under *Methodology*, I only kept *Observational*, with the options *yes* and *no*, as MCRs, by definition, are very unlikely to involve experiments of any sort. Nevertheless, given that so little is known about case reports, the options *yes* and *no* under *Observational* can help identify any other methodological approach, if any. *Statistical difference testing* and *Other advanced statistics* (under *Statistical techniques*) were also removed from the framework as these factors linked to experimental studies (Gray, 2015) and therefore, are not relevant to MCRs. Finally, factors related to research questions have been excluded under *Explicitness of research design* as case reports are based on clinical observations rather than preset research questions (Gagnier et al. 2014).

4.2.4 Setting

The two factors considered under *Setting* are the nature of journals publishing case reports, also used in the analysis of MRAs, and whether the journals are open-access or not. As reported by Rison et al. (2017), 94% of case reports are published in open-access journals (p. 2). Therefore, adding this criterion may help explain potential variations (if any) between MCRs published in open access journals and those that appear in standard subscription journals, on the one hand, and between MCRs and MRAs, on the other hand.

4.2.5 Communicative Purpose

Regarding the communicative purposes of MCRs, I decided to include both the general academic purpose and the general medical purpose. Based on the descriptions of MCRs from journal websites and literature in the medical field, case reports serve the double purpose of both informing clinical practices and supporting medical progress (e.g., Gagnier et al., 2013; Rison et al., 2017).

4.2.6 Topic

Finally, for the last characteristic included in the framework, *Topic*, I kept the same factor listed on the MRA situational analysis framework, that is, the *general topic* of MCRs, as it would be difficult to list all topics treated in case reports. Additionally, the similarities or difference in general topics may help explain some commonalities or differences in linguistics features of the two registers under scrutiny. After this presentation of the situational analysis framework for medical case reports, I now turn to the description of the situational characteristics of this register.

4.3 Situational Characteristics of Medical Case Reports

In this section, I provide a description of the situational characteristics of MCRs based on the analysis of 100 medical case reports published in five peer-reviewed journals: *BMJ Case Reports* (*BMJCR), JAMA, The Journal of Medical Case Reports (JMCR),* and *Oxford Medical Case Reports (OMCR),* and *International Medical Case Report Journal (IMCRJ).* The characteristics of the analyzed MCRs are summarized in Table 4.3. While all six categories in Table 4.3 have their importance, *Channel* and *Production circumstances* seem to reflect the main differences between the two registers being compared in the present study. I will return to these two categories for an in-depth description, after an overview of the four other categories of situational characteristics of MCRs.

4.3.1 Participants

MCRs are also primarily multi-authored and the majority of MCRs authors are practitioners. As was the case with MRAs, not all authors are involved in the actual writing of MCRs. The multiplicity of authors here does not seem to be related to multiple experiments, as was the case with MRAs, but rather to the involvement of several practitioners that may intervene either during the diagnosis phase or the treatment of the patient. This can be seen in the 'Author Contributions' or 'Contributors' section provided at the end of MCRs, as illustrated by Figure 4.6 below (underlined in red). Based on information from field-expert informants, literature in the medical field (e.g., Gagnier et al., 2014; Green & Johnson, 2006; Rison et al, 2017), and communicative purposes of MCRs described in 4.3.3., the audience of MCRs includes researchers, practitioners, medical students, and patients. Given the latter member of MCR audience, it could reasonably be expected that the language of MCRs would be more accessible to the lay person. However, MCRs are still perceived as relatively technical with the use of specialized terms that only "insiders can understand" (Helan, 2012, p. 80).

Contributors All authors served as editors of the article. NS was the main author. JM, apart from serving as mentor and adviser, also contributed significantly with the writing and formatting of the manuscript. ER took care of the patient perioperatively with respect to airway and anaesthetic management. She also reviewed and guided the description of anaesthetic techniques used in the operation. TAK interpreted the histopathological specimen and slides. He also helped procure photomicrographs and provided written captions. NS and JM took direct care of the patient presented in the case report.

Figure 4.6. *Example of 'Contributors' Section in an MCR Describing the Roles of Authors in the Management of the Patient (21BMJ77)*

Table 4.3. Situational characteristics of Medical Case Reports

Characteristics		MCR
1. Participants		
1-1. Authors	1	1
	2-4	65
	5+	34
1-2. Actual Writers	1	10
	2	25
	3-4	36
	5+	18
	not disclosed	11
1-3. Professional Training/Title	PhD (researcher)	1
	MD/MBBS (Practitioner)	92
	Both	7
1-4 Primary Audience	medical practitioners, medical students, patients	

Table 4.3. (Cont'd)

Characteristics		MCR
2. Channel		
2-1Textual layout and organization		
2-1-a. Word count	Mean word count	1,518.15
	Standard deviation	651.88
2-1-b. Use of Abstract / Summary	yes	100
	no	0
2-1-c. Visual Elements	None	0
	Tables	2
	Figures	77
	Tables & Figures	21
2-1-d. Sections/Organization	I/BCD	69
	BC(I)TODL	20
	WDWD(P)	11
	Standardized section heading	95
	Variable section heading/name	5
3. Production circumstances		
3-1. Nature of data or evidence		
3-1-a. Presence of Observed Data	yes	100
	no	0
3-1-b. Use of Numerical Evidence	yes	82
	no	18
3-1-c. Primary Presentation of Evidence	Prose with medical formula/shorthand	59
	Prose & figures with extensive description	41
3-1-d. Object of Study	unusual pathology and/or novel treatment	
3-2. Methodology		
3-2-a. Observational	yes	100
	no	0
3-2-b. Statistical Techniques	none	99
	Descriptive Statistics	1
3-3. Explicitness of Research Design		
3-3-a. Explicitness of Purpose	Direct statement	100
	Minimal / No statement	0
3-3-b. Explicitness of Citations	Within text	100
	in-notes	0
3-3-c. in-text reference to cited authors	yes	26
	no	74
4. Setting		
4-1 Nature of Journal	generalist	100
	specialized	0
4-2. Open access journal	yes	60
	no	40
5. Communicative purposes		
4-1. General academic purposes	- To report unusual pathology or novel successful treatment, usually experienced by a single patient. - To generate hypotheses for future clinical studies	
4-2. General medical purpose	To inform clinical practices	
6. Topic		
General topic	rare medical pathologies and novel successful treatments	

4.3.2 Setting

All five journals considered for this situational analysis are generalists and most of the analyzed case reports were published in open access journals. This is not surprising as three out of the five journals mentioned above are open access (*JMCR, OMCR, IJMCR*). However, there was no deliberate choice of including only two subscription journals in this study; I simply was unable to find other journals publishing MCRs that were not open access. This lends support to Rison et al.'s (2017) report that 94% of MCRs are published in open access journals. This massive publication in open access journals constitutes one of the important features of a movement termed Medicine 2.0 (Eysenbach, 2008), which among other things, facilitates access to medical information for a wider audience, including patients. This confirms that patients are members of the audience of MCRs despite the latter being still perceived as technical.

4.3.3 Communicative Purposes of MCRs

Based on the literature in the medical field on case reports, the general academic purpose of MCRs is three-fold, as case reports contribute to medical research, education, and practice. Gagnier et al. (2014) suggest that the detailed reports of unusual pathologies or novel successful treatments not only contribute to the identification of new diseases, unusual forms of common diseases, and/or adverse and beneficial effects of medications, but they may also help generate hypotheses for future clinical studies. From a pedagogical perspective, the authors further contend that the detailed descriptions in MCRs "offer a structure for case-based learning in healthcare education" and given that reported cases are from different part of the world, MCRs "may facilitate the comparison of healthcare education and delivery across cultures" (p.46). The educational purpose of MCRs extends to patients who, according to Eysenbach (2008), are "experts in experiencing" the medical conditions being reported (p.3). This inclusion of patients in the audience of MCRs probably justify the recommendations by journals to include patients' perspectives in MCRs whenever possible. For example, *BMJCR* makes the following statement in the case report template they provide to authors:

> "This [the patient's perspective] is an important section and gives the patient/next of kin the opportunity to comment on their experience. This enhances the case report and is strongly encouraged" (21*BMJCR,* Case Report Standard Template)

In addition to these academic/educational purposes, MCRs also serve the purpose of informing medical practice. According to Rison et al. (2017), MCRs provide "enough details on one or a small number of patients for clinicians to relate in their own practice" (p. 1). In view of remarks reported in research in the medical field, detailed description appears to be central to MCRs and this can be noted at different sections of MCRs, whether the focus is on the patient, the disease, or clinical investigations and diagnosis. It can be assumed that such details will mitigate the frequent use of technical terms and contribute to making the discourse of MCRs more accessible to the field outsider. Excerpt (4.8) below shows an example of a detailed description of a patient and her clinical presentation (**bolded**), as well as her history of past medication (<u>underlined</u>).

> *(4.8)* **A 57-year-old woman with a history of** chronic lymphocytic leukemia **presented with a** new rash that started 6 months after <u>she completed chemotherapy with</u> rituximab, cyclophosphamide, and cytarabine. <u>She remained on maintenance therapy with</u> monthly intravenous immunoglobulin therapy. **Dry skin on her nose and forehead developed into** more distinctive skin-colored and erythematous papules that in turn coalesced into plaques. **She had a** rough texture to her skin as well as alopecia of her eyebrows and eyelashes and the frontal aspect of her scalp. **The skin-colored papules subsequently spread to** her chest, arms, and legs. <u>She was initially treated with</u> cimetidine, topical imiquimod, salicylic acid, and hydrocortisone for presumed verruca vulgaris at an outside institution, with limited benefit. <u>She was taking no other medications</u>. (12*JAMA*69)

4.3.4 Topic

MCRs are concerned with rare medical pathologies, successful novel treatments, and unusual adverse events observed during treatments. Individual topics are varied given the multiple subfields under general medicine. Nevertheless, all MCRs have in common the rarity or uniqueness of the cases they report. These aspects of novelty and/or unusualness are emphasized in all journals' guidelines to authors reviewed in this situational analysis. For example, in its 'Guidelines for Authors', *JMCR* lays emphasis on these two aspects in these terms:

> *Journal of Medical Case Reports* will consider **any original case report** that expands the field of general medical knowledge, and **original research relating to case reports**.
>
> Case reports should show one of the following:
>
> - **Unreported or unusual** side effects or adverse interactions involving medications
> - **Unexpected or unusual** presentations of a disease
> - **New** associations or variations in disease processes
> - Presentations, diagnoses and/or management of **new and emerging** diseases
> - An **unexpected** association between diseases or symptoms
> - An **unexpected** event in the course of observing or treating a patient
> - Findings that shed **new light** on the possible pathogenesis of a disease or an adverse effect [emphases added] (*JMCR*, 2021, Guidelines for Authors)

As a response to journals' requirements of novelty and originality, case report authors often make frequent use of linguistic devices that help highlight the importance and uniqueness of the cases being reported, as shown in example (4.9).

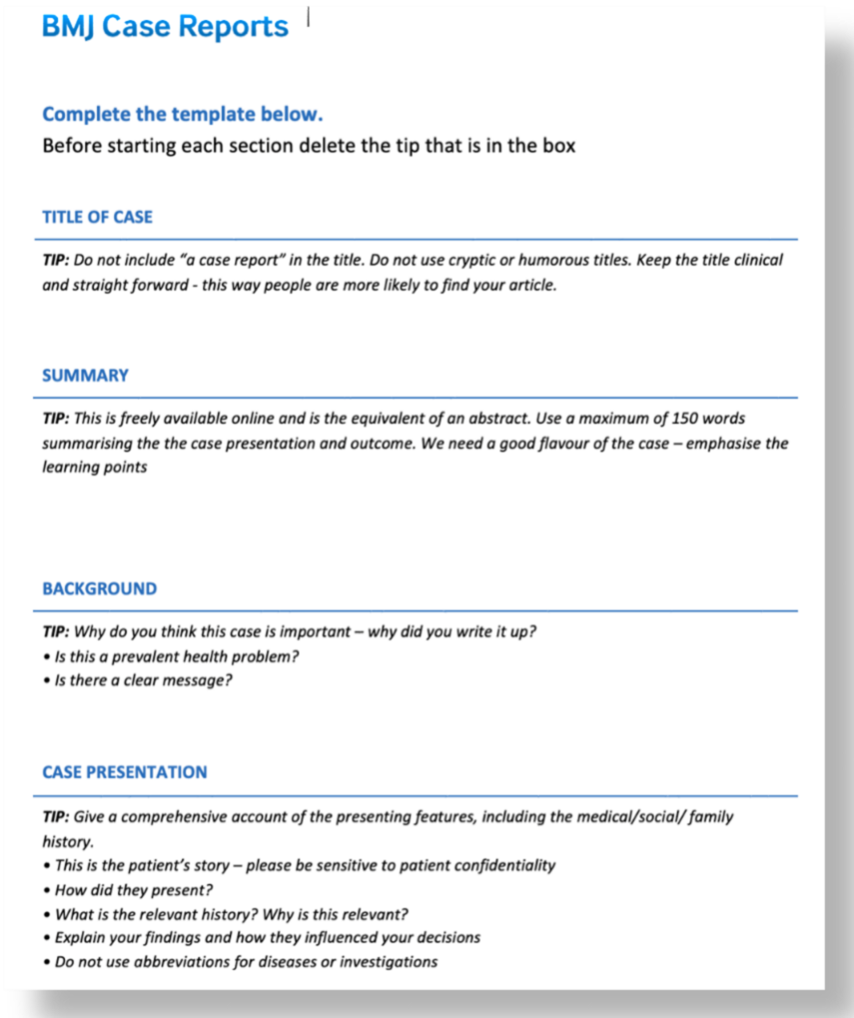
> *(4.9)* Hyperammonemic encephalopathy **is a rare and serious** adverse reaction to valproate. **Although there is** documentation of this reaction in previous reports, **very little is still known about** the exact mechanism of action. In addition, **there are no established guidelines** of the next steps needed when a patient does develop this reaction. Therefore, **this case report highlights** what is known as well as the areas of research still needed. (20*JMCR*34)

As shown in these examples, the four categories of situational characteristics described above may play a determinant role in linguistic choices made by authors of medical case reports. The next two categories (*Channel,* and *Production circumstances*) are equally, or perhaps a little more, determinant of linguistic practices in MCRs. Additionally, they constitute the main differences between MCRs and MRAs, which are discussed in section 4.4.

4.3.5 Channel

Despite the need for detailed description discussed above, the word count of MCRs is very limited (M = 1,518.15, SD = 651.88), suggesting very concise writing which, according to Gagnier et al. (2014), is "one of the appealing characteristics" of MCRs (p. 48). The authors contend that this need for conciseness requires strict textual organization to ensure smooth flow and cohesion of the information being presented. This may explain the presence of MCR templates linked to journals' guidelines to authors. These templates include a description of information to include in each section as well as writing tips. Figure 4.7. shows the first page of the template available with the *BMJCR* Instructions for Authors.

With these templates, MCRs are published in apparently very conventionalized formats. However, this conventionalization is limited to individual journals. Indeed, there seems to be a consensus on the content of MCRs but not on their textual organizations. As mentioned in 4.2.2, headings and the number of sections vary from one journal to the other. Nevertheless, MCRs remain very consistent in the type of information they include. Based on Gagnier et al.'s guidelines shown in Figure 4.5 above, case reports consist of four major sections (not including the title): the abstract or summary, the introduction or background, the case presentation, and the discussion. The type of information reported in these sections remain the same, regardless of section headings and numbers. In the sections below, I provide a description of each of the four main sections of MCRs, starting with the abstract.


BMJ Case Reports

Complete the template below.
Before starting each section delete the tip that is in the box

TITLE OF CASE

TIP: *Do not include "a case report" in the title. Do not use cryptic or humorous titles. Keep the title clinical and straight forward - this way people are more likely to find your article.*

SUMMARY

TIP: *This is freely available online and is the equivalent of an abstract. Use a maximum of 150 words summarising the the case presentation and outcome. We need a good flavour of the case – emphasise the learning points*

BACKGROUND

TIP: *Why do you think this case is important – why did you write it up?*
- *Is this a prevalent health problem?*
- *Is there a clear message?*

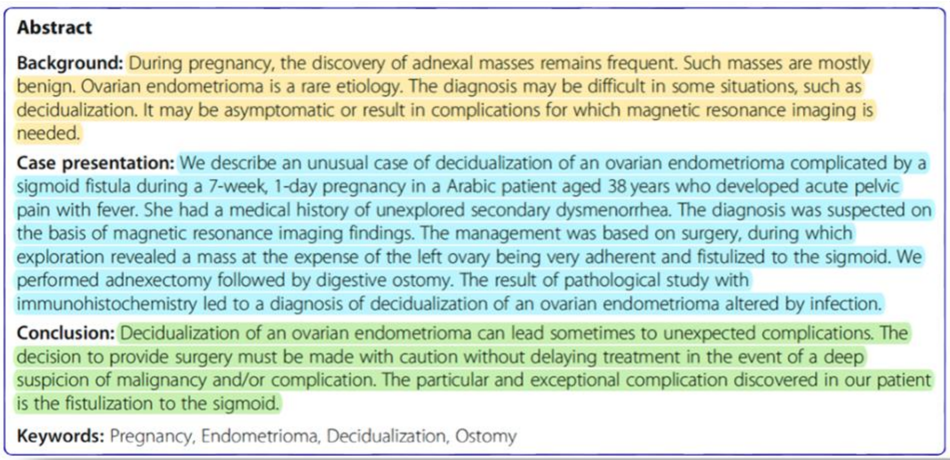
CASE PRESENTATION

TIP: *Give a comprehensive account of the presenting features, including the medical/social/ family history.*
- *This is the patient's story – please be sensitive to patient confidentiality*
- *How did they present?*
- *What is the relevant history? Why is this relevant?*
- *Explain your findings and how they influenced your decisions*
- *Do not use abbreviations for diseases or investigations*

Figure 4.7. Example of MCR Template Provided in BMJCR Guidelines for Authors

4.3.5.1 The Abstract or Summary.

The Abstract, also labelled as Summary, follows a structured format in three of the five journals consulted for this situational analysis. The two other journal, *BMJCR* and *OMCR* use one-paragraph abstracts. Regardless of the format, the abstracts or summaries include the following three points: (a) the contribution of the report, (b) the presentation of the case, including the patient's symptoms, clinical findings, diagnoses & interventions, and the main outcome, and (c) the takeaways from the case. Figures 4.8 and 4.9 show examples of the two different types of MCR abstracts. The three types of required information are color-coded with (a) in yellow, (b) in blue, and (c) green. All analyzed MCRs include an abstract and most abstracts are structured, as required by three of the five journals mentioned above. Given the type of information provided in the abstract, this section of MCRs often includes both descriptive (**bolded**) and persuasion (<u>underlined)</u> language as shown in example (4.10).



Abstract

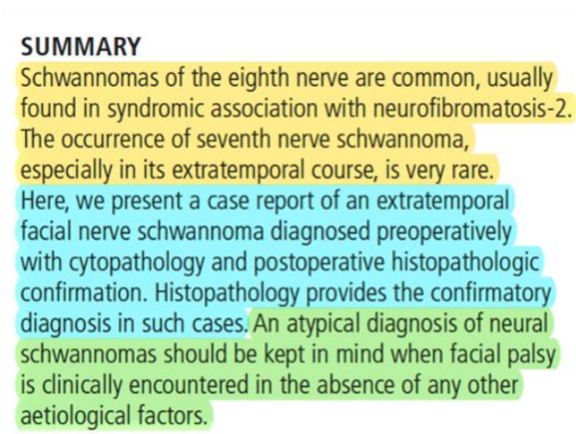
Background: During pregnancy, the discovery of adnexal masses remains frequent. Such masses are mostly benign. Ovarian endometrioma is a rare etiology. The diagnosis may be difficult in some situations, such as decidualization. It may be asymptomatic or result in complications for which magnetic resonance imaging is needed.

Case presentation: We describe an unusual case of decidualization of an ovarian endometrioma complicated by a sigmoid fistula during a 7-week, 1-day pregnancy in a Arabic patient aged 38 years who developed acute pelvic pain with fever. She had a medical history of unexplored secondary dysmenorrhea. The diagnosis was suspected on the basis of magnetic resonance imaging findings. The management was based on surgery, during which exploration revealed a mass at the expense of the left ovary being very adherent and fistulized to the sigmoid. We performed adnexectomy followed by digestive ostomy. The result of pathological study with immunohistochemistry led to a diagnosis of decidualization of an ovarian endometrioma altered by infection.

Conclusion: Decidualization of an ovarian endometrioma can lead sometimes to unexpected complications. The decision to provide surgery must be made with caution without delaying treatment in the event of a deep suspicion of malignancy and/or complication. The particular and exceptional complication discovered in our patient is the fistulization to the sigmoid.

Keywords: Pregnancy, Endometrioma, Decidualization, Ostomy

Figure 4.8. Example of MCR Structured Abstract with the three Required Types of Information (20JMCR1)



SUMMARY

Schwannomas of the eighth nerve are common, usually found in syndromic association with neurofibromatosis-2. The occurrence of seventh nerve schwannoma, especially in its extratemporal course, is very rare. Here, we present a case report of an extratemporal facial nerve schwannoma diagnosed preoperatively with cytopathology and postoperative histopathologic confirmation. Histopathology provides the confirmatory diagnosis in such cases. An atypical diagnosis of neural schwannomas should be kept in mind when facial palsy is clinically encountered in the absence of any other aetiological factors.

Figure 4.9. Example of a 1-paragraph Abstract with the Required Types of Information (21BMJ79)

(4.10) The optimal therapy for advanced thymic carcinoma **has long been controversial**. Despite that, complete (R0) resection is recommended as the first-line treatment, multidisciplinary approach including chemotherapy and radiotherapy should be considered for **patients who lost** the operation chance or received incomplete resection**. Here, we present** a **case who received** concurrent chemoradiotherapy (CCRT) after cytoreductive surgery. A complete response **was observed** and the patient **has remained** disease free for over 4 years. To our knowledge, this is the first report to demonstrate the efficacy of CCRT with cisplatin plus etoposide after incomplete surgery for advanced thymic carcinoma. (19*OMCR*15).

Additionally, given the amount of required information and the word count constraints for abstracts (150-350 words, depending on journals), MCRs authors are compelled to resort to some special techniques to condense the information they provide. The most frequently noted techniques are noun pre- and post-modifications, as illustrated in example (4.11) below, from the summary shown in Figure 4.9 above. It should be mentioned that this abstract is from an MCR published in *BMJCR* where summaries are limited to 150 words. Note the length of the bracketed noun phrase (the head noun is **bolded**) involving both pre- and post-modifications (*in italics*) that run almost the whole gamut of noun modification devices (adjectives, nouns, clause, and prepositional phrase)

(4.11) Here, we present a case report of [an *extratemporal facial nerve* **schwannoma** *diagnosed preoperatively with cytopathology and postoperative histopathologic confirmation*]. Histopathology provides the confirmatory diagnosis in such cases. (21*BMJCR*79).

4.3.5.2 The Introduction or Background.

MCR introductions, also labelled background include a brief review of relevant research and the statement of the purpose of the paper (Gagnier et al. 2014; Green & Johnson, 2006). Additionally, journal guidelines require authors to indicate the importance of their case. All this information is typically provided in three to five sentences in most analyzed MCRs. Example (4.12) shows a full introduction section with reference to the literature (underlined), indication of the importance of the case (*in italics*) and the purpose of the paper (**bolded**). Note the use of evaluative adjectives to highlight the importance and/or unusual nature of the case being reported (shaded).

> *(4.12)* Access to ultrasound at the beginning of pregnancy makes the association between pregnancy and adnexal mass an increasingly frequent situation [1]. *Endometrioma is a rare and benign etiology of adnexal mass. Clinically, the endometrioma remains difficult to recognize because it presents with no specificity*. *However, its decidualization can lead to noisy complications*. **We report a rare case of a woman with a 7-week, 1-day pregnancy with an ovarian endometrioma decidualized and fistulized to the sigmoid**.

4.3.5.3 The Case Presentation.

One of the main differences in textual organization between MCRs published in different journals pertains to this section. Based on the CARE guidelines and the description of MCRs in the medical literature (e.g., Gagnier et al., 2013; Green & Johnson, 2006; Rison et al., 2015), the case presentation is one of the key sections of a case report. Green & Johnson (2006) refer to this section as the Methods and Results section as it should provide a thorough description of the patient and the presenting condition, all clinical investigations conducted to arrive to a diagnosis, all treatments used in the management of the patient, and the outcome of the treatments. These sets of information correspond to items 5-10 (shown in Figure 4.10) of the CARE Guideline checklist. The variations noted from one journal to the other resides in whether to include all this information in one big section or to break it down into shorter subsections.

As indicated in 4.2.2, I identified three distinct formats of MCRs: I/BCD (Introduction/Background, Case presentation/Case report, and Discission), BC(I)TODL (Background, Case Presentation, (Investigation), Treatment, Outcome and Follow-up, Discussion, Learning point/take-home messages), and WDWD(P) (Report of a case, What would you do next?, Diagnosis, What to do next? Discussion, (Patient Outcome)). The majority of the analyzed MCRs are written in the I/BCD format (69%). MCRs in the I/BCB format provide all the required information for the case description in one section with slightly variable headings depending on journals. MCRs in the two other formats break down the information into multiple sections.

Patient Information	**5a**	De-identified patient specific information.
	5b	Primary concerns and symptoms of the patient.
	5c	Medical, family, and psycho-social history including relevant genetic information
	5d	Relevant past interventions with outcomes
Clinical Findings	**6**	Describe significant physical examination (PE) and important clinical findings.
Timeline	**7**	Historical and current information from this episode of care organized as a timeline
Diagnostic Assessment	**8a**	Diagnostic testing (such as PE, laboratory testing, imaging, surveys).
	8b	Diagnostic challenges (such as access to testing, financial, or cultural)
	8c	Diagnosis (including other diagnoses considered)
	8d	Prognosis (such as staging in oncology) where applicable
Therapeutic Intervention	**9a**	Types of therapeutic intervention (such as pharmacologic, surgical, preventive, self-care)
	9b	Administration of therapeutic intervention (such as dosage, strength, duration)
	9c	Changes in therapeutic intervention (with rationale)
Follow-up and Outcomes	**10a**	Clinician and patient-assessed outcomes (if available)
	10b	Important follow-up diagnostic and other test results
	10c	Intervention adherence and tolerability (How was this assessed?)
	10d	Adverse and unanticipated events

Figure 4.10. Information to Include in the Case Presentation, per the CARE Guidelines Checklist (Gagnier et al., 2014, p.49)

MCRs in the I/BCD format present the information from items 5 to 10 in the checklist following a problem-solution pattern, with a recycling trend in case of negative evaluation (Helan, 2012). In other words, Authors first describe the problem, which in the checklist corresponds to item #5 with the description of the patients, the presenting medical condition, and the relevant family and medical history. Then to report their attempts to identify the problem (the diagnosis) and solve the problem (the intervention), authors present one event (e.g., physical examinations, lab tests, CT scans, etc.) at a time, followed by its outcome until the problem is solved. The problem is solved when they reached a diagnosis, solved the patient’s medical problem, or in some cases, the patient died. The length of the Case Presentation varies from one case to the other, depending on the number of diagnostic tests conducted and subsequent therapeutic interventions. Table 4.4 illustrates this problem-solution pattern with the break-down of an entire Case Presentation section from one of the analyzed MCRs.

Table 4.4. Illustration of the Problem-Solution Pattern in the Case Presentation Section

Events	**Case Presentation section of an MCR (21*JMCR*13)**
Description of Problem (Patient, presenting condition, and past medical history)	*A 69-year-old Caucasian woman with a past medical history of low back pain, hypertension, and hyperlipidemia presented with a chief complaint of a dorsal ulnar-sided left hand-wrist mass that had been growing slowly over the previous 10 years. The patient stated that the mass had become progressively more painful over time, being particularly tender when she wore a watch. There had been no history of preceding trauma nor constitutional symptoms.*
investigation #1	*On physical exam,*
reporting result of Investigation #1	*the skin on the hands and wrist was intact with normal musculature. In the left dorsal ulnar wrist, near the extensor carpi ulnaris tendon, there was a small palpable mass that was semi-firm and not mobile; quite tender to palpation.*
Investigation #2	*A hand-wrist X-ray*
reporting result of Investigation #2	*showed marked osteoarthritic changes.*
Diagnosis #1	*An initial clinical diagnosis of a probable ganglion cyst was made,*
Intervention #1	*and the patient underwent surgical excision of the mass.*
Investigation #3	*The pathology of the soft tissue fragments*
Reporting results of Investigation #3	*revealed a biphasic neoplasm composed of spindle cells admixed with neoplastic glands (Fig. 1). No necrosis or active mitotic activity was seen. The tumor cells were positive for TLE1, focally positive for CK19, CK7, and S100, and negative for CDX2, SMA, CK20, and TTF-1 (Fig. 2).*
Challenge	*Due to tissue fragmentation, surgical margins could not be assessed; although they appeared to be involved by the neoplasm.*
Investigation #4	*FISH (fluorescence in situ hybridization) for SS18 (SYT) gene break-apart rearrangement on chromosome 18q11.2 was performed (Fig. 3),*
Reporting result of Investigation #4	*and the SYT gene rearrangement was detected in 71% of cells;*
Diagnosis #2	*thus confirming the diagnosis of synovial sarcoma.*
Investigation #5	*CT (computed tomography) of the thorax/abdomen and pelvis*
Reporting results of Investigation #5	*were without evidence of metastatic sarcoma.*
Diagnosis #3	*The tumor was classified as AJCC (American Joint Commission on Cancer) Stage IIA.*
Investigation #6	*A wide re-excision of the tumor was performed with en-bloc resection of the distal ulna.*
Reporting results of Investigation #6	*The resected tissue showed an ill-defined 1.0 × 0.5 × 0.5 cm firm mass involving the soft tissue without involving the bone.*
Investigation # 7	*The histopathologic exam*
Reporting results of Investigation #7 and **Final diagnosis**	*confirmed the prior diagnosis.*
Final Intervention	*The patient received adjuvant radiotherapy*
Outcome & Follow-up	*and had regular follow-ups for 5.5 years with no evidence of any local recurrence of the tumor or distant metastases. The timeline of the episode of care is summarized in Fig. 4.*

Linguistically, this problem-solution pattern involves a cycle of Reporting procedures → reporting of results → Reporting diagnosis/decision/intervention, until the final statement of the outcome. In the analyzed MCRs the description of procedures often involves the use of activity verbs and multiword expressions referring to medical processes and/or procedures. Often, indication of the procedure and reporting of the results are provided in the same sentence, as shown in example (4.13). This apparently contributes to the concise writing of MCRs. In such case, the reporting of the results often involves the use of communication verbs with the medical procedure or process as the subject.

(4.13) **Gross examination** revealed an enlarged uterus measuring 25 × 20 × 13 cm and weighing 3350 g, (Fig. 2), with normal bilateral fallopian tubes and ovaries. The endometrial cavity was highly enlarged, and filled with hemorrhagic villi and edematous grape-like vesicles measuring up to 1.5 cm in diameter (Fig. 3). **Microscopic examination** demonstrated a circumferential proliferation of abnormal hyperchromatic trophoblastic cells surrounding edematous hydropic villi invading the myometrium, with a few scattered trophoblastic cells within blood vessels (21*JMCR*12)

The reporting of diagnosis, decision, or intervention is typically much shorter than the reporting of results and is often done in one sentence, as shown in examples (4.14) to (4.16).

(4.14) As the patient was a postmenopausal woman with massive vaginal bleeding, **the surgical decision was** to perform total hysterectomy with bilateral salpingo-oophorectomy. (21*JMCR*12)

(4.15) Thus, the **primary differential diagnosis included** a metastatic endometrial leiomyosarcoma, a choriocarcinoma and an invasive mole.

(4.16) Then **she was started on** etoposide, cisplatin, methotrexate, actinomycin-D (EMA-EP) regimen. (20*IMCRJ*114)

As just mentioned, the majority of MCRs are written in the I/BCD format and follow the problem-solution pattern described above as in their Case Presentation sections. On the other hand, MCRs written in the BC(I)TODL and WDWD(P) format provide a more linear report of their cases, presenting each stage, from the description of the patient and his/her presenting condition to the outcome, as a separate section. The language, however, does not really differ in any of the three formats as shown in examples (4.17) and (4.18), with the description of the patient and presenting condition. Bolded texts in the same colors present the same type of information.

(4.17) **A 67-year-old Caucasian male was referred by his general practitioner to the internal medicine out-patient clinic because of hyponatremia** (127 mmol/L) found at routine laboratory examination. He had consulted his general practitioner because of abdominal pains. **His medical history revealed** colon polypectomy, an inguinal hernia, skin cancer, and reflux esophagitis. **Three years prior to this presentation his serum sodium level was 135 mmol/L.** His family history was non-contributory; he lived with his family, had a regular job, and used to engage in physical activities daily. He stopped smoking cigarettes almost 20 years ago (after 25 pack years) and did not consume alcohol or drugs. **No other symptoms or signs such as vomiting, nausea, diarrhea, altered mental status, focal neurological deficits, or palpitations were present.** (I/BCD_20*JMCR*95)

(4.18) **A 62-year-old man presented to the emergency department with 12 hours of mid-sternal chest pain** following 3 days of nausea and vomiting. **His medical history was significant for** ongoing tobacco use and newly diagnosed metastatic hepatocellular carcinoma in the context of chronic, untreated hepatitis C with associated cirrhosis (Child-Pugh B). **He met with an oncologist 3 weeks prior and received his first dose of nivolumab at the time. He was afebrile on presentation;** blood pressure was 132/95 mm Hg and heart rate was 69 beats/min. (BC(I)TODL_20*BMJ*25)

The information provided in the case presentation is typically supplemented by figures. As a result, the reader is frequently referred to these figures, and sometimes tables. As was the case with MRAs, the textual reference is done in parentheses, with minimal text, as can be seen in example (4.19) below.

(4.19) After careful discussion of risks and benefits with our patient, PPI treatment was continued and after 5 months his serum sodium level declined slightly to 131 mmol/L. During

long-term follow-up, his sodium levels were monitored regularly and stayed stable over time **(see Fig. 1)**, without any fluid restriction. (20*JMCR*95)

In this subsection, I have provided a detailed description of the Case Presentation as this section can be considered as the core of MCRs. It is the most detailed section of MCRs and the type of information it provides is what can inform clinical practices and serve for pedagogical purposes (Rison et al., 2017) or contribute to generating new research hypotheses (Gagnier et al., 2014). I now turn to the last section of the case report.

4.3.5.4 The Discussion.

The Discussion section corresponds to item # 11 in the CARE Guidelines checklist shown in Figure 4-10. It includes a discussion of previous literature related to the case presented, any conclusions drawn from the case, challenges encountered, and the main takeaways. All this information is presented in this same section in MCRs in the I/BCD and WDWD(P) format, whereas case reports in the BC(I)TODL format provide the takeaways in a separate section entitled 'Learning Points/Take-home Messages'. In the analyzed MCRs, relating previous literature to the case presented involves frequent use of passive constructions with the medical pathology (example 4.20), the successful intervention (4.21), or other elements related to the case (4.22) as the grammatical subject. The next section (4.3.6) describes the last category of MCR situational characteristics.

*(4.20) Macroscopically, **SCTATs** <u>have been reported</u> to be solid, cystic, mixture of solid and cystic, yellow to tan, and some with haemorrhage and necrosis with size ranging up to 30 cm [4, 7].* (20*OMCR*82)

*(4.21) **Sirolimus** <u>has been shown</u> to be effective in eosinophilic fasciitis, a disease belonging to the spectrum of localized scleroderma.*[3] (16*JAMA*30)

(4.22) ***Caffeine*** *has been reported to act as an activator of the sympathetic activator in central nervous system* [10]. (21*JMCR*11)

4.3.6 Production Circumstances.

As shown in Table 4.3 above, MCRs primarily report observed data. The evidence provided is primarily numerical and is presented in both prose with medical formula/shorthand and with prose supplemented with figures. As already mentioned, MCR authors often refer their readers to figures and sometimes tables elsewhere in the text. Regarding the methodology, all analyzed MCRs used an observational approach. As explained in 4.2., the decision was made to include the options *yes* and *no* in case some MCRs used a different approach, but no exception was found in the 100 analyzed MCRs. Consistent with the observational nature of MCRs, statistical techniques appear to be an exception in this register. Only one case report was found to include some basic descriptive statistics, and those statistics were used while reporting findings of previous studies. Regarding the explicitness of the research design, All MCRs include a direct statement of the purpose both in the abstract/summary and in the introduction/background. The purpose is typically announced by the phrase *(here) we report*, as in example (4.23) below. The phrase was used in 72 of the 100 analyzed MCRs.

(4.23) ***Here, we report*** *the case of a 75-year-old Japanese man with giant paratesticular liposarcoma.* (20*JMCR*98)

Regarding the citation practices, MCRs also follow the AMA style. However, this does not appear to prevent mention of cited authors in the text. Several MCRs were found to include both the names of cited authors and the superscript numbers at the end of the citations, as shown in example (4.24).

(4.24) It has been reported that majority of cases with SCTATs present with hyperestrinism, amenorrhea and postmenopausal bleeding [8]. In the literature, a few

> cases of malignant SCTATs have been reported. For instance, **Lele et al.** reported a 47-year-old female with malignant SCTAT which was bilateral [8]. **Dart et al.** reported one patient with metastasis from a series of three cases with SCTAT [9]. Recurrence and metastases tend to occur several months to years after removal of the primary tumour [8, 9]. Malignant SCTAT seems to spread mainly via the lymphatics and commonly involve the pelvic, para-aortic and supraclavicular lymph nodes [2, 8]. A recurrence rate of almost 50% has been reported in the series of **Qian** in which some of the patients had even repeated recurrences; the first recurrence was seen after 45.5 months [8].

In this section, I have attempted to provide an extensive description of the situational characteristics of medical case report. To my knowledge, no other study has conducted such analysis. Therefore, this extensive description was necessary both for the collection of a representative corpus and for the discussion of the various types of formulaic sequences instigated in the present study. I now turn to the comparison of the situational characteristics of MRAs and MCRs.

4.4 Similarities and Differences between MRAs and MCRs

The most salient differences between the two registers pertain to the situational characteristics described above in 4.3.5 and 4.3.6, that is, *Channel* and *Production Circumstances.* But before discussing these main differences, let me start with what the two registers have in common. The first and most obvious shared characteristic is that both registers belong to the same field. As such, it could be reasonably expected that they would at least have some common topics, and indeed, they do. Both MRAs and MCRs are concerned with medical pathologies and treatments and the end goal of both registers is to advance research and clinical practice in the medical field. The similarity, however, may be limited to these overarching goals and topics, as the two registers have different approaches to investigating the topics they share.

MRAs have as primary purpose to report generalizable results of research on medical pathologies and treatments, whereas MCRs are concerned with descriptions of single or few cases that are not generalizable but can directly inform clinical practice and education in the medical field. As a result, in line with previous descriptions of scientific research in the literature (e.g., Hyland, 2008a), the focus in MRAs is primarily on the research process itself to ensure generalizability of the results and replicability of the research. MRAs provide extensive descriptions of research methods and experiments, describing procedures in every step.

On the other hand, the focus in MCRs is on the patient, the correct diagnosis of the medical pathology, and the patient's response to the provided treatment. This translates, as shown in 4.3.5.3, into mere mentions of medical procedures and other investigations and a more extensive report of results that ultimately lead to a final diagnosis and/or the appropriate intervention. Compare the excerpts in Figure 4.11 and example (4.25). Figure 4.11 shows the description of one single element of the methodology in the Methods section of an MRA, while example (4.25) shows three different investigation procedures (**bolded**) being just mentioned, followed by the descriptions of their results (underlined). Given this difference in focus, it is perhaps no surprise that the majority of MCRs are written by practitioners while MRAs are typically written by researchers.

> *(4.25)* **Physical examination** revealed a palpable pelvic mass extending up to approximately 3 cm above the umbilicus. The serum beta-human chorionic gonadotropin (b-HCG) level was determined to be 542.250 mU/mL. **Pelvic ultrasonography** demonstrated an enlarged uterus the size of 24-week gestation, with a heterogeneous mass obliterating the endometrial cavity, with a vesicular appearance (Fig. 1) and normal ovaries. **Computed tomography (CT) scan of the abdominopelvic** region confirmed the presence of a well-demarcated mass measuring 25 × 20 × 13 cm with very high-density central cystic content. (MCR_20*JMCR*95)

In vitro neutralizing assays

ABAB-neutralizing activity was determined by a cell-based neutralization assay as previously described (*23*). To assess the ability of ABAB from *Sb*-ABAB culture supernatants to neutralize toxins from a broad range of clinical isolates, we used two collections of *C. difficile* strains. Most of the first collection (table S1, 14 strains) was provided by T. Lawley, except for UK1 that was provided by D. Gerding; this represents an assortment of genetically and geographically diverse clinical isolates from Europe (*84*, *85*). The second collection (table S2, 50 strains), from the Centers for Disease Control and Prevention, was obtained through Biodefense and Emerging Infections (BEI) Resources, National Institute of Allergy and Infectious Diseases, National Institutes of Health; isolates were selected to represent the diversity of strain types and geographical locations circulating in the United States during 2010–2011 (https://www.cdc.gov/hai/eip/clostridium-difficile.html). The *C. difficile* strains were cultured under anaerobic conditions for 2 days, and supernatants were collected and diluted 100 times in cell culture medium before being applied to Vero cell monolayers in a 96-well plate. The supernatants caused cell rounding within 4 hours or after overnight incubation. In some wells, 10 μl of *Sb*-ABAB supernatant (final dilution 10×) was added to the monolayers before applying *C. difficile* culture supernatants, and cell rounding was monitored after overnight culture. Neutralization was defined as protection of 100% of the cells from *C. difficile* toxin–induced cell rounding. To determine the neutralization activity of ABAB secreted in animal intestines, the intestinal lavages from ilea, ceca, and colons were diluted with PBS (10×) and filtered before mixing with TcdB (final concentration was 10 pg/ml) and applying to Vero cell monolayers in a 96-well plate. After 24 hours of incubation, cell rounding was observed by phase-contrast microscopy.

Figure 4.11. Example of Extensive Description of Research Procedures in MRAs (20Science6)

The two different approaches and foci described above, together with the textual organization and the methodology under *Channel* and *Production circumstances*, respectively, depict MRAs and MCRs as two distinct registers. The stark difference noted in the textual layout and organization of IMRD-MRAs and MCRs are indication of potential differences in writing practices. It is very unlikely for published medical research article to include the problem-solution pattern described in the Case Presentation, the core section of MCRs. Example (4.25) and Figure 4.11 can serve as illustrations of some linguistic differences that can result from the differences in both section types and contents. Reporting results of clinical investigations and describing experimental procedures certainly require different linguistic devices.

As regards the methodologies of MRAs and MCRs, the two registers differ fundamentally in two points that are indeed related. The first difference is that MRAs typically follow an experimental design while MCRs are mostly observational. As a result of this difference in research design, statistical analyses were found to be central to all analyzed MRAs with most articles featuring a *Statistical Analyses* sub-section in the *Methods* section for a description of all statistical techniques used. This description of statistical analyses involves the use of statistics jargon as in example (4.26) below.

(4.26) ***Statistical significance*** *was determined using* ***a Mann-Whitney test*** *between two groups, and* ***one-way analysis of variance (ANOVA)*** *was used to compare multiple groups. Results with* ***P < 0.05 (95% confidence interval****) were considered* ***statistically significant.*** *Subject-level data are reported in data file S1.* (20*Science*2)

On the other hand, statistical techniques are almost inexistent in MCRs. As mentioned above, only one of the 100 MCRs included some basic descriptive statistics, and that was in a report of previous studies. However, the fact that the cited case report used descriptive statistics is indication that MCRs may sometimes include basic statistics. As a result of this difference in methodology, the two registers differ in their ways of reporting results. The reporting of results in MCRs as shown in examples (4.13) and (4.25) above, frequently involves the use of medical procedures or tests as agents of communication verbs such as *reveal, show, indicate, confirm.* This practice is also found in MRAs, but additionally, the Results sections of MRAs also include reports of the statistical significance of the results being presented, as shown in example (4.27).

(4.27) *By day 6, the colon of DSS/sugar-treated mice were* ***significantly shorter (P < 0.01)*** *and exhibited extensive inflammation, crypt loss, and ulceration (*Fig. 1, L to N*, and fig. S3, A to C).* (20*Science*1)

One final observation worth mentioning is the difference in audience between the two registers. With medical students and patients as members of their audience, it can be expected that authors will strive to make MCR contents more accessible to the lay person. Indeed, such effort is encouraged by journals. The *BMJ* (2021), for example, in its ‘Tips for writing’ states: “Write as you speak. Keep it short and informal”. Therefore, despite the use of technical terms that probably cannot be avoided as they refer to specific medical processes and/or procedures, authors’ efforts to “keep it short and simple” can be noted in certain sections of MCRs. For example, MCR authors often resort to successive short and straightforward clauses/sentences when describing the patient (example 4.28) or reporting results (example 4.29)

(4.28) A 27-years-old female presented to Motahari University Hospital infertility clinic with primary infertility for 4 years in 2017. She had been referred to the infertility clinic from 1 year after her marriage. The patient had a history of dysmenorrhea and dyspareunia without any unusual bowel disease. Family history for any disease including tuberculosis was negative. (20*OMCRJ*75)

(4.29) Biologically: Mantoux test was positive, hemoglobin was at 12 g/dl, white-blood-cell count was at 15100/mm3; C reactive protein was at 115 mg/l; QuanTIFERON test was positive; and her serum ionized calcium, serum phosphate levels and serum protein electrophorese were normal. (20*IMCRJ*76)

In sum, I have described in the section the most salient similarities and differences between MCRs and MRAs. While the two registers share common topics and the end goal of advancing medical research and practice, they differ in their primary foci and audience. MRAs primarily seek to inform research while MCRs primarily focus on informing clinical practice and medical education. According to one of my informants, the audience may not be too different as every illness starts with at least one case and this case may result in a research study. But it is true that cases do have a wider audience that includes students (although students also read RAs), and patients. This difference can lead to variations in writing practices, some of which I have described above.

4.5 Conclusion

The aim of this chapter was to answer RQ1 that asked about the situational characteristics of medical research articles and medical case reports. The findings of the situational analysis of MRAs are for the most part consistent with previous descriptions of scientific research articles (e.g., Gray, 2015; Hyland, 2008a). However, some characteristics seem to be specific to the field of medicine, namely, citations in AMA style and the use of figures with extensive descriptions of the legends. Regarding the situational characteristics of MCRs, I have first presented a framework for the situational analysis of medical case reports. Then in 4.3., I have provided an extensive description of the situational characteristics of MCRs and the implications such characteristics may have in the writing of MCRs. Finally, I have discussed the similarities and differences between the two registers and the variations in writing practices and linguistic choices that can result from such differences. Both registers share overarching goals and topics, but differ fundamentally in their foci, research approaches, and to some extent, audiences. The findings reported in this section will inform the discussion of potential variations between the two registers in the use of formulaic sequences investigated in the present study.

5 CHAPTER 5. FINDINGS AND DISCUSSIONS 1: COLLOCATIONS AND MULTIWORD COLLOCATIONS IN MRAS AND MCRS

This chapter reports the findings of the use of 2-word collocations and multiword collocations of 3 to 6 words in the MRAC and the MCRC to draw comparisons of structures and functions across the two registers. The goal of the analysis of these two types of formulaic sequences is primarily to identify and describe specialized expressions that have been reported to be one of the main challenges in medical writing. The comparison of the structures and functions of collocations and multiword collocations aims at identifying the commonalities and dissimilarities between the two registers in the use of the identified sequences.

In section 5.1., the findings of each corpus are first presented and discussed individually in subsections 5.1.1. and 5.1.2., respectively, to report which collocations are used in MRAs and MCRs. Subsection 5.1.3 presents the identification and analysis of collocations shared by MRAs and MCRs. Section 5.2. and its subsections introduce the findings and discussion of the structural and functional analyses of multiword collocations in the MRAC and the MCRC, respectively; and shows the analysis of multiword collocations common to the two corpora. These sections and subsections are meant to answer the following research questions:

<u>Research Question 2a:</u> What collocations are used in MRAs and MCRs? Are they register-specific or rather shared within the same discipline?

<u>Research Question 2b</u>: What multiword collocations are used in the two registers? How do they compare in terms of structures and functions? Are they register-specific or rather shared within the same discipline?

5.1 Collocations

As explained in the Methodology Chapter, Section 3.3.1, the decision was made to use *delta P* to measure mutual expectancy and better understand the directional associations between the identified bigrams. As a result, three categories of collocations were identified from the initial list of candidates. These categories are referred to hereafter as *Bidirectionals*, *Unidirectionals 1* and *Unidirectionals 2. Bidirectionals,* as explained in 3.3.1., refer to bigrams with relatively strong mutual attraction; *Unidirectionals 1* are pairs in which $word_1$ was found to be the strongest predictor; and *Unidirectionals 2* refer to bigrams with $word_2$ as the strongest predictor. These categories are shown in Table 5.1.

Table 5.2 shows the raw number of collocations identified in the two corpora. Even though the aim was never to compare the number of sequences across the two registers, given the moderately different sizes of the two corpora, both the MRAC and the MCRC yielded a relatively similar number of collocations with a total of 334 collocations identified from the MRAC and 303 from the MCRC. Similarly, the same trend was observed regarding the number of collocations per category; with more collocations in the *Unidirectionals 1* category than in any of the two other categories in both corpora (see Table 5.2). The collocation lists in each category and each corpus were analyzed structurally and functionally, and the findings are presented and discussed in the following sections.

5.1.1 Collocations in Medical Research Articles

Collocations identified in the MRAC were analyzed both structurally and functionally. For the structural analysis, I used the framework presented in 3.2.2. and analyzed the functions of collocations based on Hyland's (2008a) framework.

Table 5.1. Collocation categories based on Predictors

Categories	Examples	Freq	Range	$\Delta P_{2\|1}$	$\Delta P_{1\|2}$	Log10
***Bidirectionals* (-0.5 ≤ Log_{10} ≤ +0.5)**	*mechanical ventilation*	45	8	0,62	0,69	-0,05
	lymph nodes	60	12	0,57	0,63	-0,04
	myocardial infarction	199	40	0,88	0,97	-0,04
	patch clamp	14	5	0,55	0,57	-0,02
	ice cold	16	9	0,35	0,36	-0,02
	colony-forming units	8	7	0,33	0,34	-0,01
	et al	236	74	0,96	0,98	-0,01
	vice versa	19	15	1,00	1,00	0,00
	bona fide	7	7	1,00	1,00	0,00
	propidium iodide	7	7	0,87	0,87	0,00
	substance abuse	7	6	0,50	0,50	0,00
	emergency department	30	12	0,51	0,50	0,01
	unstable angina	15	7	0,58	0,56	0,02
	poorly understood	25	18	0,56	0,50	0,05
	Alexa Fluor	38	22	0,95	0,84	0,05
***Unidirectionals 1* (0.51 ≤ Log_{10} ≤ 4)**	*lethally irradiated*	6	6	1.00	0.21	0.67
	folic acid	81	6	1.00	0.18	0.75
	immunosorbent assay	18	18	1.00	0.04	1.36
	transforming growth	12	9	1.00	0.04	1.43
	acetic acid	13	9	1.00	0.03	1.54
	accounted for	90	41	0.99	0.01	2.19
	accompanied by	40	30	0.99	0.01	2.29
	compensate for	11	9	0.99	0.00	3.11
	coincide with	6	6	0.99	0.00	3.45
	confined to	19	11	0.98	0.00	3.08
	belonging to	16	6	0.98	0.00	3.15
	refers to	13	8	0.98	0.00	3.24
	contributors to	12	8	0.98	0.00	3.28
	amenable to	6	6	0.98	0.00	3.58
***Unidirectionals 2* (-4 ≤ Log_{10} ≤ -0.51)**	*linked immunosorbent*	18	18	0.08	1.00	-1.10
	life expectancy	56	9	0.12	1.00	-0.94
	cerebral palsy	16	5	0.14	1.00	-0.85
	orally bioavailable	8	7	0.15	1.00	-0.83
	hemoglobin A1c	16	7	0.23	1.00	-0.65
	carbon dioxide	8	7	0.27	1.00	-0.57
	we reasoned	16	9	0.00	1.00	-2.59
	figure legends	13	13	0.00	1.00	-2.36
	per kilogram	10	7	0.01	1.00	-2.22
	per deciliter	52	12	0.03	1.00	-1.50
	a priori	27	28	0.00	0.99	-2.80
	per milliliter	46	16	0.03	0.98	-1.55
	in situ	35	20	0.00	0.97	-2.99
	in utero	38	7	0.00	0.97	-2.96
	in vitro	331	78	0.01	0.97	-2.02

Table 5.2. Raw Numbers of Collocation Types per Category in the MRAC and the MCRC

Categories	MRAC	MCRC
Bidirectionals	110	97
Unidirectionals 1	146	137
Unidirectionals 2	78	69
Totals	**334**	**303**

5.1.1.1 Structures of collocations in the MRAC.

The structural analysis of the collocations in the MRAC revealed some salient patterns in each of the three categories. Table 5.3 shows the distribution of structures across categories. The *Bidirectionals* category consisted mostly of lexical collocations that accounted for 93.67% of all identified collocations in this category. These lexical collocations are mostly noun combinations of the types *N + N* and *Adj + N* (81.09 %) and their compositions ranged from purely technical terms, defined by Nation (2001, p. 198) as terms that are "recognizably specific to a particular topic, field or discipline" (e.g., *propidium iodide, amino acids, hydrogen peroxide, dimethyl sulfoxide*), to a combination of technical and semi-technical words (e.g., *unstable angina, phylogenetic tree, lymph node, adverse events*), to pairs of semi-technical words only (e.g., *sample size, primary outcome, end point, statistical significance, Western blot*).

Such distinction may have relevant pedagogical implications as semi-technical terms, defined by Lam (2001) as "words that have one or more 'general' English meanings and which in technical contexts take on extended meanings" (p. 1), have been reported to be problematic for language learners in both the understanding and production of specialized texts (e.g., Hyland & Tse, 2007; Lam, 2001; Rezeaian, 2015).

Table 5.3. Structures of Collocations in the MRAC

	Structures	Collocations			
		Bidirectionals	*Unidirectionals1*	*Unidirectionals 2*	Examples from the MRAC
1. Lexical Collocations	Noun Combinations *(n + n / adj + n)*	**83.72%**	19.21%	34.46%	*viral load, risk factors, cytoplasmic inclusions, western blot*
	Verb combinations (*v + n / v + adj)*	0%	0%	1.28%	*is conceivable,*
	Verb-Adv combinations (*adv + v / v + adv / adv + vpp)*	4.52%	2.73%	0%	*normally distributed, randomly assigned, latently infected, virally suppressed*
	Adj-Adv combinations *(adv + adj / adv + adv)*	5.43%	2.05%	2.5%	*commercially available, statistically significant, critically ill*
Subtotals 1		**93.67%**	23.99%	38.24%	
2. Grammatical Collocations	N & function word	0.9%	20.54%	14.28%	*90th percentile, insight into, predominance of, reliance on, contraindications to, predictor of, and colleagues, for example.*
	V & function word	1.80%	**44.52%**	**37.90%**	*determine whether, confined to, belonging to, consisting of, we reasoned, we undertook*
	Adj & function word	0%	10.27%	2.59%	*irrespective of, amenable to, reminiscent of, capable of, unresponsive to, compatible with, analogous to*
Subtotals 2		2.7%	**75.33%**	**54.77%**	
3. "Borrowed" Collocations		3.63%	0.68%	6.59%	*de novo, vice versa, bona fide, in vivo, ex vivo, a priori, in situ, et al.*
Totals		100%	100%	100%	

To illustrate this point, let us look at examples (5.1) – (5.3) below. It obviously will require more than our layman understanding of *Western* and *blot* to understand the methodology referred to in each passage. In fact, based on the explanations of one expert informant, *Western blot* is a laboratory test that detects a specific protein of interest in a mixture of other proteins. What is worth noting here is that in all three excerpts below, as well as in the 60 concordance lines with *Western blot* analyzed, the expression is used without any further explanation, suggesting that it is shared knowledge in the medical community, and that probably, this is one of the “accepted terms” that Rezaeian (2015) described as crucial in medical writing. This assumption was confirmed by the field-expert informant. Thus, novice and L2-English medical writers, would certainly need to be familiarized with these types of “accepted terms”.

(5.1) ***Western blots*** *of subcellular fractions of Pggt1bfl/flLC macrophages revealed RAC1 in the membrane and cytosolic fractions, as in Pggt1bfl/+LC macrophages, which had trace amounts of nuclear RAC1.* (10*JCI*5)

(5.2) *Confirmation of successful SMPDL-3b overexpression in SMPDL-3b–transfected podocytes was obtained by* ***Western blot****.* (11*Science*4)

(5.3) *Cytokine gene expression was measured by real-time polymerase chain reaction (PCR) and oligonucleotide microarrays, cytokine release was assessed by Luminex technology, and protein phosphorylation was assessed by* ***Western blot****.* (20*Science*10)

Noun combinations were also relatively frequent in the *Unidirectionals 1* and *2* categories, even though, as shown in Table 5.3, they were not the most frequent types of collocations in these two categories. They accounted for 19.21 % and 34.46% of collocations in the *Unidirectionals 1* and *Unidirectionals 2* lists, respectively. These findings are somehow consistent with previous findings by Ackermann & Chen (2013) who reported noun combinations as being the most frequent lexical collocations in academic writing. Figure 5.1. shows the predominance of noun combinations among lexical collocations identified in the MRAC.

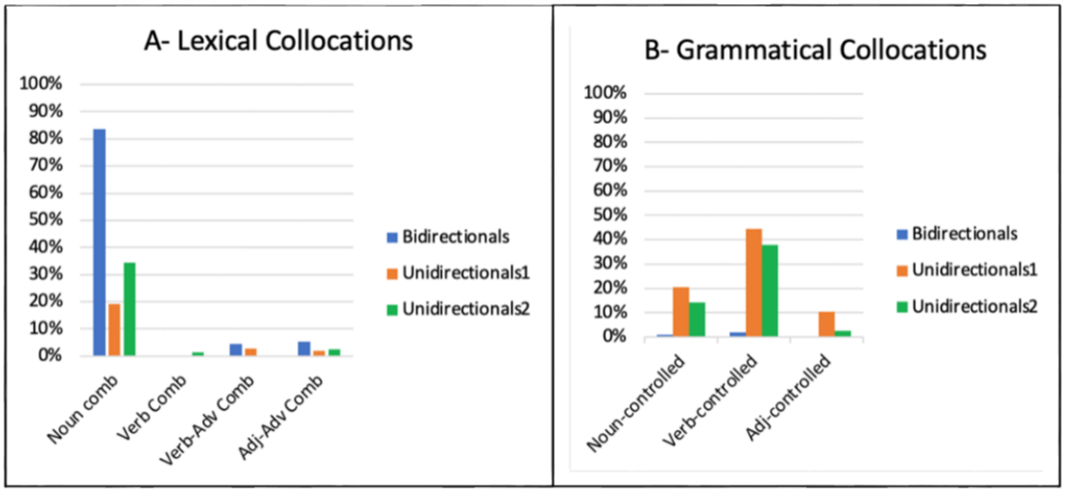


Figure 5.1. Visual Representation of Collocations Structures per Category in the MRAC

However, the use of *delta P*, the decision to include function words as possible collocates, and the subsequent division of collocations into three categories revealed that grammatical collocations were also quite frequent in the two *Unidirectionals* categories. Indeed, grammatical collocations accounted for 47% of all identified collocations in the MRAC and were the most frequent types of collocations among the *Unidirectionals 1* and *2*, where they accounted for 75.33 % of the *Unidirectionals 1* and 54.77% of the *Unidirectionals 2*. In both categories, the most frequent collocations were verb-controlled combinations (42.52% for *Unidirectionals 1* and 37.90% for *Unidirectionals 2*.), followed by noun-controlled combinations (20.54% for *Unidirectionals 1* and 14.18% for *Unidirectionals 2*). Adjective-controlled combinations were rare in *Unidirectionals 2*, but relatively frequent in *Unidirectionals 1* where they accounted for 10.27% of collocations in this category.

The examples in Table 5.3 above show that function words involved in grammatical collocations were primarily prepositions. This may be of relevance to medical writing instruction. Ackermann & Chen (2013) contend that grammatical collocations are less problematic for language learners as they are comparatively fixed and therefore, more predictable. However, research has shown that English prepositions tend to be particularly challenging for non-native of English writers (e.g., Ahn, 2013; Back, 2011; Lee et al., 2020). The pedagogical implications of the present study are discussed in Chapter 8, but for now, we can say that each of the three categories revealed distinctive collocational patterns that may need to be considered in the discussion of pedagogical implications.

As mentioned above, the *Bidirectionals* category consisted primarily of Noun Combinations. The two other categories, on the other hand, were dominated by verb-controlled collocations (see Figure 5.1), albeit in different patterns. In the *Unidirectionals 1,* verb-controlled collocations were primarily multiword verbs, that is, *Verb + Preposition/Particle (*e.g., *accounted for, accompanied by, compensate for, interacts with, counterstained with, confined to*), whereas those identified in the *Unidirectionals2* list consisted almost exclusively of infinitives. While the high frequency of infinitives may be an indication of frequent use of non-finite to-infinitive complementation in medical research articles, from a phraseological perspective, prepositional and phrasal verbs are probably of much more interest. Thus, the verb-controlled grammatical collocations in the *Unidirectionals 1 list*, together with the Noun combinations identified in all three categories, were qualitatively analyzed in context to see whether they served specific functions in the medical research articles.

5.1.1.2 Functions of collocations in the MRAC.

The functional analysis of the selected collocations was carried out using the *Concordance* and *File View* tools in *AntConc* (Anthony, 2020). The two most frequent types of collocations in the MRAC, that is, noun combinations and verb-based grammatical collocations, appeared to be mostly research-oriented expressions (see, Hyland, 2008a). Analysis in context of the 143 noun combinations identified across all three categories revealed that these collocations are used in medical research articles to primarily refer to field-specific concepts, whether it is to refer to a medical condition, the object(s) of the study and/or elements related to it (examples 5.4 & 5.5), to specify study participants (example 5.6), or refer to methodological processes (example 5.7).

(5.4) *In* ***rheumatoid arthritis****, for which remission after discontinuing medication is now the accepted goal of management, the duration of treatment during remission while continuing medication is a matter of debate. (*10*JAMA*5*)*

(5.5) *The most common adverse events of grade 3 or higher were hypertension (in 14% of the patients), an increased* ***alanine aminotransferase*** *level (in 12%), an increased* ***aspartate aminotransferase*** *level (in 10%), hyponatremia (in 6%), and lymphopenia (in 6%).* (20*NEJM*8)

(5.6) *To compare the risk of skin cancer between* ***transplant recipients*** *and* ***background population****, we used a stratified proportional hazard regression model for hazard ratio (HR) estimations.* (15*JAMA*1)

(5.7) *To parameterize and quantify this difference, tumors were dissociated, and the clonal composition was analyzed by* ***flow cytometry*** *[…].* (20*JCI*10)

Similarly, the majority of verb-based grammatical collocations in the form of multiword verbs (about 97% of all analyzed verb-based grammatical collocations) consisted of research-oriented expressions. They were primarily used to refer to methodological and experimental procedures (examples 5.8 – 5.10).

*(5.8) In an effort to **compensate for** the reduced treatment frequency, the drug dose was increased to 100 mg/kg.* (20*JCI*6)

*(5.9) Cells were fixed and **stained with** antibody specific for human collagen 1 and **counterstained with** fluorescent secondary antibody (Alex Fluor 488).* (20*Science*7)

*(5.10) After nucleofection, the cells were **resuspended in** 0.5 mL of complete RPMI medium and **transferred to** a 6-well plate **filled with** 1.5 mL of RPMI medium **supplemented with** 1 gM SCR7, an inhibitor of DNA ligase IV and nonhomologous end joining (MilliporeSigma).* (20*Science*9)

In addition to noun combinations and verb-controlled collocations, there was also, as mentioned above, a relatively substantial portion of noun-controlled and adjective-controlled grammatical collocations in the MRAC (20.54% and 10.27%, respectively). Here again, the collocations were predominantly research-oriented with functions similar to those exemplified above as can be seen in examples (5.11) and (5.12) below. Example (5.11) shows a noun-controlled collocation used to refer to an experimental procedure and (5.12) shows an adjective-controlled collocation used to describe an element of the study.

*(5.11) Fibroblasts were stimulated with TGFβ1 (10 ng/ml) and incubated with either vehicle [0.1% dimethyl sulfoxide (DMSO)] or 100 nM omipalisib, FAPL-PI3Ki1, or PI3Ki1 for 2 hours, followed by **removal of** media.* (20*Science*7)

*(5.12) Among these brain regions, the VAL and VM may be **analogous to** the human ventral intermediate nucleus of the thalamus, which has been targeted for DBS treatment.* [26-28] (20*JCI*2)

In sum, collocations identified in the MRAC, whether lexical or grammatical, were used primarily to refer to research objects and procedures. While this could be expected, as scientific writing has been described to primarily focus on empirical demonstrations and experimental results (Hyland, 2008a), these findings highlight the importance and pedagogical value of these collocations. Novice L1 and English-L2 medical writers seeking to report their research and share knowledge with their community will certainly benefit from the teaching of expressions like the ones discussed in this subsection.

5.1.2 Collocations in the Medical Case Reports

Collocations in the MCRC were structurally analyzed based on the same framework used for MRAC collocations. For the functional analysis, I used the framework explained in 3.3.2 and shown in Table 3.5.

5.1.2.1 Structures of collocations in the MCRC.

The structural analysis of collocations in medical case reports revealed the same trend as that observed in the MRAC. The distribution of structures in the three categories of collocation are shown in Table 5.4 and visualized in Figure 5.2. As was the case with collocations in medical research articles, here also, lexical collocations almost exclusively dominated the *Bidirectionals* category (95.98%), while grammatical collocations were predominant in both *Unidirectionals 1* (74.45%) and *Unidirectionals 2* (65.23%).

Overall, noun combinations (*N + N* and *Adj + N*) were the most common structure, thus illustrating their widespread use in academic writing. They accounted for 91.76 % of all *Bidirectionals,* 21.16 % of *Unidirectionals 1,* and 27.54% of *Unidirectionals 2.* Just like noun combinations in the MRAC, they involved purely technical terms (e.g., *carcinoembryonic antigen, creatine kinase, alkaline phosphatase, herpes simplex*), combination of technical and semi-technical words (e.g., *alternative diagnoses, platelet count, nasogastric tube, multiple myeloma*), and pairs of semi-technical terms (e.g., *abdominal pain, transplant recipients, vital* *signs, weight loss*). Other structures for lexical collocations were very rare in the MCRC as can be seen in Figure 5.2 with the highest occurrence being only 3.65% for verb combinations in the *Unidirectionals 1* category.

Regarding the grammatical collocations in the MCRC, the same trend as in the MRAC was again observed with a predominance of verb-controlled collocations. They accounted for 45.98% of grammatical collocations in the *Unidirectionals 1* and 39.13% in the *Unidirectionals 2.* As was the case in the MRAC, the verb-controlled collocations consisted primarily of multiword verbs in the *Unidirectionals 1* and infinitives in the *Unidirectionals 2.* They were followed by noun-controlled collocations, (28.49%), with 19.79% occurring in the *Unidirectionals 1* category in the form of *N + Prep* (e.g., *array of, coexistence of, contributor to, insight into*). This makes these collocations potential candidates for L2 writing instruction. Adjective-controlled collocations came in third position, accounting for 26.89% of grammatical collocations.

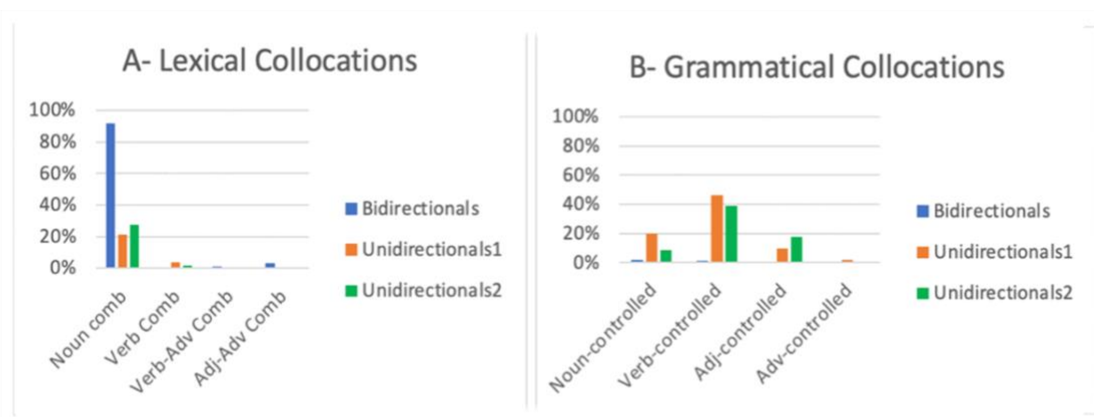


Figure 5.2. *Visual representation of Collocation Structures per Category in the MCRC*

Table 5.4 Structures of Collocations in the MCRC

		MCRC			
		Bidirectionals	*Unidirectionals 1*	*Unidirectionals 2*	Examples from the MCRC
1. Lexical Collocations	Noun Combinations (n + n / adj + n)	**91.76%**	21.16%	27.53%	*bone marrow, night sweats pleural effusions, mycophenolate mofetil, lactate dehydrogenase*
	Verb combinations (v + n / v + adj)	0%	3.65%	1.45%	*counting fingers, computerized tomography*
	Verb-Adv combinations (adv + v / v + adv / adv + vpp)	1.03%	0%	0%	*poorly differentiated*
	Adj-Adv combinations (adv + adj / adv + adv)	3.09	0.73%	0%	*critically ill, exceedingly rare*
Subttotals-1		**95.98%**	25.54%	28.98%	
2. Grammatical Collocations	N & function word	2.06%	19.79%	8.7%	*per minute, our patient completion of, lack of, for example, restoration of, on exertion, persistence of*
	V & function word	1.03%	**45.98%**	**39.13%**	*ruled out, multiply by, accounting for, consist of tends to, we believe*
	Adj & function word	0%	9.49%	17.4%	*due to, attributable to, the commonest, consistent with, compatible with, the exact.*
	Adv & function word	0%	2.1%	0%	*rather than, along with*
Subtotals 2		3.09%	**74.46%**	**65.23%**	
3. "Borrowed" Collocations (Subtotals 3)		1.03%	0%	5.79%	*en bloc, in vitro, in situ, de novo, et al.*
Grand totals		100%	100%	100%	

One salient observation was that the largest number of adjective-controlled collocations occurred in the *Unidirectionals 2* category (17.4% as opposed to 9.49% in *Unidirectionals1*), and therefore, unlike their counterparts in the MRAC, the most common function word accompanying the adjective was not a preposition, but the definite article *the* (e.g., *the rarest*, *the commonest, the highest, the entire, the exact, the largest, the best.*). Many of these pairs are superlatives. One explanation of the frequent use of superlatives may be the need for MCR authors to put emphasis on the uniqueness or unusual nature of their cases, as suggested by findings of the functional analysis discussed below.

5.1.2.2 Functions of collocations in the MCRC.

The most frequently used structures in the MCRC, that is, noun combinations and verb-based grammatical collocations, as well as the above-mentioned superlatives were analyzed in context using concordance lines generated in *Antconc* (Anthony, 2020). As mentioned above, the framework shown in Table 3.5. was used for the functional classification of MCRC collocations. Table 5.5 shows the functions served by the most frequent types of collocations in the MCRC.

Non combinations were primarily case-related, or diagnosis/intervention related; with 94.7% of them used to describe the case or refer to biological processes/elements and procedures during diagnosis or intervention. In the Description category, noun combinations were used either to introduce the medical condition or to provide details related to it, as shown in examples (5.13) and (5.14).

(5.13) A 60-year-old woman presented with ***abdominal pain*** *and jaundice for 9 months and fever for 2 months.* (21*BMJ*101)

(5.14) Jaundice was progressive, non-fluctuant, associated with severe itching and also clay-coloured stool, but not associated with Gasto-Intestinal (GI) bleed, ***abdominal distension****, encephalopathy,* ***alternative medication*** *intake, prior surgery or lump abdomen.* (21*BMJ*101)

Table 5.5. Functions of Most Frequent Types of Collocations in the MCRC

Categories	**Sub-categories**	**Examples from the MCRC**
Case-related	Subject specification	*transplant recipient*
	Description	*iliac fossa, sleep apnea, heart failure*
	Other Individuals	*general practitioner, family members, multidisciplinary team*
Diagnosis/ Intervention-related	Biological processes/elements	*flow cytometry, herpes simplex, excisional biopsy, right eye, ejection fraction, acid-fast bacilli*
	Decisions	*ruled out, referred to, elected to*
	Procedures	*thrombolytic therapy, treated with, mycophenolate mofetil*
Outcome/Follow-up – related		*adverse events, transitioned to*
Discourse organizers	Resultative signals	*owing to, accounts for, leading to, resulted in, resulting from,*
	Framing signals	*depending on, according to, based on*
	Elaboration/Clarifications	*consisting of, characterized by*
Stance Features		*the best, the highest, the rarest*

Expressions used to refer to biological processes/elements and procedures during diagnosis or intervention, were most of the time very technical and used with no further description, again, suggesting their membership to that category of "accepted terms" in the medical field. Examples (5.15) and (5.16) show some of such expressions used in the MCRC.

(5.15) A diagnosis of paradoxical cerebral embolism associated with a spontaneous venous thromboembolism and a patent foramen ovale was made. The patient benefited from ***thrombolytic therapy*** *and lifelong anticoagulation with good recovery. (20OMCR59)*

(5.16) His electrocardiogram showed sinus tachycardia with a normal PR interval. ***C-reactive protein*** *level and white cell counts were raised at 248 mg/L and 14.9 × 10^9 cells/L, respectively, and blood cultures isolated methicillin-sensitive* ***Staphylococcus aureus*** *(20OMCR33)*

Verb-controlled grammatical collocations served a wider range of functions in the MCRC. When they were diagnosis/intervention-related, they were used to indicate decisions made during diagnosis or intervention as can be seen in example (5.17) below.

(5.17) After ***ruling out*** *a pneumothorax using ultrasound, the patient was immediately commenced on CPAP set to positive end-expiratory pressure of 10 mm Hg delivering 100% oxygen.* (20BMJ4)

As discourse organizers, they served as resultative signals (example 5.18), framing signals (examples 5.19 & 5.20), or elaboration/clarification devices (examples 5.21& 5.22).

(5.18) The commonest cause of anaemia is iron deficiency, either due to nutritional deficiency or blood loss ***leading to*** *a state of absolute iron deficiency characterised by low iron stores..* (20*Lancet*1)

(5.19) There are two subsets of ALCL: Primary cutaneous ALCL which is ***confined to*** *the skin, and Systemic ALCL that affects all organs mostly the lymph nodes, with involvement of extra nodal sites including bone marrow, skin, soft tissues, lung and liver.* (19*OMCR*154)

(5.20) ***Based on*** *a large epidemiological study, the annual incidence rate of stroke among Chinese adults aged 45 to 75 years with hypertension was approximately 1.0%.* (15*JAMA*2).

(5.21) He is currently on salvage chemotherapy with a different regimen ***consisting of*** *rituximab, ifosfamide, carboplatin and etoposide (R-ICE) with close monitoring.* (20*OMCR*20)

(5.22) Two vague granuloma formations ***composed of*** *epithelioid cells aggregate,* ***surrounded by*** *a rim of lymphocytes were noted.*

Finally, the analysis of adjective-controlled collocations in the form of superlatives revealed that these expressions were either diagnosis/intervention-related, or stance features. When they were diagnosis/intervention-related, they were used to highlight the most important aspects of diagnosis findings that are likely to determine subsequent interventions, as can be seen in example (5.23) below.

(5.23) Pelvic magnetic resonance imaging (MRI) was performed to complete the exploration […] with the following findings: two cystic hemorrhagic formations of the left ovary in T1 and T2 hyperintensity of 87 mm × 78 mm […]. **The largest** cyst was in intimate contact with the sigmoid with a possible rupture of its wall at this level. […]. Thus, an exploratory laparotomy was decided on the basis of radiological suspicion of sigmoid fistulization and the worsening of the patient's clinical symptomatology. (20*JMCR*1)

As stance features, these superlatives served to justify interventions made to address the medical condition presented (example 5.24), and as mentioned earlier, to highlight the uniqueness of the case being presented (examples 5.25 & 5.26). Note how in example (5.26), the superlative is used to highlight the age group commonly affected by the medical condition, only to underscore in the subsequent sentence the stark contrast with the very unusual age of the patient in the case being presented. This need for MCR authors to put emphasis on the uniqueness of their cases seems to be motivated by journals' requirements of novelty and uniqueness, discussed in 4.3.4.

(5.24) *Our hospital set-up is neither equipped with vascular imaging machines nor staffed with specialized surgeons to do sophisticated vascular repairs, so amputation was* ***the safest*** *option for our patient*. (20*OMCR*46)

(5.25) *Our case was one of* ***the rarest*** *types, and probably the first of its kind owing to bilaterality of the lesion, which itself is rare, along with two varieties existing (type 1 on the right side and type 2 on the left side).* (20*IMCRJ*93)

(5.26) *Oligoarticular JIA is chronic arthritis begins before age 16 years with* ***the highest*** *frequency in girls aged 1–3 years […] Here, we presented a rare case of JIA in a premature baby with clinical manifestations during the neonatal period (21 days of age). (20OMJCR15)*

From a pedagogical perspective, these findings may be beneficial for medical writing instruction, in general, and writing of case reports, in particular. Indeed, one important aspect of case reports is their focus on "*unusual*, interesting, or *unique* [emphasis added] medical aspects" encountered in medical practice (Helan, 2012, p. 57). That aspect is even highlighted in journals' submission guidelines. The renowned journal *JAMA,* for example, describes cases reports in its submission guidelines as "[s]hort reports of original studies or evaluations or *unique*, *first-time* [emphases added] reports of clinical case series". In the same vein, *The Journal of Medical Case Reports,* encourages authors to include in their submission cover letter a description of "how the case report is *rare* or *unusual* as well as its *educational and/or scientific merits* [emphases added]". Therefore, it is probably safe to assume that instruction of adjective-controlled collocations, and possibly other types of formulaic sequences serving similar functions, would greatly benefit novice L1 and English L2 medical writers in quest of publication of their case reports.

In sum, collocations in medical case reports followed similar trends as those in medical research articles in terms of their structures. Noun-combinations were the predominant structure of lexical collocations whereas grammatical collocations consisted primarily of verb-controlled collocations. Noun-controlled and adjective-controlled collocations were also relatively frequent in the MCRC. While overall, all analyzed grammatical collocations involved prepositions, adjective-controlled collocations seemed to have a preference for the definite article *the* as the function word accompanying adjectives. In terms of functions, noun-controlled collocations occurred primarily in two categories: case-related and diagnosis/intervention-related. They were used in the detailed description of the case being presented and of the procedures and biological processes during diagnosis and/or interventions. Verb-controlled grammatical collocations were used as both discourse organizers and to refer to decisions made during diagnosis and/or intervention. Finally, adjective-controlled collocations in the form of superlatives were found to serve as stance features that case report authors use to lay emphasis on the importance and uniqueness of both their cases and their approaches to addressing them.

5.1.3 Comparison of Collocations in MRAs and MCRs

This section answers the second part of RQ 2a that asked whether the collocations identified in the two corpora were register-specific or shared within the same discipline. Shared collocations between the MRAC and the MCRC are listed in Table 5.6. As could be expected, the predominant structures of the shared collocations were noun combinations and verb-controlled pairs for lexical collocations, and grammatical collocations, respectively, as they were the most frequent structures in both corpora. They were followed by adjective-controlled and noun-controlled pairs, all in the Grammatical Collocations category.

The two registers share a rather limited number collocations, particularly in structural categories that were found to be predominant in the two corpora. As shown in Table 5.7, the 30 shared noun combinations, by far the most frequent structure in both corpora, represent only 21% and 21.89% of noun combinations identified in the MRAC and the MCRC, respectively. These noun combinations, as discussed above, serve distinct roles in the two registers. They were mostly research-oriented in the MRAC, and case-related and diagnosis/intervention-related in the MCRC. This distinction was observable even among shared collocations, as can be seen in the pair of examples below with the expression *flow cytometry.* In example (5.27), *flow cytometry* is used in the description of the research methodology. On the other hand, in the case report (example 5.28), the same expression is used to report diagnosis/intervention results.

(5.27) Four days after transduction, T cells were stained with CBFB-MYH11/B·40:01 pHLA tetramer and anti-CD8 mAb. CBFB-MYH11/B·40:01 tetramerpositive CD8+ T cells were sorted to greater than 95% purity, expanded, then evaluated by ***flow cytometry*** *and functional assays.* (MRAC_20*JCI*1)

(5.28) Cervical lymph node biopsy showed necrotizing histiocytic lymphadenitis. ***Flow cytometry*** *revealed B and T cells of nonclonal phenotype* (MCRC_18*JAMA*195)

Beyond the difference in functions, it was interesting to note the difference in what Hoey (2005) referred to as a word's *colligations*, that is, the grammatical environment it tends to occur in or avoid. Indeed, for most of the shared collocations, their use differed from one corpus to the other. This difference can be noticed in the two examples above. In the MRAC, *flow cytometry* often occurs in passive sentences either in a by-phrase, as in (5.27), or in a non-finite ing-clause introduced by *using*, as shown in example (5.29). In the MCRC, it mostly occurred in active sentences where it served as subject of a monotransitive verb (e.g., *showed, revealed, etc.*).

Table 5.6 Percentages of Shared Collocations between the MRAC and the MCRC

Lexical Collocations		**Grammatical Collocations**			**Borrowed**
Noun combinations	Adj-Adv Combinations	Noun-controlled	Verb-controlled	Adjective-controlled	
risk factors	*critically ill*	*paucity of*	*multiply by,*	*the same*	*et al*
chest radiograph	*twice daily*	*insight into*	*accounting for*	*the latter*	*in vitro*
bone marrow		*amounts of*	*attributed to*	*consistent with*	*in situ*
flow cytometry		*reversal of*	*accounts for*	*compatible with*	*de novo*
blood pressure		*parts of*	*tend to*	*regardless of*	
heart failure		*aspect of*	*depends on*	*capable of*	
atrial fibrillation		*for example*	*composed of*	*responsible for*	
white matter			*consisted of*	*susceptible to*	
lymph node			*ranging from*	*irrespective of*	
weight loss			*carried out*		
lymph nodes			*depend on*		
myocardial infarction			*contribute to*		
mechanical ventilation			*accompanied by*		
transplant recipients			*corresponds to*		
ethics committee			*contributes to*		
aspartate aminotransferase			*confined to*		
			regarded as		
rheumatoid arthritis			*consisting of*		
alkaline phosphatase			*led to*		
electron microscopy			*ranged from*		
folic acid			*correlated with*		
nervous system			*arising from*		
escherichia coli			*to evaluate*		
ejection fraction			*to determine*		
antiretroviral therapy			*to treat*		
adipose tissue			*to address*		
vast majority			*to assess*		
nucleic acid			*to achieve*		
hemoglobin A1c			*to detect*		
carbon dioxide					
computed tomography					

(5.29) *mRNA-mediated 3xFLAG-FOXP3 protein expression was confirmed using* ***flow cytometry*** *and immunoblotting (Figure 2E).* (13*JCI*3)

Table 5.7. Percentages of Shared Collocations by Structure in each Corpus

		Shared Raw number	**MRAC** Raw number (percent shared)	**MCRC** Raw number (percent shared)
Lexical Collocations	Noun Combinations	30	143 (21%)	137 (21.89%)
	Adj-Adv Comb	2	11 (18.18%)	4 (50%)
Grammatical Collocations	Noun-controlled	7	40 (17.5%)	33 (21.21%)
	Verb-controlled	31	97 (31.95%)	87 (35.63%)
	Adjective-controlled	10	19 (52.63 %)	27 (37.03%)
Borrowed		4	10 (40%)	5 (80%)

The expression *blood pressure* was found to be a perfect illustration of such variation. It was remarkably frequent in both corpora, occurring 336 times in the MCRC and 559 times in the MRAC. In the MCRs, 206 (61.9%) out of the 336 occurrences served to report examination findings. That function was primarily realized through copular patterns, mainly with *copular be* (43. 24%). In other cases, the measurement figures were just apposed to the expression with or without the preposition *of* (25% and 27.4%, respectively). The concordance lines in Figures 5.3 below show examples of these patterns used to report examination results in the MCRs. On the other hand, *blood pressure* was used in MRAs in various patterns including as modifier of other nouns; and more frequently, in agentless passive sentences to describe research procedures/methodology (Figure 5.4). In fact, the copular pattern shown in Figure 5.3 occurred only five times out of 559 in the entire MRAC.

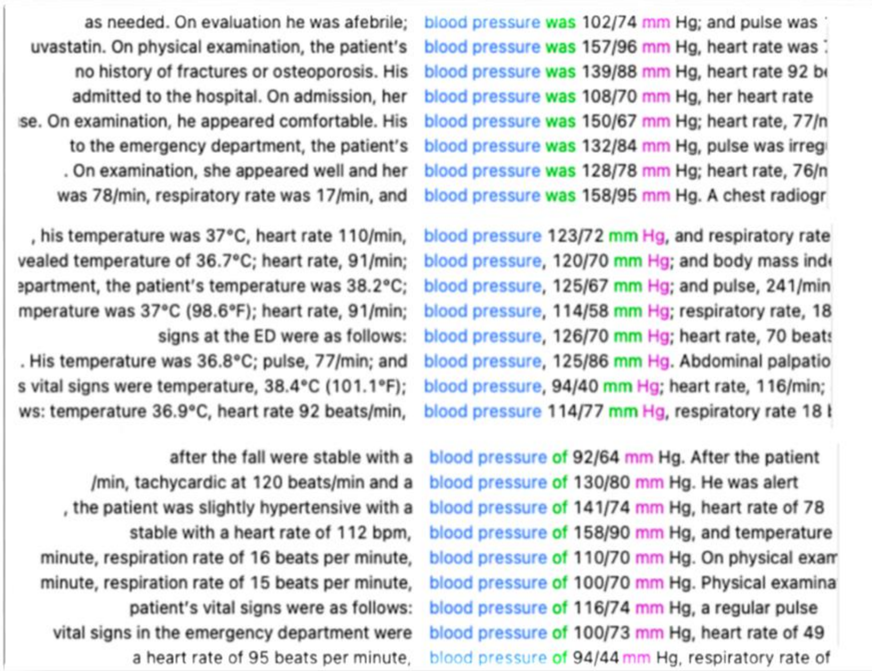

as needed. On evaluation he was afebrile; blood pressure was 102/74 mm Hg; and pulse was
uvastatin. On physical examination, the patient's blood pressure was 157/96 mm Hg, heart rate was
no history of fractures or osteoporosis. His blood pressure was 139/88 mm Hg, heart rate 92 b
admitted to the hospital. On admission, her blood pressure was 108/70 mm Hg, her heart rate
se. On examination, he appeared comfortable. His blood pressure was 150/67 mm Hg; heart rate, 77/n
to the emergency department, the patient's blood pressure was 132/84 mm Hg, pulse was irreg
. On examination, she appeared well and her blood pressure was 128/78 mm Hg; heart rate, 76/n
was 78/min, respiratory rate was 17/min, and blood pressure was 158/95 mm Hg. A chest radiogr
, his temperature was 37°C, heart rate 110/min, blood pressure 123/72 mm Hg, and respiratory rate
vealed temperature of 36.7°C; heart rate, 91/min; blood pressure, 120/70 mm Hg; and body mass ind
epartment, the patient's temperature was 38.2°C; blood pressure, 125/67 mm Hg; and pulse, 241/min
mperature was 37°C (98.6°F); heart rate, 91/min; blood pressure, 114/58 mm Hg; respiratory rate, 18
signs at the ED were as follows: blood pressure, 126/70 mm Hg; heart rate, 70 beat
. His temperature was 36.8°C; pulse, 77/min; and blood pressure, 125/86 mm Hg. Abdominal palpatio
s vital signs were temperature, 38.4°C (101.1°F); blood pressure, 94/40 mm Hg; heart rate, 116/min;
ws: temperature 36.9°C, heart rate 92 beats/min, blood pressure 114/77 mm Hg, respiratory rate 18
after the fall were stable with a blood pressure of 92/64 mm Hg. After the patient
/min, tachycardic at 120 beats/min and a blood pressure of 130/80 mm Hg. He was alert
, the patient was slightly hypertensive with a blood pressure of 141/74 mm Hg, heart rate of 78
stable with a heart rate of 112 bpm, blood pressure of 158/90 mm Hg, and temperature
minute, respiration rate of 16 beats per minute, blood pressure of 110/70 mm Hg. On physical exam
minute, respiration rate of 15 beats per minute, blood pressure of 100/70 mm Hg. Physical examina
patient's vital signs were as follows: blood pressure of 116/74 mm Hg, a regular pulse
vital signs in the emergency department were blood pressure of 100/73 mm Hg, heart rate of 49
a heart rate of 95 beats per minute, blood pressure of 94/44 mm Hg, respiratory rate of

Figure 5.3. Preferred Patterns of 'blood pressure' in the MCRC


mixed B6/129 mice. Blood pressure measurement. Blood pressure was measured in awake, chronically
non-fasting blood samples were taken. Blood pressure was measured with the use
by the American Society of Echocardiography. Blood pressure was measured in conscious mice
placebo. A greater increase in diastolic blood pressure was observed when evacetrapib, 10
discontinuation of therapy. No increase in blood pressure was observed in evacetrapib-treate
rements for 1 week before beginning experiments. Blood pressure was recorded at days 0 and 14
with telemetry units. Ten days later, blood pressure was recorded for 10 minutes every
spectral analysis of heart rate variability. Blood pressure was recorded by telemetry for 5
1994. A 20 mm Hg lower average systolic blood pressure was reported in this additional 8% (
155]). A 15 mm Hg lower average systolic blood pressure was reported between 1994 and 20
electrical spinal cord stimulation on aortic blood pressure was studied by inserting stimulation
are summarized in Table 5, and achieved blood pressure levels are shown in eFigure 2.
-tight blood-pressure control. The lower blood-pressure levels attained early in the
during routine antenatal care, regardless of blood-pressure levels before pregnancy), admissio
attempted to ensure the comparability of blood pressure levels between the treatment group
a nurse visit and 4753 had valid blood pressure readings (785 had eaten, drunk alc
delivery, and hypertension (i.e., any blood-pressure readings higher than 140/90 mm H
comparison with the following years. Three blood pressure readings were taken from each
, or smoked in the 30 min before blood pressure recording; three valid measurement
was analyzed from multiple artifact-free blood pressure records of each mouse by
norepinephrine and blood pressure were reduced. Blood pressure reductions were correlated with the
between genotypes that may have affected blood pressure regulation (data not shown). Neithe
pineal gland does not interfere with blood pressure regulation. Initially, Pdc-deficient m
the retina or pineal gland for blood pressure regulation, several conditions were

Figure 5.4. Examples of Preferred Patterns of 'blood pressure' in the MRAC

Regarding shared verb-controlled grammatical collocations, the second most frequent structure in both corpora, they showed much less variations than noun combinations; probably because they are less specific and served for the most part as discourse organizers. Nevertheless, it must be noted that over 60% of verb-controlled collocations identified in each corpus appear to be specific to register represented by the corpus, given that the shared collocations in this structural category account for only 31.95% of the MRAC list and 35.63% of the MCRC list.

To answer RQ 2a, we have seen that as has been shown by previous research of academic writing, collocations in both medical research articles and medical case reports were predominantly noun combinations. Verb-controlled grammatical collocations came as the second most frequent structure with both corpora featuring pedagogically relevant verb-preposition combination. However, the functional analysis revealed much less similarities, especially in the most frequent structural category (noun combinations) where collocations were found to serve very distinct functions in the two corpora. Additionally, the two registers shared a very limited number of collocations and even in such case, most of the shared collocations were used differently in the two registers, suggesting that the collocations identified in this study were register-specific, for the most part.

Previous research has shown variations across disciplines in the use of the same formulaic sequences (e.g., Hyland & Tse, 2007; 2009). One of the aims of this dissertation, however, is to extend the analysis beyond discipline, investigating intra-disciplinary variation across registers. The small number of shared collocations between medical research articles and medical case reports, as well as the differences found in the use of those shared collocations are indications that, as suggested by Gray (2015), discipline only tells part of the story when it comes to linguistic variations in academic writing.

I now turn to the next type of formulaic sequences: multiword collocations. It was expected that these sequences, like collocations, would also include specialized terms that could be of interest to medical writing instruction.

5.2 Multiword Collocations

This section answers RQ 2b that asked about the use of multiword collocations in MRAs and MCRs, as well as potential variations between the two registers. With the exploratory approach used in the identification of multiword collocations, sequences of up to seven words were identified in each corpus. As explained in 3.4.1, *delta P* was used to measure the collocational strength of 3-word collocations, and *MI* used for 4-word, 5-word, 6-word, and 7-word collocations. To mitigate the tendency of MI to favor low frequency terms, I used the frequency thresholds of 10 times pmw for 3-word and 4-word sequences, eight times pmw for 5-word sequences, six times pmw for 6-word and 7-word sequences, and five times pmw for longer multiword collocations. Table 5.8 shows the raw numbers of multiword collocations identified in each corpus. Overall, even though the corpora slightly differed in size, the number of types of multiword collocations was relatively similar in the two corpora with 149 in the MRAC and 155 in the MCRC.

Table 5.8. Raw Numbers of Multiword Collocations in the MRAC and the MCRC

Collocation Length	**MRAC**	**MCRC**
3-words	99	84
4-words	30	53
5-words	9	6
6-words	7	11
7-words	4	1
Totals	**149**	**155**

The identified multiword collocations were analyzed both in terms of structures and functions. For the structural analysis, I used the classification framework introduced in 3.4.2. and shown in Table 3.6. That framework comprises three main categories*: Phrasal Sequences*, *Clausal Sequences*, and *Coordinated Binominals*. The Phrasal Sequences category includes the following three subcategories:

- Complex Noun Phrases: that is, nouns that are premodified (e.g., *ambient particulate matter* ***pollution***, *the saw palmetto* ***extract***, *stem cell* ***transplantation***), postmodified (e.g., ***death*** *from cardiovascular cause*, ***activities*** *of daily living*, ***induction*** *of mixed chimerism*), or both (e.g., *heart* ***failure*** *with reduced ejection fraction*, *common terminology* ***criteria*** *for adverse events, the consolidated* ***standards*** *of reporting trials*). In all these examples, the **head nouns are bolded,** and the modifiers are underlined;
- Prepositional Phrases in the form of *Preposition + (Complex) NP* (e.g., ***with*** *no significant past medical history*, ***of*** *unknown origin*, ***in*** *close proximity*, ***in*** *nonhuman primates*). **Prepositions are bolded** and complex noun phrases are underlined; and
- Other Phrases subcategory for the very rare instances of other types of phrases (e.g., *frozen in liquid, available on request*).

The Clausal Sequences category consists of two subcategories:

- Declarative Clauses/Fragments that consist of independent clauses expressing complete statements (e.g., *function tests were normal, medical history was significant*) or clause fragments expressing incomplete statements (e.g., *physical examination revealed a, this case report highlights*), and

- Dependent Clauses/Fragments subcategory that includes dependent clauses and clause fragments, whether finite or nonfinite (e.g., *what we believe, to better understand, to be clinically meaningful*).

The third structural category is for Coordinated Binominals, defined by Biber et al. (1999, p. 1030) as "two words from the same grammatical category, coordinated by *and* or *or"*. This subcategory includes expressions such as *directly or indirectly, hematoxylin and eosin, alert and oriented*, etc., but also sequences where the coordinated elements are phrases instead of single words (e.g., *chest pain and shortness of breath*). I now turn to the findings of the structural and functional analyses of multiword collocations in the two corpora, starting with the MRAC in the section below.

5.2.1 Multiword Collocations in Medical Research Articles

5.2.1.1 Structures of Multiword Collocations in the MRAC

The structural analysis revealed that multiword collocations in the MRAC were primarily phrasal, with complex noun phrases accounting for 67.67% of 3-word collocations, 90% of 4-words, 88% of 5-words, and 100% of both 6-words and 7-words. Table 5.9 shows the distribution of structures of multiword collocations in the MRAC. As shown by Table 5.9., the longer the sequences, the less varied their structures are, with a steady increase in the proportion of complex noun phrases. Only 3-word collocations include instances of each structure. Coordinated binominals are the next most frequent structure in MRAC, accounting for 11.12% of 5-words and 10.10% of 3-words. They are closely followed by clausal sequences, with declarative clauses/fragments accounting for 10% of 4-words and 7.07% of 3-words. These findings are consistent with the claim by Biber & Gray (2016) that modern science research writing has dramatically shifted from the use of dependent clauses in favor of "a strong increase in the use of phrasal modifiers" (p. 27). The high frequency of phrasal sequences in the MRAC multiword collocations, then, can be considered as an indication of how grammatical complexity is achieved and how information is condensed and presented in medical research articles. It can also be the case that medical writers resort to this grammatical process to achieve concise writing given the limited word counts allowed by medical journals.

Table 5.9 Distribution of Categories of Multiword Collocations in the MRAC

		MRAC Multiword Collocations				
Structures		3-words	4-words	5-words	6-words	7-words
Phrasal Sequences	Complex NP	**67.67%**	**90%**	**88.88%**	**100%**	**100%**
	PP	9.10%	0%	0%	0%	0%
	other	2.02%	0%	0%	0%	0%
subtotals1		***78.78%***	***90%***	***88%***	***100%***	***100%***
Clausal sequences	Declarative Clauses/fragments	7.07%	10%	0%	0%	0%
	Dependent Clauses/Fragments	4.04%	0%	0%	0%	0%
subtotals 2		*11.11%*	*10%*	0%	0%	0%
Coordinated Binominals		10.10%	0%	11.12%	0%	0%
Totals		*100%*	*100%*	*100%*	*100%*	*100%*

Given their very high frequency, the complex noun phrases were further analyzed to better understand the noun modification processes used by authors of medical research articles. The findings are summarized in Figure 5.5 below. Noun premodification appears to be, by far, the most frequent process utilized by medical article writers. Noun premodification in the MRAC multiword collocations is achieved through multiple devices ranging from nouns, to adjectives, to *-ed* and *-ing* participials, with a predominance of nouns as modifiers, as can be seen in the list of all MRAC multiword collocations (Appendix C).

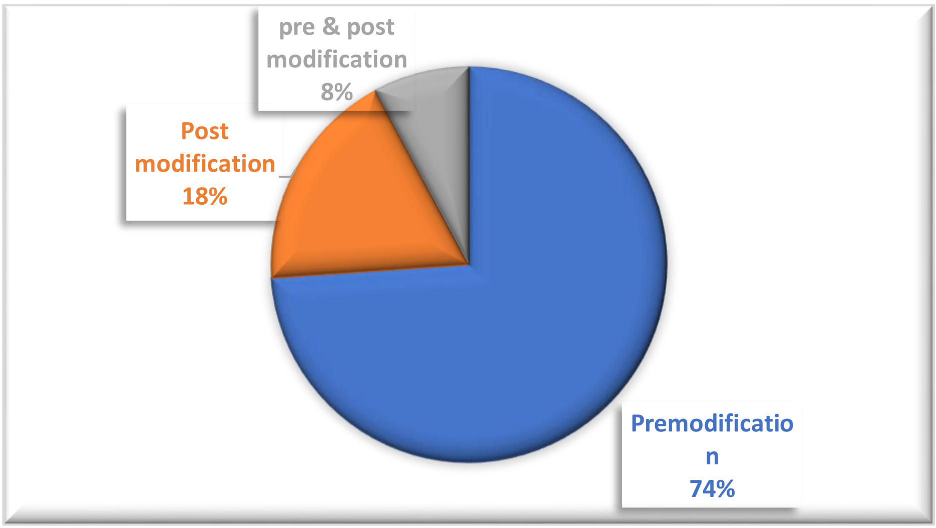


Figure 5.5. Noun Modification Processes in MRAC Multiword Collocations

Again, these findings are consistent with those in Biber & Gray (2016), who noted a steady increase in the use of nouns as noun premodifiers in modern science research writing. Noun premodification may thus be of pedagogical value in medical writing instruction. As rightly noted by Biber et al. (2020), learners need support to be able to parse “the dense packaging associated with phrasal complexity” (p. 13). This point is further discussed later in the conclusions, in Section 8.5. The next section discusses the findings of the functional analysis of multiword collocations in medical research articles.

5.2.1.2 Functions of Multiword Collocations in the MRAC

As indicated in 3.4.2, the decision was made to use Hyland’s (2008a) functional analysis framework as a starting point and then supplement it with any newly identified function. The adapted framework shown in 3.7. was used for the classification of multiword collocations. All identified sequences were analyzed in context using concordance lines in *Antconc* (Anthony, 2020). Table 5.10 shows the distribution of multiword collocations of various lengths in the major functional categories.

Table 5.10 Major Functions of MRAC Multiword Collocations by Length

	MRAC Multiword Collocations				
Major Functional Categories	3-words	4-words	5-words	6-words	7-words
Research-oriented	89.9%	100%	100%	83.33%	100%
Text-oriented	7.07%	0%	0%	16.67%	0%
Participant-oriented	3.03%	0%	0%	0%	0%

As shown in Table 5.10, the vast majority of multiword collocations in the MRAC, regardless of length, were research oriented, with the totality of 4-word, 5-word, and 7-word sequences belonging to this category. Figure 5.6 shows the distribution of research-oriented subfunctions served by multiword collocations in the MRAC. As shown in this figure, *Topic (Research Object and Related Elements)* was the most frequently served subfunction with 35% of research-oriented multiword collocations. It was closely followed by the *Procedure* subcategory (27%). The third most frequently served subfunction was *Description/Identification focus* (16%), closely followed by *Topic (Institutions*). *Location (in time/place)* and *Quantification* accounted for only 6% and 5% of research-oriented multiword collocations, respectively, suggesting that authors of medical research articles resort to some other devices (like lexical bundles, for example) to express these functions. Therefore, *Location* and *Quantification* are not further discussed in this section.

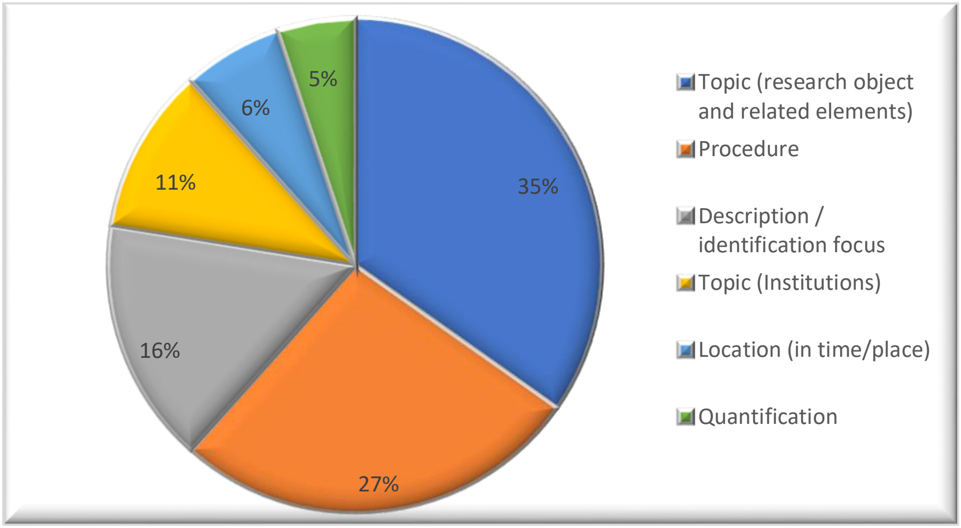


Figure 5.6. *Distribution of Research-Oriented Subfunctions Served by MRAC Multiword Collocations*

The function of *Topic (Research Object and Related Elements)* was almost exclusively served by complex noun phrases, regardless of length, with only two instances of 3-word coordinated binominals. Examples (5.30) and (5.31) below illustrate the use of complex noun phrases to refer to the object of the research (example 5.30) and elements related to the object of the research (example 5.31).

(5.30) ***Chronic obstructive pulmonary disease*** *(COPD) affects 5% to 22% of adults older than 40 years, has a lifetime risk of more than 25%, and is the third leading cause of death worldwide.* (14*JAMA*4)

(5.31) However, a cytokine storm syndrome caused by elevation of ***serum tumor necrosis factor–α*** *(TNF-α) and interleukin-2 (IL-2) triggered by intact Fc receptor (FcR)–binding anti-CD3 (FB-anti-CD3) prevents the clinical application of this conditioning regimen* (12*SCIENCE*2).

The second most frequent research-oriented function, *Procedure, was* also frequently expressed through complex noun phrases (example 5.32), but also through prepositional phrases (example 5.33).

(5.32) Between March 31, 2009, and June 2, 2014, we assessed 8820 women for eligibility and recruited 1555, with a mean BMI of 36·3 kg/m^2 (SD 4·8). 772 were randomly assigned to standard antenatal care and 783 were allocated the behavioural intervention, of which 651 and 629 women, respectively, completed an ***oral glucose tolerance test***. (15*LANCET*1)

(5.33) Total lysates of CD8$^+$ T cells after CD3/28 stimulation and treatment with 15 mM LA for the indicated time points were analyzed ***by Western blotting***. (20*Science*5)

Sequences used for *Description,* on the other hand, showed a different pattern. They were primarily 3-word coordinated binominals (example 5.34) and 3-word and 4-word clausal sequences in the form of declarative clauses/fragments (examples 5.35 & 5.36). As shown in examples (5.34) – (5.36), these sequences appear much less technical than those used to refer to research objects and procedures.

(5.34) All follow-up is through September 30, 2009. Events are the composite of ***fatal and nonfatal*** *cardiovascular events that occurred in the 12 months between each point.* (14*JAMA*3)

(5.35) The CONSORT flow diagram for both phase 1 and phase 2 is described in Figure 1. Baseline demographic and clinical characteristics ***were well balanced*** *across the 2 groups.* (20*JAMA*8)

(5.36) ***Secondary end points included*** *the duration of response, progression-free survival, and safety.* (20*NEJM*9)

Finally, sequences used to refer to institutions were predominantly complex noun phrases. The institutions they refer to appear to be internationally well-known as in the case of the *World Health Organization.* Authors of medical research articles appear to refer to those institutions primarily to validate the ethics of their procedures and research approaches, as can be seen in examples (5.37) and (5.38).

(5.37) The results of this trial prompted ***the World Health Organization*** *(WHO) and the United Nations Children's Fund (UNICEF) in 2006 to release a joint statement recommending that in malaria-endemic areas, iron supplementation (drops, syrup, or tablets) be given only to children who have anemia and are at risk of iron deficiency.*

(5.38) The protocol was completed when the infant's total serum bilirubin no longer met the criteria for study entry. Safety data were reviewed by ***the data and safety monitoring board*** *at the midpoint of the study (June 2013).* (15*NEJM*1)

In sum, this section has looked at the use of multiword collocations in medical research articles. The structural analysis revealed that multiword collocations, in their vast majority, consisted of complex noun phrases. Further analysis of those complex noun phrases revealed premodification as the most frequent process used by authors of medical research articles to create this type of phrases. Coordinated binominals and phrasal sequences were not as frequent as complex noun phrases, but they were found to be used more frequently than complex noun phrases in one of the functional subcategories. The functional analysis showed that multiword collocations in the MRAC were predominantly research-oriented sequences. While the functions of *Topic (Research Object and Related Elements), Procedure*, and *Topic (Institutions)* – first, second, and fourth most frequent research-oriented functions – were primarily realized through complex noun phrases, the third most frequent research-oriented function, *Description*, was served primarily with coordinated binominals and phrasal sequences.

5.2.2 Multiword Collocations in Medical Case Reports

5.2.2.1 Structures of Multiword Collocations in the MCRC.

Multiword collocations in the corpus of medical case reports were structurally analyzed using the same framework presented in 5.2. above. Table 5.11 shows the distribution of structural categories of multiword collocations in the MCRC. As was the case in the MRAC, phrasal sequences were also found predominant in the MCRC, accounting for 88.09%, 79.23%, 83.33%, and 63.63% of 3-word, 4-word, 5-word, and 6-word collocations, respectively. However, these proportions of phrasal multiword collocations in the MCRC are noticeably lower than expected. This is explained by a higher proportion of clausal sequences in the corpus. Indeed, clausal sequences accounted for 27.28% of 6-word sequences, 20.77% of 4-word sequences, and 16.67% of 5-words. They also represent 100% of the 7-word list, but there was only one item in that list. Still, it is worth noting that the only item in that list was a clausal sequence.

As shown in Table 5.11, clausal sequences in the MCRC were predominantly declarative clauses/fragments. The phrasal sequences, on the other hand, were largely dominated by complex noun phrases, as was the case in the MRAC. Given the latter are still highly frequent in the MCRC despite the increased use of clausal sequences, I decided to further investigate the modification processes involved in MCRC complex noun phrases to see whether authors resort to the same noun modification processes in the two registers. The findings are presented in Figure 5.7.

Table 5.11 Distribution of Structural Categories of Multiword Collocations in the MRAC

		MCRC Multiword Collocations				
Structures		3-words	4-words	5-words	6-words	7-words
Phrasal Sequences	Complex NP	84.52%	77.35%	83.33%	54.54%	0%
	PP	3.57%	1.88%	0%	9.09%	0%
	other	0%	0%	0%	0%	0%
subtotals1		*88.09%*	*79.23%*	*83.33%*	*63.63%*	*0%*
Clausal sequences	Declarative Clauses/fragments	2.38%	20.77%	16.66%	18.19%	100%
	Dependent Clauses/Fragments	1.20%	0%	0%	9.09%	0%
subtotals 2		*3.58%*	*20.77%*	16.67%	27.28%	100%
Coordinated Binominals		8.33%	0%	0%	9.09%	0%
Totals		*100%*	*100%*	*100%*	*100%*	*100%*

Complex noun phrases in the MCRC multiword collocations are almost exclusively formed through premodification, and as was previously observed in the MRAC, here also, a frequent use of nouns as noun premodifiers was noted. In longer sequences, nouns were used together with other premodification devices such as adjectives and participial premodifiers (e.g., *left ventricular ejection fraction*, *packed red blood cells*, *complete blood cell count*). In light of these findings and based on Biber & Gray's (2016) claim regarding the shift of modern scientific writing toward phrasal modifiers, it can be inferred that medical writing, in general, is fundamentally phrasal and that noun premodification plays an important role in phrasal complexity in this discipline. As mentioned above, this is precious information for both medical writing instructors and learners.

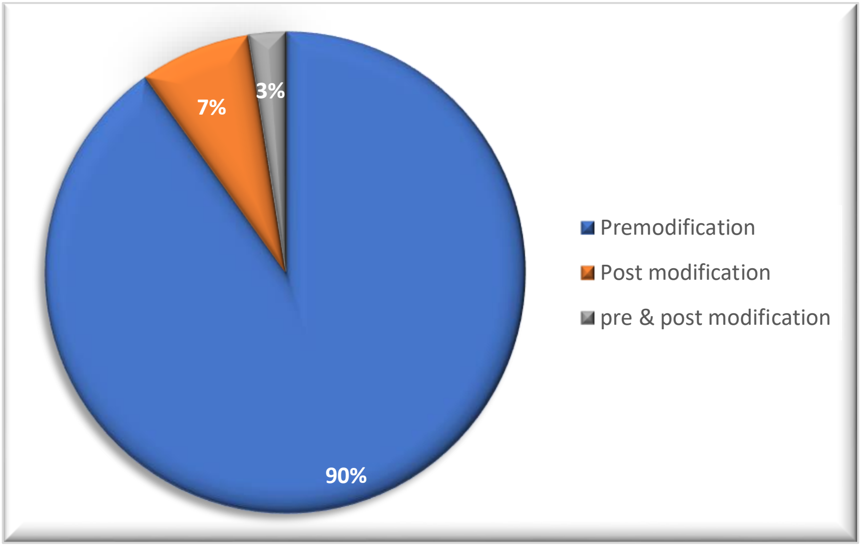


Figure 5.7 Noun Modification Processes in Multiword Collocations in the MCRC

However, as previously mentioned, there was one salient difference between the structures of multiword collocations in the MCRC and the MRAC. Clausal sequences were also relatively frequent in medical case reports, primarily in the form of declarative clauses/fragments. These were most of the time short and simple complete clauses or fragments (e.g., *function tests were normal, physical examination revealed a, vital signs were within normal limits*) that appeared to serve important functions in case reports, as will be seen in the next section below.

5.2.2.2 Functions of Multiword Collocations in the MCRC.

As mentioned in 3.4.2, for the functional classification of MCRC multiword collocations, I supplemented the initial framework developed for the collocations with some functional categories from Biber et al. (2004) and Hyland (2008a). This framework is summarized in Table 5.12. The findings of the functional analysis are shown in Table 5.13. Multiword collocations in the MCRC were almost exclusively Diagnosis/Intervention-related and Case-related. Figure 5.8. shows the distributions of the subfunctions served by the MCRC multiword collocations in each of these two categories. As shown in Figure 5.8, Description was the most frequently served function in the Case-related category, accounting for 60% of all sequences in this category. Procedure and Biological processes/elements, on the other hand, were predominant in the Diagnosis/Intervention-related category, with 44% and 32%, respectively, of all multiword collocations in this category.

Table 5.12 Functional Analysis Framework of Multiword Collocations in the MCRC

Categories	**Sub-categories**	**Examples from the MCRC**
Case-related	Subject characteristics (age, race, gender, etc.)	*year old woman, old male Caucasian*
	Location (in time/place)	*the intensive care unit, days prior to presentation*
	Description	*the most common site, heart failure with preserved ejection fraction*
	quantification	*fraction of inspired oxygen, index of suspicion*
	Other Individuals (involved with or relevant to the case)	*primary care provider, her general practitioner*
Diagnosis/ Intervention-related	Biological processes/elements	*abdomen and pelvis, right bundle branch block, blood urea nitrogen*
	Procedures	*under general anesthesia, fluorescence in situ hybridization*
	Findings	*physical examination was unremarkable, medical history was significant*
	Decisions / treatment	
	Institutions	*the World Health Organization, Centers for Disease Control and Prevention*
Outcome/Follow-up – related		N/A
Discourse organizers	Resultative signals	N/A
	Framing signals	N/A
	Transition signals	N/A
	Structuring signals	N/A
	Elaboration/Clarifications	N/A
Stance Features		*further studies are needed, highlights the importance*

Table 5.13 Major Functions of MCRC Multiword Collocations by Length

	MCRC Multiword Collocations				
Major Functional Categories	3-words	4-words	5-words	6-words	7-words
Case-related	32.55%	35.84%	60%	70%	0%
Diagnosis/ Intervention-related	61.62%	54.71%	40%	30%	100%
Outcome/Follow-up – related	0%	0%	0%	0%	0%
Discourse organizers	0%	0%	0%	0%	0%
Stance Features	2.32%	3.77%	0%	0%	0%

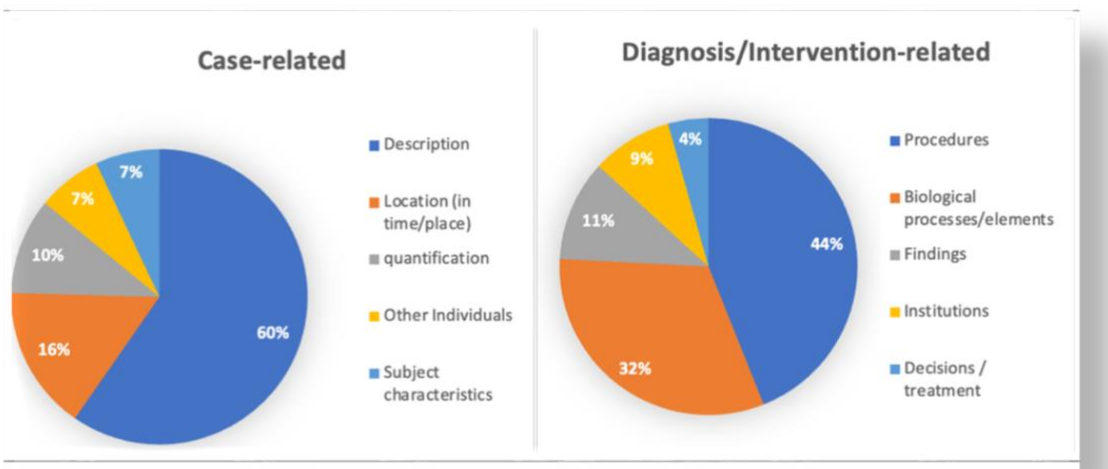


Figure 5.8 Case-related and Diagnosis-related Subfunctions of MCRC Multiword Collocations

Multiword collocations occurring in the Description subcategory were mostly used to refer to the medical condition at hand (example 5.39) and provide details in the initial presentation of the case (examples 5.40 & 5.41). Sequences serving these functions were primarily complex noun phrases (example 5.39) but also included coordinated binominals (example 5.40) and declarative clauses/fragments (example 5.41).

*(5.39) A 67-year-old man with **type 2 diabetes mellitus** and hyperlipidemia was evaluated for unintentional weight loss of 28 lb (12.7 kg) and increasing fasting blood glucose values over the past 6 months*. (19*JAMA*169)

*(5.40) He was **alert and oriented**, able to answer questions and follow commands appropriately*. (20*BMJ*4)

*(5.41) **Her medical history included** diabetes, hypertension and chronic obstructive pulmonary disease (COPD).*

Multiword sequences in the Procedures subcategory were used to refer to diagnosis and intervention procedures. They consisted exclusively of complex noun phrases and often were highly technical expressions as shown in examples (5.42) and (5.43) below.

*(5.42) Various therapies have been used for the patient, including nebulized α-interferon, umifenovir, and lopinavir/ritonavir, methylprednisolone, antibiotic therapy, biliary drainage by **endoscopic retrograde cholangiopancreatography**, continuous renal replacement therapy, and plasma exchange.* (20*IMCRJ*94)

*(5.43) We could not identify EBV DNA in the DLBCL-containing portion of the specimen by **polymerase chain reaction** (PCR) and breaks in the MALT1/18q21 gene in the MALT lymphoma-containing portion by **fluorescence in situ hybridization**.* (20*JMCR*27)

In addition to these frequent functions served by the MCRC multiword collocations, there is one Diagnosis/Intervention subcategory, *Findings*, that was not as frequent as the ones described above but displayed a particularity that may be of interest in medical case report writing instruction. Despite the predominance of complex noun phrases in all other functional categories, this category did not involve any complex noun phrase. It was served exclusively by declarative clauses/fragments that appeared as ready-made sequences for reporting the results of examination and diagnosis procedures, as shown in examples (5.44) and (5.45). This probability explains the noticeably higher proportion of phrasal sequences in the MCRC, compared to the MRAC.

(5.44) ***Physical examination revealed a*** *tall gentleman with blood pressure of 104/75, heart rate of 90 bpm and weight of 89 kg.*

(5.45) ***Magnetic resonance imaging revealed*** *bitemporal edematous lesions, greater on the left than the right sides*

Note in these examples how each clause/fragment starts with the source of the information being provided as its subject. This, according to Helan (2012), is typical of medical case report writing where authors tend to use medical technology and procedures as Agents. From a pedagogical perspective, this finding is particularly important as the declarative clauses/fragments in the *Findings* category are vehicles of required information in any medical case report. As a reminder, the functional framework was developed primarily based on all required information in a case report, per the submission guidelines of journals used in the corpus collection and according to Gagnier et al. (2014). It appears then, that novice L1 writers and L2-English medical professionals seeking to publish medical case reports, will greatly benefit from instruction of these ready-made clauses/fragments.

To complete the answer to the first part of RQ2b (*What multiword collocations are used in the two registers?),* we have seen in this section that multiword collocations in medical case reports are predominantly complex noun phrases created in their vast majority through premodification. As case-related sequences, these noun phrases were primarily used in the description of the medical cases being presented. As Diagnosis/Intervention-related, they were used to refer to procedures and biological processes/elements. Next to complex noun phrases, phrasal sequences were also frequent and appeared to specifically serve the function of reporting findings in the Diagnosis/Intervention-related category.

The findings presented in 5.2.1 and 5.2.2. already indicate some variations in the use of multiword collocations in MRAs and MCRs. To supplement the comparison, I now turn to the final section of the present chapter for an analysis of multiword collocations shared between the two registers.

5.2.3 Multiword Collocations in the MRAC and MCRC

This section answers the second part of RQ2b that asks whether identified multiword collocations are register-specific or rather shared within the same discipline. As already mentioned, the findings presented in 5.2.1 and 5.2.2. already provide some answers to this question. We have seen that multiword collocations in both registers are predominantly phrasal with a high prevalence of complex noun phrases. This was understandable given the noted shift of scientific writing from dependent complement clauses toward phrasal modifications (Biber et al., 2020; Biber & Gray, 2016). It may also be the case that medical writers use complex noun phrases in an aim to be more concise in their writing. These findings, as mentioned above, highlight the pedagogical value of including noun premodifications in medical writing syllabi and material design. However, it appears that the similarities end there. In terms of functions, it was not even possible to use the same analysis framework for MRAC and MCRC multiword collocations, given the differences in foci and organizational structure, to name only these two, discussed in the analysis of the situational characteristics of these texts in 4.3.4. and 4.3.5. As was observed with 2-word collocations, multiword collocations in the two registers served very distinct functions. It would be reasonable to expect Discourse Organizers would be shared across the two registers, but as can be seen in Tables 5.12 and 5.13, MCRC multiword collocations did not serve those functions. Perhaps upcoming chapters will shed more light on that specific point. One last step in this comparison was to identify and analyze shared multiword collocations. The findings seem to confirm what already transpired from sections 5.2.1 and 5.2.2. The two corpora shared only eleven 3-word, two 4-word, and two 6-word collocations. The shared sequences are listed in Table 5.14.

Table 5.14 Shared Multiword Collocations across the two Corpora

3-words	4-words	6-words
calcium channel blockers	*chronic obstructive pulmonary disease*	*Centers for Disease Control and Prevention*
sensitivity and specificity	*the World Health Organization*	
polymerase chain reaction		
stem cell transplantation		
intensive care unit		
written informed consent		
the United States		
acid fast bacilli		
hematoxylin and eosin		
in situ hybridization		

Given these findings, it can be said that most multiword collocations identified in the MRAC and MCRC are specific to each register. Nevertheless, I analyzed the shared sequences in context to see whether they would reveal some similarities or further variations. The analysis did not reveal noticeable variations, apart from the different uses of the sequence *intensive care unit* in the two corpora. This may be because many of the shared sequences are technical terms referring to specific medical processes (e.g., *stem cell transplantation, calcium channel blockers, in situ hybridization, polymerase chain reaction*). Examples (5.46) – (5.49) show the use of some of these technical sequences in both MRAs and MCRs.

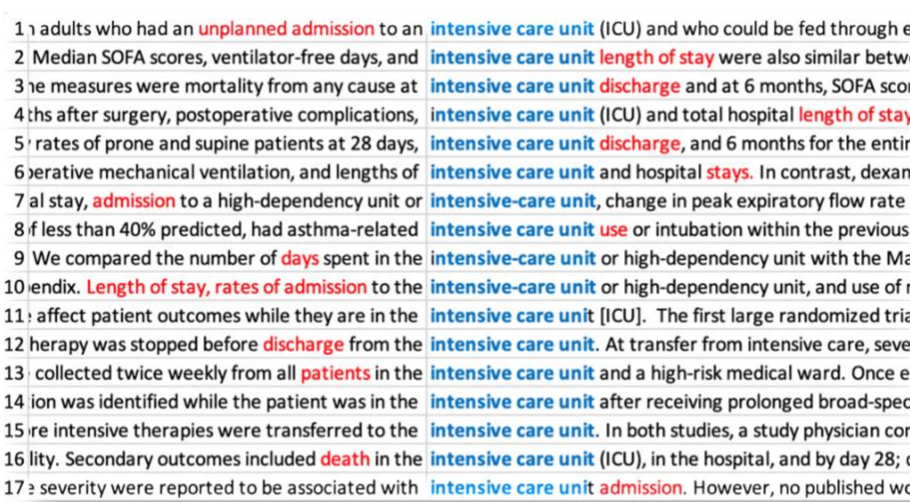
(5.46) Microscopically, using ***Hematoxylin and Eosin*** *(H&E) stained sections, the tumor was composed of atypical spindle cells and large pleomorphic epithelioid cells and multinucleated cells of osteoclastic type which were forming a new bone matrix.* (MCRC_20*IMCRJ*44)

(5.47) Lung tissues were embedded in paraffin, and 10-μm sections were prepared and stained using ***hematoxylin and eosin*** *and trichrome stain*. (MRAC_20*Science*7)

(5.48) Immunohistochemical analysis by ***in situ hybridization*** *showed strong positive (3+) HER2 expression*. (MCRC_19*OMCR*140)

(5.49) Importantly, ***in situ hybridization*** *revealed an indistinguishable cellular distribution of Notch3 mRNA between Notch3 transgenic and nontransgenic brains.* (MRAC_10*JCI*1)

The use of the sequence *intensive care unit*, on the other hand, reflects the difference in focus between the two registers, In the MRAs, the sequence is just one of the elements of the study and is mostly used as a whole unit that modifies other nouns or is part of longer strings functioning as noun modifiers, as shown in figure 5.9. The modified nouns or phrases are in red.



1	ı adults who had an unplanned admission to an	intensive care unit	(ICU) and who could be fed through e
2	Median SOFA scores, ventilator-free days, and	intensive care unit	length of stay were also similar betwe
3	ıe measures were mortality from any cause at	intensive care unit	discharge and at 6 months, SOFA scor
4	ths after surgery, postoperative complications,	intensive care unit	(ICU) and total hospital length of stay
5	' rates of prone and supine patients at 28 days,	intensive care unit	discharge, and 6 months for the entir
6	›erative mechanical ventilation, and lengths of	intensive care unit	and hospital stays. In contrast, dexam
7	al stay, admission to a high-dependency unit or	intensive-care unit	, change in peak expiratory flow rate a
8	f less than 40% predicted, had asthma-related	intensive care unit	use or intubation within the previous
9	We compared the number of days spent in the	intensive-care unit	or high-dependency unit with the Ma
10	endix. Length of stay, rates of admission to the	intensive-care unit	or high-dependency unit, and use of r
11	affect patient outcomes while they are in the	intensive care unit	[ICU]. The first large randomized tria
12	herapy was stopped before discharge from the	intensive care unit	. At transfer from intensive care, sever
13	collected twice weekly from all patients in the	intensive care unit	and a high-risk medical ward. Once ea
14	ion was identified while the patient was in the	intensive care unit	after receiving prolonged broad-spect
15	re intensive therapies were transferred to the	intensive care unit	. In both studies, a study physician con
16	lity. Secondary outcomes included death in the	intensive care unit	(ICU), in the hospital, and by day 28; d
17	severity were reported to be associated with	intensive care unit	admission. However, no published wo

Figure 5.9. Use of 'intensive care unit' in MRAs

On the other hand, in the MRAs, *intensive care unit* often refers to the setting of the case being reported. The sequence primarily occurs in passive sentences in a *to* prepositional phrase functioning as place adverbial, as shown in figure 5.10.

Figure 5.10. Main Pattern of Use of 'intensive care unit' in MCRs

5.3 Conclusion

In this chapter, I have tried to answer RQ 2a and RQ2b that asked about the use of collocations and multiword collocations. The structural and functional analyses have shown that these two types of formulaic sequences are in many cases different in essence and serve distinct functions in case reports and medical research articles. As the present study aims at a comprehensive investigation of the formulaic profiles of the two registers, the upcoming chapters will certainly shed more light on potential variations in the use of formulaic language in these two medical registers.

6 CHAPTER 6. FINDINGS AND DISCUSSIONS 2: LEXICAL BUNDLES IN MRAs AND MCRs

In this chapter, I report and discuss the use of lexical bundles in medical research articles and medical case reports. Lexical bundles, most frequent sequences of words in a corpus (Biber et al., 2004), have been referred to as the building blocks of academic writing (Cortes, 2013) and extensively described in the literature (e.g., Biber et al., 1999; Biber et at, 2004; Cortes, 2004, 2006, 2008, 2013; Hyland, 2008a, 2008b; Hyland & Jiang, 2018). In the discussion of the findings presented in this chapter, I also make comparisons with previous descriptions of bundles in academic writing.

Bearing in mind the initial goals of investigating within-discipline variations and providing a comprehensive description of the two registers under scrutiny in this study, I first present the findings from each corpus individually, before the comparisons of the structures and functions of bundles in the two registers. Therefore, section 6.1. is on the structures and functions of Bundles in medical research articles. Section 6.2. presents the findings of the structural and functional analyses of bundles in medical case reports, and section 6.3. discusses the findings of the analysis of the bundles shared between the two corpora. These sections answer the following research question and subquestions:

Research Question 2c: How are Lexical Bundles used in both registers? How do they compare to bundles previously identified in academic prose? Are there any similarities and/or differences in terms of length, structure, and function?

6.1 Lexical Bundles in Medical Research Articles

Bundles in both the MRAC and MCRC were identified using LBiaP (Lake & Cortes, forthcoming) and the exploratory approach described in 3.5, with different frequency and range thresholds established for different bundle length. Sequences from 3 to up to 11 and 8 words were identified in the MRAC and the MCRC, respectively. Table 6.1. shows the number of bundle types identified in each corpus for each bundle length. Even though the two corpora are of different sizes, the total number of bundle types from each corpus are relatively similar, albeit some slight differences depending on bundle length. The number of 3-word bundles was noticeably higher in the MRAC than in the MCRC (362 and 270 bundle types, respectively), but 4-words, 5-words, and 6-words in the MCRC slightly outnumbered those in the MRAC, which was a larger corpus.

Table 6.1. Raw Numbers of Bundle Types Identified in the MRAC and MCRC

Bundles	**MRAC**	**MCRC**
3-words	362	270
4-words	116	142
5-words	40	61
6-words	12	31
7-words	1	1
8-words	1	1
11-words	1	0
Totals	**533**	**506**

6.1.1 Structures of Bundles in the MRAC

The structural classification of bundles was made using the framework described in 3.5 and presented in Table 3.9. As explained in 3.5, the framework draws primarily from Biber et al. (1999), Hyland & Jiang (2018), and Cortes (in press), and was supplemented based on observations from the lists of bundles identified in the two corpora. This approach made it possible to classify all identified bundles, including those usually listed in the "Others" category in the literature. Table 6.2. shows the structural categories identified in the MRAC and the distribution of the major categories in the corpus is shown in Figure 6.1.

Table 6.2. Structures of Lexical Bundles in the MRAC

Main Categories	**Subcategories**	**Examples from the MRAC and MCRC**
Verb phrase-related	Passive verb	*were included in the, has been shown to, the patient was started on*
	copular be	*were eligible for, is a rare, is the most common*
	imperative	*see the supplementary appendix*
	modals	*can be a, may lead to, can also be*
Clause-related	anticipatory it	*it is possible that, it is important to*
	abstract subject	*the primary end point was, his blood pressure was*
	human subject	*all participants provided written informed consent, our patient presented with*
	external subject	*did not reveal any, confirmed the diagnosis of*
	as-fragments	*as measured by, as described previously, as shown in*
	if-fragments	
	there fragments	*there were no differences in, there was no family history of*
	(NP) + wh-fragments	*when compared with, patients who were*
	(VP) + that-fragments	*these data suggest that, we found that the*
	(VP) + to-fragments	*to assess the effect of, to confirm the diagnosis*
Noun/Preposition-related	NP with of-phrase fragment	*the onset of, a single dose of, a wide range of*
	NP with other post modifier fragments	*a man in his, death from any cause, a decrease in*
	PP with embedded of-fragment	*at the end of the, at the time of diagnosis*
	Other prepositional fragments	*in individuals with, in addition to the*
	comparative expressions	*as in this case*
adjective/ Adverb-related	AdjP /AdvP + prep + …	*consistent with the, more likely to, similar to that*
	Superlatives	*the most common,*
Function words only (3-words only)		*as in our, and in the, but not in*

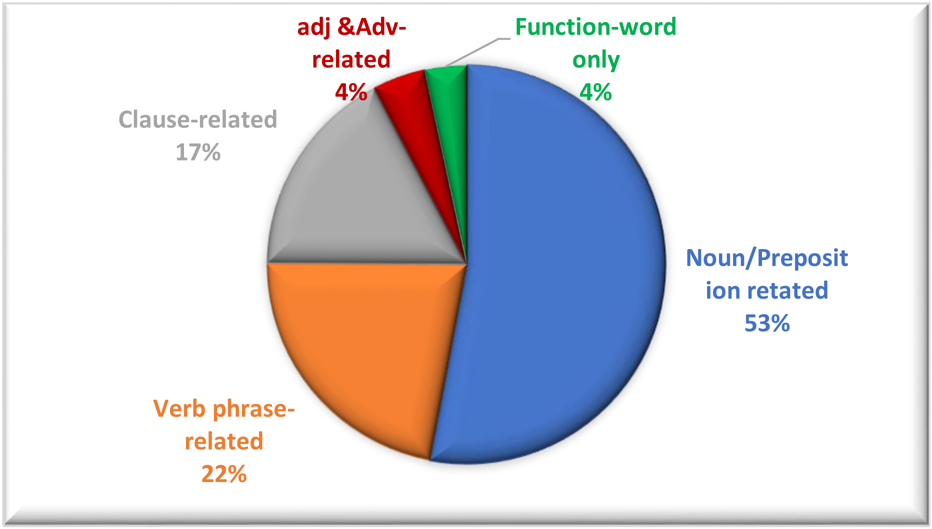


Figure 6.1. Distribution of Structural Categories in the MRAC

Overall, bundles in the MRAC were primarily Noun/Preposition-related (53%), which is consistent with previous studies that have shown the predominantly phrasal nature of academic writing (e.g., Biber et al., 1999; Cortes, 2004; Hyland, 2008a; Hyland & Jiang, 2018). The next most frequent pattern was the Verb Phrase-related category, which accounted for 22% of all identified bundles. Clause-related bundles were also relatively frequent and represented 17% of bundles identified in the MRAC. A closer look at how bundles are distributed in these three major categories shows some variations depending on bundle lengths. Table 6.3 presents the distribution of bundles in these three major categories.

As shown in Table 6.3, the 3-word and 6-word lists of bundles were NP-dominated in the Noun/Preposition-related category, while the 4-word and 5-word bundles in this category were predominantly PP-related. Together, NP-related bundles represented 64.36% of 3-word sequences in this category with 76 bundles out of the 188 (40.42%) being NPs with *of*-phrase fragments and 45 (23.63%) being NPs with other postmodifiers. The only 8-word and 11-word bundles in the MRAC were also NP-related. On the other hand, in the 4-word bundle list, PP-related bundles accounted for 66.66% of bundles in this category, with the same number of 4-

word bundles (21) in the subcategories of PPs with embedded *of*-phrases and Other Prepositional Fragments. The 5-word bundle list was also PP-dominated but included only PPs with embedded *of*-phrases which accounted for 55% of all 5-word bundles in this category.

Table 6.3. Distribution of MRAC Bundles across the Major Structural Categories

Bolded numbers represent the most frequent structures in each bundle list.

Major Categories	**Subcategories**	3-words	4-words	5-words	6-words	7-words	8-words	11-words
Noun/Preposition related	NP with of-phrase fragment	**76**	14	8	2		1	
	NP with other post modifier fragments	**45**	7		**4**	1		
	PP with embedded of-fragment	5	**21**	**11**	2			
	Other prepositional fragments	**62**	**21**					
	comparative expressions			1	1			
Total Noun/Preposition-related		*188*	*63*	*20*	*12*	*1*	*1*	*0*
Verb phrase-related	Passive verb	**69**	**23**	**8**	**2**			**1**
	copular be	6	3	3				
	imperative		1					
	modals	2						
Total verb-related		*77*	*27*	*11*	*2*	*0*	*0*	*1*
Clause-related	anticipatory it	1	1	1				
	abstract subject	6	2	2				
	human subject	7	2		1			
	external subject	8	5	2				
	as-fragments	7						
	if-fragments							
	there fragments	4	2	1				
	(NP) + wh-fragments	2	1					
	(VP) + that-fragments	**11**	**10**					
	(VP) + to-fragments	**14**	1	1				
Total clause-related		*60*	*24*	*7*	*1*	*0*	*0*	*1*

In the Verb-related category, passive-related bundles were the predominant structure across all bundle lists, accounting for 89.61% (69 out of 77) of 3-words (e.g., *was used for*, *were obtained from, were associated with, was observed in*), 85.18% (23 out of 27) of 4-words (*has been associated with*, *was added to the*, *was performed using the, were enrolled in the*), and 72.72% (8 out 11) of 5-word bundles in this category (*were randomly assigned to receive, has been shown to be, were randomly assigned in a*). The two 6-words (*associated with an increased risk of, the study was approved by the*) and the single bundle in the 11-words list (*calculated as weight in kilograms divided by height in meters squared*) were also passive constructions.

Despite the reported decline of passive constructions in scientific writing in recent years (Biber & Gray, 2016; Hyland & Jiang, 2016; 2018; Ping, 2014; Seoane, 2013), passive-related bundles appear to still be relatively frequent in medical research articles. This frequent use of passive constructions is justified in these terms by Millar et al. (2013), who specifically studied the use of passives in medical research articles: “avoidance of the passive voice is both difficult and not necessarily desirable” (p. 410). The authors further argued that the use of the passives in the Methods and Results sections help authors focus the reader’s attention on the research, and that some “semi-fixed formulae” in the passive “represent a preferred means of expressing concepts relating to medical research in general” (p. 410).

From a pedagogical perspective, this may be of interest in medical writing instruction. As rightly suggested by Millar and colleagues, given the growing recommendations from medical journals to “use the active voice whenever possible” (p. 393), novice and L2-English medical writers will certainly need guidance on when it is indeed preferable to use passive-related expressions. The passive-related bundles may then be very good candidates for instruction.

Finally, clause-related bundles showed a relatively more even distribution among the subcategories, with a slight predominance of bundles involving finite ((VP) + *that*-fragments) and non-finite ((VP) + *to*-fragments) complement fragments in the 3-words and 4-words lists. Next to these bundles there are also active clause fragments with abstract subject (e.g., *table 1 shows, previous studies suggested),* human subject (e.g., *we did not observe, we assessed the, all participants provided written informed consent*), and external subject (e.g., *play a role in, did not differ between*). I refer to the latter subcategory as "external subject" given that the subject occurs outside the bundle, which is understandable given that bundles are often not complete structures.

Together, the three types of active clause fragments mentioned above make up for 35% (21 out of 60) of 3-word clause-related bundles (e.g., *we observed a, we assessed the, did not affect, contribute to the*) and 33.33% (8 out of 24) of 4-words (e.g., *we did not observe, play a role in, had no effect on*). Four of the eight 5-word clause-related bundles were also active clause fragments (e.g., *plays a critical role in, the primary endpoint was*). The use of these types of clausal bundles may be the result of the growing recommendations to "use the active voice whenever possible". Additionally, some medical journals now encourage authors to be simple and straightforward in their writing to make it easy to read (Millar et al., 2013). The *BMJ* (2020), for example, in its 'Tips for writing' states: "Write as you speak. Keep it short and informal". Recommendations of this kind may also explain the use of these types of clausal bundles formerly found to be more frequent in spoken discourse (Biber et al., 1999).

On the other hand, clause-related bundles involving finite and non-finite complement fragments are reportedly on the decline in scientific writing (Biber & Gray, 2016; Hyland & Jiang, 2016; 2018). However, the (VP) + *that*-fragments and (VP) + *to*-fragments identified in the MRAC (e.g., *to test whether, to evaluate the, to determine whether, these findings suggest that, our data indicate that, these results demonstrate that, we hypothesized that*) appear to still serve important functions in medical research articles, some of which will be discussed in Section 6.1.2.

In sum, the findings reported here are consistent, to some extent, with what has previously been reported in the literature; that is, noun phrase-related and prepositional phrase-related are the most frequent bundle structures in academic writing. However, we have also seen that verb-related bundles with passive constructions, found on the decline elsewhere, are still relatively frequent in medical research articles. This finding is a reminder of Biber & Barbieri's (2007) caution that each discipline and register should be considered in their own right. We have also seen in MRAs the use of bundles previously described as characteristic of spoken discourse, *i.e*., clausal bundles in the form of active clause fragments. Given recent recommendations to authors in medical journals, these types of bundles may deserve some attention in medical writing instruction. After this structural classification, I now turn to the findings of the functional analysis of bundles in medical research articles.

6.1.2 Functions of Bundles in the MRAC

As indicated in 3.5, Hyland's (2008a) functional analysis framework was used as a starting point for the functional classification of lexical bundles in the MRAC. The identified bundles were analyzed in context in *Antconc* (Anthony, 2020). Only one new function, *Grammatical Only,* from Cortes's (2008, in press) functional analysis of 3-word bundles in academic writing in Spanish and English, was identified in the MRAC and added to Hyland's framework. I used the final framework shown in Table 3.10 to classify the MRAC bundles according to their discourse functions. In this process, bundles that served multiple functions were listed in multiple categories. The sequence *in patients with*, for example, functioned as both a framing signal and a structuring signal. This double function is illustrated in examples (1) and (2) below. In (1), *in patients with* is a framing signal, defining the specific group of patients, in general, for whom the proposed treatment has been successful to some degree. In (2), this bundle functions as study subject reference, referring to the two study groups. The bundle was thus listed in both categorical functions.

(6.1) ***In patients with*** *acutely decompensated cirrhosis, human serum albumin (HSA) administration has been shown to reduce inflammation […].* (20*Science*10)

(6.2) *We observed that LA increased* ***in patients with*** *AML relapse after allo-HCT but not* ***in patients with*** *AML who were in remission after allo-HCT* (20*Science*5)

Coming to the findings of the functional analysis, Table 6.4. shows the distribution of bundles across the main functional categories. Bundles in the MRAC were predominantly research-oriented, regardless of length. Research oriented bundles accounted for 60.22% of all 3-word bundles, 56.03% of 4-words, and 62.5% of 5-words. Seven out of the twelve 6-words (58.33%) and the single 8-word and 11-word bundles were also research-oriented. The rest of the bundles were almost all text-oriented with a higher proportion of 4-words in this category (43.97%), followed by 3-words (39.5%) and 5-words (35%). Five of the twelve 6-words and the only 7-word were also text-oriented.

Participant-oriented and Grammatical only were extremely rare in the MRAC. Grammatical Only accounted for only 0.83% of 3-word bundles, and Participant-oriented bundles represented 1.72% of 4-words, 2.21% of 3-words, and 5% of 5-words. Therefore, these two categories are not further discussed in this functional analysis. Hyland & Jiang (2018) found that participant-oriented bundles are on the rise in scientific writing, but that was not borne out by the findings from the MRAC. It should be noted that Hyland & Jiang's corpus did not include medical research articles and therefore, their findings may reflect only the writing practices in the disciplines (biology and electrical engineering) they chose to represent scientific writing in their study.

Table 6.4. Distribution of MRAC Bundles in the Major Functional Categories

	MRAC Bundles: Raw Number (%)						
Main Functions	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*

Research-oriented and Text-oriented bundles were further analyzed to identify the specific functions they served in these two major categories. Figure 6.2 shows the subfunctions served by the bundles in the MRAC. Description was the most frequently served function in the Research-Oriented category; with 50% of 5-word bundles (e.g., *the baseline characteristics of the, is one of the most, has been shown to be*), 43.8% of 4-words (e.g., *in the pathogenesis of, death from any cause, was defined as the, has been associated with*), 43% of 6-words (e.g., *associated with an increased risk of, a p value of less than*), and 38.07% of 3-words (e.g., *such as the, with and without, the most common, the effects of)* . The only 11-word bundle (*calculated as weight in kilograms divided by height in meters squared*) was also used for description.

The main structures of bundles in this subcategory were NP-related, PP-related, copular be, and passive-related. Bundles of three words were primarily NP-related, with 46 out of the 83 (55.42%) 3-word bundles used for description being NP-related (e.g., *the expression of, the* *effectiveness of, the change in, the hypothesis that, the ability to*). On the other hand, 4-word bundles were dominated by sequences involving the passive (e.g., *were included in the, was defined as a, was defined as the, has been associated with, was not associated with*). These passive sequences represented 11 of the 28 (39.28%) 4-word bundles used for description.

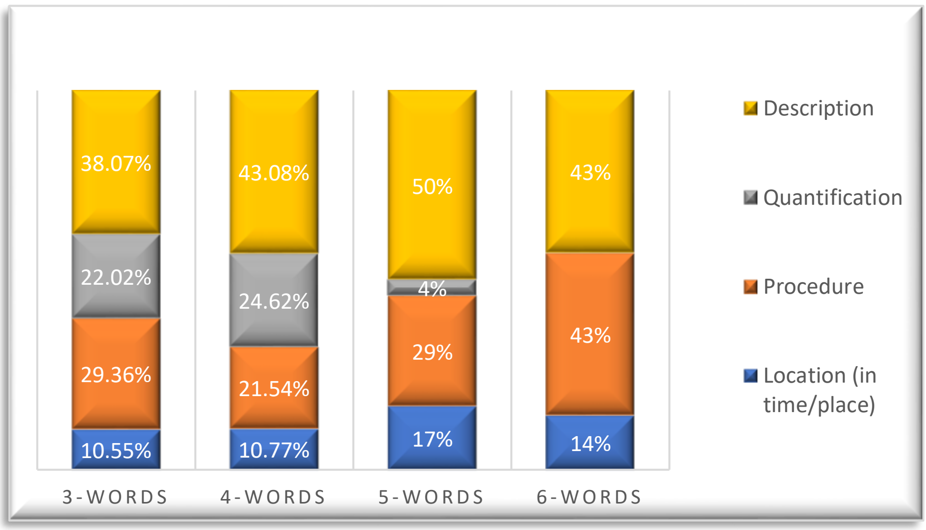


Figure 6.2. Distribution of Research-Oriented Bundles Served by MRAC Bundles

Passive-related sequences were also the most common structure in 5-word and 6-word bundles used for description, with all three 6-words and 5 out of the twelve 5-words (41.66%) in this subcategory being passive sequences. The only 11-word bundle in the MRAC and in this subcategory was also a passive construction (*calculated as weight in kilograms divided by height in meters squared*). Even the NP-dominated 3-word list included 12.04% of passive-related bundles. This is indication, as mentioned above, that passive-related bundles are still frequent in medical research articles and indeed serve important functions as was revealed by the qualitative analysis in context conducted in *Antconc* (Anthony, 2020).

Millar et al. (2013) rightly suggested that passive sequences are paramount in expressing concepts related to the medical field. However, in the description subcategory, authors resorted to passive sequences primarily to use results of previous research in the description of the object of their study (example 6.3) and elements related to it (example 6.4).

(6.3) Left ventricular hypertrophy, greater LVMI, or both ***have been shown to*** *predict CVD outcomes in both observational studies and clinical trials.* (8*JAMA*5)

(6.4) 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.* (20*JCI*7).

As for NP-related bundles in the Description subcategory, they simply were, as the name implies, descriptive bundles and served their usual purpose already described in previous research (e.g., Biber, 1999; Cortes, 2004), that is, to provide details on topics being discussed. For example, in the MRAC they were used to highlight characteristics of or provide details on various elements of the study, as in examples (6.5) to (6.7).

(6.5) Arg1 encodes the enzyme arginase, which is known as ***a marker of*** *protumorigenic (M2-polarized) macrophages/ microglia in gliomas (43) and has not yet been reported to be expressed by tumor cells at significant levels*. (20*JCI*10)

(6.6) We measured ***the expression of*** *inflammatory molecules in the colons of sugar-treated and untreated Il10*$^{-/-}$ *mice. (*20*Science*1)

(6.7) In addition, to determine whether the presence of bilirubin adversely affected ***the results of our*** *POC anemia test, we systematically added bilirubin to existing blood samples, conducted our assay, and measured our assay's Hgb levels based on spectrophotometry absorbance data.* (14*JCI*3)

The other bundle structures used for Description were copular *be* and PP-related 3-word and 4-word bundles. They served similar purposes as NP-related bundles. Additionally, copular *be* bundles were also used to make strong statements about the study and/ or its results, as in examples (6.8) and (6.9).

(6.8) *We believe this* ***is the first study*** *simultaneously to quantify lipid antigen–specific and protein antigen–specific T cells in the same individuals […]* (16*NEJM*4)

(6.9) *Thus, the lack of or delay in restoration of interstitial CD4+ T cells in the lungs of ART-treated animals in our study* ***is consistent with the*** *reactivation of TB that occurred despite ART.* (20*JCI*4)

Copular *be* was listed by Biber & Gray (2016) among common grammatical features in academic prose, and Hyland and Jiang (2018) reported an increase in the use of copular *be* bundles in scientific writing. Given the functions just described above, these types of bundles may also be of interest for medical writing instruction.

The next most frequently served research-oriented function was Procedure. It accounted for 29.36% of research-oriented 3-word bundles, 21.54% of 4-words, and 29% of 5-words. Three of the seven 6-word research-oriented bundles were also in this subcategory (*all participants provided written informed consent, wrote the first draft of the, the decision to submit for publication*). Bundles serving this function were by far dominated by passive-related sequences, mainly in the 3-word list, with 32 out of 64 (50%) sequences used to describe procedure (e.g., *were used for, were included in, was performed using, were assigned to*), the 4-word list, with 11 out of 14, that is, 78.57% (e.g., *were considered statistically significant, were excluded from the, were enrolled in the, was added to the*), and the 5-word list, with 4 out of 7 (57.14%) sequences describing procedures (*were randomly assigned to receive, were randomly assigned in a, total* *rna was isolated from*). In addition to referring to the field-related concepts mentioned by Millar and colleagues (example 6.10), passive bundles in this subcategory were also used to describe methodology procedures, as in example (6.11).

(6.10) 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.* (20*NEJM*3)

(6.11) Recurrent infections ***were treated with*** *an alternative ACT or a different drug combination; a summary is in the appendix (pp 31–32).* (20*LANCET*8)

Next to passive sequences, (VP) + to-clause fragments were also used in this subcategory, mainly to specify the purpose of the study (example 6.12) or experiments performed during the study (example 6.13).

(6.12) We aimed ***to assess the*** *safety and immunogenicity of two formulations (frozen and lyophilised) of this vaccine.* (20*LANCET*2)

(6.13) In addition, ***to determine whether the*** *presence of bilirubin adversely affected the results of our POC anemia test, we systematically added bilirubin to existing blood samples, conducted our assay, and measured our assay's Hgb levels based on spectrophotometry absorbance data*. (14*JCI*3)

Clause-related bundles with human subjects were also used to describe procedures. These bundles, though not very frequent (only 8 and only in the 3-word list), deserve some attention. They all had the first-person plural pronoun *we* as their subjects, agreeing to a certain extent to Hyland & Jiang's (2016) claim that self-mention has increased in scientific writing. Millar and colleagues (2013) suggested that in medical writing, this new trend can be seen as an attempt by authors to avoid the passive whenever possible. Furthermore, they found that self-mention was quite frequent in the Methods sections of the medical research articles where authors are indeed the “doers of the actions” and therefore, can write about “the procedures *they* followed” (p. 402). Example (6.14) and (6.15) illustrate the use of these clause-related bundles with *we* as the subject.

(6.14) For the non-dialysis outpatient visits, ***we examined the*** *total drug fees for dialysis-related drugs, total drug fees for non-dialysis-related drugs and total fees for the antihypertensive drugs*. (15*NEJM2*)

(6.15) Seventh, ***we used a*** *multiple imputation technique to impute missing values in covariates using chained equations with 10 replications.* (20*JAMA*7)

Finally, the observations from the two last functional subcategories, that is, Quantification and Location (in time and place), were consistent with what has been reported in previous studies (e.g., Biber et al., 1999; Cortes, 2004; Hyland, 2008a; Mbodj-Diop, 2015). These functions were served almost exclusively by 3-word and 4-word NP-related and PP-related bundles. As the names imply, bundles in these categories were used to quantify study-related elements (example 6.16) or specify time (example 6.17) or place (example 6.18).

(6.16) Furthermore, no difference was noted in ***the proportion of*** *large-for-gestational-age infants (the primary outcome) or in gestational weight gain, but* ***the proportion of babies*** *4 kg or heavier at birth was lower in the intervention group.* (15*LANCET*1)

(6.17) Glycated hemoglobin level was measured ***at the time of*** *randomization and at 16 weeks by a central laboratory at the University of Minnesota Advanced Research and Diagnostic Laboratory.* (20*NEJM*10)

(6.18) The Jackson Laboratory, stock no. 002287) (IFN-γ deficient) were bred in-house ***at the University of*** *Georgia Animal Facility.* (20*SCIENCE*3)

Coming now to the second main functional category served by the MRAC bundles, the Text-Oriented category, Figure 6.3 shows the distribution of the subfunctions served by the bundles identified in this category. Overall, text-oriented bundles were primarily 3-words and 4-words. The lists of 5-word and 6-word text-oriented bundles included only 14 and 5 sequences, respectively. Resultative signals were the most frequently served function, accounting for 32.17% of 3-word text-oriented bundles and 35.29% of 4-words. They were followed by Framing signals, which represented 23.78% of 3-word text-oriented bundles and 23.57% of 4-words. Eight of the fourteen 5-words (e.g., *on the basis of the, on the basis of these, in the context of the, at the discretion of the*), as well as 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.

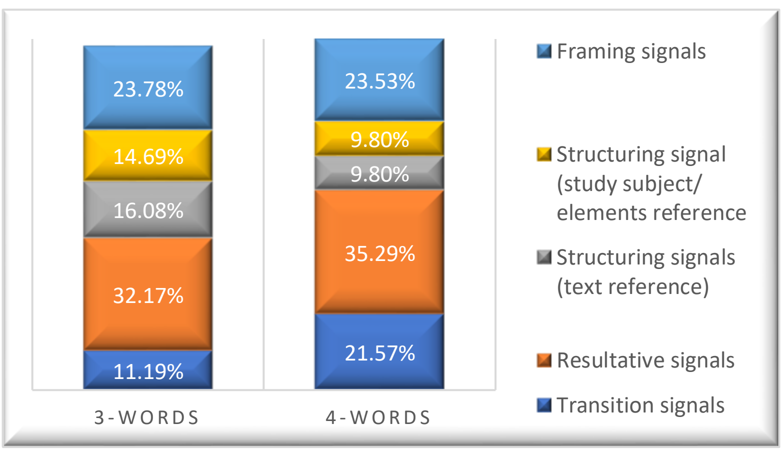


Figure 6.3. Distribution of Text-Oriented Functional Subcategories Served by MRAC Bundles

Structuring signals came in third position and were used almost in the same proportion as framing signals. Together, Structuring (text refence) and Structuring (study subject/elements reference) accounted for 30.77% of 3-word text-related bundles and 19.6% of 4-words. Additionally, One of the fourteen 5-word, three of the five 6-word text-oriented bundles, and the only 7-word bundles were structuring signals. Finally, transition signals were less frequent than the three other functional subcategories, but they still accounted for 21.57% of 4-word text-oriented bundles and 11.19% of 3-words. There were also two 5-word transition signals.

In the most frequent subcategory, that is, the Resultative signals, both 3-word and 4-word bundles were predominantly clause-related sequences in the form of (VP) + *that*-fragment (e.g., *we observed that, suggesting that the, these results demonstrate that*, *our data indicate that*). Additionally, 3-word resultative signals also included a few noun/preposition-related (e.g., *effect on the, as a result, in response to*) and passive-related bundles (e.g., *significantly associated with*, *related to the*). Resultative signal bundles in the MRAC, similar to those previously identified elsewhere in academic writing, were used to indicate causative relations between elements being discussed or presented (example 6.19) and authors' interpretations of the results of their studies (example 6.20).

(6.19) *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 […].* (15*LANCET*5)

(6.20) *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).* (20*JCI*2)

In the next most frequently served functional subcategory, framing 3-word bundles were primarily NPs with other post modifier fragments (e.g., *patients in the*, *t cells in*, *patients with a*) and PPs both with and without embedded *of*-fragments (e.g., *in terms of, in patients who, with respect to*). On the other hand, the 4-word and the two 6-word framing bundles were exclusively PPs, often with embedded *of*-fragments (e.g., *on the basis of, with the exception of, in the* *presence or absence of*). Six out of the eight 5-word framing bundles were also PPs with embedded *of*-fragments (e.g., *at the discretion of the, in the context of the*). This is consistent with previous research that found framing signals as being primarily "preposition + of-structure" (Hyland & Jiang, 2018, p. 18). However, in line with Cortes's (in press) study of 3-word bundles, the 3-word list indicated that postmodified NPs also frequently serve as framing signals. As has been shown in previous research (e.g., Cortes, 2004; Hyland, 2008a), framing signals were used to specify limits of topics and/or elements being discussed, as in examples (6.21) and (6.22) below.

(6.21) ***With the exception of*** *fatal coronary and cerebrovascular events, none of the prespecified secondary end points were reduced significantly in the low-doseaspirin group.* (8*JAMA*2)

(6.22) *In the AASK study, the primary outcome occurred in 58.1% of the* ***patients in the*** *APOL1 high-risk group and in 36.6% of those in the APOL1 low-risk group (hazard ratio in the high-risk group, 1.88; P<0.001)h.* (13*NEJM*4).

In the Structuring Signal subcategory, 4-word bundles were almost exclusively prepositional phrases (e.g., *in the present study, in the supplementary appendix*), with only one passive *(are shown in table*). On the other hand, on the 3-word list, there were relatively similar proportions of prepositional phrases (e.g., *in our study, in the appendix*), passive-related sequences (e.g., *are provided in, are shown in*), and clausal *as*-fragment bundles (e.g., *as previously described, as described above, as shown in*). This again, is a case in point with regards to Cortes's (in press) suggestion that 3-word bundles have the potential of providing "new insights to the formulaic profile of academic texts". Indeed, had the present study focused only on 4-word bundles, as has been the case in most studies of lexical bundles, the findings would likely indicate preposition-related bundles as bundles most frequently used as structuring signals in medical research articles, which obviously is not the case, based on the findings described above. That being said, the use of structuring signals in the MRAC did not differ from what has already been described in the literature, that is, to “provide readers with a cognitive roadmap” (Hyland & Jiang, 2018, p. 18). In the MRAC, they were used to direct readers to different sections of the text (example 6.23) and/or elements within the text (example 6.24) or outside of it (example 6.25).

(6.23) Samples were incubated at 37°C for 3 days, after which plaques were visualized by immunoperoxidase staining ***as described above****, and a 50% plaque-reduction neutralization titer was calculated.* (20*Science*4)

(6.24) The demographic and baseline characteristics of enrolled infants, including the results of laboratory analyses, ***are shown in Table*** *1.* (15*NEJM*1)

*(6.25) The latter contained 6, 10, 10, and 10 specimens from Global Initiative for Chronic Obstructive Lung Disease (GOLD) stage 1, 2, 3, and 4 individuals, respectively (****supplemental material available online with this article****).* (15*JCI*3)

Finally, 4-word bundles in the Transition signals subcategory were primarily preposition-related sequences where 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 has been shown in previous studies, bundles serving as transition signals were used to add information (example 6.26) or to compare/contrast elements of the study (example 6.27).

(6.26) Safety was analyzed in the populations of patients with RET-altered medullary thyroid cancer and of those with nonmedullary thyroid cancer as defined above, ***as***

well as in the *overall cohort of 531 patients who received selpercatinib by June 17, 2019. The data cutoff date was December 16, 2019.* (20*NEJM*3)

(6.27) ***Compared with the*** *low-tryptophan diet, the enriched tryptophan diet did not affect alpha diversity [...].* (20*Science*8)

The findings reported and discussed in this section partly answer RQ 2c (How are lexical bundles used in the two registers? How do they compare to bundles previously identified in academic prose? Are there any similarities and/or differences in terms of length, structure, and function?). We have seen that in accordance with findings of previous studies of bundles in academic prose, NP-related and PP-related bundles were predominant in the MRAC. However, passive bundles were also found to be frequent in the corpus. As described above, these passive bundles serve various important functions, suggesting that despite their reported decline in scientific writing, they still are staples of medical research article writing. The structural analysis also revealed a relatively frequent use of clausal bundles in the form of active clause fragments, a structure that has been reported to have increased in scientific writing. The findings of the functional analysis did not fundamentally differ from what has been already reported in previous studies of bundles in academic writing. However, as a by-product, it underscored the importance of investigating bundles of various lengths for a more comprehensive description of the formulaic profile of the register under scrutiny.

6.2 Lexical Bundles in Medical Case Reports

In this section, I report the findings of the structural and functional analyses of lexical bundles identified in the MCRC. I classified MCRC bundles structurally using the same framework I used for MRAC bundles. For the functional classification, I used the framework introduced in the Methodology chapter in section 3.5 and shown in Table 3.11.

6.2.1 Structures of Lexical Bundles in the MCRC

In this section, I report the findings of the structural and functional analyses of lexical bundles identified in the MCRC. I classified MCRC bundles structurally using the same framework I used for MRAC bundles. For the functional classification, I used the framework introduced in the Methodology chapter in section 3.5 and shown in Table 3.11.

Table 6.5. Distribution of MCRC Bundles across the Major Structural Categories

Main Categories	**Subcategories**	**3-words**	**4-words**	**5-words**	**6-words**	**7-words**	**8-words**
Noun/Preposition-related	NP with of-phrase fragment	**71**	**20**	2	2		
	NP with other post modifier fragment	11	1	2			
	PP with embedded of-fragment	1	**30**	**11**	3		
	Other prepositional fragments	**70**	**26**	4			
	Comparative expressions						
Total Noun/ Preposition-related		*153*	*77*	*19*	*5*	*0*	*0*
Verb phrase-related	Passive verb	**47**	**30**	**14**	6		
	copular be	**13**	**10**	1	2		
	imperative						
	modals	**10**					
Total VP-related		*70*	*40*	*15*	*8*	*0*	*0*
adjective/ Adverb-related	AdjP /AdvP + prep + …	**10**		4			
	Superlatives	1					
Total Adj/Adv- related		*11*	*0*	*4*	*0*	*0*	*0*
Clause-related	anticipatory it	2	2	2			
	abstract subject	7	4	3	6	1	1
	human subject	3	**13**	**11**	**9**		
	external subject	4	4	1	2		
	as-fragments						
	if-fragments						
	there fragments	7	1	4	1		
	(NP) + wh-fragments	1	0	0			
	(VP) + that-fragments	1	0	0			
	(VP) + to-fragments	5	1	2			
Total clause-related		*30*	*25*	*23*	*18*	*1*	*1*
Function words only		6					

Noun phrases were still frequent – though not as frequent as in the MRAC – and were unsurprisingly, for the most part, NPs with *of*-phrase fragments (e.g., *the onset of, removal of the, ct scan of the, an increased risk of*), as this structure has been reported to be characteristic of academic writing (Biber et al, 1999). In the MCRC, NPs with of-fragments occurred primarily in the 3-word and 4-word lists, where they accounted for 46.4% (71 out of 153) of 3-word bundles in the Noun/Preposition- related category, and 25.07% (20 out of 77) of 4-words. NPs with other postmodifiers were rare in the MCRC, with 11 on the 3-word list being their highest occurrence in one list.

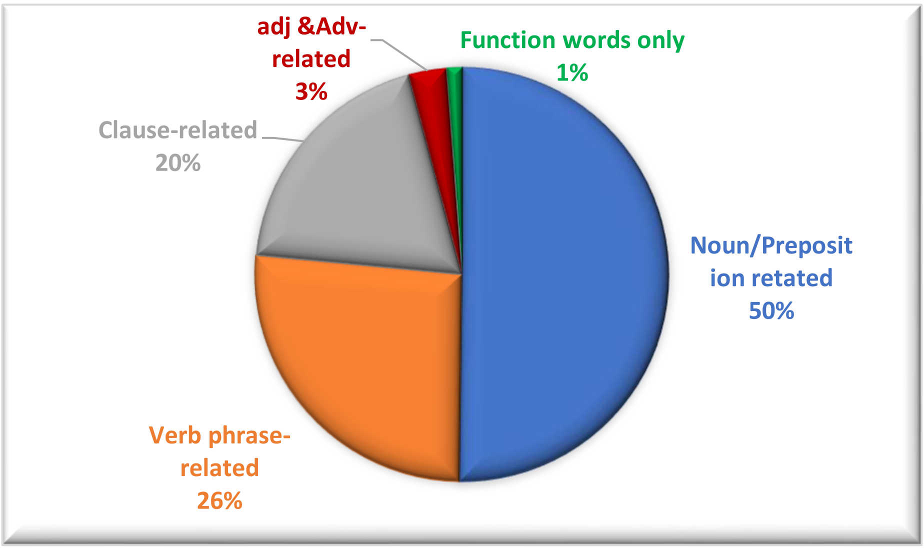


Figure 6.4. Distribution of Major Structural Categories in the MCRC

In the Verb-phrase-related category, passive-related sequences (e.g., *was diagnosed as having, she was started on, was admitted to our* hospital) were by far the most frequent structure, accounting for 67.14% (47 out of 70) 3-word VP-related bundles, 75% (30 out of 40) of 4-words, and 93.33 (14 out of 15) of 5-word bundles in this category. Six of the eight 6-word VP-related bundles were also passive sequences. Passive bundles appear then to be used even more frequently in medical case reports than in medical research articles, despite journal recommendations discussed in the previous sections. The impact of such recommendations on medical writing in the long term may be worth investigating. In the meantime, as suggested above, novice and L2-English medical writers would certainly benefit from instruction of when avoidance of passive expressions is suitable.

The two other VP-related structures were copular *be* (e.g., *is extremely rare, is a rare disease, is one of the most, is the first, is a rare*) and modals (*can lead to, may not be, can present with*). Modals occurred exclusively in the 3-word list where they represented 14.28% (10 out of 70) of VP-related bundles in that list. Copular *be* accounted for 18.57% (13 out of 70) of 3-word VP-related bundles and 25% (10 out of 40) of 4-words. One out of the fifteen 5-words and two of the eight 6-word VP-related bundles were also copular *be.* This structure will be further discussed in the functional analysis, as it appears to be very useful to authors of case reports who, as was discussed in the previous chapter, are recommended to expressly indicate the importance of the cases they present.

Finally, the clause-related category was dominated by the earlier discussed clausal bundles with human subject, abstract subject, or external subject (e.g., *our patient underwent a, we report a case of, this case highlights, presented to the emergency department*). Together, these three subgroups account for 84% (21 out of 25) of 4-word clause-related bundles, 65.21% (15 out of 23) of 5-words, and 46.66% (14 out of 30) of 3-words. The only 7-word and 8-word bundles in the MCRC were also in this group. Table 6.5 clearly shows that 'human subject' was the most frequently used structure of the three. Both passive-related bundles and these active clause fragments occurred at higher proportions in the MCRC than in the MRAC, despite the latter being a larger corpus. This finding suggests that bundles in each of these structural categories may have specific functions to serve in medical case reports. The findings of the functional analysis of these and other MCRC bundles are reported the in the section below.

6.2.2 Functions of Lexical Bundles in the MCRC

As mentioned above, I used the framework in Table 3.11 for the functional classification of MCRC bundles. Table 6.6 shows the proportions of MCRC bundles in each major functional category. The most frequently served functional category was Case-related, with the majority of bundles of all lengths occurring in that category. As shown in Table 6.6, 59.26% of 3-words (e.g., *we report a, we describe a, the nature of, is known to*), 54.95% of 4-words (e.g., *the case of a, is thought to be, this case highlights the, has been reported in*), and 70.49% of 5-words (e.g., *with a medical history of, she did not have any, is the most common site*) were case-related bundles. Twenty-five of the Thirty-one 6-words (e.g., *is the most common cause of, the patient was admitted to the, his medical history was significant for*), as well as the single sequences of 7 and 8 words (*this is the first reported case of, the key to the correct diagnosis is the*) in the MCRC were also case-related bundles.

Diagnosis & Intervention-related category bundles were the next most frequent in the MCRC. They accounted for 30.99% of 4-words (e.g., *for the diagnosis of, she was treated with, computed tomography of the*), 27.41% of 3-words (e.g., *differential diagnosis of, was transferred to, on physical examination*), and 19.67% of 5-words (e.g., *the patient was started on, the decision was made to, with a blood pressure of*). Five of the thirty-one 6-word bundles were also listed in this category (e.g., *a white blood cell count of, the patient was found to have*).

Discourse organizers were less frequent than case-related and diagnosis/intervention-related bundles, but they still accounted for 14.44% of 3-words, 13.38% of 4-words, and 11.48% of 5-words. Bundles expressing stance and engagement as well 3-word bundles in the grammatical only subcategory were very rare in the MCRC. The two most frequently served functional categories (Case-related and Diagnosis/Intervention-related) were further analyzed to describe the subfunctions served by MCRC bundles in each of these new functional categories. Additionally, Discourse Organizers were further analyzed to see whether they differed in any way from what has already been reported on bundles serving these functions.

Table 6.6. Distribution of the MCRC Bundles in the Major Functional Categories

	MCRC Bundles: Raw # (%)					
Main Functions	**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%

Figure 6.5 shows the distribution of the subfunctions served by MCRC bundles in the Case-related category. The vast majority of bundles in this category was used for description. Together, bundles used to describe the medical condition (e.g., *the pathogenesis of, the most common cause of, has been reported to*) and case subject (e.g., *a man in his, woman with a history of, she had no, our patient was*) made up for 60.01% of 3-word case-related bundles, 67.94% of 4-words, and 79.07% of 5-words. Twenty-two of the twenty-five 6-word case-related bundles (e.g., *there was no family history of, is one of the most common, presented to the emergency department with*) as well as the single bundles in the 7-word and 8-word lists (*this is the first reported case of, the key to the correct diagnosis is the*) were also used for description. Overall, bundles used for the description of the medical condition were more frequent across all bundle lists except for the 5-word list where description of the subject (44.19%) outnumbered description of the medical condition (34.88%).

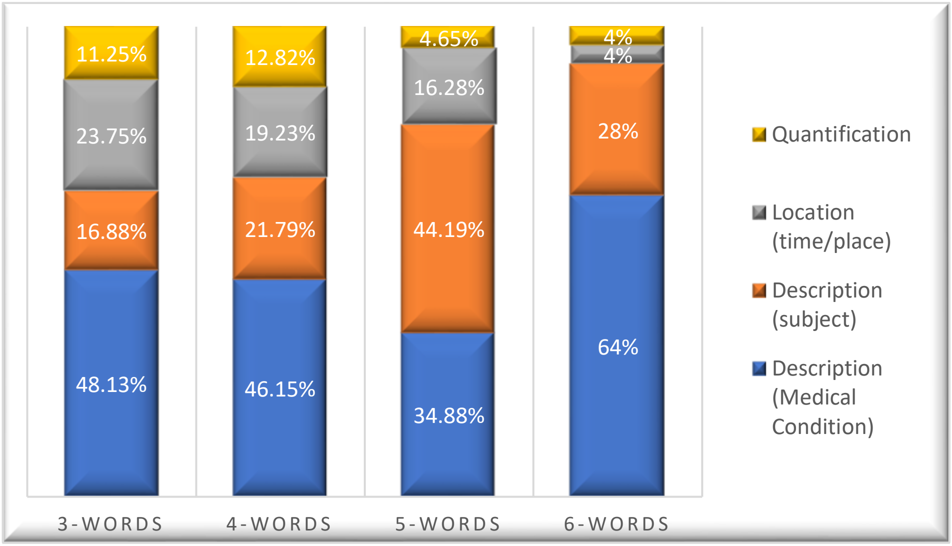


Figure 6.5. Distribution of Case-related Subcategories Served by MCRC Bundles

In the Description (medical condition) group, each bundle list appeared to be dominated by one or two structures. Three-word bundles were primarily PPs with embedded *of*-phrases and passive-related sequences (e.g., *the cause of, shortness of breath, is defined as, is caused by, is known to*), whereas 4-words were predominantly passive-related sequences and copular *be* (*have been associated with, is thought to be, is more common in, are the most common, is a very rare*). Five-word bundles were predominantly passive-related sequences (e.g., *has been reported to be, been shown in the literature*) and clause-related with human subjects (e.g., *we report a case of, we present a case of*), and 6-words were almost exclusively clause-related bundles with human subjects, all in the form of first-person plural pronoun (e.g., *we report a rare case of, we present a rare case of*). The single 7-word and 8-word bundles in the MCRC were also clause-related, but with abstract subjects (*this is the first reported case of, the key to the correct diagnosis is*).

These clausal bundles with *we* as their subjects appear to be ready-made fragments that case report authors use primarily to introduce the medical condition to be described. The main lexical verbs in these bundles are almost always *present*, *report*, or *describe* (examples 6.28 – 6.30). Note how these bundles are all sentence initial. In the medical case reports, they often occur in the very first section of the text after the abstract, regardless of headings. Sometimes, they occur in the abstract itself.

(6.28) ***We present a rare case of*** *bilateral adrenal tumors in which the left adrenocortical tumor produced cortisol and the right adrenocortical tumor secreted aldosterone, and we review literature on PA concurrent with SCS.* (20*JMCR*47)

(6.29) ***We describe a case of*** *haemophagocytic lymphohistiocytosis (HLH) secondary to disseminated histoplasmosis, which was treated with chemotherapy in addition to standard antifungal therapy.* (20*BMJ*31)

(6.30) ***We report a rare case of*** *hypertriglyceridemia which was diagnosed at 24 days after birth.* (21*JMCR*5)

NPs with embedded *of*-phrases were primarily used to provide details on the medical condition (example 6.31). Both passive-related bundles and copular *be* were also used for providing details, but passive-related sequences served also to include previous research findings in the description of the medical condition or elements related to it (example 6.32). Copular *be,* as mentioned above, was also used by authors to highlight the importance of the case report and/or the uniqueness of the case being presented. In such cases, they always include an intensifier *(*e.g., *very, extremely)*, as in example (6.33)

(6.31) ***The onset of*** *symptoms was marked by* ***the appearance of*** *a small nodule in her left breast, and an evolution marked by a rapid increase in the volume of the tumor, which motivated traditional herbal treatments of unknown nature.* (20*JMCR*92)

(6.32) Staphylococcus aureus ***has been reported to*** *cause Lemierre syndrome. Chanin et al. [6] noted 11 cases from 2002 to 2011.* (19*OMCR*160)

(6.33) Congenital hyporhinia ***is an extremely rare*** *deficiency of mid-facial embryogenesis characterised by the absence or hypoplasia of the external structures of the nose.* (20*BMJ*26)

As for bundles used to describe the subjects of the case reports, they were in their vast majority clausal sequences across all bundle lists (from 3 to 6 words) mostly with human subject (e.g., *the patient had*, *our patient had a*, *she had no history of*, *the patient had a history of*). As can be noted from these examples, the subject in these clause-related bundles were almost always *the/our patient* or the third-person singular pronouns *he/she*, and the main verb was almost exclusively the primary verb *have.* Next to these clause-related bundles with human subjects, postmodified NPs (with or without *of*-phrases) were also used in the description of the case subject (e.g., *family history of*, *a man in his*, *woman with a history of*). All these bundles were used to provide details about the case subjects, whether to describe their past and/or present health status (example 6.34), or just to provide some baseline characteristics like age, sex, race, etc. (examples 6.35 and 6.36).

(6.34) 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.* (19*JAMA*149)

(6.35) ***The patient is*** *a 59-year-old man with previously known hypertension.* (19*OMCR*151)

(6.36) ***A woman in her*** *50s presented to the emergency department in a comatose condition.* (20*JAMA*97)

The two other case-related subfunctions served by the MCRC were Location (in time and place) and Quantification. The use of bundles in these categories was similar to those described for MRAC bundles and indeed, for bundles in academic prose, in general. In the MCRC, quantification bundles were primarily NPs with embedded *of*-fragments (e.g., *a wide range of, at a dose of, a total of, the majority of*), and location bundles were predominantly PPs with or without embedded *of*-phrases. Similar to bundles in the MRAC, the MCRC bundles in these two categories were used to quantify elements related to the case being presented (examples 6.37), or to specify time (example 6.38) or place (example 6.39).

(6.37) 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.* (19*JAMA*141)

(6.38) 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.* (19*OMCR*149).

(6.39) She had brisk reflexes more ***on the left side of*** *her body.* (20*IMCRJ*49)

I now turn to the second major functional category served by the MCRC bundles, the Diagnosis/Intervention-related category. Figure 6.6 shows the different subfunctions of bundles in this category. Bundles used to report results of diagnosis and intervention procedures were the most frequent ones and represented over 40% of the 3-word, 4-word, and 5-word lists. Two of the five 6-word bundles in this category were also used to report results. The next most frequently served subfunction was Procedure, closely followed in third position by the Decisions/Outcome subcategory. All three subcategories were analyzed, regardless of their frequency as they represent new functions different from those commonly described in academic writing, in general.

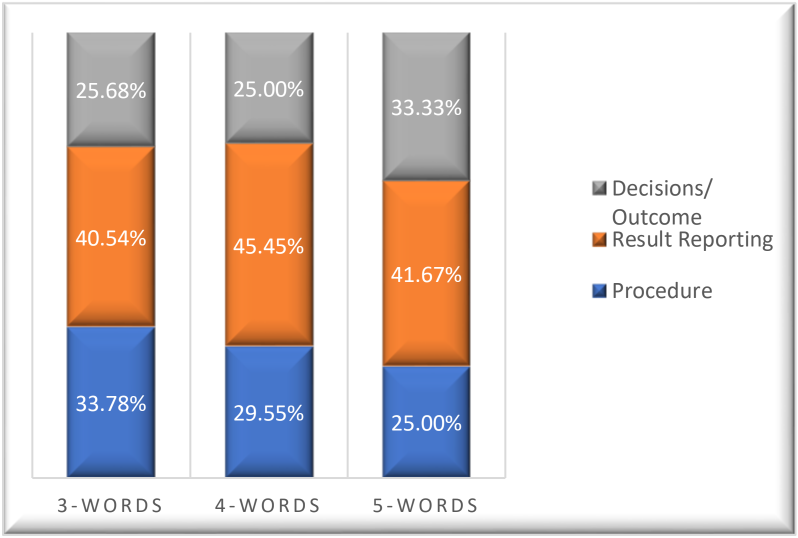


Figure 6.6. Distribution of Diagnosis/Intervention Subcategories Served by MCRC Bundles

Bundles in the Result Reporting subcategory were mostly clause-related. Of the four main bundle structures observed in this subcategory, three were clause-related; namely clausal bundles with abstract subject (*his blood pressure was*, *white blood cell count was, vital signs were*), external subject (e.g., *did not reveal any, showed no evidence of, confirmed the diagnosis of*), and *there*-fragments (e.g., *there was no evidence of, and there was no, there were no signs of*). The fourth structure frequently in this subcategory was 'Passive' (*was diagnosed as having, was found to have, the patient was diagnosed with*). These bundles were used to report (a) results of physical examination, a function served mainly by clause-related bundles with abstract subject (example 6.40); (b) results of tests and other diagnosis procedures, primarily expressed by clause- related bundles with external subject and *there*-fragments (examples 6.41 & 6.42); and (c), to report final diagnosis, mostly with passive bundles (example 6.43).

(6.40) On admission ***her blood pressure was*** *140/80 mmHg, breathing 18 breaths per minute, pulse 80 beats per minute, and temperature 36.5 °C.* (20JMCR40)

(6.41) *Screening with an ultrasonographic examination of his testicles and a computed tomographic scan of his chest, abdomen, and pelvis* ***showed no evidence of*** *tumors.* (14*JAMA*58)

(6.42) *Eye examination revealed signs of conjunctivitis, scleromalacia and mild blepharitis; but* ***there were no signs of*** *corneal injury.* (20*OMCR*60)

(6.43) *The patient was exhibiting symptoms including fever, cough and shortness of breath and* ***was found to have*** *acute pulmonary embolism.* (20*BMJ*16)

The Procedure subcategory was served primarily by 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 6.44) while PPs and clause-related bundles with human subjects were mostly used for intervention procedures (examples 6.45 & 6.46).

(6.44) *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).* (19*OMCR*165)

(6.45) ***The patient was given*** *prednisone, 1 mg/kg/d, with dose reduction across several weeks.*

(6.46) *The first line therapy consists of the use of steroids (prednisone 40 mg/daily tapered slowly)* ***in combination with*** *tamoxifen 10 mg twice daily.* (21*IMCRJ*31)

Finally, in the Decision/Outcome-related subcategory, clause-related bundles with human subjects were also the most frequent structure; with the subject almost always being *the patient* (e.g., *the patient was discharged, the patient was referred to*). The other structure frequently used in this functional subcategory was clause-related with (VP) + to-fragments (e.g., *we decided to, to confirm the diagnosis, to rule out*). Clause-related bundles with human subject were used primarily to indicate decisions made at different stages of addressing the case (example 6.47) and to report the outcome of the case (example 6.48). There were also a few instances where the outcome was expressed by a NP with a postmodifier *of*-phrase, as in example (6.49).

(6.47) *On examination, there was generalised abdominal distension with tenderness over the left iliac fossa, with no palpable masses or peritonism.* ***The patient was referred to*** *general surgery for further assessment*. (21*BMJ*66)

(6.48) *Recovery was uneventful* ***and the patient was discharged*** *3 days postoperation.* (21*BMJ*76)

(6.49) *Treatment led to* ***a complete resolution of*** *his symptoms.* (14*JAMA*60)

Clause-related bundles with *to*-fragments in this subcategory were used exclusively to provide rationales for decisions made, as can be seen in examples (6.50) and (6.51). In both examples, the decisions being justified are <u>underlined.</u>

(6.50) *However, due to the site in the nape of the neck, the extensive length of the lesion and its deep induration, the treating physician <u>decided to do a biopsy first</u>* ***to confirm the diagnosis*** *before proceeding to excision.*

(6.51) *<u>Extensive workup was done</u>* ***to rule out*** *coexisting immunological disorders, especially antiphospholipid syndrome.*

The last main category in this functional analysis is Discourse Organizers. The subfunctions in this category are the same as those listed in the Text-oriented category for bundles in the research articles. Figure 6.7 shows the distribution of these subcategories in the MCRC. Discourse organizers in the MCRC were primarily 3-word and 4-word bundles. There were only seven 5-word and one 6-word discourse organizers and only the 3-word list included bundles serving all five identified functional subcategories. There were much less bundles functioning as discourse organizers in the MCRC than in the MRAC. The most frequent function was Framing and bundles in this subcategory totaled only 21 in the entire MCRC. This probably is because of the difference in text length explained in the Situational Analysis chapter, Chapter 4. Indeed, with the very limited word count of MCRs (M = 1428.8, SD = 624), authors may not need to make frequent use of discourse organizers. It may also be the case that case report authors resort to other devices to organize their texts. For example, the subdivision of the Case Presentation section (discussed in section 4.3.5.3) into several subsections and the frequent use of bullet points can reduce the need for discourse organizers.

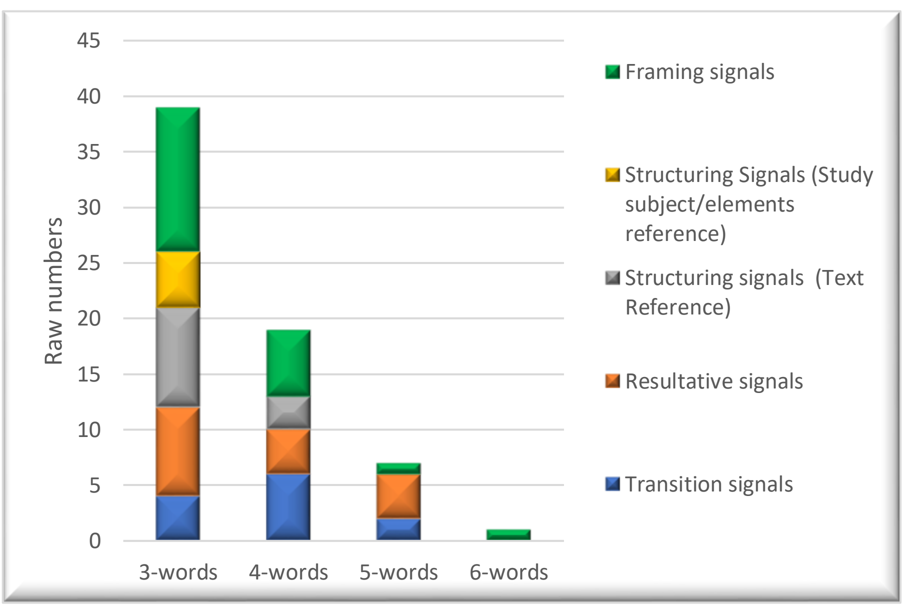


Figure 6.7. Distribution of Discourse Organizer Subcategories Served by MCRC Bundles

Discourse organizers in the MCRC served functions similar to those of bundles in the MRAC and in academic writing, in general. Examples (6.52) and (6.53) below illustrate the use of framing and resultative signals, respectively; The two most frequently served discourse organizer functions in the MCRC.

(6.52) *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.* (19*OMCR*160)

(6.53) ***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.* (21*IMCRJ*35)

This section has brought some supplementary information that helps answer RQ 2c. The structural analysis has shown that, in line with what has been reported in studies of bundles in academic prose and in the previous section on bundles in medical research articles, MCRC bundles were predominantly Noun/Preposition related. However, VP-related and clause-related bundles were also found to be very frequent in medical case reports. The functional analysis revealed that some bundles in these two structural categories (copular *be* and passive, for VP-related, and clause-related with human subject, abstract subject, and external subject) frequently served functions that appear to be specific to case reports, namely, Result Reporting, Description (subject), and Decision/Outcome-related. This already suggests some interesting differences between bundles in the MRAC and in the MCRC, and together with the comparison of shared bundles in the next section, they will complete the answer to RQ2c

6.3 Bundles Shared across the two Registers

To complete the comparison of the use of bundles in the two registers, shared bundles were identified and functionally classified. The two corpora shared just a few 3-word, 4-word and 5-word bundles. The proportions of shared bundles in each corpus are shown in Table 6.7., and Table 6.8 shows the list of shared bundles. The analysis of shared bundles did not reveal any particular differences. NP- (**bolded** in table 6.8) and PPs (underlined) were the most frequent structures of the shared bundles. This is not surprising as these structures were not only the most frequent in the MRAC and the MCRC, but as mentioned above, they have been shown to be predominant in academic writing, in general. It was also not surprising that passive-related bundles (grey-shaded in Table 6.8) were relatively frequent in the list of shared bundles as they were found to be frequent in both corpora.

Table 6.7. 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%

The functional classification of these shared bundles (Table 6.9) revealed that the functions served by shared bundles, as well as many of the bundle types in those functional categories, do not appear to be specific to any of the two registers, or even to the medical field, as they have already been identified in previous studies of lexical bundles in other academic registers and disciplines (e.g., Biber et al. 2004, Cortes, 2004; Hyland 2008a).

Table 6.8. List of Bundles Shared by the MRAC and MCRC

3-WORDS	***the importance of***	***a number of***	***a diagnosis of***	*it is possible that*
in patients with	***the lack of***	*are shown in*	***the end of***	*in the pathogenesis of*
as well as	*to our knowledge*	***the onset of***	*as a result*	***an increased risk of***
the number of	***risk factors for***	*most of the*	***the time of***	*at the same time*
the use of	*in combination with*	*found in the*	***written informed consent***	*in the setting of*
of patients with	***a combination of***	*are associated with*	***the treatment of***	*in the absence of*
the risk of	***the level of***	*found to be*	*consistent with a*	*for the treatment of*
according to the	***changes in the***	*due to the*	***quality of life***	*for a total of*
a total of	*associated with the*	*did not show*		*has been shown to*
was associated with	***the results of***	*related to the*	**4-WORDS**	*to the development of*
the presence of	***the absence of***	*could not be*	*at the time of*	***a wide range of***
based on the	*in the same*	***the combination of***	*with a history of*	*at a dose of*
the development of	*during the first*	***body mass index***	*in the context of*	
because of the	***the need for***	*it has been*	*as a result of*	**5-WORDS**
we did not	*has not been*	***a history of***	*on the other hand*	*at the end of the*
years of age	*associated with a*	*to be a*	*as well as the*	*is one of the most*
is associated with	*in which the*	***the course of***	*on the basis of*	*at the time of the*
there was a	***increased risk of***	*may not be*	*in the presence of*	*has been shown to be*
there was no	***part of the***	*of the disease*	*to be associated with*	*as well as in the*
of the patients	*and in the*	*in patients who*	*in the case of*	
with or without	*such as the*	***a series of***	*in addition to the*	
for patients with	***coronary artery disease***	***no evidence of***	*has been associated with*	
an increase in	*there were no*	*to evaluate the*	*with the use of*	
the most common	***the majority of***	*be due to*	*in the united states*	
in the first	*in the past*	*has been reported*	*by the presence of*	
in order to	*there is a*	*considered to be*	*this is the first*	
the pathogenesis of	***patients with a***	*to have a*	*in view of the*	

Other than bundles including field-related terms (e.g., of the patient, risk factors for, in the pathogenesis of) the bundles and the functions they serve are the for the most part, the same as those previously described in the literature. For example, whether a “time marker” in Cortes’s (2004) history and biology corpora, or a member of the “Location (time/place)” subcategory in Hyland’s (2008a) biology and engineering corpora or of the two medical corpora in the present study, the 4-word bundle at the time of serves the same function in all three studies: to specify time. For comparison’s sake, 17 of the 29 shared 4-word bundles occurred in the 4-word lists in Cortes (2004) and/or Hyland (2008a). These included familiar expressions such as at the time of, in the context of, as a result of, on the other hand, as well as the, on the basis of, in the presence of, to name but a few.

Table 6.9. Functions Served by Shared Bundles

	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	
Totals	*88*	*29*	*5*

On the other hand, the other bundles that include field-related terms may be considered as shared across the medical field. The analysis in context of these apparently field-specific bundles did not reveal any specific differences in their use in the two registers. This is illustrated in the two sets of examples (6.54) and (6.55) below. Example (6.54) shows the 4-word bundle *in the* *pathogenesis of* used as a Description bundle, and (6.55) shows the 3-word sequence *for patients with* used as a framing signal in both corpora.

(6.54) *Our findings provide additional support that early-life events play a critical role* ***in the pathogenesis of*** *asthma.* (MRAC_08*NEJM*3)

Factors such as obesity, hypertension, smoking, hypercholesterolemia, and diabetes play a leading role ***in the pathogenesis of*** *MI in non-hemophilic elderly men.* (MCRC_21*IMCRJ*3)

(6.55) *These results question the view that implementing SDM tools for anticoagulant treatment can improve care* ***for patients with*** *AF.* (MRAC_20*JAMA*10)

NPV is suitable ***for patients with*** *abnormal facial morphologies, excessive oropharyngeal secretions as well as patients who experience anxiety [...].* (MCRC_20*OMCR*48)

However, what needs to be highlighted is perhaps not what was shared, but rather what was *not* shared. The total number of shared bundles (122) represented less than the quarter of bundles identified in each corpus, meaning that at least 75% of bundles identified in each corpus were used only in the register represented by the corpus. This suggests that even though bundles identified in the two corpora showed similarities in term of their structures, they were in their vast majority distinct, in terms of their types. This difference was to some extend explained earlier in this chapter through the structural and functional classification of different lexical bundles identified in each corpus. Therefore, particular MCRC and MRAC bundles appear to be specific to medical case reports and medical research articles respectively. The same argument may hold regarding the similar trend noted in the structural profiles of the MCRC and MRAC bundles on the one hand, and bundles previously identified in academic prose, on the other hand. Indeed, the similarities may be limited to just their structural profiles and perhaps the few shared functions across disciplines discussed above.

6.4 Conclusion

In this chapter, I have tried to answer RQ 2c that asked about (1) the structures and functions of lexical bundles in MRAs and MCRs, (2) the similarities and differences between the two registers, and (3) the similarities and differences between bundles previously described in academic writing and the MRAC and MCRC bundles. The findings presented in the different sections of this chapter have shown that overall, the structural profiles of bundles in both registers are consistent with previous descriptions of bundles in academic writing in general, as bundles in both corpora were predominantly Noun/Preposition-related. VP-related and Cause-related bundles were found to be the second and third most frequent structural categories, respectively.

However, more in-depth analysis revealed that VP-related and Clause-related bundles were found to be used more frequently in the case reports than in the research articles. This was later explained by the findings of the functional analysis. Indeed, the functional analysis revealed that in addition to some shared functional subcategories like Description, Quantification, Procedure, and Location in time and place, MCRC bundles also served some functions that were specific to case reports, namely, Result reporting, Description (subject), and Decision/Outcome-related. These functions were served primarily by copular *be* and passive-related (VP-related), as well as clause-related bundles with human subjects, abstract subjects, and external subject. The analysis of shared bundles brought even more evidence of the differences between the two registers with less than 25% of bundles in one register occurring in the other register. This led to the conclusion that most bundles identified in each corpus were specific to the register represented by that corpus. By extension, it can be argued that there may be an overall structural profile for academic prose, in general, but not a one-size-fits-all description of the use of bundles across disciplines, or even within a same discipline.

7 CHAPTER 7. FINDINGS AND DISCUSSIONS 3: FRAMES IN MEDICAL RESEARCH ARTICLES AND CASE REPORTS

This chapter reports on and discusses the use of lexical frames, i.e., sequences of three or more words with internal variable slots, identified in the MRAC and the MCRC. As is often the case in studies of frames, the initial amount of data may cause challenges for analysis and interpretation (Stubb, 2007). The initial lists of frames identified in the two corpora included tens of thousands of sequences (22,784 and 36,837 for the MRAC and the MCRC, respectively), and these were only frames with internal slots. The lists generated by *KfNgram* (Fletcher, 2007) with both internal and external slots were much longer with hundreds of thousands of frames per list. Table 7.1 shows the number of frames in each initial list of candidates from the two corpora.

Table 7.1. Raw Numbers Initially Identified Frames in the two Corpora

Frames	**MRAC**	**MCRC**
3-grams	15,031	20,052
4-grams	3,662	12,079
5-grams	3,261	3,676
6-grams	830	1,020
Totals	**22,784**	**36,837**

As explained in section 3.6. in the Methodology chapter, I set thresholds for identified frames of various lengths to make the data manageable. The final numbers of analyzed frames are shown in Table 7.2. Even though the focus in the present study is not on a direct comparison of frequencies of formulaic sequences in the two corpora, it is worth noting the differences observed in the number of frames from each corpus. As clearly shown in both Table 7.1. and Table 7.2., frames in the MCRC outnumbered those in the MRAC despite the latter being a moderately larger corpus. This certainly warrant further investigation in future research.

Table 7.2. Raw Numbers of Analyzed Frames in the Two Corpora

Frames	**MRAC**	**MCRC**
3-grams	881	769
4-grams	306	740
5-grams	228	212
6-grams	131	196
Totals	**1546**	**1917**

The most salient difference was observed in the lists of 4-gram frames. One explanation could be the difference in statistical techniques used by MRA and MCR authors, previously described in the situational analysis of the two registers in chapter 4, section 4.3.6. Indeed, the initial MRAC 4-gram list of frames included a lot of sequences that consisted mostly of numbers (e.g., *95 ci * 1, 0 * p 0, 0 * 95 ci*). As explained in the methodology chapter, these types of statistical formulae were excluded from the list of frames to be analyzed. A second explanation was found in the structural and functional analyses of MCRC frames and will be discussed in sections 7.2. and 7.3.

Coming back to the purpose of the present chapter, i.e., the description of frames in the two registers, I analyzed all lists of sequences shown on Table 7.2, to determine the structures, variability, and predictability of frames in the two registers. The descriptions of frames in the MRAC and the MCRC are presented in sections 7.1. and 7.2., respectively. Then the similarities and differences observed across the two registers are discussed in section 7.3. As explained in section 3.6.3 in the Methodology chapter, MRAC 4-gram frames and MCRC 4-gram verb-based frames were consider for functional analysis. Section 7.4. presents the findings of the functional analyses of selected MRAC and MCRC frames and their semantic groups of fillers. Finally, section 7.5. answers the final research question that asked about the variability of semantic categories of fillers within frames. The sections described above seek to answer the following research questions and subquestions:

RQ 3a: How does the use of phrase frames compare in the two registers? Are there any variations in terms of predictability, variability, and structures?

RQ 3b How can a grouping by semantic domains of fillers inform the functional analysis of phrase frames? What are the main functions served by some of the most salient frames and their fillers in MRAs and MCRs? Are there any variations between the two registers?

Research Question 3c: How do semantic domains vary within one frame? Is there any variation between frames involving lexical bundles and those that do not?

7.1 Description of Frames in Medical Research Articles

This section first reports the structural classification and analysis of the MRAC frames on the lists presented on table 7.2. Then, the section describes the variability and predictability of the frames in this register.

7.1.1 Structures of Frames in the MRAC

All lists of frames were classified according to the three categories identified by Gray & Biber (2013) and described in the Literature Review chapter, that is, verb-based (e.g., *did not * the, were * as previously described, occurred * the*), other content word-based (e.g., *average * of, the role of * in the, a significant * in*), and function word based (e.g., *the * of, to * the, of * from, with a * of*). Figure 7.1 shows the distribution of these three categories in the MRAC. Content word-based frames were by far the predominant structure in the MRAC. This finding differs from what has previously been reported in the literature (e.g., Geluso, 2019; Gray & Biber, 2013). In their comparison of frames in academic prose and conversation, Gray & Biber

(2013, p. 122) found content word-based frames to be “fairly infrequent” in both registers, accounting for only 6-15% of all 4-gram frames occurring at least 40 times pmw in their corpora. Academic writing was found to make more frequent use of both function word-based and verb-based frames. This is only partially supported by the findings of the present structural analysis. While verb-based frames were frequent in the MRAC, representing 30% of all analyzed frames, function word-based was the least frequent structure. On the other hand, content word-based frames, previously reported as “fairly infrequent” in academic writing, were by far the predominant structure in the MRAC.

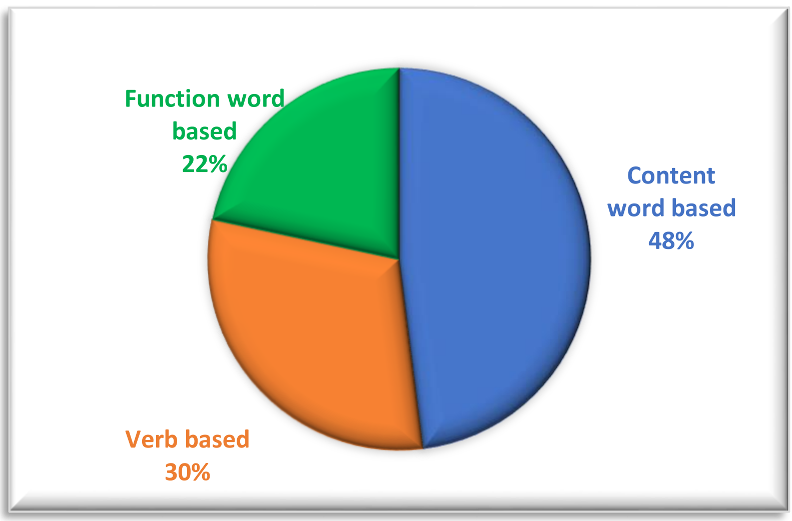


Figure 7.1. *Distribution of Frame Structures in the MRAC*

One explanation of the pervasiveness of content-word based frame in medical research articles could be the frequent use of noun pre- and post-modification in this register, discussed in chapter 5 in the analysis of multiword collocations (see section 5.2). Indeed, the list of content word-based frames in the MRAC included a large number of frames in the forms of noun phrases and/or noun phrase fragments with variable slots (e.g., *children * than 5 years, participants in the * group, the completeness and * of the, a significant * in the, patients * with, baseline * of*).

Nevertheless, when we look at the frame tokens per million words, function word-based frames can still be considered fairly frequent in medical research articles. This is somehow consistent with Gray & Biber's (2013) findings regarding the high frequency of function-word frames in academic writing, with the difference that authors of medical research articles appear to rely on a smaller number of types that are repeated throughout the papers. As a result, the identified function-word based frames in the MRAC often occurred at much higher frequencies than content word-based and verb-based. To illustrate this difference, Table 7.3. shows the top and bottom 50 3-gram frames with **function word-based bolded,** *verb-based in italics*, and content word-based shaded. Function word-based frames were not only predominant in the top 50 list (the most recurrent frames), but also, they occurred at a very high frequency with the first frame on the list occurring 8816 times pmw, compared to 1152 and 800 times pmw for the first content word-based and verb-based frames, respectively. Note that there were only six function word-based frames on the bottom 50 list. This trend was observed mainly in the 3-gram and 4-gram frame lists.

As there appeared to be some differences in frame structures depending on length, the structural correlates of frames in the different lists (Figure 7.2), are also briefly described in this section. Content-based frames more than doubled the number of both function word-based and verb-based frames in the 3-gram list and occurred in relatively similar numbers as verb-based frames in the 4-gram and 5-gram lists. The 6-gram list, however, displayed a different pattern than the other list, with verb-based frames outnumbering content word-based frames.

Table .7.3 Distribution of Structural Types in the Top and Bottom 50 3-gram Frames Occurring ≥ 40 Times PMW

Top 50			Bottom 50		
Frames	**Freq**	**Range**	**Frames**	**Freq**	**Range**
the * of	8716	193	to * cells	42	24
a * of	1658	192	**two * of**	42	31
to * the	1298	192	**with * at**	42	28
the * group	1152	92	after * treatment	41	21
of * in	1103	186	*analyzed * the*	41	34
in * with	848	172	*assigned * the*	41	25
*were * to*	800	177	**both * the**	41	32
*were * in*	792	174	course * the	41	31
*were * with*	787	168	*detected * the*	41	35
*was * in*	767	174	in * other	41	31
the * in	760	175	in * setting	41	25
in * of	699	171	in * trials	41	26
*was * to*	678	173	intervention * the	41	21
*was * by*	646	166	of * antibodies	41	13
in * to	626	159	of * dna	41	18
*the * was*	622	153	of * effects	41	26
we * that	591	145	of * funding	41	40
we * the	584	155	of * loss	41	15
*were * for*	556	164	of * stroke	41	10
*was * with*	524	156	only * the	41	35
of * with	512	155	randomised * trials	41	11
the * to	495	158	the * activity	41	18
the * for	489	161	the * epithelium	41	8
for * of	481	152	the * vaccine	41	9
of * to	465	149	to * hours	41	21
of * of	437	146	*treatment * were*	41	28
is * to	432	152	also * for	40	28
to * a	422	156	**an * between**	40	25
of * patients	419	88	*are * on*	40	34
in * the	405	143	cd4 * cell	40	14
the * that	398	141	expression * of	40	23
as * as	396	134	further * by	40	31
*was * as*	359	142	in * clinical	40	30
both * and	353	128	in * heart	40	14
*the * were*	349	130	**no * on**	40	28
*be * to*	346	145	of * human	40	28
*are * in*	345	139	*our * are*	40	35
of * or	345	143	*our * is*	40	30
we * a	344	128	*previous * have*	40	33
to * in	339	145	serum * of	40	14
to * of	337	138	substantial * of	40	32
of * was	334	144	the * design	40	35
was * for	322	136	the * prevalence	40	19
a * in	314	131	the * review	40	34
*increased * of*	314	94	**their * to**	40	31
in * patients	311	83	to * health	40	22
of * study	310	117	**to * more**	40	30
with * of	308	135	*was * when*	40	31
are * to	307	136	*were * significantly*	40	35

As for function word-based frames, they were, as mentioned above, frequent only in the 3-gram and 4-gram lists and appeared to dwindle as the frames got longer. They represented less than 10% of 228 5-gram frames and did not occur at all in the 6-gram frame list. The variability and predictability of these different structural groups of frames are presented and discussed in the next section.

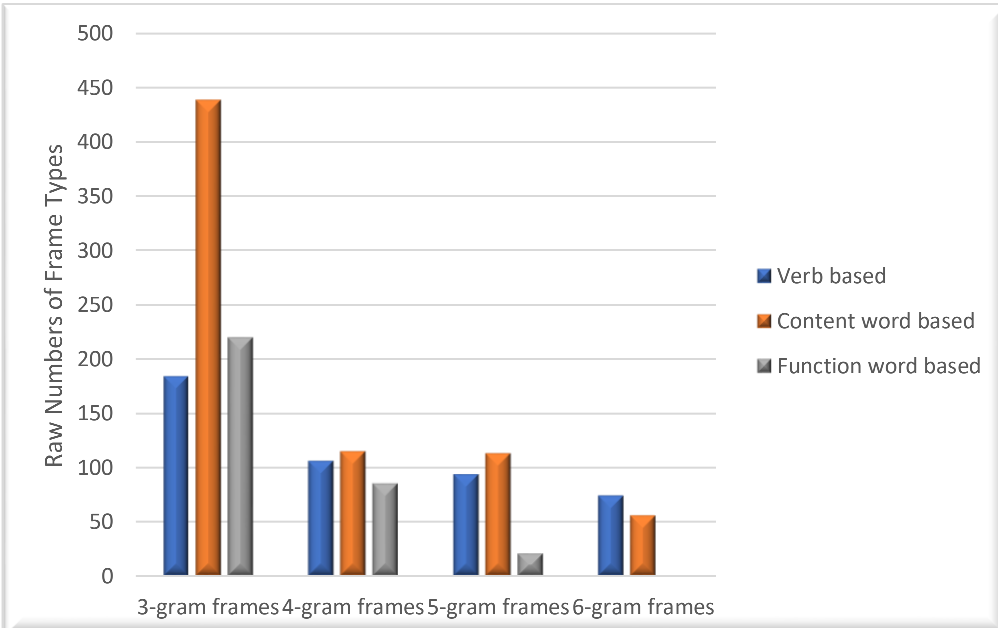


Figure 7.2 Structural Correlates of MRAC Frames by Frame Length

7.1.2 Variability and Predictability of Frames in the MRAC

As explained in section 3.6.2. in the methodology chapter, the type-token ratio scores provided by *Antgram* (Anthony, 2020) were used to determine the variability of the slots in the analyzed frames. Figure 7.3. shows the overall variability of frames in the MRAC. Overall, very few frames in the MRAC were highly variable (only 8% of all 1546 analyzed frames). A little more than half of the analyzed frames were found to be variable while fixed frames accounted for 41% of the analyzed frames. While the greatest proportion of frames in the MRAC are variable, which is consistent with the findings of Gray & Biber (2013) regarding the internal variability of frames in academic writing, it remains that there is a large number of relatively fixed frames in MRAs. Gray and Biber found that only 28% of 4-gram frames occurring at 40 times pmw in their academic writing corpus were fixed. This percentage is lower than the 41% of fixed frames reported in this section. To better understand this difference, the frames were further analyzed by frame length and structure. The findings are summarized in Figure 7.4.

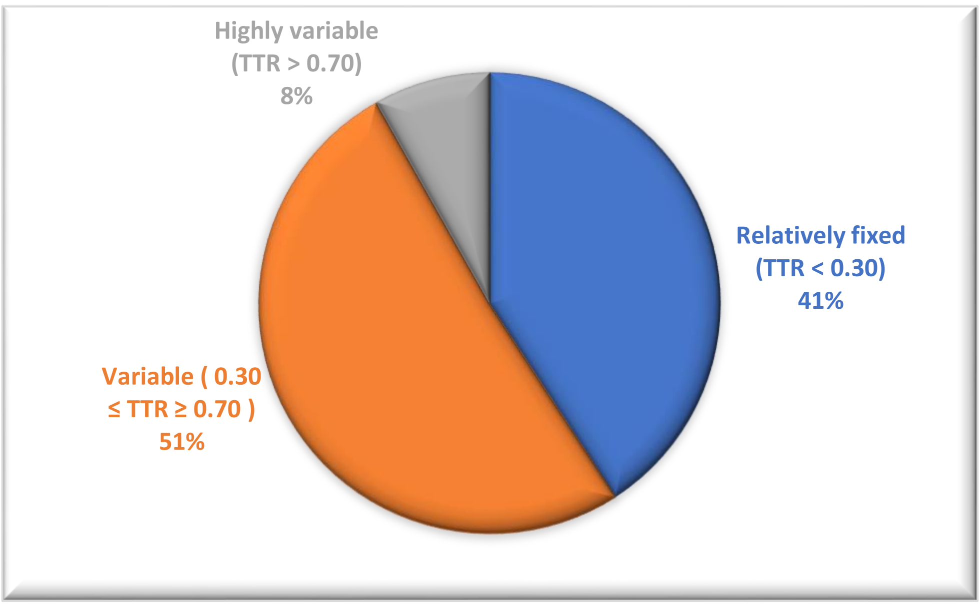


Figure 7.3. Overall Variability and Fixedness of Frames in the MRAC

Given previous findings by Gray & Biber (2013) on the variability of frames in academic writing, it could be expected that content word-based frames, the most frequent structure in the MRAC, would be primarily fixed. However, that was not the case. Relatively fixed content word-based frames accounted for approximately 45% or less of frames across all lists, meaning that over half of the content word-based frames in all lists were variable and/or highly variable. In fact, the data shown in Figure 7.4 indicates that with the exception of verb-based 6-gram frames, MRAC frames of all lengths and structures are primarily variable to highly variable, which is not surprising, given that frames in academic writing have already been described in the literature as primarily variable. Understandably, function word-based frames are the most variable ones, while content word-based and verb-based frames display similar trends, with most of them being variable and a few being highly variable. However, there still was a fairly high proportion (40 – 45%) of relatively fixed content word-based and verb-based frames across all lists. The only exception is noted in the 6-gram list where fixed verb-based frames account for 58.67% of sequences in that list.

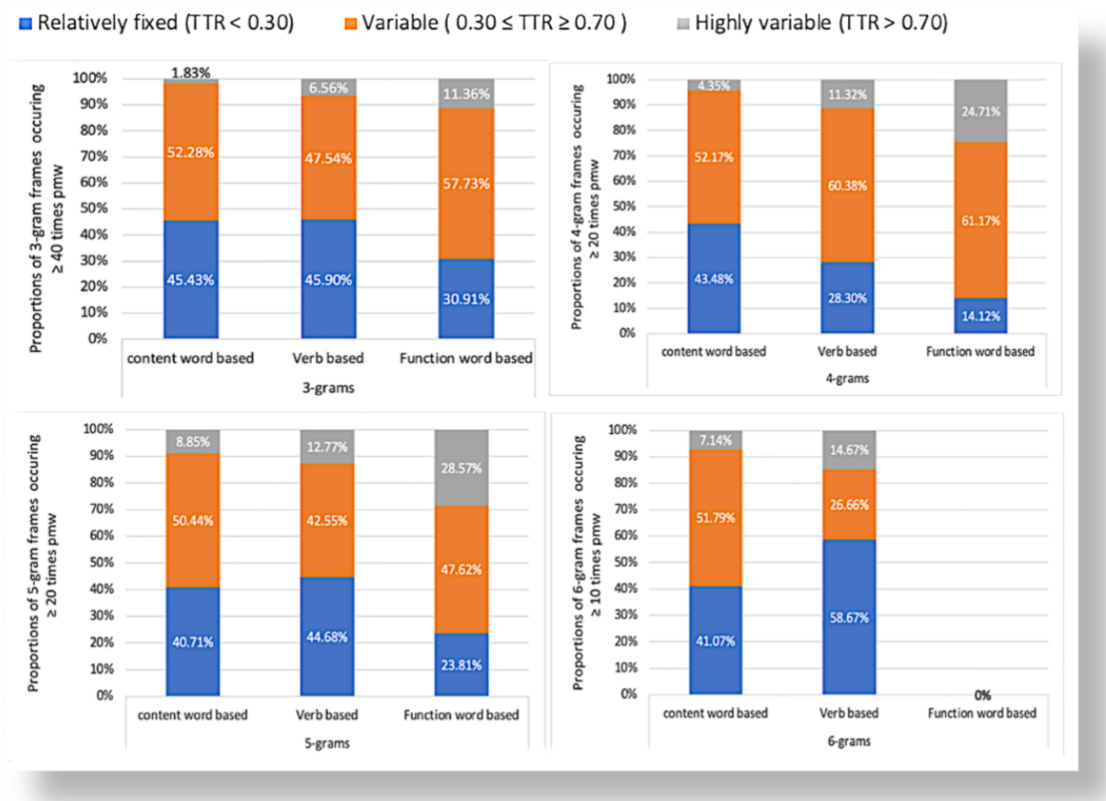


Figure 7.4. Variability and Fixedness of Frames by Length and Structure

These findings slightly deviate from previous Gray & Biber's (2013) description of frames in academic writing. The strong lexical patterning observed in the analyzed frames was described by Gray & Biber (2013) as being characteristic of conversation. This strong lexical patterning probably explains that the proportion of fixed frames is relatively higher than what could be expected in academic writing.

To further analyze the internal variability of the frames, the predictability of fillers of the variable slots was examined. As explained in section 3.6.2 in the methodology chapter, all frames were classified into three groups (High Predictability, Moderate Predictability, and Low Predictability), based on their normalized entropy scores provided by *Antgram* (Anthony, 2020). Figure 7.5 shows the proportions of frames in each of these three groups, and Figure 7.6 provides information on the predictability of frames by length and structure. The vast majority of analyzed frames (71%) has low predictability, and the detailed analysis shown in Figure 7.6 reflects the same trend. Frames of all forms and across all lists primarily have a low predictability. Given that together, variable and highly variable frames account for 60% of analyzed frames (see Figure 7.3), it was reasonably expected that most frames would have a low predictability, and this was confirmed by the results summarized in Figures 7.5 and 7.6.

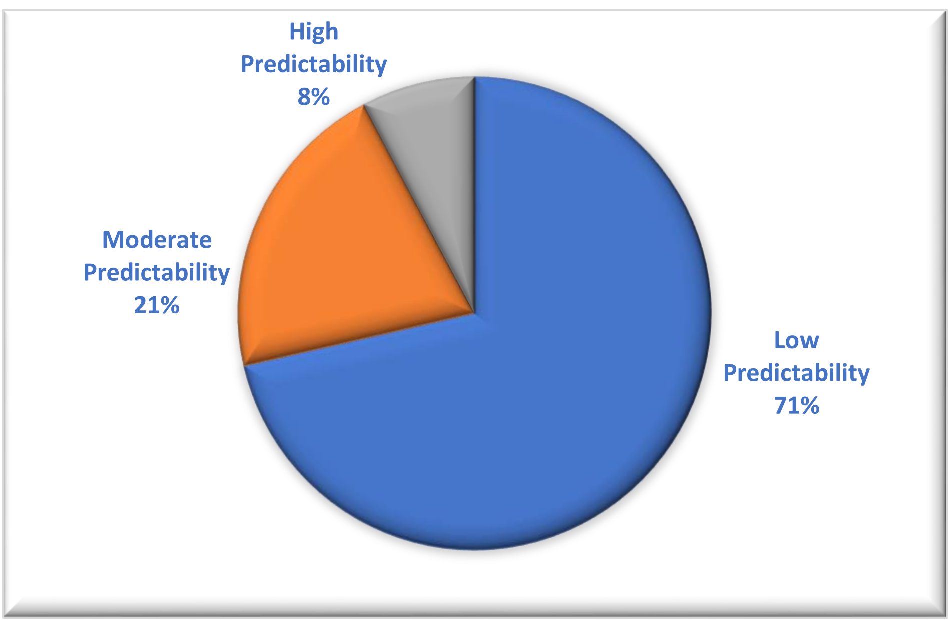


Figure 7.5. Overall Predictability of Analyzed Frames

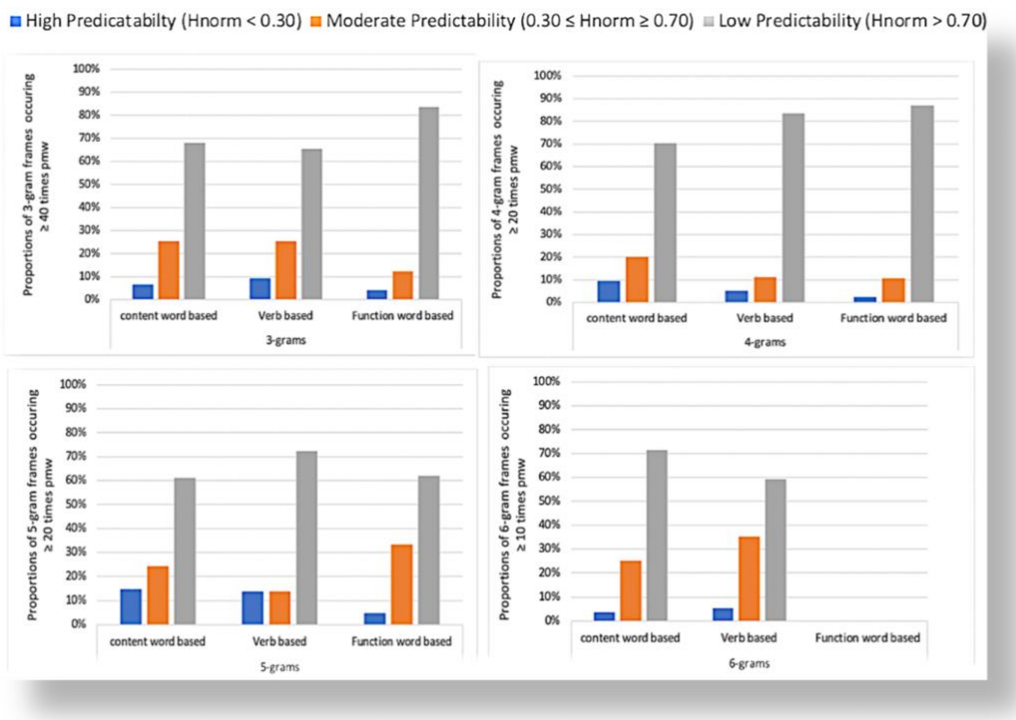


Figure 7.6. Predictability of Frames by Length and Structure

However, the very high proportion of frames with low predictability appeared to be indication that in addition to variable and highly variable frames, some relatively fixed frames could be not as predictable as would be expected. Therefore, the three predictability lists were further analyzed to find more information on the variability of frames in this list. Unsurprisingly, the high predictability group consisted exclusively of relatively fixed frames like *as * result, in * control group,* or *the presence * absence of*. Such sequences often involve bundles and are discussed later in section 7.5. The sequences with moderate predictability also were primarily fixed frames and included only 10% of variable frames, as shown in Figure 7.7. On the other hand, the low predictability list included frames of all three degrees of variability, that is, highly variable, variable, and relatively fixed, as shown in Figure 7.8. As could logically be expected, the majority of frames in the low predictability list are variable. All analyzed highly variable sequences were also found to have low predictability.

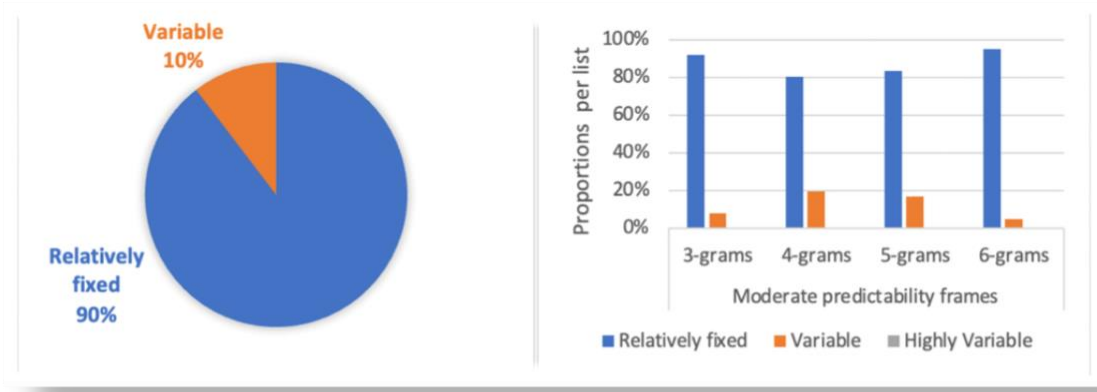


Figure 7.7. Proportions of Fixed and Variable Frames with Moderate Predictability in the MRAC

On the other hand, the presence of relatively fixed frames in the low predictability group was less expected than that of variable and highly variable frames. Yet, 18% of low predictability frames were relatively fixed, and as shown in the detailed section of Figure 7.8 (on the right), those relatively fixed frames occurred in all lists. This highlights two important points. First, it underscores the value of combining both variability and predictability measures in the analysis of the internal variability of frames. Without the entropy scores, the low predictability of these relatively fixed frames would probably have been overlooked. As rightly noted by Gray & Biber (2013) type/token ratio is sensitive to token frequency and may not reflect the actual predictability of highly frequent frames.

Indeed, inspection of the lists of low predictability fixed frames of all lengths did reveal the presence of very high frequency frames, mainly in the 3-gram and 4-gram lists. For instance, the frame *to * the* (TTR = .27, H_{norm} = .83) occurred 1298 times pmw and had 465 fillers, which explains the misleading low TTR of .27 that depicts the frame as relatively fixed. The most frequent filler (*assess*) occurred 140 times, other recurrent fillers included *determine* (99 times), *evaluate* (67 times), *estimate* (62 times), *reduce* (49 times), *examine, test,* and *compare* (45 times, each), to list only a few.

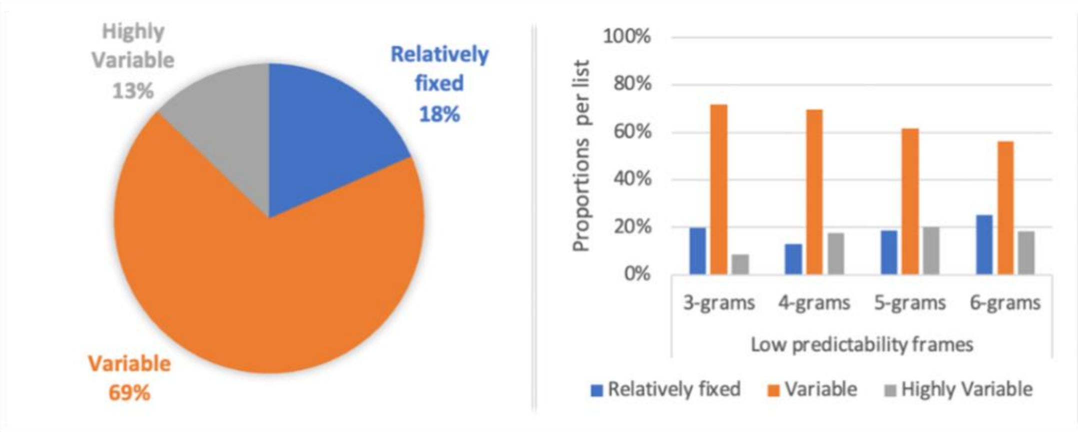


Figure 7.8. Proportions of fixed, variable, and highly variable frames with low predictability in the MRAC

The second important takeaway from these findings is that low predictability frames appear to be the primary candidates for academic writing instruction. In addition to highly variable and variable frames, the predictability measure makes it possible to identify what could be referred to as "false fixed" sequences like *to * the*. Based on its number of fillers (465) this frame is not fixed at all, and with fillers like the ones listed above, this frame and its semantic groups of fillers can be expected to serve important discourse functions in medical research articles. Therefore, in addition to variable and highly variable frames, the "false fixed" frames identified in the MRAC are worth considering for medical writing instruction.

This section has provided a description of frames in medical research articles. The structural classification revealed a predominance of content word-based frames in the MRAC. The analysis of the internal variability of frames revealed that the analyzed MRAC were primarily variable, with low predictability. Overall, these findings are consistent with Gray & Biber's (2013) description of the variability and predictability of frames in academic writing. However, the findings of the structural analysis indicate that MRAs have a lexical patterning that is stronger than what has been described for academic writing. Perhaps, such strong lexical patterning results from authors' frequent use of complex nouns phrases discussed in chapter 5, section 5.2. In terms of methodology, the inclusive approach in this study, that is, the investigation of sequences of various lengths, allowed a more detailed analysis that revealed variations in frame structures depending on lengths. Additionally, the combination of both variability and predictability measures revealed the presence of "false fixed" frames that appear to play important roles in the construction of discourse in medical research articles. The functions of frames are discussed later, in section 7.4. After this description of MRAC frames, I now turn MCRC frames.

7.2 Description of Frames in Medical Case Reports

As previously indicated on Table 7.2, 1917 frames from the MCRC qualified for analysis. These frames were all classified structurally, and their internal variability was analyzed based on their TTR and entropy scores. MCRC frame structures and internal variability are described in the next sections.

7.2.1 Structures of Frames in the MCRC

Figure 7.9 shows the distribution of MCRC frames in the three major structural groups, namely, function word-based, verb-based, and other content word-based. The same trend described above for frames in the MRAC was observed on the lists of MCRC frames. Content word-based frames were the most frequent structure, accounting for almost half of analyzed MCRC frames. Verb-based frames were the second most frequent category, followed by function word-based frames in third and last position. The proportion of function word-based frames (17%) was even lower in the MCRC than in the MRAC. These findings lend support to the previously made observations regarding the phraseological patterning of medical research articles, and suggest that perhaps, medical writing in general follows a strong lexical patterning.

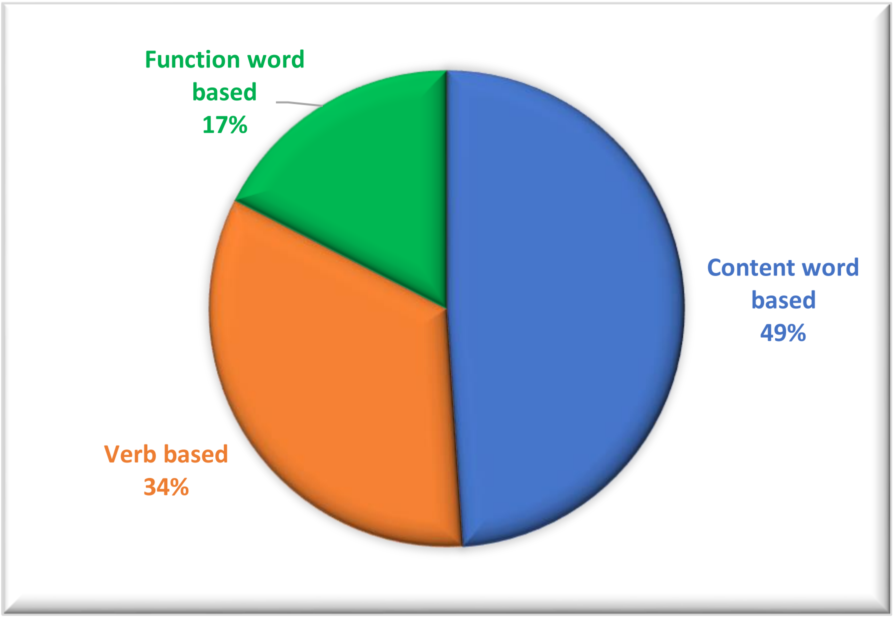


Figure 7.9. Distribution of Frame Structures in the MCRC

The detailed analysis of frames by length revealed similar trends as those observed in the MRAC, except in the 4-gram list (Figure 7.10). In the MCRC, function word-based frames occurred primarily in the 3-gram and 4-gram frame lists. Content word-based frames by far outnumbered sequences in the two other structural categories (i.e., verb-based and function word-based) in the 3-gram list and occurred at relatively similar number as verb-based frames in the 5-gram list. On the other hand, verb-based frames slightly outnumbered content word-based sequences in the 6-gram frame list, as was observed with frames of the same length in the MRAC. The 4-gram list however revealed a different picture than its counterpart in the MRAC. The number of verb-based MCRC frames is much higher in the 4-gram list than in any of the other lists, including the 4-gram list in the MRAC. This probably explains the stark difference between the MRAC and the MCRC 4-gram lists, noted in the introduction of the present chapter.

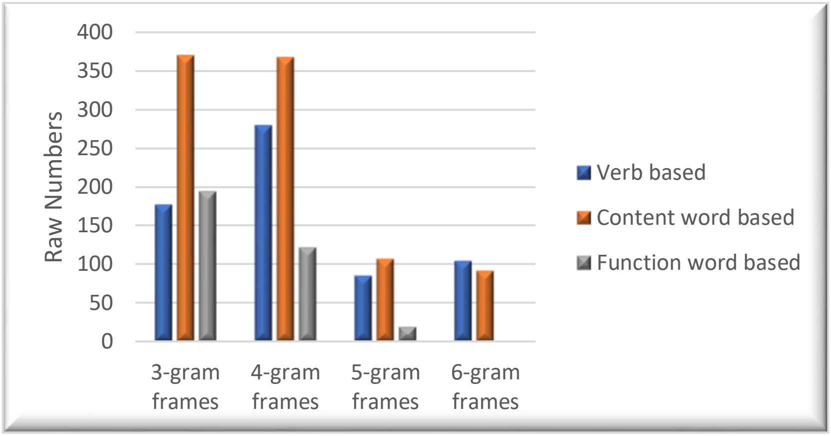


Figure 7.10. Structural Correlates of MCRC Frames by Length

An inspection of the list of 4-gram verb-based frames revealed the presence of several frames that can be linked to verb-related and clause-related bundles reported in chapter 6 to be frequent in medical case reports and to serve functions specific to this register (*e.g.*, Reporting results, Description (subject-related), Decision/Outcome). Examples of such frames and their most frequent fillers are shown in Table 7.4. Note the high frequency of these sequences. It appears then, that these apparently specific verb-based frames accounted in large part for the high number of 4-gram frames in the MCRC.

Table 7.4. Examples of Verb-based Frames Related to Functions Specific to MCRs

Frame	**Frame frequency pmw**	**Filler #1**	**Frequency of Filler #1**
*was * to the*	272	*admitted*	72
*was * in the*	200	*seen*	20
*patient was * to*	198	*transferred*	28
*he was * to*	162	*transferred*	25
*was * with a*	139	*treated*	16
*patient was * with*	130	*treated*	51
*was * to have*	122	*found*	60
*he was * with*	115	*treated*	35
*she was * to*	112	*referred*	25
*was * on the*	108	*discharged*	10
*she was * with*	105	*treated*	48
*had * history of*	96	*a*	55
*the * did not*	93	*patient*	51
*he was * on*	84	*started*	42
*was * from the*	83	*discharged*	20
*had a * of*	80	*history*	55
*was admitted * the*	80	*to*	72
*it was * that*	75	*decided*	7

7.2.2 Variability and Predictability of Frames in the MCRC

All 1917 MCRC frames were also classified into *Relatively fixed*, *Variable*, and *Highly Variable,* based on their TTRs and the same thresholds used in the classification of MRAC frames (see section 3.6.2 in the Methodology Chapter). The proportions of each type of frame are shown in Figure 7.11. The Relatively fixed frames were found to be more frequent in the MCRC than in the MRAC even though, together, variable and highly variable frames accounted for more than half of all analyzed frames. The detailed analysis by frame lengths and structures, shown in Figure 7.11, shed more light on the internal variability of MCRC frames.

The majority of variable and highly variable frames seem to be function word-based, as both content word-based and verb-based frames appear to be primarily fixed across all lists. As clearly shown in Figure 7.12, more than half of 4-gram and 5-gram content word-based sequences are relatively fixed, and so are approximately 45% of 3-gram and 6-gram frames of the same structure. Verb-based frames are apparently even more fixed, with 52.72%, 65.88%, and 64.42% of fixed sequences in the 3-gram, 5-gram, and 6-gram lists. Only the list of 4-gram verb-based frames contains a higher proportion of variable frames than fixed frames.

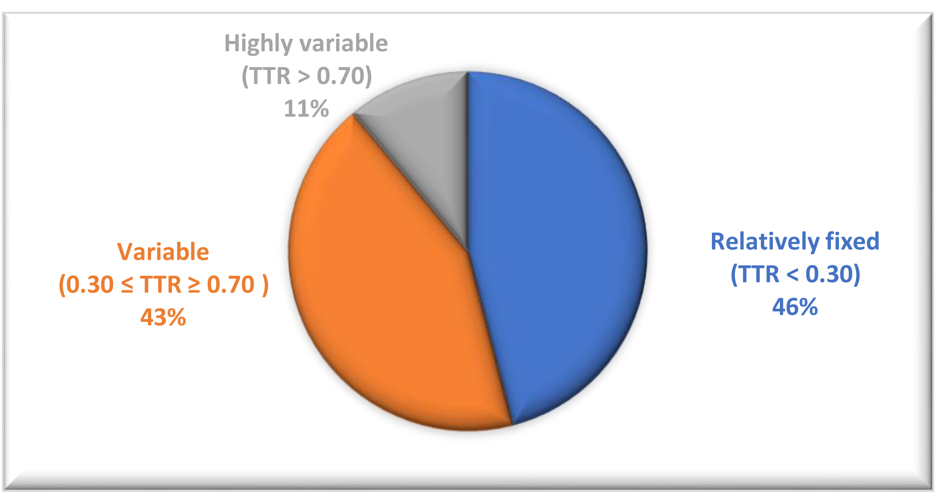


Figure 7.11. Overall Variability of Analyzed MCRC Frames

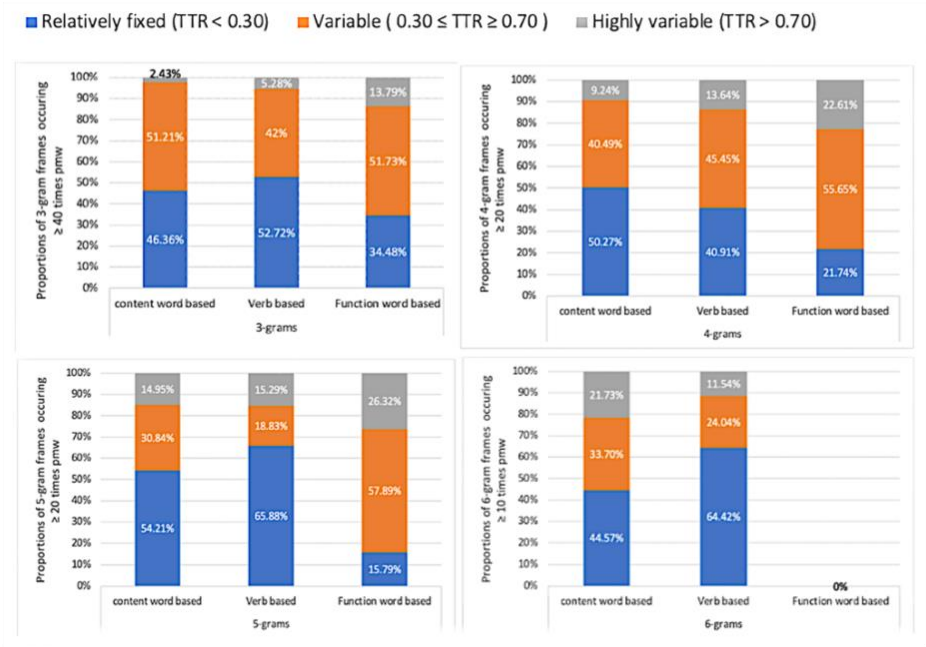


Figure 7.12. Proportions of Fixed, Variable, and Highly Variable MCRC Frames by Length and by Structure

However, these findings should be interpreted with caution as we know that TTR does not tell the whole story regarding the internal variability of frames. Frame entropy scores were needed to supplement these preliminary results regarding the internal variability of MCRC frames. The classification of frames according to their entropy scores revealed that MCRC frames primarily had a low predictability, regardless of length or structures (Figure 7.13). Together, low frequency frames from all lists accounted for over 64% of analyzed MCRC frames. Given the findings previously reported regarding MRAC frames in the low predictability group, it was expected that the list of low predictability MCRC frames would also include “false fixed” sequences. This hypothesis was borne out by the findings of the analysis of the low predictability list, shown in Figure 7.14.

As was the case in the MRAC, the low predictability group included sequences of all three variability degrees: fixed, variable, and highly variable, and as expected, the list included a large number of "false fixed" frames of various length that occurred at remarkably high frequencies in the MCRC. Some examples of "false fixed" frames and their most frequent fillers are shown in Table 7.5.

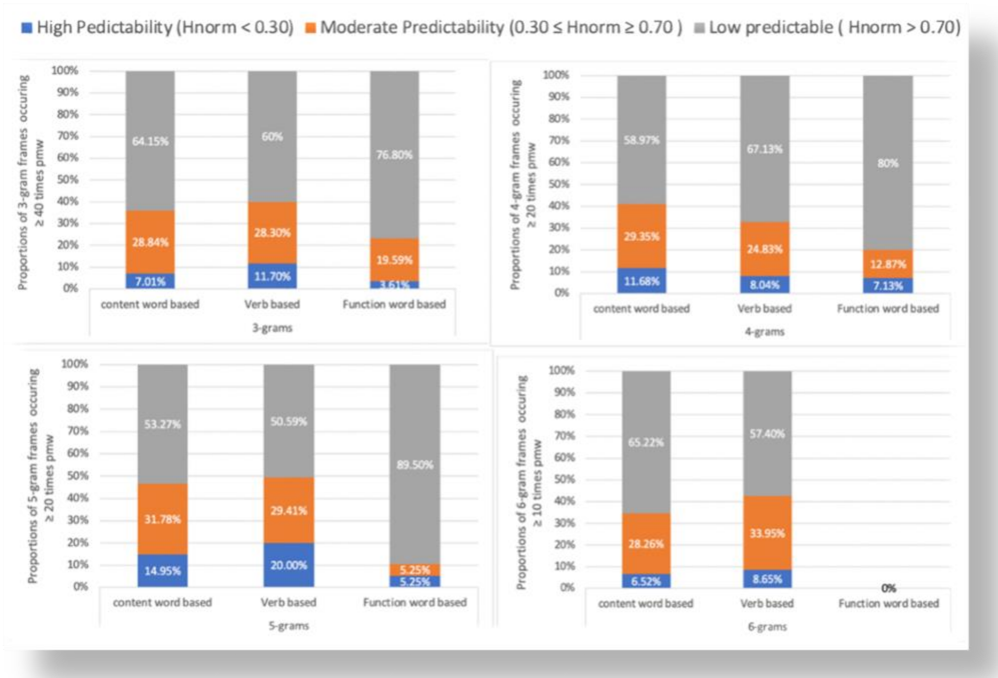


Figure 7.13. Predictability of MCRC Frames by Length and by Structure

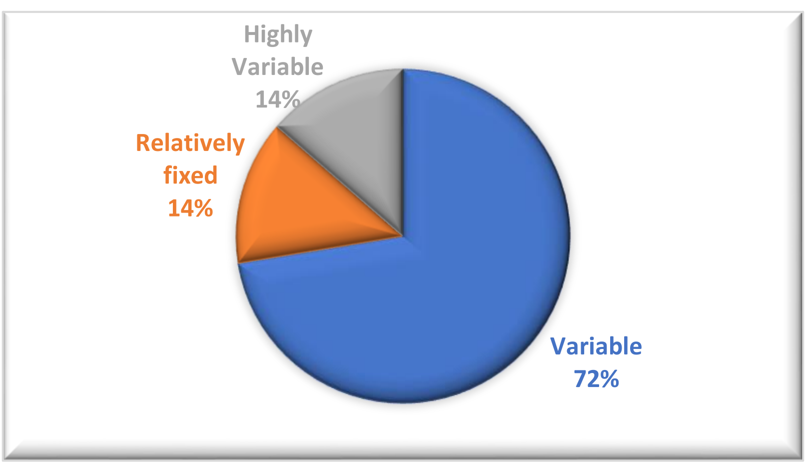


Figure 7.14. Distribution of Fixed, Variable, and Highly Variable MCRC Frames in the Low Predictability List

Examples in Table 7.5 show that in addition to occurring at high frequencies in the MCRC, these “false fixed” frames can be of any length and/or structure. Further analysis showed that in the low predictability list, “false fixed” frames accounted for 18.09% of 3-gram frames, 14.68% of 4-grams, 21.37% of 5-grams, and 27% of 6-grams. This suggests that relatively fixed frames in the MCRC account for much less than the initial 46% that resulted from the classification by TTR. The differences and similarities between the two registers are summarized in the next section.

Table 7.5. Examples of "False Fixed" Frames in the MRAC

	Frame	Frame frequency pmw	TTR	*H*norm	Filler #1	Frequency of Filler #1
3-gram frames	*the * of*	8617	0.13	0.78	*presence*	518
	*a * of*	2668	0.15	0.72	*case*	408
	*was * to*	1443	0.17	0.79	*referred*	188
	*was * with*	932	0.24	0.72	*treated*	182
	*is * to*	834	0.21	0.81	*thought*	50
	*in * of*	642	0.29	0.8	*cases*	103
	*was * for*	612	0.23	0.82	*negative*	64
	*in * cases*	327	0.25	0.8	*some*	41
	*we * the*	313	0.29	0.72	*report*	68
	*clinical * of*	294	0.28	0.83	*features*	28
	*high * of*	273	0.17	0.79	*index*	54
4-gram frames	*a * year old*	1072	0.08	0.95	*NUMBER*[#]	
	*in the * of*	996	0.21	0.74	*setting*	125
	*the * of a*	522	0.25	0.75	*case*	106
	*for the * of*	342	0.25	0.73	*treatment*	99
	*was * to the*	272	0.2	0.71	*admitted*	72
	*it is * to*	218	0.26	0.72	*important*	81
	*the first * of*	186	0.26	0.74	*report*	45
	*has been * to*	148	0.22	0.72	*reported*	27
	*it is * that*	147	0.27	0.88	*possible*	22
5-gram frames	*a * year old man*	204	0.27	0.95	*NUMBER*[#]	
	*the patient was * to*	173	0.27	0.81	*transferred*	27
	*at the * of the*	111	0.23	0.77	*level*	24
	*present * case of a*	71	0.03	1.0	*the*	36
	*a high * of suspicion*	56	0.05	0.72	*index*	40
	*has been * to be*	46	0.26	0.77	*reported*	15
	*should be * in patients*	45	0.2	0.72	*considered*	13
	medical history was * for	32	0.22	0.71	significant	17
6-gram frames	*the patient was * to the*	54	0.19	0.8	*transferred*	17
	*to convert to * per liter*	38	0.24	0.87	*microkatals*	9
	*with a * history of a*	38	0.26	0.92	*1-year*	3
	*on the * side of the*	32	0.22	0.71	*left*	15
	*we * a rare case of*	31	0.1	0.73	*present*	17
	*year old * presented to the*	30	0.2	0.76	*man*	18
	*is the most common * of*	26	0.27	0.78	*cause*	13
	*with a 2 * history of*	23	0.17	0.92	*week*	9

7.3 Similarities and Differences between Frames in the MRAC and MCRC

As mentioned above, the analysis of MRAC and MCRC frames revealed some similarities between the two registers. Both registers displayed a strong lexical patterning with a predominance of content word-based frames and the use of a relatively low number of function word-based sequences, compared to what has been previously described in the literature for academic writing. According to Gray and Biber (2013, p. 128), academic writing is “inherently linked to grammatical constructions”, which is reflected in the high frequency of function word-based frames. However, we have seen from the findings of the structural analyses that frames from both the MRAC and MCRC slightly deviate from the description of frames previously identified in academic writing.

The high frequency of content word-based frames in both the MCRC and the MRAC is consistent with the findings of multiword collocations discussed in section 4.2 of chapter 5. Both registers were found to make extensive use of noun pre- and post-modification to express field-related concepts and processes (e.g., *polymerase chain reaction, stem cell transplantation, in situ hybridization, induction of mixed chimerism*). The lists of content word-based frames from both corpora included several instances of sequences resulting from noun pre- and/or post-modifications, mainly in the 4-gram, 5-gram, and 6-gram lists (e.g., *the placebo group * ratio, a two sided * level of, woman with * history of, a * blood cell count*). Sequences like these are complex noun phrases with variable slots that offer medical writers frames to create various “stable names” for elements and processes related to their studies (e.g., *a **white/red/complete/normal** blood cell count, the **primary/combined/bivariate/composite** end point, patients with **a/no/previous/known** history of*), to use the words of Hyland & Tse (2007, p. 224).

In addition to content word-based, verb-based frames were also found to be very frequent in both registers, which is consistent with previous descriptions of frames in academic writing by Gray & Biber (2013). There was a greater number of verb-based sequences in the MCRC than in the MRAC, and this was explained by the presence of frames related to clausal and verb-related lexical bundles in medical case reports, which have already been discussed in section 6.2.2 of chapter 6. Those bundles were found to serve functions specific to medical case reports (*e.g.*, Reporting results, Description (subject-related), Decision/Outcome). Examples of such verb-related frames include *the * was diagnosed with, confirmed * presence of, the * was found to have, physical * revealed*, and *past * history was*. As mentioned above, the functions of these and other analyzed frames are discussed in sections 7.4 and 7.5.

Regarding the internal variability of frames, sequences in both registers were primarily variable with low predictability. The analysis of the low predictability lists in both corpora revealed that in addition to highly variable and variable frames, this category included relatively high proportions of "false fixed" frames. These were highly frequent frames that were in reality either variable or highly variable, making the proportions of variable and highly variable frames even higher in both corpora. To complete the comparison of the two registers, the shared frames were identified and analyzed.

The raw numbers and proportions of shared frames in each corpus are shown in Table 7.6. The two corpora shared a relatively high proportions of frames in the 3-gram and 4-gram lists, but the numbers decrease dramatically in the lists of 5-gram and 6-gram frames. Inspection of the lists of shared frames revealed that most of these sequences are primarily function word-based. This can be explained by the fact that, as reported in sections 7.1 and 7.2 above, function word-based frames in both corpora were primarily in the 3-gram and 4-gram frames lists.

Table 7.6. Percentages of Shared Frames in the MRAC and MCRC

	3-gram frames	**4-gram frames**	**5-gram frames**	**6-gram frames**
Raw numbers	296	133	23	4
Percentage from total MRAC frames	33.60%	43.46%	10.09%	3.05%
Percentage from total MCRC frames	38.40%	17.97%	10.85%	2.04%

Logically, with the predominance of function word-based sequences, shared frames were primarily variable with low predictability. Given the relatively high numbers of shared 3-gram and 4-gram frames, these two lists were further analyzed for potential similarities and/or differences in their most frequent fillers. The analysis revealed three distinct groups shown in Table 7.7. The first group consists of frames with distinct most common fillers in the MRAC and MCRC. This was the largest group and included frames that occurred at high frequencies in both corpora and were for the most part function word-based. Most fillers of frames in this group reflect the difference in study subject and focus between the two registers, described in the situational analysis in chapter 4. As explained in section 4.4, in MCRs, most descriptions center around the subject (the patient), whereas in MRAs, the focus is primarily on experimental procedures.

A couple of examples from Table 7.7. are the frames *the * is* and *the * was.* The most frequent filler for both frames in the MCRC is *patient,* whereas in the MRAC, the most frequent fillers are *study* and *trial* for *the * was* and *the * is,* respectively. Another example from the shared list is the frame *was * with a,* with the most frequent fillers being *performed* in the MRAC, and *treated* in the MCRC. This example is an illustration that even in the description of the interventions, the focus can still be on the patient. In 150 instances out of the 182 times *treated* occurred as the filler of *was * with* in the MCRC, the focus was on the patient, with the subject of the sequence *was treated with* being primarily *the patient, he,* or *she*, as shown in Figure 7.15.

Table 7.7. Examples of Shared Frames Classified by Filler Similarities/Differences

	Frame	Frame Frequency		Filler #1 (Frequency)	
		MCRC	MRAC	MCRC	MRAC
different fillers	*the * of*	8617	8716	*presence (518)*	*use (439)*
	*a * of*	2668	1658	*case (408)*	*total (232)*
	*a * in*	392	314	*role (36)*	*reduction (74)*
	*a * to*	171	125	*bridge (13)*	*response (10)*
	*in the * of*	996	657	*setting (125)*	*presence (131)*
	*the * was*	1915	622	*patient (881)*	*study (119)*
	*was * on the*	108	130	*discharged (10)*	*based (22)*
	*to * the*	1259	1298	*be (55)*	*assess (140)*
	*the * is*	579	185	*patient (59)*	*trial (53)*
same fillers	*for the * of*	342	196	*treatment (99)*	*treatment (54)*
	*it is * to*	218	70	*important (81)*	*important (25)*
	*of the * with*	124	35	*patients (15)*	*patients (21)*
	*a significant * in*	31	48	*increase (18)*	*increase (32)*
	*after the * of*	133	36	*onset (10)*	*onset (13)*
	*in * with*	1019	848	*patients (580)*	*patients (490)*
	*is * as*	181	54	*defined (46)*	*defined (11)*
	*is * by*	358	207	*characterized (98)*	*characterized (129)*
semantically close fillers	*any * of*	105	51	*evidence (24)*	*signs (7)*
	*be * as*	154	51	*considered (35)*	*interpreted (10)*
	*coronary * disease*	50	160	*artery (46)*	*heart (90)*
	*before the * of*	29	34	*onset (6)*	*start (10)*
	*a * risk of*	45	26	*high (11)*	*lower (14)*
	*a * role in*	38	27	*key (7)*	*critical (22)*
	*the * effect of*	28	64	*beneficial (5)*	*protective (8)*

The second group of shared frames consists of frames with the same #1 fillers in both corpora. In addition to familiar sequences like *at the * time* (#1 filler: *same*), *at the * of* (#1 filler: *time*), or *it is * to* (#1 filler: *important*), this group also included some more field-related sequences like *after the * of* (#1 filler: *onset*), or *for the * of* (#1 filler: *treatment*). However, unlike the first group where frames tended to occur at high frequencies in both corpora, in this groups, sequences often occurred at very different frequencies in the two corpora. This may be an indication of differences in the rhetorical choices of authors of case reports and medical research articles. The frame *it is * to*, for example, occurred 218 times pmw in the MCRC and only 70 times pmw in the MRAC. Given that the most frequent filler of *it is * to* in both corpora was "important" (22 time in the MRAC and 81 times in the MCRC), it can be speculated that perhaps, authors of case reports make more frequent use of engagement features. Indeed, other frequent fillers of *it is * to* in the MCRC included *essential* (19 times), *necessary* (10 times), *recommended, imperative,* and *crucial* (9 times, each), all forming engagement expressions when used with *it is * to.* On the other hand, the next three most frequent fillers in the MRAC were: *reasonable* (6 times), *possible* (4 times), and *difficult* (4 times).


ɔsence of systemic symptoms. He was treated with ABVD (doxorubicin [Adriamycin], bleomycin, vinbla
nd Color Doppler sonography. He was treated with amlodipine after consultation with a
ɔry is otherwise unremarkable. He was treated with amoxicillin-clavulanate, 875 mg/125 mg, twice
-convulsive status epilepticus. He was treated with anti-epileptic medication but was
a high CRP of 442 mg/L. He was treated with antibiotics and given 2 mg/kg
had a fever the first night. He was treated with antibiotics because blood culture results
d choreiform arm movements. He was treated with antidepressants without effect. There was
ical sequelae. After discharge, he was treated with combination chemotherapy with a regimen
ɔnstrated no effusion. The patient was treated with antibiotics for two weeks without
f acute LAD occlusion, the patient was treated with aspirin, clopidogrel, and high-dose
with a macular hole.8 The patient was treated with bevacizumab intravitreal injection, and macular
ous GD-TCL. The present patient was treated with brentuximab vedotin, and after 7 months
of pneumonia. Initially, the patient was treated with broad-spectrum piperacillin–tazobactam following
denocarcinoma (video 1). Patient was treated with broad-spectrum antibiotics (meropenem, ofloxacin,
n to the neurology unit where she was treated with high-dose intravenous steroids, IVIG,
lbeit serology was active, and she was treated with hydroxychloroquine, low-dose prednisolone and
CT was interpreted as normal. She was treated with hydroxyethyl starch and pentoxifylline for 7
ith neurological involvement. She was treated with infliximab and methotrexate with complete
ng throughout her right hand. She was treated with intramuscular methylprednisolone injection and a
resulted in PICU admissions. She was treated with intramuscular epinephrine, antihistamines, and st

Figure 7.15. Examples of Typical Subjects of 'was treated with' in MCRs

Finally, shared frames in the third and last group are less frequent than those in the two other groups but have the peculiarity of involving most frequent fillers that belong to the same semantic categories. That may be indication of the preferred semantic categories of sequences in this group or just authors' preferred terminology. For instance, the frame *coronary * disease,* has *artery* as its most frequent filler in the MCRC and *heart* in the MRAC. One of the field-expert informants regularly consulted in this study explained that there is no difference between the expressions formed by the frame and these two fillers (*coronary artery disease* and *coronary* *heart disease*). They "mean the same thing and are used interchangeably" (M. M. Ka, personal communication, October 19, 2021). It appears then that it just happens that in MRAs, authors used *heart* more frequently than *artery,* and the opposite happened in MCRs. As a matter of fact, the two terms are the only fillers of the frame in both corpora. The filler *artery* (86 times) is still very frequent in the MRAC, but MCR authors rarely used *heart* (4 times) as the filler of the frame *coronary * disease.*

Another example with a more variable frame is the use of *key,* and *critical* as the most frequent fillers of the frame *a * role in* in the MCRC and the MRAC, respectively. Examples (7.1) and (7.2) show the two expressions in context and clearly illustrate the similarities of the two expressions in terms of discourse function and colligation. Both sequences are in the description of a medical condition to highlight important factors related to that condition. Note that they are used as direct objects of the monotransitive verb *play.*

(7.1) Genetic factors <u>play</u> ***a key role in*** *nonsyndromic hearing impairment (NSHI) and more than 140 genes have been identified to be engaged in deafness.* (MCRC_20*IMCRJ*89)

(7.2) Previously, GABAergic Purkinje cells in the cerebellar cortex were proposed to <u>play</u> ***a critical role in*** *essential tremor.* (MRAC_20*JCI*2)

However, explaining the differences of fillers in this last group of shared frames by mere authors' preferences might be somehow simplistic. It might be the case that the choice of one filler or the other is related to factors in the textual environment of the frame (e.g., textual colligations, semantic prosody, etc.), but this is beyond the scope of the present study.

To answer RQ3a that asked about the structures, predictability, and variability of frames in the two registers, the findings discussed in sections 7.1 to 7.3 revealed that frames in both corpora are predominantly variable with low predictability despite the predominance of content word-based sequences. This strong lexical patterning of frames from both corpora led to the suggestion that perhaps beyond the two registers under scrutiny, it is medical writing in general, that encompasses that slightly deviates from the previously identified characteristics of academic writing. However, despite the similarities in phraseological patterning, the analysis of shared frames revealed that the most frequent shared sequences often differed in terms of their most frequent fillers, which reflected the differences in study subject and focus between the two registers. Shared frames with the same or semantically similar fillers in both corpora were less frequent and appeared to serve similar functions in both registers. This is a timely transition to the functions of frames and their semantic groups of fillers, to which I now turn.

7.4 Functional Analysis of Frames

This section describes functions served by selected frames and their semantic groups of fillers in MRAs and MCRs. The section also reports the findings of the analysis of frames involving bundles and the variability of semantic groups of fillers within the same frames. For reasons provided in section 3.6.3 of the Methodology chapter, only the MRAC 4-gram frames and the MCRC 4-gram verb-based frames were functionally analyzed in the present study. Selected frames and their semantic groups of fillers were functionally classified using the frameworks presented on Table 3.12 in the Methodology chapter. The next section presents the findings of the functional analysis of the selected MRAC frames.

7.4.1 Functions of Frames in the MRAC

For the functional analysis, the lists of selected frames were further cleaned. Fixed frames with high predictability and sequences with function word fillers were not included in the analysis. The final list of MRAC 4-gram frames to analyze included 202 sequences. As explained in section 3.6.3., the most frequent fillers of frames were grouped by semantic domain using the framework adapted from Biber (2006) and shown on Table 3.13. Concordance lines were generated in *Antconc* (Anthony, 2020) to analyze the frames and their fillers in context. The adapted version of Hyland's (2008a) functional taxonomy used for the analysis of bundles and the functional classification frameworks described in 3.6.3. and shown on table 3.12 were used for the functional analysis of frames. In total, 1947 fillers occurring in frames at least 5 times were manually classified by semantic categories to analyze the functions of the 202 frames.

7.4.1.1 Functions of content word-based frames

Table 7.8 shows the functions served by content word-based frames and their semantic groups of fillers, and Figure 7.16. shows the proportions of frames in each subgroup of content word-based frames serving these functions. The MRAC 4-gram content word-based frames primarily serve research-oriented functions. Description, quantification, and location in time and/or place were the predominant functions. Frames used for description are mostly determiner-initial (*the* or *a/an)*, ending with or including the preposition *of* (*the * activity of*, *a * model of*, *the * burden of*), and follow three main patterns:

- *Determiner/Noun/Pronoun-initial frame + adjective filler,*
- *Determiner/Noun/Pronoun-initial frame + noun filler,* and
- *Determiner/Noun/Pronoun-initial frame + adjective/noun filler.*

Table 7.8. Functions Served by MRAC Content Word-based Frames

Subgroups	Filler word class	Filler sematic category	Function	Examples from the MRAC
Determiner/ Noun/ pronoun-initial	Adjective	evaluative	quantification	*a * increase in (significant, substantial, striking)*
			description	*the * effects of (relative, negative, beneficial, potential, detrimental)*
		size	quantification	*a * number of (large, greater, high, limited, low)*
		relational	quantification	*the * number of (annual, absolute, average, total, median, overall)*
			description	*the * end point (primary, secondary, double, combined)*
		topical	description	*the * response of (biological, cellular, hypothalamic, therapeutic, neuronal)*
	Noun	abstract/ attribute	description	*the relative * of (strength, simplicity, risk, importance, effects)*
		technical	description	*effect of * on (rip140, rituximab, dexamethasone, screening, ivabradine)*
		quantity	location in time	*the * of treatment (end, length, start, initiation, time, course, duration)*
			quantification	*a higher * of (number, proportion, percentage, dose, prevalence, rate*
		abstract/process	Procedure	*the same * as (way, approach, method, manner)*
	Verb	communication	resultative signals	*these results * that (indicate, suggest, show, confirm)*
		existence	stance	*it * possible that (is, remains)*
Preposition-initial				
at the …	Adjective	topical, relational	framing signals	*at the * level (global, individual, molecular, transcriptional)*
for the …	Noun	technical, abstract/other	Structuring signals (subject reference)	*for the * group (iron, CBT, intervention, control, training)*
in the …	Noun	technical, abstract/other, abstract/process	Location in place	*in the * group (avelumab, aspirin, control, CBT, chlorhexidine)*
	Adjective	relational, evaluative	Location in place	*in the * population (general, overall, entire)*
		time, topical	text reference	*in the * study (present, current, previous, cardiogenic)*
		size	location in time	*in the * term (long, short, intermediate)*
of…	Noun	technical	description	*of * disease in (kidney, heart, lung)*
	Adjective	topical	description	*of * disease in (aortic, cardiovascular, metabolic),*
to-infinitive initial	Verb	mental	Procedure	*to * the effect of (assess, determine, examine, evaluate, estimate, study)*
Adj/Adv-initial	Adjective	size attributive	description	*significantly * in the (higher, longer, lower, greater, reduced)*

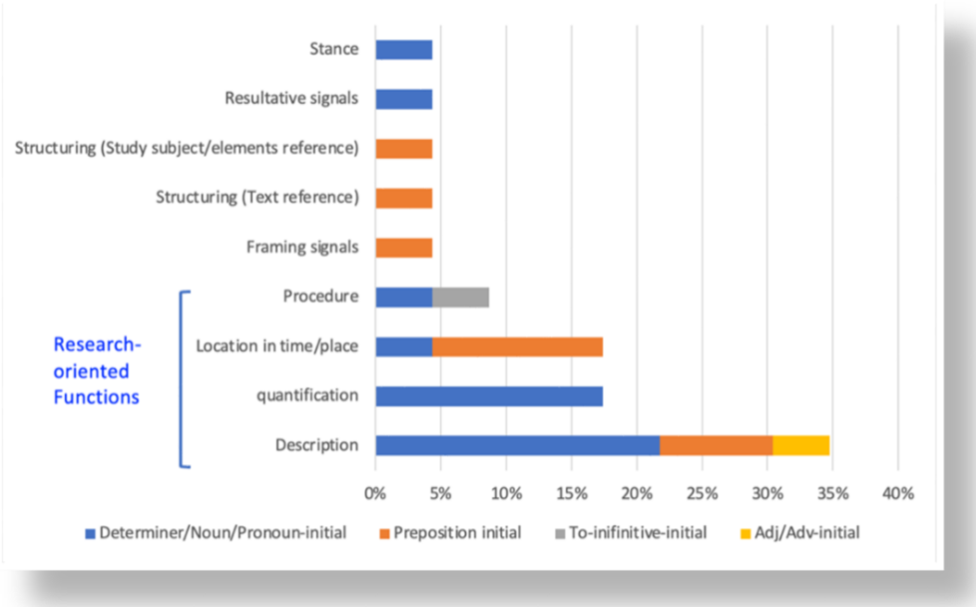


Figure 7.16. Proportions of Content-based Subgroups Serving the Identified Functions in the MRAC

Adjective fillers in the *Determiner/Noun/Pronoun-initial frame + adjective filler* pattern belong to three semantic categories: EVALUATIVE (e.g., *clear, relative, negative, beneficial, potential, positive*), RELATIONAL (e.g., *individual, immunomodulatory, cytoprotective, inhibitory, antiproliferative*), and TOPICAL (e.g., *physiological, phenotypic, cellular, biological)*.

Most frames in this pattern were found to be used with more than one of these three sematic categories. For example, the frame *the * activity of* occurs with two semantic groups of adjective fillers: RELATIONAL (*antiviral, specific, internal, residual, transcriptional*) and TOPICAL (*pharmacological, metabolic, enzymatic, biological, hepatic*). Examples (7.3) and (7.4) below show the use of the frame *the * activity of* with adjective fillers belonging to the RELATIONAL and TOPICAL semantic categories, respectively.

(7.3) It has been suggested that some of these activation steps require or are facilitated by the A subunit [(16, 17),] explaining why A-subunit mutations may affect ***the <u>specific</u> activity of*** *the associated C subunit* (11*JCI*1)

(7.4) Previous in vitro studies have shown that DENV stimulates host cells to increase the synthesis of intracellular cholesterol by upregulating ***the <u>enzymatic</u> activity of*** *3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) reductase*. (20*JCI*8)

For frames in the *Determiner/Noun/Pronoun-initial frame + noun filler* pattern, the nouns primarily belong to two semantic categories: ABSTRACT/ATTRIBUTE (e.g., *specificity, ability, sensitivity, capacity, integrity, strength, quality, effectiveness*) and TECHNICAL (e.g., *pathogenesis, immunogenicity, phenotype, onset, genotype, cells, kinetics*). Unlike frames with adjective fillers, most frames following this pattern primarily involve frequent fillers in only one semantic category. The frames may be occasionally used with other nouns, but those fillers were not included in the analysis as they occurred less than 5 times in the frames. The frame *the clinical * of*, for example, has 33 fillers in the MRAC, 22 of which belong to the semantic category ABSTRACT/Attribute (e.g., *signs, syndromes, characteristics, usefulness, applicability, value, validity, efficacy, relevance, features*). The remaining 11 fillers occurred only once, for the most part, and belonged to different semantic categories (e.g., *course, assessment, onset, question*). Examples (7.5) to (7.7) below show another frame in the pattern *Determiner/Noun/ Pronoun-initial frame + noun filler,* the sequence *effect of * on,* that almost exclusively involved fillers belonging to the TECHNICAL semantic category and mostly referring to some kind of medical intervention or treatment.

(7.5) The protective ***effect of*** ***rituximab*** ***on*** *the disruption of stress fibers observed after exposure to recurrent FSGS sera was not dependent on the regulation of the expression of vinculin, podocin, or nephrin*. (11*Science*4)

(7.6) A few studies have reported a lower rate of skin cancer in transplant recipients who were treated with sirolimus than in those treated with calcineurin inhibitors, but data focusing on the ***effect of*** ***sirolimus*** ***on*** *skin carcinomas are still limited*.

(7.7) Gera and colleagues identified a similar ***effect of*** ***iron*** ***on*** *haemoglobin (MD 7·4 g/L) but did not do meta-analysis for anaemia, instead estimating that between 37·9% and 62·3% of baseline anaemia is amenable to control by iron, less so in malaria-endemic areas.* (13*Lancet*2)

Finally, there are frames that can involve both noun and adjective fillers (*Determiner/ Noun/ Pronoun-initial frame + adjective/noun filler*). Fillers can belong to any of the semantic categories listed above, namely, EVALUATIVE, RELATIONAL, and TOPICAL for adjectives; and ABSTRACT/ATTRIBUTE and TECHNICAL for nouns. The function remains the same, regardless of filler word class or semantic category. Examples (7.8) and (7.9) show the frame *the * treatment group* used with an evaluative adjective (7.8) and a technical noun (7.9) to serve the same function of describing elements related to the study, namely, participants of the study.

(7.8) ***The aggressive treatment group*** *achieved the mean LDL-C goal of 70 mg/dL or lower and the mean SBP goal of 115 mm Hg or lower, and the group means were maintained until the end of the study.* (12*JAMA*5)

(7.9) In the period after discontinuation of ***the placebo treatment group****, an additional 1294 person-years of follow-up for assessment of incident pregnancy were accrued between the 2 PrEP groups and retention remained greater than 95%.* (14*JAMA*5).

Garner (2016) found that most frames in their study served the same function regardless of the filler of the variable slot and called for further investigation to better understand that phenomenon. They did not group the frame fillers into semantic categories, but one explanation could be that the fillers they analyzed belong to semantic categories that, when associated with a specific form of frame, serve the same function as in the examples above. In the MRAC, frames in the three patterns discussed above served the same functions as long as the fillers belonged to the identified semantic categories commonly associated with the frames. Another explanation may be the word classes of the elements of the frame. In the examples above, given that the variable slot occurs between a determiner and a noun, it can be logically expected that the fillers will be an adjective or another noun modifying the fixed noun in the frame.

Quantification, the second most frequent function served by the analyzed MRAC frames, is also expressed by Determiner/Noun/ Preposition-initial frames. Frames in this subgroup are the only content word-based sequences serving that function in the MRAC. Two patterns were observed for sequences in this subgroup used for quantification:

- *Determiner/Noun/Pronoun-initial frame + adjective filler:* when the frame already contains a noun, *and*
- *Determiner/Noun/Pronoun-initial frame + noun filler*: when the frame already contains an adjective.

Adjective fillers in the first pattern belong primarily to one of the following three semantic categories: EVALUATIVE, RELATIONAL, and SIZE (e.g., *large, greater, small, high, low*). Here also, adjectives from more than one of these categories can fill the variable slot of the same frame. Examples (7.10) to (7.12) show the frame *a * increase in* used with adjective fillers, belonging the three possible semantic categories; with a relational adjective in (7.10), a size attributive adjective in (7.11), and an evaluative adjective in (7.12).

(7.10) ***A <u>similar</u> increase in*** *the expression of cholesterol biosynthetic genes has been observed in the livers of the complete LXRα knockout, although the mechanism that accounts for the misregulation of HMG-CoA synthase in the absence of LXR remains to be determined.* (12*JCI*3)

(7.11) ***A <u>greater</u> increase in*** *diastolic blood pressure was observed when evacetrapib, 100 mg/d, was administered in combination with simvastatin, 40 mg/d, compared with simvastatin monotherapy (P = .02).*

(7.12) Hypoxia induced ***a <u>significant</u> increase in*** *ATF6LD-Cluc secretion, which was completely prevented by pretreatment with azoramide (Fig. 3C).* (15*Science*1)

Note that the noun in the frame in the three examples above is already semantically related to quantification, but that is not always the case. The frame *a * risk of*, for example, also involved adjective fillers from different semantic categories, namely, EVALUATIVE *(e.g., substantial, relative)* and SIZE (e.g., *high, low, greater*).

The second pattern of frames expressing quantification (*Determiner/Noun/Pronoun-initial frame + noun filler*) involves nouns belonging to the QUANTITY semantic category. For all analyzed sequences used for quantification, the adjective already present in the frame belongs to one of the three semantic categories above, mostly EVALUATIVE and SIZE. Examples (7.13) and (7.14) show noun fillers belonging to the QUANTITY semantic category in the frames containing an evaluative attributive adjective and a size attributive adjective, respectively.

(7.13) *ABI chromatograms were analyzed via the PolySNP PERL script, and* ***the relative*** ***proportion of*** *each virus in the dual infection was calculated by averaging the proportions of all valid single-nucleotide polymorphisms (SNPs).* (11*Science*1)

(7.14) *In Baku, Azerbaijan, prisoners were enrolled on arrival at a tuberculosis screening and treatment facility, which reports* ***a high rate of*** *multidrug resistance (25%) among patients with tuberculosis and a rate of HIV coinfection of approximately 6%.* (10*NEJM*1).

One last function served relatively frequently by content word-based frames is location in time/place. Analyzed sequences serving this function are primarily preposition-initial frames starting all with *in the.* Fillers in these frames are either adjectives or nouns occurring in the following three patterns:

- *in the …-frame + adjective (*RELATIONAL or EVALUATIVE*)* → location in place,
- *in the …- frame + noun (*TECHNICAL, ABSTRACT/Other, or ABSTRACT/Process) → location in place, and
- *in the …-frame + adjective* (SIZE) → location in time.

The only identified frame starting with *in the* and expressing location in time was *in the * term* and involved only size attributive adjective fillers. On the other hand, frames expressing location in place can involve fillers from more than one of the identified semantic categories. The frame *in the * group*, for example, is particularly productive and was used in the MRAC with both noun and adjective fillers from almost all identified semantic categories (e.g., *aspirin, insulin, placebo, control, intervention, invasive, aggressive, intensive*). Examples (7.15) and (7.16) show this frame used with both noun and adjective fillers from some of the identified semantic categories.

(7.15) *A total of 34 patients* ***in the <u>aspirin</u> group*** *and 38 patients* ***in the <u>nonaspirin</u> group*** *died from any cause (HR, 0.90; 95% CI, 0.57-1.14; log-rank test, P = .67)*. (16*JAMA*2)

(7.16) *The number of patients admitted with heart failure during follow up was 76 (mean 0·20 admissions; range 0–7)* ***in the <u>invasive</u> group.*** (20*Lancet*5).

In fact, the frame *in the * group* appeared to serve the same function regardless of the filler word class. It appears that it is the elements of the frames that already determine the function of the frame. The sequence *in the group* in itself already expresses location in place and any filler that goes in the variant slot of the frame *in the * group* will just specify the type of group. This may be another explanation of Garner's (2016) findings of frames serving the same functions regardless of fillers. The functions of some frames appear to be predetermined by the elements of the frames. This phenomenon was observed more frequently with the lexical verb-initial subgroup of verb-related frames to which I now turn.

7.4.1.2 Functions of verb-based frames.

The functions served by verb-based frames in the MRAC are presented in Table 7.9, and the subgroups primarily serving these, and other identified functions are shown in Figure 7.17. Here also, Description is the most frequently served function, closely followed by Procedure. Other functions served by verb-based frames include resultative signals, structuring signals (text reference), and engagement features. Examples of frames serving each of these functions are provided on Table 7.8. Only frames serving the two most frequent functions (description and procedure) are discussed in this section.

Verb-based frames used for description are primarily copular be-initial occurring exclusively with adjective fillers. Auxiliary-initial frames with communicative verb fillers were also frequently used in the MRAC for description. These sequences follow the five patterns listed below:

- *copular be-initial frame* + SIZE *adjective filler*,
- *copular be-initial frame* + RELATIONAL *adjective filler*,
- *copular be-initial frame*+ EVALUATIVE *adjective filler*,
- *auxiliary-initial frame* + COMMUNICATION *verb filler*, and

  auxiliary-initial frame + MENTAL *verb filler*.

Table 7.9. Functions Served by 4-gram Verb-based Frames in the MRAC

Subgroups	**Filler word class**	**Filler sematic category**	***Function***	***Examples from the MRAC***
Auxiliary-initial	Verb	activity	Procedure	*were * in the (classified, conducted, cultured, deposited,) was * by the (designed, performed, monitored)*
		mental	Procedure	*was * with the (assessed, calculated, estimated, evaluated)*
		communication, metal	description	*have been * to, has been * to (shown, reported, proposed, demonstrated, found, hypothetized)*
		communication	structuring signals (text reference)	*is * in the (reported, shown, presented, indicated), has been * in (documented, described, shown)*
		occurrence	resultative signals	*did not * the (change, alter, increase, improve, reduce)*
	Adverb	style	resultative signals	*were * associated with (inversely, moderately, independently, strongly, significantly, consistently, commonly)*
copular be-initial	Adjective	size a	description	*were * in the, was * in the (higher, elevated, lower, increased, reduced, decreased)*
		relational	description	*was * in the, were * in the (similar, different), was * to that (comparable, similar, close, identical)*
		evaluative	description	*were * for the (eligible, responsible, essential), to be * for (effective, useful, necessary,)*
			engagement features	*it is * that (crucial, important, essential, better)*
Other Lexical verbs (Mental verbs)	Noun	Abstract/attribute	procedure	*assess the * of (ability, applicability, impact, effect)*

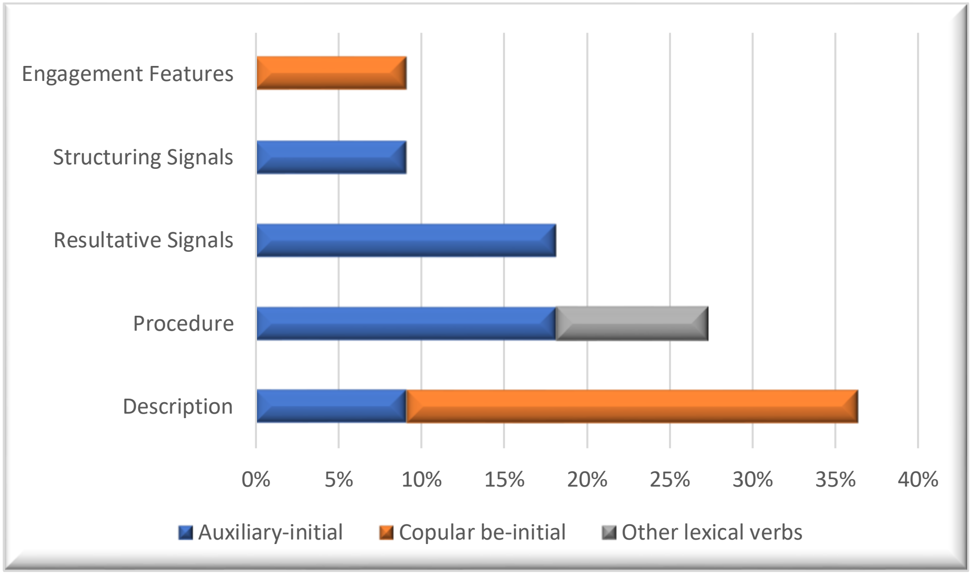


Figure 7.17. Proportions of Verb-based Subgroups of Frames Serving the Identified Functions in the MRAC

Similar to frames described above, single copular be-initial sequences can involve fillers from more than one semantic category as shown in examples (7.17) and (7.18) with the frame *to be * to*, used with adjective fillers from the RELATIONAL and EVALUATIVE semantic categories, respectively.

(7.17) Cardiogenic shock was defined as a sustained (>30-minute) episode of systolic blood pressure less than 90 mm Hg and/or a cardiac index less than 2.2 L/min/m^2 *determined* ***to be*** ***secondary*** ***to*** *cardiac dysfunction, […] to maintain blood pressure and cardiac index above those levels.* (20*JAMA*5)

(7.18) Pelvic floor muscle training (PFMT) has been reported ***to be*** ***effective*** ***to*** *cure or improve urinary incontinence symptoms in young, middle-aged, and older women with stress or mixed urinary incontinence.* (20*JAMA*1)

The same holds for fillers of auxiliary-initial frames. Examples (7.19) and (7.20) show the same frame (*have been * to*) used with verb fillers from the COMMUNICATION (7.19) and MENTAL (7.20) semantic categories to include findings of previous research in the description of elements of the study. Auxiliary-initial frames used for descriptions appear to be all related to passive construction bundles that were found to be still relatively frequent in medical texts despite their reported decline in scientific writing, as described in chapter 6. This suggests that if taught with their semantic groups of fillers, these frames can help novice L1 and L2-English medical writers make more creative use of passive expressions in situations where the passive is preferrable to the active voice.

(7.19) Systemic and topical glucocorticoids ***have been <u>reported</u> to*** *reduce symptoms, but their toxicity limits long-term use.* (20*NEJM*2)

(7.20) In most studies, Th17 cells ***have been <u>found</u> to*** *play a proatherogenic role, although some other studies came to a different conclusion (38–40).* (12*JCI*2)

Procedure is expressed by frames in the auxiliary-initial and other lexical verb-initial subgroups in the following patterns:

- *auxiliary-initial frame* + ACTIVITY *verb filler,*
- *auxiliary-initial frame* + MENTAL *verb filler,* and
- MENTAL *lexical verb-initial frame* + ABSTRACT/Attribute *noun filler*

I will start with lexical verb-initial frames as they are different from other verb-based frames. These sequences belong to the category of frames previously described, where elements of the frame seem to predetermine the function of the frame. Here, it is the semantic category of the lexical verb at the beginning of the frame that appears to control the function of the frame. All lexical verbs at the beginning of the analyzed verb-based frames, and expressing procedure belong to the MENTAL semantic category (e.g., *assess, analyze, determine, evaluate, measure, test*). The noun fillers occurring frequently enough (at least 5 times in the same frame) to be included in the analysis happen to all belong to the same semantic category, but most of these frames occurred with other less frequent noun fillers from other semantic categories (e.g., *number, length, subphenotype, diet, percentage*). Yet, the expressions formed by the frames and these less frequent fillers still served the same function of expressing procedure. To illustrate this characteristic of lexical verb-initial frames, examples (7.21) and (7.22) show the frame *determine the * of* , used with both a frequent noun filler in the ABSTRACT/Attribute semantic category (7.21) and a less frequent belonging to the QUANTITY semantic category (7.22). In both examples, the frame and its filler serve the same function of describing procedure.

(7.21) To ***determine the <u>susceptibility</u> of*** *genetic variation for this phenotype, we backcrossed mkp-1–/– mice on a 129/J/C57BL6/J background on to a pure C57BL6/J background for 8 generations, and the body weights of the C57BL6/J mkp-1–/– mice were analyzed over 6 months.* (09*JCI*1)

(7.22) After fixation and mounting, Fucci and EdU fluorescent signals were quantified only in GFP- and GFP-RIP140–positive cells in order to ***determine the <u>percentage</u> of*** *cells in G1 phase (Fucci-positive) and in S phase (EdU-positive).* (14*JCI*4)

As mentioned above, frames of this types are frequent among verb-based sequences. This finding can also have important pedagogical implications. In addition to semantic categories of fillers that occur with specific structural subgroups of frames, identifying frames that serve only one function, regardless of their fillers, can provide novice and L2-English writers with more tools in the production of texts that reflect the writing practices in their specific fields.

As for the two other patterns involving auxiliary-initial frames and expressing procedure, they function similarly to other patterns previously described. Verb fillers from more than one identified semantic category can occur in the same frame as shown in the two examples below with the frame *were * in the*. The fillers in (7.23) and (7.24) are ACTIVITY and MENTAL verbs, respectively.

(7.23) GFP-tagged ANXA11 was transiently expressed with mCherry-tagged TIA1 as an SG marker in HeLa cells, which ***were <u>cultured</u> in the*** *absence or presence of various SG inducers.*(20*Science*9)

(7.24) Patients who received invasive management after 3 days ***were <u>analyzed</u> in the*** *non-invasive group, as would be the case in an intention-to-treat analysis of a randomized trial.* (20*Lancet*5).

Like the auxiliary-initial frames used for description and previously described, these sequences are related to passive-construction bundles and therefore, may be of interest in medical writing instruction. After the discussion of the main functions served by verb-related frames in medical research articles, I now turn to the last structural category, function word-based frames, and the many functions they serve in this register.

7.4.1.3 Functions of function word-based frames.

Functions word-based frames have been reported to serve a wide range of functions in academic writing (Gray & Biber, 2015; Geluso, 2019), and the findings of the functional analysis of this category of frames in this study are consistent with that claim. Very often, the same frames serve multiple functions depending on the semantic categories of their fillers. Table 7.10 presents all the functions served by the analyzed 4-gram function word-based frames in the MRAC, and the subgroups primarily serving each function are shown in Figure 7.18.

Table 7.10. Functions Served by Function Word-based Frames in the MRAC

Subgroups	Filler word class	Filler sematic category	Function	
Determiner-initial	Noun	technical/ concrete	description	*the * of the (immunogenicity, phenotype, onset, genotype)*
		abstract/ attribute	description	*the * of the (characteristics, ability, specificity, capacity, integrity)*
		abstract/other	description	*the * of the (basis, control, risk, direction, objectives, choice)*
		quantity	quantification	*a * of the (proportion, ratio, number, part, majority, fraction)*
		quantity	location in time	*the * of the (course, date, length, peak)*
		place	location in place	*the * of the (bottom, surface, head, area, center, region, vicinity)*
		abstract/process	Procedure	*the* of the (use, design, analysis, implementation, neutralization)*
preposition-initial				
at a *of	Noun	quantity	quantification	*at a * of (dose, concentration, median density, rate, ratio)*
at the * of	Noun	quantity	location in time	*at the * of (time, end, start, beginning, midpoint, day, date)*
		place, institution	location in place	*at the * of (university, site, bottom)*
during the * of	Noun	quantity	location in time	*during the * of (course, period, phase, time)*
for * in the	Noun	animate	framing signals	*for * in the (patients, participants, women, children)*
for * of the	Noun	abstract/process	Procedure	*for * of the (establishment, inhibition, estimation, multiplexing)*
for a * of	Noun	quantity	quantification	*for a * of (maximum, mean, range, median, minimum, number, total)*
for the * of	Noun	abstract/process	Procedure	*for the * of (analysis, alignment, addition, assessment, calculation)*
	Noun	quantity	quantification	*for the * of (majority, number, percentage, range, ratio)*
	Noun	quantity	location in time	*for the * of (remainder, duration, entirety)*
	Noun	abstract/ attribute	description	*for the * of (ability, criterion, effectiveness, covariates)*
from the * of	Noun	quantity	location in time	*from the * of (date, time, middle, start)*
	Noun	technical, Place	location in place	*from the * of (blood, aortas, hearts, lungs, pancreas, vicinity, interior)*
in * of the	Noun	abstract/other	framing signals	*in * of the (view, light, terms, case)*
in the * of	Noun	abstract/other	framing signals	*in the * of (case, context, presence)*
	Noun	quantity	quantification	*in the * of (number, rate, percentage, proportion, range, rates, levels, size)*
	Noun	abstract/process	Procedure	*in the * of (development, regulation, treatment, use)*
	Noun	technical	location in place	*in the * of (feces, blood, hippocampus lungs, plasma, serum)*

Table 7.10 (Continued)

	Filler word class	Filler sematic category	Function	
in the * that	Noun	technical animate, abstract/other	framing signal	*in the * that (clusters, genes, groups, cohorts, mice, studies, literature, study)*
in * with a	Noun	place	location in place	*in * with a (countries, areas, households, sites)*
	Noun	animate, abstract/other	framing signals	*in * with a (women, participants, patients, accordance, agreement)*
in * with the	Noun	abstract/other	framing signals	*in * with the (accord, accordance, agreement, compliance, line)*
in * to the	Noun	abstract/process	transition signals	*in * to the (addition, contrast, response, comparison)*
on the * of	Noun	quantity	quantification	*on the * of (number, proportion, rate, percentage amount)*
over the * of	Noun	quantity	location in time	*over the * of (course, duration, length, lifetime, period)*
with a * of	Noun	quantity	quantification	*with a * of (range, score, prevalence, total, median, series)*
	Noun	technical	description	*with a * of (history, bmi, diagnosis, power)*
with the * of	Noun	abstract/process	Procedure	*with the * of (use, combination, support, approval, development, implementation, addition)*
	Noun	abstract/ attribute	description	*with the * of (risk, presence, goal, lack, effects, standards, loss, results, findings)*
	Noun	quantity	quantification	*with the * of (number, rest, occurrence, level, extent)*
to-infinitive initial	Verb	mental	Procedure	*to * whether the (test, determine, assess, verify, investigate)*
pronoun-initial	Verb	mental	resultative signals	*we * that the (found, demonstrated, determined, estimated, hypothesized, believe, observed)*
	Verb	communication	engagement features	*we * that the (speculate, stress, caution, argue)*
that-initial	Verb	existence, occurrence	description	*that * in the (occur, reside, were, results)*

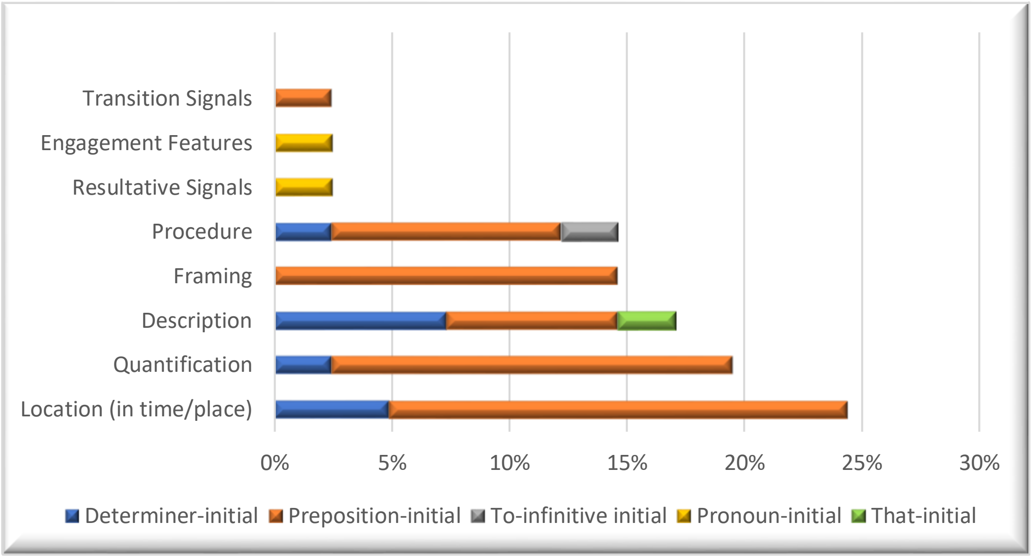


Figure 7.18. Proportions of Function Word-based Subgroups of Frames Serving the Identified Functions in the MRAC

As previously shown by Gray & Biber (2013) and Geluso (2019), function word-based frames are predominantly preposition-initial and primarily have noun fillers. Determiner-initial frames also exclusively involved noun fillers, as should logically be expected, given the frame structures. On the other hand, to-infinitive-initial, pronoun-initial and that-initial sequences all involve verb fillers. As shown in Figure 7.18, preposition-initial sequences serve the majority of identified functions. Geluso (2019) grouped all frames starting with prepositions and rooted in the frame *the * of* (e.g., *in the * of, with the ** of) into one "family" (p.123), but in the present study, I elected to list all preposition-initial frames individually. The rationale here is that while frames in the same family may involve the same word class of fillers, they do not necessarily serve the same functions when associated with a given semantic category of fillers.

For instance, the frames *over the * of* and *on the * of* both belong to the same family according to Geluso's (2019) classification, and indeed, they both occur with noun fillers belonging to the QUANTITY semantic category. Yet, the pattern *over the * of + noun filler (QUANTITY)* was found to express location in time, with fillers like *course, period, duration,* and *lifetime*; while the pattern *on the * of + noun filler (QUANTITY)* expresses quantification, with fillers like *number, proportion, rate,* and *amount.* Given that the functional analysis of frames was done with pedagogical applications in mind, the decision was made to simply list all preposition-initial frames occurring in the MRAC and identify the patterns in which they serve specific functions. This approach appears more straightforward and therefore, can be presumed to be more teachable and easier to digest for learners.

The most frequent functions are location (in time /place) and quantification, and both are primarily served by preposition-initial frames, with only a few instances of determiner-initial sequences. On the other hand, the third most frequent function, description, is equally served by both preposition-initial and determiner-initial frames. The only that-initial 4-gram frame identified in the MRAC is also used for description. Other discourse functions of function word-based frames in the MRAC include framing signals, procedure, resultative signals, engagement features, and transition signals. Examples of frames and their semantic groups of fillers serving each of these functions are provided in Table 7.10 above, and only the two most frequent functions are discussed in this section, starting with frames expressing location in time and place below.

As previously mentioned, frames used to specify location in time and place are primarily from the preposition-initial subgroup. These sequences all involve noun fillers belonging to the QUANTITY semantic category. Six frames were found to express location in time when associated with their fillers: the determiner initial *the * of the* and five preposition-initial frames, including *the * of the* (*at the * of, during the * of, for the * of, from the * of,* and *over the * of*). Frames expressing location in place involve fillers in the PLACE, INSTITUTION, and

TECHINICAL semantic categories. Four preposition-initial frames and the determiner-initial *the * of the* express location in place in the MRAC in these five patterns:

- *at the * of + noun filler* (PLACE, INSTITUTION)
- *from the + of + noun filler* (PLACE, TECHNICAL)
- *in the * of + noun filler* (TECHNICAL)
- *in * with a + noun filler* (PLACE, TECHNICAL)
- *the * of the + noun filler* (PLACE)

Examples of all frames in these patterns can be found in Table 7.10 above. The excerpts in (7.25) – (7.27) below provide examples of frames and their fillers, in context, expressing location in time (7.25) and place (7.26 and 7.27).

(7.25) ***At the <u>beginning</u> of*** *April, 2009, the medical care units of the Mexican Institute for Social Security (Instituto Mexicano del Seguro Social, IMSS) were alerted because […].* (20*Lancet*2)

(7.26) Pre-entry screening, when done ***in <u>countries</u> with a*** *prevalence of tuberculosis greater than 350 per 100 000 population seems to be within a similar range as these upon-entry and post-entry programmes*. (14*Lancet*1)

(7.27) We carried out a surface sensing of translation (SUnSET) assay and examined protein translation ***in the <u>fibroblasts</u> of*** *patients carrying the ALS-linked missense variants of ANXA11*. (20*Science*9).

Frames expressing quantification are also predominantly preposition-initial with one instance of determiner-initial, namely, the sequence *a * of the.* These frames have the peculiarity of exclusively involving noun fillers in the QUANTITY semantic category. Eight frames were found to express quantification when associated with their fillers: *a * of the, at a * of, for a * of,* *for the * of*, *in the * of*, *on the * of*, *with the * of*, *with a * of*. What transpires here is that function word-based frames expressing quantification and location in time tend to involve only noun fillers in the QUANTITY semantic category. However, one difference to note is that pairs of frames that appear to be similar (e.g., *with the * of / with a * of*, or *for a * of / for the * of*) do not necessarily involve the same fillers, even though nouns filling the variable slots all belong to the same semantic category (see Table 7.10, above). Figure 7.19 shows instances of the use of the frames *with a * of* and *with the * of* with their fillers, in context.

Figure 7.19. *Differences in Fillers and Co-texts between two Apparently Similar Frames Serving the Same Function*

Note the difference between the left co-texts of the two frames. There appears to always be a comparison between the quantities of two elements of the study when the frame involving the definite article (*with the * of*) is used. The discussion of the use of articles is beyond the scope of the present study, but this may be of relevance in writing instruction as articles have been reported to be problematic for L2 learners (Shin & Kim, 2017). The data shown in Figure 7.19 also indicates that, as already mentioned, the analysis of the textual environments of sequences can contribute to the description of formulaic language in various registers.

In sum, the findings presented in this section have shown that 4-gram frames in the MRAC serve primarily research-oriented functions, namely, description, quantification, location in time and place, and procedure. This is consistent with the findings on lexical bundle use reported in chapter 6 that MRAC bundles are primarily research oriented. After all, frames are closely related to lexical bundles. Very often, expressions formed by frames and their high frequency fillers are those identified as lexical bundles (Gray & Biber, 2013). This highlights the pedagogical value of the comprehensive analysis conducted in this section that allowed a systematic investigation of frames in all structural categories. The combination of frames and their semantic categories of fillers to serve specific discourse functions can provide learners with more choices in the paradigmatic axis of important expressions like lexical bundles. In turn, those multiple choices may help novice and L2 English writers produce more expert-like texts, as research has shown that at higher level of proficiency, writers rely less on fixed multiword units and show a more creative use of existing patterns like frames (e.g., Garner, 2016; Staples et al., 2013).

Beyond the description of the functions of frames, the functional analysis conducted in this section has helped test out the three proposed classification frameworks. As described in section 3.6.3 of the Methodology chapter, these frameworks are designed to allow a fully exploratory approach to the functional analysis of frames. With the frames as observations, the subgroups in each structural category can be adapted or adopted depending on the patterns observed on the lists of frames under scrutiny. Then the tables are populated as the researcher discovers the functions of the frames. The analyses conducted in this section have shown that these frameworks can be used for a systematic analysis of frames and their groups of semantic fillers. The verb-based framework was used for the analysis of verb-based frames in the MCRC, the findings of which are reported in the following section.

7.4.2 Functions of Verb-based Frames in Medical Case Reports

The functional analysis of the MCRC frames focused exclusively on the verb-based sequences in the 4-gram list for two main reasons. First, as previously mentioned in section 7.2., this list was found to include a large number of sequences related to bundles already described in Chapter 6 as serving discourse functions specific to medical case reports. The second reason was simply to make the data manageable. As rightly noted by Gray & Biber (2013), echoing Stubb (2007), one way of going around the challenging interpretation of large amounts of phraseological data is to “start small” (p. 111). In total, there were 740 4-gram frames, and manually classifying and analyzing the functions of each frame with its semantic categories of fillers was beyond the timeframe imparted for the completion of the present study. Nevertheless, after cleaning the list of 4-gram verb-based frames of all high predictability frames and sequences with non-lexical word fillers, there still remained 153 sequences to analyze. The functions served by 4-gram verb-based frames are listed in Table 7.11, and Figure 7.20 shows the subgroups primarily serving each of these functions.

The first observation indicating some difference between verb-based frames in the two registers is the number of subgroups listed in Table 7.11. In addition to the three structural sub-categories of verb-based frames identified in the MRAC list (*Auxiliary-initial, Copular Be-initial*, and *Other Lexical Verbs*), two other subgroups were identified in the list of MCRC verb-based 4-gram frames, namely, *Determiner/Noun/Pronoun-initial* and *Modal-initial.* The two most frequently served discourse functions, result reporting and decision/outcome, belong to the Diagnosis & Intervention-related category and are served primarily by auxiliary-initial and determiner/noun/pronoun-initial verb-based frames.

Table 7.11. Functions Served by 4-gram Verb-based Frames in the MCRC

Subgroups	Filler word class	Filler sematic category	Function	examples
Auxiliary-initial				
are * …	Verb	communication	Structuring signals (text reference)	_are * in the (described, presented, shown, listed)_
has/have been * + prep	Verb	communication, mental	Description (medical condition-related)	has been * to _(reported, shown, hypothesized, postulated)_
		Activity	Procedure	have been * with (_used, treated, performed_)
had been * with	Verb	Activity	Description (subject-related)	had been * with (_treated, vaccinated, self-medicating_)
is * + verb	Adverb	Likelihood	Description (medical condition-related)	is * seen in (_typically, commonly, usually, mostly_)
	Verb	Mental	Description (medical condition-related)	is * to be (_known, thought, estimated)_
did not * …	Verb	Existence	Description (subject-related)	did not * any (_have, experience, develop_)
	Verb	communication, mental	Result reporting	did not * the (_show, reveal, observe, indicate_)
	Verb	Activity	Procedure	was * from the (_extracted, isolated, taken, removed_)
		Causative, Communication	Decisions/Outcome	was * from the (released, discharged, referred)
was * + main verb	Adverb	Manner, Time	Decisions/Outcome	was * treated with (successfully, appropriately, initially, immediately, subsequently)
were * + prep	Verb	Mental	Result reporting	were * in the (_found, noted, detected, observed_)
copular be-initial				
is a * + noun	Adjective	Evaluative, Relational, Topical	Description (medical condition-related)	is a * condition (_rare, benign, pathological, medica, life-threatening_)
was * + prep	Adjective	Evaluative, relational	Result reporting	was * for the (_positive, negative, unusual, typical, notable_)

Table 7.11 (cont'd)

Subgroups	Filler word class	Filler sematic category	Function	examples
Determiner/Noun/Pronoun-initial				
determiner-initial	Noun	Animate	Description (subject-related)	the * was a (*patient, child, proband*)
		Technical	Description (medical condition-related)	the left * was (*pupil, cornea, adrenal, eye, breast*)
he/she-initial	Verb	Aspectual	Procedure	he was * on (*started, commenced, kept*)
it is …	Adjective	Evaluative	Engagement features	it is * for (*important, crucial, paramount, critical*)
it was…	Verb	Intention	Decisions/Outcome	it was * that (*decided, planned)*
		Mental	Result reporting	it was * that (*noted, observed, found, discovered*)
noun-initial	Verb	Activity, Causative, Communication	Decisions/Outcome	patient was * to (*admitted, released, referred, readmitted, transferred*)
	noun	Technical	Result reporting	serum * level was (*potassium, calcium, cortisol, thyroglobulin*)
	Adjective	Evaluative, Relational	Result reporting	cells were * for (*positive, negative, immunopositive, immunonegative*)
Modal-initial				
may/can/could be …	Adjective	Evaluative	stance	may be * for (*responsible, useful, helpful, beneficial*)
	Verb	Mental	stance	can be * as (*considered, viewed, regarded*)
should be…	Verb	Mental, Aspectual, Activity	Engagement features	should be * in (*considered, avoided, kept, included, performed*)
Other lexical verbs				
communication verbs	Noun	Multiple semantic categories	Result reporting	revealed the * of (*presence, proliferation, volume, co-existence, integrity, regression)*
Mental verbs	Noun	Abstract/ Attribute	Procedure	assess the * of (*cause, impact, risk, effectiveness, efficacy)*

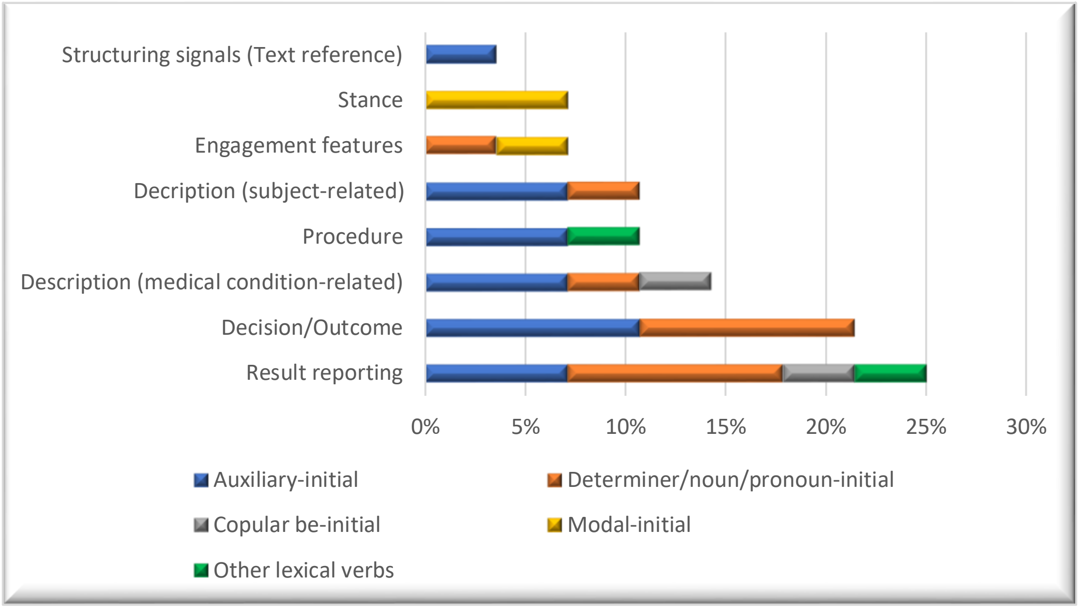


Figure 7.20. Subgroups of Verb-based Frames Serving each of the Identified Functions in the MCRC

Together with Procedure, these two functions are key in the longest section of case reports that provides information about the clinical findings, diagnoses, and interventions. As explained in the situational analysis of MCRs (see section 4.3.5.3.), this information can be under one, two or three headings, depending on journal guidelines. Under one heading, it is simply labelled *Case Presentation*, which I will use hereafter. Regardless of the headings, the information in the case presentation is presented following the problem-solution patterns described in section 4.3. with the recycling trend of solutions and negative evaluations until a successful solution is found. It is therefore unsurprising that Result reporting and Decision/ outcome were found to be the most frequently served discourse functions. The discourse function of procedure was found to be not as frequent as expected, but still, about 12% of the analyzed verb-based frames were used to serve this function. It may be possible that this function is more frequently served by frames in other structural categories (content word-based and function word-based).

MCRC verb-based auxiliary-initial frames used to report results exclusively start with *did not* and *were* and involve verb fillers in the COMMUNICATION and MENTAL semantic categories. Copular *be*-initial sequences serving this function involve EVALUATIVE and RELATIONAL adjective fillers. Frames in the Determiner/Pronoun/Noun subcategory are either an extraposed construction with the 'dummy' *it,* or noun-initial and involve MENTAL verb and TECHNICAL noun fillers, respectively. Finally, for other lexical verb-initial frames used to report results, the function of the frame is predetermined by the semantic category of the verb in the frame rather than the fillers. Lexical verbs in all analyzed sequences serving the result reporting discourse function belong to the MENTAL semantic category and involved noun fillers from various semantic domains. In sum, the following six patterns were identified for frames used to report diagnosis and intervention results.

- *Auxiliary-initial frame (did not * ...) + verb filler* (COMMUNICATION, MENTAL)
- *Auxiliary-initial frame (were * ...) + verb filler* (MENTAL)
- *Copular be-initial frame (was * ...) + adjective filler* (EVALUATIVE, RELATIONAL)
- *Determiner-initial frame ('dummy' it) + verb filler* (MENTAL)
- *Noun-initial frame + noun filler* (TECHNICAL)
- MENTAL *lexical verb-initial + noun fillers (various semantic categories)*

Table 7.11 above provides examples of frames and their fillers for each of these patterns. As was observed in section 7.4.1 above, for frames involving fillers from different semantic categories, the same frame can involve fillers from more than one category, as shown in examples (7.28) and (7.29) with the frame *did not * any* used with mental and communication verbs, respectively.

(7.28) Extensive workup ***did not <u>find</u> any*** *evidence of an underlying solid tumor or lymphoproliferative disorder.* (13*JAMA*67)

(7.29) A 2-week cardiac event monitor ***did not <u>reveal</u> any*** *evidence of arrhythmias or pathologic block at rest or with activity.* (20*JAMA*89)

Note the use of medical procedures as agents in both examples above. This is a very common practice in case reports and may deserve some attention in medical writing instruction.

Verb-based frames serving the second most frequent function, Decision/Outcome, are exclusively auxiliary-initial and determiner/pronoun/noun-initial and primarily express intervention decisions. Auxiliary-initial sequences serving this function involve verb fillers in the CAUSATIVE and COMMUNICATION semantic categories, and TIME and MANNER adverb fillers. Frames involving MANNER adverb fillers are the only sequences found to express outcome as shown in example (30) below, with the frame *was * treated with.*

(7.30) We decided to perform an endovascular repair and subsequentially the patient ***was <u>successfully</u> treated with*** *stent graft deployment, showing durable early-midterm results*. (21*IMCRJ*104)

Determiner/Pronoun/Noun-initial sequences used to report decisions are either noun-initial or pronoun-initial, with the third-person pronouns *she/he* or *it* in an extraposed construction. *She/*he-initial frames involve ASPECTUAL verb filler, while fillers in *it*-initial sequences are verbs belonging to the MENTAL semantic category. Fillers in noun-initial sequences are verbs in the ACTIVITY, CAUSATIVE, or COMMUNICATION semantic categories. In total, five patterns were identified for frames used to report decisions made during clinical intervention:

- *auxiliary-initial frame (was * + prep) + verb filler* (CAUSATIVE, COMMUNICATION),

- *auxiliary-initial frame (was * + main verb) + adverb filler* (MANNER, TIME),
- *pronoun-initial frame (she/he) + verb filler* (ASPECTUAL),
- *pronoun-initial-initial frame (it) + verb filler* (INTENTION), and
- *noun-initial frame + verb filler* (ACTIVITY, CAUSATIVE, COMMUNICATION).

Examples (7.31) to (7.33) show the auxiliary-initial, pronoun-initial, and noun-initial frames used in context with their fillers to report clinical intervention decisions. The auxiliary-initial is used with a communication verb filler (7.31), and the pronoun and noun-initial frames in (7.32) and (7.33) are used with verb fillers belonging to the INTENTION and ACTIVITY semantic categories, respectively.

(7.31) After necessary investigations, due to presence of simultaneous intrahepatic and porcelain gallbladder, the surgeon decided to perform both cholecystectomy and hepatotomy to ensure the absence of malignancy. The patient ***was referred to a*** *better-equipped center to undergo the surgery*. (20*IMCRJ*5)

*(7.32) During surgery, approximation of the quadriceps tendon (still attached to the bipartite fragment) to the superior surface of the patella resulted in excessive tension in the quadriceps tendon with the risk of failure of repair in the postoperative period. Hence****, it was decided to*** *manage by open reduction and tension band wire fixation of the superior avulsed bony fragment to enable tension-free repair of the ruptured knee extensor mechanism.* (21*BMJCR*85)

(7.33) The clinical assessment was suspected COVID-19 infection complicated with cardiorespiratory failure and the ***patient was transported to*** *a tertiary level ICU.* (20*IMCRJ*5)

Note in all these examples the succession of negative evaluations and decisions illustrating the recycling pattern of problem-solution previously mentioned. This highlights the pedagogical value of these verb-based frames that occur in what can be considered the core section of medical case reports, the case presentation. As previously mentioned, in this section of MCRs, expressions serving the two functions discussed in the present section go together with sequences used to describe procedure. Frames serving the discourse function of Procedure are not the most frequent among the analyzed verb-based frames, but they still represent a non-negligible 12%. Another observation worth mentioning is the 12% of analyzed frames used to describe the case subject (the patient). Expressions serving this function primarily occur at the beginning of the case presentation, in the few first sentences serving as the introduction of the section. These functions were not discussed in the present section, but examples of frames used for these purposes are shown on Table 7.11 above. Given their importance in MCRs, these frames and their semantic groups of fillers may be excellent candidates for medical writing instruction.

To conclude this section on the functions of 4-word verb-based frames in the MCRC, I will say that there are two main takeaways from the findings discussed above. First, we have seen that verb-based frames primarily serve two functions that are key to medical case reports, namely, Result reporting and Decision/Outcome. These frames, when taught with their semantic groups of fillers can support novice and L2-English authors in the process of case report writing. The second takeaway is that the findings discussed above provide further evidence that portrays case reports as a distinct register from medical research articles. Of the four functions specific to the core section of case presentation, only one, procedure, was found to be served by MRAC frames, as explained in section 7.4.1 above. Again, this is evidence that overall descriptions of academic writing may need to be supplemented with more targeted descriptions of different registers within the same disciplines. This is one of the goals of the present study. To complete the comparison of the use of frames in the two registers, the functions of the shared frames presented in section 7.3 above are discussed in the next section.

7.4.3 Functions of Shared Frames

As explained in section 7.3, shared frames were classified into three groups based on their most frequent fillers. As only 4-gram frames were analyzed in both registers, the decision was also made to analyze only shared 4-gram frames. After removing all frames with high predictability and those with non-lexical word fillers from the original list of shared frames, 115 sequences remained for analysis. Of these 115 frames, 49 had completely different most frequent fillers, 41 shared the same most frequent fillers, and 22 had different fillers sharing the same semantic category. This last group, though smaller than the two others, is of much relevance to the approach used in the present study to analyze the functions of frames when associated with their semantic categories of fillers. It was expected that both sequences sharing the same most frequent fillers and those with most frequent fillers belonging to the same sematic category would serve similar function in the two registers. These expectations were borne out by the findings of the functional analysis of shared frames.

Frames that share the same most frequent fillers almost always serve the same functions. Many of these frames are related to familiar lexical bundles previously identified in academic writing. For example, the frame *on the * of* involves the bundle *on the basis of,* found to serve as framing signal in previous studies (e.g., Cortes, 2004, Hyland, 2008a) and in the two registers under scrutiny as well. But beyond this bundle, the frame was also found to express other same functions in both corpora, including quantification and location in time with TIME noun fillers, location in place with PLACE noun fillers, and procedure with ABSTRACT/PROCESS noun fillers. Examples (7.34) and (7.35) show this frame with two different fillers in the TIME semantic category, expressing location in time in the MCRC (7.34) and in the MRAC (7.35)

(7.34) ***On the <u>morning</u> of*** *POD 2, the patient again rated his NRS at 9. His PCA usage including denied attempts from overnight can be seen in Fig. 1.* (MCRC_20*OMCR*77)

(7.35) *K562 cells were counted and loaded onto a Cell-Tak-pretreated (1206D69, Corning) 96-well cell culture microplate* ***on the <u>day</u> of*** *assay following the manufacturer's instructions.* (MRAC_20*JCI*9)

Similar observations were made in the group of frames involving different fillers sharing the same semantic category. Frames in this group are primarily used for description of elements of the studies, and the differences in word choices may be due to a difference in authors' preferences, as described in section 7.3, or other factors not investigated in the present study. But overall, these frames serve the same functions in the two registers. Other than description, the other functions served by frames in this group include quantification, procedure, location in time, and resultative signals. Nine of the 22 shared frames used for description in this group are verb-based frames described in section 7.4.1 as related to passive constructions used to include previous research in the description of elements of the study or case. Excerpts (7.36) and (7.37) provide examples of the frame *have been * to,* used for this purpose in both corpora, with communication verb fillers.

(7.36) *Macroscopically, SCTATs* ***have been <u>reported</u> to*** *be solid, cystic, mixture of solid and cystic, yellow to tan, and some with haemorrhage and necrosis with size ranging up to 30 cm [4, 7].* (MCRC_20*OMCR*82)

(7.37) These productively infected macrophages and myeloid cells ***have been <u>shown</u> to*** *serve as a reservoir for SIV in ART-suppressed macaques and are associated with high levels of immune activation (40).* (MRAC_20*JCI*4)

Finally, even in the group of frames with different most frequent fillers, the shared sequences were found to still serve some similar functions. The difference in the most frequent fillers does indeed indicate differences in focus between the two registers, and this difference is also observed in some of the other fillers occurring at least 5 times in the same frame. For example, the frame *was * with a* has *associated* and *treated* as its most frequent fillers in the MRAC and the MCRC, respectively. Recurrent fillers in the MCRC (e.g., *diagnosed, repaired, performed*), together with *treated,* are all activity verbs and occur in the variable slot of the frame to refer to procedures. The frame is also associated in the MCRC with causative and communication verb fillers (*admitted, discharged, referred, readmitted*) to express Decision/Outcome. However, this does not preclude the occurrence of the filler *associated* (9 times) in the MCRC, and the expression *was associated with* serves the same function of resultative signal in both corpora, as shown in examples (7.38) and (7.39) below.

(7.38) In a registry of 17 312 adults with hypertension, nondipping ***was <u>associated</u> with a*** *27% higher risk of cardiovascular events.*[10] *(*MCRC_18*JAMA*181).

(7.39) Second, the group primed with ChAd3-NSmut vector boosted better than the group primed with Ad6-NSmut: This ***was <u>associated</u> with a*** *slightly higher number of pre-boost T cells in the former group.* (MRAC_12*Science*5)

Similarly, the same frame in the MRAC also combines with activity verbs to describe procedure, just as in the MCRC. This finding suggests that frames in this group do serve some similar functions across the two registers, even though there may be some additional functions specific to one register or the other.

This section completes the answer to RQ3b that asked about the functions served by some salient frames and their fillers in MRAs and MCRs and the potential variations between the two registers. But before summarizing the findings presented so far and answering all the questions, I now turn to the last section of this chapter for a discussion of the variability of the semantic categories of fillers occurring in the same frames and frames involving bundles.

7.5 Bundles in Frames and Variability of Semantic Groups within Frames

The rationale behind looking at frames involving bundles and the variability of semantic categories within frames was that if bundles predetermine the semantic categories of the other recurring fillers in frames involving bundles, then teaching the functions of bundles together with all the other fillers in the same semantic category would be beneficial for novice L1 and L2-English writers. The answer regarding the variability of semantic categories within the same frame has already been provided as a by-product of the functional analyses of 4-gram frames in the MRAC and the MCRC. In the findings reported in sections 7.4.1 and 7.4.2 above, we have seen that structural subgroups of frames combine with multiple semantic categories of fillers to serve the same function. We have also seen that very often, fillers from more than one semantic category could occur in the variable slot of a same frame.

As explained in section 3.6.4. of the methodology section, the lists of MRAC and MCRC 4-gram frames generated by *KfNgram* (Fletcher, 2007) were used to obtain the lists of frames involving bundles. Then the semantic categories of fillers forming the bundles were compared with the semantic domains of other recurrent fillers. Two trends were observed with frames involving bundles. In high frequency sequences like function word-based frames, the presence of one or several bundles in a frame does note predetermine the semantic categories of the other fillers. For example, the most frequent frame in both the MRAC and the MCRC (*the * of the*) involved nine bundles in both corpora. The fillers that form five of the MCRC bundles share the same sematic category of QUANTITY and four of the fillers of the MRAC bundles also belong to the QUANTITY semantic category. However, that did not predetermine the semantic domains of the other recurrent fillers. Some do belong to the same semantic categories as a filler that forms a bundle, but most of these other fillers belong to different other semantic categories, as shown in Table 7.12.

Table 7.12. Examples of Frames Involving Bundles and their Other recurrent Fillers

	4-gram frame	**Freq**	**Bundles /filler (token)**	**Some other frequent variants (freq>5)**
function word-based involving bundles	*the * of the*	1729	*the end of the* (128), *the use of the* (65), *the basis of the* (56), *the results of the* (51), *the effect of the* (38), *the course of the* (36), *the time of the* (27), *the design of the* (20)	*magnitude, duration, size, nature, context, analysis, level, start, discretion, mean, sum, role, characteristics, specificity, rest, effects, etc.*
content word-based involving bundles	*in the * group*	1416	*in the placebo group* (235), *in the control group* (199), *in the intervention group* (114)	*iron, early-therapy, screening, delayed-therapy, hypertonic saline, aspirin, rampiril, preterm, etc.*
verb-based involving one bundle	*to be * to*	93	*to be due to(21)*	*related, able, secondary*

Generally, content word-based and verb-based frames include less bundles, and most of the other fillers tend to share the same semantic category with one of the fillers forming a bundle (see table 7.12). A similar trend was observed with low frequency frames that do not involve bundles. The only difference is that most of those sequences often occur at an even lower frequency than frames involving bundles and include only a few fillers recurring at least five times. For example, the frame *a large * of* in the MCRC occurs 30 times and involves 16 fillers. The two most frequent fillers *amount* (6 times) and *number* (5 times) share the same semantic category. Some of the other fillers also belong to this semantic category but occur only once or twice in the frame (e.g., *size*, *volume*, *percent,* all occurring only once). These findings do not necessarily add much to the pedagogical implications and applications of the grouping of frame fillers into the semantic categories, but they help answer RQ 3c.

7.6 Conclusion

This chapter set out to answer the following research questions.

- RQ3a: How does the use of phrase frames compare in the two registers? Are there any variations in terms of predictability, variability, and structures?
- RQ 3b: How can a grouping by semantic categories of fillers inform the functional analysis of phrase frames? What are the main functions served by frames and their fillers in MRAs and MCRs? Are there any variations between the two registers?
- RQ 3c: Do semantic categories vary within one frame? Is there any variation between frames involving lexical bundles and those that do not?

To answer RQ 3a, I have presented the findings of the analysis of frames in the MRAC and the MCRC, respectively in sections 7.1. and 7.2. These findings have revealed that structurally, the selected frames in both registers consist predominantly of content word-based and verb-based sequences. Function word-based frames that have been previously reported as being the most frequent structural category of frames in academic writing are the least represented in the two registers under scrutiny in this study. It should be mentioned, however, that the exclusion of non-unit frames, explained in chapter 3, section 3.6.1, may have affected the proportion of function word-based frames in each corpus. But even then, the proportions of content-based frames identified in both corpora remain higher than what could be expected, based on Gray & Biber's (2013) description of frames in academic writing. Regarding variability and predictability, the majority of frames in both registers were found to be variable with low predictability. The structural similarities of frames in both corpora suggest that beyond these two registers, it is medical writing that slightly deviates from the phraseological patterning of academic writing. As already mentioned, these findings highlight the importance of supplementing overall descriptions of academic writing with more fine-grained portrayals of specific registers in specific academic disciplines to better serve writing instruction in English for Academic Purposes (EAP) and English for Specific Purposes (ESP).

The answer to RQ3b needs to be broken down into two parts. The first sub-question asked how a grouping of fillers by semantic categories could inform the functional analysis of phrase frames. As mentioned in section 7.4.1, the functional analysis of MRAC 4-gram frames in all three structural categories also served to try out the proposed functional classification frameworks and showed that the frameworks do allow a systematic functional classification of frames and their semantic groups of fillers. The subsequent functional analysis of MCRC and shared frames confirmed the usefulness of these frameworks. Indeed, the grouping of fillers by semantic categories allowed the identification of multiple functions served by each of the analyzed frames. The functional analysis of frames may be problematic as these sequences do not carry much meaning per se. But once combined with their fillers, they can be analyzed in context just like other continuous sequences such as lexical bundles or multiword collocations. In addition to the systematic analysis of frames, the grouping of fillers by semantic categories offers the potential pedagogical advantage of providing learners with more options they can draw from when producing academic texts in specific registers in their disciplines.

The second and third sub-questions of RQ 3b were concerned with the functions of frames in the MCRs and MRAs and potential variations between the two registers. The findings discussed in section 7.4.1 revealed that MRAC 4-gram frames and their fillers were primarily used for research-oriented functions. The most frequent function, description, was served by frames in all three structural categories. On the other hand, the analysis in section 7.4.2 of 4-gram verb-based frames in the MCRC revealed that these sequences predominantly serve functions that are not only specific to case reports but occur in the most important section of MCRs. Even though only MCRC verb-based frames were analyzed, the stark difference between the functions of most of these frames and their counterparts in the MRAC further portrays case reports and medical research articles as two distinct registers from a phraseological perspective. The functional analysis of shared frames revealed similarities in the functions of frames in both registers, but given the relatively small proportion of shared frames, the differences between the two registers may be more prominent. The findings presented in sections 7.4.1 and 7.4.2 are further evidence that disciplinary writing is better served if within-discipline variations are investigated to supplement general descriptions of academic writing.

Finally, section 7.5 answered the pedagogically motivated RQ 3c. The findings discussed in other sections in this chapter had already revealed that most frames occur with fillers belonging to different semantic categories. The inspection of frames involving bundles lent support to these previous findings. High frequency frames often involve multiple bundles and fillers from these bundles do not necessarily share the same semantic categories with the rest of the fillers in the frame. Some of the recurrent fillers share the same semantic categories as one of the fillers that form bundles, but there are many other fillers that belong to different other categories. Overall, the answer to RQ 3c does not allow a systematic connection between fillers forming bundles and other recurrent fillers. Nevertheless, the functional classification framework proposed in this chapter and the grouping of frame fillers into semantic categories present potential pedagogical implications and applications that can be of interest in medical writing instruction. I will return to this point in the next and final chapter of this study.

8 CHAPTER 8: CONCLUSION

The main purposes of the present study were to investigate within-discipline linguistic variations and to propose a more comprehensive approach to the study of formulaic language in a register. To that end, I asked seven questions that helped me describe the situational characteristics and the formulaic profiles of medical research articles and medical case reports and discuss the observed variations between the two registers. The analyses of the formulaic sequences in the two registers often revealed structural similarities but noticeable variations in terms of the discourse functions of the sequences. Such variations reflected the differences noted in the situational characteristics of the two registers.

The comprehensive approach used in the study of the formulaic profiles of the two registers highlights the complementarity of the sequences investigated in the present study. As previously mentioned, lexical bundles constitute the building blocks of academic writing. The finding of the present study suggest that other formulaic sequences supplement these building blocks in many ways. The analyses of collocations and multiword collocations have revealed specialized expressions and sometimes ready-made phrases or clauses with potentially high pedagogical value in medical writing. The functional analyses of frames have shown that frames and their semantic groups of fillers serve the same discourse functions as lexical bundles identified in the present study. This suggests that, when taught with their semantic categories of fillers, frames can provide novice L1 and L2-English writers with more choices in the paradigmatic axis of important expressions like lexical bundles or multiword collocations.

As I provided extensive descriptions of the situational characteristics and the use of formulaic sequences in the two registers, in this final chapter, I only provide a summary of the most salient similarities and differences observed between the two registers. Then I discuss a few implications and applications, followed by the limitations of the present study and further directions for research.

8.1 Summary of the Situational Characteristics of MRAs and MCRs

In chapter 4, I answered RQ1 that asked about the situational characteristics of the two registers. Table 8.1. summarizes the most salient situational characteristics that portray the medical research article and the medical case report as two distinct registers within the medical field. The situational analyses of MCRs and MRAs revealed, as shown in Table 8.1., that both registers share the overarching communicative purpose of advancing medical research and practice and have medical pathologies and treatments as general topics. However, these similarities were found to be only at the surface level. In reality, the two registers differ fundamentally in their foci and specific communicative purposes. The focus of MRAs is on advancing medical research. This is what makes this register a "highly technical report of experiments" (Nwogu, 1997, p.119). Most of the content of MRAs is devoted to extensive descriptions of research methods and experiments and advanced statistical measures to prove the generalizability of results and ensure replicability of the research.

On the other hand, the primary focus of MCRs is on informing medical practice and education. To that end, authors of MCRs provide detailed descriptions of patients (usually only one), medical pathologies, new treatments adverse events, and outcomes of clinical interventions in an aim to provide "enough details […] for clinicians to relate in their own practice" (Rison et al., 2017, p. 1) and to "offer a structure for case-based learning in healthcare education" (Gagnier et al., 2014, p. 46). This difference in foci in the specific communicative purposes of MCRs and MRAs translates into distinct channels of production and production circumstances that explain the variations observed during the analyses of formulaic sequences in the two registers. Indeed, MRAs and MCRs fundamentally differ in their methodologies. As shown on Table 8.1., MRAs are primarily experimental with extensive use of advanced statistical techniques, while MCRs are typically observational and may occasionally use some descriptive statistics.

This difference in research design is reflected on how information is presented in the two registers. Experimental MRAs are written in the IMRD format with very detailed Methods and Results sections. As already mentioned, the extensive descriptions of methods and results are consistent with the specific purposes of MRAs to report generalizable results and replicable studies. On the other hand, case reports are typically written in three different formats, I/BCD RWDWD(P), and BC(I)TODL, but regardless of format, MCRs present similar information including diagnoses, clinical findings, interventions, outcomes, adverse events, and follow-up. The bulk of this information is provided in the Case Presentation, following a problem-solution pattern that reflects the steps and observations during the management of cases being presented. Consistent with the specific communication purposes of the register, MCR authors provide detailed descriptions of these steps and observations that can directly inform medical practice and education.

These distinct characteristics highlight the differences not only in the specific communicative purposes of the two registers, but also in their primary audiences. Both registers have the academic and medical communities as their general audience. However, MRAs are primarily addressed to researchers and to some extent, practitioners, hence the extensive descriptions of methods and results. On the other hand, MCRs have practitioners and medical students as their primary audience. As already mentioned, the detailed descriptions of case management steps and observations serve the double purposes of directly informing clinical practice and providing case-based material for medical education.

Table 8.1. Summary of Key Situational Characteristics of MRAs and MCRs

Characteristics	**MRAs**	**MCRs**
1. General Topic	Medical pathologies and treatments	
2. Communicative purposes		
2.1. General communicative Purposes	Advance medical research and inform clinical practice	
2.2. Specific communicative Purposes	Report generalizable results and/or replicable studies	- Directly inform clinical practice - Provide case-based material for medical education - Generate hypotheses for further research
3. Participants		
3.1. Number of authors	Multi-authored	Multi-authored
3.2. Professional training/Title	primarily written by researchers	Primarily written by clinical practitioners
3.3. Audience		
3.3.1. General audience	academic and medical communities (researchers, practitioners, students, and patients)	
3.3.2. Primary audience	Researchers, practitioners	Clinicians, medical students, residents, and fellows.
4. Production circumstances		
4.1. Methodology	Primarily experimental	Primarily observational
4.2. Statistical techniques	Use of descriptive statistics, statistical difference testing, and other advanced statistics	Occasional use of descriptive statistics
5. Channel		
5.1. Textual layout and organization		
5.1.1. Word count	Moderately long texts (*M* = 5,451.08, *SD* =1,800.33)	Very short texts (*M* = 1,518.15, *SD* = 651.88)
5.1.2. Sections/Organization	IMRD format with extensive descriptions of experimental procedures, statistical techniques & results, and research elements.	Three main formats (I/BCD RWDWD(P), and BC(I)TODL), but all providing similar information (diagnoses, clinical findings, interventions, outcomes, adverse events, and follow-up).

8.2 Collocations and Multiword Collocations in MRAs and MCRs

In chapter 5, I answered RQ 2a and RQ 2b that asked about the use of collocations and multiword collocations in the two registers, respectively. The analyses of collocations revealed some similarities in structures and clear differences in functions of collocations in the two registers. Collocations in both MRAs and MCRs are predominantly noun combinations (*N* + *N* and *Adj* + *N*). Such sequences primarily consist of technical and semi-technical words; the kind of "specialized terms" that Rezaeian (2015) identified as one of the main challenges in medical writing. This similarity in structures contrasts with the functions of collocations in the two registers. Consistent with findings of the situational analysis, collocations in MRAs and MCRs reflected the main foci of the two registers. MRA collocations were primarily research-oriented sequences used to refer to research procedures and elements, while the majority of MCR collocations were case-related, or diagnosis/intervention-related sequences used to describe the case and/or to refer to biological processes/elements and procedures during diagnosis or clinical intervention. This difference in functions was noted even among collocations shared by the two registers, with the same sequences serving different functions in MRAs and MCRs.

One other variation in the use of collocations in the two registers was the presence in the MCRC of adjective-controlled grammatical collocations in the form of superlatives (almost inexistent in the MRAC). Such sequences served to highlight the uniqueness of cases being reported, in response to journals' requirements of novelty and originality of cases that authors submit for publication.

For the answer to RQ 2b, the analysis of multiword collocations in the MRAs and MCRs revealed a trend similar to what was observed with collocations; that is, similarities in structures and differences in functions. Authors of both MRAs and MCRs resort primarily to noun premodification to compress information. As a result, multiword collocations in both registers were found to be primarily complex noun phrases. This linguistic feature appears to be frequently used by MCR and MRA writers for the purpose of concise writing given the amount of information they have to provide on the one hand, and the limited word count allowed by journals, on the other hand. The resulting multiword collocations, however, serve different functions in the two registers, highlighting the differences in the situational characteristics of MRAs and MCRs, as was the case with collocations. In addition to complex noun phrases, MCR multiword collocations included a noticeable proportion of declarative clauses/fragments that functioned exclusively as ready-made sequences for reporting the results of examination and diagnosis procedures (see section 4.2.2), a function that is specific to MCRs.

8.3 Lexical Bundles in MRAs and MCRs

Chapter 6 answered RQ 2c that asked about the use of lexical bundles in the two registers as well as how they compare to bundles previously identified in academic writing. Overall, the findings of the structural analysis of bundles in both MRAs and MCRs were in line with previous descriptions of these sequences in academic writing. Bundles in both registers were predominantly NP-related and PP-related. However, passive-related bundles were also found to be frequent in the corpora representing the two registers in spite of the growing recommendations from medical journals to “use active voice whenever possible”. The functional analyses revealed a different picture. While the functions of MRA bundles did not fundamentally differ from findings of previous research of bundles in academic writing, MCR bundles served several functions distinct from those of MRA bundles or other bundles previously investigated in academic writing studies. Bundles in MRAs were predominantly research-oriented, which is consistent with previous research of academic writing in hard sciences, given the focus on the description of research procedures and elements.

On the other hand, 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. But the most salient difference between MCRs and MRAs was the noticeable proportions of VP-related and clause-related bundles in MCRs. These structural categories of bundles are primarily used in MCRs to report results of diagnoses and interventions, to describe the patient, and to report the case outcome and decisions made during case management. These again, are functions specific to MCRs.

8.4 Frames in MRAs and MCRs

In chapter 7, I answered the final three research questions related to the structures, variability, and predictability of frames in the MRAs and MCRs (RQ 3a); the functions of some salient frames and their semantic groups of fillers (RQ 3b); and bundles in frames and the variability of semantic categories of fillers within frames (RQ 3c). The findings of the analyses of the structures, variability, and predictability of frames situated the two registers somewhere between academic writing and conversation. The structural analysis revealed function word-based frames, described by Gray & Biber (2013) as the predominant structural group in academic writing, were the least frequent sequences in both registers. On the other hand, content word-based frames, described as being characteristic of conversation, were found to be the most frequent structural group of frames in the MRAs and MCRs. As for the third structural group, verb-based frames, their occurrence in the two registers was more in line with previous descriptions of frames in academic writing. However, despite the high proportion of content word-based frames, which in conversation are relatively fixed, frames in both MRAs and MCRs were found to be primarily variable with low predictability, which corresponds to previous descriptions of frames in academic writing. These findings led to the conclusion that perhaps, beyond MCRs and MRAs, it is medical writing that displays characteristics of both conversation and academic writing.

Next to these shared linguistic characteristics between the two registers, the analysis of frames revealed some variations worth noting. Most shared frames had different most frequent fillers in MRAs and MCRs, reflecting the differences in foci between the two registers. Other frames had different fillers that shared the same semantic categories and appeared to serve the same functions in the two registers. Such sequences warrant further investigation to see whether some semantic prosody or colligations in the environments of the frames determine the choice of specific fillers in one register or the other.

To answer RQ 3b, I designed functional classification frameworks for the three structural categories of frames that I used to analyze the functions of selected frames and their semantic categories of fillers. The most frequently served functions by frames in the two registers did not differ from the most frequent discourse functions of lexical bundles in the respective registers. The same variations were observed with the majority of MCR verb-based frames serving functions specific to this register. A final note regarding the functional analysis of frames is that the designed frameworks indeed allow a systematic functional classification of frames and their semantic categories of fillers, which in turn can facilitate the teaching of these expressions.

Finally, the analysis of frames involving bundles yielded mixed findings. In some frames, the fillers that form the bundles share the same semantic categories with most other frequent fillers. But there were also many other frames that involved different semantic categories of fillers that were not necessarily related to the fillers that form the bundles. Nevertheless, in the first case, bundles can be taught together with the frames they occur in and the fillers sharing the same semantic categories to help learners and L2-English writer make a more creative use of formulaic language.

8.5 Summary of the Formulaic Profiles of MCRs and MRAs

As already mentioned, the proposed comprehensive approach to the study of formulaic language has highlighted the complementarity of the different formulaic sequence investigated in the present study. Biber (2009) rightly suggested that sequences in formulaic language exist on a continuum from expressions with relatively low frequency and strong mutual expectation (e.g., collocations and multiword collocations) to multiword formulaic sequences primarily characterized by their high frequency in discourse (e.g., lexical bundles, lexical frames). Therefore, the investigation of the formulaic profile of a register entails the analysis of the major sequences along the formulaic language continuum, which I attempted to accomplish by analyzing the use of collocations, multiword collocations, lexical bundles, and lexical frames in the MRAs and MCRs. Table 8.2. provides a summary of the formulaic profiles of MRAs and MCRs, with the major structural correlates of each type of formulaic sequence.

Sequences identified using a frequency-driven approach (i.e., lexical bundles and lexical frames) are those that served a wider range of discourse functions (previously summarized in Table 3.10 and 3.11 and discussed in chapters 6 and 7). As rightly noted by Cortes (2013), lexical bundles constitute “lexico-grammatical building blocks associated with the basic functions used to bind the text together” (p. 36). The findings of the functional analyses of subsets of frames in the two registers have shown that frames and their semantic categories of fillers serve similar functions as those served by lexical bundles, as also shown on Table 8.2.

Table 8.2. Summary of the Formulaic Profiles of MRAs and MCRs

Sequences	Predominant Structures: MRAs	Predominant Structures: MCRs	Major Functions of distinctive structures of MCR sequences
Collocations	- Noun combinations (e.g., *skeletal muscle, flow cytometry*) - Verb-controlled (e.g., *compensate for, filled with, resuspended in*)	- Noun combinations (e.g., *night sweats mycophenolate mofetil*) - Verb-controlled (e.g., *ruled out, consist of*) - **Adjective-controlled: mostly superlatives** (e.g., *the safest, the rarest)*	**Stance & Engagement** - Justify interventions - Highlight uniqueness/ importance of the case
Multiword Collocations	Phrasal sequences (Complex NPs) (e.g., *patients with acute heart failure, stem cell transplantation)*	- Phrasal sequences (Complex NPs) (e.g., *fine needle aspiration, heart failure with preserved ejection fraction)* - **Clausal sequences** (*function tests were normal, showed a white blood cell count of*)	**Diagnosis/intervention-related** - Reporting diagnosis and clinical test/ examination results
Lexical Bundles	- Phrasal bundles (NPs & PPs) (e.g., *the ability to, in the pathogenesis of)* - VP-related: mostly passive-related bundles (e.g., *has been shown to, was performed using*)	- Phrasal bundles (NPs & PPs) (e.g., *with a medical history of, removal of the)* - VP-related mostly passive-related bundles (e.g., *was diagnosed as having, was admitted to our hospital, has been reported to*) - **Clause-related bundles (human, abstract, and external subjects)** (e.g., *his blood pressure was, did not reveal any, the patient underwent a)*	**Diagnosis/intervention-related** - Reporting diagnosis and clinical test/ examination results - Reporting decisions/outcomes
Frames	Content word-based frames (e.g., *average * of, a significant * in*)	- Content word-based frames (e.g., *average * of, a significant * in*) - **Verb-based frames** (e.g., *she was * with, was * to the, it was * that, patient was * with*)	**Diagnosis/intervention-related** - Reporting diagnosis and clinical test/ examination results - Reporting decisions/outcomes

As such, frames and their semantic categories of fillers can be considered as supplements of the "building blocks" of academic writing as these discontinuous sequences provide writers with more options for a more creative use of fixed multiword formulaic sequences like lexical bundles.

On the other hand, sequences at the other end of the continuum constitute the relatively low frequency expressions with strong mutual expectation that could not make their way to the lists of lexical bundles, namely, collocations and multiword collocations. By definition, 2-word collocations are excluded from the lists of lexical bundles as the latter consist of sequences of three or more words. As for multiword collocations, they structurally differ from lexical bundles, and many of these sequences would not meet the high frequency thresholds set for lexical bundles. Yet, we have seen that the lists of both collocations and multiword collocations were replete with specialized sequences of technical and semi-technical lexical items that served key functions in the descriptions of the research or case in MRAs or MCRs, respectively. The use of such sequences has been reported to be not only crucial in medical writing (Rezaeian, 2015), but also challenging for L2-English authors and learners (e.g., Man et al., 2004; Hyland & Tse, 2007; Lam, 2001; Rezaeian, 2015).

In addition to these specialized expressions, multiword MCRC collocation lists also included "ready-made" clauses/fragments that served specific functions in MCRs. In fact, one constant observation is that despite similarities in the structural profiles of MRAs and MCRs, there appears to always be sequences from one or two additional structural categories (**bolded** on Table 8.2) that serve functions specific to MCRs, thus further distinguishing MCRs from MRAs.

In sum, each of the sequences analyzed in the present study makes a different contribution to the description of the formulaic profiles of MRAs and MCRs. These sequences can be viewed, as shown in figure 8.1., on a continuum representing the formulaic profile of a register. Collocations are needed for technical and semi-technical sequences and their functions in specialized texts. Multiword collocations are useful for both specialized sequences and potentially, some ready-made clauses/fragments, depending on the registers under scrutiny. Lexical bundles are at the core of any study of the formulaic profile of a register as these sequences serve the main discourse functions needed to achieve the communicative purpose of any given register and to organize the discourse in a cohesive whole. Finally, frames are the sequences that offer more flexibility in the use of fixed expressions like lexical bundles and occasionally, multiword collocations.

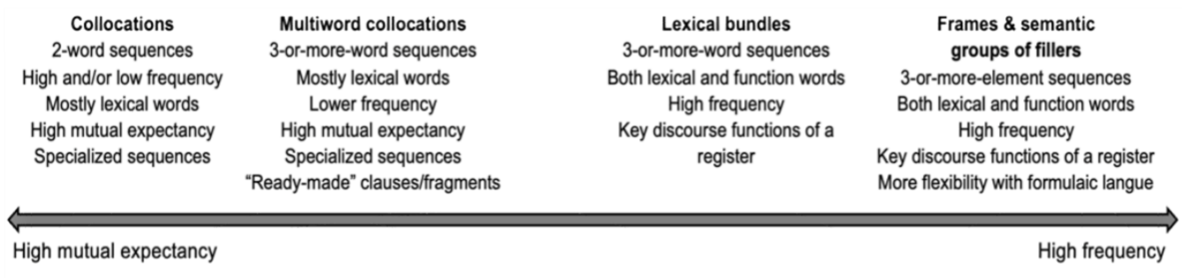


Figure 8.1. Sequences on the Continuum of the Formulaic Profile of a Register

It should be noted that Figure 8.1 does not fully reflect the frequencies of the different sequences. Frames are not necessarily more frequent than lexical bundles, and despite the lower thresholds set for the identification of collocations and multiword collocations, many of these sequences occurred at very high frequencies in the two corpora. The comprehensive approach proposed in the present study has implications for both teaching formulaic sequences in academic writing, particularly in medical writing, and for investigating sequences of various types and lengths. The next section discusses such implications.

8.6 Implications and Applications

As already mentioned, the present study has implications for both teaching and research. Indeed, the complementarity of the various sequences discussed above suggests that novice L1 and L2-English writer would probably benefit from instruction of all four types of sequences needed to produce expert-like texts. From a methodological perspective, the investigation of all these sequences in a same study entails some decisions to ensure the integrity of the lists of analyzed sequences.

8.6.1 Methodological Implications

Two methodological issues arise from a comprehensive approach like the one used in the present study. First, with the investigation of sequences of various lengths there is the possibility of many shorter sequences being embedded in longer ones. It is therefore important to attend to overlapping sequences before establishing the final lists of candidates for analysis. As explained, in chapter 3, I manually checked for overlaps for all sequences, except for 3-word to 8-word bundles. In the future, perhaps more applications like LBiaP (Cortes & Lake, Forthcoming) can help make the identification of overlapping sequences less strenuous.

The second issue relates to overlaps in lists of different types of sequences like lexical bundles and multiword collocations, or lexical bundles and frames, for example. As already mentioned, Figure 8.1 does not reflect the exact frequency of the different types of sequences. Despite the lower frequencies set for the identification of multiword collocations, several sequences occurred at high frequencies. For example, the 3-word collocation *magnetic resonance imaging* occurred 238 time in the MCRC. The sequences *the primary end point* and *intention to treat* occurred 81 and 163 times in the MRAC, respectively. Sequences like these clearly meet the higher thresholds set for the identification of lexical bundles and therefore, appeared also on the lists of bundles. To address these types of overlaps between different types of sequences, it may be necessary to set a working definition for each type of formulaic sequence from the first stages of the study. As multiword collocations consist of primarily lexical words, it is important to define the meaning of “primarily lexical words”. In the present study, I set limits for the number of function words that can occur in each multiword collocation, depending on sequence length (see section 3.4. of the Methodology chapter). These thresholds and working definitions can allow the researcher to make informed decisions regarding which sequences to keep on which list. I kept all three examples above in the multiword collocation lists as despite their high frequencies, these sequences correspond structurally to the working definition of multiword collocations I had established for this study.

As for overlaps between frames and their fillers and lexical bundles or multiword collocations, it may be necessary to identify what qualifies as a frame and which frames are worth considering for functional analysis. In the present study, I excluded relatively fixed frames and highly predictable frames from the functional analysis for two reasons. First, the majority of sequences in this category involve a very limited number of function word fillers that often follow a Zipfian distribution, with only one frequently occurring filler. For example, the frame *in * setting of* occurred 126 times in the MRAC and had only two fillers: *the* and *a.* The filler that forms the bundle *in the setting of* occurred 125 times while *in a setting of* is used only once in a corpus of over one million words. Consequently, the formulaic sequence worth analyzing, based on its frequency, is the bundles *in the setting of* and not the “frame” *in the * of.*

The second reason for excluding fixed frames with high predictability is that they may result from typos that escaped the vigilance of authors. Similar to *in * setting of,* the frame *white blood * count* occurred 73 times in the MCRC, 72 of which were with the fillers *cell.* The second filler was *cells*, and it occurred only once. Grammatically, the sequence *white blood cells count* may be inappropriate as *cell* functions as the modifier of the noun *count*, and therefore, should not be marked for plural. This suggests that *white blood * count* is not a frame as it occurred only with *cell* to form the multiword collocation *white blood cell count.*

On the other hand, highly variable and variable frames that overlap with bundles are still worth analyzing and teaching, even though the bundles have already been analyzed. As explained in chapter 7, section 7.5., high frequency variable and highly variable frames involving bundles often have other frequent fillers that share the same semantic categories as fillers forming the bundles. Thus, such frames and their other fillers are likely to provide more alternatives for the use of the involved bundles and are therefore, worth analyzing and teaching. Deciding on which frames to analyze functionally is even more important, given the always large amounts of data at hand. Based on the findings of the present study, the sequences in the low predictability lists may be the primary candidates for analysis as these lists include all highly variable frames, the vast majority of variable frames, and also the “false fixed” frames. The researcher can always supplement these initial lists with sequences with moderate predictability that do not involve function words or primarily occur with only one filler.

8.6.2 Pedagogical Implications and Applications

The first obvious pedagogical implication of the present study is that disciplinary writing instruction can be more effective if within-discipline variations are taken into account. Given the findings of the present study, it appears that it would take more than general academic writing courses to help medical students become effective writers of case reports. Even teaching how to write a medical research article is likely to be insufficient to provide students with the linguistic tools they need for writing MCRs.

The analysis of each formulaic sequence investigated in the present study may have direct pedagogical implications and applications, some of which I highlight in this section. First, the analyses of both collocations and multiword collocations underscored the importance of helping learners with noun premodification processes. The next implication relates to the frequency of passive-related bundles in both MRAs and MCRs. As mentioned in chapter 6, with the recommendations in medical journals' guidelines for authors to use the active voice, novice and L2-English writers might need guidance on when it is possible to use the active voice and when the passive voice is preferable. And finally, the grouping of frame fillers into semantic categories and the classification frameworks offer material for direct classroom applications.

As all sequences investigated in the present study contribute to the construction of medical texts, a task-based approach that brings students to integrate all four types of sequences in their writing might be well indicated. The outcome would be the completed task, but the aim would be to get students to make appropriate use of the relevant formulaic sequences. For example, students might be assigned to write a case report as a semester final evaluation. This can work particularly well if students are starting or have already started their internships. The task is explained at the beginning of the semester and different deadlines are set for the completion of the different sections of the case report. The first deadline would be for each student to identify a case to follow at their medical facilities.

Given that there have been calls in the literature for explicit teaching of formulaic sequences (e.g., Cortes, 2004; Jones and Haywood, 2004; Li & Schmitt, 2009), I would suggest task-based instruction in its weak form (Skehan, 2003), with the writing tasks assigned to complement instructor-led lectures or classroom activities. Such activities or lectures can focus on relevant formulaic sequences like clausal multiword collocations used for reporting diagnosis or results of clinical tests and/or examination, passive-related bundles used to refer to previous studies, or adjective-controlled collocations used to justify intervention or highlight the importance of the case, to name only a few. Additionally, the instructor can include in those lectures and/or activities some writing techniques like the use of complex noun phrases for more concise writing. The classroom tasks may consist of scaffolding activities going from awareness raising, to guided production, to free production, with the free production tasks being the sections of the case report.

To support students' concise writing, the instructor can start with teaching noun premodifications as this was found to be the most frequently used noun modification process in both MRAs and MCRs. Simple processes like *adjective + noun* or *noun + noun* may not be very problematic at the academic level. But the decision to start with these simple processes is left at the discretion of writing instructors who have a better knowledge of their students' levels. On the other hand, complex noun phrases may be difficult for students and L2-English users to even unpack, let alone produce (Biber & Gray, 2016). Therefore, a good starting point for awareness raising in teaching complex noun phrases could be activities that can help learners untangle the information packed in these types of noun phrases. Such activities can vary from simple reading strategies to short reformulation activities.

One strategy that I personally have used with medical students is "backward reading". This type of activities may be problematic at first, but it works most of the time to introduce students to noun premodification. The strategy consists of reading the complex noun phrase starting from the head noun and moving backwards. For example, for the phrase *oral glucose tolerance test*, the aim would be to make a meaningful sentence starting with the head noun *test* and incorporating the preceding words one at a time. Thus, an example of a coherent sentence that captures the meaning of *oral glucose tolerance test* could be *a test of the tolerance to glucose that (the test) is administered orally.* Another example with *data and safety monitoring board* would be *a board that monitors safety and data.* Once students get the gist, the instructor can provide them with a list of complex noun phrases that they can try to untangle in groups. Perhaps, together, they can come up with other strategies in cases the backward reading does not make much sense. The rationale behind this activity is to help students be aware of the amount of information that can be packed in complex noun phrases.

Once students get the idea, the instructor can introduce reformulation activities by providing strings of information that students will have to "correctly pack" in complex noun phrases. Technical complex noun phrases from the list of identified multiword collocations (Appendix C) could be a good source to select from, making sure that the selected sequences are not too topic dependent. Sequences referring to medical procedures, for example can be used in different case reports, regardless of topic (e.g., *magnetic resonance imaging, fine needle aspiration, alternative diagnostic testing approaches, etc.*). During these activities it may be important to draw students' attention to the wordiness of the provided strings of information.

The next step (not necessarily during the same class session) would be guided production activities. One possible activity could be to provide students with an excerpt of a medical case report where complex noun phrases have been identified and replaced by phrases in which the information is less compressed and more elaborated. First, students would be asked to identify the sentences or phrases that appear too wordy. This can be done in pairs of groups and reviewed as a class to ensure that the correct strings were identified (an alternative would be to directly underline the wordy sentences or phrases, depending on students' levels, as in Figure 8.2). Note that the excerpt shown in Figure 8.2. is from a medical research article, but given the task, a passage from a medical case report would be more appropriate. Then the next step would be to ask students to rewrite the excerpt by replacing the wordy phrases or sentences with complex noun phrases involving premodifications.

Sample excerpt to reformulate

Cardiovascular disease (CVD) is the leading cause of mortality and morbidity worldwide, with 80% of death related to CVD occurring in countries that have low income and middle income.[2] Although CVDs usually clinically manifest in adulthood, changes that are atherosclerotic and early are already present in childhood and adolescence. The density and severity of the lesions seem to be modulated by the presence of factors of risk of CVD that are concomitant.[3, 4] Furthermore, the presence of such factors of risk of CVD that are potentially modifiable during adolescence is a precursor of profiles of cardiovascular risk that are poor in adulthood.[5] Therefore, understanding the distribution of burden of factor of risk of CVD that is behavioural in adolescents in countries that have low income and middle incomes is of utmost importance for formulation of policy for the control and prevention of disease.

Original version (not to be provided to students)

Cardiovascular disease (CVD) is the leading cause of mortality and morbidity worldwide, with 80% of CVD-related deaths occurring in **low income and middle income countries**.[2] Although CVDs usually clinically manifest in adulthood, **early atherosclerotic changes** are already present in childhood and adolescence. The density and severity of the lesions seem to be modulated by the presence of **concomitant CVD risk factors**.[3, 4] Furthermore, the presence of such **potentially modifiable CVD risk factors** during adolescence is a precursor of poor cardiovascular risk profiles in adulthood.[5] Therefore, understanding the distribution of **behavioural CVD risk factor burden** in adolescents **in low income and middle income countries** is of utmost importance for **policy formulation** for **disease prevention and control**.

Figure 8.2. Example of Text for Guided Production of Complex Noun Phrases with Premodifications (adapted from an article in the MRAC, 15Lancet2)

As students get familiar with the noun premodification processes, they can be given a short writing task (free production stage) like a synopsis of the case they will work on, for example. The synopsis will have a very limited word count and will include at least 3 to 4 complex noun phrases involving only premodification. Post modification processes can be introduced in subsequent activities to familiarize students with both noun modification processes.

After these writing techniques, the subsequent class activities and lectures can focus directly on the types of formulaic sequences students will mostly need to complete the different sections of the case report. For example, for the Case Presentation section, instructors can choose to introduce the passive-related bundles used to report diagnosis and decisions, and clausal multiword collocations used to report results of physical examination and/or clinical tests. At this point, it can be assumed that students have already been introduced to passive-related bundles used to refer to previous research. As awareness raising activities, instructors can start with sets of concordance lines like those shown in Figure 8.3 and 8.4.

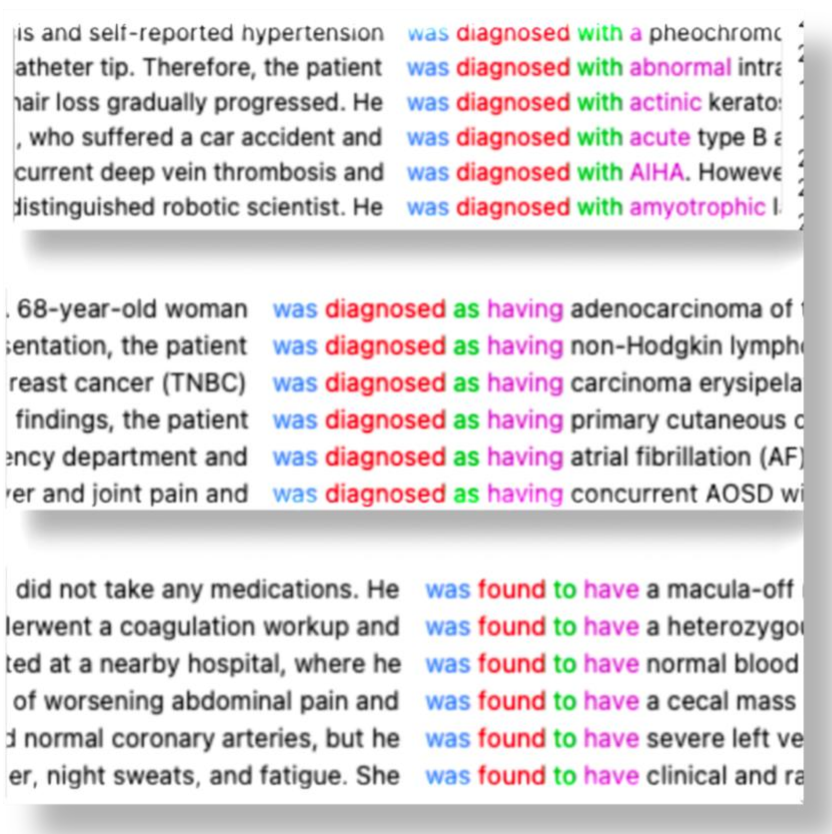


Figure 8.3. Concordance Lines of Passive-related Bundles

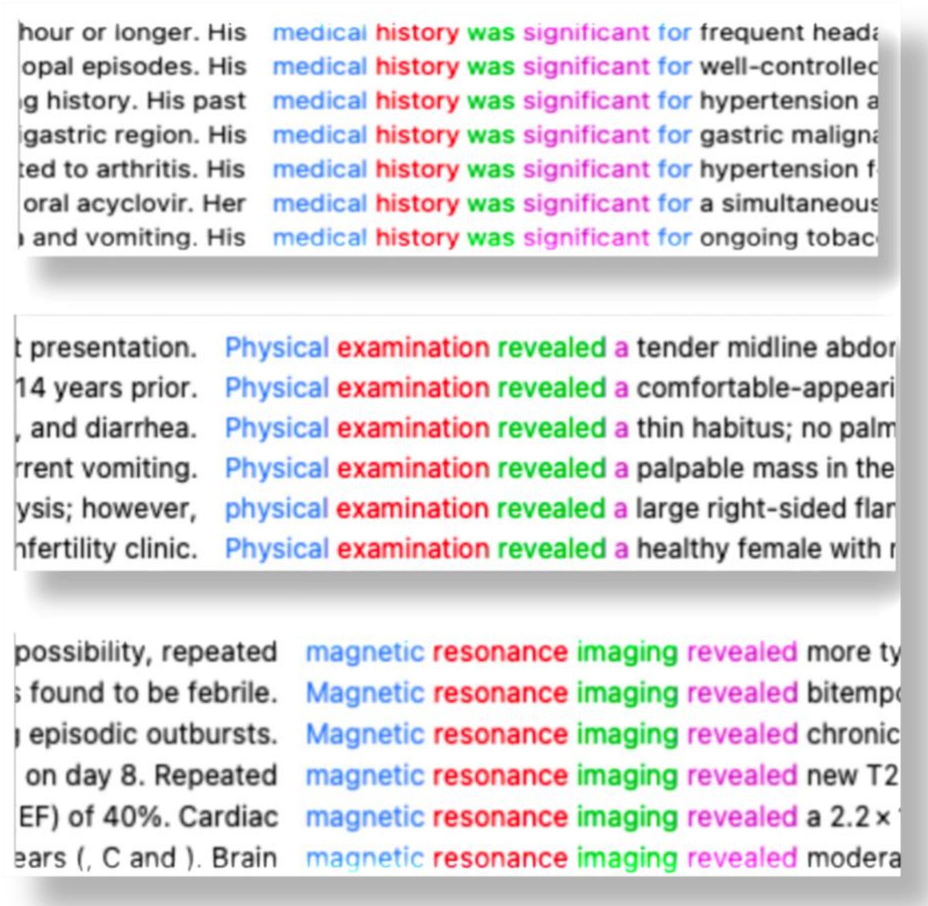


Figure 8.4. Concordance Lines of Clausal Multiword Collocations

Then in groups or pairs, students discuss and identify for what purposes (or discourse functions) passive-related and clausal multiword collocations are used. Then the activity is reviewed as a class to (1) ensure that all functions/purposes are correctly identified, and (2) bring students to notice the possibility of using medical procedures as agents in active sentences to report results and avoid using the passive.

Once students have been exposed to these different situations of use of these sequences, instructors can introduce some frames and their fillers serving similar functions. They may decide to focus only on a couple of functions so as not to overwhelm students. Instructors can then take the framework for verb-based frames, which were found to serve the selected functions, and delete all rows with other functions. Then they can delete the entries in the function column to have a framework that looks like the one shown in Table 8.3. below.

Table 8.3. Framework for the identification of Procedures and Reporting Results Functions

Subgroups	**Filler word class**	**Filler sematic category**	**Function**	**examples**
Auxiliary-initial				
were * + prep	Verb	Mental		were * in the (*found, noted, detected, observed*)
copular be-initial				
was * + prep	Adjective	Evaluative, relational		was * for the (*positive, negative, unusual, typical*)
Determiner/Noun/Pronoun-initial				
he/she-initial	Verb	Aspectual		he was * on (*started, commenced, kept)*
it was…	Verb	Mental		it was * that (*noted, observed, found, discovered*)
	noun	Technical		serum * level was (*potassium, calcium*)
	Adjective	Evaluative, Relational		cells were * for (*positive, negative, immunopositive*)
Other lexical verbs				
communication verbs	Noun	Multiple semantic categories		revealed the * of (*presence, proliferation, volume)*
Mental verbs	Noun	Abstract/ Attribute		assess the * of (*cause, impact, risk, effectiveness)*

The next step is to prepare handouts that show selected frames and some of their fillers in contexts. These can be concordance lines like the ones in Figure 8.5, showing the frame *he was * on*, or simply sentences selected from different case reports. Students are then given the framework and the handouts showing the frames and their fillers in context. In groups, they study the concordance lines or the lists of sentences and identify which frames with their semantic groups of fillers serve which functions. If instructors find that this activity is too challenging for their students, they can elect to have a lecture on these two functions and provide the completed framework to students as a reference. But having students figure out the functions might be more effective. As rightly suggested by Hunston (2002), students are more likely to remember “what they have worked to find out” (p. 170). After these activities, students can move on to writing the first draft of their Case Presentation sections and then receive feedback from both peers and the instructor.

*Figure 8.5. The Frame 'he was * on' and some its Fillers in Context*

The tasks and activities presented in this section are just some examples of materials and activities that could be used in a medical writing class for novice L1 English and L2 English writers. There may be innumerable opportunities for material and activity creation from the findings of the analyses of formulaic sequences investigated in the present study.

8.7 Some Limitations of the Study

Like many studies, the present has its limitations. Despite the comprehensive approach to the analysis of formulaic sequences, the description of the two registers could have been supplemented by an analysis of the textual environments of the formulaic sequences. Perhaps, textual colligations or semantic associations (Hoey, 2005) could have helped better explain certain observations like the differences in the choice of fillers of the same frame, particularly when the two expressions formed by the frame and the two fillers have similar meanings and/or serve the same function. From a pedagogical perspective, a description of where different sequences prefer to occur in MRAs and MCRs, and in what grammatical and semantic environments, could be of much interest to medical writing instruction. Another limitation is that for the reasons provided in 3.6.2, only subsets of frames in the two registers were analyzed functionally in the present study. There still remains the need to analyze frames of other lengths in both registers and from other structural categories for MCRs to better serve medical writing instruction.

8.8 Suggestions for Further Research

Following the limitations above, other areas to explore in the description of formulaic language in a register include the grammatical and textual colligations and the semantic associations (or prosody) of formulaic sequences. As mentioned above, such studies can have valuable pedagogical implications and applications. Still in an aim to better inform writing instruction, studies are needed that combine genre and register approaches to the description of text varieties by looking at the relationships between formulaic sequences and rhetorical moves. Very few studies have gone in that direction (e.g., Cortes, 2013; Kashiha, 2015; Le & Harrington, 2015) and have mostly focused on sections of the IMRD research article and on single types of formulaic sequences. Investigating the connection between formulaic sequences and the moves of an entire text variety could help writing instructors identify what formulaic sequences to teach to help realize each move or communication function.

As mentioned in section 1.2. of the introductory chapter, the comparison of the two registers has primarily focused on the structures and discourse functions of the formulaic sequences investigated in the present dissertation. A study like this one could be complemented with a more quantitative analysis of the expressions, based on comparable corpora. This potential investigation could experiment using various statistical procedures to look for significant differences in the number of types or tokens of the different formulaic expressions across the two registers.

Finally, in this study I have attempted to provide extensive descriptions of two registers from a formulaic language perspective and highlight potential variations that can occur within a same discipline. Though there are still many areas to explore, the findings I presented can constitute a valuable basis for medical writing instruction. It is my hope that the approach I used in the present study will motivate subsequent descriptions of other registers within other disciplines based on the analysis of frequent formulaic expressions. Such studies would be of great value for writing instruction in English for Academic Purposes (EAP) and English for Specific Purposes (ESP).

REFERENCES

Abdollahpour, Z., & Gholami, J. (2018). Building blocks of medical abstracts: frequency, functions, and structures of lexical bundles. *Asian ESP Journal*, *14*(1), 82-110.

Ackermann, K., & Chen, Y. H. (2013). Developing the Academic Collocation List (ACL) - A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes*, *12*(4), 235–247. https://doi.org/10.1016/j.jeap.2013.08.002

Ahn, S. J. (2013). An analysis of the use of phrasal verbs and prepositional verbs by Korean EFL learners. *The New Korean Journal of English Language and Literature*, *55*(2), 207-232.

Anthony, L. (2020). AntGram (Version 1.2.3) [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software

Anthony, L. (2017). AntConc (Version 3.5.2) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/software

Anthony, L. (2020). AntConc (Version 3.5.9) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/software

Back, J. (2011). Preposition errors in writing and speaking by Korean EFL learners: A corpus-based approach. *Studies in British and American Language and Literature, 99*, 227- 47.

Barlow, M. 2004. *Collocate 1.0: Locating collocations and terminology*. Houston, TX: Athelstan.

BMJ Case Report (2020, November 14). *Instructions for Authors.* https://casereports.bmj.com/pages/authors/

Biber, D. (2006). *University Language: A Corpus-Based Study of Spoken and Written Registers*. Amsterdam: John Benjamins Publishing.

Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics* 14(3): 275–311.

Biber, D., & Conrad, S. (2009). *Register, genre, and style*. Cambridge: Cambridge University Press.

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, *26*(3), 263–286.

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). *The Longman grammar of spoken and written English*. London: Longman.

Biber, D. and Gray, B., (2016). *Grammatical complexity in academic English: Linguistic change in writing*. Cambridge University Press.

Biber, D., Gray, B., Staples, S., & Egbert, J. (2020). Investigating grammatical complexity in L2 English writing research: Linguistic description versus predictive measurement. *Journal of English for Academic Purposes*, *46*, 100869. https://doi.org/10.1016/j.jeap.2020.100869

Brezina, V., McEnery, T., & Wattam, S. (2015). Collocations in context: A new perspective on collocation networks. *International Journal of Corpus Linguistics, 20*(2), 139–173. https://doi.org/10.1075/ijcl.20.2.01bre

Chapman, S., Haynes, A., Derrick, G., Sturk, H., Hall, W. D., & George, A. S. (2014). Reaching "an audience that you would never dream of speaking to": Influential public health researchers' views on the role of news media in influencing policy and public understanding. *Journal of Health Communication*, *19*(2), 260–273. https://doi.org/10.1080/10810730.2013.811327

Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology*, *14*(2), 30-49.

Chen, Q., & Ge, G. C. (2007). A corpus-based lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles (RAs). *English for Specific Purposes*, *26*, 502–514. https://doi.org/10.1016/j.esp.2007.04.003

Conrad, S., & Biber, D. (2000). Adverbial marking of stance in speech and writing. In S. Hunston & G.Thompson (Eds.), *Evaluation in text: Authorial stance and the construction of discourse* (pp. 56-73). Oxford University Press, UK.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, *23*(4), 397–423. https://doi.org/10.1016/j.esp.2003.12.001

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

Cortes, V. (2008). A comparative analysis of lexical bundles in academic history writing in English and Spanish. *Corpora*, *3*(1), 43-57.

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. (in Press). Lexical bundles in English for Academic Purposes.

Cortes, V., & Lake W. (Forthcoming). LBiaP: A solution to the problem of attaining observation independence in lexical bundle studies.

Cunningham, K. J. (2017). A phraseological exploration of recent mathematics research articles through key phrase frames. *Journal of English for Academic Purposes*, *25*, 71. https://doi.org/10.1016/j.jeap.2016.11.005

Ellis, N. C. (2007). Language acquisition as rational contingency learning. *Applied Linguistics*, *27*(1), 1–24. https://doi.org/10.1093/applin/ami038

Ellis, N. C. (2012). Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics*, *32*, 17–44. https://doi.org/10.1017/S0267190512000025

Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, *42*(3), 375–96.

Eysenbach, G. (2008). Medicine 2.0: Social networking, collaboration, participation, apomediation, and openness. *Journal of Medical Internet Research*, *10*(3), 1–13. https://doi.org/10.2196/jmir.1030

Ferguson, G. (2001). If you pop over there: A corpus-based study of conditionals in medical discourse. *English for Specific Purposes*, *20*(1), 61–82. https://doi.org/10.1016/S0889-4906(99)00027-7

Firth, J. R. (1957). *Papers in Linguistics.* London: Oxford University Press.

Fletcher, W. H. (2002–2007). *KfNgram*. Annapolis, MD: USNA.

Florek, A. G., & Dellavalle, R. P. (2016). Case reports in medical education: A platform for training medical students, residents, and fellows in scientific writing and critical thinking. *Journal of Medical Case Reports*, *10*(1), 1–3. https://doi.org/10.1186/s13256-016-0851-5

Flowerdew, J., & Forest, R. W. (2009). Schematic structure and lexico-grammatical realization in corpus-based genre analysis: The case of research in the PhD literature review. In M. Charles, S. Hunston, & D. Pecorari (Eds.). (2009). *Academic writing: At the interface of corpus and discourse*. Bloomsbury Publishing.

Fryer, D. L. (2012). Analysis of the generic discourse features of the English-language medical research article: A systemic-functional approach. *Functions of Language*, *19*(1), 5–37. https://doi.org/10.1075/fol.19.1.01fry

Fuster-Márquez, M., & Pennock-Speck, B. (2015). Target frames in British hotel websites. *International Journal of English Studies*, *15*(1), 51-69.

Gagnier, J. J., Kienle, G., Altman, D. G., Moher, D., Sox, H., Riley, D., … Tugwell, P. (2013). The CARE guidelines: Consensus-based clinical case reporting guideline development. *BMJ Case Reports*, *2*(5), 38–43. https://doi.org/10.1136/bcr-2013-201554

Garner, J. R. (2016). A phrase-frame approach to investigating phraseology in learner writing across proficiency levels. *International Journal of Learner Corpus Research*, *2*(1), 31-67.

Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in Corpus-Based Language Learning Research: Identifying, Comparing, and Interpreting the Evidence. *Language Learning*, *67*(June), 155–179. https://doi.org/10.1111/lang.12225

Geluso, J. (2019). *Frequency, semantic, and functional characteristics of discontinuous formulaic language: A learner corpus study.* (Unpublished doctoral dissertation). Iowa State University, Ames, Iowa.

Gledhill, C. (2000). The discourse function of collocation in research article introductions. *English for Specific Purposes*, *19*, 115–135. https://doi.org/10.1016/S0889-4906(98)00015-5

Goodier, C. (2008). *Purpose and identity in professional and student radiology writing: a genre based approach.* (Unpublished doctoral dissertation). University of South Africa, Pretoria, South Africa.

Grabowski, Ł. (2013). Register Variation Across English Pharmaceutical Texts: A Corpus-driven Study of Keywords, Lexical Bundles and Phrase Frames in Patient Information Leaflets and Summaries of Product Characteristics. *Procedia - Social and Behavioral Sciences*, *95*, 391–401. https://doi.org/10.1016/j.sbspro.2013.10.661

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

Gray, B. (2015). *Linguistic Variation in Research Articles: When Discipline Tells Only Part of the Story*. Amsterdam: John Benjamins Publishing Company.

Gray, B., & Biber, D. (2013). Lexical frames in academic prose and conversation. *International Journal of Corpus Linguistics*, *18*(1), 109–136. https://doi.org/10.1075/ijcl.18.1.08gra

Green, B. N., & Johnson, C. D. (2006). How to write a case report for publication. *Journal of chiropractic medicine*, *5*(2), 72-82.

Gries, S. T. (2013). 50-Something Years of Work on Collocations. *International Journal of Corpus Linguistics*, *18*(1), 137–166. https://doi.org/10.1075/ijcl.18.1.09gri

Helán, R. (2012). *Analysis of Published Medical Case Reports: Genre-Based Study*. (Unpublishes doctoral dissertation). Masaryk University, Brno, Czech Republic.

Hoey, M.P. (2005). *Lexical Priming. A new theory of words and language*. London: Routledge.

Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics*, *19* (1), 24–44. https://doi.org/10.1093/applin/19.1.24

Hyland, K. (2000). *Disciplinary Discourses: Social Interactions in Academic Writing* (Applied linguistics and language study). Harlow, England; New York: Longman.

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

Hyland, K. (2008b). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, *18*(1), 41–62.

Hyland, K., & Jiang, F. (2018). Academic lexical bundles: How are they changing? *International Journal of Corpus Linguistics*, *23*(4), 383–407. https://doi.org/10.1075/ijcl.17080.hyl

Hyland, K., & Jiang, F. (Kevin). (2016). "We must conclude that…": A diachronic study of academic engagement. *Journal of English for Academic Purposes*, *24*, 29–42. https://doi.org/10.1016/j.jeap.2016.09.003

Hyland, K., & Tse, P. (2004). Metadiscourse in academic writing: A reappraisal. *Applied Linguistics*, *25*(2), 156–177.

Hyland, K., & Tse, P. (2007). Is there an "academic vocabulary"? *TESOL Quarterly*, *41*(2), 235–253. https://doi.org/10.1002/j.1545-7249.2007.tb00058.x

Jalali, H., Eslami Rasekh, A., & Tavangar Rizi, M. (2008). Lexical bundles and intradisciplinary variation: The case of applied linguistics. *Iranian journal of language studies*, *2*(4), 447-484.

Jalali, Z. S., & Moini, M. R. (2014). Structure of lexical bundles in introduction section of medical research articles. *Procedia-Social and Behavioral Sciences, 98*, 719-726.

Jalali, Z. S., Moini, M. R., & Arani, M. A. (2015). Structural and functional analysis of lexical bundles in medical research articles: A corpus-based study. *International Journal of Information Science and Management*, *13 LA-E*(1), 51–69.

Jirapanakorn, N. (2012). How doctors report: A corpus-based contrastive analysis of reporting verbs in research article introductions published in international and Thai medical journals. *The Bangkok Medical Journal*, *4*, 39-39.

Journal of Medical Case Reports (2021, January 19). *Submission Guidelines.* https://jmedicalcasereports.biomedcentral.com/submission-guidelines

Journal of American Medical Association (2020, November14). 'Instructions for Authors'. Retrieved from https://jamanetwork.com/journals/jama/pages/instructions-for-authors

Kanoksilapatham, B. (2015). Distinguishing textual features characterizing structural variation in research articles across three engineering sub-discipline corpora. *English for Specific Purposes*, *37*, 74-86.

Kashiha, H. (2015). Recurrent formulas and moves in writing research article conclusions among native and nonnative writers. *3L: Language, Linguistics, Literature*, *21*(1), 47–59. https://doi.org/10.17576/3L-2015-2101-05

Krummes, C., & Ensslin, A. (2015). Formulaic language and collocations in German essays: from corpus-driven data to corpus-based materials. *Language Learning Journal*, *43*(1), 110–127. https://doi.org/10.1080/09571736.2012.694900

Lam, J. K. M. (2001). A study of semi-technical vocabulary in computer science texts, with special reference to ESP teaching and lexicography. *In G. James (Ed.) Research Reports*.

Laso, N. J., & John, S. (2013). An exploratory study of NNS medical writers' awareness of the collocational patterning of abstract nouns in medical discourse. *Revista Espanola de Linguistica Aplicada*, *26*, 307–331.

Lee, Y. E., Yoo, I. W. H., & Shin, Y. K. (2020). The use of English prepositions in lexical bundles in essays written by Korean university students. *Journal of English for Academic Purposes*, *45*, 100848. https://doi.org/10.1016/j.jeap.2020.100848

Li, L.-J., & Ge, G.-C. (2009). Genre analysis: Structural and linguistic evolution of the English-medium medical research article (1985–2004). *English for Specific Purposes*, *28*(2), 93–104. https://doi.org/10.1016/j.esp.2008.12.004

Lu, X., Yoon, J., & Kisselev, O. (2018). A phrase-frame list for social science research article introductions. *Journal of English for Academic Purposes*, *36*, 76–85. https://doi.org/10.1016/j.jeap.2018.09.004

Man, J. P., Weinkauf, J. G., Tsang, M., & Sin, D. D. (2004). Why do some countries publish more than others? An international comparison of research funding, English proficiency and publication output in highly ranked general medical journals. *European Journal of Epidemiology*, *19*(8), 811–817. https://doi.org/10.1023/B:EJEP.0000036571.00320.b8

Marco, M. J. L. (2000). Collocational frameworks in medical research papers: a genre-based study. *English for Specific Purposes*, *19*, 63–86. https://doi.org/10.1016/S0889-4906(98)00013-1

Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics*, *33*(3), 299–320. https://doi.org/10.1093/applin/ams010

Millar, N., Budgell, B., & Fuller, K. (2013). "Use the active voice whenever possible": The impact of style guidelines in medical journals. *Applied Linguistics*, *34*(4), 393–414. https://doi.org/10.1093/applin/ams059

Mbodj-Diop, N. B. (2016). *Lexical bundles in medical research articles: Structures and functions*. Michigan State University.

Mungra, P., & Canziani, T. (2013). Lexicographic studies in medicine : Academic Word List for clinical case histories. *Iberica*, *25*(2013), 39–62.

Nation, I. S. (2001). *Learning vocabulary in another language*. Ernst Klett Sprachen.

Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford University Press.

Nesi, H., & Basturkmen, H. (2006). Lexical bundles and discourse signalling in academic lectures. *International Journal of Corpus Linguistics*, *11*(3), 283–304. https://doi.org/10.1075/ijcl.11.3.04nes

Nwogu, K. N. (1997). The medical research paper: Structure and functions. *English for specific purposes*, *16*(2), 119-138.

O'Donnell, M. B., Römer, U., & Ellis, N. C. (2013). The development of formulaic sequences in first and second language writing. *International Journal of Corpus Linguistics*, *18*(1), 83–108. https://doi.org/10.1075/ijcl.18.1.07odo

Paquot, M. (2017). The phraseological dimension in interlanguage complexity research. *Second Language Research*, 26765831769422. http://doi.org/10.1177/0267658317694221

Peacock, M. (2012). High-frequency collocations of nouns in research articles across eight disciplines. *Iberica*, *23*(2012), 29–36.

Pérez-Llantada, C. (2014). Formulaic language in L1 and L2 expert academic writing: Convergent and divergent usage. *Journal of English for Academic Purposes*, *14*, 84–94. https://doi.org/10.1016/j.jeap.2014.01.002

Phillips, D. P., Kanter, E. J., Bednarczyk, B., & Tastad, P. L. (1991). Importance of the lay press in the transmission of medical knowledge to the scientific community. *NEJM*, *325*, 1180 – 1183.

Ping Alvin, L. (2014). The passive voice in scientific writing. The current norm in science journals. *Journal of Science Communication*, *13*(1), A03.

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

Rison, R. A., Shepphird, J. K., & Kidd, M. R. (2017). How to choose the best journal for your case report. *Journal of Medical Case Reports*, *11*(1), 1–10. https://doi.org/10.1186/s13256-017-1351-y

Römer, U. (2009). English in Academia: Does Nativeness Matter? *International Journal of English Studies*, *2*(September), 89–100.

Römer, U. (2010). Establishing the phraseological profile of a text type: The construction of meaning in academic book reviews. *English text construction*, *3*(1), 95-119.

Science Translational Medicine (2021, January19). 'Instructions for Authors of Research Articles (Initial Submission)'. Retrieved from https://www.science.org/content/page/stm-instructions-research-articles-initial-submission

Skehan, P. (2003). Task-based instruction. *Language teaching*, *36*(1), 1-14.

Seoane, E. (2013). On the conventionalisation and loss of pragmatic function of the passive in Late Modern English scientific discourse. *Journal of Historical Pragmatics*, *14*(1), 70–99. https://doi.org/10.1075/jhp.14.1.03seo

Shahriari, H. (2017). Comparing lexical bundles across the introduction, method and results sections of the research article. *Corpora*, *12*(1), 1–22. https://doi.org/10.3366/cor.2017.0107

Shin, Y. K. (2018). The Construction of English Lexical Bundles in Context by Native and Nonnative Freshman University Students. *English Teaching*, *73*(3), 115–139. https://doi.org/10.15858/engtea.73.3.201809.115

Shin, Y. K., & Kim, Y. J. (2017). Using lexical bundles to teach articles to L2 English learners of different proficiencies. *System*, *69*, 79–91. https://doi.org/10.1016/j.system.2017.08.002

Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, *31*(4), 487–512. http://doi.org/10.1093/applin/amp058

Sinclair, J. McH. (1991). *Corpus Concordance Collocation*. Oxford: Oxford University Press.

Sinclair, J. McH. (2004). *Trust the Text: Language, Corpus and Discourse*. London: Routledge.

Staples, S., Egbert, J., Biber, D., & McClair, A. (2013). Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. *Journal of English for Academic Purposes*, *12*(3), 214–225. http://doi.org/10.1016/j.jeap.2013.05.002

Stubbs, M. 2007. Quantitative data on multi-word sequences in English. The case of the word 'world'. *Text, discourse and corpora*, Hoey, M., M. Mahlberg, M. Stubbs & W. Teubert (eds). London: Continuum. 163–189.

Wang, J., Liang, S. L., & Ge, G. C. (2008). Establishment of a Medical Academic Word List. *English for Specific Purposes*, *27*, 442–458.

Williamson, P. O., & Minter, C. I. J. (2019). Exploring PubMed as a reliable resource for scholarly communications services. *Journal of the Medical Library Association*, *107*(1), 16–29. https://doi.org/10.5195/jmla.2019.433

Wray, A. (2005). *Formulaic language and the lexicon*. Cambridge University Press.

Xiao, R., & McEnery, T. (2006). Collocation, semantic prosody, and near synonymy: A cross-linguistic perspective. *Applied Linguistics*, *27*(1), 103–129. https://doi.org/10.1093/applin/ami045

Yang, A., Zheng, S. Yuan, & Ge, G. Chun. (2015). Epistemic modality in English-medium medical research articles: A systemic functional perspective. *English for Specific Purposes*, *38*, 1–10.

Yanoff , K. L. & Burg, F. D. (1988). Types of medical writing and teaching of writing. *Journal of Medical Education*, *33*(1), 30-37

APPENDICES

Appendix A: MRAC Collocations

Appendix A.1: Bidirectionals_ Relatively Strong Mutual Attraction (-0.5 ≤ Log_{10} ≤ +0.5)

		Freq	Range	$\Delta P_{2\|1}$	$\Delta P_{1\|2}$	Log_{10}
1	*viral load*	60	12	0,14	0,44	-0,50
2	*non invasive*	65	10	0,10	0,31	-0,50
3	*risk factors*	456	61	0,17	0,52	-0,48
4	*chest radiograph*	23	9	0,28	0,82	-0,47
5	*cytoplasmic inclusions*	10	8	0,14	0,37	-0,43
6	*bypass grafting*	10	8	0,29	0,77	-0,43
7	*western blot*	24	16	0,32	0,82	-0,42
8	*intestinal epithelium*	30	10	0,13	0,34	-0,41
9	*liquid nitrogen*	13	12	0,20	0,48	-0,39
10	*statistical significance*	87	71	0,15	0,37	-0,38
11	*systematic reviews*	38	14	0,30	0,72	-0,37
12	*heat inactivated*	10	8	0,14	0,30	-0,33
13	*serial dilutions*	8	8	0,15	0,32	-0,33
14	*confidence intervals*	73	44	0,25	0,52	-0,32
15	*bone marrow*	66	20	0,43	0,90	-0,32
16	*flow cytometry*	140	45	0,49	1,00	-0,31
17	*heart failure*	229	40	0,28	0,57	-0,31
18	*blood pressure*	555	57	0,39	0,79	-0,30
19	*noninferiority margin*	16	9	0,24	0,44	-0,27
20	*laser scanning*	9	7	0,19	0,35	-0,27
21	*rabbit polyclonal*	29	12	0,30	0,55	-0,26
22	*lamina propria*	9	7	0,56	1,00	-0,25
23	*atrial fibrillation*	139	9	0,56	1,00	-0,25
24	*carbon monoxide*	17	10	0,57	1,00	-0,25
25	*phosphate-buffered saline*	24	22	0,34	0,60	-0,24
26	*dimethyl sulfoxide*	10	8	0,59	1,00	-0,23
27	*cubic millimeter*	33	11	0,59	1,00	-0,23
28	*primary outcome*	216	70	0,19	0,31	-0,21
29	*de novo*	54	15	0,61	0,98	-0,21
30	*confidence interval*	106	74	0,36	0,58	-0,21
31	*white matter*	89	15	0,36	0,58	-0,20
32	*magnetic resonance*	10	5	0,53	0,84	-0,20
33	*lymph node*	42	17	0,40	0,64	-0,20
34	*decision making*	31	9	0,24	0,38	-0,20
35	*hydrogen peroxide*	10	7	0,67	1,00	-0,18
36	*buffered formalin*	9	8	0,21	0,31	-0,16
37	*extracellular matrix*	25	12	0,25	0,33	-0,13
38	*amino acids*	35	17	0,35	0,45	-0,12
39	*end point*	189	41	0,27	0,36	-0,11

40	*sample size*	158	96	0,28	0,37	-0,11
41	*predictive validity*	24	6	0,28	0,36	-0,11
42	*end points*	162	38	0,24	0,30	-0,11
43	*polyethylene glycol*	8	5	0,80	1,00	-0,10
44	*nonhuman primates*	15	6	0,58	0,71	-0,09
45	*weight loss*	214	25	0,30	0,33	-0,05
46	*mechanical ventilation*	45	8	0,62	0,69	-0,05
47	*lymph nodes*	60	12	0,57	0,63	-0,04
48	*myocardial infarction*	199	40	0,88	0,97	-0,04
49	*patch clamp*	14	5	0,55	0,57	-0,02
50	*ice cold*	16	9	0,35	0,36	-0,02
51	*colony-forming units*	8	7	0,33	0,34	-0,01
52	*et al*	236	74	0,96	0,98	-0,01
53	*vice versa*	19	15	1,00	1,00	0,00
54	*bona fide*	7	7	1,00	1,00	0,00
55	*propidium iodide*	7	7	0,87	0,87	0,00
56	*substance abuse*	7	6	0,50	0,50	0,00
57	*emergency department*	30	12	0,51	0,50	0,01
58	*unstable angina*	15	7	0,58	0,56	0,02
59	*poorly understood*	25	18	0,56	0,50	0,05
60	*Alexa Fluor*	38	22	0,95	0,84	0,05
61	*verbal autopsy*	9	5	0,38	0,33	0,06
62	*informed consent*	159	130	0,82	0,70	0,07
63	*homologous recombination*	13	7	0,35	0,28	0,10
64	*rhesus macaque*	11	5	0,37	0,28	0,11
65	*alanine aminotransferase*	30	12	0,73	0,55	0,13
66	*macaca mulatta*	9	7	1,00	0,71	0,15
67	*ethical approval*	11	10	0,37	0,26	0,16
68	*phylogenetic tree*	14	7	0,52	0,36	0,16
69	*normally distributed*	13	12	0,37	0,25	0,16
70	*cleaved caspase-3*	6	5	0,40	0,27	0,17
71	*endoplasmic reticulum*	18	12	1,00	0,67	0,18
72	*critically ill*	15	5	0,71	0,47	0,18
73	*adverse events*	428	70	0,67	0,43	0,19
74	*waist circumference*	27	10	0,73	0,47	0,20
75	*elastic lamina*	6	5	0,60	0,37	0,20
76	*bonferroni correction*	12	11	0,35	0,22	0,21
77	*oxidative stress*	95	18	0,47	0,29	0,21
78	*Hardy-Weinberg equilibrium*	11	8	1,00	0,61	0,21
79	*renin-angiotensin system*	19	7	0,47	0,28	0,23
80	*newly diagnosed*	54	10	0,52	0,30	0,25
81	*randomly assigned*	167	72	0,70	0,40	0,25
82	*transplant recipients*	38	5	0,51	0,28	0,25
83	*neutral-buffered formalin*	7	5	0,30	0,17	0,26
84	*proinflammatory cytokines*	28	13	0,35	0,19	0,26

85	*horseradish peroxidase*	7	12	0,54	0,29	0,27
86	*t cells*	1237	71	0,50	0,27	0,27
87	*bronchoalveolar lavage*	20	13	0,83	0,44	0,27
88	*body weight*	122	40	0,35	0,17	0,32
89	*determine whether*	163	80	0,42	0,20	0,33
90	*inclusion criteria*	52	41	0,31	0,14	0,33
91	*adequately powered*	16	15	0,62	0,29	0,33
92	*ethics committee*	62	68	0,62	0,28	0,34
93	*skeletal muscle*	86	10	0,92	0,42	0,35
94	*systematic review*	59	29	0,47	0,20	0,38
95	*dry ice*	9	7	0,47	0,20	0,38
96	*glycated hemoglobin*	25	7	0,86	0,35	0,39
97	*autosomal dominant*	24	6	0,63	0,25	0,40
98	*aspartate aminotransferase*	19	13	0,86	0,35	0,40
99	*carried out*	77	47	0,72	0,28	0,41
100	*plasmodium falciparum*	20	9	0,50	0,19	0,41
101	*antihypertensive drugs*	35	8	0,37	0,14	0,42
102	*90th percentile*	10	7	0,71	0,27	0,42
103	*twice daily*	42	22	0,30	0,11	0,43
104	*outpatient visits*	35	6	0,33	0,12	0,43
105	*rheumatoid arthritis*	16	6	0,84	0,31	0,44
106	*hazard ratio*	170	45	0,75	0,26	0,46
107	*mutually exclusive*	8	7	0,80	0,28	0,46
108	*guanine nucleotide*	9	9	0,32	0,10	0,48
109	*electron microscopy*	19	9	0,47	0,15	0,49
110	*alkaline phosphatase*	12	12	0,75	0,24	0,49

Appendix A.2*: *Unidirectionals1_ Word 1 as strongest predictor (0.51 ≤ Log_{10} ≤ 4)

		Freq	**Range**	**$\Delta P_{2\|1}$**	**$\Delta P_{1\|2}$**	**Log_{10}**
1	*lethally irradiated*	6	6	1,00	0,21	0,67
2	*folic acid*	81	6	1,00	0,18	0,75
3	*immunosorbent assay*	18	18	1,00	0,04	1,36
4	*transforming growth*	12	9	1,00	0,04	1,43
5	*acetic acid*	13	9	1,00	0,03	1,54
6	*accounted for*	90	41	0,99	0,01	2,19
7	*accompanied by*	40	30	0,99	0,01	2,29
8	*compensate for*	11	9	0,99	0,00	3,11
9	*coincide with*	6	6	0,99	0,00	3,45
10	*according to*	572	182	0,98	0,02	1,60
11	*confined to*	19	11	0,98	0,00	3,08
12	*belonging to*	16	6	0,98	0,00	3,15
13	*refers to*	13	8	0,98	0,00	3,24
14	*contributors to*	12	8	0,98	0,00	3,28
15	*amenable to*	6	6	0,98	0,00	3,58
16	*nervous system*	31	19	0,97	0,06	1,23
17	*consisting of*	59	39	0,96	0,00	2,89

18	*irrespective of*	37	22	0,96	0,00	3,09
19	*consists of*	24	19	0,96	0,00	3,28
20	*paucity of*	13	12	0,96	0,00	3,54
21	*reminiscent of*	10	8	0,96	0,00	3,66
22	*interquartile range*	57	30	0,95	0,12	0,91
23	*insight into*	38	30	0,95	0,04	1,38
24	*depending on*	46	39	0,95	0,01	2,01
25	*escherichia coli*	15	12	0,94	0,22	0,62
26	*ranged from*	82	50	0,94	0,01	1,85
27	*leads to*	90	52	0,94	0,00	2,38
28	*consisted of*	77	55	0,94	0,00	2,76
29	*accounts for*	19	18	0,94	0,00	2,85
30	*capable of*	51	30	0,94	0,00	2,94
31	*correspond to*	21	14	0,94	0,00	3,01
32	*ex vivo*	79	26	0,93	0,16	0,75
33	*contributes to*	54	41	0,93	0,00	2,60
34	*accounting for*	32	22	0,93	0,00	2,62
35	*regardless of*	52	34	0,93	0,00	2,92
36	*interfere with*	17	13	0,93	0,00	2,97
37	*counterstained with*	16	10	0,93	0,00	3,00
38	*tend to*	20	22	0,93	0,00	3,03
39	*unaware of*	33	23	0,93	0,00	3,12
40	*drinking water*	34	14	0,92	0,14	0,81
41	*responsible for*	107	71	0,92	0,01	2,09
42	*equipped with*	15	14	0,92	0,00	3,03
43	*ejection fraction*	21	9	0,91	0,15	0,79
44	*ranging from*	66	49	0,91	0,01	1,94
45	*fail to*	13	13	0,91	0,00	3,21
46	*deal with*	10	8	0,90	0,00	3,19
47	*corresponds to*	11	11	0,90	0,00	3,28
48	*refer to*	11	14	0,90	0,00	3,28
49	*multiply by*	26	13	0,89	0,00	2,43
50	*existence of*	13	13	0,89	0,00	3,51
51	*smooth muscle*	51	12	0,88	0,25	0,55
52	*adjusting for*	66	34	0,88	0,00	2,28
53	*failed to*	68	43	0,88	0,00	2,47
54	*attributed to*	46	36	0,88	0,00	2,65
55	*inability to*	17	99	0,88	0,00	3,08
56	*underscore the*	12	11	0,88	0,00	3,62
57	*bipolar disorder*	20	7	0,87	0,16	0,75
58	*depends on*	21	20	0,87	0,00	2,31
59	*interacts with*	15	9	0,87	0,00	3,00
60	*composed of*	41	30	0,87	0,00	3,00
61	*predominance of*	10	10	0,87	0,00	3,61
62	*rather than*	148	96	0,86	0,05	1,22
63	*commercially available*	12	8	0,86	0,02	1,61
64	*served as*	19	29	0,86	0,00	2,48
65	*unresponsive to*	7	5	0,86	0,00	3,45
66	*fluorescently labeled*	11	7	0,85	0,10	0,94

67	*logistic regression*	73	46	0,84	0,21	0,60
68	*antiretroviral therapy*	169	11	0,84	0,16	0,73
69	*suggest that*	416	165	0,84	0,05	1,26
70	*converting enzyme*	19	18	0,83	0,14	0,78
71	*usual care*	115	10	0,83	0,14	0,78
72	*associated with*	1231	219	0,83	0,07	1,06
73	*contribute to*	133	77	0,83	0,01	2,16
74	*degrees of*	19	16	0,83	0,00	3,31
75	*virally suppressed*	9	5	0,82	0,15	0,74
76	*statistically significant*	162	79	0,82	0,13	0,79
77	*led to*	173	132	0,82	0,01	2,04
78	*dendritic cells*	28	17	0,82	0,01	2,13
79	*compatible with*	15	9	0,82	0,00	2,97
80	*noninferior to*	16	6	0,82	0,00	3,08
81	*exome sequencing*	17	6	0,81	0,07	1,05
82	*analogous to*	14	11	0,81	0,00	3,12
83	*receipt of*	22	18	0,81	0,00	3,24
84	*biophysical properties*	8	5	0,80	0,08	0,99
85	*consistent with*	437	163	0,80	0,03	1,50
86	*depend on*	12	12	0,80	0,00	2,51
87	*surrounded by*	17	10	0,80	0,00	2,57
88	*reliance on*	8	7	0,80	0,00	2,69
89	*aspects of*	37	25	0,80	0,00	3,01
90	*relates to*	9	8	0,80	0,00	3,31
91	*rounds of*	15	15	0,80	0,00	3,40
92	*achievement of*	10	7	0,80	0,00	3,58
93	*regarded as*	20	13	0,79	0,00	2,42
94	*occurrence of*	67	41	0,79	0,00	2,75
95	*convincing evidence*	7	6	0,78	0,01	1,91
96	*amounts of*	67	40	0,78	0,00	2,74
97	*contraindications to*	8	8	0,78	0,00	3,36
98	*examples of*	9	8	0,78	0,00	3,61
99	*calf serum*	10	10	0,77	0,02	1,63
100	*mononuclear cells*	24	16	0,77	0,01	2,17
101	*focusing on*	17	15	0,77	0,00	2,35
102	*resuspended in*	44	29	0,77	0,00	2,80
103	*predictor of*	38	23	0,77	0,00	2,98
104	*participated in*	20	14	0,77	0,00	3,14
105	*willing to*	11	12	0,77	0,00	3,21
106	*intellectual disability*	22	9	0,76	0,15	0,71
107	*adipose tissue*	35	7	0,76	0,05	1,15
108	*emergence of*	20	13	0,76	0,00	3,26
109	*depicted in*	11	8	0,76	0,00	3,39
110	*continuation of*	12	23	0,76	0,00	3,48
111	*establishment of*	12	9	0,76	0,00	3,48
112	*hallmarks of*	8	6	0,76	0,00	3,66
113	*prostate cancer*	122	7	0,75	0,12	0,80
114	*vast majority*	6	6	0,75	0,06	1,07
115	*nucleic acid*	15	8	0,75	0,03	1,36

116	*subdivided into*	9	7	0,75	0,01	1,91
117	*serve as*	41	24	0,75	0,01	2,09
118	*respond to*	29	38	0,75	0,00	2,78
119	*relating to*	10	11	0,75	0,00	3,24
120	*aspect of*	11	11	0,75	0,00	3,51
121	*focused on*	68	50	0,74	0,01	1,73
122	*supported by*	78	54	0,74	0,01	1,87
123	*supplemented with*	68	40	0,74	0,00	2,28
124	*believed to*	19	16	0,74	0,00	2,96
125	*characterization of*	24	17	0,74	0,00	3,16
126	*supplemental figure*	612	46	0,73	0,21	0,55
127	*beats per*	11	5	0,73	0,01	2,04
128	*correlated with*	110	55	0,72	0,01	2,05
129	*attempt to*	11	10	0,72	0,00	3,18
130	*nk cells*	44	7	0,71	0,01	1,87
131	*account for*	119	85	0,71	0,01	1,93
132	*arising from*	10	8	0,71	0,00	2,65
133	*variations in*	34	21	0,71	0,00	2,87
134	*consequence of*	33	28	0,71	0,00	3,01
135	*removal of*	32	21	0,71	0,00	3,02
136	*reversal of*	23	8	0,71	0,00	3,16
137	*parts of*	21	15	0,71	0,00	3,21
138	*contrary to*	8	7	0,71	0,00	3,31
139	*displacement of*	9	5	0,71	0,00	3,58
140	*suggesting that*	210	107	0,70	0,02	1,48
141	*latently infected*	7	5	0,70	0,01	1,74
142	*lead to*	125	79	0,70	0,01	2,12
143	*comes from*	14	16	0,70	0,00	2,49
144	*susceptible to*	36	25	0,70	0,00	2,66
145	*rely on*	7	8	0,70	0,00	2,69
146	*instructed to*	13	9	0,70	0,00	3,10

Appendix A.3: Unidirectionals2_ Word 2 as strongest predictor ($-4 \leq Log_{10} \leq -0.51$)

		Freq	Range	$\Delta P_{2\|1}$	$\Delta P_{1\|2}$	Log_{10}
1	*linked immunosorbent*	18	18	0,08	1,00	-1,10
2	*life expectancy*	56	9	0,12	1,00	-0,94
3	*cerebral palsy*	16	5	0,14	1,00	-0,85
4	*orally bioavailable*	8	7	0,15	1,00	-0,83
5	*hemoglobin A1c*	16	7	0,23	1,00	-0,65
6	*carbon dioxide*	8	7	0,27	1,00	-0,57
7	*we reasoned*	16	9	0,00	1,00	-2,59
8	*figure legends*	13	13	0,00	1,00	-2,36
9	*per kilogram*	10	7	0,01	1,00	-2,22
10	*per deciliter*	52	12	0,03	1,00	-1,50
11	*a priori*	27	28	0,00	0,99	-2,80
12	*per milliliter*	46	16	0,03	0,98	-1,55

13	*in situ*	35	20	0,00	0,97	-2,99
14	*in utero*	38	7	0,00	0,97	-2,96
15	*in vitro*	331	78	0,01	0,97	-2,02
16	*the remainder*	12	11	0,00	0,96	-3,66
17	*at least*	572	186	0,09	0,94	-1,01
18	*applied biosystems*	48	29	0,23	0,94	-0,62
19	*and colleagues*	81	29	0,00	0,94	-2,72
20	*myocardial infarctions*	11	7	0,05	0,92	-1,27
21	*to visualize*	15	11	0,00	0,92	-3,15
22	*we undertook*	26	14	0,00	0,92	-2,34
23	*the same*	533	175	0,01	0,92	-1,99
24	*to neutralize*	12	7	0,00	0,91	-3,24
25	*to treat*	236	122	0,01	0,91	-1,95
26	*western blotting*	53	22	0,28	0,90	-0,51
27	*is conceivable*	9	9	0,00	0,90	-2,76
28	*cell strainer*	9	10	0,00	0,90	-2,41
29	*for example*	222	97	0,02	0,90	-1,76
30	*dose escalation*	27	7	0,03	0,90	-1,51
31	*linkage disequilibrium*	17	7	0,23	0,89	-0,58
32	*to verify*	19	18	0,00	0,89	-3,03
33	*to obtain*	66	50	0,00	0,89	-2,49
34	*as follows*	65	47	0,01	0,89	-1,96
35	*to determine*	353	135	0,02	0,89	-1,77
36	*case fatality*	15	8	0,04	0,88	-1,39
37	*policy makers*	15	7	0,16	0,88	-0,73
38	*to prepare*	9	8	0,00	0,88	-3,36
39	*as evidenced*	15	17	0,00	0,88	-2,59
40	*least squares*	20	7	0,03	0,87	-1,42
41	*to discriminate*	8	7	0,00	0,87	-3,40
42	*to infer*	8	7	0,00	0,87	-3,40
43	*the entire*	100	56	0,00	0,87	-2,70
44	*to ensure*	58	46	0,00	0,87	-2,54
45	*the latter*	64	52	0,00	0,86	-2,89
46	*to calculate*	43	36	0,00	0,86	-2,67
47	*to deliver*	26	15	0,00	0,85	-2,88
48	*mass spectrometry*	27	22	0,13	0,84	-0,80
49	*to enroll*	11	15	0,00	0,83	-3,24
50	*to assess*	325	158	0,01	0,83	-1,77
51	*both sexes*	29	13	0,02	0,83	-1,68
52	*viral loads*	14	7	0,03	0,82	-1,41
53	*were instructed*	15	10	0,00	0,82	-2,86
54	*common terminology*	9	13	0,02	0,82	-1,59
55	*flow cytometer*	13	12	0,05	0,81	-1,26
56	*to maximize*	14	10	0,00	0,81	-3,12
57	*the original*	107	58	0,00	0,81	-2,64

58	*to generate*	107	53	0,00	0,81	-2,24
59	*to evaluate*	117	68	0,01	0,81	-2,21
60	*to detect*	175	109	0,01	0,81	-2,03
61	*in vivo*	402	77	0,01	0,81	-1,85
62	*to compensate*	9	7	0,00	0,80	-3,31
63	*to optimize*	9	7	0,00	0,80	-3,31
64	*not necessarily*	15	13	0,00	0,79	-2,35
65	*cell suspensions*	19	9	0,01	0,79	-2,03
66	*to examine*	86	60	0,00	0,77	-2,32
67	*to compare*	119	79	0,01	0,77	-2,17
68	*computed tomography*	13	12	0,19	0,76	-0,60
69	*years ago*	13	12	0,01	0,76	-2,08
70	*p falciparum*	77	7	0,03	0,75	-1,43
71	*e coli*	50	17	0,08	0,75	-0,99
72	*to achieve*	98	56	0,00	0,74	-2,24
73	*aortic valve*	44	5	0,15	0,73	-0,68
74	*per liter*	38	15	0,02	0,73	-1,50
75	*to receive*	171	68	0,01	0,72	-1,99
76	*to address*	92	70	0,00	0,71	-2,25
77	*life span*	21	9	0,04	0,70	-1,21
78	*t helper*	19	13	0,01	0,70	-1,96

Appendix B: MCRC Collocations

Appendix B.1: Bidirectionals_ Relatively Strong Mutual Attraction (-0.5 ≤ Log_{10} ≤ +0.5)

		Freq	Range	$\Delta P_{2\mid 1}$	$\Delta P_{1\mid 2}$	Log_{10}
1	*even though*	49	38	0,11	0,36	-0,50
2	*acid-fast bacilli*	14	10	0,20	0,64	-0,50
3	*pleural effusions*	20	13	0,14	0,45	-0,50
4	*sleep apnea*	12	10	0,26	0,80	-0,50
5	*atrial fibrillation*	81	43	0,25	0,75	-0,48
6	*cell carcinoma*	107	42	0,10	0,30	-0,47
7	*per minute*	123	79	0,22	0,63	-0,47
8	*diabetes mellitus*	82	58	0,36	1,00	-0,44
9	*developmental milestones*	10	10	0,26	0,71	-0,43
10	*bone marrow*	207	66	0,34	0,91	-0,43
11	*lumbar puncture*	39	26	0,30	0,78	-0,42
12	*left eye*	248	62	0,12	0,30	-0,39
13	*visual acuity*	220	64	0,41	0,99	-0,39
14	*vitamin b12*	38	15	0,21	0,50	-0,38
15	*right eye*	276	62	0,14	0,34	-0,37
16	*blood pressure*	336	225	0,21	0,49	-0,36
17	*night sweats*	20	15	0,37	0,83	-0,35
18	*cell count*	177	140	0,17	0,38	-0,35

19	*risk factors*	212	121	0,22	0,48	-0,34
20	*breath sounds*	23	18	0,21	0,44	-0,33
21	*angiotensin-converting enzyme*	16	16	0,42	0,89	-0,32
22	*cranial nerves*	13	11	0,16	0,32	-0,32
23	*live births*	19	19	0,43	0,86	-0,30
24	*cytokine storm*	20	10	0,38	0,71	-0,28
25	*foreign bodies*	25	10	0,25	0,46	-0,27
26	*original magnification*	20	10	0,27	0,50	-0,26
27	*systolic murmur*	21	22	0,19	0,34	-0,25
28	*oxygen saturation*	81	66	0,37	0,64	-0,24
29	*autosomal recessive*	46	20	0,47	0,81	-0,24
30	*lymph nodes*	139	77	0,49	0,81	-0,22
31	*ct scan*	351	163	0,35	0,58	-0,22
32	*herpes simplex*	15	10	0,64	0,98	-0,18
33	*heart failure*	161	56	0,22	0,33	-0,18
34	*mycophenolate mofetil*	30	20	0,70	1,00	-0,16
35	*mechanical ventilation*	28	21	0,23	0,33	-0,15
36	*lactate dehydrogenase*	42	34	0,45	0,59	-0,12
37	*myocardial infarction*	96	45	0,48	0,63	-0,12
38	*poorly differentiated*	28	14	0,29	0,37	-0,11
39	*lymph node*	129	107	0,46	0,56	-0,09
40	*carbon dioxide*	17	13	0,85	1,00	-0,07
41	*autosomal dominant*	49	32	0,50	0,58	-0,06
42	*et al*	543	236	0,87	0,97	-0,05
43	*parenteral nutrition*	14	13	0,45	0,50	-0,04
44	*pleural effusion*	48	30	0,35	0,37	-0,04
45	*needle aspiration*	42	23	0,34	0,36	-0,03
46	*passed away*	12	11	0,32	0,34	-0,02
47	*lacrimal sac*	16	10	0,31	0,32	-0,02
48	*alternative diagnoses*	11	10	0,07	0,07	-0,02
49	*p waves*	80	29	0,29	0,30	-0,01
50	*iliac fossa*	22	10	0,38	0,37	0,01
51	*coronary artery*	62	42	0,31	0,29	0,02
52	*alkaline phosphatase*	47	31	1,00	0,94	0,03
53	*staphylococcus aureus*	34	24	0,64	0,60	0,03
54	*hemodynamic instability*	17	13	0,33	0,29	0,05
55	*ejection fraction*	92	52	0,93	0,81	0,06
56	*ethics committee*	10	13	0,62	0,53	0,07
57	*nasopharyngeal swab*	19	14	0,42	0,34	0,09
58	*multidisciplinary team*	35	25	0,30	0,23	0,10
59	*infective endocarditis*	47	17	0,54	0,42	0,11
60	*cognitive impairment*	34	19	0,32	0,25	0,11
61	*aspartate aminotransferase*	40	36	0,67	0,51	0,11
62	*computed tomography*	310	218	0,89	0,68	0,12
63	*broad-spectrum antibiotics*	48	36	0,48	0,34	0,15
64	*septic shock*	21	15	0,32	0,22	0,15

65	*intravascular coagulation*	29	11	0,36	0,25	0,16
66	*st segment*	95	27	0,58	0,39	0,17
67	*rheumatoid arthritis*	37	22	0,46	0,29	0,20
68	*restricted diffusion*	19	11	0,42	0,26	0,20
69	*weight loss*	104	72	0,33	0,18	0,25
70	*mental status*	57	27	0,35	0,19	0,25
71	*adverse effects*	77	48	0,39	0,21	0,27
72	*abdominal pain*	224	102	0,31	0,17	0,27
73	*transthoracic echocardiogram*	49	40	0,64	0,33	0,29
74	*vocal cord*	40	10	0,66	0,32	0,32
75	*electron microscopy*	18	11	0,67	0,32	0,32
76	*falciform ligament*	24	10	0,57	0,27	0,33
77	*natriuretic peptide*	22	18	1,00	0,46	0,34
78	*radiofrequency ablation*	17	10	0,77	0,34	0,36
79	*critically ill*	24	15	0,71	0,30	0,37
80	*escherichia coli*	17	14	0,94	0,40	0,38
81	*mitral valve*	64	23	0,56	0,23	0,38
82	*reference range*	107	46	0,53	0,21	0,39
83	*our patient*	943	345	0,38	0,15	0,40
84	*C-reactive protein*	120	100	0,67	0,25	0,44
85	*twice daily*	97	52	0,60	0,21	0,45
86	*predictive value*	34	17	0,61	0,22	0,45
87	*gastroesophageal reflux*	22	10	0,96	0,33	0,46
88	*dry cough*	20	12	0,39	0,14	0,46
89	*thromboembolic events*	21	1	0,34	0,12	0,46
90	*soft tissue*	175	86	0,63	0,22	0,47
91	*checkpoint inhibitors*	28	10	0,52	0,17	0,47
92	*b12 deficiency*	23	11	0,30	0,10	0,48
93	*vital signs*	119	97	0,69	0,22	0,49
94	*physical examination*	286	222	0,65	0,21	0,49
95	*cesarean section*	16	10	0,80	0,26	0,49
96	*operating room*	35	30	0,62	0,20	0,49
97	*platelet count*	88	58	0,59	0,19	0,50

Appendix B.2*: *Unidirectionals1_ Word 1 as strongest predictor (0.51 ≤ Log_{10} ≤ 4)

		Freq	**Range**	**$\Delta P_{2\|1}$**	**$\Delta P_{1\|2}$**	**Log_{10}**
1	*nasogastric tube*	14	12	1,00	0,13	0,87
2	*adrenocorticotropic hormone*	17	12	1,00	0,09	1,06
3	*photodynamic therapy*	10	10	1,00	0,01	2,19
4	*cerebrospinal fluid*	115	62	0,99	0,23	0,64
5	*ruled out*	131	87	0,98	0,26	0,58
6	*according to*	290	175	0,98	0,02	1,79
7	*owing to*	106	76	0,98	0,01	2,23

8	*nervous system*	128	81	0,97	0,29	0,53
9	*due to*	1530	505	0,97	0,08	1,07
10	*ruling out*	29	25	0,97	0,06	1,23
11	*multiply by*	76	31	0,97	0,02	1,74
12	*depending on*	77	62	0,97	0,01	1,83
13	*accounting for*	43	34	0,97	0,00	2,29
14	*characterised by*	73	44	0,96	0,02	1,75
15	*paucity of*	22	17	0,96	0,00	3,24
16	*complain of*	20	19	0,96	0,00	3,28
17	*irrespective of*	16	14	0,96	0,00	3,38
18	*consist of*	14	13	0,96	0,00	3,43
19	*carcinoembryonic antigen*	19	17	0,95	0,11	0,95
20	*folic acid*	20	15	0,95	0,07	1,11
21	*reserved for*	21	15	0,95	0,00	2,59
22	*tends to*	35	34	0,95	0,00	2,70
23	*attributable to*	32	29	0,95	0,00	2,74
24	*blurred vision*	31	19	0,94	0,10	0,96
25	*accounts for*	58	47	0,94	0,01	2,15
26	*aside from*	16	12	0,94	0,01	2,26
27	*attributed to*	83	62	0,94	0,00	2,32
28	*mistaken for*	18	15	0,94	0,00	2,66
29	*basement membrane*	14	10	0,93	0,08	1,09
30	*depends on*	56	50	0,93	0,01	1,95
31	*suggestive of*	224	149	0,93	0,01	2,22
32	*subjected to*	18	16	0,93	0,00	2,98
33	*completion of*	36	30	0,93	0,00	3,01
34	*thrombocytopenic purpura*	12	10	0,92	0,28	0,52
35	*citrullinated peptide*	12	12	0,92	0,25	0,57
36	*excisional biopsy*	23	13	0,92	0,03	1,47
37	*tend to*	49	46	0,92	0,00	2,54
38	*cardiogenic shock*	20	10	0,91	0,21	0,63
39	*characterized by*	237	157	0,91	0,06	1,22
40	*consistent with*	297	195	0,91	0,02	1,73
41	*surrounded by*	33	26	0,91	0,01	2,07
42	*lack of*	204	155	0,91	0,01	2,25
43	*composed of*	60	48	0,91	0,00	2,78
44	*consisted of*	35	30	0,91	0,00	3,01
45	*ranging from*	66	53	0,90	0,02	1,62
46	*carried out*	62	45	0,89	0,12	0,86
47	*depend on*	17	17	0,89	0,00	2,45
48	*contributed to*	52	45	0,89	0,00	2,50
49	*contribute to*	50	45	0,89	0,00	2,52
50	*compatible with*	46	34	0,89	0,00	2,52
51	*regardless of*	37	33	0,89	0,00	2,98
52	*capable of*	13	13	0,89	0,00	3,43
53	*responsible for*	73	57	0,88	0,01	2,02
54	*amount of*	72	56	0,88	0,00	2,69
55	*accompanied by*	94	79	0,87	0,02	1,60
56	*leading to*	255	187	0,87	0,01	1,80

57	*aspect of*	58	48	0,87	0,00	2,77
58	*restoration of*	20	17	0,87	0,00	3,24
59	*nucleic acid*	12	20	0,86	0,04	1,29
60	*resulted in*	123	101	0,86	0,01	2,18
61	*amphotericin B*	41	14	0,85	0,05	1,24
62	*parts of*	41	34	0,85	0,00	2,92
63	*counting fingers*	16	10	0,84	0,23	0,56
64	*pseudomembranous colitis*	10	21	0,83	0,18	0,66
65	*computerized tomography*	10	10	0,83	0,02	1,58
66	*transferred to*	97	68	0,83	0,01	2,20
67	*inability to*	29	28	0,83	0,00	2,73
68	*continues to*	28	25	0,83	0,00	2,74
69	*corresponds to*	11	10	0,83	0,00	3,14
70	*persistence of*	20	19	0,83	0,00	3,22
71	*hyperbaric oxygen*	18	11	0,82	0,08	1,00
72	*thrombolytic therapy*	33	10	0,82	0,02	1,58
73	*aware of*	84	73	0,82	0,00	2,59
74	*complained of*	57	47	0,82	0,00	2,76
75	*elected to*	15	17	0,82	0,00	3,00
76	*contributor to*	10	10	0,82	0,00	3,18
77	*gold standard*	43	40	0,81	0,20	0,61
78	*rule out*	96	77	0,81	0,19	0,63
79	*caused by*	356	213	0,81	0,08	0,99
80	*antiretroviral therapy*	22	10	0,81	0,01	1,75
81	*indicative of*	34	30	0,81	0,00	2,98
82	*contributes to*	14	13	0,81	0,00	3,03
83	*concluded that*	41	35	0,80	0,01	1,94
84	*correlate with*	18	18	0,80	0,00	2,89
85	*avoidance of*	16	15	0,80	0,00	3,30
86	*emphasizes the*	11	11	0,80	0,00	3,56
87	*chief complaint*	13	13	0,79	0,23	0,54
88	*productive cough*	19	15	0,79	0,13	0,79
89	*vast majority*	15	14	0,79	0,12	0,81
90	*constitutional symptoms*	23	18	0,79	0,02	1,71
91	*confined to*	13	11	0,79	0,00	3,05
92	*subset of*	25	22	0,79	0,00	3,10
93	*coexistence of*	19	12	0,79	0,00	3,22
94	*exploratory laparotomy*	14	10	0,78	0,17	0,65
95	*exceedingly rare*	22	21	0,78	0,02	1,63
96	*regarded as*	18	15	0,78	0,00	2,42
97	*consisting of*	40	35	0,78	0,00	2,89
98	*transitioned to*	12	11	0,78	0,00	3,08
99	*amounts of*	23	22	0,78	0,00	3,13
100	*rather than*	99	82	0,77	0,10	0,91
101	*originate from*	24	17	0,77	0,01	2,00
102	*led to*	136	151	0,77	0,01	2,02
103	*initiation of*	104	68	0,77	0,00	2,47
104	*such as*	1058	455	0,76	0,17	0,64
105	*susceptible to*	24	20	0,76	0,00	2,77

106	*proportion of*	35	33	0,76	0,00	2,94
107	*likelihood of*	28	25	0,76	0,00	3,04
108	*define the*	13	12	0,76	0,00	3,47
109	*sickle cell*	47	13	0,75	0,04	1,22
110	*ranged from*	24	21	0,75	0,01	1,98
111	*number of*	186	142	0,75	0,00	2,21
112	*correlated with*	20	18	0,75	0,00	2,82
113	*impression of*	15	10	0,75	0,00	3,30
114	*based on*	499	294	0,74	0,09	0,90
115	*followed by*	319	224	0,74	0,07	1,00
116	*arising from*	59	36	0,74	0,02	1,59
117	*referred to*	149	121	0,74	0,01	1,82
118	*attached to*	33	20	0,73	0,00	2,61
119	*stimulating hormone*	39	26	0,72	0,20	0,55
120	*hearing loss*	100	21	0,72	0,18	0,61
121	*adipose tissue*	23	11	0,72	0,03	1,40
122	*along with*	229	141	0,72	0,01	1,74
123	*resulting in*	201	157	0,72	0,01	1,89
124	*suffer from*	16	16	0,72	0,01	2,14
125	*precordial leads*	56	10	0,71	0,14	0,71
126	*insight into*	15	12	0,71	0,02	1,46
127	*apart from*	43	34	0,71	0,01	1,71
128	*originating from*	25	17	0,71	0,01	1,94
129	*foreign body*	70	18	0,70	0,15	0,66
130	*confused with*	20	18	0,70	0,00	2,79
131	*array of*	14	12	0,70	0,00	3,30
132	*creatine kinase*	18	15	0,69	0,21	0,51
133	*probability of*	29	19	0,69	0,00	2,98
134	*except for*	80	69	0,68	0,01	1,87
135	*treated with*	512	262	0,65	0,03	1,35
136	*reversal of*	11	11	0,44	0,00	3,22
137	*goal of*	10	10	0,40	0,00	3,22

Appendix B.3: Unidirectionals2_ Word 2 as strongest predictor (-4 ≤ Log₁₀ ≤ -0.51)

		Freq	Range	$\Delta P_{2\|1}$	$\Delta P_{1\|2}$	Log_{10}
1	*flow cytometry*	17	12	0,10	1,00	-1,01
2	*linked immunosorbent*	15	11	0,17	1,00	-0,77
3	*en bloc*	11	10	0,27	1,00	-0,57
4	*in vitro*	22	17	0,00	0,98	-2,98
5	*to convert*	78	35	0,00	0,98	-2,37
6	*basal ganglia*	31	15	0,28	0,97	-0,54
7	*in situ*	57	44	0,00	0,96	-2,56
8	*the rarest*	10	10	0,00	0,95	-3,67
9	*the umbilicus*	15	10	0,00	0,95	-3,50
10	*T-wave inversions*	21	14	0,07	0,95	-1,13
11	*immune checkpoint*	51	11	0,16	0,94	-0,76
12	*de novo*	12	10	0,15	0,92	-0,78

13	*the commonest*	16	10	0,00	0,90	-3,44
14	*to explore*	11	11	0,00	0,90	-3,18
15	*the same*	262	176	0,01	0,90	-2,23
16	*to ascertain*	10	10	0,00	0,89	-3,22
17	*the mainstay*	11	11	0,00	0,89	-3,18
18	*at least*	127	103	0,04	0,89	-1,36
19	*the entire*	52	43	0,00	0,88	-2,93
20	*the latter*	46	37	0,00	0,87	-2,97
21	*general practitioner*	20	19	0,06	0,87	-1,16
22	*hemoglobin a1c*	14	10	0,10	0,87	-0,94
23	*the aforementioned*	28	23	0,00	0,86	-3,18
24	*was readmitted*	20	16	0,00	0,86	-2,79
25	*to date*	105	83	0,01	0,86	-2,18
26	*to remove*	26	23	0,00	0,85	-2,78
27	*the largest*	31	25	0,00	0,84	-3,13
28	*the exact*	55	50	0,00	0,84	-2,88
29	*to determine*	73	58	0,00	0,84	-2,33
30	*blood sugar*	22	15	0,01	0,84	-1,79
31	*the opposite*	14	11	0,00	0,83	-3,47
32	*to evaluate*	97	69	0,01	0,83	-2,20
33	*to avoid*	108	84	0,01	0,83	-2,15
34	*his wife*	15	11	0,01	0,83	-2,12
35	*follow ups*	18	14	0,02	0,82	-1,62
36	*to treat*	118	129	0,01	0,81	-2,10
37	*to address*	13	11	0,00	0,80	-3,05
38	*to assess*	59	52	0,00	0,80	-2,40
39	*to relieve*	17	16	0,00	0,79	-2,94
40	*we believe*	61	38	0,03	0,79	-1,46
41	*multiple myeloma*	72	14	0,09	0,79	-0,94
42	*to diagnose*	73	59	0,00	0,78	-2,29
43	*the highest*	27	23	0,00	0,77	-3,15
44	*to confirm*	31	18	0,00	0,77	-2,23
45	*chest radiograph*	34	32	0,05	0,77	-1,19
46	*family members*	33	24	0,08	0,77	-1,01
47	*transplant recipients*	29	10	0,15	0,76	-0,70
48	*retinal detachment*	69	23	0,19	0,76	-0,60
49	*to achieve*	64	47	0,00	0,75	-2,34
50	*lower extremity*	75	36	0,11	0,75	-0,85
51	*anterior chamber*	89	23	0,19	0,75	-0,59
52	*to manage*	48	39	0,00	0,74	-2,46
53	*to alleviate*	10	10	0,00	0,73	-3,18
54	*to investigate*	32	29	0,00	0,73	-2,62
55	*for example*	65	48	0,01	0,73	-1,99
56	*to maintain*	48	40	0,00	0,72	-2,44
57	*was extubated*	16	14	0,00	0,71	-2,81
58	*abdominal distension*	24	16	0,03	0,71	-1,32
59	*respiratory distress*	72	36	0,12	0,71	-0,76
60	*plasma exchange*	48	18	0,16	0,71	-0,65
61	*the former*	12	13	0,00	0,70	-3,47

62	*to clarify*	13	11	0,00	0,70	-3,00
63	*to proceed*	13	12	0,00	0,70	-3,00
64	*to seek*	13	11	0,00	0,70	-3,00
65	*to detect*	57	49	0,00	0,70	-2,35
66	*after completing*	12	10	0,00	0,70	-2,27
67	*on exertion*	31	26	0,01	0,70	-2,09
68	*while awaiting*	16	11	0,03	0,70	-1,41
69	*white matter*	62	24	0,15	0,70	-0,66

Appendix C: Multiword Collocations

MRAC

Trigram	Freq	ΔP (A+B)	ΔP C	Log10
what we believe	12	0.67	0.19	-0.5
power to detect	44	0.71	0.21	-0.5
eastern cooperative oncology	6	1.00	0.30	-0.5
fatal and nonfatal	8	0.73	0.22	-0.5
site directed mutagenesis	6	1.00	0.32	-0.5
Wilcoxon rank sum	19	1.00	0.32	-0.5
be clinically meaningful	5	0.45	0.15	-0.5
the endoplasmic reticulum	9	1.00	0.33	-0.5
the life span	5	0.50	0.17	-0.5
accuracy and completeness	12	0.86	0.29	-0.5
the extracellular matrix	10	0.38	0.13	-0.5
was reverse transcribed	10	1.00	0.36	-0.4
sensitivity and specificity	24	0.60	0.21	-0.4
thermo fisher scientific	27	0.96	0.35	-0.4
in North America	10	0.50	0.18	-0.4
the funding source	46	0.92	0.34	-0.4
to better understand	9	0.41	0.15	-0.4
fruits and vegetables	7	0.78	0.29	-0.4
transient ischemic attack	9	0.75	0.28	-0.4
the spinal cord	9	0.47	0.18	-0.4
necrosis factor α	5	0.45	0.18	-0.4
polymerase chain reaction	40	0.98	0.38	-0.4
a blinded fashion	14	0.67	0.27	-0.4
stem cell transplantation	40	0.58	0.25	-0.4
years or older	96	0.72	0.31	-0.4
single nucleotide polymorphisms	11	0.48	0.20	-0.4
in nonhuman primates	8	0.89	0.38	-0.4
converting enzyme ace	9	0.47	0.21	-0.4
intensive care unit	21	0.60	0.27	-0.3
electronic health record	7	0.39	0.17	-0.3
embedded in paraffin	13	0.50	0.23	-0.3
directly or indirectly	5	0.56	0.26	-0.3
bovine serum albumin	15	0.52	0.25	-0.3
activated cell sorting	10	1.00	0.48	-0.3

interpreted with caution	7	0.64	0.30	-0.3
the intestinal epithelium	20	0.47	0.22	-0.3
written informed consent	112	1.00	0.49	-0.3
at room temperature	88	0.97	0.48	-0.3
by flow cytometry	65	0.94	0.46	-0.3
generalized estimating equations	8	0.62	0.31	-0.3
were as follows	32	0.82	0.44	-0.3
by western blot	15	0.37	0.21	-0.3
cell signaling technology	23	0.68	0.37	-0.3
in liquid nitrogen	11	0.73	0.41	-0.3
proof of principle	13	0.43	0.25	-0.2
proof of concept	16	0.53	0.33	-0.2
in close proximity	10	0.77	0.48	-0.2
phorbol 12 myristate	5	1.00	0.63	-0.2
the United States	73	0.75	0.47	-0.2
single nucleotide polymorphism	11	0.48	0.31	-0.2
95 confidence interval	82	0.67	0.45	-0.2
intention to treat	163	0.95	0.64	-0.2
by western blotting	23	0.57	0.39	-0.2
were well balanced	12	0.40	0.27	-0.2
and vice versa	14	1.00	0.74	-0.1
in meters squared	13	0.76	0.57	-0.1
acid fast bacilli	6	0.86	0.67	-0.1
heterotrimeric guanine nucleotide-binding	6	1.00	0.86	-0.1
randomisation and masking	12	0.52	0.46	-0.1
hematoxylin and eosin	13	0.81	0.76	0.0
in situ hybridization	19	0.54	0.51	0.0
fixed and permeabilized	7	0.37	0.35	0.0
per cubic millimeter	33	1.00	1.00	0.0
diamidino 2 phenylindole	7	1.00	1.00	0.0
nicotinamide adenine dinucleotide	5	1.00	1.00	0.0
nitric oxide synthase	9	0.36	0.37	0.0
we searched PubMed	13	0.46	0.52	0.0
enzyme linked immunosorbent	18	0.75	1.00	0.1
college of cardiology	9	0.45	0.60	0.1
available on request	5	0.31	0.42	0.1
artery bypass grafting	9	0.50	0.69	0.1
single cell suspensions	13	0.38	0.54	0.2
is worth noting	5	0.50	0.71	0.2
medical research council	10	0.34	0.50	0.2
the figure legends	11	0.55	0.85	0.2
alexa fluor 488	12	0.32	0.55	0.2
length of stay	14	0.22	0.40	0.3
New England biolabs	5	0.50	1.00	0.3
institutional review boards	36	0.40	0.90	0.3
the common terminology	6	0.21	0.55	0.4
peripheral blood mononuclear	12	0.13	0.39	0.5
the United Kingdom	20	0.21	0.61	0.5

coronary artery bypass	18	0.16	0.51	0.5
statistical analysis plan	24	0.10	0.35	0.5

			MI	Freq	Range
4-words	1	*seafood omega-3 fatty acids*	43.67	11	6
	2	*ambient particulate matter pollution*	39.45	22	7
	3	*ret-mutant medullary thyroid cancer*	37.20	20	5
	4	*de novo cholesterol synthesis*	35.19	10	5
	5	*estimated glomerular filtration rate*	34.91	19	8
	6	*chronic obstructive pulmonary disease*	31.90	23	13
	7	*oral glucose tolerance test*	31.32	10	6
	8	*minutes at room temperature*	30.23	13	9
	9	*low-income and middle-income countries*	30.22	31	5
	10	*the saw palmetto extract*	29.08	11	5
	11	*induction of mixed chimerism*	28.37	30	6
	12	*analyzed by flow cytometry*	27.60	13	8
	13	*adults with down syndrome*	27.49	20	6
	14	*the institutional review board*	27.44	14	14
	15	*secondary end points included*	27.21	13	9
	16	*plays a critical role*	27.16	12	10
	17	*myocardial infarction or stroke*	26.88	15	6
	18	*food and drug administration*	26.45	27	24
	19	*National Institutes of Health*	26.08	36	30
	20	*the World Health Organization*	25.49	18	12
	21	*participants with down syndrome*	25.21	14	6
	22	*the second heart field*	25.20	38	5
	23	*death from cardiovascular causes*	25.01	11	5
	24	*the primary end point*	23.25	81	26
	25	*time to virus escape*	23.19	10	5
	26	*the statistical analysis plan*	22.95	16	13
	27	*the blinded intervention period*	22.77	11	6
	28	*the primary outcome measure*	22.12	16	11
	29	*patients with pancreatic cancer*	20.30	11	5
	30	*risk of cardiovascular events*	19.41	10	5

			MI	Freq	Range
5-words	1	*posterior circulation territory infarctlike lesions*	57.17	13	5
	2	*tobacco smoking including second-hand smoke*	50.52	19	6
	3	*recombinant interferon beta-1b and lopinavir-ritonavir*	45.00	8	4
	4	*patients with castration-resistant prostate cancer*	35.18	16	5
	5	*animal care and use committee*	34.24	9	9
	6	*patients with acute heart failure*	31.23	13	6
	7	*risk of coronary artery disease*	31.15	16	5
	8	*patients with type 2 diabetes*	28.04	22	6
	9	*patients in the placebo group*	19.99	36	9
6-words	1	*forced expiratory volume in 1 second*	51.00	9	7
	2	*patients with ret-mutant medullary thyroid cancer*	50.59	16	5
	3	*common terminology criteria for adverse events*	50.52	9	9
	4	*the consolidated standards of reporting trials*	47.43	6	6
	5	*institutional animal care and use committee*	47.36	15	15
	6	*the data and safety monitoring board*	39.25	14	9
	7	*Centers for Disease Control and Prevention*	38.42	16	16
7-words	1	*national institute of allergy and infectious diseases*	58.22	13	7
	2	*the common terminology criteria for adverse events*	54.93	6	6
	3	*an independent data and safety monitoring board*	54.40	10	10
	4	*adverse events of grade 3 or higher*	48.88	10	6

MCRC

Trigram	Freq	Delta *P* (A + B)	Delta *P* C	Log10
the ethics committee	5	0.83	0.26	-0.5
immune checkpoint inhibitors	26	0.51	0.16	-0.5
right bundle branch	29	0.91	0.29	-0.5
index of suspicion	58	0.71	0.23	-0.49
under general anesthesia	28	0.7	0.23	-0.48
highly active antiretroviral	7	0.78	0.26	-0.48
stem cell transplantation	15	0.33	0.11	-0.48
magnetic resonance imaging	238	0.82	0.27	-0.48
dyspnea on exertion	15	1	0.34	-0.47
a timely manner	6	0.55	0.19	-0.46
deep tendon reflexes	13	0.87	0.3	-0.46
fine needle aspiration	35	0.85	0.3	-0.46
serum protein electrophoresis	14	0.7	0.25	-0.45
calcium channel blockers	10	0.59	0.21	-0.45
multiple endocrine neoplasia	9	0.9	0.32	-0.45
blood cell count	113	0.67	0.24	-0.45
kept in mind	12	0.8	0.29	-0.44
generalized tonic clonic	5	0.83	0.31	-0.43
the diagnostic workup	7	0.1	0.04	-0.43
a multidisciplinary team	20	0.36	0.13	-0.42
b cell lymphoma	42	0.37	0.14	-0.41
of unknown origin	18	0.34	0.13	-0.41
sensitivity and specificity	22	0.63	0.24	-0.41
diagnosed as having	52	0.54	0.22	-0.4
polymerase chain reaction	91	1	0.4	-0.39
beats per minute	80	0.98	0.41	-0.38
control and prevention	11	0.39	0.17	-0.37
molecular weight heparin	24	0.75	0.32	-0.37
an autosomal dominant	22	0.58	0.26	-0.35
endoscopic retrograde cholangiopancreatography	6	1	0.46	-0.34
diagnostic testing approaches	16	0.47	0.22	-0.33
written informed consent	43	1	0.47	-0.33
palms and soles	6	0.75	0.35	-0.33
nose and throat	13	0.62	0.3	-0.32
her general practitioner	5	0.45	0.22	-0.32
chronic lymphocytic leukemia	8	0.36	0.18	-0.31
college of rheumatology	6	0.37	0.19	-0.3

pars plana vitrectomy	17	0.41	0.21	-0.3
arterial blood gas	21	0.64	0.32	-0.29
acute respiratory distress	19	0.37	0.19	-0.29
giant cell arteritis	12	0.75	0.39	-0.29
a lumbar puncture	17	0.65	0.34	-0.28
beta human chorionic	9	1	0.53	-0.28
st segment elevation	43	0.45	0.24	-0.27
highlights the importance	44	0.42	0.24	-0.25
institutional review board	20	1	0.59	-0.23
abdomen and pelvis	53	0.71	0.42	-0.23
low density lipoprotein	9	0.45	0.27	-0.22
the United States	39	0.83	0.53	-0.2
activated partial thromboplastin	14	1	0.64	-0.2
inspired oxygen fio2	6	0.55	0.35	-0.19
the emergency department	185	0.8	0.53	-0.18
an autosomal recessive	15	0.39	0.26	-0.18
hands and feet	16	0.59	0.4	-0.17
hematoxylin and eosin	24	1	0.69	-0.16
inversion recovery flair	9	0.41	0.29	-0.15
shortness of breath	78	0.99	0.71	-0.14
alert and oriented	19	0.66	0.47	-0.14
intensive care unit	105	0.77	0.56	-0.14
human chorionic gonadotropin	13	0.93	0.68	-0.13
the corpus callosum	21	1	0.78	-0.11
ground glass opacities	11	0.35	0.3	-0.08
gamma glutamyl transferase	5	0.83	0.71	-0.07
in bowel habits	6	0.5	0.43	-0.07
alanine aminotransferase alt	13	0.35	0.3	-0.07
right upper quadrant	23	0.4	0.35	-0.06
programmed death ligand	6	0.55	0.5	-0.04
vitamin k antagonist	7	0.32	0.29	-0.04
granulomatosis with polyangiitis	16	0.94	0.89	-0.02
acid fast bacilli	14	0.67	0.64	-0.02
systemic lupus erythematosus	40	0.93	0.89	-0.02
blood urea nitrogen	20	0.87	0.83	-0.02
the basal ganglia	13	0.42	0.41	-0.01
toxic epidermal necrolysis	10	1	1	0
league against rheumatism	5	0.83	0.83	0
enzyme linked immunosorbent	15	0.88	1	0.05
primary care provider	7	0.32	0.44	0.14

within normal limits	79	0.64	0.88	0.14
the anterior chamber	52	0.31	0.44	0.15
range of motion	25	0.23	0.37	0.2
urea nitrogen bun	6	0.27	0.46	0.23
smooth muscle actin	18	0.35	0.67	0.28
a diagnostic dilemma	11	0.19	0.37	0.29
in situ hybridization	23	0.4	0.88	0.34

			MI	Freq	Range
4-words	1	*relative afferent pupillary defect*	39.46	11	8
	2	*epidermal growth factor receptor*	33.67	18	12
	3	*fluorescence in situ hybridization*	33.52	11	9
	4	*diffuse large b-cell lymphoma*	33.14	16	7
	5	*best corrected visual acuity*	32.98	25	20
	6	*alternative diagnostic testing approaches*	32.45	16	16
	7	*human epidermal growth factor*	32.38	10	8
	8	*right bundle branch block*	31.90	26	16
	9	*left bundle branch block*	31.46	16	8
	10	*vascular endothelial growth factor*	31.41	18	14
	11	*left ventricular ejection fraction*	30.87	24	18
	12	*packed red blood cells*	30.32	12	11
	13	*type 2 diabetes mellitus*	29.97	23	19
	14	*fraction of inspired oxygen*	29.81	10	5
	15	*the world health organization*	29.28	23	22
	16	*brain magnetic resonance imaging*	29.11	25	23
	17	*the inferior vena cava*	29.10	16	8
	18	*cardiac magnetic resonance imaging*	28.30	14	10
	19	*white blood cell count*	28.27	40	38
	20	*activities of daily living*	28.25	14	13
	21	*with reduced ejection fraction*	28.18	12	7
	22	*chronic obstructive pulmonary disease*	28.01	12	12
	23	*upper respiratory tract infection*	27.87	14	7
	24	*acute respiratory distress syndrome*	27.80	16	13
	25	*vital signs were stable*	26.69	11	11
	26	*magnetic resonance imaging revealed*	26.60	11	11
	27	*complete blood cell count*	26.48	15	15
	28	*small cell lung cancer*	26.47	13	11
	29	*the internal jugular vein*	26.40	10	7
	30	*the intensive care unit*	26.20	50	41

31	*the central nervous system*	25.42	39	35
32	*further studies are needed*	25.29	10	9
33	*the right iliac fossa*	25.04	10	5
34	*high index of suspicion*	24.47	47	35
35	*upper and lower extremities*	23.66	20	14
36	*her medical history included*	23.64	11	11
37	*upper and lower limbs*	23.21	15	13
38	*physical examination was unremarkable*	23.04	22	22
39	*function tests were normal*	22.68	10	10
40	*a bone marrow biopsy*	22.67	11	13
41	*this case report describes*	22.60	13	11
42	*a magnetic resonance imaging*	22.60	10	10
43	*blood and urine cultures*	22.54	14	13
44	*a computed tomography scan*	22.41	10	10
45	*physical examination findings were*	21.96	11	11
46	*the right lower quadrant*	21.32	10	5
47	*this case report highlights*	21.23	10	8
48	*the anterior abdominal wall*	21.12	11	5
49	*medical history was significant*	20.85	19	19
50	*renal and liver function*	20.76	10	9
51	*days prior to presentation*	20.20	10	8
52	*the most common site*	20.00	13	11
53	*physical examination revealed a*	19.20	12	13

			MI	**Freq**	**Range**
5-words	1	*multisystem inflammatory syndrome in children*	36.28	11	5
	2	*his past medical history included*	35.02	8	8
	3	*patients with human immunodeficiency virus*	34.10	9	6
	4	*a white blood cell count*	32.01	21	19
	5	*a complete blood cell count*	30.72	11	11
6-words	1	*severe acute respiratory syndrome coronavirus 2*	58.21	15	10
	2	*human epidermal growth factor receptor 2*	52.83	7	6
	3	*heart failure with preserved ejection fraction*	51.14	6	5
	4	*written informed consent was obtained from*	50.11	6	6
	5	*heart failure with reduced ejection fraction*	49.39	10	5
	6	*vital signs were within normal limits*	47.04	7	7
	7	*centers for disease control and prevention*	44.56	9	9
	8	*required to publish the case details*	42.32	7	7

	9	*the us food and drug administration*	42.31	15	12
	10	*chest pain and shortness of breath*	39.40	8	6
	11	*with no significant past medical history*	39.32	9	8
7-words	1	*showed a white blood cell count of*	44.83	11	9

Appendix D: Example of list of Frames and their Most Frequent Fillers

frame	**nber of fillers**	
the * of the	507	end, use, basis, results, course, effect, time, duration, start, magnitude, design, analysis, presence, characteristics, ability, fidelity, nature, context, size, role, importance, rest, discretion, specificity, mean, date, sensitivity, capacity, integrity, writing, sum, distribution, strength, findings, quality, proportion, effectiveness, control, effects, beginning, members, majority, shape, accuracy, sponsors, impact, expression, function, robustness, significance, efficacy, sponsor, funders, level, incidence, day, remainder, regulation, bottom, publication, criteria, midpoint, pathogenesis, complexity, composition, addition, risk, length, surface, ratio, rate, development, activity, assessment, immunogenicity, spread, details, calculation, head, validity, adequacy, abundance, support, phenotype, interpretation, area, cause, square, absence, diversity, implementation, conclusion, percentage, severity, association, goal, principles, value, activation, standards, exception, prevalence, direction, funder, objectives, onset, timing, circumference, combination, peak, coefficients, intensity, genotype, base, release, region, cells, kinetics, vicinity, progression, mechanism, density, identity, choice, analyses, performance, cost, center, neutralization, chance, safety
in the * of	325	presence, absence, context, number, pathogenesis, setting, incidence, risk, lungs, development, rate, case, regulation, treatment, prevalence, plasma, frequency, serum, percentage, proportion, expression, range, course, use, rates, analysis, management, face, form, feces, blood, induction, subgroup, control, distribution, levels, region, identification, design, event, size, subset, colons, prevention, diagnosis , country, concentration, formation, middle, majority, livers, duration, amount, brains, production, cytoplasm, fibroblasts, aorta, pancreases, interpretation, aortas, generation, declaration, vicinity, nucleus, circulation, study, extent, microenvironment, percent, assessment, progression, tissues, quality, effect, numbers, transport, ranking, mbh, frequencies, maintenance, arc-me, proportions, circle, nuclei, liver, center, cbn, shape, spleens, degree, hippocampus, modulation, pathophysiology
on the * of	129	basis, surface, day, incidence, risk, number, progression, presence, results, importance, role, effect, use, development, ability, prevalence, efficacy, effectiveness, order, amount, validity, course, detection, rate, expression, proportion, utilization, strength, percentage
of the * of	228	number, effect, percentage, use, university, effects, prevalence, lack, nature, proportion, results, burden, declaration, absence, impact, importance, ability, risk, distribution, role, contribution, efficacy, frequency, intensity, area, regulation, mechanism, initiation, duration, paucity, concentration, quality, development, ministry, frequencies,

		generation, implementation, fraction, expression, features, findings, addition, biology
to the * of	226	development, use, end, number, pathogenesis, risk, level, lack , treatment, intensity, date, expression, effect, initiation, induction, loss, presence, point, comparison, subset, rate, generation, expansion, effects, start, criteria, sera, release, activation, site, left, activity, addition, amount
with the * of	141	use, exception, addition, results, declaration, number, risk, presence, rest, combination, expression, support, findings, lack, effects, standards, provision, goal, approval, occurrence, level, development, implementation, extent, loss
for the * of	161	treatment, development, presence, detection, duration, comparison, prevention, analysis, fidelity, management, lack, purposes, purpose, effects, use, study, majority, effect, assessment, induction, number, production, percentage, design, control, identification, initiation, regulation, remainder, ability, loss, measurement, pathogenesis, association
of * in the	278	patients , participants, children, cells, decline, death, variation, screening, mice, differences, bas, stay, treatment, individuals, malaria, notch, follow-up, platelets, change, livebirths, changes, women
the * of a/an	49	use, absence, results, presence, development, effect, measurement
at the * of	46	time, end, university, start, discretion, level, beginning, site, age, onset, point, peak, midpoint, expense, day, cost, bottom, date
that the * of	127	effects, use, presence, number, effect, addition, ability, percentage, role, risk, prevalence, majority, expression, magnitude, distribution, amount, implementation, abundance, combination, administration, inhibition, level, induction, association, mechanism, absence, failure, size, suppression
was * in the	141	similar, higher, observed, found, lower, detected, lowest, performed, done, increased, included, expressed, reported, reduced, decreased, longer, placed, used, greater, shown, present, conducted
the * in the	131	patients, data, difference, change, participants, differences, reduction, emethods, changes, risk, trial, increase, value, cells, study, hospital, children
was * by the	81	approved, designed, performed, determined, provided, funded, confirmed, defined, written, supported, granted, limited, generated, characterized, recognized, identified, initiated, measured, monitored, sponsored, affected
as a * of	52	result, function, measure, consequence, percentage, marker, proportion, cause, form, source, part, mean, regulator, means
by the * of	95	presence, end , addition, university, number, lack, absence, inclusion, square, fraction, combination, detection, ratio, government, neutralisation
a * of the	75	result, member, measure, comparison, percentage, doubling, part, hallmark, third, component, subset, proportion, marker, reflection, consequence, combination
of the * in	115	patients, variation, participants, study, data, changes, differences, variability, intervention, cerebellum, height, variance
in * of the	118	view, terms, light, favor
the * of patients	53	number, proportion, percentage, subgroup, majority, management, percentages, subset, subgroups, numbers, fibroblasts, serum, group, sera, treatment

in the * study	54	present, current, cohort, initial, progress, same, hospital, hot, Australian
with a * of	65	history, bmi, combination, range, score, diagnosis, prevalence, power, median, change, series, total, mixture
we * that the	8	found, showed, speculate, stress, hypothesize, caution, demonstrate
the * of these	96	basis, results, fraction, ability, findings, use, prevalence, majority, effect, distribution, efficacy, effects, expression, interpretation, magnitude, characteristics, specificity
a * increase in	51	significant, marked, substantial, small, similar, slight, 2-fold, greater, progressive, modest, large, relative, robust
it is * that	24	possible, likely, noteworthy, conceivable, unlikely, known, plausible, estimated, notable, clear, thought
the * of this	85	results, findings, use, purpose, ability, objective, pathogenesis, magnitude, effect, strengths, basis, effectiveness, utility, effects, fidelity, feasibility, aim, context, objectives, role, impact, significance, value, cause, validity, interpretation, strength, time, mechanism, usefulness, benefits, potential
the * number of	40	total, small, median, mean, large, average, absolute, limited, size-weighted, annual, estimated, largest, highest, smallest, cumulative, optimal
these * suggest that	6	data, findings, results, observations, studies, figures
was * with the	52	performed, assessed, compared, associated, measured, determined, consistent, done, calculated, estimated, evaluated, conducted, analyzed, defined, incubated, purified, detected, synthesized
have been * to	44	shown, reported, found, proposed, due, made, linked, able, exposed, demonstrated, related
in a * of	49	subset, cohort, number, series, subgroup, variety, meta-analysis, fraction, total, population, volume, study, model
did not * the	48	affect, reduce, meet, receive, reach, increase, have, alter, change, complete, return, assess, address
has been * to	28	shown, reported, found, proposed, linked, used, suggested, difficult, estimated, hypothesized, postulated, demonstrated
were * from the	44	excluded, obtained, purchased, removed, extracted, isolated, derived, withdrawn, calculated, generated, recruited, separated, omitted, identified, eliminated, drawn, assessed, collected
an increased * of	16	risk, number, rate, frequency, prevalence, incidence, presence, amount,
as the * of	64	number, percentage, ratio, cause, presence, mean, proportion, absence, sum
was * with a	15	associated, performed, measured, assessed, done, identified, extracted
were * with the	18	associated, infected, scanned, screened, obtained, transfected, treated, stained, evaluated, isolated, generated
in * with the	30	accordance, collaboration, line, agreement, combination, compliance, conjunction, contrast, keeping, association, accord
a * reduction in	52	significant, substantial, marked, clear, population-wide, modest, striking, 30
at a * of	25	dose, concentration, median, density, rate, ratio, maximum, volume, dosage
cells were * with	37	washed, treated, transfected, infected, stained, incubated, labeled, stimulated, pulsed, loaded, fed, cocultured, pretreated, cotransfected, rinsed, costained, counterstained, coincubated, cultured
in * to the	2	contrast, addition

in the * treatment	16	aggressive, standard, treatment names (e.g., rituximab) occurring once each
is * in the	55	provided, expressed, involved, available, shown, included, found, presented, detailed, described, degraded, observed, located, increased, reabsorbed, achieved
the * effects of	59	long-term, relative, antiproliferative, inhibitory, negative, immunomodulatory, beneficial, protective, joint, observed, potential, proapoptotic, individual effects, interactive effects, detrimental, cytoprotective
was * as a	12	defined, used, included, measured
a significant * in	14	increase, reduction, decrease, difference, role, drop, improvement
are * in the	55	provided, shown, described, listed, presented, involved, present, reported, expressed, effective, important, located, found, summarised, detailed, available, outlined
the * end point	7	primary, composite, combined, bivariate
these data * that	14	suggest, indicate, demonstrate, suggested, show, demonstrated, indicated
were * for the	44	eligible, used, observed, calculated, responsible, performed, tested, corrected, assessed, recorded
from the * of	68	start, date, time, analysis, end, day
in the * to	325	presence, absence, context, number, pathogenesis, setting, incidence, risk, lungs, developmen, rate, case, regulation, treatment, prevalence, plasma, frequency, serum, percentage, proportion, expression, range, course, use, rates, analysis, management, face, form, feces, blood, induction, subgroup, control, distribution, levels, region, identification, design, event, size, subset, colons, prevention, diagnosis, country, concentration, formation, middle, majority, livers, duration, amount, brains, production
were * by the	52	approved, done, determined, calculated, estimated, judged, produced, generated
cells were * in	39	cultured, resuspended, maintained, plated, seeded, found, lysed
these results * that	12	suggest, indicate, demonstrate, show, suggested, indicated, imply
was * associated with	15	significantly, inversely, strongly, positively associated, consistently
the * for the	63	results, reasons, intervention, potential, basis
to be * to	53	related, due, able, noninferior, similar, unrelated
was * for the	40	responsible, used, observed, highest, required, obtained
has been * in	36	reported, implicated, described, shown, observed, identified, documented
significant * in the	20	increase, difference, reduction, differences, reductions, decrease
a higher * of	23	prevalence, incidence, rate, risk, proportion, intensity, percentage, number, dose, occurrence
is a * of	82	member, hallmark, marker

the * of an	49	use, absence, results, presence, development, effect, measurement
the * treatment group	9	aggressive, standard
to be * in	96	included, involved, important
are * in table	5	provided, shown, presented, summarized, detailed
mice were * with	23	treated, injected, fed, anesthetized, crossed, infected, immunized, gavaged, inoculated
the first * of	34	year, draft, dose, month, day, week
was * from the	39	obtained, calculated, removed, isolated, collected, measured, selected, extracted
did not * a	25	have, show, detect, observe, use
was * as the	21	defined, used, calculated, expressed, taken, determined, selected, estimated, chosen
was * to the	35	added, similar, applied, related, limited, confined, administered, transferred
as * by the	26	measured, determined, defined, recommended, indicated, suggested, shown, assessed, demonstrated
were * using the	37	determined, analyzed, performed, compared, estimated, calculated, prepared, evaluated using, generated
after the * of	33	onset, implementation, initiation, start, addition, end, administration
as * with the	8	compared, assessed, calculated, measured
for a * of	28	total, minimum, mean, median, score, subset, list, series
reduce the * of	28	risk, incidence, prevalence, burden, number, rate, cost, development, frequency, spread
used to * the	51	assess, estimate, identify, evaluate, determine, measure, compute, compare, calculate, detect, quantify, analyze, probe, monitor, track, test, examine
assess the * of	42	effect, effects, role, efficacy, impact, risk, ability, effectiveness
in the * care	6	usual, standard, intensive, primary
the * effect of	47	protective, joint, inhibitory, overall, beneficial, pooled, prognostic, preventive
to * the effect	31	assess, determine, examine, evaluate, estimate , study, quantify, investigate, test the effect
a * number of	25	large, small, limited, substantial, higher, significant, larger, greater, median, lower, number
be * in the	41	included, involved, found, detected, used, interpreted
been * to be	11	shown, reported, found, demonstrated
by * of the	49	use, means
our data * that	11	indicate, suggest , show, demonstrate
is * to be	20	likely, thought, unlikely, known, considered, expected, believed, assumed

the relative * of	30	risk, abundance, importance, effects, frequency, contribution, proportion
were significantly * in	15	higher, elevated, lower, increased, reduced, decreased
a high * of	17	risk, proportion, prevalence, rate, level, degree, incidence, probability
effect of * on	47	rip140, rituximab, treatment, dexamethasone, screening, aspirin, ivabradine
for * in the	49	inclusion, for patients, example, participants, malaria, participation
the * of treatment	27	end, effect, outcome, duration, type, length, start, initiation, discontinuation, course, time
was * using the	22	performed, determined, measured, calculated, assessed, isolated, estimated, extracted, compared
a large * of	19	number, proportion, cohort, effect, set, body, range, volume, fraction
in the * population	32	general, overall population, intention-to-treat population, study, per-protocol, entire
not * in the	38	included, involved, observed, available, differ, used, expressed, significant
was * by a	40	determined, increased, followed, defined, performed
was not * in	34	observed, detected, included, seen, altered, affected, detectable
were * with a	18	associated
of the * gene	39	blakpc, apc, rip140, abcg2, pdc, tnnt2, ntn1 CODE
our * suggest that	9	findings, data suggest, results, studies, study, trial
that * in the	41	occurs, results, reside, occurred
the * phase of	19	effector, treatment, early, induction, acute, late, initial
determine the * of	38	effect, number, role, efficacy, persistence, extent, rate, degree, impact
have been * in	33	reported, implicated, detected, identified, described, found, observed, examined, evaluated, shown
in the * term	4	long, short
role of * in	47	rip140, p40, mkx, sfrp5, trpa1, fgf19, gls1, hscb, tgf
the * from the	38	results, findings, data
the * group were	40	placebo, control, TREATMENT, DISIEASE
these findings * that	9	suggest, indicate, demonstrate, suggested, show, imply
as * in the	38	shown, indicated, described, prespecified, outlined, reported, specified, denoted
for the * group	31	cbt, control, experimental, training, iron, ivabradine group
it is * to	26	important, reasonable, possible, difficult, crucial, preferable, challenging, essential
the * of death	15	risk, rate, cause, date, time, rates, probability, causes
the primary * was	8	outcome, endpoint, analysis, objective, aim

was significantly * in	17	higher, lower, reduced, greater, longer, larger, decreased, elevated
were not * in	34	included, involved, observed, detected
would be * to	24	expected, required, interesting, predicted, necessary
a * rate of	31	higher, high, flow, lower, faster
a * role in	18	critical, key, central, significant, crucial, major
are * to be	14	likely, unlikely, known, believed
of * t cells	55	cd8, cd4, ma-specific, vm-specific, regulatory, lipid-specific, donor, activated, esat-6-specific, epitope-specific
significantly * in the	14	higher, lower, greater, longer, reduced, increased
was * on the	31	based, performed, dependent, selected
was * to that	8	similar, comparable
who were * to	28	assigned, unable, exposed, lost , able
a * risk of	16	lower, high, higher, reduced, low
before the * of	21	start, onset, appearance, end, initiation
can be * in	28	found, detected
did not * any	24	show, reveal, identify
here we * that	11	show, demonstrate, report, showed
is * for the	13	essential, required, critical, important, necessary
may be * to	35	due, related, necessary, linked
no significant * in	11	difference, differences, changes
of a * of	44	combination, number
the * of participants	16	number, proportion, percentage, majority, percentages
the potential * of	46	role, use, utility, effects
the primary * of	24	outcome, endpoint, analysis, prevention, objective, outcomes
was * as described	16	performed, measured, determined
during the * of	19	course, period, evolution, progression
the * group than	37	closed-loop, intervention, placebo, MEDICATION
was * to be	19	found, considered, estimated, shown, assumed t
were * in a	41	performed, randomized
an important * of	22	component, cause

reduced the * of	30	risk, expression, number, incidence
the * rate of	31	overall, lowest, high, annual, median, absolute, mean, higher, 1-year
the * role of	23	potential, functional, major, important, precise, critical
were * according to	37	performed, conducted, graded, stratified, assessed, defined, treated, analyzed, classified
a * dose of	22	single, higher, daily, loading, target, maximum, lethal, high, full, total
a * proportion of	16	high, higher, large, substantial, lower, greater
effects of * on	36	MEDICATION
studies have * that	8	shown, suggested, demonstrated, indicated, found
the * response to	22	immune, cellular, adaptive
to * the role	17	study, assess, investigate, determine, evaluate, examine
we were * to	2	able, unable
were * to be	23	found, considered, judged, predicted, expected
be * to the	24	related, due, attributed
of the * population	26	study, general, human, global, world', background, trial
our * indicate that	3	data, results, findings
study was * by	13	approved, funded, sponsored
the * activity of	28	inhibitory, transcriptional, enzymatic, pharmacological, metabolic, biological, anti-atherogenic, glycolytic
the * was approved	7	study, trial, protocol
these * indicate that	5	results, findings, data
were * to those	5	similar, comparable, equivalent, identical
at the * level	34	global, population, protein, individual, village, molecular, cellular, country
during the * phase	20	effector, induction, acute, randomized
increase the * of	20	risk, number, rate
no significant * between	7	differences, difference, interaction
our results * that	11	show, suggest, indicate
than * in the	11	patients, women, children
to that * the		example for function only, but
was * according to	25	performed, stratified, defined
was also * in	16	observed, identified, effected, higher, reported

a * model of	17	mouse, mathematical
a lower * of	17	risk, likelihood, incidence, intensity, rate, proportion
evaluate the * of	20	effect, effects, ability, effectiveness, role
in the * that	21	clusters, group
is * to the	28	related, central, similar, due, linked
of * disease in	18	CONDITIONS
to the * in	33	increase, reduction, growth, change
to the * that	17	fact, extent, hypothesis, conclusion, notion, suggestion
we also * the	33	assessed, analyzed, examined, compared, calculated, study, tested
were also * in	32	present, included, increased, observed, similar, seen, found