

# How to write medical abstracts? The rhetorical structure and phrases used in Epidemiology

*Como escrever resumos da área médica? A estrutura retórica e frases usadas na Epidemiologia*

Ana Luiza Pires de Freitas<sup>1</sup> , Ana Eliza Pereira Bocorny<sup>2</sup> 

## ABSTRACT

**Introduction:** Abstracts are critical in medical contexts. They contain formulaic building blocks called Lexical Frames (LFs), which are high-frequency word sequences with variable slots that can be formed around collocation nodes. LFs are abundant in written academic discourse, and, for this reason, have great importance for the production of abstracts. Extensive research has been conducted on formulaic language, especially on medical genres. Fewer studies, however, have focused on LFs from specialty-specific corpora (e.g., epidemiology) and their relationship with the rhetorical structure of abstracts. **Objective:** This study aims to fill this gap by describing the structure of epidemiology abstracts, presenting their rhetorical functions, and identifying the LFs that linguistically realize these functions to help researchers write more conventional abstracts. **Methods:** We put together three corpora of abstracts in the field, published in English in peer-reviewed journals, and combined genre analysis and Corpus Linguistics principles to identify the linguistic realizations of the rhetorical functions in the texts. First, the rhetorical structure was described; then, the LFs were identified and analyzed. **Results:** 92% of the texts follow a pre-established pattern, whose structure consists of five to nine sections. Eight saliently frequent nodes (study, result, method, conclusion, review, analysis, patients, and findings) around which the LFs are constructed were identified. **Conclusion:** Even though both the content and function words that make up the LFs show some variation, it is possible to notice that the LFs elicited typify the linguistic realizations of the corresponding sections' rhetorical functions and, thus, are suitable to the observation of a pattern. For that reason, the data obtained in this study were used to inform the creation of a support framework for the writing of specialty-specific medical abstracts.

**Keywords:** epidemiology; abstracts; impact factor.

## RESUMO

**Introdução:** Resumos são cruciais em contextos médicos. Tais textos contêm blocos formulaicos do discurso acadêmico, denominados Estruturas Lexicais (ELes), sequências de palavras de alta frequência compostas de termos variáveis, organizados a partir de núdulos colocacionais. Estruturas Lexicais são abundantes no discurso acadêmico escrito, tendo grande importância para a produção de resumos. Extensas pesquisas já foram conduzidas acerca de linguagem formulaica, especialmente em gêneros médicos. Poucos estudos se concentraram em Estruturas Lexicais em *corpora* de especialidades (como epidemiologia) e sua relação com a estrutura dos resumos. **Objetivo:** Este estudo busca suprir a lacuna apontada descrevendo a estrutura dos resumos de epidemiologia, apresentando suas funções retóricas e identificando as Estruturas Lexicais que realizam essas funções, para promover escrita de resumos mais convencionais. **Métodos:** Compilamos três *corpora* de resumos no campo publicados em inglês em revistas indexadas e revisadas por pares e combinamos a análise de gênero e os princípios da Linguística de *Corpus* para identificar as realizações das funções retóricas dos textos. Primeiramente, a estrutura retórica foi descrita. Em seguida, as foram identificadas, extraídas e analisadas. **Resultados:** Do total, 92% dos textos seguem um padrão, composto de cinco a nove seções. Oito núdulos colocacionais frequentes foram identificados (*study, result, method, conclusion, review, analysis, patients e findings*), em torno dos quais as Estruturas Lexicais se constroem. **Conclusão:** Embora tanto as palavras que remetem a conteúdo quanto as que se referem a funções das Estruturas Lexicais apresentem alguma variação, é possível perceber que as Estruturas Lexicais tipificam as realizações linguísticas das funções retóricas das seções correspondentes (*background & objectives, methods, results e conclusion*) e são adequadas para a observação de um padrão. Os dados obtidos foram usados para informar a criação de um quadro para a redação de resumos da especialidade.

**Palavras-chave:** epidemiologia; resumos; fator de impacto.

## INTRODUCTION

Writing plays a significant role in academic contexts as one of the central skills for a successful academic career<sup>(1)</sup>. Abstracts, in turn, are essential in medical contexts. Professionals and medical faculty rely on them to decide whether to read the full text of an article, to assess trials that they may use in clinical practice, to submit articles for publication, and to share their research at conferences. Frequently, papers are not read in their entirety, either due to lack of time or because they are not related to a specific topic being searched for. Even journal editors may not go beyond abstracts when the manuscripts are poorly written. So abstracts also fulfill the purpose of being time-savers for readers, for they allow democratic entry to

scientific advances, even when several journals are not access-free to all<sup>(2)</sup>. Besides, they can help summarize critical findings on research and enable quick retrieval. Therefore, professionals and medical faculty must write fluid abstracts and use appropriate language as informative and conventional as possible and which is recognized as authentic by the medical discourse community.

It is worth mentioning that, for the perspective adopted along with this study — that of investigations in Corpus Linguistics (CL) — fluidity and conventionality go beyond the description of possible language combinations considered acceptable by the criteria of grammatical rules. Thus, even though syntactic language guidelines go without saying, the mindset brought into account is the one of use instead. In other words, the focus is on which word associations are more recurrently used, or which combinations or patterns are more conventional in the sense of being probable and thus recognized as fluent and, as such, acceptable in the eyes of a specialized community.

Incidentally, the Brazilian Journal of Sexually Transmitted Diseases (hereinafter JBDST) has shown quite an evolution in

<sup>1</sup>Department of Education and Humanities, Universidade Federal de Ciências da Saúde de Porto Alegre – Porto Alegre (RS), Brazil.

<sup>2</sup>Department of Modern Languages, Universidade Federal do Rio Grande do Sul – Porto Alegre (RS), Brazil.

conceptual development since 1989, when it was founded. In the beginning, some articles did not feature any abstract, comprised the text in Portuguese, or included abstracts in a non-structured format. However, from the journal's second decade to the moment the JBDST became online (2013), abstracts started to be part of all articles; they came out in English and incorporated the structured format to become ever more patterned. Even though these changes also reflect the abstract genre historical advances<sup>(3)</sup>, taking the national scientific writings to a higher pattern is a desired goal. Moreover, as it stands, the BJSTD is the only open access and free of charge online periodical in its field. Thus it is only relevant to favor its textual improvement.

Furthermore, as stated above, texts that do not use conventional word combinations tend to sound less proficient and more difficult to process. Thus, the high level of conventionality ruling over academic discourse enhances text idiomaticity, as scientific language is oriented by the presence of fixed and/or relatively fixed textual blocks, which still ought to be discipline-specifically studied. Anyway, linguists, increasingly more intrigued by the patterns orienting academic language idiomaticity<sup>(4)</sup>, found out that phraseological word patterns follow either continuous or discontinuous word sequences, which constitute major formulaic building blocks of written academic discourse. The continuous type of pattern is defined as lexical bundles (LBs)<sup>(5)</sup>, as in *'in order to determine'*, *'the aim of this'* or *'further studies need to.'* The discontinuous type, on the other hand, is known as formulaic frames or Lexical Frames (LFs)<sup>(6,7)</sup>, being identified for their variable slots, as in *'the results/findings of this study suggest/ indicate that'*, or *'it is relevant/important/needed to.'* Incidentally, these combinations are formed around collocation nodes, which are words from which sequences are structured, as in *'the goal of the study'*, *'the study was to examine'* or *'were included in the study'* to mention a few collocations springing around the node *'study'*<sup>(8,9)</sup>.

As LBs and LFs make up the fabric of written academic language, they have great value for producing discipline-specific academic genres, conventionally identified templates through which discourse is constructed to fulfill the communicative purposes of social interactions<sup>(10)</sup>. Genre identification aids readers in retrieving messages conveyed in given contexts more promptly<sup>(11)</sup>. Likewise, the recognition of medical abstracts is enabled by the use of conventions in the form of recurrent rhetorical choices, which mainly involve decisions regarding the overall organization of discourse and the linguistic resources employed to reflect their communicative purposes<sup>(12)</sup>. Extensive research has been conducted on using formulaic language in written academic genres<sup>(7,13,14)</sup>, especially in medical science research article abstracts<sup>(1,2,15)</sup>. Nevertheless, fewer studies have focused on the use of LFs in abstracts of medical specialty articles and their relationship with the rhetorical structure of that academic genre. To the best of our knowledge, the study at

hand is the first investigation exploring such patterns in the medical subfield of epidemiology. Besides, it is important to mention that linguistic support for scientists seldom comes directly from applied linguists, but mainly from formal teaching settings. In that sense, shortening the way between studies in the field and the medical community might help promote writer autonomy and raise the rhetorical level of the national scientific manuscripts.

## OBJECTIVE

Based on the background mentioned above, this study aims to describe the structure of epidemiology abstracts, present their rhetorical functions, and identify the LFs that linguistically realize these functions. With that, we aim to bridge the gap between corpus research and end-users by helping professionals and medical faculty write more informative and conventional abstracts that have the rhetorical structure and language patterns used by the international medical discourse community.

## METHODS

### The corpora in the study

Data for the study were obtained through an empirical analysis of three specialized corpora, using CL principles: the Scientific Journals Corpus (SJC), which comprises three periodicals suggested by specialists in the field, namely Sexually Transmitted Infections (STI), Sexually Transmitted Diseases (STD), and the International Journal of STD & AIDS; the PLOS ONE Corpus (PLOS ONEC), originated from PLOS ONE, "an international open access online journal published by the Public Library of Science since 2006, covering all science and medicine categories"<sup>(16)</sup> and the Brazilian Journal of Sexually Transmitted Diseases Corpus (BJSTDC).

All three corpora used contain abstracts from epidemiology review and research articles written in English and published between 2003 and 2021 in peer-reviewed indexed journals. As shown in **Table 1**, SJC has 662,747 words and 1,915 texts, PLOS ONEC features 1 million words and 4,330 texts, and BJSTDC displays 83,261 words and 360 texts.

The abstracts used in the SJC corpus were manually selected from its three source periodicals. The abstracts from the PLOS ONE corpus, to that end, were crawled from the PLOS ONE platform with AntCorGen<sup>(17)</sup>, a free multiplatform corpus generating tool designed to create discipline-specific research article corpora by directly communicating with the database behind the PLOS ONE scientific journal<sup>(18)</sup>. Accordingly, the abstracts used in the BJSTD corpus were likewise manually selected from the Brazilian Journal of Sexually Transmitted Diseases (BJSTD).

**Table 1** – Numbers of corpora used in the study.

Domain	Corpus	Words with repetition (tokens)	Words without repetition (types)	Texts	Average of words per abstract
Epidemiology	SJC	662,747	21,087	1,915	346
Epidemiology	PLOS ONE	1,000,003	43,066	4,330	230
Epidemiology	BJSTD	83,261	9,010	360	231

## The methodological procedures

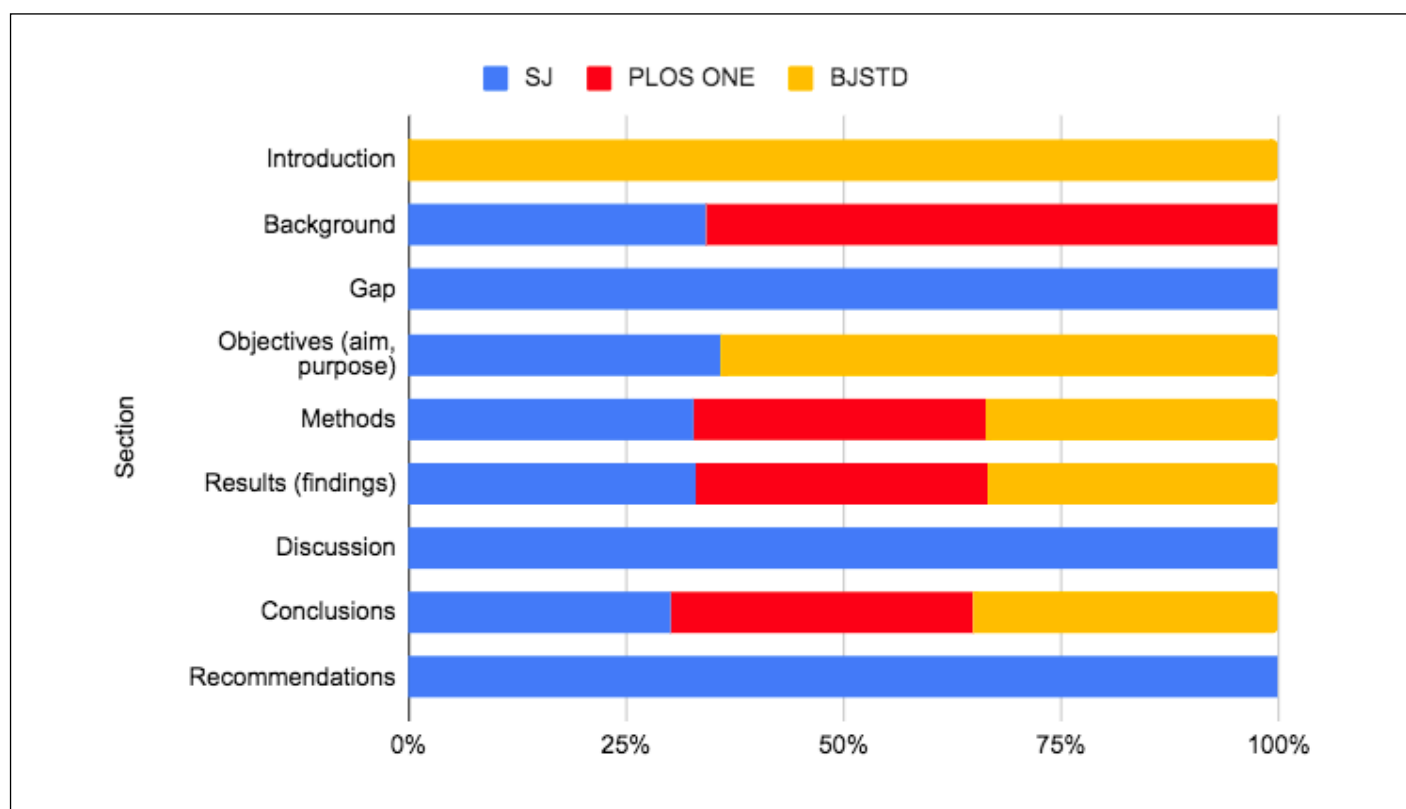
To identify the linguistic realizations of the rhetorical functions in the abstracts, we paired up genre analysis with CL principles. First, the rhetorical structure was tracked, meaning that the sections of which the texts are made up were worked out. To this aim, thirty abstracts from each corpus (SJC, PLOS ONE, and BJDSTC) were examined and described. Then, the LBs from each corpus were extracted, analyzed, and categorized. Three criteria were considered for their extraction: the extension of the word sequences (n-gram length), their frequency in the corpus per million words (minimum normalized frequency), and the number of texts in which the sequences appear along with the corpora (dispersion). According to the literature, a phraseological sequence has to occur in three to five texts<sup>(19)</sup> or in 10% of the texts to ensure they are not restricted to idiosyncrasies of an author's writing style<sup>(20)</sup>. For the three corpora (SJC, PLOS ONEC, and BJDSTC), the extraction criteria were n-gram length: 5 and minimum normalized frequency: 18 pmw. The text dispersion for the two more extensive corpora (SJC and PLOS ONEC) was 10, whereas a text dispersion of 3 was used for the BJDSTC, as it contains a much smaller number of texts. Finally, the analysis and categorization of the LBs and the construction of the LFs were optimized by the use of the most frequent noun collocation nodes<sup>(6,8)</sup>, which were adopted "as a starting point for collocation look-ups"<sup>(9)</sup>. The nodes were also used to locate the sections within the structured abstracts, facilitating the identification of their recurrent LFs. In other words, the nodes were used in two ways: in the identification of the LFs and in the detection of sections.

## RESULTS AND DISCUSSION

The rhetorical structure observation revealed that 92% of the texts analyzed use a pre-established pattern, usually suggested by the journals. It is due to such structure that the forthcoming analysis could be conducted. Incidentally, their design consists of five to nine sections whose functions are to introduce the investigation topic and present its background by making generalizations and showing its centrality, followed by the mention of a gap to be filled and a description of the purpose, methodology and results to eventually discuss the outcomes, draw conclusions by prompting future research and making recommendations, as shown in **Graph 1**. As can also be noticed, the background, objectives, methods, results and conclusions are component sections in the abstracts of most journals under study. Obiter the background is often used as an umbrella section, in which generalizations about the topic, gaps to be filled by the investigation, and sometimes its aim can be found.

The LBs extracted from the three corpora while constructing the LFs pointed out eight saliently frequent nodes: *study, result, method, conclusion, review, analysis, patients, and findings*. They were used in this process to help identify the LBs and LFs in their corresponding sections. The LFs presented in **Tables 2 to 5** comprise building blocks often used in the sections background, objectives, methods, results and conclusions, which linguistically realize the rhetorical functions expressed in the abstracts. For correlation purposes, they are presented in the analogous colors of the corpus they come from, as identified in **Graph 1**.

From the sections Background and Objectives (**Table 2**), a frequently used LF to present the purpose of the investigations is centered



**Graph 1** – Sections present in the epidemiology abstracts analyzed in this study.

around the node *study*: *The (aim, objective, purpose) of (this, the present) study was to (determine, examine, evaluate)*. This LF appeared in the three corpora, with slight lexical variation, revealing little discrepancy across journals. Such a feature could mean that this formulaic building block is already consolidated in the discipline specialty; a characteristic, which is informed by the high frequencies per million words (276, 227, 193 pmw) in the three corpora analyzed, compared to the lower frequency of the LFs in the other sections.

In **Table 3**, the data obtained from Methods show two frequently used LFs this time: *We (conducted, performed, identified) a (cross-sectional, systematic, population-based) analysis, study, review* and *A (cross-sectional, population-based cohort, prospective) study was (carried out, conducted, performed)*. As much as these LFs are used to inform how the studies were carried out, they rely upon different syntactic structures, namely the active and passive voice modes. As such, there is willingness to bring along the researchers' doings ('*We*' *conducted, did, studied*) or the understanding that it goes without saying, meaning that it is the methodological design that deserves centrality instead (*A cross-sectional study was carried out/conducted/performed*). As it stands, several theoretical descriptions of contemporary academic English lie upon a burgeoning encouragement to use the active voice<sup>(21)</sup>, as the style supposedly favored by journal editors. Nevertheless, with a metric ruled by use, the findings from our corpora reveal the adoption of the two structures as a stylistic preference, which is acceptable for publishing goal; even though frequency prevalence highlights a current tendency towards the use of the active voice (102, 52, 28 against 40, 35, 57).

In **Table 4**, the Results section is the one exhibiting more internal variation, meaning that more LFs were used to linguistically

realize the rhetorical function of presenting the outcomes of the studies. It is relevant to note that the frame pattern used in this section is different from the ones observed in the previous ones. While the LFs from the other sections were a combination of content and function words — as in *The (aim, objective, purpose) of (this, the present) study was to (determine, examine, evaluate)*, or in *We (conducted, performed, identified) a (cross-sectional, systematic, population-based) (analysis, study, review)* —, the three patterns recurrently used in the Results section start with function words (*of the, among, a total of*), followed by slots filled with numeric expressions, then a slot with nouns (as in *participants, patients, studies*). In some cases, another slot for numeric expressions is used, followed by a frequent verb, such as *were* (for example, "*Results: Of the 1,072 interviewees, 64.9% were sexually active*"); typifying the use of numeric elements (such as percentages and fractions) as a pattern for informing results in quantitative investigations in the field.

In **Table 5**, finally, in the Conclusions section, where findings of the study have to be discussed, and authors need to draw conclusions and make recommendations, a recurrent LF was spotted springing from the nodes *results* and *study*: (*Our, This/These, The) results, study, findings (suggest, show, indicate) that/the*; once again revealing a combination of function and content words. It is important to highlight that the adoption of hedging, meaning the use of cautious language to conform to a current style for academic writing, is recurrent in the three corpora, as shown by the use of *suggest* and *indicate*. Being academic studies suitable for interpretation from various perspectives, this openness in style may represent a rhetorical strategy that leaves room for improvement and development<sup>(22)</sup>.

**Table 2** – LFs extracted from the sections BACKGROUND and OBJECTIVES.

Rhetorical Function: Present the purpose of the study Node: Study			
SJ LFs 276 (pmw)	PLOS ONE LFs 227 (pmw)	BJSTD LFs 193 (pmw)	Example from SJ corpus
<i>The (aim, objective, purpose) of (this, the present) study was to (determine, evaluate, investigate, assess, examine)</i>	<i>The (aim, objective, purpose) of this study was to (examine, investigate, determine, evaluate, assess)</i>	<i>The (aim, objective) of this study was to (evaluate, describe)</i>	"The aim of this study was to examine the cross-sectional associations of factors linked to STIs among US young adults (18 - 25 years)."
The + noun + of + (this, the present) + study was to + vb			

**Table 3** – LFs extracted from the section METHODS.

Rhetorical Function: Present the methodology of the study Node: Method(s): We + study, review, analysis			
SJ LFs 102 (pmw)	PLOS ONE LFs 52 (pmw)	BJSTD LFs 28 (pmw)	Example from PLOS ONE corpus
Methods: <i>We (conducted, did, developed) a (cross-sectional, systematic, population-based) analysis, study, review</i>	Methods: <i>We (conducted, performed, identified,) a (cross-sectional, systematic, population-based) study, review, analysis</i>	Methods: <i>We (developed, conducted) a (systematic, population-based) analysis, study, review</i>	"Methods: We conducted a population-based prospective cohort study among children in 14 Taiwanese communities."
Methods: We + vb + a + modifier + noun			
Node: Method(s): A + study			
SJ LFs 40 (pmw)	PLOS ONE LFs 35 (pmw)	BJSTD LFs 57 (pmw)	Example from the BJSTD corpus
Methods: <i>A cross-sectional (study, survey) was (carried out, conducted, performed)</i>	Methods: <i>A (cross-sectional, population-based cohort, retrospective cohort) study was (conducted, carried out, performed)</i>	Methods: <i>A (cross-sectional, retrospective cohort, prospective observational) study was (conducted, carried out)</i>	"Methods: A cross-sectional study was conducted with 87 HIV seropositive patients confirmed by tests carried out previously."
Methods: A + modifier + study was + vb			

## Strengths

The main strength of our study is that it presents findings that are specific to abstracts produced within the area of epidemiology on the relation between formulaic building blocks of the academic discourse and the rhetorical structure of the target genre. These results have the potential to be used for developing pedagogical applications (e.g., an online writing support tool, learning objects) that capture linguistic variation (if any) and, therefore, guide the production of abstracts in this medical specialty. Such a tool can bridge the gap between corpus research and end-users, thus, representing an effort to encourage end-user agency by promoting access to applied language data and assisting medical researchers to increase their chances of publication in relevant journals.

## Limitations

It should be noted that this investigation was limited to a corpus compiled from abstracts of a medical specialty within a discipline. Therefore, validation of the findings or replication of the approach is suggested to other sections of research articles (e.g., introduction, methodology, results, and conclusion), to different genres (e.g.,

research proposals, lab reports), as well as to other disciplines (e.g., biology and life sciences, computer and information sciences) and specialties (e.g., cell biology, computer security).

## CONCLUSION

This study aimed to present findings on the relation between Lexical Frames (LFs) and the rhetorical structure of epidemiology review and research abstracts published in journals from the area. In particular, we wanted to present the LFs that are more frequently used to linguistically realize the rhetorical functions expressed in the abstracts of the journals under study. The findings demonstrate that the LFs chosen by the authors were used consistently across the sections of the journals, meaning that there is slight lexical variation across the three corpora analyzed, and also implying the existence of a consolidated collocational pattern based on the conventionalities of the field. In other words, the little variation across journals when it comes to the linguistic realizations of rhetorical functions means that the LFs identified are already very well established in the discipline specialty under study. Consequently and as previously mentioned, once the LFs make up the fabric of written discourse, the

Table 4 – LFs extracted from the sections RESULTS, FINDINGS.

Rhetorical Function: Present the outcomes of the study			
Node: Result(s): Of the + patients			
SJ LFs 85 (pmw)	PLOS ONE LFs 45 (pmw)	BJSTD LFs 96 (pmw)	Example from the BJSTD corpus
Results: Of the [number] (patients, participants), [number, percentage] (were, reported, had)	Results: Of the [number] (patients, participants, women), [number, percentage] (were, had)	Results: Of the [number] (patients, interviewees, women), [number, percentage] (were, had, reported)	"Results: Of the 1,072 interviewees, 64.9% were sexually active, of which 46.4% reported sexual initiation at the age of 15 or younger, and 73.2% reported inconsistent condom use."
Results: Of the + [number] + (noun), + [number, percentage] + vb			
Node: Result(s): Among + patients			
SJ LFs 135 (pmw)	PLOS ONE LFs 64 (pmw)	BJSTD LFs 88 (pmw)	Example from the SJ corpus
Results: Among [number] (women, participants, patients), [number, percentage] (were, had, tested)	Results: Among [number] (women, participants, patients), [number, percentage]	Results: Among (-, the) (women, patients), [number, percentage] were	"Results Among 29,719 syphilis patients, 23,613 (80%) were interviewed and 18,581 (63%) reported 84,224 sex partners (average, 4.5; 20,853 [25%] named and 63,371 [75%] unnamed)."
Results: Among + [number] + (noun), + [number, percentage] + were + vb			
Node: Result(s): A total of + patients			
SJ LFs 239 (pmw)	PLOS ONE LFs 185 (pmw)	BJSTD LFs 231 (pmw)	Example from the PLOS ONE corpus
Results: A total of [number] [-, modifiers] (men, patients, participants) were	Results: A total of [number] [-, modifiers] (patients, cases, studies) were	Results: A total of [number] [-, modifiers] (interviews, patients) were	"Results: A total of 4,949 SCI patients were analyzed."
Results: A total of + [number] + [-, modifiers] (noun) + were + vb			

Table 5 – LFs extracted from the section CONCLUSION(S).

Rhetorical Function: Discuss the results, draw conclusions and make recommendations			
Node: Conclusions: + results, study			
SJ LFs 96 (pmw)	PLOS ONE LFs 182 (pmw)	BJSTD LFs 106 (pmw)	Example from the SJ corpus
Conclusion(s): (Our, This/These, The) results, study, findings (suggest, show, indicate) (that, the)	Conclusion(s): (Our, This/These, The) study, results, findings (suggest, indicate, show) (that, the)	Conclusion(s): The results, findings (suggest, indicate) (that, the)	"Conclusions: Our results suggest that potential use and management of TasP may not be straightforward."
Conclusion: (Our, This/These, The) + noun + vb + (that, the)			

high frequency of the ones elicited relates them to conventionally identified templates, which may facilitate the overall organization of communication in the field.

### Participation of each author

Both authors participated equally in the writing idea, source research, writing and review of the text.

### Funding

There was no funding source.

### Conflict of interests

There are no conflicts of interests to declare.

## REFERENCES

1. Flowerdew J. English for specific academic purposes (ESAP) writing: Making the case. *Writ Pedagogy*. 2016;8(1):5-32. <https://doi.org/10.1558/wap.v8i1.30051>
2. Hanna M. How to write better medical papers. Cham: Springer; 2019.
3. Freitas ALP. Proficiência escrita em inglês especializado: estudo de corpus [thesis on the Internet]. Porto Alegre: Universidade Federal do Rio Grande do Sul; 2016 [cited on Apr 28, 2021]. Available from: <https://lume.ufrgs.br/handle/10183/148955>
4. Hong AL, Rahim HA, Hua TK, Salehuddin K. Collocations in Malaysian English learners' writing: A corpus-based error analysis. *Southeast Asian J English Language Stud*. 2019;17(9):31-44.
5. Biber D, Conrad S. Lexical bundles in conversation and academic prose. In: Hasselgard H, Oksefjell S., eds. *Out of Corpora: Studies in Honour of Stig Johansson*. Amsterdam/Atlanta, GA: Rodopi; 1999. p. 181-190.
6. Frankenberg-Garcia A, Rees G, Lew, R. Slipping through the cracks in e-lexicography. *Int J Lexicogr*. 2021;34(2):206-34. <https://doi.org/10.1093/ijl/eca022>
7. Hyland K. Disciplinary differences: language variation in academic discourses. In: Hyland K, Bondi M (eds.). *Academic discourse across disciplines*. Frankfurt: Peter Lan; 2006. p. 17-45.
8. Wray A. Formulaic language: Pushing the boundaries. Oxford: Oxford University Press; 2008.
9. Gray B, Biber D. Phraseology. In: Biber D, Reppen R, eds. *The Cambridge Handbook of English Corpus Linguistics*. Cambridge: Cambridge University Press; 2008. p. 123-145.
10. Swales JM. *Research genres: Exploration and applications*. Cambridge: Cambridge University Press; 2004.
11. Melissourgou MN, Frantzi KT. Genre identification based on SFL principles: the representation of text types and genres in English language teaching material. *Corpus Pragmatics*. 2017;1:373-92. <https://doi.org/10.1007/s41701-017-0013-z>
12. Bhatia VK. *Analysing genre: language use in professional settings*. London: Longman; 1993.
13. Asano M. Construction of medical research article corpora with AntCorGen: pedagogical implications. *English Corpus Studies*. 2018;(25):101-15.
14. Conrad S, Biber D. The frequency and use of lexical bundles in conversation and academic prose. In: Teuber W, Mahlberg M., eds. *The corpus approach to lexicography, Thematischer Teil von Lexicographica. Internationales Jahrbuch für Lexicographie*. Portland: De Gruyter; 2005. p. 56-71.
15. Coates R, Sturgeon B, Bohannon J, Pasini E. Language and publication in cardiovascular research articles. *Cardiovasc Res*. 2002;53(2):279-85. [https://doi.org/10.1016/s0008-6363\(01\)00530-2](https://doi.org/10.1016/s0008-6363(01)00530-2)
16. Fein C. Multidimensional journal evaluation of PLOS One. *Libri*. 2013;63(4):259-71. <https://doi.org/10.1515/libri-2013-0021>
17. Anthony L. AntCorGen (Version 1.1.0) [Computer Software] [Internet]. Tokyo, Japan: Waseda University, 2017 [cited on Apr 28, 2021]. Available from: <http://www.laurenceanthony.net/>
18. Anthony L. Resources for researching vocabulary. In: Webb S., ed. *The Routledge handbook of vocabulary studies*. New York: Routledge; 2019. p. 561-90.
19. Biber D, Barbieri F. Lexical bundles in university spoken and written registers. *English Specif Purp*. 2007;26(3):263-86. <https://doi.org/10.1016/j.esp.2006.08.003>
20. Hyland K. As can be seen: lexical bundles and disciplinary variation. *English Specif Purp*. 2008;27(1):4-21. <https://doi.org/10.1016/j.esp.2007.06.001>
21. Millar N, Budgell B, Fuller K. "Use the active voice whenever possible": the impact of style guidelines in medical journals. *Appl Linguist*. 2013;34(4):393-414. <https://doi.org/10.1093/applin/ams059>
22. Strunk FJ. *The elements of style*. Vancouver: Alysian; 2017.

### Address for correspondence:

**ANA LUIZA PIRES DE FREITAS**

Rua Amapá, 880 – Vila Nova

Porto Alegre (RS), Brazil

CEP: 91740-460

E-mail: [analuizaf@ufcspa.edu.br](mailto:analuizaf@ufcspa.edu.br)

Received on: 09.07.2021

Approved on: 09.13.2021

