UNIVERSITY OF SÃO PAULO
FACULTY OF PHILOSOPHY, LANGUAGES AND HUMAN SCIENCES

Rodrigo Moura Lima de Aragão

Patterns of lexis in the research article genre

A contrastive study of lexical priming in English, Portuguese and Japanese

Revised Version

São Paulo
2023

UNIVERSITY OF SÃO PAULO
FACULTY OF PHILOSOPHY, LANGUAGES AND HUMAN SCIENCES

Rodrigo Moura Lima de Aragão

Patterns of lexis in the research article genre

A contrastive study of lexical priming in English, Portuguese and Japanese

Revised Version

São Paulo
2023

UNIVERSITY OF SÃO PAULO
FACULTY OF PHILOSOPHY, LANGUAGES AND HUMAN SCIENCES

Rodrigo Moura Lima de Aragão

Patterns of lexis in the research article genre

A contrastive study of lexical priming in English, Portuguese and Japanese

Revised Version

Dissertation submitted to the Postgraduate Program in Linguistic and Literary Studies in English of the Department of Modern Languages of the Faculty of Philosophy, Languages and Human Sciences of the University of São Paulo for earning the Doctor of Letters degree.

Field of Concentration: Linguistic and Literary Studies in English

Advisor: Dr. John Blair Corbett

São Paulo
2023

Catalogação da Publicação
Serviço de Biblioteca e Documentação
Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo

Aragão, Rodrigo Moura Lima de
A659p Patterns of lexis in the research article genre: A contrastive study of lexical priming in English, Portuguese and Japanese / Rodrigo Moura Lima de Aragão; orientador John Blair Corbett - São Paulo, 2023.
366 f.

Tese (Doutorado) - Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo. Departamento de Letras Modernas. Área de concentração: Estudos Linguísticos e Literários em Inglês.

1. Linguística de Corpus. 2. Língua inglesa. 3. Língua portuguesa. 4. Língua japonesa. I. Corbett, John Blair, orient. II. Título.

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE FILOSOFIA, LETRAS E CIÊNCIAS HUMANAS

ENTREGA DO EXEMPLAR CORRIGIDO DA DISSERTAÇÃO/TESE

Termo de Anuência do (a) Orientador (a)

Nome do (a) Aluno (a): Rodrigo Moura Lima de Aragão

Data da Defesa: 02/06/2023

Nome do Prof. (a) Orientador (a): Dr. John Blair Corbett

Nos termos da legislação vigente, declaro **ESTAR CIENTE** do conteúdo deste **EXEMPLAR CORRIGIDO** elaborado em atenção às sugestões dos membros da Comissão Julgadora na sessão de defesa do trabalho, manifestando-me **plenamente favorável** ao seu encaminhamento ao Sistema Janus e publicação no **Portal Digital de Teses da USP**.

São Paulo, 06/06/2023

John B Corbett

Assinatura do (a) Orientador (a)

APPROVAL SHEET

FOLHA DE APROVAÇÃO

Name | Nome: Rodrigo Moura Lima de Aragão

Title | Título: Patterns of lexis in the research article genre: A contrastive study of lexical priming in English, Portuguese and Japanese | Padrões do léxico no gênero artigo acadêmico: Um estudo comparativo de primazia lexical em inglês, português e japonês

Dissertation submitted to the Postgraduate Program in Linguistic and Literary Studies in English of the Department of Modern Languages of the Faculty of Philosophy, Languages and Human Sciences of the University of São Paulo for earning the Doctor of Letters degree (field of concentration: Linguistic and Literary Studies in English).

Tese submetida ao Programa de Pós-Graduação em Estudos Linguísticos e Literários em Inglês do Departamento de Letras Modernas da Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo para a obtenção do título de Doutor em Letras (campo de concentração: Estudos Linguísticos e Literários em Inglês).

Approved on | Aprovado em ______________________, 2023.

Examination Committee | Banca Examinadora

Doctor | Doutor(a): ______________ Institution | Instituição: ______________

Assessment | Julgamento: ______________ Signature | Assinatura: ______________

Doctor| Doutor(a): ______________ Institution | Instituição: ______________

Assessment | Julgamento: ______________ Signature | Assinatura: ______________

Doctor | Doutor(a): ______________ Institution | Instituição: ______________

Assessment | Julgamento: ______________ Signature | Assinatura: ______________

Committee Chairman | Presidente da Banca: Dr. John Blair Corbett

Signature | Assinatura: ______________

To the tireless scientists
who have devoted strenuous efforts to find answers and develop solutions
to the COVID-19 pandemic.

To the heroic health professionals
who have selflessly dedicated themselves to providing care for patients
during the pandemic.

To the careful journalists
who have worked diligently to communicate accurate, reliable information
on the pandemic.

To the brave service workers
who have continuously faced danger to support life in the urban areas
during the pandemic.

ACKNOWLEDGMENTS

More than 20 years ago, I heard the indigenous (African?) adage that an entire village is needed to raise a child. Later, during my teacher-in-training period at São Paulo public schools, I remembered it, suspecting that disorderly students who were making classes a traumatic experience for everyone—including me—may not have had enough attention from their parents, let alone from a village! During the writing of this dissertation, the adage came into my mind again as I started thinking about how many persons are actually involved in the graduation of a single Doctor of Philosophy (PhD) student. If raising a child demands an entire village, then raising a PhD student probably demands an entire faculty. Almost certainly, there are many more persons I should thank here than my limited mind is capable of recognizing. Therefore, I apologize in advance to anyone whom I may have accidentally left out.

First, I am profoundly indebted to my advisor, Professor John B. Corbett. In the Humanities and Social Sciences, the PhD dissertation is often acknowledged as one of the most (otherwise the most) decisive works in a researcher's lifetime, a major watershed. Therefore, I always had the feeling that I would have to undergo an arduous and complex process of self-improvement and intellectual growth to produce an acceptable result. Fortunately, this process was greatly facilitated due to Professor Corbett's sensible and good-natured guidance. His enthusiastic support and penetrating insights have guided me to the point where I feel that I could bring out the best of myself.

I am also deeply indebted to Professors Marilisa Shimazumi and Luciana C. Fonseca for their valuable comments on an earlier version of this text and for their encouragement during my preliminary examination. Their observations on theoretical tenets, data collection, and analytical procedures led me to make substantial changes in my PhD research, improving it dramatically.

I am grateful to Professor Marília M. Ferreira and all the Laboratory of Academic Literacy team for sharing their knowledge and expertise on academic writing and literacy. This opened up new horizons for me.

I would like to thank the administrative staff of the Department of Modern Languages and specifically of the Postgraduate Program in Linguistic and Literary Studies in English for their invaluable assistance during the 5-year period of the doctoral program. I would also like to profusely thank the technical and administrative staff who have made possible the electronic submission of theses and dissertations at the University of São Paulo. Before that, I had some restless nights thinking about printing—and carrying—several copies of my work.

I wish to acknowledge the privilege of undertaking a PhD program in a public university in Brazil. I am immensely grateful to the population of the state of São Paulo for providing me with this opportunity through the payment of their taxes. I have made a considerable effort to complete the program as soon as possible so that another person may have the same privilege as me.

Some of my colleagues at the Itaquera College of Technology (Fatec Itaquera), in São Paulo, deserve thanks for casual but meaningful conversations on the challenges of juggling postgraduate activities with work. Anderson, Arruda, Augusto, Daniel, David, Ervaldo, Etevaldo, Fábio, Fabricio, Francisco, Jean, Luis Carlos, Maiaty, Paulo Hélio, Paulo Henrique, Rafael, Regina, Rogério: thank you very much! My students also deserve thanks for the constant cheering and for the thought-provoking questions they bring up in classes, which make me try to always put into practice—and not only advocate—life-long learning. In particular, I would like to thank my first undergraduate thesis advisees, Abner, Daniel, and Leo, for the encouraging words when I was dealing with an unexpected creative block.

I must thank my wife, Tomomi, for revising the Japanese version of the abstract. Finally, I wish to express my gratitude to my entire family for their love and support.

All the remaining mistakes are my sole responsibility.

ABSTRACT

How to cite this dissertation (American Psychological Association style): Aragão, R. M. L. de. (2023). *Patterns of lexis in the research article genre: A contrastive study of lexical priming in English, Portuguese and Japanese* (PhD dissertation, University of São Paulo).

Lexical priming is a lexicon-oriented theory formulated by the British linguist Michael Hoey that assumes that previous experiences with language prepare or prime language users to communicate in one or another way. Everyone's mind is seen as a natural, complex concordance program that continuously associates words, sounds, and even syllables with a wide variety of linguistic resources and contextual factors, from individual words to genres, domains, and situations. According to the theory, psychological priming—an associative process—would be the main force behind a given language's typical form. Different languages are expected to exhibit different patterns of lexis as a result of their users' particular associations and experiences, meaning that the mapping of patterns from one language to another would not result in standard language use in the target language. This dissertation attempts to contribute to the body of research on linguistic evidence for psychological priming by means of an investigation of research articles (RAs) in English, Portuguese, and Japanese. Using a collection of 240 RAs from 10 journals in Pediatrics and Management, several corpora and subcorpora were manually built to answer the following research questions: (1) To what extent is there evidence for genre-specificity related to psychological priming in RAs? (2) To what extent is there evidence for domain-specificity (disciplinary variation) related to psychological priming in RAs? (3) To what extent is there evidence for text-positional association related to psychological priming in RAs? (4) Do users of different languages make similar associations with semantically equivalent words in comparable contexts? With the aid of corpus analysis software programs, four research stages were carried out, each of which focused on a different question. In the first stage, specialized, single-genre corpora were compared with general, multi-genre reference corpora to extract RA-specific keywords. Highly typical keywords could be found, of which nine were selected. Collocates and semantic sets of the selected keywords were then contrasted. As a result, typical collocations and semantic associations were observed in the RA data. In the second stage, Pediatrics corpora were compared with Management corpora to extract both discipline-specific and non-discipline-specific keywords. Extremely typical as well as shared, common keywords were found; six non-discipline-specific keywords were then selected. Collocates and semantic sets of the selected keywords were contrasted, revealing distinguishing collocations and semantic associations for each discipline. In the third stage, the textual position of 18 selected discipline-specific keywords and the position of 2–4-word clusters containing the selected keywords were investigated across RA section subcorpora. While the clusters exhibited stronger textual colligational inclination, most of the keywords do not appear to be primed for use in specific parts of RAs. In the final stage, semantically equivalent, high-frequency nouns were searched in the English, Portuguese, and Japanese data. Textual position, collocates, and grammatical functions of seven selected words were then compared. As a result, textual colligations, collocations, and colligations of English and Portuguese nouns proved to be closer to each other; the associations of Japanese nouns exhibited more distinguishing features. Collectively, the findings provide strong support for Hoey's claims concerning both genre- and domain-specificity and add new layers of understanding to the existing knowledge about text-positional association and cross-linguistic variation. In addition to the theoretical contribution, the findings can be useful to (foreign) language teaching and learning for academic purposes and to academic translation as well.

Keywords: Corpus Linguistics. English for academic purposes. Portuguese for academic purposes. Academic Japanese. Primings.

RESUMO

Como citar esta tese (padrão da Associação Brasileira de Normas Técnicas): ARAGÃO, R. M. L. de. **Padrões do léxico no gênero artigo acadêmico**: Um estudo comparativo de primazia lexical em inglês, português e japonês. 2023. Tese (Doutorado em Estudos Linguísticos e Literários em Inglês) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2023.

Primazia ou aparelhamento lexical é uma teoria orientada para o léxico formulada pelo linguista britânico Michael Hoey que assume que experiências prévias com a língua preparam ou primam usuários de línguas para que se comuniquem de uma ou de outra forma. A mente dos indivíduos é vista como um programa de concordância natural e complexo que associa continuamente palavras, sons e até mesmo sílabas com uma grande variedade de recursos linguísticos e fatores contextuais, de palavras isoladas a gêneros, domínios e situações. Conforme a teoria, o aparelhamento psicológico—um processo associativo—seria a força motriz por trás da configuração típica de uma dada língua. Pressupõe-se que línguas diversas exibam padrões lexicais distintos como resultado de associações e experiências particulares de seus usuários, o que significa que o deslocamento de padrões de uma língua para outra não deve produzir um discurso natural na língua de destino. Esta tese busca contribuir com o corpo de pesquisa de evidências linguísticas do aparelhamento psicológico por meio de um estudo de artigos acadêmicos (AAs) em inglês, português e japonês. A partir de uma coleção de 240 AAs de 10 revistas acadêmicas de Pediatria e Administração, vários *corpora* e *subcorpora* foram construídos manualmente com o intuito de responder a quatro perguntas de pesquisa: (1) Em que extensão há evidência de especificidade de gênero relacionada ao aparelhamento psicológico em AAs? (2) Em que extensão há evidência de especificidade de domínio (variação disciplinar) relacionada ao aparelhamento psicológico em AAs? (3) Em que extensão há evidência de associação de posição textual relacionada ao aparelhamento psicológico em AAs? (4) Usuários de línguas diferentes fazem associações similares com palavras semanticamente equivalentes em contextos comparáveis? Com o apoio de programas de análise linguística, quatro etapas de pesquisa foram realizadas, cada qual dirigida a uma pergunta diferente. Na primeira etapa, *corpora* especializados de um único gênero foram comparados com *corpora* de referência abrangentes de vários gêneros para extrair palavras-chave específicas de AAs. Palavras-chave altamente típicas puderam ser encontradas, das quais nove foram selecionadas. Colocados e grupos semânticos das palavras-chave escolhidas foram então contrastados. Como resultado, colocações e associações semânticas típicas puderam ser observadas nos dados dos AAs. Na segunda etapa, *corpora* de Pediatria foram comparados com *corpora* de Administração para extrair tanto palavras-chave disciplinarmente típicas como palavras-chave disciplinarmente não típicas.

Palavras-chave extremamente típicas, assim como palavras-chave comuns, compartilhadas, foram encontradas; seis palavras disciplinarmente não típicas foram, então, escolhidas. Colocados e grupos semânticos das palavras-chave foram contrastados, o que revelou colocações e associações semânticas distintivas para cada disciplina. Na terceira etapa, a posição textual de 18 palavras-chave disciplinares e a posição de aglomerados de duas a quatro palavras contendo as palavras-chave escolhidas foram investigadas em *subcorpora* de seções de AAs. Ao passo que os aglomerados exibiram inclinação de coligação textual mais forte, a maior parte das palavras-chave não parece primada para uso em seções específicas de AAs. Na etapa final, substantivos semanticamente equivalentes de alta frequência foram buscados nos dados em inglês, português e japonês. Depois, a posição textual, colocados e funções gramaticais de sete palavras selecionadas foram comparadas. Como resultado, coligações textuais, colocações e coligações de substantivos ingleses e portugueses mostraram-se mais próximas entre si; as associações de substantivos japoneses exibiram mais traços distintivos. Coletivamente, os achados sustentam fortemente alegações de Hoey quanto à especificidade de gênero e de domínio e adicionam novas camadas de compreensão ao conhecimento existente de associações de posição textual e variação entre línguas. Além da contribuição teórica, os achados podem ser úteis para o ensino e aprendizagem de línguas (estrangeiras) para fins acadêmicos e para a tradução acadêmica.

Palavras-chave: Linguística de *Corpus*. Inglês para fins acadêmicos. Português para fins acadêmicos. Japonês acadêmico. Associações-primas.

要旨

本論文を引用する場合の引用・参考文献の書き方は次のとおりである（APA［アメリカ心理学会］スタイル）：Aragão, R. M. L. de.（2023).『研究論文ジャンルにおける語彙のパターン―英語，ポルトガル語，日本語をめぐるレキシカル・プライミングの比較研究―』（博士学位論文，サンパウロ大学）.

レキシカル・プライミングとは，英国の言語学者マイケル・ホーイにより提唱された語彙指向理論であり，過去の言語経験が言語使用者のコミュニケーションの基盤を整えることを想定するものである．人の頭脳は，単語，音，音節を，個々の単語からジャンル，分野，状況に至るまで，様々な言語資源や文脈要因と継続的に関連付ける，自然で複雑なコンコーダンス・プログラムと見なされる．この理論によると，連想プロセスである心理的プライミングこそが，言語の典型を形成する原動力である．このことから，異なる言語は，それらの使用者の独自の語彙の連想と経験の結果として，異なる語彙のパターンを示すと考えられる．つまり，ある言語から別の言語に語彙のパターンを応用しても，自然な言語使用を生み出すことはできないと解釈される．本博士論文は，英語，ポルトガル語，日本語で書かれた研究論文を対象とした調査により，心理的プライミングに関する言語的証拠をめぐる研究の蓄積に貢献することを目指すものである．本研究の問いは，以下の 4 点に集約される．（1）研究論文における心理的プライミングに関するジャンル特異性の証拠はどの程度存在するか．（2）研究論文における心理的プライミングに関する専門分野的差異の証拠はどの程度存在するか．（3）研究論文における心理的プライミングに関するテキスト位置と語彙との関連性の証拠はどの程度存在するか．（4）異なる言語の使用者は，類似文脈において，意味的に等価な単語に対して同様の関連付けを行うか．これらの問いに答えるために，小児科学と経営学の二つの分野の学術雑誌の 10 誌から研究論文を 240 編収集し，複数のコーパスとサブコーパスを手作業で構築した．次に，コーパス分析ソフトウェアを使用し，4 段階の研究を実施した．第 1 段階では，研究論文ジャンルの特徴語（keywords）を抽出するために，専門的な単一ジャンルのコーパスと一般的な複数ジャンルの参照コーパスを比較した．特徴度の高い語が多数見つかり，その中から 9 個を選出した．次に，それらの特徴語の共起語（collocates）および意味的関連語群（semantic sets）を対照した．その結果，研究論文ジャンルの単語同士の共起関係（collocations）と意味的連想（semantic associations）が観察された．第 2 段階では，分野特有の特徴語および非分野特有の特徴語を抽出するために，小児科学コーパスと経営学コーパスを比較した．その結果，特徴度の極めて高い語（分野特有の特徴語）と特徴度の極めて低い語（非分野特有の特徴語）が多数見つかり，6 個の非分野特有の特徴語を選出した．それらの特徴語の共起

語および意味的関連語群を対照した上で，それぞれの分野においての単語同士の共起関係と意味的連想が明らかになった．第 3 段階では，分野特有の特徴語 18 個の位置と，それらの特徴語を含む 2 語から 4 語までの単語クラスターの位置を，研究論文セクションのサブコーパスの中で調査した．クラスターはより強いテキスト上の関連性を示したが，ほとんどの特徴語は研究論文の特定の部分に現れる傾向はなさそうである．最終段階では，特に英語，ポルトガル語，日本語の違いに注目し，データから意味的に等価な高頻出名詞を検索した．その中から 7 個の名詞を選出し，テキスト位置，共起語，文法的機能を比較した．その結果，英語とポルトガル語の名詞のテキスト位置と語彙との関連（textual colligations），単語同士の共起関係，文法的関連（colligations）はある程度近いことが確認された．一方，日本語の名詞は異なる特徴を示した．以上をまとめると，本研究の結果はマイケル・ホーイの主張するジャンル・分野特異性を強く支持するものであり，テキスト位置と語彙との関連性および言語間差異に関する既存の知見に新たな理解を付加するものである．本研究の結果は，理論的な貢献に加え，学術目的の言語（外国語）教育・学習および学術翻訳に役立つであろう．

キーワード：コーパス言語学．学術目的の英語．学術目的のポルトガル語．アカデミック・ジャパニーズ．プライミング．

LIST OF RESULTING PUBLICATIONS

Aragão, R. M. L. de. (2022). 96–8–3–2: Signs of lexical priming in academic article introductions. *Linha D'Água*, *35*(2), 118–139. doi: 10.11606/issn.2236-4242.v35i2p118-139.

Aragão, R. M. L. de. (2022). Signs of psychological priming in academic articles from Pediatrics and Management. *Diálogo das Letras*, *11*, e02202. Retrieved from http://periodicos.apps.uern.br/index.php/DDL/article/view/3893.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

APA	American Psychological Association
BNC	(The) British National Corpus
CARS	Create a Research Space (model)
COCA	The Corpus of Contemporary American English
CoPEP	The Corpus of Portuguese from Academic Journals
DOAJ	Directory of Open Access Journals
enMGT	Corpus of Management research articles in English
enMGT(Cl)	Subcorpus of Management research article closing sections in English
enMGT(Mi)	Subcorpus of Management research article middle sections in English
enMGT(Op)	Subcorpus of Management research article opening sections in English
enPED	Corpus of Pediatrics research articles in English
enPED(D+C)	Subcorpus of Pediatrics research article discussion and conclusion sections in English
enPED(Intro)	Subcorpus of Pediatrics research article introductions in English
enPED(Me)	Subcorpus of Pediatrics research article method sections in English
enPED(Res)	Subcorpus of Pediatrics research article result sections in English
enRAs	Corpus of research articles in English
enTenTen15	English Web 2015 Corpus
enTenTen20	English Web 2020 Corpus

HTML	Hypertext markup language
IMRD	Introduction–Methods–Results–Discussion
jaMGT	Corpus of Management research articles in Japanese
jaMGT(Cl)	Subcorpus of Management research article closing sections in Japanese
jaMGT(Mi)	Subcorpus of Management research article middle sections in Japanese
jaMGT(Op)	Subcorpus of Management research article opening sections in Japanese
jaPED	Corpus of Pediatrics research articles in Japanese
jaPED(D+C)	Subcorpus of Pediatrics research article discussion and conclusion sections in Japanese
jaPED(Intro)	Subcorpus of Pediatrics research article introductions in Japanese
jaPED(Me)	Subcorpus of Pediatrics research article method sections in Japanese
jaPED(Res)	Subcorpus of Pediatrics research article result sections in Japanese
jaRAs	Corpus of research articles in Japanese
jaTenTen11 LUW	Japanese Web 2011 Sample Corpus with Long Unit Words
J-Stage	Japan Science and Technology Information Aggregator, Electronic
KWIC Tool	Key-Word-In-Context Tool
MI-score	Mutual Information score
PDF	Portable document format
PhD	Philosophy Doctor
ptMGT	Corpus of Management research articles in Portuguese
ptMGT(Cl)	Subcorpus of Management research article closing sections in Portuguese
ptMGT(Mi)	Subcorpus of Management research article middle sections in Portuguese

ptMGT(Op)	Subcorpus of Management research article opening sections in Portuguese
ptPED	Corpus of Pediatrics research articles in Portuguese
ptPED(D+C)	Subcorpus of Pediatrics research article discussion and conclusion sections in Portuguese
ptPED(Intro)	Subcorpus of Pediatrics research article introductions in Portuguese
ptPED(Me)	Subcorpus of Pediatrics research article method sections in Portuguese
ptPED(Res)	Subcorpus of Pediatrics research article result sections in Portuguese
ptRAs	Corpus of research articles in Portuguese
ptTenTen11	Portuguese Web 2011 Corpus
RA	Research article
SciELO Brazil	Scientific Electronic Library Online Brazil
SciELO South Africa	Scientific Electronic Library Online South Africa
USP	University of São Paulo

TABLE OF CONTENTS

GENERAL NOTES

1. American English has been employed throughout this dissertation, except in quotations from texts published in other forms of English—the excerpts were faithfully reproduced.

2. Citations and references follow the American Psychological Association (APA) style as presented by the University of São Paulo (USP, 2016), with minor adaptations due to the difference in language (the guidelines provided by USP have been designed for theses and dissertations in Portuguese).

3. With respect to paraphrasing, APA (2010) encourages authors "to provide a page or paragraph number, especially when it would help an interested reader locate the relevant passage in a long or complex text" (p. 171). This recommendation has been followed, except for information extracted from Internet websites; page numbers identify paraphrases and summaries of short passages from other sources. In addition, following the pattern provided by APA (2021), the page number is inserted after the non-literal citation, as in the following example: Strunk Jr. and White (2000) define phrase as a set of related words operating as a unit without a verb or a subject (p. 93).

4. With respect to long paraphrases, APA (2021) recommends that the original source should be cited in the first sentence. After it, provided that the text is unambiguous about the continuity of the paraphrase, there is no need to cite the source again. As the previous sentence suggests, this recommendation has also been followed.

5. Differently from the guidelines provided by USP (2016), this entire document has been prepared using both sides of the paper for ease of formatting and visual harmony; in addition, left and right margins were set to 3 cm, irrespective of the side, and page numbers were positioned at the center of the bottom of the page for the same reasons.

6. Although being printable on regular A4 paper, this document was primarily designed for colorful visualization and reading on screen. To reduce the impact of humankind on the planet, please avoid printing it.

7. Fields of knowledge or disciplines can be written in English using either capital letters (e.g., Biology, Philosophy, Engineering) or lower case (e.g., medicine, management, linguistics). In this text, capital letters have been used as a reminder that linguistic aspects may be due to disciplinary reasons.

8. Italics denote lexical examples (words, phrases, sentences) in the Roman alphabet, including those from other sources. In figures where italics have been used to display the main text, however, normal font denotes lexical examples instead. Throughout this dissertation, italics also stress technical terms in their first appearance in the text, unusual expressions, and word meanings.

9. Capitalization is used to indicate sets of semantically related words (semantic sets) as well as the dictionary form of a word (lemma). For example: *Tea*, *coffee*, *water*, *juice*, and *wine* form the semantic set DRINK. As another example: The search for the lemma WRITE in the data can be useful to understand how *write*, *wrote*, *written*, and *writing* have been used.

10. In addition to short quotations, double quotation marks denote translations into English of foreign words, phrases, and sentences, for example *muito obrigado* ("thank you very much").

11. Cited sources in other languages than English that provide an English title have been presented with the English title for ease of presentation and reading. For non-English sources without an English title, the original title is displayed together with an English translation.

12. This dissertation includes many figures and tables. No figure or table has been split between two pages. Consequently, in some cases blank spaces may look excessive. Nevertheless, the current layout benefits readability, which was preferred to economy of space.

13. Corpus analysis software programs are mentioned several times in this document. Their developers (creators), however, are cited only in specific cases. First, they are cited when the software programs are introduced in the main text. Second, they are cited when information provided by the software programs are presented either in the main text or in tables. Third, they are cited to identify the source when screenshots of the programs are shown.

1 INTRODUCTION

Two key concepts in Applied Linguistics are *genre* and *collocation*. Genre is both “a conventional category of discourse based in large-scale typification of rhetorical action” and a “rhetorical means for mediating private intentions and social exigence” (Miller, 1984, p. 163). It has also been acknowledged as a group of communicative events whose members exhibit recognizable, shared purposes (Swales, 1990, p. 58). “Genre constructs and responds to recurring situation, becoming visible through perceived patterns in the syntactic, semantic, and pragmatic features of particular texts” (Devitt, 1993, p. 580). The concept of collocation is generally associated with the following well-known quotation of John R. Firth (1957/1968): “You shall know a word by the company it keeps!” (p. 179) It is defined as “the occurrence of two or more words within a short space of each other in a text” (Sinclair, 1991, p. 170). Genre has played an influential role not only in scholarly research but also in language education (see Swales, 1990; Swales & Feak, 2004; Hyland, 2004; Bawarshi & Reiff, 2010). Collocation is one of the core concepts in Corpus Linguistics, whose applications include language teaching and learning and translation (see Berber Sardinha, 2000; Tognini-Bonelli, 2001; Sinclair, 2004; McEnery & Hardie, 2012).

A theory of language that expands the range of the founding idea behind collocations to the point that it includes genre is lexical priming. Lexical priming theory was formulated by the British linguist Michael Hoey (2004, 2005) and develops from the commonness of collocations. Its central tenet is that associations between words, sets of words, sentence patterns and other linguistic resources have been continuously made, consolidated, and changed by language users, thereby giving languages their current aspect, conventional and to some extent predictable. In Hoey’s (2011) terms:

> [W]hen we encounter language we store it much as we receive it, at least some of the time, and that repeated encounters with a word (or syllable or group of words) in a particular textual and social context, and in association with a particular genre or domain, prime us to associate that word (or syllable or group of words) with that context and that genre or domain. Each use we make of the word (etc.) and each new encounter with it either has the effect of reinforcing the priming or, if the new encounter does not conform to our previous experiences of the piece of language in question, weakens it. So when we repeatedly read the word *winter* in travel writing in the immediate context of *in* (as opposed to *over*, *through* or *within*), the experiences prime us to expect it in such a context and ultimately to reproduce the combination, especially if we write or talk about travel. (p. 155)

Lexical priming has been applied to investigate written and spoken discourse with several purposes. There are studies devoted to language learning and teaching (Ooi, 2013; Jeaco, 2015; Jantunen, 2017); others deal with humor (Partington, 2009; Goatly, 2017; Skalicky, 2018); some investigations address synonyms (Shao, 2017; Bawcon, 2017). Irrespective of the focus, most of the body of research informed by lexical priming is concerned with English. There are, however, studies on German (Pace-Sigge, 2015), Finnish (Jantunen, 2017), and Portuguese (Cunha, 2017), as well as studies involving English and German (Pace-Sigge, 2007) and English and Chinese (Li & Yang, 2017; Shao, 2017, 2018; Wang, 2018).

The appropriateness and usefulness of lexical priming as a theoretical lens for investigating different languages have been shown in previous research. Pace-Sigge (2015) states that lexical priming "can also be seen as relevant when looking at highly inflected languages like Finnish or German" (p. 4). Li and Yang (2017) claim that "lexical priming theory not only can guide English to Chinese translation practice directly, but also can offer new theoretical grounds to the contrastive study between English and Chinese as an academic discipline" (p. 137). Shao's (2018) study

> has shown that the corpus linguistic categories utilised by lexical priming can help identify similarities and differences between candidate synonyms in both English and Chinese. It not only supports the claim that lexical priming is not culture or language specific, but also demonstrates that synonymy can be described in the same way in two languages which do not have any family relations. (p. 188)

Wang (2018) states that her “project shows the potential of using LPT [lexical priming theory] to describe Chinese linguistic features, especially for grammar” (p. 310).

The present study contributes to the growing body of cross-linguistic research into lexical priming theory. It seeks linguistic evidence for priming—a psychological process (Hoey, 2005, 2013)—in the genre of the research article (RA). It consists of a priming analysis of RA data in English, Portuguese, and Japanese from two different disciplines: Pediatrics and Management. From a theoretical perspective, it aims to shed light on the relationships between (psychological) priming and genre (the RA), (psychological) priming and discipline (Pediatrics and Management), (psychological) priming and textual position (RA sections), and (psychological) priming and languages (English, Portuguese, and Japanese). From a practical perspective, it attempts to contribute to academic writing education and scholarly translation, two areas where the RA genre occupies a prominent role.

The remainder of this dissertation is organized as follows. Chapter 2 provides an overview of lexical priming theory mainly based on Hoey’s (2005) groundbreaking work, *Lexical Priming: A New Theory of Words and Language*. It details the central hypotheses proposed by the author and clarifies the notion of priming, which is at the core of the theory. Chapter 2 also presents an adapted version of the lexical priming framework that has been adopted in this study as a conceptual map. Chapter 3 focuses on the body of literature produced after the publication of *Lexical Priming*. Based on the examination of 41 sources, it shows research directions informed by lexical priming, highlighting sources relevant for the purposes of this study. In addition, Chapter 3 specifies gaps in knowledge and introduces the research questions and aims targeted here. Chapter 4 is dedicated to review studies on the RA genre in English, Portuguese, and Japanese, locating the present study in relation to the ever-growing body of research on the RA. Chapter 5 presents the linguistic data (corpora) used in this study, describing the main data sets (focus corpora) as well as additional sets used for comparison (reference corpora). It also shows the software tools used for data analysis. Chapter 6 describes the research methods used in this investigation. It presents the different research stages performed and their constituent steps. Also, Chapter 6 briefly reviews some of the statistical measures that are used in corpus research. Chapters 7, 8, 9, and 10 report the findings of our study, each chapter dealing with a different topic. Chapter 7 explores Hoey's (2005) claim that language users' associations (primings) are genre-specific. Chapter 8 is devoted to domain-specificity (disciplinary variation). Chapter 9 approaches (psychological) priming and textual position, focusing on different sections of RAs. Chapter 10 attempts to reveal differences and similarities in priming across English, Portuguese, and Japanese by means of a comparison of semantically equivalent words. Chapters 7, 8, 9, and 10 include the discussion of findings, dialoguing with relevant literature and indicating practical implications as well. Chapter 11 summarizes the conclusions of the study and describes lessons drawn from it, limitations, and suggestions for further research. To conclude, the dissertation ends with a short autobiographical postscript that describes challenges I faced as a doctoral candidate during the COVID-19 pandemic. This final, personal note serves as a reflection on the experience of doctoral study during a global crisis.

2 LEXICAL PRIMING THEORY: AN OVERVIEW

The linguistic theory of lexical priming was developed by Michael Hoey (2004, 2005) mainly combining three disciplinary areas: Psychology, Corpus Linguistics, and Cohesion Studies. Psychology (e.g., Neely, 1976, 1977; Anderson, 1983/1996) provided insights into the functioning of languages and language users' minds with the notion of priming. Corpus Linguistics (e.g., Sinclair, 1991, 2004; Stubbs, 1996) offered research techniques and concepts to examine large amounts of data and validate theoretical claims. Cohesion Studies (e.g., Winter, 1974, 1977; Halliday & Hasan, 1976; Jordan, 1980; Hoey, 1983) informed lexical priming with additional techniques and concepts that add a qualitative perspective to the overall framework.

Lexical priming assumes that everyone's mind works as a natural and complex concordance software program that generates combinations of words, senses, syntactic structures, and other resources based on previous experiences with language (Hoey, 2005). Unlike an ordinary concordancer, however, the human mind would be capable of associating language with several contextual elements, such as genres and domains. Language users would speak and write choosing words that fit with the context or they might use seemingly incompatible words or expressions to create humor.

Hoey (2005) has formulated ten central hypotheses that represent the backbone of lexical priming theory:

> 1 Every word is primed to occur with particular other words; these are its collocates.
> 2 Every word is primed to occur with particular semantic sets; these are its semantic associations.
> 3 Every word is primed to occur in association with particular pragmatic functions; these are its pragmatic associations.

> 4 Every word is primed to occur in (or avoid) certain grammatical positions, and to occur in (or avoid) certain grammatical functions; these are its colligations.
> 5 Co-hyponyms and synonyms differ with respect to their collocations, semantic associations and colligations.
> 6 When a word is polysemous, the collocations, semantic associations and colligations of one sense of the word differ from those of its other senses.
> 7 Every word is primed for use in one or more grammatical roles; these are its grammatical categories.
> 8 Every word is primed to participate in, or avoid, particular types of cohesive relation in a discourse; these are its textual collocations.
> 9 Every word is primed to occur in particular semantic relations in the discourse; these are its textual semantic associations.
> 10 Every word is primed to occur in, or avoid, certain positions within the discourse; these are its textual colligations. (p. 13)

Before proceeding to their detailed description, three points are worthwhile. First, it is important to note that the hypotheses concern words, not *lemmas*, that is, words sharing the same dictionary form. This has a considerable impact on lexical priming research, as the analytical scope often becomes narrow. It is assumed that *think* and *thinks*, for example, will exhibit different behavior in linguistic data. Second, it is important to note the meaning of the verb *prime*. *Prime* comes from the Latin *primus*, which, according to the Latin dictionary of the University of Notre Dame (2016), means “first” or “foremost.” The verb means *to prepare* (a surface for painting, for example), *to load* (a gun for firing), *to instruct* (The American Heritage Dictionary, 2012; Merriam-Webster’s Collegiate Dictionary, 2012). Therefore, “[e]very word is primed to occur” can be rewritten as “every word is *prepared* to occur.” Hoey (2004) himself defines *prime* (“*primed*”) in the following manner:

> By *primed* I mean that as a word is acquired through encounters with it in speech and writing, it is loaded with the cumulative effects of those encounters such that it is part of our knowledge of the word (along with its senses, its pronunciation and its relationship to other words in the same semantic set) that it regularly co-occurs with particular other words. (p. 23)

> Instead of saying “This word is a noun” or “This word is an adjective” I would argue we should say “This word is primed for use as a noun”. In other words, the word is loaded with the grammatical effects of our encounters with it in the same way as it is loaded with collocational effects. (p. 24)

Third, the use of the passive voice either in the quotations above or in most of the ten central hypotheses is noteworthy. The passive leads to the question as to what or who primes "this word." In Psychology and related fields, as will be clear later, another word or another known stimulus would prime it, with the primed word being called *target* and the priming word *prime*. But from the lexical priming perspective, the answer may be less straightforward, as language users are supposed to be primed in every communicative event ("encounters").

"Every word is primed to occur with particular other words; these are its collocates" (Hoey, 2005, p. 13). The first hypothesis relates to the concept of collocation, which is essential for lexical priming. Collocation is defined by Hoey (2005) as a psychological association between words observed when their co-occurrence is more frequent than could be explained in terms of random distribution (p. 5). While collocates are the co-occurring words, collocation refers to the association between them, which assumes a psychological nature in lexical priming, and so collocation should be understood differently from the way it is usually employed in Corpus Linguistics (Sinclair, 1991; Tognini-Bonelli, 2001). In Corpus Linguistics, the basis of collocation is statistical frequency of co-occurrence alone; in lexical priming theory, the psychological association is also important.

"Every word is primed to occur with particular semantic sets; these are its semantic associations" (Hoey, 2005, p. 23). Hoey (2005) claims that semantic associations exist when language users associate words with groups of semantically related words, which include collocates (p. 24). The notion of semantic association refers back to Sinclair's (1999) "semantic preference" (as cited by Hoey, 2005, p. 24), with the two terms often bearing the same meaning in the literature on lexical priming.[1] The second hypothesis is directly related to the first, because semantic sets can be considered sets of collocates assembled according to meaning. Not only are there frequently co-occurring words but also frequently co-occurring sets of semantically close words.

Although it is conceivable to investigate collocates and semantic sets independently, it seems more logical to investigate the latter sets after the study of collocates. This is because during the study of collocates the number of co-occurring words decreases through exclusion, and consequently the relevance

[1] The full reference of Sinclair's text as presented by Hoey (2005) is the following: "Sinclair, J. M. (1999) The lexical item, in E. Weigand (ed.) *Contrastive Lexical Semantics*. Amsterdam: John Benjamins." (p. 195)

of the remaining words increases, facilitating semantic analysis. Figure 2.1 illustrates the analytical process from collocates to semantic sets for the word *language*. Using Sketch Engine, an online platform developed by Lexical Computing CZ (n.d.) for linguistic research, left collocates of *language* were searched in the British National Corpus (BNC, 2007). The figure contains the 20 strongest left collocates based on logDice, a statistical measure of co-occurrence (Rychlý, 2008).

Figure 2.1 – Top 20 strongest left collocates and semantic sets of *language* (WORD + *language*).

COLLOCATES	SEMANTIC SET MEMBERS	SEMANTIC SETS
English	*ordinary*	
sign	*literary*	
spoken	*everyday*	CONTEXT OR DEGREE OF FORMALITY
foreign	*official*	
written	*common*	
natural	*sign*	
body	*spoken*	MEANS
programming	*written*	
native	*body*	
second	*foreign*	
target	*native*	NATIVENESS OR LEARNING ORDER
ordinary	*second*	
literary	*natural*	
own	*programming*	CLASS
everyday	*indexing*	
official	*English*	PEOPLE OR LANGUAGE
Welsh	*Welsh*	
sexist	*target*	—
common	*own*	—
indexing	*sexist*	—

Source: Designed by the author. Collocates extracted from the British National Corpus (2007) using Sketch Engine. The classification into semantic sets was manual.

As can be seen, most of the collocates could be grouped into sets; only three were left alone. In lexical priming terms, Figure 2.1 suggests that British users of English are primed to combine *language* with *target*, *own*, and *sexist*, as well as with words related to CONTEXT OR DEGREE OF FORMALITY, MEANS, NATIVENESS OR LEARNING ORDER, CLASS, and PEOPLE OR LANGUAGE.

It is necessary to note, however, that the psychological nature of the associations between words, meanings, and other resources involved in communication implies that textual analyses alone ultimately cannot confirm the associations. All that lexical priming research does by means of a corpus is search for reliable but indirect evidence of how language works in the human mind; conclusive proofs are out of its reach. “[T]he existence of a priming for an individual cannot be demonstrated directly from corpus evidence, because a corpus represents no one’s experience of the language” (Hoey, 2007b, p. 9). Therefore, Figure 2.1 suggests but does not attest. Even so, the strength of collocations such as *English language*, *sign language*, and *spoken language* is highly suggestive of an ingrained mental association between these collocates amongst individual speakers of English.

The third hypothesis refers to Pragmatics: “Every word is primed to occur in association with particular pragmatic functions; these are its pragmatic associations” (Hoey, 2005, p. 13). Pragmatic associations relate to links observed between a word or group of words and certain features that fulfill specific pragmatic functions. An example offered by Hoey (2005) is that of *sixty*, which is often associated with the expression of imprecision, for example *about sixty*, *almost sixty*, and *sixty or so*. Hoey (2005) states that there is no clear division between pragmatic and semantic associations. Bawcon (2017), who studied synonyms in a corpus composed of newspaper articles on the Indonesian 2004 tsunami, offers support to Hoey’s (2005) view on pragmatic associations observing that in her data *bodies* is generally neutral, being preceded often by a numeral, whereas *corpses* are accompanied by unpleasant words such as *decomposing* and *stench*, therefore fulfilling a different pragmatic function.

If the study of semantic associations is a logical sequel to the study of collocations, the study of pragmatic associations is likely to be a natural sequel to semantic analysis. Despite the overlap between one and another, it can be assumed that their difference relates to the amount of contextual information included in the analysis. While semantic associations can be perceived through an orientation toward text, pragmatic associations would be perceived through a context-oriented eye. For illustrative purposes, adjective collocates (verbs included) for the word *opinion* were sought in the Corpus of Contemporary American English (COCA; Davies, 2008–). The 20 most frequent collocates were then classified into semantic sets and classes of possible pragmatic functions. Figure 2.2 shows the results. Although only a minor disagreement possibly arises with respect to the semantic sets, the pragmatic functions probably provoke considerable controversy, since they are based on individual judgement.

Figure 2.2 – Top 20 most frequent collocates, semantic sets, and possible pragmatic functions of *opinion* (WORD + *opinion*).

COLLOCATES	SEMANTIC SETS	POSSIBLE PRAGMATIC FUNCTIONS
personal *popular* *humble* *subjective* *honest* *professional* *differing* *dissenting* *favorable* *concurring* *advisory* *informed* *judicial* *unanimous* *valid* *prevailing* *contrary* *conflicting* *editorial* *biased*	DIVERGENCE: *differing*, *dissenting*, *contrary*, *conflicting* PERSPECTIVE: *personal*, *popular*, *subjective*, *professional* POSITIVE QUALITY: *honest*, *informed*, *valid* DOMINANCE: *popular*, *unanimous*, *prevailing* SOURCE: *judicial*, *editorial* IMPORTANCE: *humble* POSITIVE EVALUATION: *favorable* CONVERGENCE: *concurring* PURPOSE: *advisory* NEGATIVE QUALITY: *biased*	**avoid responsibility:** *personal* **express disapproval:** *popular*, *subjective*, *dissenting*, *conflicting*, *biased* **avoid argument:** *humble*, *unanimous* **express approval:** *honest*, *professional*, *favorable*, *informed*, *valid* **express neutrality:** *differing*, *concurring*, *prevailing*, *contrary* **transfer responsibility:** *advisory*, *judicial*, *editorial*

Source: Designed by the author. Collocates extracted from the Corpus of Contemporary American English (Davies, 2008–). The classification into semantic sets and possible pragmatic functions was manual.

Evidently, the study of pragmatic associations requires the examination of the source data. Figure 2.3 provides 10 concordance lines for *biased opinion* extracted from COCA (Davies, 2008–). They were examined to assess whether the hypothesis that this combination is used to express disapproval is correct or not.

Figure 2.3 – Concordance lines for *biased opinion* and pragmatic functions.

LINE	FUNCTION
#1: makes this team go, " Guy said. " Obviously it's a **biased opinion**, but I think a lot of people would agree with me: If someone	**express self-criticism**
#2:: He was a brilliant forensic psychiatrist, but he was, in my **biased opinion**, an even better human being. (Photo-of-Steve-Pitt- JOSH-MANKIEWICZ) (voiceover: Steve had two sons	**express self-criticism**
#3: that women were inherently less competent to hold high-status positions, which is a **biased opinion**. UNIDENTIFIED-MAN-#2# The infamous Google memo, the most important document since the Magna Carta	**express self-criticism**
#4: an example of the best that religious belief can do, in my obviously **biased opinion**. So she worked to understand my worldview. And, you know, I	**express self-criticism**
#5: compare the McCarthy witch hunt of the 1950s to today. Talk about a **biased opinion** that lives in the world of leftism. Talk about witch hunts. All the	**express disapproval**
#6: wasn't only me, as I wouldn't have trusted in my own **biased opinion**. Authorities no less expert than David Foster, a lifelong friend of mine	**express self-criticism**
#7: 's the other guy, the one we disagree with, who holds the **biased opinion**. How, then, are we ever to get at the truth, the	**express disapproval**
#8: the same intellectual level as Creationists. " It seems to me that his **biased opinion** in this comment only highlights the fact that atheists are more hostile to people of	**express disapproval**
#9: If I were, it would only be to protect Waldorf Designs from your **biased opinion** of me, which is based on nothing but good, clean high school fun	**express disapproval**
#10: with some element of hope and resolution. Obviously, I have a very **biased opinion** about the band, but hear this... I have invested the time to enjoy	**express self-criticism**

Source: Designed by the author. Concordance lines extracted from the Corpus of Contemporary American English (Davies, 2008–).

As can be seen, *biased opinion* seems to fulfill two pragmatic functions in the lines. It may express either disapproval of other persons' views or self-criticism. The interesting aspect in these lines is that *biased opinion* unexpectedly might have a positive effect. When directed to others, it sounds quite critical—the "guy . . . who holds a biased opinion" (line 7) does not sound very agreeable. However, when directed to the speaker (writer) for self-criticism, it sounds as a friendly warning to the listener (reader) that the opinion should not be taken too seriously.

In short, then, words and expressions have variable pragmatic associations. An expression such as *biased opinion* might be associated with contexts of criticism of others or less critical qualifications of one's own position. While different contexts will prime different pragmatic associations for a single word or phrase, the proficient language user will have formed the range of associations necessary to trigger the appropriate meaning when the expression arises in a given situation.

"Every word is primed to occur in (or avoid) certain grammatical positions, and to occur in (or avoid) certain grammatical functions; these are its colligations" (Hoey, 2005, p. 13). Grammatical positions are sentence-level positions related to grammatical functions. Grammatical functions refer to those classes typically presented in morpho-syntactic descriptions of languages, such as *subject* (the person or thing that performs an action; also, the person or thing whose state is described), *verb* (the word that connects a subject and an object or complement), *verb classes* (transitive, intransitive, etc.), and *object* (the target of the action of a verb). Although Hoey (2005) introduces lexical priming as an alternative to traditional views of grammar, it is important to note that he does not propose a new grammar, even though he posits grammar as the sum of collocations, colligations, and semantic associations of words, syllables, and sounds. The fourth hypothesis means that some words are loaded in language users' minds to be used at or near the beginning of sentences in the role of subject; other words, by contrast, are prepared to be used at or near the end of sentences in the role of direct or indirect object; and so on. For example, based on Figure 2.3, it is possible to hypothesize that *biased opinion* may be primed to be preceded by either possessive pronouns or the indefinite article *a*; moreover, it is perhaps primed to occur with the function of object or complement, among others. These associations would be colligations of *biased opinion*.

Hoey (2005) states that the basic idea of colligations is that "just as a lexical item may be primed to co-occur with another lexical item, so also it may be primed to occur in or with a particular grammatical function" or "to avoid appearance in or co-occurrence with a particular grammatical function" (p. 43). As both the fourth hypothesis and these excerpts state, colligations may be either positive (co-occurrence) or negative (absence of co-occurrence). For example, *biased opinion* (Figure 2.3) may be negatively primed to occur with the function of subject, that is, to avoid this function to a certain degree, for there is only one instance in which it acts as a subject (line 8).

The fifth hypothesis proposes that "[c]o-hyponyms and synonyms differ with respect to their collocations, semantic associations and colligations" (Hoey, 2005, p. 13). Respectively, co-hyponyms and synonyms are words that can be grouped under the same class, for example, *apple* and *orange* in relation to *fruit*, and words that have similar meanings in some contexts, for example *giants* and *titans* in the sentences *I had a nightmare with giants running after me* and *I had a nightmare with titans running after me*. It is assumed that these word types differ respecting collocates, semantic sets, and co-occurring grammatical functions and positions. For illustrative purposes, an additional search for left collocates using Sketch Engine in the BNC (2007) was carried out, but this time with the word *tongue*, a part of the body that has also extended its meaning to function as a synonym of *language*. The results are shown in Figure 2.4.

Figure 2.4 – Top 20 strongest left collocates and semantic sets of *tongue* (WORD + *tongue*).

COLLOCATES	SEMANTIC SET MEMBERS	SEMANTIC SETS
native	*forked*	
forked	*potted*	
mother	*pink*	
Wolvercote	*sore*	
bellows	*flickering*	PHYSICAL ASPECT
potted	*coated*	
cat's	*swollen*	
pink	*lithe*	
adder's	*twisted*	
sharp	*cat's*	
sore	*adder's*	
flickering	*Lucy's*	OWNERSHIP
coated	*her*	
swollen	*his*	
lithe	*native*	NATIVENESS OR LEARNING ORDER
Lucy's	*mother*	
vulgar	*sharp*	DISCOURSIVE ASPECT
twisted	*vulgar*	
her	*Wolvercote*	—
his	*bellows*	—

Source: Designed by the author. Collocates extracted from the British National Corpus (2007) using Sketch Engine. The classification into semantic sets was manual.

As can be seen, although *language* and *tongue* are synonyms, they have only one common collocate on the left side: *native*. In addition, there is only one semantic set shared by them: NATIVENESS OR LEARNING ORDER. Most of *tongue*'s collocates refer to physical features of real, tangible tongues (PHYSICAL ASPECTS), being unrelated to those of *language*.

These results also relate to the sixth hypothesis: "When a word is polysemous, the collocations, semantic associations and colligations of one sense of the word differ from those of its other senses" (Hoey, 2005, p. 13). Considering that *tongue* has two main meanings, that is, *body part* and *language*, it is clear that most of the collocates in Figure 2.4 refer to body part. There are only a few that refer to language: *native*, *mother*, *sharp*, and *vulgar* (although there is room for discussion concerning *sharp* and *vulgar*, as they seem to be closer to the figurative sense of *tongue* than to the literal sense of *language*). Hoey (2005) hypothesizes that the behavior of a word with different meanings will differ according to the meaning; with respect to *tongue*, this seems to be right.

The sixth hypothesis was extensively explored by Hoey (2005) for the words *consequence* and *reason*. He compared *consequence* with the meaning of result and *consequence* with the meaning of importance showing, among other things, that while the former is primed to co-occur with semantic sets of LOGIC and NEGATIVE EVALUATION, the latter is primed to be preceded by either *any* (*any consequence*) or *of* (*of little consequence*, *of great consequence*, etc.). In addition, he compared *reason* (*cause*), *reason* (*logic*), and *reason* (*rationality*), providing frequencies of several associations for each sense, thus supporting the relationship between sense and linguistic associations. The polysemy hypothesis has three ramifications:

> 1 Where it can be shown that a common sense of a polysemous word is primed to favour certain collocations, semantic associations and/or colligations, the rarer sense of that word will be primed to avoid those collocations, semantic associations and colligations. The more common use of the word will make use of the collocations, semantic associations and colligations of the rarer word but, proportionally, less frequently.
> 2 Where two senses of a word are approximately as common as each other, they will both avoid each other's collocations, semantic associations and/or colligations.
> 3 Where either (1) or (2) do not apply, the effect will be humour, ambiguity (momentary or permanent), or a new meaning combining the two senses. (Hoey, 2005, p. 82)

These ramifications are called by Hoey (2005) as “the drinking problem hypotheses” in allusion to a joke from the film *Airplane!* (Davison, Abrahams, Zucker, & Zucker, 1980) In the film, there is a pilot called Ted Striker (played by Robert Hays) who is not allowed to fly anymore due to a “drinking problem.” The humor arises from the fact that the problem refers to his lack of motor skills to put liquid in the mouth rather than to alcohol misuse or abuse (*problem drinking*). Hoey (2005) notices that although *drinking problem* and *problem drinking* share the same collocation, only the first can be used with the meaning of physiological disorder. According to him, “[t]he more common meaning of alcoholism in effect drives the rarer meaning into a grammatical corner” (p. 82). The drinking problem hypotheses have been adopted by researchers as theoretical starting points to investigate humorous language (e.g., Goatly, 2017; Skalicky, 2018).

“Every word is primed for use in one or more grammatical roles; these are its grammatical categories” (Hoey, 2005, p. 13). Grammatical roles or categories relate to morphological classification, which groups words into classes such as nouns, verbs, and adjectives. Underlying this hypothesis is the idea that words have a relatively fluid nature. A noun or verb may act as a noun modifier, as in *research report*, *book review*; *the suffering flower*, *the smiling boy*; a verb or adjective can become a noun, as in *translating is challenging*, *paraphrasing is demanding*; *the rich and the poor*, *the good-natured*. Also, the view of grammar as a product of priming underlies the seventh hypothesis. In Hoey’s (2005) terms:

> The grammatical category we assign to a word, I want to argue, is simply a convenient label we give to the combination of (some of) the word’s most characteristic and genre-independent primings. It is in fact the outcome of other factors, not the starting point for a linguistic description. (p. 154)

From this standpoint, the fact that the preposition *of* by itself cannot be used before a noun as an adjective would not be the result of its morphological classification; instead, it would be the result of psychological associations made by English speakers and writers throughout history. Therefore, *of* would be primed for use as a preposition rather than be a preposition.

“Every word is primed to participate in, or avoid, particular types of cohesive relation in a discourse; these are its textual collocations” (Hoey, 2005, p. 13). The eighth hypothesis implies a relationship of equivalence between *text* and *discourse*. Although Hoey (2005) does not define the two terms in his explanation of lexical priming, they seem to be sometimes interchangeable (for instance, “discourse organisation” and “text organisation” are likely to share the same meaning) and sometimes slightly different, with the notion of discourse being closer to communicative context. However, if other views are considered, for instance the one by Bhatia (2004), who presents discourse as text, genre, and social practice, Hoey’s (2005) use of the term appears to be more focused on textual aspects. In fact, his previous definition of discourse, that is, that of “any stretch of spoken or written language that is felt as complete in itself” (Hoey, 1983, p. 15), appears to echo in lexical priming theory.

The eighth hypothesis raises the question as to what types of cohesive relation exist in a given text. Hoey (2005) differentiates “cohesive chains” from “cohesive links.” While the former are composed of three or more connected lexical units, the latter are limited to two and tend to be less close to the overall topic of the text. Cohesive chains and links are formed by different kinds of repetition and reference, such as “simple repetition” (*a man* → *the man*), “complex repetition” (*research* [noun] → *researched*), synonym (*the world* → *the planet*), and the use of pronouns (*the guy* → *he*) (Hoey, 2017). Some words would be primed to participate in cohesive relations (positive association); others would be primed to avoid such relations (negative association).

It is noteworthy that the eighth hypothesis stresses an important distinction between lexical priming research and Corpus Linguistics: the possibility of focusing on individual texts. Corpus Linguistics is the study of language through corpora, that is, sets of authentic texts carefully selected to find linguistic regularities (Tognini-Bonelli, 2001, p. 55). Its emphasis is on large amounts of data; a corpus is to be read vertically with the search for patterns in the surroundings of the target word or words (Tognini-Bonelli, 2001, p. 3). By contrast, the study of textual collocations may entail a single-text approach. Hoey’s (2017) detailed analysis of the popular science text “The Invisible Influence of Planet X” illustrates this point. Consequently, the lexical priming framework cannot be considered exclusively from the Corpus Linguistics perspective; there is a more subjective, qualitative aspect that must also be taken into account.

The ninth hypothesis proposes that "[e]very word is primed to occur in particular semantic relations in the discourse; these are its textual semantic associations" (Hoey, 2005, p. 13). Like the preceding one, this hypothesis also implies equivalence between text and discourse. In addition, it also relates to individual texts as the preceding hypothesis does. Textual semantic associations refer to semantic relations observed from a broader paragraph-, section- or text-level perspective. Examples of semantic relations given by Hoey (2005) are those of contrast, time sequence, and exemplification, in addition to that of problem–solution, which refers back to Winter (1974), Jordan (1980, 1984), and Hoey (1983) himself, among others.

The identification of broad semantic relations within a text can be performed in several ways. Although an exhaustive explanation of the topic is beyond the scope of this chapter, I would like to introduce three. One way is to make questions to the text as if the analyst was speaking to the writer (Hoey, 1983). For example, the relation between the two last sentences of the previous paragraph can be discovered by the question: Can you offer examples? It is a generalization–example relation. Another way is by means of observing repetition throughout the text. For example, Hoey (1983) examined a translation of "The Princess and the Pea," by Hans Andersen, of which the following two sentences were extracted:

> "(1) Once upon a time there was a prince and he wanted to marry a princess, only she had to be a *real* princess." (p. 92)
>
> "(24) So the prince took her for his wife, now he knew he had a real princess." (p. 92)[2]

[2] The full reference of the short story as presented by Hoey (1983) is the following: "'The Princess and the Pea' by Hans Andersen, translated by Reginald Spink, in *Fairy Tales and Stories* (London: Dent, 1960)." (p. 203)

According to him, the repetition of *prince*, *real princess*, and *marry* (*took for his wife* is considered an instance of repetition) signals the problem–solution structure of the entire story. A third way to identify semantic relations in a text is by looking at its vocabulary, which relates directly to the logical sequence in question. For example, Jordan (1984), based on the structural analysis of numerous texts, lists many words that show problem–solution relations. Some words that specifically signal a problem are *adverse*, *collapse*, *danger*, *depression*, *fear*, *gap*, *lack*, *suffer*, *time-consuming*, and *weakness*; some that signal a solution are *achieve*, *adapt*, *answer*, *attempt*, *develop*, *enhance*, *improvement*, *minimize*, *promote*, and *suggestion*. Words from the first group would be primed for use in the introduction and description of problems; words from the second for use in the presentation of solutions.

Finally, the tenth hypothesis states that "[e]very word is primed to occur in, or avoid, certain positions within the discourse; these are its textual colligations" (Hoey, 2005, p. 13). This hypothesis also implies a relationship of equivalence between text and discourse, as do the previous two; however, it does not rely on individual texts. As can be noted, there is an overlap between this and the fourth hypothesis (on colligations), as both relate to textual positions. The distinguishing feature is the grammatical aspect, absent here, which facilitates multiple-text analysis. Textual colligations refer to associations between words and certain positions in sentences, paragraphs, or other recognizable textual units (Hoey, 2005, p. 115). Among the several examples provided by Hoey (2005), the following one attracted my attention because, even though I grew up in Brazil, I was also primed for it (at the movies): "When I was a child, I was primed to expect narratives to have the words *The End* in text-final position, in a line of their own and always with initial capitalisation – a textual colligational priming" (p. 188).

Overall, the ten central hypotheses offer possible explanations for linguistic phenomena drawing on the notion of priming. Yet, there is no mention of either *priming* or *lexical priming* in them. Pace-Sigge (2010), in an extensive review on the roots of lexical priming, quotes Collins and Loftus (1975) to point out that the term *priming* was used first by Ross M. Quillian in the 1960s. It is informative to quote here the first paragraph of Collins and Loftus' (1975) article:

> Some years ago, Quillian (1962, 1967) proposed a spreading-activation theory of human semantic processing that he tried to implement in computer simulations of memory search (Quillian, 1966) and comprehension (Quillian, 1969). The theory viewed memory search as activation spreading from two or more concept nodes in a semantic network until an intersection was found. The effects of preparation (or priming) in semantic memory were also explained in terms of spreading activation from the node of the primed concept. Rather than a theory to explain data, it was a theory designed to show how to build human semantic structure and processing into a computer. (p. 407, note number excluded)

Therefore, the origin of the term appears to lie in the field of Artificial Intelligence. As can be seen in the quotation above, *priming* relates to preparation and activation in semantic memory. Later, as the Collins and Loftus' (1975) article itself shows, the term was adopted in Psychology, possibly spreading to other fields from there. The excerpts below, extracted from a journal article in Neuroscience and another in Psychopharmacology, describe it respectively as an unconscious phenomenon and the action of facilitating a process:

> The phenomenon of priming is considered to fall within the domain of implicit memory (Schacter, 1987; Squire, 1986, 1987). Priming occurs when the response to stimulus material is modified by its prior presentation, and this modification can occur without explicit knowledge of the prior presentation. . . . Priming takes place if the subject responds with a previously studied word, but cannot consciously remember having studied the word. (Davis et al., 1990, p. 288)

> Priming is the facilitation of processing a stimulus as a function of its prior exposure. For example, a task often used in implicit memory priming studies is word-stem completion, in which participants are presented with three letter word beginnings (e.g. BLA) and asked to complete them with the first word that comes to mind. (Boucart, Biederman, Cuervo, Danion, & Wagemans, 2002, p. 43)

In the lexical priming framework, the notion of priming was initially expanded to include the meaning of property of words:

> The focus in psycholinguistic discussion is on the relationship between the prime [the word that primes, prepares, loads] and the target [the word that is primed, prepared, loaded], rather than on the priming item per se. In the discussion that follows, however, *priming is seen as a property of the word* and what is primed to occur is seen as shedding light upon the priming item rather than the other way round. (Hoey, 2005, p. 8, italics added)

Hoey (2007b), however, reframed it, limiting the notion to persons:

> The first is that lexical priming *is a property of the person, not the word*. It is convenient sometimes to say, for example, that a word is primed to occur with a particular collocate but this is shorthand for saying that most speakers are primed to associate the word with that particular collocate. (Hoey, 2007b, p. 9, italics added)

This seems to be a delicate matter in lexical priming theory. On the one hand, the theory draws heavily on Corpus Linguistics, which investigates linguistic data. Therefore, it is natural to associate primings with words. On the other hand, it also develops from Psychology, which is concerned with mental processes. Therefore, the psychological nature of priming should not be neglected. Nevertheless, it appears that both *priming* and *primings* are primed for use with the sense of property of words, as the following quotations show:

> "The low number of new collocates of *Turk* in addition to the high proportion of consistent collocates suggests that *the priming of the term* was fairly stable throughout the seventeenth century and not subject to sustained drift." (Baker, McEnery, & Hardie, 2017, p. 53, second italics added)

> "Colligation represents thus one of the levels of description of *a word's primings*." (Cantos & Almela, 2017, p. 232, italics added)

> "In other words, the strength of synonymy among *the eleven candidate synonyms* has been shown in *their primings* with respect to their different proportions in collocations, semantic associations and colligations." (Shao, 2018, p. 134, italics added)

With this respect, the stance adopted in this text is similar to that of Patterson (2016):

> It is to be noted here that primings are primarily a psychological phenomenon within the individual language user. However, the term can also refer to the

> linguistic evidence found within a text. It is in this second case, with the first acknowledged, that primings will be referred to hereafter. (p. 245)

Priming is the psychological process of association between different language units, structures, and contextual elements. Evidence for priming, within the linguistic paradigm, however, is derived from texts. I will return to this idea later.

Another meaning for *priming* is that of associative knowledge a person or group has, which may be subject to changes. This meaning can be found in the following excerpt about the primings of the expression *time has come*:

> It is not important how strong a priming is; once it exists, it may be subject to further primings. To judge by the corpus data, in which 72 of the 93 instances of the phrase occur with *the* (77 per cent), most speakers are primed for *time has come* to occur with *the*. This then is another instance of collocation priming. Actually the sequence of primings may well be different for different people. One person may indeed, as suggested here, first become primed for the co-occurrence of *time* and *has come* and then subsequently be primed to associate *the* with *time has come*. Another, however, may initially be primed to recognize the whole phrase *the time has come* and subsequently encounter instances which might weaken the certainty of the inclusion of *the* in the phrase. It is inherent in the position that I am presenting that there is no 'right' sequence in which primings might occur. Each person is uniquely primed by a unique set of encounters with the word or group of words in question, and the routes by which we come to approximate each other's primings for that word or group of words are likely to be various. (Hoey, 2007b, p. 11)

The primings a person or group holds can change over time. Language users are continuously exposed to familiar and unfamiliar primings, which either confirm and maintain the existing associations or leads to the establishment of new ones. Hoey (2005) claims that primings may "crack," that is, be abruptly changed (as when a teacher corrects a student's utterance at school); they may "drift," that is, change smoothly (the conversion of *Google* into the verb *to google*, for example); and they can be "overridden," that is, be changed by writers or speakers on purpose to produce unusual effects.

With respect to the combination between *priming* and *lexical*, Pace-Sigge (2010) states that it was possibly brought into discussion first in Psychology by James H. Neely, whose work informed Hoey's own work. Neely's (1976, 1977) early texts, however, do not include the term *lexical priming*, only *lexical* and *priming* apart. This raises the question as to whether Hoey's (2005) lexical priming theory was informed by the psychological notion of priming or lexical priming.

Hoey and O'Donnel (2015) offer an answer to this question explaining that lexical priming is based on "a long psychological tradition into what is referred to as semantic priming and repetition priming" (p. 118). According to them, semantic and repetition priming differ in that while the first occurs when the prime and the target are generally associated in the individual's mind, the second occurs when the two words are the same. In addition, they link lexical priming theory to repetition priming:

> Repetition priming potentially provides an explanation of both semantic priming and collocation. If a listener or reader encounters two words in combination, and stores them as a combination, then the ability of one of the words to accelerate recognition of the other is explained. The point is that each time a word is heard in a particular context, that context becomes part of the language user's understanding of the word. (Hoey & O'Donnel, 2015, p. 119)

If lexical priming theory develops from semantic and repetition priming, then the unanswered question is about the inspiration for the word *lexical* in Hoey's (2004, 2005) use of *lexical priming*. Indeed, it seems that *semantic priming* and *lexical priming* were subject to indiscriminate use, as the following excerpts suggest:

> Lexical priming, or semantic priming as it is sometimes called, refers to the well-documented finding that a target word such as *doctor* is more quickly responded to in a lexical decision task or naming task when it is presented after a semantically related prime word such as *nurse* than an unrelated word such as *butter* . . . (Ratcliff, 1987, p. 483)

> Word-based priming is variously referred to as lexical priming, semantic priming, and associative priming. These terms are essentially ambiguous as to the source of the priming. Associative priming or semantic priming can be due to preexisting associations between lexical items that get triggered by words in the text or to new associations expressed by the propositions in the text. Even the term *lexical priming* is ambiguous since the priming could be due to existing intralexical associations among words in the text or to higher-level text factors that prime existing intralexical associations. Because it is important for current debates over the modularity of linguistic analyses to distinguish the source of the priming, we use the terms *word-based* and *text-based priming* to clarify

> where the activation is coming from. (Keenan, Golding, Potts, Jennings, & Aman, 1990, p. 297)

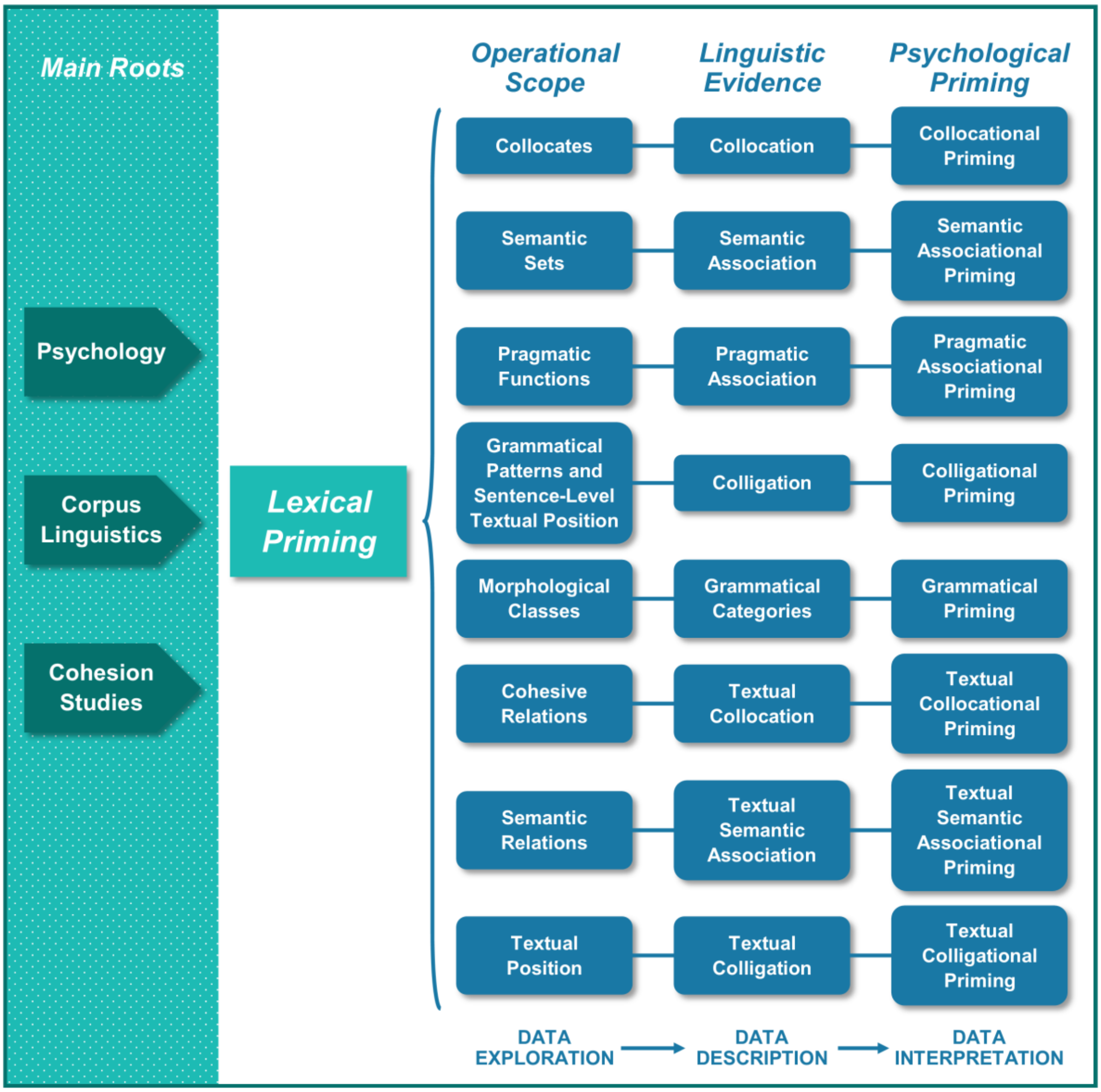
Therefore, it is likely that Hoey's (2005) use relates to the three terms in Psychology: *repetition priming*, *semantic priming*, and *lexical priming*. It should be noted, however, that the meaning that the term has assumed in Hoey's (2005) theory diverges from that of Psychology, which is narrowly associated with an observable effect:

> Lexical priming is a well-established effect in which the processing of a word under certain task conditions is facilitated if a semantically related word has been previously processed. One such task condition involves requiring a lexical decision to be made about a string of letters while the subject is simultaneously listening to a sentence. Thus, a subject might hear *The doctor decided to cure agrammatism through bloodletting*, and if shortly after hearing the word *doctor* in the sentence, that person is required to decide whether the visually presented string of letters NURSE forms a word or not, that decision will be faster than if the letter string had followed some other, unrelated, word in the sentence. (Grodzinsky, Swinney, & Zurif, 1985, pp. 75–76)

At this point, I would like to make the claim that in Hoey's (2005) lexical priming framework the notion of priming is more relevant than that of lexical priming. Indeed, in his presentation of the theory, instances of *lexical priming* are few and sparse in comparison to those of *priming* and *primings*. Moreover, many of the associations addressed by the central hypotheses are rephrased in his text with the word *priming*. For example, collocation becomes "collocational priming"; semantic association becomes "semantic associational priming"; textual colligations are referred to as "textual colligational primings."

Based on this perception and on the above presentation, a modified lexical priming framework was developed as a conceptual map for this work (Figure 2.5).

Figure 2.5 – A modified lexical priming framework.



Source: Designed by the author.

The ten central hypotheses proposed by Hoey (2005), in fact, contain eight sorts of priming: collocation, semantic association, pragmatic association, colligation, grammatical category, textual collocation, textual semantic association, and textual colligation. Their nature, however, is somewhat confusing, because primings under Hoey's (2005) paradigm are psychologically performed but linguistically suggested. In view of this, every priming was rearranged in three related items: *operational scope* (the actual focus of textual analysis), *linguistic evidence* (the linguistic associations identified in data), and *psychological priming* (the psychological associations suggested by linguistic evidence). In addition, three analytical stages were included: *data exploration*, in which the priming analyst explores data to identify associative patterns; *data description*, in which the priming analyst describes the patterns identified numerically, statistically, or textually; and *data interpretation*, in which the priming analyst interprets the linguistic evidence in terms of psychological phenomena, that is, psychological priming.

To finish this chapter, it is necessary to note that there is some debate about the relationship between corpus (linguistic evidence) and mind (the domain of psychological priming). Durrant and Doherty (2010), for example, state that “many researchers remain sceptical regarding the psycholinguistic reality of high frequency word combinations” (p. 126). In fact, there are research works devoted to exploring this relationship (e.g., Durrant & Doherty, 2010; Collins, 2019). In the present study, we acknowledge that what happens in mind and what is seen in corpus may not have such a straightforward connection as both Hoey’s (2005) exposition and Figure 2.5 might imply. Nevertheless, the experimental work needed to enhance knowledge on this topic is beyond our scope and goals. In this respect, we assume a speculative stance, in line with most research on lexical priming to date. The evidence that corpora provide for psychological phenomena may yet be circumstantial, and it must be used with caution; however, as we observed earlier, it remains highly suggestive.

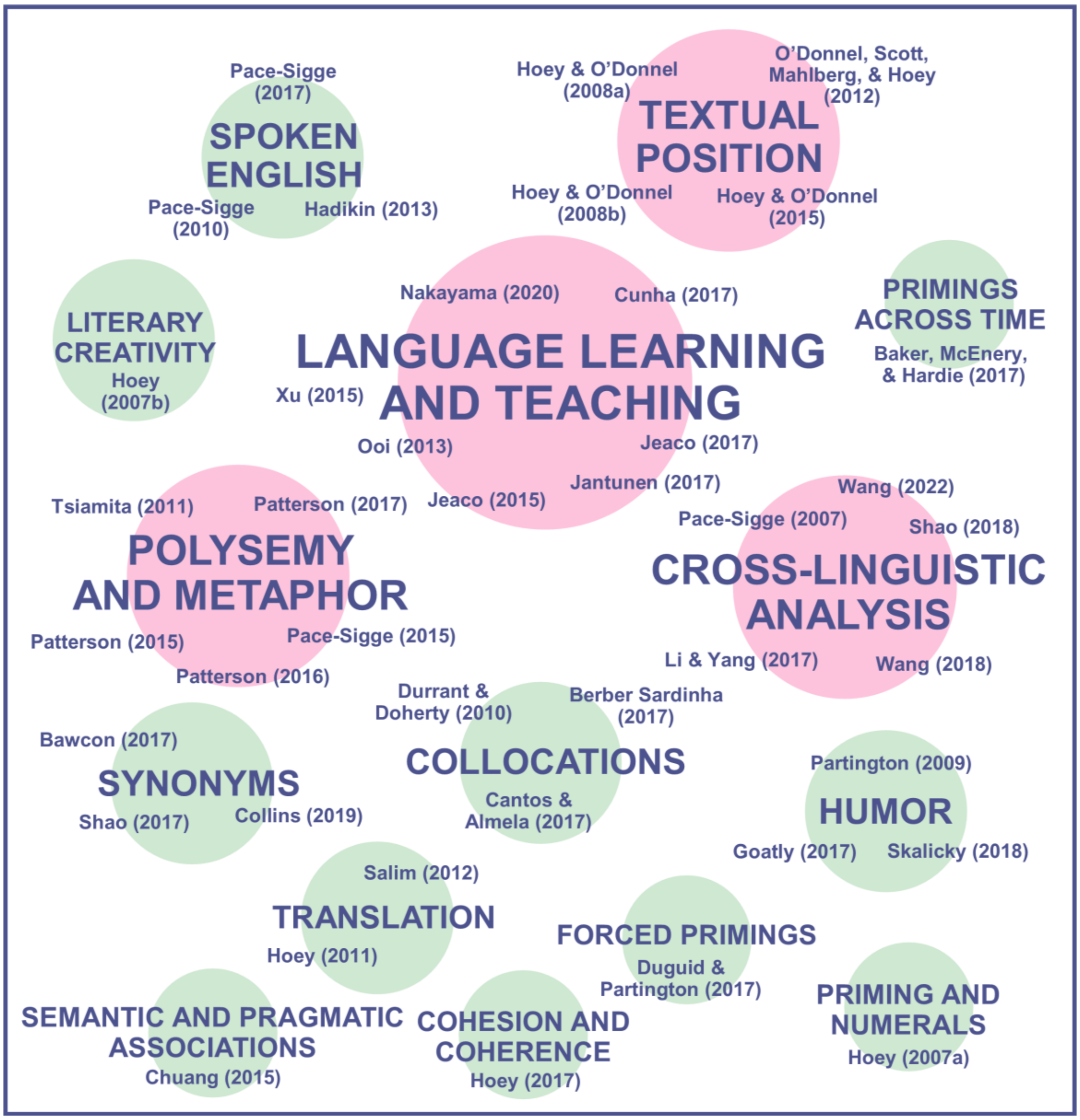
3 LEXICAL PRIMING RESEARCH: DIRECTIONS AND GAPS

The previous chapter presented an overview of lexical priming theory and introduced a modified lexical priming framework for analysis, which is adopted here as a conceptual map. The present chapter summarizes pertinent research literature on lexical priming to show directions of investigation as well as gaps in knowledge about lexical priming and its applications. Also, the description of gaps will inform the research questions and aims of the present study. This chapter puts the analyses reported in the following chapters in a broader context of research.

As the previous chapter has shown, lexical priming theory has a very wide scope. It may be applied to explore linguistic elements from syllables and sounds to genres and domains through both quantitative and qualitative approaches. Studies of lexical priming may involve the software-assisted study of large amounts of linguistic data (corpora) or manual analyses of but a few texts. Hoey's (2005) original framework, despite its theoretical basis and evidence extracted from corpora, is highly speculative in that its backbone is composed of yet to be verified hypotheses. Consequently, it leaves many open questions, which have been subject to study by Hoey (2007a, 2007b, 2011, etc.) himself and other scholars as well.

To provide an overview of the directions of research concerning lexical priming, earlier studies were examined and classified. Specifically, a total of 41 sources were grouped according to subject. Figure 3.1 shows the resulting groups surrounded by their members (represented by author names and publication years).

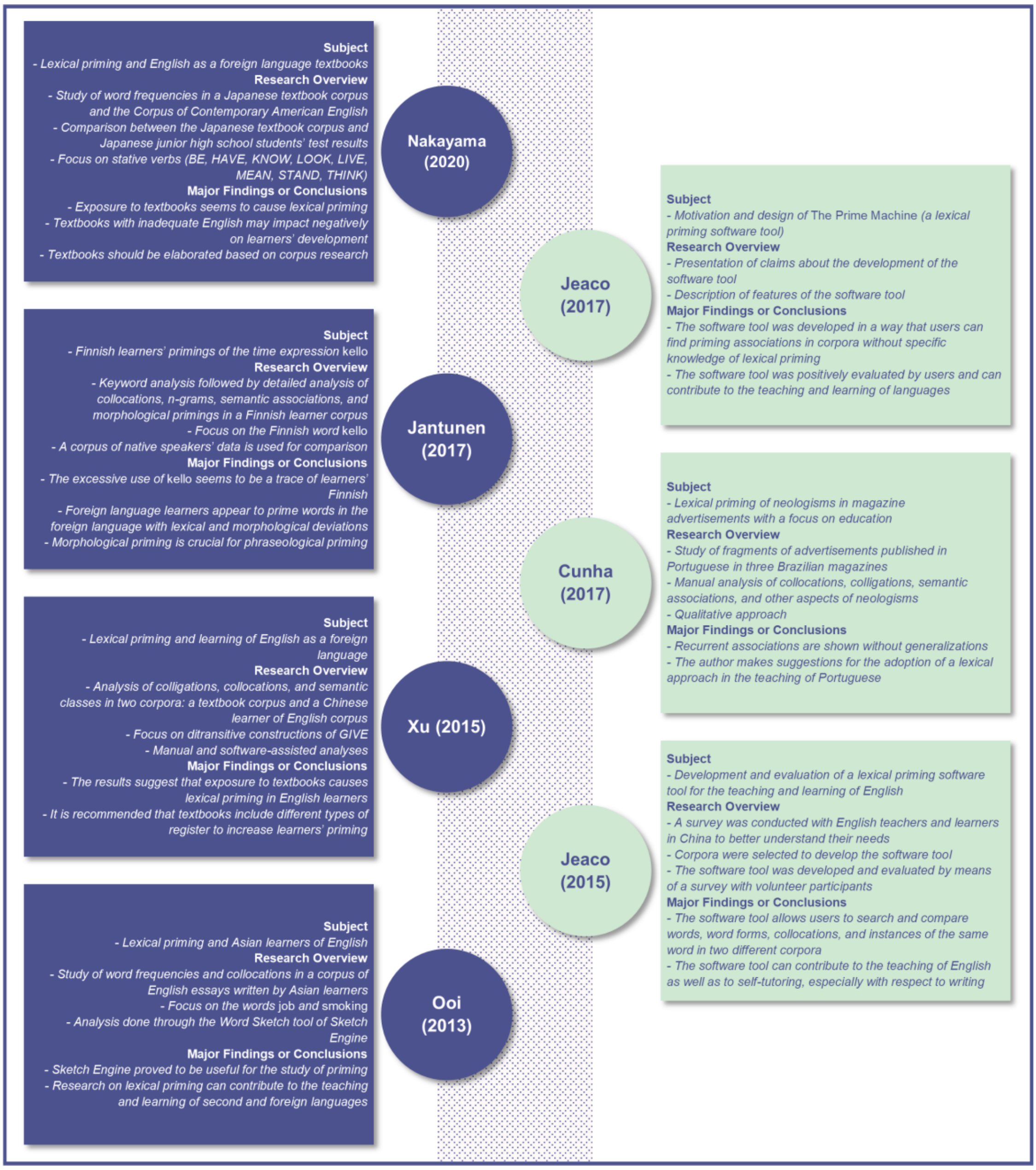
Figure 3.1 – Previous studies concerning lexical priming theory grouped according to subject.



Source: Compiled and designed by the author.

As can be seen, researchers have pursued different directions of investigation after Hoey's (2005) fundamental work. Among them, the adoption of lexical priming in studies relating to the teaching and learning of languages stands out numerically. Jeaco (2015, 2017) developed a software tool for pedagogical purposes based on lexical priming theory. While in his 2015 doctoral work he reports the development and evaluation of the tool, in the 2017 book chapter he presents its characteristics and pedagogical motivations. Ooi (2013), Xu (2015), Jantunen (2017), and Nakayama (2020) used lexical priming for corpus analysis oriented to the learning of foreign languages, more specifically English (Ooi, 2013; Xu, 2015; Nakayama, 2020) and Finnish (Jantunen, 2017). Cunha's (2017) research on neologisms in magazine advertisements aimed at the learning and teaching of Portuguese for Brazilians. Figure 3.2 shows an overview of the seven studies cited.

Figure 3.2 – Overview of the studies in the language learning and teaching group.



Source: Compiled and designed by the author.

This area of research demonstrates the recognition among scholars of the potential of lexical priming for the purposes of language education, especially but not exclusively for foreign language learning and teaching. This collective perception reinforces part of the rationale for the present study, that is, its pedagogical relevance. Studies within the lexical priming paradigm can reveal patterns that contribute to a better understanding of discourse expectations and naturalness, which is particularly worthwhile for language education. Since lexical priming can account for why some collocations seem less natural than others, the relevance to language education is obvious.

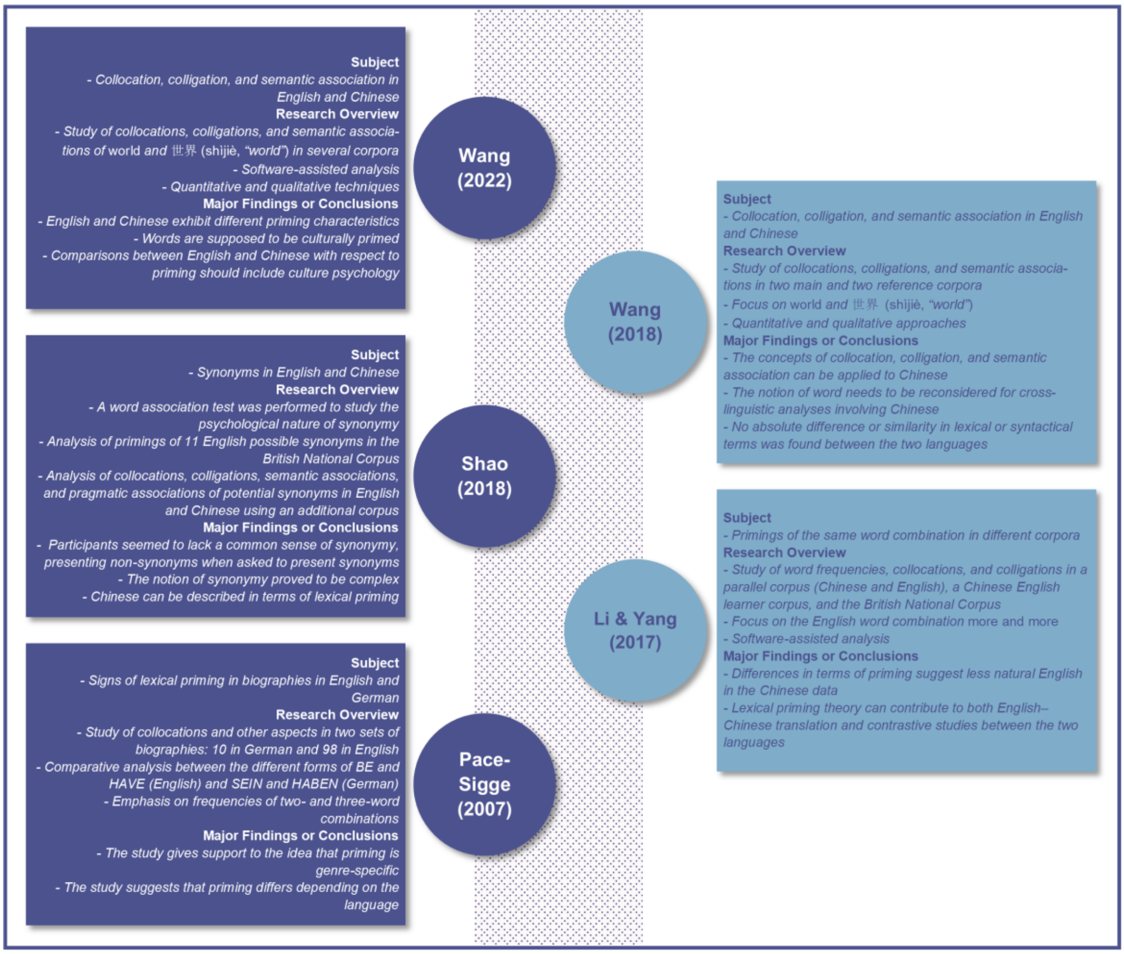
Following language learning and teaching, two research directions stand out among the others, both with five sources each: polysemy and metaphor, and cross-linguistic analysis. Tsiamita (2011) studied collocations, colligations, and semantic associations in a subcorpus of the BNC composed of 432 fiction texts. She focused on two senses of the noun *drive* (*journey* and *private road*) and on several senses of the verb *face*. It is noteworthy that she employed a discourse analysis framework to investigate the meanings of *face* (Burton, 1980 as cited by Tsiamita, 2011), thereby showing a possible way to combine quantitative and qualitative techniques.[1] Pace-Sigge (2015) analyzed primings of different senses of two German words in a corpus of political speeches. Specifically, he analyzed the primings of *Steur* (“tax”), *Steur* (“steering wheel”), *Hut* (“hat”), and *Hut* (“guard,” “care”). He observed differences in primings and concluded that lexical priming can be used to analyze German data. Patterson’s (2015, 2016, 2017) studies focused on metaphoric and non-metaphoric instances of *cultivated*, *flame*, *grew*, and *to kindle*, mainly in written texts of the nineteenth century. Her findings support Hoey’s (2005) claims about variations in primings between the different senses of the same word (that is, the sixth hypothesis), very much like Tsiamita’s (2011) and Pace-Sigge’s (2015) findings.

Cross-linguistic analysis is another research direction, with five sources in Figure 3.1. Pace-Sigge (2007) compared collocations and other aspects of BE and HAVE with those of SEIN and HABEN (the German equivalents of *be*

[1] The full reference of Burton’s (1980) work as presented by Tsiamita (2011) is the following: “Burton, D. 1980. *Dialogue and Discourse: A Sociolinguistic Approach to Modern Drama Dialogue and Naturally Occurring Conversation*. London/Boston/Henley: Routledge & Kegan Paul.” (p. 174)

and *have*). Li and Yang (2017), Shao (2018), and Wang (2018, 2022) investigated primings in English and Chinese. While Shao (2018) and Wang (2018, 2022) compared characteristics of English and Chinese by means of corresponding words (cross-linguistic equivalents), Li and Yang's (2017) analysis was indirect in that they used a parallel Chinese-English corpus to investigate the English phrase *more and more*. Figure 3.3 shows an overview of the five studies.

Figure 3.3 – Overview of the studies in the cross-linguistic analysis group.



Source: Compiled and designed by the author.

Collectively, these works demonstrate that it is possible to analyze data in languages other than English using lexical priming theory. Wang (2018), however, points out the need to reconsider the notion of word for cross-linguistic analyses involving Chinese. Chinese words can be composed of one, two, or more characters, for example 我 (*wǒ*, "I"), 我的 (*wǒ de*, "my"), and 自然度 (*zìrán* *dù*, "naturalness"), and, differently from English words, are not separated by blank spaces. Consequently, there are cases where a series of Chinese characters correspond to a single word in English (e.g., 自然度, *zìrán dù*, above). In view of this, Wang (2018) advocates that the notion of nesting should be employed as "a basic corresponding pattern and departure point for cross-linguistic study between English and Chinese" (pp. 306–307), with an important modification. Whereas Hoey (2005) proposes that nesting is a property "where the product of a priming becomes itself primed *in ways that do not apply to the individual words making up the combination*" (p. 8, italics added), Wang (2018) claims that

> the collocation primed for a *nesting* can also apply to the components constituting that combination, though differentiating in terms of collocational strength as measured by statistical scores (e.g., LogDice). A Chinese nesting is made up of a sequence of characters with a strong collocational strength of co-occurring with each other. The nesting has a featured priming that cannot apply to its components (characters); however, it can share some primings (with a weaker collocational strength) with its constituents. (p. 275)

Using the modified notion of nesting in place of that of word can change the analytical scope from formal correspondence to semantic equivalence between English and Chinese lexical items.

The studies in the cross-linguistic analysis group reveal that there are many comparisons to be made between languages using lexical priming, in addition to the exploration of data in languages different from English. To offer an additional argument in this line, 37 sources shown in Figure 3.1 were also grouped according to the language of the corpus analyzed—Durrant and Doherty's (2010), Jeaco's (2015, 2017), and Collins' (2019) works were not considered because they do not report textual analyses. Table 3.1 displays the results, demonstrating that lexical priming research has, unsurprisingly, been centered on English. There is still a huge gap concerning cross-linguistic and non-English-data analyses as well. The present study aims to fill a small part of this gap.

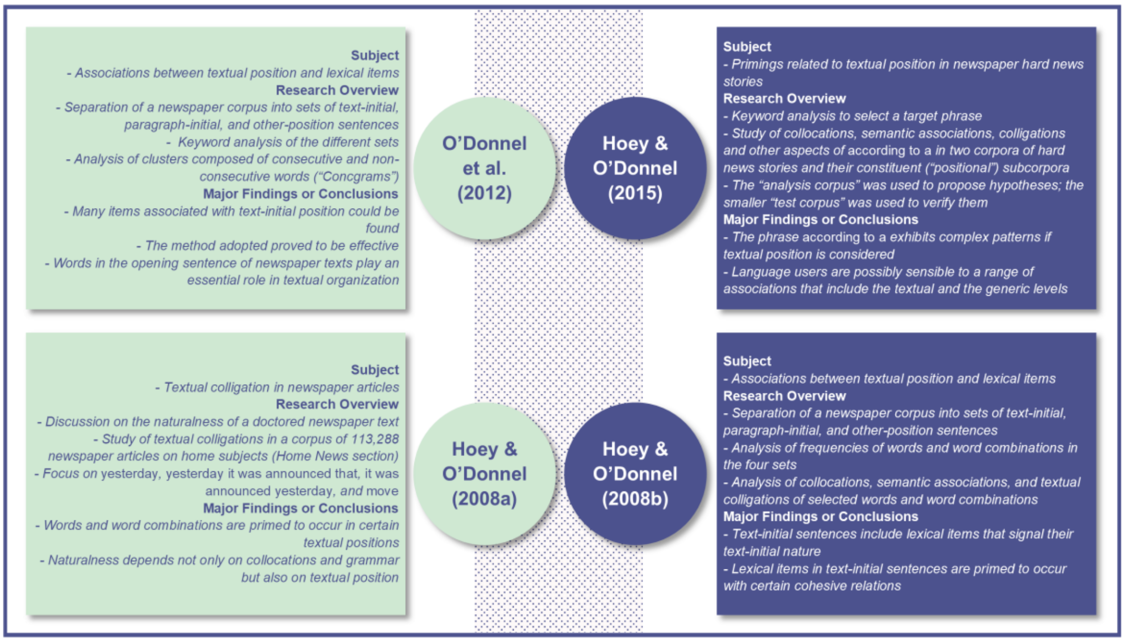
Table 3.1 – Corpus languages of research studies informed by lexical priming theory.

Corpus Language(s)	Number of Sources	Percentage
English	26	70.3
German	1	2.7
Finnish	1	2.7
Portuguese	1	2.7
English and Chinese	6	16.2
English and German	1	2.7
English and Portuguese	1	2.7
Total	37	100.0

Source: Compiled by the author.

The textual position group assembles four sources in Figure 3.1. Indeed, the relationship between textual position and priming has been extensively researched by Hoey and O'Donnel (2008a, 2008b, 2015), including a work coauthored with other colleagues (O'Donnel et al., 2012). Figure 3.4 presents an overview of these four studies.

Figure 3.4 – Overview of the studies in the textual position group.



Source: Compiled and designed by the author.

Altogether, these studies provide support for Hoey's (2005) tenth hypothesis, that is, that words are primed to occur or avoid certain positions. Also, they reveal essential steps to investigate textual colligations. One step is to build subcorpora from corpora according to overall text structure. This allows comparisons between primings of selected items according to textual position. Another important step is to use keyword analysis—a comparative technique that identifies typical words in a given data set in relation to another set—to choose target words or word combinations. As will be seen later, these steps informed part of this study; the present study also pays attention to priming with respect to where in the RA an expression occurs. The relationship between textual position and priming is established as a research line.

The studies of spoken English, synonyms, humor, and collocations can be cited to illustrate the development of the body of research informed by lexical priming in two respects. First, many researchers have sought to test or validate the assumptions of the theory. For example, Berber Sardinha (2017) attempted to verify Hoey's (2005) claims about the role of register-specificity for collocations. In order to do this, he examined collocations of the most frequent words in COCA according to register, drawing on the framework of multidimensional analysis (Biber, 1988 as cited by Berber Sardinha, 2017).[2] His results provide evidence for the assumption that language users are primed for collocations according to register. Durrant and Doherty (2010) conducted two task-based experiments in an attempt to evaluate the psychological reality of high-frequency collocations. Their results confirm the psychological existence of high-frequency collocations but suggest "that collocational priming may be restricted to word pairs which score very highly on association measures (*MI* and *t*-score)" (p. 145), which adds a layer of complexity to the whole picture. Collins (2019) used eye-tracking (a technique from Psychology that observes eye movements) to verify the psychological validity of Hoey's (2005) claim about different associational behavior among synonyms. As a conclusion, she states that "the results . . . conclude in support of Hoey's (2005) claim as part of his lexical priming theory that synonyms will differ in terms of the colligatory information stored about them" (p. 111).

Second, many scholars have sought to expand the range of lexical priming research. The studies on humor follow this line. Partington (2009)

[2] The full reference of Biber's work as provided by Berber Sardinha (2017) is the following: "Biber, D. 1988. *Variation across Speech and Writing*. Cambridge: CUP. doi: 10.1017/CBO9780511621024" (p. 226).

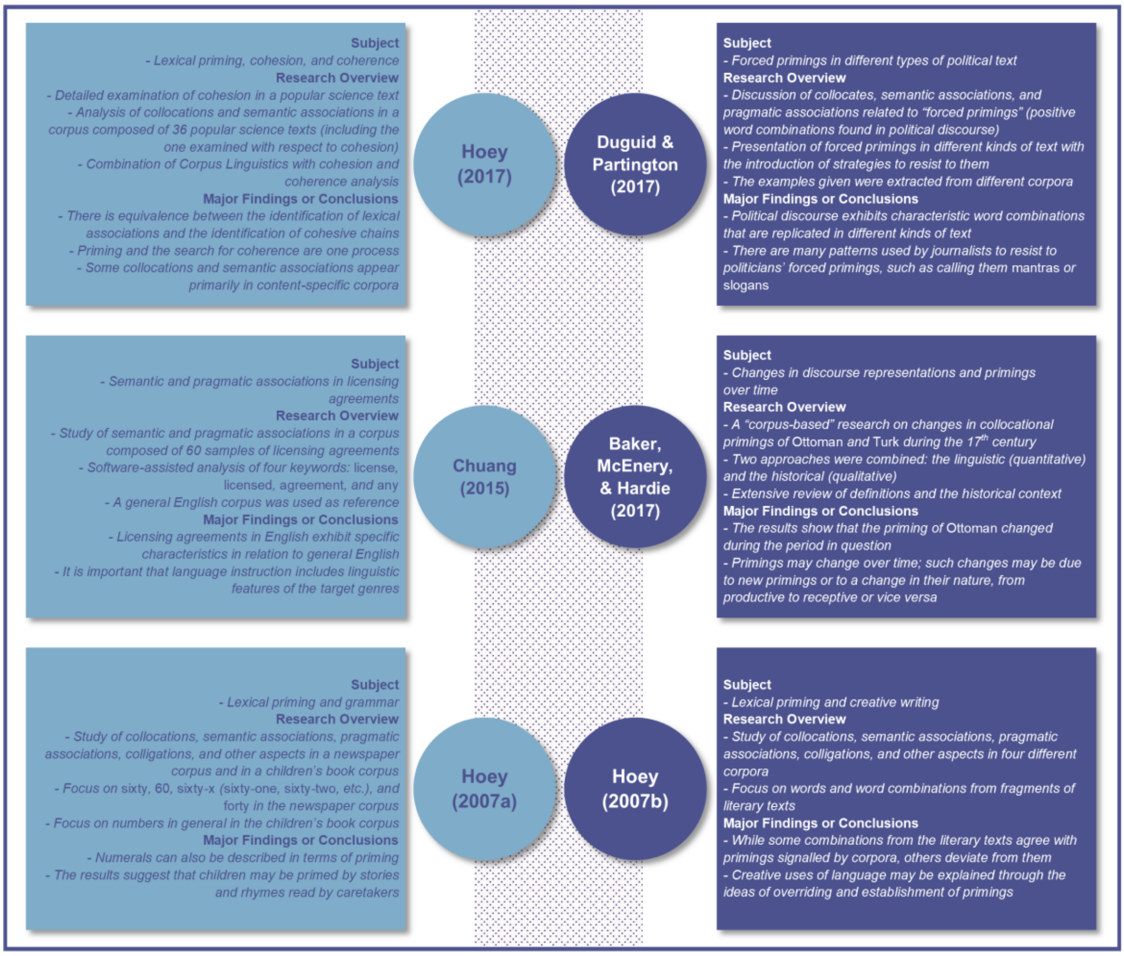
researched the structure and function of punning wordplay, making a detailed analysis of real instances extracted from a newspaper text corpus. Combining Corpus Linguistics and lexical priming, his analysis shows how lexical priming can be applied to describe and explain wordplay, especially through the idea that humor arises from differences between expected primings and the discourse as it is actually realized. Goatly (2017) examined a small collection of humorous utterances to evaluate the hypothesis that "overriding priming can account for humorous effects" (p. 57). He compared the selected utterances with collocations and other associations from the Cobuild Bank of English, concluding that the notion of overriding of priming can be used to explain humor, but there are humorous mechanisms that seem to be outside the range of lexical priming theory. Skalicky (2018) studied word frequencies, collocations, and semantic associations in a corpus of satirical newspaper headlines and in COCA. He focused on three word combinations, namely *shitty enough to*, *death tool*, and *military action*, observing that many satirical headlines exhibit low-frequency collocations and semantic associations to create humor. In line with Goatly (2017), however, Skalicky (2018) also observed that there are humorous mechanisms outside the range of lexical priming. He points out that some satirical headlines require background or knowledge of genre to be understood.

In a similar vein, the works assembled in the spoken English group also expand the range of lexical priming. Pace-Sigge (2010) analyzed collocations, colligations, semantic associations, and grammatical patterns of selected items in a corpus of spoken Liverpool English and three corpora of spoken UK English. He observed shared primings between the two forms of spoken English as well as specific primings of the Liverpool English. Hadikin (2013) studied word frequencies and collocations in two corpora of Korean English (collected in Liverpool and Seul) and two corpora of British English focusing on three high-frequency word combinations: *do you know*, *but you know*, and *and you know*. She observed differences in primings between the corpora, revealing traces of what she calls "Korean English." Finally, Pace-Sigge (2017) analyzed collocations and textual colligations of selected items in three corpora of British English. One of his findings is that textual collocations for turn-taking strategies are negatively primed for monologues. Also, he observed that language users seem to be primed to follow recognizable patterns in conversations.

Among the remaining research directions, the one on translation is of special interest for the purposes of this study. Hoey (2011) examined the translation into Portuguese of an English literary clause ("In winter Hammerfest is a thirty-hour ride by bus from Oslo," by Bill Bryson), comparing the behavior of the English and Portuguese words involved in terms of frequency, semantic association, and other aspects related to priming. He shows, for example, that while *in winter* seems to have a strong association with TIMELESS TRUTH, the corresponding Portuguese *no inverno* appears to be closer to SPECIFIC EVENT, which means that *no inverno* may not be the best choice to translate *in winter*. His analysis demonstrates how lexical priming theory can be applied to translation. Salim's (2012) subject was the translation of the Quran through a lexical priming lens. He compared collocations, semantic associations, colligations, and other features of eight words (*mercy*, *torment*, *say*, *believe*, *will*, *day*, *deeds*, and *shall*) in three different corpora: one composed of English translations of the Quran, one of an English translation of the Bible, and the BNC. As a result, he observed new kinds of priming related to punctuation, orthography, exophoric references, and pronouns. He also found that while primings in the translations of the Quran differ from those in general English, the Quran and the Bible corpora share some primings. Overall, his findings provide support for Hoey's (2005) assumptions concerning domain- and genre-specificity. These studies are of interest here because they confirm the relevance of lexical priming for translation. This is another line of reasoning behind this study.

Concerning the directions that are limited to one source, Figure 3.5 presents an overview of their corresponding studies.

Figure 3.5 – Overview of independent studies.



Source: Compiled and designed by the author.

Among the six studies, Chuang's (2015) analysis of semantic and pragmatic associations in licensing agreements is pertinent to our present purposes. The licensing agreement is described by the author as a "subgenre of contracts," but the study explores genre-specificity by comparing a specialized corpus composed of licensing agreements with a general corpus of English. It is one of the few sources listed in Figure 3.1 that effectively deals with a specific genre.

Considering all the research directions reviewed above, this work primarily situates in the cross-linguistic analysis group. Our main corpora are composed of RAs in English, Portuguese, and Japanese, and analytical procedures were chosen and carried out with these three languages as core elements. Moreover, one stage of this study is in the textual position group, as it is dedicated to the relationship between priming and position throughout RAs.

Additionally, as stated before, our rationale draws on both the language learning and teaching and the translation groups, as they represent important areas of application of lexical priming.

In addition to the diverse research trends, the literature reviewed in this chapter also indicates gaps in knowledge. One gap refers to genre-specificity. Although genre has been considered decisive for priming (Hoey 2004, 2005, 2013), lexical priming research has not focused on well-defined, specific genres. In fact, most studies reviewed in this chapter focused on large, multi-genre corpora, for example Hoey (2007a, 2007b), Patterson (2015), and Jantunen (2017). As a consequence, the relationship between priming and specific genres remains underexplored. A second gap refers to domain- or discipline-specificity. Even though domain has also been considered decisive for priming (Hoey, 2004, 2005, 2013), the relationship between priming and specific fields of knowledge or disciplines has not received much attention. A third gap concerns research on textual colligations. Although textual position represents an established area of research within the lexical priming paradigm, studies have focused chiefly on newspaper texts (Hoey & O'Donnel, 2008a, 2008b, 2015). Therefore, it remains unclear whether textual position has a similar connection with language elements found in other sources of data. A fourth and last gap is the one already cited of scarcity of priming analyses involving languages other than English. As Table 3.1 shows, although there are investigations involving German, Portuguese, Finnish, and Chinese, they are but few in number.

The present study addresses the four gaps introduced above. Specifically, we looked for answers to the following research questions: To what extent is there evidence for genre-specificity related to psychological priming in RAs? To what extent is there evidence for disciplinary variation (domain-specificity) related to psychological priming in RAs? To what extent is there evidence for text-positional associations related to psychological priming in RAs? And to what extent do users of different languages make similar associations involving semantically equivalent words in comparable situations? Additionally, these questions have led to the following four aims: (1) to assess the presence of genre-specific signs of psychological priming in RAs published in English, Portuguese, and Japanese; (2) to assess the presence of discipline-specific signs of psychological priming in RAs in English, Portuguese, and Japanese; (3) to evaluate the presence of position-specific signs of psychological priming in RAs in English, Portuguese, and Japanese; (4) to describe similarities and differences between primings of equivalent words in English, Portuguese, and Japanese.

4 THE RA GENRE: A CROSS-LINGUISTIC OVERVIEW

The previous chapter presented an overview of lexical priming research showing both investigative directions and gaps in knowledge. It also introduced our research questions and aims. The present chapter summarizes pertinent literature on the RA genre from a cross-linguitic perspective, involving sources in English, Portuguese, and Japanese. This review will situate the analyses reported in the following chapters in the context of the body of knowledge on the RA genre.

The RA is probably one of the most researched genres in the broad fields of Languages for Specific Purposes and Applied Linguistics. Although receiving several names in different languages (Figure 4.1), the RA has achieved such a high degree of standardization that texts belonging to this class can in general be easily recognized by readers who are members of the discourse community of researchers, irrespective of the label (and possibly the language).

Figure 4.1 – Common designations for the RA.

English	**Portuguese***	**Japanese****
article	*artigo acadêmico*	原著 (*gencho*)
experimental-research paper	*artigo científico*	原著論文 (*gencho ronbun*)
paper	*artigo completo*	実践研究論文 (*jissen kenkyū ronbun*)
research article	*artigo original*	研究論文 (*kenkyū ronbun*)
research paper	*artigo de pesquisa*	論文 (*ronbun*)
scholarly article	*relato de pesquisa*	
scientific paper	*trabalho completo*	

Source: Based on Aragão (2012, 2017). * English translations (top–down order): “academic article,” “scientific article,” “complete article,” “original article,” “research article,” “research report,” “complete work.” ** English translations (top–down order): “original work,” “original paper,” “applied-research paper,” “research paper,” “paper.”

A widely known and often quoted description of the RA is the one below provided by Swales (1990):

> The research article or paper . . . is taken to be a written text (although often containing non-verbal elements), usually limited to a few thousand words, that reports on some investigation carried out by its author or authors. In addition, the RA will usually relate the findings within it to those of others, and may also examine issues of theory and/or methodology. (p. 93)

The body of research on the RA genre in English is vast and continuously increasing. Two influential works are those by Bazerman (1988) and Swales (1981/2011). Bazerman (1988) made a comprehensive study of the experimental RA, performing a series of context-oriented textual analyses of RAs from several disciplines, such as Physics, Psychology, and Political Science. Textual analyses were supplemented by other research techniques, for example interviews and document research, which added a sociological layer to the study. The author characterizes scientific writing as a social practice full of constraints but also with room for change and creativity. The RA ("experimental article") is understood from a socially oriented view of genre:

> What we recognize as the genre of the experimental article embodies many . . . regularized formal and procedural elements. Genre, then, is not simply a linguistic category defined by a structured arrangement of textual features. Genre is a sociopsychological category which we use to recognize and construct typified actions within typified situations. It is a way of creating order in the ever-fluid symbolic world. (Bazerman, 1988, p. 318)

Swales (1981/2011) examined 48 RA introductions from different disciplines grouped into the Hard Sciences, the Biological/Medical field, and the Social Sciences. He described their structure in terms of *rhetorical moves*, that is, broad components oriented to context, and *rhetorical steps*, narrow components oriented to text. As a result, he observed that introductions to journal articles begin with general concerns, such as social issues or disciplinary trends, and end with particular information on the article or the study behind it. This rhetorical path was systematized into a model named Create a Research Space (CARS), which is composed of three moves: Establishing a Territory, in which authors introduce their field to the reader as lively and important, providing a short review of relevant literature; Establishing a Niche, in which authors either show alignment with previous research traditions or make a shift, developing opposing arguments against existing knowledge; and Occupying the Niche, in which authors present the article's scope by stating purposes, by outlining the research design, by summarizing findings, or by indicating the organization of the remainder of the article (Swales, 1981/2011, 1990).

Swales' (1981/2011, 1990) propositions provided the basis for what has become known as *move analysis*, a context-oriented but text-focused technique that seeks to capture rhetorical configurations and patterns. Move analysis has been used to describe entire RAs (Nwogu, 1997; Posteguillo, 1999), groups of RA sections (Ruiying & Allison, 2003), and individual RA sections as well (Crookes, 1986; Brett, 1994; Holmes, 1997; Anthony, 1999; Lim, 2006; Chahal, 2014). Overall, move analyses dedicated to RAs in English have shown similarities and differences among disciplines, indicating the relevant link between textual organization and disciplinary practices and knowledge.

Move analysis has also led to the development of what can be called *step analysis*, the study of specific components of particular genres. One example of this kind of textual microanalysis is the one by Shehzad (2011). She examined 56 RA introductions in Computer Science searching for the steps "outlining purposes," "stating the nature of the present research," and "listing research questions or hypotheses," which are described specifically in Swales and Feak's (2004) writing guide. As a result, she found few occurrences of purposes (12,5% of the sample) and research questions or hypotheses (32,1%), but the incidence of the description of research was high (85,7%). The author also described how the steps appear in the sample, listing, for example, words employed to state purposes (*aim*, *objective*, *goal*, etc.) and accompanying subject matter, such as the contrast with previous research and the indication that the present study is an extension of previous investigations.

Another line of investigation of RAs in English is the one focused on lexico-grammatical aspects. Tarone, Dwyer, Gillette, and Icke's (1981) analysis of the passive voice in two Astrophysics articles is a classic example of this line. Their main finding is presented as follows:

> [W]e have found that a count of active and passive verb forms in two professional journal papers in astrophysics shows that the active voice is used much more frequently than the passive, and, more importantly, that the active first person plural *we* verb form seems to be regularly used at strategic points in these papers. (Tarone et al., 1981, p. 135)

Other studies on lexico-grammatical aspects of RAs in English are those by Hyland (1996) on hedges in Molecular Biology, by Gledhill (2000) on phraseology in Medicine, by Harwood (2005) on the use of First-Person pronouns in Management, Computer Science, Economics, and Physics, and by Cortes (2013) on lexical bundles in introductions. Two distinguishing features between Tarone et al.'s (1981) analysis and these studies refer to data and scope. There is a shift from data expressed in number of RAs to data expressed in number of words (often big). In addition, the analytical scope has increased so that bridges between lexico-grammatical aspects and either contextual or broader textual features became a common form of exploration and description.

As shown in Chapter 3, lexical priming research has not directed its attention to the RA genre. However, it is possible to draw a parallel between research on the RA genre in English and lexical priming research, even though psychological aspects normally are outside the scope of the former. Move analysis is likely to be close to the study of textual collocations (cohesive relations), textual semantic associations (semantic relations), and textual colligations (textual position). Step analysis and investigations into lexico-grammatical features seem to be especially close to the study of collocations (collocates) and colligations (grammatical patterns and sentence-level position).

The body of research dedicated to the RA in Portuguese is far less extensive than that on the RA in the English language, but it often makes an interesting counterpoint to English-centered research. Some examples of studies on the RA in Portuguese are those by Dias and Bezerra (2013), Costa (2015), Brauer and Portela (2017), Bernardino and Pacheco (2017), and Miranda (2021). Dias and Bezerra (2013) investigated 10 RA introductions in Public Health, comparing their structure with the model proposed by Swales (1990). As a result, they found all the moves and most steps of the model in the data. The authors conclude that, despite contextual differences between their and Swales' (1981/2011, 1990) samples, the model properly accounted for the target introductions. Costa (2015) compared 10 RAs in Linguistics and 10 RAs in Medicine considering disciplinary culture. Whereas she observed textual differences apparently related to cultural variation, she noticed shared communicative strategies in the data that suggest the existence of general, cross-disciplinary academic conventions. Brauer and Portela (2017) investigated the incidence of verbs in RA introductions in Economics (44 introductions), Sanitary and Environmental Engineering (48 introductions), and Linguistics (55 introductions). With the aid of a corpus analysis software program, they observed frequent verbs and collocations, showing that verbal processes play a support role in RA introductions. Bernardino and Pacheco (2017) investigated 30 RA introductions in Nutrition, drawing on Nwogu's (1997) description of medical articles, as well as on Costa's (2015) analysis. Their main conclusion is that Nutrition introductions are similar to previous descriptions of Medicine introductions. Miranda (2021) built a RA corpus of over 12 million words to create a Brazilian Portuguese academic verb list and to compare cross-disciplinarily high-occurrence verbs. After creating the list, she examined sample concordance lines containing 10 verbs: *apresentar* ("present"), *utilizar* ("utilize," "use"), *realizar* ("perform," "carry out"), *considerar* ("consider"), *observar* ("observe"), *encontrar* ("find"), *ocorrer* ("occur"), *mostrar* ("show"), *analisar* ("analyze"), and *determinar* ("determine"). As a result, she observed general, cross-disciplinary characteristics in terms of verb usage; however, she also found disciplinary variation.

My first Master's degree research (Aragão, 2012) was also dedicated to RAs in Portuguese. It involved the examination of Brazilian journals' submission norms and the analysis of RA introductions in Botany, Chemistry, Pneumology, and Special Education (12 introductions per field). The study mapped organizational standards recommended by Brazilian journals for the writing of RAs, showing the high incidence of structures based on the Introduction–Methods–Results–Discussion (IMRD) pattern—Figure 4.2 provides a few examples. The study also revealed organizational characteristics of Portuguese RA introductions, showing that they normally answer at least two out of the three following questions: (1) What is the RA about? (2) Why is the investigation reported in the RA relevant? (3) What was aimed at or done? Answers to these questions seem to constitute the backbone of Brazilian Portuguese RA introductions.

Figure 4.2 – Examples of IMRD-based structures recommended by Brazilian journals for the writing of RAs.

Introdução, Material e Métodos, Resultados e Discussão
(Introduction, Material and Methods, Results and Discussion)

Introdução e Objetivo, Método, Resultados, Discussão
(Introduction and Aim, Method, Results, Discussion)

Introdução, Métodos, Resultados, Discussão, Conclusões
(Introduction, Methods, Results, Discussion, Conclusions)

Introdução, Métodos, Resultados, Discussão e Considerações Finais
(Introduction, Methods, Results, Discussion and Concluding Remarks)

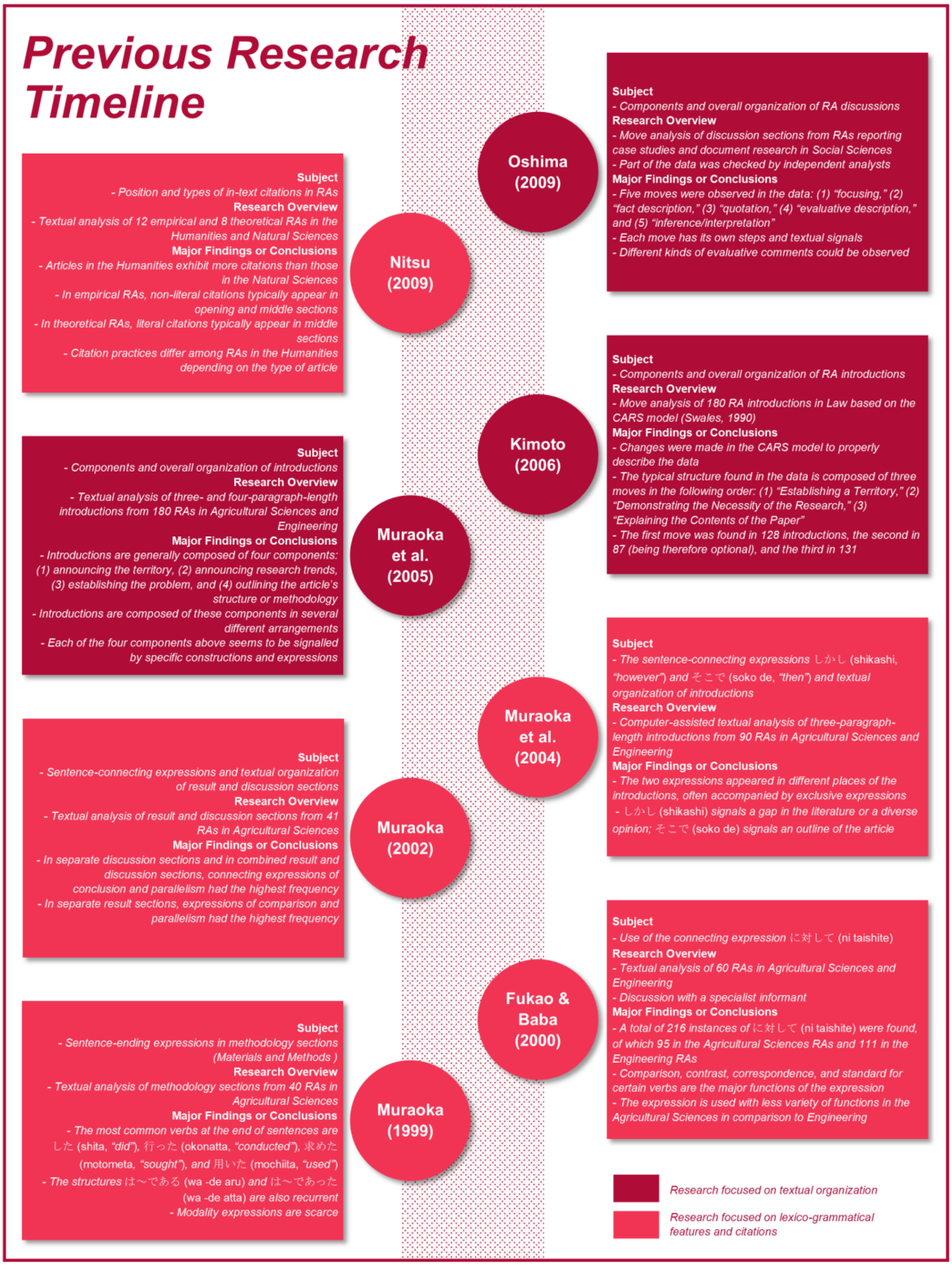
Introdução, Objetivos, Revisão da Literatura, Métodos, Resultados, Discussão, Conclusões
(Introduction, Aims, Literature Review, Methods, Results, Discussion, Conclusions)

Source: Based on Aragão (2012, p. 56).

Overall, research on RAs in Portuguese shares the same foci as the English tradition. On the one hand, it investigates organizational features and components; on the other, it deals with lexico-grammatical aspects. Therefore, again a parallel can be drawn with lexical priming research. The first approach is close to lexical priming's focus on broader, text-level associations; the second, to its focus on co-occurring words and grammatical patterns.

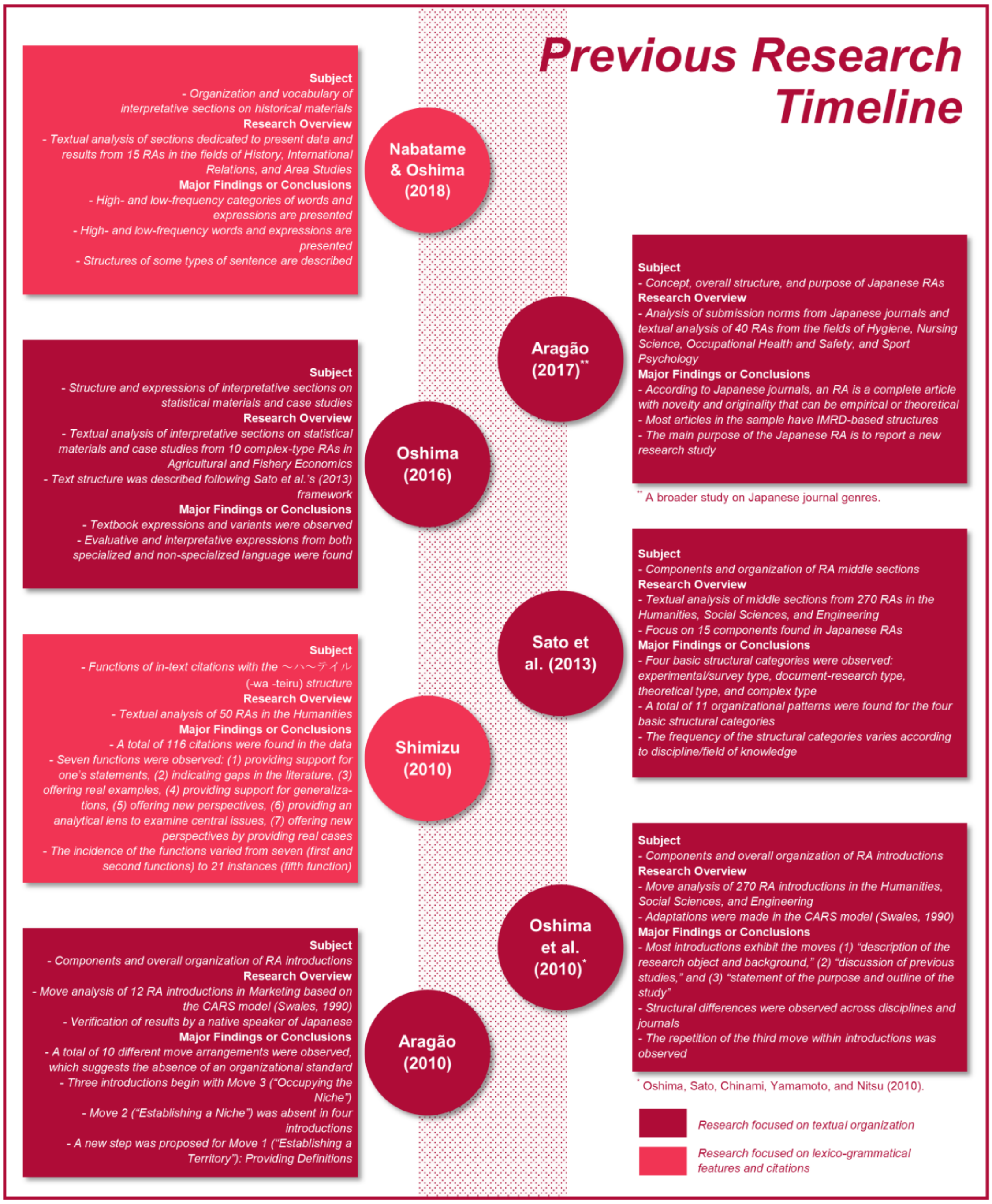
The body of research dedicated to the RA in Japanese is also less extensive than that on the RA in the English language. Its bulk has been published in few specialized journals, especially *Senmon Nihongo Kyōiku Kenkyū* (English title: *Journal of Technical Japanese Education*). Figures 4.3 and 4.4 summarize representative previous research studies in this body, including two authored by myself, one of which—my second Master's degree research (Aragão, 2017)—covered not only the RA but also other prominent genres found in Japanese journals.

Figure 4.3 – Previous studies on the RA in Japanese (Part I).



Source: Compiled and designed by the author.

Figure 4.4 – Previous studies on the RA in Japanese (Part II).



Source: Compiled and designed by the author.

As the figures above show, there are two main directions of investigation into the RA in Japanese. The first is devoted to textual organization; the second deals with lexico-grammatical features and citations. While many studies on textual organization consist of move analysis (Kimoto, 2006; Oshima, 2009; Aragão, 2010; Oshima et al., 2010), being informed therefore by Swales (1990), there are studies with other approaches to RA structure. Sato et al. (2013), for example, considered the following 15 items in their analysis of RA middle sections: 研究の対象／背景の説明 (*kenkyū no taishō*/*haikei no setsumei*, “research subject”/“background explanation”), 先行研究の検討 (*senkō kenkyū no kentō*, “examination of previous research”), 研究の目的の提示 (*kenkyū no mokuteki no teiji*, “presentation of research aims”), 研究行動の提示 (*kenkyū kōdō no teiji*, “presentation of research actions”), 研究方法の説明 (*kenkyū hōhō no setsumei*, “explanation of research methods”), 数式の提示 (*sūshiki no teiji*, “presentation of formulae”), 結果（量的データ）の提示 (*kekka* [*ryōteki dēta*] *no teiji*, “presentation of results [quantitative data]”), 結果（質的データ）の提示 (*kekka* [*shitsuteki dēta*] no *teiji*, “presentation of results [qualitative data]”), 資料（量的データ）の提示 (*shiryō* [*ryōteki dēta*] *no teiji*, “presentation of material [quantitative data]”), 資料（質的データ）の提示 (*shiryō* [*shitsuteki dēta*] *no teiji*, “presentation of material [qualitative data]”), 考察 (*kōsatsu*, “discussion”), 結論の提示 (*ketsuron no teiji*, “presentation of conclusions”), 提言 (*teigen*, “suggestions”), 研究の評価 (*kenkyū no hyōka*, “research evaluation”), and 今後の課題の提示 (*kongo no kadai no teiji*, “presentation of the next challenges”). Oshima’s (2016) analysis drew on Sato et al.’s (2013) framework.

It is noteworthy that cross-disciplinary analyses have shown disciplinary variation in terms of textual organization in Japanese RAs. Oshima et al. (2010), for example, observed that RA introductions in Resources and Material Engineering follow a more fixed pattern in comparison to introductions in Civil Engineering. They also observed structural differences between RA introductions in Social Sciences and RA introductions in Management and Economics.

Many studies in the second direction report the frequency, position, and function of selected lexical items. There seems to be a preference for connecting expressions, such as しかし (*shikashi*, “however”), 従って (*shitagatte*, “therefore”), and また (*mata*, “also”), although verbs and other word types have also been considered (e.g., Muraoka, 1999). Studies in this area also provide some evidence for disciplinary variation. Fukao and Baba (2000), for example, observed different functional behavior for the Japanese connecting expression に対して (*ni taishite*, “in contrast” or “toward,” among other translations, depending on the case) in Agricultural Sciences and Engineering RAs.

For the most part studies on lexico-grammatical features and citations involved manual analysis, which suggests the possible dominance of handmade, qualitative approaches. Only Muraoka et al. (2004) report the aid of software programs for data analysis. They used optical character recognition technology to prepare RA data, then assembled and organized the resulting text files according to sentences and paragraphs using a spreadsheet software program. After that, they examined instances of high-frequency word combinations with a software tool specifically designed for this purpose.

Overall, the same foci observed in the bodies of research on the RA in English and Portuguese can be seen in the body of research on the RA in Japanese. Therefore, once again it is possible to draw a parallel with lexical priming research. While works on textual organization can be related to textual collocation, textual colligations, and textual semantic associations, works on lexico-grammatical features and citations can be related to collocations and colligations, as well as to pragmatic associations, as the Japanese tradition often considers functional analysis.

In addition to studies dedicated to RAs published only in English, Portuguese, or Japanese, there are studies whose data set was composed of RAs in two of the three languages, more precisely either English and Portuguese or English and Japanese. Examples of comparative studies between English and Portuguese RAs are those by Rezende and Hemais (2004), Moritz, Meurer, and Dellagnelo (2008), Hirano (2009), and Romero and Joseph (2014). Rezende and Hemais (2004) investigated hedges in RAs written in Portuguese, in native-speakers’ English, and in non-native-speakers’ English (Brazilian English) from the Health Sciences. They identified and classified hedges according to section of occurrence, grammatical classes, and purposes. As a result, they found 96 hedges, with a higher frequency in introductions and discussion sections. An interesting finding of their work is that Brazilian researchers seem to avoid indicating a gap in previous research. Overall, Rezende and Hemais' (2004) study indicates that RAs written in Brazilian Portuguese and those in English exhibit differences regarding hedges. Moritz, Meurer, and Dellagnelo (2008) attempted to describe the rhetorical organization of conclusions of RAs in Applied Linguistics. Like Rezende and Hemais (2004), they also compared texts written in Portuguese, in native-speakers' English, and in non-native-speakers (Brazilian) English. By means of move analysis techniques (Swales, 1981/2011, 1990), they noticed the absence of a recurrent pattern in the data. In addition, they state that:

> In general, the results also seem to corroborate findings in the field of contrastive rhetoric as EL2 writers are portrayed as more influenced by their own linguistic and rhetorical background when favoring a more elaborate style of writing. Furthermore, they are also pictured as following certain "universal" social conventions in the organization of their messages, whose purposes are recognized as fitting the genre of RAs, otherwise their papers would certainly not have been published. (Moritz, Meurer, & Dagnello, 2008, p. 247)

Hirano (2009) investigated RAs in Applied Linguistics (specifically, English for Specific Purposes). Her sample was composed of 20 articles, of which 10 in Portuguese and 10 in English. Employing the CARS model (Swales, 1990) as her initial framework, she compared the 20 RA introductions. One of her main findings is that the introductions of RAs in Portuguese in her sample do not fit into the model; indeed, they do not seem to follow a pattern. She summarizes this point in the following manner:

> In summary, RAIs [research article introductions] in *The ESPecialist* do not seem to follow a pattern in the organization of their rhetorical moves although there seems to be a preference for the M1–M3 [Move 1–Move 3] type. In this type of structural organization, the authors establish a territory (by making topic generalizations and giving background information, for example) and then present their current study (by announcing the present research and indicating the structure of the article, for example). (Hirano, 2009, p. 244)

Romero and Joseph (2014) compared data from 223 RAs in Portuguese and 150 in English. Using a concordance software program, they searched for modality words (*may*, *should*, etc.), then classified the words found and made a comparison between Portuguese and English. The authors conclude that there is a high degree of similarity regarding modality words between the two languages, with Portuguese authors usually opting for the verb *poder* ("can") and English-language authors preferring *may* and *can*.

Comparative studies of RAs in English and Japanese are likely to be fewer than those of RAs in English and Portuguese. Indeed, only three could be found. The first is Kobayashi's (2003) analysis of RAs in Applied Linguistics. Combining move analysis (Swales, 1990) with contrastive rhetoric (Kaplan, 1966, among others), she examined 30 RAs in English and 30 RAs in Japanese to identify possible differences in logical development and in the use of cohesive strategies. As a result, she observed both similarities and differences. Although the overall structure was found to be similar, many of the Japanese texts did not exhibit an introduction. There were also differences in the order of the constituent moves of each section and in the number of "lexical signals," that is, words or phrases used very often that signal moves and steps (Kobayashi, 2003, p. 95). The Japanese RAs seem to have more lexical signals, except in discussions and background sections. The second study found is Unedaya's (2003) analysis of voice. She compared the use of active and passive voice in 10 Japanese and 10 English RAs in Applied Magnetism. Her focus was on sentences that introduce figures and tables (e.g., *Figure 1 shows the main characteristics of the RA in Japanese*). She observed that, while in English data the passive was frequent, in Japanese data it is not used to introduce figures and tables. A possible explanation for this is that subject omission is possible in Japanese but not in English. Finally, the third study is Shibata's (2013) investigation into causation. He attempted to answer the following research question: "What comprises implicit causation in Japanese academic writing and how are the linguistic aspects of Japanese different from English in terms of construing causal-effect relations?" (p. 10) In order to do this, he examined 14 RAs in Japanese and nine RAs in English, all of which were in the field of Linguistics (Second Language Acquisition), focusing on introductions. The author identified and classified causality markers; in addition, he analyzed lexico-grammatical devices and causal types. After finding similarities and differences in the data set, Shibata (2013) arrived at the following conclusion:

> Despite the similar text type or genre to which the data belong, the analysis has shown that the two languages differ not only typologically but also in terms of the lexicogrammatical patterns realizing cause and effect relations . . . This research has revealed that Japanese texts have much more structural conjunctions than English texts, whereas English texts have implicit causal semantics using logical and experiential grammatical metaphor whose causality meaning may not always be transparent. (p. 48)

The present study can be situated not only in relation to the body of research informed by lexical priming but also in relation to the body of research on the RA genre. As shown in the previous pages, research on the RA genre in English, Portuguese, and Japanese has focused on textual organization—including constituent moves and steps—and lexico-grammatical aspects. As will be clear in the following chapters, this study addresses psychological priming by investigating words, word combinations, collocations, semantic associations, colligations, and textual colligations. Therefore, it is closer to the research devoted to lexico-grammatical aspects of discourse organization. Moreover, the present study enters the existing debate about disciplinary and cross-linguistic variation in the RA genre, thereby interacting with much of the research literature reviewed in this chapter.

5 CORPORA AND TOOLS

The data set of this study is composed of 240 RAs from 10 different journals. The journals were selected based on language and subject area from the following reputable collections and databases: Directory of Open Access Journals (DOAJ, https://doaj.org), Scientific Electronic Library Online Brazil (SciELO Brazil, https://www.scielo.br), Scientific Electronic Library Online South Africa (SciELO South Africa, http://www.scielo.org.za), Scopus® (https://www.scopus.com), and the Japan Science and Technology Information Aggregator, Electronic (J-Stage, https://www.jstage.jst.go.jp). Table 5.1 introduces the selected journals.

Table 5.1 – Selected journals.

Journal	Primary Language	Field	Collection or Database
BMC Pediatrics	English	Pediatrics	DOAJ
South African Journal of Child Health	English	Pediatrics	SciELO South Africa
JAMA Pediatrics	English	Pediatrics	Scopus®
International Journal of Business Science and Applied Management	English	Management	DOAJ
South African Journal of Economic and Management Sciences	English	Management	SciELO South Africa
Academy of Management Journal	English	Management	Scopus®
Revista Paulista de Pediatria[a]	Portuguese	Pediatrics	SciELO Brazil
Revista de Administração de Empresas[b]	Portuguese	Management	SciELO Brazil
Journal of the Japanese Society of Pediatric Surgeons[c]	Japanese	Pediatrics	J-Stage
Journal of Business Management[d]	Japanese	Management	J-Stage

Source: Compiled by the author. [a] English translation: *São Paulo Journal of Pediatrics*. [b] English translation: *Journal of Business Administration*. [c] Japanese title: 日本小児外科学会雑誌 (*Nihon Shōni Geka Gakkai Zasshi*). [d] Japanese title: 日本経営学会誌 (*Nihon Keiei Gakkaishi*).

English was chosen as the main language for two reasons. The first is because it has achieved a special status in research communication around the world (Swales, 1990, 2004; Gibbs, 1995; Flowerdew, 2008), including Brazil (Rodrigues, 2011; Oliveira Jr., Schuster, & Levkowitz, 2014; Menezes & Caregnato, 2018). This status increases the need for specific instruction and teaching materials for non-anglophone researchers, which, in turn, rely heavily on the available research literature on academic English. It is necessary to advance the knowledge about academic English to offer accurate instruction for students and prospective researchers, especially non-native speakers, who find themselves in a disadvantaged position concerning writing (Flowerdew, 2008). The second reason is because lexical priming theory (Hoey, 2005) was formulated based on data in English. Consequently, it is important that data in this language be analyzed to assess its assumptions.

Portuguese and Japanese were chosen as additional target languages because both have distinguishing features that make comparisons with English a potentially effective approach for investigating differences in priming related to language, as well as for broadening the range of lexical priming theory. For example, Bennett (2010) describes academic Portuguese from Portugal in the following manner:

> In Portugal, for example, much academic production in the humanities is couched in a style that would seem to have more in common with literary discourse than scientific. Typical features include a taste for ‘copiousness’, manifested by a general wordiness and much redundancy; a preference for a high-flown erudite register (including complex syntax, lexical abstraction, etc); a propensity for indirectness, meaning that the main idea is often embedded, deferred or adorned at all ranks; and the extensive use of figurative language and other forms of subjectivity. (p. 22)

Davies (2000) points out that Japanese rhetorical preferences follow a “different set of cultural imperatives in which sociocultural factors emphasizing aesthetic qualities and empathic forms of expression associated with subjective human feelings and intuition are said to be paramount” (p. 122).

Pediatrics and Management were chosen because they represent two distinct broad areas of knowledge, respectively Health Sciences and Social Sciences, which makes them suitable to investigate differences in priming related to domain or discipline. They were also chosen because in Brazil and Japan researchers from both fields seem to work in two directions, that is, while addressing local topics, such as specific business issues and regional diseases, they appear to seek international collaboration and visibility. Although developing a full argument to support such a claim is beyond the focus of this chapter, the following passage on the internationalization of *Revista de Administração de Empresas* (RAE below), one of the journals chosen for this study, illustrates it:

> The journal went international, and internationalization became part of Brazilian educational policy. Government programs and the assessment criteria generated by *CAPES*, National Council for Scientific and Technological Development (*Conselho Nacional de Desenvolvimento Científico e Tecnológico [CNPq]*) and other state and local organizations that foster research also pressed and are still pressing for internationalization. This includes journals as the front line in this process. *RAE* started publishing material from foreign authors, included professors from other countries as referees, appointed foreigners to its Editorial Board and organized Forums which foreign authors and organizers took part in. (Bertero, 2021, p. 7)

As Table 5.1 shows, Portuguese journals were selected exclusively from SciELO Brazil, and Japanese publications exclusively from J-Stage. English journals, however, were selected from three different collections: DOAJ, SciELO South Africa, and Scopus®. Different collections were employed for journals in English to obtain a reasonably balanced representation of the language. Both DOAJ journals are based in the United Kingdom; those from SciELO South Africa are obviously based in South Africa; and the remaining two from Scopus® are based in the United States. Even though the six journals are international in terms of contributors, it was expected that their articles could reflect slightly different forms of English, making the sample more diverse.

Twenty-four articles were selected from each of the 10 journals, making up a total of 240 RAs (Table 5.2). They were selected based on publication category, period of publication, electronic availability, and technical compatibility. Documents published under designations unrelated to research processes, such as *essay* and *book review*, were not considered, since they represent genres other than the RA. Only articles published between 2011 and 2018 were considered in order to obtain a contemporary sample. Articles had to be available online in portable document format (PDF) to be included. Articles whose main text could not be electronically extracted were rejected.

Table 5.2 – Overview of the selected RAs.

Journal	Publication Category	Amount	Period
BMC Pediatrics	Research Article	24	2011–2016
South African Journal of Child Health	Research/Article	24	2013–2018
JAMA Pediatrics	Original Investigation	24	2013–2017
International Journal of Business Science and Applied Management	Unspecified	24	2011–2018
South African Journal of Economic and Management Sciences	Unspecified	24	2011–2016
Academy of Management Journal	Unspecified	24	2011–2018
Revista Paulista de Pediatria[a]	Original Article[e]	24	2016–2018
Revista de Administração de Empresas[b]	Article[f]	24	2017–2018
Journal of the Japanese Society of Pediatric Surgeons[c]	Original Article[g]	24	2013–2018
Journal of Business Management[d]	Submission Paper[h]	24	2015–2018
	Total/Overall Period	240	2011–2018

Source: Prepared by the author. [a] English translation: *São Paulo Journal of Pediatrics*. [b] English translation: *Journal of Business Administration*. [c] Japanese title: 日本小児外科学会雑誌 (*Nihon Shōni Geka Gakkai Zasshi*). [d] Japanese title: 日本経営学会誌 (*Nihon Keiei Gakkaishi*). [e] Portuguese original category: *artigo original*. [f] Portuguese original category: *artigo*. [g] Japanese original category: 原著 (*gencho*). [h] Japanese original category: 投稿論文 (*tōkō ronbun*).

Each article was stored as a separate PDF electronic file. The files were grouped into separate folders according to the journal of origin and then had their main content transferred to plain text files saved with UTF-8 encoding, which is necessary for software-assisted analysis. In cases where an article also had a hypertext markup language (HTML) version available on the Internet, the PDF file was stored as a backup, and the content was transferred directly from the HTML version. During the transference, titles, abstracts, footnotes, and endnotes were included; authors' names and affiliation, journal information, keywords, tables, figures, acknowledgements, references, and appendices were excluded. Part of the data demanded further revision due to unexpected typos; a small part of the data had to be manually inserted due to technical issues. The Japanese files had to undergo a segmentation process, as Japanese texts do not display spaces or other delimitating characters between words. The segmentation was performed with SegmentAnt, a software tool developed by Anthony (2017). For illustrative purposes, Figure 5.1 shows the result of the segmentation process of a short Japanese quotation on the notion of genre extracted from an article by Gyogi and Iwasaki (2019).

Figure 5.1 – Segmented Japanese text (example).

Original Japanese Text	「ジャンル」には様々な定義があるが，ここでは，ある社会的目的達成のための特定の言語構造や特性を有するテキストとする（例えば，Hyland 2007, p.149, Iwasaki & Kumagai 2015, p.x[2]）。
Romanized Japanese Text	“Janru”niwasamazamanateigigaaruga,kokodewa,arushakaitekimokutekitasseinotamenotokuteinogengokōzōyatokuseioyūsurutekisutotosuru(tatoeba, Hyland 2007, p.149, Iwasaki & Kumagai 2015, p.x[2]).
Segmented Japanese Text	「 ジャンル 」 に は 様々 な 定義 が ある が ， ここ で は ， ある 社会 的 目的 達成 の ため の 特定 の 言語 構造 や 特性 を 有する テキスト と する （ 例え ば ， Hyland 2007 ， p ． 149 ， Iwasaki & Kumagai 2015 ， p .x 2 ） 。
Romanized Segmented Japanese Text	“ Janru ” ni wa samazama na teigi ga aru ga , koko de wa , aru shakai teki mokuteki tassei no tame no tokutei no gengo kōzō ya tokusei o yūsuru tekisuto to suru (tatoe ba , Hyland 2007 , p . 149 , Iwasaki & Kumagai 2015 , p .x 2) .
English Translation	There are several definitions for *genre*; nevertheless, here the term will be assumed as a text prepared with specific structural and lexico-grammatical features in order to fulfill a given social purpose (e.g., Hyland, 2007, p. 149; Iwasaki & Kumagai, 2015, p. x[2]).

Source: Translated and designed by the author. The quotation comes from an article by Gyogi and Iwasaki (2019, p. 72). The segmentation was performed with SegmentAnt. The full references of the two cited works as presented by Gyogi and Iwasaki (2019) are the following: “Hyland, K. (2007) Genre pedagogy: Language, literacy and L2 writing instruction. *Journal of Second Language Writing*, 16 (3), 148–164”; “Iwasaki, N., & Kumagai, Y. (2015). *The Routledge intermediate to advanced Japanese reader: A genre-based approach to reading as a social practice (Routledge modern language readers)*. London: Routledge” (p. 84).

The software-assisted segmentation makes it possible to analyze Japanese texts through software tools. However, it should be noted that the result may not be completely accurate. In Figure 5.1, for example, I would expect to have 社会的 (*shakaiteki*, “social”) together rather than apart. Nevertheless, the alternative to software-assisted segmentation is manual segmentation, which is impracticable with relatively large amounts of data.

Therefore, despite some imprecisions, software-assisted segmentation was applied to Japanese data in this study.

The complete plain text files were divided into smaller files according to the overall organization of the RAs. To make the division, first a structural description of the data was performed. The description was done through the classification of the articles into six classes based on the degree of similarity with the IMRD pattern (Table 5.3).

Table 5.3 – Classes for structural description.

Class	Description
IMRD Pattern	Articles composed of Introduction, Methods, Results, and Discussion.
First Level Variants	Articles following the four-section IMRD pattern but with slightly different section headings (e.g., Introduction and Aims instead of Introduction; Materials and Methods instead of Methods; Discussion and Conclusions instead of Discussion).
Second Level Variants	Articles following the IMRD pattern with either one or more additional sections (a separate section for conclusions, for example) or a missing section.
Third Level Variants	Articles following the IMRD pattern with both slightly different section headings and the inclusion or exclusion of sections.
Fourth Level Variants	Articles whose overall structure relates to the IMRD pattern in a less evident form.
Other Structures	Articles whose overall structure appears to be unrelated to the IMRD pattern.

Source: Based on Aragão (2015, 2017).

The results of the structural description are shown in Table 5.4.

Table 5.4 – Overall structure of the selected RAs.

Field and Language	Structural Class					
	IMRD Pattern	1st Level Variants	2nd Level Variants	3rd Level Variants	4th Level Variants	Other Structures
Pediatrics						
English (RAs = 72)	0	8	0	64	0	0
Portuguese (RAs = 24)	23	0	1	0	0	0
Japanese (RAs = 24)	0	24	0	0	0	0
Total	23	32	1	64	0	0
Percentage	19.2	26.7	0.8	53.3	0.0	0.0
Management						
English (RAs = 72)	1	0	3	38	23	7
Portuguese (RAs = 24)	0	0	3	12	5	4
Japanese (RAs = 24)	0	0	0	3	9	12
Total	1	0	6	53	37	23
Percentage	0.8	0.0	5.0	44.2	30.8	19.2

Source: Prepared by the author.

As can be seen, whereas RAs from both Pediatrics and Management have a clear link with the IMRD pattern, RAs in Management exhibit a higher degree of variation. Because of this, different criteria were used to divide the articles. Pediatrics texts were divided into separate files according to their original organization, which in most cases means four or five files that correspond to introduction, method, result, discussion, and conclusion sections. Management RAs, by contrast, were rearranged into three files only: one for the opening section (typically, an introduction), one for middle sections (everything between the opening and the closing section), and one for the closing section (generally, an overall discussion or conclusions). During the process, one of the Pediatrics articles in Portuguese was excluded because its results and discussion were combined into a single section.

The plain text files were organized into folders to form different sets or corpora. Corpora can be classified in several ways. Stubbs and Halbe (2013) distinguish between "language corpus," "general reference corpus," and "smaller specialized corpus." According to them, a language corpus typically refers to a large collection of texts designed for linguistic analysis through computers. A general reference corpus designates a collection of texts aimed at providing a balanced and wide sample of a given language. They may be composed of hundreds of millions of words. Smaller specialized corpus can provide samples of specific genres, such as the language of RAs. Another form of classification is that between *parallel* and *comparable* corpora. While a parallel corpus is made up of texts in a given language and their respective translations into another language, a comparable corpus contains “components that are collected using the same sampling method, e.g. the *same proportions* of the texts of the *same genres* in the *same domains* in a range of *different languages* in the *same sampling period*” (McEnery & Hardie, 2012, p. 20).

The plain text files were grouped first according to language into three specialized corpora: enRAs (English), ptRAs (Portuguese), and jaRAs (Japanese). Table 5.5 shows their basic information. These sets can be considered comparable corpora, because they represent the same genre (the RA) in the same domains (Pediatrics and Management) in three languages (English, Portuguese, and Japanese) in the same period (2011–2018), despite the fact that English has been numerically favored.

Table 5.5 – Specialized, comparable corpora.

Corpus		Documents	Tokens	Words	Sentences
enRAs		144	998,238	812,947	31,602
ptRAs		48	258,115	212,163	8,047
jaRAs		48	235,482	194,020	9,432
	Total	240	1,491,835	1,219,130	49,081

Source: Compiled by the author. Corpus numbers calculated by Sketch Engine.

Next, the plain text files were grouped according to language and discipline into six smaller, more specialized corpora: enPED, enMGT, ptPED, ptMGT, jaPED, and jaMGT. Table 5.6 shows their figures. PED stands for Pediatrics; MGT stands for Management.

Table 5.6 – Specialized, comparable, discipline-specific corpora.

Corpus		Documents	Tokens	Words	Sentences
enPED		72	300,299	236,848	9,574
enMGT		72	697,939	576,099	22,028
	English Total	144	998,238	812,947	31,602
ptPED		24	82,519	69,257	2,493
ptMGT		24	175,596	142,906	5,554
	Portuguese Total	48	258,115	212,163	8,047
jaPED		24	65,461	51,917	3,024
jaMGT		24	170,021	142,103	6,408
	Japanese Total	48	235,482	194,020	9,432

Source: Compiled by the author. Corpus numbers calculated by Sketch Engine.

Finally, the plain text files of the different parts of the RAs were grouped into 21 small, highly specialized subcorpora, reflecting the overall organization of the RAs: enPED(Intro), enPED(Me), enPED(Res), enPED(D+C); ptPED(Intro), ptPED(Me), ptPED(Res), ptPED(D+C); jaPED(Intro), jaPED(Me), jaPED(Re), jaPED(D+C); enMGT(Op), enMGT(Mi), enMGT(Cl); ptMGT(Op), ptMGT(Mi), ptMGT(Cl); jaMGT(Op), jaMGT(Mi), and jaMGT(Cl).[1] Table 5.7 displays the basic information on them. It should be noted that, because they do not include articles' titles, abstracts, and notes, and because one Portuguese RA was excluded due to structural incompatibility, the figures in Table 5.7 are lower than those of the previous tables.

[1] Intro, Me, Res, and D+C are abbreviations for Introduction, Methods, Results, and Discussion and Conclusion, respectively. Op, Mi, and Cl are abbreviations for opening (section), middle (sections), and closing (section), respectively.

Table 5.7 – Highly specialized, comparable subcorpora.

Subcorpus	Files	Tokens	Words	Sentences
enPED(Intro)	72	38,560	30,986	1,074
enPED(Me)	72	75,508	60,898	2,635
enPED(Res)	72	69,235	47,925	2,066
enPED(D+C)	72	87,877	74,903	2,800
English Pediatrics Total	288	271,180	214,712	8,575
ptPED(Intro)	23	11,081	9,808	243
ptPED(Me)	23	20,403	16,928	637
ptPED(Res)	23	12,682	9,250	409
ptPED(D+C)	23	27,759	24,557	734
Portuguese Pediatrics Total	92	71,925	60,543	2,023
jaPed(Intro)	24	3,862	3,241	168
jaPED(Me)	24	6,464	5,130	351
jaPED(Res)	24	16,041	11,289	976
jaPED(D+C)	24	31,740	26,693	1,112
Japanese Pediatrics Total	96	58,107	46,353	2,607
enMGT(Op)	72	75,780	60,566	2,220
enMGT(Mi)	72	527,642	434,134	17,014
enMGT(Cl)	72	64,104	55,715	2,080
English Management Total	216	667,526	550,415	21,314
ptMGT(Op)	24	22,297	17,993	633
ptMGT(Mi)	24	132,921	107,122	4,229
ptMGT(Cl)	24	18,466	16,010	596
Portuguese Management Total	72	173,684	141,125	5,458
jaMGT(Op)	24	14,985	12,873	525
jaMGT(Mi)	24	121,203	101,141	4,629
jaMGT(Cl)	24	15,859	14,062	568
Japanese Management Total	72	152,047	128,076	5,722

Source: Compiled by the author. Subcorpus numbers calculated by Sketch Engine.

Furthermore, it should be noted that due to copyright restrictions the specialized corpora and subcorpora will not be made publicly available for other researchers. Even though most of the original data have been published under Creative Commons licenses, thus allowing wide availability (see Creative Commons, n.d.), journals from Japan and the United States have strict restrictions for use that prevent sharing the data.

In addition to the specialized corpora and subcorpora, four general reference corpora available in Sketch Engine, an already cited electronic platform developed by Lexical Computing CZ (n.d.) for linguistic research, were used in this study: English Web 2015 (enTenTen15), English Web 2020 (enTenTen20), Portuguese Web 2011 (ptTenTen11), and Japanese Web 2011 Sample (jaTenTen11 Long Unit Words [LUW]). Table 5.8 displays their basic information.

Table 5.8 – General reference corpora.

Corpus	Documents	Tokens	Words	Sentences
enTenTen15	33,655,541	15,411,682,875	13,190,556,334	688,989,861
enTenTen20	81,323,314	44,968,996,152	38,149,437,411	2,099,033,556
ptTenTen11	10,216,055	4,622,750,491	3,896,392,719	197,944,143
jaTenTen11 LUW	269,056	203,674,569	163,837,671	10,047,893

Source: Compiled by the author. Corpus numbers provided by Sketch Engine (Lexical Computing CZ, n.d.).

The four reference corpora belong to the TenTen Corpus Family, which is composed of corpora built from Internet texts with the same criteria in more than 40 languages (Lexical Computing CZ, n.d.). *TenTen* is because the members of the family are in the order of 10 billion words, that is, 10^{10} (ten to the tenth or *ten ten*); every corpus name is "formed by prefixing with the two-letter ISO-639-1 code for the language, and, optionally, suffixing with two-digits for the year of collection, to give e.g. enTenTen12 for English collected in 2012, zhTenTen for Chinese" (Jakubíček, Kilgarriff, Kovář, Rychlý, & Suchomel, 2013, p. 125).

The building process of the TenTen corpora depends on several software tools, being currently described by Lexical Computing CZ (n.d.) in the following manner:

> 1. Texts are crawled from the Internet by *Spiderling* tool, a web spider designed for linguistic purposes.
> 2. Texts are cleaned by *jusText* which removes undesirable content such as navigation links, advertisements, headers, footers, etc.
> 3. A tokenization process when texts are separated into individual positions (tokens).
> 4. *Language Filter* is used for language identification to detect and remove longer texts of different languages, but foreign words or phrases are kept (e.g. sentences with movie titles).
> 5. The *onion* tool performs deduplication [exclusion of duplicated content] on the paragraph level.
> 6. The sample texts of the biggest web domains which account for 55%–95% of all corpus texts are checked (combination of manual techniques with our standard automatic methods) and content with poor quality text and spam are removed.
> 7. Corpora are recompiled with removing poor quality texts.
> 8. The largest web domains are classified into genres (referring to writing styles) and topic (inspired by categories used by https://curlie.org/).
> 9. Then corpus texts are lemmatized and part-of-speech tagged for language for which there are tagger and lemmatizer tools are available.

10. Final checking of corpora in the interface.
11. Publishing corpora.

Even though the description above includes the classification of Internet domains into genres (eighth step), this classification is not available for any of the four reference corpora. The top-level Internet domains used to collect language data and build the corpora, however, are provided by Lexical Computing CZ (n.d.). Table 5.9 shows these domains.

Table 5.9 – Top-level domains used to build the general reference corpora.

Corpus	Internet Domains and Percentage of Documents
enTenTen15	.org (38.6%), .net (13.0%), .edu (6.1%), .uk (5.4%)
enTenTen20	.com (43.1%), .org (24.5%), .uk (10.1%), .net (6.2%), .edu (4.9%), .au (3.3%)
ptTenTen11	.br (76.0%), .pt (23.9%)
jaTenTen11 LUW	.com (49.7%), .jp (34.1%), .net (8.9%), .info (3.3%)

Source: Compiled by the author based on corpus descriptions provided by Sketch Engine (Lexical Computing CZ, n.d.).

As can be noted, the English corpora are derived from several Internet domains, of which two (.uk and .au) relate specifically to English-speaking countries—the United Kingdom and Australia. The Portuguese corpus is derived from the Brazilian and Portuguese top-level domains (.br and .pt), with a preponderance of the former; and the Japanese corpus is mainly derived from the well-known commercial .com and the Japanese .jp.

Lexical Computing CZ (n.d.) also provides the specific domains from which the language data were collected. For illustrative purposes, Table 5.10 presents the 10 domains with the highest numbers of extracted documents for each of the four corpora.

Table 5.10 – Domains with the highest numbers of extracted documents.

Corpus	Domain Names and Numbers of Documents
enTenTen15	(1) newadvent.org (10,097), (2) ufanyc.org (9,944), (3) bibliovault.org (9,044), (4) news.stv.tv (8,968), (5) buddhistchannel.tv (8,578), (6) google.as (8,274), (7) outlookafghanistan.net (8,134), (8) avhandlingar.se (8,038), (9) familysecuritymatters.org (7,956), (10) news.kuwaittimes.net (7,914)
enTenTen20	(1) en.wikipedia.org (5,324,576), (2) patents.justia.com (45,316), (3) sharewareconnection.com (44,900), (4) dearcupid.org (44,730), (5) thailand4.com (44,622), (5) theinfolist.com (40,397), (6) advicenators.com (40,183), (7) freepr101.com (38,232), (8) brightsurf.com (36,094), (10) m.reddit.com (33,739)
ptTenTen11	(1) www1.folha.uol.com.br (39,018), (2) netvasco.com.br (31,374), (3) reporternews.com.br (29,821), (4) dicio.com.br (28,903), (5) jornaldaciencia.org.br (27,665), (6) piaui.pi.gov.br (24,504), (7) tek.sapo.pt (23,442), (8) samba-choro.com.br (23,242), (9) pciconcursos.com.br (22,623), (10) artigos.netsaber.com.br (21,782)
jaTenTen11 LUW	(1) fc2.com (61,126), (2) jugem.jp (11,901), (3) so-net.ne.jp (4,246), (4) blogspot.com (3,180), (5) exblog.jp (2,827), (6) sakura.ne.jp (2,411), (7) webry.info (2,391), (8) cocolog-nifty.com (2,082), (9) ocn.ne.jp (1,576), (10) iza.ne.jp (1,114)

Source: Compiled by the author based on corpus descriptions provided by Sketch Engine (Lexical Computing CZ, n.d.).

Although the domains above account for only a small portion of the total of documents forming the reference corpora (see Table 5.8), they reveal some characteristics of these data sets. First, they show that both enTenTen15 and enTenTen20 include English texts from (about) several countries, such as Afghanistan (outlookafghanistan.net), Kuwait (news.kuwaittimes.net), and Thailand (thailand4.com). Second, they indicate that the Portuguese corpus is composed of various kinds of text, ranging from news (www1.folha.uol.com.br) to dictionary entries (dicio.com.br) and public job announcements (pciconcursos.com.br). Third, they show that the Japanese corpus contains a high amount of blog posts, as many of its domains (fc2.com, jugem.jp, blogspot.com, exblog.jp, cocolog-nifty.com) publish blogs. Collectively, the specific domains listed in Table 5.10 indicate that the four TenTen corpora are composed of open-access texts belonging to many different genres.

It is important to note that Sketch Engine provides access to many other general corpora in English, Portuguese, and Japanese, but most of them do not have the necessary specifications for some analytical tools. It provides access, for example, to Corpus Brasileiro (Berber Sardinha, Moreira Filho, & Alambert, 2008–2010), a corpus of Brazilian Portuguese composed of academic texts, news, literary essays, and political speeches, among other types of data, which would be an interesting reference corpus for Portuguese RA data. However, keyword analysis, a process that, as will be seen in the next chapter, is of particular interest here, could not be effectively performed with it due to differences in data processing. In fact, the frequencies of the keywords identified by the comparison between ptRAs and Corpus Brasileiro did not match those obtained through other tools, showing that the keyword analysis was unsuccessful. Therefore, despite the availability of additional corpora in the platform, they were not used for comparison.

Another specialized Portuguese corpus, however, was employed occasionally to obtain additional evidence: the Corpus of Portuguese from Academic Journals (CoPEP; Kuhn & Ferreira, 2018). Unlike other specialized corpora available in Sketch Engine, the data processing type of this corpus allows keyword extraction, making it suitable to look for supplementary information. Table 5.11 shows its composition.

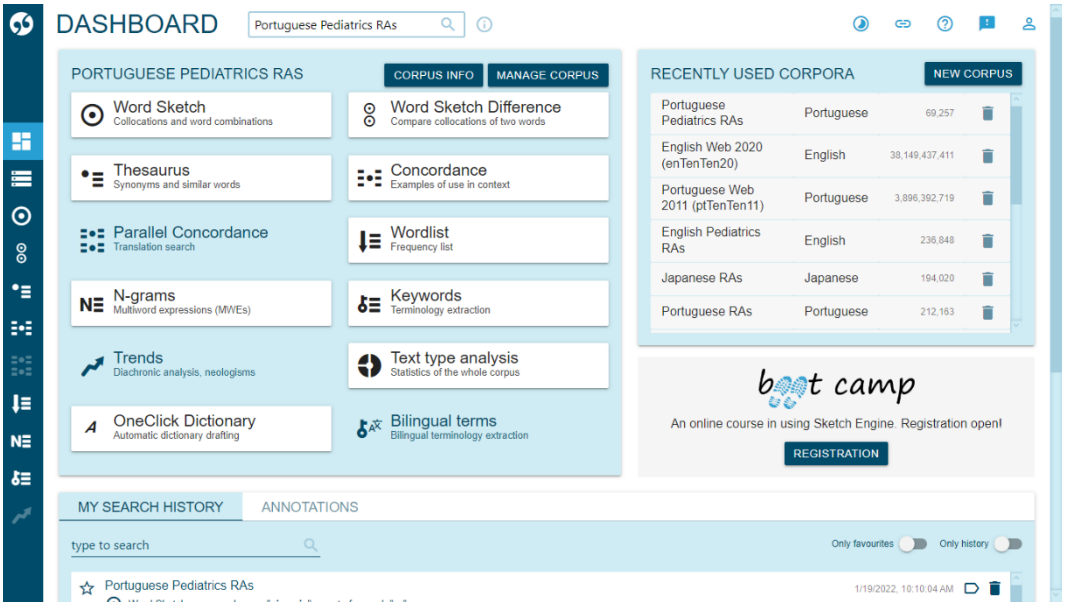
Table 5.11 – Corpus of Portuguese from Academic Journals (CoPEP).

Documents	Tokens	Words	Sentences
9,900	48,826,547	40,423,011	1,395,373

Source: Numbers provided by Sketch Engine (Lexical Computing CZ, n.d.). CoPEP was compiled by Kuhn and Ferreira (2018).

To analyze the corpora and subcorpora, two software tools were used. The first is the software tool developed by Lexical Computing CZ (n.d.), Sketch Engine, which is part of the platform with the same name. Sketch Engine analyzes corpus data stored in plain text files. The software user can upload data to its online system, creating his or her own corpora, or investigate the already available corpora, such as the ones adopted here as reference corpora. For illustrative purposes, Figure 5.2 shows the dashboard of Sketch Engine for the Portuguese Pediatrics RA corpus (ptPED). The dashboard gives access to several functions.

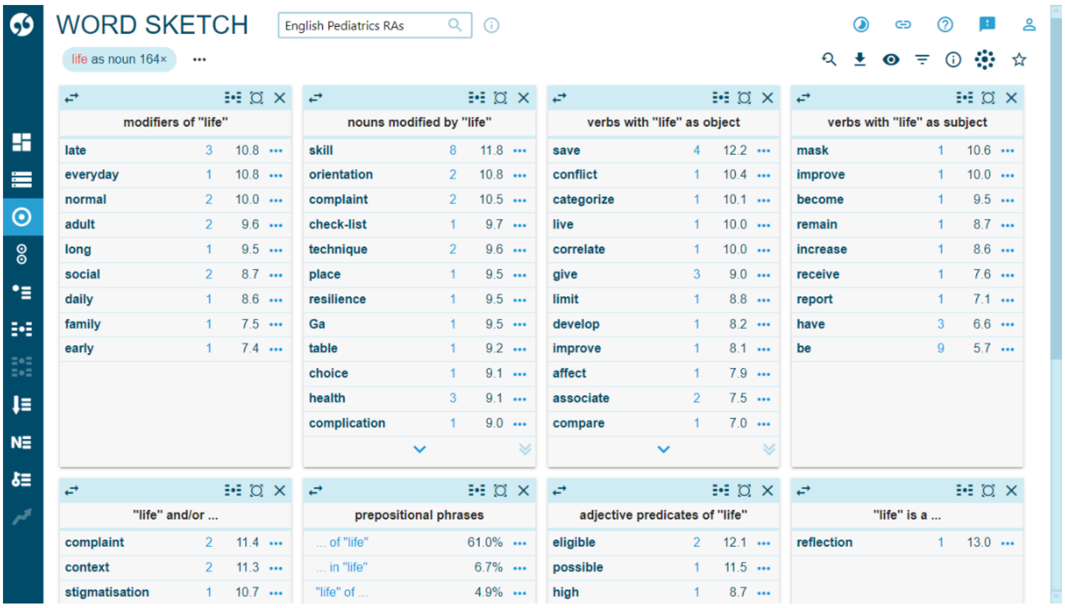
Figure 5.2 – Sketch Engine dashboard for ptPED (Portuguese Pediatrics RA corpus).



Source: Screenshot of Sketch Engine (Lexical Computing CZ, n.d.). Reproduced with permission.

One of the main functions of Sketch Engine is Word Sketch. Word Sketch generates lists of both frequent and strong collocates according to position and grammatical function. Figure 5.3 displays part of the results of the Word Sketch for the lemma LIFE in enPED.

Figure 5.3 – Word Sketch for the lemma LIFE in enPED (English Pediatrics RA corpus).

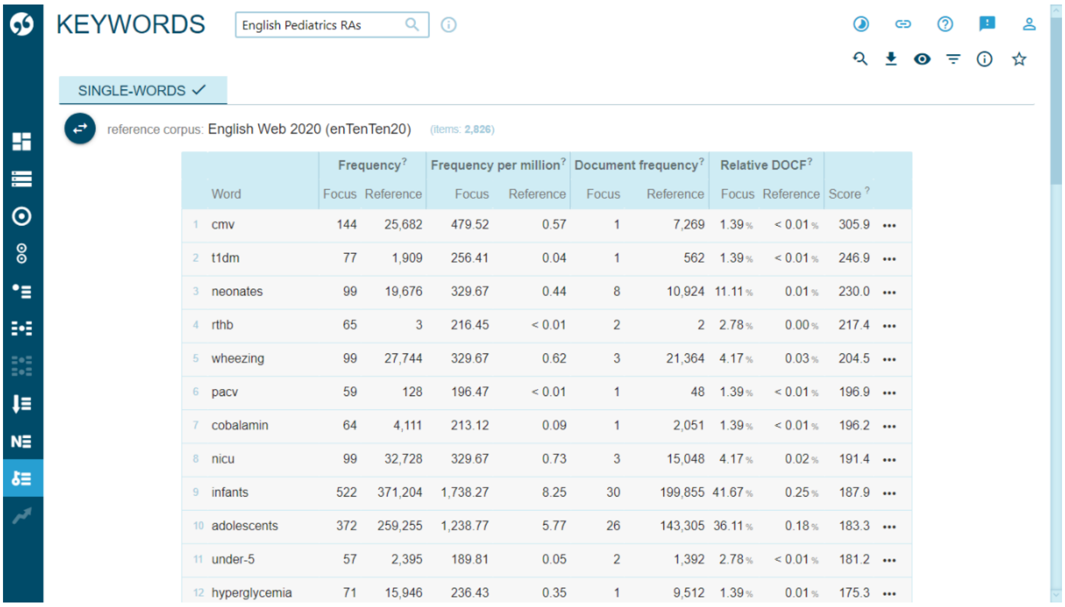


Source: Screenshot of Sketch Engine (Lexical Computing CZ, n.d.). Reproduced with permission.

According to these results, *life* and *lives* are often preceded by *late*, *everyday*, and *normal*, and are often followed by *skill*, *orientation*, and *complaint*. Also, *life* and *lives* are often the object of the verb *save*, as in *doctors save lives*. Collocates are listed according to association strength and frequency. *Late*, for example, has an association strength score of 10.8 and a frequency of 3 in the corpus. The score was calculated by logDice, a statistical measure based on the Dice coefficient (Lexical Computing Ltd., 2015).

Another important function of Sketch Engine is keyword analysis. Keyword analysis generates lists of typical words found in one set of data in relation to another (keywords). This function will be detailed in the next chapter; for now, Figure 5.4 illustrates its use with the results of the keyword analysis between enPED (focus or study corpus) and enTenTen 20 (reference corpus).

Figure 5.4 – Results of the keyword analysis function for enPED (English Pediatrics RA corpus) in relation to enTenTen20.

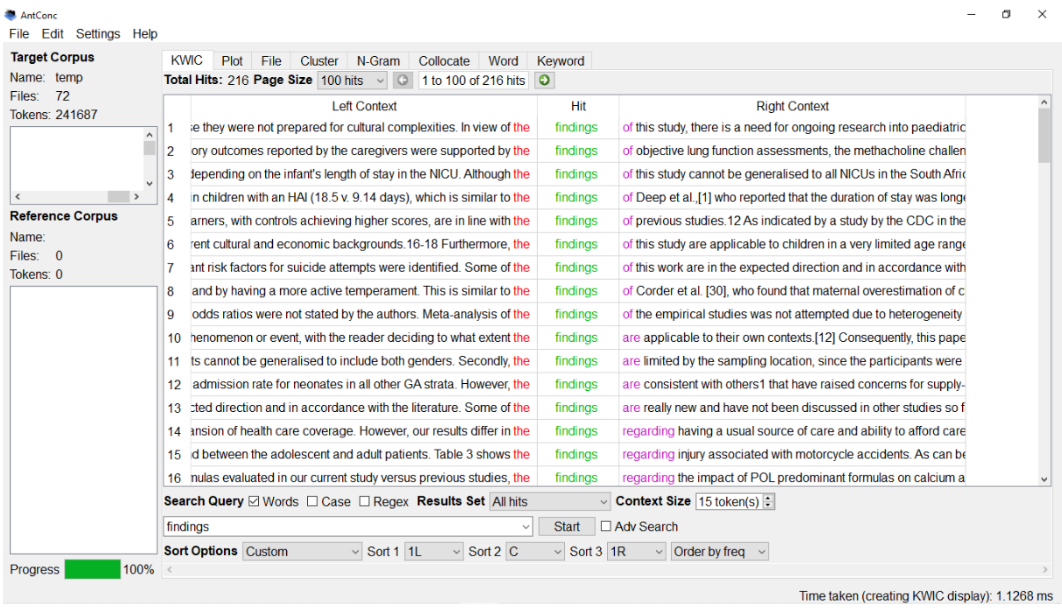


	Word	Frequency Focus	Frequency Reference	Frequency per million Focus	Frequency per million Reference	Document frequency Focus	Document frequency Reference	Relative DOCF Focus	Relative DOCF Reference	Score
1	cmv	144	25,682	479.52	0.57	1	7,269	1.39 %	< 0.01 %	305.9
2	t1dm	77	1,909	256.41	0.04	1	562	1.39 %	< 0.01 %	246.9
3	neonates	99	19,676	329.67	0.44	8	10,924	11.11 %	0.01 %	230.0
4	rthb	65	3	216.45	< 0.01	2	2	2.78 %	0.00 %	217.4
5	wheezing	99	27,744	329.67	0.62	3	21,364	4.17 %	0.03 %	204.5
6	pacv	59	128	196.47	< 0.01	1	48	1.39 %	< 0.01 %	196.9
7	cobalamin	64	4,111	213.12	0.09	1	2,051	1.39 %	< 0.01 %	196.2
8	nicu	99	32,728	329.67	0.73	3	15,048	4.17 %	0.02 %	191.4
9	infants	522	371,204	1,738.27	8.25	30	199,855	41.67 %	0.25 %	187.9
10	adolescents	372	259,255	1,238.77	5.77	26	143,305	36.11 %	0.18 %	183.3
11	under-5	57	2,395	189.81	0.05	2	1,392	2.78 %	< 0.01 %	181.2
12	hyperglycemia	71	15,946	236.43	0.35	1	9,512	1.39 %	0.01 %	175.3

Source: Screenshot of Sketch Engine (Lexical Computing CZ, n.d.). Reproduced with permission.

The second software tool is AntConc (version 4.0.3), which was developed by Anthony (2022). AntConc is a tool designed for corpus analysis in individual computers, that is, there is no need to upload data to an online system. The software user must prepare plain text files with UTF-8 encoding, so that AntConc can process the data. AntConc includes conventional corpus search functions, such as concordance lines, word lists, and collocate analysis. Its current version, moreover, allows users to create and manage corpora in order to carry out comparative analyses. For the purpose of illustration, Figure 5.5 shows concordance lines generated by AntConc for the word *findings* using enPED. In this case, both left and right collocates are highlighted; and their appearance is in decreasing order of frequency. It is possible to search for words with or without distinction between lower and upper case, as well as with the regex option, which widens the search to include every instance of the target combination of characters—for example, a search for *low* yields concordance lines with *low*, *lower*, *lowest*, *follow*, *following*, etc.

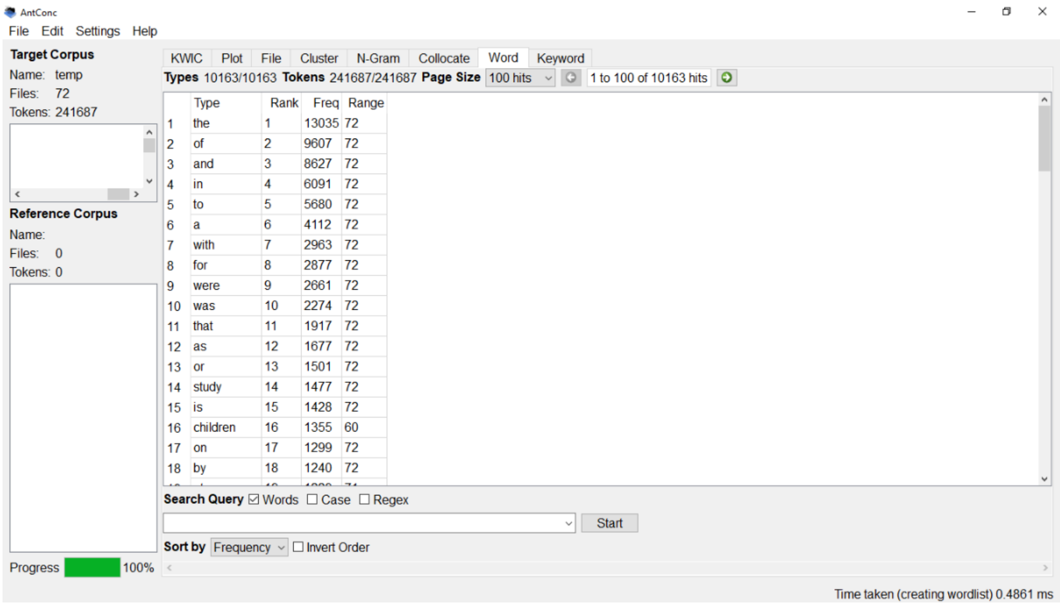
Figure 5.5 – Concordance lines for *findings* in enPED.



Source: Screenshot of AntConc (Anthony, 2022). Reproduced with permission.

Still for illustrative purposes, Figure 5.6 exhibits the word list function of AntConc using enPED. It shows the most frequent words according to frequency and range in terms of files. Because this corpus is composed of 72 files corresponding to 72 separate RAs, the range in the list can be read as the number of RAs that contain each word.

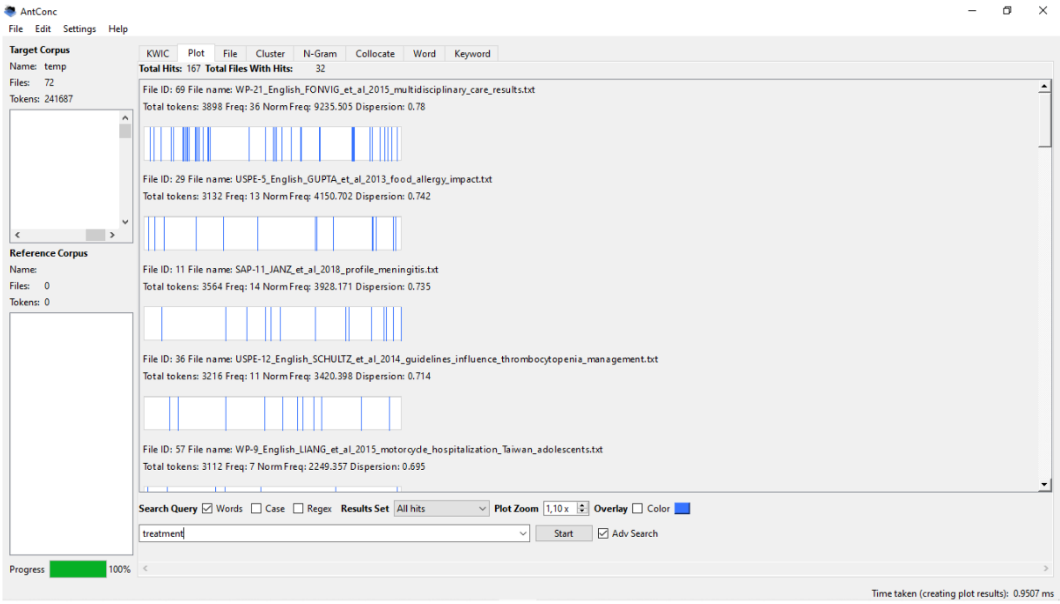
Figure 5.6 – List of the most frequent words in enPED.



Source: Screenshot of AntConc (Anthony, 2022). Reproduced with permission.

Both Sketch Engine and AntConc have been used in lexical priming research, either separately (e.g., Chuang, 2015; Cantos & Almela, 2017; Shao, 2017) or together (e.g., Wang, 2018). The major advantage of Sketch Engine over AntConc is the access its platform provides to a wide variety of big corpora in several languages, which is a valuable resource for comparative analyses. Yet, AntConc has a straightforward interface, provides accurate results, and operates offline, in addition to currently being freeware—Sketch Engine requires a paid subscription. Also, AntConc includes a useful function for analyzing textual colligations called Plot Tool, which shows the textual position and dispersion of the occurrences of a given target word. Figure 5.7 presents some graphs generated by the Plot Tool for the word *treatment* in enPED.

Figure 5.7 – Plot graphs and dispersion values for the word *treatment* in enPED.



Source: Screenshot of AntConc (Anthony, 2022). Reproduced with permission.

As will be seen in the following chapters, Sketch Engine was used in this study for cross-genre and cross-disciplinary analyses. AntConc was used to investigate textual colligational priming in RAs. Both software programs were used for cross-linguistic analysis.

6 RESEARCH PROCEDURES

As explained in Chapter 3, there are four aspects related to lexical priming theory and research of interest in the present study: genre-specificity, disciplinary variation, textual position, and cross-linguistic variation. For each aspect, a different research question and a different aim were formulated. This chapter will describe the research procedures employed in the study departing from our aims (Figure 6.1).

Figure 6.1 – Research aims.

First Aim GENRE-SPECIFICITY	To assess the presence of genre-specific signs of psychological priming in RAs in English, Portuguese, and Japanese.
Second Aim DISCIPLINARY VARIATION	To evaluate the presence of discipline-specific signs of psychological priming in RAs in English, Portuguese, and Japanese.
Third Aim TEXT POSITION	To evaluate the presence of position-specific signs of psychological priming in RAs in English, Portuguese, and Japanese.
Fourth Aim CROSS-LINGUISTIC VARIATION	To describe similarities and differences between primings of equivalent words in English, Portuguese, and Japanese.

Source: Designed by the author.

Four successive research stages were completed, each of which was dedicated to a different aim. The first stage consisted in comparing the specialized, single-genre RA corpora and the non-specialized, multi-genre reference corpora, thereby addressing genre-specificity. This stage was entirely performed with Sketch Engine since its electronic platform provides access to all the reference corpora. The stage was composed of three sub-stages: a keyword analysis to find genre-specific keywords, a collocational analysis of selected genre-specific keywords to identify possible differences in collocations between the corpora, and a semantic associational analysis concerning the same keywords to explore differences in terms of semantic associations.

Keywords are generally defined as words that appear more frequently in a given data set or corpus in comparison to another data set or reference corpus, with the difference in frequency between the first and the second being statistically significant (Baker, 2006, p. 22; Berber Sardinha, 2009, p. 193; McEnery & Hardie, 2012, p. 245). Keyword analysis, in turn, is the process by which keywords are identified; it is often adopted in lexical priming research as an initial step to find out proper words for further investigation (e.g., Salim, 2012; Jantunen, 2017; Pace-Sigge, 2017). Keyword analysis was conducted in the first stage with Sketch Engine following the criteria shown in Table 6.1.

Table 6.1 – Criteria for keyword analysis in the first stage.

Criterion	Description	Set-Up
Search Item Category	Options include “word,” “lemma,” “tag,” and “parts of speech”	Word
Focus	Ranges from 0.001 to 1,000,000 (0.001 restricts the analysis to extremely typical [rare] words in the study corpus; 1,000,000 expands it to extremely non-typical [common] words)	1.0
Minimum Frequency	Defines the minimum frequency a word must have in the study corpus to be considered a keyword	10
Maximum Frequency	Defines the maximum frequency a word may have in the study corpus to be considered a keyword	Inactive (zero)
Maximum Items	Defines the maximum number of keywords for the analysis	100
A = a	Defines whether the analysis will distinguish between lower and upper case (inactive) or not (active)	Active
At Least One Alphanumeric	Restricts the analysis to words made up of at least one letter or number	Active
Only Alphanumeric	Restricts the analysis to words made up of letters, numbers, and hyphens	Active
Include Nonwords	Includes numbers, punctuation marks, and tokens that do not start with letters (e.g., *10-year*)	Inactive

Source: Prepared by the author considering the criteria available in Sketch Engine (Lexical Computing CZ, n.d.).

The analysis was performed with “word” as the search item category, in accordance with Hoey’s (2005) premise that primings refer to words, not to lemmas. Focus was set in 1.0, because this parameter offers a good balance between relevance and range. Lower numbers restrict keywords to items found in very few RAs; higher numbers, by contrast, increase the range excessively, so that relatively common items appear. Minimum frequency was set in 10 to avoid low-frequency items. Maximum frequency was not considered pertinent, since high-frequency words such as *the* and *and* are unlikely to be listed as keywords with focus set in 1.0 (i.e., to restrict results to relatively rare items). Maximum items were limited to 100 because words after the 100th keyness position are less likely to be relevant. The remaining criteria were chosen to include words but exclude isolated numbers and punctuation marks.

Despite the criteria adopted, the results obtained included many items whose frequency was concentrated in but a few RAs. Irrespective of the degree of keyness, it was important that the words found were minimally distributed among the selected RAs—they should be *key keywords*, that is, words playing the role of keywords in a certain number of files (Berber Sardinha, 2009, p. 195). Sketch Engine includes a measure to address this aspect called *average reduced frequency*, “a modified frequency whose calculation prevents the results from being excessively influenced by a high concentration of a token in only one or small parts of the corpus” (Lexical Computing CZ, n.d.). As the files composing the English, the Portuguese, and the Japanese RA corpora correspond to separate articles, however, a minimum percentage of RAs containing the keywords seemed suitable to distinguish key keywords from non-key keywords. While the average reduced frequency is affected by different instances of the same word in the same file (as long as they are far from each other), the percentage of RAs is not. Specifically, the percentage of 10% was adopted in the first stage to distinguish key keywords from non-key keywords, that is, for being considered a key keyword the item had to appear in at least 15 out of the 144 English RAs, five out of the 48 Portuguese RAs, and five out of the 48 Japanese RAs. This percentage was chosen because it allowed to extract a manageable number of items from every set of 100 keywords.

At this point, it is necessary to explain how Sketch Engine calculates keyness scores, which form the basis for keyword analysis. The software uses the "simple maths" method (Kilgarriff, 2009; Lexical Computing Ltd., 2015), which is represented by the following formula:

$$\frac{fpm_{focus} + N}{fpm_{ref} + N}$$

The components of the formula are as follows: *fpm* stands for frequency per million of a given word; *focus* refers to the focus or study corpus; *ref* refers to the reference corpus; and *N* is the simple maths parameter, which appears in Table 6.1 under the label "focus" and is defined by the software (platform) user. The formula can be summarized as follows: With an *N* value of 1.0 or less, the higher the frequency per million of a given word in a given corpus and the lower its frequency per million in a separate reference corpus, the higher the keyness score of the word will be.

As can be noted, Sketch Engine's calculation does not involve statistical significance. It is based on the comparison of normalized word frequencies instead. In this sense, keywords found through Sketch Engine do not fully conform to the definition presented earlier. Nevertheless, they do agree with alternative views such as the following:

> Very early on, keywords were defined as 'words whose frequency is unusually high in comparison with some norm' (Scott 1996: 53). It is straightforward to derive from this definition that a keyword is identified by way of a frequency comparison. It should clearly follow, then, that an appropriate metric for keyness would reflect the size of the frequency difference, and that the larger the difference, the more 'key' a word would be. (Gabrielatos, 2018, p. 229)[1]

To illustrate how keyness scores change according to *N* values, Table 6.2 displays the application of the formula to calculate scores for three keywords found in enRAs against enTenTen15. The scores were calculated with *N* values from 0.01 to 100 using the frequency per million of the three words in both corpora.

[1] The full reference of Scott's manual as presented by Gabrielatos (2018) is the following: "Scott, M. 1996. *WordSmith Tools Manual*. Oxford: Oxford University Press." (p. 258)

Table 6.2 – Keyness score calculation with different *N* values.

Word	Frequency per Million		N	Keyness	
	FC	RC		Calculation	Score
sensemaking	159.28	0.11	100	(159.28 + 100)/(0.11 + 100) = 259.28/100.11 =	2.6
			10	(159.28 + 10)/(0.11 + 10) = 169.28/10.11 =	16.7
			1	(159.28 + 1)/(0.11 + 1) = 160.28/1.11 =	144.4
			0.1	(159.28 + 0.1)/(0.11 + 0.1) = 159.38/0.21 =	759.0
			0.01	(159.28 + 0.01)/(0.11 + 0.01) = 159.29/0.12 =	1,327.4
coopetition	145.26	0.05	100	(145.26 + 100)/(0.05 + 100) = 245.26/100.05 =	2.5
			10	(145.26 + 10)/(0.05 + 10) = 155.26/10.05 =	15,4
			1	(145.26 + 1)/(0.05 + 1) = 146.26/1.05 =	139,3
			0.1	(145.26 + 0.1)/(0.05 + 0.1) = 145.36/0.15 =	969.1
			0.01	(145.26 + 0.01)/(0.05 + 0.01) = 145.27/0.06 =	2,421.2
OCB	133.23	0.09	100	(133.23 + 100)/(0.09 + 100) = 233.23/100.09 =	2.3
			10	(133.23 + 10)/(0.09 + 10) = 143.23/10.09 =	14.2
			1	(133.23 + 1)/(0.09 + 1) = 134.23/1.09 =	123.1
			0.1	(133.23 + 0.1)/(0.09 + 0.1) = 133.33/0.19 =	701.7
			0.01	(133.23 + 0.01)/(0.09 + 0.01) = 133.24/0.10 =	1,332.4

Source: Prepared by the author. Keyword scores calculated through the simple maths formula (Lexical Computing Ltd., 2015). FC = focus corpus (enRAs). RC = reference corpus (enTenTen15). OCB = organizational citizenship behavior.

As can be observed, lower *N* values increase keyness scores and higher *N* values make them smaller, replacing typicality with commonness. As shown by Table 6.1, Sketch Engine's keyword analysis can be used to search not only typical, rare items but also shared, common words (which could be called *false keywords*). In addition, it can be seen that when *N* changes, the keyness position may also change. For example, with an *N* value below 1, *sensemaking* in Table 6.2 loses the top position.

It should be noted that there are other approaches to estimate keyness. While the simple maths method focuses on the comparison of normalized frequencies, thereby being part of the "effect-size metric" group (Gabrielatos, 2018), keyness can also be estimated by statistical significance tests, such as chi-squared and log-likelihood, which converge with the notion of keyword found in Baker (2006), Berber Sardinha (2009), and McEnery and Hardie (2012). In fact, chi-squared and log-likelihood have been widely used in keyword analyses (Pojanapunya & Todd, 2018), being normally found in corpus analysis software programs (e.g., AntConc). Chi-squared statistics are based on the "assumption that simple functions of the random variables being sampled are distributed normally or approximately normally" (Dunning, 1993, p. 62). Also, they demand big volumes of data (Dunning, 1993). Log-likelihood statistics, by contrast, are based on binomial or multinomial distribution and are suitable for investigating either small or big volumes of data (Dunning, 1993).

The keyword analysis of the first stage consisted of a series of five analyses in the following order: (1) enRAs against enTenTen15, (2) enRAs against enTenTen20, (3) ptRAs against ptTenTen11, (4) CoPEP (Kuhn & Ferreira, 2018) against ptTenTen11, and (5) jaRAs against jaTenTen 11 LUW. All the five analyses addressed genre-specificity by revealing keywords characteristic of the RA genre. In addition, the first, the third, and the fifth analyses were used to identify keywords for the subsequent collocational and semantic associational study. For this purpose, three keywords were selected for each of the three languages (nine words in total) based mainly on the amount of Pediatrics and Management RAs containing each keyword and on keyness scores. Keywords appearing in RAs from both disciplines (to avoid discipline-exclusive items) with the highest scores would be preferred. As for the Japanese keywords, however, the number of instances in jaTenTen11 LUW had also to be considered, as many keywords appear only a few times in this reference corpus.

The study of collocations was based on immediately preceding (left) and following (right) collocates with a minimum frequency in the corpus of 2. Collocates were searched with Sketch Engine's concordance function (query type: "word"; part of speech: "any" or "noun," depending on the word; with no case distinction), which provides detailed collocate lists ordered by statistical measures through its collocation option. Although Sketch Engine could find English and Portuguese collocates properly, it was not effective in listing Japanese collocates. Due to its processing characteristics, too many Japanese word fragments appeared in the results. In view of this, the Japanese collocates had to be either manually identified from concordance lines or manually filtered from collocate lists. Moreover, because of the particularities of the Japanese data, a minimum frequency of 2 or 4 in either jaRAs or jaTenTen11 LUW was adopted as the cut-off for inclusion. In the case of English and Portuguese, the study of collocations considered the top strongest collocates; in the case of Japanese, it considered the most frequent items. The strength of collocations was measured by logDice (Rychlý, 2008), a statistical measure that ranges from 0 (zero) to 14—extremely weak to exceptionally strong relationship between collocate and node. In addition, the proportion of co-occurrences between collocate and keyword in relation to the frequency of the keyword in the corpus was calculated as an additional measure for comparison.

At this point, it is useful to briefly review some measures related to collocations. Collocations can be examined by simple word counting, that is, by counting "the number of times a given word appears within, say a 5-word window to the left or right of a search term" (Baker, 2006, p. 100). The study of Japanese collocations in the first stage was done in this way but with a shorter range (-1 to +1). Collocations can also be examined by significance measures. Hunston (2002) draws attention to two of these measures, namely Mutual Information (MI) score and t-score. According to her, both calculate the number of instances of a given collocate in the designated span of the node (the Observed) and the number of instances that would be expected in that span based on the frequency of the collocate in the entire corpus (the Expected). However, while the MI-score is the result of the division of the Observed by the Expected converted to a base-2 logarithm, the t-score is the result of the subtraction of the Expected from the Observed divided by the standard deviation. Still according to Hunston (2002), the MI-score estimates the strength of a given collocation and does not depend on the corpus size; the t-score estimates the certainty of a given collocation and is dependent on the corpus size. Log-likelihood (Dunning, 1993) can also be used to calculate the strength (statistical significance) of the relationship between words and their collocates. With this respect, Evert (2005) evaluates log-likelihood against another test (Fisher's test), claiming that

> Fisher's exact test is now widely accepted in mathematical statistics as the most appropriate quantitative measurement of this significance. The log-likelihood association measure gives an excellent approximation to the *p*-values of Fisher's test and has convenient mathematical and numerical properties.

Consequently, it has recently become a *de facto* standard in the field of computational linguistics for the purpose of measuring the statistical association between words or similar entities. (p. 137)

Fisher's test was "popularised by Pedersen (1996) as an alternative to the log-likelihood measure . . . that does not have to rely on approximations" (Evert, 2005, p. 80).[2] Rychlý (2008) introduces logDice as a measure that can be consistently applied to corpora of different sizes, thus facilitating cross-corpus comparisons. LogDice is based on the Dice coefficient, which is obtained by the following formula:

$$\frac{2f_{xy}}{f_x + f_y}$$

The components of the formula are as follows: f_x refers to the number of occurrences of a given word represented by *X*; f_y refers to the number of occurrences of another word represented by *Y*; and f_{xy} is the number of co-occurrences between the two words (Rychlý, 2008; Kolesnikova, 2016).

LogDice, in turn, is the product of the sum of the Dice coefficient multiplied by a 2-base logarithm and 14:

$$14 + log_2 \frac{2f_{xy}}{f_x + f_y}$$

Rychlý (2008) explains that logDice values should be interpreted in the following way. First, the maximum of 14 is purely theoretical, as it could be obtained only when every instance of the node co-occurs with the collocate and every instance of the collocate co-occurs with the node—in general, logDice values are expected to be under 10. Second, a value of 0 (zero) shows that there is less than one co-occurrence between the node and the collocate for each 16,000 instances of either the node or the collocate. Third, negative values indicate the absence of statistical significance for the co-occurrence of

[2] The full reference of Pedersen's work as presented by Evert (2005) is the following: "Pedersen, Ted (1996). Fishing for exactness. In *Proceedings of the South-Central SAS Users Group Conference*, Austin, TX." (p. 349)

words. Lastly, for comparisons between corpora a difference of 1 point indicates double the frequency of co-occurrences (±100%) and a difference of 7 points indicates approximately one hundred times more co-occurrences (±10,000%).

Although Sketch Engine's collocation interface includes other statistics such as MI-score, t-score, and log-likelihood, logDice was chosen because of its suitability for investigating our corpora. LogDice is acknowledged as "one of the most common association measures used to detect collocations," and "its performance happens to be higher than the performance of other association measures" (Kolesnikova, 2016, p. 340). Moreover, it is noteworthy that other researchers have already used logDice as an association measure for collocational analysis, obtaining positive results (see Berber Sardinha, 2017; Cantos & Almela, 2017; Wang, 2018, 2022).

The study of semantic associations was based on semantic sets. Although there are databases that could be used as starting points or reference material to build the sets (e.g., WordNet by Princeton University, 2023), the process was manual. The semantic sets were compiled from collocate lists generated by Sketch Engine considering the first 50 strongest collocates (whenever the total of collocates was above 50) with a minimum frequency of 2 within a range of five words from the node to both the left and the right sides (-5 to +5). Initially, collocates carrying a context-independent meaning—in general, only a few—were grouped according to their meaning using different labels for each group. For example, *Anova*, *Kruskal-Wallis*, and *Manova* were put together under the label STATISTICAL TESTS. Next, the concordance lines of polysemous and context-dependent words—in general, most of the collocates—were analyzed to determine their prevalent sense and then select or create the appropriate group. For example, the examination of the concordance lines of *average* in enRAs has led to the conclusion that it refers to a kind of variance (i.e., *average variance*), therefore being suitable for the TYPES OF VARIANCE group. Verbal collocates were normally included in the ACTION AS AGENT set when the keyword played the role of subject or in the ACTION AS TARGET set when it acted as an object. Collocates that could not be assigned to a group, including most prepositions, articles, and conjunctions, were ignored. Numbers and mathematical symbols were considered, but punctuation marks were also ignored. Colors were used to facilitate the classification. The results were double-checked against the concordance lines to correct mistakes.

Both the collocational and the semantic associational analyses sought signs of psychological priming that could confirm genre-specificity, in addition to the evidence provided by genre-typical keywords. While similarities between the specialized, single-genre and the general, multi-genre corpora would deny it, differences could reinforce genre-specificity.

The second stage of the present study consisted in contrasting discipline-specific data sets, that is, the English, the Portuguese, and the Japanese Pediatrics and Management RA corpora, thus addressing disciplinary variation. This stage was also entirely performed with Sketch Engine, as it followed almost the same steps as in the previous stage. The second stage was composed of two series of keyword analyses to find both discipline-specific (first series) and non-discipline-specific (second series) keywords, a collocational analysis of selected non-discipline-specific keywords to identify possible differences in collocations between the disciplinary corpora, and a semantic associational analysis of the same keywords to investigate differences in terms of semantic associations.

The two series of keyword analysis were composed of the same six analyses: (1) enPED versus enMGT, (2) enMGT versus enPED, (3) ptPED versus ptMGT, (4) ptMGT versus ptPED, (5) jaPED versus jaMGT, and (6) jaMGT versus jaPED. The first series followed the same criteria as Table 6.1, focusing on discipline-specific, relatively rare items. The second series was performed with a change in focus from 1.0 to 1,000, thus placing emphasis on common, shared words. As in the first stage, it was important that all keywords were minimally distributed among the RAs, that is, they should be key keywords. In the first series, the presence in a minimum of 30% of RAs was adopted as a threshold to distinguish key keywords from non-key keywords. In the second, the presence in a minimum of 50% of RAs was adopted, since the change in focus produced a higher number of keywords.

For each of the three languages, two non-discipline-specific keywords were selected (six in total) based on overall frequency, distribution, meaning, and (only in the case of Japanese) syntax. The discipline-specific keywords found could not be used for collocational and semantic associational analysis in this stage because they are practically exclusive to either Pediatrics or Management data, hindering cross-disciplinary comparisons.

The study of collocations in the second stage involved two approaches. The first was the analysis of collocates with a minimum frequency in corpus of 2 occurring within a range of five words before and five words after the node (-5 to +5). The change in collocate range—in the first stage, it was limited to immediately preceding and following words (-1 to +1)—was due to two reasons. First, because it increases the number of resulting collocates, which is a desired effect to investigate small data sets such as our disciplinary corpora. Second, because it allows to overcome the Japanese particle barrier that surrounds Japanese words, probably yielding more interesting findings. Collocates were searched again with Sketch Engine's concordance function (query type: "word"; part of speech: "noun" or "any," depending on the word; with no case distinction) and its collocation option. The Japanese collocates had to be manually filtered, as many word fragments were still listed by Sketch Engine, despite the increase in range. The top 20 strongest collocates were considered (unless the total was below 20) based once more on logDice (Rychlý, 2008). The percentage of co-occurrences between collocate and keyword in relation to the frequency of the keyword in the corpus was calculated again as a further measure for comparison.

The second approach for the study of collocations in the second stage was the analysis of 2–6-word clusters containing the six selected keywords. The clusters were obtained using Sketch Engine's n-gram function with the criteria presented in Table 6.3.

Table 6.3 – Criteria adopted for cluster analysis in the second stage.

Criterion	Description	Set-Up
N-Gram Length	It is possible to choose either a number or an interval between 2 and 6	2–6
Attribute	Options include "word," "lemma," "tag," "tags," "morphemes," and "part of speech"	Word
Frequency Min	Defines the minimum frequency a n-gram must have in the corpus to be listed	5
Frequency Max	Defines the maximum frequency a n-gram can have in the corpus to be listed	Inactive (zero)
Key N-Grams	Activates the comparison between two corpora to obtain typical n-grams of a corpus in relation to another corpus	Inactive
Additional Criteria	Options include "starting with letters," "ending with letters," "starting with word," "containing word," and "ending with word"	Containing word
Nest N-Grams	Groups shorter n-grams that are part of longer n-grams	Inactive
A = a	Defines whether the analysis will distinguish between lower and upper case (inactive) or not (active)	Active
Include Nonwords	Includes numbers, punctuation marks, and tokens that do not start with letters (e.g., *10-year*)	Inactive
Exclude These Words	Excludes lexical items chosen by the user	Inactive

Source: Prepared by the author considering the criteria available in Sketch Engine (Lexical Computing CZ, n.d.).

Length was set in 2–6 to obtain the widest range possible. "Word" was chosen as attribute following again Hoey's (2005) premise that primings relate to words. Minimum frequency was set in 5 to avoid low-frequency items; maximum frequency was not considered relevant. The analysis focused on one corpus per time, so the key n-gram option was ignored. "Containing word" proved to be suitable to extract clusters with the target keywords. The distinction between lower and upper case was not desired. The remaining three options (to nest n-grams, to include nonwords, and to exclude selected words) were not necessary for the purposes of our analysis.

The study of semantic associations of the second stage was also based on semantic sets. The semantic sets were manually compiled from collocate lists generated by Sketch Engine (Microsoft Excel spreadsheets) following the same steps as in the first stage. The number of collocates used to build the sets, however, was doubled. The first 100 strongest collocates (whenever the total of collocates was above 100) within a range of five words from the node to both the left and the right sides (-5 to +5) were considered. This change was made to obtain bigger numbers of sets and collocates. Also, at this time, punctuation marks, numbers, and mathematical symbols were all excluded. Some collocates were grouped according to their isolated meanings. Part of the categories of the first stage was reused, as several collocates found in the second stage fitted well into these categories. The concordance lines of polysemous and context-dependent words—which represented the major part of the collocates—were then examined to complete the classification. Collocates that could not be assigned to a group were ignored. The resulting semantic sets were double-checked against the concordance lines to eliminate mistakes.

Both the collocational and the semantic associational analyses in the second stage sought signs of psychological priming that could confirm disciplinary specificity, in addition to the evidence provided by discipline-specific keywords. Whereas similarities between the Pediatrics RA and the Management RA corpora would deny disciplinary specificity, differences could reinforce it.

The third stage of the present study investigated the Pediatrics RA and the Management RA section subcorpora focusing on textual position. This stage was entirely performed with AntConc, as this software program includes a tool that measures word dispersion throughout text files. The third stage consisted in two distributional analyses. In the first, the textual position of selected discipline-specific keywords was investigated across the different RA section subcorpora. In the second, the position of clusters containing the selected keywords was studied in the same data sets. Three keywords were chosen for each discipline and each language (18 keywords in total) among those found in the beginning of the second stage; the selection was based mainly on overall frequency—for this stage, it was necessary that the target words would be high-frequency items.

The study of the distribution of the selected keywords started with a simple frequency count of instances using AntConc's Key-Word-In-Context (KWIC) Tool. Then the percentage of the total of instances was calculated for every subcorpus. After this, the dispersion values of the instances of the keywords were calculated using the Plot Tool. AntConc's Plot Tool offers four options of dispersion measures: Juilland's D, range, standard deviation, and standard deviation (normed). Juilland's D, the default option, was used. Its values range from 0.000 to 1.000. Whereas a value of 0.000 means that the keyword appears only in a specific, isolated part of the file, a value of 1.000 means that it occurs throughout the entire file. The dispersion values of each keyword were assembled in Microsoft Excel spreadsheets and then used to create graphs to show the distribution of the keywords according to RA section subcorpora. On the one hand, graphs exhibiting low dispersion values would indicate a strong association between a given keyword and particular textual positions; on the other, graphs exhibiting high values would suggest a weak association.

The study of the distribution of clusters started with a search for clusters. This search was performed with AntConc's Cluster Tool with the following criteria.

Table 6.4 – Criteria adopted for cluster search in the third stage.

Criterion	Description	Set-Up
Search Query	There are three options: "word" (default), "case" (case sensitive), and "regex" (only expressions separated by white spaces are considered)	Word
Cluster Size	Ranges from 1 to 25	2, 3, and 4
Min. Frequency	Defines the minimum frequency the cluster must have in the corpus to be listed	1
Min. Range	Defines the minimum number of files in which the cluster must appear to be listed	1
Search Term Position	There are three options: "on left," "on right," and "on left/right"	On left/right

Source: Prepared by the author considering the criteria available in AntConc (Anthony, 2022).

"Word" was chosen as the search query option because it is more comprehensive than the others. Cluster size was set in between 2 and 4 because this range proved to be appropriate to obtain the desired results. Clusters with more than four words seem to be very exclusive. Both minimum frequency and minimum range were set in 1 because all 2–4-word clusters were targeted. Search term position was set in “on left/right” to include every cluster containing the target keywords.

The clusters found were classified into two groups according to the data set of appearance. Clusters occurring in only one subcorpus (only one RA section) were classified as *exclusive*; clusters occurring in two or more subcorpora (two or more RA sections) were classified as *shared*. This classification was performed with Microsoft Excel’s conditional formatting rule, which identifies and highlights repeated content in spreadsheet cells. The numbers of cluster types forming each of the two classes were then computed according to size, and examples of exclusive and shared clusters were recorded separately. Although possible differences and similarities between the exclusive and shared clusters were beyond the focus of the distributional analysis, their examples were considered valuable evidence concerning textual colligational priming.

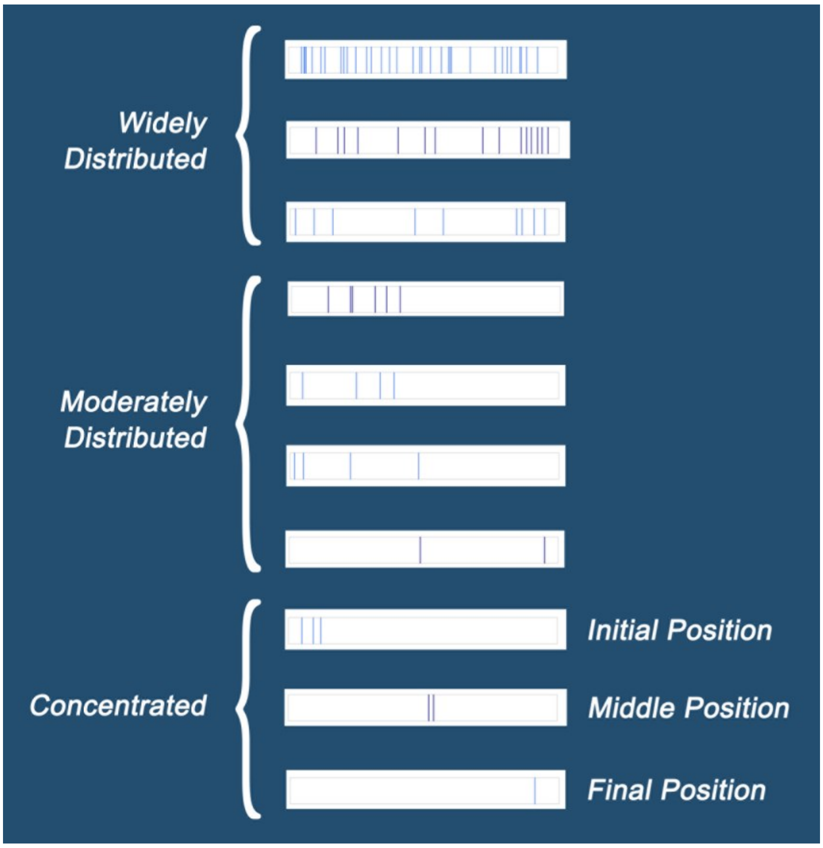
The distributional analyses in the third stage sought signs of psychological priming that could demonstrate the relevance of textual position. Keywords occurring in specific parts of the RAs, with low dispersion values, would corroborate this relevance; keywords occurring throughout the RAS, with high dispersion values, would deny it. In addition, high numbers of section-exclusive clusters would offer support for the pervasiveness of textual colligations (Hoey, 2013) in the RA genre; low numbers, by contrast, would point to the opposite.

The fourth stage of the present study consisted in comparing textual colligations, collocations, and colligations of semantically equivalent English, Portuguese, and Japanese words, addressing thereby cross-linguistic variation. This stage was performed combining Sketch Engine and AntConc. The stage had four substages: a word-list comparison to identify semantically equivalent, high-frequency words; a distributional analysis of selected words to identify possible differences in textual position; a collocational analysis to explore the existence of similar collocates; and a colligational analysis to understand whether the selected words perform similar roles at the sentence and clause levels or not.

Sketch Engine's word list function was used to find high-frequency words in enPED, enMGT, ptPED, ptMGT, jaPED, and jaMGT. This software program was used because it exports the resulting lists with both simple frequency and frequency per million tokens directly to Microsoft Excel files, which facilitates data analysis and the presentation of results. The disciplinary corpora were preferred to the bigger, RA corpora because textual colligation was examined according to RA section subcorpora, which differ depending on the discipline. The lists were manually examined, and the 50 most frequent nouns were recorded. The class of nouns was chosen because of its suitability for comparisons involving English, Portuguese, and Japanese. Other classes, such as adjectives and verbs, exhibit very distinctive characteristics, especially in Japanese. Three equivalent words were chosen from the Pediatrics lists (one per language) and three from the Management lists (one per language) based on meaning and disciplinary specificity. In addition, a partially equivalent Japanese word in Pediatrics was also included due to the limited distribution of the equivalent word initially chosen. Words representing the two disciplines were preferred, so that disciplinary variation could also be considered.

The study of the distribution of the selected words begun with a simple frequency count of instances using again AntConc's KWIC Tool. The percentage of instances per RA section subcorpus was then calculated. Next, the Plot Tool was employed to visualize the distribution of the words across RA sections. The resulting plot graphs of the English, Portuguese, and Japanese words were submitted to visual inspection and classified into the following categories: (1) *widely distributed* (plot graphs with instances in the first, second, and third parts); (2) *moderately distributed* (plot graphs with instances in the first and second, in the first and third, or in the second and third parts); and (3) *concentrated* (plot graphs with instances in the first, in the second, or in the third part only). In addition, concentrated graphs were classified into three subcategories: (1) *initial position* (plot graphs with instances in the first third only); (2) *middle position* (plot graphs with instances in the second third only); (3) *final position* (plot graphs with instances in the last third only). Figure 6.2 illustrates the classification.

Figure 6.2 – Plot classification according to word distribution.



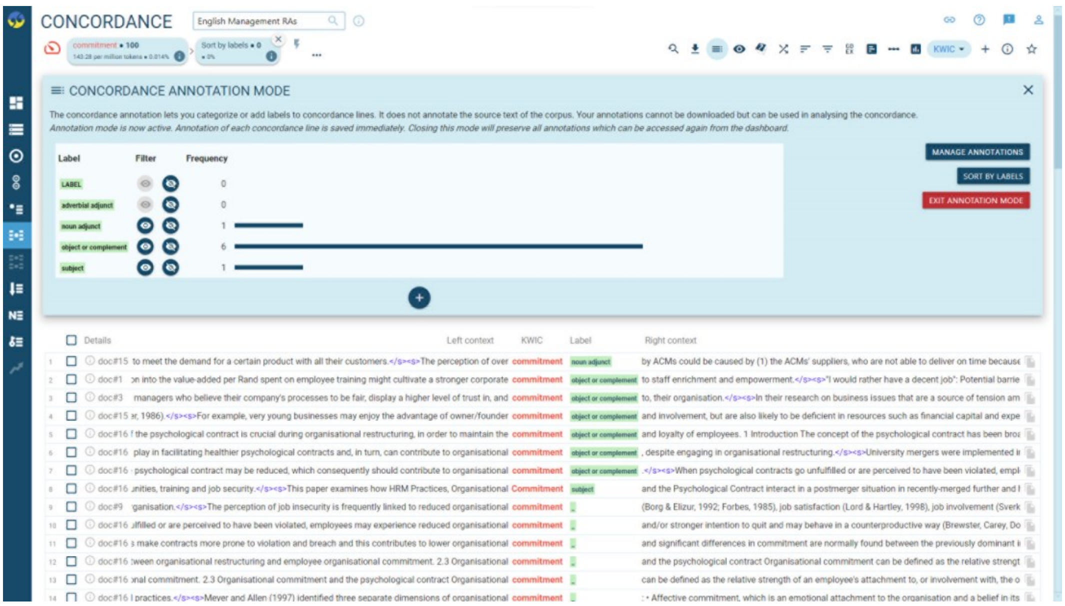
Source: Designed by the author. Plot graphs generated by AntConc.

The study of collocations was performed using Sketch Engine's concordance function (query type: "word"; part of speech: "any"; without case sensitivity) and its collocation option (attribute: "word [lowercase]" or "word" for Japanese). Collocates with a minimum frequency in corpus of 2 within a range of five words before and after the node (-5 to +5) were searched. Numbers, mathematical symbols, and punctuation marks were ignored, and Japanese word fragments were excluded. The 50 strongest collocates found (whenever the total of collocates was above 50) were classified cross-linguistically into three groups: (1) *equivalents*, (2) *semantically related*, and (3) *unrelated*. Sets of corresponding collocates such as *months* and *meses* ("months") were classified as equivalents, receiving distinguishing colors. Sets of collocates with related meanings, for example *T1DM* (type 1 diabetes mellitus) and the Portuguese *pneumonia* ("pneumonia"), were classified as semantically related, receiving tones of red or pink. The remaining collocates were classified as unrelated, being recorded without colors. Concordance lines were consulted to confirm the classification. In cases where the collocate exhibited several senses, only the main senses were considered.

Unlike the previous stages, log-likelihood was initially adopted for collocational analysis in the fourth stage, following Shao's (2017, 2018) studies of synonyms in Chinese and English. During data analysis, however, log-likelihood scores proved to be inappropriate for cross-corpus comparisons. Therefore, logDice scores were calculated for all the collocates previously found. The proportion of co-occurrences between collocate and node in relation to the overall number of occurrences of the collocate was also calculated in this stage as an additional measure for comparison.

Finally, the study of colligations was based on random samples of 100 concordance lines for each selected word, except in cases in which the total of occurrences of the word was below 100—in such cases, all lines were considered. The colligational study was performed with Sketch Engine because it includes an annotation mode in the concordance tool that facilitates data classification through tags and an automatic saving feature. Figure 6.3 displays the annotation mode.

Figure 6.3 – Annotation mode with concordance lines for *commitment* in enMGT.



Source: Screenshot of Sketch Engine (Lexical Computing CZ, n.d.). Reproduced with permission.

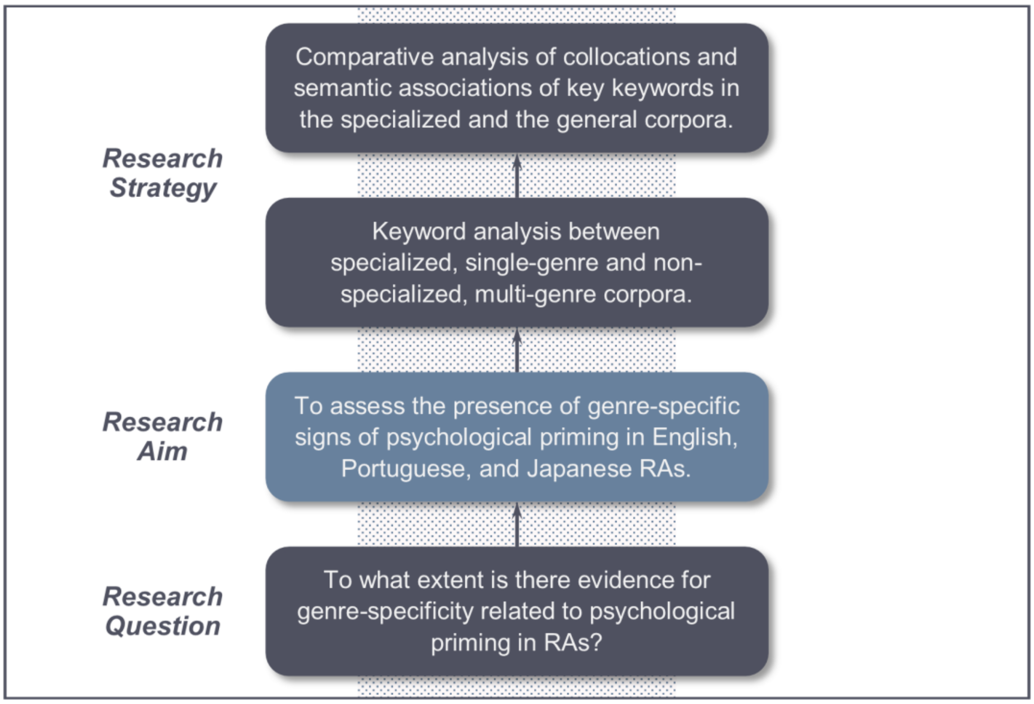
Following Hoey (2005) and Aragão (2022a), broad classes were adopted to classify the data. Words that perform the action expressed by the verb, words that receive the action in a passive structure, words that are connected to a complement by a copula (*be*, *become*, *seem*, *remain*, etc.), and words that function as the topic of the rest of the clause or sentence were classified as *subject*, irrespective of occurring in the main, in a subordinate, or in a coordinate clause. In the case of Japanese, both が (*ga*) and は (*wa*) were assumed as subject-signaling particles. The particle の (*no*, "of") was also assumed as subject-signaling when followed by a verb—as in 関心のある, *kanshin no aru*, "with interest," which holds the same meaning as 関心がある, *kanshin ga aru*, "there is interest." Words that directly or indirectly receive the action expressed by the verb (except in passive structures), words that act as the core of the complement after the copula, and words that function as objects for adjectives (for example, *the employees* in *the change was beneficial for the employees*) were classified as *object or complement*. In the case of Japanese, both を (*o*) and に (*ni*) were considered object-signaling particles, with the first indicating a direct and the second an indirect object. Words that act over the subject in passive constructions were classified as *agent of the passive* (for example, *the findings* in *the hypothesis was validated by the findings*). Words that either modify or help to modify the subject, an object, the complement, or the passive agent were classified as *noun adjunct*. In the case of Japanese, the particle の (*no*, "of") was the primary signal for this class, except when followed by a verb. Words that either contextualize or help to contextualize actions were classified as *adverb adjunct*. Words that appear in appositives, incomplete clauses within parentheses, titles, or names were classified as *appositive, parentheses-enclosed phrase, title, name*. Words that could not fit into any of the above categories were classified as *others*. The results of the classification were double-checked to ensure consistency. Despite this, Hoey's (2005) words deserve mention: "As anyone who attempts the grammatical analysis of authentic data knows, one encounters rather more cases where a correct analysis is problematic than one might anticipate on the basis of conveniently simple, made-up examples." (p. 46) Some instances could be classified into two categories, depending on data interpretation. In such cases, classification privileged clause-level over sentence-level functions. Additionally, to distinguish between the two classes of adjunct the core of the modified part was used as the main criterion. When it was a noun, the instance was classified as noun adjunct; when it was a verb, it was classified as adverbial adjunct.

The textual colligational, the collocational, and the colligational analyses in the fourth stage sought signs of psychological priming that could show whether users of different languages make similar associations when employing equivalent words in the same genre or not. While cross-linguistic similarities would suggest that they do make similar associations, differences would deny this possibility.

7 PSYCHOLOGICAL PRIMING AND THE RA GENRE

Genre has been assumed as a decisive factor for psychological priming (Hoey, 2004, 2005, 2013). Every priming in principle is genre-specific, even though there are many cross-generic primings (Hoey, 2005). Language users would be primed to recognize and replicate not only collocations but also the genres where they occur (Hoey, 2007b). Despite this, as shown in Chapter 3, lexical priming research has not directed much attention to well-defined, specific genres. Hoey's (2005) fundamental work focuses on big, multi-genre corpora: "over 95 million words of *Guardian* news and features text, supplemented by slightly more than 3 million words from the British National Corpus (written text) and 230,000 words of spoken data" (p. xi). Much of the literature reviewed in Chapter 3 shares the same focus, for example Hoey (2007a, 2007b), Patterson (2015), Cantos and Almela (2017), and Jantunen (2017). Although there are studies that do deal with recognizable genres, in most cases they explore fairly broad generic categories. Tsiamita (2011), for example, examined a subcorpus from the BNC composed of fiction texts. Ooi (2013), whose work used the International Corpus Network of Asian Learners of English, approached English learners' essays. As a matter of fact, a more definite genre-orientation is found in but few works, such as Pace-Sigge's (2007) analysis of biographies and Chuang's (2015) study of licensing agreements. Consequently, the relationship between psychological priming and specific genres has not yet been fully understood. The first stage of the present study addressed this gap, with a focus on the RA genre. Figure 7.1 exhibits the guiding elements and a summary of the research strategy adopted in this stage.

Figure 7.1 – Guiding elements and research strategy of the first stage of the present study.



Source: Designed by the author.

The keyword analysis of the first stage aimed to identify genre-specific keywords—more precisely, keywords characteristic of the RA genre—by comparing this study's specialized corpora with general corpora. It consisted of a series of five consecutive analyses. The first analysis compared the vocabulary of the single-genre English RA corpus (enRAs) with that of the multi-genre enTenTen15 corpus. Table 7.1 shows the key keywords identified through this process.

Table 7.1 – Key keywords of English RAs (enRAs versus enTenTen15).

		General Freq.		Freq. per Million		RA Freq.		
POS	Word	FC	RC	FC	RC	n	(%)	Score
1	*CI*	343	98,412	343.61	6.39	39	27.08	46.7
2	*infants*	522	159,880	522.92	10.37	30	20.83	46.1
3	*hypothesis*	438	155,268	438.77	10.07	44	30.56	39.7
4	*adolescents*	372	133,248	372.66	8.65	26	18.06	38.7
5	*organisational*	327	116,340	327.58	7.55	22	15.28	38.4
6	*bivariate*	46	4,314	46.08	0.28	17	11.81	36.8
7	*variables*	745	300,999	746.32	19.53	101	70.14	36.4
8	*organizational*	898	366,415	899.59	23.78	41	28.47	36.4
9	*hypotheses*	158	52,640	158.28	3.42	36	25.00	36.1
10	*neonatal*	158	53,187	158.28	3.45	18	12.50	35.8
11	*respondents*	512	208,312	512.90	13.52	54	37.50	35.4
12	*variance*	191	69,929	191.34	4.54	59	40.97	34.7
13	χ^2	36	1,060	36.06	0.07	17	11.81	34.7

Source: Keyword analysis performed with Sketch Engine. FC = focus corpus. RC = reference corpus. RA Freq. = number (n) and percentage (%) of RAs containing the word. CI = confidence interval.

As explained before, in this stage the presence in at least 10% of the RAs in the corpus was adopted as a threshold to distinguish key keywords from non-key keywords. Because of this, keywords in Table 7.1 do not occupy the top positions in the original list generated by Sketch Engine, which is ordered by decreasing keyness scores. The first keyword, *CI* (confidence interval), for example, is only the 48th item in the list but the first appearing in a minimum of 15 of the 144 English RAs—that is, it is the first key keyword. Also, it should be remembered that Sketch Engine's keyness scores express how much more frequent words are in a given corpus in relation to another corpus based primarily on frequencies per million. Therefore, words in Table 7.1 are ordered according to the difference in frequency between enRAs and enTenTen15.

To obtain an additional perspective, the second keyword analysis compared enRAs with the larger and newer enTenTen20 corpus. Table 7.2 shows the key keywords obtained from the process.

Table 7.2 – Key keywords of English RAs (enRAs versus enTenTen20).

		General Freq.		Freq. per Million		RA Freq.		
POS	Word	FC	RC	FC	RC	n	(%)	Score
1	*organisational*	327	190,110	327.58	4.23	22	15.28	62.9
2	*organizational*	898	666,484	899.59	14.82	41	28.47	56.9
3	*infants*	522	371,204	522.92	8.25	30	20.83	56.6
4	*adolescents*	372	259,255	372.66	5.77	26	18.06	55.2
5	*neonatal*	158	100,276	158.28	2.23	18	12.50	49.3
6	*CI*	343	273,583	343.61	6.08	39	27.08	48.6
7	*maternal*	390	345,666	390.69	7.69	27	18.75	45.1
8	*hypotheses*	158	138,973	158.28	3.09	36	25.00	38.9
9	*bivariate*	46	10,006	46.08	0.22	17	11.81	38.5
10	*Likert*	43	7,750	43.08	0.17	21	14.58	37.6
11	*hypothesis*	438	483,017	438.77	10.74	44	30.56	37.5
12	*affective*	110	92,304	110.19	2.05	17	11.81	36.4
13	*respondents*	512	593,595	512.90	13.20	54	37.50	36.2
14	*constructs*	158	152,994	158.28	3.40	28	19.44	36.2

Source: Keyword analysis performed with Sketch Engine. FC = focus corpus. RC = reference corpus. RA Freq. = number (n) and percentage (%) of RAs containing the word. CI = confidence interval.

As can be seen, there are many similarities between Tables 7.1 and 7.2. Their numbers of key keywords are close (13 versus 14); their keyness score ranges are not so different (34.7–46.7 versus 36.2–62.9); and, above all, there are 10 shared key keywords: *CI*, *infants*, *hypothesis*, *adolescents*, *organizational*, *bivariate*, *organizational*, *hypotheses*, *neonatal*, and *respondents*. The key keywords in both tables indicate that the English RAs in the sample distinguish themselves from general English mainly with words related to STATISTICS AND RESEARCH METHODOLOGY (*CI*, *hypothesis*, *bivariate*, *variables*, *hypotheses*, *respondents*, *variance*, χ^2, *Likert*, *constructs*), PEDIATRICS (*infants*, *adolescents*, *neonatal*, *maternal*), and MANAGEMENT (*organisational* and *organizational*). These semantic sets seem to represent genre-specificity for the selected English RAs.

The third keyword analysis compared the specialized, single-genre Portuguese RA corpus (ptRAs) and the general, multi-genre ptTenTen11 corpus. Table 7.3 presents the key keywords identified by the process.

Table 7.3 – Key keywords of Portuguese RAs (ptRAs versus ptTenTen11).

POS	Word	General Freq. FC	General Freq. RC	Freq. per Million FC	Freq. per Million RC	RA Freq. n	RA Freq. (%)	Score
1	*construto*	92	1,167	356.43	0.25	6	12.5	285.4
2	*lactentes*	72	3,308	278.95	0.72	7	14.6	163.2
3	*IC95*	50	1,056	193.71	0.23	10	20.8	158.5
4	*cuidadores*	62	7,631	240.20	1.65	6	12.5	91.0
5	*familiaridade*	61	8,522	236.33	1.84	5	10.4	83.5
6	*significância*	52	6,786	201.46	1.47	23	47.9	82.0
7	*et*	497	105,313	1925.50	22.78	38	79.2	81.0
8	*aleitamento*	85	15,162	329.31	3.28	7	14.6	77.2
9	*Cronbach*	19	420	73.61	0.09	8	16.7	68.4
10	*regressão*	72	15,105	278.95	3.27	15	31.3	65.6
11	*pediatras*	39	6,413	151.10	1.39	6	12.5	63.7
12	*variáveis*	295	81,461	1142.90	17.62	36	75.0	61.4
13	*fatorial*	19	1,444	73.61	0.31	6	12.5	56.9
14	*organizacional*	213	62,835	825.21	13.59	10	20.8	56.6
15	*α*	20	1,812	77.48	0.39	8	16.7	56.4
16	*qui-quadrado*	19	1,769	73.61	0.38	13	27.1	54.0
17	χ^2	14	187	54.24	0.04	5	10.4	53.1
18	*H1*	18	1,727	69.74	0.37	8	16.7	51.5
19	*gestacional*	30	6,230	116.23	1.35	6	12.5	49.9
20	*H3*	15	862	58.11	0.19	7	14.6	49.8
21	*respondentes*	23	3,774	89.11	0.82	7	14.6	49.6
22	*discriminante*	14	648	54.24	0.14	5	10.4	48.4
23	*corporativa*	132	44,653	511.40	9.66	5	10.4	48.1
24	*dummy*	13	368	50.37	0.08	7	14.6	47.6
25	*al*	456	167,559	1766.65	36.25	27	56.3	47.5
26	*mensuração*	40	10,716	154.97	2.32	6	12.5	47.0

Source: Keyword analysis performed with Sketch Engine. FC = focus corpus. RC = reference corpus. RA Freq. = number (n) and percentage (%) of RAs containing the word. An English translation of this table is provided in Appendix A.

As can be seen, there are 26 key keywords with RA frequency between 10.4 and 79.2% of RAs (threshold = 10% or five articles) and a keyness score range from 47.0 to 285.4. Most of the keywords relate to STATISTICS (*IC95,* “confidence interval of 95%,” *significância*, “significance,” *Cronbach*, *regressão*, “regression,” *variáveis*, “variables,” *fatorial*, “factorial,” *correlação*, “correlation,” *qui-quadrado*, “chi-squared,” etc.), but there are also keywords related to PEDIATRICS (*lactentes*, “nurslings,” *aleitamento*, “breast-feeding,” *pediatras*, "pediatricians," etc.) and MANAGEMENT (*organizacional*, "organizational" and *corporativa*, "corporate"). These semantic sets—much akin to the English ones—seem to represent genre-specificity for the selected Portuguese RAs.

To add to these results, the fourth analysis compared CoPEP (Kuhn & Ferreira, 2018) with ptTenTen11. Table 7.4 displays the results.

Table 7.4 – Key keywords of Portuguese journal texts (CoPEP versus ptTenTen11).

		General Freq.		Freq. per Million		DOC Freq.		
POS	Word	FC	RC	FC	RC	n	(%)	Score
1	*nº*	4,142	0	84.83	0.00	1,321	13.34	85.8
2	*Tabela_1*	3,055	0	62.57	0.00	2,027	20.47	63.6
3	*aspetos*	3,009	1,324	61.63	0.29	1,133	11.44	48.7
4	*Tabela_2*	2,293	0	46.96	0.00	1,526	15.41	48.0
5	*Figura_1*	2,200	0	45.06	0.00	1,613	16.29	46.1
6	*Tabela_3*	1,610	0	32.97	0.00	1,117	11.28	34.0
7	*et*	38,504	105,313	788.59	22.78	4,241	42.84	33.2
8	*Figura_2*	1,290	0	26.42	0.00	1,006	10.16	27.4
9	*estatisticamente*	4,167	13,541	85.34	2.93	1,378	13.92	22.0
10	*enfermagem*	29,804	129,759	610.41	28.07	1,554	15.70	21.0
11	*al*	35,054	167,559	717.93	36.25	3,707	37.44	19.3
12	*significância*	2,168	6,786	44.40	1.47	1,098	11.09	18.4
13	*variáveis*	16,495	81,461	337.83	17.62	3,164	31.96	18.2
14	*enfermeiros*	8,321	43,185	170.42	9.34	1,128	11.39	16.6
15	*verificou-se*	3,501	19,895	71.70	4.30	1,865	18.84	13.7
16	*doentes*	17,262	116,394	353.54	25.18	1,891	19.10	13.5
17	*amostra*	11,428	77,019	234.05	16.66	2,678	27.05	13.3
18	*correlação*	4,801	30,740	98.33	6.65	1,539	15.55	13.0
19	*prevalência*	5,318	35,768	108.92	7.74	1,561	15.77	12.6
20	*empírica*	1,961	10,760	40.16	2.33	1,115	11.26	12.4
21	*contextos*	5,651	39,209	115.74	8.48	2,228	22.51	12.3
22	*observou-se*	2,207	13,271	45.20	2.87	1,272	12.85	11.9
23	*utilizou-se*	1,408	7,206	28.84	1.56	1,040	10.51	11.7
24	*terapêutica*	5,423	42,303	111.07	9.15	1,479	14.94	11.0
25	*variável*	8,345	68,207	170.91	14.75	2,323	23.46	10.9
26	*questionário*	4,957	39,851	101.52	8.62	1,349	13.63	10.7

Source: Keyword analysis performed with Sketch Engine. CoPEP was compiled by Kuhn and Ferreira (2018). FC = focus corpus. RC = reference corpus. DOC Freq. = number (n) and percentage (%) of texts containing the word. An English translation of this table is provided in Appendix A.

There are again 26 key keywords but this time with frequency between 10.16 and 42.84% of documents (threshold = 10% or 990 documents) and a keyness score range from 10.7 to 85.8. The keywords relate principally to STATISTICS (*estatisticamente*, "statistically," *significância*, "significance," *variáveis*, "variables," *amostra*, "sample," *correlação*, "correlation," *prevalência*, "prevalence," and *variável*, "variable"), HEALTH (*enfermagem*, "nursing," *enfermeiros*, "male nurses," *doentes*, "sick persons," and *terapêutica*, "therapeutic"), PASSIVE VOICE (*verificou-se*, "[it] was verified," *observou-se*, "[it] was observed," and *utilizou-se*, "[it] was used"), and TEXT COMPONENTS (*Tabela_1*, "Table_1," *Tabela_2*, "Table_2," *Figura_1*, "Figure_1," *Tabela_3*, "Table_3," and *Figura_2*, "Figure_2"). Because CoPEP (Kuhn & Ferreira, 2018) is composed of several journal genres, these semantic sets do not represent genre-specificity. However, they do seem to reveal an intermediate layer between ptRAs and ptTenTen11, that is, the scholarly journal Portuguese in opposition to general Portuguese. Also, it is interesting to note that the keyness score range here (10.7–85.8) is narrower than the one obtained in the third keyword analysis (47.0–285.4). This suggests that the more genre-specific the focus corpus is, the higher the keyness scores may be.

The fifth keyword analysis compared the specialized, single-genre Japanese RA corpus (jaRAs) with the general, multi-genre jpTenTen11 LUW corpus. Table 7.5 presents the results, including two data processing mistakes: あろ (*aro*, part of sentence-ending expressions such as であろう, *dearō*, "probably") and 異 (*i*, "objection"). Both were computed separately as words possibly due to the segmentation process. However, further examination of jpTenTen11 LUW revealed numerous instances ignored by the keyword function, which invalidated their status as keywords.

Table 7.5 – Key keywords of Japanese RAs (jaRAs versus jaTenTen11 LUW).

POS	Word	General Freq. FC	General Freq. RC	Freq. per Million FC	Freq. per Million RC	RA Freq. n	RA Freq. (%)	Score
1	海外子会社	140	20	594.53	0.10	5	10.42	542.3
2	本稿	315	389	1,337.68	1.91	21	43.75	460.0
3	本症	116	25	492.61	0.12	9	18.75	439.6
4	先行研究	144	96	611.51	0.47	19	39.58	416.3
5	既存研究	78	2	331.24	0.01	13	27.08	329.0
6	施行し	113	183	479.87	0.90	19	39.58	253.3
7	保存的治療	64	20	271.78	0.10	7	14.58	248.4
8	食道閉鎖症	55	1	233.56	0.00	5	10.42	233.4
9	自験例	54	10	229.32	0.05	15	31.25	219.5
10	症例	305	1,164	1,295.22	5.72	23	47.92	193.0
11	本研究	124	404	526.58	1.98	12	25.00	176.8
12	全例	46	40	195.34	0.20	16	33.33	164.1
13	小児外科医	38	5	161.37	0.02	6	12.50	158.5
–	異	96	348	407.67	1.71	24	50.00	150.9
–	あろ	66	178	280.28	0.87	20	41.67	150.1
14	日齢	34	3	144.38	0.01	7	14.58	143.3
15	患児	42	66	178.36	0.32	15	31.25	135.5
16	当科	49	115	208.08	0.56	12	25.00	133.6
17	後方視的	30	2	127.40	0.01	15	31.25	127.2
18	本論文	40	71	169.86	0.35	7	14.58	126.7
19	得ら	26	4	110.41	0.02	17	35.42	109.3
20	胃瘻造設	25	1	106.17	0.00	6	12.50	106.6
21	分析結果	47	183	199.59	0.90	11	22.92	105.7

Source: Keyword analysis performed with Sketch Engine. FC = focus corpus. RC = reference corpus. RA Freq. = number (n) and percentage (%) of RAs containing the word. Verbs with partial inflection are double underlined. Both a Romanized transcription and an English translation of this table are provided in Appendix A.

There are 21 key keywords with RA frequency between 10.42 and 47.92% (threshold = 10% or five articles) and a keyness score range from 105.7 to 542.03. Most keywords relate to either RESEARCH METHODOLOGY AND WRITING (本稿, *honkō*, "this paper," 先行研究, *senkō kenkyū*, "previous research," 本研究, *honkenkyū*, "the present study," 分析結果, *bunseki kekka*, "analysis results," etc.) or PEDIATRICS (本症, *honshō*, "this disease," 小児外科医, *shōnigekai*, "pediatric surgeon," 胃瘻造設, *irō zōsetsu*, "gastrostomy," among others). The only item directly related to Management is 海外子会社 (*kaigai shikaisha*, “overseas subsidiary”). Collectively, the keywords above seem to represent genre-specificity for the selected Japanese RAs. The Japanese RAs in the sample distinguish themselves from general Japanese primarily by words related to RESEARCH METHODOLOGY AND WRITING as well as to PEDIATRICS. Differently from the English and Portuguese data, words related to Statistics and Management are likely to be less relevant for such distinction.

After the series of five keyword analyses, three keywords were selected for each of the three languages for the subsequent study of collocations and semantic associations. As explained in Chapter 6, this selection was restricted to keywords identified by comparing enRAs with enTenTen15, ptRAs with ptTenTen11, and jaRAs with jaTenTen11 LUW, as the other two comparisons were only supplementary. In addition, the selection was based primarily on the proportion of Pediatrics and Management RAs containing each keyword as well as on keyness scores. In principle, keywords appearing in RAs from both disciplines with the highest scores would be preferred. The selection of Japanese keywords, however, demanded also the consideration of the number of instances of the keywords in jaTenTen11 LUW, as there are many keywords whose occurrence in this reference corpus is low (see Table 7.5). Choosing a keyword with low frequency in the reference corpus would make collocational and semantic comparisons impractical. Table 7.6 reproduces the keywords shown in Tables 7.1, 7.3, and 7.5 introducing the number and proportion of RAs containing each keyword per field, as well as highlighting the selected keywords for collocational and semantic associational study with colored rectangles.

Table 7.6 – Number and proportion of RAs containing key keywords per discipline and selected keywords.

English Keyword	PED n(%)	MGT n(%)	Portuguese Keyword	PED n(%)	MGT n(%)	Japanese Keyword	PED n(%)	MGT n(%)
CI	36(92.3)	3(7.7)	*construto*	1(16.7)	5(83.3)	海外子会社	0(0.0)	5(100.0)
infants	30(100.0)	0(0.0)	*lactentes*	7(100.0)	0(0.0)	本稿	0(0.0)	21(100.0)
hypothesis	10(22.7)	34(77.3)	*IC95*	10(100.0)	0(0.0)	本症	9(100.0)	0(0.0)
adolescents	26(100,0)	0(0.0)	*cuidadores*	6(100.0)	0(0.0)	先行研究	0(0.0)	19(100.0)
organisational	0(0.0)	22(100.0)	*familiaridade*	0(0.0)	5(100.0)	既存研究	0(0.0)	13(100.0)
bivariate	13(76.5)	4(23.5)	*significância*	12(52.2)	11(47.8)	施行し	19(100.0)	0(0.0)
variables	46(45.5)	55(54.5)	*et*	15(39.5)	23(60.5)	保存的治療	7(100.0)	0(0.0)
organizational	0(0.0)	41(100.0)	*aleitamento*	7(100.0)	0(0.0)	食道閉鎖症	5(100.0)	0(0.0)
hypotheses	1(2.8)	35(97.2)	*Cronbach*	1(12.5)	7(87.5)	自験例	15(100.0)	0(0.0)
neonatal	18(100.0)	0(0.0)	*regressão*	8(53.3)	7(46.7)	症例	23(100.0)	0(0.0)
respondents	14(25.9)	40(74.1)	*pediatras*	6(100.0)	0(0.0)	本研究	0(0.0)	12(100.0)
variance	18(30.5)	41(69.5)	*variáveis*	20(55.6)	16(44.4)	全例	16(100.0)	0(0.0)
x^2	13(76.5)	4(23.5)	*fatorial*	1(16.7)	5(83.3)	小児外科医	6(100.0)	0(0.0)
–	–	–	*organizacional*	0(0.0)	10(100.0)	日齢	7(100.0)	0(0.0)
–	–	–	*α*	0(0.0)	8(100.0)	患児	15(100.0)	0(0.0)
–	–	–	*qui-quadrado*	10(76.9)	3(23.1)	当科	12(100.0)	0(0.0)
–	–	–	x^2	1(20.0)	4(80.0)	後方視的	15(100.0)	0(0.0)
–	–	–	*H1*	0(0.0)	8(100.0)	本論文	5(71.4)	2(28.6)
–	–	–	*gestacional*	6(100.0)	0(0.0)	得ら	6(35.3)	11(64.7)
–	–	–	*H3*	0(0.0)	7(100.0)	胃瘻造設	6(100.0)	0(0.0)
–	–	–	*respondentes*	0(0.0)	7(100.0)	分析結果	0(0.0)	11(100.0)
–	–	–	*discriminante*	0(0.0)	5(100.0)	–	–	–
–	–	–	*corporativa*	0(0.0)	5(100.0)	–	–	–
–	–	–	*dummy*	1(14.3)	6(85.7)	–	–	–
–	–	–	*al*	4(14.8)	23(85.2)	–	–	–
–	–	–	*mensuração*	1(16.7)	5(83.3)	–	–	–

Source: Compiled by the author. PED = Pediatrics. MGT = Management. Selected keywords are bounded by colored rectangles. Verbs with partial inflection are double underlined. CI = confidence interval. An English translation of this table is provided in Appendix A together with a Romanized transcription of the Japanese words.

As can be seen, the nine selected key keywords are the following: *hypothesis*, *variables*, and *variance*; *significância* ("significance"), *regressão* ("regression"), and *variáveis* ("variables"); 本稿 (*honkō*, "this paper"), 症例 (*shōrei*, "clinical case"), and 本論文 (*honronbun*, "this paper"). *Hypothesis* was chosen because it has one of the highest keyness scores in the English group (Table 7.1) at the same that its occurrence is not overly concentrated in RAs from either Pediatrics or Management (Table 7.6). *Variables* and *variance* were chosen because they have the best balance between the areas (Table 7.6). *Significância* was chosen because it is the strongest in terms of keyness (Table

7.3) among the well-balanced words of the group (Table 7.6). *Regressão* and *variáveis* were chosen because they have the most balanced proportions with respect to discipline after *significância* (Table 7.6). 本論文 (*honronbun*) was chosen because it appears in RAs from both fields (Table 7.6) and also because it appears 71 times in jaTenTen11 LUW (Table 7.5). Although being restricted to one area (Table 7.6), 本稿 (*honkō*) and 症例 (*shōrei*) were chosen because they have high keyness scores and appear respectively 389 and 1,164 times in the reference corpus (Table 7.5). 得ら (*era*-, "obtain," with partial inflection) was rejected because it occurs only four times in the reference corpus (Table 7.5), despite the fact that it exhibits the best balance between areas among all the Japanese key keywords.

As explained in the previous chapter, the search for collocates in the first stage of this study focused on immediately preceding (left) and following (right) items. While the software-assisted analysis was successful in finding collocates in the English and Portuguese corpora, it did not succeed in the Japanese data. The Japanese collocates were manually identified from concordance lines or manually filtered from Sketch Engine's lists.

With respect to the comparison between focus and reference corpora, two measures were considered. The first is logDice (Rychlý, 2008), which ranges from 0 (zero) to 14 (extremely weak to extremely strong association) and is calculated by Sketch Engine. As explained earlier, a difference of 1 point between corpora represents double the frequency of co-occurrences (±100%) and a difference of 7 points means roughly one hundred times more co-occurrences (±10,000%). The second measure is the proportion of co-occurrences between collocate and keyword in relation to the frequency of the keyword in the corpus, which was manually calculated. Both measures can help to understand whether shared collocates between the different corpora play a similar role or not.

The top strongest left and right collocates of *hypothesis*, *variables*, and *variance* in enRAs and enTenTen15 are listed in Tables 7.7 to 7.12.

Table 7.7 – Top 20 strongest left collocates of *hypothesis* in two corpora.

	enRAs			enTenTen15		
POS	Word	PER	LogDice	Word	PER	LogDice
1	*test*	7.76	10.38	*null*	2.85	9.18
2	*null*	2.97	9.85	*Gaia*	0.42	6.86
3	*supports*	1.60	8.85	*Rieman*	0.34	6.76
4	*supporting*	1.37	8.59	*hygiene*	0.45	6.17
5	*selling*	1.14	8.44	*testable*	0.22	6.08
6	*following*	1.60	8.07	*continuum*	0.23	5.61
7	*support*	1.60	7.73	*phylogenetic*	0.16	5.41
8	*alternative*	0.68	7.40	*plausible*	0.16	5.23
9	*corroborating*	0.46	7.22	*Sapir-Whorf*	0.11	5.19
10	*posited*	0.46	7.20	*nebular*	0.10	5.05
11	*first*	1.14	7.01	*alternative*	0.74	4.84
12	*testing*	0.46	6.93	*extraterrestrial*	0.09	4.70
13	*proposed*	0.46	6.85	*amyloid*	0.08	4.63
14	*tests*	0.46	6.82	*AGW*	0.07	4.57
15	*second*	0.68	6.81	*biophilia*	0.07	4.55
16	*for*	6.39	6.71	*lipid*	0.08	4.32
17	*our*	1.83	6.70	*statistical*	0.21	4.29
18	*market*	0.68	6.46	*overarching*	0.07	4.07
19	*through*	0.46	5.65	*unproven*	0.05	4.00
20	*this*	1.83	5.60	*evolutionary*	0.09	3.82

Source: Collocates identified with Sketch Engine. PER = number of co-occurrences between collocate and keyword divided by the overall frequency of the keyword in the corpus. Shared collocates are bounded by colored rectangles. AGW = anthropogenic global warming.

Table 7.8 – Top 20 strongest right collocates of *hypothesis* in two corpora.

	enRAs			enTenTen15		
POS	Word	PER	LogDice	Word	PER	LogDice
1	*1b*	5.71	10.75	*testing*	2.86	6.66
2	*1a*	5.71	10.75	*predicts*	0.18	5.40
3	*3*	11.19	10.07	*posits*	0.07	4.40
4	*2*	13.01	9.90	*suggests*	0.30	4.39
5	*4*	7.08	9.64	*proposes*	0.12	4.34
6	*1*	10.96	9.40	*tests*	0.39	4.31
7	*2b*	1.83	9.13	*formulation*	0.10	4.14
8	*1c*	1.37	8.79	*assumes*	0.08	3.92
9	*5*	3.65	8.59	*postulates*	0.05	3.81
10	*6*	2.74	8.49	*driven*	0.18	3.70
11	*2a*	1.14	8.45	*revisited*	0.05	3.69
12	*testing*	0.91	7.93	*generation*	0.36	3.65
13	*11a*	0.68	7.80	*concerning*	0.16	3.44
14	*tests*	0.68	7.41	*asserts*	0.04	3.18
15	*10a*	0.46	7.22	*tested*	0.12	3.12
16	*8a*	0.46	7.22	*H1*	0.03	3.09
17	*8b*	0.46	7.22	*holds*	0.14	3.07
18	*9b*	0.46	7.22	*generating*	0.07	3.05
19	*10b*	0.46	7.22	*implies*	0.05	3.01
20	*11b*	0.46	7.22	*regarding*	0.25	2.98

Source: Collocates identified with Sketch Engine. PER = number of co-occurrences between collocate and keyword divided by the overall frequency of the keyword in the corpus. Shared collocates are bounded by colored rectangles.

Table 7.9 – Top 20 strongest left collocates of *variables* in two corpora.

	enRAs			enTenTen15		
POS	Word	PER	LogDice	Word	PER	LogDice
1	*independent*	7.52	10.91	*explanatory*	0.80	7.85
2	*dependent*	5.10	10.42	*random*	1.50	7.72
3	*control*	6.58	10.31	*predictor*	0.50	7.24
4	*psychosocial*	3.36	9.96	*dependent*	1.01	7.22
5	*demographic*	2.15	9.25	*categorical*	0.49	7.22
6	*outcome*	2.01	9.02	*demographic*	0.73	7.21
7	*latent*	1.61	8.94	*confounding*	0.43	7.05
8	*continuous*	1.61	8.89	*latent*	0.37	6.72
9	*dummy*	1.34	8.70	*independent*	1.85	6.62
10	*two*	2.95	8.61	*macroeconomic*	0.33	6.55
11	*explanatory*	1.21	8.57	*continuous*	0.61	6.29
12	*categorical*	1.07	8.41	*environment*	2.40	6.06
13	*these*	4.03	8.29	*instrumental*	0.33	6.01
14	*instrumental*	0.94	8.18	*climatic*	0.23	5.97
15	*all*	2.42	8.02	*dummy*	0.20	5.93
16	*style*	0.81	7.92	*meteorological*	0.20	5.77
17	*parenting*	0.81	7.73	*socio-demographic*	0.17	5.73
18	*other*	2.01	7.62	*outcome*	0.42	5.65
19	*three*	1.07	7.61	*environmental*	1.40	5.56
20	*board*	0.81	7.54	*input*	0.52%	5.56

Source: Collocates identified with Sketch Engine. PER = number of co-occurrences between collocate and keyword divided by the overall frequency of the keyword in the corpus. Shared collocates are bounded by colored rectangles.

Table 7.10 – Top 20 strongest right collocates of *variables* in two corpora.

	enRAs			enTenTen15		
POS	Word	PER	LogDice	Word	PER	LogDice
1	*were*	10.07	8.82	*affecting*	0.24	5.32
2	*measuring*	0.81	7.77	*influencing*	0.11	4.72
3	*included*	1.07	7.65	*measured*	0.15	4.32
4	*are*	4.43	7.64	*associated*	0.36	4.18
5	*such*	0.94	6.72	*such*	1.93	3.82
6	*could*	0.67	6.64	*describing*	0.08	3.81
7	*showed*	0.40	6.63	*related*	0.38	3.67
8	*used*	0.67	6.46	*included*	0.34	3.64
9	*non-nationals*	0.27	6.43	*defined*	0.14	3.58
10	*identified*	0.40	6.41	*involved*	0.31	3.40
11	*related*	0.40	6.38	*affect*	0.11	3.36
12	*associated*	0.54	6.36	*analyzed*	0.05	3.23
13	*held*	0.27	6.35	*representing*	0.07	3.22
14	*classified*	0.27	6.32	*were*	2.26	3.19
15	*age*	0.54	6.22	*studied*	0.08	3.15
16	*that*	3.36	6.22	*examined*	0.06	3.14
17	*examined*	0.27	6.15	*derived*	0.06	3.13
18	*have*	0.94	6.11	*relating*	0.06	3.02
19	*at*	0.94	6.05	*measuring*	0.05	2.95
20	*include*	0.27	6.04	*collected*	0.07	2.82

Source: Collocates identified with Sketch Engine. PER = number of co-occurrences between collocate and keyword divided by the overall frequency of the keyword in the corpus. Shared collocates are bounded by colored rectangles.

Table 7.11 – Top 20 strongest left collocates of *variance* in two corpora.

	enRAs			enTenTen15		
POS	Word	PER	LogDice	Word	PER	LogDice
1	*average*	5.76	9.36	*phenotypic*	0.31	6.32
2	*between-individual*	2.09	9.35	*zoning*	0.51	5.93
3	*total*	6.28	9.28	*yard*	0.89	5.81
4	*conditional*	1.57	8.80	*conditional*	0.25	5.27
5	*method*	2.62	8.78	*residual*	0.22	5.06
6	*error*	1.57	8.61	*inverse*	0.17	5.00
7	*minimum*	1.57	8.57	*genetic*	0.77	4.96
8	*common-method*	1.05	8.40	*asymptotic*	0.11	4.92
9	*decreasing*	1.05	8.21	*setback*	0.14	4.89
10	*within-individual*	1.05	8.03	*unexplained*	0.10	4.65
11	*unique*	1.05	7.72	*cosmic*	0.15	4.34
12	*explained*	1.05	7.53	*noise*	0.43	4.17
13	*little*	1.05	7.50	*error*	0.58	4.07
14	*shared*	1.05	7.47	*unequal*	0.08	4.02
15	*much*	1.05	7.21	*explained*	0.53	3.87
16	*individual*	1.57	7.15	*between-study*	0.04	3.81
17	*hospital*	1.05	6.74	*wavelet*	0.05	3.74
18	*greater*	1.05	6.64	*slight*	0.14	3.54
19	*its*	1.05	6.18	*trait*	0.06	3.50
20	*of*	17.80	5.20	*total*	1.19	3.45

Source: Collocates identified with Sketch Engine. PER = number of co-occurrences between collocate and keyword divided by the overall frequency of the keyword in the corpus. Shared collocates are bounded by colored rectangles.

Table 7.12 – Strongest right collocates of *variance* in two corpora.

	enRAs			enTenTen15		
POS	Word	PER	LogDice	Word	PER	LogDice
1	*extracted*	5.76	10.59	*decomposition*	0.26	5.73
2	*inflation*	4.71	10.46	*σ*	0.17	5.53
3	*explained*	2.62	8.85	*estimation*	0.40	5.50
4	*test*	3.14	8.38	*sought*	0.93	5.23
5	*estimates*	1.05	7.85	*estimator*	0.10	4.75
6	*due*	1.05	7.10	*swaps*	0.10	4.58
7	*in*	20.42	6.06	*estimators*	0.08	4.50
8	*would*	1.05	6.04	*explained*	0.82	4.49
9	*can*	1.05	5.24	*decompositions*	0.07	4.42
10	*was*	2.62	5.07	*accounted*	0.15	4.34
11	*for*	3.66	4.75	*components*	0.77	4.27
12	*by*	1.05	3.89	*covariance*	0.06	4.16
13	*of*	5.76	3.57	*partitioning*	0.07	4.02
14	*and*	4.19	3.25	*inflation*	0.21	3.98
15	*with*	1.05	3.15	*analysis*	1.83	3.77
16	*to*	2.62	2.89	*unbiased*	0.06	3.69
17	*that*	1.05	2.65	*requests*	0.33	3.66
18	–	–	–	*reduction*	0.47	3.66
19	–	–	–	*estimates*	0.25	3.64
20	–	–	–	*homogeneity*	0.04	3.52

Source: Collocates identified with Sketch Engine. PER = number of co-occurrences between collocate and keyword divided by the overall frequency of the keyword in the corpus. Shared collocates are abounded by colored rectangles.

As can be seen, most collocates are exclusive to one or another list. While *test hypothesis*, *control variables*, and *average variance* are typical combinations from the specialized, single-genre enRAs, *extraterrestrial hypothesis*, *meteorological variables*, and *cosmic variance* are collocations found in the general, multi-genre enTenTen15. Although there are collocates shared by the two corpora, in most cases they exhibit very different logDice values, which indicates that their connections with the node have different weights. To illustrate this point, the following shared collocates can be cited (logDice values within parentheses): *alternative* before *hypothesis* (enRAs: 7.40; enTenTen15: 4.84), *tests* after *hypothesis* (enRAs: 7.41; enTenTen15: 4.31), *independent* before *variables* (enRAs: 10.91; enTenTen15: 6.62), *included* after *variables* (enRAs: 7.65; enTenTen15: 3.64), *total* before *variance* (enRAs: 9.28; enTenTen15: 3.45), and *explained* after *variance* (enRAs: 8.85; enTenTen15: 4.49). All of them constitute stronger collocations in the English RA corpus. The proportional frequencies (co-occurrences between collocate and keyword divided by the general frequency of the keyword) also differ markedly in most cases. The percentages of the above words can be shown to illustrate this point: *alternative* before *hypothesis* (enRAs: 0.68%; enTenTen15: 0.74%), *tests* after *hypothesis* (enRAs: 0.68%; enTenTen15: 0.39%), *independent* before *variables* (enRAs: 7.52%; enTenTen15: 1.85%), *included* after *variables* (enRAs: 1.07%; enTenTen15: 0.34%), *total* before *variance* (enRAs: 6.28%; enTenTen15: 1.19%), and *explained* after *variance* (enRAs: 2.62%; enTenTen15: 0.82%).

Despite what has been shown, there seem to be cross-generic collocations that exhibit similar behavior in both corpora. The combination between *null* and *hypothesis* has this characteristic, as its percentages and logDice values are relatively close: 2.97% (enRAs) versus 2.85% (enTenTen15); 9.85 (enRAs) versus 9.18 (enTenTen15).

The left and right collocates of *significância* (“significance”), *regressão* (“regression”), and *variáveis* (“variables”) are presented in Tables 7.13 to 7.18, displayed on the next pages for ease of presentation.

Table 7.13 – Left collocates of *significância* (“significance”) in two corpora.

	ptRAs			ptTenTen11		
POS	Word	PER	LogDice	Word	PER	LogDice
1	*apresentaram*	9.62	10.39	*apresentaram*	0.87	4.05
2	*de*	48.08	6.00	*demostrou*	0.04	3.69
3	*a*	19.23	5.39	*alcançou*	0.35	3.68
4	–	–	–	*extrema*	0.56	3.62
5	–	–	–	*mostraram*	0.41	3.57
6	–	–	–	*evidenciou*	0.07	3.46
7	–	–	–	*pouca*	0.63	3.17
8	–	–	–	*tamanha*	0.21	3.17
9	–	–	–	*relevada*	0.03	3.16
10	–	–	–	*detectou-se*	0.03	3.11
11	–	–	–	*alcançaram*	0.09	3.04
12	–	–	–	*somenos*	0.03	3.03
13	–	–	–	*mostrou*	0.81	3.02
14	–	–	–	*encontrou-se*	0.04	2.96
15	–	–	–	*encontrada*	0.31	2.93
16	–	–	–	*apresentou*	1.03	2.92
17	–	–	–	*atingiram*	0.13	2.90
18	–	–	–	*houve*	1.81	2.82
19	–	–	–	*observada*	0.15	2.76
20	–	–	–	*demonstraram*	0.12	2.74

Source: Collocates identified with Sketch Engine. PER = number of co-occurrences between collocate and keyword divided by the overall frequency of the keyword in the corpus. Shared collocates are bounded by colored rectangles. An English translation of this table is provided in Appendix A.

Table 7.14 – Right collocates of *significância* ("significance") in two corpora.

	ptRAs			ptTenTen11		
POS	Word	PER	LogDice	Word	PER	LogDice
1	*estatística*	17.31	11.66	*estatística*	22.27	9.03
2	*α=0,05*	3.85	10.22	*α*	0.41	6.74
3	*adotado*	3.85	9.62	*prognóstica*	0.24	6.20
4	*dos*	3.85	5.35	*adotado*	2.24	6.18
5	*de*	26.92	5.16	*limítrofe*	0.28	6.12
6	*em*	5.77	5.05	*toxicológica*	0.10	4.83
7	*da*	5.77	4.93	*sealer*	0.06	4.20
8	*e*	3.85	3.24	*pré-estabelecido*	0.07	4.16
9	–	–	–	*clínica*	1.30	4.13
10	–	–	–	*0,05*	0.10	3.92
11	–	–	–	*indeterminada*	0.06	3.77
12	–	–	–	*0*	0.04	3.58
13	–	–	–	*estatístico*	0.12	3.57
14	–	–	–	*estatistica*	0.04	3.56
15	–	–	–	*α=0,05*	0.03	3.27
16	–	–	–	*valor-p*	0.03	3.26
17	–	–	–	*p=0,05*	0.03	3.25
18	–	–	–	*dose-resposta*	0.03	3.21
19	–	–	–	*pré-fixado*	0.03	3.17
20	–	–	–	*fisiológica*	0.06	3.11

Source: Collocates identified with Sketch Engine. PER = number of co-occurrences between collocate and keyword divided by the overall frequency of the keyword in the corpus. Shared collocates are bounded by colored rectangles. An English translation of this table is provided in Appendix A.

Table 7.15 – Left collocates of *regressão* (“regression”) in two corpora.

	ptRAs			ptTenTen11		
POS	Word	PER	LogDice	Word	PER	LogDice
1	*mediante*	2.78	9.64	*acelerada*	0.06	3.30
2	*pela*	4.17	7.84	*acentuada*	0.08	3.23
3	*na*	9.72	7.19	*brutal*	0.07	3.03
4	*da*	11.11	6.34	*apresentaram*	0.16	2.65
5	*de*	43.06	6.31	*profunda*	0.15	2.54
6	*à*	2.78	5.87	*franca*	0.13	2.54
7	*a*	16.67	5.65	*ligeira*	0.04	2.45
8	*como*	2.78	5.41	*nítida*	0.03	2.34
9	–	–	–	*utilizando-se*	0.04	2.33
10	–	–	–	*tremenda*	0.03	2.32
11	–	–	–	*utilizou-se*	0.02	2.14
12	–	–	–	*espantosa*	0.02	2.12
13	–	–	–	*unitermos*	0.01	2.07
14	–	–	–	*1h30m*	0.01	2.03
15	–	–	–	*houve*	0.48	2.03
16	–	–	–	*titulada*	0.01	1.99
17	–	–	–	*progressiva*	0.04	1.94
18	–	–	–	*observaram*	0.02	1.85
19	–	–	–	*sofreria*	0.01	1.84
20	–	–	–	*acarretar*	0.02	1.75

Source: Collocates identified with Sketch Engine. PER = number of co-occurrences between collocate and keyword divided by the overall frequency of the keyword in the corpus. An English translation of this table is provided in Appendix A.

Table 7.16 – Right collocates of *regressão* ("regression") in two corpora.

	ptRAs			ptTenTen11		
POS	Word	PER	LogDice	Word	PER	LogDice
1	*múltipla*	11.11	11.57	*linear*	10.03	9.86
2	*logística*	9.72	11.38	*múltipla*	2.36	8.15
3	*linear*	2.78	9.66	*logística*	6.65	8.06
4	*univariada*	2.78	9.66	*hipnótica*	0.43	7.01
5	*indicam*	2.78	9.22	*espontânea*	1.14	6.90
6	*foi*	2.78	5.78	*multivariada*	0.23	6.05
7	*com*	4.17	5.39	*polinomial*	0.21	6.02
8	*de*	20.83	5.26	*infinita*	0.39	5.70
9	*para*	2.78	4.58	*tumoral*	0.20	5.67
10	–	–	–	*não-linear*	0.17	5.55
11	–	–	–	*civilizacional*	0.17	5.50
12	–	–	–	*quadrática*	0.10	4.96
13	–	–	–	*demográfica*	0.14	4.70
14	–	–	–	*univariada*	0.07	4.53
15	–	–	–	*autoritária*	0.11	4.44
16	–	–	–	*stepwise*	0.07	4.43
17	–	–	–	*cautelar*	0.15	4.23
18	–	–	–	*hierárquica*	0.09	4.21
19	–	–	–	*quantílica*	0.05	4.12
20	–	–	–	*terapêutica*	0.19	4.05

Source: Collocates identified with Sketch Engine. PER = number of co-occurrences between collocate and keyword divided by the overall frequency of the keyword in the corpus. Shared collocates are bounded by colored rectangles. An English translation of this table is provided in Appendix A.

Table 7.17 – Top 20 strongest left collocates of *variáveis* (noun, "variables") in two corpora.

	ptRAs			ptTenTen11		
POS	Word	PER	LogDice	Word	PER	LogDice
1	*as*	41.38	10.66	*inúmeras*	0.73	6.16
2	*das*	15.33	10.01	*múltiplas*	0.43	6.14
3	*essas*	3.07	9.45	*seguintes*	1.14	5.41
4	*às*	3.45	9.06	*diversas*	1.69	5.41
5	*três*	2.68	9.03	*destas*	0.66	5.35
6	*outras*	2.30	8.92	*outras*	4.86	5.33
7	*duas*	1.92	8.67	*essas*	1.87	5.27
8	*seis*	1.15	8.16	*dessas*	0.90	5.22
9	*possíveis*	0.77	7.74	*estas*	1.36	5.17
10	*algumas*	0.77	7.61	*tantas*	0.44	5.07
11	*quatro*	0.77	7.48	*duas*	3.38	4.99
12	*pelas*	0.77	7.48	*muitas*	2.03	4.83
13	*quais*	0.77	7.40	*algumas*	2.10	4.74
14	*nas*	0.77	6.86	*principais*	1.31	4.73
15	*para*	2.30	6.07	*as*	24.66	4.26
16	*entre*	0.77	5.84	*infinitas*	0.07	4.14
17	*como*	1.15	5.83	*determinadas*	0.13	4.12
18	*de*	5.75	5.24	*várias*	0.86	4.12
19	*com*	0.77	4.69	*diferentes*	0.82	3.98
20	*em*	0.77	4.37	*certas*	0.15	3.96

Source: Collocates identified with Sketch Engine. PER = number of co-occurrences between collocate and keyword divided by the overall frequency of the keyword in the corpus. Shared collocates are bounded by colored rectangles. An English translation of this table is provided in Appendix A.

Table 7.18 – Top 20 strongest right collocates of *variáveis* (noun, "variables") in two corpora.

	ptRAs			ptTenTen11		
POS	Word	PER	LogDice	Word	PER	LogDice
1	*independentes*	9.58	11.42	*aleatórias*	0.72	7.79
2	*dependentes*	4.21	10.35	*estudadas*	0.79	7.58
3	*categóricas*	3.07	9.93	*independentes*	1.33	7.54
4	*primárias*	1.92	9.26	*demográficas*	0.55	7.39
5	*endógenas*	1.92	9.25	*macroeconômicas*	0.49	7.27
6	*numéricas*	1.53	8.95	*explicativas*	0.50	7.23
7	*socioeconômicas*	1.53	8.93	*quantitativas*	0.46	7.08
8	*maternas*	1.53	8.91	*envolvidas*	1.00	7.05
9	*analisadas*	1.53	8.82	*categóricas*	0.38	6.95
10	*explicativas*	1.15	8.54	*analisadas*	0.59	6.91
11	*exógenas*	1.15	8.53	*contínuas*	0.39	6.75
12	*utilizadas*	1.15	8.40	*qualitativas*	0.30	6.48
13	*intervenientes*	0.77	7.96	*sócio-demográficas*	0.27	6.44
14	*manifestas*	0.77	7.96	*sociodemográficas*	0.26	6.42
15	*latentes*	0.77	7.95	*dependentes*	0.53	6.39
16	*contínuas*	0.77	7.95	*econômicas*	0.74	6.38
17	*contribuíram*	0.77	7.93	*meteorológicas*	0.28	6.32
18	*observadas*	0.77	7.92	*globais*	0.47	6.28
19	*clínicas*	0.77	7.83	*ambientais*	1.10	6.17
20	*associadas*	0.77	7.78	*psicológicas*	0.29	6.15

Source: Collocates identified with Sketch Engine. PER = number of co-occurrences between collocate and keyword divided by the overall frequency of the keyword in the corpus. Shared collocates are bounded by colored rectangles. An English translation of this table is provided in Appendix A.

As can be seen, most collocates are exclusive to one or another list once again. While *significância dos* ("significance of the"), *regressão indicam* ("regression indicate"), and *variáveis primárias* ("primary variables") are restricted to ptRAs, *significância toxicológica* ("toxicological significance"), *regressão hipnótica* ("hypnotic regression"), and *variáveis macroeconômicas* ("macroeconomic variables") are exclusive to ptTenTen11. In addition, although there are collocates shared by the two corpora, again their proportional frequencies and logDice values in general are very different according to the corpus. For illustrative purposes, the following shared collocates can be given (percentages and logDice values within parentheses): *apresentaram* (“presented”) before *significância* (ptRAS: 9.62%, 10.39; ptTenTen11: 0.87%, 4.05), *estatística* (“statistical” or “statistics”) after *significância* (ptRAs: 17.31%, 11.66; ptTenTen11: 22.27%, 9.03), *múltipla* (“multiple”) after *regressão* (ptRAs: 11.11%, 11.57; ptTenTen11: 2.36%, 8.15), *as* (“the”) before *variáveis* (ptRAs: 41.38%, 10.66; ptTenTen11: 24.66%, 4.26), and *independentes* (“independent”) after *variáveis* (ptRAs: 9.58%, 11.42; ptTenTen11: 1.33%, 7.54). It is noteworthy that (unexpectedly) not necessarily a higher proportional frequency coincides with a higher logDice value, as the case of *estatística* demonstrates.

Some collocations seem to be above genre-specificity, since the pertinent logDice values are (relatively) high in both corpora. Examples are *significância estatística* (“statistical significance”), *regressão linear* (“linear regression”), and *variáveis independentes* (“independent variables”). They are all likely to be general collocations.

It is also noteworthy that while in ptRAs *regressão* is typically preceded by prepositions such as *mediante* (“by means of”) and *pela* (“through”), in ptTenTen11 it is typically preceded by participle verbs and adjectives, for example *acelerada* (“accelerated”), *brutal* (“brutal”), and *profunda* (“deep”). Many collocates in one and another corpus are likely to refer to different senses of *regressão* (e.g., the statistical sense and the social sense), which provides support for Hoey’s (2005) hypothesis on polysemous words, according to which collocates differ depending on the sense of the word.

The left and right collocates of 本論文 (*honronbun*, “this paper”), 本稿 (*honkō*, “this paper”), and 症例 (*shōrei*, “clinical case”) are presented in Tables 7.19, 7.20, and 7.21 on the following pages.

Table 7.19 – Manually identified collocates of 本論文 (*honronbun*, "this paper") in two corpora.

Collocate	jaRAs		jaTenTen11 LUW	
	Instances	PER	Instances	PER
Left (WORD + 本論文)				
そこで	2	5.00	2	2.82
が	2	5.00	1	1.41
基づき	2	5.00	0	0.00
は	1	2.50	5	7.04
ように	0	0.00	2	2.82
Right (本論文 + WORD)				
では	18	45.00	23	32.39
の	7	17.50	10	14.08
が	5	12.50	0	0.00
は	2	5.00	26	36.62
において	2	5.00	0	0.00
で	0	0.00	4	5.63
における	0	0.00	3	4.23
を	0	0.00	3	4.23

Source: Prepared by the author. PER = percentage of instances in relation to the overall number of occurrences of 本論文 in the corpus. Includes only collocates appearing at least two times in one of the two corpora. Shared collocates are bounded by colored rectangles. The double underline indicates a verb with partial inflection. An English translation of this table is provided in Appendix A together with a Romanized transcription of the Japanese words.

Table 7.20 – Manually identified collocates of 本稿 (*honkō*, “this paper”) in two corpora.

Collocate	jaRAs		jaTenTen11 LUW	
	Instances	PER	Instances	PER
Left (WORD + 本稿)				
が	22	6.98	20	5.14
そこで	12	3.81	19	4.88
は	12	3.81	10	2.57
ため	12	3.81	2	0.51
なお	10	3.17	11	2.83
また	7	2.22	1	0.26
を	6	1.90	4	1.03
で	6	1.90	2	0.51
に	6	1.90	1	0.26
では	5	1.59	2	0.51
しかし	5	1.59	1	0.26
として	5	1.59	1	0.26
よって	5	1.59	1	0.26
踏まえ	4	1.27	0	0.00
の	1	0.32	5	1.29
ので	0	0.00	5	1.29
Right (本稿 + WORD)				
では	143	45.40%	226	58.10%
の	79	25.08%	36	9.25%
は	30	9.52%	49	12.60%
で	30	9.52%	25	6.43%
が	8	2.54%	3	0.77%
における	8	2.54%	0	0.00%
での	5	1.59%	4	1.03%
を	1	0.32%	20	5.14%
に	1	0.32%	5	1.29%
においては	0	0.00%	4	1.03%

Source: Prepared by the author. PER = percentage of instances in relation to the overall number of occurrences of 本稿 in the corpus. Includes only collocates appearing at least four times in one of the two corpora. Shared collocates are bounded by colored rectangles. The double underline indicates a verb with partial inflection. An English translation of this table is provided in Appendix A together with a Romanized transcription of the Japanese words.

Table 7.21 – Manually filtered collocates of 症例 (*shōrei*, "clinical case") in two corpora.

Collocate	jaRAs		jaTenTen11 LUW	
	Instances	PER	Instances	PER
Left (WORD + 症例)				
の	56	18.36	346	29.73
この	5	1.64	40	3.44
要する	5	1.64	0	0.00
有する	5	1.64	0	0.00
超える	4	1.31	0	0.00
難しい	0	0.00	17	1.46
珍しい	0	0.00	7	0.60
高い	0	0.00	7	0.60
で	0	0.00	5	0.43
を	0	0.00	5	0.43
同じ	0	0.00	4	0.34
その	0	0.00	4	0.34
や	0	0.00	4	0.34
Right (症例 + WORD)				
は	45	14.75	102	8.76
が	29	9.51	153	13.14
を	22	7.21	161	13.83
の	21	6.89	108	9.28
1	21	6.89	23	1.98
も	14	4.59	58	4.98
に対して	13	4.26	11	0.95
4	9	2.95	0	0.00
や	8	2.62	12	1.03
2	7	2.30	13	1.12
3	6	1.97	0	0.00
5	5	1.64	0	0.00
7	5	1.64	0	0.00
から	0	0.00	18	1.55
など	0	0.00	13	1.12
として	0	0.00	9	0.77
に対し	0	0.00	6	0.52
における	0	0.00	4	0.34

Source: Collocates identified with Sketch Engine and filtered by the author. PER = percentage of instances in relation to the overall number of occurrences of 症例 in the corpus. Includes only collocates appearing at least four times in one of the two corpora. Shared collocates are bounded by colored rectangles. An English translation of this table is provided in Appendix A together with a Romanized transcription of the Japanese words.

As can be seen, although there are many exclusive collocates, the whole picture is fundamentally different from the English and the Portuguese results. そこで (*soko de*, “then”) before 本論文 (*honronbun*), では (*de wa*, “at,” “in”) after 本論文 (*honronbun*), が (*ga*, mainly indicates the subject) before 本稿 (*honkō*), の (*no*, “of”) after 本稿 (*honkō*), の (*no*, “of”) before 症例 (*shōrei*), and に対して (*ni taishite*, “in respect to,” “while”) after 症例 (*shōrei*), for example, are all combinations found in both jaRAs and jaTenTen11 LUW. As to whether the Japanese language may be less sensible to genre-specificity than English and Japanese, however, the answer is not clear now.

Indeed, most common collocates of the search items are functional Japanese words known as *particles*. A particle can be defined as “a non-conjugating part of speech, bearing an absolute minimum of independent meaning, which attaches itself to other parts of speech and thereby places them in context” (Chino, 2008, p. 7). Japanese nouns like the selected keywords will be surrounded by particles very often, because in many cases they cannot be immediately preceded or followed by verbs or other nouns. For example, in the Japanese sentence 本論文は，上の問題を明らかにすることを目的としている (*honronbun wa, ue no mondai o akiraka ni suru koto o mokuteki to shiteiru*, “this paper aims to clear up the problem above”), 本論文 must be followed by the particle は (*wa*, which indicates the subject in the example). The shared collocates found in the three tables represent the general behavior of the three selected Japanese keywords, above genre-specificity.

Collectively, the results shown by Tables 7.19, 7.20, and 7.21 have led to two important changes for collocational analysis in this study. The first change was to increase the range in software-assisted searches for collocates, so that the Japanese particle barrier could be jumped. Thus, the focus on immediate collocates was restricted to the initial part of this study. The second change was to consider clusters as a possible supplement to collocates for the study of collocations. Software-assisted searches for clusters may reveal additional aspects on collocation, for example nesting (Hoey, 2005).

After identifying and studying their collocates, the semantic sets of the nine keywords were analyzed. As explained in the previous chapter, the semantic sets were manually compiled from collocate lists generated by Sketch Engine. Specifically, the sets were compiled from the first 50 strongest collocates (whenever the total of collocates was above 50) within a range of five words before and five words after the keyword (-5 to +5). Numbers and mathematical symbols were considered in this stage; punctuation marks were rejected. Polysemous and context-dependent words (most of the collocates) were included in a given set only after their prevalent senses were examined in the concordance lines.

The semantic sets of *hypothesis*, *variance*, and *variables* can be found in Tables 7.22, 7.23, and 7.24, together with their collocates.

Table 7.22 – Semantic sets of *hypothesis* in two corpora.

Corpus	SET (Collocates)
enRAs	ACTION AS TARGET (*proposed*, *rejected*, *supported*, *supporting*, *supports*, *test*, *tested*, *testing*)
	CAUSE OR EXPLANATION CONNECTORS (*hence*, *thus*)
	INDIVIDUAL (*altruistic*, *females*, *individuals*, *males*, *within-individual*)
	NUMERICAL LABELS (*1*, *2*, *3*, *4*, *5*, *6*, *1a*, *1b*, *1c*, *2a*, *2b*, *H1*, *H2*, *H3*)
	STATISTICS (*dummy*, *hypotheses*, *hypothesize*, *model*, *null*, *prediction*, *predicts*, *relationship*, *tests*)
enTenTen15	ACTION AS AGENT (*explain*, *posits*, *predicts*, *proposes*, *suggests*)
	ACTION AS TARGET (*confirm*, *confirmed*, *confirms*, *disprove*, *formulate*, *formulated*, *propose*, *prove*, *refute*, *reject*, *rejected*, *supported*, *supports*, *test*, *tested*, *validate*)
	ATTRIBUTES OF HYPOTHESES (*alternative*, *correct*, *null*, *plausible*, *testable*)
	RESEARCH METHODOLOGY (*empirical*, *evidence*, *experiment*, *findings*, *formulation*, *theory*)
	STATISTICS (*estimation*, *experiments*, *hypotheses*, *intervals*, *probability*, *regression*, *statistical*, *tests*, *validity*)
	WELL-KNOWN HYPOTHESES (*amyloid*, *continuum*, *Gaia*, *hygiene*, *phylogenetic*, *Riemann*, *Sapir-Whorf*)

Source: Semantic sets manually identified from collocate lists generated by Sketch Engine. Shared sets and collocates are indicated with colored boxes.

Table 7.23 – Semantic sets of *variance* in two corpora.

Corpus	SET (Collocates)
enRAs	$\boxed{\text{ACTION AS TARGET}}$ ($\boxed{\textit{explained}}$, *shared*)
	AMOUNT (*amount*, *cent*, *per*, *percent*, *ratio*, *total*)
	MATHEMATICAL SYMBOLS AND ABBREVIATIONS (*AVE*, *R2*, *VIF*, *VIFs*, *xj*, *ε*, *σ*)
	PERCENTAGE (*2.8*, *52.3*)
	$\boxed{\text{STATISTICAL TESTS}}$ ($\boxed{\textit{Anova}}$, $\boxed{\textit{Kruskal-Wallis}}$, $\boxed{\textit{Manova}}$, *test*, *Ward's*)
	$\boxed{\text{STATISTICS}}$ ($\boxed{\textit{analysis}}$, *equals*, *error*, $\boxed{\textit{estimates}}$, *factor*, *inflation*, $\boxed{\textit{mean}}$, *method*, *multicollinearity*, *statistic*, *summation*)
	$\boxed{\text{TYPES OF VARIANCE}}$ (*average*, *conditional*, *between-individual*, *common-method*, *within-individual*, *minimum*)
enTenTen15	$\boxed{\text{ACTION AS TARGET}}$ (*accounted*, $\boxed{\textit{explained}}$, *sought*)
	$\boxed{\text{STATISTICAL TESTS}}$ ($\boxed{\textit{Anova}}$, *chi-square*, $\boxed{\textit{Kruskal-Wallis}}$, $\boxed{\textit{Manova}}$, *t-test*, *t-tests*)
	$\boxed{\text{STATISTICS}}$ (*analyses*, $\boxed{\textit{analysis}}$, *bias*, *coefficient*, *correlation*, *covariance*, *decomposition*, *deviation*, *estimate*, $\boxed{\textit{estimates}}$, *estimation*, *estimator*, *estimators*, *factorial*, *Gaussian*, *homogeneity*, *inverse*, $\boxed{\textit{mean}}$, *multivariate*, *nonparametric*, *proportion*, *regression*, *repeated-measures*, *skewness*, *univariate*, *variables*, *variances*)
	$\boxed{\text{TYPES OF VARIANCE}}$ (*asymptotic*, *zoning*, *phenotypic*, *yard*, *conditional*, *additive*, *rear*, *setback*, *genetic*, *residual*)

Source: Semantic sets manually identified from collocate lists generated by Sketch Engine. Shared sets and collocates are indicated with colored boxes. AVE = average variance extracted. VIF = variance inflation factor.

Table 7.24 – Semantic sets of *variables* in two corpora.

Corpus	SET (Collocates)
enRAs	ACTION AS AGENT (*measuring*, *showed*) ACTION AS TARGET (*included*, *lagged*) AMOUNT (*four*, *three*, *two*) COLLECTION (*number*, *set*) GROUP-RELATED PREPOSITIONS (*among*, *between*, *including*) SPECIFICATION (*all*, *other*, *these*) STATISTICS (*analysis*, *bivariate*, *correlations*, *factors*, *model*, *regression*, *relationships*, *significant*, *variable*) TYPE OF VARIABLE (*categorical*, *continuous*, *control*, *demographic*, *dependent*, *dummy*, *explanatory*, *independent*, *latent*, *outcome*, *parenting*, *psychosocial*, *style*)
enTenTen15	ACTION AS TARGET (*define*, *measured*) STATISTICS (*constants*, *correlation*, *correlations*, *distributions*, *equation*, *equations*, *functions*, *linear*, *multivariate*, *parameters*, *probability*, *regression*, *relationships*, *statistical*, *values*, *variable*) TYPE OF VARIABLE (*categorical*, *climatic*, *confounding*, *contextual*, *continuous*, *demographic*, *dependent*, *discrete*, *dummy*, *environment*, *environmental*, *exogenous*, *explanatory*, *independent*, *input*, *instrumental*, *latent*, *macroeconomic*, *meteorological*, *outcome*, *predictor*, *random*, *sociodemographic*, *socio-demographic*, *socioeconomic*, *static*)

Source: Semantic sets manually identified from collocate lists generated by Sketch Engine. Shared sets and collocates are indicated with colored boxes.

As can be seen, there are two shared sets for *hypothesis* (ACTION AS TARGET and STATISTICS), four for *variance* (ACTION AS TARGET, STATISTICAL TESTS, STATISTICS, and TYPES OF VARIANCE), and three for *variables* (ACTION AS TARGET, STATISTICS, and TYPE OF VARIABLE). Most collocates that make up the sets, however, are exclusive to one or another corpus. Exceptions are collocates belonging to ACTION AS TARGET for *hypothesis*, STATISTICAL TESTS for *variance*, and TYPE OF VARIABLE for *variables*.

Collectively, the semantic sets of the English keywords demonstrate that while a given keyword may be typical of a certain genre, even though there are still semantic associations and collocations that will not be restricted to the genre. Semantic associations between *hypothesis* and STATISTICS, *variance* and STATISTICAL TESTS, *variables* and TYPE OF VARIABLE, as well as collocations between *hypothesis* and *rejected*, *variance* and *Anova*, *variables* and *independent*, illustrate this point.

It is noteworthy that the semantic sets of the English keywords reveal differences in subject matter between the specialized and the non-specialized corpora. In enTenTen15, a multi-genre, general corpus, *hypothesis* is associated with WELL-KNOWN HYPOTHESES, which refers to combinations such as *the continuum hypothesis*, *the hygiene hypothesis*, and *the Riemann hypothesis* that are likely to be part of a less specialized discourse. In enRAs, *variance* is associated with AMOUNT, which refers to collocations involving words such as *percent*, *ratio*, and *total* that might relate to empirical research reports.

The semantic sets of the Portuguese keywords *significância* ("significance"), *regressão* ("regression"), and *variáveis* ("variables") are shown with their collocates in Tables 7.25, 7.26, and 7.27.

Table 7.25 – Semantic sets of *significância* ("significance") in two corpora.

Corpus	SET (Collocates)
ptRAs	ACTION AS TARGET (*apresentaram*, *perderam*) NUMBERS AND MATHEMATICAL SYMBOLS (*5*, *95*, *%*, *<*, *0,05*, *p*, *α=0,05*) STATISTICS (*estatística*, *intervalo*, *modelo*, *nível*, *proporções*, *variáveis*)
ptTenTen11	ATTRIBUTES (*estatística*, *limítrofe*, *prognóstica*) NUMBERS AND MATHEMATICAL SYMBOLS (*0.05*, *<*, *≤*, *0,000*, *0,001*, *0,01*, *0,05*, *0,050*, *fz*, *p*, *p=0,05*, *α*) STATISTICAL TESTS (*Anova*, *Bonferroni*, *Fisher*, *Kruskal-Wallis*, *Mann-Whitney*, *não-paramétrico*, *pareado*, *Pearson*, *qui-quadrado*, *Student*, *Tukey*, *Wilcoxon*) STATISTICS (*bivariada*, *coeficientes*, *correlação*, *correlações*, *estatisticamente*, *estatístico*, *estatísticos*, *multivariada*, *nível*, *presultados*, *p-valor*, *regressão*, *Spearman*, *SPSS*, *univariada*, *valor-p*, *variância*, *variáveis*)

Source: Semantic sets manually identified from collocate lists generated by Sketch Engine. Shared sets and collocates are indicated with colored boxes. An English translation of this table is provided in Appendix A.

Table 7.26 – Semantic sets of *regressão* ("regression") in two corpora.

Corpus	SET (Collocates)
ptRAs	MEANS (*mediante*, *meio*, *métodos*, *pela*, *usada*, *utilizando*) NUMBERS AND MATHEMATICAL SYMBOLS (*5*, *F*) RESOURCES (*conhecimentos*, *software*) STATISTICS (*análise*, *análises*, *coeficiente*, *coeficientes*, *DEA*, *equação*, *estimativas*, *modelo*, *modelos*, *qui-quadrado*, *resultados*, *robustas*, *significativa*, *significativas*, *variâncias*) TYPE OF REGRESSION (*ajustada*, *binária*, *Cox*, *linear*, *logística*, *moderada*, *multinível*, *multinomial*, *múltipla*, *multivariada*, *Poisson*, *RMM*, *univariada*)
ptTenTen11	MEDICINE (*edema*, *lúteo*, *neurais*, *progressão*, *tumor*) PSYCOTHERAPY (*hipnose*, *passadas*, *terapia*, *TVP*) STATISTICS (*análise*, *Anova*, *coeficiente*, *coeficientes*, *correlação*, *discriminante*, *equação*, *equações*, *estatística*, *estimação*, *estimadores*, *estimar*, *inferência*, *lineares*, *modelos*, *recta*, *qui-quadrado*, *significância*, *variância*, *variáveis*, *variável*) TYPE OF REGRESSION (*binária*, *civilizacional*, *Cox*, *espontânea*, *hipnótica*, *infinita*, *linear*, *logística*, *multinomial*, *múltipla*, *multivariada*, *não-linear*, *Poisson*, *polinomial*, *quadrática*, *stepwise*, *tumoral*, *univariada*)

Source: Semantic sets manually identified from collocate lists generated by Sketch Engine. Shared sets and collocates are indicated with colored boxes. An English translation of this table is provided in Appendix A.

Table 7.27 – Semantic sets of *variáveis* ("variables") in two corpora.

Corpus	SET (Collocates)
ptRAs	ACTION AS TARGET (*analisadas*, *consideradas*, *utilizadas*) AMOUNT (*duas*, *três*) STATISTICS (*absolutas*, *análise*, *associação*, *bivariada*, *frequências*, *mensuração*, *modelo*, *modelos*, *padrão*, *relações*, *significância*, *significativa*, *significativas*) TYPE OF VARIABLE (*categóricas*, *confusão*, *controle*, *dependentes*, *endógenas*, *exógenas*, *independentes*, *interesse*, *maternas*, *numéricas*, *primárias*, *socioeconômicas*)
ptTenTen11	ACTION AS AGENT (*afetam*, *influenciam*, *interferem*, *determinam*) ACTION AS TARGET (*analisadas*, *avaliadas*, *consideradas*, *controladas*, *estudadas*, *selecionadas*, *utilizadas*) PERTINENT (*associadas*, *envolvidas*, *relacionadas*) STATISTICS (*correlação*, *correlações*, *equação*, *equações*, *estatisticamente*, *regressão*, *significativas*, *variável*) TYPE OF VARIABLE (*aleatórias*, *ambientais*, *antropométricas*, *categóricas*, *climáticas*, *contínuas*, *controláveis*, *demográficas*, *dependentes*, *discretas*, *econômicas*, *explicativas*, *externas*, *fisiológicas*, *globais*, *independentes*, *latentes*, *macroeconômicas*, *meteorológicas*, *psicológicas*, *qualitativas*, *quantitativas*, *sociodemográficas*, *sócio-demográficas*, *socioeconômicas*)

Source: Semantic sets manually identified from collocate lists generated by Sketch Engine. Shared sets and collocates are indicated with colored boxes. An English translation of this table is provided in Appendix A.

As can be noted, there are two shared sets for *significância* (NUMBERS AND MATHEMATICAL SYMBOLS and STATISTICS), two for *regressão* (STATISTICS and TYPE OF REGRESSION), and three for *variáveis* (ACTION AS TARGET, STATISTICS, and TYPE OF VARIABLE). While most collocates are exclusive to one or another corpus, there are many collocates of *regressão* that can be found in both ptRAs and ptTenTen11, for example *análise* ("analysis"), *coeficiente* ("coefficient"), *binária* ("binary"), and *múltipla* ("multiple").

Collectively, the semantic sets of the Portuguese keywords converge with those of the English keywords. They demonstrate that while a given keyword may be characteristic of a particular genre, there are still semantic associations and collocations that will not be exclusive to the genre. Semantic associations between *significância* and STATISTICS, *regressão* and TYPE OF REGRESSION, *variáveis* and TYPE OF VARIABLE, as well as collocations between *significância* and *nível* ("level"), *regressão* and *análise* ("analysis"), *variáveis* and *dependentes* ("dependent"), illustrate this point.

Yet, because the number of shared sets and collocates between ptRAs and ptTenTen11 is relatively low, it may be the case that RAs in Portuguese reflect a greater degree of genre-specificity in comparison to English RAs. Such an assumption, however, should be explored in further research.

The semantic sets of the Japanese keywords 本論文 (*honronbun*, "this paper"), 本稿 (*honkō*, "this paper"), and 症例 (*shōrei*, "clinical case") are displayed in Tables 7.28, 7.29, and 7.30, together with their respective collocates.

Table 7.28 – Semantic sets of 本論文 (*honronbun, "this paper"*) in two corpora.

Corpus	SET (Collocates)
jaRAs	ARTICLE CONTENT (問題, 目的, 背景, 課題, 限界) AUTHOR'S (PAPER'S) ACTIONS (基づき, 申告す[べき], 考察する) DEMONSTRATIVE FAMILY (この, そこ, その) SECTION NUMBER (*1*, *3*) SOCIETY (イノベーション, 実践共同体, 社会的側面, 組織) THOUGHT (見方, 問題意識)
jaTenTen11 LUW	ARTICLE CONTENT (内容, 提案, 結論) ATTRIBUTES (新しい, 重要)

Source: Semantic sets manually identified from collocate lists generated by Sketch Engine. All the collocates of the Japanese RAs corpus were considered (41); only collocates with positive logDice values of jaTenTen11 LUW were considered (16, in total). Shared sets and collocates are indicated with colored boxes. Verbs with partial inflection are double underlined. An English translation of this table is provided in Appendix A together with a Romanized transcription of the Japanese words.

Table 7.29 – Semantic sets of 本稿 (*honkō,* "this paper") in two corpora.

Corpus	SET (Collocates)
jaRAs	ARTICLE CONTENT (はじめ, 分析, 事例, 論点, 限界, 点, 分析結果, 課題, 先行研究)
	AUTHOR'S (PAPER'S) ACTIONS (分析する, 取り上げ, 提示し, 用い, 考える, 踏まえ)
	DEMONSTRATIVE FAMILY (この, これ, これら, そこ)
	INTERNAL REFERENCE (以上, 以下)
	MANAGEMENT (企業家的志向, 新規事業開発, 戦略グループ)
	SECTION NUMBERS (*1*, *5*)
	RESEARCH METHODS (問題意識, 実証分析, 対象, 焦点, 研究)
jaTenTen11 LUW	ARTICLE CONTENT (主題, 記述, 注, コンセプト, 経緯, 趣旨)
	AUTHOR'S (PAPER'S) ACTIONS (ご紹介し, 取り上げ, 執筆し, 紹介し, 記す, 論じ, 述べ, 述べる)
	FOCUS (主, 主に, 焦点, 注目)
	MANAGEMENT (中小企業, 企業, 同社, 開発)
	WRITING (前編, 執筆, 執筆時点, 筆者)

Source: Semantic sets manually identified from collocate lists generated by Sketch Engine. Shared sets and collocates are indicated with colored boxes. Verbs with partial inflection are double underlined. An English translation of this table is provided in Appendix A together with a Romanized transcription of the Japanese words.

Table 7.30 – Semantic sets of 症例 (*shōrei,* “clinical case”) in two corpora.

Corpus	SET (Collocates)
jaRAs	CASE DESCRIPTION (有する, 要し, 要する, 超える)
	CASE NUMBERS (*1*, *2*, *3*, *4*, *5*, *7*, *8*, *9*)
	DIAGNOSIS (疑わ, 発見さ, 認め)
	HEALTH ISSUES (本症, 来し, 気道異物, 症状, 発症し, 経験し, 食道閉鎖症)
	NUMBER OF CASES (1 例, 2 例)
	TREATMENT (保存的治療, 手術, 施行さ, 施行し, 治療, 行っ)
jaTenTen11 LUW	DIAGNOSIS (レントゲン, 確認さ, 診断, 診断さ)
	HEALTH ISSUES (記憶喪失, 先天異常症候群, 再発, 基礎疾患, 塞栓症, 急性散在性脳脊髄炎, 悪性腫瘍, 感染する, 特発性間質性肺炎, 眼球運動障害, 重篤)
	REPORT (報告さ, 報告し, 報告する, 提示し, 提示する)
	TREATMENT (*OPCAB*; 前頭頬骨縫合, 受精率, 抜歯し, 整復操作, 施行し, 栄養さ, 治療し, 経口投与, 適応できる, 除痛)

Source: Semantic sets manually identified from collocate lists generated by Sketch Engine. Verbs with partial inflection are double underlined. An English translation of this table is provided in Appendix A together with a Romanized transcription of the Japanese words.

As can be seen, there is only one shared set for 本論文 (*honronbun*), ARTICLE CONTENT, two for 本稿 (*honkō*), that is, ARTICLE CONTENT and AUTHOR’S (PAPER’S) ACTIONS, and none for 症例 (*shōrei*). Moreover, shared collocates are limited to one item: 取り上げ (*toriage*, “select,” “pick up”) for 本稿 (*honkō*).

Although this is substantially different from the collocate lists shown before, it does make sense. The collocate lists presented in Tables 7.19, 7.20, and 7.21 were exclusively made from immediately left and right collocates, while the semantic sets were made with collocates within a range of five words to both the left and the right sides, overcoming thereby the particles that surround the selected keywords. In addition, semantic sets depend on meaning, consequently functional Japanese words such as particles—that bear “an absolute minimum of independent meaning” (Chino, 2008, p. 7)—were left out.

Altogether, the semantic sets of the Japanese keywords suggest that semantic associations and related collocations in Japanese RAs tend to be genre-specific, being different from general Japanese. The comparison between English, Portuguese, and Japanese therefore leads to the provisional conclusion that between the three languages there seems to be an increasing level of genre-specificity concerning the RA, with RAs in English with lower, RAs in Portuguese with moderate, and RAs in Japanese with higher specificity. As our data set is limited, however, such a conclusion should be investigated further.

Through the study of keywords, collocates (collocations) and semantic sets (semantic associations), this chapter has revolved around genre-specificity. Our guiding question was: To what extent is there evidence for genre-specificity related to psychological priming in RAs? Our aim was to assess the presence of genre-specific signs of psychological priming in RAs in English, Portuguese, and Japanese.

Overall, the keywords, collocations, and semantic associations shown in this chapter provide evidence for genre-specificity related to psychological priming. Although there are collocations and semantic associations shared by the specialized, single-genre and the non-specialized, multi-genre corpora, the whole picture tends to reveal distinct links between keywords, collocates, and semantic sets according to genre. Collectively, the results support Hoey's (2004, 2005, 2007b, 2013) claims and agree with both Pace-Sigge's (2007) and Chuang's (2015) conclusions. Alluding to Hoey's (2005) central hypotheses, this chapter's findings can be summed up in the following way:

Every word is primed to occur with particular other words and with particular semantic sets, and these (i.e., the particular other words and semantic sets) can be expected to vary across genres.

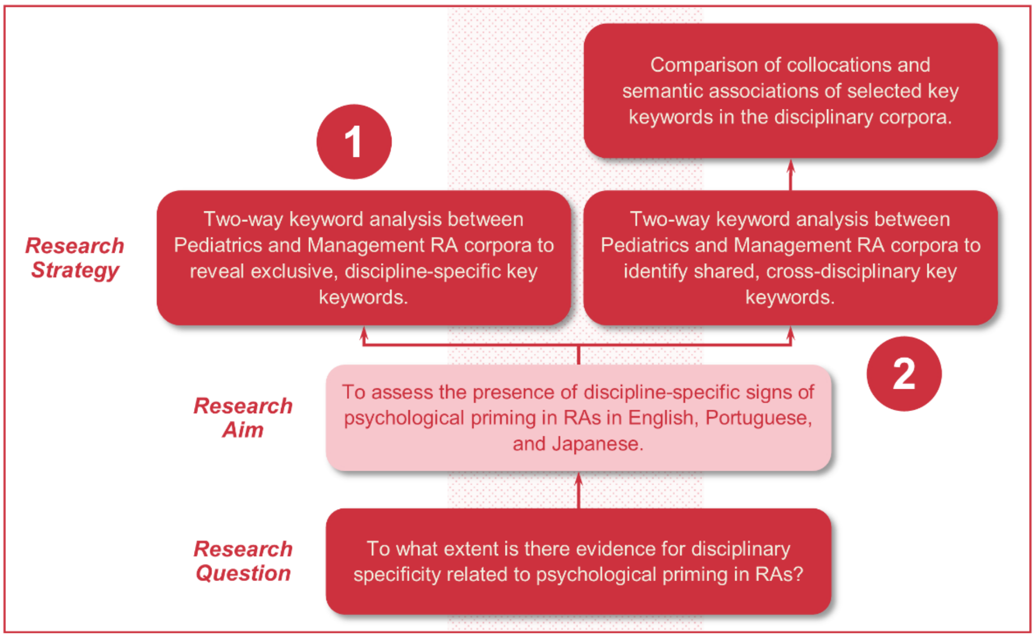
The main implications of the findings can be summarized as follows. First, they shed light on the relevance of adopting a genre-orientation for language learning and teaching, in line with the vast body of literature on genre and on the RA genre (Swales, 1990; Johns, 2003; Hyland, 2004; Rezende & Hemais, 2004; Lim, 2006; etc.). It can be assumed that not only the RA but also other genres such as the Master's thesis, the research project, and the conference paper have characteristic words, collocations, and semantic associations. Therefore, a genre-orientation can prime language learners to use words according to collective expectations, that is, established primings. Second, academic translation may also be benefited by a focus on genre. As a matter of fact, the usefulness of lexical priming for translation has already been acknowledged in previous works (Hoey, 2011; Salim, 2012; Shao, 2018; Li & Yang, 2017). Considering the findings, we understand that the combination of a lexical priming framework with a genre-orientation can possibly foster the development of a lexico-generic awareness that may contribute to naturalness in translation. A translator who is able to use genre-specific, typical word associations in the target language will possibly produce works that sound more natural to native speakers of that language.

8 PSYCHOLOGICAL PRIMING AND DISCIPLINARY VARIATION[1]

In addition to genre, domain has also been assumed as a decisive factor for psychological priming (Hoey, 2004, 2005, 2007b). Language users would be primed to write or speak with one set of linguistic resources or another depending on the domain in question. However, as shown in Chapter 3, even though lexical priming research has considered the broad domains from which language data come, the relationship between priming and specific fields of knowledge or disciplines has received little attention. To name but a few, religious texts (Salim, 2012), advertising texts (Cunha, 2017), politics texts (Duguid & Partington, 2017), and popular science texts (Hoey, 2017) were all investigated under the lens of lexical priming theory. Nevertheless, little consideration has been given by priming analysts to linguistic data from different disciplines, such as Biology, History, and Chemistry. The second stage of the present study addressed this gap, with a focus on Pediatrics and Management, the two disciplines to which the selected RAs belong. As Figure 7.1 in the previous chapter, Figure 8.1 on the next page provides the guiding elements and an outline of the research procedures adopted in this stage.

[1] Part of the content of this chapter has been presented in Aragão (2022b). However, while this chapter covers data in English, Portuguese, and Japanese, the cited article deals exclusively with data in Portuguese.

Figure 8.1 – Guiding elements and research strategy of the second stage of the present study.



Source: Designed by the author.

Two series of keyword analysis were performed. Both were composed of the same six software-assisted comparisons in the following order: (1) enPED versus enMGT, (2) enMGT versus enPED, (3) ptPED versus ptMGT, (4) ptMGT versus ptPED, (5) jaPED versus jaMGT, and (6) jaMGT versus jaPED. While the first series aimed to reveal disciplinary specificity through exclusive keywords, the second aimed to identify common keywords whose collocational and semantic associational behavior might differ depending on the discipline. Because the six corpora have a high degree of specialization, with many lexical items limited to but a single RA, the presence in at least 30% of the data was adopted as a threshold to distinguish key keywords from non-key keywords in the first series. In the second, the threshold percentage was increased to 50% because the change in focus led to a higher number of keywords, so it was necessary to increase the threshold to keep the number of identified keywords to a manageable size.

The discipline-specific key keywords identified through the comparisons between enPED and enMGT are shown in Tables 8.1 and 8.2.

Table 8.1 – Key keywords in enPED against enMGT.

POS	Word	General Freq. FC	General Freq. RC	Freq. per Million FC	Freq. per Million RC	RA Freq. n	RA Freq. (%)	Score
1	*infants*	522	0	1,738.27	0.00	30	41.67	1,739.3
2	*adolescents*	372	0	1,238.77	0.00	26	36.11	1,239.8
3	*mortality*	297	0	989.01	0.00	24	33.33	990.0
4	*infant*	183	0	609.39	0.00	22	30.56	610.4
5	*prevention*	81	0	269.73	0.00	31	43.06	270.7
6	*parents*	534	5	1,778.23	7.16	47	65.28	217.9
7	*diseases*	61	0	203.13	0.00	24	33.33	204.1
8	*CI*	339	4	1,128.87	5.73	36	50.00	167.9

Source: Keyword analysis performed with Sketch Engine. FC = focus corpus. RC = reference corpus. RA Freq. = number (n) and percentage (%) of RAs containing the word. CI = confidence interval.

Table 8.2 – Key keywords in enMGT against enPED.

POS	Word	General Freq. FC	General Freq. RC	Freq. per Million FC	Freq. per Million RC	RA Freq. n	RA Freq. (%)	Score
1	*business*	902	0	1,292.38	0.00	53	73.61	1,293.4
2	*organizational*	898	0	1,286.65	0.00	41	56.94	1,287.6
3	*firms*	758	0	1,086.05	0.00	42	58.33	1,087.1
4	*innovation*	583	0	835.32	0.00	23	31.94	836.3
5	*company*	389	0	557.36	0.00	48	66.67	558.4
6	*organisational*	327	0	468.52	0.00	22	30.56	469.5
7	*employee*	310	0	444.16	0.00	28	38.89	445.2
8	*customer*	297	0	425.54	0.00	27	37.50	426.5
9	*customers*	277	0	396.88	0.00	34	47.22	397.9
10	*competitive*	196	0	280.83	0.00	35	48.61	281.8
11	*corporate*	195	0	279.39	0.00	26	36.11	280.4
12	*theories*	124	0	177.67	0.00	28	38.89	178.7
13	*leadership*	523	1	749.35	3.33	26	36.11	173.3
14	*markets*	119	0	170.50	0.00	29	40.28	171.5
15	*scholars*	118	0	169.07	0.00	31	43.06	170.1
16	*marketing*	107	0	153.31	0.00	23	31.94	154.3
17	*managerial*	101	0	144.71	0.00	34	47.22	145.7
18	*employees*	706	2	1,011.55	6.66	47	65.28	132.2
19	*industry*	375	1	537.30	3.33	46	63.89	124.3
20	*identity*	580	2	831.02	6.66	22	30.56	108.6
21	*competitors*	74	0	106.03	0.00	23	31.94	107.0
22	*strategic*	544	2	779.44	6.66	32	44.44	101.9

Source: Keyword analysis performed with Sketch Engine. FC = focus corpus. RC = reference corpus. RA Freq. = number (n) and percentage (%) of RAs containing the word.

As can be seen, all the key keywords from both tables have keyness scores above 100.0, and the scores of the three top items reach or exceed 1,000.0. These numbers are well above those obtained from the analysis between enRAs, enTenTen15, and enTenTen20, which range from 34.7 to 62.9 (Tables 7.1 and 7.2), suggesting that domain or discipline may be a stronger factor than genre for lexical choices. It is also noteworthy that most keywords in both tables are exclusive to one or another corpus. This reinforces Hoey's (2004, 2005, 2007b) claims about domain-specificity.

The Portuguese key keywords identified by comparing ptPED and ptMGT are provided in Tables 8.3 and 8.4.

Table 8.3 – Key keywords in ptPED against ptMGT.

		General Freq.		Freq. per Million		RA Freq.		
POS	Word	FC	RC	FC	RC	n	(%)	Score
1	*mães*	116	0	1,405.74	0.00	11	45.83	1,406.7
2	*adolescentes*	106	0	1,284.55	0.00	10	41.67	1,285.6
3	*materno*	96	0	1,163.37	0.00	9	37.50	1,164.4
4	*alimentação*	78	0	945.24	0.00	8	33.33	946.2
5	*mãe*	52	0	630.16	0.00	8	33.33	631.2
6	*IC95*	50	0	605.92	0.00	10	41.67	606.9
7	*crianças*	321	1	3,890.01	5.69	20	83.33	581.2
8	*precoce*	28	0	339.32	0.00	10	41.67	340.3
9	*município*	28	0	339.32	0.00	8	33.33	340.3
10	*infantil*	27	0	327.20	0.00	10	41.67	328.2
11	*dentre*	23	0	278.72	0.00	11	45.83	279.7
12	*saudáveis*	22	0	266.61	0.00	8	33.33	267.6
13	*meninos*	22	0	266.61	0.00	8	33.33	267.6
14	*comitê*	20	0	242.37	0.00	19	79.17	243.4
15	*infância*	20	0	242.37	0.00	10	41.67	243.4
16	*apresentavam*	18	0	218.13	0.00	8	33.33	219.1
17	*clínica*	18	0	218.13	0.00	9	37.50	219.1

Source: Keyword analysis performed with Sketch Engine. FC = focus corpus. RC = reference corpus. RA Freq. = number (n) and percentage (%) of RAs containing the word. An English translation of this table is provided in Appendix B.

Table 8.4 – Key keywords in ptMGT against ptPED.

		General Freq.		Freq. per Million		RA Freq.		
POS	Word	FC	RC	FC	RC	n	(%)	Score
1	*organizacional*	213	0	1,213.01	0.00	10	41.67	1,214.0
2	*consumidores*	211	0	1,201.62	0.00	10	41.67	1,202.6
3	*empresa*	147	0	837.15	0.00	16	66.67	838.1
4	*atitude*	112	0	637.83	0.00	8	33.33	638.8
5	*consumidor*	74	0	421.42	0.00	8	33.33	422.4
6	*clientes*	72	0	410.03	0.00	9	37.50	411.0
7	*integração*	70	0	398.64	0.00	8	33.33	399.6
8	*financeiro*	58	0	330.30	0.00	12	50.00	331.3
9	*financeiros*	56	0	318.91	0.00	12	50.00	319.9
10	*percebido*	53	0	301.83	0.00	9	37.50	302.8
11	*organizacionais*	52	0	296.13	0.00	11	45.83	297.1
12	*eficiência*	45	0	256.27	0.00	15	62.50	257.3
13	*influencia*	44	0	250.58	0.00	13	54.17	251.6
14	*marketing*	43	0	244.88	0.00	11	45.83	245.9
15	*gerenciais*	41	0	233.49	0.00	9	37.50	234.5
16	*fluxo*	37	0	210.71	0.00	9	37.50	211.7
17	*equações*	33	0	187.93	0.00	9	37.50	188.9
18	*autor*	27	0	153.76	0.00	11	45.83	154.8

Source: Keyword analysis performed with Sketch Engine. FC = focus corpus. RC = reference corpus. RA Freq. = number (n) and percentage (%) of RAs containing the word. An English translation of this table is provided in Appendix B.

All the key keywords from these two additional tables also exhibit keyness scores above 100.0, and five items exceed 1,000.0. Most of the scores are higher than those observed in the comparison between ptRAs and ptTenTen11 (Table 7.3), which provides additional support for the assumption that domain or discipline may be more relevant than genre for lexical choices. Except for *crianças* ("children"), all the key keywords are exclusive to one or another corpus, which supports Hoey's (2004, 2005, 2007b) claims about domain-specificity.

The Japanese key keywords obtained from the comparison between jaPED and jaMGT are listed in Tables 8.5 and 8.6.

Table 8.5 – Key keywords in jaPED against jaMGT.

		General Freq.		Freq. per Million		RA Freq.		
POS	Word	FC	RC	FC	RC	n	(%)	Score
1	症例	305	0	4,659.26	0.00	23	95.83	4,660.3
2	本症	116	0	1,772.05	0.00	9	37.50	1,773.0
3	施行し	113	0	1,726.22	0.00	19	79.17	1,727.2
4	治療	81	0	1,237.38	0.00	16	66.67	1,238.4
5	手術	81	0	1,237.38	0.00	20	83.33	1,238.4
6	術後	63	0	962.41	0.00	15	62.50	963.4
7	自験例	54	0	824.92	0.00	15	62.50	825.9
8	症状	53	0	809.64	0.00	10	41.67	810.6
9	当科	49	0	748.54	0.00	12	50.00	749.5
10	当院	47	0	717.98	0.00	15	62.50	719.0
11	全例	46	0	702.71	0.00	16	66.67	703.7
12	患児	42	0	641.60	0.00	15	62.50	642.6
13	発症し	42	0	641.60	0.00	14	58.33	642.6
14	術式	35	0	534.67	0.00	9	37.50	535.7
15	合併症	35	0	534.67	0.00	15	62.50	535.7
16	小児	34	0	519.39	0.00	10	41.67	520.4
17	男児	32	0	488.84	0.00	13	54.17	489.8
18	本邦	31	0	473.56	0.00	13	54.17	474.6
19	後方視的	30	0	458.29	0.00	15	62.50	459.3
20	予後	29	0	443.01	0.00	10	41.67	444.0
21	診断	27	0	412.46	0.00	9	37.50	413.5
22	生後	25	0	381.91	0.00	8	33.33	382.9
23	III 結果	24	0	366.63	0.00	24	100.00	367.6
24	診断さ	24	0	366.63	0.00	10	41.67	367.6
25	II 対象	24	0	366.63	0.00	24	100.00	367.6
26	IV 考察	24	0	366.63	0.00	24	100.00	367.6
27	要旨	24	0	366.63	0.00	24	100.00	367.6
28	治療方針	24	0	366.63	0.00	9	37.50	367.6
29	新生児期	23	0	351.35	0.00	8	33.33	352.4
30	発症	23	0	351.35	0.00	9	37.50	352.4
31	報告する	20	0	305.53	0.00	16	66.67	306.5
32	疾患	20	0	305.53	0.00	11	45.83	306.5
33	施行する	17	0	259.70	0.00	9	37.50	260.7
34	児	15	0	229.14	0.00	8	33.33	230.1

Source: Keyword analysis performed with Sketch Engine. FC = focus corpus. RC = reference corpus. RA Freq. = number (n) and percentage (%) of RAs containing the word. Verbs with partial inflection are double underlined. Both a Romanized transcription and an English translation of this table are provided in in Appendix B.

Table 8.6 – Key keywords in jaMGT against jaPED.

		General Freq.		Freq. per Million		RA Freq.		
POS	Word	FC	RC	FC	RC	n	(%)	Score
1	企業	376	0	2,211.49	0.00	21	87.50	2,212.5
2	本稿	315	0	1,852.71	0.00	21	87.50	1,853.7
3	顧客	232	0	1,364.54	0.00	12	50.00	1,365.5
4	メンバー	209	0	1,229.26	0.00	8	33.33	1,230.3
5	資源	169	0	993.99	0.00	12	50.00	995.0
6	先行研究	144	0	846.95	0.00	19	79.17	848.0
7	事例	134	0	788.14	0.00	14	58.33	789.1
8	従業員	128	0	752.85	0.00	8	33.33	753.8
9	本研究	124	0	729.32	0.00	12	50.00	730.3
10	モデル	106	0	623.45	0.00	18	75.00	624.5
11	イノベーション	93	0	546.99	0.00	9	37.50	548.0
12	行動	91	0	535.23	0.00	13	54.17	536.2
13	既存研究	78	0	458.77	0.00	13	54.17	459.8
14	変数	77	0	452.89	0.00	12	50.00	453.9
15	戦略	76	0	447.00	0.00	15	62.50	448.0
16	のである	75	0	441.12	0.00	13	54.17	442.1
17	を通じて	69	0	405.83	0.00	21	87.50	406.8
18	高める	67	0	394.07	0.00	15	62.50	395.1
19	市場	67	0	394.07	0.00	14	58.33	395.1
20	意思決定	65	0	382.31	0.00	10	41.67	383.3
21	グループ	57	0	335.25	0.00	8	33.33	336.3
22	事業	52	0	305.84	0.00	11	45.83	306.8
23	プロセス	48	0	282.32	0.00	13	54.17	283.3
24	分析結果	47	0	276.44	0.00	11	45.83	277.4
25	分析し	47	0	276.44	0.00	13	54.17	277.4
26	論じ	47	0	276.44	0.00	11	45.83	277.4
27	日本企業	47	0	276.44	0.00	8	33.33	277.4
28	焦点	46	0	270.55	0.00	14	58.33	271.6
29	捉え	46	0	270.55	0.00	14	58.33	271.6
30	正	45	0	264.67	0.00	8	33.33	265.7
31	想定し	42	0	247.03	0.00	10	41.67	248.0
32	二つ	42	0	247.03	0.00	8	33.33	248.0
33	なぜ	42	0	247.03	0.00	15	62.50	248.0
34	第二	42	0	247.03	0.00	12	50.00	248.0
35	新しい	42	0	247.03	0.00	15	62.50	248.0
36	個人	42	0	247.03	0.00	11	45.83	248.0
37	そう	41	0	241.15	0.00	16	66.67	242.1
38	第一	41	0	241.15	0.00	13	54.17	242.1
39	着目し	39	0	229.38	0.00	13	54.17	230.4
40	高め	38	0	223.50	0.00	13	54.17	224.5
41	組織内	37	0	217.62	0.00	11	45.83	218.6
42	主張	37	0	217.62	0.00	9	37.50	218.6

Source: Keyword analysis performed with Sketch Engine. FC = focus corpus. RC = reference corpus. RA Freq. = number (n) and percentage (%) of RAs containing the word. Verbs with partial inflection are double underlined. Both a Romanized transcription and an English translation of this table are provided in Appendix B.

As can be seen, all the Japanese key keywords exhibit keyness scores above 200.0, and nine items exceed 1,000.0. The top items from the two tables, 症例 (*shōrei*, “clinical case”) and 企業 (*kigyō*, “business”), have impressive scores of 4,660.3 and 2,212.5, the highest among all the scores recorded during this entire study. The keyness score ranges are wider than that observed in the analysis between jaRAs and jaTenTen11 LUW (Table 7.5), which also offers support for the assumption that domain or discipline may be more decisive than genre for lexical choices. All the key keywords are exclusive to one or another corpus, thus reinforcing once again Hoey’s (2004, 2005, 2007b) claims of domain-specificity.

Collectively, the key keywords shown in Tables 8.1 to 8.6 demonstrate disciplinary specificity, especially because almost all of the keywords appear exclusively in either Pediatrics or Management data. Overall, keyness scores obtained through disciplinary comparison are higher than those obtained from the contrast between single-genre and multi-genre corpora (previous chapter), which suggests that domain or discipline may be stronger than genre as a factor for lexical choices, as already stated. Also, the results suggest that the three languages might have differing degrees of disciplinary specificity, with Japanese being more discipline-specific (or discipline-dependent) than English and Portuguese. Additional research, however, would be needed to confirm this hypothesis.

The non-specific key keywords shared by enPED and enMGT, by ptPED and ptMGT, and by jaPED and jaMGT are shown in Tables 8.7 to 8.12. Among them, six were selected for collocational and semantic associational analysis; these keywords are highlighted with colored boxes in the tables.

Table 8.7 – Highly distributed common keywords in enPED against enMGT.

		General Freq.		Freq. per Million		RA Freq. (%)		
POS	Word	FC	RC	FC	RC	FC	RC	Score
1	*were*	2,662	2,023	8,864.5	2,898.5	100.0	98.6	2.53
2	*age*	708	303	2,357.7	434.1	91.7	56.9	2.34
3	*p*	607	280	2,021.3	401.2	73.6	58.3	2.16
4	*risk*	445	161	1,481.9	230.7	76.4	51.4	2.02
5	*was*	2,276	2,410	7,579.1	3,453.0	100.0	100.0	1.93
6	*study*	1,461	1,514	4,865.2	2,169.2	100.0	100.0	1.85
7	*group*	532	409	1,771.6	586.0	73.6	77.8	1.75
8	*included*	361	202	1,202.1	289.4	95.8	72.2	1.71
9	*reported*	382	235	1,272.1	336.7	90.3	69.4	1.70
10	*associated*	476	374	1,585.1	535.9	79.2	84.7	1.68
11	*during*	461	382	1,535.1	547.3	88.9	83.3	1.64
12	*rate*	262	117	872.5	167.6	69.4	58.3	1.60
13	*at*	1,213	1,507	4,039.3	2,159.2	98.6	100.0	1.60
14	*compared*	267	131	889.1	187.7	79.2	59.7	1.59
15	*with*	2,963	4,227	9,866.8	6,056.4	100.0	100.0	1.54
16	*years*	440	426	1,465.2	610.4	86.1	97.2	1.53
17	*primary*	214	97	712.6	139.0	65.3	55.6	1.50
18	*no*	417	411	1,388.6	588.9	97.2	94.4	1.50
19	*who*	651	790	2,167.8	1,131.9	94.4	93.1	1.49
20	*population*	216	112	719.3	160.5	76.4	52.8	1.48
21	*after*	297	240	989.0	343.9	81.9	81.9	1.48
22	*received*	190	78	632.7	111.8	54.2	50.0	1.47
23	*studies*	470	526	1,565.1	753.6	87.5	97.2	1.46
24	*data*	713	918	2,374.3	1,315.3	100.0	98.6	1.46
25	*had*	597	770	1,988.0	1,103.2	95.8	88.9	1.42
26	*all*	656	870	2,184.5	1,246.5	100.0	100.0	1.42
27	*higher*	351	373	1,168.8	534.4	88.9	87.5	1.41
28	*respectively*	181	97	602.7	139.0	62.5	56.9	1.41
29	*mean*	266	247	885.8	353.9	83.3	70.8	1.39

Source: Keyword analysis performed with Sketch Engine. FC = focus corpus. RC = reference corpus. RA Freq. (%) = percentage of RAs containing the word. The colored rectangle indicates the selected keyword.

Table 8.8 – Highly distributed common keywords in enMGT against enPED.

POS	Word	General Freq.		Freq. per Million		RA Freq. (%)		Score
		FC	RC	FC	RC	FC	RC	
1	*et*	1,235	145	1,769.5	482.9	94.4	52.8	1.87
2	*research*	1,666	258	2,387.0	859.1	98.6	84.7	1.82
3	*that*	8,349	1,917	11,962.4	6,383.6	100.0	100.0	1.76
4	*they*	1,914	350	2,742.4	1,165.5	98.6	83.3	1.73
5	*can*	1,330	217	1,905.6	722.6	100.0	83.3	1.69
6	*it*	2,339	478	3,351.3	1,591.7	100.0	94.4	1.68
7	*is*	5,744	1,429	8,229.9	4,758.6	100.0	100.0	1.60
8	*its*	645	68	924.1	226.4	95.8	52.8	1.57
9	*our*	1,715	376	2,457.2	1,252.1	66.7	83.3	1.54
10	*are*	3,705	968	5,308.5	3,223.5	100.0	100.0	1.49
11	*their*	3,078	829	4,410.1	2,760.6	100.0	94.4	1.44
12	*we*	2,829	754	4,053.4	2,510.8	69.4	83.3	1.44
13	*model*	803	153	1,150.5	509.5	83.3	50.0	1.43
14	*them*	612	101	876.9	336.3	94.4	50.0	1.40
15	*do*	645	115	924.1	383.0	95.8	50.0	1.39
16	*knowledge*	614	111	879.7	369.6	81.9	58.3	1.37
17	*as*	5,577	1,677	7,990.7	5,584.4	100.0	100.0	1.37
18	*different*	815	179	1,167.7	596.1	100.0	75.0	1.36
19	*one*	1,162	290	1,664.9	965.7	100.0	88.9	1.36
20	*which*	1,765	490	2,528.9	1,631.7	100.0	98.6	1.34

Source: Keyword analysis performed with Sketch Engine. FC = focus corpus. RC = reference corpus. RA Freq. (%) = percentage of RAs containing the word. The colored rectangle indicates the selected keyword.

Table 8.9 – Highly distributed common keywords in ptPED against ptMGT.

		General Freq.		Freq. per Million		RA Freq. (%)		
POS	Word	FC	RC	FC	RC	FC	RC	Score
1	*anos*	179	85	2,169.2	484.1	91.7	95.8	2.14
2	*estudo*	364	292	4,411.1	1,662.9	100.0	100.0	2.03
3	*foram*	424	373	5,138.2	2,124.2	100.0	100.0	1.97
4	*foi*	578	540	7,004.4	3,075.2	100.0	100.0	1.96
5	*uso*	130	73	1,575.4	415.7	70.8	75.0	1.82
6	*associação*	94	47	1,139.1	267.7	79.2	62.5	1.69
7	*durante*	84	44	1,017.9	250.6	83.3	50.0	1.61
8	*maioria*	70	29	848.3	165.2	75.0	62.5	1.59
9	*média*	89	57	1,078.5	324.6	75.0	66.7	1.57
10	*presente*	84	59	1,017.9	336.0	87.5	70.8	1.51
11	*número*	84	62	1,017.9	353.1	91.7	70.8	1.49
12	*ou*	363	465	4,399.0	2,648.1	100.0	100.0	1.48
13	*teste*	66	39	799.8	222.1	54.2	58.3	1.47
14	*profissionais*	59	30	715.0	170.8	70.8	54.2	1.47
15	*qualidade*	94	88	1,139.1	501.2	62.5	70.8	1.43
16	*prática*	61	39	739.2	222.1	66.7	66.7	1.42
17	*atenção*	47	22	569.6	125.3	50.0	50.0	1.40
18	*dia*	54	35	654.4	199.3	62.5	50.0	1.38
19	*com*	914	1,364	11,076.2	7,767.8	100.0	100.0	1.38
20	*não*	477	692	5,780.5	3,940.9	100.0	100.0	1.37
21	*período*	73	66	884.6	375.9	79.2	62.5	1.37

Source: Keyword analysis performed with Sketch Engine. FC = focus corpus. RC = reference corpus. RA Freq. (%) = percentage of RAs containing the word. The colored rectangle indicates the selected keyword. An English translation of this table is provided in Appendix B.

Table 8.10 – Highly distributed common keywords in ptMGT against ptPED.

		General Freq.		Freq. per Million		RA Freq. (%)		
POS	Word	FC	RC	FC	RC	FC	RC	Score
1	*et*	452	45	2,574.1	545.3	95.8	62.5	2.31
2	*trabalho*	229	30	1,304.1	363.6	87.5	62.5	1.69
3	*suas*	201	24	1,144.7	290.8	100.0	50.0	1.66
4	*se*	631	153	3,593.5	1,854.1	100.0	95.8	1.61
5	*organização*	150	16	854.2	193.9	87.5	50.0	1.55
6	*valor*	162	23	922.6	278.7	83.3	50.0	1.50
7	*pesquisa*	347	82	1,976.1	993.7	95.8	87.5	1.49
8	*processo*	223	46	1,270.0	557.4	95.8	70.8	1.46
9	*efeito*	163	27	928.3	327.2	62.5	54.2	1.45
10	*assim*	237	51	1,349.7	618.0	95.8	83.3	1.45
11	*literatura*	132	20	751.7	242.4	95.8	58.3	1.41
12	*relação*	459	130	2,614.0	1,575.4	100.0	95.8	1.40
13	*base*	149	30	848.5	363.6	95.8	54.2	1.36
14	*esse*	272	73	1,549.0	884.6	95.8	91.7	1.35
15	*p*	181	42	1,030.8	509.0	79.2	54.2	1.35
16	*seu*	182	44	1,036.5	533.2	95.8	66.7	1.33

Source: Keyword analysis performed with Sketch Engine. FC = focus corpus. RC = reference corpus. RA Freq. (%) = percentage of RAs containing the word. The colored rectangle indicates the selected keyword. An English translation of this table is provided in Appendix B.

Table 8.11 – Highly distributed common keywords in jaPED against jaMGT.

		General Freq.		Freq. per Million		RA Freq. (%)		
POS	Word	FC	RC	FC	RC	FC	RC	Score
1	あっ	462	233	7,057.6	1,370.4	100.0	100.0	3.40
2	あり	226	225	3,452.4	1,323.4	100.0	100.0	1.92
3	方法	84	39	1,283.2	229.4	100.0	75.0	1.86
4	図	91	58	1,390.1	341.1	87.5	62.5	1.78
5	なかっ	147	146	2,245.6	858.7	100.0	100.0	1.75
6	のみ	103	81	1,573.5	476.4	75.0	87.5	1.74
7	検討し	72	35	1,099.9	205.9	95.8	66.7	1.74
8	行っ	141	146	2,154.0	858.7	91.7	91.7	1.70
9	た	2,301	3,613	35,150.7	21,250.3	100.0	100.0	1.63
10	より	218	325	3,330.2	1,911.5	100.0	100.0	1.49
11	に関して	55	50	840.2	294.1	70.8	66.7	1.42
12	で	2,319	4,194	35,425.7	24,667.5	100.0	100.0	1.42
13	うち	58	58	886.0	341.1	75.0	70.8	1.41
14	に対する	79	106	1,206.8	623.5	87.5	79.2	1.36
15	べき	52	55	794.4	323.5	91.7	70.8	1.36
16	および	81	111	1,237.4	652.9	75.0	83.3	1.35
17	なく	64	79	977.7	464.6	83.3	91.7	1.35
18	行う	89	128	1,359.6	752.8	79.2	87.5	1.35

Source: Keyword analysis performed with Sketch Engine. FC = focus corpus. RC = reference corpus. RA Freq. (%) = percentage of RAs containing the word. The colored rectangle indicates the selected keyword. Verbs with partial inflection are double underlined. Both a Romanized transcription and an English translation of this table are provided in Appendix B.

Table 8.12 – Highly distributed common keywords in jaMGT against jaPED.

POS	Word	General Freq. FC	General Freq. RC	Freq. per Million FC	Freq. per Million RC	RA Freq. (%) FC	RA Freq. (%) RC	Score
1	いう	702	60	4,128.9	916.6	100.00	62.50	2.68
2	よっ	347	26	2,040.9	397.2	100.00	50.00	2.18
3	この	611	90	3,593.7	1,374.9	100.00	83.33	1.93
4	よう	639	96	3,758.4	1,466.5	100.00	91.67	1.93
5	な	1,978	369	11,633.9	5,636.9	100.00	100.00	1.90
6	ば	352	44	2,070.3	672.2	100.00	62.50	1.84
7	う	237	25	1,393.9	381.9	95.83	50.00	1.73
8	か	568	106	3,340.8	1,619.3	100.00	75.00	1.66
9	なる	537	100	3,158.4	1,527.6	100.00	91.67	1.65
10	てき	217	34	1,276.3	519.4	100.00	70.83	1.50
11	られる	404	84	2,376.2	1,283.2	100.00	91.67	1.48
12	れる	684	162	4,023.0	2,474.8	100.00	100.00	1.45
13	これ	233	45	1,370.4	687.4	95.83	75.00	1.41
14	だ	138	20	811.7	305.5	95.83	62.50	1.39
15	こと	1,427	383	8,393.1	5,850.8	100.00	100.00	1.37

Source: Keyword analysis performed with Sketch Engine. FC = focus corpus. RC = reference corpus. RA Freq. (%) = percentage of RAs containing the word. The colored rectangle indicates the selected keyword. Both a Romanized transcription and an English translation of this table are provided in Appendix B.

As can be seen, the six key keywords chosen are the following: *study*, *research*, *estudo* ("study"), *pesquisa* ("research"), 行う (*okonau*, "perform"), and この (*kono*, "this"). *Study*, *research*, *estudo*, and *pesquisa* were chosen due to their high frequency and wide distribution across the corpora and because together they form an interesting set for comparison, both cross-disciplinarily and cross-linguistically. *Study* and *research*, very much like *estudo* and *pesquisa*, can be interchanged with each other in some contexts (synonymy); in addition, *study* and *estudo*, as well as *research* and *pesquisa*, are English–Portuguese equivalents. The Japanese verb 行う (*okonau*) was chosen based on the assumption that its surroundings may reveal disciplinary specificity, since researchers from different fields probably perform different actions. Finally, この (*kono*) was chosen because it is one of the few items in Table 8.12 that carry a meaning by itself (most items in the table need some kind of complement to express a meaning) and because, like 行う (*okonau*), its surroundings can reveal disciplinary differences—the accompanying words probably differ between one field and another.

As described earlier, both collocations and semantic associations of the six selected keywords were analyzed. Collocations were studied through the observation of strong collocates within a range of five words from the target keyword or node (-5 to +5), as well as through the comparison of 2–6-word clusters (also known as *n-grams*, *strings*, or *chunks*) containing the keyword of interest. Semantic associations were investigated through the observation of semantic sets manually compiled from collocate lists. As explained earlier, the semantic sets were compiled considering the first 100 strongest collocates (whenever the total was above 100) within a range of five words before and five words after the node (-5 to +5). The sets were based on the meaning of the collocates and were double-checked against the concordance lines to eliminate mistakes.

The top strongest collocates of *study*, *research*, *estudo*, *pesquisa*, 行う (*okonau*) and この (*kono*) are shown in Tables 8.13 to 8.18, displayed on the next pages for ease of presentation.

Table 8.13 – Top 20 strongest collocates of *study* (noun) in two corpora.

	enPED			enMGT		
POS	Word	PER	LogDice	Word	PER	LogDice
1	*this*	27.62	12.27	*this*	44.15	11.99
2	*our*	8.98	11.19	*our*	15.61	11.20
3	*in*	33.98	11.07	*present*	4.24	10.25
4	*the*	65.06	11.06	*in*	33.47	10.13
5	*was*	16.02	11.00	*was*	8.01	9.95
6	*a*	19.13	10.68	*results*	4.65	9.92
7	*present*	5.25	10.61	*findings*	3.56	9.73
8	*that*	10.64	10.55	*research*	5.54	9.73
9	*of*	30.94	10.38	*of*	38.88	9.73
10	*by*	7.11	10.30	*the*	59.07	9.66
11	*current*	4.01	10.24	*that*	15.67	9.58
12	*to*	17.82	10.22	*a*	18.62	9.52
13	*were*	10.01	10.17	*current*	2.53	9.49
14	*population*	3.80	10.08	*case*	2.67	9.48
15	*methods*	3.66	10.08	*also*	4.31	9.47
16	*period*	3.80	10.08	*purpose*	2.26	9.43
17	*limitations*	3.45	10.05	*is*	10.06	9.38
18	*from*	5.66	10.00	*for*	9.45	9.35
19	*conducted*	3.38	9.96	*conducted*	2.19	9.33
20	*is*	5.87	9.92	*we*	5.75	9.33

Source: Collocates identified with Sketch Engine. PER = number of co-occurrences between collocate and keyword divided by the overall frequency of the keyword in the corpus. Shared collocates are bounded by colored rectangles.

Table 8.14 – Top 20 strongest collocates of *research* (noun) in two corpora.

	enMGT			enPED		
POS	Word	PER	LogDice	Word	PER	LogDice
1	*future*	10.23	11.36	*committee*	7.75	11.09
2	*has*	10.66	10.97	*ethics*	7.75	11.09
3	*this*	17.10	10.75	*further*	8.91	10.83
4	*on*	16.74	10.62	*university*	5.81	10.60
5	*prior*	4.70	10.28	*future*	5.43	10.49
6	*our*	6.56	10.05	*needed*	5.04	10.36
7	*previous*	3.49	9.99	*human*	4.26	10.26
8	*empirical*	3.31	9.87	*previous*	5.81	10.24
9	*that*	16.80	9.83	*design*	4.26	9.95
10	*for*	11.68	9.80	*has*	7.36	9.93
11	*in*	23.30	9.77	*approved*	2.71	9.59
12	*study*	5.06	9.76	*qualitative*	2.71	9.56
13	*question*	2.77	9.68	*obtained*	3.10	9.53
14	*is*	10.30	9.56	*methods*	3.49	9.47
15	*of*	28.90	9.47	*question*	2.33	9.41
16	*shows*	2.41	9.43	*site*	2.33	9.36
17	*the*	44.55	9.43	*assistants*	1.94	9.27
18	*shown*	2.29	9.40	*questions*	2.33	9.27
19	*further*	2.53	9.39	*assistant*	1.94	9.26
20	*qualitative*	2.17	9.36	*is*	12.02	9.23

Source: Collocates identified with Sketch Engine. PER = number of co-occurrences between collocate and keyword divided by the overall frequency of the keyword in the corpus. Shared collocates are bounded by colored rectangles.

Table 8.15 – Top 20 strongest collocates of *estudo* (“study,” noun) in two corpora.

	ptPED			ptMGT		
POS	Word	PER	LogDice	Word	PER	LogDice
1	*presente*	16.21	12.08	*este*	23.63	12.37
2	*este*	7.97	11.21	*neste*	9.25	11.32
3	*um*	15.38	11.08	*presente*	6.16	10.71
4	*foi*	16.76	11.05	*deste*	5.48	10.60
5	*no*	16.48	10.98	*nosso*	4.11	10.21
6	*deste*	6.59	10.97	*realizado*	3.77	10.11
7	*transversal*	6.59	10.97	*contribui*	3.42	10.04
8	*neste*	6.59	10.97	*o*	31.51	9.76
9	*o*	27.75	10.81	*limitações*	2.74	9.71
10	*do*	21.70	10.80	*um*	12.67	9.68
11	*em*	19.51	10.64	*esse*	4.79	9.67
12	*nosso*	4.95	10.55	*foi*	6.85	9.62
13	*feito*	4.67	10.48	*para*	16.44	9.57
14	*objetivo*	4.95	10.45	*campo*	2.74	9.56
15	*que*	17.58	10.43	*sobre*	5.82	9.48
16	*método*	4.40	10.36	*do*	17.12	9.40
17	*com*	14.01	10.35	*no*	9.93	9.39
18	*foram*	8.24	10.28	*pesquisa*	4.45	9.38
19	*métodos*	4.12	10.27	*propõe*	2.05	9.33
20	*realizado*	4.12	10.26	*al*	4.79	9.27

Source: Collocates identified with Sketch Engine. PER = number of co-occurrences between collocate and keyword divided by the overall frequency of the keyword in the corpus. Shared collocates are bounded by colored rectangles. An English translation of this table is provided in Appendix B.

Table 8.16 – Top 20 strongest collocates of *pesquisa* ("research," noun) in two corpora.

	ptMGT			ptPED		
POS	Word	PER	LogDice	Word	PER	LogDice
1	*desta*	9.91	11.51	*comitê*	27.50	12.82
2	*pré-compra*	7.58	11.19	*ética*	22.50	12.48
3	*desenvolvimento*	9.91	11.06	*pelo*	25.00	11.64
4	*nesta*	6.12	10.86	*seres*	8.75	11.35
5	*esta*	6.41	10.83	*humanos*	8.75	11.33
6	*curso*	6.41	10.69	*universidade*	7.50	10.99
7	*inspiração*	4.96	10.57	*aprovada*	5.00	10.61
8	*serviços*	4.96	10.06	*faculdade*	5.00	10.51
9	*foi*	7.29	9.86	*participar*	5.00	10.36
10	*resultados*	5.54	9.81	*presente*	7.50	10.23
11	*da*	19.24	9.75	*desta*	3.75	10.19
12	*construto*	3.21	9.70	*projeto*	3.75	10.09
13	*como*	9.33	9.51	*recente*	3.75	10.08
14	*modelo*	4.08	9.51	*trata-se*	3.75	10.05
15	*o*	22.74	9.50	*esta*	3.75	10.03
16	*no*	9.33	9.49	*nossa*	3.75	9.99
17	*sobre*	5.25	9.47	*da*	38.75	9.89
18	*verificação*	2.33	9.46	*realizada*	3.75	9.86
19	*nossa*	2.33	9.43	*feita*	3.75	9.82
20	*na*	8.16	9.43	*uma*	16.25	9.72

Source: Collocates identified with Sketch Engine. PER = number of co-occurrences between collocate and keyword divided by the overall frequency of the keyword in the corpus. Shared collocates are bounded by colored rectangles. An English translation of this table is provided in Appendix B.

Table 8.17 – Top 20 strongest (filtered) collocates of 行う (*okonau*, “perform”) in two corpora.

	jaPED			jaMGT		
POS	Word	PER	LogDice	Word	PER	LogDice
1	ことができ	8.99	11.23	分析	14.06	11.03
2	こと	33.71	11.02	意思決定	5.47	10.21
3	喉頭気管分離術	7.87	10.94	説明	3.91	10.10
4	方針	5.62	10.55	積極的	3.91	10.04
5	でき	5.62	10.50	検証	3.13	9.84
6	多期的手術	4.49	10.29	ことができる	3.91	9.71
7	を	84.27	10.06	円滑	2.34	9.47
8	陽圧換気	3.37	10.05	比較	2.34	9.37
9	積極的	4.49	10.03	こと	24.22	9.35
10	安全	3.37	9.91	を	96.09	9.32
11	乳児期	3.37	9.88	重要	3.91	9.32
12	胃瘻造設	3.37	9.75	際に	2.34	9.31
13	か	5.62	9.71	上	2.34	9.31
14	十分	3.37	9.65	焦点	2.34	9.14
15	待機的	2.25	9.49	組織変革	3.91	9.09
16	哺乳	2.25	9.49	べき	2.34	9.07
17	な	11.24	9.48	検討	2.34	9.05
18	退院後	2.25	9.48	いう	10.16	9.00
19	人工呼吸器	2.25	9.48	より	5.47	8.98
20	原則	2.25	9.46	加え	2.34	8.95

Source: Collocates identified with Sketch Engine and filtered by the author. PER = number of co-occurrences between collocate and keyword divided by the overall frequency of the keyword in the corpus. Shared collocates are bounded by colored rectangles. Verbs with partial inflection are double underlined. An English translation of this table is provided in Appendix B together with a Romanized transcription of the Japanese words.

Table 8.18 – Top 20 strongest (filtered) collocates of この (*kono*, “this”) in two corpora.

POS	jaMGT Word	jaMGT PER	jaMGT LogDice	jaPED Word	jaPED PER	jaPED LogDice
1	よう	22.42	11.81	よう	17.78	11.46
2	点	6.87	10.76	論文	4.44	10.43
3	は	43.54	10.48	申告す	4.44	10.39
4	な	16.86	10.35	方法	6.67	10.14
5	ため	4.75	9.84	うち	5.56	10.11
6	なる	4.58	9.64	尚	3.33	9.96
7	が	23.40	9.64	こと	15.56	9.92
8	こと	8.02	9.62	ため	7.78	9.82
9	を	26.51	9.61	な	13.33	9.74
10	論理	2.62	9.61	場合	5.56	9.71
11	と	18.99	9.57	症例	11.11	9.70
12	から	5.40	9.54	もの	4.44	9.65
13	本稿	3.44	9.54	中	3.33	9.61
14	も	7.20	9.51	5 例	3.33	9.53
15	の	42.39	9.44	得る	2.22	9.46
16	結果	2.95	9.44	利点	2.22	9.45
17	組織	3.27	9.41	なる	4.44	9.43
18	事業	2.13	9.33	のみ	4.44	9.41
19	プロジェクト	2.29	9.27	ことができる	2.22	9.40
20	しかし	2.29	9.26	予想さ	2.22	9.39

Source: Collocates identified with Sketch Engine and filtered by the author. PER = number of co-occurrences between collocate and keyword divided by the overall frequency of the keyword in the corpus. Shared collocates are bounded by colored rectangles. Verbs with partial inflection are double underlined. An English translation of this table is provided in Appendix B together with a Romanized transcription of the Japanese words.

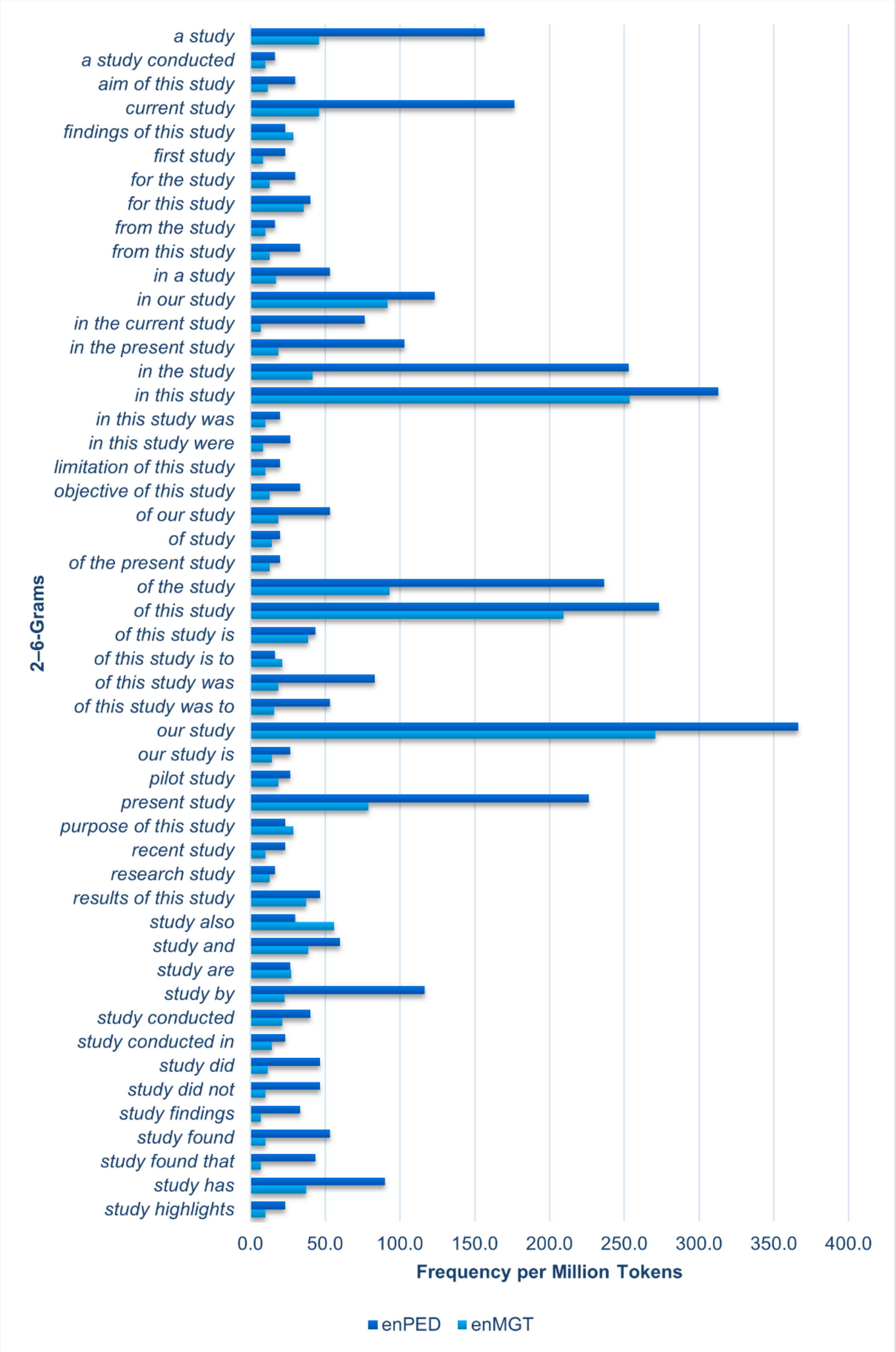
As can be noted, only *study* and *estudo* have more than 10 collocates shared by the Pediatrics and the Management data. Most collocates of the other four keywords are exclusive to either Pediatrics or Management. While in Pediatrics the English word *research* is accompanied by *committee*, *ethics*, and *design*, in Management it is accompanied by *prior*, *empirical*, and *our*. While in Pediatrics the Portuguese word *pesquisa* is accompanied by *projeto* (“project”), *humanos* (“humans”), and *participar* (“participate”), in Management it is accompanied by *desenvolvimento* (“development”), *serviços* (“services”), and *resultados* (“results”). While in Pediatrics the Japanese word 行う (*okonau*) is accompanied by 乳児期 (*nyūjiki*, “suckling stage”), 方針 (*hōshin*, “policy”), and 陽圧換気 (*yōatsu kanki*, “positive pressure ventilation”), in Management it is accompanied by 分析 (*bunseki*, “analysis”), 説明 (*setsumei*, “explanation”), and 比較 (*hikaku*, “comparison”). This indicates that keywords shared by different disciplines are unlikely to have the same collocations in texts within those disciplines.

However, unlike the results of the previous chapter, there seems to be some convergence in logDice values between the corpora for shared keywords, irrespective of their relative frequency (number of co-occurrences between collocate and keyword divided by the number of occurrences of the keyword). To illustrate this point, the following shared collocates can be cited (logDice values within parentheses): *this* (enPED: 12.27; enMGT: 11.99), *future* (enMGT: 11.36; enPED: 10.48), *deste* (“of this”; ptPED: 10.97; ptMGT: 10.60), *esta* (“this”; ptMGT: 10.83; ptPED: 10.03), 積極的 (*sekkyokuteki*, “actively”; jaPED: 10.03; jaMGT: 10.04), and よう (*yō*, “like”; jaMGT: 11.81; jaPED: 11.46).

The second approach adopted to investigate collocations in the second stage of this study was the analysis of 2–6-word clusters containing the selected keywords. As explained earlier, the cluster analysis was limited to word combinations with a minimum frequency of 5 in the corpora. In enPED, a total of 229 clusters with the word *study* were found, of which 100 (43.7%) are shared by enMGT. In enMGT, a total of 241 clusters with *study* were identified, the proportion of shared clusters being 41.5%. That is, 56.3% of the clusters found in enPED and 58.5% of those found in enMGT are exclusive. Examples of exclusive clusters in the English Pediatrics corpus are *study population* (44 instances; freq. per million: 146.5), *study design* (21 instances; freq. per million: 69.9), *cohort study* (19 instances; freq. per million: 63.3), *study was approved by* (17 instances; freq. per million: 56.6), *intervention study* (16 instances; freq. per million: 53.3), and *study sample* (14 instances; freq. per million: 46.6). Examples of exclusive clusters in the English Management corpus are *case study* (31 instances; freq. per million: 44.4), *this study’s* (28 instances; freq. per million: 40.1), *study contributes* (16 instances; freq. per million: 22.9), *study* *shows* (16 instances; freq. per million: 22.9), *study aims* (15 instances; freq. per million: 21.5), and *study makes* (14 instances; freq. per million: 20.1).

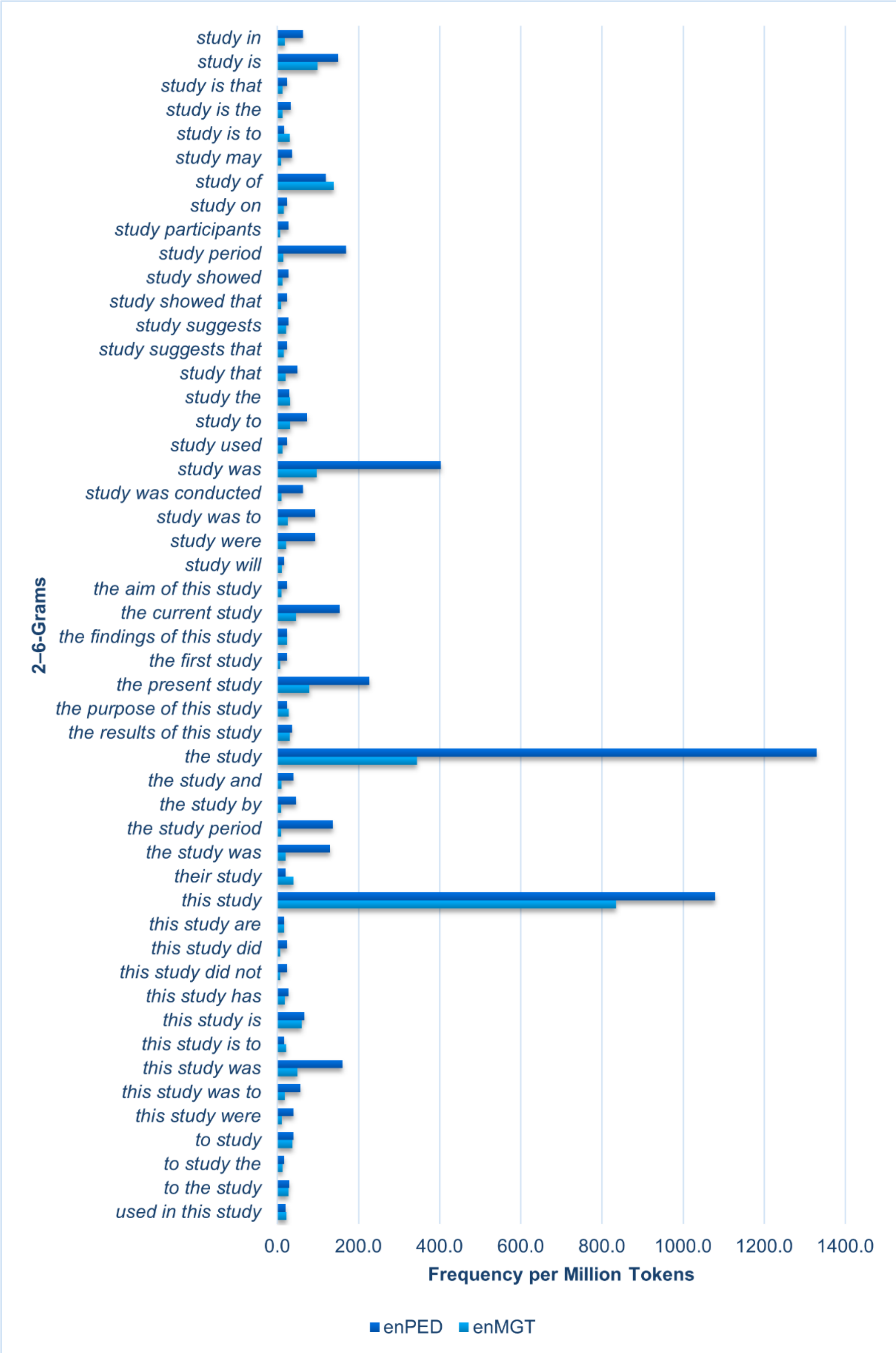
Some examples of shared clusters are *a study*, *aim of this study*, *current study*, *in this study*, *our study*, *pilot study*, *study was conducted*, *the current study*, *this study*, and *this study was*. Figures 8.2 and 8.3 (displayed on the next pages for ease of presentation) show the frequency per million of all the 100 shared clusters.

Figure 8.2 – Frequency per million tokens of shared 2–6-word clusters with *study* in two corpora (first part).



Source: Designed by the author.

Figure 8.3 – Frequency per million tokens of shared 2–6-word clusters with *study* in two corpora (second part).



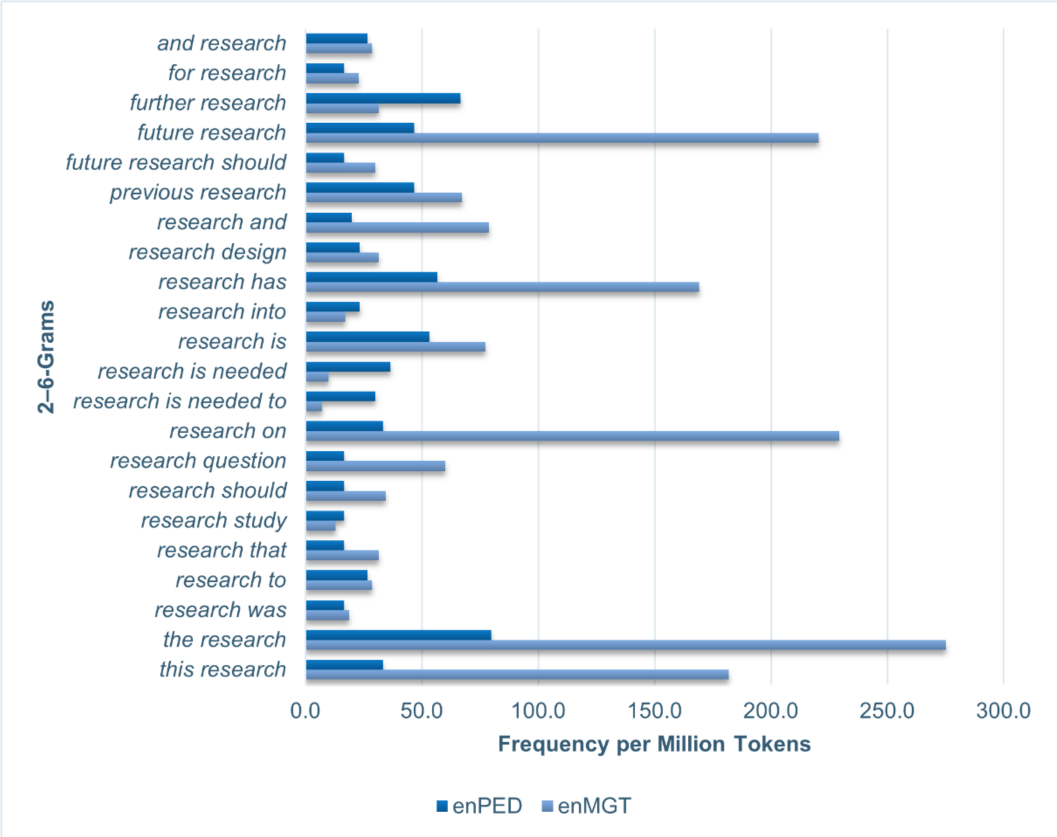
Source: Designed by the author.

As can be seen from the figures, most of the shared clusters exhibit marked differences in frequency according to the corpus, with few exceptions, for example *used in this study* (enPED: 20.0; enMGT: 22.9), *to the study* (enPED: 30.0; enMGT: 27.2), and *the findings of this study* (enPED: 23.3; enMGT: 24.4). The results suggest that, even when cross-disciplinary word combinations exist, they are likely to be preferred by authors from one or another discipline. For example, Pediatrics authors seem to be primed to write *the current study* or *the present study* more often than Management authors. Management authors, by contrast, seem to be primed to write *study also* more often than Pediatrics authors.

As for the word *research*, a total of 35 clusters containing it were found in enPED and a total of 214 in enMGT. There are 22 clusters shared by the two corpora, so that 27.1% of the Pediatrics clusters and 89.7% of those from Management are exclusive. Examples of exclusive clusters in the English Pediatrics corpus are *research ethics* (19 instances; freq. per million: 63.3), *research ethics committee* (15 instances; freq. per million: 50.0), *health research* (7 instances; freq. per million: 23.3), *further research is needed* (6 instances; freq. per million: 20.0), *human research ethics* (6 instances; freq. per million: 20.0), and *research assistant* (5 instances; freq. per million: 16.7). Examples of exclusive clusters in the English Management corpus are *prior research* (71 instances; freq. per million: 101.7), *our research* (63 instances; freq. per million: 90.3), *empirical research* (48 instances; freq. per million: 68.8), *research shows* (33 instances; freq. per million: 47.3), *qualitative research* (22 instances; freq. per million: 31.5), and *research and development* (20 instances; freq. per million: 28.7).

Examples of shared clusters are *previous research*, *research design*, *research question*, *research study*, *the research*, and *this research*. Figure 8.4 shows the normalized frequencies of the 22 shared clusters.

Figure 8.4 – Frequency per million tokens of shared 2–6-word clusters with *research* in two corpora.



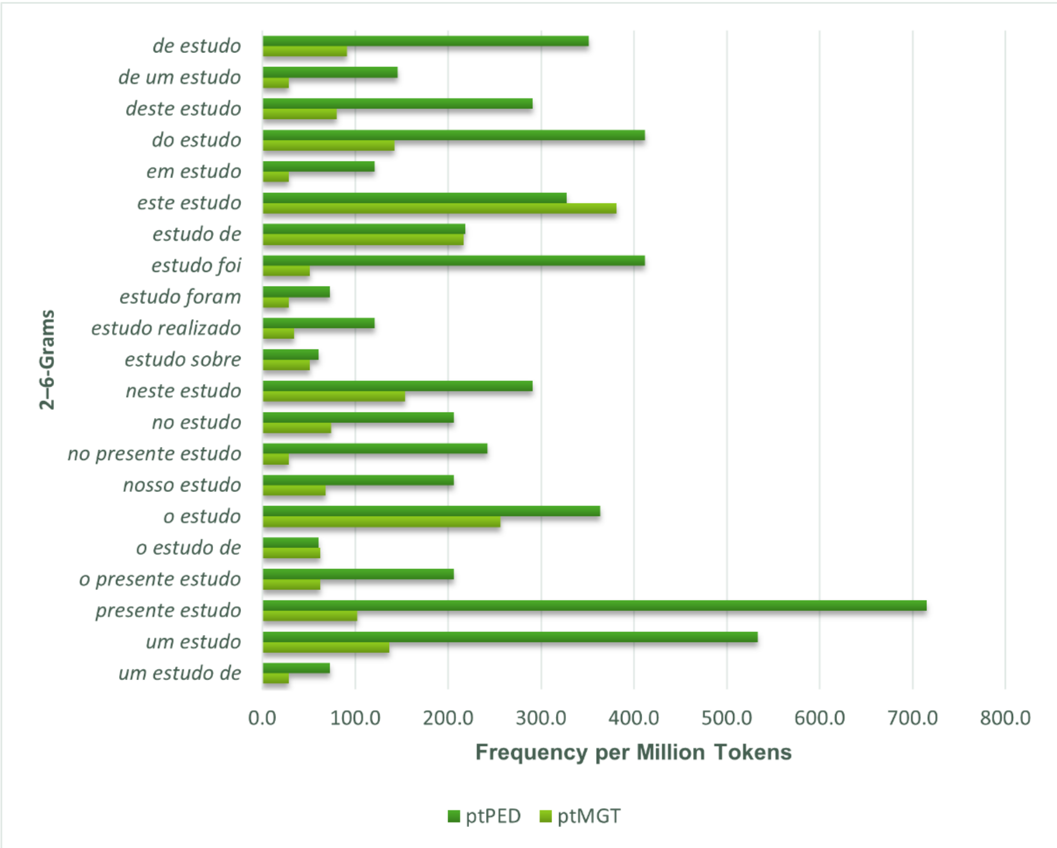
Source: Designed by the author.

As can be seen, in most cases, the clusters exhibit markedly different frequencies in enPED and enMGT, suggesting again that even when the same word combination is used in two different disciplines, it may still be more characteristic of one or another. For example, whereas Pediatrics authors are likely to be primed to write *further research*, Management authors appear to be primed to use *future research* instead. Only few clusters exhibit similar normalized frequencies, for example *research was* (enPED: 16.7; enMGT: 18.6).

In ptPED, 67 clusters with the word *estudo* were found, of which 21 (31.3%) are shared by ptMGT. In ptMGT, a total of 31 clusters were identified; therefore, 67.7% of its total also appears in ptPED. In other words, 68.7% of the clusters found in ptPED and 32.3% of those found in ptMGT are exclusive. The following combinations are examples of clusters found only in ptPED: *do presente estudo* ("of the present study"; 22 instances; freq. per million: 266.6), *estudo transversal* ("transversal study"; 13 instances; freq. per million: 157.5), *tempo de estudo* ("time of study"; 13 instances; freq. per million: 157.5), *estudo descritivo* ("descriptive study"; 7 instances; freq. per million: 84.8), and *o objetivo deste estudo* ("the aim of this study"; 6 instances; freq. per million: 72.7). The following ones are examples of clusters restricted to ptMGT: *estudo contribui* ("study contributes"; 7 instances; freq. per million: 39.9), *campo de estudo* ("field of study"; 7 instances; freq. per million: 39.9), *para este estudo* ("for this study"; 7 instances; freq. per million: 39.9), *estudo propõe* ("study proposes"; 6 instances; freq. per million: 34.2), and *estudo realizado por* ("study carried out by"; 5 instances; freq. per million: 28.5).

Examples of shared clusters are *de estudo* ("of study"), *deste estudo* ("of this study"), and *este estudo* ("this study"). The frequencies per million tokens of the shared clusters are listed in Figure 8.5.

Figure 8.5 – Frequency per million tokens of shared 2–6-word clusters with *estudo* ("study," noun) in two corpora.



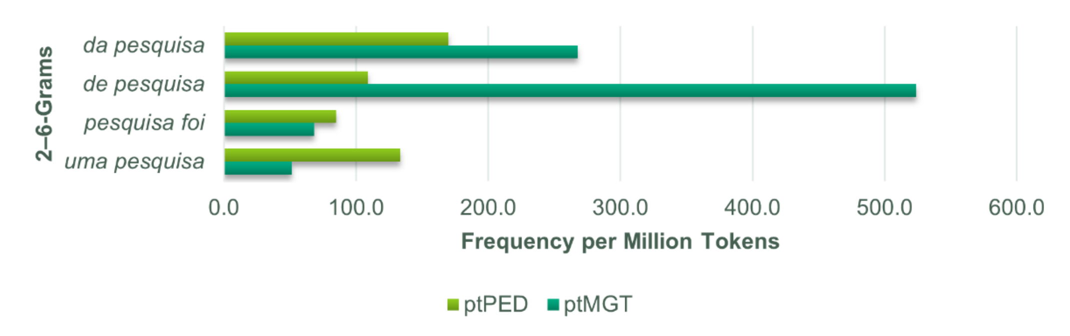
Source: Designed by the author. English translations (top–down order): "of study," "of a study," "of this study," "of the study," "under study," "this study," "study of," "study was," "study were," "study carried out," "study on," "in this study," "in the study," "in the present study," "our study," "the study," "the study of," "the present study," "present study," "a study," and "a study of."

As can be noted, except for *estudo de* ("study of"; ptPED: 218.1; ptMGT: 216.4) and *o estudo de* ("the study of"; ptPED: 60.6; ptMGT: 62.6), the clusters exhibit clear differences in frequency, which indicates that also in Portuguese even when the same word combination is used in different disciplines, it is likely to be used in different proportions. An author in the field of Pediatrics seems to be more predisposed to write *o presente estudo* ("the present study") than someone in Management.

In ptPED, 16 clusters with the word *pesquisa* could be found, of which four (25%) also appear in ptMGT. The total of clusters containing *pesquisa* in ptMGT corresponds to 50. Therefore, 75% of the clusters found in ptPED and 92% of those found in ptMGT are exclusive. Examples of clusters restricted to ptPED are *ética em pesquisa* ("research ethics"; 18 instances; freq. per million: 218.1), *comitê de ética em pesquisa* ("research ethics committee"; 17 instances; freq. per million: 206.0), *pelo comitê de ética em pesquisa* ("by the research ethics committee"; 14 instances; freq. per million: 169.7), *ética em pesquisa da* ("research ethics of/from"; 10 instances; freq. per million: 121.2), and *presente pesquisa* ("present research"; 5 instances; freq. per million: 60.6). Some of the clusters that are exclusive to ptMGT are the following: *pesquisa e desenvolvimento* ("research and development"; 28 instances; freq. per million: 159.5), *a pesquisa* ("the research"; 45 instances; freq. per million: 256.3), *desta pesquisa* ("of this research"; 34 instances; freq. per million: 193.6), *nesta pesquisa* ("in this research"; 21 instances; freq. per million: 119.6), and *modelo de pesquisa* ("research model"; 5 instances; freq. per million: 28.5).

The four shared clusters are *da pesquisa* ("of the research"), *de pesquisa* ("of research"), *pesquisa foi* ("research was"), and *uma pesquisa* ("a research study"). Their frequencies per million are shown in Figure 8.6.

Figure 8.6 – Frequency per million tokens of shared 2–6-word clusters with *pesquisa* (“research,” noun) in two corpora.



Source: Designed by the author. English translations (top–down order): “of the research,” “of research,” “research was,” and “a research study.”

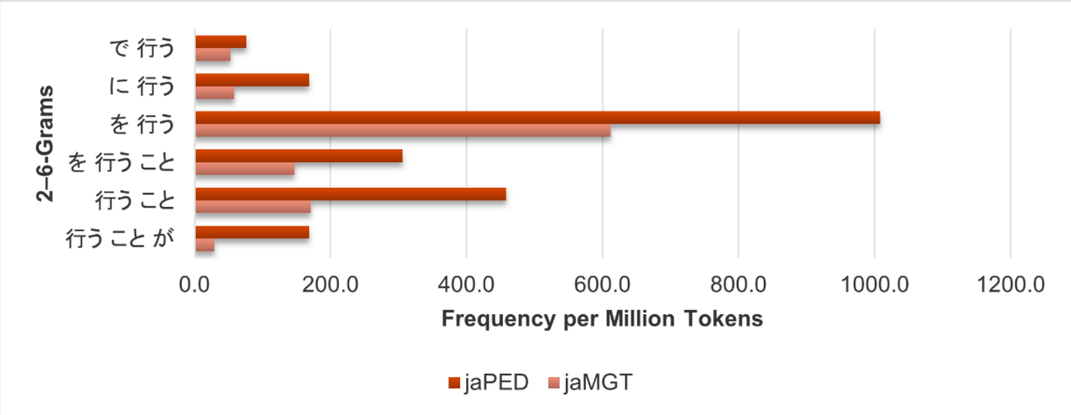
As can be observed, once more, the rule seems to be the existence of markedly different frequencies for the same cluster in different disciplines.

Concerning the Japanese verb 行う (*okonau*), a total of 15 clusters with it were observed in jaPED. The number of clusters found in jaMGT is 18. The total of exclusive clusters is nine (60%) for jaPED and 12 (66.7%) for jaMGT. Examples of exclusive clusters from the Pediatrics corpus are を行うことが (*o okonau koto ga*, “performing [something] . . .”; 6 instances; freq. per million: 91.7), に行うこと (*ni okonau koto*, “performing [something] in [some way]”; 6 instances; freq. per million: 91.7), 行うことと (*okonau koto to*, “performing . . .”; 5 instances; freq. per million: 76.4), 喉頭気管分離術を行う (*kōtōkikan bunrijutsu o okonau*, “perform a laryngotracheal separation surgery”; 5 instances; freq. per million: 76.4), 行う方針 (*okonau hōshin*, “policy to be adopted”; 5 instances; freq. per million: 76.4), and 行うことができた (*okonau koto ga dekita*, “it was possible to perform”; 5 instances; freq. per million: 76.4). Examples of exclusive clusters from the Management corpus are 分析を行う (*bunseki o okonau*, “perform an analysis”; 14 instances; freq. per million: 82.3), 意思決定を行う (*ishi kettei o okonau*, “perform a decision-making [process]”; 7 instances; freq. per million: 41.2), 説明を行う (*setsumei o okonau*, “give an explanation”; 5 instances; freq. per million: 29.4), 分析を行うこと (*bunseki o okonau koto*, “performing an analysis”; 5 instances; freq. per million: 29.4), 行うことができる (*okonau koto ga dekiru*, “it is possible to perform”; 5 instances; freq. per million:

29.4), and を行うことも (*o okonau koto mo*, “performing [something] also”; 5 instances; freq. per million: 29.4).

Figure 8.7 presents the normalized frequencies of the six shared clusters, namely で行う (*de okonau*, “perform by/with/through [something]”), に行う (*ni okonau*, “perform in [some way]”), を行う (*o okonau*, “perform [something]”), を行うこと (*o okonau koto*, “performing [something]”), and 行うことが (*okonau koto ga*, “performing . . .”).

Figure 8.7 – Frequency per million tokens of shared 2–6-word clusters with 行う (*okonau*, “perform”) in two corpora.



Source: Designed by the author. Romanized transcriptions and English translations (top–down order): *de okonau* (“perform at/through”), *ni okonau* (“perform in/on”), *o okonau* (“perform something”), *o okonau koto* (“the action of performing something”), *okonau koto* (“the action of performing”), *okonau koto ga* (“the action of performing” as the subject).

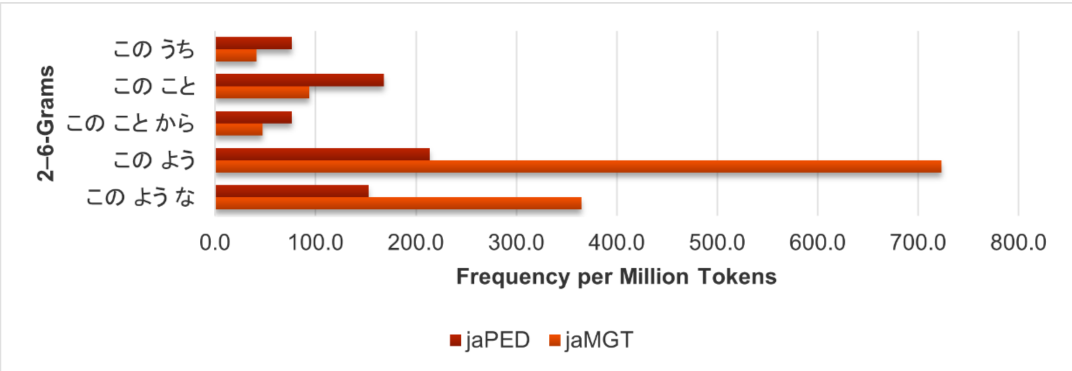
As can be noted, except for で行う (jaPED: 76.4; jaMGT: 52.9), all of the six shared clusters exhibit markedly different frequencies.

With respect to この (*kono*), seven clusters were found in jaPED and 43 in jaMGT. While exclusive clusters correspond to only two in the Pediatrics corpus (28.6%), they correspond to 38 in the Management corpus (88.4%). The two exclusive clusters found in jaPED are この症例 (*kono shōrei*, “this [clinical] case”; 5 instances; freq. per million: 76.4) and この方法 (*kono hōhō*, “this method”; 5 instances; freq. per million: 76.4). Examples of exclusive clusters in jaMGT are この点 (*kono ten*, “this point”; 36 instances; freq. per million: 211.7), このため (*kono tame*, “because of this”; 15 instances; freq. per million: 88.2), こ

のプロジェクト (*kono purojekuto*, “this project”; 12 instances; freq. per million: 70.6), このモデル (*kono moderu*, “this model”; 11 instances; freq. per million: 64.7), この事例 (*kono jirei*, “this case”; 10 instances; freq. per million: 58.8), and この場合 (*kono baai*, “this situation”; 10 instances; freq. per million: 58.8).

Figure 8.8 shows the frequencies per million tokens of the five shared clusters, namely このうち (*kono uchi*, “among these,” “between them”), このこと (*kono koto*, “this thing”), このことから (*kono koto kara*, “from this thing,” “based on this”), このよう (*kono yō*, “this way”), and このような (*kono yō na*, “[something] like this”).

Figure 8.8 – Frequency per million tokens of shared 2–6-word clusters with この (*kono*, “this”) in two corpora.



Source: Designed by the author. Romanized transcriptions and English translations (top–down order): *kono uchi* (“among these,” “between them”), *kono koto* (“this,” “this thing”), *kono koto kara* (“based on this thing,” “from this”), *kono yō ni* (“in this way”), *kono yō na* (“like this”).

As can be seen, the shared clusters exhibit again clearly different normalized frequencies depending on the corpus (and discipline).

Overall, the results of the analysis of 2–6-word clusters confirm those obtained from the typical collocate analysis. Non-specific keywords found in language data from different disciplines often combine with different words according to the discipline. Moreover, whenever the same word combination appears in data from two disciplines, it seems that it occurs with different frequencies. Altogether, the results give support to Hoey’s (2004, 2005, 2007b) claims of domain-specificity in lexical priming.

Semantic associations, as explained earlier, were investigated through manually assembled semantic sets. The sets were compiled from collocate lists generated by Sketch Engine based on the meaning of the collocates. While some words had context-independent meanings (e.g., *T1DM*, type 1 diabetes mellitus), most of them were context-dependent, so that their classification was based primarily on the analysis of concordance lines—for example, the collocates *I* and *my* were assigned to THE AUTHOR set because they refer to the article author in the pertinent lines. Only the first 100 strongest collocates were considered for keywords with more than 100 collocates, but all were considered for keywords with less than 100 collocates. As explained earlier, punctuation marks, numbers and mathematical symbols, all of which were listed as collocates by the software tool, were ignored. Target collocates were restricted to a range of five words before and five words after the node (-5 to +5). The minimum frequency of each collocate was 2.

The semantic sets compiled for the words *study* and *research* are shown in Tables 8.19 and 8.20. The presentation follows the same pattern as in Chapter 7, with shared sets and collocates enclosed by colored boxes.

Table 8.19 – Semantic sets of *study* (noun) in two corpora.

Corpus	SET (Collocates)
enPED	ACTION AS AGENT (*examined*, *found*)
	ACTION AS TARGET (*approved*, *conducted*)
	ARTICLE CONTENT (*aim*, *conclusion*, *conclusions*, *design*, *discussion*, *findings*, *limitation*, *limitations*, *methods*, *objective*, *results*)
	ATTRIBUTES OF STUDY (*previous*, *recent*)
	NUMBERS IN LETTERS (*one*, *two*)
	PEDIATRICS (*age*, *children*, *formula*, *formulas*, *hospital*, *infants*, *parents*, *patients*, *years*)
	RESEARCH METHODOLOGY AND STATISTICS (*association*, *data*, *eligible*, *group*, *large*, *participants*, *period*, *phase*, *population*, *research*, *sample*, *significant*, *size*, *studies*, *subjects*)
	THE PRESENT STUDY (*current*, *our*, *present*, *this*)
	TYPE OF STUDY (*cohort*, *cross-sectional*, *intervention*, *observational*, *prospective*, *retrospective*)
enMGT	ACTION AS AGENT (*aims*, *contributes*, *makes*, *provides*, *shows*, *suggests*)
	ARTICLE CONTENT (*aim*, *contributions*, *findings*, *implications*, *limitation*, *limitations*, *purpose*, *results*)
	MANAGEMENT (*job*, *performance*, *strategic*)
	NUMBERS IN LETTERS (*one*, *two*)
	ORDINAL NUMBERS OR ADVERBS (*first*, *second*)
	RESEARCH AIM (*determine*, *investigate*)
	RESEARCH METHODOLOGY AND STATISTICS (*analysis*, *approach*, *context*, *data*, *literature*, *measures*, *period*, *research*, *respondents*, *sample*, *support*, *theoretical*, *theory*, *variables*)
	THE PRESENT STUDY (*current*, *our*, *present*, *this*)
	TYPE OF STUDY (*case*, *empirical*, *exploratory*, *field*, *longitudinal*, *pilot*, *qualitative*)

Source: Semantic sets manually compiled from collocate lists generated by Sketch Engine. Shared sets and collocates are indicated with colored boxes.

Table 8.20 – Semantic sets of *research* (noun) in two corpora.

Corpus	SET (Collocates)
enMGT	ACTION AS TARGET (*focused*, *shown*, *shows*, *suggested*, *suggests*) ARTICLE CONTENT (*design*, *findings*, *implications*, *limitations*, *methodology*, *question*, *questions*, *results*) ATTRIBUTES OF RESEARCH (*existing*, *further*, *future*, *most*, *much*, *previous*, *prior*, *recent*, *some*) MANAGEMENT (*activity*, *development*, *identity*, *organizational*, *status*, *women*, *work*) RESEARCH AIM OR DIRECTION (*examine*, *explore*, *investigate*) RESEARCH INSTITUTIONS (*centre*, *centres*, *public*) RESEARCH METHODOLOGY AND STATISTICS (*approach*, *area*, *context*, *data*, *field*, *framework*, *literature*, *model*, *sample*, *studies*, *study*, *theoretical*, *theory*) THE PRESENT RESEARCH (*current*, *our*, *present*, *this*) TYPE OF RESEARCH (*empirical*, *qualitative*)
enPED	ACTION AS AGENT (*examined*, *found*, *shown*, *suggested*) ARTICLE CONTENT (*design*, *findings*, *implications*, *methods*, *question*, *questions*) ATTRIBUTES OF RESEARCH (*existing*, *further*, *future*, *previous*, *prior*) ETHICS IN RESEARCH (*approved*, *board*, *committee*, *consent*, *ethics*, *health*, *human*, *informed*, *medical*, *obtained*, *reviewed*) PEDIATRICS (*activity*, *families*, *parents'*, *physical*, *T1DM*) RESEARCH AIM OR DIRECTION (*address*, *explore*, *investigate*) RESEARCH INSTITUTIONS (*agency*, *criminological*, *department*, *healthcare*, *institute*, *Saxony*, *sciences*, *scientific*, *Sesalmaul*, *university*) RESEARCH METHODOLOGY AND STATISTICS (*area*, *assistant*, *assistants*, *data*, *project*, *site*, *staff*, *study*, *subject*, *survey*, *team*) THE AUTHOR (*I*, *my*) TYPE OF RESEARCH (*descriptive*, *qualitative*)

Source: Semantic sets manually compiled from collocate lists generated by Sketch Engine. Shared sets and collocates are indicated with colored boxes. T1DM = type 1 diabetes mellitus.

As can be seen, *study* and *research* do have semantic sets shared by the two corpora; yet, shared collocates are less common. It seems that even if a given English word appears in different disciplines associated with the same semantic set, the individual members of the set tend to be different. Most collocates from the tables illustrate this, but a prominent example is that of TYPE OF STUDY for the word *study*. Although *study* relates to this semantic set in both enPED and enMGT, the set members found in one and another corpus are different, reflecting different traditions of research. In the Pediatrics corpus, a given study may be *cross-sectional*, *observational*, *prospective*, or *retrospective*; in the Management corpus, it may be *exploratory*, *longitudinal*, or *qualitative*. Another prominent example is that of RESEARCH METHODOLOGY AND STATISTICS for the word *research*. In the Pediatrics corpus, *research* relates to either a team or team members—*assistant*, *assistants*, *staff*, *team*. In the Management corpus, it has a particular relationship with theory—*framework*, *literature*, *model*, *studies*, *theory*, *theoretical*. These differences also reflect distinct traditions of research. In Pediatrics, joint research projects are likely to be the norm; in Management, theoretical concerns play a major role.

Despite this, it must be noted that there are general semantic associations involving either all or almost all the same collocates. Two examples are the associations between *study* and THE PRESENT STUDY and between *research* and ATTRIBUTES OF RESEARCH. In both corpora, *study* co-occurs with *current*, *our*, *present*, and *this*; in both corpora, *research* is accompanied by *existing*, *further*, *future*, *previous*, and *prior*. This finding agrees with some previous studies on the RA genre, for example the one by Miranda (2021), which indicates that there are general, cross-disciplinary lexical characteristics in addition to disciplinary variation.

The semantic sets of the Portuguese words *estudo* and *pesquisa* are shown in Tables 8.21 and 8.22.

Table 8.21 – Semantic sets of *estudo* (noun, "study") in two corpora.

Corpus	SET (Collocates)
ptPED	ACTION AS AGENT (*analisou*, *avaliou*, *mostrou*) ACTION AS TARGET (*aprovado*, *feito*, *realizado*) ARTICLE CONTENT (*discussão*, *limitações*, *método*, *métodos*, *objetivo*, *resultados*) PEDIATRICS (*adolescentes*, *anos*, *CAPS*, *crianças*, *idade*, *lactentes*, *mães*, *TEA*) RESEARCH METHODOLOGY AND STATISTICS (*abordagem*, *achados*, *amostra*, *associação*, *comitê*, *dados*, *desenho*, *participantes*, *participar*, *participaram*, *período*, *prevalência*, *variáveis*) THE PRESENT STUDY (*deste*, *este*, *neste*, *nosso*, *presente*) TYPE OF STUDY (*descritivo*, *multicêntrico*, *observacional*, *piloto*, *revisão*, *transversal*)
ptMGT	ACTION AS AGENT (*considera*, *contribui*, *identificou*, *mostra*, *propõe*, *revela*, *sugere*, *utilizou*) ARTICLE CONTENT (*conclusão*, *conclusões*, *contribuições*, *discussão*, *implicações*, *limitações*, *método*, *metodologia*, *objetivo*, *resultado*, *resultados*, *resumo*) MANAGEMENT (*comportamento*, *consumidores, indústria*) RESEARCH AIMS (*compreender*, *entender*, *identificar*) RESEARCH METHODOLOGY AND STATISTICS (*abordagem*, *amostras*, *campo*, *dados*, *desenho*, *efeito*, *evidências*, *fatores*, *hipóteses*, *modelo*, *natureza*, *objeto*, *participantes*, *pesquisa*, *pesquisadores*, *respondentes*, *teoria*, *variáveis*) THE PRESENT STUDY (*deste*, *este*, *neste*, *nosso*, *presente*) TYPE OF STUDY (*caso*, *eventos*, *piloto*)

Source: Semantic sets manually compiled from collocate lists generated by Sketch Engine. Shared sets and collocates are indicated with colored boxes. An English translation of this table is provided in Appendix B.

Table 8.22 – Semantic sets of *pesquisa* (noun, "research") in two corpora.

Corpus	SET (Collocates)
ptMGT	ARTICLE CONTENT (*contribuição*, *hipótese*, *hipóteses*, *implicações*, *metodologia*, *objetivo*, *questão*, *resultado*, *resultados*)
	MANAGEMENT (*consumidor*, *consumidores*, *gestores*, *serviços*)
	RESEARCH METHODOLOGY AND STATISTICS (*abordagem*, *amostra*, *categorias*, *construto*, *contexto*, *dados*, *estudo*, *método*, *modelo*, *responder*, *sujeitos*, *teórica*, *teórico*, *variáveis*)
	THE PRESENT RESEARCH (*desta*, *esta*, *nesta*, *nossa*, *presente*)
	TYPE OF RESEARCH (*qualitativa*, *pré-compra*, *bibliográfica*, *curso*, *campo*)
ptPED	ACTION AS TARGET (*feita*, *realizada*)
	ETHICS IN RESEARCH (*aprovada*, *aprovado*, *aprovou*, *assinatura*, *CEP*, *comitê*, *envolvendo*, *ética*, *humanos*, *seres*)
	PEDIATRICS (*aleitamento*, *crianças*, *enfermeiras*, *hospital*, *mães*, *materno*, *medicina*, *saúde*)
	RESEARCH INSTITUTIONS (*centro*, *faculdade*, *instituição*, *universidade*)
	RESEARCH METHODOLOGY AND STATISTICS (*aceitaram*, *dados*, *estudo*, *informações*, *participar*, *prevalência*, *projeto*, *qualitativa*, *resultados*, *transversal*)
	SCOPE (*Brasil*, *nacional*)
	THE PRESENT RESEARCH (*desta*, *esta*, *nossa*, *presente*)

Source: Semantic sets manually compiled from collocate lists generated by Sketch Engine. Shared sets and collocates are indicated with colored boxes. An English translation of this table is provided in Appendix B.

The Portuguese semantic associations include five shared sets for *estudo* and two for *pesquisa*. While shared collocates are not so many, there are two semantic sets whose members are either the same or almost the same: THE PRESENT STUDY and THE PRESENT RESEARCH. The first is formed by *deste* ("of this"), *este* ("this"), *neste* ("in this"), *nosso* ("our"), and *presente* ("present") in both corpora. The second set is formed by *desta* ("of this"), *esta* ("this"), *nesta* ("in this"), *nossa* ("our"), and *presente* ("present") in ptMGT, and it is formed by *desta* ("of this"), *esta* ("this"), *nossa* ("our"), and *presente* ("present") in ptPED. This finding also agrees with previous studies on the RA genre, for example the already cited study by Miranda (2021).

It is noteworthy that there are many similarities between the associations indicated by Tables 8.19 and 8.20 and those suggested by Tables 8.21 and

8.22. Collocations and semantic associations of the English–Portuguese equivalents are very similar, indicating that there might be corresponding primings between English and Portuguese. For example, with respect to *study* and *estudo*, the following collocates from enPED and ptPED are English–Portuguese equivalents: *aim* and *objetivo*, *approved* and *aprovado*, *data* and *dados*, *discussion* and *discussão*, *limitations* and *limitações*. Also, *contributes* and *contribui*, *contributions* and *contribuições*, *present* and *presente*, and *shows* and *mostra* are equivalents found in enMGT and ptMGT. With respect to *research* and *pesquisa*, examples of English–Portuguese equivalents are the following: *question* and *questão*, *methodology* and *metodologia*, *results* and *resultados* (enMGT and ptMGT); *approved* and *aprovada* or *aprovado*, *committee* and *comitê*, *ethics* and *ética* (enPED and ptPED).

Finally, the semantic sets compiled for 行う (*okonau*) and この (*kono*) are presented in Tables 8.23 and 8.24.

Table 8.23 – Semantic sets of 行う (*okonau,* “carry out”) in two corpora.

Corpus	SET (Collocates)
jaPED	FEASIBILITY (可能, 困難)
	NEED FOR THE ACTION (べき, 必要)
	MODE (安全, 待機的, 積極的)
	PEDIATRICS (乳児期, 人工呼吸器, 保存的加療, 哺乳, 喉頭気管分離術, 多期的手術, 手術, 手術中, 方針, 施設, 栄養, 根治手術, 根治術, 治療, 管理, 胃瘻造設, 術式, 観察, 評価, 負荷試験, 退院後, 針先, 鈍的剥離, 陽圧換気)
jaMGT	DEMONSTRATIVE FAMILY (この, その)
	NEED FOR THE ACTION (べき, 必要, 必要不可欠, 重要)
	MODE (円滑, 積極的, 計画的)
	MANAGEMENT (意思決定, チーム, プロジェクト, メンバー, 事業展開, 交渉, 仕事, 他部門, 企業, 企業家的志向, 問題解決, 変革, 投資, 海外子会社, 組織, 組織変革, 行動, 調整, 議論, 賃金管理)
	RESEARCH METHODOLOGY AND WRITING (*fsQCA*, 仮説, 分析, 実証分析, 実証研究, 対象, 本稿, 検討, 検証, 構成概念, 比較, 焦点, 用い, 考え, 説明, 課題)

Source: Semantic sets manually compiled from collocate lists generated by Sketch Engine. Shared sets and collocates are indicated with colored boxes. The double underline indicates a verb with partial inflection. An English translation of this table is provided in Appendix B together with a Romanized transcription of the Japanese words.

Table 8.24 – Semantic sets of この (*kono*, "this") in two corpora.

Corpus	SET (Collocates)
jaMGT	TWO THINGS (2つ, 二つ)
	MANAGEMENT (お墨付き, プロジェクト, 事業, 企業家的志向, 取り組み, 変革, 意思決定, 技術, 新規事業, 状況, 組織, 背景, 製品)
	RESEARCH METHODOLOGY AND WRITING (アプローチ, モデル, 事例, 先行研究, 分類, 図, 指摘し, 捉え, 本稿, 概念, 点, 理由, 用い, 研究, 研究群, 結果, 考え, 考え方, 論じ, 論理, 踏まえ)
jaPED	AMONG (うち, 中)
	MEDICAL CASE NUMBERS (1例, 2例, 5例)
	PEDIATRICS (低出生体重児, 方法, 早期, 症例, 経験し, 胃瘻, 腹膜外剥離操作)
	RESEARCH METHODOLOGY AND WRITING (図, 表, 論文, 申告す, 予想さ, 考え)
	THE PRESENT RESEARCH AND AUTHORS (今回, 我々)
	TIME EXPRESSIONS (以前, 時, 時点)

Source: Semantic sets manually compiled from collocate lists generated by Sketch Engine. Shared sets and collocates are indicated with colored boxes. Verbs with partial inflection are double underlined. An English translation of this table is provided in Appendix B together with a Romanized transcription of the Japanese words.

As can be seen, there are just two semantic sets shared by jaPED and jaMGT in Table 8.23 (NEED FOR THE ACTION and MODE) and only one in Table 8.24 (RESEARCH METHODOLOGY AND WRITING). Moreover, shared collocates are limited to five: べき (*beki*, "must"), 必要 (*hitsuyō*, "need"), 積極的 (*sekkyokuteki*, "actively"), 図 (*zu*, "figure"), and 考え (*kangae*, "think," "thought"). These results provide strong support for Hoey's (2004, 2005, 2007b) claims of domain-specificity and suggest that Japanese may be more marked as to discipline than English and Portuguese, although the difference may be due to the fact that the target Japanese words (a verb and a demonstrative pronoun) do not belong to the same class as the English and Portuguese words (nouns). This assumption, however, should be investigated further.

Overall, this chapter has provided evidence for disciplinary specificity related to psychological priming in RAs. Exclusive key keywords, characteristic collocations, and distinct semantic associations are all signs of discipline-specific priming. Alluding again to Hoey's (2005) central hypotheses, this chapter's findings can be summarized in the following manner:

Every word is primed to occur with particular other words and with particular semantic sets, and these (i.e., the particular other words and semantic sets) can be expected to vary according to the pertinent specific domain or discipline.

The main implications of the findings can be summarized as follows. First, they support the need of adopting a disciplinary orientation for language learning and teaching, as already suggested by Hyland (2008) for English for academic purposes instruction. It can be supposed that there are not only Pediatrics- and Management-specific collocations and semantic associations but also primings exclusive to Engineering, Physics, Chemistry, and other disciplines. An approach that ignores disciplinary variation may prime learners to use language associations that sound unnatural or imprecise in their own disciplines, even though there are primings that appear to be above disciplinary specificity. Second, the same thought can be applied to the translation of academic texts. It seems necessary that the process of translating academic texts be disciplinarily grounded to produce standard writing in the target language. A translator who works with RAs in Pediatrics should not take his or her established primings for granted, applying them to the translation of a text in another field. It would be necessary that he or she become acquainted with the particular associations found in the second field. This view converges with Hoey's (2011) observations on translation. Third, the findings suggest that discipline may be more relevant than genre for building useful corpora for language education and translation, since there seems to be a stronger connection between lexical typicality and discipline than between lexical typicality and genre. In this sense, initiatives such as that of Kuhn (2017) and Kuhn and Ferreira (2018), who compiled a multi-genre corpus of journal texts in Portuguese according to knowledge areas, are extremely valuable.

9 PSYCHOLOGICAL PRIMING AND TEXTUAL POSITION

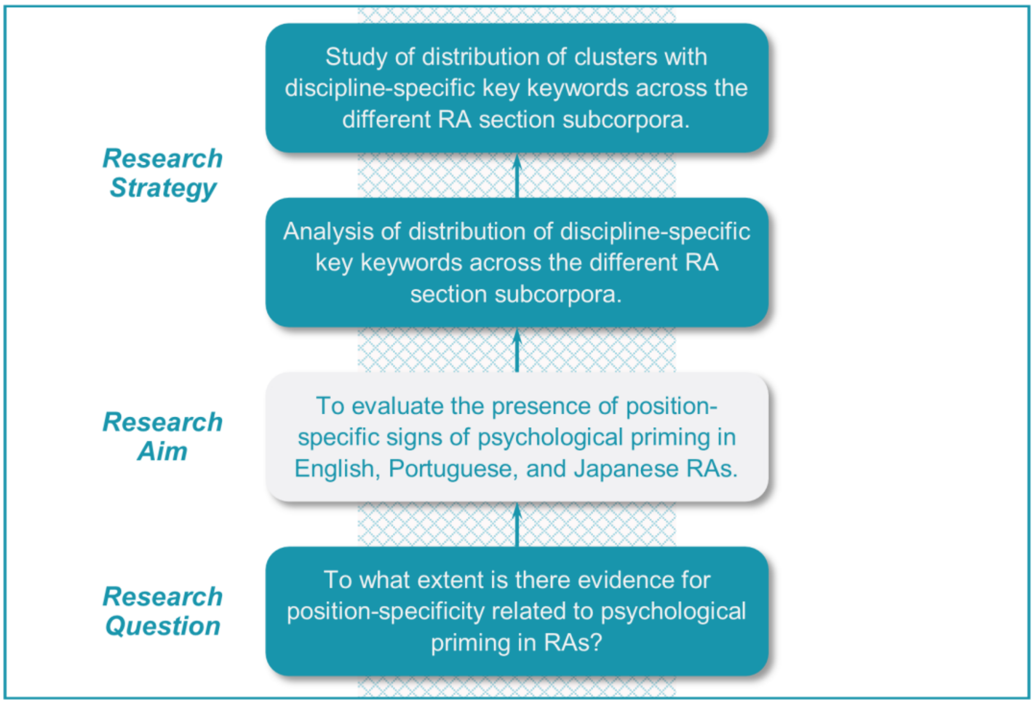
Another relevant factor for psychological priming is textual position (Hoey, 2004, 2005, 2013). Hoey (2013) states that readers and listeners subconsciously notice associations between, on the one hand, words, word combinations, or syllables and, on the other hand, particular positions in a discourse or text (p. 3344). Language users would be primed to use certain words (or word combinations, or syllables) in specific positions as the result of numerous previous encounters with those words (or combinations, or syllables) in those positions. As shown in Chapter 3, textual position has received careful attention in lexical priming research. Hoey (2013) summarizes the existing body of research dedicated to it as follows:

> The one [discourse-oriented claim] that has been most thoroughly investigated is that of textual colligation. In an AHRC-funded [Arts-and-Humanities-Research Council] research project, Hoey, Mahlberg, O'Donnell, and Scott have demonstrated the pervasiveness of textual colligation in newspaper writing, identifying hundreds of words and word combinations which are associated with specific textual positions relatively rarely; it has yet to be shown whether other genres, spoken or written, manifest the same pervasiveness (Hoey & O'Donnell, 2007, 2008a, 2008b, 2009; Mahlberg & O'Donnell, 2008). (p. 3344)[1]

[1] The references of the cited sources as they appear in Hoey (2013) are the following: "Hoey, M., & O'Donnell, M. B. (2007). Death to the topic sentence: How we really paragraph. In L. Yiu-nam (Ed.), *Selected papers from the sixteenth international symposium on English teaching* (pp. 60–76). Taipei, Republic of China: English Teachers' Association/ROC; Hoey, M., & O'Donnell, M. B. (2008a). The beginning of something important: Corpus evidence on the text beginnings of hard news stories. In B. Lewandowska-Tomaszczyk (Ed.), *PALC 2007 (Practical applications in language corpora, 7,* pp. 189–212). New York, NY: Peter Lang; Hoey, M., & O'Donnell, M. B. (2008b). Lexicography, grammar and textual position. *International Journal of Lexicography, 21*(3), 293–309; Hoey, M., & O'Donnell, M. B. (2009). The chunking of newspaper text. In M. Shiro, P. Bentivoglio, & F. D. Erlich (Eds.), *Haciendo discurso: Homenaje a Adriana Bolívar* (pp. 433–52). Caracas: Comision de Estudios de Postgrado, Universidad Central de Venezuela"; "Mahlberg, M., & O'Donnell, M. B. (2008, September). *A fresh view of the structure of hard news stories*. Paper presented at the 19th European Systemic Functional Linguistics Conference and Workshop. Retrieved September 13, 2011 from http://scidok.sulb.uni-saarland.de/volltexte/2008/1700/" (pp. 3346–3347).

Part of the final sentence is of special interest here: “it has yet to be shown whether other genres, spoken or written, manifest the same pervasiveness” (Hoey, 2013, p. 3344). The third stage of the present study addressed this gap, with a focus on the RA genre. Figure 9.1 shows the research question and aim that have guided this stage and outlines the research strategy.

Figure 9.1 – Guiding elements and research strategy of the third stage of the present study.



Source: Designed by the author.

The study of textual colligation focused on a few selected discipline-specific key keywords chosen from Tables 8.1 to 8.6 for their high occurrence in the corpora, a prerequisite for identifying patterns. Before introducing them, however, it is necessary to make a preliminary observation. Unlike the previous stages of this study, in the present, third stage AntConc was employed. During the data analysis, it was noticed that some numbers did not match those obtained earlier with Sketch Engine. This can be illustrated by the following two tables, which display the total of instances of all the discipline-specific key keywords presented earlier calculated by both software programs.

Table 9.1 – Numbers of instances of Pediatrics key keywords calculated by two software programs.

Keyness Position	enPED			ptPED			jaPED		
	Word	SE	AC	Word	SE	AC	Word	SE	AC
1	*infants*	522	534	*mães*	116	118	症例	305	504
2	*adolescents*	372	382	*adolescentes*	106	114	本症	116	143
3	*mortality*	297	297	*materno*	96	99	施行し	113	0
4	*infant*	183	219	*alimentação*	78	79	治療	81	398
5	*prevention*	81	84	*mãe*	52	54	手術	81	316
6	*parents*	534	575	*IC95*	50	0	術後	63	130
7	*diseases*	61	61	*crianças*	321	331	自験例	54	50
8	*CI*	339	339	*precoce*	28	29	症状	53	91
9	–	–	–	*município*	28	29	当科	49	49
10	–	–	–	*infantil*	27	28	当院	47	59
11	–	–	–	*dentre*	23	23	全例	46	85
12	–	–	–	*saudáveis*	22	24	患児	42	46
13	–	–	–	*meninos*	22	22	発症し	42	0
14	–	–	–	*comitê*	20	20	術式	35	59
15	–	–	–	*infância*	20	23	合併症	35	0
16	–	–	–	*apresentavam*	18	19	小児	34	271
17	–	–	–	*clínica*	18	19	男児	32	44
18	–	–	–	–	–	–	本邦	31	32
19	–	–	–	–	–	–	後方視的	30	0
20	–	–	–	–	–	–	予後	29	67
21	–	–	–	–	–	–	診断	27	145
22	–	–	–	–	–	–	生後	25	40
23	–	–	–	–	–	–	III 結果	24	0
24	–	–	–	–	–	–	診断さ	24	0
25	–	–	–	–	–	–	II 対象	24	0
26	–	–	–	–	–	–	IV 考察	24	0
27	–	–	–	–	–	–	要旨	24	13
28	–	–	–	–	–	–	治療方針	24	0
29	–	–	–	–	–	–	新生児期	23	0
30	–	–	–	–	–	–	発症	23	166
31	–	–	–	–	–	–	報告する	20	0
32	–	–	–	–	–	–	疾患	20	81
33	–	–	–	–	–	–	施行する	17	0
34	–	–	–	–	–	–	児	15	77

Source: Numbers calculated by Sketch Engine (SK) and AntConc (AC). Words whose difference between the totals is above 10 are indicated by colored boxes. Japanese verbs with partial inflection are double underlined. An English translation of this table is provided in Appendix C together with a Romanized transcription of the Japanese words.

Table 9.2 – Numbers of instances of Management key keywords calculated by two software programs.

Keyness Position	enPED			ptPED			jaPED		
	Word	SE	AC	Word	SE	AC	Word	SE	AC
1	*business*	902	964	*organizacional*	213	214	企業	376	1043
2	*organizational*	898	937	*consumidores*	211	211	本稿	315	316
3	*firms*	758	867	*empresa*	147	149	顧客	232	348
4	*innovation*	583	592	*atitude*	112	112	メンバー	209	290
5	*company*	389	480	*consumidor*	74	83	資源	169	356
6	*organisational*	327	332	*clientes*	72	72	先行研究	144	0
7	*employee*	310	355	*integração*	70	70	事例	134	217
8	*customer*	297	335	*financeiro*	58	58	従業員	128	15
9	*customers*	277	314	*financeiros*	56	56	本研究	124	116
10	*competitive*	196	197	*percebido*	53	53	モデル	106	234
11	*corporate*	195	200	*organizacionais*	52	55	イノベーション	93	116
12	*theories*	124	124	*eficiência*	45	45	行動	91	208
13	*leadership*	523	528	*influencia*	44	44	既存研究	78	0
14	*markets*	119	120	*marketing*	43	43	変数	77	174
15	*scholars*	118	120	*gerenciais*	41	41	戦略	76	312
16	*marketing*	107	107	*fluxo*	37	37	のである	75	0
17	*managerial*	101	101	*equações*	33	33	を通じて	69	0
18	*employees*	706	868	*autor*	27	27	高める	67	67
19	*industry*	375	397	–	–	–	市場	67	229
20	*identity*	580	588	–	–	–	意思決定	65	0
21	*competitors*	74	76	–	–	–	グループ	57	228
22	*strategic*	544	545	–	–	–	事業	52	291
23	–	–	–	–	–	–	プロセス	48	157
24	–	–	–	–	–	–	分析結果	47	0
25	–	–	–	–	–	–	分析し	47	0
26	–	–	–	–	–	–	論じ	47	47
27	–	–	–	–	–	–	日本企業	47	2
28	–	–	–	–	–	–	焦点	46	47
29	–	–	–	–	–	–	捉え	46	53
30	–	–	–	–	–	–	正	45	50
31	–	–	–	–	–	–	想定し	42	0
32	–	–	–	–	–	–	二つ	42	41
33	–	–	–	–	–	–	なぜ	42	41
34	–	–	–	–	–	–	第二	42	50
35	–	–	–	–	–	–	新しい	42	41
36	–	–	–	–	–	–	個人	42	65
37	–	–	–	–	–	–	そう	41	20
38	–	–	–	–	–	–	第一	41	54
39	–	–	–	–	–	–	着目し	39	0
40	–	–	–	–	–	–	高め	38	35
41	–	–	–	–	–	–	組織内	37	2
42	–	–	–	–	–	–	主張	37	64

Source: Numbers calculated by Sketch Engine (SK) and AntConc (AC). Words whose difference between the totals is above 10 are indicated by colored boxes. Japanese verbs with partial inflection are double underlined. An English translation of this table is provided in Appendix C together with a Romanized transcription of the Japanese words.

As can be seen, discrepancies arose in all corpora but mainly in the English and the Japanese data. To understand why they arose, concordance lines generated by both Sketch Engine and AntConc were examined. As a result, it was observed that AntConc includes English instances that are not computed by Sketch Engine. For example, while there are 43 instances of *parents’* in the concordance lines generated by AntConc for the word *parents* in enPED, not a single instance of it appears in the lines generated by Sketch Engine. Another example is that of *organizational* in the Management corpus. Whereas AntConc displays 37 instances of *inter-organizational* within its results for the word, Sketch Engine does not consider any of these instances. For illustrative purposes, Figure 9.2 highlights a concordance line containing *parents’* displayed by AntConc for the search word *parents* in enPED.

Figure 9.2 – Concordance lines for *parents* in enPED.

atopic disease and anemia in children (NHIS and NHANES) and	parents (	NHIS) are presented in eTable 2 in the Supplement. Definition
middle socio-economic status, and 84.7% of them lived with their	parents.	Of the 2 097 participants, 1 009 (48.1%) admitted to skipping at least one
should be validated in different geographic and demographic samples of	parents.	Parental acceptance of childhood vaccines is eroding. Nonmedical exemption
›attle" (P05). Blood glucose monitoring requires vigilance and commitment from	parents	Parents described their continuous commitment to blood glucose monitoring,
P08). Hypoglycaemia is challenging and a cause of concern for	parents	Parents were aware that physical activity came with the
at Northern Arizona University. Written informed consent was obtained from	parents.	Participants Pairs of parents (mother or father) and infants (
medical visits and hospitalizations) and indirect (missed work days for	parents).	Previous studies estimate a total burden of up to $3.8
assessment of such dimensions, including research comparing children's and	parents’	QoL (Quality of Life) across several health conditions, and
of oral health awareness was created for learners, educators and	parents.	Responses from all (100%) focus group participants emphasised that the
phase of the study upon written informed consent from their	parents.	Study feedings The two study formulas evaluated in this
assess psychological and physical maltreatment and neglect of children by	parents.	The psychometric characteristics of the instrument, including its reliability,
sent informational material to the schools including consent forms for	parents.	The study was announced by a letter sent to
nd 15 corresponding subthemes. Themes are supported by verbatim quotes from	parents.	Theme 1 Conflict between careful planning and spontaneous activity Parents
been explored primarily through qualitative analyses of focus groups with	parents [2, 8, 9, 10, 11, 12].	This literature has generally focused on parent-identified barriers

Source: Screenshot of AntConc (Anthony, 2022). Reproduced with permission. Red rectangle added.

In addition, it was observed that AntConc considers separately Japanese items that are computed as single words by Sketch Engine. For example, while AntConc processes 中小企業 (*chūshō kigyō*, “small and medium-sized businesses”) as 中小 (*chūshō kigyō*, “small and medium-sized”) and 企業 (*kigyō*, “business”), thus increasing the number of instances of 企業 (*kigyō*), Sketch Engine processes 中小企業 (*chūshō kigyō*) as an individual word that is different from 企業 (*kigyō*). Although the same segmented files were uploaded to both AntConc and Sketch Engine, they process data differently. This explains, for example, why there are 47 instances of 分析結果 (*bunseki kekka*, "analysis results") for Sketch Engine and none for AntConc. AntConc does not recognize it as a word. The difference in data processing also explains why AntConc displays much higher frequencies for words such as 企業 (*kigyō*, difference of 667 instances) and 症例 (*shōrei*, difference of 199 instances). AntConc recognizes lexical items computed by Sketch Engine as Japanese-word components as separate words.

There arises the question as to what the implications of this technical difference in analytical procedures would be for this study. Insofar as each software tool processes the language data automatically and systematically, the results from each one can be considered reliable. Each, however, has its own protocols that account for certain differences in the results presented. For valid interpretation, each can be used independently for analysis, but the numerical and statistical findings produced by the two tools cannot be immediately compared.

Having clarified the implications of the technical differences between the two software tools, the selected key keywords can be presented. In total, 18 discipline-specific key keywords from Tables 8.1 to 8.6 were selected: *infants*, *adolescents*, *parents*; *mães* ("mothers"), *adolescentes* ("adolescents"), *crianças* ("children"); 症例 (*shōrei*, "clinical case"), 本症 (*honshō*, "this disease"), 手術 (*shujutsu*, "surgery"); *business*, *organizational*, *firms*; *organizacional* ("organizational"), *consumidores* ("consumers"), *empresa* ("company"); 企業 (*kigyō*, "business"), 本稿 (*honkō*, "this paper"), and 顧客 (*kokyaku*, "customer," "customers"). In all cases, the three keywords with the highest frequencies in Tables 8.1 to 8.6 were chosen, except for 手術 (*shujutsu*). This keyword was preferred to 施行し (*shikkō shi*, "perform"), whose frequency was higher, because 施行し (*shikkō shi*) is a partially inflected verb form, thus being less suitable for analysis. 手術 (*shujutsu*) was also preferred to 治療 (*chiryō*, "treatment"), whose number of instances in jaPED is the same, because it appeared in more Japanese RAs (20 against 16 articles for 治療, *chiryō*).

Textual colligation was explored via two distributional analyses. The first focused on occurrences of the selected keywords across the different sections of the RAs. The number of instances of each keyword was recorded using AntConc's KWIC Tool, and the percentages were calculated for every subcorpus. The dispersion of the instances of the keywords was then measured using AntConc's Plot Tool with Juilland's D, and the resulting values were employed to create graphs to show the distribution of the words in the data sets. The second analysis focused on occurrences of clusters containing the selected keywords. Clusters composed of two, three, and four words were searched in all the RA section subcorpora. Every cluster found was then classified as *exclusive* (occurs in only one subcorpus) or *shared* (occurs in two or more subcorpora), and the numbers of members of each group were calculated. Additionally, examples of both exclusive and shared clusters were collected as evidence related to textual colligational priming.

The absolute and relative frequencies of the Pediatrics key keywords are shown in Tables 9.3 (English), 9.4 (Portuguese), and 9.5 (Japanese). As explained in Chapter 5, the RA section subcorpora are smaller than the previous corpora because they do not include titles, abstracts, and notes, and because one Pediatrics RA in Portuguese was excluded due to structural incompatibility. Therefore, the totals found in these and following tables are a little lower than those from the AntConc columns from Tables 9.1 and 9.2.

Table 9.3 – Instances of discipline-specific key keywords in English Pediatrics RA section subcorpora.

	enPED(Intro)		enPED(Me)		enPED(Res)		enPED(D+C)	
Keyword (Total)	n	(%)	n	(%)	n	(%)	n	(%)
parents (522)	76	14.6	115	22.0	156	29.9	175	33.5
infants (458)	74	16.2	99	21.6	135	29.5	150	32.8
adolescents (322)	59	18.3	53	16.5	52	16.1	158	49.1

Source: Prepared by the author with the aid of AntConc.

Table 9.4 – Instances of discipline-specific key keywords in Portuguese Pediatrics RA section subcorpora.

	ptPED(Intro)		ptPED(Me)		ptPED(Res)		ptPED(D+C)	
Keyword (Total)	n	(%)	n	(%)	n	(%)	n	(%)
crianças (288)	50	17.36	66	22.92	51	17.71	121	42.01
mães (107)	11	10.28	21	19.63	23	21.50	52	48.60
adolescentes (98)	16	16.33	22	22.45	17	17.35	43	43.88

Source: Prepared by the author with the aid of AntConc. English translations (top–down order): "children," "mothers," and "adolescents."

Table 9.5 – Instances of discipline-specific key keywords in Japanese Pediatrics RA section subcorpora.

	jaPED(Intro)		jaPED(Me)		jaPED(Res)		jaPED(D+C)	
Keyword (Total)	n	(%)	n	(%)	n	(%)	n	(%)
症例 (461)	34	7.4	43	9.3	159	34.5	225	48.8
本症 (130)	23	17.7	7	5.4	10	7.7	90	69.2
手術 (260)	15	5.8	35	13.5	69	26.5	141	54.2

Source: Prepared by the author with the aid of AntConc. Romanized transcriptions and English translations (top–down order): *shōrei* ("clinical case"), *honshō* ("this disease"), *shujutsu* ("surgery").

As can be seen, all the Pediatrics key keywords appear well-distributed across the different sections of the RAs. The results suggest that the selected English, Portuguese, and Japanese keywords are primed to occur in all parts of the body of RAs. Although there are differences among the subcorpora, most of them seem to reflect differences in length among sections rather than a predisposition to occur in one or another section. As shown in Table 5.7, the subcorpora composed of discussion and conclusion sections in Pediatrics have more tokens and words than the others. Therefore, it was expected that the occurrence of the selected keywords would be concentrated in the bigger data sets, as they effectively did. Two specific primings suggested by the figures above, however, involve 手術 (*shujutsu*) and 本症 (*honshō*). While the former seems to be negatively primed to appear in introductions, the latter is likely to be negatively primed to occur in method sections.

The absolute and relative frequencies of the Management key keywords are presented in Tables 9.6 (English), 9.7 (Portuguese), and 9.8 (Japanese).

Table 9.6 – Instances of discipline-specific key keywords in English Management RA section subcorpora.

Keyword (Total)	enMGT(Op) n	enMGT(Op) (%)	enMGT(Mi) n	enMGT(Mi) (%)	enMGT(Cl) n	enMGT(Cl) (%)
business (927)	118	12.7	713	76.9	96	10.4
organizational (892)	178	20.0	585	65.6	129	14.5
firms (787)	117	14.9	562	71.4	108	13.7

Source: Prepared by the author with the aid of AntConc.

Table 9.7 – Instances of discipline-specific key keywords in Portuguese Management RA section subcorpora.

Keyword (Total)	ptMGT(Op) n	ptMGT(Op) (%)	ptMGT(Mi) n	ptMGT(Mi) (%)	ptMGT(Cl) n	ptMGT(Cl) (%)
organizacional (202)	43	21.3	134	66.3	25	12.4
consumidores (209)	43	20.6	138	66.0	28	13.4
empresa (147)	11	7.4	122	82.4	15	10.1

Source: Prepared by the author with the aid of AntConc. English translations (top–down order): "organizational," "consumers," and "company."

Table 9.8 – Instances of discipline-specific key keywords in Japanese Management RA section subcorpora.

Keyword (Total)	jaMGT(Op) n	jaMGT(Op) (%)	jaMGT(Mi) n	jaMGT(Mi) (%)	jaMGT(Cl) n	jaMGT(Cl) (%)
企業 (943)	157	16.6	697	73.9	89	9.4
本稿 (265)	56	21.1	141	53.2	68	25.7
顧客 (316)	32	10.1	230	72.8	54	17.1

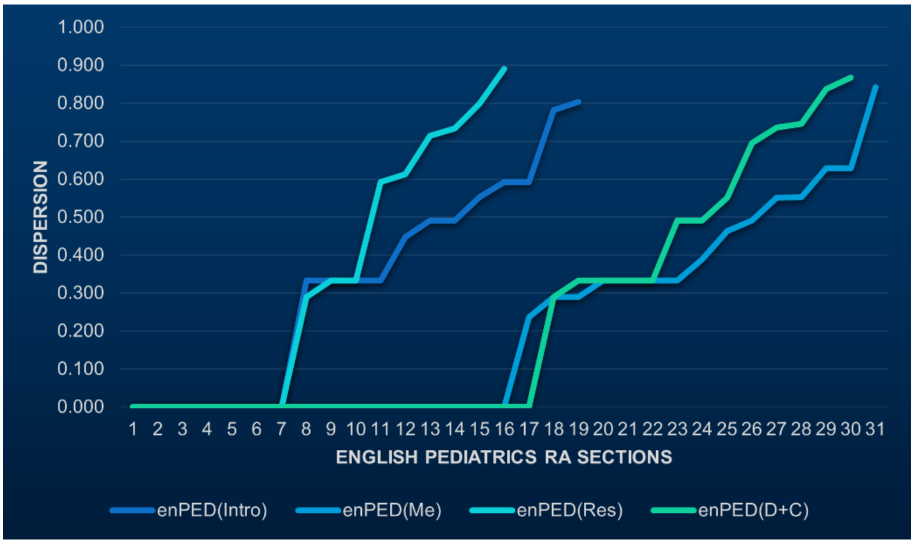
Source: Prepared by the author with the aid of AntConc. Romanized transcriptions and English translations (top–down order): *kigyō* ("business"), *honkō* ("this paper"), *kokyaku* ("client," "customer," "buyer").

All the Management key keywords also appear well-distributed across the different parts of the RAs. In addition, the differences in distribution seem to reflect again differences in length among the subcorpora (see Table 5.7). An interesting finding is that 本稿 (*honkō*, "this paper"), whose occurrence was expected to be higher in opening and closing sections, appeared mainly in middle sections. Because 本稿 (*honkō*) signals external–internal transitions (Aragão, 2022a), this finding reveals that such transitions might be common in the main parts of Management RAs in Japanese.

In addition to the number of instances, the analysis of distribution considered the dispersion of the key keywords across the RA section subcorpora. By using AntConc's Plot Tool, the dispersion (Juilland's D) values of the instances of the keywords for each individual subcorpus component were obtained and then arranged in graphs. As explained before, a dispersion value of 0.000 means that the keyword occurs only in a single, specific part of the section; a value of or near 1.000 means that it occurs throughout the section. It can be assumed that values between 0.000 and 0.299 indicate a localized distribution of the word, values between 0.300 and 0.599 indicate a moderate distribution of the word, and values between 0.600 and 1.000 indicate a wide distribution of the word. Here, we assume that values below 0.300 suggest a textual colligation between the word and a given textual position.

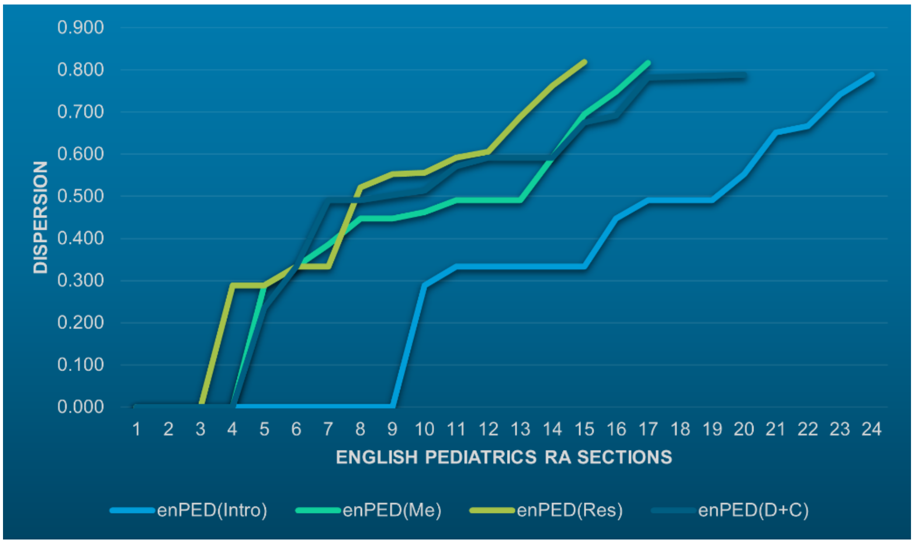
The dispersion graphs of the Pediatrics key keywords are displayed in Figures 9.3 to 9.11. Figures 9.3 to 9.5 show the resulting graphs for the English keywords; Figures 9.6 to 9.8, those for the Portuguese keywords; and Figures 9.9 to 9.11, those for the Japanese keywords.

Figure 9.3 – Dispersion of *parents* in English Pediatrics RA section subcorpora.



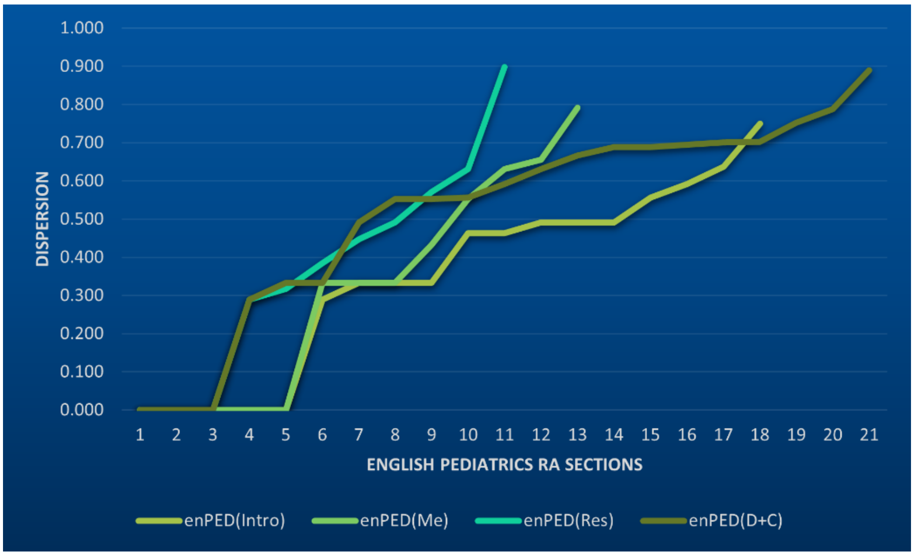
Source: Designed by the author.

Figure 9.4 – Dispersion of *infants* in English Pediatrics RA section subcorpora.



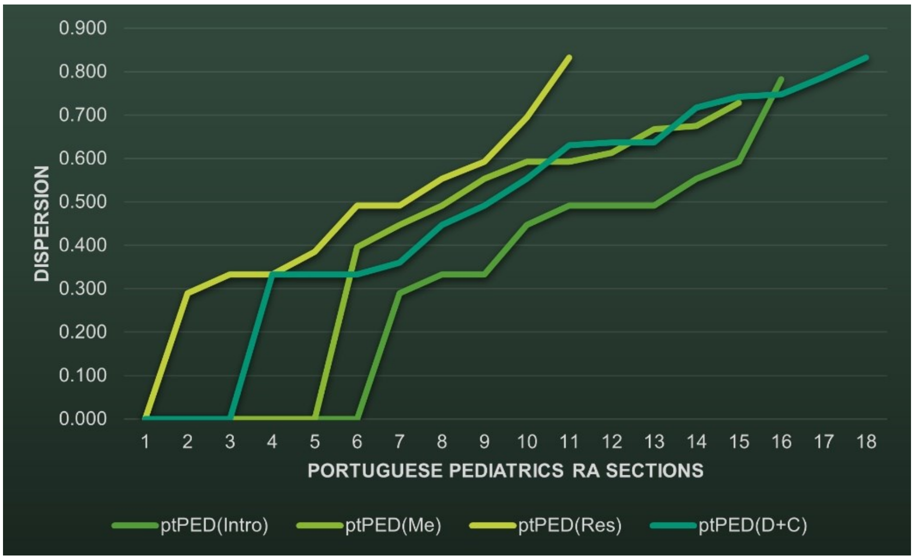
Source: Designed by the author.

Figure 9.5 – Dispersion of *adolescents* in English Pediatrics RA section subcorpora.



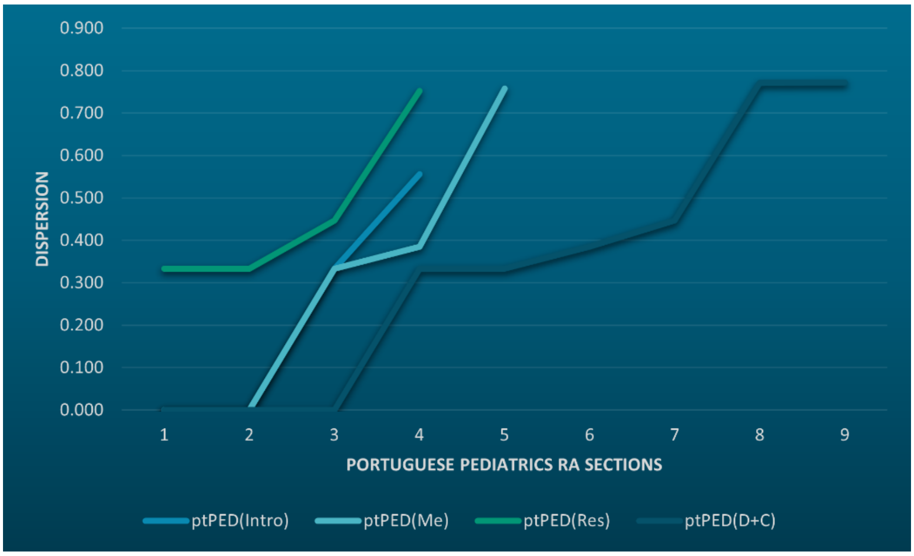
Source: Designed by the author.

Figure 9.6 – Dispersion of *crianças* ("children") in Portuguese Pediatrics RA section subcorpora.



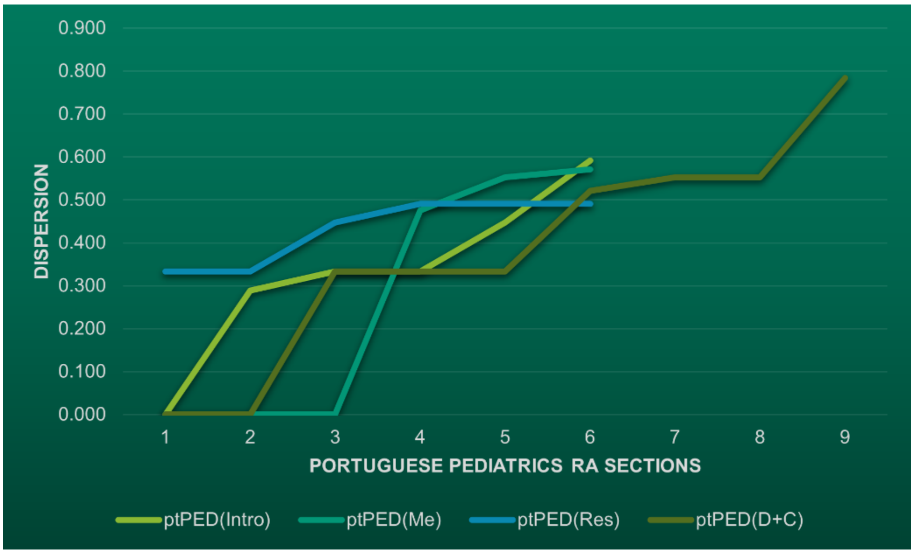
Source: Designed by the author.

Figure 9.7 – Dispersion of *mães* ("mothers") in Portuguese Pediatrics RA section subcorpora.



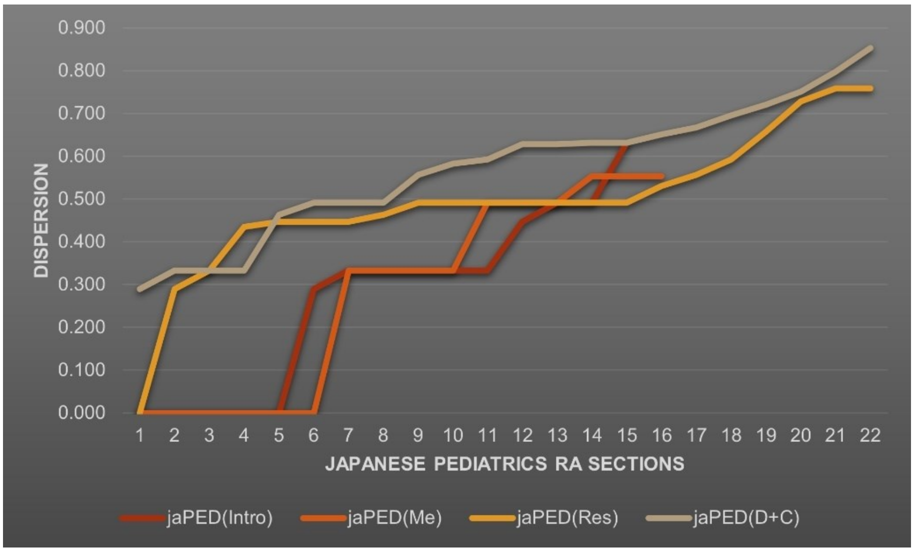
Source: Designed by the author.

Figure 9.8 – Dispersion of *adolescentes* ("adolescents") in Portuguese Pediatrics RA section subcorpora.



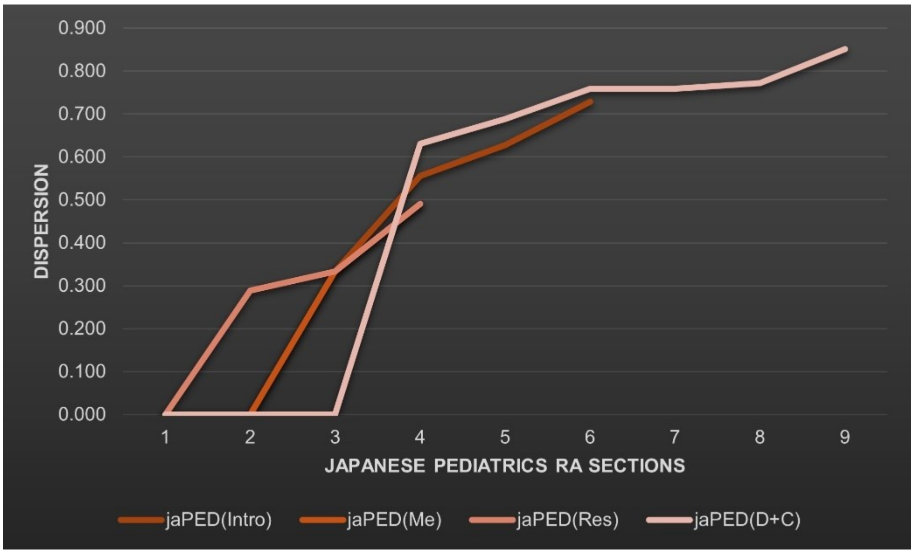
Source: Designed by the author.

Figure 9.9 – Dispersion of 症例 (*shōrei*, "clinical case") in Japanese Pediatrics RA section subcorpora.



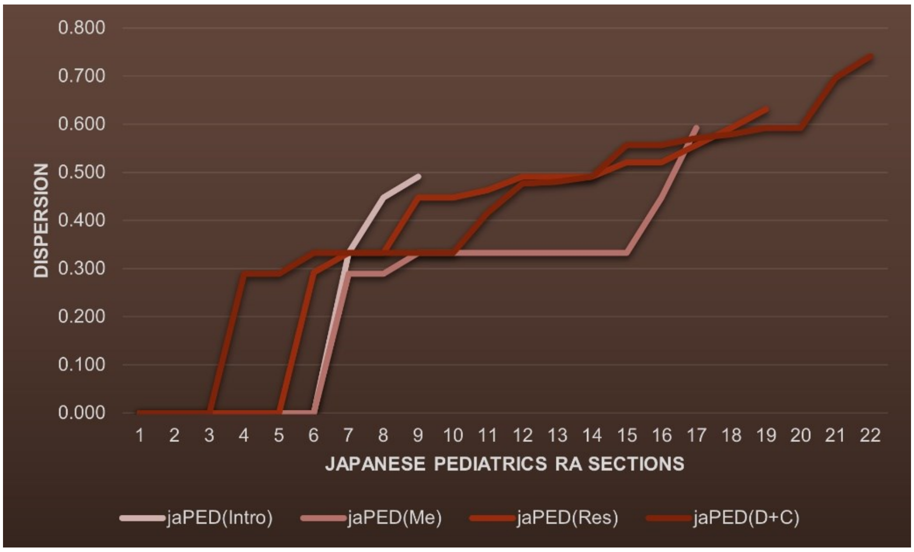
Source: Designed by the author.

Figure 9.10 – Dispersion of 本症 (*honshō*, “this disease”) in Japanese Pediatrics RA section subcorpora.



Source: Designed by the author.

Figure 9.11 – Dispersion of 手術 (*shujutsu*, “surgery”) in Japanese Pediatrics RA section subcorpora.

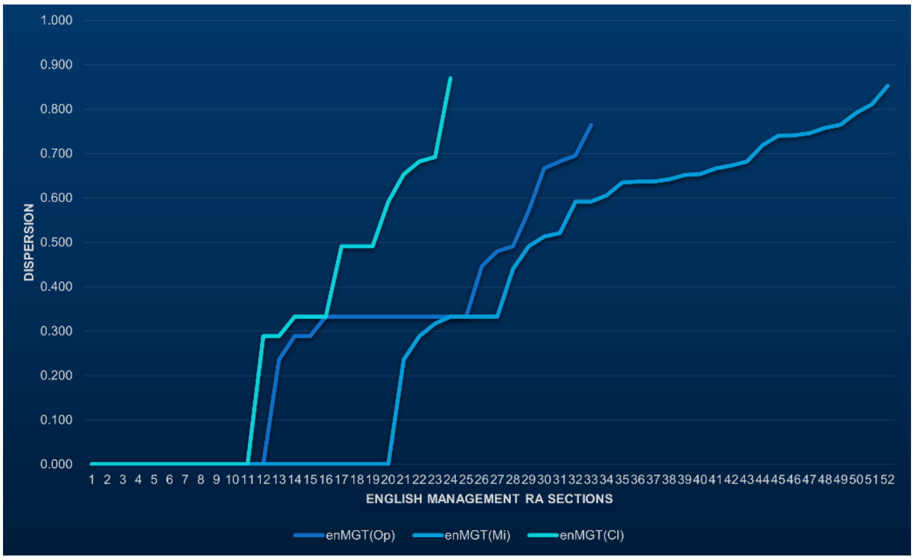


Source: Designed by the author.

Altogether, the graphs show that most of the Pediatrics keywords occur in RA sections with either a moderate or wide distribution. Most keywords are unlikely to have a preferred place for appearance. The only exception is *parents*, whose graphs suggest that there may be a preferred position in the method, discussion, and conclusion sections (Figure 9.3). Visual inspection of the plot graphs generated by AntConc, however, showed the absence of a preferred position for this keyword too.

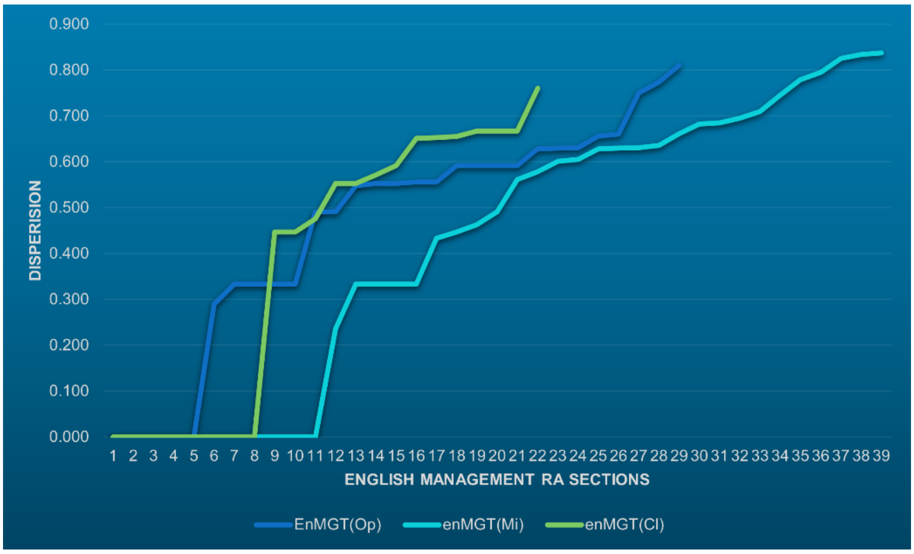
The dispersion graphs of the Management key keywords are presented in Figures 9.12 to 9.20. Figures 9.12 to 9.14 show the resulting graphs for the English keywords; Figures 9.15 to 9.17, those for the Portuguese keywords; and Figures 9.18 to 9.20, those for the Japanese keywords.

Figure 9.12 – Dispersion of *business* in English Management RA section subcorpora.



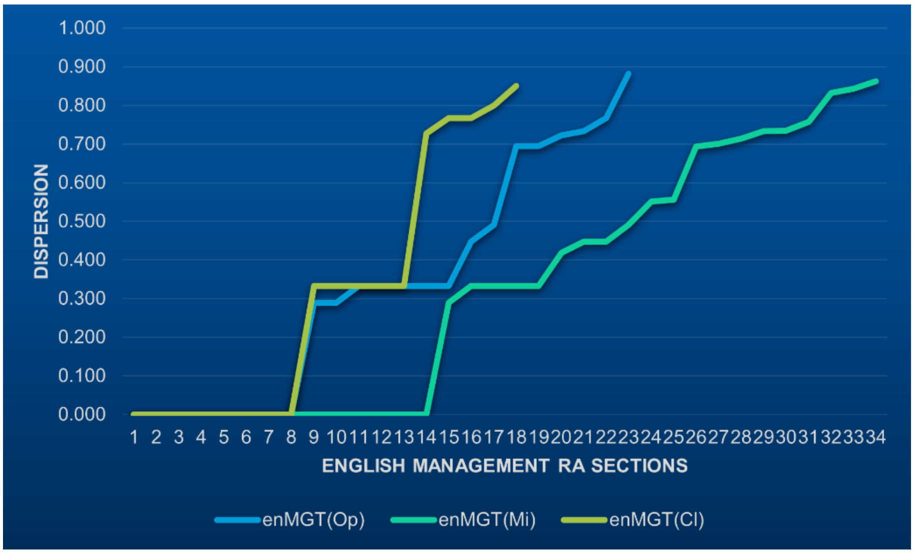
Source: Designed by the author.

Figure 9.13 – Dispersion of *organizational* in English Management RA section subcorpora.



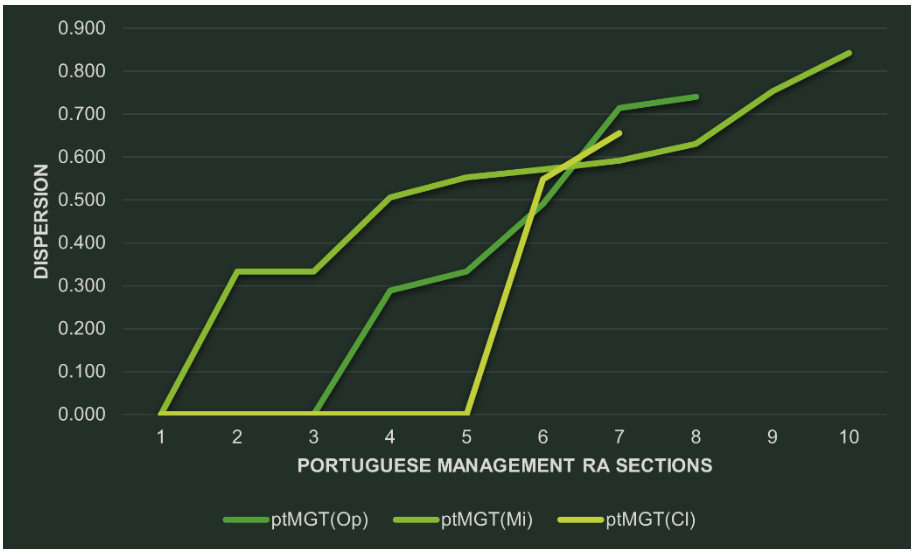
Source: Designed by the author.

Figure 9.14 – Dispersion of *firms* in English Management RA section subcorpora.



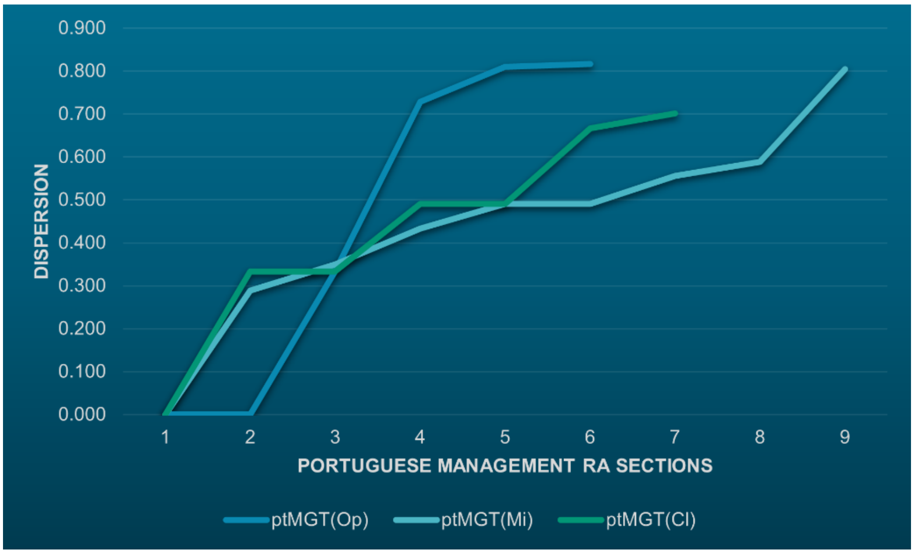
Source: Designed by the author.

Figure 9.15 – Dispersion of *organizacional* ("organizational") in Portuguese Management RA section subcorpora.



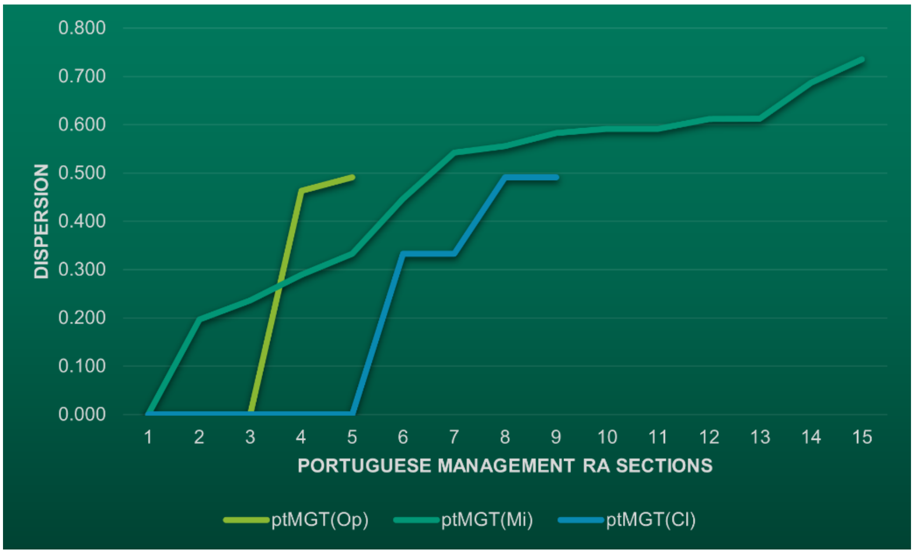
Source: Designed by the author.

Figure 9.16 – Dispersion of *consumidores* ("consumers") in Portuguese Management RA section subcorpora.



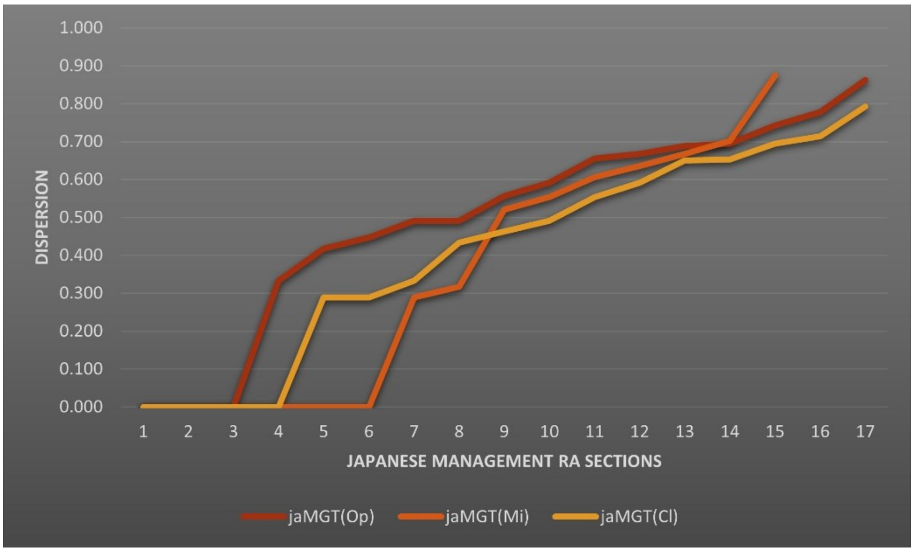
Source: Designed by the author.

Figure 9.17 – Dispersion of *empresa* ("company") in Portuguese Management RA section subcorpora.



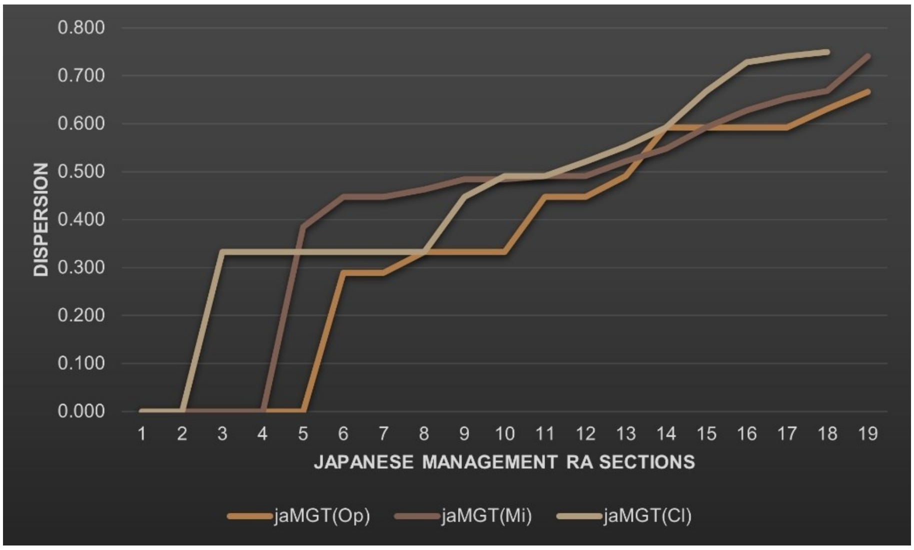
Source: Designed by the author.

Figure 9.18 – Dispersion of 企業 (*kigyō*, "business") in Japanese Management RA section subcorpora.



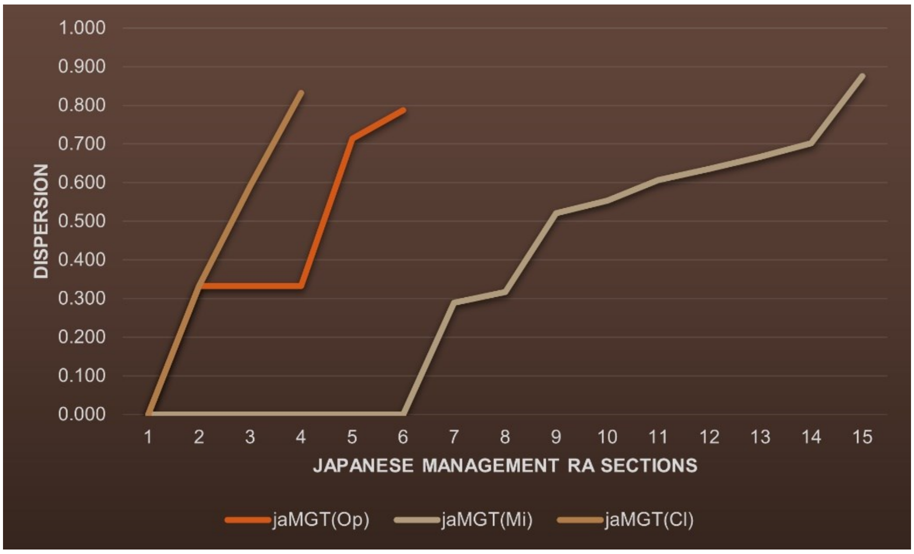
Source: Designed by the author.

Figure 9.19 – Dispersion of 本稿 (*honkō*, "this paper") in Japanese Management RA section subcorpora.



Source: Designed by the author.

Figure 9.20 – Dispersion of 顧客 (*kokyaku*, "customer") in Japanese Management RA section subcorpora.



Source: Designed by the author.

The graphs show that most of the Management keywords also occur in RA sections with a moderate or wide distribution. The only exceptions are *organizacional* ("organizational") and 顧客 (*kokyaku*, "customer"). While the former seems to be primed for use in specific positions of closing sections (Figure 9.15), the latter seems to be primed for use in specific positions of middle sections (Figure 9.20). Visual inspection of plot graphs generated by

AntConc, however, showed again the absence of a preferred position for these two keywords. Even though their instances occur in restricted areas, their places vary according to the section (article) of origin. To illustrate this, the plots of *organizacional* in ptMGT(Cl) and 顧客 (*kokyaku*) in jaMGT(Mi) are provided in Figures 9.21 and 9.22.

Figure 9.21 – Dispersion plots of *organizacional* ("organizational") in Portuguese Management RA closing sections.

Source: Plot graphs generated by AntConc.

Figure 9.22 – Dispersion plots of 顧客 (*kokyaku*, “customer”) in Japanese Management RA middle sections.

Source: Plot graphs generated by AntConc.

In sum, the results shown by Tables 9.3 to 9.8 and Figures 9.3 to 9.22 demonstrate that all the 18 selected discipline-specific key keywords occur across the different sections of RAs without a clear association with textual position. Seemingly, the selected keywords are primed to be used throughout the RAs.

As explained in the beginning of this chapter, the second distributional analysis focused on clusters. Cluster lists composed of 2–4-word combinations with the selected keywords were generated by AntConc for each of the RA section subcorpus and then were compared using Microsoft Excel. By means of its conditional formatting rule, the clusters were classified into two groups. Exclusive types refer to clusters that occur only in one subcorpus (therefore, in one RA section); shared types refer to clusters occurring in two or more subcorpora (therefore, in two or more RA sections). The results for the Pediatrics keywords are shown in Tables 9.9 (English), 9.10 (Portuguese), and 9.11 (Japanese).

Table 9.9 – Clusters with English Pediatrics key keywords.

			Exclusive Types		Shared Types	
Keyword	Cluster Size	Cluster Types	n	(%)	n	(%)
	2 words	366	276	75.4	90	24.6
parents	3 words	821	754	91.8	67	8.2
	4 words	950	919	96.7	31	3.3
	2 words	278	199	71.6	79	28.4
infants	3 words	672	616	91.7	56	8.3
	4 words	812	789	97.2	23	2.8
	2 words	214	168	78.5	46	21.5
adolescents	3 words	475	436	91.8	39	8.2
	4 words	560	538	96.1	22	3.9

Source: Prepared by the author with the aid of AntConc.

Table 9.10 – Clusters with Portuguese Pediatrics key keywords.

Keyword	Cluster Size	Cluster Types	Exclusive Types		Shared Types	
			n	(%)	n	(%)
crianças	2 words	149	113	75.8	36	24.2
	3 words	424	384	90.6	40	9.4
	4 words	495	466	94.1	29	5.9
mães	2 words	71	52	73.2	19	26.8
	3 words	174	161	92.5	13	7.5
	4 words	200	193	96.5	7	3.5
adolescentes	2 words	69	47	68.1	22	31.9
	3 words	159	144	90.6	15	9.4
	4 words	183	177	96.7	6	3.3

Source: Prepared by the author with the aid of AntConc. English translations (top–down order): "children," "mothers," and "adolescents."

Table 9.11 – Clusters with Japanese Pediatrics key keywords.

Keyword	Cluster Size	Cluster Types	Exclusive Types		Shared Types	
			n	(%)	n	(%)
症例	2 words	169	112	66.3	57	33.7
	3 words	575	502	87.3	73	12.7
	4 words	760	724	95.3	36	4.7
本症	2 words	65	37	56.9	28	43.1
	3 words	181	160	88.4	21	11.6
	4 words	219	207	94.5	12	5.5
手術	2 words	145	89	61.4	56	38.6
	3 words	372	336	90.3	36	9.7
	4 words	447	428	95.7	19	4.3

Source: Prepared by the author with the aid of AntConc. Romanized transcriptions and English translations (top–down order): *shōrei* ("clinical case"), *honshō* ("this disease"), and *shujutsu* ("surgery").

The figures above reveal that the bigger the cluster size, the more exclusive the cluster seem to be. Exclusive 2-word cluster types range from a minimum of 56.9% (本症, *honshō*, "this disease") to a maximum of 78.5% (*adolescents*); exclusive 3-word cluster types range from 87.3% (症例, *shōrei*, "clinical case") to 92.5% (*mães*, "mothers"); and exclusive 4-word cluster types range from 94.1% (*crianças*, "children") to 97.2% (*infants*). Collectively, the results suggest that word combinations containing Pediatrics keywords are primed to occur in specific parts of RAs.

For illustrative purposes, examples of both exclusive and non-exclusive clusters containing Pediatrics key keywords are provided in Figures 9.23 to 9.31. Figures 9.23 to 9.25 show examples with the English keywords; Figures 9.26 to 9.28 show clusters with the Portuguese keywords; and Figures 9.29 to 9.31, clusters with the Japanese keywords.

Figure 9.23 – Examples of exclusive and shared clusters with *parents*.

CLUSTER	enPED(Intro)	enPED(Me)	enPED(Res)	enPED(D+C)
therefore parents	•			
parents agreed		•		
parents believed			•	
parents may				•
as parents			•	•
both parents		•	•	•
parents with	•	•	•	•
parents and educators	•			
parents were asked		•		
parents indicated having			•	
parents believe their				•
and their parents		•		•
parents and caregivers	•	•	•	
parents of children	•	•	•	•
parents are defined as	•			
parents were categorized into		•		
parents believed that speech			•	
parents expressed concerns about				•
parents and health professionals			•	•
parents of a deceased	•	•	•	•

Source: Designed by the author. Clusters identified with AntConc.

Figure 9.24 – Examples of exclusive and shared clusters with *infants*.

CLUSTER	enPED(Intro)	enPED(Me)	enPED(Res)	enPED(D+C)
infants need	•			
infants during		•		
infants after			•	
infants among				•
breastfed infants	•		•	•
all infants	•	•	•	•
infants and toddlers	•			
infants for eligibility		•		
total of infants			•	
infants had significantly				•
among the infants		•	•	
hospital born infants	•	•		•
infants who were	•	•	•	•
infants at risk of	•			
infants in the intervention		•		
infants assessed in the			•	
infants born with other				•
infants and young children	•			•
low birth weight infants	•		•	•

Source: Designed by the author. Clusters identified with AntConc.

Figure 9.25 – Examples of exclusive and shared clusters with *adolescents*.

CLUSTER	enPED(Intro)	enPED(Me)	enPED(Res)	enPED(D+C)
adolescents living	•			
selected adolescents		•		
adolescents fulfilled			•	
adolescents would				•
adolescents attending	•	•		•
adolescents aged	•	•	•	•
addressed pregnant adolescents	•			
sample of adolescents		•		
adolescents reported strong			•	
adolescents are identified				•
adolescents and adults		•		•
children and adolescents	•	•	•	•
adolescents and their friends	•			
adolescents were asked to		•		
adolescents perceiving a better			•	
substantial number of adolescents				•
adolescents who did not			•	•
adolescents who feel that	•		•	•
adolescents with chronic diseases		•		•
have shown that adolescents	•			•

Source: Designed by the author. Clusters identified with AntConc.

Figure 9.26 – Examples of exclusive and shared clusters with *crianças* ("children").

CLUSTER	ptPED(Intro)	ptPED(Me)	ptPED(Res)	ptPED(D+C)
crianças existem	•			
crianças diagnosticadas		•		
crianças mostraram			•	
crianças hospitalizadas				•
algumas crianças	•			•
crianças atendidas	•			•
crianças com	•	•	•	•
crianças em risco	•			
crianças em acompanhamento		•		
metade das crianças			•	
crianças que apresentavam				•
anemia em crianças	•	•	•	•
crianças e adolescentes	•	•		•
crianças menores de	•	•	•	•
crianças com menos de	•			
crianças com deficiência intelectual		•		
crianças com idade média			•	
crianças com baixo peso				•
crianças menores de cinco	•			•
guia alimentar para crianças	•	•	•	•

Source: Designed by the author. Clusters identified with AntConc. An English translation of this figure is provided in Appendix C.

Figure 9.27 – Examples of exclusive and shared clusters with *mães* ("mothers").

CLUSTER	ptPED(Intro)	ptPED(Me)	ptPED(Res)	ptPED(D+C)
mães são	•			
mães participaram		•		
mães tinham			•	
mães gostariam				•
as mães	•	•	•	•
das mães	•	•	•	•
por mães	•	•		•
mães e gestantes	•			
mães com comprometimento		•		
expectativas das mães			•	
mães com gravidez				•
conhecimento das mães		•	•	•
mães e cuidadores	•			•
maioria das mães			•	•
alimentação saudável para mães	•			
foi constituída por mães		•		
foram incluídas as mães			•	
a motivação das mães				•
a maioria das mães			•	•
mães e aos cuidadores	•			•

Source: Designed by the author. Clusters identified with AntConc. An English translation of this figure is provided in Appendix C.

Figure 9.28 – Examples of exclusive and shared clusters with *adolescentes* ("adolescents").

CLUSTER	ptPED(Intro)	ptPED(Me)	ptPED(Res)	ptPED(D+C)
adolescentes brasileiras	●			
envolvem adolescentes		●		
adolescentes ativos			●	
adolescentes apresentavam				●
adolescentes atletas	●	●	●	●
adolescentes brasileiros	●	●		●
adolescentes praticantes			●	●
sugerem que adolescentes	●			
composto por adolescentes		●		
características dos adolescentes			●	
adolescentes com depressão				●
adolescentes atletas amadores	●	●	●	●
adolescentes de ambos	●	●		
prevalência de adolescentes			●	●
adolescentes em situações de	●			
adolescentes com menos de		●		
adolescentes de maior nível			●	
alta prevalência de adolescentes				●
adolescentes praticantes de artes			●	●
amostra final de adolescentes		●	●	

Source: Designed by the author. Clusters identified with AntConc. An English translation of this figure is provided in Appendix C.

Figure 9.29 – Examples of exclusive and shared clusters with 症例 (*shōrei*, "clinical case").

CLUSTER	jaPED(Intro)	jaPED(Me)	jaPED(Res)	jaPED(D+C)
症例 によって	●			
症例 および		●		
全 症例			●	
高い 症例				●
ある 症例		●	●	●
この 症例			●	●
の 症例	●	●	●	●
症例 も 報告	●			
症例 を 治療		●		
症例 の 概要			●	
症例 の 報告				●
で の 症例	●	●		●
症例 が 報告	●			●
症例 に対し て	●	●	●	●
症例 の 中 に	●			
症例 を 対象 と		●		
症例 の 詳細 を			●	
症例 に 比べ て				●
その よう な 症例		●	●	●
施行 し た 症例	●	●	●	●

Source: Designed by the author. Clusters identified with AntConc. An English translation of this figure is provided in Appendix C together with a Romanized transcription of the Japanese words.

Figure 9.30 – Examples of exclusive and shared clusters with 本症 (*honshō*, "this disease").

CLUSTER	jaPED(Intro)	jaPED(Me)	jaPED(Res)	jaPED(D+C)
本症 について	•			
本症 へ		•		
例 本症			•	
本症 が				•
が 本症	•			•
に 本症	•	•	•	•
本症 と	•	•	•	•
本症 と は	•			
本症 へ の		•		
本症 による 死亡			•	
本症 で は				•
本症 と 診断	•	•		•
本症 の 特徴	•			•
本症 の 発生	•			•
まで に も 本症	•			
本症 の リスク を		•		
この 時点 で 本症			•	
本症 の 特徴 を				•
のみ で は 本症		•		•
本症 を 発症 し	•		•	•

Source: Designed by the author. Clusters identified with AntConc. An English translation of this figure is provided in Appendix C together with a Romanized transcription of the Japanese words.

Figure 9.31 – Examples of exclusive and shared clusters with 手術 (*shujutsu*, "surgery").

CLUSTER	jaPED(Intro)	jaPED(Me)	jaPED(Res)	jaPED(D+C)
手術 自体	•			
手術 治療		•		
手術 以降			•	
心臓 手術				•
の 手術	•	•	•	•
初回 手術		•	•	•
手術 の	•	•		•
手術 が 第一	•			
手術 の 手順		•		
手術 を 希望			•	
手術 時 に				•
なかっ た 手術			•	•
手術 を 行う	•	•		•
手術 療法 を		•	•	•
手術 が 第一 選択	•			
手術 を 同時 に		•		
手術 時 の 平均			•	
手術 を 受け た				•
を 認め た 手術			•	•
手術 時間 出血 量		•	•	•

Source: Designed by the author. Clusters identified with AntConc. An English translation of this figure is provided in Appendix C together with a Romanized transcription of the Japanese words.

The results of the classification of the Management clusters are displayed in Tables 9.12 (English), 9.13 (Portuguese), and 9.14 (Japanese).

Table 9.12 – Clusters with English Management key keywords.

			Exclusive Types		Shared Types	
Keyword	Cluster Size	Cluster Types	n	(%)	n	(%)
	2 words	519	409	78.8	110	21.2
business	3 words	1,339	1,255	93.7	84	6.3
	4 words	1,645	1,601	97.3	44	2.7
	2 words	472	348	73.7	124	26.3
organizational	3 words	1,380	1,275	92.4	105	7.6
	4 words	1,642	1,596	97.2	46	2.8
	2 words	569	449	78.9	120	21.1
firms	3 words	1,244	1,161	93.3	83	6.7
	4 words	1,462	1,425	97.5	37	2.5

Source: Prepared by the author with the aid of AntConc.

Table 9.13 – Clusters with Portuguese Management key keywords.

			Exclusive Types		Shared Types	
Keyword	Cluster Size	Cluster Types	n	(%)	n	(%)
	2 words	110	84	76.4	26	23.6
organizacional	3 words	257	235	91.4	22	8.6
	4 words	329	313	95.1	16	4.9
	2 words	109	74	67.9	35	32.1
consumidores	3 words	285	252	88.4	33	11.6
	4 words	340	314	92.4	24	7.1
	2 words	108	92	85.2	16	14.8
empresa	3 words	240	225	93.8	15	6.3
	4 words	276	269	97.5	7	2.5

Source: Prepared by the author with the aid of AntConc. English translations (top–down order): “organizational,” “consumers,” and “company.”

Table 9.14 – Clusters with Japanese Management key keywords.

			Exclusive Types		Shared Types	
Keyword	Cluster Size	Cluster Types	n	(%)	n	(%)
企業	2 words	314	228	72.6	86	27.4
	3 words	1,148	1,020	88.9	128	11.1
	4 words	1,535	1,458	95.0	77	5.0
本稿	2 words	101	65	64.4	36	35.6
	3 words	274	238	86.9	36	13.1
	4 words	428	391	91.4	37	8.6
顧客	2 words	134	96	71.6	38	28.4
	3 words	360	316	87.8	44	12.2
	4 words	463	433	93.5	30	6.5

Source: Prepared by the author with the aid of AntConc. Romanized transcriptions and English translations (top–down order): *kigyō* ("clinical case"), *honkō* ("this paper"), and *kokyaku* ("customer," "client," "buyer").

The figures above reveal the same pattern as before. That is, the bigger the cluster size, the more exclusive the cluster is likely to be. Exclusive 2-word cluster types range from a minimum of 64.4% (本稿, *honkō*, "this paper") to a maximum of 85.2% (*empresa*, "company"); exclusive 3-word cluster types range from 86.9% (again 本稿, *honkō*) to 93.8% (again *empresa*); and exclusive 4-word cluster types range from 91.4% (once more, 本稿, *honkō*) to 97.5% (*empresa* and *firms*). Altogether, the results suggest that clusters with Management keywords are primed to occur in defined parts of RAs.

Once again for illustrative purposes, examples of exclusive and non-exclusive clusters containing the selected Management key keywords are listed in Figures 9.32 to 9.40. Figures 9.32 to 9.34 provide clusters with the English keywords; Figures 9.35 to 9.37 display clusters with the Portuguese keywords; and Figures 9.38 to 9.40, clusters with the Japanese keywords.

Figure 9.32 – Examples of exclusive and shared clusters with *business*.

CLUSTER	enMGT(Op)	enMGT(Mi)	enMGT(Cl)
business failures	•		
established business		•	
help business			•
business against	•	•	
business context	•	•	
business environment	•	•	•
business ethics	•	•	•
business activities in	•		
business and industry		•	
general family business			•
a real business	•	•	
a small business	•	•	
business leaders in	•	•	•
business and social science	•		
business age and mean		•	
business women has set			•
age of a business	•	•	•
business managers and also		•	•
performance of the business		•	•
of students and business	•	•	•

Source: Designed by the author. Clusters identified with AntConc.

Figure 9.33 – Examples of exclusive and shared clusters with *organizational*.

CLUSTER	enMGT(Op)	enMGT(Mi)	enMGT(Cl)
organizational analysis	•		
organizational challenges		•	
organizational values			•
about organizational		•	•
an organizational	•	•	•
between organizational	•	•	•
by organizational	•	•	•
organizational psychology field	•		
organizational roles that		•	
engaging in organizational			•
aspects of organizational	•	•	
effects on organizational	•	•	•
organizational learning and compliance	•		
organizational and environmental levels		•	
organizational learning theory by			•
a kind of organizational	•	•	
a variety of organizational		•	•
organizational relations and innovation	•		•
the literature on organization	•	•	•

Source: Designed by the author. Clusters identified with AntConc.

Figure 9.34 – Examples of exclusive and shared clusters with *firms*.

CLUSTER	enMGT(Op)	enMGT(Mi)	enMGT(Cl)
firms operating	•		
firms eliminated		•	
firms simply			•
accounting firms	•	•	
asymmetric firms		•	•
established firms	•	•	•
firms are able	•		
patents from firms		•	
at private firms			•
between two firms		•	•
demands on firms		•	•
firms have a	•	•	•
firms to pursue	•	•	•
firms face obstacles to	•		
firms account for the		•	
firms prefer to structure			•
constrained founder run firms	•	•	•
firms are likely to	•	•	
firms with high competitive		•	•

Source: Designed by the author. Clusters identified with AntConc.

Figure 9.35 – Examples of exclusive and shared clusters with *organizacional* ("organizational").

CLUSTER	ptMGT(Op)	ptMGT(Mi)	ptMGT(Cl)
gestão organizacional	•		
impacto organizacional		•	
eficácia organizacional			•
apoio organizacional	•	•	•
comprometimento organizacional	•	•	•
mudança organizacional		•	•
da cultura organizacional	•		
do comprometimento organizacional		•	
a realidade organizacional			•
baixo suporte organizacional	•	•	
no nível organizacional	•	•	•
o desempenho organizacional	•	•	•
organizacional de empresas industriais	•		
escala de apoio organizacional		•	
desempenho no nível organizacional			•
atributos de nível organizacional		•	•
comportamento de cidadania organizacional		•	•
comportamentos de cidadania organizacional	•	•	•
laboral e comprometimento organizacional	•	•	•
percepção de apoio organizacional		•	•

Source: Designed by the author. Clusters identified with AntConc. An English translation of this figure is provided in Appendix C.

Figure 9.36 – Examples of exclusive and shared clusters with *consumidores* (“consumers”).

CLUSTER	ptMGT(Op)	ptMGT(Mi)	ptMGT(Cl)
consumidores estão	•		
consumidores imunes		•	
consumidores possam			•
aos consumidores	•	•	
como consumidores	•	•	
consumidores de	•	•	•
pelos consumidores	•	•	•
estudo de consumidores	•		
mentes dos consumidores		•	
em alguns consumidores			•
aspirações dos consumidores	•	•	•
comportamento dos consumidores	•	•	•
consumidores mais jovens		•	•
preferência dos consumidores	•	•	•
consumidores estão dispostos a	•		
consumidores a pagar um		•	
ação coletiva de consumidores			•
aos olhos dos consumidores	•	•	
as aspirações dos consumidores	•	•	•
e comportamento dos consumidores	•	•	•

Source: Designed by the author. Clusters identified with AntConc. An English translation of this figure is provided in Appendix C.

Figure 9.37 – Examples of exclusive and shared clusters with *empresa* (“company”).

CLUSTER	ptMGT(Op)	ptMGT(Mi)	ptMGT(Cl)
empresa nas	•		
empresa é		•	
empresa com			•
a empresa		•	•
cada empresa		•	•
uma empresa	•	•	•
empresa de varejo	•		
empresa de computação		•	
situação da empresa			•
com a empresa		•	•
de uma empresa	•	•	•
empresa socialmente responsável		•	•
aos executivos da empresa	•		
empresa na promoção da		•	
gestor sobre a empresa			•
consumidores veem uma empresa		•	•
de investir na empresa	•	•	
dos produtos da empresa		•	•
em relação à empresa	•	•	
lojas de uma empresa	•	•	

Source: Designed by the author. Clusters identified with AntConc. An English translation of this figure is provided in Appendix C.

Figure 9.38 – Examples of exclusive and shared clusters with 企業 (*kigyō*, “business”).

CLUSTER	jaMGT(Op)	jaMGT(Mi)	jaMGT(Cl)
中小 企業	•		
企業 者		•	
中国 企業			•
が 企業	•	•	•
な 企業	•	•	•
の 企業	•	•	•
でき ない 企業	•		
企業 の 価値		•	
特定 の 企業			•
す なわち 企業		•	•
で は 企業	•	•	•
は 日本 企業	•	•	•
企業 が 経済 の	•		
企業 と の 活動		•	
の よう な 企業			•
こう し た 企業	•	•	
によって 他 の 企業	•	•	
の 著名 な 企業	•	•	•
企業 に おい て	•	•	•
企業 を 対象 と	•	•	•

Source: Designed by the author. Clusters identified with AntConc. An English translation of this figure is provided in Appendix C together with a Romanized transcription of the Japanese words.

Figure 9.39 – Examples of exclusive and shared clusters with 本稿 (*honkō*, “this paper”).

CLUSTER	jaMGT(Op)	jaMGT(Mi)	jaMGT(Cl)
より 本稿	•		
および 本稿		•	
ただし 本稿			•
さらに 本稿		•	•
しかし 本稿	•	•	
また 本稿	•	•	•
本稿 の 目的	•		
本稿 が 注目		•	
本稿 と 異なる			•
そこ で 本稿	•	•	
本稿 で は	•	•	•
本稿 の 分析	•	•	•
本稿 で 分析 する	•		
本稿 で この 論理		•	
本稿 の 主 な			•
本稿 で は この	•	•	
本稿 で は 以上	•	•	•
本稿 で 明らか に		•	•
本稿 の 分析 結果		•	•
本稿 の 課題 と		•	•

Source: Designed by the author. Clusters identified with AntConc. An English translation of this figure is provided in Appendix C together with a Romanized transcription of the Japanese words.

Figure 9.40 – Examples of exclusive and shared clusters with 顧客 (*kokyaku*, “customer”).

CLUSTER	jaMGT(Op)	jaMGT(Mi)	jaMGT(Cl)
特定 顧客	•		
顧客 にとって		•	
求める 顧客			•
さまざまな 顧客		•	•
と 顧客	•	•	•
の 顧客	•	•	•
同じ 顧客		•	•
顧客 の 評価	•		
目的 は 顧客		•	
本研究 は 顧客			•
こと から 顧客		•	•
多様 な 顧客		•	•
顧客 と の	•	•	•
顧客 の 声		•	•
顧客 へ の 依存	•		
顧客 と の 関係		•	
本研究 で は 顧客			•
の よう に 顧客		•	•
顧客 と 資源 の	•	•	•
顧客 の 範囲 は	•	•	•

Source: Designed by the author. Clusters identified with AntConc. An English translation of this figure is provided in Appendix C together with a Romanized transcription of the Japanese words.

Overall, this chapter has provided partial evidence for Hoey’s (2004, 2005, 2013) claims concerning textual colligation (textual colligational priming). On the one hand, the results do not suggest strong association between individual discipline-specific keywords and specific parts of RAs, whether in English, Portuguese, or Japanese. Based on our results, the pervasiveness of textual colligations does not appear to manifest in the RA genre. It must be noted, however, that there may be other words strongly primed to occur in specific sections of RAs. Genre analysis, for example, has indicated that words related to IMPORTANCE, INTEREST and PROMINENCE, such as *important*, *interest*, *concerns*, and *well-known*, often appear in RA introductions (see Swales, 1981/2011, 1990; Samraj, 2002; etc.). Therefore, the absence of a strong relationship between the selected keywords and textual position may be a characteristic of these particular keywords rather than a general rule. On the other hand, the results show a different arrangement for clusters that contain the keywords. It seems that 2–4-word clusters occur more often in particular parts of RAs than throughout the entire RAs. In addition, the longer the cluster, the more localized it appears to be. This leads to the following provisional conclusion, inspired once again by Hoey's (2005) hypotheses:

Not every word is primed to occur in certain positions within the discourse; however, whenever two or more words combine or nest, they are likely to occur in rather restricted textual areas.

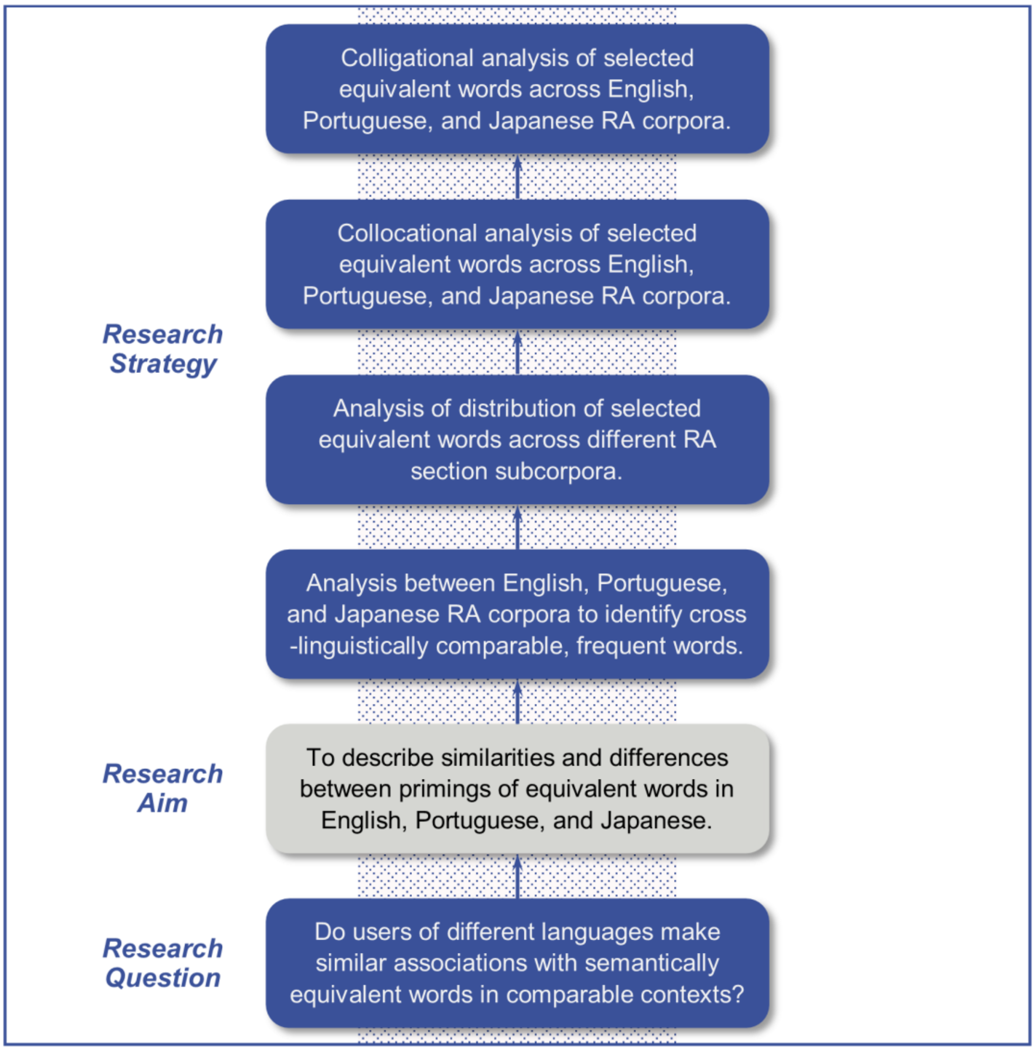
The main implications of this chapter's findings can be summarized as follows. First, concerning language learning and teaching, it is possible that academic writing students may benefit from some understanding about associations between specific strings of words and particular textual positions. As Hyland (2008) points out, "[m]ulti-word expressions are an important component of fluent linguistic production and a key factor in successful language learning" (p. 4); in addition to knowing these expressions, it may be useful to know where to use them, specifically in academic writing. Second, the translation of scholarly texts may also benefit from the understanding about possible associations between chunks of words and positions in text. If a translator can distinguish between general use and section-specific clusters, possibly he or she will be able to enhance the shift from the original lexical arrangement to the translation work (balance). This seems to be especially useful for translations from Japanese into English or Portuguese, because they may require more attention to word strings than to individual words.

10 PSYCHOLOGICAL PRIMING ACROSS LANGUAGES

Lexical priming research has explored data not only in English but also in other languages, such as German (Pace-Sigge, 2015), Portuguese (Cunha, 2017), and Finnish (Jantunen, 2017). The existing body of research has successfully demonstrated that lexical priming can be applied as a theoretical framework to investigate different languages. There are also cross-linguistic studies, such as those between English and German (Pace-Sigge, 2007) and English and Chinese (Shao, 2018; Wang, 2018, 2022), which collectively have shown the appropriateness of lexical priming for comparative analyses across languages. Despite this, studies involving languages other than English remain scarce. The extent to which users of different languages make similar associations in comparable situations is not yet fully understood. The fourth stage of the present study addressed this gap, with a focus on the comparison between English, Portuguese, and Japanese, the three languages in which the selected RAs have been published. While the previous chapters have already shown some cross-linguistic variations, [1] this chapter is exclusively dedicated to differences and similarities in priming associations of cross-linguistically comparable words. Figure 10.1 shows the guiding elements and an overview of the procedures of the final stage of this study.

[1] Chapter 7, for example, suggests that there may be different levels of genre-specificity between the three languages, with RAs in Japanese having higher specificity. Findings in Chapter 8 indicate that Japanese discipline-specific keywords might be more exclusive than those from English and Portuguese.

Figure 10.1 – Guiding elements and research strategy of the fourth stage of the present study.



Source: Designed by the author.

The identification of cross-linguistically equivalent, high-frequency words was done with Sketch Engine's word list function. Separate word lists were generated for enPED, enMGT, ptPED, ptMGT, jaPED, and jaMGT, and then the 50 most frequent nouns of each list were manually extracted and assembled according to the discipline. Tables 10.1 and 10.2 show the results and highlight the selected nouns for the three languages and two disciplines.

Table 10.1 – Top 50 most frequent nouns in three Pediatrics corpora.

	enPED			ptPED			jaPED		
POS	Word	Freq.	Norm. Freq.	Word	Freq.	Norm. Freq.	Word	Freq.	Norm. Freq.
1	*study*	1,461	4,865.2	*estudo*	364	4,411.1	症例	305	4,659.3
2	*children*	1,222	4,069.3	*crianças*	321	3,890.0	考え	151	2,306.7
3	*health*	773	2,574.1	*saúde*	211	2,557.0	必要	120	1,833.2
4	*care*	732	2,437.6	*idade*	197	2,387.3	本症	116	1,772.0
5	*data*	713	2,374.3	*anos*	179	2,169.2	場合	106	1,619.3
6	*age*	708	2,357.7	*peso*	171	2,072.3	報告	100	1,527.6
7	*p*	607	2,021.3	*ser*	166	2,011.7	図	91	1,390.1
8	*child*	585	1,948.1	*dados*	138	1,672.3	女性医師	86	1,313.8
9	*parents*	534	1,778.2	*relação*	130	1,575.4	方法	84	1,283.2
10	*group*	532	1,771.6	*variáveis*	124	1,502.7	施設	84	1,283.2
11	*infants*	522	1,738.3	*mães*	116	1,405.7	治療	81	1,237.4
12	*results*	481	1,601.7	*análise*	112	1,357.3	手術	81	1,237.4
13	*studies*	470	1,565.1	*resultados*	112	1,357.3	結果	74	1,130.4
14	*months*	447	1,488.5	*meses*	107	1,296.7	もの	73	1,115.2
15	*risk*	445	1,481.9	*estudos*	106	1,284.6	保存的治療	64	977.7
16	*years*	440	1,465.2	*sono*	106	1,284.6	例	63	962.4
17	*mothers*	402	1,338.7	*adolescentes*	106	1,284.6	表	55	840.2
18	*patients*	383	1,275.4	*criança*	104	1,260.3	食道閉鎖症	55	840.2
19	*adolescents*	372	1,238.8	*prevalência*	103	1,248.2	自験例	54	824.9
20	*hospital*	359	1,195.5	*pacientes*	100	1,211.8	症状	53	809.6
21	*table*	345	1,148.9	*alimentos*	98	1,187.6	Fogarty カテーテル	51	779.1
22	*participants*	342	1,138.9	*qualidade*	94	1,139.1	検討	50	763.8
23	*CI*	339	1,128.9	*associação*	94	1,139.1	当科	49	748.5
24	*analysis*	336	1,118.9	*média*	89	1,078.5	対象	49	748.5
25	*birth*	318	1,058.9	*tempo*	88	1,066.4	当院	47	718.0
26	*level*	304	1,012.3	*nível*	87	1,054.3	可能性	47	718.0
27	*time*	301	1,002.3	*aleitamento*	85	1,030.1	子供	47	718.0
28	*mortality*	297	989.0	*amostra*	85	1,030.1	臍帯ヘルニア	47	718.0
29	*school*	296	985.7	*tratamento*	85	1,030.1	TEF	47	718.0
30	*association*	286	952.4	*número*	84	1,017.9	後	47	718.0
31	*rates*	286	952.4	*atividade*	83	1,005.8	全例	46	702.7
32	*n*	284	945.7	*pesquisa*	82	993.7	困難	46	702.7
33	*number*	276	919.1	*perda*	81	981.6	GER	45	687.4
34	*activity*	272	905.8	*medicamentos*	79	957.4	他	44	672.2
35	*control*	269	895.8	*alimentação*	78	945.2	p	44	672.2
36	*mean*	266	885.8	*risco*	77	933.1	患児	42	641.6
37	*sample*	264	879.1	*fatores*	76	921.0	目的	40	611.1
38	*rate*	262	872.5	*doenças*	76	921.0	小児外科医	38	580.5
39	*support*	261	869.1	*informações*	75	908.9	有無	36	549.9
40	*days*	259	862.5	*forma*	74	896.8	AW	36	549.9
41	*research*	258	859.1	*tabela*	73	884.6	術式	35	534.7
42	*status*	257	855.8	*período*	73	884.6	合併症	35	534.7
43	*levels*	255	849.2	*meio*	72	872.5	小児	34	519.4
44	*factors*	254	845.8	*lactentes*	72	872.5	平均	34	519.4
45	*outcomes*	248	825.8	*frequência*	72	872.5	際	34	519.4
46	*family*	242	805.9	*maioria*	70	848.3	傾向	34	519.4
47	*intervention*	242	805.9	*sexo*	67	811.9	日齢	34	519.4
48	*variables*	238	792.5	*teste*	66	799.8	男児	32	488.8
49	*groups*	234	779.2	*desenvolvimento*	64	775.6	腫瘍	31	473.6
50	*period*	218	725.9	*participantes*	64	775.6	本邦	31	473.6

Source: Nouns extracted from word lists generated by Sketch Engine. Numbers calculated by Sketch Engine. Colored rectangles highlight selected words. CI = confidence interval. An English translation of this table is provided in Appendix D together with a Romanized transcription of the Japanese words.

Table 10.2 – Top 50 most frequent nouns in three Management corpora.

	enMGT			ptMGT			jaMGT		
POS	Word	Freq.	Norm. Freq.	Word	Freq.	Norm. Freq.	Word	Freq.	Norm. Freq.
1	*research*	1,666	2,387.0	*relação*	459	2,614.0	こと	1,427	8,393.1
2	*work*	1,590	2,278.1	*marca*	395	2,249.5	企業	376	2,211.5
3	*study*	1,514	2,169.2	*resultados*	352	2,004.6	組織	352	2,070.3
4	*performance*	1,181	1,692.1	*pesquisa*	347	1,976.1	本稿	315	1,852.7
5	*status*	1,011	1,448.6	*estudo*	292	1,662.9	影響	305	1,793.9
6	*data*	918	1,315.3	*modelo*	288	1,640.1	考え	264	1,552.7
7	*business*	902	1,292.4	*empresas*	278	1,583.2	もの	260	1,529.2
8	*results*	840	1,203.5	*desempenho*	242	1,378.2	研究	239	1,405.7
9	*model*	803	1,150.5	*meio*	240	1,366.8	結果	239	1,405.7
10	*organizations*	792	1,134.8	*trabalho*	229	1,304.1	顧客	232	1,364.5
11	*job*	791	1,133.3	*processo*	223	1,270.0	場合	231	1,358.7
12	*time*	771	1,104.7	*análise*	216	1,230.1	メンバー	209	1,229.3
13	*firms*	758	1,086.1	*dados*	214	1,218.7	必要	206	1,211.6
14	*employees*	706	1,011.5	*consumidores*	211	1,201.6	知識	195	1,146.9
15	*level*	681	975.7	*Brasil*	193	1,099.1	点	180	1,058.7
16	*market*	679	972.9	*produtos*	190	1,082.0	組織変革	172	1,011.6
17	*analysis*	672	962.8	*estudos*	188	1,070.6	資源	169	994.0
18	*knowledge*	614	879.7	*p*	181	1,030.8	範囲	166	976.3
19	*relationship*	598	856.8	*desenvolvimento*	179	1,019.4	契約社員	162	952.8
20	*information*	592	848.2	*serviços*	178	1,013.7	分析	154	905.8
21	*process*	590	845.3	*mercado*	177	1,008.0	先行研究	144	847.0
22	*innovation*	583	835.3	*variáveis*	171	973.8	海外子会社	140	823.4
23	*role*	583	835.3	*administração*	170	968.1	仮説	135	794.0
24	*identity*	580	831.0	*efeito*	163	928.3	事例	134	788.1
25	*firm*	553	792.3	*valor*	162	922.6	プロジェクト	130	764.6
26	*change*	544	779.4	*vez*	158	899.8	関係	130	764.6
27	*findings*	543	778.0	*organização*	150	854.2	従業員	128	752.8
28	*leaders*	538	770.8	*base*	149	848.5	本研究	124	729.3
29	*table*	536	768.0	*imagem*	148	842.8	他	123	723.4
30	*number*	535	766.5	*empresa*	147	837.1	効果	123	723.4
31	*management*	532	762.2	*informações*	146	831.5	可能性	118	694.0
32	*value*	530	759.4	*processos*	143	814.4	対象	110	647.0
33	*studies*	526	753.6	*capacidade*	140	797.3	要因	108	635.2
34	*leadership*	523	749.3	*comportamento*	137	780.2	モデル	106	623.5
35	*experience*	518	742.2	*recursos*	136	774.5	問題	104	611.7
36	*feedback*	514	736.5	*ações*	134	763.1	業績	101	594.0
37	*women*	512	733.6	*critérios*	132	751.7	進化	101	594.0
38	*use*	510	730.7	*usuários*	132	751.7	成果	99	582.3
39	*factors*	509	729.3	*literatura*	132	751.7	イノベーション	93	547.0
40	*service*	507	726.4	*entrevistados*	130	740.3	行動	91	535.2
41	*variables*	507	726.4	*nível*	129	734.6	日本	90	529.3
42	*literature*	500	716.4	*país*	129	734.6	異	90	529.3
43	*cent*	499	715.0	*organizações*	124	706.2	正社員	88	517.6
44	*people*	489	700.6	*gestão*	123	700.5	概念	81	476.4
45	*context*	489	700.6	*IDORT*	122	694.8	変革	79	464.6
46	*effect*	477	683.4	*intenção*	121	689.1	デザイン部門	78	458.8
47	*respondents*	474	679.1	*forma*	121	689.1	既存研究	78	458.8
48	*members*	471	674.8	*práticas*	120	683.4	変数	77	452.9
49	*example*	460	659.1	*acordo*	117	666.3	製品	77	452.9
50	*organization*	450	644.8	*relações*	116	660.6	成長	77	452.9

Source: Nouns extracted from word lists generated by Sketch Engine. Numbers calculated by Sketch Engine. Colored rectangles highlight selected words. An English translation of this table is provided in Appendix D together with a Romanized transcription of the Japanese words.

As can be seen, the words chosen are *children*, *crianças* ("children"), 子供 (*kodomo*, "child" or "children"), *organization*, *organização* ("organization"), 組織 (*soshiki*, "organization"), and 患児 (*kanji*, "child patient"). The first six words were chosen because of their equivalence in meaning and disciplinary representativeness, in addition to their frequency in the corpora. 患児 (*kanji*) was included as an additional target because 子供 (*kodomo*) occurs in only one RA and because of the semantic proximity of 患児 (*kanji*) with *children* and *crianças*.

The distributional analysis of the seven selected words started with a simple frequency count using AntConc's KWIC Tool. This step was followed by the calculation of the percentages of occurrences of the words in each of the RA section subcorpus. Finally, the analysis was completed with the classification of the plot graphs generated by AntConc showing the distribution of the occurrences of the words across the RA sections.

The absolute and relative frequencies of each selected word are presented in Tables 10.3 (Pediatrics) and 10.4 (Management), according to RA sections. The overall numbers are different from those found in Tables 10.1 and 10.2 due to differences in data processing between AntConc, employed this time, and Sketch Engine, employed earlier.

Table 10.3 – Instances of cross-linguistically comparable words plus 患児 (*kanji*, "child patient") across Pediatrics RA sections.

	Introduction		Methods		Results		Discussion & Conclusion	
Word (Total)	n	(%)	n	(%)	n	(%)	n	(%)
children (1,177)	259	22.0	238	20.2	253	21.5	427	36.3
crianças (288)	50	17.4	66	22.9	51	17.7	121	42.0
子供 (43)	0	0.0	2	4.7	26	60.5	15	34.9
患児 (41)	3	7.3	4	9.8	7	17.1	27	65.9

Source: Prepared by the author with the aid of AntConc.

Table 10.4 – Instances of cross-linguistically comparable words across Management RA sections.

	Opening Section		Middle Sections		Closing Section	
Word (Total)	n	(%)	n	(%)	n	(%)
organization (539)	74	13.7	393	72.9	72	13.4
organização (150)	21	14.0	117	78.0	12	8.0
組織 (995)	122	12.3	757	76.1	116	11.7

Source: Prepared by the author with the aid of AntConc.

As can be seen, *children* and *crianças* exhibit a similar distribution across Pediatrics RA sections, with more instances occurring in discussion and conclusion sections. 子供 (*kodomo*), by contrast, seems to be positively primed to occur in sections of results and negatively primed to avoid introductions. 患児 (*kanji*) is likely to be primed for use mainly in discussion and conclusion sections. *Organization*, *organização*, and 組織 (*soshiki*) exhibit a similar distribution across Management RA sections, which suggests more similarity in behavior for these words.

The results of the analysis of the plot graphs are summarized in Tables 10.5 (Pediatrics) and 10.6 (Management) on the following pages. The *n* values refer to the number of graphs, which, in turn, correspond to the number of RA sections where the selected words occur. For example, *children* appear in 50 introductions, of which 20 (40%) exhibit it in all three thirds of the graphs, 15 (30%) in two thirds, and the remaining 15 (30%) in only one third.

Table 10.5 – Overview of plot graphs for cross-linguistically comparable words plus 患児 (*kanji*, "child patient") across Pediatrics RA sections.

Section and Class	*children* n	*children* (%)	*crianças* n	*crianças* (%)	子供 n	子供 (%)	患児 n	患児 (%)
Introduction								
Broadly Distributed	20	40.0	2	12.5	0	–	0	–
Moderately Distributed	15	30.0	7	43.8	0	–	1	50.0
Concentrated	15	30.0	7	43.8	0	–	1	50.0
Initial Position	*4*	*8.0*	*1*	*6.3*	*0*	–	*0*	–
Middle Position	*5*	*10.0*	*2*	*12.5*	*0*	–	*1*	*50.0*
Final Position	*6*	*12.0*	*4*	*25.0*	*0*	–	*0*	–
Total	**50**	**100.0**	**16**	**100.0**	**0**	**–**	**2**	**100.0**
Methods								
Broadly Distributed	13	26.0	5	33.3	0	–	0	–
Moderately Distributed	18	36.0	5	33.3	0	–	0	–
Concentrated	19	38.0	5	33.3	1	100.0	4	100.0
Initial Position	*12*	*24.0*	*1*	*6.7*	*0*	–	*1*	*25.0*
Middle Position	*4*	*8.0*	*3*	*20.0*	*1*	*100.0*	*1*	*25.0*
Final Position	*3*	*6.0*	*1*	*6.7*	*0*	–	*2*	*50.0*
Total	**50**	**100.0**	**15**	**100.0**	**1**	**100.0**	**4**	**100.0**
Results								
Broadly Distributed	15	40.5	4	36.4	1	100.0	0	–
Moderately Distributed	13	35.1	4	36.4	0	–	2	66.6
Concentrated	9	24.3	3	27.3	0	–	1	33.4
Initial Position	*3*	*8.1*	*2*	*18.2*	*0*	–	*1*	*33.4*
Middle Position	*2*	*5.4*	*1*	*9.1*	*0*	–	*0*	–
Final Position	*4*	*10.8*	*0*	–	*0*	–	*0*	–
Total	**37**	**100.0**	**11**	**100.0**	**1**	**100.0**	**3**	**100.0**
Discussion & Conclusion								
Broadly Distributed	29	56.9	7	38.9	0	–	0	–
Moderately Distributed	14	27.5	5	27.8	1	100.0	4	28.6
Concentrated	8	15.7	6	33.3	0	–	10	71.4
Initial Position	*3*	*5.9*	*1*	*5.6*	*0*	–	1	*7.1*
Middle Position	*2*	*3.9*	*2*	*11.1*	*0*	–	1	*7.1*
Final Position	*3*	*5.9*	*3*	*16.7*	*0*	–	8	*57.1*
Total	**51**	**100.0**	**18**	**100.0**	**1**	**100.0**	**14**	**100.0**

Source: Prepared by the author based on plot graphs generated by AntConc.

Table 10.6 – Overview of plot graphs for cross-linguistically comparable words across Management RA sections.

Section and Class	*organization* n	*organization* (%)	*organização* n	*organização* (%)	組織 n	組織 (%)
Opening Section						
Broadly Distributed	5	27.8	2	22.2	7	41.2
Moderately Distributed	8	44.4	1	11.1	5	29.4
Concentrated	5	27.8	6	66.7	5	29.4
Initial Position	*2*	*11.1*	*3*	*33.3*	*1*	*5.9*
Middle Position	*1*	*5.6*	*2*	*22.2*	*1*	*5.9*
Final Position	*2*	*11.1*	*1*	*11.1*	*3*	*17.6*
Total	**18**	**100.0**	**9**	**100.0**	**17**	**100.0**
Middle Sections						
Broadly Distributed	11	33.3	5	26.3	10	52.6
Moderately Distributed	13	39.4	7	36.8	7	36.8
Concentrated	9	27.3	7	36.8	2	10.5
Initial Position	*5*	*15.2*	*4*	*21.1*	*1*	*5.3*
Middle Position	*1*	*3.0*	*3*	*15.8*	*1*	*5.3*
Final Position	*3*	*9.1*	*0*	–	*0*	–
Total	**33**	**100.0**	**19**	**100.0**	**19**	**100.0**
Closing Section						
Broadly Distributed	3	16.7	6	31.6	7	50.0
Moderately Distributed	6	33.3	6	31.6	2	14.3
Concentrated	9	50.0	7	36.8	5	35.7
Initial Position	*1*	*5.6*	*4*	*21.1*	*2*	*14.3*
Middle Position	*1*	*5.6*	*3*	*15.8*	*0*	–
Final Position	*7*	*38.9*	*0*	–	*3*	*21.4*
Total	**18**	**100.0**	**19**	**100.0**	**14**	**100.0**

Source: Prepared by the author based on plot graphs generated by AntConc.

As can be seen, *children* and *crianças* seem to be primed for use in all RA sections. 患児 (*kanji*) appears to be primed for use mainly in the final part of discussion and conclusion sections. As for 子供 (*kodomo*), because its instances are limited to but a single RA, there is no conclusion to be drawn except that it may be inclined to be left out of Pediatrics RAs. The more technical 患児 (*kanji*) is perhaps preferred by Japanese Pediatrics RA authors.

With respect to *organization*, *organização*, and 組織 (*soshiki*), the plot graphs demonstrate that these words basically occur throughout all parts of the Management RAs. The only candidate for favorite place is the final part of the closing section of English RAs. The plot graphs suggest that whenever *organization* occurs in a small area of a given Management RA closing section, it will probably be found in the final part of the section.

The collocational analysis was performed using Sketch Engine's concordance function and its collocation option. As explained in Chapter 6, the 50 strongest collocates with a minimum frequency in corpus of 2 within a range of five words from the node (-5 to +5) were extracted and then classified cross-linguistically into three groups, namely *equivalents*, *semantically related*, and *unrelated*. Even though log-likelihood was used to identify the collocates, logDice scores were calculated later to compare the results across the three languages. Additionally, the proportion of co-occurrences between collocate and node against the total number of occurrences of the node was also calculated in this stage.

The top strongest collocates of *children*, *crianças*, 子供 (*kodomo*), and 患児 (*kanji*) are shown in Tables 10.7 to 10.10. While Tables 10.7 and 10.9 provide the original lists in decreasing order of log-likelihood scores, Tables 10.8 and 10.10 display the same results but with logDice scores for cross-linguistic comparison. Blue, green, yellow, brown, and gray tones indicate cross-linguistic equivalents, for example *with* and *com* ("with"). The red and pink tones indicate semantically related collocates, more specifically English and Portuguese words that express health conditions, for example *sick* and *saudável* ("healthy"), English and Japanese words related to family, for example *parents* and 家族 (*kazoku*, "family"), and Portuguese and Japanese words that represent medical care activities: *internações* ("hospitalization") and 経管栄養 (*keikan eiyō*, "tube feeding").

Table 10.7 – Collocates of cross-linguistically comparable Pediatrics words (*children*, *crianças*, and 子供, *kodomo*).

	enPED			ptPED			jaPED (Filtered)		
POS	Collocate	PER	Log-LKH	Collocate	PER	Log-LKH	Collocate	PER	Log-LKH
1	*of*	48.3	2,449.798	*de*	55.1	604.621	いる	61.7	423.390
2	*in*	32.9	1,652.038	*em*	26.5	376.997	女性医師	61.7	333.395
3	*with*	23.2	1,343.015	*anos*	15.0	332.996	の	83.0	184.738
4	*the*	36.1	1,251.080	*com*	22.7	324.254	25 名	14.9	108.185
5	*and*	27.1	985.348	*menores*	10.9	306.040	で	53.2	103.883
6	*to*	20.4	792.089	*das*	15.0	272.829	ない	27.7	95.270
7	*their*	9.1	584.846	*as*	17.8	271.866	既婚女性医師	8.5	45.904
8	*for*	12.4	520.013	*e*	24.6	218.123	は	34.0	45.813
9	*who*	7.4	486.182	*dois*	8.7	207.544	あり	14.9	40.287
10	*aged*	4.8	466.841	*que*	17.1	184.870	について	8.5	35.641
11	*months*	5.6	374.163	*a*	20.9	153.943	家庭	6.4	33.052
12	*T1DM*	3.4	346.601	*para*	13.1	138.809	24 名	4.3	29.043
13	*adolescents*	5.0	345.040	*entre*	8.7	126.521	女性外科医	6.4	24.374
14	*were*	9.3	343.450	*nas*	5.3	102.901	持た	4.3	23.500
15	*years*	5.2	338.337	*o*	12.1	87.421	男性外科医	4.3	22.317
16	*young*	3.8	334.806	*um*	7.8	86.220	が	21.3	22.098
17	*parents*	5.3	327.162	*internações*	3.4	85.073	持つ	4.3	20.053
18	*a*	10.6	322.077	*às*	4.7	81.435	非常勤医師	4.3	19.045
19	*mothers*	4.6	297.198	*foram*	7.2	81.077	対応	4.3	17.005
20	*that*	7.4	285.115	*pneumonia*	3.4	79.072	考え	6.4	14.377
21	*are*	5.6	273.041	*TEA*	3.7	77.756	に	19.1	14.266
22	*among*	4.2	271.371	*prevalência*	4.4	74.780	うち	4.3	11.720
23	*ASD*	3.0	254.558	*por*	7.2	73.246	における	4.3	10.070
24	*age*	4.7	238.753	*saudável*	2.8	69.459	よう	4.3	9.729
25	*these*	3.8	196.154	*maior*	4.7	69.088	なる	4.3	9.570
26	*have*	4.1	194.167	*do*	9.3	68.801	ある	6.4	7.966
27	*study*	5.0	178.593	*alimentar*	3.1	66.540	を	12.8	7.562
28	*than*	3.6	178.531	*número*	3.7	65.278	も	6.4	7.253
29	*as*	5.2	176.889	*respiratórias*	2.8	65.121	より	4.3	6.593
30	*on*	4.7	172.903	*nos*	4.4	63.959	や	4.3	5.584
31	*by*	4.6	172.209	*doenças*	3.4	60.108	と	6.4	2.295
32	*activity*	2.8	172.168	*guia*	2.5	59.384	–	–	–
33	*had*	3.4	164.260	*anemia*	3.1	58.713	–	–	–
34	*physical*	2.5	163.769	*não*	5.9	55.677	–	–	–
35	*younger*	1.8	155.082	*hospitalizadas*	1.6	55.571	–	–	–
36	*sick*	1.4	149.950	*alimentação*	3.1	52.072	–	–	–
37	*from*	4.1	148.008	*características*	2.5	47.539	–	–	–
38	*undervaccinated*	1.2	131.743	*acompanhadas*	1.6	47.211	–	–	–
39	*HIV-infected*	1.1	129.894	*adolescentes*	3.1	45.783	–	–	–
40	*or*	4.1	125.370	*meses*	3.1	45.594	–	–	–
41	*number*	2.2	123.022	*uma*	5.0	45.160	–	–	–
42	*age-appropriately*	0.9	114.317	*idade*	3.7	44.614	–	–	–
43	*older*	1.4	111.887	*amamentadas*	1.2	44.445	–	–	–
44	*this*	3.4	108.082	*no*	5.6	43.037	–	–	–
45	*all*	2.7	107.499	*econômicas*	1.6	41.747	–	–	–
46	*health*	2.9	107.126	*até*	2.5	41.717	–	–	–
47	*more*	2.6	104.421	*CAPS*	1.9	41.692	–	–	–
48	*be*	3.0	103.253	*todas*	2.2	41.380	–	–	–
49	*household*	1.6	102.644	*foi*	5.3	40.356	–	–	–
50	*PA*	1.1	100.248	*dos*	4.7	39.636	–	–	–

Source: Collocates identified with Sketch Engine. PER = number of co-occurrences between collocate and node divided by the overall frequency of the node in the corpus. Colors indicate semantically equivalent (blue, green, yellow, and others) or related (red) words. T1DM = type 1 diabetes mellitus. ASD = autism spectrum disorder or adequate sleep duration. PA = physical activity or palmitic acid. An English translation of this table is provided in Appendix D together with a Romanized transcription of the Japanese words.

Table 10.8 – Collocates of cross-linguistically comparable Pediatrics words with logDice scores.

	enPED			ptPED			jaPED (Filtered)		
POS	Collocate	PER	LogDice	Collocate	PER	LogDice	Collocate	PER	LogDice
1	*of*	48.3	10.81	*de*	55.1	10.21	いる	61.7	13.55
2	*in*	32.9	10.83	*em*	26.5	10.95	女性医師	61.7	12.80
3	*with*	23.2	11.12	*anos*	15.0	11.62	の	83.0	8.45
4	*the*	36.1	9.99	*com*	22.7	10.92	25名	14.9	12.08
5	*and*	27.1	10.11	*menores*	10.9	11.54	で	53.2	8.44
6	*to*	20.4	10.22	*das*	15.0	11.27	ない	27.7	10.71
7	*their*	9.1	10.79	*as*	17.8	11.00	既婚女性医師	8.5	11.19
8	*for*	12.4	10.25	*e*	24.6	9.90	は	34.0	7.64
9	*who*	7.4	10.64	*dois*	8.7	11.15	あり	14.9	9.71
10	*aged*	4.8	10.48	*que*	17.1	10.25	について	8.5	10.79
11	*months*	5.6	10.38	*a*	20.9	9.57	家庭	6.4	10.80
12	*T1DM*	3.4	10.01	*para*	13.1	10.15	24名	4.3	10.39
13	*adolescents*	5.0	10.29	*entre*	8.7	10.50	女性外科医	6.4	10.34
14	*were*	9.3	9.91	*nas*	5.3	10.40	持た	4.3	10.33
15	*years*	5.2	10.28	*o*	12.1	9.48	男性外科医	4.3	10.30
16	*young*	3.8	10.12	*um*	7.8	10.00	が	21.3	7.21
17	*parents*	5.3	10.24	*internações*	3.4	10.01	持つ	4.3	10.22
18	*a*	10.6	9.65	*às*	4.7	10.17	非常勤医師	4.3	10.17
19	*mothers*	4.6	10.14	*foram*	7.2	9.98	対応	4.3	10.02
20	*that*	7.4	9.88	*pneumonia*	3.4	9.98	考え	6.4	8.96
21	*are*	5.6	10.01	*TEA*	3.7	10.04	に	19.1	6.70
22	*among*	4.2	10.04	*prevalência*	4.4	10.08	うち	4.3	9.29
23	*ASD*	3.0	9.81	*por*	7.2	9.83	における	4.3	8.92
24	*age*	4.7	9.92	*saudável*	2.8	9.74	よう	4.3	8.84
25	*these*	3.8	9.73	*maior*	4.7	10.00	なる	4.3	8.80
26	*have*	4.1	9.72	*do*	9.3	9.45	ある	6.4	7.57
27	*study*	5.0	9.54	*alimentar*	3.1	9.83	を	12.8	6.44
28	*than*	3.6	9.64	*número*	3.7	9.92	も	6.4	7.40
29	*as*	5.2	9.50	*respiratórias*	2.8	9.72	より	4.3	7.95
30	*on*	4.7	9.54	*nos*	4.4	9.93	や	4.3	7.62
31	*by*	4.6	9.54	*doenças*	3.4	9.83	と	6.4	5.96
32	*activity*	2.8	9.54	*guia*	2.5	9.57	–	–	–
33	*had*	3.4	9.56	*anemia*	3.1	9.76	–	–	–
34	*physical*	2.5	9.46	*não*	5.9	9.61	–	–	–
35	*younger*	1.8	9.12	*hospitalizadas*	1.6	8.97	–	–	–
36	*sick*	1.4	8.80	*alimentação*	3.1	9.68	–	–	–
37	*from*	4.1	9.42	*características*	2.5	9.49	–	–	–
38	*undervaccinated*	1.2	8.62	*acompanhadas*	1.6	8.96	–	–	–
39	*HIV-infected*	1.1	8.53	*adolescentes*	3.1	9.58	–	–	–
40	*or*	4.1	9.23	*meses*	3.1	9.58	–	–	–
41	*number*	2.2	9.21	*uma*	5.0	9.46	–	–	–
42	*age-appropriately*	0.9	8.19	*idade*	3.7	9.57	–	–	–
43	*older*	1.4	8.76	*amamentadas*	1.2	8.66	–	–	–
44	*this*	3.4	9.14	*no*	5.6	9.31	–	–	–
45	*all*	2.7	9.17	*econômicas*	1.6	8.95	–	–	–
46	*health*	2.9	9.17	*até*	2.5	9.42	–	–	–
47	*more*	2.6	9.14	*CAPS*	1.9	9.17	–	–	–
48	*be*	3.0	9.13	*todas*	2.2	9.32	–	–	–
49	*household*	1.6	8.90	*foi*	5.3	9.28	–	–	–
50	*PA*	1.1	8.50	*dos*	4.7	9.34	–	–	–

Source: Collocates identified with Sketch Engine. PER = number of co-occurrences between collocate and node divided by the overall frequency of the node in the corpus. Colors indicate semantically equivalent (blue, green, yellow, and others) or related (red) words. T1DM = type 1 diabetes mellitus. ASD = autism spectrum disorder or adequate sleep duration. PA = physical activity or palmitic acid. An English translation of this table is provided in Appendix D together with a Romanized transcription of the Japanese words.

Table 10.9 – Collocates of cross-linguistically comparable Pediatrics words (*children* and *crianças*) and 患児 (*kanji*, “child patient”).

	enPED			ptPED			jaPED (Filtered)		
POS	Collocate	PER	Log-LKH	Collocate	PER	Log-LKH	Collocate	PER	Log-LKH
1	of	48.3	2,449.798	de	55.1	604.621	の	78.6	149.158
2	in	32.9	1,652.038	em	26.5	376.997	を	50.0	85.603
3	with	23.2	1,343.015	anos	15.0	332.996	は	50.0	79.493
4	the	36.1	1,251.080	com	22.7	324.254	に	50.0	76.401
5	and	27.1	985.348	menores	10.9	306.040	と	33.3	53.057
6	to	20.4	792.089	das	15.0	272.829	が	28.6	33.695
7	their	9.1	584.846	as	17.8	271.866	で	28.6	32.127
8	for	12.4	520.013	e	24.6	218.123	や	14.3	31.143
9	who	7.4	486.182	dois	8.7	207.544	家族	7.1	27.441
10	aged	4.8	466.841	que	17.1	184.870	身体的特徴	4.8	25.685
11	months	5.6	374.163	a	20.9	153.943	伴う	7.1	25.069
12	T1DM	3.4	346.601	para	13.1	138.809	消化器外科疾患	4.8	21.870
13	adolescents	5.0	345.040	entre	8.7	126.521	合わせ	4.8	20.513
14	were	9.3	343.450	nas	5.3	102.901	に対して	7.1	19.554
15	years	5.2	338.337	o	12.1	87.421	経管栄養	4.8	19.505
16	young	3.8	334.806	um	7.8	86.220	する	7.1	19.469
17	parents	5.3	327.162	internações	3.4	85.073	へ	7.1	18.910
18	a	10.6	322.077	às	4.7	81.435	原疾患	4.8	18.034
19	mothers	4.6	297.198	foram	7.2	81.077	有する	4.8	17.463
20	that	7.4	285.115	pneumonia	3.4	79.072	必要	7.1	16.411
21	are	5.6	273.041	TEA	3.7	77.756	GERD	4.8	15.275
22	among	4.2	271.371	prevalência	4.4	74.780	状態	4.8	15.275
23	ASD	3.0	254.558	por	7.2	73.246	1 例	7.1	14.498
24	age	4.7	238.753	saudável	2.8	69.459	その	7.1	13.771
25	these	3.8	196.154	maior	4.7	69.088	対象	4.8	12.848
26	have	4.1	194.167	do	9.3	68.801	症状	4.8	12.533
27	study	5.0	178.593	alimentar	3.1	66.540	多く	4.8	11.366
28	than	3.6	178.531	número	3.7	65.278	および	4.8	10.844
29	as	5.2	176.889	respiratórias	2.8	65.121	場合	4.8	9.787
30	on	4.7	172.903	nos	4.4	63.959	として	4.8	8.571
31	by	4.6	172.209	doenças	3.4	60.108	よる	4.8	8.217
32	activity	2.8	172.168	guia	2.5	59.384	も	7.1	7.870
33	had	3.4	164.260	anemia	3.1	58.713	こと	4.8	4.961
34	physical	2.5	163.769	não	5.9	55.677	–	–	–
35	younger	1.8	155.082	hospitalizadas	1.6	55.571	–	–	–
36	sick	1.4	149.950	alimentação	3.1	52.072	–	–	–
37	from	4.1	148.008	características	2.5	47.539	–	–	–
38	undervaccinated	1.2	131.743	acompanhadas	1.6	47.211	–	–	–
39	HIV-infected	1.1	129.894	adolescentes	3.1	45.783	–	–	–
40	or	4.1	125.370	meses	3.1	45.594	–	–	–
41	number	2.2	123.022	uma	5.0	45.160	–	–	–
42	age-appropriately	0.9	114.317	idade	3.7	44.614	–	–	–
43	older	1.4	111.887	amamentadas	1.2	44.445	–	–	–
44	this	3.4	108.082	no	5.6	43.037	–	–	–
45	all	2.7	107.499	econômicas	1.6	41.747	–	–	–
46	health	2.9	107.126	até	2.5	41.717	–	–	–
47	more	2.6	104.421	CAPS	1.9	41.692	–	–	–
48	be	3.0	103.253	todas	2.2	41.380	–	–	–
49	household	1.6	102.644	foi	5.3	40.356	–	–	–
50	PA	1.1	100.248	dos	4.7	39.636	–	–	–

Source: Collocates identified with Sketch Engine. PER = number of co-occurrences between collocate and node divided by the overall frequency of the node in the corpus. Colors indicate semantically equivalent (blue, green, yellow, and others) or related (red and pink) words. T1DM = type 1 diabetes mellitus. ASD = autism spectrum disorder or adequate sleep duration. PA = physical activity or palmitic acid. An English translation of this table is provided in Appendix D together with a Romanized transcription of the Japanese words.

Table 10.10 – Collocates of cross-linguistically comparable Pediatrics words and 患児 (*kanji*, “child patient”) with logDice scores.

	enPED			ptPED			jaPED (Filtered)		
POS	Collocate	PER	LogDice	Collocate	PER	LogDice	Collocate	PER	LogDice
1	*of*	48.3	10.81	*de*	55.1	10.21	の	78.6	8.22
2	*in*	32.9	10.83	*em*	26.5	10.95	を	50.0	8.25
3	*with*	23.2	11.12	*anos*	15.0	11.62	は	50.0	8.04
4	*the*	36.1	9.99	*com*	22.7	10.92	に	50.0	7.93
5	*and*	27.1	10.11	*menores*	10.9	11.54	と	33.3	8.18
6	*to*	20.4	10.22	*das*	15.0	11.27	が	28.6	7.48
7	*their*	9.1	10.79	*as*	17.8	11.00	で	28.6	7.38
8	*for*	12.4	10.25	*e*	24.6	9.90	や	14.3	9.22
9	*who*	7.4	10.64	*dois*	8.7	11.15	家族	7.1	10.63
10	*aged*	4.8	10.48	*que*	17.1	10.25	身体的特徴	4.8	10.51
11	*months*	5.6	10.38	*a*	20.9	9.57	伴う	7.1	10.44
12	*T1DM*	3.4	10.01	*para*	13.1	10.15	消化器外科疾患	4.8	10.42
13	*adolescents*	5.0	10.29	*entre*	8.7	10.50	合わせ	4.8	10.36
14	*were*	9.3	9.91	*nas*	5.3	10.40	に対して	7.1	9.76
15	*years*	5.2	10.28	*o*	12.1	9.48	経管栄養	4.8	10.30
16	*young*	3.8	10.12	*um*	7.8	10.00	する	7.1	9.75
17	*parents*	5.3	10.24	*internações*	3.4	10.01	へ	7.1	9.67
18	*a*	10.6	9.65	*às*	4.7	10.17	原疾患	4.8	10.19
19	*mothers*	4.6	10.14	*foram*	7.2	9.98	有する	4.8	10.14
20	*that*	7.4	9.88	*pneumonia*	3.4	9.98	必要	7.1	9.25
21	*are*	5.6	10.01	*TEA*	3.7	10.04	*GERD*	4.8	9.89
22	*among*	4.2	10.04	*prevalência*	4.4	10.08	状態	4.8	9.89
23	*ASD*	3.0	9.81	*por*	7.2	9.83	1 例	7.1	8.88
24	*age*	4.7	9.92	*saudável*	2.8	9.74	その	7.1	8.74
25	*these*	3.8	9.73	*maior*	4.7	10.00	対象	4.8	9.49
26	*have*	4.1	9.72	*do*	9.3	9.45	症状	4.8	9.43
27	*study*	5.0	9.54	*alimentar*	3.1	9.83	多く	4.8	9.18
28	*than*	3.6	9.64	*número*	3.7	9.92	および	4.8	9.06
29	*as*	5.2	9.50	*respiratórias*	2.8	9.72	場合	4.8	8.79
30	*on*	4.7	9.54	*nos*	4.4	9.93	として	4.8	8.45
31	*by*	4.6	9.54	*doenças*	3.4	9.83	よる	4.8	8.35
32	*activity*	2.8	9.54	*guia*	2.5	9.57	も	7.1	7.41
33	*had*	3.4	9.56	*anemia*	3.1	9.76	こと	4.8	7.27
34	*physical*	2.5	9.46	*não*	5.9	9.61	–	–	–
35	*younger*	1.8	9.12	*hospitalizadas*	1.6	8.97	–	–	–
36	*sick*	1.4	8.80	*alimentação*	3.1	9.68	–	–	–
37	*from*	4.1	9.42	*características*	2.5	9.49	–	–	–
38	*undervaccinated*	1.2	8.62	*acompanhadas*	1.6	8.96	–	–	–
39	*HIV-infected*	1.1	8.53	*adolescentes*	3.1	9.58	–	–	–
40	*or*	4.1	9.23	*meses*	3.1	9.58	–	–	–
41	*number*	2.2	9.21	*uma*	5.0	9.46	–	–	–
42	*age-appropriately*	0.9	8.19	*idade*	3.7	9.57	–	–	–
43	*older*	1.4	8.76	*amamentadas*	1.2	8.66	–	–	–
44	*this*	3.4	9.14	*no*	5.6	9.31	–	–	–
45	*all*	2.7	9.17	*econômicas*	1.6	8.95	–	–	–
46	*health*	2.9	9.17	*até*	2.5	9.42	–	–	–
47	*more*	2.6	9.14	*CAPS*	1.9	9.17	–	–	–
48	*be*	3.0	9.13	*todas*	2.2	9.32	–	–	–
49	*household*	1.6	8.90	*foi*	5.3	9.28	–	–	–
50	*PA*	1.1	8.50	*dos*	4.7	9.34	–	–	–

Source: Collocates identified with Sketch Engine. PER = number of co-occurrences between collocate and node divided by the overall frequency of the node in the corpus. Colors indicate semantically equivalent (blue, green, yellow, and others) or related (red and pink) words. T1DM = type 1 diabetes mellitus. ASD = autism spectrum disorder or adequate sleep duration. PA = physical activity or palmitic acid. An English translation of this table is provided in Appendix D together with a Romanized transcription of the Japanese words.

As can be seen, while the columns for *children* and *crianças* are full of colors, the columns for 子供 (*kodomo*) and 患児 (*kanji*) are almost entirely white. The English and Portuguese collocates are relatively close to each other; the Japanese ones, however, contain distinguishing features: untranslatable particles, such as が (*ga*, which signals the subject) and を (*o*, which signals a direct object), and characteristic content words, such as 女性外科医 (*josei gekai*, “female surgeon”) and 必要 (*hitsuyō*, “need”). It is noteworthy that most of the English–Portuguese equivalents have similar logDice scores, for example, in Table 10.8, *in* and *em* (10.83 versus 10.95), *with* and *com* (11.12 versus 10.92), *ASD* and *TEA* (9.81 versus 10.04), and *all* and *todas* (9.17 and 9.32). The Japanese collocates, by contrast, often exhibit lower values than their English and Portuguese counterparts. This is illustrated by the pair *of* and の (*no*), also in Table 10.8, whose scores are respectively 10.81 and 8.45. Despite this, semantically related collocates in the three languages have relatively close logDice values, for example *T1DM* (type 1 diabetes mellitus), *pneumonia* (“pneumonia”), and *GERD* (“gastroesophageal reflux disease”), whose scores are 10.01, 9.98, and 9.89 (Table 10.10).

The top strongest collocates of *organization*, *organização*, and 組織 (*soshiki*) are shown in Tables 10.11 (log-likelihood) and 10.12 (logDice), on the following two pages. The green, yellow, orange, and purple tones indicate cross-linguistic equivalents, such as *members* and *membros* (“members”) and *the*, *a* (“the”), and *o* (“the”). The red and pink tones indicate semantically related collocates. Collocates with the #F08095 pink ( ), such as *behaviour* and *comprometimento* (“commitment”), refer to the individual dimension. Collocates with the #E1191E red ( ), such as *scandal-stricken*, *departamentos* (“departments”), and 組織慣性 (*soshiki gansei*, “organizational inertia”), refer to the organizational dimension. Collocates with the #941013 red ( ), such as *needs* and *desempenho* (“performance”), refer to both the individual and the organizational dimension (in some cases, they refer to employees or customers; in others, they refer to organizations themselves). Finally, collocates with the #C31736 red ( ), such as *theory* and *científica* (“scientific”), refer to the theoretical, abstract dimension.

Table 10.11 – Collocates of cross-linguistically comparable Management words (*organization*, *organização*, and 組織, *soshiki*).

POS	enMGT			ptMGT			jaMGT (Filtered)		
	Collocate	PER	Log-LKH	Collocate	PER	Log-LKH	Collocate	PER	Log-LKH
1	*the*	67.1	1,279.282	*da*	33.3	251.924	の	64.8	788.217
2	*of*	32.2	476.203	*a*	41.3	238.316	は	43.5	584.888
3	*to*	29.6	468.348	*de*	39.3	169.651	に	45.7	530.434
4	*an*	15.6	400.559	*e*	28.0	135.963	が	34.9	416.282
5	*in*	23.3	362.972	*uma*	18.0	130.297	を	32.1	327.014
6	*and*	24.9	323.625	*para*	17.3	103.515	で	26.4	296.817
7	*a*	17.8	259.462	*na*	14.0	97.954	強い	7.4	224.231
8	*their*	10.0	199.759	*que*	17.3	78.496	慣性	5.1	202.570
9	*for*	10.2	155.762	*trabalho*	8.0	76.763	や	11.6	177.228
10	*at*	6.4	143.228	*racional*	4.7	72.535	と	19.6	171.415
11	*that*	10.9	133.862	*do*	14.0	67.215	能力開発	5.4	165.974
12	*within*	4.7	132.644	*com*	12.0	67.071	も	11.4	142.402
13	*or*	6.9	126.424	*dentro*	4.7	55.152	な	12.5	134.722
14	*implicated*	2.2	125.994	*em*	11.3	52.234	弱い	3.7	126.597
15	*members*	4.2	121.325	*pela*	6.0	51.640	組織慣性	4.0	126.577
16	*as*	8.2	108.463	*o*	13.3	47.331	衰退	3.1	121.772
17	*with*	7.3	106.364	*trabalhador*	2.7	44.273	ある	9.9	104.267
18	*on*	6.9	97.347	*membros*	3.3	42.163	よる	5.7	103.584
19	*by*	5.8	88.414	*é*	7.3	38.773	従業員	4.5	102.489
20	*employees*	3.6	83.707	*à*	6.7	38.237	想定	3.4	98.600
21	*is*	6.9	78.975	*essa*	4.7	37.613	有効性	3.4	94.894
22	*from*	4.9	78.844	*sua*	4.7	34.876	この	5.7	74.560
23	*this*	5.8	77.825	*dos*	7.3	34.387	寄与する	2.3	74.373
24	*can*	4.0	76.187	*científica*	2.0	34.139	よう	5.7	72.848
25	*career*	2.2	72.790	*das*	6.0	32.683	意図	2.6	71.674
26	*scandal-stricken*	1.3	71.612	*comprometi-mento*	3.3	31.377	として	5.1	69.341
27	*focal*	1.8	63.617	*agir*	2.0	30.247	見方	2.0	68.740
28	*be*	4.7	62.937	*capacidades*	2.7	26.431	における	4.0	68.242
29	*theory*	2.4	59.857	*instituto*	2.7	26.192	から	6.0	67.775
30	*whole*	1.6	56.328	*funcionários*	2.7	25.516	硬直性	1.7	62.818
31	*may*	3.1	55.276	*contribuições*	2.0	25.257	その	5.1	58.737
32	*are*	4.7	54.953	*por*	6.0	25.157	エンプロイアビリティ保障	2.3	58.725
33	*it*	3.8	52.022	*como*	6.0	25.113	形態変化	1.1	57.795
34	*functioning*	1.1	48.528	*posição*	2.0	24.674	いう	4.8	53.573
35	*more*	3.3	47.768	*suas*	3.3	24.332	ため	4.0	52.018
36	*theorists*	0.9	46.450	*afetivo*	2.7	24.318	に対する	2.6	50.259
37	*such*	2.7	45.397	*expectativas*	2.0	23.894	欧米	1.4	49.523
38	*its*	2.2	44.780	*essencialmente*	1.3	22.746	戦略	2.3	48.192
39	*one*	2.7	44.460	*comprometidos*	1.3	22.746	他	2.6	47.539
40	*theorists'*	0.7	44.100	*funcionário*	1.3	22.746	こと	5.7	43.460
41	*work*	2.9	42.506	*desempenho*	3.3	22.522	著名	1.7	43.080
42	*not*	3.1	40.698	*virtuoso*	1.3	21.563	意思決定	2.0	42.480
43	*behaviour*	1.6	40.212	*permanecem*	1.3	21.563	権限	1.4	40.654
44	*what*	2.0	39.884	*Farace*	1.3	20.657	変革	2.0	39.663
45	*these*	2.9	39.466	*sentem*	1.3	19.921	成立する	1.1	37.141
46	*specific*	1.8	39.141	*um*	5.3	19.487	保証	1.1	37.141
47	*needs*	1.6	38.867	*administrativa*	1.3	19.301	組成し	0.9	37.106
48	*how*	2.4	38.807	*obrigações*	1.3	19.301	欠く	0.9	37.106
49	*social*	2.0	38.241	*ele*	2.0	19.144	集中し	1.1	36.050
50	*employee*	1.6	36.437	*departamentos*	1.3	18.765	へ	2.6	35.431

Source: Collocates identified with Sketch Engine. PER = number of co-occurrences between collocate and node divided by the overall frequency of the node in the corpus. Colors indicate semantically equivalent (green, yellow, orange, and purple) or related (red and pink) words. An English translation of this table is provided in Appendix D together with a Romanized transcription of the Japanese words.

Table 10.12 – Collocates of cross-linguistically comparable Management words with logDice scores.

	enMGT			ptMGT			jaMGT (Filtered)		
POS	Collocate	PER	LogDice	Collocate	PER	LogDice	Collocate	PER	LogDice
1	*the*	67.1	8.18	*da*	33.3	9.46	の	64.8	9.29
2	*of*	32.2	7.82	*a*	41.3	8.56	は	43.5	9.74
3	*to*	29.6	8.02	*de*	39.3	7.86	に	45.7	9.35
4	*an*	15.6	9.57	*e*	28.0	8.21	が	34.9	9.49
5	*in*	23.3	8.01	*uma*	18.0	9.36	を	32.1	9.15
6	*and*	24.9	7.57	*para*	17.3	8.79	で	26.4	9.39
7	*a*	17.8	7.88	*na*	14.0	9.24	強い	7.4	10.95
8	*their*	10.0	8.71	*que*	17.3	8.10	慣性	5.1	10.62
9	*for*	10.2	7.99	*trabalho*	8.0	10.02	や	11.6	10.01
10	*at*	6.4	8.92	*racional*	4.7	10.39	と	19.6	8.89
11	*that*	10.9	7.51	*do*	14.0	8.24	能力開発	5.4	10.58
12	*within*	4.7	9.44	*com*	12.0	8.61	も	11.4	9.57
13	*or*	6.9	8.42	*dentro*	4.7	10.07	な	12.5	9.27
14	*implicated*	2.2	9.46	*em*	11.3	8.14	弱い	3.7	10.14
15	*members*	4.2	9.40	*pela*	6.0	9.58	組織慣性	4.0	10.21
16	*as*	8.2	7.65	*o*	13.3	7.61	衰退	3.1	9.94
17	*with*	7.3	7.85	*trabalhador*	2.7	9.69	ある	9.9	9.18
18	*on*	6.9	7.79	*membros*	3.3	9.79	よる	5.7	9.99
19	*by*	5.8	7.95	*é*	7.3	8.42	従業員	4.5	10.09
20	*employees*	3.6	8.83	*à*	6.7	8.59	想定	3.4	9.96
21	*is*	6.9	7.36	*essa*	4.7	9.28	有効性	3.4	9.94
22	*from*	4.9	8.04	*sua*	4.7	9.10	この	5.7	9.41
23	*this*	5.8	7.67	*dos*	7.3	8.15	寄与する	2.3	9.47
24	*can*	4.0	8.37	*científica*	2.0	9.30	よう	5.7	9.37
25	*career*	2.2	9.07	*das*	6.0	8.45	意図	2.6	9.57
26	*scandal-stricken*	1.3	8.73	*comprometi-mento*	3.3	9.36	として	5.1	9.41
27	*focal*	1.8	8.92	*agir*	2.0	9.26	見方	2.0	9.30
28	*be*	4.7	7.64	*capacidades*	2.7	9.24	における	4.0	9.61
29	*theory*	2.4	8.67	*instituto*	2.7	9.23	から	6.0	9.18
30	*whole*	1.6	8.77	*funcionários*	2.7	9.19	硬直性	1.7	9.09
31	*may*	3.1	8.12	*contribuições*	2.0	9.17	その	5.1	9.13
32	*are*	4.7	7.37	*por*	6.0	7.89	エンプロイアビリティ保障	2.3	9.38
33	*it*	3.8	7.64	*como*	6.0	7.89	形態変化	1.1	8.53
34	*functioning*	1.1	8.43	*posição*	2.0	9.15	いう	4.8	9.05
35	*more*	3.3	7.69	*suas*	3.3	8.87	ため	4.0	9.20
36	*theorists*	0.9	8.16	*afetivo*	2.7	9.11	に対する	2.6	9.33
37	*such*	2.7	7.97	*expectativas*	2.0	9.13	欧米	1.4	8.83
38	*its*	2.2	8.23	*essencialmente*	1.3	8.73	戦略	2.3	9.26
39	*one*	2.7	7.93	*comprometidos*	1.3	8.73	他	2.6	9.28
40	*theorists'*	0.7	7.76	*funcionário*	1.3	8.73	こと	5.7	8.53
41	*work*	2.9	7.71	*desempenho*	3.3	8.71	著名	1.7	9.00
42	*not*	3.1	7.50	*virtuoso*	1.3	8.72	意思決定	2.0	9.10
43	*behaviour*	1.6	8.39	*permanecem*	1.3	8.72	権限	1.4	8.79
44	*what*	2.0	8.14	*Farace*	1.3	8.71	変革	2.0	9.06
45	*these*	2.9	7.56	*sentem*	1.3	8.71	成立する	1.1	8.50
46	*specific*	1.8	8.24	*um*	5.3	7.62	保証	1.1	8.50
47	*needs*	1.6	8.34	*administrativa*	1.3	8.70	組成し	0.9	8.11
48	*how*	2.4	7.80	*obrigações*	1.3	8.70	欠く	0.9	8.11
49	*social*	2.0	8.06	*ele*	2.0	8.90	集中し	1.1	8.50
50	*employee*	1.6	8.24	*departamentos*	1.3	8.69	へ	2.6	8.95

Source: Collocates identified with Sketch Engine. PER = number of co-occurrences between collocate and node divided by the overall frequency of the node in the corpus. Colors indicate semantically equivalent (green, yellow, orange, and purple) or related (red and pink) words. An English translation of this table is provided in Appendix D together with a Romanized transcription of the Japanese words.

As can be inferred from the colors, the three Management words have more affinities between each other than the Pediatrics words. Cross-linguistically equivalent words are many and generally have close logDice values. Some examples with their respective scores are the following: *of* (8.16), *de* (7.86), and の (*no*, 9.29); *for* (7.99), *para* (8.79), and ため (*tame*, 9.20); *employees* (8.83) and *funcionários* (9.19); *members* (9.40) and *membros* (9.79); *essa* (9.28) and その (*sono*, 9.13). Among them, only の (*no*) and *for* exhibit a difference of more than 0.50 in relation to their equivalents. Furthermore, the amount of semantically related words exceeds by far the amounts observed in the Pediatrics lists. While Tables 10.11 and 10.12 contain 35 words, Tables 10.7 and 10.8 (子供, *kodomo*) contain only seven, and Tables 10.9 and 10.10 (患児, *kanji*), 17 words. The logDice scores range from 8.39 to 9.36 among the individual dimension group members ( ), from 8.69 to 10.95 among the organizational dimension group members ( ), from 8.34 to 9.07 among the individual and organizational dimension group members ( ), and from 7.76 to 10.39 among the theoretical dimension group members ( ), which shows that semantically related words from different languages may exhibit a difference of more than 2.0 in logDice scores. Yet altogether the results above suggest that *organization*, *organização*, and 組織 (*soshiki*) share similar semantic gravitational fields.

The final part of the fourth stage of this study consisted in a colligational analysis at the clause- and sentence-level based on concordance lines sampled (generated) by Sketch Engine. Seven broad categories were used to classify the instances of the target words: (1) *subject*, (2) *object or complement*, (3) *agent of the passive*, (4) *noun adjunct*, (5) *adverbial adjunct*, (6) *appositive, parentheses-enclosed phrase, title, name*, and (7) *others*. Table 10.13 shows the results for the Pediatrics words.

Table 10.13 – Functions of cross-linguistically comparable Pediatrics words and 患児 (*kanji*, "child patient").

	children		*crianças*		子供		患児	
Category	n	(%)	n	(%)	n	(%)	n	(%)
Subject	37	37.0	24	24.0	40	85.1	17	40.5
Object or Complement	11	11.0	6	6.0	5	10.6	5	11.9
Noun Adjunct	33	33.0	39	39.0	2	4.3	13	31.0
Agent of the Passive	1	1.0	0	–	0	–	0	–
Adverbial Adjunct	13	13.0	23	23.0	0	–	7	16.7
Appositive, Parentheses-Enclosed Phrase, Title, Name	3	3.0	4	4.0	0	–	0	–
Others	2	2.0	4	4.0	0	–	0	–
Total	100	100.0	100	100.0	47	100.0	42	100.0

Source: Prepared by the author based on concordance lines extracted by Sketch Engine.

As can be seen, *children* seems to be primed for use as either subject or part of a noun adjunct. *Crianças*, by contrast, shows an inclination toward adjuncts. 子供 (*kodomo*) is primed for use as subject, generally followed by the particle の (*no*, "of") and a verb in such cases as 子供のいる女性医師 (*kodomo no iru josei ishi*, "female doctors with children") and 子供のいない女性医師 (*kodomo no inai josei ishi*, "female doctors without children"), which can be segmented as 子供のいる (*kodomo no iru*, "with children" or "there are children") plus 女性医師 (*josei ishi*, "female doctors") and 子供のいない (*kodomo no inai*, "without children" or "there are no children") plus 女性医師 (*josei ishi*). 患児 (*kanji*) is primed for use as both subject and part of a noun adjunct. As subject, it is often preceded by a verb and an object, for example *GERD* を発症した患児 (*GERD o hasshō shita kanji*, "child patient who developed gastroesophageal reflux disease").

The results concerning the Management words can be found in Table 10.14.

Table 10.14 – Functions of cross-linguistically comparable Management words.

Category	*organization*		*organização*		組織	
	n	(%)	n	(%)	n	(%)
Subject	9	9.0	13	13.0	15	15.0
Object or Complement	21	21.0	16	16.0	15	15.0
Noun Adjunct	35	35.0	42	42.0	42	42.0
Agent of the Passive	1	1.0	6	6.0	0	–
Adverbial Adjunct	26	26.0	17	17.0	25	25.0
Appositive, Parentheses-Enclosed Phrase, Title, Name	6	6.0	4	4.0	2	2.0
Others	2	2.0	2	2.0	1	1.0
Total	100	100.0	100	100.0	100	100.0

Source: Prepared by the author based on concordance lines extracted by Sketch Engine.

Colligations involving *organization*, *organização*, and 組織 (*soshiki*) seem to be relatively alike. However, whereas all three words are primed for use in adjuncts, *organization* is more inclined to be used as an object or complement rather than *organização* and 組織 (*soshiki*). Some verbs preceding *organization* as object are *align*, *benefit*, *trust*, and *perceive*. *Organização* shows a stronger inclination to function as agent of the passive, always preceded by *pela* ("by"). 組織 (*soshiki*) works as both subject and object (complement) in an equal proportion, being followed by the particles が (*ga*) and は (*wa*) when it works as subject and mainly by を (*o*) and に (*ni*) when it works as object or complement.

Do users of different languages make similar associations with semantically equivalent words in comparable contexts? Overall, the answer provided by this chapter is both yes and no. There are more likenesses between English and Portuguese than between English and Japanese or Portuguese and Japanese. Moreover, there are more likenesses among Management words than among Pediatrics words. Considering Portuguese and English, it seems that in many cases RA authors do make similar associations with semantically equivalent words, such as *children* and *crianças* and *organization* and *organização*. Japanese has such distinguishing features that its collocations and colligations appear to follow other standards. Considering the two disciplines, however, it seems that Management authors, even the Japanese, do make similar associations with semantically equivalent words.

Alluding once more to Hoey's (2005) central hypotheses, the results reported in this chapter can be summarized in the following way:

Cross-linguistically equivalent words are primed to occur in, or avoid, certain positions in discourse, to occur with particular other words, for use in one or more grammatical roles; and these (i.e., the positions in discourse, the other words, and the grammatical roles) may be similar or not, depending on the languages and domains involved.

The main implications of the findings can be described as follows. For language learning and teaching, they shed light on the need for a holistic approach. Whereas *children* and *crianças* ("children") seem to be common in RA writing, the same cannot be said about the Japanese 子供 (*kodomo*, "child" or "children"). It is likely that 患児 (*kanji*, "child patient"), which does not have an exact equivalent in English and Portuguese, is the preferred choice for the kind of discourse in question. In addition, the textual position, collocates, and grammatical functions of the three Japanese words exhibit distinguishing features in relation to English and Portuguese, despite some commonalities. In view of this, a speaker of English or Portuguese who is learning Japanese and overfocus on individual words will possibly produce lexically dislocated texts. The same might happen with Japanese native speakers who learn either English or Portuguese. If they place a great emphasis on individual words, the texts they produce may become lexically dislocated. With respect to the translation of academic texts, the same line of thought can be applied. There is the need for a holistic approach, since the behavior of cross-linguistically equivalent words may not be the same, depending on the languages and disciplines involved. Therefore, it seems important that translators of academic texts check not only whether the word in the target language matches that from the source but also whether its associations with textual position, surrounding words, and grammatical roles agree with the patterns found in reference texts originally written in the target language while having some degree of correspondence with the source.

11 CONCLUDING REMARKS

The present study has investigated signs of psychological priming in the RA genre. By means of a corpus-driven, multi-stage approach, four aspects were addressed: genre-specificity, disciplinary variation, text-positional association, and cross-linguistic variation. Our main conclusions can be summarized as follows.

The findings shown in Chapter 7 provide strong support for Hoey's (2004, 2005, 2013) claims about genre-specificity. Genre-specific key keywords were identified from comparisons between the specialized, single-genre and the general, multi-genre corpora in English, Portuguese, and Japanese, demonstrating that RAs in the three languages contain typical words. Moreover, collocations and semantic associations of selected genre-specific keywords differ according to the type of corpus in question, whether specialized, single-genre or general, multi-genre. This study's results agree with those by Pace-Sigge (2007) and Chuang (2015), who have also confirmed genre-specificity. There is evidence for genre-specificity related to psychological priming in RAs.

As explained in Chapter 5, the descriptions of the reference corpora do not include their constitutive genres, but it is possible to infer part of them based on the specific domains provided by these descriptions. In fact, the domain names listed in Table 5.10 lead to several types of text, such as news, blog posts, dictionary entries, public job announcements, short book reviews, religion-related articles, and promotional texts. Despite the fact that the three main reference corpora were built in 2011 and 2015, it is possible to assume that the types above represent a sample of the genres that formed enTenTen15, ptTenTen11, and jaTenTen11 LUW. The results presented in Chapter 7, therefore, suggest that writers of RAs make distinguishing associations in comparison to writers of news, blog posts, job announcements, and book reviews, among other genres.

The findings in Chapter 8 provide solid support for Hoey's (2004, 2005, 2007b) claims concerning domain-specificity. Key keywords exclusive to either Pediatrics or Management arose from comparisons involving enPED, enMGT, ptPED, ptMGT, jaPED, and jaMGT. Furthermore, collocations and semantic associations of selected non-discipline-specific keywords differ according to the corpus and discipline. Pediatrics data exhibit certain patterns; Management data, others. This study's results add to previous priming analyses that considered broad domains (Salim, 2012; Cunha, 2017; Duguid & Partington, 2017) suggesting that corpora in specific fields of knowledge or disciplines have a high degree of typicality. There is evidence for disciplinary variation related to psychological priming in RAs. The results indicate that RA writers in one field and in another make distinct associations.

The findings in Chapter 9 provide partial support for Hoey's (2004, 2005, 2013) claims with respect to textual colligation. Overall, discipline-specific key keywords do not appear to have typical positions in RAs. Based on our findings, it seems that the pervasiveness of textual colligations does not manifest itself in the RA genre with respect to discipline-specific keywords. Whenever such words combine with other items or nest, however, they do tend to appear in exclusive RA parts, thus exhibiting traces of textual colligation. This study's results add to previous research dedicated to text-positional associations (Hoey & O'Donnel, 2008a, 2008b, 2015; O'Donnel et al., 2012) by showing the relationship between textual position and selected words across three languages and two disciplines. There is evidence for position-specificity related to psychological priming in RAs, but it relates to clusters or word combinations rather than to individual words. It is likely that RA writers make associations between particular clusters and specific parts of RAs.

The findings in Chapter 10 agree with those reported by other researchers who have applied lexical priming to the study of languages other than English (Pace-Sigge, 2007; Jantunen, 2017; Wang, 2018) in that they confirm that lexical priming can be used with this purpose. Furthermore, the findings suggest that users of languages relatively close such as English and Portuguese tend to make similar associations in comparable situations, while users of languages markedly different such as English and Japanese or

Portuguese and Japanese tend to make different associations in comparable situations, despite some points of convergence. There is evidence for both cross-linguistic similarities and differences related to psychological priming in RAs. Depending on the languages involved and the disciplines considered, RA writers of different languages may make similar associations.

As shown in Chapter 3, the body of research related to lexical priming has been growing. However, to the best of our knowledge, the present study is the only dedicated to a single genre in three languages and two disciplines. Due to its exclusive scope, there are some lessons that can be drawn from it. They are as follows.

The first lesson refers to the study of collocations in Japanese. Most of the analyses performed had a span of 10 words, that is, they considered five words to the left and five words to the right of the node (-5 to +5). To include verbs that occur far apart from the subject (something very common in Japanese), it may be interesting to extend the span to 20 (-10 to +10) or even 40 words (-20 to +20). It is noteworthy that data processing methods such as that of Sketch Engine may acknowledge word fragments as words, which makes the adoption of a broad span also a possible solution to reach words beyond fragments surrounding the node. As for cross-linguistic investigations, however, the impact on results of such an increase in span should be taken into account, as a span between 20 and 40 words may not work well with languages such as English and Portuguese. In the face of this, it may be convenient to adopt a fluid criterion for span based on language, for example a span of 10 words for English (-5 to +5) and one of 20 for Japanese (-10 to +10), provided that there is some kind of justifiable correspondence between the spans.

The second lesson refers to data processing. Corpus analysis software programs acknowledge English and Portuguese words mainly by their surrounding spaces. As shown in Chapter 5, Japanese texts must go through a segmentation process so that these programs recognize Japanese words and analyze them as such. However, the segmentation process has certain characteristics that interfere with further data analysis. To improve cross-linguistic studies involving Japanese, it would be effective to follow a semantic rather than a morphological criterion for Japanese segmentation, which, however, might demand different technologies. Artificial intelligence, recently defined in a specialized editorial as “the automation of cognition” (Abbas, 2021, p. 95), may be the solution to this issue.

The third lesson refers to corpus building. In this study, corpus building has considered language, country of origin, discipline, genre, publication period, availability, and technical specifications. Among them, language, discipline, and genre were the primary criteria. Genre identification was based on manual examination of journals’ electronic pages, tables of contents, and texts. Although this process is time-consuming, it was necessary to avoid the undesirable inclusion of other journal genres, such as the editorial and the book review. Manual examination has proved to be a critical step for priming analysis focused on well-defined genres, even though it may be demanding.

The fourth lesson refers to the scope of lexical priming. Jantunen (2017) has pointed to the need to expand its scope so that particular aspects of Finnish can be considered. The analysis of Japanese data has demonstrated the need to consider two additional categories of priming. Japanese texts are made of four writing systems: 漢字 (*kanji*), a complex script primarily composed of thousands of Chinese characters; 平仮名 (*hiragana*), a Japanese syllabary composed of 46 letters; 片仮名 (*katakana*), another Japanese syllabary also composed of 46 letters; and the Roman alphabet. The same word can be written with different systems, and each written form will have its own conventions and preferred contexts for use. To illustrate this point, some additional searches were performed with Sketch Engine leading to the following results. In jaRAs, there are 47 instances of the *kanji*-based 子供 (*kodomo*, “child,” “children”), one instance of the *hiragana*-based こども (*kodomo*), and one instance of the *kanji-hiragana*-mixed 子ども (*kodomo*); in jaTenTen11 LUW, by contrast, there are 35,671 instances of 子供 (*kodomo*), 2,006 instances of こども (*kodomo*), and 16,309 instances of 子ども (*kodomo*). In jaRAs, there are zero instances of the *kanji*-based 即ち (*sunawachi*, “namely,” “that is”) and four instances of the *hiragana*-based すなわち (*sunawachi*); in jaTenTen11 LUW, by contrast, there are 1,230 instances of 即ち (*sunawachi*) and 4,686 instances of

すなわち (*sunawachi*). Japanese users possibly associate written forms with genre and domain, another type of priming. Furthermore, Japanese has different levels of politeness with marked forms in both speech and writing. While sentences in academic Japanese normally end with short-form verbs (known as 普通形, *futsū-kei*, "normal form," or 辞書形, *jisho-kei*, "dictionary form"), sentences in general Japanese often end with long-form verbs (known as 丁寧形, *teinei-kei*, "polite form," or ます形, *masu-kei*, "*masu* form"). To illustrate this point, again additional searches were done with Sketch Engine generating the following figures. In jaRAs, there are two instances of the short-form 調べた (*shirabeta*, "investigated") and zero instances of the long-form 調べました (*shirabemashita*, "investigated"); in jaTenTen11 LUW, by contrast, there are 1,979 instances of 調べた (*shirabeta*) and 422 instances of 調べました (*shirabemashita*). Japanese users possibly associate politeness levels with genre and situation, which is another kind of priming to be considered.

The fifth lesson refers to data analysis. As Chapters 7 to 10 have shown, two approaches were employed for data analysis in this study. The first was used in collocational, semantic associational, and colligational analyses and was based on manual examination of collocate lists and concordance lines. The second was used in collocational and textual colligational analyses and was based on cluster lists as well as on dispersion values and plot graphs. While the former involved direct contact with data, the latter consisted in an indirect approach. The indirect approach proved to be useful to deal with large amount of data while focusing on local aspects.

The present study has limitations that may be overcome in further research. First, it has dealt with collocations, semantic associations, colligations, and textual colligations, but it has not covered other kinds of association, for example textual collocations. Consequently, the extent to which its conclusions also apply to other classes of priming remains unknown. In fact, there may be classes less impacted by discipline. Grammatical categories (priming), for example, is one kind that may be less subject to disciplinary variation, since it is arguable that grammatical choices may bear a stronger relationship with genre-or register-specificity. Future priming analyses of RA corpora could investigate one or more of the unexplored classes to deepen the understanding of the topic.

Another limitation relates to discipline. Even though differences between Pediatrics and Management data have become clear, it is not possible to conclude that data in other fields, such as Biology, Engineering, and Chemistry, have the same degree of differentiation. Further priming analyses could investigate RAs from other disciplines to expand the understanding of disciplinary variation.

Following the mainstream line of lexical priming research (Hoey, 2007a, 2007b; Pace-Sigge, 2010; Patterson, 2015; etc.), the present study focused on corpus data, which are a source of clues for psychological processes. No attempt was made to effectively explore psychological associations, unlike Durrant and Doherty (2010) and Shao (2018), who performed psycholinguistic experiments. Consequently, the psychological dimension of the results can be inferred but not confirmed. Further research could explore such dimension by means of proper psycholinguistic methods and techniques to add to our results.

A last limitation of this study refers to its applications. Although potential applications of the results for both language education and translation have been presented in Chapters 7 to 10, they lack empirical support. Researchers in the fields of Education and Translation Studies could perform quasi-experiments, case studies, mixed-method studies, or action research projects (see Shadish, Cook, & Campbell, 2002; Creswell, 2012; Yin, 2018) to investigate the practical value of the findings of the present research for language learners and translation practitioners as well. Such future studies could provide an important contribution to the understanding of the applications of lexical priming.

POSTSCRIPT: CHALLENGES FACED BY A PHD CANDIDATE DURING THE COVID-19 PANDEMIC[1]

In the beginning of 2018, a few months after being accepted as a doctoral student by the Postgraduate Program in Linguistic and Literary Studies in English at the University of São Paulo, I read Canagarajah's (2002) *A Geopolitics of Academic Writing*. At the time, I felt myself privileged to be able to carry out research studies in Brazil under fairly favorable conditions. Despite all the serious problems the country had been facing in the previous years, it still appeared to be far from Canagarajah's (2002) description of Sri Lanka. Is it really possible to develop research projects in the kind of harsh environment that Canagarajah (2002) described?

Time passed, and, at the beginning of 2020, COVID-19 spread across Brazil, reaching São Paulo, the state where I live. Suddenly, everything shifted to the virtual world. The undergraduate classes I had been teaching became Internet-based; meetings were also transferred to the digital environment; research activities were done almost exclusively online—libraries remained closed for more than 2 years. Above all, we started to live with an invisible enemy, with death persistently surrounding us.

Personally, I still felt privileged because scholars and students were able to do their job remotely, thus having less chance of being infected. However, the pandemic affected everyone, including higher education professionals, researchers, and students. The situation in São Paulo was different from the one in which Canagarajah (2002) found himself many years ago; nevertheless,

[1] This postscript adds a personal note to the conventionally impersonal genre of the research dissertation. The justification for doing so lies in the fact that the research was undertaken during the extraordinary events of a global pandemic.

the feeling of unease may have been similar. Under this kind of circumstance, one starts doubting whether his or her academic aspirations can be fulfilled.

Here, I would like to share three challenges I faced as a Philosophy Doctor (PhD) candidate during the pandemic. This brief account contextualizes the main text of the doctoral dissertation and might be helpful for others confronting similar challenges, especially postgraduate students, even in the post-pandemic world.

The first challenge relates to time management. In the first months of the COVID-19 pandemic, I did not feel that I was working at home; instead, I had the feeling that I was living at work. The change in work mode, in addition to the intermittent disappearance of people and vehicles from the streets, skewed my perception of time, sometimes giving me the impression that I was inside the movie *Groundhog Day* (Albert & Ramis, 1993), where Phil Connors (Bill Murray) lives every day the same day.

I followed health experts' advice, staying at home as much as possible. As a result, the places where I sleep, eat, study, learn, research, write, work, teach, talk, shop, and rest became one—that is, my home. Consequently, housekeeping activities increased drastically, and my normal pace of life was disrupted. It took about 4 months for me to adapt to the new routine. Eventually, I realized that the amount of time I spent going to and coming from work was almost the same amount I needed for cooking, washing dishes, tidying up the house, and so forth. Therefore, neither had I lost nor gained time. Time management was the key to getting things done under the new circumstances.

In a normal routine, our activities depend on the open spaces and buildings where we are. The classroom is the space for teaching (or learning); the library is the space for reading; the restaurant is the space for eating and chatting; the mall is the site for shopping; the meeting room is the site for discussing (sometimes listening); the coffee shop is the space for drinking; etc. In the new routine, this relationship became fuzzy because the home turned into the main stage for almost every single act. My strategy to overcome this challenge was to mentally replicate the original division marked by physical settings, carefully managing time spent in each activity.

Sleeping, cooking, eating, and working hours were already well-defined, because the quarantine did not directly affect these aspects of my life. Time periods devoted to other housekeeping activities, research, rest, and leisure, however, were overlapping. As every week I was resting too much, I decided to create *sacred islands of research* in my weekly schedule, that is, periods of time primarily devoted to research activities. My goal was to spend at least 20 hours per week focused on research (my average before the pandemic), but I could not reach this goal without overcoming other obstacles. It is one thing to rationally understand what must be done; another completely different thing is to actually do it.

The second challenge was about getting into an acceptable state of mind. The COVID-19 pandemic had been not only a health issue but also a battle between information and disinformation, where everybody seemed to lose. Keeping a peaceful mind was fairly difficult because, in addition to the consternation caused by the virus itself, there was the perplexity generated by other people's irrational behavior and widespread ignorance. In the initial stage of the pandemic, I could not calm myself down; basically, my mind had turned into an echo of the sharp criticism expressed by experts in the media about public health policies against COVID-19.

Gradually, however, I realized that to achieve the necessary inner peace, I should throw out expectations about external changes and focus on my own path and responsibilities. This could remind me of the reasons why I started the PhD project in the very beginning as well as of my role in society. I chose to concentrate my efforts on trying to produce an original, stimulating piece of research and become someone worthy of the title of Doctor, despite my flaws. Developing a vaccine, discovering the best method for COVID-19 prevention, or changing the people's behavior was completely beyond my reach. Therefore, worrying about these things was nothing but waste of mental energy. It was necessary to focus on the alterable, the reachable, the possible—not on their opposites.[2]

---

[2] Even though, I must admit that teachers of reading and writing around the globe—myself included—share some of the responsibility for the current state of affairs with respect to the spread of fake news, disinformation campaigns, and Internet scams. Societies comprised of critical readers and writers capable of seeing through, around, and beyond words certainly find

Finally, the third challenge refers to creativity. One needs to receive input from a wide variety of sources before being able to do or write something interesting. Sources relate to theories, research discoveries, lectures, meetings, and group discussions, as well as to less obvious matters, such as walks, shopping, travel, and everyday conversations. The COVID-19 pandemic severely limited human activities, thus reducing the number and variety of sources for creative output. As a consequence, the issue of lack of creativity or block arose. Between 2020 and 2021, for the first time in my life I experienced a creative block. Ideas did not flow, which made me increasingly anxious about the doctoral program. Although I believed that the issue was not a matter of time but one of inner demons, eventually I had to request the postponement of the final text's submission deadline because of the block. As a matter of fact, it took some months for me to find a way to overcome it.

Steve Jobs (2005), in his famous commencement speech at Stanford University, tells three stories of his life revealing different factors behind success. One of the stories is about connecting the dots. He shares some experiences he had that did not appear to be valuable in the beginning but later proved to be so. For example, he took calligraphy classes in his short period at college, and many years later this experience became a useful input for the design of computer typography. He says that one should trust that the dots will connect in the future, following his or her intuition.

My solution for the creative block was not the product of a rational process. Instead, it had to do with my intuition and the feeling that somehow the dots would connect later, as in Jobs' (2005) story. At the end of 2019, I became interested in photography and started to look for cameras and photography classes. In the beginning of 2020, I started to take pictures regularly, and, during the pandemic, I read books and studied—mainly through online learning—different kinds of photography, such as portrait, still life, and macro. Although I could not stop feeling guilty about putting time and efforts into something seemingly unrelated to the doctoral program, I thought that this might be useful in some way.

---

themselves in a better position to deal with the different forms of misleading and deceitful communication.

Between April and May of 2021, my block was getting worse because even when I had plenty of time I did not move forward in my research work. Then, when I was thinking about what I should do to recover my ability to research and write smoothly, it occurred to me that taking pictures could be inexplicably a good idea. After that moment, I started to take pictures using photography as a creative anchor to stabilize and revitalize my research and writing processes. Thanks to photography, I was finally able to continue my project and write the necessary texts. I overcame the block and regained my powers. In the end, I felt myself like Superman in the sky after being for a long time on the ground in the presence of a piece of Kryptonite. The dots connected.

Originally written between June 20 and July 4, 2021.
Revised on July 23, 2022.
Final revision made on April 17, 2023.

REFERENCES

Abbas, H. (2021). Editorial: What is artificial intelligence? *IEEE Transactions on Artificial Intelligence*, *2*(2), 94–95. doi: 10.1109/TAI.2021.3096243.

Albert, T. (Producer), & Ramis, H. (Producer and Director). (1993). *Groundhog Day* [Motion picture]. United States: Columbia Pictures.

American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: American Psychological Association.

American Psychological Association. (2021). *Paraphrasing*. Retrieved from https://apastyle.apa.org/style-grammar-guidelines/citations/paraphrasing.

Anderson, J. R. (1983/1996). *The architecture of cognition*. New York, NY: Psychology Press.

Anthony, L. (1999). Writing research article introductions in Software Engineering: How accurate is a standard model? *IEEE Transactions on Professional Communication*, *42*(1), 38–46.

Anthony, L. (2017). *SegmentAnt* (v. 1.1.3) [Computer software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net.

Anthony, L. (2022). *AntConc* (v. 4.0.3) [Computer software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net.

Aragão, R. M. L. de. (2010). Introductions of Marketing research articles written in Japanese. *Signum: Estudos da Linguagem*, *13*(2), 101–112. doi: 10.5433/2237-4876.2010v13n2p101.

Aragão, R. M. L. de. (2012). *Modelos para a estruturação de artigos científicos: Um estudo de instruções aos autores a introduções de artigos de revistas da Scientific Electronic Library Online do Brasil* (Dissertação de mestrado, Universidade de São Paulo). "Models for the organization of scientific articles: A study of instructions to authors, articles, and article introductions from journals in the Scientific Electronic Library Online Brazil (Master's thesis, University of São Paulo)." doi: 10.11606/D.8.2012.tde-22082012-101602.

Aragão, R. M. L. de. (2015). A view into *gencho ronbun*: To what extent do Japanese original articles follow the IMRD pattern? Paper presented at the

International Symposium on Japanese as a Global Language (unnumbered pages). São Paulo, SP.

Aragão, R. M. L. de. (2017). *Fundamental categories of documents from Japanese scholarly periodicals* (Master's thesis, University of São Paulo). doi: 10.11606/D.8.2017.tde-20102017-191222.

Aragão, R. M. L. de. (2022a). 96–8–3–2: Signs of lexical priming in academic article introductions. *Linha D'Água*, *35*(2), 118–139. doi: 10.11606/issn.2236-4242.v35i2p118-139.

Aragão, R. M. L. de. (2022b). Signs of psychological priming in academic articles from Pediatrics and Management. *Diálogo das Letras*, *11*, e02202. Retrieved from http://periodicos.apps.uern.br/index.php/DDL/article/view/3893.

Baker, H., McEnery, T., & Hardie, A. (2017). A corpus-based investigation into English representations of Turks and Ottomans in the early modern period. In M. Pace-Sigge & K. J. Patterson (Eds.), *Lexical priming: Applications and advances* (pp. 41–66). Amsterdam, Netherlands: John Benjamins.

Baker, P. (2006). *Using corpora in discourse analysis*. London, UK: Continuum.

Bawarshi, A. S., & Reiff, M. J. (2010). *Genre: An introduction to history, theory, research, and pedagogy*. West Lafayette, IN: Parlor Press LLC and The WAC Clearinghouse.

Bawcom, L. (2017). Lexical priming and the selection and sequencing of synonyms. In M. Pace-Sigge & K. J. Patterson (Eds.), *Lexical priming: Applications and advances* (pp. 121–139). Amsterdam, Netherlands: John Benjamins.

Bazerman, C. (1988). *Shaping written knowledge: The genre and activity of the experimental article in science*. Madison, WI: The University of Wisconsin Press.

Bennett, K. (2010). Academic discourse in Portugal: A whole different ballgame? *Journal of English for Academic Purposes*, *9*(1), 21–32.

Berber Sardinha, T. (2000). Corpus Linguistics: History and problematization. *DELTA*, *16*(2), 323–367. doi: 10.1590/S0102-44502000000200005.

Berber Sardinha, T. (2009). *Pesquisa em Linguística de Corpus com Wordsmith Tools*. "Research in Corpus Linguistics with Wordsmith Tools." Campinas, SP: Mercado de Letras.

Berber Sardinha, T. (2017). Lexical priming and register variation. In M. Pace-Sigge & K. J. Patterson (Eds.), *Lexical priming: Applications and advances* (pp. 189–229). Amsterdam, Netherlands: John Benjamins.

Berber Sardinha, T., Moreira Filho, J. L., & Alambert, E. (2008–2010). *Corpus Brasileiro*. "Brazilian Corpus." Available from https://www.sketchengine.eu/.

Bernardino, C. G., & Pacheco, J. T. S. (2017). A socio-rhetorical analysis of introductions in original articles of the Nutrition area's disciplinary culture. *Fórum Linguístico*, *14*(1), 1749–1766. doi: 10.5007/1984-8412.2017v14n1p1749.

Bertero, C. O. (2021). Sixty years of RAE: A journey of criticism, resistance and reinvention. *Revista de Administração de Empresas*, *61*(3), 1–8. doi: 10.1590/S0034-759020210303x.

Bhatia, V. K. (2004). *Worlds of written discourse: A genre-based view*. London, UK: Continuum.

Boucart, M., Biederman, I., Cuervo, C., Danion, J.-M., & Wagemans, J. (2002). Effect of benzodiazepines on structural and conceptual/lexical priming. *Psychopharmacology*, *165*, 43–50. doi: 10.1007/s00213-002-1243-7.

Brauer, K. C. N., & Portela, K. C. A. (2017). Verbal processes analysis of scientific article introductions of three areas: Economic, Environmental and Sanitary Engineering and Linguistics according to systemic functional linguistics. *Holos*, *1*, 104–115. doi: 10.15628/holos.2017.4904.

Brett, P. (1994). A genre analysis of the results section of Sociology articles. *English for Specific Purposes*, *13*(1), 47–59.

Canagarajah, A. S. (2002). *A geopolitics of academic writing*. Pittsburgh, PA: University of Pittsburgh Press.

Cantos, P., & Almela, M. (2017). Colligational effects of collocation: Lexically-conditioned dependencies between modification patterns of the noun *cause*. In M. Pace-Sigge & K. J. Patterson (Eds.), *Lexical priming: Applications and advances* (pp. 231–249). Amsterdam, Netherlands: John Benjamins.

Chahal, D. (2014). Research article introductions in Cultural Studies: A genre analysis exploration of rhetorical structure. *The Journal of Teaching English for Specific and Academic Purposes*, *2*(1), 1–20.

Chino, N. (2008). *All about particles: A handbook of Japanese function words*. Tokyo, Japan: Kodansha International.

Chuang, W.-c. (2015). Lexical primings of licensing agreements: A preliminary investigation. *International Journal of English Linguistics*, *5*(4), 1–11. doi: 10.5539/ijel.v5n4p1.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428.

Collins, D. A. W. (2019). *Investigating lexical priming using eye tracking data – some "glad news"* (Master's thesis, University of Birmingham). Retrieved from https://etheses.bham.ac.uk/id/eprint/9759/.

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*, 33–43.

Costa, R. L. S. da. (2015). *Culturas disciplinares e artigos acadêmicos experimentais: Um estudo comparativo da descrição sociorretórica* (Dissertação de mestrado, Universidade Estadual do Ceará). "Disciplinary cultures and experimental scholarly articles: A comparative study of socio-rhetorical descriptions (Master's dissertation, Ceará State University)." Retrieved from https://siduece.uece.br/siduece/trabalhoAcademicoPublico.jsf?id=84074.

Creative Commons. (n.d.). *About CC licenses*. Retrieved from https://creativecommons.org/about/cclicenses/.

Creswell, J. W. (2012). *Education research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed). Boston, MA: Pearson.

Crookes, G. (1986). Towards a validated analysis of scientific text structure. *Applied Linguistics*, *7*(1), 57–70.

Cunha, A. L. da. (2017). *Análise do* priming *lexical das composições sintagmáticas em textos publicitários: Por uma metodologia em que o léxico seja o ponto central do ensino de língua portuguesa* (Tese de doutorado, Universidade Federal de Minas Gerais). "Analysis of lexical priming of syntagmatic compositions in advertising texts: Toward a methodology centered on the lexicon for the teaching of Portuguese (PhD thesis, Federal University of Minas Gerais)." Retrieved from http://hdl.handle.net/1843/LETR-AN6KNY.

Davies, M. (2008–). *The Corpus of Contemporary American English (COCA)*. Available from https://www.english-corpora.org/coca/.

Davies, R. J. (2000). *The academic writing of Japanese students of English: Contrastive rhetoric and its implications for an integrated approach to composition pedagogy* (PhD thesis, University of Wales). Retrieved from http://e.bangor.ac.uk/id/eprint/4216.

Davis, H. P., Cohen, A., Gandy, M., Colombo, P., VanDusseldorp, G., Simolke, N., & Romano, J. (1990). Lexical priming deficits as a function of age. *Behavioral Neuroscience*, *104*(2), 288–297.

Davison, J. (Producer), Abrahams, J., Zucker, D., & Zucker, J. (Executive Producers and Directors). (1980). *Airplane!* [Motion picture]. United States: Paramount Pictures.

Devitt, A. J. (1993). Generalizing about genre: New conceptions of an old concept. *College Composition and Communication*, *44*(4), 573–586.

Dias, F. G. R., & Bezerra, B. G. (2013). Rhetorical analysis of introductions in Public-Health-related articles. *Horizontes de Linguística Aplicada*, *12*(1), 163–182. doi: 10.26512/rhla.v12i1.1238.

Duguid, A., & Partington, A. (2017). Forced lexical primings in transdiscoursive political messaging: How they are produced and how they are received. In M. Pace-Sigge & K. J. Patterson (Eds.), *Lexical priming: Applications and advances* (pp. 67–92). Amsterdam, Netherlands: John Benjamins.

Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, *19*(1), 61–74.

Durrant, P., & Doherty, A. (2010). Are high-frequency collocations psychologically real? Investigating the thesis of collocational priming. *Corpus Linguistics and Linguistic Theory 6*(2), 125–155. doi: 10.1515/CLLT.2010.006.

Evert, S. (2005). *The statistics of word cooccurrences: Word pairs and collocations* (PhD dissertation, Universität Stuttgart). Retrieved from https://elib.uni-stuttgart.de/bitstream/11682/2573/1/Evert2005phd.pdf.

Firth, J. R. (1957/1968). A synopsis of linguistic theory, 1930–55. In F. R. Palmer (Ed.), *Selected papers of J. R. Firth 1952–1959* (pp. 168–205). London, UK: Longmans.

Flowerdew, J. (2008). Scholarly writers who use English as an additional language: What can Goffman's "stigma" tell us? *Journal of English for Academic Purposes*, *7*, 77–86.

Fukao, Y., & Baba, M. (2000). Analysis of "-ni taishite" in agricultural and technological papers. *Senmon Nihongo Kyōiku Kenkyū*, *2*, 14–21. Retrieved from http://stje.kir.jp/download/02_14.pdf.

Gabrielatos, C. (2018). Keyness analysis: Nature, metrics and techniques. In C. Taylor & A. Marchi (Eds.), *Corpus approaches to discourse: A critical review* (pp. 225–258). Abingdon, UK: Routledge.

Gibbs, W. W. (1995). Lost science in the Third World. *Scientific American*, *273*(2), 92–99.

Gledhill, C. (2000). The discourse function of collocation in research article introductions. *English for Specific Purposes*, *19*, 115–135.

Goatly, A. (2017). Lexical priming in humorous discourse. *European Journal of Humour Research*, *5*(1), 52–68. doi: 10.7592/EJHR2017.5.1.goatly.

Grodzinsky, Y., Swinney, D., & Zurif, E. (1985). Agrammatism: Structural deficits and antecedent processing disruptions. In M.-L. Kean (Ed.), *Agrammatism* (pp. 65–81). Orlando, FL: Academic Press. doi: 10.1016/B978-0-12-402830-2.50007-9.

Gyogi, E., & Iwasaki, N. (2019). Genre-based teaching of mimetics in the beginner-level classroom: Translating TV commercials. *Journal of Japanese Language Teaching*, *174*, 71–85. doi: 10.20721/nihongokyoiku.174.0_71.

Hadikin, G. S. (2013). You know Korean English? Lexical priming in short strings of Korean spoken English. *International Journal of Language Studies*, *7*(1), 59–78.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London, UK: Longman.

Harwood, N. (2005). '*Nowhere has anyone attempted . . . In this article I aim to do just that*': A corpus-based study of self-promotional *I* and *we* in academic writing across four disciplines. *Journal of Pragmatics*, *37*, 1207–1231.

Hirano, E. (2009). Research article introductions in English for Specific Purposes: A comparison between Brazilian Portuguese and English. *English for Specific Purposes*, *28*, 240–250.

Hoey, M. (1983). *On the surface of discourse*. London, UK: George Allen & Unwin.

Hoey, M. (2004). The textual priming of lexis. In G. Aston, S. Bernardini, & D. Stewart (Eds.), *Corpora and language learners* (pp. 21–41). Amsterdam, Netherlands: John Benjamins.

Hoey, M. (2005). *Lexical priming: A new theory of words and language*. Abingdon, UK: Routledge.

Hoey, M. (2007a). Grammatical creativity: A corpus perspective. In M. Hoey, M. Mahlberg, M. Stubbs, & W. Teubert, *Text, discourse, and corpora: Theory and analysis* (pp. 31–56). London, UK: Continuum.

Hoey, M. (2007b). Lexical priming and literary creativity. In M. Hoey, M. Mahlberg, M. Stubbs, & W. Teubert, *Text, discourse, and corpora: Theory and analysis* (pp. 7–30). London, UK: Continuum.

Hoey, M. (2011). Lexical priming and translation. In A. Kruger, K. Wallmach, & J. Munday (Eds.), *Corpus-based translation studies: Research and applications* (pp. 153–168). London, UK: Bloomsburry.

Hoey, M. (2013). Lexical priming. In C. A. Chapelle (Ed.), *The encyclopedia of Applied Linguistics* (pp. 3342–3347). Oxford, UK: Wiley-Blackwell.

Hoey, M. (2017). Cohesion and coherence in a content-specific corpus. In M. Pace-Sigge & K. J. Patterson (Eds.), *Lexical priming: Applications and advances* (pp. 3–40). Amsterdam, Netherlands: John Benjamins.

Hoey, M., & O'Donnell, M. B. (2008a). Lexicography, grammar, and textual position. *International Journal of Lexicography*, *21*(3), 293–309.

Hoey, M., & O'Donnell, M. B. (2008b). The beginning of something important: Corpus evidence on the text beginnings of hard news stories. In B. Lewandowska-Tomaszcyk (Ed.), *Corpus Linguistics, computer tools, and* *applications – State of the art: PALC 2007, Vol. 17* (pp. 189–212). Frankfurt am Main, German: Peter Lang.

Hoey, M., & O'Donnell, M. B. (2015). Examining associations between lexis and textual position in hard news stories, or *according to a study by…* In N. Groom, M. Charles, & S. John (Eds.), *Corpora, grammar, and discourse: In honour of Susan Hunston* (pp. 117–144). Amsterdam, Netherlands: John Benjamins.

Holmes, R. (1997). Genre analysis, and the Social Sciences: An investigation of the structure of research article discussion sections in three disciplines. *English for Specific Purposes*, *16*(4), 321–337.

Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge, UK: Cambridge University Press.

Hyland, K. (1996). Writing without conviction? Hedging in science research articles. *Applied Linguistics*, *17*(4), 433–454.

Hyland, K. (2004). *Genre and second language writing*. Ann Arbor, MI: The University of Michigan Press.

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, *27*, 4–21.

Jakubíček, M., Kilgarriff, A., Kovář, V., Rychlý, P., & Suchomel, V. (2013). The TenTen Corpus Family. Paper presented at the *7th International Corpus Linguistics Conference CL2013* (pp. 125–127). Lancaster, UK. Retrieved from https://ucrel.lancs.ac.uk/cl2013/doc/CL2013-ABSTRACT-BOOK.pdf.

Jantunen, J. H. (2017). Lexical and morphological priming: A holistic phraseological analysis of the Finnish time expression *kello*. In M. Pace-Sigge & K. J. Patterson (Eds.), *Lexical priming: Applications and advances* (pp. 253–271). Amsterdam, Netherlands: John Benjamins.

Jeaco, S. (2015). *The Prime Machine: A user-friendly corpus tool for English language teaching and self-tutoring based on the lexical priming theory of language* (PhD thesis, University of Liverpool). Retrieved from https://livrepository.liverpool.ac.uk/2014579/.

Jeaco, S. (2017). Concordancing lexical primings: The rationale and design of a user-friendly corpus tool for English language teaching and self-tutoring based on the lexical priming theory of language. In M. Pace-Sigge & K. J. Patterson (Eds.), *Lexical priming: Applications and advances* (pp. 273–296). Amsterdam, Netherlands: John Benjamins.

Jobs, S. (2005). Commencement address on June 12, 2005. *Stanford News*, June 12, 2005. Retrieved from https://news.stanford.edu/2005/06/14/jobs-061505/.

Johns, A. M. (2003). Genre and ESL/EFL composition instruction. In B. Kroll (Ed.), *Exploring the dynamics of second language writing* (pp. 195–217). New York, NY: Cambridge University Press.

Jordan, M. P. (1980). Short texts to explain problem–solution structures – and vice versa: An analytical study of English prose to show the relationship between clear writing and clear thinking. *Instructional Science*, *9*, 221–252.

Jordan, M. P. (1984). *Rhetoric of everyday English texts*. London, UK: George Allen & Unwin.

Kaplan, R. B. (1966). Cultural thought patterns in inter-cultural education. *Language Learning*, *16*, 1–20.

Keenan, J. M, Golding, J. M., Potts, G. R., Jennings, T. M., & Aman, C. J. (1990). Methodological issues in evaluating the occurrence of inferences. *Psychology of Learning and Motivation*, *25*, 295–312. doi: 10.1016/S0079-7421(08)60261-3.

Kilgarriff, A. (2009). Simple maths for keywords. Paper presented at the *Corpus Linguistics Conference CL2009* (unnumbered pages). Liverpool, UK. Retrieved from https://www.sketchengine.eu/wp-content/uploads/2015/04/2009-Simple-maths-for-keywords.pdf.

Kimoto, K. (2006). Analyzing the textual structure of Law paper introductions on civil law, commercial law, and intellectual property law. *Senmon Nihongo Kyōiku Kenkyū*, *8*, 19–26. Retrieved from http://stje.kir.jp/download/08_19.pdf.

Kobayashi, K. (2003). *A genre analysis of English and Japanese research articles in the field of Applied Linguistics: A contrastive study* (PhD thesis, Universiti Malaya). Retrieved from http://studentsrepo.um.edu.my/304/.

Kolesnikova, O. (2016). Survey of word co-occurrence measures for collocation detection. *Computación y Sistemas*, *20*(3), 327–344. doi: 10.13053/CyS-20-3-2456.

Kuhn, T. Z. (2017). *A design proposal of an online corpus-driven dictionary of Portuguese for university students* (PhD thesis, University of Lisbon). Retrieved from http://hdl.handle.net/10451/32013.

Kuhn, T. Z., & Ferreira, J. P. (2018). *CoPEP – The Corpus of Portuguese from Academic Journals* (v. 1.4). Available from https://www.sketchengine.eu/.

Lexical Computing CZ. (n.d.). *Sketch Engine*. Brno, Czech Republic: Lexical Computing CZ. Available from https://www.sketchengine.eu/.

Lexical Computing Ltd. (2015). *Statistics used in Sketch Engine*. Retrieved from https://www.sketchengine.eu/wp-content/uploads/ske-statistics.pdf.

Li, D., & Yang, X. (2017). A corpus-based study of *more and more* and its environments: Insights from lexical priming theory. *Journal of Beijing International Studies University*, *39*(1), 116–126. doi: 10.12002/j.bisu.084.

Lim, J. M. H. (2006). Methods sections of Management research articles: A pedagogically motivated qualitative study. *English for Specific Purposes*, *25*(3), 282–309.

McEnery, T., & Hardie, A. (2012). *Corpus Linguistics: Method, theory and practice*. Cambridge, UK: Cambridge University Press.

Menezes, S. D., & Caregnato, S. E. (2018). Brazilian scientific production in Chemistry between 2004 and 2013: Analysis of articles indexed in the Web of Science. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da informação*, *23*(53), 25–38. doi: 10.5007/1518-2924.2018v23n53p25.

Merriam-Webster's Collegiate Dictionary (11th ed.). (2012). Springfield, MA: Merriam-Webster.

Miller, C. R. (1984). Genre as social action. *Quarterly Journal of Speech*, *70*, 151–167.

Miranda, M. V. (2021). *Transitividade no vocabulário acadêmico em português brasileiro: Uma análise baseada em artigos científicos* (Tese de doutorado, Universidade Federal de Minas Gerais). "Transitivity in Brazilian Portuguese academic vocabulary: An analysis based on scientific articles (PhD thesis, Federal University of Minas Gerais)." Retrieved from http://hdl.handle.net/1843/35551.

Moritz, M. E. W., Meurer, J. L., & Dellagnelo, A. K. (2008). Conclusions as components of research articles across Portuguese as a native language, English as a native language and English as a foreign language: A contrastive genre study. *The ESPecialist*, *29*(2), 233–253.

Muraoka, T. (1999). Commonly used Japanese sentence ending expressions and sentence patterns, as stated in 'Materials and Methods', in Agricultural Science papers. *Senmon Nihongo Kyōiku Kenkyū*, *1*, 16–23. Retrieved from http://stje.kir.jp/download/01_16.pdf.

Muraoka, T. (2002). Connectives and paragraph organization of 'Results and Discussion', in Japanese Agricultural Science papers. *Senmon Nihongo Kyōiku Kenkyū*, *4*, 27–34. Retrieved from http://stje.kir.jp/download/04_27.pdf.

Muraoka, T., Yoneda, Y., Chinami, K., Nishina, K., Fukao, Y., & Otani, S. (2005). An analysis of the structure and expressions of the introductions of Agricultural and Engineering Science papers written in Japanese. *Senmon Nihongo Kyōiku Kenkyū*, *7*, 21–28. Retrieved from http://stje.kir.jp/download/07_21.pdf.

Muraoka, T., Yoneda, Y., Otani, S., Goto, K., Fukao, Y., & Chinami, K. (2004). Conjunctions and logical development: A study of introductory sections to

Agricultural and Engineering Science papers written in Japanese. *Senmon Nihongo Kyōiku Kenkyū*, *6*, 41–48. Retrieved from http://stje.kir.jp/download/06_41.pdf.

Nabatame, T., & Oshima, Y. (2018). Sentence structure and expressions in interpretative descriptions of historical materials found in academic papers of History, International Relations, and Area Studies. *Senmon Nihongo Kyōiku Kenkyū*, *20*, 19–26. doi: 10.11448/jtje.20.19.

Nakayama, S. (2020). Do EFL textbooks in Japan cause lexical priming? In P. Clements, A. Krause, & R. Gentry (Eds.), *Teacher efficacy, learner agency* (pp. 273–281). Tokyo, Japan: JALT. doi: 10.37546/JALTPCP2019-32.

Neely, J. H. (1976). Semantic priming and retrieval from lexical memory: Evidence for facilitatory and inhibitory processes. *Memory & Cognition*, *4*(5), 648–654.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, *106*(3), 226–254.

Nitsu, N. (2009). 論文の引用に関する基礎的調査と引用モデルの試案. *Ronbun no in'yō ni kansuru kisoteki chōsa to in'yō moderu no shian*. "Basic research and a draft citation model for citations in academic papers." *Akademikku Japanīzu Jānaru*, *1*, 65–74. Retrieved from http://academicjapanese.jp/dl/ajj/65-74-nitsu.pdf.

Nwogu, K. N. (1997). The medical research paper: Structure and functions. *English for Specific Purposes*, *16*(2), 119–138.

O'Donnell, M. B., Scott, M., Mahlberg, M., & Hoey, M. (2012). Exploring text-initial words, clusters and concgrams in a newspaper corpus. *Corpus Linguistics and Linguistic Theory*, *8*(1), 73–101. doi: 10.1515/cllt-2012-0004.

Oliveira Jr., O. N., Schuster, E., & Levkowitz, H. (2014). Preface. In E. Schuster, H. Levkowitz, & O. N. Oliveira Jr. (Eds.), *Writing scientific papers in English successfully: Your complete roadmap* (pp. 5–7). Andover, MA: Hyprtek.

Ooi, V. (2013). Lexical priming and Asian learners of English. *Learner Corpus Studies in Asia and the World*, *1*, 31–41. Retrieved from http://www.lib.kobe-u.ac.jp/infolib/meta_pub/G0000003kernel_81006673.

Oshima, Y. (2009). Analysis of the discussions of case studies and historical document-based studies in Social Science. *Senmon Nihongo Kyōiku Kenkyū*, *11*, 15–22. doi: 10.11448/jtje.11.15.

Oshima, Y. (2016). Sentence structure and expressions in interpretations of statistical materials and case studies found in "complex" type academic papers in Agricultural/Fishery Economics. *Senmon Nihongo Kyōiku Kenkyū*, *18*, 29–36. doi: 10.11448/jtje.18.29.

Oshima, Y., Sato, S., Chinami, K., Yamamoto, F., & Nitsu, N. (2010). An analysis of discourse structure of introductions of academic papers in various disciplines. *Senmon Nihongo Kyōiku Kenkyū*, *12*, 27–34. Retrieved from https://www.jstage.jst.go.jp/article/jtje/12/0/12_0_27/_pdf.

Pace-Sigge, M. (2007). Traces of lexical priming in English and German uses of *be* and *have* forms in biographies. Paper presented at the *Corpus Linguistic Conference CL2007* (Article #52). Birmingham, UK. Retrieved from https://www.birmingham.ac.uk/documents/college-artslaw/corpus/conference-archives/2007/52Paper.pdf.

Pace-Sigge, M. (2017). Can lexical priming be detected in conversation turn-taking strategies? In M. Pace-Sigge & K. J. Patterson (Eds.), *Lexical priming: Applications and advances* (pp. 93–117). Amsterdam, Netherlands: John Benjamins.

Pace-Sigge, M. T. L. (2010). *Evidence of lexical priming in spoken Liverpool English* (PhD thesis, University of Liverpool). Retrieved from https://livrepository.liverpool.ac.uk/7873/.

Pace-Sigge, M. T. L. (2015). Applying the concepts of lexical priming to German polysemantic words. Paper presented at *Corpus Linguistics 2015* (pp. 264–266). Lancaster, UK. Retrieved from http://ucrel.lancs.ac.uk/cl2015/doc/CL2015-AbstractBook.pdf.

Partington, A. S. (2009). A linguistic account of wordplay: The lexical grammar of punning. *Journal of Pragmatics*, *41*, 1794–1809. doi: 10.1016/j.pragma.2008.09.025.

Patterson, K. (2015). *Lexical priming and metaphor: Application of the theory of lexical priming to metaphoric language* (PhD thesis, University of Liverpool). Retrieved from https://livrepository.liverpool.ac.uk/id/eprint/3001530.

Patterson, K. J. (2016). The analysis of metaphor: To what extent can the theory of lexical priming help our understanding of metaphor usage and comprehension? *Journal of Psycholinguistic Research*, *45*, 237–258.

Patterson, K. J. (2017). Lexical priming and metaphor – Evidence of nesting in metaphoric language. In M. Pace-Sigge & K. J. Patterson (Eds.), *Lexical priming: Applications and advances* (pp. 141–161). Amsterdam, Netherlands: John Benjamins.

Pojanapunya, P., & Todd, R. W. (2018). Log-likelihood and odds ratio: Keyness statistics for different purposes of keyword analysis. *Corpus Linguistics and Linguistic Theory*, *14*(1), 133–167. doi: 10.1515/cllt-2015-0030.

Posteguillo, S. (1999). The schematic structure of Computer Science research articles. *English for Specific Purposes*, *18*(2), 139–160.

Princeton University. (2023). *WordNet* [Lexical database of English]. Available from https://wordnet.princeton.edu/.

Ratcliff, J. E. (1987). The plausibility effect: Lexical priming or sentential processing? *Memory & Cognition*, *15*(6), 482–496.

Rezende, P. A. de, & Hemais, B. (2004). Contrastive analysis of scientific articles in the area of Health. *The ESPecialist*, *25*(2), 131–152.

Rodrigues, L. O. C. (2011). Publishing more, or better? The Olympic anteater. *Psicologia USP*, *22*(2), 457–472. doi: 10.1590/S0103-65642011005000014.

Romero, T. R. de S., & Joseph, N. L. de L. (2014). Typological patterns across the interpersonal metafunction modality in agricultural scientific articles: Differences and similarities both in Portuguese and English. *DELTA*, *30*(1), 115–136. doi: 10.1590/S0102-44502014000100007.

Ruiying, Y., & Allison, D. (2003). Research articles in Applied Linguistics: Moving from results to conclusions. *English for Specific Purposes*, *22*(4), 365–385.

Rychlý, P. (2008). A lexicographer-friendly association score. In P. Sojka & A. Horák (Eds.), *Proceedings of Recent Advances in Slavonic Natural Language Processing* (pp. 6–9). Brno, Czech Republic: Masaryk University.

Salim, H. Y. A. (2012). *Collocation and other lexical relationships in translations of the Quran: A corpus based application of lexical priming theory to a unique theological text* (Doctoral thesis, University of Liverpool). Retrieved from https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.632426.

Samraj, B. (2002). Introductions in research articles: Variations across disciplines. *English for Specific Purposes*, *21*(1), 1–17.

Sato, S., Oshima, Y., Nitsu, N., Yamamoto, F., Chinami, K., & Yamaji, N. (2013). Types of discourse structures in academic articles and their occurrence in different disciplines: An analysis of 270 articles in Humanities, Social Science and Engineering. *Nihongo Kyōiku*, *154*, 85–99. doi: 10.20721/nihongokyoiku.154.0_85.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin Company.

Shao, J. (2017). Teaching near-synonyms more effectively: A case study of “happy” words in Mandarin Chinese. In M. Pace-Sigge & K. J. Patterson (Eds.), *Lexical priming: Applications and advances* (pp. 163–185). Amsterdam, Netherlands: John Benjamins.

Shao, J. (2018). *Synonymy and lexical priming—A cross-linguistic investigation of synonymy from corpus and psycholinguistic perspectives*. (PhD thesis, University of Liverpool). Retrieved from https://livrepository.liverpool.ac.uk/3021737/.

Shezad, W. (2011). Outlining purposes, stating the nature of the present research, and listing research questions or hypotheses in academic papers. *Journal of Technical Writing and Communication*, *41*(2), 139–160.

Shibata, M. (2013). *A contrastive systemic functional analysis of causality in Japanese and English academic articles* (Master of Arts thesis, Marshall University). Retrieved from https://mds.marshall.edu/etd/473.

Shimizu, M. (2010). 「〜ハ〜テイル」引用文の論文中での使用目的. *"~Wa~teiru" in'yōbun no ronbunchū de no shiyō mokuteki*. "Purposes of use of citation sentences with the structure *-wa -teiru* in academic papers." *Akademikku Japanīzu Jānaru*, *2*, 90–100. Retrieved from http://academicjapanese.jp/dl/ajj/90-100.pdf.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford, UK: Oxford University Press.

Sinclair, J. (2004). *Trust the text: Language, corpus and discourse*. London, UK: Routledge.

Skalicky, S. (2018). Lexical priming in humorous satirical newspaper headlines. *Humor*, *31*(4), 583–602. doi: 10.1515/humor-2017-0061.

Strunk Jr., W., & White, E. B. (2000). *The elements of style* (4th ed.). Boston, MA: Pearson.

Stubbs, M. (1996). *Text and corpus analysis: Computer assisted studies of language and culture*. Oxford, UK: Blackwell.

Stubbs, M., & Halbe, D. (2013). Corpus Linguistics: Overview. In C. A. Chapelle (Ed.), *The encyclopedia of Applied Linguistics* (pp. 1377–1379). Oxford, UK: Wiley-Blackwell.

Swales, J. M. (1981/2011). *Aspects of article introductions: A reissue of Aston ESP research reports No.1 with a new introduction*. Ann Arbor, MI: University of Michigan Press.

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

Swales, J. M. (2004). *Research genres: Explorations and applications*. New York, NY: Cambridge University Press.

Swales, J. M., & Feak, C. B. (2004). *Academic writing for graduate students: Essential tasks and skills* (2nd ed.). Ann Arbor, MI: The University of Michigan Press.

Tarone, E., Dwyer, S., Gillette, S., & Icke, V. (1981). On the use of the passive in two Astrophysics journal papers. *The ESP Journal*, *1*(2), 123–140.

The American Heritage Dictionary (5th ed.). (2012). New York, NY: Dell.

The British National Corpus (BNC) (v. 3, XML Edition). (2007). Distributed by Oxford University Computing Services on behalf of the BNC Consortium. Available from http://www.natcorp.ox.ac.uk/.

The University of Notre Dame (2016). *Latin dictionary and grammar aid*. Retrieved from https://archives.nd.edu/latgramm.htm.

Tognini-Bonelli, E. (2001). *Corpus Linguistics at work*. Amsterdam, Netherlands: John Benjamins.

Tsiamita, F. (2011). *Disambiguating meaning: An examination of polysemous words within the framework of lexical priming* (PhD thesis, University of Liverpool). Retrieved from https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.569456.

Unedaya, K. (2003). A comparative study of the use of passive sentences in Applied Magnetic experimental papers in Japanese and English: Focusing on sentences to present figures and tables. *Senmon Nihongo Kyōiku Kenkyū*, *5*, 29–36. Retrieved from http://stje.kir.jp/download/05_29.pdf.

University of São Paulo. Universidade de São Paulo. Sistema Integrado de Bibliotecas da USP. (2016). *Diretrizes para apresentação de dissertações e teses da USP: Parte II (APA)* (3a ed.). "USP guidelines for the presentation of theses and dissertations: Part II (APA) (3rd ed.)." São Paulo, SP: SIBiUSP. doi: 10.11606/9788573140576.

Wang, L. (2018). *A comparative investigation of collocation, colligation and semantic association in English and Chinese: A test of Chinese on the applicability of three major concepts as defined in lexical priming theory* (PhD thesis, University of Liverpool). doi: 10.17638/03032915.

Wang, L. (2022). Contrasting the primings between English and Chinese: To advance Hoey's theory of lexical priming from the perspective of culture psychology. *Frontiers in Psychology*, *13*, 893120. doi: 10.3389/fpsyg.2022.893120.

Winter, E. O. (1974). *Replacement as function of repetition: A study of some of its principal features in the clause relations of contemporary English* (PhD dissertation, The University of London). Retrieved from https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.260792.

Winter, E. O. (1977). A clause-relational approach to English texts: A study of some predictive lexical items in written discourse. *Instructional Science*, *6*, 1–92.

Xu, Q. (2015). Lexical priming effects of textbooks on EFL learners' use of *give*. *English Language Teaching*, *8*(10), 123–132. Retrieved from https://files.eric.ed.gov/fulltext/EJ1078776.pdf.

Yin, R. K. (2018). *Case study research and applications: Design and methods*. London, UK: Sage.

APPENDIX A – ENGLISH TRANSLATIONS OF CHAPTER 7 TABLES

This appendix presents English translations of tables from Chapter 7 that have Portuguese and Japanese words. Table A.1 below shows the correspondence between the original and the translated tables. An exclusive table for Romanized transcriptions of Japanese is provided in one case due to space limitations in the original layout.

Table A.1 – Correspondence between original and translated tables.

Original Table	Translated Table
7.3	A.2
7.4	A.3
7.5	A.4 (Transcription) and A.5 (Translation)
7.6	A.6
7.13	A.7
7.14	A.8
7.15	A.9
7.16	A.10
7.17	A.11
7.18	A.12
7.19	A.13
7.20	A.14
7.21	A.15
7.25	A.16
7.26	A.17
7.27	A.18
7.28	A.19
7.29	A.20
7.30	A.21

Source: Prepared by the author.

Table A.2 – Translation of Table 7.3 (Key keywords of Portuguese RAs [ptRAs versus ptTenTen11]).

		General Freq.		Freq. per Million		RA Freq.		
POS	Word	FC	RC	FC	RC	n	(%)	Score
1	*construct*	92	1,167	356.43	0.25	6	12.5	285.4
2	*nurslings*	72	3,308	278.95	0.72	7	14.6	163.2
3	*95% CI*	50	1,056	193.71	0.23	10	20.8	158.5
4	*caregivers*	62	7,631	240.20	1.65	6	12.5	91.0
5	*familiarity*	61	8,522	236.33	1.84	5	10.4	83.5
6	*significance*	52	6,786	201.46	1.47	23	47.9	82.0
7	*et*	497	105,313	1925.50	22.78	38	79.2	81.0
8	*breastfeeding*	85	15,162	329.31	3.28	7	14.6	77.2
9	*Cronbach*	19	420	73.61	0.09	8	16.7	68.4
10	*regression*	72	15,105	278.95	3.27	15	31.3	65.6
11	*pediatricians*	39	6,413	151.10	1.39	6	12.5	63.7
12	*variables*	295	81,461	1142.90	17.62	36	75.0	61.4
13	*factorial*	19	1,444	73.61	0.31	6	12.5	56.9
14	*organizational*	213	62,835	825.21	13.59	10	20.8	56.6
15	α	20	1,812	77.48	0.39	8	16.7	56.4
16	*chi-squared*	19	1,769	73.61	0.38	13	27.1	54.0
17	χ^2	14	187	54.24	0.04	5	10.4	53.1
18	*H1*	18	1,727	69.74	0.37	8	16.7	51.5
19	*gestational*	30	6,230	116.23	1.35	6	12.5	49.9
20	*H3*	15	862	58.11	0.19	7	14.6	49.8
21	*respondents*	23	3,774	89.11	0.82	7	14.6	49.6
22	*discriminant*	14	648	54.24	0.14	5	10.4	48.4
23	*corporate*	132	44,653	511.40	9.66	5	10.4	48.1
24	*dummy*	13	368	50.37	0.08	7	14.6	47.6
25	*al*	456	167,559	1766.65	36.25	27	56.3	47.5
26	*measurement*	40	10,716	154.97	2.32	6	12.5	47.0

Source: Keyword analysis performed with Sketch Engine. FC = focus corpus. RC = reference corpus. RA Freq. = number (n) and percentage (%) of RAs containing the word. 95% CI = confidence interval of 95%.

Table A.3 – Translation of Table 7.4 (Key keywords of Portuguese journal texts [CoPEP versus ptTenTen11]).

		General Freq.		Freq. per Million		DOC Freq.		
POS	Word	FC	RC	FC	RC	n	(%)	Score
1	*no.*	4,142	0	84.83	0.00	1,321	13.34	85.8
2	*Table_1*	3,055	0	62.57	0.00	2,027	20.47	63.6
3	*aspects*	3,009	1,324	61.63	0.29	1,133	11.44	48.7
4	*Table_2*	2,293	0	46.96	0.00	1,526	15.41	48.0
5	*Figure_1*	2,200	0	45.06	0.00	1,613	16.29	46.1
6	*Table_3*	1,610	0	32.97	0.00	1,117	11.28	34.0
7	*et*	38,504	105,313	788.59	22.78	4,241	42.84	33.2
8	*Figure_2*	1,290	0	26.42	0.00	1,006	10.16	27.4
9	*statistically*	4,167	13,541	85.34	2.93	1,378	13.92	22.0
10	*nursing*	29,804	129,759	610.41	28.07	1,554	15.70	21.0
11	*al*	35,054	167,559	717.93	36.25	3,707	37.44	19.3
12	*significance*	2,168	6,786	44.40	1.47	1,098	11.09	18.4
13	*variables*	16,495	81,461	337.83	17.62	3,164	31.96	18.2
14	*male nurses*	8,321	43,185	170.42	9.34	1,128	11.39	16.6
15	*(it) was verified*	3,501	19,895	71.70	4.30	1,865	18.84	13.7
16	*sick persons*	17,262	116,394	353.54	25.18	1,891	19.10	13.5
17	*sample*	11,428	77,019	234.05	16.66	2,678	27.05	13.3
18	*correlation*	4,801	30,740	98.33	6.65	1,539	15.55	13.0
19	*prevalence*	5,318	35,768	108.92	7.74	1,561	15.77	12.6
20	*empirical*	1,961	10,760	40.16	2.33	1,115	11.26	12.4
21	*contexts*	5,651	39,209	115.74	8.48	2,228	22.51	12.3
22	*(it) was observed*	2,207	13,271	45.20	2.87	1,272	12.85	11.9
23	*(it) was used*	1,408	7,206	28.84	1.56	1,040	10.51	11.7
24	*therapeutic*	5,423	42,303	111.07	9.15	1,479	14.94	11.0
25	*variable*	8,345	68,207	170.91	14.75	2,323	23.46	10.9
26	*questionnaire*	4,957	39,851	101.52	8.62	1,349	13.63	10.7

Source: Keyword analysis performed with Sketch Engine. CoPEP was compiled by Kuhn and Ferreira (2018). FC = focus corpus. RC = reference corpus. DOC Freq. = number (n) and percentage (%) of texts containing the word.

Table A.4 – Transcription of Table 7.5 (Key keywords of Japanese RAs [jaRAs versus jaTenTen11 LUW]).

		General Freq.		Freq. per Million		RA Freq.		
POS	Word	FC	RC	FC	RC	n	(%)	Score
1	*kaigai shikaisha*	140	20	594.53	0.10	5	10.42	542.3
2	*honkō*	315	389	1,337.68	1.91	21	43.75	460.0
3	*honshō*	116	25	492.61	0.12	9	18.75	439.6
4	*senkō kenkyū*	144	96	611.51	0.47	19	39.58	416.3
5	*kizon kenkyū*	78	2	331.24	0.01	13	27.08	329.0
6	*shikō shi-*	113	183	479.87	0.90	19	39.58	253.3
7	*hozonteki chiryō*	64	20	271.78	0.10	7	14.58	248.4
8	*shokudō heisashō*	55	1	233.56	0.00	5	10.42	233.4
9	*jikenrei*	54	10	229.32	0.05	15	31.25	219.5
10	*shōrei*	305	1,164	1,295.22	5.72	23	47.92	193.0
11	*honkenkyū*	124	404	526.58	1.98	12	25.00	176.8
12	*zenrei*	46	40	195.34	0.20	16	33.33	164.1
13	*shōni gekai*	38	5	161.37	0.02	6	12.50	158.5
–	*i*	96	348	407.67	1.71	24	50.00	150.9
–	*aro*	66	178	280.28	0.87	20	41.67	150.1
14	*nichirei*	34	3	144.38	0.01	7	14.58	143.3
15	*kanji*	42	66	178.36	0.32	15	31.25	135.5
16	*tōka*	49	115	208.08	0.56	12	25.00	133.6
17	*kōhōshiteki*	30	2	127.40	0.01	15	31.25	127.2
18	*honronbun*	40	71	169.86	0.35	7	14.58	126.7
19	*era-*	26	4	110.41	0.02	17	35.42	109.3
20	*irō zōsetsu*	25	1	106.17	0.00	6	12.50	106.6
21	*bunseki kekka*	47	183	199.59	0.90	11	22.92	105.7

Source: Keyword analysis performed with Sketch Engine. FC = focus corpus. RC = reference corpus. RA Freq. = number (n) and percentage (%) of RAs containing the word. Verbs with partial inflection are double underlined.

Table A.5 – Translation of Table 7.5 (Key keywords of Japanese RAs [jaRAs versus jaTenTen11 LUW]).

		General Freq.		Freq. per Million		RA Freq.		
POS	Word	FC	RC	FC	RC	n	(%)	Score
1	*overseas subsidiary*	140	20	594.53	0.10	5	10.42	542.3
2	*this paper*	315	389	1,337.68	1.91	21	43.75	460.0
3	*this disease*	116	25	492.61	0.12	9	18.75	439.6
4	*previous research*	144	96	611.51	0.47	19	39.58	416.3
5	*existing research*	78	2	331.24	0.01	13	27.08	329.0
6	*carry out*	113	183	479.87	0.90	19	39.58	253.3
7	*conservative treatment*	64	20	271.78	0.10	7	14.58	248.4
8	*esophageal obstruction*	55	1	233.56	0.00	5	10.42	233.4
9	*medical case study*	54	10	229.32	0.05	15	31.25	219.5
10	*clinical case*	305	1,164	1,295.22	5.72	23	47.92	193.0
11	*the present study*	124	404	526.58	1.98	12	25.00	176.8
12	*all instances*	46	40	195.34	0.20	16	33.33	164.1
13	*pediatric surgeon*	38	5	161.37	0.02	6	12.50	158.5
–	*objection*	96	348	407.67	1.71	24	50.00	150.9
–	*part of sentence-ending expressions*	66	178	280.28	0.87	20	41.67	150.1
14	*age in days*	34	3	144.38	0.01	7	14.58	143.3
15	*child patient*	42	66	178.36	0.32	15	31.25	135.5
16	*this department*	49	115	208.08	0.56	12	25.00	133.6
17	*retrospective*	30	2	127.40	0.01	15	31.25	127.2
18	*this paper*	40	71	169.86	0.35	7	14.58	126.7
19	*obtain*	26	4	110.41	0.02	17	35.42	109.3
20	*gastrostomy*	25	1	106.17	0.00	6	12.50	106.6
21	*analysis results*	47	183	199.59	0.90	11	22.92	105.7

Source: Keyword analysis performed with Sketch Engine. FC = focus corpus. RC = reference corpus. RA Freq. = number (n) and percentage (%) of RAs containing the word. Verbs with partial inflection are double underlined.

Table A.6 – Translation of Table 7.6 (Number and proportion of RAs containing key keywords per discipline and selected keywords).

English Keyword	PED	MGT	Portuguese Keyword	PED	MGT	Japanese Keyword	PED	MGT
	n(%)	n(%)		n(%)	n(%)		n(%)	n(%)
CI	36(92.3)	3(7.7)	*construct*	1(16.7)	5(83.3)	*kaigai shikaisha* ("overseas subsidiary")	0(0.0)	5(100.0)
infants	30(100.0)	0(0.0)	*nurslings*	7(100.0)	0(0.0)	*honkō* ("this paper")	0(0.0)	21(100.0)
hypothesis	10(22.7)	34(77.3)	*95% CI*	10(100.0)	0(0.0)	*honshō* ("this disease")	9(100.0)	0(0.0)
adolescents	26(100,0)	0(0.0)	*caregivers*	6(100.0)	0(0.0)	*senkō kenkyū* ("previous research")	0(0.0)	19(100.0)
organisational	0(0.0)	22(100.0)	*familiarity*	0(0.0)	5(100.0)	*kizon kenkyū* ("existing research")	0(0.0)	13(100.0)
bivariate	13(76.5)	4(23.5)	*significance*	12(52.2)	11(47.8)	*shikō shi-* ("carry out")	19(100.0)	0(0.0)
variables	46(45.5)	55(54.5)	*et*	15(39.5)	23(60.5)	*hozonteki chiryō* ("conservative treatment")	7(100.0)	0(0.0)
organizational	0(0.0)	41(100.0)	*breastfeeding*	7(100.0)	0(0.0)	*shokudō heisashō* ("esophageal obstruction")	5(100.0)	0(0.0)
hypotheses	1(2.8)	35(97.2)	*Cronbach*	1(12.5)	7(87.5)	*jikenrei* ("medical case study")	15(100.0)	0(0.0)
neonatal	18(100.0)	0(0.0)	*regression*	8(53.3)	7(46.7)	*shōrei* ("clinical case")	23(100.0)	0(0.0)
respondents	14(25.9)	40(74.1)	*pediatricians*	6(100.0)	0(0.0)	*honkenkyū* ("the present study")	0(0.0)	12(100.0)
variance	18(30.5)	41(69.5)	*variables*	20(55.6)	16(44.4)	*zenrei* ("all instances")	16(100.0)	0(0.0)
χ^2	13(76.5)	4(23.5)	*factorial*	1(16.7)	5(83.3)	*shōni gekai* ("pediatric surgeon")	6(100.0)	0(0.0)
–	–	–	*organizational*	0(0.0)	10(100.0)	*nichirei* ("age in days")	7(100.0)	0(0.0)
–	–	–	*α*	0(0.0)	8(100.0)	*kanji* ("child patient")	15(100.0)	0(0.0)
–	–	–	*chi-squared*	10(76.9)	3(23.1)	*tōka* ("this department")	12(100.0)	0(0.0)
–	–	–	χ^2	1(20.0)	4(80.0)	*kōhōshiteki* ("retrospective")	15(100.0)	0(0.0)
–	–	–	*H1*	0(0.0)	8(100.0)	*honronbun* ("this paper")	5(71.4)	2(28.6)
–	–	–	*gestational*	6(100.0)	0(0.0)	*era-* ("obtain")	6(35.3)	11(64.7)
–	–	–	*H3*	0(0.0)	7(100.0)	*irō zōsetsu* ("gastrostomy")	6(100.0)	0(0.0)
–	–	–	*respondents*	0(0.0)	7(100.0)	*bunseki kekka* ("analysis results")	0(0.0)	11(100.0)
–	–	–	*discriminant*	0(0.0)	5(100.0)	–	–	–
–	–	–	*corporate*	0(0.0)	5(100.0)	–	–	–
–	–	–	*dummy*	1(14.3)	6(85.7)	–	–	–
–	–	–	*al*	4(14.8)	23(85.2)	–	–	–
–	–	–	*measurement*	1(16.7)	5(83.3)	–	–	–

Source: Compiled by the author. PED = Pediatrics. MGT = Management. Selected keywords are bounded by colored rectangles. Verbs with partial inflection are double underlined. CI = confidence interval. 95% CI = confidence interval of 95%.

Table A.7 – Translation of Table 7.13 (Left collocates of *significância* [“significance”] in two corpora).

	ptRAs			ptTenTen11		
POS	Word	PER	LogDice	Word	PER	LogDice
1	*presented (pl.)*	9.62	10.39	*presented (pl.)*	0.87	4.05
2	*of*	48.08	6.00	*demonstrated (s.)*	0.04	3.69
3	*the*	19.23	5.39	*achieved (s.)*	0.35	3.68
4	–	–	–	*extreme*	0.56	3.62
5	–	–	–	*showed (pl.)*	0.41	3.57
6	–	–	–	*evidenced*	0.07	3.46
7	–	–	–	*little*	0.63	3.17
8	–	–	–	*such*	0.21	3.17
9	–	–	–	*highlighted*	0.03	3.16
10	–	–	–	*(it) was detected*	0.03	3.11
11	–	–	–	*achieved (pl.)*	0.09	3.04
12	–	–	–	*inferior*	0.03	3.03
13	–	–	–	*showed (s.)*	0.81	3.02
14	–	–	–	*(it) was found*	0.04	2.96
15	–	–	–	*found*	0.31	2.93
16	–	–	–	*presented (s.)*	1.03	2.92
17	–	–	–	*reached*	0.13	2.90
18	–	–	–	*there was*	1.81	2.82
19	–	–	–	*observed*	0.15	2.76
20	–	–	–	*demonstrated (pl.)*	0.12	2.74

Source: Collocates identified with Sketch Engine. PER = number of co-occurrences between collocate and keyword divided by the overall frequency of the keyword in the corpus. Shared collocates are bounded by colored rectangles.

Table A.8 – Translation of Table 7.14 (Right collocates of *significância* ["significance"] in two corpora).

	ptRAs			ptTenTen11		
POS	Word	PER	LogDice	Word	PER	LogDice
1	*statistical/ statistics*	17.31	11.66	*statistical/ statistics*	22.27	9.03
2	*α=0.05*	3.85	10.22	*α*	0.41	6.74
3	*adopted*	3.85	9.62	*prognostic*	0.24	6.20
4	*of the (pl.)*	3.85	5.35	*adopted*	2.24	6.18
5	*of*	26.92	5.16	*bordering*	0.28	6.12
6	*in*	5.77	5.05	*toxicological*	0.10	4.83
7	*of the (s.)*	5.77	4.93	*sealer*	0.06	4.20
8	*and*	3.85	3.24	*pre-established*	0.07	4.16
9	–	–	–	*clinical*	1.30	4.13
10	–	–	–	*0.05*	0.10	3.92
11	–	–	–	*indeterminate*	0.06	3.77
12	–	–	–	*0*	0.04	3.58
13	–	–	–	*statistical*	0.12	3.57
14	–	–	–	*statistical/ statistics*	0.04	3.56
15	–	–	–	*α=0.05*	0.03	3.27
16	–	–	–	*p-value*	0.03	3.26
17	–	–	–	*p=0.05*	0.03	3.25
18	–	–	–	*dose-response*	0.03	3.21
19	–	–	–	*prefixed*	0.03	3.17
20	–	–	–	*physiological*	0.06	3.11

Source: Collocates identified with Sketch Engine. PER = number of co-occurrences between collocate and keyword divided by the overall frequency of the keyword in the corpus. Shared collocates are bounded by colored rectangles.

Table A.9 – Translation of Table 7.15 (Left collocates of *regressão* ["regression"] in two corpora).

POS	ptRAs			ptTenTen11		
	Word	PER	LogDice	Word	PER	LogDice
1	*by means of*	2.78	9.64	*accelerated*	0.06	3.30
2	*through*	4.17	7.84	*marked*	0.08	3.23
3	*in the*	9.72	7.19	*brutal*	0.07	3.03
4	*of the*	11.11	6.34	*presented*	0.16	2.65
5	*of*	43.06	6.31	*deep*	0.15	2.54
6	*to the*	2.78	5.87	*frank*	0.13	2.54
7	*the*	16.67	5.65	*fast*	0.04	2.45
8	*how*	2.78	5.41	*sharp*	0.03	2.34
9	–	–	–	*using*	0.04	2.33
10	–	–	–	*tremendous*	0.03	2.32
11	–	–	–	*(it) was used*	0.02	2.14
12	–	–	–	*frightening*	0.02	2.12
13	–	–	–	*uniterms*	0.01	2.07
14	–	–	–	*1hr.30min.*	0.01	2.03
15	–	–	–	*there was*	0.48	2.03
16	–	–	–	*titled*	0.01	1.99
17	–	–	–	*progressive*	0.04	1.94
18	–	–	–	*observed*	0.02	1.85
19	–	–	–	*would suffer*	0.01	1.84
20	–	–	–	*entail*	0.02	1.75

Source: Collocates identified with Sketch Engine. PER = number of co-occurrences between collocate and keyword divided by the overall frequency of the keyword in the corpus.

Table A.10 – Translation of Table 7.16 (Right collocates of *regressão* [“regression”] in two corpora).

	ptRAs			ptTenTen11		
POS	Word	PER	LogDice	Word	PER	LogDice
1	*multiple*	11.11	11.57	*linear*	10.03	9.86
2	*logistical*	9.72	11.38	*multiple*	2.36	8.15
3	*linear*	2.78	9.66	*logistical*	6.65	8.06
4	*univariate*	2.78	9.66	*hypnotic*	0.43	7.01
5	*indicate*	2.78	9.22	*spontaneous*	1.14	6.90
6	*was*	2.78	5.78	*multivariate*	0.23	6.05
7	*with*	4.17	5.39	*polynomial*	0.21	6.02
8	*of*	20.83	5.26	*infinite*	0.39	5.70
9	*to (for)*	2.78	4.58	*tumoral*	0.20	5.67
10	–	–	–	*nonlinear*	0.17	5.55
11	–	–	–	*civilizational*	0.17	5.50
12	–	–	–	*quadratic*	0.10	4.96
13	–	–	–	*demographic*	0.14	4.70
14	–	–	–	*univariate*	0.07	4.53
15	–	–	–	*authoritarian*	0.11	4.44
16	–	–	–	*stepwise*	0.07	4.43
17	–	–	–	*precautionary*	0.15	4.23
18	–	–	–	*hierarchical*	0.09	4.21
19	–	–	–	*quantile*	0.05	4.12
20	–	–	–	*therapeutic*	0.19	4.05

Source: Collocates identified with Sketch Engine. PER = number of co-occurrences between collocate and keyword divided by the overall frequency of the keyword in the corpus. Shared collocates are bounded by colored rectangles.

Table A.11 – Translation of Table 7.17 (Top 20 strongest left collocates of *variáveis* [noun, “variables”] in two corpora).

	ptRAs			ptTenTen11		
POS	Word	PER	LogDice	Word	PER	LogDice
1	*the*	41.38	10.66	*numerous*	0.73	6.16
2	*of the*	15.33	10.01	*multiple*	0.43	6.14
3	*those*	3.07	9.45	*following*	1.14	5.41
4	*to the*	3.45	9.06	*several*	1.69	5.41
5	*three*	2.68	9.03	*of these*	0.66	5.35
6	*other*	2.30	8.92	*other*	4.86	5.33
7	*two*	1.92	8.67	*those*	1.87	5.27
8	*six*	1.15	8.16	*of those*	0.90	5.22
9	*possible*	0.77	7.74	*these*	1.36	5.17
10	*some*	0.77	7.61	*many*	0.44	5.07
11	*four*	0.77	7.48	*two*	3.38	4.99
12	*by the*	0.77	7.48	*many*	2.03	4.83
13	*what*	0.77	7.40	*some*	2.10	4.74
14	*in the*	0.77	6.86	*main*	1.31	4.73
15	*to (for)*	2.30	6.07	*the*	24.66	4.26
16	*between*	0.77	5.84	*infinite*	0.07	4.14
17	*how*	1.15	5.83	*determinate*	0.13	4.12
18	*of*	5.75	5.24	*various*	0.86	4.12
19	*with*	0.77	4.69	*different*	0.82	3.98
20	*in*	0.77	4.37	*certain*	0.15	3.96

Source: Collocates identified with Sketch Engine. PER = number of co-occurrences between collocate and keyword divided by the overall frequency of the keyword in the corpus. Shared collocates are bounded by colored rectangles.

Table A.12 – Translation of Table 7.18 (Top 20 strongest right collocates of *variáveis* [noun, "variables"] in two corpora).

	ptRAs			ptTenTen11		
POS	Word	PER	LogDice	Word	PER	LogDice
1	*independent*	9.58	11.42	*random*	0.72	7.79
2	*dependent*	4.21	10.35	*studied*	0.79	7.58
3	*categorical*	3.07	9.93	*independent*	1.33	7.54
4	*primary*	1.92	9.26	*demographic*	0.55	7.39
5	*endogenous*	1.92	9.25	*macroeconomic*	0.49	7.27
6	*numerical*	1.53	8.95	*explanatory*	0.50	7.23
7	*socioeconomic*	1.53	8.93	*quantitative*	0.46	7.08
8	*maternal*	1.53	8.91	*involved*	1.00	7.05
9	*analyzed*	1.53	8.82	*categorical*	0.38	6.95
10	*explanatory*	1.15	8.54	*analyzed*	0.59	6.91
11	*exogenous*	1.15	8.53	*continuous*	0.39	6.75
12	*used*	1.15	8.40	*qualitative*	0.30	6.48
13	*intervening*	0.77	7.96	*socio-demographic*	0.27	6.44
14	*manifest*	0.77	7.96	*sociodemographic*	0.26	6.42
15	*latent*	0.77	7.95	*dependent*	0.53	6.39
16	*continuous*	0.77	7.95	*economic*	0.74	6.38
17	*contributed*	0.77	7.93	*meteorological*	0.28	6.32
18	*observed*	0.77	7.92	*global*	0.47	6.28
19	*clinical*	0.77	7.83	*environmental*	1.10	6.17
20	*associated*	0.77	7.78	*psychological*	0.29	6.15

Source: Collocates identified with Sketch Engine. PER = number of co-occurrences between collocate and keyword divided by the overall frequency of the keyword in the corpus. Shared collocates are bounded by colored rectangles.

Table A.13 – Translation of Table 7.19 (Manually identified collocates of 本論文 [*honronbun*, "this paper"] in two corpora).

Collocate	jaRAs		jaTenTen11 LUW	
	Instances	PER	Instances	PER
Left (WORD + 本論文)				
soko de ("then," "in view of this")	2	5.00	2	2.82
ga (indicates the subject, among other functions)	2	5.00	1	1.41
motozuki ("based on")	2	5.00	0	0.00
wa (indicates the topic, among other functions)	1	2.50	5	7.04
yōni ("like," "in this way")	0	0.00	2	2.82
Right (本論文 + WORD)				
de wa ("at," "in")	18	45.00	23	32.39
no ("of")	7	17.50	10	14.08
ga (indicates the subject, among other functions)	5	12.50	0	0.00
wa (indicates the topic, among other functions)	2	5.00	26	36.62
ni oite ("at," "in")	2	5.00	0	0.00
de ("by," "through," "at")	0	0.00	4	5.63
ni okeru ("in")	0	0.00	3	4.23
o (indicates the target of an action)	0	0.00	3	4.23

Source: Prepared by the author. PER = percentage of instances in relation to the overall number of occurrences of 本論文 in the corpus. Includes only collocates appearing at least two times in one of the two corpora. Shared collocates are bounded by colored rectangles. The double underline indicates a verb with partial inflection.

Table A.14 – Translation of Table 7.20 (Manually identified collocates of 本稿 [*honkō*, “this paper”] in two corpora).

Collocate	jaRAs		jaTenTen11 LUW	
	Instances	PER	Instances	PER
Left (WORD + 本稿)				
ga (indicates the subject, among other functions)	22	6.98	20	5.14
soko de (“then,” “in view of this”)	12	3.81	19	4.88
wa (indicates the topic, among other functions)	12	3.81	10	2.57
tame (“for,” “in order to,” “because”)	12	3.81	2	0.51
nao (“further,” “yet”)	10	3.17	11	2.83
mata (“also,” “again”)	7	2.22	1	0.26
o (indicates the target of an action)	6	1.90	4	1.03
de (“by,” “through,” “at”)	6	1.90	2	0.51
ni (“at,” “in”)	6	1.90	1	0.26
de wa (“at,” “in”)	5	1.59	2	0.51
shikashi (“however”)	5	1.59	1	0.26
toshite (“as,” “supposing that”)	5	1.59	1	0.26
yotte (“consequently”)	5	1.59	1	0.26
fumae (“based on”)	4	1.27	0	0.00
no (“of”)	1	0.32	5	1.29
no de (“because,” “since”)	0	0.00	5	1.29
Right (本稿 + WORD)				
de wa (“at,” “in”)	143	45.40%	226	58.10%
no (“of”)	79	25.08%	36	9.25%
wa (indicates the topic, among other functions)	30	9.52%	49	12.60%
de (“by,” “through,” “at”)	30	9.52%	25	6.43%
ga (indicates the subject, among other functions)	8	2.54%	3	0.77%
ni okeru (“in”)	8	2.54%	0	0.00%
de no (“with,” of + by)	5	1.59%	4	1.03%
o (indicates the target of an action)	1	0.32%	20	5.14%
ni (“at,” “in”)	1	0.32%	5	1.29%
ni oite wa (“in,” “at,” with emphasis)	0	0.00%	4	1.03%

Source: Prepared by the author. PER = percentage of instances in relation to the overall number of occurrences of 本稿 in the corpus. Includes only collocates appearing at least four times in one of the two corpora. Shared collocates are bounded by colored rectangles. The double underline indicates a verb with partial inflection.

Table A.15 – Translation of Table 7.21 (Manually filtered collocates of 症例 [*shōrei*, "clinical case"] in two corpora).

Collocate	jaRAs Instances	jaRAs PER	jaTenTen11 LUW Instances	jaTenTen11 LUW PER
Left (WORD + 症例)				
no ("of")	56	18.36	346	29.73
kono ("this")	5	1.64	40	3.44
yōsuru ("need," "require")	5	1.64	0	0.00
yūsuru ("have," "own")	5	1.64	0	0.00
koeru ("exceed," "transcend")	4	1.31	0	0.00
muzukashii ("difficult")	0	0.00	17	1.46
mezurashii ("rare")	0	0.00	7	0.60
takai ("high")	0	0.00	7	0.60
de ("by," "through," "at")	0	0.00	5	0.43
o (indicates the target of an action)	0	0.00	5	0.43
onaji ("the same," "identical")	0	0.00	4	0.34
sono ("that")	0	0.00	4	0.34
ya ("and," "or")	0	0.00	4	0.34
Right (症例 + WORD)				
wa (indicates the topic, among other functions)	45	14.75	102	8.76
ga (indicates the subject, among other functions)	29	9.51	153	13.14
o (indicates the target of an action)	22	7.21	161	13.83
no ("of")	21	6.89	108	9.28
1	21	6.89	23	1.98
mo ("too," "also")	14	4.59	58	4.98
ni taishite ("in respect to," "while")	13	4.26	11	0.95
4	9	2.95	0	0.00
ya ("and," "or")	8	2.62	12	1.03
2	7	2.30	13	1.12
3	6	1.97	0	0.00
5	5	1.64	0	0.00
7	5	1.64	0	0.00
kara ("from")	0	0.00	18	1.55
nado ("etc.")	0	0.00	13	1.12
toshite ("as," "supposing that")	0	0.00	9	0.77
ni taishi ("in respect to," "against")	0	0.00	6	0.52
ni okeru ("in")	0	0.00	4	0.34

Source: Collocates identified with Sketch Engine and filtered by the author. PER = percentage of instances in relation to the overall number of occurrences of 症例 in the corpus. Includes only collocates appearing at least four times in one of the two corpora. Shared collocates are bounded by colored rectangles.

Table A.16 – Translation of Table 7.25 (Semantic sets of *significância* ["significance"] in two corpora).

Corpus	SET (Collocates)
ptRAs	ACTION AS TARGET (*presented*, *lost*)
	NUMBERS AND MATHEMATICAL SYMBOLS (*5*, *95*, *%*, *<*, *0.05*, *p*, *α=0.05*)
	STATISTICS (*statistical/statistics*, *interval*, *model*, *level*, *proportions*, *variables*)
ptTenTen11	ATTRIBUTES (*statistical/statistics*, *bordering*, *prognostic*)
	NUMBERS AND MATHEMATICAL SYMBOLS (*0.05*, *<*, ≤, *0.000*, *0.001*, *0.01*, *0.05*, *0.050*, *fz*, *p*, *p=0.05*, *α*)
	STATISTICAL TESTS (*Anova*, *Bonferroni*, *Fisher*, *Kruskal-Wallis*, *Mann-Whitney*, *non-parametric*, *paired*, *Pearson*, *chi-squared*, *Student*, *Tukey*, *Wilcoxon*)
	STATISTICS (*bivariate*, *coefficients*, *correlation*, *correlations*, *statistically*, *statistical*, *statistical/statisticians*, *multivariate*, *level*, *p results*, *p-value*, *regression*, *Spearman*, *SPSS*, *univariate*, *p-value*, *variance*, *variables*)

Source: Semantic sets manually identified from collocate lists generated by Sketch Engine. Shared sets and collocates are indicated with colored boxes. SPSS = Statistical Package for the Social Sciences.

Table A.17 – Translation of Table 7.26 (Semantic sets of *regressão* ["regression"] in two corpora).

Corpus	SET (Collocates)
ptRAs	MEANS (*by means of*, *medium*, *methods*, *through*, *used*, *employing*) NUMBERS AND MATHEMATICAL SYMBOLS (*5*, *F*) RESOURCES (*knowledge*, *software*) STATISTICS (*analysis*, *analyses*, *coefficient*, *coefficients*, *DEA*, *equation*, *estimates*, *model*, *models*, *chi-squared*, *results*, *robust*, *significant*, *significant*, *variances*) TYPE OF REGRESSION (*adjusted*, *binary*, *Cox*, *linear*, *logistic*, *moderate*, *multilevel*, *multinomial*, *multiple*, *multivariate*, *MMR*, *Poisson*, *univariate*)
ptTenTen11	MEDICINE (*edema*, *luteum*, *neural*, *progression*, *tumor*) PSYCOTHERAPY (*hypnosis*, *past*, *therapy*, *PLT*) STATISTICS (*analysis*, *Anova*, *coefficient*, *coefficients*, *correlation*, *discriminant*, *equation*, *equations*, *statistical/statistics*, *estimation*, *estimators*, *estimate, inference*, *linear*, *models*, *straight line*, *chi-squared*, *significance*, *variance*, *variables*, *variable*) TYPE OF REGRESSION (*binary*, *civilizational*, *Cox*, *spontaneous*, *hypnotic*, *infinite*, *linear*, *logistic*, *multinomial*, *multiple*, *multivariate*, *nonlinear*, *Poisson*, *polynomial*, *quadratic*, *stepwise*, *tumoral*, *univariate*)

Source: Semantic sets manually identified from collocate lists generated by Sketch Engine. Shared sets and collocates are indicated with colored boxes. DEA = data envelopment analysis. MMR = moderated multiple regression. PLT = past life therapy.

Table A.18 – Translation of Table 7.27 (Semantic sets of *variáveis* ["variables"] in two corpora).

Corpus	SET (Collocates)
ptRAs	ACTION AS TARGET (*analyzed*, *considered*, *employed*) AMOUNT (*two*, *three*) STATISTICS (*absolute*, *analysis*, *association*, *bivariate*, *frequencies*, *measurement*, *model*, *models*, *standard*, *relations*, *significance*, *significant*, *significant*) TYPE OF VARIABLE (*categorical*, *confounding*, *control*, *dependent*, *endogenous*, *exogenous*, *independent*, *interest*, *maternal*, *numerical*, *primary*, *socioeconomic*)
ptTenTen11	ACTION AS AGENT (*affect*, *influence*, *interfere*, *determine*) ACTION AS TARGET (*analyzed*, *evaluated*, *considered*, *controlled*, *studied*, *selected*, *employed*) PERTINENT (*associated*, *involved*, *related*) STATISTICS (*correlation*, *correlations*, *equation*, *equations*, *statistically*, *regression*, *significant*, *variable*) TYPE OF VARIABLE (*random*, *environmental*, *anthropometric*, *categorical*, *climatic*, *continuous*, *controllable*, *demographic*, *dependent*, *discrete*, *economic*, *explanatory*, *external*, *physiological*, *global*, *independent*, *latent*, *macroeconomic*, *meteorological*, *psychological*, *qualitative*, *quantitative*, *sociodemographic*, *socio-demographic*, *socioeconomic*)

Source: Semantic sets manually identified from collocate lists generated by Sketch Engine. Shared sets and collocates are indicated with colored boxes.

Table A.19 – Translation of Table 7.28 (Semantic sets of 本論文 [*honronbun*, “this paper”] in two corpora).

Corpus	SET (Collocates)
jaRAs	ARTICLE CONTENT (*mondai*, “issue,” “problem,” “question”; *mokuteki*, “aim”; *haikei*, “background”; *kadai*, “challenge,” “matter”; *genkai*, “limitation”) AUTHOR’S (PAPER’S) ACTIONS (*motozuki*, “draw on,” “be based on”; *shinkokusu*[*beki*], “report,” “must be reported”; *kōsatsu suru*, “discuss”) DEMONSTRATIVE FAMILY (*kono*, “this”; *soko*, “there”; *sono*, “that”) SECTION NUMBER (*1*, *3*) SOCIETY (*inobēshon*, “innovation”; *jissen kyōdōtai*, “community of practice”; *shakaiteki sokumen*, “social aspect,” “the social side”; *soshiki*, “organization”) THOUGHT (*mikata*, “viewpoint”; *mondai ishiki*, “problem awareness”)
jaTenTen11 LUW	ARTICLE CONTENT (*naiyō*, “content”; *teian*, “suggestion”; *ketsuron*, “conclusion”) ATTRIBUTES (*atarashii*, “new”; *jūyō*, “important”)

Source: Semantic sets manually identified from collocate lists generated by Sketch Engine. All the collocates of the Japanese RAs corpus were considered (41); only collocates with positive logDice values of jaTenTen11 LUW were considered (16, in total). Shared sets and collocates are indicated with colored boxes. Verbs with partial inflection are double underlined.

Table A.20 – Translation of Table 7.29 (Semantic sets of 本稿 [*honkō*, "this paper"] in two corpora).

Corpus	SET (Collocates)
jaRAs	ARTICLE CONTENT (*hajime*, "introduction"; *bunseki*, "analysis"; *jirei*, "case"; *ronten*, "the point"; *genkai*, "limitation"; *ten*, "point"; *bunseki kekka*, "analysis results"; *kadai*, "challenge," "matter"; *senkō kenkyū*, "previous research") AUTHOR'S (PAPER'S) ACTIONS (*bunseki suru*, "analyze"; *toriage*, "select," "pick up"; *teiji shi*, "present"; *mochii*, "use," "employ"; *kangaeru*, "think"; *fumae*, "draw on," "be based on") DEMONSTRATIVE FAMILY (*kono*, "this"; *kore*, "this"; *korera*, "these"; *soko*, "there") INTERNAL REFERENCE (*ijō*, "above"; *ika*, "below") MANAGEMENT (*kigyōkateki shikō*, "entrepreneurial orientation"; *shinki jigyō kaihatsu*, "development of new business"; *senryaku gurūpu*, "strategy group") SECTION NUMBERS (*1*, *5*) RESEARCH METHODS (*mondai ishiki*, "problem awareness"; *jisshō bunseki*, "empirical analysis"; *taishō*, "target," "subject"; *shōten*, "focus"; *kenkyū*, "research")
jaTenTen11 LUW	ARTICLE CONTENT (*shudai*, "subject," "theme"; *kijutsu*, "description," "account"; *chū*, "note"; *conseputo*, "concept"; *keii*, "details"; *shushi*, "aim," "purpose") AUTHOR'S (PAPER'S) ACTIONS (*go shōkai shi*, "introduce"; *toriage*, "select," "pick up"; *shippitsu shi*, "write"; *shōkai shi*, "introduce"; *shirusu*, "write down," "mention"; *ronji*, "argue," "deal with"; *nobe*, "state," "mention"; *noberu*, "state," "mention") FOCUS (*omo*, "main"; *omo ni*, "mainly"; *shōten*, "focus"; *chūmoku*, "attention") MANAGEMENT (*chūshō kigyō*, "small and medium-sized enterprises"; *kigyō*, "business"; *dōsha*, "the same firm"; *kaihatsu*, "development") WRITING (*zenpen*, "the first part," "the first volume"; *shippitsu*, "writing"; *shippitsu jiten*, "the time of writing"; *hissha*, "the writer," "the author")

Source: Semantic sets manually identified from collocate lists generated by Sketch Engine. Shared sets and collocates are indicated with colored boxes. Verbs with partial inflection are double underlined.

Table A.21 – Translation of Table 7.30 (Semantic sets of 症例 [*shōrei*, "clinical case"] in two corpora).

Corpus	SET (Collocates)
jaRAs	CASE DESCRIPTION (*yūsuru*, "have," "own"; *yōshi*, "need"; *yōsuru*, "need"; *koeru*, "exceed") CASE NUMBERS (*1*, *2*, *3*, *4*, *5*, *7*, *8*, *9*) DIAGNOSIS (*utagawa*, "suspect"; *hakken sa*, "discover"; *mitome*, "notice," "consider") HEALTH ISSUES (*honshō*, "this disease"; *kitashi*, "provoke," "produce"; *kidō ibutsu*, "foreign body in airway"; *shōjō*, "symptom," "condition"; *hasshō shi*, "appearance of symptoms"; *keiken shi*, "experience"; *shokudō heisashō*, "esophageal atresia") NUMBER OF CASES (*1 rei*, "case 1"; *2 rei*, "case 2") TREATMENT (*hozonteki chiryō*, "conservative treatment"; *shujutsu*, "surgery"; *shikō sa*, "carry out"; *shikō shi*, "carry out"; *chiryō*, "treatment"; *okonatt*, "do," "conduct")
jaTenTen11 LUW	DIAGNOSIS (*rentogen*, "X-ray"; *kakunin sa*, "confirm"; *shindan*, "diagnosis"; *shindan sa*, "diagnose") HEALTH ISSUES (*kioku sōshitsu*, "amnesia," "memory loss"; *senten ijō shōkōgun*, "syndrome of congenital abnormalities"; *saihatsu*, "recurrence"; *kiso shikkan*, "underlying disease"; *sokusenshō*, "embolism"; *kyūsei sanzaisei nōseki zuien*, "acute disseminated encephalomyelitis"; *akushō shuyō*, "malignant tumor"; *kansen suru*, "get infected"; *toppatsusei kanshitsusei haien*, "idiophatic interstitial pneumonia"; *gankyū undo shōgai*, "ocular motility disorder"; *jūtoku*, "serious") REPORT (*hōkoku sa*, "report"; *hōkoku shi*, "report"; *hōkoku suru*, "report"; *teiji shi*, "present"; *teiji suru*, "present") TREATMENT (*OPCAB*; *zentō hoobone hōgō*, "frontozygomatic suture"; *juseiritsu*, "fertilization rate"; *basshi shi*, "extract a tooth"; *seifuku sōsa*, "manipulation to correct the position of bones and tissue"; *shikō shi*, "carry out"; *eiyō sa*, "nutrition"; *chiryō shi*, "treat"; *keikō tōyaku*, "oral administration"; *tekiō dekiru*, "can adapt to"; *jotsū*, "desensitization," "pain relief")

Source: Semantic sets manually identified from collocate lists generated by Sketch Engine. Verbs with partial inflection are double underlined. OPCAB = off-pump coronary artery bypass.

APPENDIX B – ENGLISH TRANSLATIONS OF CHAPTER 8 TABLES

This appendix presents English translations of tables from Chapter 8 that have Portuguese and Japanese words. Table B.1 below shows the correspondence between the original and the translated tables. Exclusive tables for Romanized transcriptions of Japanese are provided in four cases due to space limitations in the original layouts.

Table B.1 – Correspondence between original and translated tables.

Original Table	Translated Table
8.3	B.2
8.4	B.3
8.5	B.4 (Transcription) and B.5 (Translation)
8.6	B.6 (Transcription) and B.7 (Translation)
8.9	B.8
8.10	B.9
8.11	B.10 (Transcription) and B.11 (Translation)
8.12	B.12 (Transcription) and B.13 (Translation)
8.12	B.14
8.16	B.15
8.17	B.16
8.18	B.17
8.21	B.18
8.22	B.19
8.23	B.20
8.24	B.21

Source: Prepared by the author.

Table B.2 – Translation of Table 8.3 (Key keywords in ptPED against ptMGT).

POS	Word	General Freq.		Freq. per Million		RA Freq.		Score
		FC	RC	FC	RC	n	(%)	
1	*mothers*	116	0	1,405.74	0.00	11	45.83	1,406.7
2	*adolescents*	106	0	1,284.55	0.00	10	41.67	1,285.6
3	*maternal*	96	0	1,163.37	0.00	9	37.50	1,164.4
4	*feeding*	78	0	945.24	0.00	8	33.33	946.2
5	*mother*	52	0	630.16	0.00	8	33.33	631.2
6	*95% CI*	50	0	605.92	0.00	10	41.67	606.9
7	*children*	321	1	3,890.01	5.69	20	83.33	581.2
8	*early*	28	0	339.32	0.00	10	41.67	340.3
9	*municipality*	28	0	339.32	0.00	8	33.33	340.3
10	*infant*	27	0	327.20	0.00	10	41.67	328.2
11	*among*	23	0	278.72	0.00	11	45.83	279.7
12	*healthy*	22	0	266.61	0.00	8	33.33	267.6
13	*boys*	22	0	266.61	0.00	8	33.33	267.6
14	*committee*	20	0	242.37	0.00	19	79.17	243.4
15	*childhood*	20	0	242.37	0.00	10	41.67	243.4
16	*presented*	18	0	218.13	0.00	8	33.33	219.1
17	*clinical*	18	0	218.13	0.00	9	37.50	219.1

Source: Keyword analysis performed with Sketch Engine. FC = focus corpus. RC = reference corpus. RA Freq. = number (n) and percentage (%) of RAs containing the word. 95% CI = confidence interval of 95%.

Table B.3 – Translation of Table 8.4 (Key keywords in ptMGT against ptPED).

POS	Word	General Freq.		Freq. per Million		RA Freq.		Score
		FC	RC	FC	RC	n	(%)	
1	*organizational*	213	0	1,213.01	0.00	10	41.67	1,214.0
2	*consumers*	211	0	1,201.62	0.00	10	41.67	1,202.6
3	*company*	147	0	837.15	0.00	16	66.67	838.1
4	*attitude*	112	0	637.83	0.00	8	33.33	638.8
5	*consumer*	74	0	421.42	0.00	8	33.33	422.4
6	*clients*	72	0	410.03	0.00	9	37.50	411.0
7	*integration*	70	0	398.64	0.00	8	33.33	399.6
8	*financial*	58	0	330.30	0.00	12	50.00	331.3
9	*financial*	56	0	318.91	0.00	12	50.00	319.9
10	*perceived*	53	0	301.83	0.00	9	37.50	302.8
11	*organizational*	52	0	296.13	0.00	11	45.83	297.1
12	*efficiency*	45	0	256.27	0.00	15	62.50	257.3
13	*influence*	44	0	250.58	0.00	13	54.17	251.6
14	*marketing*	43	0	244.88	0.00	11	45.83	245.9
15	*managerial*	41	0	233.49	0.00	9	37.50	234.5
16	*flow*	37	0	210.71	0.00	9	37.50	211.7
17	*equations*	33	0	187.93	0.00	9	37.50	188.9
18	*author*	27	0	153.76	0.00	11	45.83	154.8

Source: Keyword analysis performed with Sketch Engine. FC = focus corpus. RC = reference corpus. RA Freq. = number (n) and percentage (%) of RAs containing the word.

Table B.4 – Transcription of Table 8.5 (Key keywords in jaPED against jaMGT).

		General Freq.		Freq. per Million		RA Freq.		
POS	Word	FC	RC	FC	RC	n	(%)	Score
1	*shōrei*	305	0	4,659.26	0.00	23	95.83	4,660.3
2	*honshō*	116	0	1,772.05	0.00	9	37.50	1,773.0
3	*shikō shi*	113	0	1,726.22	0.00	19	79.17	1,727.2
4	*chiryō*	81	0	1,237.38	0.00	16	66.67	1,238.4
5	*shujutsu*	81	0	1,237.38	0.00	20	83.33	1,238.4
6	*jutsugo*	63	0	962.41	0.00	15	62.50	963.4
7	*jikenrei*	54	0	824.92	0.00	15	62.50	825.9
8	*shōjō*	53	0	809.64	0.00	10	41.67	810.6
9	*tōka*	49	0	748.54	0.00	12	50.00	749.5
10	*tōin*	47	0	717.98	0.00	15	62.50	719.0
11	*zenrei*	46	0	702.71	0.00	16	66.67	703.7
12	*kanji*	42	0	641.60	0.00	15	62.50	642.6
13	*hasshō shi*	42	0	641.60	0.00	14	58.33	642.6
14	*jutsushiki*	35	0	534.67	0.00	9	37.50	535.7
15	*gappeishō*	35	0	534.67	0.00	15	62.50	535.7
16	*shōni*	34	0	519.39	0.00	10	41.67	520.4
17	*danji*	32	0	488.84	0.00	13	54.17	489.8
18	*honpō*	31	0	473.56	0.00	13	54.17	474.6
19	*kōhōshiteki*	30	0	458.29	0.00	15	62.50	459.3
20	*yogo*	29	0	443.01	0.00	10	41.67	444.0
21	*shindan*	27	0	412.46	0.00	9	37.50	413.5
22	*seigo*	25	0	381.91	0.00	8	33.33	382.9
23	*III kekka*	24	0	366.63	0.00	24	100.00	367.6
24	*shindan sa*	24	0	366.63	0.00	10	41.67	367.6
25	*II taishō*	24	0	366.63	0.00	24	100.00	367.6
26	*IV kōsatsu*	24	0	366.63	0.00	24	100.00	367.6
27	*yōshi*	24	0	366.63	0.00	24	100.00	367.6
28	*chiryō hōshin*	24	0	366.63	0.00	9	37.50	367.6
29	*shinseijiki*	23	0	351.35	0.00	8	33.33	352.4
30	*hasshō*	23	0	351.35	0.00	9	37.50	352.4
31	*hōkoku suru*	20	0	305.53	0.00	16	66.67	306.5
32	*shikkan*	20	0	305.53	0.00	11	45.83	306.5
33	*shikō suru*	17	0	259.70	0.00	9	37.50	260.7
34	*ji*	15	0	229.14	0.00	8	33.33	230.1

Source: Keyword analysis performed with Sketch Engine. FC = focus corpus. RC = reference corpus. RA Freq. = number (n) and percentage (%) of RAs containing the word. Verbs with partial inflection are double underlined.

Table B.5 – Translation of Table 8.5 (Key keywords in jaPED against jaMGT).

POS	Word	General Freq. FC	General Freq. RC	Freq. per Million FC	Freq. per Million RC	RA Freq. n	RA Freq. (%)	Score
1	*clinical case*	305	0	4,659.26	0.00	23	95.83	4,660.3
2	*this disease*	116	0	1,772.05	0.00	9	37.50	1,773.0
3	*carry out*	113	0	1,726.22	0.00	19	79.17	1,727.2
4	*treatment*	81	0	1,237.38	0.00	16	66.67	1,238.4
5	*surgery*	81	0	1,237.38	0.00	20	83.33	1,238.4
6	*postoperative*	63	0	962.41	0.00	15	62.50	963.4
7	*medical case study*	54	0	824.92	0.00	15	62.50	825.9
8	*a symptom*	53	0	809.64	0.00	10	41.67	810.6
9	*this department*	49	0	748.54	0.00	12	50.00	749.5
10	*this clinic*	47	0	717.98	0.00	15	62.50	719.0
11	*all instances*	46	0	702.71	0.00	16	66.67	703.7
12	*child patient*	42	0	641.60	0.00	15	62.50	642.6
13	*appearance of symptoms*	42	0	641.60	0.00	14	58.33	642.6
14	*operative method*	35	0	534.67	0.00	9	37.50	535.7
15	*complications*	35	0	534.67	0.00	15	62.50	535.7
16	*an infant*	34	0	519.39	0.00	10	41.67	520.4
17	*a boy*	32	0	488.84	0.00	13	54.17	489.8
18	*Japan*	31	0	473.56	0.00	13	54.17	474.6
19	*retrospective*	30	0	458.29	0.00	15	62.50	459.3
20	*prognosis*	29	0	443.01	0.00	10	41.67	444.0
21	*diagnosis*	27	0	412.46	0.00	9	37.50	413.5
22	*after birth*	25	0	381.91	0.00	8	33.33	382.9
23	*III results*	24	0	366.63	0.00	24	100.00	367.6
24	*diagnose*	24	0	366.63	0.00	10	41.67	367.6
25	*II subject*	24	0	366.63	0.00	24	100.00	367.6
26	*IV discussion*	24	0	366.63	0.00	24	100.00	367.6
27	*summary*	24	0	366.63	0.00	24	100.00	367.6
28	*treatment policy*	24	0	366.63	0.00	9	37.50	367.6
29	*neonatal period*	23	0	351.35	0.00	8	33.33	352.4
30	*appearance of symptoms*	23	0	351.35	0.00	9	37.50	352.4
31	*report*	20	0	305.53	0.00	16	66.67	306.5
32	*disease*	20	0	305.53	0.00	11	45.83	306.5
33	*carry out*	17	0	259.70	0.00	9	37.50	260.7
34	*child*	15	0	229.14	0.00	8	33.33	230.1

Source: Keyword analysis performed with Sketch Engine. FC = focus corpus. RC = reference corpus. RA Freq. = number (n) and percentage (%) of RAs containing the word. Verbs with partial inflection are double underlined.

Table B.6 – Transcription of Table 8.6 (Key keywords in jaMGT against jaPED).

		General Freq.		Freq. per Million		RA Freq.		
POS	Word	FC	RC	FC	RC	n	(%)	Score
1	*kigyō*	376	0	2,211.49	0.00	21	87.50	2,212.5
2	*honkō*	315	0	1,852.71	0.00	21	87.50	1,853.7
3	*kokyaku*	232	0	1,364.54	0.00	12	50.00	1,365.5
4	*membā*	209	0	1,229.26	0.00	8	33.33	1,230.3
5	*shigen*	169	0	993.99	0.00	12	50.00	995.0
6	*senkō kenkyū*	144	0	846.95	0.00	19	79.17	848.0
7	*jirei*	134	0	788.14	0.00	14	58.33	789.1
8	*jūgyōin*	128	0	752.85	0.00	8	33.33	753.8
9	*honkenkyū*	124	0	729.32	0.00	12	50.00	730.3
10	*moderu*	106	0	623.45	0.00	18	75.00	624.5
11	*inobēshon*	93	0	546.99	0.00	9	37.50	548.0
12	*kōdō*	91	0	535.23	0.00	13	54.17	536.2
13	*kizon kenkyū*	78	0	458.77	0.00	13	54.17	459.8
14	*hensū*	77	0	452.89	0.00	12	50.00	453.9
15	*senryaku*	76	0	447.00	0.00	15	62.50	448.0
16	*no dearu*	75	0	441.12	0.00	13	54.17	442.1
17	*o tsūjite*	69	0	405.83	0.00	21	87.50	406.8
18	*takameru*	67	0	394.07	0.00	15	62.50	395.1
19	*shijō*	67	0	394.07	0.00	14	58.33	395.1
20	*ishi kettei*	65	0	382.31	0.00	10	41.67	383.3
21	*gurūpu*	57	0	335.25	0.00	8	33.33	336.3
22	*jigyō*	52	0	305.84	0.00	11	45.83	306.8
23	*purosesu*	48	0	282.32	0.00	13	54.17	283.3
24	*bunseki kekka*	47	0	276.44	0.00	11	45.83	277.4
25	*bunseki shi*	47	0	276.44	0.00	13	54.17	277.4
26	*ronji*	47	0	276.44	0.00	11	45.83	277.4
27	*nihon kigyō*	47	0	276.44	0.00	8	33.33	277.4
28	*shōten*	46	0	270.55	0.00	14	58.33	271.6
29	*torae*	46	0	270.55	0.00	14	58.33	271.6
30	*shō/masa*	45	0	264.67	0.00	8	33.33	265.7
31	*sōtei shi*	42	0	247.03	0.00	10	41.67	248.0
32	*futatsu*	42	0	247.03	0.00	8	33.33	248.0
33	*naze*	42	0	247.03	0.00	15	62.50	248.0
34	*dai ni*	42	0	247.03	0.00	12	50.00	248.0
35	*atarashii*	42	0	247.03	0.00	15	62.50	248.0
36	*kojin*	42	0	247.03	0.00	11	45.83	248.0
37	*sō*	41	0	241.15	0.00	16	66.67	242.1
38	*dai ichi*	41	0	241.15	0.00	13	54.17	242.1
39	*chakumoku shi*	39	0	229.38	0.00	13	54.17	230.4
40	*takame*	38	0	223.50	0.00	13	54.17	224.5
41	*soshiki nai*	37	0	217.62	0.00	11	45.83	218.6
42	*shuchō*	37	0	217.62	0.00	9	37.50	218.6

Source: Keyword analysis performed with Sketch Engine. FC = focus corpus. RC = reference corpus. RA Freq. = number (n) and percentage (%) of RAs containing the word. Verbs with partial inflection are double underlined.

Table B.7 – Translation of Table 8.6 (Key keywords in jaMGT against jaPED).

POS	Word	General Freq.		Freq. per Million		RA Freq.		Score
		FC	RC	FC	RC	n	(%)	
1	*business*	376	0	2,211.49	0.00	21	87.50	2,212.5
2	*this paper*	315	0	1,852.71	0.00	21	87.50	1,853.7
3	*customer*	232	0	1,364.54	0.00	12	50.00	1,365.5
4	*member*	209	0	1,229.26	0.00	8	33.33	1,230.3
5	*resource*	169	0	993.99	0.00	12	50.00	995.0
6	*previous research*	144	0	846.95	0.00	19	79.17	848.0
7	*case*	134	0	788.14	0.00	14	58.33	789.1
8	*employee*	128	0	752.85	0.00	8	33.33	753.8
9	*this study*	124	0	729.32	0.00	12	50.00	730.3
10	*model*	106	0	623.45	0.00	18	75.00	624.5
11	*innovation*	93	0	546.99	0.00	9	37.50	548.0
12	*behavior*	91	0	535.23	0.00	13	54.17	536.2
13	*existing research*	78	0	458.77	0.00	13	54.17	459.8
14	*variable*	77	0	452.89	0.00	12	50.00	453.9
15	*strategy*	76	0	447.00	0.00	15	62.50	448.0
16	*of*	75	0	441.12	0.00	13	54.17	442.1
17	*through*	69	0	405.83	0.00	21	87.50	406.8
18	*increase*	67	0	394.07	0.00	15	62.50	395.1
19	*market*	67	0	394.07	0.00	14	58.33	395.1
20	*decision making*	65	0	382.31	0.00	10	41.67	383.3
21	*group*	57	0	335.25	0.00	8	33.33	336.3
22	*industry*	52	0	305.84	0.00	11	45.83	306.8
23	*process*	48	0	282.32	0.00	13	54.17	283.3
24	*analysis results*	47	0	276.44	0.00	11	45.83	277.4
25	*analyze*	47	0	276.44	0.00	13	54.17	277.4
26	*discuss*	47	0	276.44	0.00	11	45.83	277.4
27	*Japanese company*	47	0	276.44	0.00	8	33.33	277.4
28	*focus*	46	0	270.55	0.00	14	58.33	271.6
29	*capture*	46	0	270.55	0.00	14	58.33	271.6
30	*certain, correct*	45	0	264.67	0.00	8	33.33	265.7
31	*assume*	42	0	247.03	0.00	10	41.67	248.0
32	*two*	42	0	247.03	0.00	8	33.33	248.0
33	*why*	42	0	247.03	0.00	15	62.50	248.0
34	*second*	42	0	247.03	0.00	12	50.00	248.0
35	*new*	42	0	247.03	0.00	15	62.50	248.0
36	*individual*	42	0	247.03	0.00	11	45.83	248.0
37	*like that*	41	0	241.15	0.00	16	66.67	242.1
38	*first*	41	0	241.15	0.00	13	54.17	242.1
39	*focus on*	39	0	229.38	0.00	13	54.17	230.4
40	*increase*	38	0	223.50	0.00	13	54.17	224.5
41	*within the organization*	37	0	217.62	0.00	11	45.83	218.6
42	*claim*	37	0	217.62	0.00	9	37.50	218.6

Source: Keyword analysis performed with Sketch Engine. FC = focus corpus. RC = reference corpus. RA Freq. = number (n) and percentage (%) of RAs containing the word. Verbs with partial inflection are double underlined.

Table B.8 – Translation of Table 8.9 (Highly distributed common keywords in ptPED against ptMGT).

		General Freq.		Freq. per Million		RA Freq. (%)		
POS	Word	FC	RC	FC	RC	FC	RC	Score
1	*years*	179	85	2,169.2	484.1	91.7	95.8	2.14
2	*study*	364	292	4,411.1	1,662.9	100.0	100.0	2.03
3	*were*	424	373	5,138.2	2,124.2	100.0	100.0	1.97
4	*was*	578	540	7,004.4	3,075.2	100.0	100.0	1.96
5	*use*	130	73	1,575.4	415.7	70.8	75.0	1.82
6	*association*	94	47	1,139.1	267.7	79.2	62.5	1.69
7	*during*	84	44	1,017.9	250.6	83.3	50.0	1.61
8	*majority*	70	29	848.3	165.2	75.0	62.5	1.59
9	*average*	89	57	1,078.5	324.6	75.0	66.7	1.57
10	*present*	84	59	1,017.9	336.0	87.5	70.8	1.51
11	*number*	84	62	1,017.9	353.1	91.7	70.8	1.49
12	*or*	363	465	4,399.0	2,648.1	100.0	100.0	1.48
13	*test*	66	39	799.8	222.1	54.2	58.3	1.47
14	*professional/ professionals*	59	30	715.0	170.8	70.8	54.2	1.47
15	*quality*	94	88	1,139.1	501.2	62.5	70.8	1.43
16	*practice*	61	39	739.2	222.1	66.7	66.7	1.42
17	*attention*	47	22	569.6	125.3	50.0	50.0	1.40
18	*day*	54	35	654.4	199.3	62.5	50.0	1.38
19	*with*	914	1,364	11,076.2	7,767.8	100.0	100.0	1.38
20	*no*	477	692	5,780.5	3,940.9	100.0	100.0	1.37
21	*period*	73	66	884.6	375.9	79.2	62.5	1.37

Source: Keyword analysis performed with Sketch Engine. FC = focus corpus. RC = reference corpus. RA Freq. (%) = percentage of RAs containing the word. The colored rectangle indicates the selected keyword.

Table B.9 – Translation of Table 8.10 (Highly distributed common keywords in ptMGT against ptPED).

		General Freq.		Freq. per Million		RA Freq. (%)		
POS	Word	FC	RC	FC	RC	FC	RC	Score
1	*et*	452	45	2,574.1	545.3	95.8	62.5	2.31
2	*job/work*	229	30	1,304.1	363.6	87.5	62.5	1.69
3	*yours/theirs*	201	24	1,144.7	290.8	100.0	50.0	1.66
4	*if*	631	153	3,593.5	1,854.1	100.0	95.8	1.61
5	*organization*	150	16	854.2	193.9	87.5	50.0	1.55
6	*value*	162	23	922.6	278.7	83.3	50.0	1.50
7	*research*	347	82	1,976.1	993.7	95.8	87.5	1.49
8	*process*	223	46	1,270.0	557.4	95.8	70.8	1.46
9	*effect*	163	27	928.3	327.2	62.5	54.2	1.45
10	*thus*	237	51	1,349.7	618.0	95.8	83.3	1.45
11	*literature*	132	20	751.7	242.4	95.8	58.3	1.41
12	*relation*	459	130	2,614.0	1,575.4	100.0	95.8	1.40
13	*basis/base*	149	30	848.5	363.6	95.8	54.2	1.36
14	*this*	272	73	1,549.0	884.6	95.8	91.7	1.35
15	*p*	181	42	1,030.8	509.0	79.2	54.2	1.35
16	*your/his*	182	44	1,036.5	533.2	95.8	66.7	1.33

Source: Keyword analysis performed with Sketch Engine. FC = focus corpus. RC = reference corpus. RA Freq. (%) = percentage of RAs containing the word. The colored rectangle indicates the selected keyword.

Table B.10 – Transcription of Table 8.11 (Highly distributed common keywords in jaPED against jaMGT).

		General Freq.		Freq. per Million		RA Freq. (%)		
POS	Word	FC	RC	FC	RC	FC	RC	Score
1	*att*	462	233	7,057.6	1,370.4	100.0	100.0	3.40
2	*ari*	226	225	3,452.4	1,323.4	100.0	100.0	1.92
3	*hōhō*	84	39	1,283.2	229.4	100.0	75.0	1.86
4	*zu*	91	58	1,390.1	341.1	87.5	62.5	1.78
5	*nakatt*	147	146	2,245.6	858.7	100.0	100.0	1.75
6	*nomi*	103	81	1,573.5	476.4	75.0	87.5	1.74
7	*kentō shi*	72	35	1,099.9	205.9	95.8	66.7	1.74
8	*okonatt*	141	146	2,154.0	858.7	91.7	91.7	1.70
9	*ta*	2,301	3,613	35,150.7	21,250.3	100.0	100.0	1.63
10	*yori*	218	325	3,330.2	1,911.5	100.0	100.0	1.49
11	*ni kanshite*	55	50	840.2	294.1	70.8	66.7	1.42
12	*de*	2,319	4,194	35,425.7	24,667.5	100.0	100.0	1.42
13	*uchi*	58	58	886.0	341.1	75.0	70.8	1.41
14	*ni taisuru*	79	106	1,206.8	623.5	87.5	79.2	1.36
15	*beki*	52	55	794.4	323.5	91.7	70.8	1.36
16	*oyobi*	81	111	1,237.4	652.9	75.0	83.3	1.35
17	*naku*	64	79	977.7	464.6	83.3	91.7	1.35
18	*okonau*	89	128	1,359.6	752.8	79.2	87.5	1.35

Source: Keyword analysis performed with Sketch Engine. FC = focus corpus. RC = reference corpus. RA Freq. (%) = percentage of RAs containing the word. The colored rectangle indicates the selected keyword. Verbs with partial inflection are double underlined.

Table B.11 – Translation of Table 8.11 (Highly distributed common keywords in jaPED against jaMGT).

POS	Word	General Freq. FC	General Freq. RC	Freq. per Million FC	Freq. per Million RC	RA Freq. (%) FC	RA Freq. (%) RC	Score
1	sentence-ending expression; *there is*	462	233	7,057.6	1,370.4	100.0	100.0	3.40
2	connective; *there is*	226	225	3,452.4	1,323.4	100.0	100.0	1.92
3	*method*	84	39	1,283.2	229.4	100.0	75.0	1.86
4	*figure*	91	58	1,390.1	341.1	87.5	62.5	1.78
5	*absence*; negation (past)	147	146	2,245.6	858.7	100.0	100.0	1.75
6	*only*	103	81	1,573.5	476.4	75.0	87.5	1.74
7	*investigate*	72	35	1,099.9	205.9	95.8	66.7	1.74
8	*perform*	141	146	2,154.0	858.7	91.7	91.7	1.70
9	past tense indicator	2,301	3,613	35,150.7	21,250.3	100.0	100.0	1.63
10	*than*, *from*	218	325	3,330.2	1,911.5	100.0	100.0	1.49
11	*concerning*	55	50	840.2	294.1	70.8	66.7	1.42
12	*at*, *through*	2,319	4,194	35,425.7	24,667.5	100.0	100.0	1.42
13	*among*	58	58	886.0	341.1	75.0	70.8	1.41
14	*for*, *to*, *against*	79	106	1,206.8	623.5	87.5	79.2	1.36
15	*must*	52	55	794.4	323.5	91.7	70.8	1.36
16	*and*	81	111	1,237.4	652.9	75.0	83.3	1.35
17	*absence*; negation	64	79	977.7	464.6	83.3	91.7	1.35
18	*perform*	89	128	1,359.6	752.8	79.2	87.5	1.35

Source: Keyword analysis performed with Sketch Engine. FC = focus corpus. RC = reference corpus. RA Freq. (%) = percentage of RAs containing the word. The colored rectangle indicates the selected keyword. Verbs with partial inflection are double underlined.

Table B.12 – Transcription of Table 8.12 (Highly distributed common keywords in jaMGT against jaPED).

		General Freq.		Freq. per Million		RA Freq. (%)		
POS	Word	FC	RC	FC	RC	FC	RC	Score
1	*iu*	702	60	4,128.9	916.6	100.00	62.50	2.68
2	*yott*	347	26	2,040.9	397.2	100.00	50.00	2.18
3	*kono*	611	90	3,593.7	1,374.9	100.00	83.33	1.93
4	*yō*	639	96	3,758.4	1,466.5	100.00	91.67	1.93
5	*na*	1,978	369	11,633.9	5,636.9	100.00	100.00	1.90
6	*ba*	352	44	2,070.3	672.2	100.00	62.50	1.84
7	*u*	237	25	1,393.9	381.9	95.83	50.00	1.73
8	*ka*	568	106	3,340.8	1,619.3	100.00	75.00	1.66
9	*naru*	537	100	3,158.4	1,527.6	100.00	91.67	1.65
10	*teki*	217	34	1,276.3	519.4	100.00	70.83	1.50
11	*rareru*	404	84	2,376.2	1,283.2	100.00	91.67	1.48
12	*reru*	684	162	4,023.0	2,474.8	100.00	100.00	1.45
13	*kore*	233	45	1,370.4	687.4	95.83	75.00	1.41
14	*da*	138	20	811.7	305.5	95.83	62.50	1.39
15	*koto*	1,427	383	8,393.1	5,850.8	100.00	100.00	1.37

Source: Keyword analysis performed with Sketch Engine. FC = focus corpus. RC = reference corpus. RA Freq. (%) = percentage of RAs containing the word. The colored rectangle indicates the selected keyword.

Table B.13 – Translation of Table 8.12 (Highly distributed common keywords in jaMGT against jaPED).

		General Freq.		Freq. per Million		RA Freq. (%)		
POS	Word	FC	RC	FC	RC	FC	RC	Score
1	*say*, *mean*, *called*	702	60	4,128.9	916.6	100.00	62.50	2.68
2	*depends on*, *due to*, *based on*	347	26	2,040.9	397.2	100.00	50.00	2.18
3	*this* (it must be followed by a noun)	611	90	3,593.7	1,374.9	100.00	83.33	1.93
4	*like*, among other meanings	639	96	3,758.4	1,466.5	100.00	91.67	1.93
5	adjective particle	1,978	369	11,633.9	5,636.9	100.00	100.00	1.90
6	*if*, *when*	352	44	2,070.3	672.2	100.00	62.50	1.84
7	without meaning by itself	237	25	1,393.9	381.9	95.83	50.00	1.73
8	indicates questions and uncertainty, among other functions	568	106	3,340.8	1,619.3	100.00	75.00	1.66
9	*become*	537	100	3,158.4	1,527.6	100.00	91.67	1.65
10	without meaning by itself	217	34	1,276.3	519.4	100.00	70.83	1.50
11	indicates possibility or passive voice	404	84	2,376.2	1,283.2	100.00	91.67	1.48
12	indicates possibility or passive voice	684	162	4,023.0	2,474.8	100.00	100.00	1.45
13	*this*	233	45	1,370.4	687.4	95.83	75.00	1.41
14	affirmative sentence-ending expression	138	20	811.7	305.5	95.83	62.50	1.39
15	*thing*, *the act of*	1,427	383	8,393.1	5,850.8	100.00	100.00	1.37

Source: Keyword analysis performed with Sketch Engine. FC = focus corpus. RC = reference corpus. RA Freq. (%) = percentage of RAs containing the word. The colored rectangle indicates the selected keyword.

Table B.14 – Translation of Table 8.15 (Top 20 strongest collocates of *estudo* [“study,” noun] in two corpora).

	ptPED			ptMGT		
POS	Word	PER	LogDice	Word	PER	LogDice
1	*present*	16.21	12.08	*this*	23.63	12.37
2	*this*	7.97	11.21	*in this*	9.25	11.32
3	*a/an*	15.38	11.08	*present*	6.16	10.71
4	*was*	16.76	11.05	*of this*	5.48	10.60
5	*in the*	16.48	10.98	*our*	4.11	10.21
6	*of this*	6.59	10.97	*performed*	3.77	10.11
7	*transversal*	6.59	10.97	*contributes*	3.42	10.04
8	*in this*	6.59	10.97	*the*	31.51	9.76
9	*the*	27.75	10.81	*limitations*	2.74	9.71
10	*of the*	21.70	10.80	*a/an*	12.67	9.68
11	*in/at*	19.51	10.64	*that*	4.79	9.67
12	*our*	4.95	10.55	*was*	6.85	9.62
13	*done*	4.67	10.48	*for/to*	16.44	9.57
14	*objective*	4.95	10.45	*field*	2.74	9.56
15	*that/which*	17.58	10.43	*about*	5.82	9.48
16	*method*	4.40	10.36	*of the*	17.12	9.40
17	*with*	14.01	10.35	*in the*	9.93	9.39
18	*were*	8.24	10.28	*research*	4.45	9.38
19	*methods*	4.12	10.27	*proposes*	2.05	9.33
20	*performed*	4.12	10.26	*al*	4.79	9.27

Source: Collocates identified with Sketch Engine. PER = number of co-occurrences between collocate and keyword divided by the overall frequency of the keyword in the corpus. Shared collocates are bounded by colored rectangles.

Table B.15 – Translation of Table 8.16 (Top 20 strongest collocates of *pesquisa* ["research," noun] in two corpora).

	ptMGT			ptPED		
POS	Word	PER	LogDice	Word	PER	LogDice
1	*of this*	9.91	11.51	*committee*	27.50	12.82
2	*pre-purchase*	7.58	11.19	*ethics*	22.50	12.48
3	*development*	9.91	11.06	*by the*	25.00	11.64
4	*in this*	6.12	10.86	*beings*	8.75	11.35
5	*this*	6.41	10.83	*humans*	8.75	11.33
6	*course*	6.41	10.69	*university*	7.50	10.99
7	*inspiration*	4.96	10.57	*approved*	5.00	10.61
8	*services*	4.96	10.06	*school*	5.00	10.51
9	*was*	7.29	9.86	*paricipate*	5.00	10.36
10	*results*	5.54	9.81	*present*	7.50	10.23
11	*of the*	19.24	9.75	*of this*	3.75	10.19
12	*construct*	3.21	9.70	*project*	3.75	10.09
13	*how*	9.33	9.51	*recent*	3.75	10.08
14	*model*	4.08	9.51	*it is*	3.75	10.05
15	*the*	22.74	9.50	*this*	3.75	10.03
16	*in/at the*	9.33	9.49	*our*	3.75	9.99
17	*about*	5.25	9.47	*of the*	38.75	9.89
18	*verification*	2.33	9.46	*performed*	3.75	9.86
19	*our*	2.33	9.43	*done*	3.75	9.82
20	*in/at the*	8.16	9.43	*a/an*	16.25	9.72

Source: Collocates identified with Sketch Engine. PER = number of co-occurrences between collocate and keyword divided by the overall frequency of the keyword in the corpus. Shared collocates are bounded by colored rectangles.

Table B.16 – Translation of Table 8.17 (Top 20 strongest [filtered] collocates of 行う [*okonau*, "perform"] in two corpora.

	jaPED			jaMGT		
POS	Word	PER	LogDice	Word	PER	LogDice
1	*koto ga deki* ("is possible to")	8.99	11.23	*bunseki* ("analysis")	14.06	11.03
2	*koto* ("thing")	33.71	11.02	*ishi kettei* ("decision making")	5.47	10.21
3	*kōtō kikan bunri jutsu* ("laryngotracheal separation surgery")	7.87	10.94	*setsumei* ("explanation")	3.91	10.10
4	*hōshin* ("policy")	5.62	10.55	*sekkyokuteki* ("actively")	3.91	10.04
5	*deki* ("can")	5.62	10.50	*kenshō* ("inspection")	3.13	9.84
6	*takiteki shujutsu* ("surgery performed several times")	4.49	10.29	*koto ga dekiru* ("is possible to")	3.91	9.71
7	*o* (indicates the target of the action)	84.27	10.06	*enkatsu* ("smoothly")	2.34	9.47
8	*yōatsu kanki* ("positive pressure ventilation")	3.37	10.05	*hikaku* ("comparison")	2.34	9.37
9	*sekkyokuteki* ("actively")	4.49	10.03	*koto* ("thing")	24.22	9.35
10	*anzen* ("safety")	3.37	9.91	*o* (indicates the target of the action)	96.09	9.32
11	*nyūjiki* ("infancy")	3.37	9.88	*jūyō* ("important")	3.91	9.32
12	*irō zōsetsu* ("gastrostomy")	3.37	9.75	*sai ni* ("when")	2.34	9.31
13	*ka* (indicates question and uncertainty, among other functions)	5.62	9.71	*ue* ("above")	2.34	9.31
14	*jūbun* ("enough")	3.37	9.65	*shōten* ("focus")	2.34	9.14
15	*taikiteki* ("on standby")	2.25	9.49	*soshiki henkaku* ("organizational transformation")	3.91	9.09
16	*honyū* ("breast-feeding")	2.25	9.49	*beki* ("must")	2.34	9.07
17	*na* (adjective particle)	11.24	9.48	*kentō* ("investigation")	2.34	9.05
18	*tai'in go* ("after discharge")	2.25	9.48	*iu* ("say")	10.16	9.00
19	*jinkō kokyūki* ("mechanical ventilator")	2.25	9.48	*yori* ("than," "from")	5.47	8.98
20	*gensoku* ("principle")	2.25	9.46	*kuwae* ("add")	2.34	8.95

Source: Collocates identified with Sketch Engine and filtered by the author. PER = number of co-occurrences between collocate and keyword divided by the overall frequency of the keyword in the corpus. Shared collocates are bounded by colored rectangles. Verbs with partial inflection are double underlined.

Table B.17 – Translation of Table 8.18 (Top 20 strongest [filtered] collocates of この [*kono*, “this”] in two corpora).

	jaMGT			jaPED		
POS	Word	PER	LogDice	Word	PER	LogDice
1	*yō* (“like”)	22.42	11.81	*yō* (“like”)	17.78	11.46
2	*ten* (“point”)	6.87	10.76	*ronbun* (“paper”)	4.44	10.43
3	*wa* (indicates the topic, among other functions)	43.54	10.48	*shinkoku su* (“declare”)	4.44	10.39
4	*na* (adjective particle)	16.86	10.35	*hōhō* (“method”)	6.67	10.14
5	*tame* (“for,” “because”)	4.75	9.84	*uchi* (“among”)	5.56	10.11
6	*naru* (“become”)	4.58	9.64	*nao* (“further”)	3.33	9.96
7	*ga* (indicates the subject, among other functions)	23.40	9.64	*koto* (“thing”)	15.56	9.92
8	*koto* (“thing”)	8.02	9.62	*tame* (“for,” “because”)	7.78	9.82
9	*o* (indicates the target of an action)	26.51	9.61	*na* (adjective particle)	13.33	9.74
10	*ronri* (“logic”)	2.62	9.61	*baai* (“case”)	5.56	9.71
11	*to* (“and,” “when”)	18.99	9.57	*shōrei* (“clinical case”)	11.11	9.70
12	*kara* (“from”)	5.40	9.54	*mono* (“object,” “thing”)	4.44	9.65
13	*honkō* (“this paper”)	3.44	9.54	*uchi/naka* (“among”)	3.33	9.61
14	*mo* (“also,” “too”)	7.20	9.51	*go rei* (“five cases”)	3.33	9.53
15	*no* (“of”)	42.39	9.44	*eru* (“obtain”)	2.22	9.46
16	*kekka* (“result”)	2.95	9.44	*riten* (“advantage”)	2.22	9.45
17	*soshiki* (“organization”)	3.27	9.41	*naru* (“become”)	4.44	9.43
18	*jigyō* (“business”)	2.13	9.33	*nomi* (“only”)	4.44	9.41
19	*purojekuto* (“project”)	2.29	9.27	*koto ga dekiru* (“is possible to”)	2.22	9.40
20	*shikashi* (“however”)	2.29	9.26	*yosō sa* (“be expected”)	2.22	9.39

Source: Collocates identified with Sketch Engine and filtered by the author. PER = number of co-occurrences between collocate and keyword divided by the overall frequency of the keyword in the corpus. Shared collocates are bounded by colored rectangles. Verbs with partial inflection are double underlined.

Table B.18 – Translation of Table 8.21 (Semantic sets of *estudo* [noun, "study"] in two corpora).

Corpus	SET (Collocates)
ptPED	ACTION AS AGENT (*analyzed*, *assessed*, *showed*) ACTION AS TARGET (*approved*, *done*, *performed*) ARTICLE CONTENT (*discussion*, *limitations*, *method*, *methods*, *objective*, *results*) PEDIATRICS (*adolescents*, *years*, *PCCs*, *children*, *age*, *nurslings*, *mothers*, *ASD*) RESEARCH METHODOLOGY AND STATISTICS (*approach*, *findings*, *sample*, *association*, *committee*, *data*, *design*, *participants*, *participate*, *participated*, *period*, *prevalence*, *variables*) THE PRESENT STUDY (*of this*, *this*, *in this*, *our*, *present*) TYPE OF STUDY (*descriptive*, *multicenter*, *observational*, *pilot*, *review*, *transversal*)
ptMGT	ACTION AS AGENT (*considers*, *contributes*, *identified*, *shows*, *proposes*, *reveals*, *suggests*, *employed*) ARTICLE CONTENT (*conclusion*, *conclusions*, *contributions*, *discussion*, *implications*, *limitations*, *method*, *methodology*, *objective*, *result*, *results*, *abstract*) MANAGEMENT (*behavior*, *consumers, industry*) RESEARCH AIMS (*comprehend*, *understand*, *identify*) RESEARCH METHODOLOGY AND STATISTICS (*approach*, *samples*, *field*, *data*, *design*, *effect*, *pieces of evidence*, *factors*, *hypotheses*, *model*, *nature*, *object*, *participants*, *research*, *researchers*, *respondents*, *theory*, *variables*) THE PRESENT STUDY (*of this*, *this*, *in this*, *our*, *present*) TYPE OF STUDY (*case*, *events*, *pilot*)

Source: Semantic sets manually compiled from collocate lists generated by Sketch Engine. Shared sets and collocates are indicated with colored boxes. PCC = psychosocial care center. ASD = autism spectrum disorder.

Table B.19 – Translation of Table 8.22 (Semantic sets of *pesquisa* [noun, "research"] in two corpora).

Corpus	SET (Collocates)
ptMGT	ARTICLE CONTENT (*contribution*, *hypothesis*, *hypotheses*, *implications*, *methodology*, *objective*, *question*, *result*, *results*) MANAGEMENT (*consumer*, *consumers*, *managers*, *services*) RESEARCH METHODOLOGY AND STATISTICS (*approach*, *sample*, *categories*, *construct*, *context*, *data*, *study*, *method*, *model*, *answer* [v.], *subjects*, *theoretical*, *theoretical*, *variables*) THE PRESENT RESEARCH (*of this*, *this*, *in this*, *our*, *present*) TYPE OF RESEARCH (*qualitative*, *pre-purchase*, *bibliographic*, *course*, *field*)
ptPED	ACTION AS TARGET (*done*, *performed*) ETHICS IN RESEARCH (*approved*, *approved*, *approved*, *sign*, *REC*, *committee*, *involving*, *ethics*, *humans*, *beings*) PEDIATRICS (*breast-feeding*, *children*, *nurses*, *hospital*, *mothers*, *maternal*, *medicine*, *health*) RESEARCH INSTITUTIONS (*center*, *school*, *institution*, *university*) RESEARCH METHODOLOGY AND STATISTICS (*accepted*, *data*, *study*, *pieces of information*, *participate*, *prevalence*, *project*, *qualitative*, *results*, *transversal*) SCOPE (*Brazil*, *national*) THE PRESENT RESEARCH (*of this*, *this*, *our*, *present*)

Source: Semantic sets manually compiled from collocate lists generated by Sketch Engine. Shared sets and collocates are indicated with colored boxes. REC = research ethics committee.

Table B.20 – Translation of Table 8.23 (Semantic sets of 行う [*okonau*, "carry out"] in two corpora).

Corpus	SET (Collocates)
jaPED	FEASIBILITY (*kanō*, "possible"; *konnan*, "troublesome") NEED FOR THE ACTION (*beki*, "must"; *hitsuyō*, "need") MODE (*anzen*, "safefy"; *taikiteki*, "on standby"; *sekkyokuteki*, "actively") PEDIATRICS (*nyūjiki*, "infancy"; *jinkō kokyūki*, "mechanical ventilator"; *hozonteki karyō*, "conservative treatment"; *honyū*, "breast-feeding"; *kōtō kikan bunri jutsu*, "laryngotracheal separation surgery"; *takiteki shujutsu*, "surgery performed several times"; *shujutsu*, "surgery"; *shujutsuchū*, "during the surgery"; *hōshin*, "policy"; *shisetsu*, "facilities"; *eiyō*, "nutrition"; *konji shujutsu*, "radical surgery"; *konjijutsu*, "radical operation"; *chiryō*, "treatment"; *kanri*, "management," "control"; *irō zōsetsu*, "gastrostomy"; *jutsushiki*, "technique"; *kansatsu*, "observation"; *hyōka*, "evaluation"; *fuka shiken*, "load test"; *tai'ingo*, "after discharge"; *harisaki*, "needle tip"; *donteki hakuri*, "blunt exfoliation," "blunt peeling"; *yōatsu kanki*, "positive pressure ventilation")
jaMGT	DEMONSTRATIVE FAMILY (*kono*, "this"; *sono*, "that") NEED FOR THE ACTION (*beki*, "must"; *hitsuyō*, "need"; *hitsuyō fukaketsu*, "absolutely essential," "critical"; *jūyō*, "important") MODE (*enkatsu*, "smoothly"; *sekkyokyokuteki*, "actively"; *keikakuteki*, "planned") MANAGEMENT (*ishikettei*, "decision making"; *chīmu*, "team"; *purojekuto*, "project"; *menbā*, "member"; *jigyō tenkai*, "business development"; *kōshō*, "negotiation"; *shigoto*, "job"; *tabumon*, "other departments"; *kigyō*, "business"; *kigyōkateki shikō*, "entrepreneurial intention"; *mondai kaiketsu*, "problem solving"; *henkaku*, "transformation"; *tōshi*, "investment"; *kaigai shikaisha*, "overseas subsidiary"; *soshiki*, "organization"; *soshiki henkaku*, "organizational transformation"; *kōdō*, "behavior"; *chōsei*, "adjustment"; *giron*, "debate," "dispute"; *chingin kanri*, "wage management") RESEARCH METHODOLOGY AND WRITING (*fsQCA*; *kasetsu*, "hypothesis"; *bunseki*, "analysis"; *jisshō bunseki*, "empirical analysis"; *jisshō kenkyū*, "empirical research"; *taishō*, "subject"; *honkō*, "this paper"; *kentō*, "investigation"; *kenshō*, "inspection"; *kōsei gainen*, "construct"; *hikaku*, "comparison"; *shōten*, "focus"; *mochii*, "use"; *kangae*, "thought," "think"; *setsumei*, "explanation"; *kadai*, "challenge")

Source: Semantic sets manually compiled from collocate lists generated by Sketch Engine. Shared sets and collocates are indicated with colored boxes. The double underline indicates a verb with partial inflection. fsQCA = fuzzy-set qualitative comparative analysis.

Table B.21 – Translation of Table 8.24 (Semantic sets of この [*kono*, “this”] in two corpora).

Corpus	SET (Collocates)
jaMGT	TWO THINGS (*futatsu*, “two”; *futatsu*, “two”) MANAGEMENT (*osumitsuki*, “authorization,” “certificate”; *purojekuto*, “project”; *jigyō*, “business”; *kigyōtekika shikō*, “entrepreneaurial intention”; *torikumi*, “attempt”; *henkaku*, “transformation”; *ishi kettei*, “decision making”; *gijutsu*, “technology”; *shinki jigyō*, “new business”; *jōkyō*, “situation”; *soshiki*, “organization”; *haikei*, “background”; *seihin*, “product”) RESEARCH METHODOLOGY AND WRITING (*apurōchi*, “approach”; *moderu*, “model”; *jirei*, “case”; *senkō kenkyū*, “previous research”; *bunrui*, “classification”; *zu*, “figure”; *shiteki shi*, “point out”; *torae*, “capture”; *honkō*, “this paper”; *gainen*, “concept”; *ten*, “point”; *riyū*, “reason”; *mochii*, “use”; *kenkyū*, “research”; *kenkyūgun*, “research group”; *kekka*, “result”; *kangae*, “thought,” “think”; *kangaekata*, “way of thinking”; *ronji*, “discuss”; *ronri*, “logic”; *fumae*, “based on”)
jaPED	AMONG (*uchi*, “among”; *uchi*/*naka*, “among”) MEDICAL CASE NUMBERS (*ichi rei*, “one case”; *ni rei*, “two cases”; *go rei*, “five cases”) PEDIATRICS (*teishussei taijū ji*, “low birth weight infant”; *hōhō*, “method”; *sōki*, “early”; *shōrei*, “clinical case”; *keiken shi*, “experience” (v.); *irō*, “gastrostoma”; *fukumakugai hakuri sōsa*, “extraperitoneal separation operation”) RESEARCH METHODOLOGY AND WRITING (*zu*, “figure”; *hyō*, “table”; *ronbun*, “paper”; *shinkoku su*, “report,” “declare”; *yosō sa*, “be expected”; *kangae*, “thought,” “think”) THE PRESENT RESEARCH AND AUTHORS (*konkai*, “this time”; *wareware*, “we”) TIME EXPRESSIONS (*izen*, “before”; *toki*, “time”; *jiten*, “point of time”)

Source: Semantic sets manually compiled from collocate lists generated by Sketch Engine. Shared sets and collocates are indicated with colored boxes. Verbs with partial inflection are double underlined.

APPENDIX C – ENGLISH TRANSLATIONS OF CHAPTER 9 TABLES AND FIGURES

This appendix presents English translations of tables and figures from Chapter 9 that have Portuguese and Japanese words. Table C.1 below shows the correspondence between the original and the translated tables and figures. Romanized transcriptions are provided together with the English translations in the two tables and in all figures that contain Japanese words.

Table C.1 – Correspondence between original and translated tables.

Original Table or Figure	Translated Table or Figure
Table 9.1	Table C.2
Table 9.2	Table C.3
Figure 9.26	Figure C.1
Figure 9.27	Figure C.2
Figure 9.28	Figure C.3
Figure 9.29	Figure C.4
Figure 9.30	Figure C.5
Figure 9.31	Figure C.6
Figure 9.35	Figure C.7
Figure 9.36	Figure C.8
Figure 9.37	Figure C.9
Figure 9.38	Figure C.10
Figure 9.39	Figure C.11
Figure 9.40	Figure C.12

Source: Prepared by the author.

Table C.2 – Translation of Table 9.1 (Numbers of instances of Pediatrics key keywords calculated by two software programs).

Keyness Position	enPED			ptPED			jaPED		
	Word	SE	AC	Word	SE	AC	Word	SE	AC
1	*infants*	522	534	*mothers*	116	118	*shōrei* ("clinical case")	305	504
2	*adolescents*	372	382	*adolescents*	106	114	*honshō* ("this disease")	116	143
3	*mortality*	297	297	*maternal*	96	99	*shikō shi* ("carry out")	113	0
4	*infant*	183	219	*feeding*	78	79	*chiryō* ("treatment")	81	398
5	*prevention*	81	84	*mother*	52	54	*shujutsu* ("surgery")	81	316
6	*parents*	534	575	*95% ci*	50	0	*jutsugo* ("postoperative")	63	130
7	*diseases*	61	61	*children*	321	331	*jikenrei* ("medical case study")	54	50
8	*ci*	339	339	*early*	28	29	*shōjō* ("a symptom")	53	91
9	–	–	–	*municipality*	28	29	*tōka* ("this department")	49	49
10	–	–	–	*infant*	27	28	*tōin* ("this clinic")	47	59
11	–	–	–	*among*	23	23	*zenrei* ("all instances")	46	85
12	–	–	–	*healthy*	22	24	*kanji* ("child patient")	42	46
13	–	–	–	*boys*	22	22	*hasshō shi* ("appearance of symptoms")	42	0
14	–	–	–	*committee*	20	20	*jutsushiki* ("operative method")	35	59
15	–	–	–	*childhood*	20	23	*gappeishō* ("complications")	35	0
16	–	–	–	*presented*	18	19	*shōni* ("an infant")	34	271
17	–	–	–	*clinical*	18	19	*danji* ("a boy")	32	44
18	–	–	–	–	–	–	*Honpō* ("Japan")	31	32
19	–	–	–	–	–	–	*kōhōshiteki* ("retrospective")	30	0
20	–	–	–	–	–	–	*yogo* ("prognosis")	29	67
21	–	–	–	–	–	–	*shindan* ("diagnosis")	27	145
22	–	–	–	–	–	–	*seigo* ("after birth")	25	40
23	–	–	–	–	–	–	*III kekka* ("III results")	24	0
24	–	–	–	–	–	–	*shindan sa* ("diagnose")	24	0
25	–	–	–	–	–	–	*II taishō* ("II subject")	24	0
26	–	–	–	–	–	–	*IV kōsatsu* ("IV discussion")	24	0
27	–	–	–	–	–	–	*yōshi* ("summary")	24	13
28	–	–	–	–	–	–	*chiryō hōshin* ("treatment policy")	24	0
29	–	–	–	–	–	–	*shinsei jiki* ("neonatal period")	23	0
30	–	–	–	–	–	–	*hasshō* ("appearance of symptoms")	23	166
31	–	–	–	–	–	–	*hōkoku suru* ("report")	20	0
32	–	–	–	–	–	–	*shikkan* ("disease")	20	81
33	–	–	–	–	–	–	*shikō suru* ("carry out")	17	0
34	–	–	–	–	–	–	*ji* ("child")	15	77

Source: Numbers calculated by Sketch Engine (SK) and AntConc (AC). Words whose difference between the totals is above 10 are indicated by colored boxes. Japanese verbs with partial inflection are double underlined.

Table C.3 – Translation of Table 9.2 (Numbers of instances of Management key keywords calculated by two software programs).

Keyness Position	enPED			ptPED			jaPED		
	Word	SE	AC	Word	SE	AC	Word	SE	AC
1	*business*	902	964	*organizational*	213	214	*kigyō* ("business")	376	1043
2	*organizational*	898	937	*consumers*	211	211	*honkō* ("this paper")	315	316
3	*firms*	758	867	*company*	147	149	*kokyaku* ("customer")	232	348
4	*innovation*	583	592	*attitude*	112	112	*menbā* ("member")	209	290
5	*company*	389	480	*consumer*	74	83	*shigen* ("resource")	169	356
6	*organisational*	327	332	*clients*	72	72	*senkō kenkyū* ("previous reserach")	144	0
7	*employee*	310	355	*integration*	70	70	*jirei* ("case")	134	217
8	*customer*	297	335	*financial*	58	58	*jūgyōin* ("employee")	128	15
9	*customers*	277	314	*financial*	56	56	*honkenkyū* ("this study")	124	116
10	*competitive*	196	197	*perceived*	53	53	*moderu* ("model")	106	234
11	*corporate*	195	200	*organizational*	52	55	*inobēshon* ("innovation")	93	116
12	*theories*	124	124	*efficiency*	45	45	*kōdō* ("behavior")	91	208
13	*leadership*	523	528	*influences*	44	44	*kizon kenkyū* ("existing research")	78	0
14	*markets*	119	120	*marketing*	43	43	*hensū* ("variable")	77	174
15	*scholars*	118	120	*managerial*	41	41	*senryaku* ("strategy")	76	312
16	*marketing*	107	107	*flow*	37	37	*no dearu* ("of")	75	0
17	*managerial*	101	101	*equations*	33	33	*o tsūjite* ("through")	69	0
18	*employees*	706	868	*author*	27	27	*takameru* ("increase")	67	67
19	*industry*	375	397	–	–	–	*shijō* ("market")	67	229
20	*identity*	580	588	–	–	–	*ishi kettei* ("decision making")	65	0
21	*competitors*	74	76	–	–	–	*gurūpu* ("group")	57	228
22	*strategic*	544	545	–	–	–	*jigyō* ("industry")	52	291
23	–	–	–	–	–	–	*purosesu* ("process")	48	157
24	–	–	–	–	–	–	*bunseki kekka* ("analysis results")	47	0
25	–	–	–	–	–	–	*bunseki shi* ("analyze")	47	0
26	–	–	–	–	–	–	*ronji* ("discuss")	47	47
27	–	–	–	–	–	–	*nihon kigyō* ("Japanese company")	47	2
28	–	–	–	–	–	–	*shōten* ("focus")	46	47
29	–	–	–	–	–	–	*torae* ("capture")	46	53
30	–	–	–	–	–	–	*shō/masa* ("certain," "correct")	45	50
31	–	–	–	–	–	–	*sōtei shi* ("assume")	42	0
32	–	–	–	–	–	–	*futatsu* ("two")	42	41
33	–	–	–	–	–	–	*naze* ("why")	42	41
34	–	–	–	–	–	–	*dai ni* ("second")	42	50
35	–	–	–	–	–	–	*atarashii* ("new")	42	41
36	–	–	–	–	–	–	*kojin* ("individual")	42	65
37	–	–	–	–	–	–	*sō* ("like that")	41	20
38	–	–	–	–	–	–	*dai ichi* ("first")	41	54
39	–	–	–	–	–	–	*chakumoku shi* ("focus on")	39	0
40	–	–	–	–	–	–	*takame* ("increase")	38	35
41	–	–	–	–	–	–	*shoshiki nai* ("within the organization")	37	2
42	–	–	–	–	–	–	*shuchō* ("claim")	37	64

Source: Numbers calculated by Sketch Engine (SK) and AntConc (AC). Words whose difference between the totals is above 10 are indicated by colored boxes. Japanese verbs with partial inflection are double underlined.

Figure C.1 – Translation of Figure 9.26 (Examples of exclusive and shared clusters with *crianças* [“children”]).

CLUSTER	ptPED(Intro)	ptPED(Me)	ptPED(Res)	ptPED(D+C)
children there are	•			
children diagnosed		•		
children showed			•	
hospitalized children				•
some children	•			•
treated children	•			•
children with	•	•	•	•
children under risk	•			
following up children		•		
half of the children			•	
children who had				•
anemia in children	•	•	•	•
children and adolescents	•	•		•
children below	•	•	•	•
children with less than	•			
children with intellectual disability		•		
children with an average age			•	
low-weight children				•
children below five	•			•
food guide for children	•	•	•	•

Source: Designed by the author. Clusters identified with AntConc.

Figure C.2 – Translation of Figure 9.27 (Examples of exclusive and shared clusters with *mães* [“mothers”]).

CLUSTER	ptPED(Intro)	ptPED(Me)	ptPED(Res)	ptPED(D+C)
mothers are	•			
mothers participated		•		
mothers had			•	
mothers would like				•
the mothers	•	•	•	•
of the mothers	•	•	•	•
by mothers	•	•		•
mothers and pregnant women	•			
mothers with damage		•		
mothers’ expectations			•	
mothers whose pregnancy				•
mothers’ awareness		•	•	•
mothers and caregivers	•			•
majority of the mothers			•	•
healthy food for mothers	•			
was composed of mothers		•		
mothers were included			•	
mothers’ motivation				•
the majority of the mothers			•	•
mothers and to the caregivers	•			•

Source: Designed by the author. Clusters identified with AntConc.

Figure C.3 – Translation of Figure 9.28 (Examples of exclusive and shared clusters with *adolescentes* ["adolescents"]).

CLUSTER	ptPED(Intro)	ptPED(Me)	ptPED(Res)	ptPED(D+C)
Brazilian female adolescents	•			
involve adolescents		•		
active adolescents			•	
adolescents had				•
athlete adolescents	•	•	•	•
Brazilian adolescents	•	•		•
adolescent practitioners			•	•
suggest that adolescents	•			
composed of adolescents		•		
characteristics of the adolescents			•	
adolescents with depression				•
amateur athlete adolescents	•	•	•	•
adolescents of both	•	•		
prevalence of adolescents			•	•
adolescents in situation of	•			
adolescents with less than		•		
adolescents of higher level			•	
high prevalence of adolescents				•
adolescents who practice arts			•	•
final sample of adolescents		•	•	

Source: Designed by the author. Clusters identified with AntConc.

Figure C.4 – Translation of Figure 9.29 (Examples of exclusive and shared clusters with 症例 [*shōrei*, “clinical case”]).

CLUSTER	jaPED(Intro)	jaPED(Me)	jaPED(Res)	jaPED(D+C)
shōrei ni yotte (“through the case,” “based on the case”)	•			
shōrei oyobi (“case and”)		•		
zen shōrei (“all cases”)			•	
takai shōrei (“high . . . case”)				•
aru shōrei (“case with . . .”)		•	•	•
kono shōrei (“this case”)			•	•
no shōrei (“case of”)	•	•	•	•
shōrei mo hōkoku (“also report a case . . .”)	•			
shōrei o chiryō (“treatment . . . case . . .”)		•		
shōrei no gaiyō (“outline of the case”)			•	
shōrei no hōkoku (“case report”)				•
de no shōrei (“case with/by”)	•	•		•
shōrei ga hōkoku (“report . . . case”)	•			•
shōrei ni taishite (“in respect to the case”)	•	•	•	•
shōrei no naka ni (“among the cases”)	•			
shōrei o taishō to (“case . . . as target”)		•		
shōrei no shōsai o (“details of the case . . .”)			•	
shōrei ni kurabete (“compared to . . . case”)				•
sono yō na shōrei (“case like this”)		•	•	•
shikō shita shōrei (“case in which . . . was performed”)	•	•	•	•

Source: Designed by the author. Clusters identified with AntConc.

Figure C.5 – Translation of Figure 9.30 (Examples of exclusive and shared clusters with 本症 [*honshō*, "this disease"]).

CLUSTER	jaPED(Intro)	jaPED(Me)	jaPED(Res)	jaPED(D+C)
honshō ni tsuite ("about this disease")	•			
honshō e ("toward this disease")		•		
rei honshō ("example this disease")			•	
honshō ga ("this disease . . . ")				•
ga honshō (". . . this disease")	•			•
ni honshō (". . . this disease")	•	•	•	•
honshō to ("this disease and")	•	•	•	•
honshō to wa ("this disease is")	•			
honshō e no (". . . against this disease")		•		
honshō ni yoru shibō ("death due to this disease")			•	
honshō de wa ("in this disease")				•
honshō to shindan ("this disease and diagnosis")	•	•		•
honshō no tokuchō ("features of this disease")	•			•
honshō no hassei ("appearance of this disease")	•			•
made ni mo honshō ("until [now] this disease")	•			
honshō no risuko o (". . . the risk of this disease")		•		
kono jiten de honshō ("at this point, this disease")			•	
honshō no tokuchō ("features of this disease . . .")				•
nomi de wa honshō ("only . . . this disease")		•		•
honshō o hasshō shi ("start to develop this disease")	•		•	•

Source: Designed by the author. Clusters identified with AntConc.

Figure C.6 – Translation of Figure 9.31 (Examples of exclusive and shared clusters with 手術 [*shujutsu*, “surgery”]).

CLUSTER	jaPED(Intro)	jaPED(Me)	jaPED(Res)	jaPED(D+C)
shujutsu jitai (“surgery itself”)	•			
shujutsu chiryō (“surgical treatment”)		•		
shujutsu ikō (“after surgery”)			•	
shinzō shujutsu (“heart surgery”)				•
no shujutsu (“surgery of”)	•	•	•	•
shokai shujutsu (“first-time surgery”)		•	•	•
shujutsu no (“surgery’s”)	•	•		•
shujutsu ga dai ichi (“surgery is [the] first”)	•			
shujutsu no tejun (“surgical procedures”)		•		
shujutsu o kibō (“wish surgery”)			•	
shujutsu ji ni (“at the time of surgery”)				•
nakatta shujutsu (“surgery without . . . “)			•	•
shujutsu o okonau (“perform surgery”)	•	•		•
shujutsu ryōhō o (“. . . surgical treatment”)		•	•	•
shujutsu ga dai ichi sentaku (“surgery is the first choice”)	•			
shujutsu o dōji ni (“surgery . . . at the same time”)		•		
shujutsu ji no heikin (“average at the time of surgery”)			•	
shujutsu o uketa (“had surgery”)				•
o mitometa shujutsu (“surgery that provided [a diagnosis]”)			•	•
shujutsu jikan shukketsu ryō (“bleeding amount during surgery”)		•	•	•

Source: Designed by the author. Clusters identified with AntConc.

Figure C.7 – Translation of Figure 9.35 (Examples of exclusive and shared clusters with *organizacional* [“organizational”]).

CLUSTER	ptMGT(Op)	ptMGT(Mi)	ptMGT(Cl)
organizational management	•		
organizational impact		•	
organizational effectiveness			•
organizational support	•	•	•
organizational commitment	•	•	•
organizational change		•	•
of the organizational culture	•		
of the organizational commitment		•	
the organizational reality			•
low organizational support	•	•	
at the organizational level	•	•	•
the organizatinal performance	•	•	•
organizational of industrial companies	•		
scale of organizational support		•	
performance at the organizational level			•
attributes at the organizational level		•	•
organizational citizenship behavior		•	•
organizational citizenship behavior patterns	•	•	•
working and organizational commitment	•	•	•
organizational support perception		•	•

Source: Designed by the author. Clusters identified with AntConc.

Figure C.8 – Translation of Figure 9.36 (Examples of exclusive and shared clusters with *consumidores* [“consumers”]).

CLUSTER	ptMGT(Op)	ptMGT(Mi)	ptMGT(Cl)
consumers are	•		
consumers immune (to)		•	
consumers can			•
to the consumers	•	•	
how/as consumers	•	•	
consumers of	•	•	•
by the consumers	•	•	•
study of consumers	•		
consumers' minds		•	
in some consumers			•
consumers' aspirations	•	•	•
consumers' behavior	•	•	•
younger consumers		•	•
consumers' preference	•	•	•
consumers are willing to	•		
consumers to pay a/an		•	
collective action of consumers			•
from the viewpoint of the consumers	•	•	
the aspirations of the consumers	•	•	•
and consumers' behavior	•	•	•

Source: Designed by the author. Clusters identified with AntConc.

Figure C.9 – Translation of Figure 9.37 (Examples of exclusive and shared clusters with *empresa* [“company”]).

CLUSTER	ptMGT(Op)	ptMGT(Mi)	ptMGT(Cl)
company at the	•		
company is		•	
company with			•
the company		•	•
each company		•	•
a company	•	•	•
retail company	•		
computer company		•	
company’s situation			•
with the company		•	•
of a company	•	•	•
socially responsible company		•	•
to the company’s executives	•		
company in the promotion of the		•	
manager about the company			•
consumers see a company		•	•
of investing in the company	•	•	
of the products of the company		•	•
in relation to the company	•	•	
a company’s shops	•	•	

Source: Designed by the author. Clusters identified with AntConc.

Figure C.10 – Translation of Figure 9.38 (Examples of exclusive and shared clusters with 企業 [*kigyō*, “business”]).

CLUSTER	jaMGT(Op)	jaMGT(Mi)	jaMGT(Cl)
chūshō kigyō (“small and medium-sized enterprises”)	•		
kigyō sha (“entrepreneur”)		•	
chūgoku kigyō (“Chinese corporations”)			•
ga kigyō (“. . . company”)	•	•	•
na kigyō (“[adjective] business”)	•	•	•
no kigyō (“business of”)	•	•	•
dekinai kigyō (“companies that cannot”)	•		
kigyō no kachi (“business value”)		•	
tokutei no kigyō (“specific company”)			•
sunawachi kigyō (“that is, business”)		•	•
de wa kigyō (“at . . . business”)	•	•	•
wa nihon kigyō (“. . . Japanese corporations”)	•	•	•
kigyō ga keizai no (“business . . . of economy”)	•		
kigyō to no katsudō (“actions with the company . . .”)		•	
no yō na kigyō (“businesses like . . .”)			•
kō shita kigyō (“companies like . . .”)	•	•	
ni yotte ta no kigyō (“due to . . . other companies”)	•	•	
no chomei na kigyō (“famous company of”)	•	•	•
kigyō ni oite (“in the company”)	•	•	•
kigyō o taishō to (“[having a/that] company as target”)	•	•	•

Source: Designed by the author. Clusters identified with AntConc.

Figure C.11 – Translation of Figure 9.39 (Examples of exclusive and shared clusters with 本稿 [*honkō*, "this paper"]).

CLUSTER	jaMGT(Op)	jaMGT(Mi)	jaMGT(Cl)
yori honkō ("based on . . . this paper")	•		
oyobi honkō ("and this paper")		•	
tadashi honkō ("but this paper")			•
sara ni honkō ("in addition, this paper")		•	•
shikashi honkō (however, this paper")	•	•	
mata honkō ("also, this paper")	•	•	•
honkō no mokuteki ("this paper's aim")	•		
honkō ga chūmoku ("this paper focuses")		•	
honkō to kotonaru ("differently from this paper")			•
soko de honkō ("then, this paper")	•	•	
honkō de wa ("in this paper")	•	•	•
honkō no bunseki ("this paper's analysis")	•	•	•
honkō de bunseki suru ("in this paper . . . analyze")	•		
honkō de kono ronri ("in this paper, this logic")		•	
honkō no omo na ("this paper's main")			•
honkō de wa kono ("in this paper, this")	•	•	
honkō de wa ijō ("in this paper . . . above")	•	•	•
honkō de akiraka ni ("in this paper . . . clear")		•	•
honkō no bunseki kekka ("this paper's analysis results")		•	•
honkō no kadai to ("this paper's challenge")		•	•

Source: Designed by the author. Clusters identified with AntConc.

Figure C.12 – Translation of Figure 9.40 (Examples of exclusive and shared clusters with 顧客 [*kokyaku*, "customer"]).

CLUSTER	jaMGT(Op)	jaMGT(Mi)	jaMGT(Cl)
tokutei kokyaku ("specific customer")	•		
kokyaku ni totte ("for the customer")		•	
motomeru kokyaku ("customer seeking")			•
samazama na kokyaku ("many customers")		•	•
to kokyaku ("and customers")	•	•	•
no kokyaku ("customers of")	•	•	•
onaji kokyaku ("the same customer")		•	•
kokyaku no hyōka ("customer evaluation")	•		
mokuteki wa kokyaku ("goal is . . . customers")		•	
honkenkyū wa kokyaku ("this study . . . customers")			•
koto kara kokyaku ("from this . . . customers")		•	•
tayō na kokyaku ("various customers")		•	•
kokyaku to no (". . . with/of customers")	•	•	•
kokyaku no koe ("customer opinion")		•	•
kokyaku e no izon ("customer dependency")	•		
kokyaku to no kankei ("relationship with customers")		•	
honkenkyū de wa kokyaku ("in this study . . . customers")			•
no yō ni kokyaku ("this way . . . customers")		•	•
kokyaku to shigen no ("of customers and resources")	•	•	•
kokyaku no han'i wa ("the range of customers . . .")	•	•	•

Source: Designed by the author. Clusters identified with AntConc.

APPENDIX D – ENGLISH TRANSLATIONS OF CHAPTER 10 TABLES

This appendix presents English translations of tables from Chapter 10 that have Portuguese and Japanese words. Table D.1 below shows the correspondence between the original and the translated tables. Romanized transcriptions are provided for Japanese words in all the translated tables.

Table D.1 – Correspondence between original and translated tables.

Original Table	Translated Table
Table 10.1	Table D.2
Table 10.2	Table D.3
Table 10.7	Table D.4
Table 10.8	Table D.5
Table 10.9	Table D.6
Table 10.10	Table D.7
Table 10.11	Table D.8
Table 10.12	Table D.9

Source: Prepared by the author.

Table D.2 – Translation of Table 10.1 (Top 50 most frequent nouns in three Pediatrics corpora).

	enPED			ptPED			jaPED		
POS	Word	Freq.	Norm. Freq.	Word	Freq.	Norm. Freq.	Word	Freq.	Norm. Freq.
1	*study*	1,461	4,865.2	*study*	364	4,411.1	*shōrei* (“clinical case”)	305	4,659.3
2	*children*	1,222	4,069.3	*children*	321	3,890.0	*kangae* (“thought”)	151	2,306.7
3	*health*	773	2,574.1	*health*	211	2,557.0	*hitsuyō* (“need”)	120	1,833.2
4	*care*	732	2,437.6	*age*	197	2,387.3	*honshō* (“this disease”)	116	1,772.0
5	*data*	713	2,374.3	*years*	179	2,169.2	*baai* (“case”)	106	1,619.3
6	*age*	708	2,357.7	*weight*	171	2,072.3	*hōkoku* (“report”)	100	1,527.6
7	*p*	607	2,021.3	*being (n.)/be*	166	2,011.7	*zu* (“figure”)	91	1,390.1
8	*child*	585	1,948.1	*data*	138	1,672.3	*josei ishi* (“female doctor”)	86	1,313.8
9	*parents*	534	1,778.2	*relation*	130	1,575.4	*hōhō* (“method”)	84	1,283.2
10	*group*	532	1,771.6	*variables*	124	1,502.7	*shisetsu* (“facilities”)	84	1,283.2
11	*infants*	522	1,738.3	*mothers*	116	1,405.7	*chiryō* (“treatment”)	81	1,237.4
12	*results*	481	1,601.7	*analysis*	112	1,357.3	*shujutsu* (“surgery”)	81	1,237.4
13	*studies*	470	1,565.1	*results*	112	1,357.3	*kekka* (“result”)	74	1,130.4
14	*months*	447	1,488.5	*months*	107	1,296.7	*mono* (“thing”)	73	1,115.2
15	*risk*	445	1,481.9	*studies*	106	1,284.6	*hozonteki chiryō* (“conservative treatment”)	64	977.7
16	*years*	440	1,465.2	*sleep*	106	1,284.6	*rei* (“example”)	63	962.4
17	*mothers*	402	1,338.7	*adolescents*	106	1,284.6	*hyō* (“table”)	55	840.2
18	*patients*	383	1,275.4	*child*	104	1,260.3	*shokudō heisashō* (“esophageal obstruction”)	55	840.2
19	*adolescents*	372	1,238.8	*prevalence*	103	1,248.2	*jikenrei* (“medical case study”)	54	824.9
20	*hospital*	359	1,195.5	*patients*	100	1,211.8	*shōjō* (“symptoms”)	53	809.6
21	*table*	345	1,148.9	*food*	98	1,187.6	*Fogarty katēteru* (“Fogarty catheter”)	51	779.1
22	*participants*	342	1,138.9	*quality*	94	1,139.1	*kentō* (“investigation”)	50	763.8
23	*CI*	339	1,128.9	*association*	94	1,139.1	*tōka* (“this department”)	49	748.5
24	*analysis*	336	1,118.9	*average/mean*	89	1,078.5	*taishō* (“subject”)	49	748.5
25	*birth*	318	1,058.9	*time*	88	1,066.4	*tōin* (“this clinic”)	47	718.0
26	*level*	304	1,012.3	*level*	87	1,054.3	*kanōsei* (“possibility”)	47	718.0
27	*time*	301	1,002.3	*breast-feeding*	85	1,030.1	*kodomo* (“children”)	47	718.0
28	*mortality*	297	989.0	*sample*	85	1,030.1	*saitai herunia* (“umbilical hernia”)	47	718.0
29	*school*	296	985.7	*treatment*	85	1,030.1	*TEF*	47	718.0
30	*association*	286	952.4	*number*	84	1,017.9	*ato* (“later”)	47	718.0
31	*rates*	286	952.4	*activity*	83	1,005.8	*zenrei* (“all instances”)	46	702.7
32	*n*	284	945.7	*research*	82	993.7	*kon'nan* (“difficulty”)	46	702.7
33	*number*	276	919.1	*loss*	81	981.6	*GER*	45	687.4
34	*activity*	272	905.8	*medicines*	79	957.4	*ta* (“others”)	44	672.2
35	*control*	269	895.8	*feeding*	78	945.2	*p*	44	672.2
36	*mean*	266	885.8	*risk*	77	933.1	*kanji* (“child patient”)	42	641.6
37	*sample*	264	879.1	*factors*	76	921.0	*mokuteki* (“goal”)	40	611.1
38	*rate*	262	872.5	*diseases*	76	921.0	*shōni gekai* (“pediatric surgeon”)	38	580.5
39	*support*	261	869.1	*pieces of information*	75	908.9	*umu* (“presence and absence”)	36	549.9
40	*days*	259	862.5	*form*	74	896.8	*AW*	36	549.9
41	*research*	258	859.1	*table*	73	884.6	*jutsushiki* (“technique”)	35	534.7
42	*status*	257	855.8	*period*	73	884.6	*gappeishō* (“complications”)	35	534.7
43	*levels*	255	849.2	*middle/medium*	72	872.5	*shōni* (“infant”)	34	519.4
44	*factors*	254	845.8	*nurslings*	72	872.5	*heikin* (“average,” “mean”)	34	519.4
45	*outcomes*	248	825.8	*frequency*	72	872.5	*sai* (“occasion”)	34	519.4
46	*family*	242	805.9	*majority*	70	848.3	*keikō* (“trend”)	34	519.4
47	*intervention*	242	805.9	*sex*	67	811.9	*nichirei* (“age in days”)	34	519.4
48	*variables*	238	792.5	*test*	66	799.8	*danji* (“boy”)	32	488.8
49	*groups*	234	779.2	*development*	64	775.6	*shuyō* (“tumor”)	31	473.6
50	*period*	218	725.9	*participants*	64	775.6	*honpō* (“Japan”)	31	473.6

Source: Nouns extracted from word lists generated by Sketch Engine. Numbers calculated by Sketch Engine. Colored rectangles highlight selected words. CI = confidence interval. TEF = tracheoesophageal fistula. GER = gastroesophageal reflux. AW = anterior wrapping.

Table D.3 – Translation of Table 10.2 (Top 50 most frequent nouns in three Management corpora).

	enMGT			ptMGT			jaMGT		
POS	Word	Freq.	Norm. Freq.	Word	Freq.	Norm. Freq.	Word	Freq.	Norm. Freq.
1	*research*	1,666	2,387.0	*relation*	459	2,614.0	*koto* (“thing”)	1,427	8,393.1
2	*work*	1,590	2,278.1	*brand*	395	2,249.5	*kigyō* (“business”)	376	2,211.5
3	*study*	1,514	2,169.2	*results*	352	2,004.6	*soshiki* (“organization”)	352	2,070.3
4	*performance*	1,181	1,692.1	*research*	347	1,976.1	*honkō* (“this paper”)	315	1,852.7
5	*status*	1,011	1,448.6	*study*	292	1,662.9	*eikyō* (“influence”)	305	1,793.9
6	*data*	918	1,315.3	*model*	288	1,640.1	*kangae* (“thought”)	264	1,552.7
7	*business*	902	1,292.4	*companies*	278	1,583.2	*mono* (“thing”)	260	1,529.2
8	*results*	840	1,203.5	*performance*	242	1,378.2	*kenkyū* (“research”)	239	1,405.7
9	*model*	803	1,150.5	*middle/medium*	240	1,366.8	*kekka* (“result”)	239	1,405.7
10	*organizations*	792	1,134.8	*job/work*	229	1,304.1	*kokyaku* (“client”)	232	1,364.5
11	*job*	791	1,133.3	*process*	223	1,270.0	*baai* (“case”)	231	1,358.7
12	*time*	771	1,104.7	*analysis*	216	1,230.1	*menbā* (“member”)	209	1,229.3
13	*firms*	758	1,086.1	*data*	214	1,218.7	*hitsuyō* (“need”)	206	1,211.6
14	*employees*	706	1,011.5	*consumers*	211	1,201.6	*chishiki* (“knowledge”)	195	1,146.9
15	*level*	681	975.7	*Brazil*	193	1,099.1	*ten* (“point”)	180	1,058.7
16	*market*	679	972.9	*products*	190	1,082.0	*soshiki henkaku* (“organizational transformation”)	172	1,011.6
17	*analysis*	672	962.8	*studies*	188	1,070.6	*shigen* (“resource”)	169	994.0
18	*knowledge*	614	879.7	*p*	181	1,030.8	*han'i* (“range”)	166	976.3
19	*relationship*	598	856.8	*development*	179	1,019.4	*keiyaku shain* (“contract employee”)	162	952.8
20	*information*	592	848.2	*services*	178	1,013.7	*bunseki* (“analysis”)	154	905.8
21	*process*	590	845.3	*market*	177	1,008.0	*senkō kenkyū* (“previous research”)	144	847.0
22	*innovation*	583	835.3	*variables*	171	973.8	*kaigai shikaisha* (“overseas subsidiary”)	140	823.4
23	*role*	583	835.3	*administration*	170	968.1	*kasetsu* (“hypothesis”)	135	794.0
24	*identity*	580	831.0	*effect*	163	928.3	*jirei* (“case”)	134	788.1
25	*firm*	553	792.3	*value*	162	922.6	*purojekuto* (“project”)	130	764.6
26	*change*	544	779.4	*time*	158	899.8	*kankei* (“relation”)	130	764.6
27	*findings*	543	778.0	*organization*	150	854.2	*jūgyōin* (“employee”)	128	752.8
28	*leaders*	538	770.8	*base*	149	848.5	*honkenkyū* (“this study”)	124	729.3
29	*table*	536	768.0	*image*	148	842.8	*ta* (“others”)	123	723.4
30	*number*	535	766.5	*company*	147	837.1	*kōka* (“effect”)	123	723.4
31	*management*	532	762.2	*pieces of information*	146	831.5	*kanōsei* (“possibility”)	118	694.0
32	*value*	530	759.4	*processes*	143	814.4	*taishō* (“subject”)	110	647.0
33	*studies*	526	753.6	*capacity*	140	797.3	*yōin* (“factor”)	108	635.2
34	*leadership*	523	749.3	*behavior*	137	780.2	*moderu* (“model”)	106	623.5
35	*experience*	518	742.2	*resources*	136	774.5	*mondai* (“problem”)	104	611.7
36	*feedback*	514	736.5	*actions*	134	763.1	*gyōseki* (“achievements”)	101	594.0
37	*women*	512	733.6	*criteria*	132	751.7	*shinka* (“evolution”)	101	594.0
38	*use*	510	730.7	*users*	132	751.7	*seika* (“results”)	99	582.3
39	*factors*	509	729.3	*literature*	132	751.7	*inobēshon* (“innovation”)	93	547.0
40	*service*	507	726.4	*interviewees*	130	740.3	*kōdō* (“behavior”)	91	535.2
41	*variables*	507	726.4	*level*	129	734.6	*Nihon* (“Japan”)	90	529.3
42	*literature*	500	716.4	*country*	129	734.6	*i* (“distinction”)	90	529.3
43	*cent*	499	715.0	*organizations*	124	706.2	seishain (“full-time employee”)	88	517.6
44	*people*	489	700.6	*management*	123	700.5	*gainen* (“concept”)	81	476.4
45	*context*	489	700.6	*IDORT*	122	694.8	*henkaku* (“transformation”)	79	464.6
46	*effect*	477	683.4	*intention*	121	689.1	*dezain bumon* (“design department”)	78	458.8
47	*respondents*	474	679.1	*form*	121	689.1	*kizon kenkyū* (“existing research”)	78	458.8
48	*members*	471	674.8	*practices*	120	683.4	*hensū* (“variable”)	77	452.9
49	*example*	460	659.1	*agreement*	117	666.3	*seihin* (“product”)	77	452.9
50	*organization*	450	644.8	*relations*	116	660.6	*seichō* (“growth”)	77	452.9

Source: Nouns extracted from word lists generated by Sketch Engine. Numbers calculated by Sketch Engine. Colored rectangles highlight selected words. IDORT = Institute for Rational Work Organization.

Table D.4 – Translation of Table 10.7 (Collocates of cross-linguistically comparable Pediatrics words [*children*, *crianças*, and 子供, *kodomo*]).

	enPED			ptPED			jaPED (Filtered)		
POS	Collocate	PER	Log-LKH	Collocate	PER	Log-LKH	Collocate	PER	Log-LKH
1	*of*	48.3	2,449.798	*of*	55.1	604.621	*iru* (“be”)	61.7	423.390
2	*in*	32.9	1,652.038	*in/at*	26.5	376.997	*josei ishi* (“female doctor”)	61.7	333.395
3	*with*	23.2	1,343.015	*years*	15.0	332.996	*no* (“of”)	83.0	184.738
4	*the*	36.1	1,251.080	*with*	22.7	324.254	*25 mei* (“25 persons”)	14.9	108.185
5	*and*	27.1	985.348	*younger*	10.9	306.040	*de* (“at,” “with”)	53.2	103.883
6	*to*	20.4	792.089	*of the*	15.0	272.829	*nai* (absence)	27.7	95.270
7	*their*	9.1	584.846	*the*	17.8	271.866	*kikon josei ishi* (“married female doctor”)	8.5	45.904
8	*for*	12.4	520.013	*and*	24.6	218.123	*wa* (topic particle)	34.0	45.813
9	*who*	7.4	486.182	*two*	8.7	207.544	*ari* (“be”)	14.9	40.287
10	*aged*	4.8	466.841	*that*	17.1	184.870	*ni tsuite* (“on,” “about”)	8.5	35.641
11	*months*	5.6	374.163	*the*	20.9	153.943	*katei* (“household”)	6.4	33.052
12	*T1DM*	3.4	346.601	*to/for*	13.1	138.809	*24 mei* (“24 persons”)	4.3	29.043
13	*adolescents*	5.0	345.040	*among/between*	8.7	126.521	*josei gekai* (female surgeon)	6.4	24.374
14	*were*	9.3	343.450	*in the*	5.3	102.901	*mota* (“carry”)	4.3	23.500
15	*years*	5.2	338.337	*the*	12.1	87.421	*dansei gekai* (“male surgeon”)	4.3	22.317
16	*young*	3.8	334.806	*a/an/one*	7.8	86.220	*ga* (subject particle)	21.3	22.098
17	*parents*	5.3	327.162	*hospitalization*	3.4	85.073	*motsu* (“carry”)	4.3	20.053
18	*a*	10.6	322.077	*to the*	4.7	81.435	*hijōkin ishi* (“part-time doctor”)	4.3	19.045
19	*mothers*	4.6	297.198	*were*	7.2	81.077	*taiō* (“dealing with”)	4.3	17.005
20	*that*	7.4	285.115	*pneumonia*	3.4	79.072	*kangae* (“thought,” “think”)	6.4	14.377
21	*are*	5.6	273.041	*ASD*	3.7	77.756	*ni* (“in”)	19.1	14.266
22	*among*	4.2	271.371	*prevalence*	4.4	74.780	*uchi* (“while”)	4.3	11.720
23	*ASD*	3.0	254.558	*by/for*	7.2	73.246	*ni okeru* (“in”)	4.3	10.070
24	*age*	4.7	238.753	*healthy*	2.8	69.459	*yō* (“like”)	4.3	9.729
25	*these*	3.8	196.154	*bigger*	4.7	69.088	*naru* (“become”)	4.3	9.570
26	*have*	4.1	194.167	*of the*	9.3	68.801	*aru* (“be,” sentence-ending expression)	6.4	7.966
27	*study*	5.0	178.593	*alimentary*	3.1	66.540	*o* (object particle)	12.8	7.562
28	*than*	3.6	178.531	*number*	3.7	65.278	*mo* (“also”)	6.4	7.253
29	*as*	5.2	176.889	*respiratory*	2.8	65.121	*yori* (“than”)	4.3	6.593
30	*on*	4.7	172.903	*in the*	4.4	63.959	*ya* (“or”)	4.3	5.584
31	*by*	4.6	172.209	*diseases*	3.4	60.108	*to* (multifunction particle)	6.4	2.295
32	*activity*	2.8	172.168	*guide*	2.5	59.384	–	–	–
33	*had*	3.4	164.260	*anemia*	3.1	58.713	–	–	–
34	*physical*	2.5	163.769	*no/not*	5.9	55.677	–	–	–
35	*younger*	1.8	155.082	*hospitalized*	1.6	55.571	–	–	–
36	*sick*	1.4	149.950	*feeding*	3.1	52.072	–	–	–
37	*from*	4.1	148.008	*characteristics*	2.5	47.539	–	–	–
38	*undervaccinated*	1.2	131.743	*accompanied*	1.6	47.211	–	–	–
39	*HIV-infected*	1.1	129.894	*adolescents*	3.1	45.783	–	–	–
40	*or*	4.1	125.370	*months*	3.1	45.594	–	–	–
41	*number*	2.2	123.022	*a/an/one*	5.0	45.160	–	–	–
42	*age-appropriately*	0.9	114.317	*age*	3.7	44.614	–	–	–
43	*older*	1.4	111.887	*breast-fed*	1.2	44.445	–	–	–
44	*this*	3.4	108.082	*in the*	5.6	43.037	–	–	–
45	*all*	2.7	107.499	*economic*	1.6	41.747	–	–	–
46	*health*	2.9	107.126	*until*	2.5	41.717	–	–	–
47	*more*	2.6	104.421	*CAPS*	1.9	41.692	–	–	–
48	*be*	3.0	103.253	*all/every*	2.2	41.380	–	–	–
49	*household*	1.6	102.644	*was*	5.3	40.356	–	–	–
50	*PA*	1.1	100.248	*of the*	4.7	39.636	–	–	–

Source: Collocates identified with Sketch Engine. PER = number of co-occurrences between collocate and node divided by the overall frequency of the node in the corpus. Colors indicate semantically equivalent (blue, green, yellow, and others) or related (red) words. T1DM = type 1 diabetes mellitus. ASD = autism spectrum disorder or adequate sleep duration (English corpus only). PA = physical activity or palmitic acid. CAPS = Psychosocial Care Center.

Table D.5 – Translation of Table 10.8 (Collocates of cross-linguistically comparable Pediatrics words with logDice scores).

	enPED			ptPED			jaPED (Filtered)		
POS	Collocate	PER	LogDice	Collocate	PER	LogDice	Collocate	PER	LogDice
1	of	48.3	10.81	of	55.1	10.21	iru ("be")	61.7	13.55
2	in	32.9	10.83	in/at	26.5	10.95	josei ishi ("female doctor")	61.7	12.80
3	with	23.2	11.12	years	15.0	11.62	no ("of")	83.0	8.45
4	the	36.1	9.99	with	22.7	10.92	25 mei ("25 persons")	14.9	12.08
5	and	27.1	10.11	younger	10.9	11.54	de ("at," "with")	53.2	8.44
6	to	20.4	10.22	of the	15.0	11.27	nai (absence)	27.7	10.71
7	their	9.1	10.79	the	17.8	11.00	kikon josei ishi ("married female doctor")	8.5	11.19
8	for	12.4	10.25	and	24.6	9.90	wa (topic particle)	34.0	7.64
9	who	7.4	10.64	two	8.7	11.15	ari ("be")	14.9	9.71
10	aged	4.8	10.48	that	17.1	10.25	ni tsuite ("on," "about")	8.5	10.79
11	months	5.6	10.38	the	20.9	9.57	katei ("household")	6.4	10.80
12	T1DM	3.4	10.01	to/for	13.1	10.15	24 mei ("24 persons")	4.3	10.39
13	adolescents	5.0	10.29	among/between	8.7	10.50	josei gekai (female surgeon)	6.4	10.34
14	were	9.3	9.91	in the	5.3	10.40	mota ("carry")	4.3	10.33
15	years	5.2	10.28	the	12.1	9.48	dansei gekai ("male surgeon")	4.3	10.30
16	young	3.8	10.12	a/an/one	7.8	10.00	ga (subject particle)	21.3	7.21
17	parents	5.3	10.24	hospitalization	3.4	10.01	motsu ("carry")	4.3	10.22
18	a	10.6	9.65	to the	4.7	10.17	hijōkin ishi ("part-time doctor")	4.3	10.17
19	mothers	4.6	10.14	were	7.2	9.98	taiō ("dealing with")	4.3	10.02
20	that	7.4	9.88	pneumonia	3.4	9.98	kangae ("thought," "think")	6.4	8.96
21	are	5.6	10.01	ASD	3.7	10.04	ni ("in")	19.1	6.70
22	among	4.2	10.04	prevalence	4.4	10.08	uchi ("while")	4.3	9.29
23	ASD	3.0	9.81	by/for	7.2	9.83	ni okeru ("in")	4.3	8.92
24	age	4.7	9.92	healthy	2.8	9.74	yō ("like")	4.3	8.84
25	these	3.8	9.73	bigger	4.7	10.00	naru ("become")	4.3	8.80
26	have	4.1	9.72	of the	9.3	9.45	aru ("be," sentence-ending expression)	6.4	7.57
27	study	5.0	9.54	alimentary	3.1	9.83	o (object particle)	12.8	6.44
28	than	3.6	9.64	number	3.7	9.92	mo ("also")	6.4	7.40
29	as	5.2	9.50	respiratory	2.8	9.72	yori ("than")	4.3	7.95
30	on	4.7	9.54	in the	4.4	9.93	ya ("or")	4.3	7.62
31	by	4.6	9.54	diseases	3.4	9.83	to (multifunction particle)	6.4	5.96
32	activity	2.8	9.54	guide	2.5	9.57	–	–	–
33	had	3.4	9.56	anemia	3.1	9.76	–	–	–
34	physical	2.5	9.46	no/not	5.9	9.61	–	–	–
35	younger	1.8	9.12	hospitalized	1.6	8.97	–	–	–
36	sick	1.4	8.80	feeding	3.1	9.68	–	–	–
37	from	4.1	9.42	characteristics	2.5	9.49	–	–	–
38	undervaccinated	1.2	8.62	accompanied	1.6	8.96	–	–	–
39	HIV-infected	1.1	8.53	adolescents	3.1	9.58	–	–	–
40	or	4.1	9.23	months	3.1	9.58	–	–	–
41	number	2.2	9.21	a/an/one	5.0	9.46	–	–	–
42	age-appropriately	0.9	8.19	age	3.7	9.57	–	–	–
43	older	1.4	8.76	breast-fed	1.2	8.66	–	–	–
44	this	3.4	9.14	in the	5.6	9.31	–	–	–
45	all	2.7	9.17	economic	1.6	8.95	–	–	–
46	health	2.9	9.17	until	2.5	9.42	–	–	–
47	more	2.6	9.14	CAPS	1.9	9.17	–	–	–
48	be	3.0	9.13	all/every	2.2	9.32	–	–	–
49	household	1.6	8.90	was	5.3	9.28	–	–	–
50	PA	1.1	8.50	of the	4.7	9.34	–	–	–

Source: Collocates identified with Sketch Engine. PER = number of co-occurrences between collocate and node divided by the overall frequency of the node in the corpus. Colors indicate semantically equivalent (blue, green, yellow, and others) or related (red) words. T1DM = type 1 diabetes mellitus. ASD = autism spectrum disorder or adequate sleep duration (English corpus only). PA = physical activity or palmitic acid. CAPS = Psychosocial Care Center.

Table D.6 – Translation of Table 10.9 (Collocates of cross-linguistically comparable Pediatrics words [*children* and *crianças*] and 患児 [*kanji*, “child patient”]).

	enPED			ptPED			jaPED (Filtered)		
POS	Collocate	PER	Log-LKH	Collocate	PER	Log-LKH	Collocate	PER	Log-LKH
1	*of*	48.3	2,449.798	*of*	55.1	604.621	*no* (“of’)	78.6	149.158
2	*in*	32.9	1,652.038	*in/at*	26.5	376.997	*o* (object particle)	50.0	85.603
3	*with*	23.2	1,343.015	*years*	15.0	332.996	*wa* (topic particle)	50.0	79.493
4	*the*	36.1	1,251.080	*with*	22.7	324.254	*ni* (“in’)	50.0	76.401
5	*and*	27.1	985.348	*younger*	10.9	306.040	*to* (multifunction particle)	33.3	53.057
6	*to*	20.4	792.089	*of the*	15.0	272.829	*ga* (subject particle)	28.6	33.695
7	*their*	9.1	584.846	*the*	17.8	271.866	*de* (“at,” “with”)	28.6	32.127
8	*for*	12.4	520.013	*and*	24.6	218.123	*ya* (“or”)	14.3	31.143
9	*who*	7.4	486.182	*two*	8.7	207.544	*kazoku* (“family”)	7.1	27.441
10	*aged*	4.8	466.841	*that*	17.1	184.870	*shintaiteki tokuchō* (“physical characteristics”)	4.8	25.685
11	*months*	5.6	374.163	*the*	20.9	153.943	*tomonau* (“accompany”)	7.1	25.069
12	*T1DM*	3.4	346.601	*to/for*	13.1	138.809	*shōkaki geka shikkan* (“gastroenterological surgery disease”)	4.8	21.870
13	*adolescents*	5.0	345.040	*among/between*	8.7	126.521	*awase* (“align”)	4.8	20.513
14	*were*	9.3	343.450	*in the*	5.3	102.901	*ni taishite* (“in contrast with,” “toward”)	7.1	19.554
15	*years*	5.2	338.337	*the*	12.1	87.421	*keikan eiyō* (“tube feeding”)	4.8	19.505
16	*young*	3.8	334.806	*a/an/one*	7.8	86.220	*suru* (“do”)	7.1	19.469
17	*parents*	5.3	327.162	*hospitalization*	3.4	85.073	*e* (“to,” “for”)	7.1	18.910
18	*a*	10.6	322.077	*to the*	4.7	81.435	*genshikkan* (“primary disease”)	4.8	18.034
19	*mothers*	4.6	297.198	*were*	7.2	81.077	*yūsuru* (“have”)	4.8	17.463
20	*that*	7.4	285.115	*pneumonia*	3.4	79.072	*hitsuyō* (“need”)	7.1	16.411
21	*are*	5.6	273.041	*ASD*	3.7	77.756	*GERD*	4.8	15.275
22	*among*	4.2	271.371	*prevalence*	4.4	74.780	*jōtai* (“condition”)	4.8	15.275
23	*ASD*	3.0	254.558	*by/for*	7.2	73.246	*1 rei* (“one example”)	7.1	14.498
24	*age*	4.7	238.753	*healthy*	2.8	69.459	*sono* (“that”)	7.1	13.771
25	*these*	3.8	196.154	*bigger*	4.7	69.088	*taishō* (“subject”)	4.8	12.848
26	*have*	4.1	194.167	*of the*	9.3	68.801	*shōjō* (“symptom”)	4.8	12.533
27	*study*	5.0	178.593	*alimentary*	3.1	66.540	*ooku* (“man””)	4.8	11.366
28	*than*	3.6	178.531	*number*	3.7	65.278	*oyobi* (“and”)	4.8	10.844
29	*as*	5.2	176.889	*respiratory*	2.8	65.121	*baai* (“case”)	4.8	9.787
30	*on*	4.7	172.903	*in the*	4.4	63.959	*toshite* (“as”)	4.8	8.571
31	*by*	4.6	172.209	*diseases*	3.4	60.108	*yoru* (“depend”)	4.8	8.217
32	*activity*	2.8	172.168	*guide*	2.5	59.384	*mo* (“also”)	7.1	7.870
33	*had*	3.4	164.260	*anemia*	3.1	58.713	*koto* (“thing”)	4.8	4.961
34	*physical*	2.5	163.769	*no/not*	5.9	55.677	–	–	–
35	*younger*	1.8	155.082	*hospitalized*	1.6	55.571	–	–	–
36	*sick*	1.4	149.950	*feeding*	3.1	52.072	–	–	–
37	*from*	4.1	148.008	*characteristics*	2.5	47.539	–	–	–
38	*undervaccinated*	1.2	131.743	*accompanied*	1.6	47.211	–	–	–
39	*HIV-infected*	1.1	129.894	*adolescents*	3.1	45.783	–	–	–
40	*or*	4.1	125.370	*months*	3.1	45.594	–	–	–
41	*number*	2.2	123.022	*a/an/one*	5.0	45.160	–	–	–
42	*age-appropriately*	0.9	114.317	*age*	3.7	44.614	–	–	–
43	*older*	1.4	111.887	*breast-fed*	1.2	44.445	–	–	–
44	*this*	3.4	108.082	*in the*	5.6	43.037	–	–	–
45	*all*	2.7	107.499	*economic*	1.6	41.747	–	–	–
46	*health*	2.9	107.126	*until*	2.5	41.717	–	–	–
47	*more*	2.6	104.421	*CAPS*	1.9	41.692	–	–	–
48	*be*	3.0	103.253	*all/every*	2.2	41.380	–	–	–
49	*household*	1.6	102.644	*was*	5.3	40.356	–	–	–
50	*PA*	1.1	100.248	*of the*	4.7	39.636	–	–	–

Source: Collocates identified with Sketch Engine. PER = number of co-occurrences between collocate and node divided by the overall frequency of the node in the corpus. Colors indicate semantically equivalent (blue, green, yellow, and others) or related (red and pink) words. T1DM = type 1 diabetes mellitus. ASD = autism spectrum disorder or adequate sleep duration (English corpus only). PA = physical activity or palmitic acid. CAPS = Psychosocial Care Center. GERD = gastroesophageal reflux disease.

Table D.7 – Translation of Table 10.10 (Collocates of cross-linguistically comparable Pediatrics words and 患児 [*kanji*, “child patient”] with logDice scores).

	enPED			ptPED			jaPED (Filtered)		
POS	Collocate	PER	LogDice	Collocate	PER	LogDice	Collocate	PER	LogDice
1	*of*	48.3	10.81	*of*	55.1	10.21	*no* (“of”)	78.6	8.22
2	*in*	32.9	10.83	*in/at*	26.5	10.95	*o* (object particle)	50.0	8.25
3	*with*	23.2	11.12	*years*	15.0	11.62	*wa* (topic particle)	50.0	8.04
4	*the*	36.1	9.99	*with*	22.7	10.92	*ni* (“in”)	50.0	7.93
5	*and*	27.1	10.11	*younger*	10.9	11.54	*to* (multifunction particle)	33.3	8.18
6	*to*	20.4	10.22	*of the*	15.0	11.27	*ga* (subject particle)	28.6	7.48
7	*their*	9.1	10.79	*the*	17.8	11.00	*de* (“at,” “with”)	28.6	7.38
8	*for*	12.4	10.25	*and*	24.6	9.90	*ya* (“or”)	14.3	9.22
9	*who*	7.4	10.64	*two*	8.7	11.15	*kazoku* (“family”)	7.1	10.63
10	*aged*	4.8	10.48	*that*	17.1	10.25	*shintaiteki tokuchō* (“physical characteristics”)	4.8	10.51
11	*months*	5.6	10.38	*the*	20.9	9.57	*tomonau* (“accompany”)	7.1	10.44
12	*T1DM*	3.4	10.01	*to/for*	13.1	10.15	*shōkaki geka shikkan* (“gastroenterological surgery disease”)	4.8	10.42
13	*adolescents*	5.0	10.29	*among/between*	8.7	10.50	*awase* (“align”)	4.8	10.36
14	*were*	9.3	9.91	*in the*	5.3	10.40	*ni taishite* (“in contrast with,” “toward”)	7.1	9.76
15	*years*	5.2	10.28	*the*	12.1	9.48	*keikan eiyō* (“tube feeding”)	4.8	10.30
16	*young*	3.8	10.12	*a/an/one*	7.8	10.00	*suru* (“do”)	7.1	9.75
17	*parents*	5.3	10.24	*hospitalization*	3.4	10.01	*e* (“to,” “for”)	7.1	9.67
18	*a*	10.6	9.65	*to the*	4.7	10.17	*genshikkan* (“primary disease”)	4.8	10.19
19	*mothers*	4.6	10.14	*were*	7.2	9.98	*yūsuru* (“have”)	4.8	10.14
20	*that*	7.4	9.88	*pneumonia*	3.4	9.98	*hitsuyō* (“need”)	7.1	9.25
21	*are*	5.6	10.01	*ASD*	3.7	10.04	*GERD*	4.8	9.89
22	*among*	4.2	10.04	*prevalence*	4.4	10.08	*jōtai* (“condition”)	4.8	9.89
23	*ASD*	3.0	9.81	*by/for*	7.2	9.83	*1 rei* (“one example”)	7.1	8.88
24	*age*	4.7	9.92	*healthy*	2.8	9.74	*sono* (“that”)	7.1	8.74
25	*these*	3.8	9.73	*bigger*	4.7	10.00	*taishō* (“subject”)	4.8	9.49
26	*have*	4.1	9.72	*of the*	9.3	9.45	*shōjō* (“symptom”)	4.8	9.43
27	*study*	5.0	9.54	*alimentary*	3.1	9.83	*ooku* (“man””)	4.8	9.18
28	*than*	3.6	9.64	*number*	3.7	9.92	*oyobi* (“and”)	4.8	9.06
29	*as*	5.2	9.50	*respiratory*	2.8	9.72	*baai* (“case”)	4.8	8.79
30	*on*	4.7	9.54	*in the*	4.4	9.93	*toshite* (“as”)	4.8	8.45
31	*by*	4.6	9.54	*diseases*	3.4	9.83	*yoru* (“depend”)	4.8	8.35
32	*activity*	2.8	9.54	*guide*	2.5	9.57	*mo* (“also”)	7.1	7.41
33	*had*	3.4	9.56	*anemia*	3.1	9.76	*koto* (“thing”)	4.8	7.27
34	*physical*	2.5	9.46	*no/not*	5.9	9.61	–	–	–
35	*younger*	1.8	9.12	*hospitalized*	1.6	8.97	–	–	–
36	*sick*	1.4	8.80	*feeding*	3.1	9.68	–	–	–
37	*from*	4.1	9.42	*characteristics*	2.5	9.49	–	–	–
38	*undervaccinated*	1.2	8.62	*accompanied*	1.6	8.96	–	–	–
39	*HIV-infected*	1.1	8.53	*adolescents*	3.1	9.58	–	–	–
40	*or*	4.1	9.23	*months*	3.1	9.58	–	–	–
41	*number*	2.2	9.21	*a/an/one*	5.0	9.46	–	–	–
42	*age-appropriately*	0.9	8.19	*age*	3.7	9.57	–	–	–
43	*older*	1.4	8.76	*breast-fed*	1.2	8.66	–	–	–
44	*this*	3.4	9.14	*in the*	5.6	9.31	–	–	–
45	*all*	2.7	9.17	*economic*	1.6	8.95	–	–	–
46	*health*	2.9	9.17	*until*	2.5	9.42	–	–	–
47	*more*	2.6	9.14	*CAPS*	1.9	9.17	–	–	–
48	*be*	3.0	9.13	*all/every*	2.2	9.32	–	–	–
49	*household*	1.6	8.90	*was*	5.3	9.28	–	–	–
50	*PA*	1.1	8.50	*of the*	4.7	9.34	–	–	–

Source: Collocates identified with Sketch Engine. PER = number of co-occurrences between collocate and node divided by the overall frequency of the node in the corpus. Colors indicate semantically equivalent (blue, green, yellow, and others) or related (red and pink) words. T1DM = type 1 diabetes mellitus. ASD = autism spectrum disorder or adequate sleep duration (English corpus only). PA = physical activity or palmitic acid. CAPS = Psychosocial Care Center. GERD = gastroesophageal reflux disease.

Table D.8 – Translation of Table 10.11 (Collocates of cross-linguistically comparable Management words [*organization*, *organização*, and 組織, *soshiki*]).

	enMGT			ptMGT			jaMGT (Filtered)		
POS	Collocate	PER	Log-LKH	Collocate	PER	Log-LKH	Collocate	PER	Log-LKH
1	*the*	67.1	1,279.282	*of the*	33.3	251.924	*no* (“of”)	64.8	788.217
2	*of*	32.2	476.203	*the*	41.3	238.316	*wa* (topic particle)	43.5	584.888
3	*to*	29.6	468.348	*of*	39.3	169.651	*ni* (“in”)	45.7	530.434
4	*an*	15.6	400.559	*and*	28.0	135.963	*ga* (subject particle)	34.9	416.282
5	*in*	23.3	362.972	*a/an/one*	18.0	130.297	*o* (object particle)	32.1	327.014
6	*and*	24.9	323.625	*to/for*	17.3	103.515	*de* (“at,” “with”)	26.4	296.817
7	*a*	17.8	259.462	*in the*	14.0	97.954	*tsuyoi* (“strong”)	7.4	224.231
8	*their*	10.0	199.759	*that*	17.3	78.496	*kansei* (“inertia”)	5.1	202.570
9	*for*	10.2	155.762	*work*	8.0	76.763	*ya* (“or”)	11.6	177.228
10	*at*	6.4	143.228	*rational*	4.7	72.535	*to* (“and,” among others)	19.6	171.415
11	*that*	10.9	133.862	*of the*	14.0	67.215	*nōryoku kaihatsu* (“ability development”)	5.4	165.974
12	*within*	4.7	132.644	*with*	12.0	67.071	*mo* (“also”)	11.4	142.402
13	*or*	6.9	126.424	*within*	4.7	55.152	*na* (adjective particle)	12.5	134.722
14	*implicated*	2.2	125.994	*in*	11.3	52.234	*yowai* (“weak”)	3.7	126.597
15	*members*	4.2	121.325	*by the*	6.0	51.640	*soshiki kansei* (“organizational inertia”)	4.0	126.577
16	*as*	8.2	108.463	*the*	13.3	47.331	*suitai* (“decline”)	3.1	121.772
17	*with*	7.3	106.364	*worker*	2.7	44.273	*aru* (“be,” sentence-ending expression)	9.9	104.267
18	*on*	6.9	97.347	*members*	3.3	42.163	*yoru* (“depend,” “based on”)	5.7	103.584
19	*by*	5.8	88.414	*is*	7.3	38.773	*jūgyōin* (“employee”)	4.5	102.489
20	*employees*	3.6	83.707	*to the*	6.7	38.237	*sōtei* (“hypothesis”)	3.4	98.600
21	*is*	6.9	78.975	*this*	4.7	37.613	*yūkōsei* (“effectiveness”)	3.4	94.894
22	*from*	4.9	78.844	*his/her/their/its*	4.7	34.876	*kono* (“this”)	5.7	74.560
23	*this*	5.8	77.825	*of the*	7.3	34.387	*kiyo suru* (“contribute”)	2.3	74.373
24	*can*	4.0	76.187	*scientific*	2.0	34.139	*yō* (“like”)	5.7	72.848
25	*career*	2.2	72.790	*of the*	6.0	32.683	*ito* (“intention’)	2.6	71.674
26	*scandal-stricken*	1.3	71.612	*commitment*	3.3	31.377	*toshite* (“as’)	5.1	69.341
27	*focal*	1.8	63.617	*act*	2.0	30.247	*mikata* (“viewpoint”)	2.0	68.740
28	*be*	4.7	62.937	*capacities*	2.7	26.431	*ni okeru* (“in”)	4.0	68.242
29	*theory*	2.4	59.857	*institute*	2.7	26.192	*kara* (“from”)	6.0	67.775
30	*whole*	1.6	56.328	*employees*	2.7	25.516	*kōchokusei* (“rigidity”)	1.7	62.818
31	*may*	3.1	55.276	*contributions*	2.0	25.257	*sono* (“that”)	5.1	58.737
32	*are*	4.7	54.953	*by/for*	6.0	25.157	*enpuroiabirity hoshō* (“employability security”)	2.3	58.725
33	*it*	3.8	52.022	*as*	6.0	25.113	*keitai henka* (“morphological change”)	1.1	57.795
34	*functioning*	1.1	48.528	*position*	2.0	24.674	*iu* (“say”)	4.8	53.573
35	*more*	3.3	47.768	*his/her/their/its*	3.3	24.332	*tame* (“to,” “for,” “because”)	4.0	52.018
36	*theorists*	0.9	46.450	*affective*	2.7	24.318	*ni taisuru* (“regarding”)	2.6	50.259
37	*such*	2.7	45.397	*expectations*	2.0	23.894	*Ōbei* (“Europe and America”)	1.4	49.523
38	*its*	2.2	44.780	*essentially*	1.3	22.746	*senryaku* (“strategy”)	2.3	48.192
39	*one*	2.7	44.460	*committed*	1.3	22.746	*ta* (“other[s]”)	2.6	47.539
40	*theorists’*	0.7	44.100	*employee*	1.3	22.746	*koto* (“thing”)	5.7	43.460
41	*work*	2.9	42.506	*performance*	3.3	22.522	*chomei* (“famous”)	1.7	43.080
42	*not*	3.1	40.698	*virtuous*	1.3	21.563	*ishi kettei* (“decision making”)	2.0	42.480
43	*behaviour*	1.6	40.212	*stay*	1.3	21.563	*kengen* (“authority”)	1.4	40.654
44	*what*	2.0	39.884	*Farace*	1.3	20.657	*henkaku* (“change”)	2.0	39.663
45	*these*	2.9	39.466	*feel*	1.3	19.921	*seiritsu suru* (“establish”)	1.1	37.141
46	*specific*	1.8	39.141	*a/an/one*	5.3	19.487	*hoshō* (“guarantee”)	1.1	37.141
47	*needs*	1.6	38.867	*administrative*	1.3	19.301	*sosei shi* (“constitute”)	0.9	37.106
48	*how*	2.4	38.807	*duties*	1.3	19.301	*kaku* (“lack”)	0.9	37.106
49	*social*	2.0	38.241	*he*	2.0	19.144	*shūchū shi* (“concentrate”)	1.1	36.050
50	*employee*	1.6	36.437	*departments*	1.3	18.765	*e* (“to,” “for”)	2.6	35.431

Source: Collocates identified with Sketch Engine. PER = number of co-occurrences between collocate and node divided by the overall frequency of the node in the corpus. Colors indicate semantically equivalent (green, yellow, orange, and purple) or related (red and pink) words.

Table D.9 – Translation of Table 10.12 (Collocates of cross-linguistically comparable Management words with logDice scores).

	enMGT			ptMGT			jaMGT (Filtered)		
POS	Collocate	PER	LogDice	Collocate	PER	LogDice	Collocate	PER	LogDice
1	*the*	67.1	8.18	*of the*	33.3	9.46	*no* ("of")	64.8	9.29
2	*of*	32.2	7.82	*the*	41.3	8.56	*wa* (topic particle)	43.5	9.74
3	*to*	29.6	8.02	*of*	39.3	7.86	*ni* ("in")	45.7	9.35
4	*an*	15.6	9.57	*and*	28.0	8.21	*ga* (subject particle)	34.9	9.49
5	*in*	23.3	8.01	*a/an/one*	18.0	9.36	*o* (object particle)	32.1	9.15
6	*and*	24.9	7.57	*to/for*	17.3	8.79	*de* ("at," "with")	26.4	9.39
7	*a*	17.8	7.88	*in the*	14.0	9.24	*tsuyoi* ("strong")	7.4	10.95
8	*their*	10.0	8.71	*that*	17.3	8.10	*kansei* ("inertia")	5.1	10.62
9	*for*	10.2	7.99	*work*	8.0	10.02	*ya* ("or")	11.6	10.01
10	*at*	6.4	8.92	*rational*	4.7	10.39	*to* ("and," among others)	19.6	8.89
11	*that*	10.9	7.51	*of the*	14.0	8.24	*nōryoku kaihatsu* ("ability development")	5.4	10.58
12	*within*	4.7	9.44	*with*	12.0	8.61	*mo* ("also")	11.4	9.57
13	*or*	6.9	8.42	*within*	4.7	10.07	*na* (adjective particle)	12.5	9.27
14	*implicated*	2.2	9.46	*in*	11.3	8.14	*yowai* ("weak")	3.7	10.14
15	*members*	4.2	9.40	*by the*	6.0	9.58	*soshiki kansei* ("organizational inertia")	4.0	10.21
16	*as*	8.2	7.65	*the*	13.3	7.61	*suitai* ("decline")	3.1	9.94
17	*with*	7.3	7.85	*worker*	2.7	9.69	*aru* ("be," sentence-ending expression)	9.9	9.18
18	*on*	6.9	7.79	*members*	3.3	9.79	*yoru* ("depend," "based on")	5.7	9.99
19	*by*	5.8	7.95	*is*	7.3	8.42	*jūgyōin* ("employee")	4.5	10.09
20	*employees*	3.6	8.83	*to the*	6.7	8.59	*sōtei* ("hypothesis")	3.4	9.96
21	*is*	6.9	7.36	*this*	4.7	9.28	*yūkōsei* ("effectiveness")	3.4	9.94
22	*from*	4.9	8.04	*his/her/their/its*	4.7	9.10	*kono* ("this")	5.7	9.41
23	*this*	5.8	7.67	*of the*	7.3	8.15	*kiyo suru* ("contribute")	2.3	9.47
24	*can*	4.0	8.37	*scientific*	2.0	9.30	*yō* ("like")	5.7	9.37
25	*career*	2.2	9.07	*of the*	6.0	8.45	*ito* ("intention')	2.6	9.57
26	*scandal-stricken*	1.3	8.73	*commitment*	3.3	9.36	*toshite* ("as')	5.1	9.41
27	*focal*	1.8	8.92	*act*	2.0	9.26	*mikata* ("viewpoint")	2.0	9.30
28	*be*	4.7	7.64	*capacities*	2.7	9.24	*ni okeru* ("in")	4.0	9.61
29	*theory*	2.4	8.67	*institute*	2.7	9.23	*kara* ("from")	6.0	9.18
30	*whole*	1.6	8.77	*employees*	2.7	9.19	*kōchokusei* ("rigidity")	1.7	9.09
31	*may*	3.1	8.12	*contributions*	2.0	9.17	*sono* ("that")	5.1	9.13
32	*are*	4.7	7.37	*by/for*	6.0	7.89	*enpuroiabiríty hoshō* ("employability security")	2.3	9.38
33	*it*	3.8	7.64	*as*	6.0	7.89	*keitai henka* ("morphological change")	1.1	8.53
34	*functioning*	1.1	8.43	*position*	2.0	9.15	*iu* ("say")	4.8	9.05
35	*more*	3.3	7.69	*his/her/their/its*	3.3	8.87	*tame* ("to," "for," "because")	4.0	9.20
36	*theorists*	0.9	8.16	*affective*	2.7	9.11	*ni taisuru* ("regarding")	2.6	9.33
37	*such*	2.7	7.97	*expectations*	2.0	9.13	*Ōbei* ("Europe and America")	1.4	8.83
38	*its*	2.2	8.23	*essentially*	1.3	8.73	*senryaku* ("strategy")	2.3	9.26
39	*one*	2.7	7.93	*committed*	1.3	8.73	*ta* ("other[s]")	2.6	9.28
40	*theorists'*	0.7	7.76	*employee*	1.3	8.73	*koto* ("thing")	5.7	8.53
41	*work*	2.9	7.71	*performance*	3.3	8.71	*chomei* ("famous")	1.7	9.00
42	*not*	3.1	7.50	*virtuous*	1.3	8.72	*ishi kettei* ("decision making")	2.0	9.10
43	*behaviour*	1.6	8.39	*stay*	1.3	8.72	*kengen* ("authority")	1.4	8.79
44	*what*	2.0	8.14	*Farace*	1.3	8.71	*henkaku* ("change")	2.0	9.06
45	*these*	2.9	7.56	*feel*	1.3	8.71	*seiritsu suru* ("establish")	1.1	8.50
46	*specific*	1.8	8.24	*a/an/one*	5.3	7.62	*hoshō* ("guarantee")	1.1	8.50
47	*needs*	1.6	8.34	*administrative*	1.3	8.70	*sosei shi* ("constitute")	0.9	8.11
48	*how*	2.4	7.80	*duties*	1.3	8.70	*kaku* ("lack")	0.9	8.11
49	*social*	2.0	8.06	*he*	2.0	8.90	*shūchū shi* ("concentrate")	1.1	8.50
50	*employee*	1.6	8.24	*departments*	1.3	8.69	*e* ("to," "for")	2.6	8.95

Source: Collocates identified with Sketch Engine. PER = number of co-occurrences between collocate and node divided by the overall frequency of the node in the corpus. Colors indicate semantically equivalent (green, yellow, orange, and purple) or related (red and pink) words.