

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Faculdade de Letras
Programa de Pós-Graduação em Estudos Linguísticos

Janailton Mick Vitor da Silva

(UN)SURPRISINGLY, STYLE CAN BE FOUND IN ACADEMIC WRITING:

towards characterizing Douglas Biber's metadiscourse style

Belo Horizonte
2024

Janailton Mick Vitor da Silva

**(UN)SURPRISINGLY, STYLE CAN BE FOUND IN ACADEMIC WRITING:
towards characterizing Douglas Biber's metadiscourse style**

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Estudos Linguísticos (PosLin), da Faculdade de Letras (FALE), da Universidade Federal de Minas Gerais (UFMG), como requisito parcial para obtenção do Título de Doutor em Estudos Linguísticos.

ÁREA DE CONCENTRAÇÃO: 1 –
Linguística Teórica e Descritiva.

LINHA DE PESQUISA: 1D – Estudos
Linguísticos Baseados em *Corpora*.

ORIENTADORA: Dra. Deise Prina Dutra.

Belo Horizonte
2024

S586u

Silva, Janailton Mick Vitor da.

(Un)surprisingly, style can be found in academic writing : [recurso eletrônico]: towards characterizing Douglas Biber's metadiscourse style/ Janailton Mick Vitor da Silva. – 2024.
1 recurso online (176 f.: il., tabs., color., p&b.): pdf.

Orientadora: Deise Prina Dutra.

Área de concentração: Linguística Teórica e Descritiva.

Linha de pesquisa: Estudos da Linguísticos Baseados em Corpora.

Tese (doutorado) – Universidade Federal de Minas Gerais, Faculdade de Letras.

Bibliografia: f. 129-137.

Apêndices: f. 138-154.

Anexos: f. 155-176.

1. Linguística aplicada – Teses. 2. Linguística de corpus – Teses. 3. Redação acadêmica – Teses. 4. Linguagem e línguas – Estilo – Teses. I. Dutra, Deise Prina. II. Universidade Federal de Minas Gerais. Faculdade de Letras. III Título.

CDD : 418



UNIVERSIDADE FEDERAL DE MINAS GERAIS

FOLHA DE APROVAÇÃO

**(UN)SURPRISINGLY, STYLE CAN BE FOUND IN ACADEMIC WRITING: towards
characterizing Douglas Biber's metadiscourse style**

JANAILTON MICK VITOR DA SILVA

Tese submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ESTUDOS LINGÜÍSTICOS, como requisito para obtenção do grau de Doutor em ESTUDOS LINGÜÍSTICOS, área de concentração LINGÜÍSTICA TEÓRICA E DESCRITIVA, linha de pesquisa Estudos Linguísticos Baseados em Corpora.

Aprovada em 04 de setembro de 2024, pela banca constituída pelos membros:

Prof(a). Deise Prina Dutra - Orientadora
UFMG

Prof(a). Antonio Paulo Berber Sardinha
PUC-SP

Prof(a). Ana Eliza Pereira Bocorny
UFRGS

Prof(a). Danilo Duarte Costa
UFVJM

Prof(a). Lucia de Almeida Ferrari
UFMG

Belo Horizonte, 04 de setembro de 2024.



Documento assinado eletronicamente por **Deise Prina Dutra, Professora do Magistério Superior**, em 05/09/2024, às 18:04, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Antônio Paulo Berber Sardinha, Usuário Externo**, em 08/09/2024, às 09:29, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Ana Eliza Pereira Bocorny, Usuária Externa**, em 08/09/2024, às 11:26, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Lucia de Almeida Ferrari, Professora do Magistério Superior**, em 09/09/2024, às 10:18, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Danilo Duarte Costa, Usuário Externo**, em 09/09/2024, às 16:13, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3473892** e o código CRC **6BD7B828**.

ACKNOWLEDGEMENTS

The development of this doctoral investigation has been possible through the meaningful and qualitative support of many individuals and institutions. I am therefore deeply thankful to:

- all my family and friends, for all the support and for understanding my physical absence while I have diasporically pursued my undergraduate and graduate education nationwide over the past decade;
- all my former advisors, teachers, and professors, for providing me with everlasting teachings across my academic career;
- the advisor of this dissertation, Deise Prina Dutra, for all the guidance through the corpus linguistics world over the last (pandemic) years;
- the members of the Specialized and Learner Corpora Study Group (*Grupo de Estudos de Corpora Especializados e de Aprendizes – GECEA*), for the rich debates and exchange of knowledge in the monthly meetings;
- the defense committee members, for assessing this work and offering it insightful comments;
- Federal University of Minas Gerais (*Universidade Federal de Minas Gerais – UFMG*), for continuing to offer free, public, and high-quality education;
- Faculty of Languages, Literature, and Linguistics (*Faculdade de Letras – FALE*) and the Graduate Program in Linguistic Studies (*Programa de Pós-Graduação em Estudos Linguísticos – PosLin*), for offering a renowned, high-level graduate program and for the financial support to participate in academic conferences;
- the officers of *FALE/PosLin*, for promptly answering e-mails and carrying out administrative work effectively.

ABSTRACT

The main goal of this doctoral dissertation is to investigate how metadiscourse language features help shape Douglas Biber's style in research articles published by peer-reviewed journals. In this investigation, style is defined as the particular way a scholar makes variation in the use of language features in academic genres and registers for specific purposes, resulting in the expression of voice(s), authorial identity(ies), stance(s), and potential effects in texts and audiences. A recurring feature of academic register is metadiscourse in two broad dimensions: interactive and interactional. While the former refers to scholar's patterns of discourse organization, the latter regards the expression of stance and establishment of engagement with the reading audience. In this research, metadiscourse may pave the way for the characterization of an author's style in academic writing. To achieve the main goal of this doctoral dissertation, a corpus linguistics methodological framework is designed and implemented. This approach involves the compilation of two corpora through the *AntConc* software suite: a study corpus (*Biber Corpus*), and a reference corpus (*Corpus of Journal Articles in Applied Linguistics*). In addition, the *Keyword* tool is used to compare the study corpus against the reference corpus, and the *KWIC* (*Key-Word-In-Context*), *Cluster*, and *Collocate* tools are employed to examine metadiscourse usage in context. The data retrieved from this comparison is assessed following the application of the *Stylistic Metadiscourse Model of Interpersonal Interaction in Academic Writing*. The results illustrate that, although Biber uses interactive and interactional metadiscourse features, his style is mainly interactive while writing research articles. In doing so, he expresses his voice and discloses a variety of authorial identities, mainly presenting himself as supportive and confident. Furthermore, he (un)surprisingly makes individual stylistic choices aimed at organizing discourse so as to better accommodate his participating audience's interests, rhetorical expectations, and reading comprehension processing abilities. Finally, this study shows potential implications for the understanding of how an expert academic writer crafts an authorial style that may empower novice and other expert writers to craft a style that involves the audience in meaningful reading and writing experiences through academic texts.

Keywords: Style. Metadiscourse. Corpus linguistics. Individual linguistic variation. Academic writing.

RESUMO

O objetivo principal desta tese de doutorado é investigar como recursos linguístico-metadiscursivos ajudam a moldar o estilo de Douglas Biber na escrita de artigos de pesquisa publicados por periódicos revisados por pares. Nesta investigação, estilo é definido como a maneira particular como um acadêmico faz variações no uso de recursos linguísticos em gêneros e registros acadêmicos para propósitos específicos, resultando na expressão de voz(es), identidade(s) autoral(ais), postura(s) e potenciais efeitos em textos e leitores(as). Uma característica recorrente do registro acadêmico é o metadiscurso, que se divide em duas dimensões amplas: interativa e interacional. Enquanto a primeira se refere aos padrões de organização do discurso acadêmico, a segunda diz respeito à expressão de postura e a criação de engajamento com o público leitor. Nesta pesquisa, o metadiscurso pode abrir caminho para a caracterização do estilo de um autor na escrita acadêmica. Para atingir o objetivo principal desta tese, uma estrutura metodológica da Linguística de *Corpus* é criada e implementada. Essa abordagem envolve a compilação de dois *corpora* por meio do pacote de programas disponíveis no *AntConc*: um *corpus* de estudo (*Biber Corpus*) e um *corpus* de referência (*Corpus of Journal Articles in Applied Linguistics*). Além disso, a ferramenta *Keyword* é usada para comparar o *corpus* de estudo com o *corpus* de referência, e as ferramentas *KWIC* (*Key-Word-In-Context*), *Cluster* e *Collocate* são empregadas para examinar o uso do metadiscurso em contexto. Os dados oriundos dessa comparação são avaliados seguindo a aplicação do *Modelo Estilístico-Metadiscursivo de Interação Interpessoal na Escrita Acadêmica*. Os resultados ilustram que, embora Biber utilize recursos metadiscursivos interativos e interacionais, seu estilo é predominantemente interativo ao escrever artigos de pesquisa. Ao fazer isso, ele expressa sua voz e revela uma variedade de identidades autorais, principalmente se apresentando como solidário e confiante. Além disso, ele (não)surpreendentemente faz escolhas estilísticas individuais com o objetivo de organizar o discurso de modo a melhor acomodar os interesses, as expectativas retóricas e as habilidades de processamento de compreensão leitora do público participante. Por fim, este estudo mostra implicações para a compreensão de como um autor acadêmico experiente cria um estilo autoral que pode empoderar novos e experientes autores a criar um estilo que envolva o público em experiências significativas de leitura e escrita por meio de textos acadêmicos.

Palavras-chave: Estilo. Metadiscurso. Linguística de *Corpus*. Variação linguística individual. Escrita acadêmica.

LIST OF ILLUSTRATIONS

Figure 1. Continuum of academic knowledge: disciplinary variation	33
Figure 2. Key resources of academic interaction	47
Figure 3. Example of a tagged journal RA from <i>BC</i>	69
Figure 4. Screenshot of <i>AntConc</i> keyword tool.....	76
Figure 5. Stylistic metadiscourse model of interpersonal interaction in academic writing.....	79

LIST OF TABLES

Table 1. Google Scholar metrics of Douglas Biber's paper citations	19
Table 2. Taxonomy of academic articles published in journals	35
Table 3. Framework for describing the situational characteristics of academic journal registers	36
Table 4. Literacy brokers.....	37
Table 5. Distinguishing grammatical features in academic prose	40
Table 6. Distinguishing grammatical features in applied linguistics RAs	41
Table 7. Checklist of linguistic categories for fiction style analysis	42
Table 8. Major semantic categories used to express stance	49
Table 9. Taxonomy for bundles functional analysis	52
Table 10. Overview of <i>BC</i>	59
Table 11. Situational characteristics of the study corpus RAs	60
Table 12. Style guidelines of <i>BC</i> journals publishing companies	62
Table 13. Definitions and layout of a reference corpus.....	64
Table 14. Overview of <i>CorJAAL</i>	64
Table 15. Situational characteristics of the reference corpus RAs per journal.....	66
Table 16. Example of text identification tags in each corpus.....	68
Table 17. <i>AntConc</i> default keyness values	75
Table 18. <i>AntConc</i> default collocate values	78
Table 19. Semantic classification of single-word verbs	80
Table 20. Number of positive keywords	83
Table 21. Categorization of keywords.....	84
Table 22. Number of stylistic metadiscourse interactive and interactional keywords across <i>BC</i>	86
Table 23. Categorization of interactive positive keywords from <i>BC</i>	88
Table 24. Keywords functioning as <i>transition markers</i> across <i>BC</i>	89
Table 25. Keywords functioning as <i>evidentials</i> across <i>BC</i>	91
Table 26. Semantic classification of evidential verbs and their normed frequencies in <i>BC</i>	94
Table 27. Keywords functioning as <i>code glosses</i> across <i>BC</i>	99
Table 28. Keywords functioning as <i>frame markers</i> across <i>BC</i>	102
Table 29. Frequencies of <i>frame markers</i> across <i>BC</i>	103
Table 30. Categorization of interactional positive keywords from <i>BC</i>	107

Table 31. Keywords functioning as <i>boosters</i> across <i>BC</i>	109
Table 32. Semantic classification of booster verbs and their normed frequencies in <i>BC</i>	110
Table 33. Keywords functioning as <i>hedges</i> across <i>BC</i>	113
Table 34. Keywords functioning as <i>attitude markers</i> across <i>BC</i>	115
Table 35. Keywords functioning as <i>endophoric and directive markers</i> across <i>BC</i>	118
Table 36. Number of stylistic metadiscourse keyword types and tokens.....	122
Table 37. Interactive dimension	138
Table 38. Interactional dimension	139
Table 39. Topics and impact factors of the <i>CorJAAL</i> journals	145
Table 40. List of authors whose texts compose <i>CorJAAL</i> separated by journal	146
Table 41. Log-likelihood critical values and associated p-values (Rayson; Berridge; Francis, 2004).....	155
Table 42. List of all positive keywords generated by <i>AntConc</i>	156

LIST OF ACRONYMS

BC	Biber Corpus
CARS	Create-a-Research-Space
CL	Corpus Linguistics
CorJAAL	Corpus of Journal Articles in Applied Linguistics
CV	Curriculum Vitae
IF	Impact Factor
IMRD	Introduction-Method-Result-Discussion
KWIC	Key-Word-In-Context
MD	Multi-Dimensional
MI	Mutual Information
RA	Research Article
RAs	Research Articles
TESOL	Teaching English to Speakers of Other Languages
WST	WordSmith Tools

TABLE OF CONTENTS

CHAPTER 01 – An introduction to this study	14
1.1 An academic writer’s style	14
1.2 Research problem and rationale	17
1.3 Hypotheses, research questions, and objectives	20
CHAPTER 02 – A theoretical overview of style	23
2.1 A broad definition of style.....	23
2.1.1 Style as academic writer’s individual language variation	24
2.1.2 Style as the expression of academic writer’s voice(s), authorial identity(ies), and stance(s)	26
2.1.3 Style as sensitive to social context(s): the case of academic and linguistic norms and conventions	30
2.1.4 Style as writer’s expression of purposes and effects: the case of metadiscourse	41
2.2 Style detected in corpora: corpus stylistics.....	53
CHAPTER 03 – The methodological detection of style	59
3.1 An overview of the study corpus and the reference corpus.....	59
3.1.1 The study corpus.....	59
3.1.2 The reference corpus	63
3.1.3 The compilation of the study and reference corpora	67
3.2 Methodological procedures for data extraction and analysis	70
3.2.1 Introducing the <i>Keyword tool</i> and its tests and metrics.....	70
3.2.2 Selecting keywords for analysis	75
3.2.3 The <i>Stylistic Metadiscourse Model of Interpersonal Interaction in Academic Writing</i>	79
CHAPTER 04 – Douglas Biber’s metadiscourse style	83
4.1 Overview of the keywords results	83
4.2 Biber’s interactive metadiscourse style	87
4.2.1 Transitions	89
4.2.2 Evidentials	91
4.2.3 Code glosses	98
4.2.4 Frame markers	102
4.3 Biber’s interactional metadiscourse style	106
4.3.1 Boosters	108

4.3.2 Hedges	113
4.3.3 Attitude markers	115
4.3.4 Endophoric and directive markers	117
CHAPTER 05 – Final remarks	121
5.1 Revisiting hypotheses, research questions, and objectives.....	121
5.2 Laying out research limitations and foreseeing the future	125
REFERENCES	129
APPENDIX A – Stylistic metadiscourse model of interpersonal interaction in academic writing: interactive dimension.....	138
APPENDIX B – Stylistic metadiscourse model of interpersonal interaction in academic writing: interactional dimension.....	139
APPENDIX C – List of journals that published Biber’s texts	142
APPENDIX D – List of bibliographic references of the texts composing <i>BC</i>	143
APPENDIX E – Topics and impact factors of the <i>CorJAAL</i> journals	145
APPENDIX F – List of authors whose texts compose <i>CorJAAL</i>	146
APPENDIX G – List of bibliographic references of the texts composing <i>CorJAAL</i>	149
ANNEX A – Log-likelihood critical values and associated p-values	155
ANNEX B – List of all positive keywords generated by <i>AntConc</i>	156

CHAPTER 01 – An introduction to this study¹

This chapter introduces a brief definition of style and lays out the research problem, rationale, hypotheses, research questions, and objectives of this doctoral research,

1.1 An academic writer's style

In our everyday life, when we claim that someone has style or is stylish, we promptly assume that they have preferences for peculiar clothing, music, or food, and that such peculiarities set them apart from other people. We may also think that they have particular ways of using words and structures from a given language that can be distinguished from other people's language uses, leaving their fingerprints in texts and revealing a style that can be tied to them. For instance, the way someone employs an adjective in a predicative and not in an attributive position may reveal a personal linguistic preference, to the same degree as when they favor a particular word or expression over others. Therefore, these language choices, tied to other consistent decisions, subsidize the identification of an individual's language style.

Implicit in the word *way* mentioned in the previous paragraph lies the notion of style. In this doctoral investigation, style is investigated in the context of academic writing. It is hereby comprehended as an academic writer's fingerprint, defined by a remarkable preference for some linguistic items and/or structures over others, identified consistently in a variety of texts through comparisons made to other authors' language preferences. Style is also comprehended as an individual's personal form of expression whereby their voice(s), authorial identity(ies), and stance(s) are similarly revealed. Style discloses how academic writers use language in given contexts and for particular purposes, capable of bringing about potential effects in texts and audiences, including the creation of interpersonal bonds with readers.

These bonds are of at least two types, interactive and interactional. According to Hyland (2005), writers can be interactive if they demonstrate awareness of a participating audience by shaping and constraining discourse to meet this audience's needs. In this case, they favor a

¹ Throughout this text, the following stylistic typographical norms are adopted, unless otherwise stated:

- Double quotation marks are used for quoting in the body of the text;
- Single quotation marks are used for marking quotes within quoted authors' citations;
- Italic types are used for highlighting words and terms in foreign languages; quoted authors' emphasis within citations; corpus-related, computerized tools and programs; functional (sub-)categories from our style model; (sub-)corpora names; and our emphasis.
- Underlined types are used for marking metadiscourse single words and lexical bundles.

discourse-organized approach. Furthermore, scholars can be interactional when they are engaging, intrude and comment on their messages, make their views explicit, convey evaluations and judgements, and express solidarity. In this case, they put forth a participant-oriented perspective.

In texts, these interpersonal bonds have been investigated from different angles. Metadiscourse is the way chosen in this dissertation, for it refers to scholar's discourse organization means and stance and engagement patterns towards either discourse or reading audience (Hyland, 2004). In this regard, we understand metadiscourse as part of author's style in academic writing, since scholars may present other types of style, even potentially non-metadiscoursal, when writing in and/or out of academia. This definition adds to long-standing discussions which have surrounded the concept of style: "the elusiveness of style lies in the age-old dispute as to whether style is extrinsic or intrinsic in nature; whether it is the icing on the cake – in effect, an optional extra – or whether it is an inherent part of the cake itself." (Burke, 2014, p. 24).

It is on this broad understanding of style that the main goal of this doctoral research is based. Here, we aim at unveiling the metadiscourse academic writing style of a prominent applied linguist, Douglas Biber, based on frequent and consistent use of metadiscourse language features across journal research articles (RAs)² and in comparison to other authors' articles. In this regard, we subscribe to McMenamin (2002) when he claims that language also varies at the individual level. Furthermore, specifically in the writing context, we advocate that "all writing carries information about the writer [...]", and that "writers cannot avoid projecting an impression of themselves and how they stand in relation to their arguments, their community and their readers." (Hyland, 2005, p. 53).

Biber has been chosen as protagonist in this style investigation because of his distinguishing contributions to corpus linguistics (CL) research in a variety of topics, especially in register studies. According to his personal curriculum vitae (CV)³, he has been a Regents' Professor of Applied Linguistics at Northern Arizona University, in the United States, since 2000. His education background includes a Bachelor of Science degree in Geophysics from

² In this dissertation, the terms *journal article* and *research article* are considered synonyms because the papers which served the analysis of Biber's style are research articles published in different peer-reviewed academic journals. Moreover, this understanding is similar to Gray's (2015, p. 7), in that "[...] research articles are generally grouped into a single register that is defined as texts which report on research and are published in academic journals".

³ Available at: <https://dougibiber.weebly.com/> Access on: 10 Apr. 2023.

The Pennsylvania State University (1974), a Master of Arts degree in Linguistics from The University of Texas at Arlington (1977), and both Master of Arts (1982) and Doctor of Philosophy degrees (1984) in Linguistics from University of Southern California. His CV also shows multiple professional experiences; honorary degrees, awards, fellowships, and grants; short-term teaching and research appointments; and comprehensive lists of published books, monographs, and academic articles. This background thus represents a distinguishing career, with the author even figuring among the world's top scientists, according to the 2022 Stanford list of the World's Top 2% Scientists (Ioannidis, 2023).

In an interview with Gray, Biber (2013, p. 174) defines himself as a linguist, and not a corpus linguist: "I see myself primarily as a linguist. [...] And I don't actively dislike being called a corpus linguist – it's just that it seems odd to me when I really think about it." When he is questioned about how he became acquainted with CL, he describes his experiences in other areas and with mentors at different institutions:

As an undergraduate, I had developed a strong background in science (with a degree in geophysics from Penn State University [...]). I gradually shifted my interests to issues in sociolinguistics, focusing especially on the analysis of spoken versus written discourse. But I had no intention of building on my background in science and technology. Two mentors at USC [University of Southern California] helped me to change that. First, Ed Purcell helped me to develop stronger technical skills. Ed taught me both statistical analysis as well as advanced computer programming skills. [...] A second Ed – Ed Finegan – was central to my development as a corpus linguist, and as a researcher and writer in general. Ed was my dissertation chair and completely supportive of my general interests in spoken and written discourse. [...] he is the one who really got me into corpus linguistics. (Biber, 2013, p. 362-363).

As for the use of corpora in research, he admits the following:

I believe that corpora should be analyzed in quantitative terms using the best statistical methods available. But I also believe that corpus analyses should fundamentally be linguistic investigations, that quantitative findings must be interpreted and explained in linguistic terms, and that those quantitative patterns should be exemplified in specific texts. (Biber, 2013, p. 365).

In this regard, a corpus-based methodological framework is proposed in this doctoral research, as will be better described in Chapter 03. The core of this investigation departs from the empirical side of the language, the *performance*, which is in fact produced by language users and becomes a sample of natural data available in a corpus format, promoting scholarly investigations (McEnery; Wilson, 2001). Specifically, we draw on *AntConc*, version 4.2.0, designed by Lawrence Anthony (2022a), through *Key-Word-In-Context (KWIC)*, *Cluster*, *Collocate*, and *Keyword* tools.

In addition, style should be revealed by CL statistical significance tests inbuilt in such programs as *AntConc* through comparisons between study and reference corpora. In this

context, I have compiled two corpora, a study (*Biber Corpus - BC*) and a reference one (*Corpus of Journal Articles in Applied Linguistics - CorJAAL*). The linguistic data generated by the program was submitted to corpus-based quantitative and qualitative linguistic analyses following the application of the *Stylistic Metadiscourse Model of Interpersonal Interaction in Academic Writing* (Hyland, 2005; 2008a; 2009; Biber, 2006; Biber *et al.*, 2021), which will be explained in Chapter 03.

At last, in this dissertation, I would like to make explicit my personal preference for alternating between first-person pronouns (in singular and plural forms) and other language forms to refer to you, my prime audience. All these choices have not been merely subconscious, but represent a well-thought-out desire to mark my authorial style and voice more noticeably. When I make use of first-person singular pronouns *I*, *my*, and *me*, I bluntly hold on to the thoughts and notions I have provided here and am fully responsible for stating them. When I would rather opt for inclusive, first-person plural forms *we*, *our*, and *us*, I am both including other voices in this work (e.g., advisor's, cited scholars', qualifying exam and defense committee members', etc.) and also inviting you, as audience, to participate in the agency and knowledge construction. I proceed next with introducing the research problem, rationale, hypotheses, research questions, and objectives of this study.

1.2 Research problem and rationale

The paramount reason underlying this doctoral investigation stems from the issue of stylistic profiling raised, but not fully accounted for, by Crystal (2013). The scholar imported from clinical linguistics an approach to studying language disability: grammatical profiling. According to the author, since it had been possible to “[...] identify people by singling out their distinguishing features, and presenting them in a coherent manner [...]”, it seemed similarly possible “[...] to identify the most salient features of emerging grammatical structure in a language-handicapped person, and to present these also in a systematic and clinically illuminating way.” (Crystal, 2013, p. 221). To address this issue, the scholar laid out a series of language areas to be considered in this endeavor, including such elements as phonetics, phonology, graphetics, graphology, grammar, and semantics. Differently from the interests of clinical linguistics, Crystal (2013) used those elements as a basis for identifying salient linguistic structures that were distinctive in legal and newspaper writing instead. According to the scholar, linguistic profiles provide insightful support in at least five ways:

1. They are, firstly, conveniences: their design enables the user to bring together into a single place a great deal of relevant data which would otherwise be fragmented on cards, notes or the like.
2. Secondly, the data are organized in such a way that significant patterns emerge quickly.
3. Thirdly, when information from a sample is plotted on a profile chart, it is immediately apparent (in terms of the categories represented on the chart) not only what *is* in the sample but also what is not - in clinical terms, often a more significant factor.
4. Fourthly, the fact that an attempt has been made to choose and grade only the most important features makes the task of learning to use a profile relatively easy some training schools, in fact, eventually used the procedure as a way of introducing students to English grammar.
5. And fifthly, profiles have a clear numerical dimension, which makes them good sources of input to statistical or computational procedures. In a subject where numerical precision is rated highly, profiles have a natural place. (Crystal, 2013, p. 222-223, emphasis in original).

Although Crystal's (2013) pathway to understanding individuals' and groups' linguistic profiles seemed innovative, his investigation did not focus on academic writers' profile. In view of this limitation, we have addressed the following problem in this research:

- since we know little about an individual's linguistic style, how does the peculiar way an academic writer employs metadiscourse features in a variety of journal articles reveal their authorial linguistic style in academic writing?

In Crystal's (2013) study, many were the language features proposed. Here, however, the focus is given to metadiscourse, which is linguistically realized by some of Crystal's (2013) features. In this dissertation, metadiscourse has been selected as a macro language feature for two major reasons. The methodological reason lies in the fact that "[...] stylistic analysis must be very selective indeed: some studies concentrate on just one feature, and others on a mere handful of features" (Leech; Short, 2007, p. 55). The theoretical reason is grounded in the fact that metadiscourse is considered a feature of academic writing register and has long been used in different academic genres, like RAs (Hyland, 2004; 2005; 2009), and is also one of the most prominent approaches to analyzing academic writing (Pearson; Abdollahzadeh, 2023).

In this regard, to better address the issue of academic writer's style in academia, it is necessary to limit the focus to a macro functional language feature: metadiscourse. As will be seen, metadiscourse is realized by functional subcategories (*transition markers, frame markers, evidential markers, code glosses, hedges, boosters, attitude markers, endophoric and directive markers*, among others), which become materialized by grammatical features, including adjectives (in the case of *attitude markers*), verbs (in *evidentials* and *boosters*), adverbs (in *transitions*), etc., as will be detailed in Chapter 03. Furthermore, metadiscourse has often been investigated with respect to its use in academic genres, commonly ignoring the issue of

authorial style. In this doctoral investigation, because “styles are normally described for the texts within a register or genre” (Biber; Conrad, 2019, p. 18), we attempt to show that metadiscourse is not only characteristic of particular genres or registers, but also constitutes an academic writer’s stylistic repertoire when he writes RAs. Hence, metadiscourse is a way to determine an academic writer’s style.

As previously mentioned, Biber has been selected as protagonist because of his prominence in applied linguistics, particularly in CL, investigating how different language components (lexis, grammar) are used in academic registers. This fact adds to our interest in understanding how his single-authored texts, which have had a great impact among so many scholars in the field, are written and mirror his style. This notoriety has been detected largely based on citations his academic papers have received over the past five years (since 2019), setting him apart from other similarly prominent scholars within CL (Ioannidis, 2023). The Google Scholar metrics⁴ illustrated in Table 1 subsidize our choice of author:

Table 1. Google Scholar metrics of Douglas Biber’s paper citations

Scholar	Areas of interest self-informed to Google Scholar	n. of citations		h-index		i10-index	
		all	since 2019	all	since 2019	all	since 2019
Douglas Biber ⁵	linguistics, corpus linguistics, English grammar, register	80,251	29,841	92	67	227	182

Source: Adapted from Google Scholar (2024).

These results highlight the professor’s substantial contribution to academia, especially to the area of applied linguistics, making him a notorious figure not only as a researcher, but also a writer. Given the fact that scholars from around the world have dedicated time to studying *what* Biber has written over the past four decades, as attested by the citation figures above, it seems relevant to proceed with another research interest: to study the *way* he writes within academia.

Another reason why it is important to carry out this investigation stems from a wider necessity to study writer’s style in the academic context. One of the criticisms raised by theoretical linguists is that traditional stylistics has long been concerned with literary texts and, therefore, literary authors (Jeffries; McIntyre, 2010). As Atkinson (2003, p. 202) stated, “[...] literary writers are not the sole purveyors of written voice or style. Academic writers also exist

⁴ The information provided in this chart has been downloaded on September 05th, 2024. To understand how these metrics are calculated, check the following: <https://scholar.google.com/intl/en/scholar/metrics.html#metrics>

⁵ <https://scholar.google.com/citations?hl=pt-BR&user=mdWIU4MAAAAJ>

who can doubtless[ly] be identified on the basis of their style [...].” Furthermore, Bacon (2013) challenges a classical dichotomy in the field of style in academic writing, particularly in the context of teaching style to undergraduate students. This dichotomy holds that a scholar should either write with style or write for the academy, as if they could not write stylistically within academia. For this reason, we believe it is imperative to investigate a scholar’s style in the academic setting.

At last, by carrying out this investigation, we hope to refute the myth of a dry, impersonal discourse in academic writing (Hyland, 2002a). Although we subscribe to the view that one’s writing results from a myriad of voices and discourses (Bakhtin, 1981; Ivanič, 1998; Tardy, 2012), our focus will be directed to the way Biber has written (and employed his multiples voices, including his *individualized voice*, as Tardy highlights), especially because he, as the main author of the texts analyzed, has claimed authorship (an example is his solo signature stamped on the pages of his RAs) for his ideas by means of specific metadiscourse language patterns. After all, style is the individual’s voice (also being construed in and deriving from the social) in the collectivity.

1.3 Hypotheses, research questions, and objectives

As McIntyre and Walker (2019, p. 121) advocated, “[...] testing our intuitions is one of the prime purposes of doing corpus analysis. Again, this means using our intuitions to generate hypotheses and research questions.” In this regard, my intuitions as a reader of Biber’s texts and now as a researcher investigating his style have helped me formulate some hypotheses, research questions, and objectives to be addressed in this dissertation.

There are two main hypotheses underlying this investigation. The first, if confirmed, may trigger the second. Because metadiscourse is a feature of academic writing, I assume that Biber, like any other academic writer, makes metadiscourse language choices which consistently add up to his larger language repertoire and help constitute his academic writing style when he writes RAs. If this holds true, given Biber’s background experience in the hard sciences and paradigmatic bias towards quantitative and statistical methods (Biber, 2013), the second hypothesis holds that Biber’s metadiscourse choices are more concerned with discourse organization (interactive style) than with stance and engagement (interactional style), which is a strong characteristic of texts written in the hard sciences (Hyland, 2005). In this regard, Biber’s texts may deviate from currently published texts in applied linguistics, an area which is

located in the soft pole of an academic knowledge continuum (Hyland, 2009) and favors a more interactional approach (Hyland, 2004; 2005; 2009).

Based on these two hypotheses, I have formulated the following research questions:

➤ Research questions:

1. What kind of words appear as positively key⁶ in Douglas Biber's RAs when compared to RAs written by other applied linguistics scholars?
2. Are there any keywords that can be classified as metadiscoursal? If so, what is their number of raw and normed values after categorized according to the *Stylistic Metadiscourse Model of Interpersonal Interaction in Academic Writing* (Hyland, 2005; 2008a; 2009; Biber, 2006; Biber *et al.*, 2021)?
3. Does Biber employ interactive metadiscourse language features? If so, how does he organize discourse in his RAs?
4. Does Biber use interactional metadiscourse language features? If so, how does he express stance and establish engagement with readers in his RAs?
5. In general, is Biber's metadiscourse style more interactive or interactional?

Based on these research questions, I have formulated the following objectives:

➤ General objective:

- i) To investigate how metadiscourse language features help shape Douglas Biber's linguistic style in the writing of RAs.

➤ Specific objectives:

- i) To identify positive keywords by comparing *BC* and *CorJAAL*.
- ii) To classify metadiscourse keywords from Biber's RAs.
- iii) To assess how Biber organizes discourse through interactive metadiscourse language features.
- iv) To examine how Biber expresses stance and establishes engagement with readers by means of interactional metadiscourse language features.

⁶ According to Scott (2022), "a word which is *positively* key occurs *more often* than would be expected by chance in comparison with the reference corpus".

- v) To categorize Biber's metadiscourse style as either/both interactive or/and interactional.

Besides this Introduction Chapter, this dissertation also consists of the following chapters. Chapter 02, *A theoretical overview of style*, introduces a general definition of style within the scope of this doctoral research based on existing concepts of style proposed by scholars in different fields of academia. Later, Chapter 03, *The methodological detection of style*, introduces the methodological steps to carry out this investigation. Afterwards, Chapter 04, *Douglas Biber's metadiscourse style*, presents functional analyses of the author's style. It begins with an overview of the keywords results by presenting the types of words identified as key from the comparison between *BC* and *CorJAAL*, and then proceeds with the presentation and discussion of linguistic data following the application of the *Stylistic Metadiscourse Model of Interpersonal Interaction in Academic Writing* (Hyland, 2005; 2008a; 2009; Biber, 2006; Biber *et al.*, 2021). Then, Chapter 05, *Final remarks*, revisits the hypotheses, research questions, and objectives of this doctoral investigation, lays out limitations of the study, and offers suggestions for future research. At last, references, appendices, and annexes are provided at the end.

CHAPTER 02 – A theoretical overview of style

This chapter introduces a general definition of style within the scope of this doctoral research based on existing concepts of style proposed by scholars in different fields of academia. Our definition includes such issues as author's individual language variation, author's expression of voice(s), authorial identity(ies), and stance(s), style as sensitive to academic and linguistic norms and conventions, and metadiscourse language choices used in academic texts.

2.1 A broad definition of style

Style has long been an elusive concept, especially due to its polysemic meaning deriving from an apparent disagreement among scholars as to how it can be properly defined (Burke, 2014). Nevertheless, our definition of style derives mainly from Leech and Short (2007, p. 9), who define it as “[...] the way in which language is used in a given context, by a given person, for a given purpose, and so on”, and McMenamin (2002, p. 126): “style in written language refers to the variable ways that language is used in certain genres, periods, situations, and individuals.” These definitions imply that every choice made in a language comes from a subject (or a group of people) who, in a social and specific context, seeks to impact their audience in a certain way. Building on these major definitions, we attempt to define style within the context of academic writing as:

- style is the particular way a scholar varies the use of language in academic genres and registers for specific purposes, resulting in the expression of voice(s), authorial identity(ies), stance(s), and potential effects in texts and audiences.

This definition of style, though, encompasses some underlying elements. They are all interrelated, but for the sake of pedagogical reasons, they will be discussed separately and in depth in this chapter. Before that, let us split this major definition into smaller parts.

When we argue that style is the particular way an individual makes variation in the use of language in academia, we mean that scholars, here synonymously understood as authors or writers, show consistent preference for particular language items and/or structures over others, representing distinctive individual variation in the register used within the academic context in academic genres.

Moreover, this individual variation may also give room for the expression of academic writers' voice(s), authorial identity(ies), and stance(s) in academic registers and genres, as they inevitably leave marks and traces on what they write. This expression, however, does not represent sole and individual voice(s), identity(ies), and stance(s) per se because, as socially engaged subjects, authors take part in discourse practices enacted in and out of academia, which is a social milieu consisting of specific principles and traditions.

Furthermore, writers' personal form of expression in academia is not simply random, for it needs to abide by academic and linguistic norms, including conventions that regulate academic discourse and represent linguistic disciplinary variation. Language used by academics is imprinted in a myriad of text varieties, such as RAs, which reveals how style is also construed in and derived from the social discourse communities to which authors, readers, and other agents belong in academia.

Scholars' style is adopted with the intention of reaching particular purposes and causing potential effects in texts and audiences. However, since writers are concerned with the expression of research-related facts, information, data, viewpoints, among other meanings, they may also end up creating interpersonal bonds with audiences by means of interactive and interactional metadiscourse features. It is precisely through metadiscourse language features that we identify Biber's style in academic RAs in this doctoral research.

As can be seen, style is a thorough and complex concept, but we hope to narrow it down given the specificity of what is being investigated herein. This is to say that working definitions different from ours might equally serve the comprehension of this multifaceted concept and suit the goals of other scholars' research. The following subsections provide insightful discussions about the underlying elements of our definition of style.

2.1.1 Style as academic writer's individual language variation

Style is the major concern of the area called stylistics, which can be broadly "[...] defined as the (linguistic) study of style" (Leech; Short, 2007, p. 11). Traditionally, the center of style research has been on the linguistic study of both literary texts and literary language since stylistics' inception. However, there have been shifts in such a tradition, with attention now given to non-literary texts and language issues such as discourse, gender, and ideology, focusing on "[...] the study of the ways in which meaning is created through language in literature as well as in other types of text" (Nørgaard; Busse; Montoro, 2010, p. 1).

Nevertheless, attention has also been paid to “[...] the scientific study of patterns of variation in written language [...], the language of a single individual (idiolect), resulting in a description of his or her identifying linguistic characteristics.” (McMenamin, 2002, p. 176). In this regard, style is comprehended as an individual’s or group’s “[...] personal form of expression [...]”⁷ (Garcia, 2006, p. 123, our translation⁸), or “[...] the definition of a personality in linguistic terms”⁹ (Câmara Junior, 1978, p. 13), or yet “[...] a kind of thumb-print that is expressed in a range of linguistic - as well as non-linguistic - features.” (Baker, 2000, p. 245). Hence, Crystal (2013) states that the focus of stylistics would be on linguistic distinction (e.g., foregrounding), on what is peculiar to an individual or social group that separates them (in terms of language use) from another individual or group. This definition suggests that “writers, like composers or painters, often develop styles that are recognizable and distinctive” (Elbow, 1994, p. 13), for style concerns distinctiveness (Irvine, 2001). As for the study of an individual’s patterns of variation (or their idiolect), McMenamin adds:

Stylistic variation is also reflected in the idiolect of individuals who speak the same dialect and language, thus reflecting the individual characteristics of speakers and writers. [...] The range of linguistic variation in the writing style of a single person is identifiable as his or her idiolectal style, reflected as interwriter variation between that writer and other members of the same speech community, or as intrawriter variation occurring within that writer as a single member of the speech community. (McMenamin, 2002, p. 128).

It is particularly in the individual aspect of language variation that lies the central interest of this doctoral investigation. In this regard, style

is a reflection of group or individual variation in written language. Individual variation is a result of the writer’s choices of one form out of the array of all available forms. Choices represent variations within a norm (different ‘correct’ ways of saying the same thing), deviations from a norm (mistakes), and idiosyncrasies (author-specific forms). The style of a writer is demonstrated by his or her unique aggregate set of grammatical patterns, which is usually the result of the writer’s recurrent (habitual) use of some or all of the forms in the set. (McMenamin, 2002, p. 130).

Similarly, Malmkjaer and Carter (2002, p. 510) define style as “[...] a consistent occurrence in the [spoken and written] text of certain items and structures, or types of items and structures, among those offered by the language as a whole.” This is to say that writers choose items and structures among many others likewise available in the language, but for particular reasons they have limited their use to some over others. As can be seen, central to both scholars’

⁷ “[...] *forma pessoal de expressão* [...]” (GARCIA, 2006, p. 123).

⁸ All translations in this dissertation are ours, unless otherwise stated.

⁹ “[...] *a definição de uma personalidade em termos lingüísticos*” (CÂMARA JUNIOR, 1978, p. 13).

definitions are the issues of recurrence (or consistency) and choice of language items and structures:

Recurrent reflects the writer's subconscious habits of regularly choosing one form over others to say the same thing. *Choices* are of two types: variation within a norm or deviation from a norm. Variation within a norm refers to choices among grammatical or acceptable ('correct') forms, e.g., *twenty-six* / *twenty six* / *26*. Deviation from a norm refers to choices that include ungrammatical or unacceptable ('incorrect') forms, e.g., *I might go.* / *I could go.* / *I might could go* / *I might could did go.* (McMenamin, 2002, p. 126).

As we reflect upon this definition, in the context of this doctoral research, we seek to show how Biber often employs some items over others in more than one text. Also, we do not expect his style to deviate from the norm as understood in McMenamin's (2002) quote, especially because his texts have been submitted to extensive proofreading and editing stages and published without these kinds of mistakes reported. In fact, we expect his style to vary within the norm of academic writing in terms of favoring certain metadiscourse features over others while writing RAs, as will be discussed in 2.1.3 and 2.1.4.

2.1.2 Style as the expression of academic writer's voice(s), authorial identity(ies), and stance(s)

Writer's (self-)representation in texts is also related to such concepts as voice, writer's authorial identity, and stance. Although these concepts are interrelated, they are not considered to be synonyms in this dissertation, as will be seen across this subsection. Hyland (2002b, p. 1092) asserts that "academic writing, like all forms of communication, is an act of identity: it not only conveys disciplinary 'content' but also carries a representation of the writer." If we consider that "[...] all writing contains 'voice' [...]" (Hyland, 2008b, p. 144), we can then assume that

the creation of an authorial persona is clearly also an act of personal choice, where the influence of individual personality, confidence, experience, and ideological preference all enter the mix to influence our style. (Hyland, 2008b, p. 158).

As can be seen, some aspects influence a writer's style. For instance, style is inextricably related to *voice*, especially because an important dimension of texts is voice. "Many of the textual features that people describe in terms of voice can also be described as matters of *style*" (Elbow, 2007, p. 177). From a theoretical perspective, voice can be understood from three broad dimensions, namely individual, social, and dialogic (Tardy, 2012). The individual perspective is tied to the notions of style introduced in subsection 2.1.1, and so voice and style would be synonyms, used to refer to the same phenomena: writer's unique and individual way of using

language; writer's individual imprint or stamp on a text; writer's accounts as distinguishable from other writer's. In summary,

individualized voice is often described as a writer's unique and recognizable imprint, associated with *authenticity*, *resonance*, *authoritativeness*, and *authorial presence* within a text. It may relate to notions of *style* and *stance* and is often considered to be carried out through particular linguistic features. More generally, individual aspects of voice contribute to a broader conception of a writer's self-representation within a text [...] (Tardy, 2012, p. 37).

Hyland (2002a; 2002b) holds that writing is a means of self-representation, which is constructed both individually and socially. Therefore, it is relevant to understand voice from a social perspective as well. This is to say that no writing takes shape without a social context, for it needs to be produced within a particular circumstance, however minor it be. For this reason, "[...] 'social voice' is associated more with the disciplinary or other social groups to which the writing and writer are linked." (Tardy, 2012, p. 37). Similarly, the social voice "[...] expressed (or perceived) is not simply the property of the author but constructed by the social worlds that the author works within." (p. 39).

Since voice is responsive to the social context where it is conveyed, we can infer that no individual possesses an entire voice of their own. However, as Tardy (2012) affirmed, the way people mingle and operationalize others' voices in discourse is what characterizes their voice. Bakhtin's (1981) notion of *heteroglossia* is rather useful here, for "language is not a neutral medium that passes freely and easily into the private property of the speaker's intentions; it is populated – overpopulated – with the intentions of others" (Bakhtin, 1981, p. 294).

An easy way to identify this multi-voicedness aspect of authorial voice in academic texts is through citations (Hyland, 2004; 2005). This is an important tool in academic discourse, which moves careers and sanctions what knowledge can be legitimated, at times even turning authors into accommodators (Casanave, 2003). These voices come "[...] from other members of the disciplinary or academic community, historical or contemporary actors and experts or even aggregate voices expressing ideas circulating in and around communities." (Silver, 2012, p. 204). In a personal account of this practice, Casanave (2003, p. 161) shares that:

We learned that being a scholar in academic settings in general and writing for publication in particular means situating our work, and our identities, within bodies of literature. I learned that even though the authors I cite are rarely people I have met in person, they are people with whom I need to have 'conversations' if I wish to contribute my own voice to my field.

However, careful consideration should be given to what may be claimed as fully one's own voice here. In fact, as Devine (2021, p. 32) alerts:

[...] none of us can claim complete, authentic, original authorship. Even our most 'original' ideas are usually recycled older ideas, dusted off for contemporary purposes, or ideas from different sources, brought together to make a new amalgam.

This social perspective is also aligned with Ivanič's (1998) interrelated aspects of writer identity, involving the identity of an actual writer producing a particular text (*autobiographical self, discursial self, self as author*), and the abstract, prototypical identities available in the socio-cultural context of writing (*possibilities for self-hood*). One's identity is socially constructed within ideological and power structures, which also means that the various ways of self-formation and expression are not fixed and can change over time. This is to say that "[...] each individual takes on an identity in relation to the [discourse] communities they come into contact with" (Ivanič, 1998, p. 83), because "a writer's identity is not individual and new, but constituted by the discourses s/he adopts" (Ivanič, 1998, p. 86). These aspects can be revealed in Casanave's (2003) personal account on authorial publication. She admits to portraying herself as having at least three different selves and that people (like herself) in real life can be rather different from how they picture themselves in texts:

One indication for me that I am not the only one whose discursial self represents me only partially is that sometimes when I think I 'know' a published writer in print, then meet that person, I feel as though I have met two different people. [...] I construct three narratives of myself as an academic writer that differ in the ways I talk about participating in my communities of academic practice. The selves in each narrative are constructed through discourse for different purposes, effects, and audiences, and each represents me only partially. In the first narrative, I portray myself as a card-toting Community Member. In the second narrative, I portray myself as a Boundary Pusher. In the third narrative, I might be labeled the Cynic at the Sidelines. (Casanave, 2003, p. 158-159).

Lastly, the dialogic perspective on voice concerns, as the term suggests, the dialogue between both individual and social perspectives previously introduced. For Tardy (2012, p. 39), "[...] voice is neither controlled solely by writers nor determined by the social worlds within which they write – it is subject to and a result of both writer and social context." This notion also includes another individual who is similarly part of this circumstance: the reader. Therefore, "a dialogic view draws our attention [...] to the reader and to the ways *in which writer and reader interact*, co-constructing voice at a particular space and time" (Tardy, 2012, p. 40). In this way, writers also construct an image of themselves "[...] in the light of their estimates of their expected reader(s) interpretive practices, and of the various power

asymmetries between them.” (Ivanič, 1998, p. 102). This phenomenon can be understood by the features of interactive and interactional metadiscourse to be elaborated in subsection 2.1.4.

Another related concept to voice(s) and authorial identity(ies) is *stance*. In fact, there is much overlap in the use of the terms *stance* and *voice*. Hewings (2012, p. 188) offers an overview of the issues raised in these areas: “*stance*: research which both focuses on the textual characteristics of writing involved in persuasion, evaluation, and judgements; *voice*: research which takes a wider perspective and in which the construction of writer identity is foregrounded.” According to the author, stance would focus on persuasion, evaluation, and judgements, while voice on writer identity.

Stance is one among many other linguistic mechanisms used to address writers’ personal feelings and assessments but has long been referred to by a myriad of labels, including “‘evaluation’; ‘intensity’; ‘affect’; ‘evidentiality’; ‘hedging’; ‘modality’ and ‘stance’” (Biber, 2006, p. 87). The expression of stance is very important in academia, for “in claiming a right to be heard, and to have their work taken seriously, writers must simultaneously take a stance towards what they discuss and display competence as disciplinary insiders.” (Hyland; Jiang, 2019, p. 131). Stance is expressed in many ways, including grammatical devices, word choice, and paralinguistic devices to convey evaluation, affect, certainty, among other meanings, as will be introduced in subsection 2.1.4.

Stance is also involved with the issue of (im)personality in academic writing. Elbow (2007, p. 13) states that “scientists, for example, often write their research and scholarly arguments in a way that seeks to be impersonal.” In addition,

The traditional ‘scholarly’ paper has been written as if by an automaton: no hint of person or personality is allowed to emerge. This style is at its most obvious in the worlds of science and medicine, and, oddly enough, history. It is at its best an attempt to be ‘objective’ – that is, unbiased. At its worst it is an attempt to conceal subjectivity and bias. (Devine, 2021, p. 34).

However, some claim that there is no such way scholars may be impersonal in their texts, because occasionally “[...] speakers and writers in university registers seem more concerned with the expression of stance than with the communication of ‘facts’.” (Biber, 2006, p. 87). Therefore, “a ‘scientific’ paper, which endeavours to suppress all trace of personality or idiosyncrasy, is still taking on the persona of ‘scientific narrator’.” (Devine, 2021, p. 31).

To this end, individuals employ stance markers with the intention to “[...] express [their] personal feelings, attitudes, value judgments, or assessments [...]” towards what is being said or written or to another proposition available in the discourse (Biber *et al.*, 2021, p. 958). This

way of expression is similarly related to author's authority¹⁰ or sense of inferiority¹¹ in texts (Ivanič, 1998). Hyland (2002b) advocates that such authoritativeness serves to gain credibility in academia.

As has been seen in this subsection, writer's style in academic texts can reveal their voice(s), authorial identity(ies), and stance(s). These three concepts may be seen as overlapping, however, they complement each other. Take, for example, the following case: when authors express levels of (un)certainty or yet evaluation towards text propositions, not only are they taking a stance toward what is being written, but also disclosing their authorial persona (or voice), and yet confident, authoritative authorial identities. Style would thus result from authors' consistent language choices (through adverbs, adjectives, etc.) that disclose their stance, voice, and authorial identity expression. Nevertheless, one should not disregard the fact that such a style is dependent on social regulations, as will be discussed in the next subsection.

2.1.3 Style as sensitive to social context(s): the case of academic and linguistic norms and conventions

As briefly introduced, writer's style in academia represents individual variation and can reveal their voice(s), authorial identity(ies), and stance(s). Consistent linguistic choices that form scholar's repertoire, however, do not denote mere standalone decisions on the part of writers, because they may have been conditioned by outside norms. As a matter of fact,

the writer's 'choice' of available alternate forms is often determined by external conditions and then becomes the conscious, semiconscious, subconscious, or (usually) unconscious result of habitually using one form instead of another. (McMenamin, 2002, p. 177).

Such external conditions are understood here as the norms that influence writer's choices, like academic and linguistic norms, which, in turn, end up shaping writer's repertoire of language patterns that become available for prompt use in (sub-/semi-/un-)conscious ways

¹⁰ Devine (2021, p. 35) states that "the word 'author' is historically bound up with the word 'authority', and perhaps cannot ever completely escape that origin."

¹¹ Ivanič (1998, p. 88) distinguishes authority from inferiority as follows: "In the case of writing within the university, that is the identity of a person with authority. This is a crucial insight, because the one thing that characterizes most of the writers I worked with was a sense of *inferiority*, a lack of confidence in themselves, a sense of powerlessness, a view of themselves as people without knowledge, and hence without authority. For some, this was the legacy of a working-class background. For others, it was associated with age or gender; for all, it was associated with previous failure in the education system and an uncertainty as to whether they had the right to be members of the academic community at all. On the other hand, there are some who bring authority of different types into the academic institution from different domains, such as business, local politics or parenthood: authority which often goes unrecognized by the academic community". (Ivanič, 1998, p. 88).

over time. There are academic norms that may condition writer's stylistic choices, and they are better understood by looking at the bivalent concept of social context: "[...] the immediate interaction between real individuals, and the norms of the cultures in which they are operating" (Ivanič, 1998, p. 77). In other words, whereas the first aspect involves the context of the situation (social interaction), concerning the interaction among academic members (i.e., writers, peers, proofreaders, editors, peer reviewers, publishers, readers, etc.), the second considers the context of culture (social construction), regarding major academic disciplinary and community conventions to which academic members respond. Let us begin with the last aspect.

Academic discourse "[...] constructs the social roles and relationships which create academics and students and which sustain the universities, the disciplines, and the creation of knowledge itself." (Hyland, 2009, p. 1). As part of this broader context of culture where scholars write, academia is seen as an enterprise nowadays, because of the entrepreneurial, business nature of research and teaching, aligned with competition, productivity, and hierarchy principles (Hyland, 2009). One of these aspects involves paper publication, which is necessary for new discoveries, insights, and inventions to have significance in the academic realm.

As a matter of fact, "[...] academic publication is now an enormous industry that dominates the professional lives of academics across the globe" (Hyland, 2022, p. 15). This scenario comprises three major companies (e.g., Reed Elsevier, Springer, and John Wiley) that account for over 40% of all articles published (Hyland, 2022), 5.5 million researchers, 2,000 publishers, and 17,500 research/higher education institutions (Lillis; Curry, 2010). There are at least four reasons for this growth: "[...] i) technological change, ii) the emergence of open access business models, iii) an increase in active researchers, and iv) the importance of publishing to academic careers." (Hyland; Jiang, 2019, p. 5).

Hyland (2022) states that academia has surrendered to capitalism, globalization, and commercialization forces, where "funding bodies, governments, and cash-conscious universities are increasingly insistent on involvement with the private sector and in applied work" (Hyland, 2022, p. 18). A new efficiency culture is thriving: productivity is measured on the basis of the amount of papers scholars publish and the number of citations they receive. English has even been chosen as the international language of science (Hyland, 2009; 2022; Hyland; Jiang, 2019; Lillis; Curry, 2010; Wood, 2001). In this context, quantitative productivity is synonymous to good and proper knowledge and seen as the main, if not the only, reason for grant funding and career advancement. Nonetheless, three major problems arise from this

embellished publication practice, including plagiarism, salami slicing of studies, and paper retractions (Hyland, 2022). Furthermore,

the establishment of an institutional appraisal culture, by focusing on the number of papers published by individuals and universities, discourages the sharing of resources and ideas and replaces it with a highly competitive environment where each individual or research group is working for him- or herself. (Hyland, 2022, p. 22).

This trend in paper publication also illustrates how academic knowledge is created nowadays. “To a large extent, academic discourse has evolved as a means of funding, constructing and evaluating knowledge.” (Hyland, 2009, p. 10). Knowledge takes a strenuous path to get from researchers to the audience, for it needs to abide by academic conventions inherent to the publication realm:

This involves a long process of convincing editors, reviewers and peers to accept a claim as interesting or valid, drawing on approved and familiar discourses to do so. A view must be framed within a context of what is already accepted and using an argument carefully crafted for a particular audience. Ultimately a theory prevails because it is presented in a way which academics recognize as persuasive: knowledge, in other words, is what people can be persuaded to accept [...] (Hyland, 2009, p. 2).

As can be seen, academic knowledge results from the situations in which it is created and is dependent on disciplinary argument, affiliation, and agreement-making. This assumption illustrates how scholar’s decisions subscribe to community’s discourse conventions (Wood, 2001) and that “[...] it is this intellectual climate which determines the problems they investigate, the methods they employ, the results they see and the ways they write them up” (Hyland, 2009, p. 12).

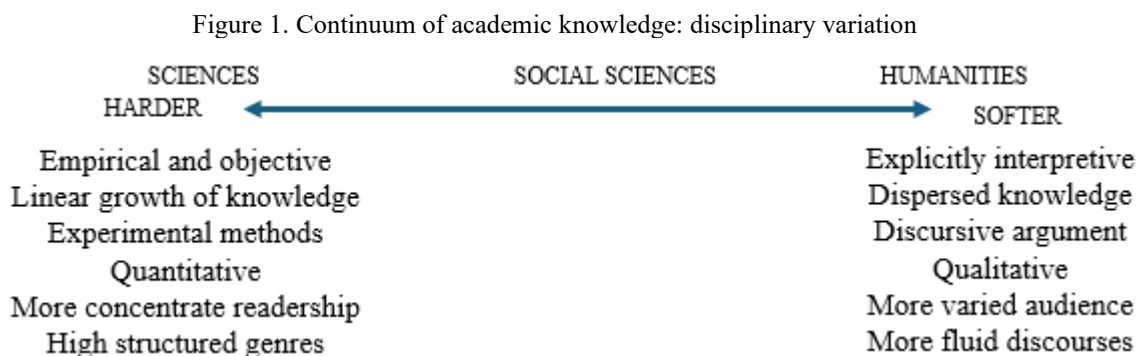
Discourse communities can be macro, like the area or discipline itself, or micro, such as research groups in a department. As Ivanič (1998, p. 81) recalls, “[...] each academic community will have its own, socially constructed norms and conventions for writing [...]”, although they are neither universal nor static and are partial to change. For instance, “the humanities, like history and philosophy, [...] employ abstraction rather than technicality, moving from instances to generalizations by gradually shifting away from particular contexts.” (Hyland, 2009, p. 8). Although there are common practices across fields in academia, like “[...] acknowledging sources, rigorous testing, intellectual honesty, [...] there are differences which are likely to be more significant than such broad similarities” (Hyland, 2004, p. 11).

These rhetorical conventions allocate disciplines in more precise domains. This division, however, is not bereft of problems, since it can be seen as antithetical, reductionist, dichotomic, or ideologically loaded (Hyland, 2005), especially owing to the almost dissolution of scholarship boundaries between (especially different) disciplines nowadays, or yet to

historical, geographic and other social changes they undergo over time (Hyland, 2009). Nonetheless, there is a way to understand similarities and differences between disciplines that participate in discourse domains:

Obviously such broad distinctions cannot capture the full complexity of disciplinary differences, and may only be acceptable at a general level of analysis, but they do provide a useful basis for identifying dimensions of variability between these fields. If the hard-soft distinction is conceived as a continuum rather than as unidimensional scales, I believe it offers a convenient way of examining general similarities and differences between fields without positing rigidly demarcated categories. (Hyland, 2005, p. 30).

In this regard, the author proposes a way to understand disciplines that belong to soft and hard discourse domains through a continuum scheme of academic knowledge: “by seeing the hard-soft distinction as a continuum, rather than as unidimensional scales, it can offer a useful way of examining contrasts while avoiding rigidly demarcated categories.” (Hyland, 2009, p. 64). The continuum of academic knowledge, proposed in Hyland (2009, p. 63), is illustrated in Figure 1:



Source: Adapted from Hyland (2009, p. 63).

We can see in the image that the two major discourse domains have particular features that illustrate how texts in their disciplines convey knowledge. Nevertheless, this continuum helps us see that disciplines from a particular discourse domain may also share characteristics that are promptly associated with disciplines from an opposite domain, and vice-versa. For instance, texts in harder disciplines may be similarly interpretive, while texts in softer disciplines may be equally empirical. This assumption also seems valid for social sciences, which can mutually share features from both poles. Another example comes from Prina Dutra and Berber Sardinha (2021), who discovered that the dimension of interpretive elaboration is what mostly characterizes RAs in applied linguistics as opposed to RAs in chemistry and biology. However, depending on the methodology used, RAs in applied linguistics were similarly quantitative, and not only qualitative.

As can be seen, these socially agreed conventions connect texts with disciplines, and this brings us to the academic genre that has long received much attention and prestige in the scholarship (Hyland, 2009): research article. In fact, “genres and disciplines interact in an interesting and dynamic manner, especially in academic discourses” (Bhatia, 2002, p. 31). Genres are

[...] essentially defined in terms of the use of language in conventionalised communicative settings, which give expression to a specific set of communicative goals of specialised disciplinary and social groups, which in turn establish relatively stable structural forms and, to some extent, even constrain the use of lexicogrammatical resources. Often such constraints can also be attributed to variations in disciplinary practices. (Bhatia, 2002, p. 23).

In this regard, the greatest boost of today’s academia marketplace lies in the publication of RAs in specialized, peer-reviewed journals. As a result, “[...] the massive increase in the number of journals have encouraged a shift away from book publishing, a relentless drive toward ever greater specialization, the strengthening of journal status hierarchies [...]” (Hyland, 2022, p. 17).

In fact, the research article (RA) is at the foundation of scientific knowledge and has been analyzed from different perspectives, including the Introduction – Method – Result – Discussion (IMRD) rhetorical structure proposed by Swales (1990). The author offers a Create-a-Research-Space (CARS) model with three moves that represent a number of characteristics of RA introductions. Move, in genre analysis, “[...] is a discoursal or rhetorical unit that performs a coherent communicative function in a written or spoken discourse (Swales, 2004, p. 228), which is mainly a functional unit, but can be grammatically realized by clauses and sentences. These moves reveal:

the need to re-establish in the eyes of the discourse community the significance of the research field itself; the need to situate the actual research in terms of that significance; and the need to show how this niche in the wider ecosystem will be occupied and defended. (Swales, 1990, p. 142).

In addition, Method shows more variation and indeterminacy in structure, not often following the classic Given-New paradigm. Disciplinary variation may help illustrate such differences. For example, hard sciences favor explicitness by providing careful step-by-step description, rooted in anaphoric reference and lexical repetition. Differently, physical and life sciences tend to be “[...] enigmatic, swift, presumptive of background knowledge, not designed for easy replication, and with little statement of rationale or discussion of the choices made” (Swales, 1990, p. 170).

At last, Results and Discussions can be substituted by other sections often labeled as Conclusions, Implications, or Applications. Many of them ambiguously report on the same issue, with Results describing results while Discussions redescrives them. In comparison to Introductions, Discussions moved “[...] from stating the results themselves, to placing them within the established literature, to reviewing their general significance.” (Swales, 1990, p. 173).

The IMRD rhetorical structure has received great attention in academia and has been revised in another text (Swales, 2004), with some changes, for instance, in the move structure applied to RA introduction analysis. The author also recognizes that, although RA is a genre with slow changes, it should now be seen as a macro genre encompassing theoretical and experimental papers, including theory pieces, review articles, experimental or data-based RAs, and shorter communications (i.e., squibs, forum pieces, shorter articles, conference preprints, notes, etc.).

From a register-perspective, where an analysis of linguistic characteristics common in a text variety is combined with an analysis of the situation where this variety is used (Biber; Conrad, 2019), RAs are deemed different from other registers (and genres). They

present more focused, specialized information about new, developing knowledge in the field to expert readers. [...] differ in terms of the discipline within which they fall, the journals they occur in, who their authors are, the types of methods that are used to conduct the research being reported on [...] (Gray, 2015, p. 7).

Gray (2015) discovers that published journal writing has three ‘meta’ article types and proposes the following taxonomy as seen in Table 2:

Table 2. Taxonomy of academic articles published in journals

Taxonomy	Features
Empirical	- analyze some type of observed data; - include quantitative, qualitative, and mixed methods research; a) quantitative: include experimental/quasi-experimental and observational research; b) qualitative: analyze any observed data (not quantitative); include ethnographies, case studies, focus groups, interviews, and field observations; c) mixed: use a nearly equal focus on quantitative and qualitative data analysis.
Theoretical	- discuss matters of theory within the discipline; - do not analyze any empirical data; - focus on explaining and extending key premises in the discipline. - involve three subtypes: author interpretation, logic-based, and general theoretical.
Evaluative	- offer critique or summary of the state of the field, a particular article, book or product; - involve three main types: book or product review, commentary/forum, review.

Source: Adapted from Gray (2015, p. 32-33).

RAs can also be analyzed from a register-perspective based on the elements that constitute its situational context, including participants, relations among participants, channel,

processing circumstances, setting, communicative purposes, and topic (Biber; Conrad, 2019, p. 40). For RAs, Gray (2015) proposes the following framework for describing the situational characteristics of academic journal registers as seen in Table 3:

Table 3. Framework for describing the situational characteristics of academic journal registers

Major elements	Features	Sub-features
1. Participants	A. Writer	1 (single); 2–4 (small group); 5+ (large group)
2. Textual layout & organization	A. Length	mean page length; standard deviation of page length; page length range
	B. Headings	none; un-numbered; numbered
	C. Use of abstracts	yes / no
	D. Visual elements	none; tables; figures; equations
	E. Sections / organization	IMRD; IMRD with varied order; other; standardized section headings; varying section headings/names
3. Setting	A. Nature of journal	generalist; specialized
4. Subject/Topic	A. General topic of the discipline	varied (based on informal survey of a range of sources)
5. Purpose	A. General academic purpose	varied (based on initial register survey)
6. Nature of data or evidence	A. Presence of observed data	yes / no
	B. Use of numerical data	yes / no
	C. Primary presentation of evidence	extensive prose; quantitative displays; mathematical formulas
	D. Object of study	varied (based on trends in the corpus)
7. Methodology	A. General method type	observational; experimental; n/a (for theoretical articles)
	B. Statistical techniques	n/a (for theoretical and qualitative articles); descriptive statistics; statistical difference testing; other advanced statistics
8. Explicitness of research design	A. Explicitness of purpose	direct statement; indirect; no discernible statement
	B. Explicitness of research questions	direct statement; indirect; no discernible statement
	C. Explicitness of citations	within the text; in footnotes/endnotes
	D. Explanation of evidence	extensive; mention; no discernible statement
	E. Explanation of procedures	extensive; mention; no discernible statement

Source: Adapted from Gray (2015, p. 56-57).

Furthermore, the prominence of RAs in the academic realm is also observed by two other aspects, literacy brokering (peer reviewing) and knowledge establishment (Hyland, 2009), which mutually affect each other. Literacy brokering refers to

[...] all the different kinds of direct intervention by different people, other than named authors, in the production of texts and to track what is actually involved [...] We use the term literacy brokering rather than other less ideologically loaded notions (mediation, sponsoring, shaping) as a way of explicitly acknowledging that intervention in academic texts is not a neutral enterprise, but rather involves participants of unequal status and power. (Lillis; Curry, 2010, p. 88).

Brokering activity is more frequent in journal articles than in other academic genres. This task is carried out by literacy brokers, who are people who significantly impact academic texts, including friends, editors, peer reviewers, academic peers, and translators, illustrating the interaction and cooperation among members of what Ivanič (1998) defines as the context of the situation. As Swales (2004, p. 218) advocates:

The standard RA is, of course, in its final form a *product*, but that product is patently the outcome of a complex *process*. A manuscript is most often slowly gestated, undergoes several drafts, and receives multiple inputs from coauthors, colleagues, reviewers, and editors.

According to Lillis and Curry (2010, p. 93), there are two major types of literacy brokers as seen in Table 4:

Table 4. Literacy brokers

Language brokers	Academic brokers
- demonstrate knowledge and expertise of the English language; - include professional language brokers (have professional backgrounds and qualifications), such as translators, proofreaders and English-language specialists (English teachers); - include informal language brokers (have personal relationship with the author), such as friends, spouses, other family members; - represent 17% of all brokers reported.	- work at universities or research institutes; - can be subclassified into three types: a) general academic: is not from the same disciplinary area as the author(s); b) disciplinary expert: shares the same disciplinary background and interests as the author(s); c) sub-disciplinary specialist: is from the same specialist field as the author(s). - can also include people according to their role/function in the chronological steps of a text production: a) pre-submission: academic colleague; b) post-submission: journal peer reviewer or editor; - represent 83% of all brokers reported.

Source: Adapted from Lillis and Curry (2010, p. 93).

Peer reviewers are a type of academic brokers (or gatekeepers) and they are responsible for “[...] the process by which research output is subjected to scrutiny and critical assessment by individuals who are experts in those areas” (Hames, 2012, p. 16). This process ensures paper quality and determines, from a big pool of submissions, papers whose research has been well planned, carried out, and written up (Paltridge, 2017). In addition, RA prominence is also attributed to

[...] the prestige attached to a genre which restructures the processes of thought and research it describes to establish a discourse for scientific fact-creation. Language becomes a form of technology which attempts to present interpretations and position participants in particular ways as a means of establishing knowledge. (Hyland, 2009, p. 68).

In this regard, the RAs writers produce reveal “[...] the ways [they] present their topics, signal their allegiances, and stake their claims represent[ing] careful negotiations with, and

considerations of, their colleagues.” (Hyland, 2004, p. 1). These cultural practices affect writers, readers, and the communities of which they are part:

The ways that writers chose to represent themselves, their readers and their world, how they seek to advance knowledge, how they maintain the authority of their discipline and the processes whereby they establish what is to be accepted as substantiated truth, a useful contribution and a valid argument are all culturally-influenced practical actions and matters for community agreement. (Hyland, 2004, p. 11).

From the process of RA submission to publication, many are the steps involved:

This process commences when an author submits an article to the journal. The journal editor will send it out for review, ask for revisions to be made to the article then send it out for review, or reject the article without sending it out for review (a ‘desk reject’). If the article goes out for review, the reviewers will be asked to recommend publishing it as is (this is extremely rare), accepting the article with minor amendments (this is also rare), accepting it with major amendments (this is more common), revise and resubmit the article and be reviewed again, or reject. A paper can sometimes go through several rounds of reviews until it is published and involve a number of communications among editors, authors, and reviewers as it moves through the various stages of the review process. (Paltridge, 2017, p. 23).

This long, strenuous process is not unproblematic. On the one hand, it can take academia to be seen as “[...] a fair, consistent, and impartial enterprise” (Hyland, 2022, p. 18) and influence scholars’ appointments, promotions, and grant funding (Hyland, 2020). On the other hand, it can also involve several delays (long time from submission to publication), bias (reviewers may ask authors to modify analyses), disrespect (reviewers may be rude to authors in the comments), and skullduggery (reviewers may favor their friends or themselves, review their peer’s work unfairly, or worse, take advantage of ideas from the papers) (Hyland, 2020). Braine (2003, p. 105) gives his testimony about these issues:

The world of academic publications is often baffling and frustrating. Its inhabitants include unsympathetic and authoritarian editors, and reviewers who can be crass and infuriatingly demanding. They will sometimes request the impossible – that the research design be changed, that more subjects be added, or that the scope be extended – from an article written well after the research had been completed.

One of the issues mentioned by Braine (2003) was the need to make textual changes. In the process of allegiance to community and disciplinary academic discourse practices, texts undergo structural changes, including additions, deletions, reformulation, sentence-level changes, adherence to publishing conventions, among others (Lillis; Curry, 2010, p. 89-91). Consequently, “[...] texts don’t exist in isolation but are part of the communicative routines of social communities.” (Hyland, 2009, p. 10).

These changes respond to community and disciplinary academic discourse practices, which shape much of how scholars write in academia, because “[...] the rhetorical conventions

of each text will reflect something of the epistemological and social assumptions of the author's disciplinary culture" (Hyland, 2004, p. 11). In fact,

Writing as a member of a discipline involves crafting texts in a way that insiders can see as 'doing biology' or 'doing sociology' and this both restricts how something can be said and authorizes the writer as someone competent to say it. [...] students [and expert writers] learn what counts as good writing through an understanding of their discipline and the conventions and genres regarded as effective means for representing knowledge in that discipline. (Hyland, 2009, p. 6).

This practice stems from a broader understanding that language has a tripartite function (Halliday; Matthiessen, 2014), consisting of three connected metafunctions (ideational, interpersonal, textual), seen "[...] as a 'social semiotics', that is, a model of linguistic meaning-making as a social phenomenon influencing and influenced by the context in which it occurs" (Nørgaard; Busse; Montoro, 2010, p. 3). In this regard, language shows variation according to the context where it is used, and thus scholars should beware of such norms while writing. For example, there are two major kinds of linguistic norms, descriptive and prescriptive, which McMenamin (2002) defines as:

A descriptive norm reflects speakers' and writers' knowledge of what is acceptable or grammatical in language form and function. A prescriptive norm is the social sense of what in language use is correct and appropriate. Linguistic norms are not static; they evolve over time in a social, cultural, and geographic community of speakers and writers. Linguistic norms are implicit in language use and in the perceptions of speakers and writers and are often explicitly described in dictionaries and grammars. (McMenamin, 2002, p. 126).

Within the scope of corpus-based descriptive linguistics and from a register-perspective (Biber; Conrad, 2019), academic prose is deemed to have a distinguished register of its own in comparison to conversation, fiction, and newspaper (Biber, 2006; Gray, 2015). In this regard, the *Longman Grammar of Spoken and Written English* (Biber *et al.*, 1999) identified prevailing grammatical features of books and RAs from a number of academic disciplines, including biology, chemistry, medicine, sociology, education, and law/history/politics. For instance, three word classes are prevalent: nouns, adjectives, and prepositions. Also, texts show greater use of "[...] nouns as pre-modifiers of a head noun, or appositive noun phrases post-modifying a head noun" (Biber; Gray, 2016, p. 26). Therefore, grammatical complexity is phrasal rather than clausal, making structures inexplicit in meaning (Biber; Gray, 2016; Biber, 2023). All in all, other features are summarized in Table 5:

Table 5. Distinguishing grammatical features in academic prose

Major features	Commonly (under-)used sub-features
Nouns, noun phrases, pronouns, determiners	absence of pronouns; specific pronouns: <i>this</i> and generic <i>one</i> ; plural nouns; nominalizations; anaphoric expressions; definite article <i>the</i> ; demonstrative determiners; noun phrases with modifiers; noun phrases with pre-modifiers; nouns as pre-modifiers; noun phrases with post-modifiers; noun phrases with multiple post-modifiers; noun <i>and/or</i> noun binomial phrases.
Adjectives and adjective phrases	attributive adjectives; specific predicative adjectives; derived adjectives.
Verbs and verb phrases	copula <i>be</i> and copular verb <i>become</i> ; “existence” verbs; specific lexical verbs; specific prepositional verbs; verbs with inanimate subjects; derived verbs; simple aspect; present tense verb phrases; passive voice; specific passive verbs
Adverbs and adverbials	specific adverbs; specific amplifiers; specific degree adverbs; linking adverbials; purpose and concessive adverbials.
Dependent clause features	relative clauses with the relative pronoun <i>which</i> ; participle clauses as post-modifiers in noun; noun complement clauses as post-modifiers in noun phrases; abstract noun + <i>of</i> + <i>ing</i> -clause; extraposed <i>that</i> -clauses; extraposed <i>to</i> -clauses; subject predicative <i>to</i> -clause; <i>ing</i> -clauses controlled by adjective predicates; concessive adverbial clauses.
Other features	<i>of</i> -phrases; prepositional phrases as post-modifiers in noun phrases; stance noun + <i>of</i> -phrase; <i>that/those</i> + <i>of</i> -phrase; preposition + <i>which</i> in relative clauses with adverbial gaps; selected coordination tags: <i>and so on, etc.</i> ; quantifier <i>each</i> ; semi-determiners <i>same, other, certain, and such</i> ; dual gender reference: <i>he or she, his or her, he/she</i> ; lexical bundles with noun phrases and/or prepositional phrases.

Source: Adapted from Biber *et al.* (1999); Biber (2006, p. 16-18).

When considering grammatical features that are mostly characteristic to RAs, Hyland (2021, p. 38) lists some predominant patterns:

[...] research articles are now more informationally focused, increasingly contain present tense, provide more explicit in-text reference, and use less abstract language. Writers are giving greater attention to cohesion with both more cases of demonstrative *this* and with more of these structures containing an attending noun to help readers follow the thread. They are also citing massively more often and giving less prominence to those they cite, even if they are citing themselves, with more references to co-authors. [...] academic writing is becoming more uniform and less formulaic in its use of lexical bundles as while both the range and frequency of bundles have risen, variation in their use and the proportion they comprise of total words have declined. In terms of interaction, explicit markers of stance and authorial attitude has declined although authorial self-mention has massively increased, particularly the use of exclusive *we*. Explicit engagement with readers has also dropped significantly, especially in the soft fields, and we could find no evidence of a significant rise in ‘informality’¹² beyond an increase in the use of authorial self-mention.

¹² According to Hyland and Jiang (2019, p. 209), “informality in academic writing is the expression of a more personal tenor and this implies a closer relationship to readers, a willingness to negotiate claims and a positive attitude towards subjectivity. It also indicates the use of certain rhetorical characteristics which can be identified and counted, allowing us to measure changes in academic argument.” Informality can be realized by a set of features, including first-person and second-person pronouns, unattended anaphoric pronouns, split infinitives, sentence initial conjunctions or conjunctive adverbs, sentence final preposition, listing expressions, contractions, direct questions, and exclamations (Hyland; Jiang, 2019, p. 213).

Furthermore, RAs in applied linguistics also have a distinguished register, as summarized in Table 6:

Table 6. Distinguishing grammatical features in applied linguistics RAs

Major features	Commonly used sub-features
Nouns	process nouns; nouns as nominal premodifiers
Verbs	finite verb complement clauses; <i>that</i> -clause (factive verbs); to-clauses (verbs of desire); passive voice (agentless); activity verbs; mental verbs
Adverbs	finite adverbials
Other clauses types	finite relative clauses; non-finite complement clauses
Other features	academic involvement and elaboration features; informational density features; overt empiricism features; human focus; balanced use of present and past tense

Source: Adapted from Gray (2015, p. 172; 174).

These language patterns illustrate some of the conventions inherent to academic texts in a particular discipline, resulting from larger practices of how scholars commonly write and in turn how they should employ particular language features to abide by linguistic and academic writing disciplinary conventions. As has been discussed in this subsection, writer's stylistic individual variation in academia is also conditioned by linguistic and academic norms. As constituents of discourse academic communities and domains, writers need to grapple with available social norms, principles, and beliefs that may equally shape their style. This is not to say, however, that they are static and unchangeable. As McMenamin (2002) stated, all norms are subjected to shift over time to adequate to the communicative needs of language users. In the following subsection, we will see how writer's stylistic choices can be used to express purposes and effects in texts, particularly those which foster the creation of interpersonal bonds with audiences by means of interactive and interactional metadiscourse language features, which include some of the prevailing linguistic features of academic prose and RA registers just presented (Biber *et al.*, 1999; Biber, 2006; Gray, 2015; Hyland, 2021).

2.1.4 Style as writer's expression of purposes and effects: the case of metadiscourse

When scholars write academic texts, they opt for particular language items and structures to share with the academic community research-related facts, information, data, viewpoints, among other meanings, as discussed in the previous subsections. Scholars' style similarly reveals their intention to reach particular purposes and cause potential effects in texts and audiences (Leech; Short, 2007). In a seminal work, Bell (2001) claims that authors shift their style to acquiesce to audience needs:

1. Style is what an individual speaker does with a language in relation to other people. [...]

2. Speakers design their style primarily for and in response to their audience. [...]
3. Speakers have a fine-grained ability to design their style for a range of different addressees, as well as for other audience members.
4. Style-shifting according to topic or setting derives its meaning and direction of shift from the underlying association of topics or settings with typical audience members. [...] (Bell, 2001, p. 141-148).

Similarly, as elaborated in subsection 2.1.2, voice and stance illustrate stylistic choices that also exert influence on readers:

Their importance lies in the role they play in negotiating meanings and agreement with interlocutors and the impression of a writer they make on those readers. The realization of stance and voice reveals a diversity and vibrancy shaped by conformity to community conventions and the expression of personal proclivities: they ‘read distinctive’ because choices represent individual preferences accountable to disciplinary practices. (Guinda; Hyland, 2012, p. 19).

Many are the language items and structures authors can use to reach purposes and effects in texts and ultimately characterize their style, including a myriad of “[...] phonological, morphological, syntactic, lexical, semantic, discursive resources of the language [are used] to express, orally or in writing, thoughts, feelings, opinions etc.”¹³ (Henriques, 2011, p. 27). Leech and Short (2007) have proposed a checklist of linguistic categories for stylistic analysis applied to fictional texts, but which might similarly be used in other contexts, as can be seen in Table 7:

Table 7. Checklist of linguistic categories for fiction style analysis

Categories	Language (sub-)features
Lexical categories	General (formal / informal vocabulary; idiomatic expressions or notable expressions; rare / specialized vocabulary etc.)
	Nouns (abstract / concrete; proper names; collective nouns etc.)
	Adjectives (attributive / predicative; gradable / non-gradable; used to describe physical, psychological, visual attributions etc.)
	Verbs (intransitive / transitive; static / dynamic; used to refer to states, perceptions, speech acts etc.)
	Adverbs (of mode, place, direction, degree; conjuncts / disjuncts etc.)
Grammatical categories	Sentence types (declarative / interrogative / exclamatory etc.)
	Sentence complexity (simple / complex structure; average sentence length; dependent / independent etc.)
	Clause types (relative, adverbial; finite / non-finite; that-clauses / wh-clauses / -ing clauses / -ed clauses etc.)
	Clause structure (position of elements: initial / median / final; frequency of use of elements such as objects, complements, adverbs etc.)
	Noun phrases (simple / complex; most complex part: pre-modifying and post-modifying elements etc.)
	Verb phrases (use of phrasal verbs; particular use of present / past tense and perfect / progressive aspect etc.)
	Other phrases types (prepositional / adverb / adjective etc.)

¹³ “[...] recursos fonológicos, morfológicos, sintáticos, lexicais, semânticos, discursivos da língua para expressar, oralmente ou por escrito, pensamentos, sentimentos, opiniões, etc.” (Henriques, 2011, p. 27).

	Word classes (more specific uses of prepositions, conjunctions, pronouns, interjections etc.)
	Other grammatical constructions (comparative / superlative / parenthetical / coordinative constructions etc.)
Figures of speech	Grammatical and lexical (anaphora, parallelism, antithesis; etc.)
	Phonological schemes (rhythm, alliteration, assonance, etc.)
	Tropes (deviations from the standard norm and their effects; neologisms; similes; etc.)
Context and cohesion	Context: external relations of a text or a part of a text (social relation between its participants, including author-reader, character-character; participants' sharing of knowledge and assumptions; etc.).
	Cohesion: ways in which one part of a text is linked to another (the ways in which sentences are connected; the internal organization of the text).

Source: Adapted from Leech and Short (2007, p. 61-64).

As can be seen, Table 7 has just introduced a comprehensive list of elements for style analysis, but not all these language features are contemplated in this doctoral research. As will be analyzed in Chapter 04, only features that create interpersonal bonds with audiences by means of interactive and interactional metadiscourse mechanisms will be considered. These language features are introduced in APPENDICES A and B and are part of what we hereby call the *Stylistic Metadiscourse Model of Interpersonal Interaction in Academic Writing*.

This model is not entirely new. In fact, it is a combination of Hyland's (2005, p. 48-54) *interpersonal model of metadiscourse*, Hyland's (2008b, p. 13-14) *taxonomy for bundles functional analysis*, Hyland's (2009, p. 74-78) *key resources of academic interaction* and, when feasible, Biber's (2006, p. 92-93) and Biber *et al.*'s (2021, p. 961-967) *lexico-grammatical and semantic features used for stance analyses*, with adaptations to suit our research aims. Very few changes have been made to Hyland's (2005; 2009) original model, as will be illustrated in Chapter 03. It should be stressed that Hyland's model has often been used in a variety of corpus-based studies (Pearson; Abdollahzadeh, 2023), thus highlighting its importance for this ongoing investigation.

As mentioned in Chapter 01, it must be made explicit that the reason for applying this model to our style analysis is basically twofold. Firstly, metadiscourse is a functional feature of academic writing and has long been used in different academic genres, like RAs (Hyland, 2004; 2005; 2009). In this regard, we intend to show how Biber makes variation within the norm, understood as the common practice of employing different interactive and interactional metadiscourse mechanisms. Secondly, it turns out that, after excluding text-aboutness words and other keywords, the remaining number of keywords identified in our searches were those that Hyland, Biber, and many other scholars have long investigated under the labels *metadiscourse* or *stance*. However, prior to presenting this model, it is necessary to become

familiarized with some theoretical notions on *metadiscourse* and *stance markers*, which are fundamental for understanding such a model to be presented in Chapter 03.

Like the term *style*, *metadiscourse* is also a fuzzy one, which has frequently been comprehended as either “discourse about discourse” or “talk about talk” (Hyland, 2005, p. 16), notions that only encompass aspects of texts themselves and ignore interpersonal elements. Broadly speaking, metadiscourse is defined as aspects of a particular text, such as language features, “[...] which explicitly refer to the organisation of the discourse or the writer’s stance towards either its content or the reader.” (Hyland, 2004, p. 109).

The term has been in use for more than seven decades to date. In fact, metadiscourse was coined by Zellig Harris in 1959, with the intent “to offer a way of understanding language in use, representing a writer’s or speaker’s attempts to guide a receiver’s perception of a text.” (Hyland, 2005, p. 3). Like the term *style*, which pays great homage to poetics and rhetoric, metadiscourse research has also been carried out in composition, reading, rhetoric, and text structure. Furthermore, much research has currently been executed in the fields of discourse analysis and language education (e.g., English as a Second Language, English as a Foreign Language, academic instruction, among others).

Similarly, Vande Kopple (1985) defines metadiscourse as some linguistic material which adds no propositional information to the text, but rather points to the author’s presence therein. Moreover, Crismore (1983, p. 2) refers to metadiscourse as “the author’s intrusion into the discourse, either explicitly or non-explicitly, to direct rather than inform, showing readers how to understand what is said and meant in the primary discourse and how to ‘take’ the author”. In this regard, recognizing that writing (and speaking) in academic milieu is a type of social engagement, Hyland (2005, p. 4) sees metadiscourse as a mediator between (con)text, reader, and writer, for

it illuminates some aspects of how we project ourselves into our discourses by signalling our attitude towards both the content and the audience of the text. [...], a writer is able not only to transform what might otherwise be a dry or difficult text into coherent, reader-friendly prose, but also to relate it to a given context and convey his or her personality, credibility, audience-sensitivity and relationship to the message. (Hyland, 2005, p. 4).

The absence or removal of metadiscourse-marking language features in a text may characterize it as impersonal, uninteresting, and difficult to comprehend on the part of the readers. Therefore, it is necessary that writers maintain such elements to help their reading audience. The concept of audience has equally been multifaceted in the area, which could mutually involve readers, peer reviewers, proofreaders, editors, among other people who

eventually interact with the text. Nonetheless, writing with an audience in mind entails looking at least at five parameters which underlie the author's task:

1. One parameter of reader influence on the writing is the number of persons who are expected to read the text. [...]
2. A second parameter of audience variation is the extent to which readers are known or unknown. [...]
3. Along a third parameter, that of status, the writing will vary according to whether the reader has a higher status, an equal status, or a lower status than the writer. [...]
4. As a fourth parameter, the extent of shared background knowledge will influence the writing to a considerable degree; that is, readers with a high degree of shared background knowledge are likely to influence the writing in particular ways. [...]
5. As a final parameter, the extent of specific topical knowledge shared by the reader and the writer will influence the writing. [...] (Grabe; Kaplan, 1996, p. 207-211).

As can be seen, writing with an audience in mind involves a variety of aspects, including the number of readers who are supposed to read the text, the extent to which readers are known or unknown to writers, the status of readers, and the degree of shared background and specific knowledge readers possess about the topic. With this in mind, writers may opt for stylistic choices that can include metadiscourse elements in their academic texts. For that, they should equally acknowledge the existence of three underlying key metadiscourse principles:

1. that metadiscourse is distinct from propositional aspects of discourse;
2. that metadiscourse refers to aspects of the text that embody writer-reader interactions;
3. that metadiscourse refers only to relations which are internal to the discourse. (Hyland; Tse, 2004, p. 159).

Let us understand each one of these aspects. The notion of *proposition* broadly concerns all types of knowledge and information that exists about external reality, such as thoughts, actors, states of affairs, etc. When compared with metadiscourse, Hyland (2005, p. 23) states that “the role of metadiscourse is therefore to signal the writer's communicative intent in presenting propositional matter”. In addition, not only does metadiscourse support such a propositional content, but it also becomes the “[...] means by which propositional content is made coherent, intelligible and persuasive to a particular audience.” (Hyland, 2005, p. 39).

As the nature of metadiscourse is essentially dialogic, the second key principle concerns the basic understanding that “[...] all metadiscourse is interpersonal in that it takes account of the reader's knowledge, textual experiences and processing needs and that it provides writers with an armoury of rhetorical appeals to achieve this” (Hyland; Tse, 2004, p. 161). Moreover, it involves two main types of interaction, namely *interactive* and *interactional*, terms whereby

Hyland (2005) pays homage to the works of Thompson (2001) and Thompson and Thetela (1995), essential to his model as we will see soon.

Finally, key principle 03 from the quote also stems slightly from the first principle, in that metadiscourse is concerned with what happens within the text. “An internal relation thus connects events in the account and is solely communicative, while an external relation refers to those situations themselves.” (Hyland; Tse, 2004, p. 165). The authors mention that the use of connectives, such as *therefore*, *in contrast*, *furthermore*, among others, may be used with both references. When used to connect steps in an exposition and organize the discourse as an argument, they are internal markers, but when used to connect activities in the world outside the text, they function as external, displaying experiences as a series of events.

As can be seen, metadiscourse, as part of the language discursial toolkit used by writers, presents information through the organization of the text itself at the same time as engages with readers as to how they should understand it. In this regard, Hyland (2005) proposes an *interpersonal model of metadiscourse*, based on a functional approach to language (Halliday; Matthiessen, 1999). This model encompasses two major dimensions, interactive and interactional. The *interactive dimension* regards the extent to which writers have empathy with their readers, better organizing the discourse so that they can fulfill the demands of the audience and make sure readers retrieve their objectives. The *interactional dimension* is aimed at writers’ direct, explicit intrusions along the discourse, making their epistemic and attitudinal stances explicit to the audience. This is a more dialogic connection between reader and writer, engaging the public in the discussion and inviting it to co-construct the texts. These dimensions present particular corresponding functional categories (interactive = *transitions*, *frame markers*, *endophoric markers*, *evidentials*, *code glosses*; interactional = *hedges*, *boosters*, *attitude markers*, *self-mentions*, *engagement markers*). In summary:

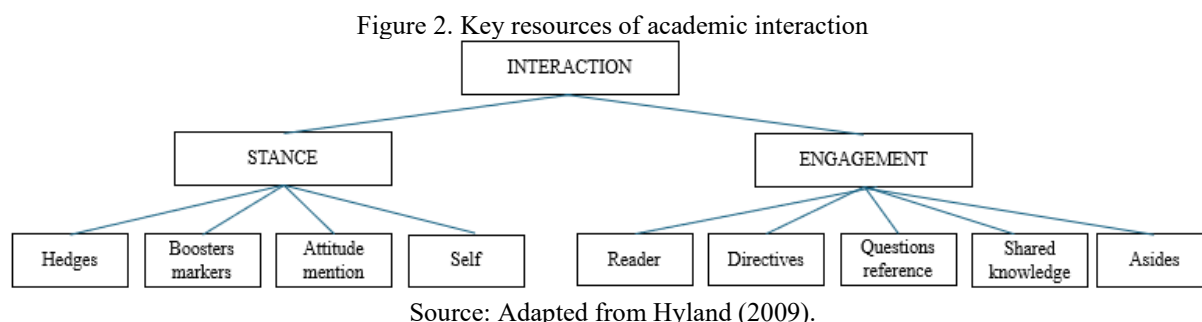
The interactive dimension. This concerns the writer’s awareness of a participating audience and the ways he or she seeks to accommodate its probable knowledge, interests, rhetorical expectations and processing abilities. The writer’s purpose here is to shape and constrain a text to meet the needs of particular readers, setting out arguments so that they [readers] will recover the writer’s preferred interpretations and goals. The use of resources in this category therefore addresses ways of organizing discourse, rather than experience, and reveals the extent to which the text is constructed with the readers’ needs in mind.

The interactional dimension. This concerns the ways writers conduct interaction by intruding and commenting on their message. The writer’s goal here is to make his or her views explicit and to involve readers by allowing them to respond to the unfolding text. This is the writer’s expression of a textual ‘voice’, or community-recognized personality, and includes the ways he or she conveys judgements and overtly aligns him- or herself with readers. Metadiscourse here is essentially evaluative and engaging, expressing solidarity, anticipating objections and responding to an

imagined dialogue with others. It reveals the extent to which the writer works to jointly construct the text with readers. (Hyland, 2005, p. 50-52, emphasis in original).

In this same study, Hyland identified patterns of disciplinary variation as far as the use of metadiscourse interactive and interactional features are concerned. In comparison to soft disciplines, hard sciences tended to produce more interactive texts, which confirmed the author's hypothesis that "[...] the sciences tend to produce more impersonal, or at least less reader-inclusive, texts." (Hyland, 2005, p. 144). These results also illustrated that writers in the humanities and social sciences often took more explicitly involved and personal stances than those in the hard sciences.

In another book, *Academic discourse*, Hyland (2009) revisits his earlier model and splits the interactional dimension into two, by making the differences between stance from engagement more explicit. Now, stance encompasses all the four previous elements (*hedges*, *boosters markers*, *attitude mention*, and *self*) from the interactional dimension, whereas engagement consists of the following: *reader*, *directives*, *questions reference*, *shared knowledge*, and *asides*. Based on Hyland (2009, p. 74), Figure 2 illustrates this new division:



As seen in Figure 2, Hyland (2009, p. 74) distinguishes stance from engagement, although both are still forms of interaction. For him, while stance is an attitudinal, writer-oriented resource, concerning writer's "[...] textual 'voice' or community recognized personality [...]" and "[...] the ways writers present themselves and convey their judgements, opinions and commitments", engagement is an alignment resource which regards "[...] the ways that writers rhetorically recognize the presence of their readers to actively pull them along with the argument, include them as discourse participants, and guide them to interpretations". (Hyland, 2009, p. 74).

In a more recent study, Hyland and Jiang (2019) discovered new trends in the use of some of these metadiscourse features. For example, the use of citations through *evidentials* in applied linguistics has increased 270%, hence illustrating how the discipline has grown from an emerging sub-field since the 1960s. The number of self-citations has also augmented in

sociology, applied linguistics, and electrical engineering, but plummeted in biology. Furthermore, as for stance features, there have interestingly been “[...] rises in hedges in engineering and biology, falls in the use of boosters by all disciplines but engineering, steady uses of attitude markers and, excepting applied linguistics, substantial rises in self-mention.” (Hyland; Jiang, 2019, p. 133). Moreover, as for engagement features, decreases have been noticed in applied linguistics and sociology, but increases in biology and electrical engineering. More specifically, asides and references to shared knowledge have dropped steadily in all four disciplines, especially in applied linguistics; questions and directives have remained stable, except in biology (questions have doubled) and electrical engineering (directives have increased); reference to readers has declined in all disciplines, except electrical engineering.

As briefly introduced in subsection 2.1.3 above, stance is one among many other language features used to express writers’ personal feelings, attitudes, assessments, value judgements, having been labeled differently in literature over time (Biber, 2006). Hyland, for instance, has adopted the term *hedges/hedging*, and has extensively published papers about the subject, such as an entire book only on the topic: *Hedging in scientific research articles* (Hyland, 1998a). On the other hand, Biber (2006) and Biber *et al.* (2021) have studied the same phenomenon under the term *stance*, considering it as a hypernym, which encompasses smaller categories realized in more specific lexico-grammatical features.

Biber (2006) and Biber *et al.* (2021) have found that grammatical devices, word choice, and paralinguistic devices mark stance. Adverbials and complement clauses controlled by verbs, adjectives, and nouns are the two most common grammatical ways to convey stance in discourse. On the one hand, while stance adverbials “express the attitude or assessment of the speaker/writer with respect to the proposition contained in the main clause”, such as in “*Unfortunately* it’s true”, complement clauses “with the main clause verb or adjectival predicate can express speaker stance with respect to the proposition in the complement clause”, like in “*I really doubt that* the check is there” (Biber *et al.*, 2021, p. 958, emphasis in original). In addition, word/lexical choice alone is also found to mark stance, as in “They’re very *nice*, cats are” (Biber *et al.*, 2021, p. 958). Adjectives (*good, lovely, nice*) and verbs in simple declarative forms (*like, love, want*) are often evaluative in nature. Specifically in academic prose, adjectives were commonly found as predicative (*difficult, important, likely, necessary, possible, true*) and attributive (*appropriate, good/best, important, practical, useful*).

In all cases observed, however, interpretation stemmed from the addressee’s ability to assess both context and shared background in which the stance lexical expressions were used. The authors complement: “such value-laden words differ from grammatical stance devices in

that they do not provide an attitudinal or evaluative frame for some other proposition.” (Biber *et al.*, 2021, p. 958).

Biber (2006) notes that paralinguistic devices vary according to the language modality: in speech, they consist of pitch, intensity, and duration, including modifications of articulation, like hissing or whispering. They are often accompanied by other non-linguistic features, such as body position, facial expressions, and gestures. In writing, instead, the devices used are more limited, with emphasis given to typographical styles through italics, bold face, or underlining.

Modals (including semi-modals) are the most common stance-marking grammatical device, showing different distributions across registers, but are more regularly used in conversation than in academic registers. In terms of frequency, (semi-)modals are then followed by complement clauses and adverbials, even in academic prose. Furthermore, it is also interesting to note that

[...] stance markers are prevalent in academic writing, especially given the general lack of first person involvement in that register. That is, authors of academic articles and books rarely refer to themselves overtly. However, it is not at all uncommon to find personal attitudes and estimates of likelihood expressed in academic writing through impersonal stance devices such as modal verbs, adverbials, and extraposed complement clauses. (Biber *et al.*, 2021, p. 972).

All in all, these grammatical devices accomplish at least three semantic functions, namely epistemic, attitudinal, and styles of speaking, as presented in Table 8:

Table 8. Major semantic categories used to express stance

Semantic categories	Definitions
Epistemic	convey writer's/speaker's certainty (or doubt), actuality, precision, limitation, or the source of knowledge or the perspective from which the information is given
Attitudinal	report writer's/speaker's personal attitudes, feelings, evaluations, judgements, emotions
Styles of speaking	present writer's/speaker's comments on the communication itself

Source: Adapted from Biber *et al.* (2021, p. 964-967).

Biber *et al.* (2021) call attention to the fact that the distinction between attitudes and emotions is generally fuzzy, since many language items can be used to convey both meanings. Moreover, attitudinal stance markers are more limited grammatically and appear with less frequency than epistemic markers.

Stance markers can also be identified as being attributed to the speaker or writer. Biber *et al.* (2021) state that many are overtly attributed, especially through first-person pronouns; other markers, instead, implicitly express stance without identifying the speaker or writer, with frequent use of modal verbs for this purpose; there are also some other markers which are ambiguous as to whether they are reporting the stance of the speaker/writer or of some third

party. It is interesting to add that Biber (2006) has excluded second-person and third-person pronouns from his study, although he recognized that such markers are used to express some kind of attitude or evaluation, but not necessarily reflect the speaker's or writer's personal stance. Differently, Hyland (2009) has included such features under the engagement category.

Among various forms available, formulaic language is another linguistic feature whereby writers may reveal their style, figuring as one of several style-marking lexicogrammatical features (Leech; Short, 2007). The understanding of the term *formulaic language* is rather ambiguous in academic literature, since it has commonly been referred to by different terms. For instance, Scott (2022) opts for the term *clusters*, which he defines thus:

Clusters are words which are found repeatedly together in each others' company, in sequence. They represent a tighter relationship than collocates, more like multi-word units or groups or phrases. (I call them *clusters* because *groups* and *phrases* already have uses in grammar and because simply being found together in software doesn't guarantee they are true multi-word *units*.) Language is phrasal and textual. [...] Words keep company: the extreme example is idiom where they're bound tightly to each other, but all words have a tendency to cluster together with some others. These clustering relations may involve colligation (e.g. the relationship between *depend* and *on*), collocation, and semantic prosody (the tendency for *cause* to come with negative effects such as *accident*, *trouble*, etc.).

Scott (2022) recognizes that the option for one term over another represents scholars' preference. However, he warns that issues like frequency thresholds, register, and genre are crucial to defining such multi-word units. Authors like Biber and Hyland, for instance, have preferred the term *lexical bundles*. In fact, Scott's definition of clusters is no different from Biber's and Hyland's concept of lexical bundles. Their main difference, however, lies in structural size: while the latter authors concentrate on word strings that range from three to six elements, Scott expands to include clusters that have only two elements (e.g., *once upon*). All in all, here's a summary of lexical bundles definitions:

I refer to these multi-word sequences as 'lexical bundles', defined simply as the most frequent recurring sequences of words. (Biber, 2006, p. 133).

Essentially, these are words which follow each other more frequently than expected by chance, helping to shape text meanings and contributing to our sense of distinctiveness in a register. [...] What I shall call 'bundles', or frequently recurrent strings of uninterrupted word-forms, thus appear to represent a psychological association between words and reflect a very real part of users' communicative experiences. (Hyland, 2008b, p. 5).

Lexical bundles are recurrent expressions, regardless of their idiomaticity, and regardless of their structural status. That is, lexical bundles are simply sequences of word forms that commonly go together in natural discourse. To make the scope of our investigation more manageable, a lexical bundle is defined here as a recurring sequence of three or more words. Shorter bundles are often incorporated into more than one longer lexical bundle. (Biber *et al.*, 2021, p. 982).

As can be seen, the authors agree that lexical bundles or clusters are frequently occurring uninterrupted word sequences which are incorporated in users' communicative experiences. The issue of frequency, as previously mentioned, is pivotal. In fact, Biber (2006, p. 134) admits that "lexical bundles are identified using a frequency-driven approach [...]", guiding researchers to set particular cut-off points for their proper identification. Many are the minimum thresholds adopted, such as: 40 times per million in at least 20 texts (Biber, 2006); 20 times per million in at least 10% of texts (Biber, 2006; Hyland; Jiang, 2019); 10 times per million in at least five texts (Biber *et al.*, 2021).

The last point to be mentioned about lexical bundles regards two of their defining characteristics (idiomaticity and grammatical structures):

First, most lexical bundles are not idiomatic in meaning. Rather, the meanings of bundles like *do you want to* or *I don't know what* are transparent from the individual words. [...] most lexical bundles bridge two structural units: they begin at a clause or phrase boundary, but the last words of the bundle are the beginning elements of a second structural unit (Biber, 2006, p. 134).

In addition to clusters and lexical bundles, other terms have also been adopted in the literature to refer to formulaic language, including "[...] chunks, collocations, fixed expressions, formulaic language, idioms, lexical bundles, multiword units, prefabricated units, set phrases" (Frankenberg-Garcia, 2018, p. 93). However, although these terms are interconnected, we adopt the term *lexical bundles* as synonymous to *clusters* as part of this large superordinate *formulaic language* for a few reasons in this dissertation.

Firstly, bundles or clusters are defined by Biber (2006), Hyland (2008b), Biber *et al.* (2021), and Scott (2022) as frequently occurring uninterrupted word sequences, consisting of various elements that include at least two words. They are often found in different registers, including conversation, but our focus here underlies their noticeable and important presence in academic writing (Biber, 2006; Biber *et al.*, 2021; Hyland, 2008a; Frankenberg-Garcia, 2018), particularly in RAs, and how they help shape author's style (Leech; Short, 2007).

Furthermore, our preference for the term is also aligned with the basic sense that "the choice of one word conditions the choice of the next, and of the next again" (Sinclair, 2004, p. 19). In other words, the co(n)text of occurrence of a word (node) determines its surrounding to-be-partnered words (collocates), thus forming bundles or clusters. Associated with this notion lies the co-occurrence principle, which is observed when "[...] two [or more] word-forms are found together (co-occur) more often than chance would predict" (Scott; Tribble, 2006, p. 33). In this regard, Hyland (2008a) has proposed a taxonomy for bundles functional analysis, which

he developed based on Biber's (2006) framework, comprising three main categories as seen in Table 9:

Table 9. Taxonomy for bundles functional analysis

Broad categories	Subcategories and uses	4-word bundles
Research-oriented: help structure writers' activities and experiences within the real world	Location: indicate time/place	<i>at the beginning of; at the same time; in the present study</i>
	Procedure: show how both experiments and research are conducted	<i>the use of the; the role of the; the purpose of the; the operation of the</i>
	Quantification: quantify or mark existence	<i>the magnitude of the; a wide range of; one of the most</i>
	Description: describe research objects/contexts, specify aspects of models/equipment/materials or aspects of the research environment	<i>the structure of the; the size of the; the surface of the</i>
	Topic: refer to the research field	<i>in the Hong Kong; the currency board system</i>
Text-oriented: concern the organization of the text	Transition signals: establish additive or contrastive links between elements	<i>on the other hand; in addition to the; in contrast to the</i>
	Resultative signals: mark inferential or causative relations between elements	<i>as a result of, it was found that; these results suggest that</i>
	Structuring signals: organize stretches of discourse or direct reader elsewhere in text	<i>in the present study; in the next section; as shown in figure</i>
	Framing signals: situate arguments by specifying limiting conditions	<i>in the case of; with respect to the; on the basis of; in the presence of; with the exception of</i>
Participant-oriented: focus on the writer or reader of the text	Stance features: convey writers' attitudes and evaluations	<i>are likely to be; may be due to; it is possible that</i>
	Engagement features: address readers directly	<i>it should be noted that; as can be seen</i>

Source: Adapted from Hyland (2008a, p. 13-14).

In view of this brief theoretical introduction to the language features which build stylistic metadiscourse in academic writing, we have combined Hyland's (2005; 2008a; 2009), Biber's (2006) and Biber *et al.*'s (2021) contributions on the topic and proposed the *Stylistic Metadiscourse Model of Interpersonal Interaction in Academic Writing*, as can be seen in

Chapter 03 and in depth in APPENDICES A and B. I now proceed with offering an overview of how CL has been operationalized for stylistic analysis within the field of corpus stylistics.

2.2 Style detected in corpora: corpus stylistics

More recently, perhaps envisaging a third turn in the area of stylistics, newer branches have taken shape, especially motivated by the emergence of technological tools and the multimodal, semiotic nature of texts. Corpus stylistics, for instance, has drawn on CL tools to help compile and analyze large portions of texts, such as patterns of linguistic phenomena, which had earlier been unfeasible to study given the resources available. As Wynne (2006, p. 223) puts it, “corpus linguistics is opening up new vistas for the study of language, and there are interesting similarities in the approaches of stylistics and corpus linguistics”. As a matter of fact, back in 2010 Jeffries and McIntyre (2010) had speculated that the future of stylistics would lie in areas like corpus stylistics (for the analysis of authorial style, for example). This foreseen future has now come.

According to McIntyre and Walker (2019, p. 13, emphasis in original), “the term *corpus stylistics* has found its way into popular usage to describe the kind of work that uses corpus tools in the service of stylistic analysis.” Biber (2011, p. 20) states that “corpus-stylistic research is widely considered to be a recent innovation: an extension of earlier linguistic corpus-based studies, which focused on more traditional issues of lexical and grammatical variation and use”. A similar interest between stylistics and its offshoot corpus stylistics is the form-function interdependence in textual analysis (Nørgaard; Busse; Montoro, 2010).

Booming in the first decade of the 21st century, corpus stylistics “[...] is concerned with the application of corpus methods to the analysis of literary texts by relating linguistic description with critical interpretation” (Mahlberg, 2013, p. 1282). The scholar proposes that the area be informed by a philological circle, encompassing three connected elements: linguistic description, critical interpretation, and literary appreciation.

Nevertheless, the use of computerized tools in the study of literature has not actually been that novel (Biber, 2011; Mahlberg, 2013; McIntyre; Walker, 2019). In fact, the first studies date back to the 1980s. Much research at the time had concentrated on words, collocations, and idioms (statistically retrieved by computer programs) used by fictional characters, and how these language uses shaped the characters and helped the audience better understand them. For instance, Semino and Short (2004) investigated speech, writing, and thought presentation in an

annotated corpus of written (non-)fictional narratives (fiction, news reports, and (auto-)biographies).

As has been the case with defining stylistics, there are also some drawbacks involved in the definition of corpus stylistics. McIntyre and Walker (2019) list a few reasons for that, since there is: i) a narrow definition of the term, as if research were done only to analyzing literary texts using corpora tools and methods; ii) an assumption that traditional stylistics lacks rigor for not being corpus-based; iii) a limited view of the macro area stylistics, as though it were only concerned with the linguistic study of literature.

Because corpus stylistics research demands a focus on linguistic style and the use of stylistic theories and analytical frameworks, it differs from CL: “we define corpus stylistics as the application of theories, models and frameworks from stylistics in corpus analysis.” (McIntyre and Walker, 2019, p. 15). In addition:

What corpus linguistics offers to stylistics is a set of theories and analytical methods for exploring large quantities of language data. What stylistics offers to corpus linguistics is a set of theories, models and analytical frameworks that may be deployed in the interpretation of corpus data and corpus-derived results. (McIntyre; Walker, 2019, p. 315).

Furthermore, corpus stylistics “[...] is concerned with discerning patterns in language use through the study of large quantities of language data [...]” (McIntyre; Walker, 2019, p. 19). This data can be compiled in the form of corpora for style analysis in literary and non-literary texts. For that, the emergence of large corpora and statistical methods have substituted a common practice of obtaining stylistic traits through scholars’ semi-subjective approaches: “this is the capacity that corpora have to shed light on what constitutes the background against which linguistic items can be foregrounded.” (McIntyre; Walker, 2019, p. 13). Similarly, Stubbs (2005, p. 5, emphasis added) asserts the following: “an understanding of the background of the usual and everyday – what happens millions of times – is necessary in order to understand the *unique*.”

In this sense, our notion of style converses quite well with Wynne (2006), Biber (2011) and Mahlberg (2013), but reasonably with McIntyre and Walker (2019), because of the authors’ concern on applying stylistics theories, models, and frameworks to corpus analysis (these scholars focused on literary texts, which might justify their choice of stylistics-specific theories). Here, we apply metadiscourse linguistic theory (Hyland, 2005; 2008a; 2009; Hyland; Tse, 2004; Thompson, 2001; Thompson; Thetela, 1995) to analyze style based on corpus-obtained language data. Will this way of analyzing our data envisage a new term in the area, corpus-based *style* approach, perhaps?

In either case, style detection has commonly been operationalized in basically two macro ways. Academics have selected some language features from a much larger repertoire of an academic writer's set of linguistic items and/or structures, since focusing on all or many features would be an onerous, if not an impossible task to carry out in detail (Leech; Short, 2007). In addition, style has been detected by the comparison between at least two corpora, a study and a reference corpus, the former consisting of a set of texts written by a particular writer and the latter containing a set of texts written by another writer or group of writers. This procedure yields the identification of particular linguistic items and/or structures, since distinctiveness often comes out based on what is peculiar to an author in comparison to another. This operationalization has been a common tradition in other fields of stylistics as well, like translation stylistics (Baker, 2000; Malmkjaer, 2003; Saldanha, 2011). In this regard, Biber (2011, p. 16) upholds that

most corpus-stylistic studies focus on the distribution of words to identify textual features that are especially characteristic of an author, particular text, or even a single character within a play or novel. Three major methodological approaches have been used to study the stylistic relevance of such word distributions: 'keyword' analysis, identifying typical extended lexical phrases, and collocational analysis.

In the case of keyword analysis mentioned in Biber (2011), Culpeper (2002, p. 12) uses the term *style-markers* as synonymous to *keywords*: “style-markers, then, are words whose frequencies differ significantly from their frequencies in a norm. [...] this is precisely the principle used to identify ‘keywords’.” As for Scott (2022), “key words are those whose frequency is unusually high in comparison with some norm.”

Norms are to be defined prior to commencing a style investigation: “since variation presupposes a norm, it is necessary to establish the norm to describe variation within or from it. [...] Any social, geographic, or situational norm can be used as the basis for describing variation in writing.” (McMenamin, 2002, p. 182). As will be later introduced in Chapter 03, what constitutes an author's style in this doctoral investigation is also determined by statistical significance calculated by a computer program, *AntConc*. This is to say that, should we look at the data from a non-corpus perspective, different elements might strike up as stylistic. Therefore, we subscribe to the view that, “style, then, is a matter of ‘frequencies’, ‘probabilities’ and ‘norms’.” (Culpeper, 2002, p. 12).

An example of a study using keywords for style analysis comes in Shepherd and Berber-Sardinha (2013), in which they report that the main characters' names are mostly key in chapter one, while other words have important keyness, such as the abstract noun *life*. They also notice that bi-grams and three-grams are more frequent in the corpus and conclude by reinforcing that

the future of corpus stylistics studies should lie in the use of a digital toolkit that aims to bridge the gap between stylisticians and corpus linguists so as to produce relevant scholarship regarding text's overall style and meaning.

Outside the literary domain, author's style through metadiscourse features is described in Hyland (2002a), who addresses the importance of employing first-person pronouns as a linguistic mechanism to build one's authorial voice in academic texts. In his research, he aims to challenge a common myth that academic prose is dry and impersonal by studying the use of author pronouns such as *I*, *me*, *my*, *we*, *us*, and *our* in two corpora. The first corpus is made of "240 published journal articles, 30 from each of eight disciplines" (p. 352), written by academic experts, and the second one is composed by "48 project reports written by final-year Hong Kong undergraduates in six fields" (p. 353).

By using *Wordpilot 2000* software, he observes that the (under-)use of pronouns is closely associated with the discipline. The results from the first corpus point out that expert writers in the hard sciences and engineering use fewer pronouns (downplaying their personal role) than those in the humanities and social sciences (showing a stronger identity). The results from the 2nd corpus, on the other hand, show the novice academics' preference for passive voice, mostly motivated by cultural reasons. Interviews conducted with the participants have shown that such students, so as not to conflict with beliefs and practices which favored collectivist forms of self-representation, avoided the use of first-person pronouns. Swales (1990) note that in the earliest configurations of RAs, there was a tendency to rely on these pronouns as a way to project personal honesty and modesty.

Another similar study is reported in Hyland (2002b), in which the scholar investigates the notion of identity in Hong Kong second language learner's texts through students' (non)use of first-person pronouns and possessives, *I*, *me*, *my*, *we*, *us*, and *our*. For that, he compiles a corpus of 630,000 words, consisting of project reports in 8 different disciplines. By using *Wordpilot 2000* software, he compares this corpus to a larger reference corpus made of 240 RAs written by expert writers in 8 disciplines, totaling 1.3 million words. Hyland also uses interviews with students and supervisors as part of his analytical toolkit.

He observes that the most common author reference was *I*, and other first-person singular pronouns represented 60% of the total. However, when comparing both corpora, expert writers intervene four times more with the first-person pronouns, especially in soft disciplines. Additionally, when Hyland (2002b) investigates the discourse functions of authorial reference, once again he notices that professional writers make more solid personal commitment to what they are writing, both taking more risks and gaining credibility from the audience. Surprisingly,

students use first-person pronouns more often in the acknowledgements section of their reports. In conclusion, Hyland (2002b) notes that students would rather avoid authorial reference in their texts as compared to the expert writers.

In another study, Hyland's (2008b) initial work under the label *style*, similar to our ongoing doctoral research, reports on how John Swales uses language features to both position himself, present ideas, and interact with readers. For that, Hyland (2008b) compiles a corpus of 342,500 words comprising 14 single-authored papers and most chapters from Swales' three main books, published along 15 years. He later compares it to a broader applied linguistics reference corpus consisting of 710,000 words (66 RAs and 25 chapters from 12 books). He submits both corpora to frequency, keyword, and concordance analyses using *WordSmith Tools* (WST), v. 4.

By using the program, Hyland (2008b) obtains wordlists of syntactically complete units, including prepositional and noun phrases, as well as sentence stems. Later, he inputs this data into keyword tool to obtain key words and phrases. He then submits all results to pragmatic and functional analyses. Initially, Hyland (2008b) detects that both single words (e.g., *research*, *genre*) and multi-word units (e.g., *a genre based approach*, *English as a Second Language*) represent the themes of Swales' works, which is not so surprising. Not until Hyland (2008b) observes keyword results does he notice more accurate indicators of words and clusters characteristic of Swales' texts, such as "genre(s), would seem to, the University of Michigan, Non-native Speakers of English" (p. 145), among many others.

Additionally, it is especially from non-content words and phrases that Hyland (2008b) observes more pragmatic, discoursal features of Swales' work differently from the themes, such as *self-mention and reflection* (e.g., use of first-person singular pronouns); *hedges*, *mitigation*, and *attitude* (e.g., use of words such as *suggest*, *perhaps*, *suspect*; use of *remarkable*, *dedicated*, *surprising*, *highly*); *reader engagement* (e.g., use of *the reader*, *we*, *so let us*); and metadiscourse (e.g., use of *in terms of*, *in the next section*). Hyland (2008b) concludes by conceiving of Swales' writing style as highly personal, modest, and interactive.

Swales' writing style is further investigated in another work published by Hyland (2010). Comparisons are made between texts written by Swales and another renowned applied linguist, Deborah Cameron. By aiming to explore the issue of authorial identity in academic writing, Hyland (2010) builds two study corpora, one made of Cameron's 21 single-authored articles (125,000 words) and another of Swales' 14 single-authored articles (342,000 words). He compares them to a reference corpus consisting of 75 RAs and 25 book chapters in applied linguistics (750,000 words). Hyland (2010) uses WST, v. 4, and applies three programs to his

data analysis: *Wordlist, Concord, and Keywords*. The programs generate single words, three-grams and four-grams clusters, on which he conducts his research.

Initially, Hyland (2010) observes that the most common content words in Cameron's texts are "[...] women, language, gender, men, social, linguistic, talk, people, discourse, and work [...]" (p. 168), while in Swales' texts are "[...] research, genre(s), English, discourse, language, academic, writing, and students [...]". These findings stress that such words pertain to the academic interests of the two academics and the papers composing the corpora.

Through *Keywords* output by WST, Hyland (2010) reveals that, by establishing truths, challenging contrary positions, and creating solidarity with readers, Cameron projects a confident, radical, and combative identity (by using *nouns; it is + adj. + to infinitive clauses; is; conjunctions; stance adverbials*; etc.). Swales' rhetorical choices, however, build a more self-effacing, conciliatory, and interpersonal identity through *self-mention, hedges*, and engagement features (by using *personal pronouns; verbs; adverbs; adjectives*; etc.).

Hyland (2010) concludes that Cameron and Swales have different authorial identities and pertain to wider social and cultural practices. While the former has a combative style, even positioning herself in the media and in other popular genres, the latter shows a quieter, more urbane, and gently self-deprecating style.

Another way to operationalize author's style investigation is proposed in Biber and Conrad (2019). Their outline allows for a focus on any language feature found in text excerpts, observing their distribution throughout the variety. Their research has shown that "[...] variation directly reflects the varying communicative constraints and affordances associated with different contexts, as well as the varying goals of people who communicate in these contexts." (Grieve, 2023, p. 5). In this doctoral research, we understand that academic writers often make choices from a wide language repertoire within a specific situational context (academia). Since "styles are normally described for the texts within a register or genre" (Biber; Conrad, 2019, p. 18), and "[...] corpus-based analyses of linguistic form are only the first step of a corpus-stylistic study, providing the foundation for subsequent functional interpretation" (Biber, 2011, p. 21), we understand that author's choices, in comparison to other writers', are functional and help build interpersonal interaction between themselves and readers across texts. All these scholars, though, abide by the communicative constraints and affordances associated with different contexts of language use. In this regard, the approach to be taken in this doctoral research will be presented next in Chapter 03.

CHAPTER 03 – The methodological detection of style

In this research, CL is being used as an approach (Tognini-Bonelli, 2001) to conduct linguistic analyses in a study corpus compared to a reference corpus. Here, CL “[...] constitutes a methodology which underpins and informs the use of a number of analytical methods, including, among others, frequency profiling, collocation and keyness” (McIntyre; Walker, 2019, p. 133). Precisely, we follow a corpus-based approach towards corpus exploration, for it is “as a methodology that uses corpus evidence mainly as a repository of examples to expound, test or exemplify given theoretical statements” (Tognini-Bonelli, 2001, p. 10). Based on these assertions, the methodological steps to carry out this investigation are presented in depth across this chapter.

3.1 An overview of the study corpus and the reference corpus

This step entailed compiling two corpora, a study and a reference corpus, using multiple tools and programs. Berber Sardinha’s (2004) prerequisites towards corpus compilation were followed, including representativeness, authenticity of the texts produced in natural language, and text compilation following pre-defined criteria according to the research goals.

3.1.1 The study corpus

The first corpus compiled was the study corpus, *BC*. It consists of 25 single-authored RAs, mainly quantitative empirical (Gray, 2015), written by Douglas Biber and published in a variety of peer-reviewed scientific journals. These texts span a time period of 35 years, having been published between 1984 and 2019. We divided this time span into three different time slots, each representing roughly a 10-year period, so as to better check whether Biber’s stylistic choices have shifted over time. All in all, Table 10 presents an overview of these texts and their time range, as well as the number of types and tokens as calculated by *AntConc* (Anthony, 2022a):

Table 10. Overview of *BC*

<i>Biber Corpus - BC</i>					
Sub-register	Time range	Time slots	N. of texts	Types	Tokens
Journal RAs	1984-2019 (35 years)	• (1) 1984 to 1993 = 13 texts; • (2) 1995 to 2005 = 6 texts; • (3) 2006 to 2019 = 6 texts	25	5,263	144,719

Source: Elaborated by the author.

It should be noted that, during this time period, Biber published more than 200 texts, including books, book chapters, reviews, monographs, and proceeding papers, especially in collaboration with other scholars, as listed on his personal website CV¹⁴. However, due to the focus of this investigation, only 25 single-authored journal RAs were selected for the study corpus compilation. The additional 74 journal RAs he published in co-authorship were discarded from *BC*, as well as other texts. Regarding this collaborative work and authorship attribution, Biber (2013, p. 375) states:

I like collaborating with a coauthor and working as part of a collaborative team. [...] But I somehow have ended up doing as much collaborative research as individual research. [...] The first author listed on my books and articles is the person who takes the lead in the analysis, and who actually does most of the writing. But all authors are usually fully involved in the analysis process and help with commenting and revising the paper.

In the study corpus, there is some variation in the 25 RAs from the classic IMRD structure (Swales, 1990; 2004), with papers showing all or some of these sections. There are papers which contain other post-textual elements, like appendix, bio, acknowledgments, and notes. Oftentimes the titles of these elements shift from those of IMRD, typically adjusting to the goal of each article. In some papers, some textual elements are combined together (i.e., summary and conclusion). In spite of these differences, the following elements commonly appear: abstract, keywords, introduction, theoretical background, methods, results, summary, conclusion, appendix, references, bio, acknowledgments, and notes. Moreover, the situational characteristics of *BC* journal articles (Gray, 2015) is provided in Table 11:

Table 11. Situational characteristics of the study corpus RAs

Major elements	Features	Biber Corpus
1. Participants	A. Writer	only Biber (single)
2. Textual layout & organization	A. Length	4 to 42 pages
	B. Headings	un-numbered and numbered
	C. Use of abstracts	yes
	D. Visual elements	tables, figures, and equations
	E. Sections / organization	IMRD, other (theoretical RAs) and varying section headings/names
3. Setting	A. Nature of journals	generalist and specialized
4. Subject/Topic	A. General topic of the discipline	varied
5. Purpose	A. General academic purpose	varied
6. Nature of data or evidence	A. Presence of observed data	yes
	B. Use of numerical data	yes (for quantitative RAs)
	C. Primary presentation of evidence	extensive prose, quantitative displays, and mathematical formulas
	D. Object of study	varied

¹⁴ Available at: <https://dougibiber.weebly.com/> Access on: 10 Apr. 2023.

7. Methodology	A. General method type	quantitative and qualitative (empirical RAs); general theoretical (theoretical RAs)
	B. Statistical techniques	advanced and descriptive statistics (quantitative empirical RAs); none for theoretical RAs
8. Explicitness of research design	A. Explicitness of purpose	direct and indirect
	B. Explicitness of research questions	indirect and no discernible statement
	C. Explicitness of citations	within the text and in footnotes/endnotes
	D. Explanation of evidence	extensive
	E. Explanation of procedures	extensive

Source: Elaborated by the author.

As previously mentioned, *BC* consists of journal RAs only, mainly because of three major reasons. First, Biber's greatest number of single-authored texts are RAs, so, in much the same way as Hyland's (2008b; 2010) and McIntyre and Walker's (2019) sample corpora, ours is considered an appropriate sample to investigate Biber's style in comparison to his contemporaries'. Second, an earlier version of this dissertation presented at the qualifying exam ruled out results concerning differences in the use of metadiscourse features according to genre and register, since most of the language choices recurred across different books and book chapters. Hence, excluding these two registers and including only journal RAs in the current sample would not pose a problem, especially because these three academic registers share particular features in comparison to other registers (Biber; Gray, 2016) and have been produced in similar situations (Biber; Conrad, 2019). Third, a great amount of metadiscourse research focuses on RAs (Pearson; Abdollahzadeh, 2023). Last, it is much easier to track Biber's prominence as a scholar (and potentially as an academic writer) on the basis of RAs citations (Lillis; Curry, 2010) calculated by Google Scholar metrics and other sources (Ioannidis, 2023). These RAs were published in more than a dozen different journals, as presented in APPENDICES C and D.

The journals are edited by different scholarly publishers. Assuming that style guidelines could potentially influence Biber's language choices (Lillis; Curry, 2010), Table 12 presents a summary of key points regarding the use of language across four major publishing companies. These publishers account for 50% of the published articles in *BC*. Their style sheets were assessed by answering the following major question:

- i) Do the journal's style guidelines provide specific orientation about how scholars should use English for making citations, providing *code glosses* and cross-references (i.e., introducing/referring to illustrations and examples), and addressing the research/paper with special language usage?

Specifically, we intended to verify whether:

- i) in the case of citations, writers should favor integral or non-integral formats and use specific vocabulary (like verbs such as *discover*, *analyze*, *identify*, etc.) to introduce/refer to them;
- ii) in the case of *code glosses*, scholars ought to favor any of the following: *e.g.*; *i.e.*; *for instance*; *for example*; etc.;
- iii) in the case of cross-references (illustrations like *tables*, *figures*, *images*, etc.; paper sections and subsections) and examples, writers should introduce/refer to them with specific vocabulary (including verbs such as *see*, *note*, *check*, or bundles like *as noted above*, *it should be noted that*; etc.);
- iv) in the case of other language uses, scholars ought to favor and/or avoid particular vocabulary/structures (like *personal* or *reader pronouns*, *evaluative adjectives/adverbs*, etc.).

Table 12. Style guidelines of BC journals publishing companies

Journals publishing companies	Citations	Code glosses	Cross-references	Other language uses
De Gruyter Mouton ¹⁵	- integral and non-integral allowed	none observed	- should be explicitly referred to in the text (<i>see</i> , <i>cf.</i> , etc.); - tables: contain info presented in rows and columns; - figures: graphs, line drawings, photographs, etc. - no preference for a particular form observed	none observed
John Benjamins Publishing Company ¹⁶	- integral and non-integral allowed	none observed	none observed	- spelling must follow a consistent English variety
Oxford University Press ¹⁷	- integral and non-integral allowed	- forms like <i>e.g.</i> and <i>i.e.</i> are allowed	- figures and tables should be numbered and mentioned in the text, like: <i>see Section 2.5</i>	- use the system which you are most accustomed to, but be consistent;

¹⁵ Available at:

https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.degruyter.com%2Fpublication%2Fjournal_key%2FALR%2FdownloadAsset%2FALR_Mouton_journal_stylesheet.pdf&psig=AOvVaw2Zjx-zPpa03Sgt6Y-f_O6B&ust=1713869358761000&source=images&cd=vfe&opi=89978449&ved=0CAgQr5oMahcKEwiA55Kj09WFAxUAAAAAHQAAAAAQBA Access on: 22 Apr. 2024.

¹⁶ Available at: <https://benjamins.com/series/ijcl/styleguide.pdf> Access on: 22 Apr. 2024.

¹⁷ Available at: https://academic.oup.com/dsh/pages/general_instructions Access on: 22 Apr. 2024.

		- no preference for a particular form observed	<i>see Appendix I</i> <i>see Fig. 1</i> - no preference for a particular form observed	- use <i>and</i> instead of &
Wiley ¹⁸	- integral and non-integral allowed	- forms like <i>e.g.</i> are allowed - no preference for a particular form observed	- examples and figures should be numbered consecutively and incorporated in the text - no preference for a particular form observed	none observed

Source: Elaborated by the author.

Biber's journal RAs discuss a variety of topics in corpus linguistics, including approaches to carry out corpus-based research (i.e., multi-dimensional analysis); analysis of multi-word patterns in speech and writing; variability in the form and use of English complement clauses; comparison of British and American writing; typology of English texts; co-occurrence patterns among collocations; grammatical marking of stance; spoken/written differences; investigations on Somali; register and genre variation. Now, let us learn about the reference corpus.

3.1.2 The reference corpus

In addition to the study corpus, a reference corpus was also compiled, *CorJAAL*, consisting of texts published by three journals: Applied Linguistics, System, and TESOL (Teaching English to Speakers of Other Languages) Quarterly. For style investigation, it is important to have a comparison parameter to illustrate what is peculiar to a writer's language repertoire and is probably not so to other academic writers'. A reference corpus works as a norm to which the study corpus will be compared, but as a relative norm, as "[...] an absolute norm for English cannot be relied on [...]" (Leech; Short, 2007, p. 41). Therefore, since Biber writes RAs in this area, the purpose is to investigate how he varies language within this expected relative norm as far as the use of key metadiscourse elements is concerned.

It should be added, however, that the "the issue of reference corpus selection is far from decided" (Scott; Tribble, 2006, p. 64), because many scholars have attempted to establish the best layout for a reference corpus, without reaching much consensus. Nevertheless, some agree

¹⁸ Available at:

<https://onlinelibrary.wiley.com/page/journal/14679582/homepage/forauthors.html> Access on: 22 Apr. 2024.

that researchers seeking to build a reference corpus should bear in mind such issues as size, genres, registers, content, research goals, among others, for the layout of the corpus may influence the research results. Overall, four statements are summarized in Table 13:

Table 13. Definitions and layout of a reference corpus

Scholars	Insights on reference corpus
Berber Sardinha (2000)	A reference corpus should be five times as large as the study corpus and be more suitable in terms of contents in relation to the study corpus.
Scott and Tribble (2006)	A reference corpus should be an appropriate sample (i.e., a large one, preferably many thousands of words) of the language in which the study corpus is written. In terms of the composition, the research purpose is fundamental to consider.
Scott (2009)	A reference corpus layout should consider four aspects: size (the larger, the better); date; texts; genre-specific.
Geluso and Hirsch (2019)	A reference corpus of a similar register will result in words very particular to the study corpus, while a reference corpus of a different register will likely return many words that are more general to the register that the study corpus represents.

Source: Elaborated by the author.

Taking into account the suggestions of the authors in Table 13, the reference corpus of this dissertation combines various elements, including size (five times larger than the study corpus); register (academic writing); sub-register and genre (empirical and theoretical RAs); diversity of authors; three different journals; variety of themes, topics, and fields of the RAs within applied linguistics.

In terms of size, Table 14 presents an overview of the reference corpus and its time range, as well as the number of types and tokens as calculated by *AntConc*:

Table 14. Overview of *CorJAAL*

<i>Corpus of Journal Articles in Applied Linguistics - CorJAAL</i>				
Sub-register	Time range	N. of texts	Types	Tokens
Journal RAs (<i>JA</i>)	1984-2019 (35 years)	126 (42 RAs per journal)	21,311	819,612

Source: Elaborated by the author.

In terms of genre and register, the choice of selecting only RAs published in journals represents a way to control variables that might otherwise impact the results. In our context, however, focusing on other academic genres such as books and book chapters would not pose potential differences because they are often produced in similar situational contexts (Biber; Gray, 2016; Biber; Conrad, 2019). Also, a great amount of metadiscourse research focuses on RAs (Pearson; Abdollahzadeh, 2023). In addition, as to Geluso and Hirsch (2019), when a

reference corpus contains a similar register¹⁹ to the study corpus, the results from a keyword search are very particular to the study corpus and not to the register. This is to say that the words identified characterize the study corpus (and potentially the author's style, as we will show) and not the register. This is not to say, on the other hand, that register may not influence someone's individual style, since research like Grieve's (2023) has shown evidence on this issue. Nevertheless, this doctoral study is not working on stylometric analysis as did Grieve to prove or measure the extent to which register influences one's style. Herein, we assume that there is potential influence, but we focus on how Biber has made metadiscourse choices within a specific genre (RA) and in a particular (sub-)register (academic writing in applied linguistics RAs) under all the previously introduced linguistic and academic norms and conventions laid out in Chapter 02. However, the degree of this register-style influence is encouraged to be measured in future studies.

CorJAAL consists of empirical (qualitative and quantitative) and theoretical RAs. In the case of the empirical RAs, there is a balanced amount of both qualitative and quantitative texts published in Applied Linguistics, but more qualitative RAs in the other two journals. As for their structure, there is some noticeable variation among the 126 texts from the classic IMRD structure (Swales, 1990; 2004), with papers showing all or some of the common sections. Some theoretical papers published in System and TESOL Quarterly, for instance, do not present all these sections. Other papers, however, contain additional post-textual elements to the IMRD structure, like appendix, bio (except for those published in Applied Linguistics), acknowledgments, and notes. Particularly in the papers published in Applied Linguistics, some sections are combined together (i.e., summary and conclusion). Oftentimes the titles of the elements shift from those of IMRD, typically adjusting to the goal of each article. Despite these differences, the following elements are commonly identified: abstract, keywords, introduction, theoretical background, methods, results, summary, conclusion, appendix, references, bio, acknowledgments, and notes. The situational characteristics of *CorJAAL* journal articles following Gray's (2015) framework is provided in Table 15:

¹⁹ When the authors made this claim, they referred to the following academic sub-register: academic writing in applied linguistics. In their research, the authors made different comparisons. To reach the conclusion that keywords characterize the study corpus and not the register, they compared their study corpora (made of papers published by faculty members from two US universities, including book chapters and journal publications) to a reference corpus of the same sub-register: research articles in the discipline of applied linguistics.

Table 15. Situational characteristics of the reference corpus RAs per journal

Major elements	Features	Applied Linguistics	System	TESOL Quarterly
1. Participants	A. Writer	single and small groups	single and small groups	single and small groups
2. Textual layout & organization	A. Length	14 to 35 pages	5 to 23 pages	5 to 37 pages
	B. Headings	un-numbered and numbered	none, un-numbered, and numbered	un-numbered and numbered
	C. Use of abstracts	yes	yes	yes
	D. Visual elements	tables, figures, and equations	tables and figures	tables and figures
	E. Sections / organization	IMRD, other (theoretical RAs), and varying section headings/names	IMRD, other (theoretical RAs), and varying section headings/names	IMRD, other (theoretical RAs), and varying section headings/names
3. Setting	A. Nature of journal	generalist	generalist	generalist
4. Subject/ Topic	A. General topic of the discipline	varied	varied	varied
5. Purpose	A. General academic purpose	varied	varied	varied
6. Nature of data or evidence	A. Presence of observed data	yes	yes	yes
	B. Use of numerical data	yes (in quantitative RAs)	yes (in quantitative RAs)	yes (in quantitative RAs)
	C. Primary presentation of evidence	extensive prose, quantitative displays, and mathematical formulas	extensive prose, quantitative displays, and mathematical formulas	extensive prose, quantitative displays, and mathematical formulas
	D. Object of study	varied	varied	varied
7. Methodology	A. General method type	quantitative and qualitative (in empirical RAs); general theoretical (in theoretical RAs)	quantitative and qualitative (in empirical RAs); general theoretical (in theoretical RAs)	quantitative and qualitative (in empirical RAs); general theoretical (in theoretical RAs)
	B. Statistical techniques	advanced and descriptive statistics (in quantitative empirical RAs); none for theoretical RAs	advanced and descriptive statistics (in quantitative empirical RAs); none for theoretical RAs	descriptive statistics (in quantitative empirical RAs); none for theoretical RAs
8. Explicitness of research design	A. Explicitness of purpose	direct	direct	direct
	B. Explicitness of research questions	direct and indirect	direct and indirect	direct and indirect
	C. Explicitness of citations	within the text and in footnotes/ endnotes	within the text and in footnotes/ endnotes	within the text and in footnotes/ endnotes
	D. Explanation of evidence	extensive	extensive	extensive
	E. Explanation of procedures	extensive	extensive	extensive

Source: Elaborated by the author.

The journals and RAs which compose *CorJAAL* were selected based on the following criteria:

➤ Criteria for selecting journals:

- Figure among top 20 with the highest impact factor (IF)²⁰ throughout 2022 and across a 5-year period according to Clarivate's (2023) Journal Citation Reports;
- Publish RAs on a variety of themes, topics, and fields within applied linguistics.

➤ Criteria for selecting RAs from these journals:

- Have been published over a period of 35 years, the same time lapse of *BC* RAs (from 1984 to 2019);
- Show the highest number of citations in the volumes and issues published in the same years as the *BC* texts.

The topics, themes, and fields of the texts within applied linguistics published in the journals are introduced in APPENDIX E. In addition, as seen in APPENDICES F and G, a plethora of authors and their texts compose *CorJAAL*, which serve a consistent diversity of authorship against which Biber's texts were compared.

3.1.3 The compilation of the study and reference corpora

In terms of corpus compilation, all study and reference corpora texts have been downloaded from the web and saved to the research computer folders. I converted the *.pdf* files into *.txt* documents (UTF-8) using free online programs, such as *I love Pdf*²¹, *Ocr 2 Edit*²², and

²⁰ IF has figured as the dominant bibliometric for evaluating academic output for decades, and this was the major reason why it was chosen to select journals for the ongoing research. However, we recognize that it raises many problems, especially when it is wrongly used by institutions to assess articles' quality and prevent scholars from getting jobs, tenure, or grants. In general, a common critique of IF is that "[...] citations are a shallow measure of research quality or impact" (Lillis; Curry, 2010, p. 9). In addition, Lillis and Curry (2010) notice that journals with higher IF receive more submissions; authors make immoderate self-citations and deliberate citations of other scholars publishing in the same journal; journals may pressure academics to cite papers from the same journal and limit scholars' citations to rival journals; university library subscriptions favor journals with higher IF. Furthermore, Waltman and Traag (2020) uphold that, under certain conditions, the number of citations an article receives is a more accurate indicator of the value of the article than IF. Hyland (2022, p. 18) also notes that "[...] a few highly cited papers can dramatically increase the impact of average or less significant papers."

²¹ Available at: <https://www.ilovepdf.com/> Access on: 10 Apr. 2023.

²² Available at: <https://www.ocr2edit.com/> Access on: 10 Apr. 2023.

*Online 2 PDF*²³. By using basic automated *Microsoft Word* tools, I edited the texts to correct typos and other conversion-resulting errors, such as in “distribu-tion → distribution”, so as to ensure text readability on *AntConc*. Later, I withheld in brackets (< >) pieces of text from the original papers, including: in-text short and long citations (only in the case of *BC*); titles of paper sections and subsections (e.g., <Introduction>); information related to Biber and other authors (e.g., <Douglas Biber Northern Arizona University, Flagstaff>). Because only the IMRD sections from *BC* and *CorJAAL* were used for the corpora compilation, I opted to delete such post-textual elements as authors’ bios, notes, and acknowledgements, as well as images, tables, graphs, and lists of bibliographical works cited in the references section. For all these cases, I inserted the tag <DELETED EXCERPT> in place. Moreover, I added text-related identification tags at the beginning of each file, as exemplified in Table 16:

Table 16. Example of text identification tags in each corpus

<i>BC</i>	<i>CorJAAL</i>
<text id = BC.JA.1984.01> <title = > <author = >	<text id = CorJAAL.SYS.1984.02> <title = > <author = >
BC = Biber Corpus	CorJAAL = Corpus of Journal Articles in Applied Linguistics
JA = Journal RAs (subcorpus)	SYS = abbreviation for the journal System (other abbreviations include: AL = Applied linguistics; TQ = TESOL Quarterly)
Year of publication = 1984	Year of publication = 1984
Order of texts per year ²⁴ = 01	Order of texts per year = 02
Title = the title of each RA is typed	Title = the title of each RA is typed
Author = Biber’s name is provided	Author = the name of each author is provided

Source: Elaborated by the author.

Lastly, I saved all the edited *.docx* files into *.txt* (UTF-8) format. All these procedures were similarly followed in the reference corpus compilation. An example of a tagged journal article from *BC* and *CorJAAL* is illustrated in Figures 3 and 4:

²³ Available at: <https://online2pdf.com/> Access on: 10 Apr. 2023.

²⁴ This information was added in the tag because more than one paper was published in the same year.

Figure 3. Example of a tagged journal RA from *BC*

Arquivo Editar Exibir

<text id = BC.JA.1984.01>
 <title = PRAGMATIC ROLES IN CENTRAL SOMALI NARRATIVE DISCOURSE>
 <author = Douglas Biber University of Southern California>

<abstract>
 This paper compares the marking of pragmatic roles in Central Somali oral narrative discourse and elicited question-answer pairs to illustrate the claim that information structure must be studied within the context of particular discourse genres. The study of pragmatic roles in Central Somali is especially interesting because clause-level focus is explicitly marked through the particle *yaa*. The functions of this particle in elicitation question-answer pairs and narrative discourse are compared, and it is shown that elicitation data exhibit only a few of the possible functions of *yaa*, viz. as a marker of assertive and contrastive focus. In contrast, narrative discourse provides examples of *yaa* as a marker of both event-clause focus and discourse topic. This result is discussed within the context of discourse coherence and is shown to be not as surprising as it first appears. In addition, narrative focus constructions (defined as the most salient section of new information in a narrative text) are shown to be formally well-defined and functionally important in giving coherence to a narrative, although no counterpart has been found in elicitation data. In conclusion, it is noted that pragmatic roles should be studied in a broad range of discourse genres in addition to elicited question-answer data, since each genre may illustrate different functions of the same constructions.

<1. Introduction>
 By definition, pragmatic roles, <e.g.> focus and topic, must be studied within a discourse context, since they mark the status of information with respect to the preceding or following text. This necessity is acknowledged in descriptive linguistics through the use of elicited question-answer pairs in which the new information or the choice of topic in the answer is forced by the wording of the question <[Comrie 1981:57]>. Even within generative grammar, with its traditional emphasis on the sentence as the maximum syntactic unit of interest, the marking of focus has been discussed in terms of question-answer pairs, <e.g. Jackendoff 1972:229ff>. Data of this type, however, exhibit only a small range of the

Source: Elaborated by the author.

Figure 4. Example of a tagged journal RA from *CorJAAL*.

Arquivo Editar Exibir

<text id = CorJAAL.SYS.Art.1984.002>
 <title = USING INPUT METHODS TO IMPROVE WRITING SKILLS>
 <author = MARY SCHLEPPEGRELL>

<abstract>
 This paper describes two special purposes English language programs for Egyptian economists. One used a communicative approach consisting of conversation sessions and extensive writing, and the other focused on developing the receptive skills of listening and reading. Writing ability was tested, using an essay task, and the group that had concentrated on the receptive skills had greater gains in writing ability than the group that had focused on the productive skills of speaking and writing throughout the course. This gives added support to the input hypothesis of language acquisition and comprehension-based methods of language teaching.

<INTRODUCTION>
 This study compares two methods of improving the writing abilities of students in a special purposes English course. It compares a strategy which stressed the productive skills of writing and speaking with a comprehension-based strategy focusing on the receptive skills of listening and reading. The purpose of the study was to test the hypothesis that focusing on receptive skills is the most effective way to promote language acquisition.

Source: Elaborated by the author.

After carrying out these steps, the text files were inserted into *AntConc*, version 4.2.0. *AntConc* is a software suite which contains nine tools. According to Anthony (2022b, p. 1),

AntConc is a freeware, multiplatform tool for carrying out corpus linguistics research, introducing corpus methods, and doing data-driven language learning. It runs on any computer running Microsoft Windows (built on Win 10), MacOS (built on Mac Catalina), and Linux (built on Linux Mint). It is developed in *Python* and *Qt* using the *PyInstaller* compiler to generate executables for the different operating systems. It uses SQLite as the underlying database.

Having presented the steps for the study and reference corpora compilation, I proceed next with introducing the methodological procedures undertaken in the doctoral investigation.

3.2 Methodological procedures for data extraction and analysis

This subsection provides details of how keywords were obtained from *AntConc* to serve the stylistic metadiscourse analysis aimed at in this doctoral research.

3.2.1 Introducing the *Keyword tool* and its tests and metrics

Gabrielatos (2018) draws attention to the fact that most scholars have wrongfully used both terms *keywords* and *keyness* as interchangeable. Although these words are related, they ought not to be taken as synonymous. “Definitions of the terms *keyness* or *keyword* have tended to conflate their nature with the proposed metric for measuring keyness” (p. 229).

As a matter of fact, it was with another useful computer program, WST, that keyness analysis²⁵ was first incorporated in the mid-to-late 1990s. In Scott (2022), the concept of keyness is presented as follows:

A word is said to be ‘key’ if

- a) it occurs in the text at least as many times as the user has specified as a minimum frequency
- b) its frequency in the text when compared with its frequency in a reference corpus is such that the statistical probability as computed by an appropriate procedure is smaller than or equal to a p value specified by the user
- c) in addition, the strength of keyness must be at least as great as the minimum log ratio set by the user. (Scott, 2022).

According to *AntConc* Help Manual, the *Keyword Tool*

[...] shows words that appear unusually frequently in the target corpus in comparison with the words in the reference corpus based on a statistical measure (i.e., ‘keywords’). These words can be considered to be characteristic of the target corpus. The settings can also be changed to show words that appears unusually infrequently

²⁵ Gabrielatos (2018, p. 228) admits that “it needs to be clarified that using ‘keyword’ as a default term to refer to the linguistic unit of focus in a keyness analysis is both restricted and restricting.” For this reason, he opts for the term *keyness analysis*, which, for him, encompasses not only the analysis of words per se, but of what he names as *key items* (e.g., multi-words; linguistic units, especially if the corpus is lemmatized, annotated for grammatical, syntactic, or semantic categories). In addition, he places particular focus on the metrics (e.g., effect size metrics over statistical significance tests such as log-likelihood) used in such analyses. Despite that, I will carry on opting for the terms *keyword(s)* and *keyword(s) analysis* (except when I am noticeably quoting or referring to Gabrielatos’ term), even if this may include multi-word units, because my attention is given to the combined quantitative-qualitative values of such keywords over methodological disputes or allegiances as to what test or metric is better than another.

in the target corpus compared with the reference corpus (i.e., ‘negative keywords’.) (Anthony, 2022b, p. 3).

To obtain such keywords, the program compares two wordlists automatically generated by *AntConc Word tool*. In this doctoral research, one wordlist comes from *BC*, which is the study corpus – or target corpus, as named by Anthony (Anthony, 2022b) –, and the second one from *CorJAAL*, or reference corpus. The results displayed by *Keyword Tool* are classified based on their statistical significance and labeled as either positive or negative.²⁶ In fact, according to Scott (2022),

a word which is *positively* key occurs *more often* than would be expected by chance in comparison with the reference corpus.
A word which is *negatively* key occurs *less often* than would be expected by chance in comparison with the reference corpus. (Scott, 2022).

The program displays keywords based on the comparison of frequency patterns, showcasing both positive and negative words. In other words, in our case, by comparing both study and reference corpora, positive keywords are found more unusually frequently in *BC* than in *CorJAAL*, while negative keywords are noticed more unusually frequently in *CorJAAL* than in *BC*. For the purposes of the keyword analysis carried out in this dissertation, only positive keywords are analyzed, because they show more statistical significance in *BC* in comparison to the reference corpus.

Scott (2022) goes on to explain that what a human being identifies as key may not exactly be considered as such by the program. Moreover, Gabrielatos (2018) adds that a keyness analysis is a way to establish aboutness, especially in the earliest studies, as we will see below. In addition to such aboutness-related keywords, however, there are two other types of keywords generated by the program:

First, there will be proper nouns. Proper nouns are often key in texts, though a text about racing could wrongly identify as key, names of horses which are quite incidental to the story. This can be avoided by specifying a higher Minimum Frequency.

Second, there are key words that human beings would recognise. The program is quite good at finding these, and they give a good indication of the text’s ‘aboutness’. [...]

Third, there are high-frequency words like *because* or *shall* or *already*. These would not usually be identified by the reader as key. They may be key indicators more of style than of ‘aboutness’. [...] (Scott, 2022).

²⁶ I must make explicit that such adjectives are not being used on a judgmental basis, in that one word is better (positive/superior) than another (negative/inferior).

As for this third type of keywords, which is the focus of this doctoral investigation, Bondi (1996, p. 10-11) adds that grammatical words (*of, which*) and phrases (*in terms of, is equal to, as a whole, etc.*) characterize both text and authorial style:

The corpus of current economics articles, on the other hand, highlights grammatical words: *of* and *which*. *Which* is mainly to be attributed to frequent use of relative defining and non-defining clauses [...] The presence of *of* reflects a preference for nominal postmodification against noun + noun constructions. [...] Both *which* and *of* can be seen as matters of individual style. [...] These [key phrases] can again be considered matters of individual style, but they also point at important features of the genre and of the writer's authorial identity, highlighting that we are dealing with scientific argument that favours logical structures.

Grammatical words like *of* and *in*, identified as key in Geluso and Hirsch's (2019) study, function similarly in Biber and Gray (2016). They are the two most common prepositions in prepositional phrases used as postmodifiers (Biber *et al.*, 2021), creating more informationally dense structures, such as in noun postmodification: "Add to our understanding *of* a behaviorist definition" and "Incomplete knowledge refers to the gaps *in* the learner's knowledge" (Geluso; Hirsch, 2019, p. 230). For the authors, this is typical of the academic writing register, including applied linguistics.

Scott (2022) mentions that keyness is calculated by statistical tests using a p-value (probability value) as a reference:

the p-value is that used in standard chi-square and other statistical tests. This value ranges from 0 to 1. A value of .01 suggests a 1% danger of being wrong in claiming a relationship, .05 would give a 5% danger of error. In the social sciences a 5% risk is usually considered acceptable. (Scott, 2022).

In addition, Brezina (2018, p. 12) makes explicit the conventional cut-off point used in these tests:

[...] 0.05 or 5% is the conventional cut-off point which can be imagined as the risk we are willing to take when inferring from the sample to the population [...]. If we are willing to take only a smaller risk than 5%, we can decide on the p-value cut-off point 0.01 (1%) or even 0.001 (0.01%).

These tests are carried out to test alternative and null hypotheses and check if there is statistical significance based on the results. For Brezina (2018, p. 12), "based on the p-value we can decide whether we have enough evidence in the corpus (sample) to reject the null hypothesis." Let us take the example of the following hypotheses: i) alternative hypothesis (H_1): *Biber and CorJAAL authors differ in the use of the word 'registers'*; ii) null hypothesis (H_0): *Biber and CorJAAL authors do NOT differ in the use of the word 'registers'*. The null hypothesis is basically the reverse of the alternative hypothesis. "To put it very simply, the null hypothesis states that there is nothing special going on in the corpus or corpora we analyse, e.g.

there is no difference between two (sub)corpora.” (Brezina, 2018, p. 12). As the goal here is to reject the null hypothesis and show that there is noticeable difference in the use of particular metadiscourse words between Biber and *CorJAAL* authors, we assume that both positive and negative keywords do not merely occur by chance, as they reveal the extent to which Biber crafts a metadiscourse style in comparison to *CorJAAL* authors. Brezina (2018, p. 12) adds:

If the p-value is small enough, usually smaller than 0.05, i.e. 5%, we reject the null hypothesis and conclude that the observed difference is unlikely to be due to chance and therefore the result is *statistically significant*. This means that the difference observed in the corpus (sample) is likely to be a true difference in the population (all language use).

These results are obtained by calculating log-likelihood and dice coefficient (effect size measure). As for the log-likelihood test (a.k.a. G^2 or LL), Smith and Hardie (2009) state:

If the log likelihood for your result is greater than 6.63, the probability of the result - i.e. the difference between the two corpora - happening by chance is less than 1%. So we can be 99% certain that the result actually means something. This is usually expressed as $p < 0.01$.

If the log likelihood is 3.84 or more, the probability of it happening by chance is less than 5%. So we are 95% certain of the result. This is expressed as $p < 0.05$.

Log-likelihood critical values and associated p-values adopted on *AntConc* are presented in ANNEX A. As Gabrielatos (2018, p. 231) highlights, “[...] the lower the *p*-value, the higher the statistical significance”. The use of this test, especially in WST, has apparently propelled the inception of the term *keyword*: “[...] it is not unlikely that the wording of the definition of keywords was influenced by (or reflected) the choice of the particular statistical significance metric in WST, log-likelihood [...]” (Gabrielatos, 2018, p. 229). He goes on to remind us of who developed the test and for what primary goal, especially tied to Anthony’s (2022b, p. 3) “unusually infrequently” usage of keywords: “Dunning [...] developed the log-likelihood test in order to accurately identify the statistical significance of rare events [...]” (Gabrielatos, 2018, p. 229).

Regarding the effect size measure, Brezina (2018, p. 14) states:

Effect size in descriptive statistics is a standardized measure, that is a measure comparable across different studies [...], that expresses the practical importance of the effect observed in the corpus or corpora. For example, if we establish by a statistical test [...] that two groups of speakers (e.g. men and women) differ from each other in the use of a particular linguistic variable, i.e. there is a statistically significant difference between these two groups, we still need to see *how large this difference is* and *whether it is practically important*.

By distinguishing both log-likelihood and effect size measures, McIntyre and Walker (2019, p. 159-161) summarize:

[...] while log-likelihood tells us how much evidence there is for there being a difference between the two corpora [study and reference], it says nothing about the *size* of that difference (it may be, for example, that we obtain a significant result as a consequence of having a large corpus). [...] For this we need to consider effect size.

It should be noted that *AntConc* also makes available a combination of both log-likelihood and effect size measures in the *Keyword* tool. However, the issue of the best test and metric for keyword analysis is far from decided. Some scholars like Gabrielatos (2018) favor the use of effect size measures for keyword analysis, although he admits that, depending on the type of analysis to be carried out, there should be methodological changes. He claims that dice, available on *AntConc*, is not an appropriate effect size measure for keyword analysis. Other researchers like Pojanapunya and Watson Todd (2016) take a more balanced approach towards this issue, claiming that these two statistics are suited for different research purposes. In Pojanapunya and Watson Todd's (2016) investigation, while keywords obtained from log-likelihood were relatively common in general use and served genre-related research, those generated by effect size (odds ratio) highlighted more specialized words and suited critically-oriented purposes. In this doctoral research, since we have opted to use *AntConc*, we subscribe to Anthony (2022b, p. 14, emphasis added):

Appropriate effect size measures are still being debated in the field, so *the default setting is to show all values for this measure*. With the default settings, keywords are ranked according to their likelihood measure scores. This equates to ranking keywords according to p-values, which raises several questions/problems. However, *it is the current standard in the field and results tend to show that ranking by likelihood leads to more intrinsically intuitive results than those generated when an effect size measure is used*. The current selection of likelihood measures and effect size measures are inspired by the work of Andrew Hardie of Lancaster University.

Anthony (2024) reiterated this claim in a more recent post on *AntConc Google Group* as he answered to my inquiries on the same topic:

There is no established effect size measure that the field generally agrees with. [...] Each effect size measure will generate values on a different scale so they are not comparable. [...] There is no general agreement on how to interpret effect size measures. [...] For keyword analysis, I don't recommend using an effect size measure to rank the keywords. This is why the default setting in *AntConc* is to rank by log-likelihood (a statistical test, which is not an effect size measure, but may have an effect size component to it). [...] As each effect size measure exaggerates/weights different types of values (some weight low frequencies higher and others weight high frequency values higher), the results will vary dramatically depending on your choice of effect size measure. This is one reason why I don't recommend using them for keywords [...]

In this regard, aligned with Anthony's (2022b; 2024) statements above and as will be explained in the next subsection, the keywords from our doctoral research will be initially sorted out based on likelihood measure scores automatically generated by *AntConc* following its

default values. At last, we also agree with Brezina (2018), for whom not all tests and metrics available are God's truths, since they are all recommendations and, because of that, can generate different results and serve various purposes according to the researcher's choice (Pojanapunya; Watson Todd, 2016). In view of our hypotheses, research questions, and objectives, our investigation departs from the results generated based on these scores, but extends to a contextual keyword analysis:

[...] it should be borne in mind that a *table* of keywords does not in itself constitute a keyword *analysis*. The table of keywords is the beginning of the analysis, or more accurately, the part prior to the analysis. The actual analysis is what happens when we try to work out why keywords appear, which involves consideration of context both within the corpus and outside it. (Gillings; Mautner; Baker, 2023, p. 36, emphasis in original).

3.2.2 Selecting keywords for analysis

Based on the definitions presented in subsection 3.2.1, we now introduce the steps taken to select keywords for analysis.

1. Setting default keyness values on *AntConc* to obtain positive keywords

This step entailed setting default keyness values available in the program prior to obtaining positive keywords, as illustrated in Table 17:

Table 17. *AntConc* default keyness values

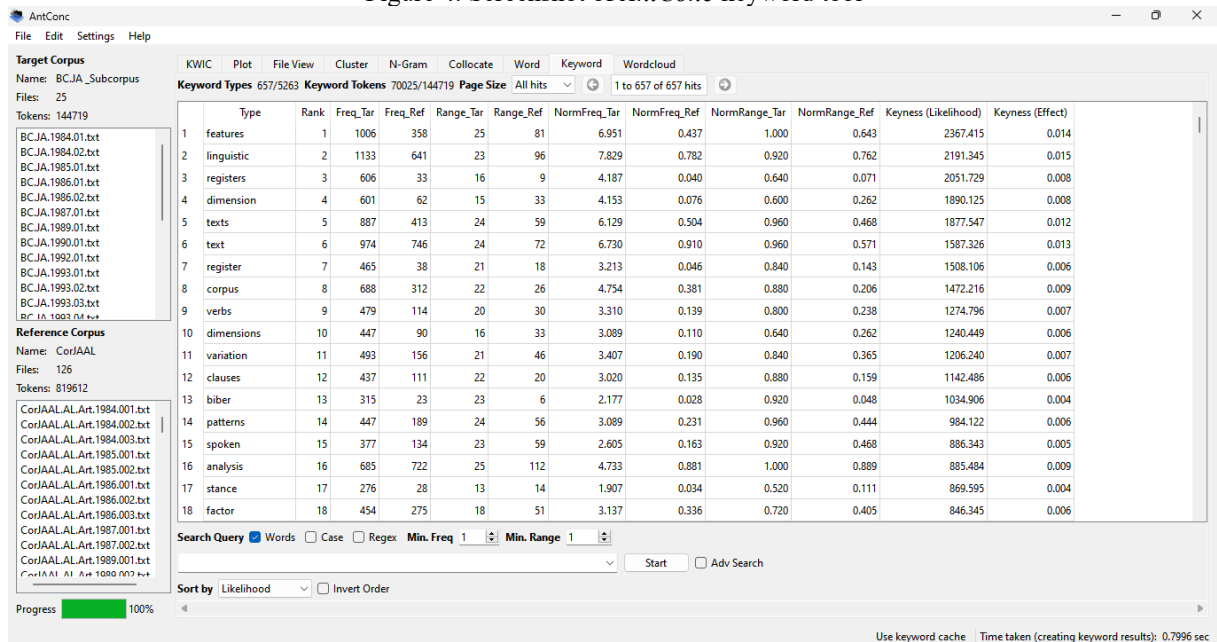
Tests/metrics	Measures	Threshold
Likelihood	Log-Likelihood 4-term	$p < 0.05$ 3.84 with Bonferroni ²⁷
Effect Size	Dice coefficient ²⁸	All Values

Source: Elaborated by the author.

After setting these values, both *BC* and *CorJAAL* were loaded in the program by using the *Corpus Manager* function. *BC* was used as the target corpus and *CorJAAL* as the reference. Upon selecting both corpora, the button *Start* in the *Keyword tool* was pressed, and a list of keywords was shown on screen, as seen in Figure 4:

²⁷ Napierala (2012, p. 1) highlights that “the Bonferroni correction is an adjustment made to P values when several dependent or independent statistical tests are being performed simultaneously on a single data set. To perform a Bonferroni correction, divide the critical P value (α) by the number of comparisons being made. For example, if 10 hypotheses are being tested, the new critical P value would be $\alpha/10$. The statistical power of the study is then calculated based on this modified P value. The Bonferroni correction is used to reduce the chances of obtaining false-positive results (type I errors) when multiple pair wise tests are performed on a single set of data.”

²⁸ Scott (2022) states that the dice coefficient used as an effect size measure varies from 0 to 1.

Figure 4. Screenshot of *AntConc* keyword tool

Source: Elaborated by the author.

As can be seen, Figure 4 shows an unarguably long list of keywords (657). However, it should be remembered that setting particular thresholds is troublesome, since this choice may limit the number of keywords retrieved by the program, not to mention that other potentially important keywords may be overlooked in a more specific manual analysis by the researcher (Gabrielatos, 2018; McIntyre; Walker, 2019), especially if cut-off points are too strict. With this in mind, it was necessary to choose what keywords to analyze, which brings us to the next procedure.

2. Selecting keywords for contextual analysis and eliminating other keywords

This step consisted of selecting some keywords for contextual analysis. Inevitably, this incurred in excluding other keywords. As will be explained in this subsection, these decisions were mainly guided by two reasons: theoretical, following guidance from other scholars; and research-oriented, according to the research questions and objectives of this dissertation.

Traditionally, researchers have limited themselves to analyzing only “top 10, 50 or 100 keywords” (Brezina, 2018, p. 85). Although dealing with long lists of keywords is time-consuming and may lead to errors, selecting some keywords over others may overlook important features of the corpus. In this regard, “a crucial next step when dealing with lists of key items [...] is working out which ones, if any, help most to answer your research questions.” (McIntyre; Walker, 2019, p. 283).

A good starting point for a keyness analysis is to try to *group similar keywords into categories*, then perhaps to select a few keywords for *a more detailed analysis*, via their concordance lines and collocates. [...] Instead, analyses are likely to be most engaging if they focus on *a set of connected keywords* which contribute something original to a coherent and convincing narrative. We would recommend that keywords which tell us *something of unexpected or non-obvious meaning* should take precedence over those which simply confirm what is already known. The analysis should ideally feature *illustrative examples* from the corpus texts as opposed to simply reciting frequencies and keyness scores. (Gillings; Mautner; Baker, 2023, p. 36, emphasis added).

Bearing in mind these suggestions and Anthony's (2022b; 2024), I ranked the keywords by likelihood and began grouping them into three major categories: text-aboutness; non-text-aboutness metadiscourse; non-metadiscourse and non-text-aboutness. Only the non-text-aboutness metadiscourse keywords underwent contextual analysis in this dissertation. It should be mentioned that all this selection included keywords that were not only on top 100, but also above this classical threshold (Brezina, 2018). For this reason, the issue of sorting by likelihood or effect size did not bear much importance, because all the keywords were eventually categorized, including those which had shown varying likelihood and effect size values.

As previously mentioned (Gabrielatos, 2018; Pojanapunya; Watson Todd, 2016; Scott, 2022), the first type of keywords usually regard themes, topics, and other information about the texts (i.e., methodological approaches, data, results, discussion, conclusion, etc.). In our research, they were mainly nouns and adjectives. Later on, more emphasis was placed on the second type of keywords by carrying out contextual analysis of non-text-aboutness metadiscourse keywords (Gillings; Mautner; Baker, 2023). These comprised both content and function words, including verbs, adverbs, adjectives, and prepositions. Because they were the focus of this dissertation, they were categorized in more detail, as will be introduced in Chapter 04. It should be mentioned that previous research has already upheld the use of grammatical words as authorial style markers (Bondi, 1996; Hyland, 2008b; Scott, 2022; Grieve, 2023; Gillings; Mautner; Baker, 2023). At last, other keywords that were neither text-aboutness nor part of the stylistic model of metadiscourse were discarded, because a style analysis must be very selective (Leech; Short, 2007).

Contextual analyses were carried out using three other tools available on *AntConc*: *KWIC*, *Cluster*, and *Collocate*:

KWIC (Key-Word-In-Context) Tool

This tool shows search results in a concordance or 'KWIC' (Key-Word-In-Context) format. This allows you to see how words and phrases are commonly used in a corpus of texts.

Cluster Tool

The tool shows contiguous (together in a sequence) word patterns based on the search condition. This allows you to see common phrases that appear in the target texts.

Collocate Tool

This tool shows words that appear frequently within a certain distance of the search term (i.e., collocates). This allows you to find which words co-occur with other words in a corpus. (Anthony, 2022b, p. 2-3, emphasis in original)

Specifically for the *Collocate tool*, the default values were maintained, and many of them were the same as used in the *Keyword Tool*. In fact, as Anthony (2023, p. 1-2) states:

Collocation measures are analogous to keyness measures in that they are used to identify and rank the degree to which words in a corpus appear unusually frequently in a context span surrounding a target word (i.e., collocate with the target word) compared with their occurrence in the corpus as a whole.

The default collocate values adopted by the program are introduced in Table 18:

Table 18. *AntConc* default collocate values

Tests/metrics	Measures	Threshold
Likelihood	Log-Likelihood	$p < 0.05$ 3.84 with Bonferroni
Effect Size	MI (Mutual Information)	All Values

Source: Elaborated by the author.

McIntyre and Walker (2019) highlight that MI figures among the most frequently used collocational significance tests. MI score is calculated “[...] by dividing the observed frequency by the expected frequency and determining the base 2 logarithm of the result [...]”. (p. 166). This score determines “[...] the number of times that the base number 2 needs to be multiplied by itself in order to arrive at whatever the observed frequency divided by the expected frequency happens to be”. (McIntyre; Walker, 2019, p. 166-167). If the score is higher than 3, it indicates a strong collocation.

We noted that some keywords collocated with other words (some even key), forming lexical bundles, illustrating Sinclair’s (2004) co-occurrence principle and that isolated tokens may be little informative (Berber Sardinha, 2004). To better locate these bundles, we used the previously introduced *Cluster tool* and adopted Biber *et al.*’s (2021, p. 985) cut-off points. To be considered a bundle in this investigation, the expression should appear in at least five (05) texts (minimum range) and recur at least 10 times (minimum frequency). Another issue was raised even when adopting this cut-off point as the program returned many bundles: they needed to be classified as metadiscoursal according to our model, otherwise they would be deleted from the analysis.

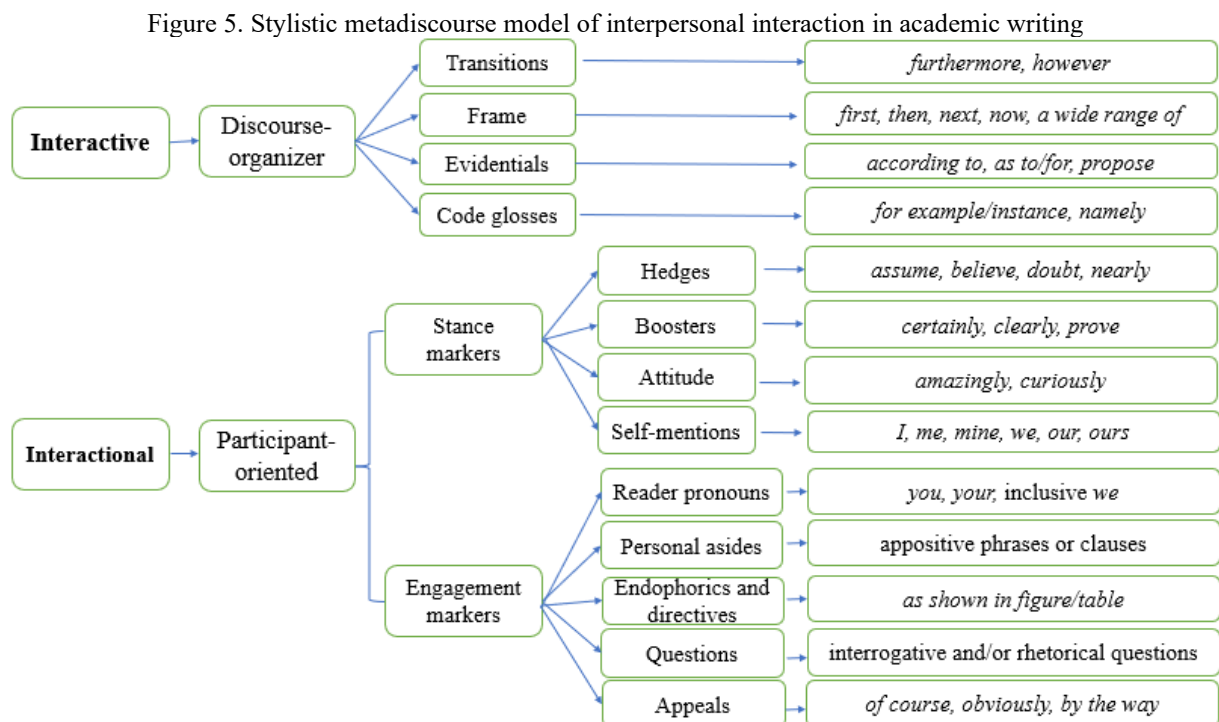
As will be introduced in Chapter 04, when reporting on the quantitative use of metadiscourse keywords, normalized frequencies will be shown. This type of frequency is obtained automatically when you click on the *Show Normalized Values* icon in most of *AntConc*

tool settings. I have set the value to 1,000 words in the *Statistics global settings*, although this option can be altered at any time to suit the user's preference. This calculus is exemplified in Biber (2006): the number of types is divided by the number of tokens and the result is multiplied by one thousand (1000/normed rate). For instance, take the case of the keyword *registers*: it occurs 606 times across *BC* (which has 144,719 tokens): $(606 / 144,719) * 1000 = 4.187$. So the normed rate of the keyword *registers* in *BC* is 4.187 per 1 thousand words.

Upon setting these parameters and following the steps described in this section, I proceed next with introducing the *Stylistic Metadiscourse Model of Interpersonal Interaction in Academic Writing*, based on Hyland (2005; 2008a; 2009), Biber (2006), and Biber *et al.* (2021).

3.2.3 The Stylistic Metadiscourse Model of Interpersonal Interaction in Academic Writing

Figure 5 below briefly introduces the model split into two major dimensions: *interactive* and *interactional*. Each dimension has their specific functional categories (*discourse-organizer* and *participant-oriented*, respectively), with each also including corresponding functional subcategories (*transitions*, *frame markers*, etc.). A fuller version is provided in APPENDICES A and B.



Source: Adapted from Hyland (2005; 2008a; 2009), Biber (2006), and Biber *et al.* (2021).

The examples of words, clauses, and phrases provided in Figure 5 have been taken directly from the authors' original models. The application of this model to our corpora data will be presented in Chapter 04 only. It should be stressed that metadiscourse, as understood in this model and elsewhere across this dissertation, is a macro functional language feature, which consists of functional subcategories, such as *transitions*, *frame markers*, among others. These are realized by grammatical features, such as adjectives (in the case of *attitude markers*), verbs (in *evidentials* and *boosters*), adverbs (in *transitions*), etc.

I need to highlight, however, that this classification is not entirely uncomplicated, because many of the subcategories are ambiguous and could easily be interchangeable. Words, phrases, and clauses can equally be allocated in more than one subcategory, depending on the contexts in which they are used and the collocates with which they combine. After all, language items realize different functions and can be classified in distinct subcategories.

For instance, Biber (2006) includes some verbs into the category he labels as *speech act and other communication verbs*, because, for him, “[...] they distance the speaker/writer from the truth of the proposition in the complement clause” (Biber, 2006, p. 95), like the verbs *help*, *say*, *tell*, among others. If we look at Hyland's (2005) model, these verbs could be included in two different categories: interactive (*evidentials*) and interactional (*boosters / attitude markers*). Hyland (2005, p. 51-52) states that “evidentials distinguish *who* is responsible for a position and while this may contribute to a persuasive goal, it needs to be distinguished from the writer's *stance* towards the view, which is coded as an interpersonal feature”. Considering both issues, some speech act and other communication verbs from Biber (2006) have been categorized as *evidentials* when they explicitly referred to an evident distance between writer and proposition, pointed the readers to specific information from other texts, and led writers to report on but not take a stance towards cited scholars' research, findings, and quotations.

Since many keywords from our research corpora are verbs, Table 19 provides a semantic classification of single-word verbs based on Biber *et al.* (2021, p. 360-363):

Table 19. Semantic classification of single-word verbs

Semantic domains	Functions	Examples
activity	denote actions and events that could be associated with choice; express non-volitional actions, events, or static relations; express volitional activities	<i>bring, buy, carry, come, give, go, leave, move, open, run, take, work</i>
communication	involve communication activities (considered a special subcategory of activity verbs)	<i>ask, announce, call, discuss, explain, say, shout, speak, state, suggest, talk, tell, write</i>
mental	denote activities and states experienced by humans; include cognitive and emotional meanings, attitudes or	<i>think, know, love, want, see, taste, read, hear, believe, doubt, know,</i>

	desires, perception, receipt of communication	<i>remember, understand, enjoy, fear, feel, hate, like, prefer, suspect</i>
facilitation or causation	indicate that some person or inanimate entity brings about a new state of affairs	<i>allow, cause, enable, force, help, let, require, permit</i>
simple occurrence	primarily report events (typically physical events) that occur apart from any volitional activity	<i>become, change, happen, develop, grow, increase, occur</i>
existence or relationship	report a state that exists between entities, a particular state of existence or a particular relationship between entities	<i>be, seem, appear, exist, live, stay, contain, include, involve, represent</i>
aspectual	characterize the stage of progress of some event or activity	<i>begin, continue, finish, keep, start, stop</i>

Source: Adapted from Biber *et al.* (2021, p. 360-363).

Although not identified as key in our corpus, modal verbs nonetheless pose problems to the introduced model, and so some explanations should be given. Biber *et al.* (2021, p. 970) state that “modal verbs are problematic, since they have many meanings and are often ambiguous. However, most occurrences of modal verbs can be considered as indicating either epistemic or attitudinal stance [...]”. For example, *will* and *be going to* are reasonably comprehended as marking epistemic stance when referring to predictions that will become true sometime in the future, and so have been included in the *hedges* subcategory. We have also included modal verbs and other modality words in three subcategories, according to the prevailing meaning they conveyed in their original context, namely *hedges*, *boosters*, and *attitude markers*. This selection also included modals like *can* and *could*, but only when they expressed possibility.

Biber (2006) and Biber *et al.* (2021) label one type of stance marker as *style of speaking*, arguing that it presents speaker’s or writer’s comments on the communication itself. This understanding is no different from Hyland’s interactional feature. Therefore, they are considered synonyms in the proposed model, and all stance markers labeled as *style of speaking* are made part of interactional subcategories, such as *hedges*, *boosters*, and *attitude markers*. For instance, style-speaking stance markers like *honestly* and *frankly* are visibly evaluative in practice, so it seems unwise to exclude them from the *attitude markers* subcategory and simply label them as *style*.

Another confusion is made when personal pronouns are used, especially due to their collocates. Such examples as *my purpose is*, *(as) I argue here*, *I hope to persuade* and *let us return to* were originally labeled as *frame markers* in Hyland (2005), but because of the use of the pronouns *me*, *I*, *us*, they could equally be included in *self-mentions*. In all such cases, the discourse acts of both convincing the audience of the research goals and propositions and

indicating topic shifts seemed prevalent, even if including *self-mention* personal pronouns within, and thus the original classification was followed.

Finally, from the *engagement markers* category, *endophoric markers* also set ambiguous relationships with *directives*. Both resources, however, serve a polysemic purpose: writers direct readers to particular points in the text and influence them to reach specific understandings and conclusions. For this reason, we have removed *endophoric markers* from its original *interactive* feature and placed it within the *interactional engagement markers*, together with *directives*. Upon introducing this model, I proceed next with presenting and discussing our data following the application of the *Stylistic Metadiscourse Model of Interpersonal Interaction in Academic Writing*, based on Hyland (2005; 2008a; 2009), Biber (2006), and Biber *et al.* (2021).

CHAPTER 04 – Douglas Biber’s metadiscourse style

This chapter presents functional analyses of Douglas Biber’s metadiscourse style. It begins with an overview of the keywords results by presenting the types of words identified as key from the comparison between *BC* and *CorJAAL*, and then proceeds with the presentation and discussion of linguistic data following the application of the *Stylistic Metadiscourse Model of Interpersonal Interaction in Academic Writing*.

4.1 Overview of the keywords results

In this section, we start by answering the first research question posed in the Introduction:

1. What kind of words appear as positively key in Douglas Biber’s RAs when compared to RAs written by other applied linguistics scholars?

Overall, Table 20 illustrates the number of positive keywords resulting from the comparison of *BC* to *CorJAAL*.

Table 20. Number of positive keywords

Types	Tokens
657	70,025

Source: Elaborated by the author.

As mentioned in the previous chapter, these keywords were ranked by likelihood (Anthony, 2022b; 2024). From this total, 54 were excluded because they only occurred in one text. As we understand here, for words to be considered stylistic, one of the requirements is that they should be consistently used by authors in more than one text. Consistency, here, is not synonymous to frequency, but to dispersion (the degree an item recurs throughout texts in a given corpus). Furthermore, the remainder of the keywords equaled 603, which were split into three major categories, as illustrated in Table 21. The full list of all these words can be found in Table 42 in ANNEX B.

Table 21. Categorization of keywords

Keyword categories	Types	Percentage
Text-aboutness	449	74,46%
Metadiscourse	126	20,90%
Non-text-aboutness and non-metadiscourse	28	4,64%
TOTAL	603	100%

Source: Elaborated by the author.

As can be seen, almost 75% of the keywords retrieved by the program are of the text-aboutness type. This is not an unexpected finding, though, since previous research has found that this category of keywords tends to be highly frequent (Gabrielatos, 2018; Geluso; Hirsch, 2019; Scott, 2022). For example, in top 10 in *BC*, the first most frequently used text-aboutness keywords are *linguistic*, *features*, *text*, and *texts*. When we expand the search span to the first 100 keywords, it is not uneasy to find such nouns and adjectives as *corpus*, *analysis*, *registers*, *dimensions*, *variation*, *written*, *patterns*, *factor*, *spoken*, among others. Unarguably, these keywords regard themes, topics, and other relevant information in Biber's texts, including methodological approaches, data analysis, and discussion of results. These patterns illustrate what has long been one of the scholar's main research interests over the last decades: the application of his multi-dimensional (MD) approach (Biber, 1988) to the study of spoken and written language variation in different registers, such as news, fiction, conversation, academic writing, etc. In an interview with Bethany Gray, Biber (2013) asserts that his interest in register variation emerged from the corpus studies:

[...] I was repeatedly confronted with findings that demonstrated the importance of register differences, so I became 'interested'! [...] Thus, my preoccupation with the importance of register (or genre) variation dates back to a very early stage! [A register perspective] describes patterns of language use in all spoken (and written) registers, and it describes patterns of language use in terms of the full inventory of lexicogrammatical characteristics. [...] because register has often been disregarded by linguists of many persuasions, I have been pretty persistent in arguing for the importance of register. (Biber, 2013, p. 359-361).

This interest is made evident in the word choices across *BC*. Many of these keywords are also collocates to one another, such as *linguistic features*, *MD analysis*, *register variation*, *spoken register(s)*, *written register(s)*, etc. As can be seen in some extracts²⁹ highlighted below, Biber mentions elements included in an analysis (1), urges for a need in corpus analyses (2), and explains what MD studies entail (3 and 4):

²⁹ All the cited corpus extracts in this subsection and elsewhere across this dissertation are faithful reproductions from the texts and have been taken directly from *BC* using the *KWIC* tool on *AntConc*.

1. In the present case, a factor analysis of forty linguistic features in the texts of the LOB and London-Lund corpora was done. (BC.JA.1987.01.txt)
2. Corpus-based analyses of register variation obviously need to be extended in several ways. (BC.JA.1993.04.txt)
3. MD studies have demonstrated that no single parameter or dimension is adequate in itself to capture the full range of variation among registers in a language. (BC.JA.2019.01.txt)
4. The overall comparison of spoken and written registers in the 1988 MD analysis requires consideration of all dimensions of variation, which each define a different set of relations among spoken and written registers. (BC.JA.2014.01.txt)

The second most frequently occurring keywords fell into the metadiscourse type, which will be detailed later. Before that, let us explore keywords of the third type as shown in Table 21. These are words that figure as neither text-aboutness nor metadiscourse, accounting for almost 5% of the findings. Words of this kind include demonstrative pronouns (*these*), prepositions (*versus*, *among*), multifunction words (*each* as pronoun, adverb, adjective), abbreviations (WH), among others. For instance, Biber refers to two approaches to be followed in corpus design (5), relations which exist between text types in given MD-derived dimensions (6), MD-based studies conducted in different languages other than English (7), and an algorithm used to identify a particular type of relative clause (8):

5. These two approaches would address the question of what people actually do with language on a regular basis. (BC.JA.1993.03.txt)
6. These derived variables (known as factor scores) enable further interpretation of the textual dimensions underlying the factors and examination of the relations among the text types along *these* dimensions. (BC.JA.1985.01.txt)
7. After all, each of these studies is based on the linguistic features found in that language, and the corpora analysed include the distinctive registers found in that language. (BC.JA.2014.01.txt)
8. The algorithm for WH relative clauses depends on an accurate identification of all nouns, adverbs, and verbs—in absolute terms and then matched in the correct sequence. (BC.JA.1995.01.txt)

Furthermore, as for the second research question of this doctoral investigation:

2. Are there any keywords that can be classified as metadiscoursal? If so, what is their number of raw and normed values after categorized according to the *Stylistic Metadiscourse Model of Interpersonal Interaction in Academic Writing*?

As illustrated in Table 21 above, metadiscourse keywords account for over 20% among those retrieved as positive by *AntConc*. This number refers to two subtypes of metadiscourse keywords: readily metadiscoursal and potentially metadiscoursal. The difference among them

is that, in the former case, some keywords were promptly identified as metadiscoursal, such as the transition marker *thus*, which will be introduced in section 4.2.1. On the other hand, in the latter case, some keywords were candidates to metadiscourse, such as *example*. On its own, *example* is not metadiscoursal, but it becomes so when collocating with *for*, forming *for example*, as will be introduced in section 4.2.3. In this regard, we identified 159 metadiscourse items, including single keywords and keyword-derived bundles, besides the initial 126 keywords. All in all, these results also confirm the first hypothesis of this study, in that not only is metadiscourse a feature of academic writing, but also part of an author's stylistic language repertoire. The total number of types and tokens of key language items categorized according to the *Stylistic Metadiscourse Model of Interpersonal Interaction in Academic Writing* is presented in Table 22:

Table 22. Number of stylistic metadiscourse interactive and interactional keywords across *BC*

Interactive	types	raw tokens	normed tokens
<i>transitions</i>	7	1,248	8,62
<i>evidentials</i>	62	1,036	7,16
<i>code glosses</i>	4	899	6,21
<i>frame markers</i>	22	736	5,09
subtotal	95	3,919	27,08
Interactional	types	raw tokens	normed tokens
<i>boosters (stance)</i>	43	2,267	15,66
<i>attitude markers (stance)</i>	8	730	5,04
<i>endophoric and directive markers (engagement)</i>	12	226	1,56
<i>hedges (stance)</i>	1	45	0,31
subtotal	64	3,268	22,58
TOTAL	159	7,187	49,66

Source: Elaborated by the author.

As can be seen in Table 22, Biber's texts show more instances of interactive over interactional key metadiscourse features. For example, for the former case, they show more *transition markers* than *evidentials*, *code glosses*, and *frame markers*. For the latter, they favor the marking of stance through *boosters*, *attitude markers*, and *hedges*, with some engagement patterns through *endophoric and directive markers*. It is important to note that no positive keyword that could be categorized as either *self-mention*, *reader pronoun*, *personal aside*, *question* or *appeal to shared knowledge* was found key, and that is the reason why these subcategories do not appear in Table 22.

Let us compare these results to Hyland's (2005, p. 92; 144). When looking only at the RA irrespective of the discipline, Hyland's (2005) findings indicate that there is a tendency to apply more interactive items, favoring, respectively, *transitions*, *code glosses*, *evidentials*, and *frame markers* (Hyland, 2005, p. 92). However, when considering disciplinary variation together with genre, RAs written in the area of applied linguistics favored an interactional over an interactive approach, showing a preference, respectively, for *hedges*, *attitude markers*, *boosters*, *self-mentions*, and *engagement markers* (Hyland, 2005, p. 144). Furthermore, if we compare our corpus results to Hyland and Jiang's (2019), we notice that Biber's choices of *boosters* differ from the current trend of applied linguistics texts, suggesting that Biber's style is even more interactive than other RAs written in the field. However, his use of *directives* and underuse of *self-mention*, *asides*, and *references to shared knowledge* comply with what has been commonplace in the area, implying that his style may be as similarly interactional as others conveyed in RAs in the field.

Given this overview of the keywords results, as of the next subsection, more will be discussed about Biber's interactive and interactional language uses. To continue, we will answer the third research question in the next section.

4.2 Biber's interactive metadiscourse style

In this section, we answer the third research question posed in the Introduction:

3. Does Biber employ interactive metadiscourse language features? If so, how does he organize discourse in his RAs?

As previously shown in section 4.1, Biber employs interactive metadiscourse features, such as *transition markers*, *evidentials*, *code glosses*, and *frame markers*. Table 23 exhibits how these discourse organizers subcategories are grammatically realized through adverbs, adjectives, conjunctions, verbs, proper nouns, prepositions, and bundles. It is worth mentioning that these features often chosen by Biber illustrate many of the distinguishing features commonly used in different academic writing registers, including RAs (Biber *et al.*, 1999; Biber, 2006; Gray, 2015). Examples from each of those metadiscourse discourse organizers subcategories will be discussed in the subsections 4.2.1 to 4.2.4.

Table 23. Categorization of interactive positive keywords from *BC*

Interactive Metadiscourse Feature	
Discourse organizers subcategories	Linguistic features and examples
<i>Transitions</i>	• Adverb / adjective / conjunction / bundles: • addition: <i>in addition;</i> • consequence: <i>thus;</i> • comparison / contrast: <i>however; while; rather; rather than; similar to; in contrast.</i>
<i>Evidentials</i>	• Verbs: <i>occur; represents; representing; represent; show; shows; shown; identify; identified; include; included; illustrate; illustrates; interpreted; analyze; analyzed; presents; reflect; considered; describe; describing; carried; investigate; uncovered; analyzing; disregarded; characterize; compares; uncover; compute.</i> • Proper nouns: <i>Biber; Finegan; Chafe; Conrad; Johansson; Halliday; Sinclair; Francis; Hymes; Donnell; Gray; Poole; Reppen; Altenberg; Blankenship; Beaman; Gorsuch; Szmrecsanyi; Tottie; Akinnaso; Besnier; Nukulaelae; Renouf; Svartvik; Hared; Longacre; Mair; Bergen; Haan; Mahlberg; Comrie; Geisler; Kay; Kucera.</i>
<i>Code glosses</i>	• Adverb / preposition / bundles / others: <i>for example; including; e.g., i.e.</i>
<i>Frame markers</i>	• Adverb: <i>finally.</i> • Bundles: <i>with respect to; the extent to which; the present study; in the case of; a wide range of; in this case; in the use of; the present paper; a broad range of; a large number of; on the basis of; the full range of; in the present study; the following sections; to a lesser extent; the present analysis; of the present study; in the present paper; in the present case; as the basis for; the present section.</i>

Source: Elaborated by the author.

Given the plethora of language choices illustrated in Table 23, in this section³⁰ we will see how Biber organizes discourse in his RAs and, as a consequence, favors an interactive metadiscourse style. Precisely, we will assess how Biber's individual stylistic choices establish contrastive, comparative, additive, and consequence relations between stretches of discourse (through *transitions*); refer readers to information from other texts, reporting on other scholars' research findings with no stance taken towards them (through *evidentials*); provide readers with extra information about the text by rephrasing, explaining, or elaborating on what had just been discussed (through *code glosses*); sequence, label, and shift arguments, making the discourse visible to the reading audience (through *frame markers*). As will be seen, these rhetorical practices reveal Biber's "[...] awareness of a participating audience and the ways [he] seeks to accommodate its probable knowledge, interests, rhetorical expectations and processing abilities." (Hyland, 2005, p. 50).

³⁰ The metadiscourse features introduced in subsections 4.2.1 to 4.2.4 do not follow their order of appearance as illustrated in Figure 5. Rather, they are presented here in accordance with how frequent they were used across *BC* in comparison to *CorJAAL* (cf. Table 22).

4.2.1 Transitions

Table 24 displays the raw and normed frequencies and both text and time ranges of *transition markers* used in *BC*. As will be seen in this subsection, Biber establishes more contrastive and comparative relations between clauses by favoring the use of *however*, *while*, *in contrast*, *rather than*, *rather*, and *similar to*. He also chooses to convey additive meanings by employing *in addition*, followed by the expression of consequence through *thus*. Furthermore, as you can see in Table 24, some markers are bundles, namely *in contrast*, *rather than*, *similar to*, *in addition*. In our first assessment of the keyword list, though, *contrast*, *rather*, *than*, *similar*, and *addition* were identified as key, but when we checked for their collocates, we observed that they formed bundles and were metadiscoursal. To avoid inflated results, the frequency of single metadiscourse keywords was excluded from that in which they were part of lexical bundles. For example, the keyword *rather* occurred 192 times in *BC*, but 89 times alone and 96 times as part of the bundle *rather than*, while 7 times functioning as non-metadiscoursal. This same procedure was executed in all cases where keywords were metadiscoursal alone and as part of lexical bundles.

Table 24. Keywords functioning as *transition markers* across *BC*

<i>Transition markers</i>	Frequency		Range	
	Raw	Normed	Text	Time slot
however	290	2,00	25	1,2,3
thus	264	1,82	25	1,2,3
while	209	1,44	23	1,2,3
in contrast	157	1,08	23	1,2,3
rather than	96	0,66	24	1,2,3
rather	89	0,61	25	1,2,3
in addition	78	0,54	24	1,2,3
similar to	65	0,45	18	1,2,3
TOTAL	1,248	8,62	-	-

Source: Elaborated by the author.

Transition markers are a common resource used to establish different types of relations between main clauses (Hyland, 2005). In *BC*, a few *transition markers* are employed to establish at least three kinds of relation: addition, consequence, and comparison/contrast. Biber's stylistic choices recur in almost all 25 texts and in all three time slots, proving his consistent decision for these markers across his academic production and time.

Contrastive and comparative relations between clauses are more often noticed. For instance, *while* is used 1,44 times per 1 thousand words. Biber *et al.* (2021) show that *while* is mainly used as a subordinator to refer to time and establish concession/contrast relationships between clauses. In *BC*, the use of this conjunction as an adverbial subordinator to mark

concession/contrast is noticed in 94% of the cases, proper of an interactive metadiscourse style (Hyland, 2005). The remainder accounts for its use as a time reference subordinator, but this frequency was ignored from the values in Table 24. As seen in excerpt 9, Biber contrasts information related to positive features in the main clause to information on negative features in the adverbial clause. The choice of *while* here illustrates how the author is aware of audience's processing abilities to grasp the type of comparison established between the clauses:

9. [...] the positive features were indicative of an interactional focus (questions, first and second person pronouns, present tense, reduced forms, and subordinate forms) while the negative features indicated a content focus (type/token ratio and word length). (BC.JA.1986.01.txt)

Furthermore, the only key consequence marker as linking adverbial is *thus*, used 1,82 times per 1,000 words. Academic prose has preference for this marker, reflecting “[...] the characteristic choice of this register to mark the links between ideas overtly, as these arguments are developed” (Biber *et al.*, 2021, p. 872). Across *BC*, as illustrated in example 10, by using *thus*, Biber establishes a point in discourse where he transitions between two sentences: the first, in which he states that there is a specific procedure to extract variance in two different factors, and the second one, in which he begins explaining how the extraction occurs in the first factor. In example 11, the scholar uses *thus* to link the clauses that explain the nature of corpus-driven studies and how these should be seen. In both cases, this form of transition helps accommodate the participating audience's comprehension need:

10. This procedure extracts the maximum amount of shared variance among the variables in the data for each factor. Thus, the first factor extracts the maximum amount of shared variance in the data, i.e., the largest grouping of co-occurrences in the data. The second factor then extracts the maximum amount of shared variance from the tokens left over after the first factor has been extracted, and so on. (BC.JA.1985.01.txt)
11. [...] corpus-driven studies often incorporate some corpus-based methods, and thus they might be best considered as hybrids. (BC.JA.2009.01.txt)

In the first case, the use of *thus* is in initial position, and in the second in medial position. As a matter of fact, 56% of *thus* appears in initial position across *BC*. Biber's preference for *thus* and in this position stresses an individual stylistic trait, particularly if we consider that *thus* could have been replaced by *therefore* or *hence*.

At last, the only key language item to convey addition in *BC* is *in addition* used as linking adverbial 0,54 times per 1,000 words. In example 12, Biber compares two text types. After laying out similarities between them, he shows differences. In example 13, the author provides additional characteristics of cluster eight in analysis. In both cases, it is possible to

note how he chooses to list particular details of the entities analyzed so as to base his arguments on and direct readers' attention to them:

12. In summary, both text type 3 and text type 4 are expository, extremely informational, and explicit in reference. The difference between them relates to their use of an abstract, technical style. [...] In addition, the differences between these types seem to reflect attitudinal preferences. (BC.JA.1989.01.txt)
13. But there are, in addition, oblique object arguments which are formally marked in Somali, although not by means of nominal inflections. (BC.JA.1984.02.txt)

Another aspect to consider when looking at *in addition* is the position in which it occurs in the sentences. In almost 77% of the cases, it appears in initial position, highlighting “[...] explicitly the connection between units of discourse at the point when the connection is [...] being made” (Biber *et al.*, 2021, p. 883). Biber’s preference for *in addition* and in this initial position also denotes an individual stylistic preference, especially considering the fact that, in initial position, *besides*, *moreover*, or *furthermore* could have been used in place: “in academic prose, individual author preferences are also influential in the choices among linking adverbials” (Biber *et al.*, 2021, p. 881).

As seen in this subsection, it was possible to observe how Biber organizes discourse in his RAs by establishing contrastive, comparative, additive, and consequence relations between clauses through *transition markers*. Consequently, he favors an interactive metadiscourse style which helps him make connections between arguments and meet readers’ comprehension and interpretive needs, displaying a supportive and helpful authorial identity. Next in subsection 4.2.2, it will be explained how the author opts to refer to information from other texts through *evidentials*.

4.2.2 Evidentials

Table 25 displays the raw and normed frequencies and both text and time ranges of *evidentials* employed in BC. As will be seen, Biber often opts for proper nouns and verbs to point readers to information from other texts, reporting on scholars’ research and findings without taking a stance towards them.

Table 25. Keywords functioning as *evidentials* across BC

<i>Evidentials</i>	Frequency		Range	
	Raw	Normed	Text	Time slot
Biber	315	2,18	23	1,2,3
Finegan	45	0,31	14	1,2,3
Chafe	34	0,23	6	1,2,3
Conrad	32	0,22	9	2,3
identify	30	0,21	13	1,2,3

Johansson	28	0,19	13	1,2,3
identified	25	0,17	9	1,2,3
Halliday	24	0,17	8	1,2,3
shows	24	0,17	9	1,2,3
carried	23	0,16	10	1,2,3
Sinclair	22	0,15	9	1,2,3
Francis	17	0,12	9	1,2,3
represent	17	0,12	11	1,2,3
Hymes	16	0,11	6	1,3
shown	15	0,10	7	1,2
Donnell	14	0,10	4	1,2
analyze	14	0,10	9	1,2,3
Gray	13	0,09	3	3
Poole	13	0,09	4	1
Reppen	13	0,09	8	2,3
Altenberg	12	0,08	7	1,2,3
Blankenship	12	0,08	2	1
show	12	0,08	6	1,2,3
analyzed	12	0,08	4	1,2,3
describe	11	0,08	8	1,2,3
uncovered	11	0,08	5	1,2,3
analyzing	10	0,07	3	1,3
Beaman	10	0,07	4	1,2
Gorsuch	10	0,07	3	1
include	10	0,07	6	1,2,3
Szmrecsanyi	10	0,07	2	3
Tottie	10	0,07	8	1,2
Akinnaso	8	0,06	2	1
Besnier	8	0,06	7	1,2
Nukulaelae	8	0,06	4	1,2,3
Renouf	8	0,06	3	1,2,3
Svartvik	8	0,06	8	1,2
occur	8	0,06	7	1,2,3
describing	8	0,06	4	1,2,3
investigate	8	0,06	5	1,3
disregarded	7	0,05	3	2,3
Hared	7	0,05	4	1,2
Longacre	7	0,05	2	1,3
Mair	7	0,05	4	2,3
Bergen	6	0,04	5	1
Haan	6	0,04	3	2
Mahlberg	6	0,04	2	3
interpreted	6	0,04	5	1,2,3
considered	6	0,04	6	1,2,3
characterize	5	0,03	4	1,2
compares	5	0,03	4	1,2
Comrie	5	0,03	2	1
Geisler	5	0,03	3	2
Kay	5	0,03	3	1,2
Kucera	5	0,03	5	1
reflect	4	0,03	3	1,2
representing	4	0,03	3	1,2
illustrate	3	0,02	2	2,3
uncover	3	0,02	2	1
compute	2	0,01	2	2,3
presents	2	0,01	2	1,2
represents	2	0,01	2	2

TOTAL	1,036	7,16	-	-
--------------	-------	------	---	---

Source: Elaborated by the author.

Although proper nouns and verbs recur in almost all *BC* texts, differently from transitions, Biber's stylistic choices of particular *evidentials* are not consistent in all *BC* texts, most appearing in less than 30% of all texts and in the first and second time slots. This result can be explained by the fact that the majority of keywords functioning as *evidentials* in *BC* represent proper nouns, that is, the names of authors Biber cited across his texts, occurring 5,18 times per 1,000 words altogether. The reason why some names do not appear in all texts is logical and undeniable: some authors were cited only in the texts where their ideas were relevant for Biber's argumentation. The same issue applies to verbs, which overall recur 1,98 per 1,000 words, used only in the texts where they were necessary for Biber's report on cited work. All in all, these patterns illustrate what was similarly identified in Hyland (2005), in that *evidentials* are employed with persuasive and argumentative purposes, especially in the form of citations and self-citations, thus characterizing an interactive metadiscourse style.

In general, the use of citations represents a need "[...] to gaining approval of new claims by providing persuasive support for arguments and demonstrating the novelty of assertions" (Hyland, 2005, p. 106). Never before have authors made so many citations. Today, they are "[...] the currency of the scholarly economy and having one's work recognized and referenced by others is an increasingly valued commodity in today's fiercely competitive academic world" (Hyland; Jiang, 2019, p. 72). This pattern is also a disciplinary convention of applied linguistics and other fields in the soft and hard sciences, illustrating how the rhetorical conventions identified in texts (Swales, 1990; 2004) reflect aspects of the social and epistemological assumptions of the disciplinary culture of which the author is part.

Despite these trends, no journal publishing company makes specific determination regarding which type of citation ought to be favored, nor which evidential verbs Biber should choose, as was earlier introduced in Chapter 03. In *BC*, *evidentials* through citations and self-citations were better identified by proper nouns and verbs. As for verbs, a semantic classification based on Biber *et al.* (2021) and their normed frequencies (1,98 times overall) follow in Table 26:

Table 26. Semantic classification of evidential verbs and their normed frequencies in *BC*

Activity	Communication	Mental	Existence	Occurrence
0,82	0,13	0,56	0,42	0,06
analyze; analyzed; analyzing; carried; investigate; show; shows; shown	describe; describing; mark	compares; compute; considered; disregarded; identify; identified; interpreted;	characterize; illustrate; illustrates; include; included; presents; reflect; represents; representing; represent; uncover; uncovered	occur

Source: Elaborated by the author.

As can be seen, Biber employs more activity verbs, followed by mental, existence, communication, and occurrence verbs. This finding reveals that the author's style favors the expression of actions and events across RAs, particularly through activity verbs. In addition, we can see how these language choices corroborate with Biber *et al.* (2021, p. 365), in that “in academic prose, existence verbs are almost as common as activity verbs”.

Some verbs from Table 26 also introduce/refer to citations Biber makes in his texts, including self-citations, illustrating Biber's individual preference for some verbs over others. This strategy of self-reference in RAs “[...] is obviously an important means of demonstrating one's disciplinary credentials and credibility.” (Hyland, 2001, p. 213). Nowadays, while this practice has been seen as inflating citation rates and a way for author's vanity and self-promotion, with more specialized publications, more co-authored papers, and fiercer competition to publish texts, “[...] self-citation may now be a more significant rhetorical and tactical tool in the struggle for visibility and authority.” (Hyland; Jiang, 2019, p. 91).

In *BC*, in fact, the proper noun *Biber* also figures on the keyword list, and is the most used proper noun, appearing in almost all texts and time slots. This result illustrates both Scott's (2022) claim that the keyword tool also reveals words of this kind and Hyland and Jiang's (2019) finding that self-citation discloses scholars who have a long history of engagement in an area. For instance, in example 14 below, by using the activity verb *analyze*, Biber makes readers aware of a study he conducted with Hared in which the scholars carried out the action of *analyzing* the early development of six press genres:

14. Biber and Hared analyze the early development of six press genres with respect to several dimensions of variation. (*BC.JA.1993.02.txt*)

Biber also chooses mental verbs in his texts. In example 15, he takes readers to some studies carried out by scholars who denoted the mental activity of *identifying* underlying dimensions of stylistic variation in English writing:

15. A similar methodology has been used by Carroll and Marckworth and Baker to identify underlying dimensions of stylistic variation in English writing. (BC.JA.1986.01.txt)

As can be seen, this conventionalized community practice of placing one's own work within the body of existing literature through citations is an important instrument in academic discourse. Although this may not be claimed as Biber's own voice here, the way he chooses to locate his own investigation in the grounds of already published research is what characterizes his individualized voice.

Biber also makes reference to a study conducted by some authors through the existence verb *illustrates*. In example 16, by quoting these scholars, Biber leads readers to published research that reveals how multi-dimensional scaling is used:

16. The use of multivariate statistical techniques for lexical knowledge acquisition was first proposed by Bindi et al. In particular, that study illustrates the use of Multidimensional Scaling to identify the relations among semantically related word types (e.g., piccolo, corto, breve) and their collocates (or 'word mates'). (BC.JA.1993.01.txt)

In most of the cases, the keyword *Biber* collocates with other proper nouns, like another key one, *Finegan*. In 99% of the times, the keyword *Finegan* accompanies the keyword *Biber*, such as in the following quote in which Biber cites a study that he had carried out with Ghadessy and Finegan and used corpus-based techniques for register investigation. In example 17, the choice of the existence verb *include* takes the audience to recognize the particular state of existence between the authors' studies and other recent edited collections:

17. In recent years, most analysts studying registers have begun to use corpus-based techniques; recent edited collections with studies of this kind include Ghadessy and Biber and Finegan. (BC.JA.1996.01.txt)

The identification of *Finegan* as keyword may be expected. In fact, the Edward (or Ed) Finegan referred to in the quotes has been of much importance to Biber himself, as reported in an interview (Biber, 2013). Biber's (2013) claim that he enjoys cooperating with coauthors and working in collaborative teams is also shown in other findings across *BC*, representing his authorial identity of a card-toting community member. Other authors' names appear as collocates to the keyword *Biber*, including *Conrad*, *Reppen*, *Gray*, *Egbert*, among others. In example 18, Biber uses the communication verb *describe* to address the (under-)use of

(semi-)modals in a study carried out by himself, Conrad, and Reppen, and in doing so he equally persuades the audience to accept the validity of his assertions:

18. Biber, Conrad, and Reppen describe a general decline in the use of modals, accompanied by a rise in the use of semi-modals, over the past three centuries. (BC.JA.2004.01.txt)

As can be seen, the use of evidential verbs in citations and self-citations provides persuasive support for arguments and demonstrates novelty of assertions, search for authority and visibility, and allegiance to disciplinary conventions. In some cases, the same verbs are used in both citations and self-citations with persuasive and argumentative reasons. In example 19, Biber persuades readers to see that the most common verbs in English occur in two tenses, and in example 20, he presents the frequency of occurrence of (lexical) verbs in a given register. The choice of the occurrence verb *occur* in both cases, together with examples and numerical results, report factual events that were identified in at least two different studies:

19. For example, some of the most common verbs in English occur most of the time in the simple present tense (eg., think, know, want, mean), while others occur more often in the simple past tense (eg., said, came, took) (Kennedy) (BC.JA.2012.01.txt).
20. However, this is not at all the case: there are only 63 lexical verbs that occur more than 500 times per million words in a register, and only 12 verbs occur more than 1,000 times per million words in the LSWE Corpus (Biber et al.) (BC.JA.2005.01.txt).

Another interesting pattern in the practice of making citations and self-citations across *BC* concerns the place in the sentences where these references to cited authors are inserted. Swales (1990, p. 148) differentiates two forms:

An integral citation is one in which the name of the researcher occurs in the actual citing sentence as some sentence-element; in a non-integral citation, the researcher occurs either in parenthesis or is referred to elsewhere by a superscript number or via some other device.

In other words, integral citations are those where the cited author's name occurs within the citing sentence and outside parentheses, while non-integral forms place the author's name inside parentheses or by superscript numbers outside the citing sentence. As was seen in Chapter 03, no journal publishing company makes specific determination as to which citation format should be favored, but there is some consequence regarding the degree of emphasis assigned to the cited author's claim when either type is chosen:

The use of one form rather than the other appears to reflect a decision to give greater emphasis to either the reported author or the reported message. The use of a reporting verb to introduce the work of other researchers is also a significant rhetorical choice. (Hyland, 2004, p. 23).

For instance, in *BC*, the keyword *Biber* appears in non-integral self-citations in 68,6% of the cases, which shows that Biber places more emphasis on the message over himself as author. As Hyland and Jiang (2019) discovered, authors nowadays have preferred a non-integral form of citation, allowing themselves less scope for evaluating cited material. In example 21, Biber cites several studies which claimed that texts and registers can undergo any level of generalization analysis:

21. Further, it has been pointed out repeatedly that texts and registers can be analyzed at any level of generalization (e.g., Biber, 1988: chapters 8 and 9; Biber, 1990; Biber, 1994; Biber and Finegan, 1991). (*BC.JA.1995.01.txt*)

In *BC*, however, Biber stresses his cited works in other moments through integral self-citations and activity verbs like *show*. For instance, in example 22, while highlighting his authoritative authorial identity, the author cites a study he carried out with Conrad in which they showed that registers and individual styles can be differentiated by pervasive grammatical features:

22. [...] as Biber and Conrad *show*, registers and individual styles are distinguished by the use of pervasive grammatical features [...] (*BC.JA.2011.01.txt*).

As has been seen, the common practice of using citations reveals allegiance to specific traditions and beliefs held by academic discourse communities regarding what entails carrying out research and what constitutes scientific investigation (Hyland, 2004; 2009). As for the use of self-citation, it is also noticed Biber's attempt to make his presence more visible to the audience, both in integral and non-integral ways, boosting his authoritative authorial identity while building a persuasive and an argumentative rhetoric. Hyland and Jiang (2019, p. 106) state that self-citation is similarly "[...] a device for the management of self-presentation, is an indexical marker of 21st-century professional academic writing." Other reasons for incorporating this feature in texts may include the following:

[...] the circumstances which motivate writers to cite their own work are varied and complex, involving psychological, rhetorical, and social factors that contain elements of confidence, experience and self-promotion. [...] Self-citation can help to emphasize these links one has to one's colleagues through an engagement in a common literature and the professional intimacy one shares with a set of current disciplinary problems. Participation in published research is perhaps the strongest demonstration a writer can make to establish his or her claim to be seen as an important player in a field and to have work taken seriously. (Hyland, 2001, p. 213-214).

The use of different verbs, especially activity verbs, and proper nouns to build an interactive style here also denotes Biber's attempt to seek credibility from readers for stating his claims through *evidentials*. However, it is particularly through verbs that we can see his

interactive metadiscourse style more fiercely represented, because the use of proper nouns to refer to cited authors is actually a common convention of academia. What distinguishes the proper nouns cited in *BC* from those in *CorJAAL* is merely the fact that Biber cited particular authors more often than *CorJAAL* writers. This conventionalized practice of using proper nouns to make citations, though, discloses that Biber's individualized authorial voice also results from a mingling of other people's voices and discourses available in society, previously understood by Bakhtin's (1981) notion of *heteroglossia*. Furthermore, this practice reveals how Biber has taken identities related to the discourse communities of which he is part, also showcasing that his identities are not wholly individual and new, but rather constituted by the discourses he has enacted in his texts. In doing so, he discloses a card-toting community membership and an authoritative authorial identity while referring readers to information from other texts and reporting on other scholars' (and oftentimes his own) research findings.

Next in subsection 4.2.3, it will be explained how the author opts to supply additional information to the readers through *code glosses*. Before that, it should be highlighted that, as you will verify in subsection 4.3.1, many of the verbs classified here as *evidentials* are also categorized as *boosters*, thus pertaining to two different categories, *interactive* and *interactional*. This classification has stemmed from an intricate, contextual analysis of the verbs, for they differed in functions along *BC*. To proceed with this polysemic, ambiguous classification, Hyland's (2005) suggestion was adopted. Therefore, the verbs included in the *evidentials* subcategory and commented on in this subsection visibly highlight Biber's reports on research that he has either directly or indirectly cited along his texts, while his voices mingle and converse with thoughts, ideas, claims, and other notions formerly voiced by quoted scholars.

4.2.3 Code glosses

Table 27 displays the raw and normed frequencies and both text and time ranges of *code glosses* found in *BC*. As will be seen, Biber opts for adverbs and a preposition to "supply additional information by rephrasing, explaining, or elaborating on what has been said, to ensure the reader is able to recover the writer's intended meaning." (Hyland, 2005, p. 52). Biber's *code glosses* stylistic choices recur in almost all 25 texts and in all three time slots, proving his consistent decision for these markers across his academic production and time. Furthermore, as you can see in the table, one marker is a lexical bundle, *for example*. In our first assessment of the keyword list, *example* was identified as key, but when we checked for

its collocates, we observed that it was frequently preceded by the preposition *for*, forming the bundle *for example*.

Table 27. Keywords functioning as *code glosses* across *BC*

<i>Code glosses</i>	Frequency		Range	
	Raw	Normed	Text	Time slot
e.g.	352	2,43	24	1,2,3
for example	340	2,35	25	1,2,3
i.e.	108	0,75	20	1,2,3
including	99	0,68	22	1,2,3
TOTAL	899	6,21	-	-

Source: Elaborated by the author.

For example, *e.g.*, and *i.e.* are exemplification linking adverbials which convey appositive meanings in academic prose, “[...] used as connectors for examples that support more general claims, and with restatements that clarify previous statements.” (Biber *et al.*, 2021, p. 873). The markers *e.g.* and *i.e.* are Latin abbreviations and stand for *exempli gratia* and *id est* (American Heritage Dictionary of the English Language, 2011), respectively. As was seen in Chapter 03, some journal publishing companies mention the possibility to include them in the text, but without determining a favorite form. In this regard, the choice of one linking adverbial over another reveals a writer’s individual stylistic preference. As a matter of fact, “[...] the appositional adverbials *for example*, *for instance*, and *e.g.* exhibit more variability by author.” (Biber *et al.*, 2021, p. 881, emphasis in original).

As stated in Biber *et al.* (2021, p. 882), “examples given with *e.g.* illustrate background information rather than the main ideas; as a result, many occurrences are within parentheses.” Across *BC*, Biber inserted *e.g.* within parentheses in 96,6% of the cases. In excerpt 23, he provides readers with examples of verb agreement with collective nouns and different choices of prepositions placed within parentheses:

23. They note verb agreement with collective nouns (e.g., the public are vs. the public is) and different choices of prepositions (e.g., different to/from, and to live in/ on a street). (*BC.JA.1987.01.txt*)

Like *e.g.*, *i.e.* is another linking adverbial used by Biber. Some similar patterns are equally observed. For example, in 65,8% of the cases *i.e.* comes within parentheses. In example 24, Biber provides the audience with additional information about word length and type/token ratio:

24. Consider Factor 1: the cluster of features with negative weights includes only word length (i.e., longer words) and type/token ratio (i.e., a more varied vocabulary for a given length of text). (*BC.JA.1986.02.txt*)

E.g. also appears outside parentheses, although in 3,4% of the cases. In excerpt 25, you can see how Biber provides the participating audience with the mention of two tenses in the running sentence outside parentheses:

25. Speech and writing differ in their use of verb tenses and aspects, e.g., the past and present tenses. (BC.JA.1986.02.txt)

In 34,2% of the cases, *i.e.* is found outside parentheses. In example 26, Biber gives supplementary information about corpus designs by explaining what two of them are like:

26. Three of the possible overall designs are organized around text production, text reception, and texts as products. The first two of these are demographically organized at the top level; i.e., individuals are selected from a larger demographic population, and then these individuals are tracked to record their language use. (BC.JA.1993.03.txt)

Both *e.g.* and *i.e.* are also used to follow scholars' names in non-integral citations, including self-citations. For instance, *e.g.* is followed by author's names (*Biber, Sinclair, O'Donnell, de Hann, Hunston, Salem*, among others) in 20,5% of the cases, while *i.e.* is used in this way only once. In the excerpts below, Biber supplies extra bibliographical references from which the audience can obtain further information about factor analysis (example 27) and his MD model (example 28):

27. Factor analysis can be used in either an exploratory fashion (e.g., Biber, 1988) or for theory-based 'confirmatory' analyses (e.g., Biber, 1992). (BC.JA.1993.03.txt)
28. Watson's critique has three main points, as summarized in the following quotation: Biber's model (i.e., Biber, 1988) suffers from several serious inadequacies. (BC.JA.1995.01.txt)

In a few cases, however, *e.g.* is not offhandedly added to the sentence, but is also preceded by the mental verb *see* in 7,7% of the times. The use of this mental verb is similar to other patterns that will be discussed in subsection 4.2.4 regarding *endophoric markers and directives*. In example 29, Biber points readers to two bibliographical references from which they can obtain more information about an approach used to investigate lexical bundles:

29. Previous studies of 'lexical bundles' in spoken and written registers have adopted this approach (see, e.g., Biber *et al.* 1999: Chapter 13; Biber *et al.* 2004). (BC.JA.2009.01.txt)

A further pattern observed is the higher use of *e.g.* and *i.e.* in medial position, also reflecting a similar finding to Biber *et al.* (2021, p. 882). In the following excerpts, Biber exemplifies two most frequently recurring sequences of words in a register (example 30) and

explains steps undertaken in factor score computing, mentioning first what language features compose positive weights and then those with negative weights (example 31):

30. More recently, such research has been carried out using the construct of ‘lexical bundle’: the most frequently recurring sequences of words in a register (e.g., I don’t know if, I just wanted to). (BC.JA.2009.01.txt)
31. A factor score is computed by summing, for a given text, the number of occurrences of the features having salient weights on that factor. Thus the score for Factor 3 would be computed by adding the number of past tense forms, perfect aspect forms, and 3rd person pronouns (i.e., the features with positive weights), and then subtracting the number of present tense forms and adjectives (i.e., the features with negative weights). (BC.JA.1986.02.txt)

As seen thus far, Biber helps his reading audience by providing them with additional information so they can better grasp the ongoing argumentation, which displays the author’s facilitator authorial identity. Also, considering both *code glosses* abbreviations, Biber shows individual preference for *e.g.* over *i.e.*, with the former accompanying cited author’s names and appearing within parentheses. Another linking adverbial used is *for example*, more often than *i.e.* In excerpt 32, *for example* is used to introduce an additional explanation about how the dimension 1 score is computed, also providing the audience with some linguistic features grouped on this dimension:

32. Conceptually, a dimension score represents a simple sum of all linguistic features grouped on a dimension. For example, the Dimension 1 score is computed by adding together the frequencies of contractions, demonstrative pronouns, pronoun it, first person pronouns, present tense verbs, etc. [...] (BC.JA.2019.01.txt)

Another pattern observed is the higher use of *for example* in initial position, also resonating with Biber *et al.* (2021). Representing 80% of all cases, this linking adverbial introduces further information right at the beginning of the sentences, while the remainder illustrates extra information added in the middle of the sentences. In the excerpts below, Biber provides readers with some nouns that control two types of complement causes (example 33) and patterns that characterize register differences (example 34):

33. However, most nouns can control only one or the other type of complement clause. For example, the nouns hope, possibility, and fact can control only a that-clause, while the nouns attempt, opportunity, and effort can control only a to-clause. (BC.JA.1999.01.txt)
34. Further, there are important register differences; for example, the pattern begin + complement clause is especially characteristic of conversation, while intransitive begin is more likely to occur in news and academic prose. (BC.JA.2005.01.txt)

In addition to these three linking adverbials, Biber shows individual preference for *including*, all in medial position. In excerpt 35, the author opts for this conjunction to give

readers the information related to two topics of investigation on written academic discourse within English for Specific Purposes/English for Academic Purposes studies:

35. Although most ESP/EAP studies have focused on written academic discourse, more recently researchers have begun to turn their attention to university classroom discourse, including the study of stance and evaluation. (BC.JA.2006.01.txt)

As seen in this subsection, Biber's individual use of different linking adverbials to organize discourse by supplying additional information to readers, explaining and elaborating on what has been discussed promotes an interactive metadiscourse style whereby he helps his reading audience better grasp the information provided along the discourse and displays a facilitator authorial identity. In addition to these features, the author employs *frame markers* to refer to research discourse acts, sequences, and stages, as will be presented next in subsection 4.2.4.

4.2.4 Frame markers

Table 28 displays the raw and normed frequencies and both text and time ranges of *frame markers* used in *BC*. As will be seen, Biber refers to research discourse acts, sequences, and stages by favoring the use of different three-word and four-word bundles and an adverb. Biber's *frame markers* stylistic choices recur in half of all his texts and in most of all three time slots, proving his consistent decision for these markers across his academic production and time. Furthermore, as you can see in Table 28, except for *finally*, the majority of markers are lexical bundles. In our first assessment of the keyword list, however, *respect*, *extent*, *present*, *case*, *wide*, *range*, *broad*, *use*, *full*, *paper*, *basis*, *section*, and *sections* were identified as key, but when we checked for their collocates, we observed that they formed bundles and were metadiscoursal.

Table 28. Keywords functioning as *frame markers* across *BC*

<i>Frame markers</i>	Frequency		Range	
	Raw	Normed	Text	Time slot
with respect to	176	1,22	19	1,2,3
finally	73	0,50	22	1,2,3
the extent to which	47	0,32	16	1,2,3
the present study	42	0,29	11	1,2,3
in the case of	42	0,29	11	1,2,3
a wide range of	41	0,28	18	1,2,3
in this case	35	0,24	18	1,2,3
in the use of	33	0,23	11	1,2,3
the present paper	28	0,19	16	1,2,3
a broad range of	27	0,19	11	1,2

a large number of	26	0,18	11	1,2,3
on the basis of	25	0,17	11	1,2,3
the full range of	24	0,17	10	1,2,3
in the present study	21	0,15	9	1,3
the following sections	14	0,10	12	1,2,3
to a lesser extent	14	0,10	9	1,2,3
the present analysis	13	0,09	6	1,2
of the present study	12	0,08	6	1,2,3
in the present paper	11	0,08	6	1,2
in the present case	11	0,08	7	1,3
as the basis for	11	0,08	8	1,2,3
the present section	10	0,07	6	1,2
TOTAL	736	5,09	-	-

Source: Elaborated by the author.

Overall, *frame markers* are employed in *BC* to “[...] sequence, label, predict and shift arguments, making the discourse recognizable to readers [...]” (Hyland, 2005, p. 51). As a matter of fact, most of the frame marker bundles in *BC* are text-oriented, used to organize the information and arguments laid out in the texts, “[...] display the discursively crafted and rhetorically machined nature of the RA genre, providing warrants, connecting ideas, directing readers around the text, drawing conclusions and specifying limitations.” (Hyland; Jiang, 2019, p. 123). There are, however, research-oriented bundles, employed to structure Biber’s activities or experiences within the real world. Based on Hyland (2008a), a functional classification of frame marker bundles from *BC* and their normed frequencies (all used 4,58 times per 1,000 words) can be seen in Table 29:

Table 29. Frequencies of *frame markers* across *BC*

Research-oriented (1,46)		Text-oriented (3,12)		
Quantification (1,24)	Description (0,23)	Framing signals (2,07)	Structuring signals (1,4)	
the extent to which	in the use of	with respect to	the present study	
a wide range of		in the case of	the present paper	
a broad range of		in this case	in the present study	
a large number of		on the basis of	the following sections	
the full range of		as the basis for	the present analysis	
to a lesser extent		in the present case		of the present study
				in the present paper
				the present section

Source: Elaborated by the author.

As seen in Table 29, the most frequently used marker is *with respect to*, employed as a text-oriented framing signal to “situate arguments by specifying limiting conditions” (Hyland, 2008a, p. 14-15). It is utilized 1,22 times per 1,000 words, mainly in medial position. It also forms four-word bundles, including *with respect to the*, appearing mostly in medial position.

In example 36, Biber makes explicit to readers that his paper aims at developing a typology of English texts considering a five-dimensional model of variation. In excerpt 37, he specifies the focus and the limitation of his study: to show differences between British and American writing regarding two of the dimensions he had previously introduced (dimension 1: interactive versus edited text; dimension 2: abstract versus situated content):

36. This paper develops a typology of texts in English with respect to a five-dimensional model of variation. (BC.JA.1989.01.txt)
37. The present study shows that there are systematic linguistic differences between British and American writing with respect to the first two of these dimensions. (BC.JA.1987.01.txt)

Furthermore, in example 38, Biber stresses the specifying limiting conditions of another study. As inferred from Biber's assertion, a comprehensive MD analysis of the results would take more space than available, and so it was necessary to specify the patterns to be analyzed. There are also other phrases that help the participating audience grasp the author's specifying focus in the paper, including "*there is not space*", "*for the sake of illustration*", "*I focus on two patterns*", displaying Biber's facilitator authorial identity more visibly:

38. There is not space in the present treatment to fully describe the results of this MD analysis. Rather, for the sake of illustration, I focus on two patterns coming out of this study: the relations among general university registers with respect to Dimension 1, and the relations among academic disciplines with respect to Dimension 2. (BC.JA.2019.01.txt)

Moreover, another text-oriented bundle used to organize stretches of discourse is *the present study*, appearing mainly in initial position. It also forms four-word bundles, such as *in the present study* (mainly in medial position) and *of the present study* (mostly in medial position). In BC, Biber employs these bundles to outline the goal of the studies (examples 39 and 40) and the linguistic features a factor analysis is based on (example 41). The choice of these bundles illustrates the author's way of organizing pieces of discourse, promptly leading the readers to grasp the boundaries of the papers:

39. The present study uses a theory-based statistical approach to investigate the dimensions of discourse complexity in English, analyzing the distribution of 33 surface linguistic markers of complexity across 23 spoken and written registers. (BC.JA.1992.01.txt)
40. Furthermore, the factor analysis in the present study is based on linguistic features measured in fifteen different spoken and written text types, so that the textual dimensions identified are representative of the range of spoken and written textual variation. (BC.JA.1986.01.txt)
41. Section 3, then, introduces the T2K-SWAL Corpus and the four registers that are the focus of the present study: classroom teaching, class management talk,

textbooks, and written course management language (e.g., syllabi).
(BC.JA.2006.01.txt)

Furthermore, the research-oriented bundle *in the use of* is employed 0,23 times per 1,000 words to specify aspects investigated within the research, also assisting readers to achieve this perception. In example 42, Biber sets out to explore the historical change of stance-marking patterns. He lays out three major patterns, and the second of them presupposes a decline in the use of some grammatical devices:

42. [...] (2) the grammatical system for the expression of stance could undergo change, showing an overall decline in the use of some grammatical devices, replaced by an increase in the use of other devices; [...] (BC.JA.2004.01.txt)

Another research-oriented bundle used to express the discourse acts of quantifying and marking existence across texts is the frame marker *a wide range of*. These uses make readers aware of elements or phenomena that are being assembled or counted for, promoting a better understanding of arguments laid out in the ongoing discourse. In excerpt 43, although not referring to a specific and determined amount of elements, Biber marks the existence of linguistic features by using the bundle *a wide range of*:

43. The sections above have surveyed a wide range of linguistic features used to express stance meanings in university registers. (BC.JA.2006.01.txt)

In addition to bundles, Biber also shows individual preference for the single-word adverb *finally*, using it 0,50 times per 1,000 words. Similarly, it is employed as linking adverbial to “[...] structure the information in academic prose, giving readers clear signposts of where they are in the text” (Biber *et al.*, 2021, p. 873). For instance, as can be seen in example 44, the author reaches the conclusion section of the RA and lays out several issues that require further investigation. Upon raising all these issues, he introduces the last one: the need for more research on complexity in Western and non-Western languages other than English. In doing so, he makes detectable to readers that this regards the last issue to be mentioned:

44. Finally, the analysis here is based only on the distribution of linguistic features in English texts; further research is required to uncover the dimensions of complexity in other Western and non-Western languages. (BC.JA.1992.01.txt)

All uses of *finally* come in initial position. There are only four cases where this frame marker is followed by transition marker *and* as a sentence initial conjunction, seen as a feature of informality in academic writing. Together with this coordinator, there is an additive meaning provided to the sentence. It is possible to perceive how the author sequences his arguments in BC. For instance, in excerpt 45, Biber lists three characteristics of lexical bundles that set them

apart from other kinds of formulaic expressions. After providing two features, the author adds the last one by introducing it with *and finally*:

45. First, lexical bundles are by definition extremely common (in contrast to most idioms and many ‘grammar patterns’, which tend to be rare). Second, most lexical bundles are not idiomatic in meaning and not perceptually salient. For example, the meanings of bundles like *do you want to* or *I don’t know what* are transparent from the individual words. *And finally*, lexical bundles usually do not represent a complete structural unit. (BC.JA.2009.01.txt)

As seen in this subsection, Biber’s individual use of bundles and a single-word adverb to organize discourse promoted the sequencing, labeling, and shifting of arguments, favoring an interactive metadiscourse style whereby he also displays a facilitator authorial identity. It must also be mentioned that he had a great array of forms available to him to build research and text-oriented discourses, but the interchange among them seems also illustrative of allegiances to rhetorical conventions of the RA genre. Next in section 4.3, the focus of the analysis will be given to the metadiscourse feature the author used with different purposes: to make his views explicit through stance markers and involve the audience in the unfolding text through engagement markers. As will be seen, the language features analyzed next have fostered an interactional metadiscourse style.

4.3 Biber’s interactional metadiscourse style

In this section, we answer the fourth research question posed in the Introduction:

4. Does Biber use interactional metadiscourse language features? If so, how does he express stance and establish engagement with readers in his RAs?

As previously shown in section 4.1, Biber employs interactional metadiscourse features. To express stance, he uses *boosters*, *attitude markers*, and *hedges*. To establish engagement with readers, he opts for *endophoric markers* and *directives*. Overall, Table 30 exhibits how these participant-oriented subcategories were realized by key language items through adverbs, adjectives, verbs, and bundles. It is worth mentioning that these features often chosen by Biber in his texts illustrate many of the distinguishing features commonly used in different academic writing registers, including RAs. Examples from each of those metadiscourse participant-oriented subcategories will be discussed in the subsections 4.3.1 to 4.3.4.

Table 30. Categorization of interactional positive keywords from *BC*

Interactional Metadiscourse Feature		
Participant-oriented systems of interaction	Sub-Categories	Linguistic features and examples
<i>Stance markers</i>	<i>Boosters (epistemic)</i>	• Adverb: <i>reliably</i>. • Verbs: <i>occur; occurring; represented; represents; representing; represent; show; shows; shown; identify; identified; include; included; mark; differ; illustrate; illustrates; illustrated; interpreted; analyze; analyzed; presents; reflect; reflecting; considered; describe; describing; required; carried; investigate; resulting; characterized; considering; reflects; comprises; marks; analyzing; uncover; combines; compares; compute; characterize.</i>
	<i>Hedges (epistemic)</i>	• Adverb: <i>nearly</i>.
	<i>Attitude markers (attitudinal)</i>	• Adjectives: <i>particular; important; adequate; surprising; primary; distinctive.</i> • Adverbs: <i>especially; strikingly.</i> • Bundle: <i>in particular.</i>
<i>Engagement markers</i>	<i>Endophoric markers and directives</i>	• Bundles: <i>figure shows; table shows; sample illustrates; table presents; see table; as noted above; table summarizes; see section; shown in table; figure plots; given in table; summarized in table.</i>

Source: Elaborated by the author.

In view of the plethora of language choices illustrated in Table 30, in this section³¹ we will see how Biber expresses stance and establishes engagement with readers in his RAs and, as a consequence, favors an interactional metadiscourse style. More specifically, we will explore how Biber's individual stylistic choices convey levels of certainty (through *boosters*) and uncertainty (through *hedges*) about his propositions; make his views explicit, comment on the message, and express evaluations and emotions through *attitude markers*; and directly involve readers in the ongoing discourse by overtly pointing them to other parts in the texts through *endophoric markers and directives*. As will be seen, these rhetorical practices are “[...] essentially evaluative and engaging, expressing solidarity, anticipating objections [...], reveal[ing] the extent to which [Biber] works to jointly construct the text with readers”. (Hyland, 2005, p. 51).

³¹ The metadiscourse features introduced in subsections 4.3.1 to 4.3.4 do not follow their order of appearance as illustrated in Figure 5. Rather, they are presented here in accordance with how frequent they were used across *BC* in comparison to *CorJAAL* (cf. Table 22).

I should also mention, however, that Biber's style could have been more interactional if other metadiscourse features had been identified as positively key in *BC*. In other words, the corpus was bereft with statistically significant results of *self-mentions*, *reader pronouns*, *personal asides*, *questions*, and *appeals to shared knowledge*. These are more self-expressing and engaging features, particularly two of them, which will be commented on in the next paragraph. This finding equally illustrates the recent trend of texts in applied linguistics in the decrease of engagement features, as well as "[...] an increase in more quantitative studies which restrict opportunities for overt stance-taking." (Hyland; Jiang, 2019, p. 134). Biber's texts fit this quantitative paradigm and thus prove Hyland and Jiang's (2019) assumption.

Self-mentions are by far the most powerful means of self-representation in academic writing, revealing "[...] the degree of explicit author presence in the text measured by the frequency of first-person pronouns and possessive adjectives [...]" (Hyland, 2005, p. 53). This decision is not merely by chance, because "the presence or absence of explicit author reference is generally a conscious choice by writers to adopt a particular stance and a contextually situated authorial identity." (Hyland, 2005, p. 53). Another feature absent in *BC*, *reader pronouns*, "[...] explicitly address readers, either to focus their attention or include them as discourse participants", serving to "[...] highlight or downplay the presence of readers in the text." (Hyland, 2005, p. 53). The style guidelines of *BC* journals publishing companies posed no restriction for the use of *self-mentions* or *reader pronouns*, suggesting Biber's decision to avoid using them altogether or using them less than expected.

Despite the absence of these features, Biber's texts show significant use of others. Let us proceed next with exploring how the author manages to express stance and establish engagement with readers in his RAs, accomplishing an interactional metadiscourse style. For that, this section presents and discusses a few examples (at least one) from each participant-oriented subcategory as used in their respective context of occurrence across *BC*.

4.3.1 Boosters

Table 31 displays the raw and normed frequencies and both text and time ranges of *boosters* utilized in *BC*. As will be discussed, Biber opts for verbs and a single adverb to stress his confidence in attesting claims in his texts. Biber's *boosters* stylistic choices recur in over 50% of all 25 texts and in almost all three time slots, proving his consistent decision for these markers across his academic production and time.

Table 31. Keywords functioning as *boosters* across *BC*

<i>Boosters</i>	Frequency		Range	
	Raw	Normed	Text	Time slot
represent	170	1,17	24	1,2,3
occur	165	1,14	25	1,2,3
show	140	0,97	24	1,2,3
identified	118	0,82	22	1,2,3
identify	105	0,73	21	1,2,3
include	99	0,68	24	1,2,3
shows	97	0,67	24	1,2,3
considered	94	0,65	21	1,2,3
required	79	0,55	19	1,2,3
represents	75	0,52	19	1,2,3
included	72	0,50	18	1,2,3
interpreted	62	0,43	17	1,2,3
reflect	59	0,41	18	1,2,3
occurring	57	0,39	21	1,2,3
illustrate	57	0,39	16	1,2,3
representing	55	0,38	15	1,2,3
shown	52	0,36	23	1,2,3
differ	51	0,35	19	1,2,3
represented	48	0,33	16	1,2,3
mark	48	0,33	17	1,2,3
analyze	48	0,33	15	1,2,3
reflecting	48	0,33	15	1,2,3
analyzed	46	0,32	16	1,2,3
describe	41	0,28	18	1,2,3
investigate	32	0,22	13	1,2,3
characterized	30	0,21	10	1,2,3
describing	30	0,21	15	1,2,3
considering	28	0,19	16	1,2,3
reflects	27	0,19	10	1,3
resulting	26	0,18	18	1,2,3
comprises	24	0,17	12	1,2,3
illustrated	24	0,17	15	1,2,3
marks	23	0,16	12	1,2,3
illustrates	21	0,15	19	1,2,3
carried	19	0,13	14	1,2,3
reliably	16	0,11	4	1,2,3
analyzing	14	0,10	10	1,2,3
characterize	14	0,10	11	1,2
uncover	13	0,09	8	1,2,3
combines	12	0,08	5	1,3
compares	11	0,08	10	1,2,3
compute	10	0,07	7	1,2,3
presents	5	0,03	5	1,2
TOTAL	2,267	15,66	-	-

Source: Elaborated by the author.

Boosters, like *hedges*, convey epistemic modality, but there is a plausible difference between them, especially in terms of either opening (*hedges*) or closing (*boosters*) space for readers to reach specific conclusions while interpreting texts. Despite the fact that *boosters* tend to limit the variability of viewpoints on the part of readers, they still end up joining together

writers and readers, as both agents are part of the same discourse community and, to a certain extent, agree with one another:

Hedges imply, then, that a statement is based on plausible reasoning rather than certain knowledge, and *allow readers the freedom to dispute it*. At first glance, boosters seem to contradict such conciliatory and defensive tactics. *They emphasize the force of propositions and display commitment to statements, thereby asserting the writer's conviction and restricting the negotiating space available to the reader*. But while an apparently risky tactic, *boosters nevertheless allow writers to strategically engage with colleagues, effecting interpersonal solidarity and membership of a disciplinary in-group*. (Hyland, 1998b, p. 353).

As we saw in subsection 4.2.2, some verbs classified as *evidentials* are now categorized as *boosters* here. The main differences between their uses are two-fold: booster verbs state the writers' stance towards their views and claims; booster verbs are endophoric in nature, pointing readers to some internal aspects of the authors' research in the genre and building solidarity with the audience. In this regard, a semantic classification of booster verbs based on Biber *et al.* (2021) and their normed frequencies (all occurring 15,55 times per 1,000 words) follow in Table 32:

Table 32. Semantic classification of booster verbs and their normed frequencies in *BC*

Activity	Communi- cation	Mental	Existence	Occurrence	Causative
3,18	0,49	2,96	6,67	1,71	0,55
analyze; analyzed; analyzing; carried; combines; investigate; show; shows; shown	describe; describing	compares; compute; considered; considering; identify; identified; interpreted;	characterize; characterized; comprises; differ; mark; illustrate; illustrates; illustrated; include; included; presents; marks; reflect; reflects; reflecting; represents; representing; represent; represented; uncover	occur; occurring; resulting	required

Source: Elaborated by the author.

As can be seen in Table 32, Biber favors existence verbs, followed by activity, mental, occurrence, causative, and communication verbs. Biber's stance expression for existence verbs serves to report states that exist between entities, representing the state of existence or a particular relationship between entities. In example 46 in *BC*, Biber explains what a comparison corpus is used for within stylistic studies. By opting for the verb *represent*, he expresses certainty about the state of *representing* that this type of corpus is submitted to because of the

way it functions within a stylistic investigation. Moreover, the author inhibits the negotiating space available to readers by claiming that this corpus purpose is to *represent*, but nothing else, illustrating Biber's choice of one form out of the array of all available forms:

46. The comparison corpus is used to represent 'typical' patterns of use, making it possible to empirically identify distinctive linguistic patterns in the target corpus that depart from those typical patterns. (BC.JA.2011.01.txt)

In the conclusion section of another text, Biber revisits and reiterates what has been his purpose throughout the research article: to show that corpora of different sizes can be useful (example 47). For that, he chooses the activity verb *show* to denote the action of *showing* while ascertaining the usefulness of such corpora. Furthermore, he prevents readers from identifying other research goals other than the one already laid out. At the same time, however, this limitation offers readers interpersonal solidarity, preventing them from going astray:

47. Rather, my purpose has been to show that existing corpora are adequate in many respects [...] (BC.JA.1990.01.txt)

Elsewhere in *BC* (example 48), while Biber admits the systematicity and importance of lexico-grammatical association patterns, he nonetheless recognizes some limitation in interpreting these findings only this way, making explicit to the audience that additional reflections must be made. For that, he employs the mental verb *interpreted* in a passive voice structure to denote the interpreting action that must be executed by someone, either himself as researcher/writer or another scholar investigating the same subject, an apparent effort to engage with the disciplinary community, displaying a supportive authorial identity. At the same time, Biber restricts his readership to dispute other possibilities for this case besides the urgent need to interpret it:

48. Although patterns such as these must be interpreted much more fully, the present section has illustrated the systematicity and importance of these association patterns in describing the use of related grammatical features. (BC.JA.1996.01.txt)

Biber's certainty and tendency to restrict disputing alternatives to his propositions can also be recovered through the occurrence verb *occur*. In example 49, the author reports on his research findings concerning fillers that complete the lexical frame *the * of the*. These results are even shown in an illustration followed by the statement, proving that Biber's claim is accurate:

49. In most cases, the high frequency patterns in academic writing are extremely productive. For example, there are numerous fillers that occur in the highest frequency frame the * of the, and as Table 4 shows, the most common of these

fillers accounts for only 2% of the total occurrences of the pattern.
(*BC.JA.2009.01.txt*)

At the end of another publication (example 50), Biber urges for the need for additional research on use factors and relations between functions and association patterns. To make this demand, he opts for the causative verb *required*, especially in passive voice structures, also indicating that someone else (or perhaps himself) should carry on the research he started. This choice signals the author's certainty for his claim, displaying a confident authorial identity, as can also be seen by the attitude marker *obviously* at the beginning of the sentence. Although it is not obvious who is *required* to undertake additional research, this absence may also be seen as an attempt to engage with the disciplinary community by inviting it to follow on Biber's investigation:

50. Obviously, future research is required to investigate the relative importance of use factors and the ways in which particular functional considerations relate to particular kinds of association patterns. (*BC.JA.1996.01.txt*)

Another way of ascertaining Biber's claims is through the communication verb *describe*. In example 51, the author lays out three features of a text-linguistic approach to linguistic variation, also limiting how his readership may see such an approach. To address issues about the first feature, he admits that the research goal of this approach entails *describing* the characteristics of texts, and not *identifying* or *illustrating* them, actions which belong to other verb semantic categories:

51. The text-linguistic approach to linguistic variation differs in all three respects: the research goal is to describe the characteristics of texts, rather than the characteristics of a linguistic feature; each text constitutes an observation; and the quantitative findings represent the rates of occurrence of linguistic features in texts rather than the proportional preference for a linguistic variant in comparison to other variants. (*BC.JA.2012.01.txt*)

Biber also uses a single adverb to stress his confident, assertive authorial identity in *BC*. In example 52, to highlight two important issues not considered by his analysis, he opts for the booster adverb *reliably* to refer to particular texts that could truthfully represent mean scores calculated on the basis of six linguistic features compared across 10-text samples:

52. Secondly, this analysis addressed how many texts were needed to reliably represent mean scores, but did not address the representation of linguistic diversity in registers. (*BC.JA.1993.03.txt*)

Given this overall picture on the use of *boosters* in *BC*, Biber's individual boosting choices disclose an interactional metadiscourse style whereby the author stresses the strength of his propositions, displays commitment and declares his conviction to statements, even if this

entails, on the one hand, restricting the negotiating space available to the participating audience and, on the other hand, offering it interpersonal solidarity occasionally. These choices are also responsive to the rhetorical conventions of the genre and illustrate how the author expresses stance in *BC* to reveal confident, assertive, and supportive authorial identities. Next in subsection 4.3.2, the focus will be given to the feature the author uses to convey uncertainty about his propositions, *hedges*, which is the counterpart of *boosters*.

4.3.2 Hedges

Table 33 displays the raw and normed frequencies and both text and time ranges of *hedges* utilized in *BC*. As will be seen, Biber opts for only one adverb to show “a lack of complete commitment to the truth value of an accompanying proposition [...] or a desire to express that commitment categorically” (Hyland, 1998a, p. 1). This stylistic choice recurs in almost all *BC* texts and in all three time slots, proving the author’s consistent decision for this marker across his academic production and time.

Table 33. Keywords functioning as *hedges* across *BC*

<i>Hedge</i>	Frequency		Range	
	Raw	Normed	Text	Time slot
nearly	45	0,31	19	1,2,3
TOTAL	45	0,31	-	-

Source: Our elaboration

As a mechanism of author’s self-effacing, epistemic modality, hedging is realized by only one key adverb in *BC*: *nearly*. This result resonates with Biber *et al.*’s (2021) claim that *hedges* are considered a sub-class of adverbs and that, like conversation, academic prose shows a heavy reliance on single adverbs as stance devices, especially those indicating epistemic stance. Although only one single hedging adverb was found key, this result still represents Biber *et al.*’s (2021) finding that authors in academic prose tend to favor this grammatical device for the expression of epistemic modality, like Biber himself across his RAs.

Hyland (1998a) advocates for the scientific investigation of *hedges* in academia because they are essential to building the rhetoric of academic discourse, especially when it comes to writer’s providing interpretative statements on paper’s theory, method, or data:

Hedging enables writers to express a perspective on their statements, to present unproven claims with caution and to enter into a dialogue with their audiences. It is therefore a substantial means by which scientists confirm their membership of the scientific community [...]. (Hyland, 1998a, p. 6).

In *BC*, the only key hedge is *nearly*, often collocating with such words as *all*, *identical*, and *every*, as well as with numbers and percentage. These patterns demonstrate that “hedges

are also very common with numbers, measurements, and quantities. These forms are also called *approximators*, and typically function as modifiers of numerical or other quantifying expressions.” (Biber *et al.*, 2021, p. 554, emphasis in original). His choice of *nearly* over other available forms in the language also depicts an individual stylistic preference.

For instance, in example 53, Biber is unsure as to whether specialized dimensions arise from all MD analyses, and in doing so enters into a dialogue with his audience to dispute this claim. In the meantime, he reveals hesitant and dialogic authorial identities:

53. And to some extent, this expectation is met, with specialized dimensions emerging in nearly all MD analyses. (BC.JA.2019.01.txt)

In example 54, Biber is cautious to determine that every discourse domain of English and other languages and cultures favors the presence of a narrative dimension, allowing his readership to refute whether this claim holds true:

54. Similar to the oral/literate dimensions, a narrative dimension emerges in the MD analysis of nearly every discourse domain of English, and in nearly every language/ culture. (BC.JA.2014.01.txt)

Moreover, when reporting on numerical results from two studies, Biber hedges his commitment to, respectively, providing absolute values for model improvement (example 55) and two syntactic functions of prepositional phrases (example 56). It can be seen that the author opens possibilities for his readers to notice how approximate (and not absolute) the results are:

55. It represents a significant improvement over all other models and a relatively strong improvement in Delta over other models – nearly a 5% improvement over the 4-D model, and nearly a 10% improvement over the 1-D model. (BC.JA.1992.01.txt)
56. For editorials, this table shows that there is nearly a 50/50 split for prepositional phrases attached as nominal versus verbal modifiers [...] (BC.JA.1993.04.txt)

Given this overall picture, Biber’s individual hedging choice of *nearly* uncovers an interactional metadiscourse style by which he refrains from committing himself to the truth values of his propositions. At the same time, however, he allows readers to respond to the unfolding text, opening up possibilities for (re)assessing the claims and propositions. This stylistic choice also reveals the author’s hesitant and dialogic authorial identities. Next in subsection 4.3.2, the focus will be allocated to the last feature the author uses with a stance-marking purpose: *attitude markers*.

4.3.3 Attitude markers

Table 34 displays the raw and normed frequencies and both text and time ranges of *attitude markers* utilized in *BC*. As will be examined, Biber opts for adjectives and adverbs to overtly comment on the message and express evaluations and emotions in the meantime. This stylistic choice recurs in over 60% of *BC* texts and in almost all three time slots, proving the author's consistent decision for these markers across his academic production and time. Furthermore, as you can see in the table, there is one keyword that works as metadiscoursal both on its own and as part of a lexical bundle: *particular*.

Table 34. Keywords functioning as *attitude markers* across *BC*

<i>Attitude markers</i>	Frequency		Range	
	Raw	Normed	Text	Time slot
important	228	1,58	24	1,2,3
particular	180	1,24	25	1,2,3
especially	100	0,69	18	1,2,3
primary	67	0,46	19	1,2,3
adequate	59	0,41	16	1,2,3
surprising	48	0,33	16	1,2,3
in particular	29	0,20	21	1,2,3
strikingly	19	0,13	11	1,2,3
TOTAL	730	5,04	-	-

Source: Our elaboration

Differently from *hedges* and *boosters*, which are epistemic in nature and convey degrees of (un)certainly and (un)likelihood, *attitude markers* are attitudinal. This is to say that they are used to express writers' affective attitudes, emotions, evaluations, and feelings to/on a proposition, conveying surprise, agreement, importance, obligation, frustration, etc. (Biber *et al.*, 2021; Biber, 2006; Hyland, 2005). As a result, this self-expression helps construct an interactional metadiscourse style in texts. Compared to frequently used epistemic stylistic choices, Biber's underuse of *attitude markers* appears to reflect academic prose register: "overall, attitudinal stance markers are less common than epistemic markers [...]." (Biber *et al.*, 2021, p. 966).

Given the words previously shown in Table 34, it is perceivable that Biber favors the expression of evaluations and emotions while overtly commenting on the messages, echoing his voice plainly to the readers. In the former case, he opts for the adjectives *important*, *particular*, *primary*, the adverb *especially*, and the bundle *in particular* to convey importance; and the adjective *adequate* to transmit a sense of suitability. In the latter, he chooses the adjective *surprising* and the adverb *strikingly* to express his surprise towards claims.

Let us begin the discussion by Biber's expression of importance, which unveils his evaluative authorial identity. As seen in example 57, the author opts for the adjective *important* to highlight that textual co-occurrence patterns between frequent nouns, adjectives, and prepositional phrases are relevant for the identification of linguistic characteristics of registers and styles. It is thus possible to detect the way Biber comments on his message and leaves his fingerprint for readers to spot, also representing an individual stylistic choice:

57. For example, frequent nouns, adjectives, and prepositional phrases commonly co-occur in academic prose texts, working together to provide a dense integration of information. Textual co-occurrence patterns such as these are important in identifying the salient linguistic characteristics of registers and styles. (BC.JA.1996.01.txt)

The expression of importance is also recognized by the adverb *especially*. In some cases, this adverb is even followed by the stance adjective *important*. In other cases, though, it collocates with other stance adjectives (*noteworthy*, *robust*, *conservative*) that did not return key by *AntConc*, but which can help us sense Biber's evaluative authorial identity. In example 58, the author recognizes that, out of all dimensions of MD analysis often identified in many different studies, two bear especial importance because they help build basic human communication:

58. Two such dimensions are especially important, regardless of the discourse domain: a dimension associated with oral vs. literate discourse, and a dimension associated with narrative discourse. The robustness of narrative dimensions across languages and discourse domains indicates that this rhetorical mode is basic to human communication, whether in speech or in writing. (BC.JA.2014.01.txt)

Another evaluative stance adjective used to transmit a sense of suitability in *BC* is *adequate*. In example 59, Biber refutes some criticisms and leads readers to grasp that existing computer-based corpora are suitable for many studies of linguistic variation.

59. However, the findings reported here argue against the claim that existing computer-based corpora are unsuited for the study of linguistic variation; instead, they indicate that, with respect to the above questions, the corpora are adequate for many studies of linguistic variation across texts. (BC.JA.1990.01.txt)

The expression of emotion in *BC* is observed through the stance *attitude markers* *surprising* and *strikingly*, disclosing Biber's emotional authorial identity. In example 60, Biber is fully surprised with some research results because they may have been influenced by the layout of the corpus used. Again, it is possible to see how Biber comments on his message and emits his voice for readers to hear:

60. It is perhaps not surprising that Dimension 1 in the original 1988 MD analysis was strongly associated with the oral/literate opposition, given that the corpus in that study ranged from spoken conversational texts to written expository texts. (BC.JA.2014.01.txt)

In another use, the adjective *surprising* is modified by the adverb *especially* presented earlier. Below, we can perceive how Biber's evaluative and emotional authorial identities are revealed in the sentence, with particular stress on the former. As can be seen in example 61, a common assumption about syntactic complexity held by many linguists is weakened when research revealed unexpected findings. Once again, Biber's interactional choice evidently discloses how he feels about the results found:

61. The oral style relies on clauses to construct discourse – including a dense use of dependent clauses. In contrast, the complexity of the literate style is phrasal. This finding, replicated across languages and across discourse domains, is especially surprising, because it runs counter to assumptions about syntactic complexity held by many linguists. (BC.JA.2019.01.txt)

Furthermore, emotion is also conveyed through the adverb *strikingly*. As seen in example 62, Biber's stance choice of this word reveals his surprise towards particular lexico-grammatical patterns detected in some analyses:

62. These analyses show that strikingly different lexico-grammatical patterns are associated with each complement clause type and with each register, and that those associations can be explained in terms of the typical topics and communicative purposes of each register. (BC.JA.1999.01.txt)

Given this overall picture on the use of stance-marking *attitude markers* in *BC*, Biber's individual stylistic attitudinal choices unveil an interactional metadiscourse style whereby the author overtly comments on the messages, taking the participating audience to recognize his evaluative and emotional authorial identities. Lastly, in subsection 4.3.4, the focus will be allocated to the only feature the author uses with a more overt engaging purpose: *endophoric and directive markers*.

4.3.4 Endophoric and directive markers

Table 35 displays the raw and normed frequencies and both text and time ranges of *endophoric and directive markers* employed in *BC*. As will be examined, Biber opts for bundles to explicitly direct readers to particular points in his RAs. This stylistic choice recurs in over 50% of *BC* texts and in almost all three time slots, proving the author's consistent decision for these markers across his academic production and time. Furthermore, as you can see in the table, all markers are lexical bundles. In our first assessment of the keyword list, however, *figure*, *shows*, *table*, *sample*, *illustrates*, *presents*, *above*, *section*, *shown*, and *summarizes* were

identified as key, but when we checked for their collocates, we observed that they formed bundles and were metadiscoursal.

Table 35. Keywords functioning as *endophoric and directive markers* across BC

<i>Endophoric and directive markers</i>	Frequency		Range	
	Raw	Normed	Text	Time slot
figure shows	49	0,34	11	1,2,3
table shows	28	0,19	14	1,2,3
sample illustrates	23	0,16	8	1,2,3
table presents	20	0,14	8	1,2
see table	20	0,14	13	1,2,3
as noted above	17	0,12	10	1,2,3
table summarizes	15	0,10	9	1,3
see section	12	0,08	8	1,2,3
shown in table	11	0,08	7	1,3
figure plots	11	0,08	8	1,2,3
given in table	10	0,07	9	1,3
summarized in table	10	0,07	8	1,2,3
TOTAL	226	1,56	-	-

Source: Our elaboration

Prior to illustrating some examples of *endophoric and directive markers* from BC, it is important to remember that, as previously justified in Chapter 02, *endophoric markers* were originally classified as an interactive feature by Hyland (2005), but in a revisited version of the model, the author excluded it altogether and created *directives* (Hyland, 2009). However, we hereby understand that there is a connection between *endophoric markers* (originally interactive) and *directives* (originally interactional), for they have the same purpose: to show perceptible assessments of the material and the audience, direct readers to particular points in the text, and influence them to reach specific understandings and conclusions (Hyland, 2005). For this reason, in this dissertation, *endophoric markers* are a type of *interactional engagement markers*, together with *directives*, for both involve “[...] rhetorically positioning the audience, pulling readers into the discourse at critical points, predicting possible objections and guiding them to particular interpretations.” (Hyland, 2005, p. 54). These features are also concerned with the genre conventions (Swales, 1990; 2004). In sum:

Directives instruct the reader to perform an action or to see things in a way determined by the writer. [...] they are signalled mainly by imperatives (like *consider*, *note*, and *imagine*), modals of obligation addressed to the reader (e.g. *must*, *should*, *ought*) [...] Directives are the only engagement feature which occur more frequently in papers in the sciences and engineering. (Hyland, 2005, p. 154).

Endophorics [...] refer the reader to explanatory or related material elsewhere in the text. They represent the writer’s assessment of both the material and the audience, relating the propositions to the reader’s assumed ability to process, and accept, the ongoing argument. In the soft fields this largely means reinforcing an argument by ensuring the reader has immediate access to relevant data or arguments located elsewhere in the text [...] Endophorics, however, are overwhelmingly a feature of

writing in the hard disciplines, where the referent is usually a nearby table or graph. (Hyland, 2005, p. 156-157).

In *BC*, Biber explicitly directs readers to other parts of the texts through such bundles as *see table* and *see section*. It should be noted that the journal publishing companies' style guidelines encouraged authors to explicitly make cross-references along the texts, with two publishers suggesting the use of mental verbs like *see*. In this regard, Biber's choices above seem to have undergone style guidelines influences, suggesting that he may have ended up abiding by academic conventions administered by literacy brokers and gatekeepers, including peer reviewers. Despite that, an interesting pattern across *BC* is that all uses of *see table* and *see section* come in parentheses and mostly at the end of sentences, implying that these directives behave similarly to non-integral citations, receiving minor focus and attention from the author. For instance, through textual acts, Biber discloses a supportive authorial identity by directing readers to *see*: a) a list of texts used to compare British and American writing, classified according to genre (example 63); and b) a section where it has been discussed how some dimensions characterized English and some discourse domains (example 64):

63. Each dimension will be discussed briefly with respect to both written and spoken genres, but the focus here is on the 796 British and American written texts taken from the LOB and Brown corpora, representing the same nine written genres in each dialect (*see table 3*). (*BC.JA.1987.01.txt*)
64. Subsequent MD studies have shown that some of these dimensions turn out to be peculiar to English or to particular discourse domains (*see Section 5*). (*BC.JA.2014.01.txt*)

Moreover, the keyword *shows* often collocates with *figure* and *table*. In example 65, we can notice that Biber motivates readers to look at an illustration in the text, in which a marked decline with respect to the use of modals in different registers over a 5-decade period is depicted. The author's choice regards a textual act to be undertaken by readers:

65. The study here provides additional evidence that modals have undergone a recent decline in use. As *Figure 4 shows*, modals have undergone a marked decline in all registers, especially over the last 50 years. (*BC.JA.2004.01.txt*)

As can be seen, the use of *endophoric and directive markers* demonstrates an involvement between different semiotic resources available in the writer's texts to which he makes explicit reference. This connection can also be identified in example 66, proving that "[...] writing is a constant switching between written and diagrammatic representation because scientific concepts are typically semiotic hybrids which combine verbal, mathematical and visual elements." (Hyland, 2005, p. 157). The way Biber refers to data in tables illustrates this hybrid combination of elements. Again, readers are encouraged to achieve a textual act by

looking at a specific table in the article which presents numerical data on mean score and standard deviation:

66. Table 4 presents the mean score and standard deviation of seven linguistic features in the pilot corpus, together with the computed tolerable error for each feature. It can be seen that the tolerable error ranges from 9.03 for nouns (which have a mean of 180.5) to 0.13 for conditional clauses (which have a mean of only 2.5) (*BC.JA.1993.03.txt*)

A further textual act is fostered in another excerpt in *BC*. In example 67, Biber leads readers to another part of the text, although not certainly where, encouraging them to note that text-linguistic register studies have now unveiled previously undiscovered language patterns. This choice also signals an individual stylistic decision:

67. Second, as noted above, text-linguistic register studies have also been important because they have uncovered patterns of use that were not even anticipated in previous research. (*BC.JA.2019.01.txt*)

As seen in this subsection, Biber's preference for some *endophoric and directive markers* demonstrates how he engages with readers by leading them to particular text extracts, some even made of hybrid semiotic resources (*figures, tables*). The author's stylistic engaging choices disclose a supportive authorial identity and unveil an interactional metadiscourse style by which he explicitly bonds with the participating audience so that it can promptly recognize relevant information referred to in the RAs. In some cases, though, these choices may have been influenced by literacy brokers.

Considering what has been discussed in this Chapter 04, I now proceed with making Final Remarks, answering the last research question and summarizing the ways in which Biber has constructed a metadiscourse style through interactive and interactional dimensions, with a predilection for an interactive style.

CHAPTER 05 – Final remarks

This last chapter revisits the hypotheses, research questions, and objectives of this doctoral investigation, lays out limitations of the study, and offers suggestions for future research.

5.1 Revisiting hypotheses, research questions, and objectives

In this section, we revisit the hypotheses, research questions, and objectives. To (re)start with, the first hypothesis laid out that Biber would consistently make metadiscourse language choices in his RAs that could constitute his academic writing style. This hypothesis was confirmed, especially when we consider the answers provided to the first and second research questions:

1. What kind of words appear as positively key in Douglas Biber's RAs when compared to RAs written by other applied linguistics scholars?

More than 75% of the keywords retrieved by *AntConc* are of the text-aboutness type, but this is not an unexpected discovery, since the keyword tool usually identifies this kind of word more frequently (Gabrielatos, 2018; Geluso; Hirsch, 2019; Scott, 2022). Furthermore, metadiscourse keywords account for about 20% among those retrieved as positive. The remainder refers to keywords that are neither text-aboutness nor metadiscourse, representing almost 5% of the findings. By answering this question, we have also fulfilled the first specific objective: to identify positive keywords by comparing *BC* and *CorJAAL*.

2. Are there any keywords that can be classified as metadiscoursal? If so, what is their number of raw and normed values after categorized according to the *Stylistic Metadiscourse Model of Interpersonal Interaction in Academic Writing*?

The number of metadiscourse keyword types equals 126, but this total does not consider other language items that were equally categorized as metadiscoursal after we checked for the collocates of the initial 126 keywords and observed that some of them were part of lexical bundles. Therefore, these initial results shifted. We also classified them into the functional metadiscourse subcategories proposed in our stylistic metadiscourse model. After this categorization, the number of raw and normed values can once again be retrieved in Table 36.

By responding to this question, we have also achieved the second specific objective: to classify metadiscourse keywords from Biber's RAs.

Table 36. Number of stylistic metadiscourse keyword types and tokens

Stylistic metadiscourse features	types	tokens	normed
Interactive	95	3,919	27,08
Interactional	64	3,268	22,58
Total	159	7,187	49,66

Source: Our elaboration

Upon confirming the first hypothesis, attention was given to addressing the second one. Because of Biber's background experience in the hard sciences and paradigmatic bias towards quantitative and statistical methods (Biber, 2013), I assumed that his metadiscourse style would be more interactive than interactional, which is a strong characteristic of texts written in the hard sciences (Hyland, 2005). Although it cannot be wholly proven that Biber's background and paradigmatic favoritism have entirely influenced his style, his stylistic choices reveal, to a certain extent, a commonplace tradition of texts written in the hard sciences, particularly when we ponder the answers previously provided to the third and fourth research questions, and below, in this final chapter, answer the fifth question.

3. Does Biber employ interactive metadiscourse language features? If so, how does he organize discourse in his RAs?

Biber has drawn on *transition markers*, *evidentials*, *code glosses*, and *frame markers* to organize discourse in his RAs, which results in the creation of an interactive metadiscourse style. For that, he shows individual preference for peculiar adverbs, adjectives, conjunctions, verbs, proper nouns, prepositions, and lexical bundles.

To organize discourse, the author's individual stylistic choices across texts and time establish contrastive, comparative, additive, and consequence relations between clauses (through *transitions*); refer readers to information from other texts, reporting on other scholars' research findings with no stance taken towards them (through *evidentials*); provide readers with extra information about the text by rephrasing, explaining, or elaborating on what has just been discussed (through *code glosses*); and sequence, label, and shift arguments, making the discourse recognizable to the reading audience (through *frame markers*). His preference for an interactive style suggests that he writes on complex themes by following a logical, straightforward pathway.

More specifically, Biber's use of *transitions* has allowed him to make connections between arguments so as to meet readers' comprehension and interpretive needs, displaying a supportive and helpful authorial identity. In addition, the author's *evidentials* choices have illustrated how he seeks credibility from and shows authority to readers by: i) making citations and self-citations; ii) locating his research within the body of existing literature; iii) echoing his voice and knitting other's voices to his own; iv) making his presence more visible to the audience in integral and non-integral patterns; and v) disclosing a card-toting community membership and an authoritative authorial identity. Furthermore, Biber's preference for particular *code glosses* has helped his reading audience better grasp the information provided along the discourse, fostering a facilitator authorial identity. At last, the author's varying uses of *frame markers* have better guided the participating audience throughout discourse, making visible an equally facilitator authorial identity.

By answering this question, we have also fulfilled the third specific objective: to assess how Biber organizes discourse through interactive metadiscourse features.

4. Does Biber use interactional metadiscourse language features? If so, how does he express stance and establish engagement with readers in his RAs?

Biber has used *boosters*, *attitude markers*, and *hedges* to express stance, and *endophoric markers and directives* to establish engagement with readers. Consequently, these patterns result in the creation of an interactional metadiscourse style. For that, he demonstrates individual preference for particular adverbs, adjectives, verbs, and bundles.

As seen across *BC*, the author's individual stylistic choices across texts and time convey levels of certainty (through *boosters*) and uncertainty (through *hedges*) about his propositions; make his views explicit, comment on the message, and express evaluations (importance and suitability) and emotions (surprise) through *attitude markers*; and directly involve readers in the ongoing discourse by overtly pointing them to other parts in the RAs through *endophoric markers and directives*.

More specifically, Biber's use of *boosters* has stressed the strength of his propositions and displayed his commitment and conviction to statements, revealing confident, assertive, and supportive authorial identities. Moreover, the author's *hedge* choice has demonstrated bivalent decisions on his part: he refrains from committing himself to some claims, but also allows the audience to respond to the unfolding text, opening up possibilities for (re)assessing the propositions. Together, these choices have disclosed the author's hesitant and dialogic authorial

identities. Furthermore, Biber's adoption of *attitude markers* has allowed him to overtly comment on the discourse through the expression of evaluations and emotions, leading the participating audience to perceive his evaluative and emotional authorial identities. At last, the author's preference for some *endophoric and directive markers* have demonstrated how he engages with readers by leading them to promptly recognize relevant information located in figures, tables, sections, and other parts of the RAs, disclosing a supportive authorial identity.

By replying to this question, we have also accomplished the fourth specific objective: to examine how Biber expresses stance and establishes engagement with readers by means of interactional metadiscourse features. Given the answers to questions three and four provided above, we now reply to the last one.

5. In general, is Biber's metadiscourse style more interactive or interactional?

In view of the analyses carried out in this dissertation to answer all four previous research questions and fulfill related objectives, Biber's metadiscourse style is, in fact, more interactive. The author's preference for this approach distinguishes himself from all the other authors whose texts compose *CorJAAL*. His style has even remained stable for the past 35 years and been identified in all his single-authored RAs.

Although the difference in metadiscourse usage is relatively small (4,5) between both dimensions, Biber uses the interactive feature slightly more than the interactional one. This fact illustrates what he confessed to Gray in an interview (Biber, 2013) and suggests that his background in the hard sciences and paradigmatic bias towards statistical and quantitative methods might have influenced his interactive style. Future studies, though, could measure this background-style influence in more defined ways.

Another related point should not be disregarded here. Even though the author's stylistic preference for interactive features is more salient in his RAs, another part of his metadiscourse style is undeniably interactional. This is to say that the author integrates both of these metadiscourse dimensions to construct texts which seem to be easy to read and grasp. In other words, both dimensions coexist in his RAs to create meaningful pathways that will support the audience along the reading.

Furthermore, as we revisit both the research problem inspired by Crystal's (2013) stylistic profiling endeavor proposed a decade ago and the definition of style posed in Chapter 02, we recognize that Biber's profile is formed by a variety of frequent and consistent language usage, which illustrates distinguishing and particular ways of employing the academic register

in different RAs for logically organizing discourse, expressing stances, and engaging himself with his reading audience. These ways have fostered the expression of voices, authorial identities, and brought about (mainly) interactive and interactional effects in texts and readers.

By answering this last question, we have also fulfilled the last specific objective: to categorize Biber's metadiscourse style as either/both interactive or/and interactional. Upon revisiting the hypotheses, research questions, and objectives of this doctoral investigation, we finally proceed with laying out limitations of the study and offering suggestions for future research.

5.2 Laying out research limitations and foreseeing the future

Notwithstanding the accomplishment of our research objectives and answers to the research questions, seven issues have been raised. They might, though, be resolved or attended to in future research.

The first aspect lies in the degree register may have influenced the author's style, as mentioned in Chapter 03. In this dissertation, we have acknowledged that register may affect how authors craft their style, especially if one considers that no one writes in a social vacuum, much less Biber or myself, where different elements come into play at the time of writing. So it is arguable that language features characteristic of the applied linguistics RAs register may have an impact on author's stylistic decisions. Therefore, the degree of this register-style influence is encouraged to be measured in future studies, perhaps adopting stylometric analytical methods and tools as suggested in Grieve's (2023) research.

The second aspect involves genre influences. Our analysis of the study and reference corpora has disregarded peculiar characteristics of genre sections (Swales, 1990; 2004) and how these could more bluntly impact Biber's choices. Again, the hypothesis that genre features might impact an author's style also proceeds, especially because Biber has specifically written RAs, which, like any other genre inside (and outside) academia, have peculiar features of their own. Therefore, the degree of this genre-style influence is persuaded to be measured in future studies, starting perhaps with splitting the macro corpora into smaller subcorpora, each consisting of respective RA sections following the IMRD pattern.

The third aspect involves corpus design. As mentioned, *BC* only includes RAs, because previous research (Geluso; Hirsch, 2019) has already ruled out potential stylistic differences motivated by different genres written in the same register. However, if we had focused on more informal and engaging academic genres, like interviews, the author's choices might have been

different, perhaps more interactional. This is also the case of focusing on non-academic genres, like recipes, medicine leaflets, movie and theater screenplays, etc. Future research could focus on how the same author writes inside and outside academia. An interesting suggestion would be on another scholar's style, perchance Noam Chomsky's, where his style as an academic and a political activist could be assessed.

Furthermore, *CorJAAL* may not have been the best reference corpus, but has nonetheless proven so on the basis of the research goals. Unarguably, the best layout of a reference corpus is an unsolved issue and has continued to foster additional discussions. Nevertheless, as previous research (Geluso; Hirsch, 2019) has shown, a different layout might affect the findings. Therefore, future studies could compare our study corpus to other existing reference corpora, if not a reference corpus compiled especially for the research itself, as was the case here, some other corpora, such as the *Corpus of Contemporary American English*, the *British National Corpus*, etc.

The fourth aspect is similarly concerned with corpus design. Across this dissertation, we have reiterated that other people may have made language decisions for Biber throughout RA publication phases. We understand that a style analysis is comprehensive and that gatekeepers/literacy brokers may have influenced Biber's stylistic choices. However, our attention has been particularly directed to the texts resulting from such editing and publishing processes, since tracking other people's editions to Biber's texts might have been a daunting, if not an impossible task to carry out here. Nevertheless, future studies could compile a corpus that consists of different versions of the same text and examine how the author's final style shifts throughout these editing phases. In doing so, we might understand more specifically how academic and language brokers directly impact an author's style.

The fifth aspect equally concerns methodology. We chose to investigate style based on keywords given their proven effectiveness for style investigation (Shepherd; Berber-Sardinha, 2013; Hyland, 2008b; 2010). However, the issue of the best metrics and tests used in keyword analysis is just as unresolved as is the reference corpus layout debate (Anthony, 2022b; 2024; Gabrielatos, 2018; Pojanapunya; Watson Todd, 2016). In this research, we have combined these resources available on a free corpus program, *AntConc*, which has also figured as a useful computer program in identifying language items that revealed themselves stylistic within the scope of this investigation. This use has illustrated Shepherd and Berber-Sardinha's (2013) desire that the future (now this very present) of corpus stylistics studies should lie in the use of a digital toolkit that could bridge the gap between stylisticians and corpus linguists. Nevertheless, like the research carried out by Pojanapunya and Watson Todd (2016), future

studies might compare keyword lists generated separately by likelihood and effect size measures and see how they illustrate potentially different stylistic authorial choices.

The sixth aspect regards the *Stylistic Metadiscourse Model of Interpersonal Interaction in Academic Writing* we have proposed here, noticeably as a critical reading of other existing models and functional analytical frameworks for assessing academic texts (Hyland, 2005; 2008a; 2009; Biber, 2006; Biber *et al.*, 2021). The main framework on which this model was based was initially designed by Hyland (2005) to evaluate how scholars often build interpersonal interaction with their readers throughout academic texts. Again, this was not a model firstly conceived for authorial style description and assessment. We hereby have decided to put together his model and other frameworks because, after excluding text-aboutness and non-text-aboutness words from our *AntConc*-generated keywords list, the remainder was classified as either *metadiscourse* or *stance*. These two issues have long been investigated by many scholars, but never so neatly and precisely as have Hyland and Biber. For this reason, we have opted for their models/frameworks because they seemed feasible to pursue style description and analysis.

However, because we have not used theories, models, or frameworks from stylistics, but from linguistics, we assume that we have envisaged a new perspective in the area: a corpus-based *style* approach. This assumption is no different from Crystal's (2013, p. 223) assertion that each style investigation develops its own approach, "[...] which is often as idiosyncratic as the characteristics of the style it investigates". In either case, though, future studies could fulfill two goals: 1) apply the *Stylistic Metadiscourse Model of Interpersonal Interaction in Academic Writing* to assess another author's style in academia; 2) apply this model to assess another author's style outside academia. These studies could, on the one hand, illustrate how the model carries on disclosing an authorial style in academia, and, on the other hand, demonstrate whether the model may be similarly applied to exploring author's style outside the academic realm.

Finally, the last issue derives from the sixth and concerns our precise focus to metadiscourse keywords. Leech and Short (2007), for example, advocate that a style analysis must be more specific to one feature or a set of them. As was the case here, our goal was reduced to metadiscourse features. This decision, obviously motivated by the constraints and limits of a doctoral investigation, may have overlooked other potentially relevant features that could equally illustrate Biber's style in academia. For instance, the text-aboutness and non-text-aboutness and non-metadiscourse keywords were discarded. Future research, though, could investigate these two other keyword types and check whether an author's style may be revealed

by them. A more precise suggestion could be on how scholars write about the same topic. Biber is not the only scholar who has investigated register, but it would be a worthwhile endeavor to assess how he and other scholars (i.e., Egbert, Gray, Conrad, etc.) write about this same topic, comparing the peculiar ways each one chooses specific text-aboutness nouns over others to address the same issue.

Despite and also because of all these aspects raised, we have come to conclude that Biber's style has proven to be mainly interactive while writing RAs, debunking the myth of a dry, impersonal discourse in academic writing. Once again, in view of our choice of this metadiscourse theoretical toolkit, *AntConc* keyword tool and its inbuilt tests and metrics, metadiscourse keywords, and full RAs (and not specific sections or other academic genres and registers), the author's style has therefore been identified as interactive. This is to say that style is not absolute or is revealed outside social and contextual constraints, but also derives therefrom. Should we have made decisions for other theories, methods, keywords, etc., the author's style might have been identified otherwise.

Given the theoretical and methodological parameters adopted in this dissertation, nevertheless, Biber's style has shown to be metadiscoursally interactive, whereby he (un)surprisingly has made individual metadiscourse stylistic choices, targeted to fulfill pre-determined purposes and effects, which, in his texts, aimed at organizing discourse so as to better accommodate his participating audience's interests, rhetorical expectations, and reading comprehension processing abilities. Finally, by taking on Crystal's (2013) early proposal, this study shows potential implications for the understanding of how an expert academic writer crafts an authorial style that may empower novice and other expert writers to craft a style that involves the audience in meaningful reading and writing experiences through academic texts.

REFERENCES

ANTHONY, Lawrence. *AntConc (Version 4.2.0) [Computer Software]*. Tokyo: Waseda University, 2022a. Available at: <https://www.laurenceanthony.net/software/antconc/> Access on: 10 Apr. 2023.

ANTHONY, Lawrence. *AntConc (Windows, MacOS, Linux) Build 4.2.0 [Help Manual]*. Tokyo: Waseda University, 2022b. Available at: <https://www.laurenceanthony.net/software/antconc/releases/AntConc420/help.pdf> Access on: 10 Apr. 2023.

ANTHONY, Lawrence. Common statistics used in corpus linguistics. Tokyo: Waseda University, 2023. Available at: https://www.laurenceanthony.net/resources/statistics/common_statistics_used_in_corpus_linguistics.pdf Access on: 11 Sep. 2024.

ANTHONY, Lawrence. Different keyness results when applying different effect size measure settings. Mountain View, 16-18 May 2024. Google: AntConc-Discussion Google Group. Available at: <https://groups.google.com/g/antconc/c/YfaUQ2CJWTE/m/6MEh1LZNBQAJ> Access on: 04 June 2024.

ATKINSON, Dwight. Writing for publication/writing for public execution: on the (personally) vexing notion of (personal) voice. In: CASANAVE, Christine Pearson; VANDRICK, Stephanie (ed.). *Writing for scholarly publication: behind the scenes in language education*. Mahwah / London: Lawrence Erlbaum Associates Publishers, 2003. p. 189-210.

BACON, Nora. Style in academic writing. In: DUNCAN, Mike; VANGURI, Star Medzerian (ed.). *The centrality of style*. Anderson: Parlor Press, 2013. p. 173-190.

BAKER, Mona. Towards a methodology for investigating the style of a literary translator. *Target*, Amsterdam/Philadelphia, v. 12, n. 2, p. 241-266, 2000.

BAKHTIN, Mikail. *The dialogic imagination: four essays by M. M. Bakhtin*. Austin: University of Texas Press, 1981.

BELL, Allan. Back in style: reworking audience design. In: ECKERT, Penelope; RICKFORD, John R. (ed.). *Style and sociolinguistic variation*. Cambridge: Cambridge University Press, 2001. p. 139-169.

BERBER SARDINHA, Antonio Paulo. Comparing corpora with WordSmith Tools: how large must the reference corpus be? In: Annual Meeting of the Association for Computational Linguistics, 38, 2000, Hong Kong. *Proceedings of the Workshop on Comparing Corpora...* Hong Kong: Hong Kong University of Science and Technology, 2000. p. 7-13.

BERBER SARDINHA, Antonio Paulo. *Linguística de corpus*. Barueri: Manole, 2004.

BHATIA, Vijay K. A generic view of academic discourse. In: FLOWERDEW, John (ed.). *Academic discourse*. Oxon/New York: Pearson Education Limited/Routledge, 2002. p. 21-39.

BIBER, Douglas. *Variation across speech and writing*. Cambridge: Cambridge University Press, 1988.

BIBER, Douglas. *University language: A corpus-based study of spoken and written registers*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2006.

BIBER, Douglas. Corpus linguistics and the study of literature: Back to the future? *Scientific Study of Literature*, Amsterdam, v. 1, n. 1, p. 15-23, 2011.

BIBER, Douglas. Interview with Douglas Biber. [Interview given to] Bethany Gray. *Journal of English Linguistics*, Thousand Oaks, v. 41, n. 4, p. 359-379, 2013.

BIBER, Douglas *et al.* *Longman grammar of spoken and written English*. England: Longman, 1999.

BIBER, Douglas *et al.* *Grammar of spoken and written English*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2021.

BIBER, Douglas; CONRAD, Susan. *Register, genre, and style*. 2. ed. Cambridge: Cambridge University Press, 2019.

BIBER, Douglas; GRAY, Bethany. *Grammatical complexity in academic English: linguistic change in writing*. Cambridge: Cambridge University Press, 2016.

BONDI, Marina. Perspectives on keywords and keyness: an introduction. In: BONDI, Marina; SCOTT, Mike (ed.). *Keyness in texts*. Amsterdam / Philadelphia: John Benjamins Publishing Company, 1996. p. 1-18.

BRAINE, George. Negotiating the gatekeepers: the journey of an academic article. In: CASANAVE, Christine Pearson; VANDRICK, Stephanie (ed.). *Writing for scholarly*

publication: behind the scenes in language education. Mahwah / London: Lawrence Erlbaum Associates Publishers, 2003. p. 87-108.

BREZINA, Vaclav. *Statistics in corpus linguistics: a practical guide*. Cambridge/New York: Cambridge University Press, 2018.

BURKE, Michael. Rhetoric and poetics: The classical heritage of stylistics. In: BURKE, Michael (ed.). *The Routledge Handbook of Stylistics*. London/New York: Routledge, 2014. p. 11-30.

CÂMARA JUNIOR, João Mattoso. *Contribuição à estilística portuguesa*. 3. ed. Rio de Janeiro: Ao livro técnico, 1978.

CASANAVE, Christine Pearson. Narrative braiding: constructing a multistrand portrayal of self as writer. In: CASANAVE, Christine Pearson; VANDRICK, Stephanie (ed.). *Writing for scholarly publication: behind the scenes in language education*. Mahwah / London: Lawrence Erlbaum Associates Publishers, 2003. p. 157-174.

CLARIVATE. *Journal Citation Reports*. London: Clarivate, 2023.

COUPLAND, Nikolas. *Style: language variation and identity*. Cambridge/New York: Cambridge University Press, 2007.

CRISMORE, Avon. *Metadiscourse: what is it and how is it used in school and non-school social science texts*. Urbana-Champaign: University of Illinois, 1983.

CRYSTAL, David. Stylistic profiling. In: AIJMER, Karin; ALTENBERG, Bengt (ed.). *English corpus stylistics: Studies in honour of Jan Svartvik*. New York / London: Routledge, 2013. p. 221-238.

CULPEPER, Jonathan. Computers, language and characterisation: An Analysis of six characters in *Romeo and Juliet*'. In: MELANDER-MARTTALA, Ulla; ÖSTMAN, Carin.; KYTÖ, Merja (ed.). *Conversation in life and in literature: Papers from the ASLA Symposium*. Uppsala: Universitetsstryckeriet/Association Suedoise de Linguistique Appliquee, 2002. p.11-30.

DEVINE, Nesta. Standpoint, style and self in writing. In: STEWART, Georgina Tuari; DEVINE, Nesta; BENADE, Leon (ed.). *Writing for publication: liminal reflections for academics*. Singapore: Springer, 2021. p. 31-40.

E.G. In: American Heritage Dictionary of the English Language. New York: HarperCollins Publishers LLC, 2011. 5.ed. Available at: <https://www.thefreedictionary.com/e.g>. Access on: 19 Apr. 2024.

ELBOW, Peter. What do we mean when we talk about voice in texts? In: Yancey, Kathleen Blake (ed.). *Voices on voice: perspectives, definitions, inquiry*. Urbana: National Council of Teachers of English, 1994. p. 1-35.

ELBOW, Peter. Voice in writing again: embracing contraries. *College English*, Amherst, v. 7, p.168-188, 2007.

FRANKENBERG-GARCIA, Ana. Investigating the collocations available to EAP writers. *Journal of English for Academic Purposes*, Amsterdam, v. 35, p. 93-104, 2018.

GABRIELATOS, Costas. Keyness analysis: nature, metrics and techniques. In: TAYLOR, Charlotte; MARCHI, Anna (ed.). *Corpus approaches to discourse: a critical review*. London/New York: Routledge, 2018. p. 225-258.

GARCIA, Othon Moacir. *Comunicação em prosa moderna: aprenda a escrever, aprendendo a pensar*. 26. ed. Rio de Janeiro: Editora FGV, 2006.

GELUSO, Joe; HIRCH, Roz. The reference corpus matters: comparing the effect of different reference corpora on keyword analysis. *Register Studies*, Amsterdam/Birmingham, v. 1, n. 2, p. 209–242, 2019.

GILLINGS; Mathew; MAUTNER, Gerlinde; BAKER, Paul. *Corpus-assisted discourse studies*. Cambridge/New York/Port Melbourne/New Delhi/Singapore: Cambridge University Press, 2023.

GOOGLE SCHOLAR. *Douglas Biber*. Menlo Park: Google.com, 2024. Available at: <https://scholar.google.com/citations?hl=pt-BR&user=mdWIU4MAAAAJ> Access on: 05 Sep. 2024.

GRABE, William; KAPLAN, Rober. *Theory and practice of writing: an applied linguistic perspective*. Harlow: Longman, 1996.

GRAY, Bethany. *Linguistic variation in research articles: when discipline tells only part of the story*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2015.

GRIEVE, Jack. Register variation explains stylometric authorship analysis. *Corpus Linguistics and Linguistic Theory*, Berlin, v. 19, n. 1, 1-31, 2023.

I.E. In: American Heritage Dictionary of the English Language. New York: HarperCollins Publishers LLC, 2011. 5.ed. Available at: <https://www.thefreedictionary.com/i.e>. Access on: 19 Apr. 2024.

IOANNIDIS, John. October 2023 data-update for “Updated science-wide author databases of standardized citation indicators”. *Elsevier Data Repository*, Amsterdam, 4 Oct. 2023. Available at: <https://elsevier.digitalcommonsdata.com/datasets/btchxktzyw/6> Access on: 12 Aug. 2024.

HALLIDAY, Michael Alexander Kirkwood.; MATTHIESSEN, Christian. *Construing experience through meaning: a language-based approach to cognition*. London: Cassell, 1999.

HALLIDAY, Michael Alexander Kirkwood.; MATTHIESSEN, Christian. *Halliday's introduction to functional grammar*. 4. ed. London/New York: Routledge, 2014.

HAMES, Irene. Peer review in a rapidly changing landscape. In: CAMPBELL, Robert; PENTZ, Ed; BORTHWICK, Ian (ed.). *Academic and professional publishing*. Cambridge: Chandos Publishing, 2012. p. 15-52.

HENRIQUES, Claudio Cezar. *Estilística e discurso: estudos produtivos sobre texto e expressividade*. Rio de Janeiro: Elsevier, 2011.

HYLAND, Ken. *Hedging in scientific research articles*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 1998a.

HYLAND, Ken. Boosting, hedging and the negotiation of academic knowledge. *Text*, Berlin, v. 18, n. 3, p. 349-382, 1998b.

HYLAND, Ken. Humble servants of the discipline? Self-mention in research articles. *English for Specific Purposes*, Amsterdam, v. 20, n. 3, p. 207-226, 2001.

HYLAND, Ken. Options of identity in academic writing. *ELT Journal Volume*, Oxford, v. 56, n. 4, p. 351-358, 2002a.

HYLAND, Ken. Authority and invisibility: authorial identity in academic writing. *Journal of Pragmatics*, Amsterdam, v. 34, p. 1091-1112, 2002b.

HYLAND, Ken. *Disciplinary discourses: social interactions in academic writing*. Ann Arbor: The University of Michigan Press, 2004.

HYLAND, Ken. *Metadiscourse: Exploring Interaction in writing*. London/New York: Continuum International Publishing Group, 2005.

HYLAND, Ken. As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, Amsterdam, v. 27, n. 1, p. 4-21, 2008a.

HYLAND, Ken. Small bits of textual material: A discourse analysis of Swales' writing. *English for Specific Purposes*, Amsterdam, v. 27, n. 2, p. 143-160, 2008b.

HYLAND, Ken. *Academic Discourse: English in a global context*. London/New York: Continuum International Publishing Group, 2009.

HYLAND, Ken. Community and individuality: performing identity in applied linguistics. *Written Communication*, Thousand Oaks, v. 27, n. 2, p. 159-188, 2010.

HYLAND, Ken. Peer review: Objective screening or wishful thinking? *Journal of English for Research Publication Purposes*, Amsterdam, v. 1, n. 1, p. 51-65, 2020.

HYLAND, Ken. A reflection on some of my recent research. *Sciendo*, Warsaw/Boston, v. 15, n. 1, p. 35-43, 2021.

HYLAND, Ken. The scholarly publishing landscape. In: HANGANU-BRESCH Cristina; ZERBE, Michael J.; CUTRUFELLO, Gabriel; MACI Stefania M. (ed.). *The Routledge handbook of scientific communication*. London/New York: Routledge, 2022. p. 15-25.

HYLAND, Ken; JIANG, Feng (Kevin). *Academic discourse and global publishing: disciplinary persuasion in changing times*. London/New York: Routledge, 2019.

HYLAND, Ken; TSE, Polly. Metadiscourse in academic writing: a reappraisal. *Applied Linguistics*, Oxford, v. 25, n. 2, p. 156-77, 2004.

IRVINE, Judith T. "Style" as distinctiveness: the culture and ideology of linguistic differentiation. In: ECKERT, Penelope; RICKFORD, John R. (ed.). *Style and sociolinguistic variation*. Cambridge: Cambridge University Press, 2001. p. 21-43.

IVANIČ, Roz. *Writing and identity: the discoursal construction of identity in academic writing*. Amsterdam/Philadelphia: John Benjamins Publishing, 1998.

JEFFRIES, Lesley; MCINTYRE, Dan. *Stylistics*. Cambridge: Cambridge University Press, 2010.

LEECH, Geoffrey; SHORT, Mick. *Style in fiction: a linguistic introduction to English fictional prose*. Harlow: Pearson/Longman, 2007.

LILLIS, Theresa; CURRY, Mary Jane. *Academic writing in a global context: the politics and practices of publishing in English*. London/New York: Routledge, 2010.

MAHLBERG, Michaela. Corpus analysis of literary texts. In: CHAPPELLE, Carol (ed.). *The encyclopedia of applied linguistics*. New Jersey: Blackwell Publishing, 2013. p. 1282-1289.

MALMKJAER, Kirsten. What happened to God and the angels: an exercise in translational stylistics. *Target*, Amsterdam/Birmingham, v. 15, p. 37-58, 2003.

MALMKJAER, Kirsten; CARTER, Ronald. Stylistics. In: MALMKJAER, Kirsten (ed.) *The linguistics encyclopedia*. 2. ed. London/New York: Routledge, 2002. p. 510-520.

MCENERY, Tony; WILSON, Andrew. *Corpus linguistics: an introduction*. 2. ed. Edinburgh: Edinburgh University Press, 2001.

MCINTYRE, Dan; WALKER, Brian. *Corpus stylistics: theory and practice*. Edinburgh: Edinburgh University Press Ltd, 2019.

MCMENAMIN, Gerald R. (ed.) *Forensic Linguistics: advances in forensic stylistics*. Boca Raton: CRC Press, 2002.

NAPIERALA, Matthew. What Is the Bonferroni Correction? *American Academy of Orthopaedic Surgeons*, Illinois, n. 2, 1 Apr. 2012, 2012. Available at: https://docs.ufpr.br/~giolo/LivroADC/Material/S3_Bonferroni%20Correction.pdf Access on: 12 May 2023.

NØRGAARD, Nina; BUSSE, Beatrix; MONTORO, Rocío. *Key terms in stylistics*. London/New York: Continuum International Publishing Group, 2010.

PEARSON, William S.; ABDOLLAHZADEH, Esmaeel. Metadiscourse in academic writing: a systematic review. *Lingua*, Amsterdam, v. 263, p. 1-32, 2023.

POJANAPUNYA, Punjaporn; WATSON TODD, Richard. Log-likelihood and odds ratio: Keyness statistics for different purposes of keyword analysis. *Corpus Linguistics and Linguistic Theory*, Berlin, v. 14, n. 1, 133-167, 2016.

PRINA DUTRA, Deise; BERBER SARDINHA, Tony. A multi-dimensional typology of English research article sections. In: THE AMERICAN ASSOCIATION FOR APPLIED LINGUISTICS CONFERENCE (AAAL), 2021, virtual conference. *Proceedings* [...] Marietta: AAAL, 2021.

RAYSON, Paul; BERRIDGE, Damon; FRANCIS, Brian. Extending the Cochran Rule for the comparison of word frequencies between corpora. In: PURNELLE, G  rald; FAIRON, C  drick; DISTER, Anne (ed.). *Le poids des mots: Proceedings of the 7th International Conference on Statistical Analysis of Textual Data (JADT 2004)*. Louvain-la-Neuve: Presses universitaires de Louvain, 2004. p. 926–936.

SALDANHA, Gabriela. Translator style: methodological considerations. *The Translator*, UK, v. 17, n. 1, p. 25-50, 2011.

SCOTT, Mike. In search for a bad reference corpus. In: ARCHER, Dawn. *What's in a word-list?: investigating word frequency and keyword extraction*. Farnham/ Burlington: Ashgate, 2009. p. 79-92.

SCOTT, Mike. *WordSmith Tools manual*. Stroud: Lexical Analysis Software Ltd., 2022. v. 8. Available at: https://lexically.net/downloads/version_64_8/HTML/index.html Access on: 03 Jun. 2022.

SCOTT, Mike; TRIBBLE, Christopher. *Textual patterns: key words and corpus analysis in language education*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2006.

SEMINO, Elena; SHORT, Mick. *Corpus Stylistics: Speech, writing and thought presentation in a corpus of English writing*. London/New York: Routledge, 2004.

SHEPHERD, Tania M. G.; BERBER SARDINHA, Tony. A rough guide to doing corpus stylistics. *Matraga*, Rio de Janeiro, v. 20, n. 32, p. 66-89, 2013.

SMITH, Nick; HARDIE, Andrew. *Testing for significance: log likelihood*. Lancaster: Lancaster University, 2009. Available at: https://www.lancaster.ac.uk/fss/courses/ling/corpus/blue/108_4.htm Access on: 08 May 2023.

- SINCLAIR, John. *Trust the text: language corpus and discourse*. London: Routledge, 2004.
- SILVER, Mark. Voice and stance across disciplines in academic discourse. In: HYLAND, Ken; GUINDA, Carmen (ed.). *Stance and voice in written academic genres*. London: Palgrave Macmillan, 2012. p. 202-217.
- STUBBS, Michael. Conrad in the computer: examples of quantitative stylistic methods. *Language and Literature*, Thousand Oaks, v. 14, n. 1, p. 5-24, 2005.
- SWALES, John M. *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press, 1990.
- SWALES, John M. *Research genres: explorations and applications*. Cambridge: Cambridge University Press, 2004.
- TARDY, Christine. Current conceptions of voice. In: HYLAND, Ken; GUINDA, Carmen (ed.). *Stance and voice in written academic genres*. London: Palgrave Macmillan, 2012. p. 34-47.
- THOMPSON, Geoff. Interaction in academic writing: learning to argue with the reader. *Applied Linguistics*, Oxford, v. 22, n. 1, p. 58-78, 2001.
- THOMPSON, Geoff; THETELA, Puleng. The sound of one hand clapping: the management of interaction in written discourse. *TEXT*, Berlin, v. 15, n. 1, 103-27, 1995.
- TOGNINI-BONELLI, Elena. *Corpus linguistics at work*. Amsterdam/Birmingham: John Benjamins Publishing Company, 2001.
- VANDE KOPPLE, William. Some exploratory discourse on metadiscourse. *College Composition and Communication*, New York/ Ann Arbor, v. 36, n. 1, p. 82-93, 1985.
- WALTMAN, Ludo; TRAAG, Vincent A. Use of the journal impact factor for assessing individual articles: Statistically flawed or not? *F1000Research*, London, v. 9, p. 1-29, 2000.
- WOOD, Alistair. International scientific English: the language of research scientists around the world'. In: FLOWERDEW, John; PEACOCK, Matthew (ed.). *Research perspectives on English for Academic Purposes*. Cambridge: Cambridge University Press, 2001. p. 71-83.
- WYNNE, Martin. Stylistics: corpus approaches. In: BROWN, Keith; ASHER, Ronald (ed.). *Encyclopedia of language and linguistics*. 2. ed. Elsevier, 2006. p. 223-226.

APPENDIX A – Stylistic metadiscourse model of interpersonal interaction in academic writing: interactive dimension

Table 37. Interactive dimension

Functional category: discourse-organizer (research-oriented and text-oriented)		
Functional subcategories	Functions	Linguistic features and examples
<i>Transitions</i>	express additive, causative, and contrastive relations between main clauses	Adverbials, conjunctions, and lexical bundles: addition: <i>and; furthermore; in addition to (the); moreover; by the way; etc.</i> consequence: <i>thus; therefore; consequently; so; etc.</i> comparison / contrast: <i>similarly; likewise; equally; in the same way; correspondingly; in contrast (to the); however; but; on the contrary; on the other hand; with the exception of; nevertheless; anyway; in any case; etc.</i>
<i>Frame markers</i>	refer to research/discourse acts, sequences, stages, or topic shifts	Adverbials, conjunctions, bundles: <i>at the beginning of; a wide range of; at the same time; (in/of) the present study; first; then; next; to summarize; in conclusion; in sum; by way of introduction; finally; to conclude; (as) I argue here; my purpose is; the paper proposes; I hope to persuade; there are several reasons why; well; right; OK; now; let us return to; the use / role / purpose / operation / magnitude / structure / size / surface of (the); it was found (that); these results suggest (that); (in) the case of; with respect to (the); on the basis of; etc.</i>
<i>Evidentials</i>	refer to information from other texts, reporting on but not taking a stance towards other scholars' research and findings	Adverbials: <i>according to; as for/as to; etc.</i> Verbs + <i>that</i>-clause: <i>announce; argue; assert; claim; contend; declare; emphasize; explain; mention; promise; propose; recommend; remark; require; respond; say; state; seek; tell; etc.</i> Verbs + <i>to</i>-clause: <i>allow; ask; be said; claim; enable; help; instruct; invite; manage; remind; request; show; teach; tell; urge; warn; etc.</i>
<i>Code glosses</i>	supply additional information by rephrasing, explaining, or elaborating on what has been said	Adverbials and other lexical bundles: <i>e.g.; for example/instance; in other words; namely; such as; that is; (in) the case of; etc.</i>

Source: Adapted from Hyland (2005; 2008a; 2009), Biber (2006), and Biber *et al.* (2021).

APPENDIX B – Stylistic metadiscourse model of interpersonal interaction in academic writing: interactional dimension

Table 38. Interactional dimension

Functional category: participant-oriented		
a) Stance markers		
Functional subcategories	Functions	Linguistic features and examples
<i>Hedges (epistemic)</i>	withhold writer's commitment to a proposition, allowing information to be presented as an opinion rather than a fact and opening that position to negotiation	(Semi-)Modal verbs: prediction/volition: <i>be going to; may; might; will; would; shall.</i> Adverbials and lexical bundles: <i>about; apparently; kind of; in most cases/instances; maybe; perhaps; possibly; predictably; probably; reportedly; roughly; sort of; etc.</i> Verbs + <i>that</i>-clause: <i>assume; believe; doubt; gather; guess; hypothesize; imagine; predict; presuppose; presume; reckon; seem; speculate; suggest; suppose; suspect; think; etc.</i> Verbs + <i>to</i>-clause: <i>appear; attempt; assume; believe; consider; expect; find; forget; happen; hesitate; imagine; judge; know; learn; presume; pretend; remember; see; seem; suppose; tend; try; etc.</i> Adjectives + <i>that</i>-clause: <i>apparent; doubtful; likely; possible; probable; unlikely; etc.</i> Adjectives + <i>to</i>-clause: <i>likely; prone; unlikely; etc.</i> Nouns + <i>that</i>-clause: <i>assumption; belief; hypothesis; idea; implication; impression; notion; opinion; possibility; presumption; suggestion; etc.</i> Bundles: <i>are likely to be; may be due to; it is possible that; etc.</i>
<i>Boosters (epistemic)</i>	emphasize writer's certainty and confident voice, strengthening an argument but closing down possible alternatives to a proposition	Modal verbs: necessity/obligation: <i>must; can; could; should; (had) better; have to; got to; ought to.</i> Adverbials and lexical bundles: <i>actually; always; certainly; clearly; definitely; evidently; generally; it is clear that; indeed; inevitably; in fact; never; no doubt; of course; obviously; really; technically; typically; truthfully; undoubtedly; usually; without doubt; etc.</i> Verbs + <i>that</i>-clause: <i>conclude; demonstrate; determine; discover; find; know; learn; mean; notice; observe; prove; realize; recognize; remember; see; show; understand; etc.</i> Verbs + <i>to</i>-clause: <i>agree; choose; decide; mean; prefer; enable; instruct; manage; oblige; order; permit; prompt; require; etc.</i> Adjectives + <i>that</i>-clause: <i>certain; clear; confident; convinced; correct; evident; false; impossible; inevitable; obvious; positive; right; sure; true; well-known; etc.</i> Adjectives + <i>to</i>-clause: <i>apt; certain; due; guaranteed; liable; sure; etc.</i> Nouns + <i>that</i>-clause: <i>assertion; claim; conclusion; contention; conviction; discovery; fact; grounds; knowledge; observation; principle; realization; reason; result; statement; view; etc.</i>
<i>Attitude markers (attitudinal)</i>	express writer's affective attitudes, emotions, evaluations to/on a proposition, conveying surprise, agreement, importance,	Adverbials: <i>admittedly; amazingly; astonishingly; confidentially; conveniently; curiously; hopefully; even worse; frankly; fortunately; honestly; importantly; ironically; mainly; primarily; rightly; sadly; surprisingly; unfortunately; etc.</i> Verbs + <i>that</i>-clause: <i>agree; anticipate; complain; concede; ensure; expect; fear; feel; forget; hope; mind; prefer; pretend; require; wish; worry; etc.</i>

	obligation, frustration, etc.	Verbs + <i>to</i>-clause: (desire/intention/decision/effort): agree; allow; choose; decide; encourage; fail; hate; hesitate; help; hope; intend; like; love; need; plan; persuade; prefer; refuse; want; wish; etc. Adjectives + <i>that</i>-clause: afraid; amazed; amazing; appropriate; aware; concerned; conceivable; crucial; disappointed; encouraged; essential; fortunate; glad; happy; hopeful; logical; imperative; inconceivable; incredible; interesting; lucky; necessary; nice; noteworthy; odd; pleased; remarkable; ridiculous; shocked; surprised; surprising; strange; unacceptable; unfortunate; worried; etc. Adjectives + <i>to</i>-clause: afraid; anxious; ashamed; bad/worse; careful; convenient; determined; difficult; disappointed; eager; easier; easy; eligible; embarrassed; essential; glad; good/better/best; happy; hard; hesitant; important; (im)possible; interesting; (in)appropriate; inclined; necessary; nice; obliged; pleased; prepared; proud; puzzled; ready; relieved; reluctant; silly; smart; sorry; surprised; surprising; stupid; tough; worried; reasonable; (un)able; useful; useless; unreasonable; (un)willing; wise; wrong; etc. Nouns + <i>that</i>-clause: feeling; hope; thought; etc. Nouns + <i>to</i>-clause: agreement; decision; desire; failure; inclination; intention; obligation; opportunity; plan; promise; proposal; reluctance; responsibility; right; tendency; threat; wish; willingness; etc.
<i>Self-mentions (epistemic)</i>	make explicit reference to the writer him/herself	First-person pronouns and possessive adjectives: <i>I; me; mine; we; our; ours</i> ; etc.
b) Engagement markers		
Functional subcategories	Functions	Linguistic features and examples
<i>Reader pronouns</i>	explicitly build relationship with reader	Pronouns: <i>you; your</i> ; inclusive <i>we</i> ; etc. Interjections: <i>you may notice</i> ; etc.
<i>Personal asides</i>	interrupt the argument to offer a comment on what has been said	Use of appositive phrases or clauses in between dashes or parentheses.
<i>Endophoric and directive markers</i>	direct readers to another part of the text or to another text (textual acts); direct readers how to carry out some action in the real-world (physical acts); instruct readers how to interpret an argument, explicitly encouraging them to note, concede or consider some argument or claim in the text (cognitive acts)	Imperative verbs and lexical bundles: <i>see Figure 2; note; consider; refer to the next section; as noted above; as shown in figure; as can be seen; it should be noted that</i> ; etc. Modal verbs: <i>should; must; have to</i> ; etc.
<i>Questions</i>	invite engagement, encourage curiosity, bring interlocutors	Use of interrogative and/or rhetorical questions .

	into the discourse arena	
<i>Appeals to shared knowledge</i>	ask readers to recognize something as familiar or accepted, identify themselves with particular views, and assign to them a role in creating the argument.	Adverbials: <i>of course; obviously; by the way; etc.</i>

Source: Adapted from Hyland (2005; 2008a; 2009), Biber (2006), and Biber *et al.* (2021).

APPENDIX C – List of journals that published Biber's texts

1. American Speech
2. Applied Linguistics Forum News
3. Computational Linguistics
4. Computers and the Humanities
5. Corpus Linguistics and Linguistic Theory
6. Discourse Processes
7. Discourse Studies
8. Indonesian Journal of English Language Teaching
9. International Journal of Corpus Linguistics
10. Journal of African Languages and Linguistics
11. Journal of English for Academic Purposes
12. Journal of Historical Pragmatics
13. Language
14. Languages in Contrast
15. Linguistics
16. Literary and Linguistic Computing
17. Register Studies
18. Scientific Study of Literature
19. Studia Linguistica
20. Studies in African Linguistics
21. Text.

APPENDIX D – List of bibliographic references of the texts composing *BC*

1. BIBER, D. A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, [S. I.], v. 14, n. 3, p. 275–311, 2009.
2. BIBER, D. A register perspective on grammar and discourse: variability in the form and use of English complement clauses. *Discourse Studies*, [S. I.], v. 1, n. 2, p. 131–150, 1999.
3. BIBER, D. A textual comparison of British and American writing. *American Speech*, [S. I.], v. 62, n. 2, p. 99, 1987.
4. BIBER, D. A typology of English texts. *Linguistics*, [S. I.], v. 27, n. 1, p. 3–44, 1989.
5. BIBER, D. Co-occurrence patterns among collocations: a tool for corpus-based lexical knowledge acquisition. *Computational Linguistics*, [S. I.], v. 19, n. 3, p. 531–538, 1993a.
6. BIBER, D. Corpus linguistics and the study of English grammar. *Indonesian Journal of English Language Teaching*, [S. I.], v. 1, n. 1, 2005a.
7. BIBER, D. Corpus linguistics and the study of literature: Back to the future?. *Scientific Study of Literature*, [S. I.], v. 1, n. 1, p. 15–23, 2011.
8. BIBER, D. Historical patterns for the grammatical marking of stance: A cross-register comparison. *Journal of Historical Pragmatics*, [S. I.], v. 5, n. 1, p. 107–136, 2004.
9. BIBER, D. Investigating language use through corpus-based analyses of association patterns. *International Journal of Corpus Linguistics*, [S. I.], v. 1, n. 2, p. 171–197, 1996.
10. BIBER, D. Investigating macroscopic textual variation through multifeature/multidimensional analyses. *Linguistics*, [S. I.], v. 23, n. 2, 1985.
11. BIBER, D. Methodological issues regarding corpus-based analyses of linguistic variation. *Literary and Linguistic Computing*, [S. I.], v. 5, n. 4, p. 257–269, 1990.
12. BIBER, D. On the complexity of discourse complexity: A multidimensional analysis. *Discourse Processes*, [S. I.], v. 15, n. 2, p. 133–163, 1992a.
13. BIBER, D. On the investigation of spoken/written differences. *Studia Linguistica*, [S. I.], v. 40, n. 1, p. 1–21, 1986a.
14. BIBER, D. On the role of computational, statistical, and interpretive techniques in multi-dimensional analyses of register variation: A reply to Watson. *Text & Talk*, [S. I.], v. 15, n. 3, p. 341–370, 1995.
15. BIBER, D. Pragmatic roles in central Somali narrative discourse. *Studies in African Linguistics*, [S. I.], v. 15, n. 1, p. 1–26, 1984a.
16. BIBER, D. Register as a predictor of linguistic variation. *Corpus Linguistics and Linguistic Theory*, [S. I.], v. 8, n. 1, p. 9–37, 2012.
17. BIBER, D. Representativeness in Corpus Design. *Literary and Linguistic Computing*, [S. I.], v. 8, n. 4, 1993b.
18. BIBER, D. Spoken and written textual dimensions in English: resolving the contradictory findings. *Language*, [S. I.], v. 62, n. 2, p. 384, 1986b.
19. BIBER, D. Stance in spoken and written university registers. *Journal of English for Academic Purposes*, [S. I.], v. 5, n. 2, p. 97–116, 2006.
20. BIBER, D. Text-linguistic approaches to register variation. *Register Studies*, [S. I.], v. 1, n. 1, p. 42–75, 2019.
21. BIBER, D. The diachronic development of preverbal case markers in Somali. *Journal of African Languages and Linguistics*, [S. I.], v. 6, n. 1, 1984b.

22. BIBER, D. The multi-dimensional approach to linguistic analyses of genre variation: An overview of methodology and findings. *Computers and the Humanities*, [S. I.], v. 26, n. 5–6, p. 331–345, 1992b.
23. BIBER, D. Using multi-dimensional analysis to explore cross-linguistic universals of register variation. *Languages in Contrast*, [S. I.], v. 14, n. 1, p. 7–34, 2014.
24. BIBER, D. Using register-diversified corpora for general language studies. *Computational Linguistics*, [S. I.], v. 19, n. 2, p. 219–241, 1993c.
25. BIBER, D. What can corpus linguistics tell us about English grammar. *Applied Linguistics Forum News*, [S. I.], v. 26, n. 1, 2005b.

APPENDIX E – Topics and impact factors of the *CorJAAL* journalsTable 39. Topics and impact factors of the *CorJAAL* journals

Journals	IF (2022)	IF (5-year)	Topics
Applied linguistics ³²	3.6	5.6	- connections between scholarly discourses, theories, and research methods from a range of areas; - current, cutting-edge theory and practice in applied linguistics; - institutional practices in education, the workplace, healthcare, legal systems, and the media; - language-related problems and solutions in real-world contexts.
System ³³	6	6	- applications of educational technology and applied linguistics to problems of foreign language teaching and learning; - learning and teaching of all languages as second or foreign languages in all countries.
TESOL Quarterly ³⁴	3.2	4.665	- implications and applications of research in anthropology, applied and theoretical linguistics, communication education, English education (including reading and writing theory) psycholinguistics, psychology, first and second language acquisition, sociolinguistics; - instructional methods, materials, and techniques; - issues in research and research methodology; - language planning; - professional preparation, curriculum design and development; - professional standards; - psychology and sociology of language learning and teaching; - testing and evaluation.

Source: Elaborated by the author.

³² Available at: <https://academic.oup.com/applij/pages/About> Access on: 09 Apr. 2024.³³ Available at: <https://www.sciencedirect.com/journal/system/about/aims-and-scope> Access on: 09 Apr. 2024.³⁴ Available at: <https://onlinelibrary.wiley.com/page/journal/15457249/homepage/productinformation.html> Access on: 09 Apr. 2024.

APPENDIX F – List of authors whose texts compose *CorJAAL*Table 40. List of authors whose texts compose *CorJAAL* separated by journal

Journals	Authors
Applied linguistics	1. Adriam Holliday 2. Amy Snyder Ohta 3. Brigitta Busch 4. Craig Chaudron 5. Daniel Andrea 6. Deborah Cameron 7. Dee Gardner 8. Diane Larsen-Freeman 9. Dianne Wall 10. Dilin Liu 11. Durk Gorter 12. Elite Olshtain 13. Evangeline Marlos Varonis 14. Fred Genesee 15. Graham Crookes 16. Greg Myers 17. Helen Basturkmen 18. J. Charles Alderson 19. Jack C. Richards 20. James Jason Legatto 21. James P. Lantolf 22. Jasone Cenoz 23. Jean W. Bodman 24. Joan Kelly Hall 25. John M. Norris 26. John Read 27. Judit Kormos 28. Kata Csizer 29. Kees De Bot 30. Ken Hyland 31. Kenji Hakuta 32. Kevin R. Gregg 33. Larry Vandergrift 34. Lei Lei 35. Lourdes Ortega 36. Lydia White 37. Manfred Pienemann 38. Marianne Verhallen 39. Mark Davies 40. Miriam Eisenstein 41. Nicos C. Sifakis 42. Nigel Harwood 43. Norbert Schmitt 44. Paul Nation 45. Pauline Foster 46. Peter Skehan 47. Peter D. Macintyre 48. Polly Tse 49. R L. Allwright 50. Richard W. Schmidt 51. Rob Schoonen 52. Robin Goulden 53. Rod Ellis 54. Ron Martinez 55. Sandra S. Fotos 56. Shawn Loewen

	57. Shoshana Blum-Kulka 58. Susan Gass 59. Teresa Pica 60. Thom Kiddle 61. Wen-Ta Tseng 62. William Frawley 63. Zoltan Dornyei
System	1. Adel Abu Radwan 2. Alan Hirvela 3. Alastair Pennycook 4. Albert T.Y. Wong 5. Anthony Bruton 6. Brian Paltridge 7. Bruce Roberts 8. Carla Meskill 9. Cecilia Tong-Fredericks 10. Christopher Jones 11. Clare Mar-Molinero 12. Cynthia White 13. David Carver 14. Donald M. Morrison 15. Eri Anan 16. G. Imlah 17. Gerard M. Willems 18. Gillian Lord 19. Graeme Storer 20. J. B. H. Du Boulay 21. Jean-Marc Dewaele 22. Jian-E Peng 23. Judith A. Burry-Stock 24. Kaiqi Shao 25. Ken Hyland 26. Lara Lomicka 27. Laura J. Nicholson 28. Lawrence Jun Zhang 29. Lew Barnett 30. Lonnie Turbeet 31. Ludo Beheydt 32. Marie-Jose Gremmo 33. Mark Warschauer 34. Martin Lamb 35. Mary Schleppegrell 36. Muhammad Rahimi 37. Murray J. Munro 38. Natasha Anthony 39. Nourollah Zarrinabadi 40. Paul A. Kusel 41. Paul Nation 42. Pidlip Riley 43. Rebecca L. Oxford 44. Reinhard Pekrun 45. Rob Batstone 46. Rod Ellis 47. Ron I. Thomson 48. S.M. Conrad 49. Simon Borg 50. Susan Gray 51. Teresa Mihwa Chung 52. Tim Johns 53. Tracey M. Derwing

	54. Vicky Wright 55. William Savage 56. William Littlewood 57. Yan Jiang
TESOL Quarterly	1. A. Suresh Canagarajah 2. Alastair Pennycook 3. Andrew D. Cohen 4. Awad El Karim M. Ibrahim 5. Bonny Norton Peirce 6. Catherine Doughty 7. Chris Davison 8. Christoph A. Hafner 9. Clyde A. Warden 10. Constant Leung 11. Daniel M. Horowitz 12. Deborah J. Short 13. Eli Hinkel 14. Elite Olshtain 15. Elke Peters 16. Ellen Block 17. Ellen L. Block 18. Elsa Roberts Auerbach 19. Gloria Park 20. Graham Crookes 21. Huo-Tsan Chang 22. John Field 23. John M. Green 24. Judit Kormos Kata Csizér 25. Judy F. Chen 26. Karen E. Johnson 27. Karen L. Porreca 28. Mary Jane Curry 29. Michael H. Long 30. Muriel Saville-Troike 31. Murray J. Munro 32. N. S. Prabhu 33. Nancy H. Hornberger 34. Naoko Morita 35. Ofelia Garcia 36. Patricia A. Porter 37. Paula R. Golombek 38. Rebecca Oxford 39. Richard Young 40. Roberta G. Abraham 41. Roberta J. Vann 42. Rod Ellis 43. Scott Grapin 44. Taco Justus Homburg 45. Teresa Pica 46. Theresa Lillis 47. Thomas K. Ricento 48. Tracey M. Derwing 49. Virginia P. Collier 50. Vivian Zamel 51. Xiaofei Lu

Source: Our elaboration

APPENDIX G – List of bibliographic references of the texts composing *CorJAAL***Applied Linguistics**

1. ALDERSON, J. C.; WALL, D. Does washback exist?. *Applied Linguistics*, Oxford, v. 14, n. 2, p. 115–129, 1993.
2. ALLWRIGHT, R. L. The importance of interaction in classroom language learning. *Applied Linguistics*, Oxford, v. 5, n. 2, p. 156–171, 1984.
3. BASTURKMEN, H. Teachers' stated beliefs about incidental focus on form and their classroom practices. *Applied Linguistics*, Oxford v. 25, n. 2, p. 243–272, 2004.
4. BLUM-KULKA, S.; OLSHTAIN, E. Requests and apologies: a cross-cultural study of speech act realization patterns (CCSARP). *Applied Linguistics*, Oxford, v. 5, n. 3, p. 196–213, 1984.
5. BOT, K. D. A bilingual production model: Levelt's 'speaking' model adapted. *Applied Linguistics*, Oxford, v. 13, n. 1, p. 1–24, 1992.
6. BUSCH, B. The linguistic repertoire revisited. *Applied Linguistics*, Oxford, v. 33, n. 5, p. 503–523, 2012.
7. CAMERON, D. Language, gender, and sexuality: current issues and new directions. *Applied Linguistics*, Oxford, v. 26, n. 4, p. 482–502, 2005.
8. CENOZ, J.; GENESEE, F.; GORTER, D. Critical analysis of CLIL: taking stock and looking forward. *Applied Linguistics*, Oxford, v. 35, n. 3, p. 243–262, 2014.
9. CHAUDRON, C.; RICHARDS, J. C. The effect of discourse markers on the comprehension of lectures. *Applied Linguistics*, Oxford, v. 7, n. 2, p. 113–127, 1986.
10. CROOKES, G. Towards a validated analysis of scientific text structure. *Applied Linguistics*, Oxford, v. 7, n. 1, p. 57–70, 1986.
11. FOSTER, P.; OHTA, A. S. Negotiation for meaning and peer assistance in second language classrooms. *Applied Linguistics*, Oxford, v. 26, n. 3, p. 402–430, 2005.
12. FOTOS, S. S. Consciousness raising and noticing through focus on form: grammar task performance versus formal instruction. *Applied Linguistics*, Oxford, v. 14, n. 4, p. 385–407, 1993.
13. FRAWLEY, W.; LANTOLF, J. P. Second language discourse: a Vygotskian perspective. *Applied Linguistics*, Oxford, v. 6, n. 1, p. 19–44, 1985.
14. GARDNER, D.; DAVIES, M. A new academic vocabulary list. *Applied Linguistics*, Oxford, v. 35, n. 3, p. 305–327, 2014.
15. GOULDEN, R.; NATION, P.; READ, J. How large can a receptive vocabulary be?. *Applied Linguistics*, Oxford, v. 11, n. 4, p. 341–363, 1990.
16. GREGG, K. R. Krashen's monitor and Occam's razor. *Applied Linguistics*, Oxford, v. 5, n. 2, p. 79–100, 1986.
17. GREGG, K. R. Taking explanation seriously; or, let a couple of flowers bloom. *Applied Linguistics*, Oxford, v. 14, n. 3, p. 276–294, 1993.
18. HALL, J. K. The role of oral practices in the accomplishment of our everyday lives: the sociocultural dimension of interaction with implications for the learning of another language. *Applied Linguistics*, Oxford, v. 14, n. 2, p. 145–166, 1993.
19. HARWOOD, N. 'We do not seem to have a theory ... the theory I present here attempts to fill this gap': inclusive and exclusive pronouns in academic writing. *Applied Linguistics*, Oxford, v. 26, n. 3, p. 343–375, 2005.
20. HOLLIDAY, A. Small cultures. *Applied Linguistics*, Oxford, v. 20, n. 2, p. 237–264, 1999.

21. HYLAND, K. Academic attribution: citation and the construction of disciplinary knowledge. *Applied Linguistics*, Oxford, v. 20, n. 3, p. 341–367, 1999.
22. HYLAND, K. Writing without conviction? hedging in science research articles. *Applied Linguistics*, Oxford, v. 17, n. 4, p. 433–454, 1996.
23. HYLAND, K.; TSE, P. Metadiscourse in academic writing: a reappraisal. *Applied Linguistics*, Oxford, v. 25, n. 2, p. 156–177, 2004.
24. KENJI, H.; D'ANDREA, D. Some properties of bilingual maintenance and loss in Mexican background high-school students. *Applied Linguistics*, Oxford, v. 13, n. 1, p. 72–99, 1992.
25. KORMOS, J.; KIDDLE, T.; CSIZER, K. Systems of goals, attitudes, and self-related beliefs in second-language-learning motivation. *Applied Linguistics*, Oxford, v. 32, n. 5, p. 495–516, 2011.
26. LARSEN-FREEMAN, D. The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics*, Oxford, v. 27, n. 4, p. 590–619, 2006.
27. LEI, L.; LIU, D. Research trends in applied linguistics from 2005 to 2016: a bibliometric analysis and its implications. *Applied Linguistics*, Oxford, v. 40, n. 3, p. 540–561, 2019.
28. MACINTYRE, P. D.; LEGATTO, J. J. A dynamic system approach to willingness to communicate: developing an idiodynamic method to capture rapidly changing affect. *Applied Linguistics*, Oxford, v. 32, n. 2, p. 149–171, 2011.
29. MARTINEZ, R.; SCHMITT, N. A phrasal expressions list. *Applied Linguistics*, Oxford, v. 33, n. 3, p. 299–320, 2012.
30. MYERS, G. The pragmatics of politeness in scientific articles. *Applied Linguistics*, Oxford, v. 10, n. 1, p. 1–35, 1989.
31. NORRIS, J. M.; ORTEGA, L. Towards an organic approach to investigating CAF in instructed SLA: the case of complexity. *Applied Linguistics*, Oxford, v. 30, n. 4, p. 555–578, 2009.
32. PICA, T. Second-language acquisition, social interaction, and the classroom. *Applied Linguistics*, Oxford, v. 8, n. 1, p. 3–21, 1987.
33. PIENEMANN, M. Is Language teachable? psycholinguistic experiments and hypotheses. *Applied Linguistics*, Oxford, v. 10, n. 1, p. 52–79, 1989.
34. SCHMIDT, R. W. The role of consciousness in second language learning. *Applied Linguistics*, Oxford, v. 11, n. 2, p. 129–158, 1990.
35. SIFAKIS, N. C. ELF Awareness in english language teaching: principles and processes. *Applied Linguistics*, Oxford, v. 40, n. 2, p. 288–306, 2019.
36. SKEHAN, P. Modelling second language performance: integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, Oxford, v. 30, n. 4, p. 510–532, 2009.
37. SWALES, J. Citation analysis and discourse analysis. *Applied Linguistics*, Oxford, v. 7, n. 1, p. 39–56, 1986.
38. TSENG, W.-T.; DÖRNYEI, Z.; SCHMITT, N. A new approach to assessing strategic learning: the case of self-regulation in vocabulary acquisition. *Applied Linguistics*, Oxford, v. 27, n. 1, p. 78–102, 2006.
39. VANDERGRIFT, L. Relationships among motivation orientations, metacognitive awareness and proficiency in l2 listening. *Applied Linguistics*, Oxford, v. 26, n. 1, p. 70–89, 2005.
40. VARONIS, E. M.; GASS, S. Non-native/non-native conversations: a model for negotiation of meaning. *Applied Linguistics*, Oxford, v. 6, n. 1, p. 71–90, 1985.
41. VERHALLEN, M. Lexical knowledge of monolingual and bilingual children. *Applied Linguistics*, Oxford, v. 14, n. 4, p. 344–363, 1993.

42. WHITE, L. Against comprehensible input: the input hypothesis and the development of second-language competence. *Applied Linguistics*, Oxford, v. 8, n. 2, p. 95–110, 1987.

System

1. BATSTONE, R.; ELLIS, R. Principled grammar teaching. *System*, Amsterdam, v. 37, n. 2, p. 194–204, 2009.
2. BEHEYDT, L. The semantization of vocabulary in foreign language learning. *System*, Amsterdam, v. 15, n. 1, p. 55–67, 1987.
3. BORG, S. The impact of in-service teacher education on language teachers' beliefs. *System*, Amsterdam, v. 39, n. 3, p. 370–380, 2011.
4. BRUTON, A. Is CLIL so beneficial, or just selective? Re-evaluating some of the research. *System*, Amsterdam, v. 39, n. 4, p. 523–532, 2011.
5. CARVER, D. Plans, learner strategies and self direction in language learning. *System*, Amsterdam, v. 12, n. 2, p. 123–131, 1984.
6. CHUNG, T. M.; NATION, P. Identifying technical vocabulary. *System*, Amsterdam, v. 32, n. 2, p. 251–263, 2004.
7. CONRAD, S. M. The importance of corpus-based research for language teachers. *System*, Amsterdam, v. 27, n. 1, p. 1–18, 1999.
8. DERWING, T. M.; THOMSON, R. I.; MUNRO, M. J. English pronunciation and fluency development in Mandarin and Slavic speakers. *System*, Amsterdam, v. 34, n. 2, p. 183–193, 2006.
9. ELLIS, R. Interpretation-based grammar teaching. *System*, Amsterdam, v. 21, n. 1, p. 69–78, 1993.
10. GRAY, S. From principles to practice: teachers' uptake of principles from instructed language learning to plan a focus on language form in content lessons. *System*, Amsterdam, v. 37, n. 4, p. 570–584, 2009.
11. GREMMO, M.-J.; RILEY, P. Autonomy, self-direction and self access in language teaching and learning: The history of an idea. *System*, Amsterdam, v. 23, n. 2, p. 151–164, 1995.
12. HIRVELA, A. The case against CAI: A reply to John Higgins. *System*, Amsterdam, v. 17, n. 1, p. 61–65, 1989.
13. HYLAND, K. ESL computer writers: What can we do to help?. *System*, Amsterdam, v. 21, n. 1, p. 21–30, 1993.
14. HYLAND, K.; ANAN, E. Teachers' perceptions of error: The effects of first language and experience. *System*, Amsterdam, v. 34, n. 4, p. 509–519, 2006.
15. IMLAH, W. G.; DU BOULAY, J. B. H. Robust natural language parsing in computer-assisted language instruction. *System*, Amsterdam, v. 13, n. 2, p. 137–147, 1985.
16. JIANG, Y.; DEWAELE, J.-M. How unique is the foreign language classroom enjoyment and anxiety of Chinese EFL learners?. *System*, Amsterdam, v. 82, p. 13–25, 2019.
17. JOHNS, T. Micro-concord: A language learner's research tool. *System*, Amsterdam, v. 14, n. 2, p. 151–162, 1986.
18. JONES, C. It's not so much the program, more what you do with it: The importance of methodology in call. *System*, Amsterdam, v. 14, n. 2, p. 171–178, 1986.
19. KENNY, B. For more autonomy. *System*, Amsterdam, v. 21, n. 4, p. 431–442, 1993.
20. KUSEL, P. A. Rhetorical approaches to the study and composition of academic essays. *System*, Amsterdam, v. 20, n. 4, p. 457–469, 1992.
21. LAMB, M. Integrative motivation in a globalizing world. *System*, Amsterdam, v. 32, n. 1, p. 3–19, 2004.

22. LITTLEWOOD, W. "Autonomy": an anatomy and a framework. *System*, Amsterdam, v. 24, n. 4, p. 427–435, 1996.
23. LOMICKA, L.; LORD, G. A tale of tweets: analyzing microblogging among language learners. *System*, Amsterdam, v. 40, n. 1, p. 48–63, 2012.
24. MAR-MOLINERO, C.; WRIGHT, V. Languages and open learning in higher education. *System*, Amsterdam, v. 21, n. 2, p. 245–255, 1993.
25. MESKILL, C.; ANTHONY, N. Foreign language learning with CMC: forms of online instructional discourse in a hybrid Russian class. *System*, Amsterdam, v. 33, n. 1, p. 89–105, 2005.
26. MORRISON, D. M. GAPPER: A microcomputer-based learning game. *System*, Amsterdam, v. 12, n. 2, p. 169–180, 1984.
27. NATION, P. Improving speaking fluency. *System*, Amsterdam, v. 17, n. 3, p. 377–384, 1989.
28. OXFORD, R. L.; BURRY-STOCK, J. A. Assessing the use of language learning strategies worldwide with the ESL/EFL version of the Strategy Inventory for Language Learning (SILL). *System*, Amsterdam, v. 23, n. 1, p. 1–23, 1995.
29. PALTRIDGE, B. Writing up research: A systemic functional perspective. *System*, Amsterdam, v. 21, n. 2, p. 175–192, 1993.
30. PENG, J.-E. Towards an ecological understanding of willingness to communicate in EFL classrooms in China. *System*, Amsterdam, v. 40, n. 2, p. 203–213, 2012.
31. PENNYCOOK, A. Critical pedagogy and second language education. *System*, Amsterdam, v. 18, n. 3, p. 303–314, 1990.
32. RADWAN, A. A. The effectiveness of explicit attention to form in language learning. *System*, Amsterdam, v. 33, n. 1, p. 69–87, 2005.
33. SAVAGE, W.; STORER, G. An emergent language program framework: Actively involving learners in needs analysis. *System*, Amsterdam, v. 20, n. 2, p. 187–199, 1992.
34. SCHLEPPEGRELL, M. Using input methods to improve writing skills. *System*, Amsterdam, v. 12, n. 3, p. 287–292, 1984.
35. SHAO, K.; PEKRUN, R.; NICHOLSON, L. J. Emotions in classroom language learning: What can we learn from achievement emotion research?. *System*, Amsterdam, v. 86, p. 102121, 2019.
36. TONG-FREDERICKS, C. Types of oral communication activities and the language they generate: a comparison. *System*, Amsterdam, v. 12, n. 2, p. 133–145, 1984.
37. WARSCHAUER, M.; TURBEE, L.; ROBERTS, B. Computer learning networks and student empowerment. *System*, Amsterdam, v. 24, n. 1, p. 1–14, 1996.
38. WHITE, C. Expectations and emergent beliefs of self-instructed language learners. *System*, Amsterdam, v. 27, n. 4, p. 443–457, 1999.
39. WILLEMS, G. M. Communication strategies and their significance in foreign language teaching. *System*, Amsterdam, v. 15, n. 3, p. 351–364, 1987.
40. WONG, A. T. Y. Writers' mental representations of the intended audience and of the rhetorical purpose for writing and the strategies that they employed when they composed. *System*, Amsterdam, v. 33, n. 1, p. 29–47, 2005.
41. ZARRINABADI, N. Communicating in a second language: Investigating the effect of teacher on learners' willingness to communicate. *System*, Amsterdam, v. 42, p. 288–295, 2014.
42. ZHANG, L. J.; RAHIMI, M. EFL learners' anxiety level and their beliefs about corrective feedback in oral communication classes. *System*, Amsterdam, v. 42, p. 429–439, 2014.

TESOL Quarterly

1. AUERBACH, E. R. Reexamining English only in the ESL classroom. *TESOL Quarterly*, Hoboken, v. 27, n. 1, p. 9, 1993.
2. BLOCK, E. L. See how they read: comprehension monitoring of L1 and L2 readers. *TESOL Quarterly*, Hoboken, v. 26, n. 2, p. 319, 1992.
3. BLOCK, E. The comprehension strategies of second language readers. *TESOL Quarterly*, Hoboken, v. 20, n. 3, p. 463, 1986.
4. CANAGARAJAH, A. S. Teacher development in a global profession: an autoethnography. *TESOL Quarterly*, Hoboken, v. 46, n. 2, p. 258–279, 2012.
5. CHEN, J. F.; WARDEN, C. A.; CHANG, H.-T. Motivators that do not motivate: the case of Chinese EFL learners and the influence of culture on motivation. *TESOL Quarterly*, Hoboken, v. 39, n. 4, p. 609, 2005.
6. COHEN, A. D.; OLSHTAIN, E. The production of speech acts by EFL learners. *TESOL Quarterly*, Hoboken, v. 27, n. 1, p. 33, 1993.
7. COLLIER, V. P. Age and rate of acquisition of second language for academic purposes. *TESOL Quarterly*, Hoboken, v. 21, n. 4, p. 617, 1987.
8. COLLIER, V. P. How long? a synthesis of research on academic achievement in a second language. *TESOL Quarterly*, Hoboken, v. 23, n. 3, p. 509, 1989.
9. CURRY, M. J.; LILLIS, T. Multilingual scholars and the imperative to publish in English: negotiating interests, demands, and rewards. *TESOL Quarterly*, Hoboken, v. 38, n. 4, p. 663, 2004.
10. DAVISON, C.; LEUNG, C. Current issues in English language teacher-based assessment. *TESOL Quarterly*, Hoboken, v. 43, n. 3, p. 393–415, 2009.
11. DERWING, T. M.; MUNRO, M. J. Second language accent and pronunciation teaching: a research-based approach. *TESOL Quarterly*, Hoboken, v. 39, n. 3, p. 379, 2005.
12. ELLIS, R. Current issues in the teaching of grammar: an SLA perspective. *TESOL Quarterly*, Hoboken, v. 40, n. 1, p. 83, 2006.
13. ELLIS, R. The structural syllabus and second language acquisition. *TESOL Quarterly*, Hoboken, v. 27, n. 1, p. 91, 1993.
14. FIELD, J. Intelligibility and the listener: the role of lexical stress. *TESOL Quarterly*, Hoboken, v. 39, n. 3, p. 399, 2005.
15. GARCÍA, O. Emergent bilinguals and TESOL: what's in a name? *TESOL Quarterly*, Hoboken, v. 43, n. 2, p. 322–326, 2009.
16. GRAPIN, S. Multimodality in the new content standards era: implications for English learners. *TESOL Quarterly*, Hoboken, v. 53, n. 1, p. 30–55, 2019.
17. GREEN, J. M.; OXFORD, R. A closer look at learning strategies, L2 proficiency, and gender. *TESOL Quarterly*, Hoboken, v. 29, n. 2, p. 261, 1995.
18. HAFNER, C. A. Embedding digital literacies in English language teaching: students' digital video projects as multimodal ensembles. *TESOL Quarterly*, Hoboken, v. 48, n. 4, p. 655–685, 2014.
19. HINKEL, E. Current perspectives on teaching the four skills. *TESOL Quarterly*, Hoboken, v. 40, n. 1, p. 109, 2006.
20. HOMBURG, T. J. Holistic evaluation of ESL compositions: can it be validated objectively?. *TESOL Quarterly*, Hoboken, v. 18, n. 1, p. 87, 1984.
21. HOROWITZ, D. M. What professors actually require: academic tasks for the ESL classroom. *TESOL Quarterly*, Hoboken, v. 20, n. 3, p. 445, 1986.
22. IBRAHIM, A. E. K. M. Becoming black: rap and hip-hop, race, gender, identity, and the politics of ESL learning. *TESOL Quarterly*, Hoboken, v. 33, n. 3, p. 349, 1999.

23. JOHNSON, K. E.; GOLOMBEK, P. R. The transformative power of narrative in second language teacher education. *TESOL Quarterly*, Hoboken, v. 45, n. 3, p. 486–509, 2011.
24. KORMOS, J.; CSIZÉR, K. The interaction of motivation, self-regulatory strategies, and autonomous learning behavior in different learner groups. *TESOL Quarterly*, Hoboken, v. 48, n. 2, p. 275–299, 2014.
25. LONG, M. H.; CROOKES, G. Three approaches to task-based syllabus design. *TESOL Quarterly*, Hoboken, v. 26, n. 1, p. 27, 1992.
26. LONG, M. H.; PORTER, P. A. Group work, interlanguage talk, and second language acquisition. *TESOL Quarterly*, Hoboken, v. 19, n. 2, p. 207, 1985.
27. LU, X. A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, Hoboken, v. 45, n. 1, p. 36–62, 2011.
28. MORITA, N. Negotiating participation and identity in second language academic communities. *TESOL Quarterly*, Hoboken, v. 38, n. 4, p. 573, 2004.
29. PARK, G. "I am never afraid of being recognized as an NNES": one teacher's journey in claiming and embracing her nonnative-speaker identity. *TESOL Quarterly*, Hoboken, v. 46, n. 1, p. 127–151, 2012.
30. PEIRCE, B. N. Social identity, investment, and language learning. *TESOL Quarterly*, Hoboken, v. 29, n. 1, p. 9, 1995.
31. PENNYCOOK, A. Borrowing others' words: text, ownership, memory, and plagiarism. *TESOL Quarterly*, Hoboken, v. 30, n. 2, p. 201, 1996.
32. PENNYCOOK, A. Introduction: critical approaches to TESOL. *TESOL Quarterly*, Hoboken, v. 33, n. 3, p. 329, 1999.
33. PENNYCOOK, A. The concept of method, interested knowledge, and the politics of language teaching. *TESOL Quarterly*, Hoboken, v. 23, n. 4, p. 589, 1989.
34. PETERS, E. The effect of imagery and on-screen text on foreign language vocabulary learning from audiovisual input. *TESOL Quarterly*, Hoboken, v. 53, n. 4, p. 1008–1032, 2019.
35. PICA, T.; YOUNG, R.; DOUGHTY, C. The impact of interaction on comprehension. *TESOL Quarterly*, Hoboken, v. 21, n. 4, p. 737, 1987.
36. PORRECA, K. L. Sexism in current ESL textbooks. *TESOL Quarterly*, Hoboken, v. 18, n. 4, p. 705, 1984.
37. PRABHU, N. S. There is no best method-why?. *TESOL Quarterly*, Hoboken, v. 24, n. 2, p. 161, 1990.
38. RICENTO, T. K.; HORNBERGER, N. H. Unpeeling the onion: language planning and policy and the ELT professional. *TESOL Quarterly*, Hoboken, v. 30, n. 3, p. 401, 1996.
39. SAVILLE-TROIKE, M. What really matters in second language learning for academic achievement?. *TESOL Quarterly*, Hoboken, v. 18, n. 2, p. 199, 1984.
40. SHORT, D. J. Assessing integrated language and content instruction. *TESOL Quarterly*, Hoboken, v. 27, n. 4, p. 627, 1993.
41. VANN, R. J.; ABRAHAM, R. G. Strategies of unsuccessful language learners. *TESOL Quarterly*, Hoboken, v. 24, n. 2, p. 177, 1990.
42. ZAMEL, V. Responding to student writing. *TESOL Quarterly*, Hoboken, v. 19, n. 1, p. 79, 1985.

ANNEX A – Log-likelihood critical values and associated p-values

Table 41. Log-likelihood critical values and associated p-values (Rayson; Berridge; Francis, 2004)

Log-likelihood critical value	p-value	Percentage level of confidence in a significant result	Percentage probability that the result is due to chance
3.84	0.05	95	5
6.63	0.01	99	1
10.82	0.001	99.9	0.1
15.13	0.0001	99.99	0.01

Source: Adapted from McIntyre and Walker (2019, p. 158).

ANNEX B – List of all positive keywords generated by *AntConc*Table 42. List of all positive keywords generated by *AntConc*

Type	Freq_Tar	Freq_Ref	Range_Tar	NormFreq_Tar	NormFreq_Ref	Keyness (Likelihood)	Keyness (Effect)
features	1006	358	25	6.951	0.437	2.367.415	0.014
linguistic	1133	641	23	7.829	0.782	2.191.345	0.015
registers	606	33	16	4.187	0.040	2.051.729	0.008
dimension	601	62	15	4.153	0.076	1.890.125	0.008
texts	887	413	24	6.129	0.504	1.877.547	0.012
text	974	746	24	6.730	0.910	1.587.326	0.013
register	465	38	21	3.213	0.046	1.508.106	0.006
corpus	688	312	22	4.754	0.381	1.472.216	0.009
verbs	479	114	20	3.310	0.139	1.274.796	0.007
dimensions	447	90	16	3.089	0.110	1.240.449	0.006
variation	493	156	21	3.407	0.190	1.206.240	0.007
clauses	437	111	22	3.020	0.135	1.142.486	0.006
biber	315	23	23	2.177	0.028	1.034.906	0.004
patterns	447	189	24	3.089	0.231	984.122	0.006
spoken	377	134	23	2.605	0.163	886.343	0.005
analysis	685	722	25	4.733	0.881	885.484	0.009
stance	276	28	13	1.907	0.034	869.595	0.004
factor	454	275	18	3.137	0.336	846.345	0.006
types	460	304	20	3.179	0.371	817.693	0.006
are	2050	5299	25	14.165	6.465	806.256	0.027
written	477	370	22	3.296	0.451	770.081	0.007
informational	195	5	18	1.347	0.006	694.775	0.003
prose	220	46	21	1.520	0.056	604.756	0.003
genres	225	56	12	1.555	0.068	591.301	0.003
co	230	69	19	1.589	0.084	572.131	0.003

multi	202	42	14	1.396	0.051	556.010	0.003
analyses	276	145	23	1.907	0.177	552.286	0.004
these	1062	2419	25	7.338	2.951	535.115	0.014
dimensional	157	8	14	1.085	0.010	534.259	0.002
cluster	164	19	11	1.133	0.023	506.402	0.002
fiction	153	18	20	1.057	0.022	471.276	0.002
versus	169	36	17	1.168	0.044	462.412	0.002
corpora	178	47	22	1.230	0.057	460.042	0.002
clause	211	94	16	1.458	0.115	454.407	0.003
grammatical	359	400	22	2.481	0.488	442.378	0.005
type	313	314	24	2.163	0.383	420.597	0.004
textual	193	93	17	1.334	0.113	401.765	0.003
occur	249	198	25	1.721	0.242	395.345	0.003
conversation	247	203	21	1.707	0.248	383.696	0.003
distribution	147	42	22	1.016	0.051	371.157	0.002
common	264	255	22	1.824	0.311	365.305	0.004
nouns	146	48	18	1.009	0.059	352.460	0.002
complement	128	29	13	0.884	0.035	344.821	0.002
among	341	469	24	2.356	0.572	343.837	0.005
across	256	257	21	1.769	0.314	343.752	0.004
md	94	2	4	0.650	0.002	337.828	0.001
sample	227	205	17	1.569	0.250	330.198	0.003
represent	187	129	24	1.292	0.157	324.091	0.003
respect	179	120	19	1.237	0.146	315.385	0.002
parameters	125	36	15	0.864	0.044	314.830	0.002
each	520	1065	24	3.593	1.299	313.501	0.007
underlying	174	115	21	1.202	0.140	309.060	0.002
bundles	84	1	4	0.580	0.001	308.129	0.001
characteristics	195	157	18	1.347	0.192	307.046	0.003

constructions	134	56	20	0.926	0.068	296.193	0.002
shows	218	218	24	1.506	0.266	293.596	0.003
associated	194	169	22	1.341	0.206	289.513	0.003
samples	175	131	14	1.209	0.160	288.705	0.002
verb	176	138	20	1.216	0.168	281.945	0.002
functional	148	94	19	1.023	0.115	268.745	0.002
studies	415	812	22	2.868	0.991	268.522	0.006
tense	142	84	18	0.981	0.102	267.809	0.002
discourse	363	652	23	2.508	0.795	265.705	0.005
loadings	82	7	10	0.567	0.009	264.331	0.001
occurrence	116	45	17	0.802	0.055	263.944	0.002
present	305	482	25	2.108	0.588	263.122	0.004
lexical	241	308	20	1.665	0.376	261.671	0.003
somali	78	5	9	0.539	0.006	259.750	0.001
speeches	68	0	13	0.470	0.000	257.970	0.001
differences	295	470	22	2.038	0.573	252.004	0.004
range	260	372	24	1.797	0.454	251.267	0.004
relatively	180	172	22	1.244	0.210	251.064	0.002
adjectives	96	25	17	0.663	0.031	249.052	0.001
eg	352	656	24	2.432	0.800	244.640	0.005
statistical	128	74	18	0.884	0.090	244.264	0.002
relative	163	142	25	1.126	0.173	243.223	0.002
prepositional	79	12	17	0.546	0.015	232.641	0.001
occurring	110	53	21	0.760	0.065	228.950	0.002
pronouns	156	140	17	1.078	0.171	227.904	0.002
letters	94	35	14	0.650	0.043	217.178	0.001
clusters	71	9	8	0.491	0.011	216.004	0.001
contrast	178	202	24	1.230	0.246	215.740	0.002
style	136	114	15	0.940	0.139	208.404	0.002

functions	164	177	23	1.133	0.216	207.533	0.002
score	137	118	15	0.947	0.144	206.045	0.002
markedly	70	11	10	0.484	0.013	204.776	0.001
frequent	135	116	21	0.933	0.142	203.374	0.002
reportage	67	9	16	0.463	0.011	201.809	0.001
situational	100	52	13	0.691	0.063	200.991	0.001
scores	200	279	17	1.382	0.340	198.577	0.003
association	84	30	9	0.580	0.037	197.024	0.001
noun	101	56	18	0.698	0.068	196.814	0.001
elaboration	86	33	15	0.594	0.040	196.478	0.001
narrative	181	232	14	1.251	0.283	195.922	0.002
overall	175	217	23	1.209	0.265	195.585	0.002
passives	64	8	14	0.442	0.010	195.163	0.001
major	171	213	22	1.182	0.260	190.291	0.002
sequences	125	106	11	0.864	0.129	190.029	0.002
speech	253	457	23	1.748	0.558	183.674	0.003
distributions	59	7	19	0.408	0.009	181.457	0.001
contractions	52	2	15	0.359	0.002	180.808	0.001
exposition	56	5	9	0.387	0.006	179.475	0.001
collocations	62	11	6	0.428	0.013	176.893	0.001
associations	83	41	12	0.574	0.050	170.820	0.001
finegan	45	0	14	0.311	0.000	170.710	0.001
modals	61	12	8	0.422	0.015	170.070	0.001
weights	49	2	5	0.339	0.002	169.660	0.001
wh	50	3	20	0.345	0.004	167.597	0.001
pattern	128	133	21	0.884	0.162	167.130	0.002
multidimensional	52	5	9	0.359	0.006	165.008	0.001
passive	98	71	15	0.677	0.087	164.923	0.001
modal	67	21	7	0.463	0.026	164.290	0.001

marking	73	31	19	0.504	0.038	160.299	0.001
lob	44	1	9	0.304	0.001	157.650	0.001
nominalizations	45	2	11	0.311	0.002	154.818	0.001
word	345	839	22	2.384	1.024	152.938	0.005
quantitative	96	76	17	0.663	0.093	152.799	0.001
controlling	66	25	5	0.456	0.031	151.511	0.001
adverbs	70	31	12	0.484	0.038	151.078	0.001
section	144	192	24	0.995	0.234	149.829	0.002
literate	54	11	8	0.373	0.013	149.324	0.001
british	88	64	13	0.608	0.078	147.745	0.001
sampling	66	27	5	0.456	0.033	147.105	0.001
lexico	48	6	7	0.332	0.007	146.368	0.001
expository	54	13	12	0.373	0.016	143.151	0.001
abstract	88	68	9	0.608	0.083	142.266	0.001
million	72	39	15	0.498	0.048	141.906	0.001
extremely	83	59	16	0.574	0.072	141.283	0.001
marked	87	68	19	0.601	0.083	139.622	0.001
genre	95	85	8	0.656	0.104	139.065	0.001
vs	68	35	9	0.470	0.043	137.321	0.001
variationist	36	0	4	0.249	0.000	136.566	0.000
lund	36	0	13	0.249	0.000	136.566	0.000
proportional	40	2	3	0.276	0.002	136.310	0.001
collocational	45	6	7	0.311	0.007	135.715	0.001
restricted	72	43	22	0.498	0.052	135.093	0.001
identify	135	188	21	0.933	0.229	134.256	0.002
identifies	63	29	17	0.435	0.035	133.756	0.001
groupings	52	14	14	0.359	0.017	133.606	0.001
phrases	99	100	22	0.684	0.122	132.227	0.001
based	334	861	23	2.308	1.050	131.130	0.005

rare	59	25	13	0.408	0.031	129.667	0.001
academic	309	779	21	2.135	0.950	127.217	0.004
several	166	290	25	1.147	0.354	126.070	0.002
along	101	112	17	0.698	0.137	124.871	0.001
person	141	218	17	0.974	0.266	124.774	0.002
identified	143	226	22	0.988	0.276	123.286	0.002
grouped	57	26	15	0.394	0.032	121.490	0.001
frequency	192	382	21	1.327	0.466	120.991	0.003
chafe	34	1	6	0.235	0.001	120.222	0.000
broadcasts	34	1	9	0.235	0.001	120.222	0.000
computational	47	13	12	0.325	0.016	119.805	0.001
computed	52	20	15	0.359	0.024	118.687	0.001
function	151	259	23	1.043	0.316	117.486	0.002
frequencies	79	70	17	0.546	0.085	116.446	0.001
figure	134	211	18	0.926	0.257	116.033	0.002
similar	181	357	23	1.251	0.436	115.588	0.002
past	126	189	23	0.871	0.231	115.471	0.002
different	415	1237	24	2.868	1.509	114.383	0.006
techniques	101	124	21	0.698	0.151	113.923	0.001
attributive	30	0	9	0.207	0.000	113.804	0.000
illustrates	56	29	19	0.387	0.035	112.762	0.001
likelihood	50	20	10	0.345	0.024	112.423	0.001
sub	65	45	12	0.449	0.055	112.381	0.001
distinctions	61	39	15	0.422	0.048	110.341	0.001
newspaper	50	21	11	0.345	0.026	110.278	0.001
systematic	70	57	19	0.484	0.070	109.361	0.001
interpreted	68	53	17	0.470	0.065	109.322	0.001
writing	314	856	20	2.170	1.044	108.593	0.004
analyze	62	42	15	0.428	0.051	108.555	0.001

american	84	89	10	0.580	0.109	107.922	0.001
used	418	1278	25	2.888	1.559	107.280	0.006
macroscopic	32	2	4	0.221	0.002	106.828	0.000
representing	59	38	15	0.408	0.046	106.290	0.001
investigations	56	33	15	0.387	0.040	105.802	0.001
complementary	45	16	18	0.311	0.020	105.707	0.001
news	47	19	11	0.325	0.023	105.242	0.001
slots	34	4	2	0.235	0.005	104.705	0.000
example	367	1086	25	2.536	1.325	103.156	0.005
extent	133	229	23	0.919	0.279	102.956	0.002
semi	41	12	6	0.283	0.015	102.736	0.001
differing	44	16	14	0.304	0.020	102.528	0.001
show	152	290	24	1.050	0.354	102.021	0.002
conrad	32	3	9	0.221	0.004	101.891	0.000
colloquial	29	1	12	0.200	0.001	101.567	0.000
most	379	1142	24	2.619	1.393	101.354	0.005
dialects	38	9	7	0.263	0.011	101.172	0.001
representative	53	31	15	0.366	0.038	100.520	0.001
dense	30	2	8	0.207	0.002	99.491	0.000
nominal	49	25	16	0.339	0.031	99.353	0.001
situated	59	43	11	0.408	0.052	98.921	0.001
press	75	77	16	0.518	0.094	98.869	0.001
consideration	63	51	20	0.435	0.062	98.807	0.001
approach	241	610	23	1.665	0.744	98.360	0.003
johansson	28	1	13	0.193	0.001	97.842	0.000
framing	37	9	4	0.256	0.011	97.809	0.001
thus	264	701	25	1.824	0.855	97.022	0.004
counts	64	55	16	0.442	0.067	96.386	0.001
feature	119	198	19	0.822	0.242	96.274	0.002

large	151	299	23	1.043	0.365	95.840	0.002
represents	77	85	19	0.532	0.104	95.563	0.001
prepositions	41	15	16	0.283	0.018	95.326	0.001
particles	25	0	3	0.173	0.000	94.836	0.000
characterization	40	14	12	0.276	0.017	94.486	0.001
characterizations	27	1	9	0.187	0.001	94.120	0.000
factors	169	369	17	1.168	0.450	91.467	0.002
driven	52	35	2	0.359	0.043	91.381	0.001
stylistic	40	16	10	0.276	0.020	89.936	0.001
population	69	72	4	0.477	0.088	89.766	0.001
linguistically	57	46	10	0.394	0.056	89.579	0.001
relations	104	165	17	0.719	0.201	89.234	0.001
elaborated	41	18	11	0.283	0.022	88.804	0.001
deviation	44	23	5	0.304	0.028	88.206	0.001
kinds	99	153	18	0.684	0.187	87.637	0.001
personal	130	248	19	0.898	0.303	87.261	0.002
complementizer	23	0	6	0.159	0.000	87.249	0.000
lgswe	23	0	5	0.159	0.000	87.249	0.000
documents	43	22	13	0.297	0.027	87.073	0.001
slot	29	4	3	0.200	0.005	86.935	0.000
single	119	214	23	0.822	0.261	86.880	0.002
table	205	514	22	1.417	0.627	85.331	0.003
devices	67	72	10	0.463	0.088	85.067	0.001
specialized	42	22	10	0.290	0.027	84.114	0.001
subordination	60	57	12	0.415	0.070	84.029	0.001
two	412	1365	25	2.847	1.665	82.407	0.006
communicative	166	382	18	1.147	0.466	81.774	0.002
adequate	59	57	16	0.408	0.070	81.579	0.001
tagger	25	2	5	0.173	0.002	81.227	0.000

aspect	89	134	16	0.615	0.163	81.205	0.001
loading	32	9	9	0.221	0.011	81.162	0.000
nearly	45	29	19	0.311	0.035	81.039	0.001
shared	84	121	14	0.580	0.148	80.533	0.001
tagging	29	6	5	0.200	0.007	79.891	0.000
plots	23	1	11	0.159	0.001	79.260	0.000
distributional	23	1	9	0.159	0.001	79.260	0.000
perfect	45	31	13	0.311	0.038	78.024	0.001
representativeness	26	4	6	0.180	0.005	76.369	0.000
salient	55	54	15	0.380	0.066	75.107	0.001
adverbial	39	23	13	0.269	0.028	73.651	0.001
categories	134	290	18	0.926	0.354	73.628	0.002
participial	21	1	8	0.145	0.001	71.851	0.000
modifiers	22	2	9	0.152	0.002	70.337	0.000
commonly	66	87	19	0.456	0.106	69.453	0.001
adjective	39	26	11	0.269	0.032	68.909	0.001
domain	61	75	13	0.422	0.092	68.702	0.001
fixed	54	58	4	0.373	0.071	68.588	0.001
fictional	20	1	7	0.138	0.001	68.152	0.000
edited	28	9	14	0.193	0.011	68.088	0.000
in	4431	21905	25	30.618	26.726	68.001	0.052
london	37	23	13	0.256	0.028	67.957	0.001
quite	88	153	22	0.608	0.187	67.240	0.001
referential	26	7	7	0.180	0.009	66.800	0.000
negative	98	185	13	0.677	0.226	66.792	0.001
design	86	148	9	0.594	0.181	66.606	0.001
high	181	481	21	1.251	0.587	66.371	0.002
similarities	42	34	11	0.290	0.041	65.868	0.001
continuous	42	34	8	0.290	0.041	65.868	0.001

methodological	51	54	15	0.352	0.066	65.553	0.001
conditional	32	16	9	0.221	0.020	65.488	0.000
standard	85	148	13	0.587	0.181	64.822	0.001
typical	63	85	16	0.435	0.104	64.744	0.001
official	42	35	10	0.290	0.043	64.600	0.001
interactiveness	17	0	12	0.117	0.000	64.487	0.000
clefts	17	0	8	0.117	0.000	64.487	0.000
emphatics	17	0	9	0.117	0.000	64.487	0.000
nonabstract	17	0	2	0.117	0.000	64.487	0.000
editorials	19	1	7	0.131	0.001	64.459	0.000
typology	26	8	2	0.180	0.010	64.130	0.000
pronoun	49	51	14	0.339	0.062	63.876	0.001
occurrences	43	38	18	0.297	0.046	63.496	0.001
together	94	179	21	0.650	0.218	63.255	0.001
illustrated	45	43	15	0.311	0.052	62.741	0.001
styles	51	57	17	0.352	0.070	62.618	0.001
comparison	80	137	21	0.553	0.167	62.358	0.001
rates	29	13	5	0.200	0.016	62.265	0.000
reference	89	165	17	0.615	0.201	62.256	0.001
event	51	58	8	0.352	0.071	61.673	0.001
case	198	565	24	1.368	0.689	61.179	0.003
longman	25	8	11	0.173	0.010	60.882	0.000
illustrate	60	82	16	0.415	0.100	60.844	0.001
nonnarrative	16	0	4	0.111	0.000	60.694	0.000
varieties	38	30	12	0.263	0.037	60.583	0.001
primary	67	102	19	0.463	0.124	60.353	0.001
interpretation	82	147	14	0.567	0.179	60.117	0.001
etc	74	123	18	0.511	0.150	59.936	0.001
analyzed	58	78	16	0.401	0.095	59.804	0.001

combinations	34	23	10	0.235	0.028	59.574	0.000
distinctive	30	16	6	0.207	0.020	59.566	0.000
functionally	19	2	12	0.131	0.002	59.516	0.000
multivariate	26	10	11	0.180	0.012	59.340	0.000
moderately	26	10	10	0.180	0.012	59.340	0.000
formulaic	64	95	2	0.442	0.116	59.340	0.001
conversations	46	49	10	0.318	0.060	58.831	0.001
phrasal	32	20	8	0.221	0.024	58.601	0.000
reflecting	48	54	15	0.332	0.066	58.598	0.001
detailed	60	86	20	0.415	0.105	57.832	0.001
reduced	52	65	14	0.359	0.079	57.650	0.001
plot	23	7	10	0.159	0.009	56.928	0.000
postnominal	15	0	6	0.104	0.000	56.901	0.000
fa	15	0	2	0.104	0.000	56.901	0.000
presents	48	56	15	0.332	0.068	56.739	0.001
adverbials	59	85	9	0.408	0.104	56.551	0.001
collocation	18	2	5	0.124	0.002	55.928	0.000
pronominal	18	2	6	0.124	0.002	55.928	0.000
statistics	37	32	11	0.256	0.039	55.472	0.001
voice	49	61	7	0.339	0.074	54.536	0.001
interactive	69	118	14	0.477	0.144	53.878	0.001
variable	63	101	9	0.435	0.123	53.368	0.001
defined	82	160	13	0.567	0.195	53.211	0.001
donnell	14	0	4	0.097	0.000	53.107	0.000
drama	24	10	3	0.166	0.012	53.099	0.000
complexity	141	371	13	0.974	0.453	52.866	0.002
particular	209	648	25	1.444	0.791	51.448	0.003
designs	22	8	4	0.152	0.010	51.261	0.000
while	222	705	23	1.534	0.860	50.841	0.003

paper	98	220	19	0.677	0.268	50.493	0.001
certainty	28	18	6	0.193	0.022	50.491	0.000
surprising	48	64	16	0.332	0.078	49.927	0.001
words	361	1321	24	2.494	1.612	49.719	0.005
extraposed	13	0	3	0.090	0.000	49.314	0.000
reppen	13	0	8	0.090	0.000	49.314	0.000
subordinators	13	0	7	0.090	0.000	49.314	0.000
use	491	1925	24	3.393	2.349	49.245	0.007
description	71	133	21	0.491	0.162	48.929	0.001
broadcast	16	2	4	0.111	0.002	48.786	0.000
grouping	31	25	9	0.214	0.031	48.738	0.000
automated	18	4	4	0.124	0.005	48.720	0.000
confirmatory	18	4	7	0.124	0.005	48.720	0.000
probabilistic	18	4	7	0.124	0.005	48.720	0.000
directive	17	3	3	0.117	0.004	48.555	0.000
differ	51	74	19	0.352	0.090	48.501	0.001
considerations	35	34	13	0.242	0.041	48.187	0.000
general	173	514	25	1.195	0.627	48.051	0.002
analytical	34	32	13	0.235	0.039	47.948	0.000
include	109	266	24	0.753	0.325	47.935	0.002
domains	45	60	10	0.311	0.073	46.806	0.001
normalized	19	6	7	0.131	0.007	46.471	0.000
ie	108	267	20	0.746	0.326	46.263	0.001
epistemic	27	19	3	0.187	0.023	46.230	0.000
frames	30	25	5	0.207	0.031	46.141	0.000
novels	14	1	6	0.097	0.001	46.084	0.000
previous	101	243	20	0.698	0.296	45.687	0.001
persuasive	24	14	3	0.166	0.017	45.578	0.000
blankenship	12	0	2	0.083	0.000	45.520	0.000

taggers	12	0	3	0.083	0.000	45.520	0.000
factorial	12	0	4	0.083	0.000	45.520	0.000
probabilities	15	2	4	0.104	0.002	45.236	0.000
correlates	15	2	9	0.104	0.002	45.236	0.000
conjuncts	15	2	6	0.104	0.002	45.236	0.000
usually	79	168	16	0.546	0.205	44.705	0.001
here	169	512	25	1.168	0.625	44.465	0.002
dialect	23	13	8	0.159	0.016	44.384	0.000
variants	24	15	6	0.166	0.018	43.950	0.000
progressive	37	43	8	0.256	0.052	43.887	0.001
within	187	590	18	1.292	0.720	43.681	0.003
five	89	206	14	0.615	0.251	43.358	0.001
persuasion	18	6	6	0.124	0.007	43.240	0.000
summarizes	23	14	12	0.159	0.017	42.719	0.000
factive	17	5	2	0.117	0.006	42.531	0.000
francis	17	5	9	0.117	0.006	42.531	0.000
poole	13	1	4	0.090	0.001	42.434	0.000
collocates	19	8	4	0.131	0.010	41.860	0.000
strikingly	19	8	11	0.131	0.010	41.860	0.000
biographies	11	0	5	0.076	0.000	41.727	0.000
proportionally	11	0	2	0.076	0.000	41.727	0.000
characteristic	41	56	17	0.283	0.068	41.600	0.001
mark	49	79	17	0.339	0.096	41.221	0.001
addition	98	245	24	0.677	0.299	41.022	0.001
interpretations	39	52	11	0.269	0.063	40.565	0.001
both	299	1099	25	2.066	1.341	40.400	0.004
considered	100	255	21	0.691	0.311	40.149	0.001
purpose	90	219	22	0.622	0.267	39.805	0.001
simple	65	132	16	0.449	0.161	39.649	0.001

diversified	12	1	3	0.083	0.001	38.794	0.000
altenberg	12	1	7	0.083	0.001	38.794	0.000
lancaster	12	1	8	0.083	0.001	38.794	0.000
valency	15	4	2	0.104	0.005	38.644	0.000
literary	26	23	5	0.180	0.028	38.363	0.000
empirically	24	19	13	0.166	0.023	38.192	0.000
gray	13	2	3	0.090	0.002	38.184	0.000
tottie	10	0	8	0.069	0.000	37.933	0.000
tuvaluan	10	0	5	0.069	0.000	37.933	0.000
cooccurring	10	0	3	0.069	0.000	37.933	0.000
beaman	10	0	4	0.069	0.000	37.933	0.000
gorsuch	10	0	3	0.069	0.000	37.933	0.000
szmrecsanyi	10	0	2	0.069	0.000	37.933	0.000
represented	52	94	16	0.359	0.115	37.683	0.001
historical	47	79	9	0.325	0.096	37.526	0.001
basis	90	227	22	0.622	0.277	36.981	0.001
comprises	24	20	12	0.166	0.024	36.913	0.000
especially	100	266	18	0.691	0.325	36.576	0.001
variant	20	13	6	0.138	0.016	35.844	0.000
intuitions	17	8	5	0.117	0.010	35.745	0.000
interpretive	22	17	8	0.152	0.021	35.560	0.000
structural	76	180	15	0.525	0.220	35.443	0.001
diachronic	14	4	5	0.097	0.005	35.338	0.000
figures	41	65	11	0.283	0.079	35.202	0.001
tags	11	1	6	0.076	0.001	35.168	0.000
causative	11	1	5	0.076	0.001	35.168	0.000
poles	11	1	8	0.076	0.001	35.168	0.000
of	6207	32413	25	42.890	39.547	35.117	0.068
particle	13	3	3	0.090	0.004	34.847	0.000

microscopic	13	3	2	0.090	0.004	34.847	0.000
perspective	86	219	18	0.594	0.267	34.627	0.001
descriptions	40	63	17	0.276	0.077	34.615	0.001
unmarked	19	12	8	0.131	0.015	34.596	0.000
label	24	22	7	0.166	0.027	34.513	0.000
genitive	9	0	3	0.062	0.000	34.140	0.000
swal	9	0	2	0.062	0.000	34.140	0.000
intransitive	9	0	3	0.062	0.000	34.140	0.000
exploratory	26	27	5	0.180	0.033	33.954	0.000
mean	113	326	17	0.781	0.398	33.933	0.002
few	88	229	22	0.608	0.279	33.807	0.001
however	290	1103	25	2.004	1.346	33.657	0.004
involvement	37	56	12	0.256	0.068	33.551	0.001
shown	108	308	23	0.746	0.376	33.407	0.001
extreme	23	21	9	0.159	0.026	33.171	0.000
lesser	19	13	12	0.131	0.016	33.072	0.000
uncover	16	8	8	0.111	0.010	32.743	0.000
dependent	52	104	13	0.359	0.127	32.486	0.001
token	26	29	10	0.180	0.035	31.976	0.000
empirical	72	175	16	0.498	0.214	31.915	0.001
replicated	15	7	7	0.104	0.009	31.655	0.000
tagged	15	7	4	0.104	0.009	31.655	0.000
opposite	25	27	11	0.173	0.033	31.605	0.000
important	228	833	24	1.575	1.016	31.588	0.003
romance	10	1	4	0.069	0.001	31.557	0.000
bundle	10	1	2	0.069	0.001	31.557	0.000
size	52	106	10	0.359	0.129	31.527	0.001
preference	29	37	9	0.200	0.045	31.517	0.000
compute	12	3	7	0.083	0.004	31.484	0.000

tag	12	3	4	0.083	0.004	31.484	0.000
summing	11	2	10	0.076	0.002	31.215	0.000
algorithms	11	2	3	0.076	0.002	31.215	0.000
controlled	35	54	9	0.242	0.066	31.037	0.000
numerous	35	54	17	0.242	0.066	31.037	0.000
prominence	17	11	3	0.117	0.013	30.544	0.000
distinct	42	76	13	0.290	0.093	30.392	0.001
besnier	8	0	7	0.055	0.000	30.347	0.000
nukulaelae	8	0	4	0.055	0.000	30.347	0.000
akinnaso	8	0	2	0.055	0.000	30.347	0.000
svartvik	8	0	8	0.055	0.000	30.347	0.000
lexicographic	8	0	4	0.055	0.000	30.347	0.000
predicative	8	0	4	0.055	0.000	30.347	0.000
parameter	20	17	7	0.138	0.021	30.346	0.000
fuller	19	15	12	0.131	0.018	30.290	0.000
highly	74	187	18	0.511	0.228	30.283	0.001
summarized	24	26	10	0.166	0.032	30.262	0.000
describing	38	64	15	0.263	0.078	30.261	0.001
larger	49	99	16	0.339	0.121	30.129	0.001
above	108	321	20	0.746	0.392	29.957	0.001
matrix	23	24	6	0.159	0.029	29.918	0.000
positive	97	277	11	0.670	0.338	29.901	0.001
oblique	13	5	2	0.090	0.006	29.669	0.000
resulting	37	62	18	0.256	0.076	29.657	0.001
cross	45	87	11	0.311	0.106	29.604	0.001
reflect	63	149	18	0.435	0.182	29.459	0.001
opposition	21	20	5	0.145	0.024	29.351	0.000
reliably	16	10	4	0.111	0.012	29.299	0.000
frequently	74	190	19	0.511	0.232	29.271	0.001

forms	155	523	21	1.071	0.638	29.085	0.002
referents	17	12	6	0.117	0.015	29.053	0.000
concordancing	17	12	2	0.117	0.015	29.053	0.000
retention	17	12	2	0.117	0.015	29.053	0.000
distributed	30	43	13	0.207	0.052	28.913	0.000
addressees	12	4	2	0.083	0.005	28.826	0.000
characterized	32	49	10	0.221	0.060	28.628	0.000
wide	51	109	20	0.352	0.133	28.615	0.001
temporal	22	23	7	0.152	0.028	28.573	0.000
subordinate	22	23	8	0.152	0.028	28.573	0.000
four	108	327	22	0.746	0.399	28.455	0.001
purposes	77	204	19	0.532	0.249	28.418	0.001
sets	43	83	18	0.297	0.101	28.356	0.001
extents	11	3	6	0.076	0.004	28.154	0.000
infinitives	11	3	6	0.076	0.004	28.154	0.000
synchronic	11	3	5	0.076	0.004	28.154	0.000
descriptive	36	62	13	0.249	0.076	27.849	0.000
computerized	10	2	5	0.069	0.002	27.770	0.000
participles	10	2	3	0.069	0.002	27.770	0.000
weight	18	15	3	0.124	0.018	27.684	0.000
specificity	17	13	9	0.117	0.016	27.661	0.000
sinclair	22	24	9	0.152	0.029	27.576	0.000
maximally	13	6	4	0.090	0.007	27.566	0.000
defining	30	45	13	0.207	0.055	27.484	0.000
stable	24	29	7	0.166	0.035	27.471	0.000
uses	68	173	23	0.470	0.211	27.432	0.001
plus	21	22	11	0.145	0.027	27.228	0.000
markers	48	102	17	0.332	0.124	27.193	0.001
taken	79	216	18	0.546	0.264	27.101	0.001

expression	54	123	15	0.373	0.150	27.094	0.001
marks	23	27	12	0.159	0.033	27.034	0.000
restrictions	14	8	7	0.097	0.010	26.867	0.000
sports	14	8	6	0.097	0.010	26.867	0.000
coefficients	19	18	3	0.131	0.022	26.662	0.000
clausal	19	18	4	0.131	0.022	26.662	0.000
focus	160	558	23	1.106	0.681	26.643	0.002
relating	34	58	11	0.235	0.071	26.632	0.000
hared	7	0	4	0.048	0.000	26.553	0.000
longacre	7	0	2	0.048	0.000	26.553	0.000
lswe	7	0	4	0.048	0.000	26.553	0.000
mair	7	0	4	0.048	0.000	26.553	0.000
microanalyses	7	0	3	0.048	0.000	26.553	0.000
noninformational	7	0	2	0.048	0.000	26.553	0.000
partington	7	0	3	0.048	0.000	26.553	0.000
sermons	7	0	6	0.048	0.000	26.553	0.000
ambiguities	17	14	4	0.117	0.017	26.356	0.000
compares	16	12	10	0.111	0.015	26.353	0.000
interactional	38	71	7	0.263	0.087	26.281	0.001
reliance	21	23	11	0.145	0.028	26.234	0.000
entire	34	59	13	0.235	0.072	26.040	0.000
circumstances	32	53	12	0.221	0.065	26.031	0.000
required	79	220	19	0.546	0.268	25.928	0.001
absolute	20	21	13	0.138	0.026	25.883	0.000
line	60	148	14	0.415	0.181	25.815	0.001
than	396	1675	25	2.736	2.044	25.726	0.005
distinguishing	25	34	11	0.173	0.041	25.478	0.000
statistically	25	34	12	0.173	0.041	25.478	0.000
rather	192	709	25	1.327	0.865	25.417	0.003

stability	14	9	2	0.097	0.011	25.245	0.000
infrequent	14	9	5	0.097	0.011	25.245	0.000
primarily	49	110	17	0.339	0.134	25.241	0.001
broad	38	73	12	0.263	0.089	25.236	0.001
corresponding	24	32	10	0.166	0.039	24.962	0.000
omission	20	22	4	0.138	0.027	24.893	0.000
pairs	32	55	10	0.221	0.067	24.820	0.000
management	50	115	4	0.345	0.140	24.643	0.001
factual	12	6	6	0.083	0.007	24.557	0.000
comparisons	35	65	14	0.242	0.079	24.417	0.000
subordinator	8	1	5	0.055	0.001	24.393	0.000
with	1107	5316	25	7.649	6.486	24.213	0.015
finally	73	203	22	0.504	0.248	24.042	0.001
structurally	13	8	8	0.090	0.010	24.005	0.000
quantitatively	13	8	7	0.090	0.010	24.005	0.000
promise	17	16	3	0.117	0.020	23.973	0.000
distinguished	26	39	15	0.180	0.048	23.819	0.000
marker	14	10	3	0.097	0.012	23.758	0.000
foundation	15	12	9	0.104	0.015	23.707	0.000
conclusions	31	54	9	0.214	0.066	23.622	0.000
consistently	31	54	12	0.214	0.066	23.622	0.000
comprehensive	30	51	10	0.207	0.062	23.604	0.000
characterize	19	21	11	0.131	0.026	23.551	0.000
underpinnings	11	5	8	0.076	0.006	23.478	0.000
carried	42	90	14	0.290	0.110	23.461	0.001
further	121	405	22	0.836	0.494	23.343	0.002
considerably	27	43	13	0.187	0.052	23.054	0.000
involved	81	239	15	0.560	0.292	22.914	0.001
central	59	153	14	0.408	0.187	22.837	0.001

cooccur	6	0	2	0.041	0.000	22.760	0.000
haan	6	0	3	0.041	0.000	22.760	0.000
discontinuous	6	0	3	0.041	0.000	22.760	0.000
mahlberg	6	0	2	0.041	0.000	22.760	0.000
bergen	6	0	5	0.041	0.000	22.760	0.000
oslo	6	0	5	0.041	0.000	22.760	0.000
sections	33	62	18	0.228	0.076	22.642	0.000
hymes	16	15	6	0.111	0.018	22.629	0.000
brown	44	99	14	0.304	0.121	22.573	0.001
physical	31	56	11	0.214	0.068	22.485	0.000
intervening	13	9	5	0.090	0.011	22.473	0.000
generalized	17	18	7	0.117	0.022	21.849	0.000
labeled	17	18	8	0.117	0.022	21.849	0.000
including	99	319	22	0.684	0.389	21.650	0.001
romantic	9	3	3	0.062	0.004	21.620	0.000
statistic	11	6	3	0.076	0.007	21.603	0.000
halliday	24	37	8	0.166	0.045	21.301	0.000
preferred	29	52	9	0.200	0.063	21.250	0.000
attitudinal	21	29	10	0.145	0.035	21.062	0.000
sampled	8	2	3	0.055	0.002	20.989	0.000
renouf	8	2	3	0.055	0.002	20.989	0.000
emphases	7	1	5	0.048	0.001	20.850	0.000
disregarded	7	1	3	0.048	0.001	20.850	0.000
values	51	130	11	0.352	0.159	20.488	0.001
formality	10	5	8	0.069	0.006	20.464	0.000
construction	48	119	11	0.332	0.145	20.438	0.001
considering	29	54	16	0.200	0.066	20.154	0.000
investigate	40	91	13	0.276	0.111	20.114	0.001
medical	19	25	2	0.131	0.031	20.028	0.000

concerning	43	102	16	0.297	0.124	19.991	0.001
uncovered	11	7	5	0.076	0.009	19.946	0.000
preposition	11	7	5	0.076	0.009	19.946	0.000
combines	12	9	5	0.083	0.011	19.765	0.000
bases	12	9	6	0.083	0.011	19.765	0.000
describe	52	137	18	0.359	0.167	19.432	0.001
have	629	2924	25	4.346	3.568	19.410	0.008
disregard	9	4	5	0.062	0.005	19.392	0.000
overtly	15	16	7	0.104	0.020	19.161	0.000
full	52	138	16	0.359	0.168	19.116	0.001
expressions	39	90	6	0.269	0.110	19.102	0.001
cline	5	0	5	0.035	0.000	18.967	0.000
collocate	5	0	3	0.035	0.000	18.967	0.000
subregisters	5	0	3	0.035	0.000	18.967	0.000
sociolinguists	5	0	2	0.035	0.000	18.967	0.000
nonfictional	5	0	3	0.035	0.000	18.967	0.000
lore	5	0	2	0.035	0.000	18.967	0.000
subcorpora	5	0	2	0.035	0.000	18.967	0.000
comrie	5	0	2	0.035	0.000	18.967	0.000
kucera	5	0	5	0.035	0.000	18.967	0.000
kay	5	0	3	0.035	0.000	18.967	0.000
geisler	5	0	3	0.035	0.000	18.967	0.000
uncorrelated	5	0	2	0.035	0.000	18.967	0.000
reflects	27	50	10	0.187	0.061	18.912	0.000
analyzing	24	41	10	0.166	0.050	18.763	0.000
humanities	18	24	7	0.124	0.029	18.721	0.000
set	109	377	21	0.753	0.460	18.713	0.002