| _ | | | | _ | | _ | | _ | | | _ | | _ | | _ | _ | - |
|--|---|---|---|--|--|---------------------------------|---|--|-------------------------------------|--|---|--|---|--|---|--|---|
|), which prepares students at tertiary level for further a |) Program (Level 10 with 80%) and then enter the univ |) which means that I must be knowledgeable in all aca |) author Aylin Graves provides a set of lesson plans to |) coursework taken at Florida SouthWestern State Col | program).⊲/s> <s>He is also academic director of UGA</s> |) B1+ INTERMEDIATE - components |) program.⊲/s> <s>l teach level 2 writing every mornin(</s> |).>>Classes consist of International students fror |) series from author, Aylin Graves. |) learners face, and what teaching staff and lecturers r |) courses.>She has spent many hours in the cl |), translation, interpreting, education quality assessme |) courses in reading, listening, writing, and research be |) researchers and practitioners to provide support for r |) ELT (Enhanced Language Training) ESP (English for |) researchers, such as Christison and Krahnke, 1986; | |

ENGLISH FOR ACADEMIC PURPOSES REFLECTIONS, DESCRIPTION & PEDAGOGY SIMONE SARMENTO, ROZANE REBECHI, Marine Laísa Matte (org.)

| survey. |
|--|
| ernational students at colleges and universities EAP (|
| onventions. |
| is required. |
| interests include second language acquisition, EAP (|
| es, and adjunct professor for E. |
| Edward de Chazal explains the challenges that EAP (|
| s article provides a guide to the award-winning EAP (|
| hing English on the BU campus in the EAP program (|
| rted my second year of teaching at BU with the EAP (|
| y that Continue reading → |
| or at Emory University's Candler School of Theology (|
| not apply; however, six credits of college-level EAP (|
| ents is: |
| / students. |
| on, Canada so that a student can complete our EAP (|
| e for learning English. |

ENGLISH FOR ACADEMIC PURPOSES Reflections, description & pedagogy

SIMONE SARMENTO Rozane Rebechi Marine Laísa Matte (org.)

Porto Alegre • 2024 • 1ª edição



Conselho Editorial

Cristiane Tavares – Instituto Vera Cruz/SP Daniela Mussi – UFRJ Idalice Ribeiro Silva Lima – UFTM Joanna Burigo – Emancipa Mulher Leonardo Antunes - UFRGS Lucia Tennina – UBA Luis Augusto Campos – UERJ Luis Felipe Miguel – UnB Maria Amelia Bulhões – UFRGS Regina Dalcastagnè – UnB Regina Zilberman – UFRGS Renato Ortiz - Unicamp Ricardo Timm de Souza – PUCRS Rodrigo Saballa de Carvalho – UFRGS Rosana Pinheiro Machado – University College Dublin Susana Rangel – UFRGS Winnie Bueno – Winnieteca

copyright © 2024 Simone Sarmento, Rozane Rebechi, Marine Laísa Matte

Projeto gráfico e edição: Editora Zouk Revisão: Simone Sarmento, Rozane Rebechi, Marine Laísa Matte Imagem da capa: SKELL







direitos reservados à Editora Zouk r. Cristóvão Colombo, 1343 sl. 203 90560-004 – Floresta – Porto Alegre – RS – Brasil f. 51. 3024.7554

www.editorazouk.com.br

From specialized corpus to the EAP classroom: integrating authentic data into materials design

Ana Eliza Pereira Bocorny (UFRGS) Ana Luiza Freitas (UFCSPA) Rozane Rodrigues Rebechi (UFRGS)

Introduction

Almost two decades ago, Sinclair (2004a) anticipated that corpus-based language teaching would revolutionize language pedagogy. After all, relying on empirical evidence enables the design of pedagogical applications based on authentic input, providing teachers and researchers with an actual perspective of how language works. Today, the positive impact of corpus-based approaches to additional language learning and teaching is undeniable (Boulton & Cobb, 2017; Boulton, 2021; Karlsen, 2021; Anthony, 2022a; O'Keeffe, 2022).

Despite the importance of corpus linguistics as a means of identifying authentic language use and the fact that many studies (Flowerdew, 2009, 2013, 2014; Gray et al., 2020; Charles & Frankenberg-Garcia, 2021) suggest integrating corpus data into English for Academic Purposes¹ (EAP) pedagogy, the use of authentic data in language classrooms around the world is still incipient (Kavanagh, 2021; Poole, 2020; Pérez-Paredes, 2019). Moreover, according to Römer (2006: 122), "there is still a strong resistance towards corpora from the side of students, teachers, and materials writers."

¹ The term English for Academic Purposes (EAP) refers to the English which is needed to study or conduct research in the academic context. Although it is often associated with non-native speakers of the language, EAP has extended also to native speakers who are faced with writing essays, presenting papers, reading articles, etc. (Charles, 2013).

Previous studies have suggested that "lack of time, group sizes, and technological obstacles" (Kavanagh, 2021: 2) could be standing in the way between corpus data and the language classroom. Poole (2020: 1) reports that although teachers embrace the use of corpus, they also reveal "emergent tensions regarding the use of ready-made corpus activities and the key affordances of discovery, authenticity, and autonomy often forwarded in support of corpus pedagogy." Breyer (2011: 207) claims that the lack of "(classroom) user-friendly concordancing software" was mentioned by teachers as one of the hurdles to the smooth adoption of corpora as language learning input. Other reasons identified by Mukherjee (2004: 243) had to do with the fact that not enough teachers were acquainted with "the basic foundations, implications, and applications of Corpus Linguistics."

Ranging from the context of graduate and undergraduate students from the Federal University of Rio Grande do Sul (UFRGS), this contribution arose from the needs of Brazilian pre-service and in-service EAP novice teachers when designing EAP writing course materials with corpus data at the Center of Languages for Academic Purposes (CLA)². After being introduced to corpus linguistics principles and methods, these novice teachers were asked to design a Pedagogical Unit (PU), i.e., a set of learning activities sequenced together to promote advances in learning, for a given EAP course where selected language features would be taught within the context of a given academic genre. Those teachers were then asked to extract and analyze said language data and integrate it into their EAP materials.

Having this said, the aim of this chapter is twofold: (i) help EAP teachers better understand corpus linguistics methods for the extraction of language data from specialized corpora and (ii) show how said language data can be used in the design of EAP writing course materials through a pedagogical model that combines corpus and genre-based approaches.

The first section – 'Combining corpus and genre-based approaches' - reviews the literature on corpus and genre-based approaches to language learning and teaching and on pedagogical models that combine

² CLA website: https://www.ufrgs.br/cla/

both approaches. Section 2 – 'The design of EAP materials' - describes the framework suggested in the study for designing EAP materials and presents a step-by-step guide on extracting and integrating corpus data into materials used for EAP writing courses. Finally, we finish the chapter with some final considerations and suggestions for further studies.

Combining corpus and genre-based approaches

Corpus Linguistics

According to Sinclair (1991: 171), "a corpus corresponds to a collection of natural texts chosen to characterize a state or variety of language". For Biber and Conrad (1999: 4), the notion of corpus is naturally approached from the perspective of register: "a collection of spoken or written texts, organized by the register and codified for other discursive considerations, comprises a corpus." McEnery and Hardie (2012: 1) define corpus linguistics as "an area which focuses upon a set of procedures, or methods, for studying language." As such, it can be applied to different areas.

Two central concepts are pillars of the field: the empiricist approach and the view of language as a probabilistic system. The empiricist system is based on the fact that knowledge originates from data organized in the form of a corpus. The view of language as a probabilistic system stems from the epistemological basis of the field, according to which linguistic traits do not happen randomly. Nevertheless, it is possible to point out and quantify patterns of regularity, highlighting a correlation between such traits and the situational contexts of use. From these patterns, it can be recognized that a language is not limited to empty spaces arbitrarily filled. Instead, the linguistic environment acts on the co-selection of lexical items. Within a linguistic environment, a given item prefers another one. This way, language is seen as a non-arbitrarily motivated and functional system of potential choices. These aspects refer to the issue of usage patterns and, therefore, to the idiomatic principle postulated by Sinclair (1991).

Let us take an example from the corpus used to extract linguistic data in this text. 'The aim of this study' is a sequence whose continuity is limited by a word within the verb category 'be' followed by the preposition 'to', confirming a preference of academic textual genres/records (Hyland, 2008; Biber & Conrad, 1999) for a greater incidence of this association of words. Thus, the phrase above is expected to precede 'is not' or 'was to'.

Although the literature proposes many definitions for what constitutes a corpus (such as Atkins et al., 1992; Francis, 1992; Kennedy, 1998; McEnery et al., 2006), the consensus is that it should comprise:

1. Authentic Linguistic Data;

2. Readable Computer Segments;

3. Specially Organized Language Portions;

4. Texts Capable of Representing a Particular Language or Variety of Language.

For this chapter, a corpus is roughly understood as a set of machine-readable texts compiled with the aim to provide answers to specific research questions (McEnery & Hardie, 2012). To achieve these goals, a corpus should be built under well-defined criteria.

Corpus-Based Pedagogy

Since John Sinclair's seminal work on corpus research led to the use of corpus-based approaches (Sinclair, 1987, 1991, 2004b), corpus linguistics has always been connected with language teaching. Contributions such as Gavioli (2005), O'Keeffe et al. (2007), Aijmer (2009), Flowerdew (2012), and Cotos (2014), among others, all followed the principles of adopting empirical data to boost language learning. Hence corpus-based pedagogy is the application of corpus linguistics's foundations to facilitate the teaching and learning of additional languages springing from authentic occurrences of language.

Among the advantages of adopting corpora for language teaching are the possibilities of explaining the differences in the uses of words and linguistic forms, among other traits, based on the probability of occurrence in specific contexts (Biber et al., 1998), as intuition alone could not explain these facts (Sinclair, 1991). As pointed out by Shepherd (2009: 152), the analytical enterprise "cannot depend on the researcher's intuitions, since human beings tend to recognize what is not typical more often than what is standardized". Corpora, therefore, are used to generate empirical knowledge about languages. Besides, using corpora for pedagogical purposes can disclose solutions to language queries that have not been dealt with otherwise. Furthermore, the use of corpora can highlight frequency patterns of words and language structures, and such patterns can be used to teach and create or improve teaching materials.

The most common tools used in corpus analysis for pedagogical purposes are concordancing programs, understood as text search engines with sorting functions, as will be demonstrated in the 'Step-by-step guide' to 'The design of EAP materials' below. Currently, among the most popular concordancing programs are WordSmith Tools (Scott, 2020), Sketch Engine (Kilgarriff et al., 2004), and AntConc 4.1 (Anthony, 2022b). As they are queried, these tools enable users to get in contact with "a collection of the occurrences of a word-form, each in its textual environment" (Sinclair, 1991: 32).

By using corpora for teaching purposes, users are empowered, as this approach holds the potential to foster autonomous and personalized learning (Boulton & Cobb, 2017; McEnery & Wilson, 1997). That happens because, on the one hand, the adoption of corpora encourages discoveries. Corpora can be employed, for example, to have students explore patterns of specific language features that stand out from the concordance lines. On the other hand, exploring language corpora by employing software enables learners within the same class to focus on different language features. Furthermore, corpus-based pedagogy can lead learners themselves to draw conclusions about language use and its principles.

Data Driven Learning (DDL)

As Boulton (2021: 9) affirms, "Data-driven learning (DDL) typically involves language learners consulting corpus data, either directly or via prepared materials, to answer questions about language." Some alleged benefits of using DDL are that it stimulates learners' autonomy and increases language awareness (Boulton, 2007). As for teachers, the use of DDL allows for a change of roles from a lecturer to "a co-ordinator of student-initiated research" (Johns, 1991: 3). Nevertheless, the change of roles mentioned by Johns (1991) does not come without challenges, such as learning how to compile and extract language data from a corpus or how to include the language data extracted into the materials designed for EAP courses in a meaningful and contextualized way. Besides, employing DDL implies choosing which approach to be used, whether direct DDL, through hands-on activities (where you teach your learners how to look for information in the corpus) or indirect DDL, an approach through which you (teacher) previously extract the language data yourself and include them into pedagogical units.

Corpus processing systems like Sketch Engine (Kilgarriff et al., 2004), WordSmith Tools (Scott, 2020), AntConc (Anthony, 2022b), and #LancsBox v6 (Brezina et al., 2020) can be of great help. They usually offer varied resources to extract language features, such as lists of words, keywords, and n-grams. In Sketch Engine (SE), it is also possible to use Corpus Query Language (CQL) to create special search syntaxes or queries to look for more complex grammatical and lexical patterns (see 'Description of the EAP writing course', Table 5, for examples of language features and ways to retrieve them from the corpus using CQL queries). The smart search option available in #LancsBox v6 (henceforth, LancsBox) software package is another option for extracting more complex language patterns. Pérez-Llantada (2022), for example, uses the LancsBox smart search option to retrieve passive voice forms from four corpora.

To cater to the challenges mentioned above, in this chapter we provide EAP teachers with a step-by-step guide on retrieving and integrating corpus data into materials designed for EAP writing courses through indirect DDL. At this moment, we chose to focus on indirect DDL because we considered its simplicity an asset to encourage novice EAP teachers in their pursuits of work with corpus-based pedagogy.

Genre, Genre-Analysis, Move-Analysis and Genre-Based Pedagogy

Bhatia (1993: 13) defines genre as "a recognizable communicative event characterized by a set of communicative purpose(s) identified and mutually understood by the members of the professional or academic community in which it regularly occurs". For Swales (1990, 1994), these characteristics are organized from models that shape the structure of the text and guide specialists of the discursive communities in terms of content and style choices. While guiding members, these models are, at the same time, delimited by their motivations regarding the schematic formatting of the manuscript.

When Swales (1990) introduced criteria for defining the academic genre, he also established an organizational description of the conventions for introducing academic articles, which would become widespread. The structure, known as the Create a Research Space (CARS) model, comprises the description of the segments³ that perform specific functions in the text, called rhetorical moves.

Next, we present the CARS model, as adapted from Swales (1990: 141), set into three moves that cover specific steps:

Move 1 – Establish the Territory
 Step 1: Establish the importance of research and/or
 Step 2: Make generalizations about the topic and/or
 Step 3: Review the literature

2. Move 2 – Establish the Niche
Step 1a: Counterargue or
Step 1b: Indicate gap(s) in already established knowledge or
Step 1c: Raise questions or
Step 1d: Continue the tradition

³ Various labels have been used to refer to the information units observed from this format: moves and steps (Swales, 1990), moves and sub-moves (Santos, 1999), moves and subfunctions (Motta-Roth, 1995), moves and strategies (Araújo, 1999) and rhe-torical units (Meurer, 1997).

3. Move 3 – Occupy the Niche
Step 1a: Outline the goals or
Step 1b: Submit the survey or
Step 2: Present the main results or
Step 3: Indicate the structure of the article.

The models for the rhetorical structure of genres are not prescriptions but classifications for didactic purposes. Therefore, as mentioned above, they are subject to variations that derive from the characteristics of the different research areas. According to Biber and Conrad (2009), academic texts do not encompass universal characteristics, but may vary situationally, given their publication conditions. However, the traits we recognize as the most constant show us what is most relevant and conventional to the user's discursive community in question. Likewise, such traits indicate what should be prioritized, as this investigation aims to highlight.

Genre pedagogy, genre-based pedagogy, and genre-based approach are some of the names given to the framework comprised of a set of assumptions, strategies, and practices for EAP teaching and learning that have as a premise the need to communicate a message to a particular audience in an appropriate way using discourse genres (for example, research papers, webinars, abstracts).

Swales's (1990: 9) genre pedagogy, as described in his seminal book *Genre Analysis: English in academic and research settings*, "rests on a pragmatic concern to help people, both non-native and native speakers, to develop their academic, communicative competence". It is essential to mention that, even though genre pedagogy has its origins in academic settings, the approach is used to teach different discourse genres.

Pedagogical Models Combining Corpus and Genre-Based Approaches

According to Charles (2020), even though corpus methods and genre analysis share a close connection, applications of such approaches for teaching purposes are not so frequent in practice. In said applications, both the target genre and the language features to be taught play a fundamental role. While the target genre serves as the starting point and the context within which language features are built-in, the language data extracted from the corpus reveal patterns that are conventionally used by experts of the discourse community of a given discipline. Therefore, the language features to be taught should be selected according to their relevance to the chosen genre and students' needs.

As reported by Moreno and Swales (2018), the identification of linguistic features characterizing the various rhetorical moves of different genres for pedagogical purposes has been reported in many studies as the main aim of move analysis (for example, Cortes, 2013; Cotos et al., 2017; Kanoksilapatham, 2005; Le & Harrington, 2015; Swales, 1981). Moreno and Swales (2018: 41) highlight that filling the "function-form gap" involves "establishing the most salient types of text items, or patterns, occurring in a specific rhetorical context in an RA, or any other genre, that may lead a competent reader to interpret a given communicative function in a highly predictable manner". Few research methodologies and pedagogical models, though, have managed to converge these two analytic paradigms: the top-down, which involves investigations into "the rhetorical composition of texts through Swalesian (1981, 2004) move analysis", and the bottom-up, which refers to "investigations into the linguistic characteristics of texts through analysis of lexical, phraseological, grammatical, and lexico-grammatical patterns of use" (Gray et al., 2020: 261). Charles (2007: 289), for example, suggested reconciling top-down (discourse analysis) and bottom-up (corpus investigation) approaches as she presents EAP writing materials designed through "a pedagogic approach which combines discourse analysis with corpus investigation".

As the pedagogical model described above sets the scene for the EAP teaching and learning framework to be suggested in this chapter, it is essential to remember that another gap needs to be filled: the one between corpus linguistics and teaching practice. It is also noteworthy that initial decisions should be made in EAP course planning and materials design. An essential first step is to carry out a needs analysis in order to know the students' background (e.g., their language proficiency level, their background knowledge in the discipline they work with), their learning preferences (e.g., using inductive or deductive methods), as well as what they expect

and need from the course⁴. Also, decisions about which genre (e.g., oral presentation, research article), section (e.g., abstract, introduction, methodology, results), discipline (e.g., Nursing, Physics, Applied Linguistics), and language skill(s) (e.g., reading, listening, writing, speaking) the EAP course will focus on, need to be made. Information about the course to be taught and its target audience allows for defining clear and achievable learning objectives based on the learners' prior knowledge, skills, needs, preferences, and expectations. The choice of an appropriate methodology, the selection and design of materials, the feedback between learners and teachers, and the construction of knowledge that will be a consequence of this process are essential elements for designing and implementing EAP courses. It is always important to remember that course and materials design are not linear processes. Figure 1 shows an interplay between actions and procedures involved in implementing an EAP course, being the design of materials one of them:



Figure 1. Stages involved in the process of designing and implementing an EAP course

⁴ See Viana et al. (2018) for a detailed overview of types of information that can be gathered in a needs analysis, the likely sources to be examined and methods that can be employed.

The design of EAP materials

Framework

Schneuwly and Dolz (2004: 51) define didactic sequences⁵ as "a sequence of teaching modules, organized together to improve a given language practice." The authors advocate for having genres as the basis for organizing didactic sequences. With the genre as a starting point, the process of knowledge construction is scaffolded by tasks, activities, and exercises⁶ designed according to specific guiding principles (Bocorny & Welp, 2021: 1601-1602), ultimately achieving pre-established learning objectives within a specific time frame.

For the design of activities with online corpora, Reppen (2010: 43) suggests a checklist with general guidelines;

- Have a clear idea of the point that you want to teach;
- Select the corpus that is the best resource for your lesson;
- Explore the corpus completely for the point you want to teach;
- Make sure that your directions are complete and easy to follow;
- Make sure that your examples focus on the point that you are teaching;
- Provide a variety of ways for interacting with the materials;
- Use a variety of exercises types;
- If you are using computers, *always* have an alternative plan or activity in the event of computer glitches.

In coursebooks, a pedagogical unit can be the focus of one or more classes, and its structure tends to be the same throughout the book. Table 1 shows the structure of the pedagogical unit and the section titles used in the EAP writing course presented as an example in this chapter:

⁵ In this study, the terms 'didactic sequences' and 'pedagogical units' are considered equivalent in meaning.

⁶ In this study, the term 'task' is used as a didactic plan to produce a communicative response from participants, comprising one or more sets of activities. The terms 'activity' and 'exercise', in turn, are considered equivalent in meaning, and, for this reason, they are used interchangeably in the sense of segments that make up a task.

| PEDAGOGICAL UNIT | SECTION TITLES OF A |
|--|-------------------------------------|
| STRUCTURE | PEDAGOGICAL UNIT |
| Context of use, purpose and definition | 1) Activate previous knowledge |
| Characteristics of the genre | 2) Learn about key characteristics |
| Rhetorical structure | 3) Find the parts |
| Language features | 4) Know important language features |
| | 5) Analyze examples |
| Dreduction of conne | 6) Write the first draft |
| Production of genre | 7) Get feedback |
| | 8) Write the final draft |

Table 1. Pedagogical unit structure for an EAP writing course

Welp et al. (2019: 6) list guiding principles to orient teachers in planning and designing general English teaching materials. Those principles were adapted by Bocorny and Welp (2021: 1601-1602) to guide the design of EAP materials:

- 1. Learning objectives should be established based on the knowledge area and academic needs of the group of learners the tasks are aimed at;
- 2. Target genres should be academically relevant and coherent with the established learning objectives;
- 3. Selected texts should be authentic and representative of social practices and genres that circulate in the academic context;
- 4. Tasks should offer the learners opportunities to use the language proper to the texts produced in the learners' domain and raise awareness on such use in a contextualized way;
- 5. Tasks dealing with linguistic resources should take into account the frequency of lexical and discursive items present in academic texts in the learners' area of knowledge;
- 6. Tasks' order and statements should be organized in a way to promote progress and scaffold learning;
- 7. Tasks should provoke relevant interactions between learners and texts, learners and learners and learners and teacher;
- 8. Task performance should provide meaningful learning opportunities and achieve results beyond the classroom.

Specifically, when it comes to the design of EAP materials within a framework that combines corpus and genre-based pedagogies, two elements are key: knowing the rhetorical structure of the target genre and identifying language features that are relevant to the genre that is being taught, considering the learners' prior knowledge, skills, needs, and expectations (see 'Corpus Linguistics' and 'Genre, Genre-Analysis, Move-Analysis and Genre-Based Pedagogy' above for details on both elements). In particular, it is vital to identify the language features used to realize the functions expressed in genre moves and steps. Moreno and Swales (2018: 40) mentioned that "A widely shared aspiration of move analysts has been to identify the linguistic features characterizing the various RA moves not only in English but also across languages."

A checklist for planning and designing EAP materials within a corpus and genre-based framework is proposed in the next section having in mind these two major elements, along with the guidelines suggested by Reppen (2010) and the principles put forward by Welp et al. (2019) and used by Bocorny and Welp (2021).

Step-by-step guide

This section is organized as a guide to be used by novice EAP teachers when designing materials within the proposed pedagogical model that combines corpus and genre-based approaches. We use the first five guiding principles suggested by Welp et al. (2019) and adapted by Bocorny and Welp (2021) as a checklist to be followed. Next, we provide brief explanations and describe some associated actions for each of the five first principles. Finally, examples of the proposed actions are presented, considering an EAP writing course for producing Health Sciences structured abstracts.

Description of the EAP writing course

As can be seen in Table 2, structured abstracts are the target genre of the course, which is aimed at upper-intermediate (B2, C1) Health Sciences graduate students and researchers. The course is to be taught online with a

total of 16 hours divided into 8 hours of synchronous activities and 8 hours of asynchronous activities:

| Name of the course | Written production of structured abstracts in the area of Health Sciences |
|-------------------------------|--|
| Target genre | Structured abstracts |
| Target section | All sections |
| Students level of proficiency | Upper-intermediate (B2, C1) |
| Students level of education | Tertiary level (graduate students) |
| Course modality | Online |
| Length of the course | 4 week course (16 hours: 8 hours of synchronous activities and 8 hours of asynchronous activities) |

Table 2. Description of the EAP writing course

PRINCIPLE 1. Learning objectives should be established based on the knowledge area and academic needs of the group of learners the tasks are aimed at

EXPLANATION: A learning objective is a description of what the learner should be able to do upon successful completion of an educational step (for example, course, task, exercise/activity) over a period of time. Clearly defined learning objectives specify the knowledge, skills, and/or attitudes the learner will gain from the educational step so that such aspects can be assessed later on.

EXAMPLE: As can be seen in Table 3, there are two types of learning objectives for the course described: (i) the course learning goal, which is the outcome that is expected after its successful conclusion (being able to produce a structured abstract in the area of Health Sciences to be submitted to a journal in the area) and (ii) the learning goal of each class. The fruitful accomplishment of each of these goals is verifiable through implementing pedagogical tasks:

| Learning objective of the course | By the end of this course, participants should be able to pro- duce a structured abstract in the area of Health Sciences to be submitted to a journal in the area. |
|-------------------------------------|--|
| Learning objective of class 1 | By the end of this class, participants should be able to under- stand what a structured abstract is and in which contexts it is used in the area of Health Sciences. |
| Learning objective of class 2 | By the end of this class, participants should be able to rec- ognize the rhetorical structure of a structured abstract in the area of Health Sciences. |
| Learning objective of class 3 | By the end of this class, participants should be able to use lan- guage features relevant to producing a structured abstract in the area of Health Sciences. |
| Learning objective of class 4 | By the end of this class, participants should be able to produce the first draft of a structured abstract in the area of Health Sciences. |

Table 3. Learning objectives for course and classes

PRINCIPLE 2. The target genres should be academically relevant and coherent with the established learning objectives

EXPLANATION: The target genre is the one that is going to be worked with along the course. As it has already been mentioned (see 'Framework'), within the framework proposed, two elements are central: knowing the rhetorical structure of the target genre and identifying relevant language features. Many patterns representing the rhetorical structure of academic genres can be found in the literature. Can et al. (2016: 4), for example, present the rhetorical structure of abstracts within Applied Linguistics, as shown in Figure 2:

| Abstract Moves (Pho [20]) | Function/Description | Question Asked | Move Labels along with Abbreviations in the Present Study |
|----------------------------|--|----------------------------------|---|
| Situating the research | setting the scene for the current research | What is known in the field? | introduction (I) |
| Presenting the research | stating the purpose of the study, research questions and hypotheses | What is the study about? | purpose (P) |
| Describing the methodology | describing the materials, subjects, variables, procedures, etc. | How was the research done? | methods (M) |
| Summarizing the findings | reporting the main findings of the study | What did the researcher find? | results (R) |
| Discussing the research | | | |
| (a) | interpreting the results/findings and/or giving recommendations | What do the results mean? | discussion (D-a) |
| (b) | no discussions or recommendations | | pseudo-discussion (D-b) |

Figure 2. Rhetorical structure of Applied Linguistics abstracts. From Can et al. (2016: 4)

The rhetorical structure of a given genre can also be obtained by using: (i) text structure analyzers like AntMover (Anthony, 2003); (ii) rhetorical tagging or rhetorical move-step coding (Bondi, 2022; Berdanier, 2019; Gray et al., 2020; Yoon & Casal, 2020a; 2020b; Geluso, 2019) or, concerning structured abstracts, (iii) the section headings, as suggested by Freitas and Bocorny (2021).

EXAMPLE: The target genre of the course described is structured abstracts, that is, abstracts that "describe a study using specific content headings rather than paragraph format" (Stevenson & Harrison, 2009: 1). Figure 3 exemplifies the rhetorical structure aimed at in a writing course for structured abstracts in health sciences:

ABSTRACT

Introduction: The Brazilian Ministry of Health had planned face-to-face workshops for professional training about the Clinical Protocols and Therapeutic Guidelines for Comprehensive Care for People with Sexually Transmitted Infections for the year 2020. Due to the COVID-19 pandemic, the workshops were cancelled, and a new strategy was adopted: virtual meetings, called Webinars — Clinical Protocols and Therapeutic Guidelines for Comprehensive Care for People with Sexually Transmitted Infections 2020. Objective: To report the experience at the Ministry of Health in online training about the clinical protocol and therapeutic guidelines for comprehensive care for people sexually transmitted infections for health professionals in 2020. Methods: The webinars were held in partnership with the Brazilian Society of Sexually Transmitted Diseases and the Pan American Health Organization. Each chapter of the Clinical Protocols and Therapeutic Guidelines for Comprehensive Care for People with Sexually Transmitted Infections — 2020 was converted into a webinar, with the participation of at least three experts, two speakers, and a moderator. Results: In total, 16 webinars were presented, covering topics such as sexually transmitted infections surveillance, prevention, diagnosis, treatment, public policies, and sexual violence. The initiative had more than 77,000 hits, with an average of 4,900 hits per webinar and the topic "syphilis" being the most accessed. The event reached all 27 federative units of Brazil, as well as 27 other countries. About 500 questions were received from the audience and answered during the sessions and/or through a document published later on by the Ministry of Health. Conclusion: Given the high number of hits and inquiries received, we can conclude that health professionals remained engaged in the topic of sexually transmitted infections during the pandemic. This experience shows the great potential of innovative methods for distance learning to promote continuing education, including a series of we

Figure 3. Example of a structured abstract in Health Sciences. From Gaspar et al. (2022: 2)

The example of the rhetorical structure frequency distribution shown in Figure 4 was extracted from three corpora of structured abstracts in the area of Epidemiology using the section headings, as suggested by Freitas and Bocorny (2021). To obtain the rhetorical structure shown in Figure 4, the following CQL was used in Sketch Engine: <s> []{1,3} [word=":"]:



Figure 4. Rhetorical structure of Epidemiology structured abstracts. From Freitas and Bocorny (2021: 3)

As seen in Figure 4, the section headings in all the three corpora are Methods, Results/Findings, and Conclusions, and in two corpora, Background and Objectives (aim, purpose). The procedure for identifying SECTION HEADINGS used in this study is described below.

PROCEDURE 1:

- 1) Go to Sketch Engine
- 2) Select the corpus you want to work with
- 3) Go to Concordance
- 4) Select Advanced
- 5) Click on CQL
- 6) Paste the CQL <s> []{1,3} [word=":"]
- 7) Click on GO

The results from **PROCEDURE 1** are shown in Figure 5. These headings can be categorized into families representing the sections of the structured abstracts of the discipline under study:

| 136 | () doc#3620 ionitoring the cancer burden in this at-risk population. | Aims: | This retrospective case-control study was aimed at identifying |
|-------|--|-------------------|--|
| 137 | () doc#3632 nd lymphoma risk and medical radiation associations. -/s>s> | Aims: | The European Senior Program (ESP) aims to avoid waiting lis |
| 138 | () doc#3779 vices, which is amplified by the COVID-19 pandemic. | Aims: | We studied the pattern of spatial association between post-aci |
| 139 🔲 | (i) doc#3810 roke among different CKD stages are not well known. | Aims: | We aimed to investigate whether the severity of CKD would in |
| 140 | (i) doc#3829 or from those observed in day-to-day clinical practice. | Aims: | To compare the risk of stroke/systemic embolism (S/SE) and r $\begin{tabular}{l} \begin{tabular}{l} \end{tabular}$ |
| 141 | $()$ doc#3914 ${\it y}$ become more prevalent with increasing prematurity. | Aims: | To investigate the association between PCOS and extremely ${\ensuremath{\tt r}}$ |
| 142 | (i) doc#4036 inal fusion surgery without additional adverse effects. | Aims: | The aims of this study were to examine the prevalence of host |
| 143 🔲 | doc#2011 and weight loss and thus worsening the quality of life. | Aims and methods: | Our aim was to find correlations from a multicentre database r |
| 144 | () doc#2776 Hate with the degree of liver fibrosis in these patients. | Aims and methods: | To investigate the accuracy of noninvasive scoring systems in |
| 145 | (i) doc#2966 3.SGPS data were collected for all drug shops. | Analysis: | Quantitative data were analyzed using SPSS for descriptive st |
| 146 | (i) doc#0 Id link between inflammation and cancer progression. | Author Summary: | Cancer progression has been depicted as a linear process, du |
| 147 | (i) doc#1 Ied design of solid tumor immunotherapy in the clinic. | Author Summary: | Among the many potential drugs explored within the scope of |
| 148 | (i) doc#2 / to effectively utilize existing drugs for new purposes. | Author Summary: | The combination of distinct drugs in combinatorial therapy can |
| 149 | (i) doc#3 tegies that best inhibit diverse tumor cell populations. | Author Summary: | Immunologic surveillance is a function of the immune system \boldsymbol{v} $\begin{tabular}{c} \end{tabular}$ |
| 150 | (i) doc#4 us disease, computer viruses, or ecological networks. | Author Summary: | WHO/CDC recommendations prioritize influenza vaccinations |
| 151 | (i) doc#5 died, for both influenza and other infectious diseases. | Author Summary: | The spread of infectious diseases can be inhibited by both vac |
| 152 | (i) doc#6 Ily assessed and compared with previous pandemics. | Author Summary: | The ever-increasing availability of timely, large-scale clinical e |
| 153 | (i) doc#7 Int on spatial interactions between metastatic lesions. | Author Summary: | We used mathematical modelling to formalize the standard the $\begin{tabular}{c} \end{tabular}$ |
| 154 | () doc#4038 but may also pave the way for interventional studies. | Author Summary: | Malaria remains a major source of morbidity and mortality thro |
| 155 | () doc#3468 n patients with N1 HNC when combined with surgery. | Backgroud: | Resection is still the only potentially curative treatment for pati |
| 156 | doc#3479 women living with advanced breast cancer in Ghana. | Background & aim: | In clinical practice, transarterial chemoembolization (TACE) ha |

Figure 5. Section heading of the structured abstracts being studied

PRINCIPLE 3. The selected texts should be authentic and representative of social practices and genres that circulate in the academic context

EXPLANATION: An authentic and representative sample of texts to extract language data to inform materials design can be obtained in existing freely-available corpora (for example, COCA⁷, MICUSP⁸, CODISSAE⁹). However, suppose you want to design a pedagogical unit of a genre (or section of a genre) that is not available in the existing freely-available corpora. In that case, you can compile your corpus using tools like AntCorGen (Anthony, 2022b)¹⁰ or Sketch Engine (Kilgarriff, 2004)¹¹. AntCorGen, for example, is very useful for designing tasks and exercises for discipline and section-specific EAP writing courses on research articles or abstracts, that is, EAP courses that focus on one of the sections of research articles within a particular discipline. Now, suppose you want to work with a more specific genre within a particular area. In that case, you may have to compile your corpus manually and upload it to a tool that will enable language data extraction.

EXAMPLE: Three corpora were compiled for the course on the **Written Production of Health Sciences Structured Abstracts**. As described by Freitas and Bocorny (2021), the corpora comprise abstracts from Epidemiology articles published in peer-reviewed indexed journals between 2003 and 2021. Their characteristics are represented in Table 4:

⁷ https://www.english-corpora.org/coca/

⁸ http://micusp.elicorpora.info/

⁹ https://drive.google.com/drive/folders/145ZFPOUuCwvTWFirM-

lqG1vGbD-1g7p7o?usp=sharing

¹⁰ https://www.laurenceanthony.net/software/antcorgen/

¹¹ https://www.sketchengine.eu/blog/build-a-corpus-from-the-web/

| Domain | Corpus | Words with repeti- tion (tokens) | Words with- out repeti- tion (types) | Texts | Average words per abstract |
|--------------|----------|---|---|-------|----------------------------------|
| Epidemiology | SJC | 662,747 | 21,087 | 1,915 | 346 |
| Epidemiology | PLOS ONE | 1,000.003 | 43,066 | 4,330 | 230 |
| Epidemiology | BJSTD | 83,261 | 9,010 | 360 | 231 |

Table 4. Numbers of corpora used in the study. From Freitas and Bocorny (2021: 2)

PRINCIPLE 4. The tasks should offer the learners opportunities to use the language proper to the texts produced in the learners' domain and promote reflections on such use in a contextualized way

EXPLANATION: After compiling the corpus that will be used to inform the design of tasks and exercises within a pedagogical unit, it is time to choose a language feature (or language features) that will be focused on. Said language feature needs to be proper and relevant to the texts produced in the learners' knowledge area. The decision on which language features to focus on in EAP courses can challenge novice EAP teachers. Some of these features have been addressed in different studies as relevant for producing academic genres. Swales and Feak (2009), for example, mention tenses (past tense x simple present tense), passive voice, metadiscoursal expressions, lexical bundles, 'that' clauses, reporting verbs, pronouns (I, we). Kanoksilapatham (2005) refers to passive constructions, past tense, 'that' clauses, and metatextual devices. Table 5 provides examples of language features and ways of retrieving them from corpora using SE CQL queries. It is important to emphasize that the previous identification of language features elicited by learners as relevant also works as a compass needle pointing to what to focus on.

| Language feature | Way to extract | | |
|------------------|--|--|--|
| to be analyzed | language feature using SE CQL queries | | |
| | Passive voice: []{1,5} [tag="VBD.*" tag="VBG" tag="VBN" tag="VBP" tag="VBZ"] [tag="VVN"] | | |
| Sentence voice | Passive voice in each section of a structured abstract: <s> []{1,3} [word=":"] []{1,5} [tag="VBD.*" tag="VBG" tag="VBN" tag="VBP" tag="VBZ"] [tag="VVN"]</s> | | |
| | Obs: It is possible to FILTER the results obtained in the previ- ous search by section heading or specific words (for example, the word 'by') to obtain concordance lines with passive voice in section CONCLUSION of a structured abstract followed | | |
| | by the word 'by'. See Appendix 5 for results. | | |
| Pronouns (I, we) | Pronouns in each section of a structured abstract: <s> []{1,3} [word=":"] [lemma="we" lemma="I"]</s> | | |
| | Lexical bundles in each section of a structured abstract | | |
| | <s> []{1,3} [word=":"] []{1,4} [word="study"] []{1,4}</s> | | |
| Lexical Bundles | | | |
| | Obs: In this case, the word 'study' can be replaced by any of the | | |
| | collocation nodes identified in the wordlist (see Figure 11) | | |

Table 5. Some language features and ways of retrieving them from corpora using SE CQL queries.

Some of these language features are easier to extract and analyze. Imagine that one of your students wants to know whether to use 'I' or 'we'¹² when writing structured abstracts. Simply checking the wordlist for pronouns will show that, in our study corpus, 'we' occurs 3,345 times per million words (pmw) while 'I' occurs 95 times (pmw). If your students want to know which pronoun is more conventional in the different sections of structured abstracts in initial position, after the section heading (for example, 'CONCLUSION: We concluded that'), it is possible to use the CQL <s> []{1,3} [word=":"] [lemma="we" | lemma="I"]. All the 1,037 concordance

¹² Previous research has explored the role of personal pronouns in academic writing (Henderson & Barr, 2010; Martínez, 2005; Hyland, 2002). According to Hyland (2002), a solid authorial identity that refers to authors taking 'ownership' for their work has to do with the use of self-reference in active voice constructions (where personal pronouns are used) as opposed to the anonymity of passive forms.

lines obtained with this query show section headings followed by the pronoun 'we'. This information could orient an exercise on authorial identity (see footnote 11) and on the use of pronouns in a course on writing structured abstracts.

EXAMPLE: For the course on **Written Production of Structured Abstracts in Health Sciences**, the language feature selected was Lexical Frames (LFs), that is, discontinuous sequences of words forming a structure around variable slots (Gray & Biber, 2013). According to Gray and Biber (2013), written academic discourse relies primarily on LFs. For this reason, that language feature has great pedagogical importance in written academic genres.

PRINCIPLE 5. Tasks dealing with linguistic resources should take into account the frequency of lexical and discursive items present in academic texts in the learner's area of knowledge

EXPLANATION: The lexical and discursive items selected as language features should be conventional. In other words, they should reveal the language used by the expert discourse community of a given discipline.

EXAMPLE: Learning about tools that can facilitate the teacher's access to linguistic data obtained from corpora might help bridge the gap between corpus linguistics and language teaching (Cheng, 2010). Different methodologies (for example, bundles-to-frames approach and fully inductive approach¹³) and tools (for example, AntGram 0.0.3 (Anthony, 2017), AntConc 4.1 (Anthony, 2022b)¹⁴, WordSmith Tools 8.0 (Scott, 2000), KfNgram 1.3.1

¹³ Bundles-to-frames approach (Biber, 2009; Römer, 2010) and fully inductive approach (Gray & Biber, 2013) are methodological procedures for identifying LFs in a corpus. While, according to Gray and Biber (2013), the former starts by finding the most frequent continuous lexical sequences in a register and then analyzes the sequences to determine if they are associated with discontinuous lexical frames with variable slots, the latter "directly identifies the full set of discontinuous sequences in a corpus" (Gray & Biber, 2013: 111).

¹⁴ The use of different versions of AntConc implies the impossibility of extracting certain data related to Lexical Frames.

(Fletcher, 2012)) have been suggested for the extraction of LFs. AntConc 4.1 is, in our opinion, the most user-friendly tool for extracting LFs. Figure 6 shows the LFs extracted from the corpus of Health Sciences RA structured abstracts with AntConc 4.1 (Anthony, 2022b). The criteria used for the extraction was: n-gram size = 6, open slots = 2, minimum frequency = 60, minimum range = 20.

| 10 1371_journal_pbio_1 1 i+ 10 1371_journal_pbio_1 2 p+ 10 1371_journal_pbio_1 3 the 10 1371_journal_pbio_1 4 this 10 1371_journal_pbio_1 5 the 10 1371_journal_pbio_1 5 the 10 1371_journal_pbio_1 1 6 10 1371_journal_pbio_1 7 7 | i i + i p the + of + of + study was study + to + the + of this + was | 1 2 3 4 | 1596 1385 1379 | 530 1299 | 0.155 | 0.682 | | | | |
|--|--|------------------|----------------------|-------------|-------|-------|-------|-------|-------|-------|
| 10_1371_journal_pbio_1 2 p + 10_1371_journal_pbio_1 3 the 10_1371_journal_pbio_1 4 this 10_1371_journal_pbio_1 5 the 10_1371_journal_pbio_1 6 the 10_1371_journal_pbio_1 7 n the | p the + of + of + study was study + to + the + of this + was | 2 3 4 | 1385 1379 | 1299 | 0.021 | | | | | |
| 10_1371_journal_pbio_(3 the 10_1371_journal_pbio_(4 this 10_1371_journal_pbio_(5 the 10_1371_journal_pbio_(6 the 10_1371_journal_pbio_(10 6 the 10_1371_journal_pbio_(7 p.re 6 | + of + study was study + to + the + of this + was | 3 | 1379 | | 0.021 | 0.65 | | | | |
| 10 1371_journal_pbio_ 4 this 10 1371_journal_pbio_ 5 the 10 1371_journal_pbio_ 6 the 10 1371_journal_pbio_ 7 n | study + to + the + of this + was | 4 | | 1377 | 0.011 | 0.399 | | | 0.003 | 0.4 |
| 10_1371_journal_pbio_ 10_1371_journal_pbio_ 10_1371_journal_pbio_ 10_1371_journal_pbio_ 7 pre | + of this + was | | 1251 | 1250 | | | 0.01 | 0.518 | | |
| 10_1371_journal_pbio_ 10_1371_journal_pbio_ 10_1371_journal_pbio_ 7 p.re | | 5 | 1234 | 1231 | 0.015 | 0.394 | | | | |
| 10_1371_journal_pbio | + + this study was | 6 | 1171 | 1169 | 0.014 | 0.403 | 0.008 | 0.03 | | |
| | sults p + + of | 7 | 1002 | 1002 | | | | | 0.11 | 0.403 |
| 0_1371_journal_pbio_ 8 aim | of + study + to | 8 | 982 | 982 | | | 0.005 | 0.371 | | |
| 10_1371_journal_pbio_ 9 aim | of + + was to | 9 | 897 | 897 | | | 0.004 | 0.436 | 0.011 | 0.123 |
| 10_1371_journal_pbio_ 10 aim | of this + + to | 10 | 863 | 863 | | | | | 0.013 | 0.148 |
| 10_1371_journal_pbio_ 11 the | aim of + + was | 11 | 859 | 859 | | | | | 0.005 | 0.431 |
| 10_1371_journal_pbio_ 12 aim | + + study was to | 12 | 857 | 857 | 0.004 | 0.016 | 0.005 | 0.442 | | |
| 10_1371_journal_pbio13aim | + this study + to | 13 | 823 | 823 | 0.005 | 0.066 | | | | |
| 10 1371 journal pbio 14 the | aim + + study was | 14 | 822 | 822 | | | 0.004 | 0.017 | 0.005 | 0.436 |
| 0_1371_journal_pbio_ 15 p+ | n a + of | 15 | 790 | 774 | 0.023 | 0.37 | | | | |
| .0_1371_journal_pbio_ | | 10 | 771 | 754 | 0.023 | 0.57 | | | | |

Figure 6. LFs extracted with AntConc 4.1 described in PROCEDURE 2. From Anthony (2022b)

PROCEDURE 2:

- 1) Open AntConc 4.1
- 2) Upload the corpus you want to work with
- 3) Click on N-Gram
- 4) Select the extraction criteria (in this extraction we used n-gram size = 6, open slots = 2, minimum frequency = 60, minimum range = 20).
- 5) Click on START

The results show the most recurrent LFs in this corpus. It is possible to see that the most frequent units are those that linguistically express the rhetorical function 'presenting the aim of the study'. If you double-click on

one of the LFs (for example, 'this study + to + the'), you can see the unit in context, as shown in Figure 7:

| Target Corpus | | | KWIC Plot File C | luster N-Gram Collocate Word | Keyword |
|--|------|-----------------------------------|--------------------------------|-----------------------------------|--|
| Name: temp | Tota | Hite: 1351 Page Size 100 bits | 1 to 100 of 1251 bit | | |
| Files: 15070 | 1018 | | | | |
| Tokens: 4009256 | | File | Left Context | Hit | Right Context |
| _10_1371_journal_pmed 10_1371_journal_pmed | 5 | _10_1371_journal_pmed_0040290.txt | s unknown. The objective of | this study was to assess the | absolute risk of venous |
| _10_1371_journal_pmed | 6 | _10_1371_journal_pmed_1000057.txt | ociated Q-waves. The aim of | this study was to investigate the | prevalence and prognosis ass |
| _10_1371_journal_pmed | 7 | _10_1371_journal_pmed_1000194.txt | rials (RCTs). The objective of | this study was to evaluate the | external validity of published |
| _10_1371_journal_pmed | 8 | _10_1371_journal_pmed_1000339.txt | oopulations. The objective of | this study was to quantify the | overall impact of lifestyle- |
| _10_1371_journal_pmed | 9 | _10_1371_journal_pmed_1001140.txt | es (GBD) studies. The aim of | this study was to compare the | population burden of injuries |
| _10_1371_journal_pmed _10_1371_journal_pmed | 10 | _10_1371_journal_pmed_1001505.txt | ional challenges. The aim of | this study was to investigate the | rates of first diagnosis |
| _10_1371_journal_pmed | 11 | _10_1371_journal_pmed_1001599.txt | prediction. The objective of | this study was to evaluate the | relationship between OSA-re |
| _10_1371_journal_pmed | 12 | _10_1371_journal_pmed_1001709.txt | l subsidies. The objective of | this study was to measure the | effect of the TSC |
| _10_1371_journal_pmed | 13 | _10_1371_journal_pmed_1002368.txt | alidated. The primary aim of | this study was to evaluate the | Stockholm and Helsinki CT |
| _10_1371_journal_pmed | 14 | _10_1371_journal_pmed_1002392.txt | ric disorder. The purpose of | this study was to estimate the | incidence of postpartum affe |
| _10_1371_journal_pmed 10_1371_journal_pmed | 15 | _10_1371_journal_pmed_1002543.txt | developed. The main aim of | this study was to compare the | association between 35 frailt |
| _10_1371_journal_pmed | 16 | _10_1371_journal_pmed_1002625.txt | up periods. The objective of | this study was to investigate the | association between adheren |
| _10_1371_journal_pmed | 17 | _10_1371_journal_pmed_1002833.txt | s for health. The objective of | this study was to explore the | broad clinical effects of |
| _10_1371_journal_pmed | 18 | _10_1371_journal_pmed_1002844.txt | nce system. The objective of | this study was to determine the | existence and magnitude of |
| _10_1371_journal_pmed | 19 | _10_1371_journal_pmed_1003142.txt | rrently unknown. The aim of | this study was to investigate the | clinical impact of this |
| _10_1371_journal_pmed 10_1371_journal_pmed | 20 | _10_1371_journal_pmed_1003366.txt | stroke survivors. The aim of | this study was to estimate the | trends over time in |
| _10_1371_journal_pmed | 21 | _10_1371_journal_pmed_1003504.txt | n societies. The objective of | this study was to quantify the | risk of several adverse |
| _10_1371_journal_pmed _10_1371_journal_pmed | 22 | _10_1371_journal_pntd_0001214.txt | Background The aim of | this study was to investigate the | relationship between prior <i< td=""></i<> |
| _10_1371_journal_pmed | 23 | _10_1371_journal_pntd_0005270.txt | other countries. The aim of | this study was to explore the | characteristics and prognosti |
| _10_1371_journal_pmed _10_1371_journal_pmed _10_1371_journal_pmed _10_1371_journal_pmed | Sear | ch Query 🗹 Words 🗌 Case 🗌 Regex R | esults Set All hits | Context Size 10 token(s) | |
| _10_1371_journal_pmed | this | study + to + the | | Start Adv Search | |
| _10_1371_journal_pmed | Sort | Options Sort to right Options | 1 1R ᅌ Sort 2 2R | Sort 3 3R COrder by fre | 9 ᅌ |
| Progress | | | | | |

Figure 7. LF 'this study + to + the' in context. From Anthony (2022a)

The LFs extracted with AntConc 4.1 can 'inspire' the creation of a CQL that could be used in SE to identify the LFs used in the different sections of the structured abstracts. For example, the LF 'the + of + study was' can lead to the following CQL [lemma="the"] [tag="N.*"] [lemma="of"] [lemma="of"] [lemma="study"] [tag="VB.*"] [lemma="to"] [tag="V.*"]. To extract the LF in different sections of structured abstracts, this CQL should contain <s> []{1,3} [word=":"]. Hence, the CQL becomes: <s> [] {1,3} [word=":"] [lemma="the"] [tag="N.*"] [lemma="of"] [lemma="this"] [lemma="the"] [tag="N.*"] [lemma="this"] [lemma="this] [lemma="t

Another way of identifying recurrent LFs in sections of structured abstracts is by having collocation nodes as a starting point. Following Flowerdew (2013), Freitas and Bocorny (2021) used a combination of lexical and phraseological elements to extract LFs from Epidemiology RA structured abstracts. A list of frequent noun collocation nodes was used "as a starting point for collocation look-ups" (Frankenberg-Garcia et al., 2021: 208). As can be seen in Figure 8, the five most frequent nouns in the Epidemiology PLOS ONE study corpus were 'patient', 'risk', 'study', 'cancer', and 'result'. Collocation nodes could also be found in other word classes, like verbs, adjectives, adverbs, and prepositions:

| W | ORDLIST | Med&HealSci1mill - Vivian |] (1) | | G | D 🕜 📕 | ■ ≗ |
|----|------------------------|---|---------|--------------------|---|------------------|-----|
| no | UN (16,409 items I 402 | 2,475 total frequency) | | | ર 🛓 | • - (|) ☆ |
| | Lemma | Frequency Per Million $^{?}$ \downarrow | DOCF ? | Lemma | Frequency Per Million $^{?}$ \downarrow | DOCF ? | |
| 1 | patient | 11,678.11 | 3,043 | 251 curve | 198.90 | 201 | |
| 2 | risk | 4,591.83 | 1,938 | 252 prescription | 198.16 | 123 | |
| 3 | study | 4,584.41 | 3,014 | 253 classification | 196.68 | 184 | |
| 4 | p | 4,560.66 | 1,789 | 254 region | 196.68 | 176 | |
| 5 | cancer | 4,359.53 | 1,395 … | 255 category | 196.68 | 147 | |
| 6 | ci | 3,440.72 | 1,431 | 256 program | 195.19 | 157 | |
| 7 | result | 3,026.59 | 3,245 | 257 endpoint | 195.19 | 163 | |
| 8 | group | 2,885.57 | 1,296 | 258 efficacy | 193.71 | 199 | |
| 9 | disease | 2,683.70 | 1,529 | 259 growth | 193.71 | 141 | |
| 10 | method | 2,645.11 | 3,118 | 260 community | 192.97 | 156 | |
| 11 | year | 2,550.85 | 1,717 | 261 experience | 192.22 | 154 | |
| 12 | conclusion | 2,432.10 | 3,234 | 262 hemorrhage | 190.74 | 124 | |
| 13 | analysis | 2,277.73 | 1,927 | 263 prediction | 190.00 | 158 | |
| 14 | factor | 2,270.31 | 1,489 | 264 adherence | 189.25 | 90 | |
| 15 | treatment | 2,259.18 | 1,296 | 265 safety | 188.51 | 184 | |
| | | | | | | | |

Figure 8. Noun wordlist for the Health Sciences PLOS ONE study corpus. From Kilgarriff et al. (2004)

Using Sketch Engine and searching for concordance lines with the lemma 'study' as a noun, it is possible to retrieve language data that could be easily integrated into exercises to be used in the course **Written Production** of Structured Abstracts in the Area of Health Sciences. Figure 9 shows the results:

| С | ٦C | | DANCE | Med8 | HealSci1m | nill - Vivian | ٩ | . i | | | | | | | | | | | | Ð | 0 | P | 2 |
|----|----|--|---------------------------|----------------|----------------|----------------|-------------|---|------------------------|---------------------|-------|----------|------------|------------|---------|--|---------|----------|---------|---------------|-----------|--------------------------|---|
| 0 | C | emma study • 6 4,584.41 per million | 5,177 tokens • 0.46% 1 | ۶ | | | | | ঽ | ± | | 0 | 3 > | : = | Ŧ | GD | F | ••• | ıL | | + | () | ☆ |
| | | Details | | | | | | Left co | ntext | KWI | | Right | context | | | | | | | | | | |
| 1 | | i doc#0 | and thus would | d occur lat | te in cance | r progressio | on. | >Howeve | r, recer | nt stud | ies : | show th | nat can | cer cell | s disse | eminat | e ear | ly, imp | olying | the existe | nce of a | a differe | |
| 2 | | i doc#0 | is mutations that | at confer t | he ability to | o metastasia | | >Howeve | r, recer | nt stud | ies : | show th | nat can | cer cell | s disse | eminat | e ear | y, bef | ore su | ich mutati | ons car | n accur | |
| 3 | | i doc#1 | tic (PK) and ph | armacody | namic (PD |) data were | acquired | from a pr | reclinica | al <mark>stu</mark> | dy a | applyin | g syste | emic IL- | 21 the | rapy ir | n muri | ne so | lid ca | ncers. | <s>We</s> | develo | |
| 4 | | i doc#1 | its strength as a | in antican | cer agent l | nas been es | tablished | in severa | al anima | al stud | ies , | respo | nse rat | es in m | elanor | na anc | d rena | l cell (| carcin | oma patie | nts ren | nain lov | |
| 5 | | i doc#1 | d its optimal clin | ical use.< | /s> <s>Our</s> | model inte | grates da | ta from pi | reclinica | al stud | ies I | under o | liverse | IL-21 ti | reatme | ent sett | tings, | and w | vas va | lidated by | extens | sive exp | |
| 6 | | i doc#3 | ⇒ of actively pro | liferating a | and functio | ning immun | e system | cells. <th><mark>≪s></mark>Oι</th> <th>ır stu</th> <th>dy (</th> <th>contrib</th> <th>utes to</th> <th>an und</th> <th>erstan</th> <th>ding o</th> <th>fcand</th> <th>er de</th> <th>velop</th> <th>ment durin</th> <th>ng imm</th> <th>une sys</th> <th></th> | <mark>≪s></mark> Oι | ır stu | dy (| contrib | utes to | an und | erstan | ding o | fcand | er de | velop | ment durin | ng imm | une sys | |
| 7 | | i doc#6 | irst affected p | opulation | R0 and p | C could be 1 | well estim | ated. | - <s>Thi</s> | s stu | dy i | provide | is a cle | ar refer | ence i | n this f | two-di | imens | ional | space aga | uinst wh | nich futu | |
| 8 | | i doc#7 | >To improve | our know | ledge of it | s dynamics, | we cond | ucted a m | nodellin | g stu | dy I | based | on mul | ti-moda | data | from a | n orth | otopio | c muri | ne experir | mental | system | |
| 9 | | i doc#9 | the utilization of | health ca | re services | s as a proxy | for symp | toms, the | preser | nt stu | dy s | seeks t | o dete | rmine w | hen e | arly sy | mptor | ns of | childh | ood cance | er are s | een in | |
| 10 | | i doc#9 | neral practice.< | :/s> <s>Me</s> | ethods: A p | opulation-b | ased mat | ched com | parativ | e stu | dy v | was co | nducte | d using | natior | wide r | registi | y data | a. | <s>As ca:</s> | ses, all | childre | |
| 11 | | i doc#9 | em (CNS) tumo | urs had m | iore contac | ts than othe | er childrer | n during ti | ne entir | e stu | dy I | period. | | >The u | se of p | ractice | e-bas | ed dia | gnost | ic tests an | id the n | umber | |
| 12 | | () doc#9 | ase four to five | months b | efore canc | er diagnosi | s. <s></s> | Conclusio | ons: Th | e stu | dy s | shows | that ex | cess he | ealth c | are use | e, a p | roxy f | or syn | nptoms of | childho | od can | |
| 13 | | i doc#11 | <s>Background</s> | I: Most me | eta-analyse | es include d | ata from o | one or mo | ore sma | ll stud | ies t | that, in | dividua | illy, do r | ot hav | e pow | er to | detect | t an in | tervention | effect. | <s< th=""><th></th></s<> | |
| 14 | | i doc#11 | t. | elative infl | uence of a | dequately p | owered a | nd unden | powere | d stud | ies i | in publi | shed n | neta-an | alyses | has n | ot pre | vious | ly bee | n explored | d. | s>We | |
| 15 | | (i) doc#11 | y been explored | 1. | We examir | ne the distril | oution of p | power ava | ailable i | n <u>stud</u> | ies 1 | within r | neta-a | nalyses | publis | hed in | Coch | nrane | reviev | vs, and inv | vestiga | te the ir | |
| 16 | | i doc#11 | ublished in Coch | nrane revi | ews, and ir | nvestigate ti | ne impact | of under | powere | d stud | ies (| on met | a-analy | sis res | ults. | s> <s></s> | Metho | ids an | id Fin | dings: For | 14,886 | meta- | |
| 17 | | i doc#11 | of binary outcom | ies from 1 | ,991 Coch | rane review | s, we cale | culated p | ower pe | r stu | dy v | within e | each m | eta-ana | lysis.< | :/s> <s:< th=""><th>>We o</th><th>define</th><th>d ade</th><th>quate pow</th><th>ver as a</th><th>50% p</th><th></th></s:<> | >We o | define | d ade | quate pow | ver as a | 50% p | |
| 18 | | i doc#11 | : reduction. | <s>In a su</s> | ubset of 1, | 107 meta-ar | nalyses in | cluding 5 | or mor | e stud | ies 1 | with at | least th | vo adeo | quately | powe | red a | nd at I | least | one under | powere | d, resu | |

Figure 9. Concordance lines with the lemma 'study' as a noun. From Kilgarriff et al. (2004)

PROCEDURE 3:

- 1) Open Sketch Engine
- 2) Select the corpus you want to work with
- 3) Choose Concordance
- 4) Select Advanced
- 5) Click on lemma, in Query type
- 6) Click on noun, in Part of speech
- 7) Write 'study' (or any other recurrent collocation node) under Lemma
- 8) Press GO

Figure 10 illustrates the search for 'study':

Figure 10. SE interface for PROCEDURE 3. From Kilgarriff et al. (2004)

The results obtained with PROCEDURE 3 can be filtered for each structured abstract recurrent section heading: (METHODS, RESULTS/ FINDINGS, CONCLUSIONS BACKGROUND, and OBJECTIVES/AIM/ PURPOSE). For example, Figure 11 shows the filtered results of concordance lines with the lemma 'study' for the section CONCLUSIONS:

| СС | | ICORD | ANCE | Med&HealSci1mill - V | ivian Q | i | | | | | e | o (| 2 | . 2 |
|----|---------|---|---------------------|---|------------------------|----------------------|------------|------------------|-----------------|-------------------|-------------------|----------|-----------|---------|
| ٩ | 16 4 | mma study • 6,1 584.41 per million tol | 177 kens • 0.46% | filter Conclusion (-33,+KV 141.01 per million tokens • 0.014 | /IC) • 190 × % 1 | ۴ | | | | | | | | |
| | | | | | | ર : | <u>•</u> ≡ | o 🦉 X | ₹ ₹ | GD 🖬 🚥 | ıl. KWI | • | + (| 0 ☆ |
| | | Details | | | | Left context | KWIC | Right context | | | | | | |
| 1 | | i) doc#9 | e four to five m | onths before cancer diag | nosis. <s>C</s> | onclusions : The | study | shows that exc | ess health ca | re use, a proxy | for symptom | s of ch | ildhood | car 📄 |
| 2 | | i) doc#53 | ygy with an area | a under the ROC curve o | f 0.85. <s>C</s> | conclusions : Our | study | confirmed seve | eral factors as | sociated with n | ormal liver hi | stology | , includ | ing 🚡 |
| 3 | | i doc#70 | 3 was 1.47 (95% | 6 CI, 1.13 to 1.92; p = 0.0 | 0045). <s>C</s> | onclusions : This | study | showed an incr | reased risk of | developing IHI | D in young pa | tients v | with nev | vly i 🐚 |
| 4 | | i doc#77 | ependently ass | ociated with an increase | d risk. <s>C</s> | onclusions : This | study | showed that HI | Ds, which are | widely used in | South Korea | in the | winter s | iea: 🚡 |
| 5 | | i doc#87 | I4; 0.36–0.53) c | ompared with low persis | tence. <s>C</s> | conclusions : Our | study | reinforces the t | penefits of AH | I medications in | n routine clinic | cal prac | ctice an | d hi 🚡 |
| 6 | | i doc#100 | lyses were not | performed for any of the | other outcomes | due to scarcity of | studies | | lusions : The | targeted interv | rentions aimir | ig to im | iprove r | nat: 🚡 |
| 7 | | i) doc#132 | ition, sex and a | ge differences were obse | erved. <s>C</s> | onclusions : This | study | confirms the as | ssociation bet | ween cholangio | ocarcinoma a | nd sevi | eral les | s es 👕 |
| 8 | | i) doc#168 | ndrome, allergie | es, endometriosis, and as | thma. <s>C</s> | conclusions : Our | study | results indicate | d an associat | tion between hy | yperthyroidisr | n and E | BPS/IC. | |
| 9 | | i) doc#183 | with more home | ogeneous overall surviva | l rate. <s>C</s> | onclusions : This | study | defines that the | e lymph node: | s ratio is an ind | ependent pro | gnostic | c factor | for 🕞 |
| 10 | | i) doc#221 | 42; 95% CI, 1.0 | 19-1.84) than older GD p | atients. | Conclusion : This | study | found an increa | ased risk of C | VD in patients | diagnosed wi | th GD. | <s></s> | •The 🖺 |
| 11 | | i doc#310 | ecayed, missed | and filled teeth (DMFT) | values. | Conclusion : This | study | revealed that c | hronic period | ontitis, tooth me | obility, furcatio | on invo | lvemen | t an 🖺 |
| 12 | | i) doc#342 | are compared to | o fertile women with adeo | uate care. | SConclusions : | Study | findings sugge | st that adequa | ate prenatal ca | re can reduce | the ris | k of ad | vers 📄 |
| 13 | | i doc#373 | er type accordir | ng to the primary sites of | NETs. <s>C</s> | onclusions : Our | study | showed that th | e risk of seco | nd cancer follo | wing NETs is | increas | sed, esp | oeci 📄 |
| 14 | | i doc#376 | significant whe | n all covariates were adj | usted. <s>C</s> | onclusions : This | study | relieves the co | ncern of a bla | dder cancer ris | k associated | with hu | uman in | suli р |
| 15 | | i doc#489 | ders was 1.93 (| 95% Cl, 1.16–3.20; p = 0 |).0110). | Conclusion : Our | study | showed an incr | reased risk of | developing isc | hemic stroke | in your | ng patie | nts 🚡 |
| 16 | | i) doc#494 | to 2.27±0.68 mm | m at the 3- to 5-year follo | w-up. <s>C</s> | onclusions : This | study | provides clinica | al and angiogi | raphic results fr | rom a large p | opulatio | on of pa | ıtier 👕 |
| 17 | | i doc#505 | ation of the pres | ent meta-analysis is the | non-randomizat | tion of all included | studies | | lusions : RPN | A appears to be | e an efficient a | alternat | tive to C |)PN 📄 |

Figure 11. Filtered results of concordance lines with the lemma 'study' for the section CONCLUSIONS. From Kilgarriff et al. (2004)

PROCEDURE 4 presents the steps for filtering data:

- 1) Use the results obtained with PROCEDURE 3 (search for the lemma 'study', as a noun)
- 2) Click on the Filter icon, as shown in Figure 12:

| С | NC | ICORE | DANCE | Med&Heals | Sci1mill - Vivian | Q | 0 | | | | | | | | | | | Θ | 0 | | 2 |
|----|----------|-------------------------------------|---------------------------|--------------------|--------------------|--------------|-----------------|------------|---------|----------|------------|-----------|---|--|---------------|----------|----------|----------|-----------|------------|---|
| C | le 4, | mma study • 6 584.41 per million | 5,177 tokens • 0.46% 1 | ۶ | | | ର | . <u>+</u> | : ≕ | 0 | a x | F | Ŧ | GD EX | 8 | (| ĸ | WIC - | + | () | ☆ |
| | | Details | | | | | Left context | k | WIC | Right | context | | | | | | | | | | |
| 1 | | i doc#0 | and thus would | l occur late in c | ancer progressi | on. | However, rec | ent s | studies | show t | hat can | cer cell | s disse | eminat | e early | , imply | ing the | exister | nce of a | differe | |
| 2 | | i doc#0 | is mutations that | t confer the ab | ility to metastasi | ze. | However, rec | ent s | studies | show t | hat can | cer cell | s disse | eminat | e early | , befor | e such | mutatic | ons car | accur | |
| 3 | | i doc#1 | tic (PK) and pha | armacodynamic | c (PD) data were | e acquired | from a preclin | ical | study | applyir | ng syste | mic IL- | 21 the | rapy in | murin | e solid | cance | rs. | <s>We</s> | develo | |
| 4 | | i doc#1 | its strength as a | n anticancer ag | gent has been e | stablished | in several ani | mal s | studies | , respo | onse rat | es in m | elanon | na and | renal | cell ca | rcinom | a patier | nts rem | ain lov | |
| 5 | | i doc#1 | d its optimal clin | ical use. | >Our model inte | egrates dat | ta from preclin | ical s | studies | under | diverse | IL-21 t | reatme | ent sett | ings, a | ind wa | s valida | ated by | extens | ive exp | |
| 6 | | i doc#3 | of actively prol | iferating and fu | nctioning immu | ne system | cells. | Our | study | contrib | outes to | an und | erstan | ding of | cance | er deve | lopme | nt durin | g immu | ine sys | |
| 7 | | i doc#6 | First affected p | opulation, R0 a | nd pC could be | well estim | ated. | his | study | provid | es a cle | ar refe | rence i | n this t | wo-dir | nensio | nal spa | ice agai | inst wh | ich futu | |
| 8 | | i doc#7 | Solution >> To improve | our knowledge | of its dynamics | , we cond | ucted a model | ling | study | based | on mult | ti-moda | l data i | from a | n ortho | otopic r | nurine | experin | nental s | system | |
| 9 | | i doc#9 | the utilization of | health care se | rvices as a prox | y for symp | toms, the pres | ent | study | seeks | to deter | mine w | hen ea | arly sy | mptom | is of ch | ildhoo | d cance | r are s | een in | |
| 10 | | i doc#9 | neral practice.< | /s> <s>Methods</s> | s: A population-t | based mate | ched compara | tive | study | was co | onducte | d using | nation | wide r | egistry | data. | | As cas | es, all | childre | |
| 11 | | i doc#9 | em (CNS) tumo | urs had more o | ontacts than oth | er childrer | n during the er | itire | study | period | .s | >The u | se of p | ractice | -base | d diagr | nostic t | ests and | d the n | umber | |
| 12 | | (i) doc#9 | ase four to five | months before | cancer diagnos | is. <s></s> | Conclusions: | The | study | shows | that ex | cess he | ealth ca | are use | e, a pro | oxy for | sympt | oms of o | childho | od can | |
| 13 | | i doc#11 | <s>Background</s> | : Most meta-an | alyses include of | data from o | one or more si | nall s | studies | that, in | idividua | lly, do r | not hav | e pow | er to d | etect a | n inter | vention | effect. | <td></td> | |
| 14 | | i doc#11 | t. | lative influence | of adequately | powered a | nd underpowe | red s | studies | in publ | lished n | neta-an | alyses | has n | ot prev | viously | been e | xplored | l. | s>We | |
| 15 | | i doc#11 | y been explored | l. | camine the distr | ibution of p | ower availabl | e in 🚦 | studies | within | meta-ai | nalyses | publis | hed in | Cochi | ane re | views, | and inv | estigat | e the i | |
| 16 | | i doc#11 | ublished in Coch | irane reviews, a | and investigate | the impact | of underpowe | red s | studies | on me | ta-analy | vsis res | ults. <td>s><s></s></td> <td>Nethoo</td> <td>is and</td> <td>Finding</td> <td>gs: For</td> <td>14,886</td> <td>meta-</td> <td></td> | s> <s></s> | Nethoo | is and | Finding | gs: For | 14,886 | meta- | |
| 17 | | i doc#11 | of binary outcom | es from 1,991 | Cochrane review | vs, we cald | culated power | per | study | within | each m | eta-ana | alysis.< | :/s> <s:< td=""><td>We d</td><td>efined</td><td>adequa</td><td>ate pow</td><td>er as ≥</td><td>50% p</td><td></td></s:<> | We d | efined | adequa | ate pow | er as ≥ | 50% p | |
| 18 | | (i) doc#11 | reduction. | <s>In a subset</s> | of 1,107 meta-a | nalyses in | cluding 5 or m | ore s | studies | with at | least tv | vo adeo | quately | powe | red an | d at lea | ast one | underp | owere | d, resu | |

Figure 12. Filtering data in SE. From Kilgarriff et al. (2004)

- 3) Select Advanced
- 4) Click on lemma, in Query type
- 5) Click on noun, in Part of speech
- 6) Write 'Conclusion', under Lemma
- 7) Press GO

Figure 13 illustrates the search:

| Med&HealSci1mill - Vivian | Q (j | | | | | | | | | | æ | 0 | | 2 |
|---------------------------|-------|-------|-----|---|------------|---|---|----------|-----|------------|------------------|--------|------------|---|
| F | | ঽ | ± ≡ | Ø | @ X | = | Ŧ | GD EX | 3 - | • 11 | | + | () | ☆ |
| | | | | | | | | | | | | | | × |
| | | | | | | | | | | | | | | |
| <u>•</u> | | | | | | | | | | Quick filt | ers: HIDE SUE | в-нітѕ | | |
| | | | | | | | | | | ONLY | 1ST HIT IN | | MENT | |
| Part of speech | Concl | usion | | | | | | | | | | | | |
| any | Conci | usion | | | | | | | | | | | | |
| adjective | 🗸 A = | a? | | | | | | | | | | | | |
| adverb | | | | | | | | | | | | | | |
| conjunction | | | | | | | | | | | | | | |
| noun | | | | | | | | | | | | | | |
| preposition | | | | | | | | | | | | | | |

Figure 13. SE interface for PROCEDURE 4. From Kilgarriff et al. (2004)

If you want to organize the results obtained with PROCEDURE 4, you can click on the icon SORT (to the left of the FILTER icon). The results obtained are shown in Figure 14:

| 59 | $(i)\ doc\#1804\ ing\ according\ to\ our\ meta-regression\ analysis/s>-conclusions : Our$ | study | shows that high expression of FAK is associated with a worse OS in patier $\overline{\begin{tabular}{c} \end{tabular}}$ |
|----|--|-------|---|
| 60 | i) doc#3127 tion about family histories often escapes notice.-conclusion : Our | study | shows that though generally considered a sporadic disease, the presence $\begin{tabular}{c} \begin{tabular}{c} \end{tabular}$ |
| 61 | (i) doc#623 ed by other types of NA or combination therapy. | study | suggested benefits of adjuvant NA therapy following curative treatment of |
| 62 | (i) doc#702 apillary thyroid cancer in both NLST and PLCO. <s>Conclusion : Our</s> | study | suggests that certain medical encounters, such as those using low-dose h |
| 63 | i) doc#2319 il, 1.63-16.25 vs OR, 4.2; 95% Cl, 1.56-11.34).Conclusion : Our | study | suggests that statin therapy may prevent the progression of symptomatic I $\begin{tabular}{c} \begin{tabular}{c} \end{tabular}$ |
| 64 | (i) doc#2524 ons for TV viewing did not show a clear pattern.Conclusion : Our | study | suggests that pre- and post-diagnosis physical activity is associated with ${\rm Ir}$ |
| 65 | (i) doc#2689 ter in the N0-2 or in the N3 subgroup analysis.Conclusions : Our | study | suggests that SCLN metastasis is not a prognostic factor in locally advanc |
| 66 | (i) doc#3795 'e) among the studied major chronic diseases.Conclusions : Our | study | suggests that the costs associated with treating cancer account for a low \mathfrak{g} $\begin{tabular}{ll} \end{tabular}$ |
| 67 | (i) doc#2203 ie women's arrival at the participating hospitals. <s>Conclusion : The</s> | study | demonstrated a lower maternal near-miss incidence ratio compared to pre |
| 68 | $\textcircled{\sc 0}$ doc#3490 inces, slower speeds, and higher frequencies.conclusions : The | study | found physically active lung cancer patients, although with metastatic cont $\begin{tabular}{c} \end{tabular}$ |
| 69 | (i) doc#2829)4; RR 2.43, 95% C.I.: 2.27-2.60, respectively).Conclusion : The | study | has highlighted the presence of significant differences in the quality of EOI $\begin{tabular}{c} \begin{tabular}{c} \end{tabular}$ |
| 70 | (i) doc#2958 re significantly associated with overall survival.conclusion : The | study | has clearly demonstrated that survival rate for CRC patients at KATH, Ghz $\begin{tabular}{c} \end{tabular}$ |
| 71 | $\textcircled{\sc 0}$ doc#3504 and top North-Eastern corridor of the country.conclusions : The | study | has confirmed common modifiable risk factors of two cardiovascular disea $\begin{tabular}{c} \end{tabular}$ |
| 72 | (i) doc#3568 ars obtained water from non-improved source.SConclusions : The | study | has demonstrated that children in Nigeria are not only exposed to the risk |
| 73 | (i) doc#643 ids ratio: 8.07; 95% CI: 5.14–12.68; P<0.001).SConclusions : The | study | identified high incidence of intraoperative CAs with high mortality in older $\ensuremath{\mathfrak{l}}$ |
| 74 | (i) doc#3279 I of health care utilization by pregnant women.SConclusions : The | study | identifies relevant social determinants for the utilisation of antenatal care, $ \;[\begin{tabular}{c} \end{tabular} $ |
| 75 | (i) doc#1426 on was avoidable in 17/30 (56.7%) of children.SConclusions : The | study | presents a novel methodology, examining quality of care across an entire : |
| 76 | (i) doc#3469 us with Richmond Agitation Scale after surgery.conclusion : The | study | results reveal that postoperative anaemia is not only a frequent postsurgic $\begin{tabular}{c} \hline \end{tabular}$ |
| 77 | \bigcirc doc#1015 g, while the national recommended level is 6g/s>-conclusion : The | study | revealed outdated and inadequate treatment and health education for hyp |
| 78 | (i) doc#2951 63) and 1.74 (95% CI: 1.28-2.36), respectively.Conclusion : The | study | revealed a low level of maternal knowledge of danger signs and BP/CR ar |
| 79 | i) doc#3356 pregnancy [AOR = 13.94: 95% CI 4.39, 24.27]. | study | revealed maternal sociodemographic factors, short birth space, lack of ant $\begin{tabular}{c} \begin{tabular}{c} \end{tabular}$ |
| 80 | () doc#4005 srsons in the least wealthy regions of Germany.conclusion : The | study | revealed and confirmed some profound risk factors of SVI/B at both the in |

Figure 14. Sorting data in SE. From Kilgarriff et al. (2004)

A more direct way of finding recurrent LBs (and afterwards the LFs) in the sections of structured abstracts is to use Corpus Query Language (CQL) syntaxes. The CQL <s> []{1,3} [word=":"] []{1,4} [word="study"] [] {1,4}, for example, extracts all the collocations that occur in the sections of

the structured abstracts that have 'study' as a collocation node. In this case, the collocation node 'study' can be replaced by any of the collocation nodes identified in the wordlists extracted from the corpus. Figure 15 shows the results when using this CQL:

| С | DN | | DANCE | Med&HealSci1 | mill - Vivi | an | ৎ ় |) | | | | | | | | | | | | e | 0 | | ڪ |
|----|---------|---|--|----------------------|-------------|------------|-----------|------------|---------|--------|----------|---------|--------|-------|-------|-------|-------|------|--------|---------------------------|-------------|----------|---|
| C | С 3, | QL <s> []{1,3} 532.01 per million</s> | [word=":"] []{1,4} [word tokens = 0.35% | d="study"] []{1,4} • | 4,759 | ۴ | | হ | ŧ | =: | • | 2 | × | - | Ŧ | GD | F | ••• | ıl. | | + | () | ☆ |
| | | Details | Left | context | | | | | KWI | С | | | | | | | | | Righ | context | | | |
| 1 | | i doc#9 | n in general pract | ice. | | Methods | s: A popu | lation-b | ased n | natch | ed con | npara | tive s | tudy | was | | | | condu | cted using | nation | wide re | |
| 2 | | i doc#9 | n in general pract | ice. | Meti | nods: A pe | opulation | n-based r | natche | d co | nparat | ive s | tudy | was o | ond | ucted | | | using | nationwide | e regist | ry data | |
| з | | i doc#9 | n in general pract | ice. <s></s> | Method | ls: A popu | lation-ba | ased mat | ched o | ompi | arative | stud | ly was | s con | duct | ed us | ing | | natior | wide regis | try data | a. | |
| 4 | | (i) doc#9 | n in general pract | ice. <s> Meti</s> | nods: A p | opulation | -based n | natched | compa | rative | study | was | conc | lucte | d usi | ng na | ation | vide | regist | ry data. <mark>⊲/s</mark> | > <s>A:</s> | s cases | |
| 5 | | i doc#9 | re cancer diagno | sis. | | | c | Conclusio | ons: T | he stu | idy sh | ows | | | | | | | that e | xcess heal | th care | use, a | |
| 6 | | () doc#9 | re cancer diagno | sis. | | | Co | nclusion | s: The | stud | show | rs tha | ıt | | | | | | exces | s health ca | are use | , a prox | |
| 7 | | i doc#9 | re cancer diagno | sis. <s></s> | | | Conclu | isions: T | he stu | dy sh | ows th | at ex | cess | | | | | | health | care use, | a prox | y for sy | |
| 8 | | () doc#9 | re cancer diagno | sis. | | C | onclusio | ns: The s | study : | shows | s that e | xces | s hea | lth | | | | | care u | ise, a prox | y for sy | mptom | |
| 9 | | i doc#15 | ndergo lung surg | ery. <s></s> | | | Ob | ojective: | The ai | m of t | his stu | ıdy is | • | | | | | | to dev | elop a sho | rt-term | benefi | |
| 10 | | () doc#15 | ndergo lung surg | ery. | | | Obje | ective: T | he aim | of th | is stuc | ly is t | to | | | | | | devel | op a short- | term be | eneficia | |
| 11 | | i doc#15 | ndergo lung surg | ery. <s></s> | | | Objectiv | e: The a | im of t | his st | udy is | to de | evelo | • | | | | | a sho | rt-term ben | eficial | prograr | |
| 12 | | () doc#15 | ndergo lung surg | ery. <s></s> | | | Objective | e: The air | m of th | is stu | idy is t | o de | velop | a | | | | - | short- | term benel | icial pr | ogram | |
| 13 | | i doc#18 | evated risk of can | cer. <s></s> | | | 0 | bjectives | : Our | study | objec | tives | | | | | | 1 | were | to characte | rize th | e spect | |
| 14 | | () doc#18 | ∋vated risk of can | cer. <s></s> | | | Obje | ectives: C | Our stu | idy ob | jectiv | es we | ere | | | | | | to cha | racterize t | he spe | ctrum c | |
| 15 | | i doc#18 | evated risk of can | cer. <s></s> | | | Objec | tives: Ou | ur stud | ly obj | ectives | s wer | e to | | | | | | chara | cterize the | spectr | um of n | |
| 16 | | (i) doc#18 | ∋vated risk of can | cer. <s></s> | | Obj | ectives: | Our stud | ly obje | ctives | were | to ch | aract | erize | | | | | the sp | ectrum of | neopla | sia in V | |
| 17 | | i doc#27 | preceding the st | udy. <s></s> | | | | Metho | ds: Th | is stu | dy wa | 8 | | | | | | | a cros | s-sectiona | l surve | y of wo | |
| 18 | | (i) doc#27 | preceding the st | udy. <s></s> | | | | Method | ls: Thi | s stuc | ly was | a | | | | | | | cross | sectional s | survey | of wom | |
| 19 | | i) doc#27 | preceding the sti | udy. | | | Method | ds: This : | study | was a | cross | -sect | ional | | | | | | surve | y of women | n who g | ave bi | |

Figure 15. Results for the CQL <s> []{1,3} [word=":"] []{1,4} [word="study"] [] {1,4}. From Kilgarriff et al. (2004)

PROCEDURE 5:

- 1) Open Sketch Engine
- 2) Select the corpus you want to work with
- 3) Go to Concordance
- 4) Select Advanced
- 5) Click on CQL, in Query type
- 6) Paste the CQL <s> []{1,3} [word=":"] []{1,4} [word="study"] []{1,4} under CQL
- 7) Press GO
- 8) Click on KWIC (to organize the results alphabetically)

| СС | NCORDANCE | Med&HealSci1mill - 1 | /ivian Q |] () | | | | | | | | | Ð | 0 | | 2 |
|--------|---|--|---------------|---------|----------|-------------|---------------|-----|----|-----|------------|-----------------------------|---|--------|----|---|
| ۲ | CQL <\$> []{1,3} [word=":"] []{1,4} [word=":"] []{1, | ord="study"] []{1,4} • 4,759 | Sort word × | 0 | 4 | | | | | | | | | | | |
| | | | | | থ 🖣 | = | • | 9 × | E. | - 8 | 2 | | il. KWIC - | + | () | ☆ |
| 2 | CHANGE CRITERIA | | | | | | | | | | | | | | | × |
| | ADVANCED | ABOUT | | | | | | | | | | | | | | |
| Q | uery type () simple lemma phrase word character CGL | COL <s> []{1,3} [wor insert [] () () COL BUILDER [] Default attribute ? Iemma</s> | d=":"] []{1,4 |) [word | d="stud; | r") [] # | {1,4} TAGS | 8 | 2 | | 5 ° | QL 1: Q an int corpus | Supplex cor. troduction to guage SKECH ENGINE | nanual | 2 | O |
| Su | boopus ⑤ one (the whole cor ▼ ☆ + Filter context ③ ∨ Text types ? ∨ | Macro ³ none | | | G | D | <u> </u> | ÷ • | | | | | | | | |

Figure 16. SE interface for extracting LBs from sections of the structured abstracts using CQL <s> []{1,3} [word=":"] []{1,4} [word="study"] []{1,4}. From Kilgarriff et al. (2004)

The results in Figure 16 indicate that collocations with 'study' occur across sections of these structured abstracts. These results can also be filtered for each section identified as part of the rhetorical structure of the abstracts under study. For example, as shown in Figure 17, collocations with the word 'study' occur 440 times in the section CONCLUSION in the corpus of Health Sciences:

| ANCE Med&HealSci1m | ill - Vivian Q (i) | ප 🕐 🖪 🔮 |
|--|--|---|
| lord=":"] []{1,4} [word="study"] []{1,4} • 4,7 kens • 0.35% | 759 \$ filter Conclusion (-33,+KWIC) • 440 \$ Sort word × \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ | |
| | 역 🛓 ≕ 💿 🍕 % ≓ ≂ 🛱 🖥 | ••• 🗈 KWIC 🕶 + 🛈 🕁 |
| Left context | KWIC 1 | Right context |
|) tients were still living. | Conclusion : Data of this study suggest | that the APSCU may have a re |
|) tients were still living. | Conclusion : Data of this study suggest that | the APSCU may have a releva 下 |
|) tients were still living. | Conclusion : Data of this study suggest that the | APSCU may have a relevant rt |
|) tients were still living. | Conclusion : Data of this study suggest that the APSCU | may have a relevant role for m |
|) rgical group p<0.001. | Conclusion : From the current study findings | , we concluded that there is no $\begin{tabular}{c} \begin{tabular}{c} \end{tabular}$ |
|) rgical group p<0.001. | Conclusion : From the current study findings, | we concluded that there is no (|
|) rgical group p<0.001. | Conclusion : From the current study findings, we | concluded that there is no diffe |
|) rgical group p<0.001. | Conclusion : From the current study findings, we concluded | that there is no difference in sa |
| r nconspicuous per se. | Conclusion : In conclusion, our study demonstrated | that an integrated approach in |
| r nconspicuous per se. | Conclusion : In conclusion, our study demonstrated that | an integrated approach involvir |
| r nconspicuous per se. | Conclusion : In conclusion, our study demonstrated that an | integrated approach involving (|
| r nconspicuous per se. | Conclusion : In conclusion, our study demonstrated that an integrated | approach involving clinicians a 📗 |
| ? :M group (p = 0.005). | Conclusion : In conclusion, this study shows | that immunosuppressive treatr |
| ? :M group (p = 0.005). | Conclusion : In conclusion, this study shows that | immunosuppressive treatment |
| | Acceleration of the construction white calculated and a first terminal construction | to characteristic Territoria and Marka and E |

Figure 17. Collocations with the word 'study' filtered for the section CONCLUSION. From Kilgarriff et al. (2004)

The collocations extracted with the node 'study' filtered for the section CONCLUSIONS show different LFs that can be used in exercises. An example is the LF shown in Table 6, below:

| * | * | study | * | that |
|----------------------|------|-------|-----------------|------|
| - | The | | showed (68x) | |
| The results of (25x) | Our | | shows (48x) | |
| | This | | suggests (54x) | |
| | | | suggested (6x) | |
| | | | indicates (24x) | |
| | | | indicated (8x) | |

Table 6. LF with the node 'study'

As can be seen in Table 6, the LF *(**The, Our, This**) **study** *(**show(ed**), **suggests, indicates**) is a chunk of language that can be taught as an option to be used at the beginning of the section CONCLUSION(S) in structured abstracts in Health Sciences. 'The results of' precedes some of the sentences where this LF occurs. 'Showed' is the most recurrent slot filler after the collocation node 'study'. The procedure of filtering, shown in Figure 12, can be done with the other sections of structured abstracts to identify LFs to be

included in exercises with the LFs that are recurrent in different sections of structured abstracts.

Concluding remarks

As aforementioned, this chapter drew from the needs of Brazilian pre-service and in-service EAP novice teachers, graduate and undergraduate students from the Federal University of Rio Grande do Sul (UFRGS), all teachers at CLA (Center of Languages for Academic Purposes). While the COVID-19 pandemic obliged us to stay home for two years and two months, we held weekly online pedagogical meetings. During these meetings, we reported and reflected upon our online classroom experiences, to find solutions to problems that we had never faced before. Moreover, we discussed language learning and teaching theories. Finally, we planned courses and classes. However, above all, we tried to figure out how corpus linguistics and genre studies could guide us to design materials to help our students, the Brazilian academic community, to write more conventional academic texts. The insights that came up from these meetings guided the writing of this chapter.

During this period, we identified that novice EAP teachers were not confident using corpus linguistics to inform their teaching practice, even though this approach has been proved effective by many scholars. With this gap in mind, we created a framework drawing on the principles proposed by Welp et al. (2019) and adapted by Bocorny and Welp (2021) to design EAP materials combining corpus and genre-based pedagogies. In this chapter, we introduced a step-by-step guide to help teachers to retrieve and integrate corpus data into materials designed for EAP writing courses through indirect DDL. Moreover, we provided explanations and descriptions of actions for each of the five first principles. Besides exemplifying those actions, we had in mind an EAP writing course for producing Health Sciences structured abstracts.

The COVID-19 pandemic is now over (or so we believe), and we are back to on-site classes. Nevertheless, we are glad to say that we genuinely believe we have all become more skilled and knowledgeable teachers. Although we had a particular group of teachers in mind to produce this study, we believe that the insights it led to can be generalized. Even so, further studies could focus on work with a more significant sample of teachers, both from the secondary and tertiary levels. Above all, we expect this contribution will help to bridge the gap between corpus linguistics and EAP materials design.

Acknowledgement

For the shared route and mutual support, we wish to thank the CLA teachers for sticking together and supporting each other in a genuine case of scaffolding.

References

Aijmer, K. (2009). (Ed.) Corpora and language teaching. John Benjamins Publishing.

Anthony, L. (2003). *AntMover* (Version 1.1.0) [Computer Software]. Tokyo, Japan: Waseda University. https://www.laurenceanthony.net/software

Anthony, L. (2017). *AntGram* (Version 0.0.3) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.antlab.sci.waseda.ac.jp/

Anthony, L. (2019). *AntCorGen* (Version 1.2.0) [Computer Software]. Tokyo, Japan: Waseda University. https://www.laurenceanthony.net/software

Anthony, L. (2022a). International Perspectives on Corpus Technology for Language Learning - The University of Queensland Seminar Series). Addressing the challenges of data-driven learning through corpus tool design: An introduction to AntConc 4 [Video]. https://languages-cultures.uq.edu.au/event/session/8171

Anthony, L. (2022b). *AntConc* (Version 4.1) [Computer Software]. Tokyo, Japan: Waseda University. https://www.laurenceanthony.net/software

Araújo, A. D. (1999). Uma análise de organização discursiva de resumos na área de Educação. Revista do GELNE – *Grupo de Estudos Linguísticos do Nordeste, 1*, 26-30.

Atkins, S., Clear, J. & Oslter, N. (1992). Corpus design criteria. *Literary and Linguistic Computing*, *7*, 1-16.

Berdanier, C. G. (2019). Genre maps as a method to visualize engineering writing and argumentation patterns. *Journal of Engineering Education*, *108*(3), 377-393.

Bhatia, V. K. (1993). *Analysing Genre: Language Use in Professional Settings*. London: Longman.

Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multiword patterns in speech and writing. *International Journal of Corpus Linguistics*, 14 (3), 275-311.

Biber, D., Conrad, S. & Reppen, R. (1998). *Corpus linguistics Investigating language structure and use*. Cambridge: Cambridge University Press.

Biber, D. & Conrad, S. (1999). Lexical bundles in conversation and academic prose. In H. Hasselgard & S. OKSEFJELL (Eds.). *Out of Corpora: Studies in Honour of Stig Johansson* (pp. 181-190). BRILL.

Biber, D. & Conrad, S. (2009). *Register, genre, and style*. Cambridge: Cambridge University Press.

Bocorny, A. E. P. & Welp, A. K. D. S. (2021). O desenho de tarefas pedagógicas para o ensino de inglês para fins acadêmicos: conquistas e desafios da Linguística de Corpus. *Revista de estudos da linguagem, 29(2),* 1589-1638.

Bondi, M. (2022). Comparable corpora in cross-cultural genre studies: Tools for the analysis of CSR reports. *Corpus Linguistics and Translation Tools for Digital Humanities: Research Methods and Applications*, 37-63.

Boulton, A. (2007). But where's the proof? The need for empirical evidence for data-driven learning. In *Proceedings of the BAAL Annual Conference 2007*, p. 13-16.

Boulton, A. (2021). Research in data-driven learning. *Beyond Concordance Lines: Corpora in language education*, *102*, 9-34.

Boulton, A. & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language learning*, 67(2), 348-393.

Brezina, V., Weill-Tessier, P. & McEnery, T. (2020). #LancsBox 5.x and 6.x [software]. http://corpora.lancs.ac.uk/lancsbox

Breyer, Y. (2011). *Corpora in Language Teaching and Learning. Potential, Evaluation, Challenges.* Peter Lang.

Can, S., Karabacak, E. & Qin, J. (2016). Structure of moves in research article abstracts in applied linguistics. *Publications*, *4*(*3*), 23.

Charles, M. (2007). Reconciling top-down and bottom-up approaches to graduate writing: Using a corpus to teach rhetorical functions. *Journal of English for academic purposes*, 6(4), 289-302.

Charles, M. (2013). English for academic purposes. *The handbook of English for specific purposes*, 137-153.

Charles, M. (2020). Combining genre analysis and corpus consultation in class: Using do-it-yourself corpora to explore the literature review. *Approaches to Specialized Genres*, 243-258.

Charles, M. & Frankenberg-Garcia, A. (2021). Introduction: Dichotomies and debates in corpora and ESP/EAP writing. In M. Charles, A. & Frankenberg-Garcia (Eds.). *Corpora in ESP/EAP Writing Instruction* (pp. 1-10). Routledge.

Cheng, W. (2010). What can a corpus tell us about language teaching?. In M. McCarthy & A. O'Keeffe (Eds.) *The Routledge handbook of corpus linguistics* (pp. 319-332). Routledge.

Cortes, V. (2013). The purpose of this study is to: Connecting lexical bundles and moves in research article introductions. *Journal of English for academic purposes*, *12*(1), 33-43.

Cotos, E. (2014). *Genre-based automated writing evaluation for L2 research writing: From design to evaluation and enhancement.* New York, NY: Palgrave Macmillan.

Cotos, E., Haufman, S. & Link, S. (2017). A move/step model for methods sections: Demonstrating Rigour and Credibility. *English for Specific Purposes*, *46*, 90-106.

Fletcher, W. H. (2012). *kfNgram* (Version 1.3.1). Retrieved from http://kwicfinder. com/kfNgram/kfNgramHelp.html

Flowerdew, L. (2009). Applying corpus linguistics to pedagogy: A critical evaluation. *International journal of corpus linguistics*, *14*(3), 393-417.

Flowerdew, L. (2012). *Corpus and Language Education*. Basingstoke: Palgrave Macmillan.

Flowerdew, L. (2013). Corpus-based research and pedagogy in EAP: From lexis to genre. *Language Teaching*, 48(1), 99-116.

Flowerdew, L. (2014). Corpus-based analyses in EAP. In J. Flowerdew & C. Candlin (Eds.). *Academic discourse* (pp. 105-124). Routledge.

Francis, W. N. (1992). Language Corpora BC. In Svartvik, J. [ed.] *Directions in Corpus Linguistics. Proceedings of Nobel Symposium 82*, Stockholm. Berlin/ New York, p. 17-32.

Frankenberg-Garcia, A., Lew, R., Rees, G. Roberts, J.C., Sharma, N. & Butcher, P. (2021). *ColloCaid* (around 30 thousand academic English collocations and examples of collocations in context curated from corpora of expert academic English), open access at http://www.collocaid.uk/

Freitas, A. L. P. & Bocorny, A. E. P. (2021). How to write medical abstracts? The rhetorical structure and phrases used in Epidemiology. *Brazilian Journal of Sexually Transmitted Diseases*, 33, 1-6.

Gavioli, L (2005). *Exploring corpora for ESP learning* (pp. 1-176). Amsterdam: John Benjamins.

Gaspar, P. C., Santos, A. S. D. dos., Santana, L. B., Aragón, M. G., Machado, N. M. da S., López, M. A. A., Passos, M. R. L., Pereira, G. F. M. & Miranda, A. E. (2022). The fight against sexually transmitted infections cannot stop in the COVID-19 era: a brazilian experience in online training for sexually transmitted infections guide-lines. *Brazilian Journal of Sexually Transmitted Diseases*, 34. https://doi.org/10.5327/DST-2177-8264-20223404

Geluso, J. (2019). *Frequency, semantic, and functional characteristics of discontinuous formulaic language: A learner corpus study.* Master's dissertation, Iowa State University.

Gray, B. & Biber, D. (2013). Lexical frames in academic prose and conversation. *International journal of corpus linguistics*, *18*(1), 109-136.

Gray, B., Cotos, E. & Smith, J. (2020). Combining rhetorical move analysis with multi-dimensional analysis: Research writing across disciplines. Advances in corpus-based research on academic writing: Effects of discipline, register, and writer expertise, 137-168.

Henderson, A. & Barr, R. (2010). Comparing indicators of authorial stance in psychology students' writing and published research articles. *Journal of Writing Research*, *2*(2), 245-264.

Hyland, K. (2002). Authority and invisibility: Authorial identity in academic writing. *Journal of pragmatics*, *34*(8), 1091-1112.

Hyland, K. (2008). Academic clusters: text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, 18(1), 41–61.

Johns, T. (1991). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. *English Language Research Journal*, *4*, 27–45.

Karlsen, P. H. (2021). *Teaching and Learning English through Corpus-based Approaches in Norwegian Secondary Schools: Identifying Obstacles and a Way Forward*. Doctoral thesis. Inland Norway University of Applied Sciences.

Kanoksilapatham, B. (2005). Rhetorical structure of biochemistry research articles. *English for specific purposes, 24*(3), 269-292.

Kavanagh, B. (2021). Bridging the Gap from the Other Side: How Corpora Are Used by English Teachers in Norwegian Schools. *Nordic Journal of English Studies*, 20(1), pp.1–35. DOI: http://doi.org/10.35360/njes.522

Kennedy, G. (1998). An introduction to Corpus Linguistics. New York: Longman.

Kilgarriff, A., Rychlý, P., Smrž, P. & Tugwell, D. (2004) The sketch engine. *Proceedings* of the 11th EURALEX International Congress: 105-116.

Le, T. N. P. & Harrington, M. (2015). Phraseology used to comment on results in the discussion section of applied linguistics quantitative research articles. *English for Specific Purposes*, *39*, 45-61.

Martínez, I. A. (2005). Native and non-native writers' use of first person pronouns in the different sections of biology research articles in English. *Journal of second language writing*, *14*(3), 174-190.

McEnery, T. & Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.

McEnery, T., Xiao, R. & Tono, Y. (2006). *Corpus-Based Language Studies*. USA/Canada: Routledge.

McEnery, T. & Wilson, A. (1997). Teaching and Language Corpora. *ReCALL*, 9(1), 5-14.

Meurer, J. L. (1997). Esboço de um modelo de produção de textos. *In* J. L. Meurer & D. Motta-Roth (Eds.). *Parâmetros de textualização*. (pp. 14- 27). Santa Maria: Editora da UFSM.

Moreno, A. I. & Swales, J. M. (2018). Strengthening move analysis methodology towards bridging the function-form gap. *English for Specific Purposes*, *50*, 40-63.

Motta-Roth, D. (1995). *Rhetorical Features and Disciplinary Cultures: A Genre-Based Study of Academic Book Review in Linguistics, Chemistry and Economics*. Doctoral thesis. Universidade Federal de Santa Catarina.

Mukherjee, J. (2004). Bridging the gap between applied corpus linguistics and the reality of English language teaching in Germany. In U. Connor & T. A. Upton, T. A. (Eds.). *Applied Corpus Linguistics*, 239-250. Amsterdam: Rodopi.

O'Keeffe, A. (2022). Data-driven learning and second language acquisition – it's time to connect. [Video]. School of Languages and Cultures. https://languages-cultures. uq.edu.au/event/session/7987

O'Keeffe, A., McCarthy, M. & Carter, R. (2007). *From corpus to classroom: language use and language teaching*. Cambridge: Cambridge University Press.

Pérez-Llantada, C. (2022). Online Data Articles: The Language of Intersubjective Stance in a Rhetorical Hybrid. *Written Communication*. https://doi. org/10.1177/07410883221087486

Pérez-Paredes, P. (2019). The pedagogic advantage of teenage corpora for secondary school learners. *Data-Driven Learning for the Next Generation*, 67-87.

Poole, R. (2020). "Corpus can be tricky": revisiting teacher attitudes towards corpus-aided language learning and teaching. *Computer Assisted Language Learning*, 1–22. doi:10.1080/09588221.2020.1825

Reppen, R. (2010). Using corpora in the language classroom. Cambridge University Press.

Römer, U. (2006). Pedagogical Applications of Corpora: Some Reflections on the Current Scope and a Wish List for Future Developments. *Zeitschrift für Anglistik und Amerikanistik*, 54(2), 121-134. https://doi.org/10.1515/zaa-2006-0204

Römer, U. (2010). Establishing the phraseological profile of a text type: The construction of meaning in academic book reviews. *English Text Construction*, *3*(1), 95-119.

Santos, A. R. (1999). *Metodologia científica: a construção do conhecimento*. Rio de Janeiro: DP & A Editora.

Schneuwly, J. & Dolz, B. (2004). *Gêneros orais e escritos na escola*. Campinas: Mercado de Letras.

Scott, M. (2020). WordSmith Tools version 8. Stroud: Lexical Analysis Software.

Shepherd, T. M. (2009). O Estatuto da Linguística de *Corpus*: Metodologia ou Área da Linguística? *Matraga*, *16*(24), 150-172.

Sinclair, J. (1987). *Collins Cobuild English Language Dictionary: Helping Learners with Real English.* Heinle ELT.

Sinclair, J. (1991). Corpus, Concordance, Collocation. Oxford: OUP.

Sinclair, J. (2004a). 'Introduction'. *In* J. M. Sinclair (Ed.) *How to Use Corpora in Language Teaching* (pp. 1-13). Amsterdam and Philadelphia: John Benjamins.

Sinclair, J. (2004b). *Trust the text: Language, Corpus and Discourse.* London/ New York: Routledge.

Stevenson, H. A. & Harrison, J. E. (2009). Structured abstracts: Do they improve citation retrieval from dental journals?. *Journal of orthodontics*, *36*(1), 52-60.

Swales, J. (1981). *Aspects of article introductions*. Birmingham: Language Studies Unit. University of Aston [Aston ESP Research Reports 1].

Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.

Swales, J. (1994). The writing of research articles introduction. *Written Communication*, 4(2), 175-191.

Swales, J. (2004). *Research genres: Exploration and applications*. Cambridge: Cambridge University Press.

Swales, J. & Feak, C. B. (2009). *Abstracts and the writing of abstracts* (Vol. 2). University of Michigan Press ELT.

Viana, V., Bocorny, A. & Sarmento, S. (2018). *Teaching English for Specific Purposes*. *ELT Development Series*. TESOL Press.

Welp, A. K., Didio, Á. R. & Finkler, B. (2019). Questões contemporâneas no cinema e na literatura: o desenho de uma sequência didática para o ensino de inglês como língua adicional. *BELT-Brazilian English Language Teaching Journal*, *10*(2), e35861-e35861.

Yoon, J. & Casal, J. E. (2020a). P-frames and rhetorical moves in applied linguistics conference abstracts. *Advances in corpus-based research on academic writing: Effects of discipline, register, and writer expertise*, 95, 282-305.

Yoon, J. & Casal, J. E. (2020b). Rhetorical structure, sequence, and variation: A step-driven move analysis of applied linguistics conference abstracts. *International Journal of Applied Linguistics*, 30(3), 462-478.

Appendix I - Checklist for planning and designing an EAP course using a framework that combines corpus and genre-based pedagogies

| | Know learners' language proficiency level |
|-------------------------------|--|
| | Know learner's level of instruction or position (e.g. |
| | undergraduate, graduate master, graduate doctor's, |
| | professor) |
| Information about learners | Know discipline learner works with |
| | Know learners' needs |
| | Know learners' wants |
| | Know learners' expectations |
| | Select the target genre |
| | Select the target section (may not apply) |
| Information about the | Select the target skill(s) |
| course | Know how many and which disciplines (multiple or |
| | single) you will be working with |
| Dianning the source | Set learning objectives |
| Planning the course | Select methodology and approach |
| Select materials | Find existing materials |
| | Find the target-genre rhetorical structure in the liter- |
| | ature or describe it |
| | Decide which language features are worth working |
| | within the academic context in which the target |
| Design materials that are | genre is used and considering all the previously col- |
| corpus-based, genre (sec- | lected information |
| tion) and discipline specific | Compile a genre (section) and discipline specific |
| | corpus |
| | Extract language data from the corpus |
| | Use said language data to design tasks, exercises, ac- |
| | tivities within the context of the target genre |

Appendix II - Example of completed checklist for the course **Written Production of Health Sciences Structured Abstracts**

| | Language proficiency level | B2, C1 |
|---|--|--|
| Information about learners | Learner level of instruction or position (e.g. undergraduate, graduate master, graduate doc- tor's, professors) | Graduate students |
| | Discipline, specialty learners works with | Health sciences |
| | Target genre | Structured abstracts |
| Information about the course | Target section (may not apply) | Background and ob- jectives, method, re- sults, conclusion |
| | Target skill(s) | Written production |
| | Discipline (multiple or single) | Single discipline |
| Rhetorical structure of the target genre | Found in the literature or de- scribed by the teacher | Described by the teacher |
| Language feature(s) worth working within the context of the target genre | Lexical Frames | The first LF after the section name |
| Methodology | Combination of corpus and genre-based approaches | |