

DESIGNING A PHRASE BANK FOR ACADEMIC LEARNING AND TEACHING: A EUROPEAN PORTUGUESE CASE STUDY

Micaela AGUIAR¹ , Sílvia ARAÚJO² 

Article history: Received 30 January 2024; Revised 4 July 2024; Accepted 15 October 2024; Available online 10 December 2024; Available print 30 December 2024.

©2024 Studia UBB Philologia. Published by Babeş-Bolyai University.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ABSTRACT. *Designing a Phrase Bank for Academic Learning and Teaching: A European Portuguese Case Study.* This work focuses on the development and application of an academic phrase bank for European Portuguese. Our goal is to present an overview of this resource, discuss its potential replication in other languages, and explore its application in the classroom for teaching academic language. As the first of its kind resulting from corpora-based research, our phrase bank represents a contribution to the emerging field study of academic discourse in Portugal and a useful resource for students, teachers, and researchers. The paper begins by outlining the process of constructing the

-
- ¹ **Micaela AGUIAR** is a researcher at the Centre for Humanistic Studies at the University of Minho. She earned her PhD in Language Sciences with a specialization in Discourse Analysis in 2021 from the same institution. Her doctoral research explored the construction of “Presidential images in inauguration speeches over 100 years of the Portuguese Republic.” After completing her PhD, she contributed to the PortLinguE project at the University of Minho as a contracted researcher. This initiative focused on developing a Multilingual Portal for Specialized Languages, leveraging open data for cross-linguistic information retrieval. Email: maguiar60@gmail.com.
- ² **Sílvia ARAÚJO** is an Associate Professor in the Department of Romance Studies at the University of Minho, with expertise in Language Sciences and Digital Humanities. She has spearheaded major FCT-funded projects, including “Per-Fide: Portuguese with Six Different Languages” and “PortLinguE: Multilingual Portal for Specialized Languages.” A prolific scholar, she has numerous indexed publications and supervises master’s and PhD research in areas such as corpus linguistics, natural language processing, and AI in education. She also plays a key role in academic activities, (co)organizing conferences, (co)editing journals and books, and serving as a reviewer for various academic platforms. She directs the Master’s Degree in Digital Humanities and leads the Digital Humanities Research Group at the Center for Humanistic Studies. Additionally, she offers accredited training on integrating technology into education through the University of Minho’s Center for Innovation and Development of Teaching and Learning. Email: saraujo@elach.uminho.pt.

phrase bank, providing insights into the selection and categorization of formulaic academic phrases. We explore its accessibility options, including a web version and integration with a search engine powered by a Large Language Model. The paper then focuses on the findings and presents practical implementations of the phrase bank in educational settings. Outlined are two classroom activities incorporating the phrase bank, undertaken by master's students. The first activity entailed the revision of a group portfolio's introduction, employing the phrase bank to heighten awareness of communicative functions and expressions within introductions. The second activity involved peer assessment, where students evaluated their peers' literature reviews. This involved identifying communicative functions, analyzing, and rephrasing expressions related to authors and studies using the phrase bank. To facilitate the assessment, students utilized a choice board to offer constructive and considerate feedback to their classmates.

Keywords: *academic writing, academic phrase bank, academic writing assistant, teaching strategies, improving classroom teaching*

REZUMAT. Proiectarea unei bănci de fraze pentru învățarea și predarea academică: un studiu de caz portughez european. Acest articol prezintă o înțelegere cuprinzătoare a dezvoltării și aplicării unei bănci de expresii academice concepută special pentru limba portugheză europeană. Scopul nostru este să oferim o prezentare generală a acestei resurse, să discutăm posibilitatea replicării sale în alte limbi și să explorăm aplicarea sa în sala de clasă pentru predarea limbajului academic. Ca primul de acest fel, rezultat al unei cercetări extinse, banca noastră de expresii reprezintă o contribuție la domeniul emergent al studiului discursului academic din Portugalia și o resursă utilă pentru studenți, profesori și cercetători. Articolul începe prin conturarea procesului de construire a băncii de expresii, oferind perspective asupra selecției și categorizării expresiilor academice standardizate. Explorăm opțiunile sale de accesibilitate, inclusiv o versiune web și integrarea cu un motor de căutare alimentat de un Model de Limbaj Extins. Apoi, articolul își îndreaptă atenția către implementarea practică a băncii de expresii în mediile educaționale. Sunt descrise două activități de clasă care integrează banca de expresii, realizate de studenți de master. Prima activitate a implicat revizuirea introducerii unui portofoliu de grup, utilizând banca de expresii pentru a accentua conștientizarea funcțiilor de comunicare și a expresiilor din introduceri. A doua activitate a implicat evaluarea colegilor prin intermediul unui proces de evaluare reciprocă, în care studenții au analizat revizuirile literare ale colegilor. Acest proces a inclus identificarea funcțiilor de comunicare, analizarea și reformularea expresiilor legate de autori și studii cu ajutorul băncii de expresii. Pentru a facilita evaluarea, studenții au folosit o placă de opțiuni pentru a oferi feedback constructiv și considerat colegilor lor.

Cuvinte-cheie: *scriere academică, banca de fraze academice, asistent de scriere academică, strategii de predare, îmbunătățirea predării la clasă*

Contextualization

Many students in higher education struggle with academic writing. While the issue is widely recognized worldwide (Defazio et al. 2010), the situation takes on a specific context in Portugal. Access to higher education in Portugal became more widely available only after the dictatorship ended on April 25, 1974. During the 1960s, Portugal grappled with a substantial illiteracy rate, reaching 33.1% among the population, being particularly higher among women, where only 1 in 4 possessed reading and writing skills (Pordata). Despite substantial investments in literacy programs, Portugal still lagged other European countries in 2020, ranking last among the 23 OECD nations (Barros and Volta e Pinto 2022). In 2023, Portugal recorded its highest-ever enrollment in higher education, boasting a total of 446,000 students (Observador 2023). The substantial rise in student enrollment has presented a challenge for higher education institutions and teachers (Ferrão Tavares and Pereira 2012), who find themselves unprepared to foster the development of both disciplinary and transdisciplinary academic language skills among students.

Several researchers point out that one of the main drivers of the lack of academic skills among students is the deficiencies in language education at earlier levels (Gouveia 2019). Language curricula frequently prioritize literary texts over academic writing (Carvalho 2014), which can result in inadequate preparation for the challenges of higher education. By introducing academic writing skills and practices at the primary and secondary levels, students can be better equipped to fulfill the requirements of higher education. However, implementing such changes can be difficult as it requires significant governmental educational reforms.

Some higher education institutions have sought to tackle this challenge by providing courses in academic writing. Nevertheless, not all institutions extend such offerings, and those that do may impose fees, leading to a disparity in access to academic literacy skills (Preto-Bay 2004). The use of language is a critical factor in determining whether students can participate fully in academic communities. In fact, as Hyland (2018) points out, gaining competence in specialized discourses is the means through which newcomers and outsiders attain entry into these communities. Those who lack the necessary academic literacy skills may find it difficult to understand and engage with the academic discourse, which can limit their opportunities for academic and professional advancement (Short and Fitzsimmons-Doolan 2007). Ensuring equitable access to academic literacy skills instruction is imperative for all students, regardless of background or financial means, to succeed in higher education and beyond.

Given this backdrop, it is understandable that academic discourse studies in Portugal are at an early stage of development, with significant strides being made only in recent years, as highlighted by the organization of the first National Meeting on Academic Discourse (ENDA) in 2019. Just like the field it is grounded in — English for Academic Purposes (EPA) — academic discourse researchers in Portugal come predominantly from applied linguistics, but also include researchers from the spheres of academic literacy and cultural studies, delving into discourses spanning from higher education and teacher training to earlier levels of education (Gouveia 2019).

Still, given English's dominance as the primary language of science, the question arises: why study and create resources for academic language in Portuguese? Upon consulting national repositories, with exceptions for fields like medicine and computer science, it becomes apparent that the predominant language for most master's dissertations and doctoral theses continues to be Portuguese. Moreover, a noteworthy trend emerges as the enrollment of foreign students in higher education has seen a substantial twofold increase over the past six years (Leiria 2023). Firstly, the discourse on the crucial affirmation of Portuguese as a language of science has never been more pronounced, particularly in an international (and even national) scientific environment markedly influenced by English (Assunção 2019, 190). Secondly, a definitive stride in advancing Portuguese as a language of science involves dedicating efforts to construct resources that enhance scientific, academic, and technical literacy. This investment is essential for higher education students destined to become future researchers and contributors to scientific knowledge.

With the goal of being replicated in various languages, this paper seeks to outline the process of construction, development, and pedagogical implementations of an academic phrase bank in European Portuguese. As mentioned above, given the emerging nature of the field in academic discourse studies, there is a lack of resources focused on academic language in European Portuguese. This recourse stands as a valuable contribution to the field, supporting educators in teaching academic writing while also functioning as a standalone reference for native Portuguese speakers and those learning Portuguese as a second or foreign language. This is one element of the wider PortLinguE research project (PTDC/LLT-LIG/31113/2017), which is backed by European funding. The project focuses on developing a range of tools for specialised languages, including academic language, with the aim of using open research data from national institutional repositories to create free and accessible educational resources for the academic community.

The paper is structured in the following way. The background section presents an introduction to the central concepts discussed, specifically academic formulaic language and transdisciplinary scientific lexicon. The subsequent section outlines the various phases involved in constructing the phrase bank. Following that, we delve into the efforts to make the resource accessible, both in web format and as an academic writing assistant, driven by Large Language Models. Moving on, we explore the practical applications of this resource in the classroom, detailing two activities implemented with MA students. Finally, we conclude the paper by summarizing the study's findings and outlining potential areas for future research.

Why focus on academic formulaic language?

Wray (2002) suggests that native speakers are inclined to use formulaic language, asserting that these expressions are essential for communication to be effective. In academic communication, a large part of language is formulaic. Indeed, Hyland (2008) highlights that approximately 21% of the 5.3 million words in the academic section of the Longman Spoken and Written English corpus consist of formulaic language.

When it comes to academic language learning, Li and Schmitt (2009), among others, note that the absence of these expressions could be a sign of a lack of experience in using the language in academic contexts. The problem is that using academic formulas is not a natural skill for native speakers (Pérez-Llantada 2014). For L2 speakers, overcoming this can significantly impede achieving native-like proficiency.

Speakers' inclination toward using formulaic language is evident and widespread, yet studying and pinpointing it proves to be a complex task. Wray (2009) suggests that embracing a "fuzzy" definition of formulaic language allows for greater inclusivity. Consequently, our goal is not to delve into theoretical debates regarding the definition of formulaic language, but rather to adopt a definition that guides our extraction and categorization efforts.

We opted to identify academic language that extends beyond disciplinary borders, aiming to construct a phrase bank, that can be valuable to a diverse audience within the academic and scientific community. For that purpose, the concept of a transdisciplinary scientific lexicon (Drouin 2012) stands out for its interest, referring to the non-terminological lexicon widely used across diverse scientific disciplines and the scientific community. According to Drouin (2012), this concept includes diverse types of expressions, ranging from simple, referential polylexical sequences (like collocations and fixed expressions) to discourse polylexical sequences (which involve recurrent

expressions employed to organize speech) and interpersonal polylexical sequence (expressions used for conveying pragmatic or modal functions). It also includes semantic-rhetorical routines, described by Toutin and Jacques (2018) as typical scientific writing utterances that serve specific rhetorical functions.

Semantic-rhetorical routines showcase a higher degree of semantic, syntactic, and lexical complexity and diversity in comparison to other types of expressions. These units frequently take the form of complete utterances centred around a predicate. Unlike, for instance, lexical bundles (Biber and Barbieri 2007; Hyland 2008; and Hyland and Jiang 2018) these units are not easily identified using corpus analysis software. It's tough to extract these expressions (semi-) automatically because of their complexity and diverse forms; however, we have already made efforts in this direction (Aguiar et al. 2023). For these reasons, we opted to gather these types of expressions along with the functions they serve within a corpus of scientific papers. This corpus laid the foundation for creating the academic phrase bank, the construction process of which will be detailed in the next section.

Creating the phrase bank

For those looking to boost the quality of their academic writing, an academic phrase bank could be a helpful resource for students, researchers, and scholars. These linguistic resources are organised into large categories that are typically linked to the IMRAD structure (Introduction, Methodology, Results, and Discussion). The phrases in the bank are commonly used in academic discourse to serve various purposes, such as reporting the results of a study, describing the sample of an experiment, introducing a topic, or summarising findings.

Students can find various online academic phrase banks to improve their writing skills. For instance, the Ref-N-Write Academic Phrasebank helps with English academic writing, while the Dictionnaire des expressions from Base ARTES provides a multilingual phrase bank catering to several languages. Portuguese writers can find a multilingual Portuguese phrase bank at Bab.la.

Our phrase bank is unique as it is, to our knowledge, the first-ever European Portuguese phrase bank that has been developed based on corpora-driven scientific research work. This means that our phrase bank is a result of an analysis of authentic academic texts and offers a comprehensive and reliable resource for writers to enhance the quality of their academic discourse in European Portuguese.

Background

Our phrase bank was modeled after the University of Manchester's Academic Phrasebank, which was also established through research efforts. Developed by John Morley in 2004, the Manchester Phrasebank remains one of the most widely used resources for academic English.

What began as a collection of 100 postgraduate dissertations from the University of Manchester, the Academic Phrasebank corpus now includes academic materials from various other sources. Swales' (1990) genre analysis, which highlights rhetorical patterns in research article introductions, plays a significant role in shaping the Academic Phrasebank. His concept of a "move" — a section of text that serves a distinct communicative function — is a fundamental organizing principle within the Phrasebank's framework. Morley gathered the phrases used to make these rhetorical moves and subsequently simplified them by removing their specific contextual content.

In line with Morley's methodology, we have developed a three-step approach, which includes compiling a corpus of authentic academic texts, identifying, categorising, and extracting the rhetorical moves and expressions, and then simplifying them for ease of use.

Corpus

We began by compiling a corpus of 40 research papers from four scientific areas. This included 10 articles each from Life and Health Sciences, Exact and Engineering Sciences, Natural and Environmental Sciences, and Social Sciences and Humanities. The corpus was restricted to 10 samples from each area due to the time constraints of the research project. The objective of this work was to organize the phrase bank in a way that would facilitate future expansion with additional scientific articles and other academic-scientific genres. These papers were randomly collected from RepositóriUM, the repository of the University of Minho, and from Repositório Aberto, the repository of the University of Porto. Our selection criteria required the articles to be available on an open access basis, classified as scientific papers in the repositories, and written in European Portuguese.

To align with the pedagogical purpose of the phrase bank, we used scientific articles as our corpus, as they are a widely recognized genre in academic research and can serve as an example for up-and-coming researchers and undergraduate students. While the current corpus of scientific articles provides a useful foundation for our work, we recognize that academic writing

encompasses a much broader range of genres. We intend to enhance the corpus in the future by incorporating various academic genres, such as dissertations, theses, research projects, and defences. Different academic genres present distinct linguistic and discursive patterns, each with its own conventions. By incorporating these into the corpus, we help users build a stronger grasp of effective communication and writing within diverse academic contexts.

Categorization

We adapted Morley's typology, which was originally designed for theses and dissertations, to develop our methodology, since as Toutin and Jacques (2018) suggest, the scientific lexicon found in academic communication is genre dependent. We grouped the phrase into six main categories: introduction, literature review, methodology, results, discussion, and conclusions. Rastier's (2001) concept of included genres informs these categories, where included genres refer to text sections, like introductions or conclusions, commonly found in academic works such as scientific papers, doctoral theses, or master's dissertations. We chose these categories because we aim to create a pedagogical tool that not only presents academic phrases but also teaches students about the most common functions of each category.

Annotation of the corpus was carried out by the authors using MAXQDA, a tool that supports both mixed methods and quantitative analysis. We referred to Morley's proposed questions to determine whether the phrases should be included: (1) Does it serve a communicative purpose in the academic text? (2) Does it contain formulaic expressions or collocations? (3) Are the content words generic, and does the combination sound natural to a native speaker? (Morley 2004) The analysis of the corpus allowed us to identify the key rhetorical moves used in each of the six main categories. In total, we found 51 rhetorical moves that were used to achieve specific communicative purposes across the six categories: introduction (8 moves), literature review (8 moves), methodology (13 moves), results (10 moves), discussion (7 moves), and conclusion (5 moves). From there, we extracted the phrases used to accomplish each move, resulting in a dataset of 939 phrases.

The Introduction section (Table 1) is used to explain the importance of the topic in general, identify gaps in knowledge or scarcity of previous studies, point out controversies, state the objective and object of the work, indicate research questions, explain the importance of the study, and outline the structure of the work.

Table 1. Introduction's Moves with Examples

Introduction's Moves	Example
Explain the topic's importance (in general)	A preocupação com a problemática de x é cada vez mais crescente... [The concern with the issue of x is growing increasingly...]
Explain the lack of previous studies or gaps in knowledge	Uma das razões que pode explicar a falta de tratamento que x tem recebido na literatura é o facto de y... [One of the reasons that may explain the lack of attention x has received in the literature is the fact that y...]
Identify controversies or lack of consensus in the field of study	Embora exista consenso quanto a x, é ainda polémica a definição de y... [Although there is consensus regarding x, the definition of y remains controversial...]
Identify the aim of the work	Foi objetivo deste trabalho proceder a x... [The objective of this work was to carry out x...]
State the object of study	É sobre este objeto que nos debruçaremos de forma particular... [It is on this subject that we will focus in detail...]
Indicate the research questions	Colocámos como hipóteses de trabalho que... [We proposed as working hypotheses that...]
Explain the importance of the study	Este estudo poderá ter uma contribuição importante para x... [This study could make a significant contribution to x...]
Indicate the structure of the work	Este artigo está estruturado da seguinte forma... [This article is structured as follows...]

In the Literature Review (Table 2), writers provide general comments on the literature, identify consensus/disagreement, mention studies in chronological order or with particular focus, reference methods and methodologies, designate objects or topics previously studied, refer to ideas of specific authors, and indicate the position of specific authors.

Table 2. Literature Review's Moves with Examples

Literature Review's Moves	Example
Make general comments on the literature	A investigação neste campo também tem apontado para a existência de x... [Research in this field has also pointed to the existence of x...]
Identify consensus/disagreement in the literature	De forma semelhante, Smith et al. (2000) descreveram que... [Similarly, Smith et al. (2000) described that...]
Refer to studies focusing on chronology	Estudos realizados na última década relevam que... [Studies conducted in the last decade reveal that...]
Mention specific studies	Estudos anteriores demonstraram que x... [Previous studies have demonstrated that x...]
Refer to previously proposed methods and methodologies	Existem descritos na literatura inúmeros métodos para x... [Numerous methods for x are described in the literature...]
Mention objects or topics previously studied	Estudos demonstraram que a frequência de x é relativamente baixa... [Studies have shown that the frequency of x is relatively low...]
Refer to the ideas of specific authors	Smith (1984) e Jones (2010) descrevem as várias etapas de x... [Smith (1984) and Jones (2010) describe the various stages of x...]
Indicate the positions of specific authors	É esta a posição assumida por Smith (2008) ... [This is the position taken by Smith (2008) ...]

The Methodology section (Table 3) is used to identify the type of study, indicate the use of methods established by specific authors, use standards and recommendations, indicate methodological limitations, point out inclusion/exclusion criteria, characterise the sample, describe procedures for data collection and questionnaires, indicate the purpose of a procedure, describe instruments used, describe actions performed, and describe statistical procedures.

Table 3. Methodology's Moves with Examples

Methodology's Moves	Example
Identify the type of study conducted	O estudo que realizámos é observacional... [The study we conducted is observational...]
Indicate the use of methods established by specific authors	O sistema de pontuação utilizado foi uma adaptação do método de XYZ utilizado por Smith e Jones (2000) ... [The scoring system used was an adaptation of the XYZ method employed by Smith and Jones (2000) ...]
Indicate the use of standards and recommendations	Os autores adotaram as normas XYZ para a redação do presente estudo... [The authors adopted XYZ standards for the writing of this study...]
Indicate the methodological limitations of the study	Uma das suas principais limitações prende-se com... [One of its main limitations relates to...]
Outline the inclusion/exclusion criteria of the study	Como critérios de exclusão para o estudo considerámos... [As exclusion criteria for the study, we considered...]
Characterize the sample	A amostra deste estudo descritivo foi de conveniência... [The sample for this descriptive study was convenience-based...]
Describe procedures related to data collection	A recolha de dados decorreu durante x, no período do mês y a mês z... [Data collection took place during x, from month y to month z...]
Describe procedures related to questionnaires	É um instrumento composto por 26 itens que tem como objetivo medir y... [It is an instrument consisting of 26 items designed to measure y...]
Indicate the purpose of a procedure	Para medir este efeito, calculou-se a métrica x para cada um dos atributos... [To measure this effect, the metric x was calculated for each of the attributes...]
Describe the instruments used	Realizou-se um teste piloto através da simulação de uma entrevista... [A pilot test was conducted through the simulation of an interview...]
Describe the actions taken	Os procedimentos foram previamente uniformizados... [The procedures were standardized in advance...]
Describe sequences of actions	Numa primeira fase, os ensaios foram realizados... [In the initial phase, the trials were conducted...]
Describe statistical procedures	Efetúámos a análise estatística com recurso ao software... [We performed the statistical analysis using the software...]

In the Results section (Table 4), the authors carefully present the study's findings, highlighting both positive and negative outcomes, indicate trends in results, report survey results, refer to percentages, emphasise aspects, give examples or excerpts of analysis, and present data in tables, charts, or figures.

Table 4. Results' Moves with Examples

Results' Moves	Example
Indicate study results	Estes resultados mostram que... [These results show that...]
Present results with caution	Os resultados obtidos sugerem que... [The results obtained suggest that...]
Indicate negative results	Conclui-se que não existiu diferença estatisticamente significativa em x... [It is concluded that there was no statistically significant difference in x...]
Indicate positive results	Verifica-se uma correlação positiva entre x e y... [A positive correlation between x and y is observed...]
Point out trends in the results	Os resultados demonstraram uma tendência decrescente... [The results demonstrated a decreasing trend...]
Report questionnaire/survey results	Os participantes foram unânimes em considerar que... [The participants were unanimous in considering that...]
Refer to percentages in the results	Verificou-se que em x % da amostra considera... [It was found that x% of the sample considers...]
Highlight an aspect	Um outro aspeto importante a ter em conta é... [Another important aspect to consider is...]
Provide examples/excerpts of analysis	Os seguintes excertos são exemplificativos... [The following excerpts are illustrative...]
Present data in tables/figures/graphs	A Figura 1 apresenta os resultados de... [Figure 1 presents the results of...]

The Discussion section (Table 5) is used to compare results obtained with results from other studies, comment on results, summarise research questions, indicate unexpected results, provide possible explanations, state hypotheses, and make recommendations for new practices or policies.

Table 5. Discussion's Moves with Examples

Discussion's Moves	Example
Compare the results obtained with results from other studies	Estes resultados estão de acordo com alguns estudos realizados... [These results are consistent with some studies conducted...]
Comment on the results obtained	Esta observação é interessante à luz de... [This observation is interesting in light of...]
Revisit research questions	Foram estas as perguntas às quais se pretendeu dar resposta... [These were the questions we aimed to answer...]
Indicate unexpected results	Esperavam-se correlações mais significativas... [More significant correlations were expected...]
Provide possible explanations for the obtained results	Estes resultados podem ser explicados por... [These results may be explained by...]
Present hypotheses	Coloca-se a hipótese de... [The hypothesis is raised that...]
Make recommendations for new practices or policies	Os resultados obtidos devem incentivar a implementação de... [The results obtained should encourage the implementation of...]

Finally, the Conclusions section (Table 6) is used to emphasise the importance of the study, reaffirm objectives, summarise main conclusions, acknowledge limitations, and make recommendations for future work.

Table 6. Conclusions' Moves with Examples

Conclusions' Moves	Example
Emphasize the importance of the study	Este estudo surge como "inovador" ao centrar-se em... [This study is considered "innovative" as it focuses on...]
Reaffirm the study's objectives	O presente estudo teve como principal objetivo... [The main objective of this study was to...]
Summarize the main conclusions of the study	Em conclusão, os aspetos mais prementes identificados neste estudo foram... [In conclusion, the most pressing aspects identified in this study were...]
Acknowledge the study's limitations	As principais limitações do presente estudo são o facto de... [The main limitations of this study are that...]
Make recommendations for future work	Futuramente, serão de grande utilidade estudos que incidam sobre... [In the future, studies focusing on... will be highly valuable]

Sifting

Once the phrases were extracted, they were simplified by removing specific content or replacing it with more general terms. This revision helped make the phrases more accessible for students, ensuring they could easily incorporate them into their writing. The main goal of this process was to emphasise the phraseology element and facilitate its use by students in their writing, while simultaneously discouraging plagiarism. Therefore, where necessary, the phrases were carefully sifted from their original academic context and, in some cases, content words were replaced with substitutes for exemplification purposes. Here, we can observe the distinction between the original sentence in 1 and the simplified version provided in 2.

1. *Uma das razões que pode explicar a falta de tratamento* que os róticos têm recebido na literatura do Português Europeu é o facto de, efetivamente, serem sons problemáticos.
[*One of the reasons that may explain the lack of treatment* that rhotic sounds have received in European Portuguese literature is that they are, in fact, problematic sounds.]
2. *Uma das razões que pode explicar a falta de tratamento* que x tem recebido na literatura é o facto de y
[*One of the reasons that may explain the lack of treatment* that x has received in the literature is the fact that y]

As the examples that follow highlight, we used placeholder names for proper nouns (3) and substituted content words with unknowns (4), such as x, y, and z, whenever necessary.

3. *Smith e Jones (2015) sublinham que...*
[*Smith and Jones (2015) point out that...*]
4. *x foi avaliado com base na especificação europeia...*
[*x was assessed using the European specification...*]

Accessing the phrase bank

The online platform Lang2Science provides access to the phrase bank. Our project is dedicated to creating various resources to improve academic and specialized language skills and we aim to make these resources widely available. The phrase bank is available in two formats: a web format for browsing and a searchable format for quick and easy retrieval of specific phrases.

Web format

The web format of the phrase bank on the Lang2Science online platform is designed to provide users with an easy and user-friendly way to access and browse the database of academic phraseologies. The web format includes a well-organised and structured interface, featuring an intuitive navigation system that will allow users to browse through the various categories and subcategories of the phrase bank. The categories are organised in a hierarchical manner, with the main categories such as introduction, methodology, and conclusion being easily accessible from the main page.

Users can find the relevant rhetorical moves and their corresponding phrases within each category. The phrases are displayed in an intuitive format, ensuring users can quickly identify the one that works best for them.

With this format, users can dive into the phrase bank and discover not only the different phrases and their usage but also gain insights into the macro-structure of scientific articles and the types of rhetorical moves typical to each section of this genre.

Apresentação da etapa	O que é?	+ Explicar a importância do tópico (em geral)
Banco de Frases	Introdução	+ Explicar a escassez de estudos anteriores ou lacunas no conhecimento
	Revisão da Literatura	+ Identificar controvérsias ou falta de consenso na área de estudo
	Metodologia	+ Indicar o objetivo do trabalho
	Resultados	+ Indicar o objeto de estudo
	Discussão	+ Indicar as questões de investigação
	Conclusão	+ Explicar a importância do presente estudo
		+ Indicar a estrutura do trabalho

Figure 1. Web-format of the Academic Phrase Bank

Academic Writing Assistant

The searchable format of our phrase bank is available through our Academic Writing Assistant, which simplifies the process of searching through our phrase bank. The Writing Assistant is the name we've given to one of the

functions of a search engine we developed within our research project (Araújo et al. 2023). At the heart of our Academic Writing Assistant, is the BERT (Devlin 2019) deep learning language model. BERT, short for Bidirectional Encoder Representations from Transformers, is a widely used pre-trained model in Natural Language Processing, created by Google’s AI team. It handles tasks like question answering, summarization, and sentence prediction. BERT’s ability to understand context, trained through masked language modeling and next sentence prediction (Lutkevich 2020), makes it effective for semantic search, which interprets user intent.

Keyword-based searches dominate most phrase banks, which restrict results to content with the exact words provided by the user. For example, if a user searches for the phrase “the authors highlight that,” the results will only show instances of the word “authors” in its singular or plural form. However, by utilizing the BERT model, the Writing Assistant understands the context of the user’s query and can provide the most relevant phrases, even if they don't match the exact words used. This helps users with limited academic vocabulary by offering a wider range of academic language options, ultimately improving their writing. Figure 2 depicts the engine’s interface and organisation, which are user-friendly.

autor	REVISÃO DE LITERATURA ▾	TUDO ▾
<p>Defende o autor que... Categoria: Revisão da Literatura Sub-Categoria: Referir ideias de autores em específico Área: Ciências Sociais e Humanidades Tipo: Artigo</p>	<p>Defende o autor que, no contexto europeu, as oportunidades e os constrangimentos das políticas nacionais e dos mercados regionais de habitação ajudam a explicar as estratégias adotadas localmente (Van Gen, 2010).</p>	
<p>Argumenta o autor que...</p>	<p>Argumenta, por isso, o autor que a compreensão dos tipos dos envolvidos no arrendamento privado.</p>	
<p>Smith transforma o conceito...</p>	<p>Marx transforma o conceito, discernindo aí dois aspectos fundamentais:</p>	
<p>Smith insiste em x...</p>	<p>[...] se a teoria marxista tradicional considerava sobretudo o lado do “repressivo”, Althusser insiste na dimensão ideológica:</p>	
<p>Smith descreve x como...</p>	<p>Cruz-Ferreira descreve os fonemas róticos do PE como /s/ e /ʃ/, tendo como base a norma de uma falante lisboeta de meia idade.</p>	
<p>À semelhança de Smith, também Jones considera...</p>	<p>À semelhança de Tinto, também considera o envolvimento académico, traduzido na dedicação e empenho nas atividades académicas, e o envolvimento institucional.</p>	

Figure 2. Writing Assistant interface (displaying the findings of a query)

Searching for an expression is as easy as using the search bar, which causes the screen to split into two sections. Expressions on the left are shown in a generic form, making it easier for users to integrate them into their own writing without plagiarizing. The right side provides these expressions in their full, original context. Clicking on a search result reveals the specific part of the text (e.g., introduction, literature review, methods, discussion, results, or conclusion) where the phrase is located. Users can also apply filters to target specific sections, such as the introduction or results, making it easier to find the most relevant academic expressions. The Writing Assistant also explains the communicative role of the expression in the text, helping users understand how it is used in academic contexts and its proper application. Finally, the Writing Assistant provides extra context, such as the academic genre (limited to scientific papers) and the related subject area. This helps clarify the expression's meaning and ensures it is used correctly within specific academic fields.

Bringing the phrase bank to the classroom

The phrase bank was developed as a self-directed tool for students and researchers, designed to empower autonomous use and learning. However, we understand the crucial role of teachers in academic language instruction. That is why we emphasise the importance of developing learning scenarios where educators can actively employ the phrase bank in their teaching strategies. The phrase bank holds immense potential as an enriching pedagogical resource, catering to both native students and those who are learning Portuguese as a foreign language. Maximising the potential of the phrase bank in pedagogical settings requires diverse approaches. To illustrate its practical use in the classroom, we will outline two classroom activities conducted in MA classes utilising the phrase bank.

Classroom activity 1: Rewriting the Introduction of a Group Portfolio

In the first-class activity, students practised using academic phrases to improve their writing skills and better express specific ideas in the introduction of their group's portfolio.

This classroom activity was conducted with MA students enrolled in the Corpus Linguistics course of the MA in Multilingual Communication and Translation programme at University of Minho. The task began with students being instructed to write a preliminary draft of the introduction for their subject portfolio. They relied solely on their existing genre knowledge of introductions to complete this initial task. The students were divided into six groups and given 45 minutes of in-class time to work on their drafts. In the

subsequent phase, students were provided with a statement for the rewriting activity. The statement included a brief informative paragraph about introductions and detailed instructions for the exercise. The exercise prompted students to select relevant sentences from the phrase bank to rewrite their portfolio introductions, focusing on four specific functions: explaining the importance of the topic/theme, indicating the purpose of the work, indicating the object of study, and indicating the structure of the work. The students then completed the rewriting task at home and submitted their revised introductions in the following class session.

To implement this activity in-class, here are the steps to follow:

1. Instruct the students to write the first draft of their assignment, using only their existing genre knowledge of that type of text. Divide the students into groups and allocate 45 minutes of in-class time for this task.
2. Provide the students with a statement outlining the rewriting activity. This statement should include a concise paragraph providing the genre-specific information, along with clear instructions for the exercise. Provide students with genre-appropriate phrases and expressions from the phrase bank to rewrite their work.
3. Allow the students to complete the rewriting task at home and request that they submit their revised introductions during the subsequent class session.
4. Provide feedback to the students based on the analysis of the two versions of the introductions. Highlight areas of improvement and offer guidance on effective language use in academic writing.

We conducted an analysis of the two versions of the introductions, focusing on the presence of communicative functions and the use of formulaic language. In the initial version of the introductions, all six groups successfully included the functions of indicating the object of study and indicating the structure. Five groups incorporated the function of indicating the objective, and merely two groups addressed the importance of the theme. In terms of formulaic language, the groups demonstrated systematic use of formulaic language to express the function of indicating the structure of the work, but none of the groups utilised formulaic language to convey the function of indicating the object of study. During the initial stage of the experiment, the groups utilised formulaic language in approximately 66.67% of their expressions. After completing the rewriting task, there was a noticeable surge in formulaic language usage, accounting for 87.5% of all expressions employed. Additionally, it was observed that most groups successfully integrated the typical communicative functions of

the introduction. It should be highlighted that the groups employed various formulaic expressions for different functions, leading to greater diversity in the expressions used across the groups. The rewriting activity also had an impact on the length of the introductions. On average, the number of words in the groups' texts increased by 47.21% following the rewriting task.

Classroom activity 2: Peer Assessment of a Research Proposal's Literature Review

The second classroom activity involved a guided peer assessment activity aimed at acquainting students with the typical communicative functions and corresponding expressions found in literature reviews, by using the phrase bank. Peer assessment empowers students to assume accountability and control over their own learning, fosters the development of their evaluative skills, and promotes the exchange of constructive feedback among peers (Emam et al., 2019). Yu (2020) observes that peer feedback in academic writing helps students enhance their self-evaluation skills and plays a crucial role in a dialogic process, cultivating an understanding of and proficiency in good academic writing.

This classroom activity was conducted with MA students enrolled in the Project Management course of the MA in Digital Humanities programme at University of Minho. The duration of this activity was approximately 60 minutes, allowing sufficient time for the exchange of feedback and meaningful discussions among students. As part of their research proposal writing process, students composed literature reviews. They engaged in the peer assessment activity by following the specific tasks outlined below.

To implement this activity in class, here are the steps to follow:

1. Divide the class into pairs or small groups.
2. Instruct each student to bring their literature review to the session.
3. Provide students with a statement outlining the peer assessment activity.
4. Have students refer to the phrase bank to acquaint themselves with the common communicative functions within literature reviews and to explore exemplars of expressions conveying these functions.
5. Instruct students to identify these communicative functions in their peer's work, highlighting and recording specific examples.
6. Prompt students to observe and interrogate the richness and variety of expressions used by their classmates when referring to authors and studies (e.g., "according to the author," "as noted by Smith (2000)," "most studies indicate").

7. After this reflection, direct students to propose alternative formulations for at least five of these expressions. Our students used the phrase bank in the format of the Writing Assistant, although a static version of a phrase bank would work equally well.
8. Next, task students with evaluating their peer's literature review (in a minimum of 100 words), guided by a choice board. Creating a supportive and respectful atmosphere where students feel comfortable sharing their feedback and insights is essential. To achieve this, teachers can use a choice board that offers a range of prompts and options to guide feedback, encompassing both meaningful praise and constructive improvement recommendations. The choice board operates by directing students to choose two prompts from the peer feedback choice board, enabling them to offer their classmates specific, meaningful, and considerate feedback. Below, we have adapted Tucker's (2021) model to suggest a choice board that can be used in this activity.
 - Greatest Strength: Identify the strongest aspect of this work. In particular, what is strong about it? Why do you think that aspect is particularly good or well done? How did this aspect positively affect the overall quality of the work?
 - Tiny Tweaks: Identify one aspect of this work that would benefit from a small adjustment, modification or improvement. How would reworking this aspect of the work affect the overall quality of the work? Do you have specific recommendations on how your colleague could improve that aspect of their work?
 - Celebrate Surprises: What surprised you about your colleague's work? Was there any aspect of this work that was original, creative or captivating? Describe what you liked most about it.
 - Hungry for More: Identify one aspect of this work that needs to be developed further. What would you have liked more information on? In what way would more development and detail have strengthened this work?
 - Mind Blown: Identify an aspect of this work that you loved and didn't consider in your own work. Were there any ideas or approaches that your colleague used that you would like to incorporate into your work? How could you incorporate these ideas into your own work?
 - Clarifying Confusion: When reviewing your colleague's work, was there anything unclear, confusing or raised doubts? Can you identify specific elements of the project that would benefit from clearer language and/or more explanation? (see Tucker 2021)
9. Finally, guide students to read their peer's assessment of their work and participate in discussions with their classmates, elucidating the rationale behind their recommendations.

Through the exchange and assessment of their literature reviews, students engaged in a process of peer review and recommendation, focusing on the integration of relevant academic phrases and expressions from the phrase bank to improve clarity, coherence, and academic rigor in their peers' texts.

When questioned, most students found their peers' feedback helpful and indicated their intention to integrate it into their work. One student remarked, "Getting feedback from my colleagues gives me a different view of my work and helps me improve." Regarding the Writing Assistant, most students thought it was easy to use, with one stating, "I think the Writing Assistant is a valuable tool for academic writing, as it assists in eliminating redundancies and provides guidance during moments of discursive confusion."

Conclusion

In conclusion, this paper has offered an exploration of the creation, design, accessibility, and practical use of the European Portuguese Academic Phrase Bank in the classroom, with the aim of facilitating its replication in other languages. Moving forward, there are several avenues for future work and research. To begin with, there is potential for ongoing expansion and refinement of the phrase bank to encompass a wider array of academic genres, including but not limited to theses, dissertations, and research proposals. Furthermore, our intention is to persist in the development and testing of pedagogical activities incorporating the phrase bank, following the principles of active learning. We will seek collaboration of fellow educators interested in implementing this resource in their classes. Considering the significant demand for specialized instruction in academic Portuguese among foreign students, specifically, we aim to collaborate with instructors of Portuguese as a Second/Foreign Language. This stems from our belief that the phrase bank holds great promise as a resource for classes with both native and non-native language learners.

Funding Statement

This work was carried out within the scope of the PortLinguE (PTDC/LLT-LIG/31113/2017) project, financed by FEDER under Portugal 2020 and by national funds through Fundação para a Ciência e a Tecnologia, I.P. (FCT, I.P.).

WORKS CITED

- Aguiar, Micaela, José Monteiro, and Sílvia Araújo. 2023. "Semi-automatic Extraction of Academic Transdisciplinary Phraseology: Expanding a Corpus of Categorized Academic Phraseology Using BERT Machine Learning Models (Short Paper)." *Proceedings*. <http://ceur-ws.org> ISSN, 1613, 0073.
- Araújo, Sílvia, Aguiar, Micaela and Monteiro, José. 2023. "aBERTUM: A Portuguese Search Engine for Scientific and Academic Language." *International Conference on Advanced Research in Technologies, Information, Innovation and Sustainability*, 88-102. Cham: Springer Nature Switzerland.
- Assunção, Carlos. 2019. "Internacionalização da Língua Portuguesa como Língua de Ciência: do Passado para o Futuro." In *Políticas da Língua, da Comunicação e da Cultura no Espaço Lusófono*, edited by Moses. L. Martins & Isabel Macedo, 185-197. Húmus/CECS.
- Barros, Rui and Volta e Pinto, José. "Ditadura Democracia Portugal." *Público*. March 24, 2022. <https://www.publico.pt/interactivo/ditadura-democracia-portugal>
- Biber, Douglas, and Barbieri, Federica. 2007. "Lexical Bundles in University Spoken and Written Registers." *English for Specific Purposes* 26: 263-286.
- Carvalho, José António Brandão. 2014. "Literacia académica: da escola básica ao ensino superior-uma visão integradora." *Letras & Letras* 29, no 2.
- Defazio, Joseph., Jones, Josette, Tennant, Felisa, and Hook, Sara Anne. 2010. "Academic Literacy: The Importance and Impact of Writing across the Curriculum – a Case Study." *Journal of the Scholarship of Teaching and Learning* 10, no 2: 34-47.
- Drouin, Patrick. 2007. "Identification Automatique du Lexique Scientifique Transdisciplinaire." *Revue française de linguistique appliquée*, 2, no XII: 45-64. <https://doi.org/10.3917/rfla.122.0045>
- Emam, Maii, Taha, Dina, & ElSayad, Zeyad. 2019. "Collaborative Pedagogy in Architectural Design Studio: A Case Study in Applying Collaborative Design." *Alexandria Engineering Journal* 58, no 1: 163-170. <https://doi.org/10.1016/j.aej.2018.03.005>
- Estrela, Antónia & Sousa, Otlia. 2011. "Competência Textual à Entrada no Ensino Superior." *Revista de Estudos da Linguagem* 19, no 1: 247-267. <https://doi.org/10.17851/2237-2083.19.1.247-267>
- Ferrão Tavares, Clara, & Pereira, Álvares Luísa (2012). "Apresentação: Literacias Académicas Multimodais." *Intercompreensão* 16: 5-10.
- Gouveia, Carlos. M. (2019). "Como se faz uma Disciplina: Mapas de Conhecimento e Distinções Operacionais sobre o que é o Discurso Académico enquanto Objeto de Estudo." In *Discurso Académico: Uma Área Disciplinar em Construção*, edited by Fausto Caels, Luís Filipe Barbeiro e Joana Vieira Santos, 19-43. CELGA-ILTEC: Universidade de Coimbra/Escola Superior de Educação e Ciências Sociais, Politécnico de Coimbra.
- Hyland, Ken. 2008. "As Can Be Seen: Lexical Bundles and Disciplinary Variation." *English for Specific Purposes* 27, no 1: 4-21. <https://doi.org/10.1016/j.esp.2007.06.001>

- Hyland, Ken. 2018 "Sympathy for the Devil? A Defence of EAP." *Language Teaching* 51, no. 3. 383-99. <https://doi.org/10.1017/S0261444818000101>
- Hyland, Ken and Jiang, Feng. 2018. "Academic Lexical Bundles: How Are They Changing?" *International Journal of Corpus Linguistics*, 23, no 4: 383-407. <https://doi.org/10.1075/ijcl.17080.hyl>
- Jonsmoen, Kari Mari, & Greek, Marit. 2017. "Lecturers' Text Competencies and Guidance towards Academic Literacy." *Educational Action Research* 25, no 3: 354-369. <https://doi.org/10.1080/09650792.2016.1178156>
- Leiria, Isabel. 2023. "Ensino Superior: o Número de Estudantes Estrangeiros mais do que Duplicou em apenas Seis Anos." *Expresso*. May 7, 2023. <https://expresso.pt/sociedade/2023-05-07-Ensino-Superior-o-numero-de-estudantes-estrangeiros-mais-do-que-duplicou-em-apenas-seis-anos-45c34e67>
- Li, Jie, & Schmitt, Nobert. 2009. "The Acquisition of Lexical Phrases in Academic Writing: a Longitudinal Case Study." *Journal of Second Language Writing* 18, no 2: 85-10. <https://doi.org/10.1016/j.jslw.2009.02.001>
- Lutkevich, Ben. 2020. "BERT Language Model." *TechTarget*. January, 2020. <https://www.techtarget.com/searchenterpriseai/definition/BERT-language-model>
- Morley, John. 2004. "Academic Phrasebank". <https://www.phrasebank.manchester.ac.uk/about-academic-phrasebank/>
- Observador. 2023. "Portugal com o maior Número de sempre de Alunos do Ensino Superior". October 2, 2023. <https://observador.pt/2023/10/02/portugal-com-o-maior-numero-de-sempre-de-alunos-do-ensino-superior/>
- Pérez-Llantada, Carmen. 2014. "Formulaic Language in L1 and L2 Expert Academic Writing: Convergent and Divergent Usage." *Journal of English for Academic Purposes* 14: 84-94. <https://doi.org/10.1016/j.jeap.2014.01.002>
- Pintrich, Paul R. 2002. "The Role of Metacognitive Knowledge in Learning, Teaching, and Assessing." *Theory into Practice* 41, no 4: 219-225. https://doi.org/10.1207/s15430421tip4104_3
- Pordata. (n.d.). "Taxa de Analfabetismo segundo os Censos Total e por Sexo." <https://www.pordata.pt/Portugal/Taxa+de+analfabetismo+segundo+os+Censos+total+e+por+sexo-2517>
- Preto-Bay, Ana Maria. 2004. "The Social-Cultural Dimension of Academic Literacy Development and the Explicit Teaching of Genres as Community Heuristics." *The Reading Matrix* 4, no 3: 86-117.
- Rastier, François. 2001. *Arts et Sciences du Texte*. PUF: Presses Universitaires de France.
- Short, Deborah J. and Shannon Fitzsimmons-Doolan. 2007. "Double the Work: Challenges and Solutions to Acquiring Language and Academic Literacy for Adolescent English Language Learner." *A Report to Carnegie Corporation of New York*. Washington, DC: Alliance for Excellent Education.
- Swales, John. 1990. *Genre Analysis: English in Academic and Research Settings*. Cambridge University Press.

- Tardy, Christine, Buck, Rachel Hall, Jacobson, Brad, LaMance, Rachel, Pawlowski, Madelyn, Slinkard, Jennifer and Vogel, Stefan. 2022. "It's Complicated and Nuanced': Teaching Genre Awareness in English for General Academic Purposes." *Journal of English for Academic Purposes* 57, no 100966.
<https://doi.org/10.1016/j.jeap.2022.101117>
- Tucker, Catlin. 2021. "Peer Feedback: Making It Meaningful."
<https://catlintucker.com/2021/04/peer-feedback/>
- Tutin, Agnès and Jacques, Marie-Paule. 2018. Le Lexique Scientifique Transdisciplinaire: une Introduction. In *Lexique Transversal et Formules Discursives des Sciences Humaines* edited by Tutin, Agnès & Jacques, Marie-Paule, 1-26. ISTE Editions.
- Wray, Alison. 2002. *Formulaic Language and the Lexicon*. Cambridge University Press.
- Wray, Alison. 2009. "Identifying Formulaic Language. Persistent Challenges and New Opportunities." In *Formulaic Language. (Vol. 1: Distribution and historical change)* edited by Corrigan, Roberta L., Moravcsik, Edith A., Oulali, Hamid and Wheatley, Kathleen M., 27-51. Amsterdam: John Benjamins.
- Yu, Shulin. 2021. "Giving Genre-based Peer Feedback in Academic Writing: Sources of Knowledge and Skills, Difficulties and Challenges." *Assessment & Evaluation in Higher Education* 46, no 1: 36-53.