



Revista de Políticas Públicas da UFPE, Recife/PE, v. 11, n. 2, ano 2026. ISSN-e 2595-5535

Corpus Linguistics and AI Tools for Academic English Development in EMI Contexts

Maria Valesia Silva da Silva¹
Sabrina Bonqueves Fadanelli²

Abstract: English as a Medium of Instruction (EMI) has become a central strategy for the internationalization of higher education institutions, particularly in contexts where English is not the official language. However, EMI presents significant challenges for both students and teachers, including difficulties with specialized vocabulary, academic discourse conventions, listening comprehension, cognitive overload, communicative anxiety, and insufficient institutional and pedagogical support. This article discusses these challenges and proposes an integrated pedagogical approach based on Corpus Linguistics (CL) and Artificial Intelligence (AI) resources to support the development of academic communication skills in EMI environments. Drawing on literature from EMI, Corpus Linguistics, and Natural Language Processing, the study presents a set of freely available tools—including AntConc, ChatGPT, the Manchester Academic Phrasebank, MICUSP, Grammarly, QuillBot, and Text-to-Speech applications—and discusses their pedagogical applications for vocabulary acquisition, academic writing, paraphrasing, self-editing, oral communication, and rhetorical awareness. A step-by-step pedagogical sequence is proposed to guide learners from terminology identification to academic writing revision and oral presentation. The article also critically examines the limitations of these technologies, emphasizing the need for teacher guidance, corpus literacy, ethical AI use, and institutional support. It concludes that, although AI and corpus-based tools cannot replace systematic language instruction, they can significantly enhance academic communication in EMI contexts when integrated into carefully designed pedagogical practices.

Keywords: English as a Medium of Instruction (EMI); Corpus Linguistics; Artificial Intelligence; Academic English; Higher Education.

¹ Mestre em Aquisição da Linguagem pela Universidade Federal do Rio Grande do Sul (UFRGS), com licenciatura em Línguas Inglesa e Portuguesa e suas Licenciaturas (PUC/UCS). Atua como professora de Língua Inglesa, Estratégias de Comunicação Oral, Línguas Estrangeiras para Crianças, Tecnologias e Metodologias Aplicadas ao Ensino de Línguas Estrangeiras, Estágios Supervisionados no Ensino Fundamental e Médio, e Coordenação do UCS Writing Center na Universidade de Caxias do Sul (UCS). Participa da formação de professores da Instituição em English as a Medium of Instruction – EMI- mvssilval@ucs.br

² Doutora em Teorias Linguísticas do Léxico pela UFRGS, mestre em Linguística Aplicada e professora credenciada no PPG Letras e Cultura da Universidade de Caxias do Sul. Participa da formação de professores da UCS em English as a Medium of Instruction – EMI – sbfadane@ucs.br

Resumo: O Ensino de Conteúdos por Meio do Inglês (English as a Medium of Instruction – EMI) tem se consolidado como uma importante estratégia de internacionalização das instituições de ensino superior, especialmente em países onde o inglês não é língua oficial. Entretanto, sua implementação impõe desafios significativos para estudantes e professores, tais como dificuldades relacionadas ao domínio da terminologia especializada, das convenções dos gêneros acadêmicos, da compreensão oral, da sobrecarga cognitiva, da ansiedade comunicativa e da insuficiência de apoio institucional e formação docente específica. Este artigo discute esses desafios e propõe uma abordagem pedagógica integrada baseada em ferramentas da Linguística de Corpus e da Inteligência Artificial para favorecer o desenvolvimento das competências de comunicação acadêmica em contextos EMI. Com base em estudos da área de EMI, Linguística de Corpus e Processamento de Linguagem Natural, são apresentadas ferramentas gratuitas, como AntConc, ChatGPT, Manchester Academic Phrasebank, MICUSP, Grammarly, QuillBot e aplicativos de conversão de texto em fala, destacando suas aplicações no ensino e na aprendizagem de vocabulário técnico, escrita acadêmica, paráfrase, revisão linguística, produção oral e consciência retórica. Além disso, é proposta uma sequência pedagógica que integra essas ferramentas desde a identificação da terminologia especializada até a produção e apresentação oral de textos acadêmicos. Conclui-se que, embora essas tecnologias não substituam o ensino sistemático da língua nem políticas institucionais de apoio ao EMI, seu uso crítico e pedagogicamente orientado pode contribuir significativamente para o desenvolvimento da comunicação acadêmica em inglês e para o fortalecimento das práticas de internacionalização no ensino superior.

Palavras-chave: Ensino por Meio do Inglês (EMI); Linguística de Corpus; Inteligência Artificial; Inglês Acadêmico; Internacionalização do Ensino Superior.

INTRODUCTION

The acquisition of academic communication skills in English, as well as the learning of English terminology related to technical and scientific fields, are fundamental for students to successfully integrate into university environments and technical and scientific areas that use English as a Medium of Instruction (EMI). This article aims to reflect on the challenges students face in relation to written and oral communication in EMI scenarios, as well as to suggest the application of Corpus Linguistics (CL) tools and resources in conjunction with Artificial Intelligence to facilitate the development of academic English communication skills in EMI environments.

According to Macaro et al. (2018), EMI environments are defined as those in which English is used to teach academic subjects in countries or jurisdictions where English is not an official language. In Brazil, a reasonable number of public and private universities adopt EMI as a way to promote the internationalization of the institution, since the practice of EMI can help participants improve their English language proficiency and also facilitate the offering of subjects to foreigners who wish to study at the university as an exchange activity (Gimenez et al, 2018).

CL, seen as an approach, has as its main characteristic empirical analysis, based on existing patterns in natural texts, making use of computerized tools and depending on qualitative and quantitative approaches (Biber, 1988; Berber-Sardinha, 2004). Unlike more traditional approaches, which determine that language should be analyzed according to previously determined principles, CL first observes the behavior of language in its natural habitat (the text) and then theorizes about it. Through the CL approach, it is easier to observe patterns in the language under analysis. Gries (2006, 2009) and Thompson and Hunston (2006) argue that examining language use means finding patterns of occurrence within contextual factors. Since the recognition of what is not typical occurs much more frequently than what is standardized, CL then enables the recognition of these patterns with a more homogeneous view.

Corpus Linguistics often operates in cooperation with the field of Natural Language Processing, which in turn interacts with other areas belonging to AI, such as Information Technology and Computational Linguistics. Natural Language Processing (NLP) helps computers understand, interpret, and manipulate human language, seeking to bridge the gap between human communication and computer understanding. Some tools that will be described later in this chapter come from this cooperation.

The following section discusses the challenges that students may encounter in EMI environments; the reflection then continues with the presentation of some resources that can serve as tools for developing Academic English communication skills in EMI contexts.

CHALLENGES FACED BY STUDENTS IN EMI CONTEXTS IN RELATION TO WRITTEN AND ORAL COMMUNICATION

Considering the various scenarios in which English as a Medium of Instruction (EMI) is applied, it is essential to address the main obstacles faced by students from different perspectives, namely: linguistic, cognitive, affective, and institutional.

In the sphere of linguistic and cognitive obstacles, as discussed by Masrai, Milton, El-Dakhs et al (2021), one of the key points is *the study of specialized vocabulary*, since students in technical fields need to learn specific terms that are not part of everyday English vocabulary. Often, they know the concepts in their native language but cannot recognize them in English. This lexical gap compromises their ability to read technical texts, produce written work, and understand academic lectures or presentations. In addition, technical vocabulary makes up a considerable portion of the lexicon found in specialized texts, which is often neglected in general English (GE) courses.

Also, based on research by Swales (1990), a second aspect within this sphere is *the lack of knowledge of the structures of academic text genres*. This topic tends to be a significant challenge for students to produce or even improve academic texts, such as abstracts, technical articles, reports, oral presentations, among other genres, which follow different conventions in academic English, whether comparing them to informal English used in everyday communication; or comparing them to the models of academic texts in the mother tongue, in this case, Portuguese. Such genres require not only mastery of complex linguistic structures, but also an understanding of the discursive practices of the scientific community, something that requires time, practice, and adequate guidance until it becomes an easy field for students to navigate.

Furthermore, a third element must be integrated into the sphere of linguistic and cognitive barriers, namely *the complexity of listening comprehension*. Especially in EMI classes, as Dang (2022) points out, teachers may—and are expected to—have very peculiar accents, proficiency levels, and ways of organizing speech, making listening and listening comprehension more challenging. This linguistic diversity, although representative of the reality of an internationalized environment, can cause confusion and insecurity among students, especially when there are prosodic variations, elisions, and the use of unfamiliar idiomatic expressions. In addition, the absence of visual aids or complementary support materials—such as transcripts, slides, or subtitles—intensifies the cognitive effort required to follow oral classes dense in technical content. In such situations, many students end up adopting a more passive stance, hesitating to ask for clarification for fear of exposure or delaying the progress of the class. Listening comprehension in EMI, therefore, is not limited to linguistic mastery, but requires the development of active listening skills, familiarity with academic spoken language, and the creation of strategies to deal with the fluidity of speech in real time. This highlights the importance of more inclusive teaching practices, such as the use of clear language, strategic pauses, rephrasing, and multimodal support by teachers.

To complete this first group of barriers, it is important to consider the *cognitive overload* created by having to learn technical content while decoding a foreign language, in this case, English. There is consensus among authors that students have to deal with large amounts of information and experience difficulties in searching, selecting, organizing, and mentally processing both technical content and linguistic decoding, as pointed out by Sigolo and Casarin (2024) in their literature review article. Dang (2022) complements this discussion by highlighting that the specific lexical choices and syntactic complexity of EMI classes can generate cognitive overload, especially when students need to both

understand the foreign language, here focusing on English, and process the technical content of the subjects. In this context, the development of specific language skills, combined with the improvement of metacognitive learning strategies, emerges as a way to mitigate these challenges and encourage student engagement.

In a second sphere of difficulties, it is necessary to mention the affective challenges, often linked to the students' own personalities. It is quite common to find a certain *communicative anxiety* in EMI environments, when students feel frightened and insecure about exposing themselves orally in English, fearing the judgment of others and teachers. Furthermore, insecurity due to not mastering technical English can make students feel out of place in an EMI environment. This topic is addressed, among other researchers, by Suzuki (2017), who identified high levels of anxiety among undergraduate students at Japanese universities during their first months of exposure to courses taught entirely in English.

This anxiety is not limited to public speaking, but also extends to understanding questions, formulating answers under pressure, and the constant fear of making mistakes. This phenomenon is known as *foreign language classroom anxiety* and encompasses, among other factors, *communication apprehension*, the fear of communicating orally in a foreign language—and the fear of negative evaluation. In EMI contexts, these feelings can be intensified by the fact that English is being used not as an object of study, but as a means of acquiring technical knowledge. Thus, language is both an instrument and an obstacle, which can negatively affect students' self-esteem, motivation, and academic engagement. Therefore, overcoming communicative anxiety requires greater familiarity with the language, as well as pedagogical strategies that promote a welcoming environment, with an emphasis on active listening, valuing effort, and gradually building communicative confidence.

The last set of challenges faced by academics and other students in EMI situations relates to institutional and pedagogical issues, which are addressed by Macaro et al (2018). *The lack of institutional support* is one of the crucial points for students. Few educational institutions offer continuous language support to students in EMI courses, specialized tutoring, or academic writing and reading workshops in English. The absence of systematic support initiatives compromises not only students' academic performance but also their retention and participation in international programs, courses, and events. In many cases, the choice of EMI is based on institutional internationalization guidelines, but it does not translate into structured training or language support policies, creating a mismatch between institutional discourse and the reality of teaching and learning.

Finally, the most complex obstacle is the *lack of teacher training* (Kobayashi; Matsuda 2025). In Brazil, most higher education teachers still feel limited and unprepared to teach classes in English, as they feel that this may compromise the quality of their explanations and the clarity of the content. The training offered to teachers, when available, may focus on language proficiency, leaving aside essential aspects of EMI pedagogy, such as adaptation of materials, use of accessible language, linguistic-discursive mediation, and *scaffolding* strategies. This training gap can have a negative impact on teaching practice, undermining the quality of classes and the level of student learning. In addition, teacher unpreparedness and resistance in EMI contexts end up being reproduced and naturalized at other educational levels, since teachers in training are influenced by these ineffective practices. This highlights the urgency of institutional policies that not only encourage EMI as a tool for internationalization but also ensure real conditions for its quality implementation through the valorization and continuous training of teachers and effective support for students.

The studies discussed here point out that English proficiency alone is not a guarantee of active participation and academic success for students in EMI environments. Lack of mastery of specialized vocabulary and unfamiliarity with the elements that constitute written and oral academic genres impose significant barriers. In addition, the cognitive overload that occurs due to the simultaneity of learning subject content and processing the English language compromises comprehension and academic performance. At the same time, communicative anxiety and insecurity when speaking in another language directly affect classroom interaction and student autonomy. Finally, both the lack of continuous institutional support and the absence of specific pedagogical training for teachers in EMI contexts reveal the urgent need for institutional policies that integrate linguistic and methodological development into internationalized higher education practices.

It is important to note that internationalization involves various sectors of HEIs, it is a process of construction involving many hands, and "requires structural changes, with the inclusion of the international dimension in the organizational structure of the institution, investment in team development, and the involvement of all members of the academy, whether they are teachers, researchers, students, or managers" (Stallivieri & Gonçalves, 2015). In recent years, there has been an increase in opportunities for classes to be taught in English at HEIs. (Baumvol & Sarmiento, 2016; Martinez, 2016; Gimenez et al., 2018). In some cases, this has been in summer courses, in others in partnerships with HEIs in other countries, or in subjects within the course curriculum itself. However, in some institutions, the task is left to the professors who are asked to teach their technical domain through English without any specific

training, because the institutional belief is that to begin teaching using EMI all the professor must do is switch languages (Kobayashi; Matsuda 2025). It is due to all these pitfalls that we believe the Corpus Linguistics and AI tools presented here may assist, even if just a little, the professors who are placed under these circumstances. In times of great developments in the field of Artificial Intelligence, as mentioned earlier, it is essential to consider how AI can assist, together with Corpus Linguistics, in the process of academic writing and in the teaching of scientific domains' terminology in EMI environments.

AI presents several significant challenges to EMI environments. One major concern is the growing reliance on AI writing and translation tools, which may undermine the development of genuine English language proficiency, as students can bypass the productive struggle required to build communicative competence (Chun et al, 2025). This raises questions about academic integrity, since it becomes increasingly difficult for educators to distinguish between a student's authentic language output and AI-generated text. Additionally, AI tools may reinforce inequalities (OECD 2024), as students with better access to advanced technologies gain an unfair advantage over peers with limited resources. Finally, educators themselves face the challenge of adapting their pedagogical approaches and assessment strategies to remain meaningful and rigorous in an environment where AI assistance is increasingly pervasive and difficult to regulate.

The next section brings ideas on how to use some tools that can facilitate the learning of Academic English communication skills in virtual or face-to-face EMI environments.

TOOLS THAT FACILITATE THE TEACHING/LEARNING OF ACADEMIC ENGLISH COMMUNICATION SKILLS IN EMI ENVIRONMENTS

The tools presented below are all free and come from research at the intersection of Corpus Linguistics and Natural Language Processing, a branch of Artificial Intelligence.

The first challenge described above referred to the difficulty students have in identifying and acquiring technical vocabulary in EMI contexts. Our suggested resource for addressing this difficulty involves understanding the Principle of Idiomaticity.

Sinclair (1991, p. 1106) argues that one of the formative principles of the language system is that of idiomaticity: "a language user has at their disposal a large number of pre-constructed phrases that constitute unique choices, even if they appear to be analyzable into segments." One of the bases for the principle of idiomaticity is the statement that Firth (1957, p. 117) postulated in his analysis of meaning and its contextual nature, that a word is known by the company it keeps. Part of the process of acquiring vocabulary in a foreign language means knowing which combinations of words are most frequent in a

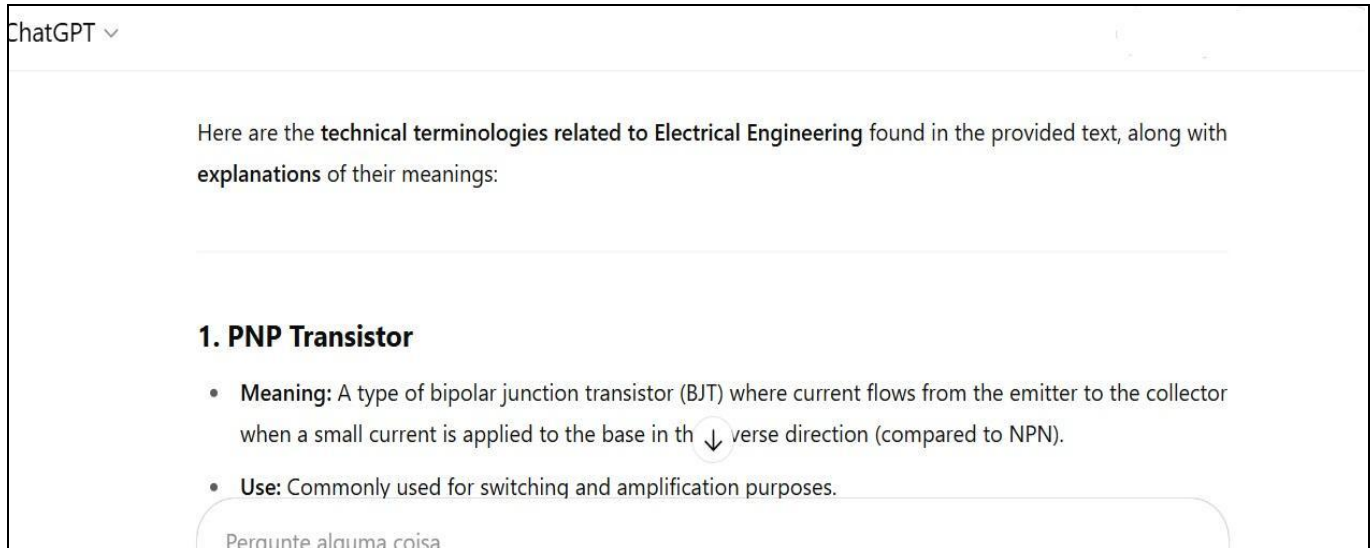
given technical-scientific context; and what the terminology of that technical-scientific domain is, which will consist of the principle of idiomaticity.

Therefore, to build knowledge about the technical vocabulary and lexical combinations of a scientific field, both students and teachers can use the following tools: a keyword generator and a concordance tool. AntConc is a corpus analysis toolkit designed by Laurence Anthony (ANTHONY, 2026). It is equipped with a concordance tool (which indicates which words combine with which), word frequency list generators, and a tool used to extract technical terms, the Keyword function (which provides the keywords of a corpus by comparing the study corpus with another corpus at least five times larger – called the reference corpus, which is found in the tool itself).

With AntConc, teachers can instruct students to compile a corpus of texts containing technical terms related to their area of research. After uploading this corpus into the tool, students can search for the most frequent word combinations that occur in the documents and generate a list of keywords that, because they occur much more frequently in technical texts than in general English texts, denote the most specific terminology specific to the technical domain in question. From there, students and teachers can work together to identify examples that appear in the texts that make up the corpus and build their own terminological units and combinations for their chosen field. Real examples of the use of this tool for this purpose can be found in studies such as Fadanelli (2017), Fadanelli and Monzón (2017), and Tcacenco and Fadanelli (2022).

Another more automated way to identify the technical terms in a text belonging to a technical-scientific domain, in addition to obtaining more detailed explanations of what they mean in the domain (not the translation), is to use ChatGPT, with the following prompt: *“Find the technical terminology related to (technical area) in the text, providing an explanation of their meanings,”* asking the AI tool to search for technical terms in a given text in a given technical area and provide definitions of what these terms are. The teacher can use the information generated by ChatGPT to set up an exercise in which the student must relate the technical term to its explanation, and the student can obtain more information about the function of that terminology in their field of study. Below is a figure (Figure 1) that illustrates how ChatGPT can provide interesting information that aids in the acquisition of vocabulary and specific concepts related to this vocabulary, as well as its use. Obviously, students should receive guidance from the teacher on the information found to verify that the tool has provided the explanation correctly:

Figure 1



Source: ChatGPT

In the figure, ChatGPT states: "Here are the technical terms related to Electrical Engineering in the text, along with explanations of their meanings." It provides the explanation that a PNP transistor is a type of bipolar junction transistor where currents flow from the emitter to the collector when a low current is applied to the base in the reverse direction (compared to NPN), and that its use is commonly associated with circuit breakers and amplifier functions.

In an EMI context, helping students identify rhetorical moves in academic texts (Swales, 1990) can help boost the production and comprehension of academic texts from different scientific domains. To facilitate this identification, we recommend two tools: the Manchester Phrasebank and MICUSP.

The Manchester Phrasebank³ is a free online repository of ready-made academic phrases organized by discursive functions (e.g., introduction, comparison, definition, explanation of results, conclusion). It offers sentence templates commonly used in academic articles, reports, essays, and theses. The original corpus from which the phrases were "harvested" consists of 100 postgraduate dissertations completed at the University of Manchester. However, phrases from academic articles drawn from a wide range of disciplines continue to be incorporated. Phrasebank helps students and teachers formulate ideas with appropriate linguistic structures, which is essential in courses taught in English, especially in academic and scientific contexts. One of the skills developed by the tool is the recognition of patterns of discourse in academic English, which improves not only writing but also the comprehension of academic texts and

³ Available at: <https://www.phrasebank.manchester.ac.uk/>. Accessed July 2025.

oral presentations. Teachers can use Phrasebank to prepare teaching materials, provide linguistic feedback to students, and plan writing tasks based on academic functions.

MICUSP⁴, which stands for "Michigan Corpus of Upper-Level Student Papers," presents on its platform text genres produced by graduate students at the University of Michigan in the United States; The texts belong to four main areas: Social Sciences, Biological and Health Sciences, Humanities and Arts, and Physical Sciences and Mathematics (Fadanelli; Da Silva 2021). With this tool, it is possible to identify different types of textual genres and filter searches according to area, as well as search for specific phrases (Bevilacqua 2005) or terminology that can be found in these texts. Fadanelli and Da Silva (2021) propose comparing discourse markers that express contrast between different areas using the tool, which is very useful for producing academic texts with vocabulary and linguistic constructions more specific to each scientific area. The same application is possible in activities focused on EMI, and we also suggest another more specific possibility for use: MICUSP allows for the comparison of how rhetorical movements are employed in different disciplines, revealing commonalities and differences. To do this, simply select the textual characteristics you want to analyze, the textual genres, and the areas to compare in the boxes on the left side of the main screen, and the tool will bring up examples of texts from these areas so that the comparison can be made. Other possibilities for using the tool are described by Wang (2022), who in his research demonstrates how mechanisms used by authors to reduce the strength of their arguments (modality) are more commonly used in more argumentative areas than those guided by data collection; and by Aull (2019), who describes the linguistic differences used in the positioning of authors in explanatory and argumentative texts.

To address affective challenges, two other tools that may be very helpful for teachers and learners in EMI contexts: *Grammarly*⁵ and *Quillbot*⁶. For learners, Grammarly serves as an on-demand writing assistant, providing real-time feedback on grammar, punctuation, vocabulary, and tone. Rather than simply correcting errors, it can help students understand why a particular construction is incorrect, functioning as a form of automated corrective feedback that complements classroom instruction. This is especially valuable in EMI settings where students may have limited opportunities for individualized language support from their instructors, who are often content specialists rather than language teachers.

QuillBot, on the other hand, offers learners a paraphrasing and summarizing tool that can support reading comprehension and academic writing development. When used responsibly, it can help students

⁴ Available at: <https://micusp.elicorpora.info/main>. Accessed in June 2026.

⁵ Available at: <https://www.grammarly.com/> Accessed in June 2026

⁶ Available at: <https://quillbot.com/pt/> Accessed in June 2026

rephrase complex source material in ways that are more accessible or serve as a model for how ideas can be expressed differently in English. For EMI learners who struggle with reformulating content-heavy texts, QuillBot can act as a scaffold, bridging the gap between understanding an idea and expressing it in academically appropriate English. However, learners must be guided to use it as a learning aid rather than a shortcut, engaging critically with the suggested reformulations rather than accepting them uncritically.

From a teaching perspective, Grammarly can be integrated into the writing process as a pre-submission tool, encouraging students to self-edit and reflect on their language use before handing in assignments. Teachers in EMI environments — who may not feel confident or equipped to provide detailed language feedback — can direct students to Grammarly as a first layer of linguistic support, freeing up the instructor's feedback for content and argumentation. Some educators have also used Grammarly's reports and insights to identify common error patterns across a class, informing targeted mini lessons on recurring language issues relevant to their discipline.

Teachers can similarly harness QuillBot to design productive language activities, such as asking students to compare their own paraphrases with QuillBot's output and evaluate which version is more accurate or appropriate in a given academic context. This kind of critical comparison encourages metalinguistic awareness and develops the paraphrasing skills that are central to academic writing.

In order to work with difficulties regarding oral communication, our suggestion is the use of AI Text-to-Speech (TTS) tools. They convert written text into spoken audio with natural-sounding voices. For teachers, these tools can be valuable in lecture preparation, especially for EMI professors who feel uncertain about their spoken delivery. A teacher can convert written lecture notes or slide content into audio to review how the material sounds in English before delivering it live. By using AI-generated audio, teachers can create new and more interactive learning tools, maximizing their time for student interaction. They can also use TTS to produce supplementary audio content for their courses — narrated slides, recordings of key concepts, or listening exercises — without needing a studio or a native speaker.

For students, TTS provides powerful educational assistance, particularly foreign-language learners who are working to integrate written and spoken expressions in a new tongue. For learners who struggle with reading, TTS tools have been shown to improve reading comprehension, information retention, and decoding. In an EMI context, a student can listen to technical readings at their own pace, slow down complex passages, or listen repeatedly until the content sinks in — reducing the cognitive load of processing both new content and a foreign language simultaneously.

All these suggestions bring us to a final suggestion of pedagogical sequence that professors may be able to organize: the main objective of this pedagogical sequence is have the students prepare and write a

text in English regarding their findings about a specific technical/scientific domain, and after presenting the text to the group, using English. The sequence could be made of steps:

Step 1: The student uses AntConc to identify the most frequent technical terms in their specialized corpus.

Step 2: They use ChatGPT to obtain contextualized definitions of those terms.

Step 3: They use the Manchester Phrasebank to find rhetorical structures to incorporate those terms into their text.

Step 4: They use MICUSP to compare their text with texts from their field produced by graduate students.

Step 5: They use Quillbot to help them to compare their own paraphrasing and check if they can be improved.

Step 6: They use Grammarly for a final linguistic review.

Step 7: They use TTS tools to help prepare for the presentation.

Nevertheless, all tools mentioned here come with important caveats for EMI educators: they must establish clear guidelines around their use to prevent over-reliance, ensure that institutional academic integrity policies are respected, and remain aware that neither tool fully accounts for discipline-specific discourse conventions, which are a core feature of authentic EMI communication. Other possible pitfalls that surround the sequence are:

- AntConc requires a degree of corpus literacy that most content professors have not been trained in. Without guidance on how to build a representative corpus and interpret frequency lists, the tool can produce misleading results. A professor might, for instance, over-focus on high-frequency general academic words rather than the discipline-specific terms that actually matter most in their lectures.
- ChatGPT might generate inaccurate or outdated definitions. A professor who is not already confident in the subject matter may not catch these errors. There is also the risk of oversimplification — AI-generated definitions tend to smooth over the nuance and ongoing debates that characterize disciplinary knowledge. Professors should treat ChatGPT as a starting point for reflection, never as an authoritative source.
- Over-reliance on the Manchester Phrasebank can produce writing and speech that feels formulaic or disconnected from the professor's natural voice. Phrases that work well on a paper may sound stiff and unnatural when spoken aloud in a classroom.

- MICUSP contains texts written by upper-level undergraduate and graduate students — not professional academics or published researchers. This means it is a useful benchmark for student-level academic writing, but it may not fully reflect the conventions expected. Professors need to be aware of this distinction and select the appropriate comparison texts for their specific purpose. Additionally, MICUSP is weighted toward certain disciplines, so coverage across all fields is uneven.
- Grammarly is calibrated primarily toward general written English and tends to push users toward a standardized, often American English norm. It frequently misidentifies disciplinary conventions as errors — for instance, the passive voice, which is standard in many scientific fields, is often flagged unnecessarily.
- As for Quillbot, the risk is that it tends to prioritize surface-level lexical substitution over genuine conceptual reformulation, which means its paraphrases can obscure or subtly distort the original meaning. In disciplinary contexts where precision matters — engineering, medicine, law — this is a serious concern. There is also the issue that Quillbot is not discipline-aware, so it may replace a technical term with a near-synonym that carries different connotations within the field.
- Stress patterns, pauses for emphasis, and interactive cues (essential for clear and effective communication) are poorly represented in TTS output. Also, TTS tools do not reveal whether the person can actually pronounce the terms fluently, manage their breath and pace under pressure, or recover when they lose their place.

FINAL CONSIDERATIONS

The challenges faced by students in English as a Medium of Instruction (EMI) environments are multiple and interrelated. Mastery of specialized technical vocabulary, familiarity with academic genre conventions, listening comprehension in diverse accent contexts, and the cognitive overload of processing content and language simultaneously all represent significant linguistic and cognitive barriers. These are compounded by affective obstacles — particularly communicative anxiety and the fear of negative evaluation — as well as by persistent institutional gaps, such as the absence of structured language support for students and the lack of specific EMI pedagogical training for teachers. This article aimed at presenting AI and Corpus Linguistics tools that could address some of the difficulties teachers and students face when working with communication in writing and speaking in English as a Medium of Instruction environments.

The Corpus Linguistics and Artificial Intelligence tools presented here represent a concrete step toward addressing these difficulties. AntConc and ChatGPT offer complementary pathways into specialized vocabulary acquisition: the former provides empirically grounded frequency and keyword data drawn from authentic disciplinary texts, while the latter offers contextualized definitional support that can be scaffolded into classroom activities. The Manchester Phrasebank and MICUSP target the structural dimension of academic literacy, helping students and teachers identify the rhetorical conventions of different genres and disciplines, reducing the insecurity that arises from unfamiliarity with academic English discourse patterns. For the affective dimension, Grammarly and QuillBot function as low-stakes, on-demand support tools that allow students to engage with their own writing more autonomously, gradually building confidence without the fear of immediate judgment. Finally, Text-to-Speech tools extend this support to oral communication, enabling both students and teachers to prepare for the spoken demands of EMI with greater self-assurance. The pedagogical sequence proposed here integrates these tools into a coherent workflow, moving students from vocabulary identification through text production and revision to oral presentation.

It is important, however, to recognize the limitations of these tools and of the approach proposed here. None of the resources discussed can substitute for systematic, sustained language instruction or for institutional policies that genuinely support both teachers and students in EMI environments. Educators adopting this pedagogical sequence must therefore invest in critical, guided implementation, which includes establishing clear learning objectives, transparent guidelines around academic integrity, and ongoing reflection on what each tool can and cannot do.

Future research should focus on empirically evaluating the outcomes of pedagogically sequenced AI and corpus-assisted instruction in real EMI classrooms, attending to variables such as disciplinary areas, students' prior language proficiency, and institutional context. As Artificial Intelligence continues to evolve rapidly, the tools available to EMI educators will inevitably change, making ongoing critical engagement with their possibilities and limitations a core component of responsible EMI pedagogy. Higher education institutions, in turn, are called upon to move beyond adopting EMI as an internationalization strategy and to commit to the structural conditions — training, support, and inclusive policies — that make quality EMI education a more feasible reality.

REFERENCES

- AULL, L. L. Linguistic Markers of Stance and Genre in Upper-Level Student Writing. **Written Communication**, v.36, n.2, 2019, p.267-295.
- ANTHONY, L. (2026). *AntConc* (Version 4.4.1). Tokyo, Japan: Waseda University.
Available at: <https://www.laurenceanthony.net/software/AntConc>
Accessed June 2026
- BAUMVOL L.K. SARMENTO S. **Internationalization at home and the use of English as a medium of instruction**. Echoes, Florianópolis, 2016.
- BERBER-SARDINHA, T. **Corpus Linguistics**. São Paulo: Manole, 2004.
- BEVILACQUA, C. R. Phraseology: perspective of common language and specialized language. **Revista Língua & Literatura**, v. 7, n.10-11, 2005, p. 73-86.
- BIBER, D. **Variation across Speech and Writing**. Cambridge: Cambridge University Press, 1988.
- CHUN, C. K. W.; SUM, K. W.; SHIU, C. W. H. (2025). Coping with the Challenge from English-Medium Instruction (EMI): What Learning Strategies Matter and How? In: **11th International Conference on Higher Education Advances (HEAd'25)**. Valencia, 17-20 June 2025. Available at: <https://riunet.upv.es/server/api/core/bitstreams/cdd4fda4-f126-4160-b951-ec969cde4646/content>
Accessed in June 2026.
- DANG, T. N. Y. Vocabulary in academic lectures. **Journal of English for Academic Purposes**. 58, Special issue: Lecture discourse and lecturer training. 2022. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S1475158522000431?via%3Dihub> . Accessed June 2026.
- FADANELLI, S. B. **Didactic-Pedagogical Terminography: methodology for the development of resources aimed at teaching English for specific purposes**. Doctoral Thesis - Graduate Program in Letters, UFRGS, 2017.
- FADANELLI, S. B. MONZÓN, A. J. Textual genres datasheet and scientific article in ESP classes: lexical-statistical surveys for educational purposes. **Domínios de Lingu@gem**, v. 11, n. 2, 2017, p. 351-378.
- FADANELLI, S. B. DA SILVA, M. V. S. Building Academic Vocabulary Acquisition Skills in English Academic Writing with the aid of Corpus Linguistics: the UCS Writing Center Initiative.— **Revista Linguagem em Foco**, v. 13, n. 4, 2021, p. 155-170.
- FIRTH, J.R. **Applications of General Linguistics**. Transactions of the Philological Society Volume 56, Issue 1, 1957.
- GIMENEZ et al. **Guide To English as a Medium of Instruction in Brazilian Higher Education Institutions**. FAUBAI: British Council, 2018. Available at: <https://faubai.org.br/britishcouncilfaubaiguide2018.pdf>
Accessed: June 2026.
- GRIES, S. T. **Some proposals towards more rigorous Corpus Linguistics**. Zeitschrift für Anglistik und Amerikanistik, v. 54, n.2, 2006, p. 191–202.
- _____. What is Corpus Linguistics? **Language and Linguistics Compass**, v. 3, 2009, p. 1–17.

KOBAYASHI, E. MATSUDA, E. Revisão Sistemática do EMI: uma investigação do contexto brasileiro. *Travessias*. Cascavel, v. 19, n. 3, p. 1-21, set./dez. 2025. Available at: https://e-revista.unioeste.br/index.php/travessias/pt_BR/article/view/36228

Accessed June 2026.

MARTINEZ, R. English as Medium of Instruction (EMI) in Brazilian Higher Education: Challenges and Opportunities. In: FINARDI, K. *English in Brazil: Views, Policies and Programs*. Londrina: Eduep, 2016, p. 191–228.

MORLEY, J. The Manchester Academic Phrasebank. The University of Manchester, 2023. Available at <https://www.phrasebank.manchester.ac.uk/>

Accessed June 2026.

OECD. The Potential Impact Of Artificial Intelligence on Equity and Inclusion in Education. OECD Publishing, August 2024. Available at: https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/08/the-potential-impact-of-artificial-intelligence-on-equity-and-inclusion-in-education_0d7e9e00/15df715b-en.pdf#page=33.10

Accessed June 2026.

STALLIVIERI, L. BIRCH GONÇALVES, R. **New Pedagogical Proposals for the Development of Subjects Taught in a Foreign Language in Multicultural Classrooms**. *Journal of Management Sciences*, vol. 17, no. 41, April, 2015, pp. 130-142. Federal University of Santa Catarina, Santa Catarina, Brazil.

MACARO, E. et al. A systematic review of English Medium Instruction in higher education. **Language Teaching**, v. 51 (1), pp. 36-76.

MASRAI, A.; MILTON, J. EL-DAKHS, D.A.S. et al. Measuring the contribution of specialist vocabulary knowledge to academic achievement: disentangling effects of multiple types of word knowledge. **Asian Journal of Second Foreign Language Education**. 6, 8, 2021.

SIGOLO, B.O.O.; CASARIN, H.C.S. Contributions of cognitive load theory to understanding information overload: a literature review. **Digital Journal of Library Science and Information Science**. Campinas, SP, v. 22, 2024.

SINCLAIR, J. **Corpus, Concordance, Collocation**: Describing English Language. Oxford: Oxford University Press, 1991.

SUZUKI, N. Foreign Language Speaking Anxiety in the English-Medium Instruction (EMI) Program in a Japanese University. **CASELE Research Bulletin**. v.47, 2017, p. 11-21.

SWALES, J. **Genre Analysis**: English in Academic and Research Settings. Cambridge: Cambridge University Press, 1990.

TCACENCO, L. FADANELLI, S. B. A comparison of syntactic patterns in Brazilian Portuguese and English: an example with comparable corpora of museum texts. **Letras**, v. 62, 2022, p. 5–21.

THOMPSON, G; HUNSTON, S. (eds.). **System and corpus**: exploring connections. London: Equinox. 2006.

WANG, X. Hedging in academic writing: Cross-disciplinary comparisons in the Michigan Corpus of Upper-Level Student Papers (MICUSP). **JALTCALL Publications**, v. 1 2021, p.125–142.