revistadelta

https://dx.doi.org/10.1590/1678-460X202541474063

Articles

Corpus Linguistics and Artificial Intelligence

Linguística de Corpus e Inteligência Artificial

Tony Berber Sardinha*

ABSTRACT

This article argues that a register-based Multi-Dimensional (MD) description is a suitable route for characterizing AI-generated language in corpus linguistics. The argument is illustrated with two sample studies: a grammar-oriented investigation that applies traditional MD analysis to English-as-a-foreign-language textbook texts and a discourse-oriented analysis that relies on lexical MD analysis to explore AI-generated pop music lyrics. In both cases, the results reveal sharp differences between AI-generated and human language. In the EFL texts, AI-written texts are more informational, abstract, and impersonal whereas human texts display interpersonal awareness, stance, and engagement. In the pop lyrics, AI generates moralized empowerment discourses that recast historical conflicts depicted in rap music as generalized virtue narratives. In both cases, AI demonstrates signs of register deficit (a limited awareness of register variation due to shallow knowledge of the linguistic constituency of human registers) and register metamorphosis (generation of texts that resemble one register on the surface but are realized linguistically as another).

Keywords: *Corpus Linguistics; Artificial Intelligence; Register; Multidimensional Analysis*

RESUMO

O presente artigo argumenta que a descrição linguística baseada em registro (variedades textuais) e em Análise Multidimensional (AMD) constitui um caminho adequado para caracterizar a linguagem gerada por IA no âmbito da Linguística de Corpus. O argumento é ilustrado por meio de dois estudos

* Pontifícia Universidade Católica de São Paulo – PUC-SP. São Paulo, Brasil. https://orcid.org/0000-0001-8815-1521. E-mail: tonycorpuslg@gmail.com

prévios: um que aplica a AMD Tradicional a textos de livros didáticos de Inglês como Língua Estrangeira (ILE; estudo de orientação gramatical) e outro que utiliza a AMD Lexical para explorar letras de músicas pop geradas por IA (estudo de orientação discursiva). Em ambos os casos, os resultados revelam diferenças marcantes entre a linguagem gerada por IA e a linguagem humana. Na pesquisa sobre ILE, os textos produzidos por IA são mais informacionais, abstratos e impessoais, enquanto os textos humanos demonstram consciência interpessoal, posicionamento e engajamento. No estudo sobre letras de música pop, a IA gera discursos de empoderamento moralizado que convertem os conflitos históricos retratados no rap como se fossem narrativas generalizadas de virtude. Em ambos os casos, a IA demonstra sinais de déficit de registro (incompreensão da variação de registro devido a conhecimento limitado da constituição linguística dos registros usados na comunicação humana) e de metamorfose de registro (tendência de geração de textos com aparência de um registro e constituição linguística de outro).

Palavras-chave: *Linguística de Corpus; Inteligência Artificial; Registro; Análise Multidimensional*

1. Introduction

This article advances the claim that research on AI-generated language benefits from a register-based Multi-Dimensional (MD) analysis and applies that framework to the description of the grammar and discourse of AI-produced texts. Two sample studies illustrating this application—namely, Shimazumi and Berber Sardinha (2025) and Delfino and Berber Sardinha (2025)—are presented. These studies share the following common goals: (1) to demonstrate how corpus linguistics can be applied to describe the linguistic characteristics of texts generated by large language models (LLMs); (2) to identify the dimensions of variation underlying AI-generated texts through MD analysis; (3) to determine the corresponding dimensions of variation for a corpus of human-authored texts using the same analytical procedure; and (4) to compare AI-generated and human-authored texts based on their distributions across these dimensions.

This register- and MD-based approach rests on the methodological assumption that linguistic variation is best described relative to specific registers rather than to general language patterns (Biber, 2012). Based on this premise, comparisons between AI and human language should hold register constant so as to narrow the range of possible linguistic realizations, in turn

revistadelta
41.4
2025

allowing the resulting dimensions to reflect genuine register distinctions rather than incidental variations (Berber Sardinha, 2024; Collentine & Berber Sardinha, in press).

2. Corpus Linguistics and Register

According to Biber et al. (1998), corpus-based analysis is defined by four essential characteristics. First, it is empirical, relying on the observation and analysis of actual patterns of use in natural texts. Second, it requires a large and principled collection of naturally occurring language data (a corpus) as the foundation for analysis. Third, it makes extensive use of computational tools, both automatic and interactive, to handle large datasets and facilitate complex analysis. Finally, it involves both quantitative and qualitative methods, combining frequency-based measurements with interpretive insights. These characteristics allow corpus-based studies to identify patterns not otherwise observable, process large amounts of linguistic data consistently, and ensure analytical transparency. Importantly, the approach goes beyond simple frequency counts by emphasizing the need for functional interpretations of linguistic patterns, thereby contributing to our understanding of how language is actually used.

Biber and Conrad (2009) define register as a variety of language that is functionally associated with a specific situational context. This definition highlights that language varies according to the purpose of communication, the relationship among those participating in the communication, the channel of communication (spoken or written), and participants' level of shared background knowledge. Registers are determined not by subject matter, but by the functional demands imposed on language by the situation. These demands lead to consistent patterns of linguistic choices across texts that share the same communicative context. For example, a classroom lecture, a legal contract, and a casual conversation differ in terms of purpose, audience, and mode; these contextual differences are reflected in the texts' grammatical and lexical features.

The concept of register rests on the observation that linguistic features co-occur in predictable ways when the communicative context is held constant. Instead of analyzing features in isolation, the goal is to identify groups of features that appear together because they are functionally related to the situation. For instance, the frequent use of nouns, long noun phrases,

and passives in academic writing arises not from stylistic choice, but from the need to express information densely and impersonally. Similarly, personal pronouns, contractions, and active verbs tend to occur in conversation, reflecting its interactive and real-time nature. These feature groupings are not random; they reflect how language is adapted to serve communicative functions under specific conditions.

Registers are not recognized on the basis of isolated linguistic items, but through co-occurring groups of features that are functionally related to the communicative situation (Biber & Conrad, 2009). For instance, in conversation, interactive features such as second-person pronouns, contractions, and discourse markers are prevalent. In contrast, written academic prose favors nominalizations, attributive adjectives, and passive constructions, reflecting its informational and non-interactive nature. The link between situation and linguistic form is not prescriptive but descriptive: Language adapts to the functional requirements of the context, and register captures this adaptation through the empirical observation of feature co-occurrence across texts. This understanding of register serves as the foundation for corpus-based analyses of functional language variation.

To investigate such context-sensitive co-occurrence patterns, a register-focused corpus linguistic analysis must be designed around appropriate observational units. The choice of unit, whether individual linguistic features, entire texts, or full corpora, directly affects the kinds of generalizations that can be made about language use. In particular, research aiming to describe how communicative situations shape recurring feature combinations must adopt a design that reflects register as a functionally motivated linguistic phenomenon rather than as an arbitrary collection of features or topics.

Reflecting this concern, Biber (2012) distinguishes among three types of corpus-based research designs, each defined by the unit of analysis it adopts. Type A studies focus on individual linguistic features, treating each token (e.g., every instance of a relative clause or verb form) as an observation. This design is typically used to study proportional preferences among linguistic variants (such as active vs. passive constructions) and is analyzed using categorical, non-parametric techniques. However, Biber notes that this design cannot provide information about how often these features occur across texts or registers. As it disregards the distribution of features at the text level, it is unsuited for characterizing register-based variation.

revistadelta
41.4
2025

Type B studies treat individual texts as the unit of analysis. This design allows researchers to calculate frequency rates for each text and explore how features co-occur within and across communicative contexts. It is the foundation for multi-dimensional approaches, such as those Biber developed in his earlier work, to register analyses. Because Type B studies reflect the situational conditioning of language use and permit parametric statistical analyses (e.g., factor analysis), Biber considers this design essential for investigating register as a predictor of linguistic variation.

Type C studies, in contrast, use the entire corpora or subcorpora as the unit of observation. Although they can reveal general patterns, their limited number of observations prevents the use of inferential statistics. For this reason, Biber argues that Type C designs are inadequate for generalizing findings beyond the sampled corpora. In sum, for studies concerned with register, only Type B designs allow for empirically grounded comparisons of linguistic behavior across contexts. For this reason, the sample studies reported herein operationalize a Type B design.

3. Generative Artificial Intelligence, Register Deficit, and Register Metamorphosis

Generative AI systems produce language by estimating the most probable next word based on the sequence of words that precedes it. In this way, it capitalizes on the principle that language use is a system of probabilities (Berber Sardinha, 2000b; Halliday, 1991), where many possible linguistic choices are possible, but they do not occur with equal frequency. Hence, to determine the probability of a given feature, one must observe empirical frequencies in a corpus. AI systems "learn" how to use language by mapping these probabilities from corpora and using the most probable choices to generate fluent idiomatic text.

This mechanism can be understood through Sinclair's (1991) idiom principle, according to which speakers (and, by extension, systems trained on language data) typically do not construct utterances word by word using grammatical rules alone (what Sinclair calls the open-choice principle). Instead, they rely on pre-constructed phrases and common word combinations that are stored and retrieved as units. These collocational patterns are not necessarily idioms in the traditional sense but are repeated co-occurrences that have become conventionalized in language use.

Generative AI models follow this principle by favoring high-frequency, conventional lexical sequences when producing text. Because the model learns by observing large-scale distributions, it acquires and reproduces patterns that are statistically probable (i.e., those that instantiate the idiom principle). For example, it tends to produce combinations like "strong argument," "deep concern," or "take into account" because these sequences appear often and are predictable in context. Consequently, generated texts are characterized by collocational regularity and a preference for phrases already common in the training corpora.

LLMs are trained on texts scraped from the internet, which are not typically organized or annotated for register. This situation presents a fundamental limitation when it comes to learning text varieties. Registers are defined not by topic or source alone, but by recurring configurations of linguistic features shaped by situational variables such as communicative purpose, mode, and audience. Because LLMs are exposed to these configurations only incidentally and without explicit contextual labeling, their capacity to internalize register distinctions is limited.

This limitation is particularly pronounced for underrepresented registers. Although academic texts, news articles, Wikipedia entries, and blog posts are abundantly available in online data, many situationally defined text varieties, such as spoken classroom interaction, personal letters, and student essays, are either absent or drastically underrepresented. As a result, LLMs disproportionately learn the linguistic patterns of high-frequency registers and may approximate other varieties by analogy, drawing from what is statistically most likely across the entire text base rather than what is typical of specific registers.

Without register-level annotations, LLMs are unable to develop an awareness of boundaries between varieties or simulate text productions that are faithful to specific registers. This limitation raises methodological and epistemological concerns, especially when these models are evaluated for their ability to generate or understand texts across different communicative contexts (Berber Sardinha, 2024). Unlike a corpus-based register analysis, which begins with clearly defined situational categories and analyzes co-occurring linguistic features accordingly, LLM training conflates all input into a largely undifferentiated whole (a "blob"). Consequently, the output may reflect generic forms of language that are not representative of the registers of interest.

revistadelta
41.4
2025

The issue is not simply whether the training data contain enough examples of a given variety, but whether the model has access to the contextual information needed to associate linguistic choices with communicative functions. Register variation is not random; it reflects systematic associations between situational parameters and language use. Because LLMs are trained without annotations that mark these parameters, they do not learn to associate specific linguistic patterns with particular communicative settings. Instead, they rely on overall frequency and co-occurrence across the entire dataset, which often results in generic "whole language" modeling that does not capture register distinctions.

To refer to this problem of insufficient text variety awareness in LLMs, we propose the term *register deficit*, which highlights the widespread absence of register-specific knowledge in LLM training data. Because LLMs are trained on massive, unlabeled corpora where reliable register distinctions are often missing, they cannot learn to associate patterns of linguistic co-occurrence with specific registers.

Thus, when tasked to produce a particular register, LLMs tend to produce texts from different registers, a process that Shimazumi and Berber Sardinha (2025) refer to as *register metamorphosis*. This phenomenon refers to the way AI-generated texts can alter or distort the register features typical of human communication, which is particularly evident when LLMs generate text that superficially resembles one particular register, yet diverges markedly in terms of its linguistic makeup and communicative function. For instance, AI-generated conversations often mimic dialogic forms (e.g., question–answer sequences) without engaging in a genuine interpersonal exchange, resulting in a simulated interaction that lacks authentic pragmatic engagement, leading to conversations that resemble interviews, FAQ lists, or scripted performances. This shift produces a kind of "frankenregister," or a hybridized or distorted register that may appear plausible on the surface but does not fulfill the functions expected of the genre or situation it imitates.

This process of register distortion has already been discussed in the context of EFL teaching materials. Gilmore (2004, p. 363) points out how EFL textbook dialogs often appear contrived due to overly tidy turn-taking, a complete absence of hesitation or repair phenomena, and abrupt shifts in topic with no response to what has been said. The resulting interaction lacks the spontaneity and contingency that characterize real conversation and instead produces an artificial and decontextualized exchange. Similarly,

Carter (1998, p. 47) observes that these dialogs reflect an idealized communicative world aligned with a "can-do" ethos, where speakers are invariably polite, problems are absent, and the structure resembles a quiz show or courtroom interrogation more than an actual interpersonal exchange.

Because of this lack of naturalness, such language teaching texts are considered artificial, as are texts generated by LLMs, which often exhibit a similar displacement from genuine register contexts. In the case of LLMs, this distortion arises from the absence of access to situational parameters or authentic interactional cues. In contrast, textbook authors intentionally produce these effects in order to foreground structurally convenient language, such as particular grammatical constructions or vocabulary items. In both cases, the resulting register features do not correspond to those found in naturally occurring settings.

4. AI Grammar: The Case of EFL Textbooks

To look at AI grammar, we turn to Shimazumi and Berber Sardinha (2025), who compared texts used as writing models in EFL textbooks produced by both human textbook authors and AI. The human-authored texts comprise the English Language Teaching Texts (ELTT) corpus, while the AI-generated texts form the AI-ELTT corpus. Both corpora consist of the same registers, with the same text counts per register.

In the instructional materials, these texts function as model texts, that is, examples used in writing activities to serve as templates for producing specific text varieties. They appear in both EFL exam preparation textbooks and general language teaching textbooks. In the first case, the textbooks present the texts as good examples of exam writing tasks, typically having received high scores from examiners. As such, they serve as reference points for examiner training, helping to standardize the identification and assessment of high-quality student writing. In the second case, the books present the texts as examples of particular registers that students are expected to emulate in their own writing, usually as part of learning objectives in classroom activities. In both contexts, the texts serve as models in the sense that they are seen (by examiners or by students) as exemplary instances of writing that fulfill the intended pedagogical or evaluative purposes.

revistadelta
41.4
2025

The motivation for the study is that teachers are increasingly using AI-generated texts for such tasks as preparing activities for classroom use (Ash, 2025), which include generating example texts that meet their class goals. Traditionally, textbook texts, such as model essays, dialogs, or descriptive paragraphs, are typically decontextualized and purpose-built not to replicate authentic communicative exchange, but rather to illustrate specific grammatical structures, lexical items, or text types. Their primary function is not to reflect naturally occurring language, but to serve as controlled exemplars for instructional use. As such, they are intentionally fabricated to highlight certain linguistic features, often simplifying or exaggerating register traits to meet pedagogical aims. From that perspective, incorporating AI-generated texts into the classroom would not pose a problem, as both they and textbook texts are artificially constructed to fulfill instructional purposes. If anything, AI-generated texts may offer the additional benefit of being customized by the teacher to fit their specific classroom needs, unlike pre-written textbook materials.

Thus, the comparison between human-authored EFL textbook texts and AI-generated texts is not a comparison of natural versus artificial language use, but rather of two forms of artifice: one human, the other machine. If the AI can reliably produce texts that replicate the register features of human-produced textbook examples, then its artificiality may be functionally equivalent to that of pedagogical texts authored by humans. If not, then instructional AI-generated language must be treated as a distinct form of constructed language from that found in textbook text models.

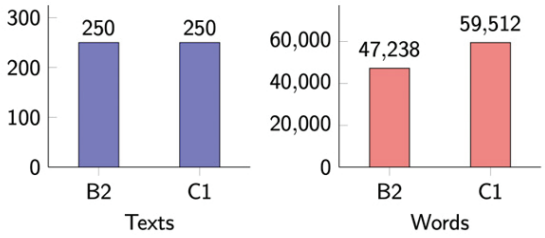
Either way, both AI-generated and human-produced artificial texts contrast with naturally occurring texts. Natural texts are those that emerge from real-life communicative situations, such as conversations, interviews, and emails. These texts are contextually grounded and reflect the constraints and purposes of communication for a specific context. In contrast, artificial texts (whether produced by AI or human authors) are deliberately constructed outside of an actual communicative event. The difference is that human-authored artificial texts are created by individuals who have direct experience with, or at least some understanding of, the communicative contexts that they aim to simulate. These texts may even be simplified adaptations of naturally occurring language. AI-generated texts, on the other hand, are necessarily artificial, as their author (the AI agent) has never participated in, nor directly observed, the real-world communicative situations the texts are

meant to represent. The texts are generated without situational awareness or social interaction. They are simulations of language that do not originate from lived experience or communicative necessity.

If a high degree of similarity is observed, it suggests that both types of model texts represent comparable forms of artificiality. If not, it would point to distinct types of artificiality between human-produced and machine-generated registers.

The ELTT corpus captures a sample of the language of EFL textbooks and proficiency exam preparation materials. It includes 500 texts drawn from 43 different sources, the majority of which are print-based (91%), with a smaller portion (9%) having been collected online. These sources span a 25-year period (1996–2021) and reflect a global, international market targeting young adult and adult learners (see Table 1). The corpus text count is balanced across two proficiency levels (B2 and C1, 50% each; see Figure 1), despite the difference in word count. The sample averages 213.7 words per text, totaling 106,840 words. The texts follow communicative, task-based, prescriptive, and school-centered approaches, with some being corpus-informed. The corpus includes 16 different registers (Table 1) and focuses on British English and General English varieties, with materials published by major educational publishers and websites.

Figure 1 – Breakdown of the ELTT Corpus by Proficiency Level



revistadelta
41.4
2025

Table 1 – Breakdown of the ELTT Corpus by Register

Register	Text Count	Word Count
Essay	93	22,476
Summary	62	10,700
Formal letter	60	12,410
Informal letter	59	10,927
Report	58	12,942
Review	49	11,599
Article	28	5,616
Story	24	4,363
Proposal	23	6,246
Informal email	15	3,057
Formal email	9	1,179
Description	6	1,358
Other	6	1,694
CV	4	839
Brochure	3	777
Instruction	3	657
Total	**500**	**108,840**

Figure 2 – Breakdown of the ELTT (Human) and AI-ELTT (AI) Corpora by Register Word Count

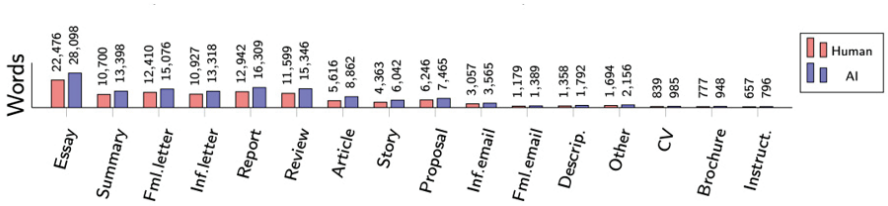


Figure 2 presents the distribution of word counts across registers in the ELTT (human) and AI-ELTT (AI) corpora. The AI corpus contains more words in most registers, particularly *Summaries* (13,398 vs. 10,700), *Informal Letters* (15,076 vs. 10,927), and *Reports* (16,390 vs. 12,942). In contrast, the human corpus includes more words from *Essays*, with 22,476 compared to 20,988 for AI. Some register categories, such as *Proposals* and *Articles*, are more evenly represented. Registers with fewer texts (e.g., *CVs*, *Brochures*, *Instructions*) show relatively small differences.

The generation of the AI text models was carried out as follows. The original textbook prompts were often insufficient on their own as they were either too brief to convey the full pedagogical context or relied on images that the AI could not interpret reliably. This situation arises from the fact that textbook model texts are typically the endpoint of a whole lesson and, therefore, the task assumes a great deal of information not spelled out in the activity prompt itself. We did try out these textbook task prompts as AI prompts, but the results were not satisfactory: The AI did not have sufficient information to produce a reliable counterpart to the textbook model.

To address this problem, the authors implemented a two-part prompt system. In Prompt #1, the AI was asked to read an actual textbook model text and produce a task prompt suitable for instructional use. In this way, the AI included in that prompt the information it deemed necessary for someone to carry out the task. Then, after a several-hour delay and after being told to forget the source text used with Prompt #1, the AI was given Prompt #2, which told the AI to read that task prompt and generate a new model text. This process was carried out through API calls using ChatGPT-3.5.

To conduct a traditional MD analysis, researchers generally follow a series of steps (Egbert & Staples, 2019; Friginal & Hardy, 2019). The process begins with corpus design and collection, a stage that requires the compilation of a representative sample of texts covering the register (or registers) under investigation. Because the text is the unit of observation, each file must contain a complete and self-contained text. The number of texts must exceed the number of linguistic variables by at least a factor of five to ensure statistical adequacy. Alongside the texts, researchers record situational metadata such as communicative purpose and production setting, as these contextual variables are subsequently used for interpreting the factors.

The next stage involves identifying, tagging, and counting linguistic features. In traditional MD analysis, these features are typically lexico-grammatical, including syntactic, structural, and morphological elements that can be automatically retrieved through part-of-speech tagging or specialized tagging tools. Frequency counts are produced for each feature in every text, and the results are normalized (usually per 1,000 words) to account for variations in the text length. The normalized frequencies are then standardized into z-scores so that features contribute equally to the subsequent factor scores.

revistadelta
41.4
2025

The main statistical procedure in an MD analysis is factor analysis, which is a multivariate statistical procedure that identifies groups of correlated linguistic features—that is, they tend to appear together within and across texts. These groups, or factors, represent empirically derived patterns of linguistic co-occurrence. Features with low communalities are excluded, the appropriate number of factors is determined from the scree plot, and a rotated factor solution (commonly using Promax rotation) is obtained to allow intercorrelations among factors.

The factors are then interpreted qualitatively as dimensions of variation. Each dimension has two poles, defined by the linguistic features that load positively and negatively on the factor, which correspond to distinct communicative functions. Traditionally, these two poles (or ends of the continuum of variation) are called "positive" and "negative," but these terms carry no judgment or implication of worth; they could be switched around if needed. Through careful inspection of many texts (usually high-scoring texts) and the consideration of situational contexts, researchers assign descriptive labels that capture the communicative orientation underlying the dimension.

Finally, dimension scores are computed for each text by summing the standardized scores of the features associated with each factor. These scores quantify the position of each along each dimension and can be statistically compared across registers using analyses of variance (ANOVAs) or tested for predictive accuracy using discriminant function analysis. In the end, the set of dimensions represents a quantitative model of register variation that describes how registers differ from and are similar to each other.

It is important to stress two fundamental properties of dimensions. First, as previously noted, texts are distributed continuously along each dimension, rather than being classified into one pole or the other. Figure 3 illustrates this property: the distribution is based on the scores assigned to each text, which are ranked along a continuum from one pole to the other. Second, all dimensions are present in every text, which is what gives the approach its multi-dimensional character. However, these dimensions are present to varying degrees, reflecting the distinct linguistic profile of each text. Figure 4 sketches this principle: Different sizes of text segments represent different degrees of dimension incidence. However, in reality, dimensions are not confined to discrete segments; rather, different dimensions can occur within the same sentence.

Figure 3 – Distribution of Texts Along the Two Poles of a Dimension

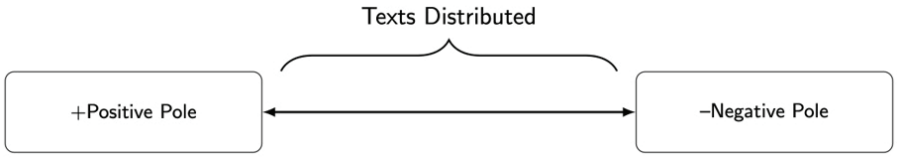
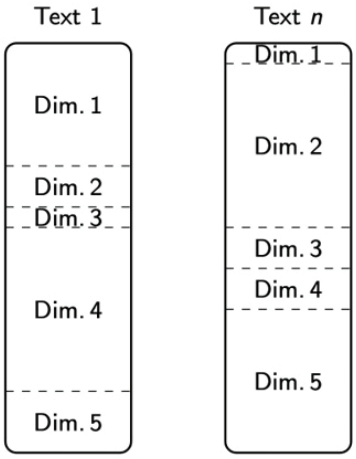


Figure 4 – Incidence of Dimensions in Individual Texts



The MD analysis of the (human-authored) ELTT corpus resulted in five dimensions (Shimazumi & Berber Sardinha, 2023), only the first of which is presented here due to space constraints. This dimension, which was labeled "Persuasion, speaker engagement, and personal opinion vs. Expression of analysis and technical information," comprises the factor loadings shown in Table 2.

revistadelta
41.4
2025

Table 2 – Dimension 1 Loadings

Positive Features	
Feature	**Loading**
Sum stance adverbs	.57962
Emphatics	.52316
Contraction	.50323
First-person pronoun/possessive	.45759
Verb (not including auxiliary verbs)	.45175
Pronoun "it"	.42205
Place adverbials	.38777
That-deletion	.38573
Second-person pronoun/possessive	.33459

Negative Features	
Feature	**Loading**
Nominalization	-.57626
Word length	-.57618
Preposition	-.50109
Abstract/process nouns	-.42603

On the positive side of the dimension, the highest loadings are for stance adverbs and emphatics, which indicate a strong presence of speaker attitude and evaluation. These features are commonly found in texts that aim to persuade or express personal opinion. Other high-loading features, such as contractions, first- and second-person pronouns, and place adverbials, reflect an informal, interactive style where the speaker or writer is actively engaged with the audience. The use of "it," main verbs, and *that*-deletion enhances the conversational tone typically associated with such registers as argumentative essays, personal letters, and oral presentations.

In contrast, the negative pole includes features typical of abstract, formal, and information-focused texts. High loadings for nominalization and abstract/process nouns indicate a preference for dense, noun-based expressions that are characteristic of academic and technical registers. These features condense actions or processes into compact noun phrases, thereby increasing information density while reducing personal involvement. Furthermore, longer words and prepositions suggest greater lexical specificity and syntactic elaboration. Such features are common in scientific

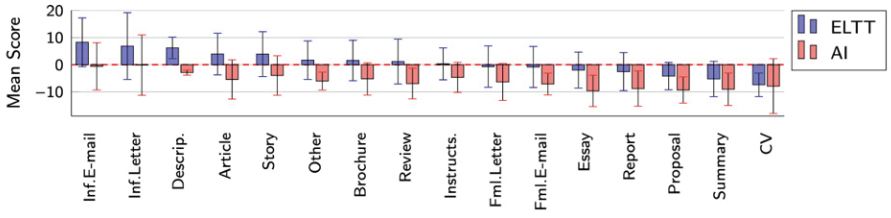
articles, reports, and other technical or expository texts that prioritize clarity, objectivity, and precision over personal stance or engagement.

Thus, this dimension reflects a functional opposition between speaker involvement and engagement on one end and analytical detachment and technical precision on the other. Texts that score positively are oriented toward persuasion and interpersonal communication whereas those scoring negatively are oriented toward the impersonal transmission of information.

The mean dimension scores for Dimension 1 reveal sharp differences between the human- and AI-generated texts (Figure 4). On the positive pole of Dimension 1, human-authored texts show higher mean scores across the board. These are contexts where writers are expected to express opinions, engage readers, or adopt a personal tone. In contrast, AI-generated texts are concentrated nearer to zero or extend further toward the negative pole across nearly all registers. This includes registers where human writers show personal involvement, indicating that AI fails to fully realize the subjective or persuasive features that characterize those registers. Instead, AI-generated texts tend to adopt a more neutral or formal style even in situations where informality or stance-taking would be instructionally adequate.

At the negative end of the dimension, both AI- and human-generated texts show similar scores in registers such as essays, reports, and CVs, where impersonal and technical exposition is typically expected. However, the AI-produced texts tend to be even more impersonal than the human-produced texts, exaggerating the expository style. The general pattern across both poles is that AI-created texts underrealize the positive features of Dimension 1 and overrealize the negative ones, thereby resulting in a discourse that is consistently more formal, abstract, and impersonal than that produced by human materials writers.

Figure 5 – Mean Dimension 1 Scores (Persuasion, speaker engagement, and personal opinion vs. Expression of analysis and technical information)



revistadelta
41.4
2025

Example 1 illustrates how AI reversed the polarity of a text (C1 level informal letter). The human-generated texts score positively on average on Dimension 1, meaning they express "persuasion, speaker engagement, and personal opinion"; however, this particular AI-produced text, like many other AI-created texts, scored negatively, thereby expressing analysis and technical information instead.

> **Example 1**
> Zoos have always sparked a controversial **debate among** conservationists, animal lovers, and administrators alike. According **to** the author, zoos are not black and white entities but, instead, a nuanced **amalgamation of** pros and cons. Commendably, zoos contribute **to education, conservation,** and **research**, thus offering an invaluable platform **for** learning and preserving various species. However, **despite** these apparent benefits, zoos also face intense **criticism**, evident **in** their perceived impact **on** an animal's **lifespan** and behavior. Advocates **of** zoos contend that they provide an unmatched learning **environment**, allowing millions to **encounter** species (…)

In this example, the frequent use of nominalizations such as *deterioration*, *erosion*, and *expectation* shifts the focus away from personal experience or agency and toward abstract processes. The presence of abstract nouns such as *communication*, *compromise*, and *responsibility* reinforces this distancing effect. In turn, prepositions integrate information by combining nouns into larger informational units (e.g., platform *for* learning; impact *on* an animal's lifespan; advocates *of* zoos). These features all point away from persuasion or speaker involvement and toward the expression of analysis and technical information, resulting in a text that is markedly impersonal and detached.

5. AI Discourse: The Case of Song Lyrics

In this section, we turn to an exploration of the ways in which generative AI produces or reproduces discourse. As these models are trained on vast textual corpora, they inevitably absorb patterns of discourse circulating in their input data, later reshaping them through instruction-based fine-tuning. What remains unclear, however, is which discourses are being incorporated and how they are rearticulated when the model is asked to generate texts belonging to specific registers.

First, we need to define what we mean by discourse, as this term has different meanings in linguistic and social theory. We see discourses as

"socially shared, socially situated representations of real-world phenomena conveyed implicitly through language use" (Berber Sardinha & Fitzsimmons-Doolan, 2025, p. 4). Similarly, Burr (1995, p. 48) defines discourse as "the set of meanings, representations, [and] statements that together produce a particular version of events," emphasizing that multiple, competing discourses coexist. Hajer (1993, p. 44) calls discourse "an ensemble of ideas, concepts, and categorizations that are produced, reproduced, and transformed in a particular set of practices," while for Baker and McEnery (2015, p. 5), discourses are "ways of looking at the world, of constructing objects and concepts in certain ways." These definitions have in common the notion that discourses are not individual opinions, but collective interpretive systems that both enable and constrain what can be said or thought about a phenomenon. Because they are socially situated, discourses emerge from practice and experience, and because they are shared, they carry authority in shaping meaning.

Referring to AI-generated discourse as we do is possible because LLMs are trained on millions of texts that already instantiate human discourses and, when prompted, they output text embedded in those discourses, particularly those that were predominant in the input data. In addition, through human supervision in the training process, LLMs are trained to respond in particular ways and to suppress certain manners of expression, which steers them to prefer some underlying discourses over others.

LLMs have no agency, but agency is not a prerequisite for discourse production as discourses are the product of social and historical conditions and circulate independently of human beings' intentions. Therefore, a machine without consciousness can still generate text that fits within particular discourses and participates in ideological reproduction, even if the generator lacks awareness of that process.

When AI systems generate language, they generate the linguistic materiality through which discourses become visible, thereby placing discourses in motion. They do not create them intentionally, but rather passively by responding to a user's prompts. Nonetheless, once released, these texts enter social circulation: They are read, shared, and responded to, thereby influencing ongoing meaning production.

From this exposition, it should be clear that discourse is a different construct from topic. As Hall (1992, p. 92) observes, discourses "provide a language for talking about—i.e. a way of representing—a particular kind

revistadelta
41.4
2025

of knowledge about a topic." Hence, topics refer to what is talked about, an immediate, surface-level manifestation of subject matter. In contrast, discourse refers to why and how a topic is talked about as well as who speaks, when, and under what conditions. In other words, discourse encompasses the social and historical context that gives a topic its meaning. Thus, "[w]hen statements about a topic are made within a particular discourse, the discourse makes it possible to construct the topic in a certain way. It also limits the other ways in which the topic can be constructed" (Hall, 1992, p. 92).

To illustrate how we can employ the MD analysis to detect AI-generated discourses in a specific register, we report on a previous study that engaged AI to generate English pop music lyrics (Delfino & Berber Sardinha, 2025). In a previous MD analysis, Delfino (2022) unveiled the four major discourse dimensions in human pop music, which include (1) a hedonistic discourse associated with materialism and superficiality; (2) a relational discourse, centered on alterity and interpersonal dynamics; (3) a spiritual discourse, linked to mysticism and transcendence; and (4) an idealistic discourse, expressing romanticism and personal quest.

To explore AI-generated discourse in pop lyrics, Delfino and Berber Sardinha (2025) created a comparable corpus of human and AI-produced pop music lyrics (see Table 3). This corpus is evenly split into two major sources: 2,000 human-authored lyrics and 2,000 AI-generated lyrics. The AI lyrics are further broken down equally into four source models: ChatGPT-4, Gemini, Llama2, and an uncurated version of Llama2. Human texts total more than 640,000 words with a mean length of 320.09 words and a relatively high standard deviation of 202.00, indicating considerable variability in length. For AI, despite the equal number of texts from each model (500), the total word count and mean text length vary considerably. Gemini produces the longest texts on average (363.77 words), followed by ChatGPT-4 (336.01). In contrast, Llama2 and its uncurated version yield much shorter lyrics (205.33 words).

By musical genre, the corpus is evenly divided among country, pop, rap, rock, and soul, with 800 texts each. Rap lyrics are markedly longer than those of other genres, with an average length of 498.03 words and a high standard deviation of 238.19. In comparison, country, pop, rock, and soul texts are more similar in length, ranging from approximately 244 to 274 words on average, with lower standard deviations.

Table 3 – Music Corpus Design

Source	Texts	Words	Mean Text Length	Std. Dev.
Human	2,000	640,180	320.09	202.00
ChatGPT-4	500	168,007	336.01	111.45
Gemini	500	181,887	363.77	198.55
Llama2	500	124,253	248.51	63.24
Llama2 Unc.	500	102,666	205.33	75.92
Total	**4,000**	**1,216,993**	**304.25**	**174.33**

Genre	Texts	Words	Mean Text Length	Std. Dev.
Country	800	195,778	244.72	109.10
Pop	800	219,338	274.17	127.75
Rap	800	398,424	498.03	238.19
Rock	800	203,667	254.58	95.15
Soul	800	199,786	249.73	102.98
Total	**4,000**	**1,216,993**	**304.25**	**174.33**

Unlike the previous study, which used the traditional version of MD analysis (Berber Sardinha, 2000a; Biber, 1988), the study reported here adopts lexical MD analysis (LMDA), which is specifically designed to identify latent discursive constructs, such as ideologies and representations, through patterns of lexical co-occurrence within and across texts (Berber Sardinha & Fitzsimmons-Doolan, 2025). Although both approaches share the foundational assumptions and statistical procedures of MD analysis, LMDA differs by focusing exclusively on lexical features rather than on grammatical features. In LMDA, actual words (e.g., lemmas) are treated as variables, and the dimensions that emerge through factor analysis are interpreted not as communicative functions, but as macro-level discourse formations that reveal underlying discourse structures encoded in lexical choice. LMDA thus shifts the interpretive focus from functional variation (as in traditional MD analysis) to discursive variation.

The dimensions identified by each approach are both sets of correlated linguistic features, but the parameters revealed by each are different. In traditional MD analysis, dimensions correspond to underlying functional parameters of grammatical variation. They emerge from correlations among lexico-grammatical features, such as morphological classes, tense, semantic categories, and clause structure, and reflect communicative functions associated with particular registers. Each dimension represents a functional

revistadelta
41.4
2025

continuum along which texts vary depending on how language is used to achieve situational goals.

In LMDA, in contrast, dimensions correspond to macro-level discursive constructs. They are derived from patterns of lexical co-occurrence and point to latent discourse structures rather than communicative functions. The dimensions do not correspond to register-based variation in a functional sense, but rather to the discourses, representations, or ideologies embedded in the lexical choices of texts. Although both types of dimensions rely on the same statistical logic, their interpretive frameworks and the kinds of meaning they reveal are fundamentally distinct.

The LMDA was conducted based on the following steps. First, the corpus was tagged using the TreeTagger (1994) to identify lemmas for each word. Lexical features were then selected based on their frequency and dispersion across texts through keywords. The selected features were counted in each text and saved as a matrix of normalized frequencies. Pairwise correlations among the lexical variables were then computed to identify co-occurrence patterns across texts, serving as the input for an initial factor analysis, which extracted preliminary factors representing sets of co-occurring lemmas. After that, the factor solution was subjected to rotation to improve interpretability, and weakly loading features were dropped to retain only the most relevant contributors to each factor. Next, qualitative micro analyses were conducted on a large number of texts; based on these analyses, interpretive labels were assigned to each dimension to capture the discursive constructs reflected in the lexical groupings. Finally, dimension scores were computed for each text by adding the standardized counts of lexical variables that loaded on each factor, which enabled comparisons across the corpus categories along each dimension.

As mentioned, to select the items to be included in the LMDA, a keyword analysis (Egbert & Biber, 2019; Scott, 1997, 2000) was carried out to determine the key lemmas associated with each source and musical genre combination. Each pair (e.g., Human+Rock, ChatGPT+Rap, Gemini+Soul) was treated as a target corpus, which was compared against a reference corpus consisting of all remaining texts in the corpus. The purpose was to identify lemmas that were statistically overrepresented (positive keywords) in each target corpus relative to the rest of the corpus.

The process began by computing normalized text counts for each lemma, that is, the number of individual texts in which each lemma appeared.

These text-level counts were used for the statistical comparison: If a lemma occurred in significantly more texts in the target than in the reference, it was classified as a positive keyword. If it occurred in fewer texts or in comparable distributions across both corpora, it was not considered a keyword. Log-likelihood statistics were used to determine keyness; these statistics compare observed and expected frequencies in categorical data.

To illustrate how human-authored and AI-generated lyrics differ with respect to the underlying discourses, we focus on a single dimension from this study (i.e., Dimension 1), labeled as "Discourse of Empowerment and Resilience in Rap Narrative vs. Discourse of Tragic Romance, Pursuit of Happiness, Love, and Loneliness." A portion of the loadings for this dimension appears in Table 4.

Table 4 – Excerpt of Loadings for Dimension 1

Positive pole: Discourse of Empowerment and Resilience in Rap Narrative
voiceless (1.19), pioneer (1.18), storyteller (1.18), warrior (1.16), empower (1.13), healer (1.12), justice (1.11), vivid (1.10), barrier (1.10), quo (1.10), wordsmith (1.09), norm (1.08), injustice (1.08), survivor (1.08), challenge (1.07), boundary (1.06), dreamer (1.05), renegade (1.04), weapon (1.04), syllable (1.03), inspire (1.02), unity (1.02), fighter (1.01), defy (1.00), lyrical (1.00), unheard (1.00), status (1.00), knowledge (1.00), sharp (.99), captivate (.99) […]
Negative pole: Discourse of Tragic Romance, Pursuit of Happiness, Love, and Loneliness
loving (-.60), telephone (-.57), wonderful (-.56), lovely (-.55), sweetheart (-.55), no (-.54), hotel (-.54), whoa (-.54), buddy (-.53), afternoon (-.53), daylight (-.53), convince (-.52), bedroom (-.52), promised (-.52), beach (-.52), velvet (-.51), suicide (-.51), boat (-.50), satisfied (-.50), cigarette (-.50), pickin (-.49), sexy (-.49), drinkin (-.49), petal (-.49), waltz (-.49), hug (-.49), pourin (-.48), ashamed (-.48), loneliness (-.48), drag (-.48) […]

The positive pole is characterized by lexemes such as *voiceless, warrior, pioneer, survivor,* and *empower*, which construct a discourse of personal and collective assertion. These terms project an image of the speaker as both artist and agent of change, someone who channels struggle into creativity and transforms adversity into strength. The lexical co-occurrence patterns suggest that rap operates here as a vehicle for reclaiming voice and agency. Words like *justice, unity,* and *knowledge* indicate a moral and communal orientation while *storyteller, lyrical,* and *wordsmith* point to a self-conscious reflection on artistic identity. Together, these lexical elements express a discourse in which self-expression is framed as empowerment, resistance,

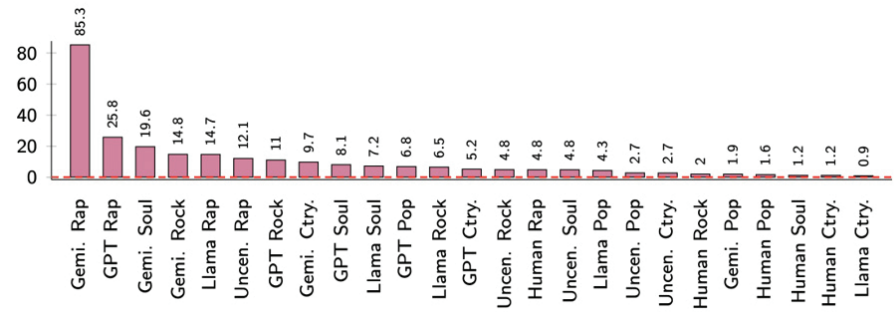
revistadelta
41.4
2025

and social awareness, revealing the ideological function of rap as both self-narration and cultural critique.

At the opposite pole, the lexemes *loving, sweetheart, loneliness, ashamed,* and *suicide* index an emotional world of intimacy, yearning, and loss. This discourse turns inward, focusing on affective experience rather than collective identity. It evokes the fragility of romantic attachment and the emotional toll of separation through recurrent references to places (*hotel, beach, bedroom*) and sensations (*velvet, cigarette, drinkin', pourin*'). The tone is nostalgic and melancholic, tied to the pursuit of happiness and the pain of its absence. Whereas the positive pole draws energy from struggle and transformation, this negative pole dwells on emotional exhaustion and the search for meaning through love and companionship. The contrast between the two poles captures a broader opposition between self-assertive, socially engaged expression and introspective, affect-driven reflection.

Dimension 1 scores (Figure 6) show a clear preference for the discourse of empowerment and resilience identified at the positive pole. The highest values are found in Gemini Rap (85.29) and ChatGPT Rap (25.75), which stand out sharply from all other categories. This distinction indicates that the lexical patterns most strongly associated with empowerment are an artifact of rap lyrics generated by large language models, especially Gemini. The high scores suggest that AI produces rap discourse in a way that is very different from human rap. This discourse of struggle, defiance, and self-assertion is quite rare in Human Rap, whose score is about 20 times lower than that of Gemini Rap. The AI model that comes closest to the human result is Llama Uncensored, which is licensed to output profanity; even so, it is more than twice as loaded on this discourse as Human Rap.

Figure 6 – Mean Dimension 1 Scores



Example 2 (gemini_rap_072.txt, score = 122) illustrates how this discourse of empowerment and resilience is manifested in an excerpt of a Gemini-generated rap lyric.

Example 2
Spittin' **truth** and passion, no compromise
My **lyrics paint vivid pictures,** like a **master** with hues
Unveiling stories that the world might choose
I'm a **storyteller,** a weaver of **tales**
Unraveling the tapestry of life's travails
My **words resonate,** like **thunder** in the night
Igniting hearts, as
I bring **truth** to light
With each **verse,**
I **unveil** a hidden **truth**
A glimpse into my soul, my **struggles,** my youth
I lay **bare** my scars, my **triumphs,** my pain
In **hopes** that my **words** might **inspire** and **sustain**
I'm a **warrior** in this **lyrical** arena

The example shows that the speaker adopts the persona of an artist-warrior. Repeated references to *truth, storyteller, verse,* and *wordsmith* construct a poetic voice of someone aware of the power of language as a medium of revelation and influence. The performer is positioned as a bearer of insight and emotion, "unveiling stories" and "igniting hearts," which connects artistic production to moral and affective renewal. The repeated emphasis on truth and light constructs AI rap as an instrument for uncovering hidden realities, creating this notion of singing expression as an expression of struggle and overcoming. Words such as *struggles, scars, triumphs,* and *pain* highlight vulnerability, yet they are immediately countered by terms like *inspire, sustain,* and *warrior*, which reassert hope and self-discovery. This alternation between an underspecified reality and inner strength creates this distorted discourse of self-empowerment.

The association of rap music with this discourse reveals a form of AI reductionism, as AI reinterprets rap from a political battle against discrimination into moralized perseverance and lyrical heroism. The song constructs the rapper as a "warrior" and "truth-teller" fighting a subjective personal struggle. AI reduces social struggle into the moral archetype of the poetic fighter overcoming adversity, recasting oppression as a metaphorical

revistadelta
41.4
2025

battle rather than a material and historical one—a politically hollow representation of empowerment.

In contrast, the discourse of "Tragic Romance, Pursuit of Happiness, Love, and Loneliness" (negative pole) is illustrated by the human rock lyric excerpt in Example 3 (human_rock_074.txt, score = -4; "I Don't Want to Miss a Thing" by Aerosmith). This romantic rock lyric is a stark contrast to the positive pole of the dimension as it communicates an intense focus on love, longing, and vulnerability. The speaker expresses a desire to remain perpetually present with a loved one, even forsaking sleep to avoid missing shared moments. This obsessive devotion and emotional dependency evoke themes of romantic idealization and existential loneliness, reinforcing the melancholic and affective qualities that define this pole. The emphasis on treasuring intimacy reflects a pursuit of happiness that is fragile and always at risk of loss.

Example 3
I could stay **awake**
Just to hear you breathin'
Watch you smile while you are sleepin'
While you're far away and dreamin'
I could spend my life in this sweet surrender
I could stay lost in this moment forever
Every moment spent with you is a moment I treasure
Don't wanna close my eyes, I don't wanna fall **asleep**
'Cause I'd miss you, **babe**

6. Discussion

6.1. Register Characteristics of AI-generated Language

The tendency of AI-generated texts to be informationally dense and less interactive than expected has been noted in the literature. Here, we turn to studies on student writing, a register within the same educational domain as our sample study (Shimazumi & Berber Sardinha, 2025), to demonstrate that, despite differences in method and scope, these studies converge on a similar characterization of AI-generated language.

Goulart et al. (2024) conducted an additive MD analysis based on Biber's (1988) work, combining it with situational analysis (Goulart, 2024) to verify the extent to which AI-generated student essays differed

from authentic student writing. The authors annotated the corpus for non-linguistic situational characteristics like audience, communicative purpose, and text layout to establish the contextual parameters. Repeated-measures ANOVAs and post-hoc tests were used to compare dimension scores across registers and sources, enabling the authors to evaluate the extent to which AI-generated assignments replicated the situational and linguistic properties of human student texts.

The analysis revealed that AI-generated texts only partially replicated the situational and linguistic features of student-authored assignments. In terms of situational characteristics, AI-written texts used more headings but included fewer references and in-text citations; communicative purposes were narrower, with AI essays adopting an explanatory mode, whereas student texts reflected a wider range of rhetorical intents. Personal narratives were the most similar across sources, with AI being more capable of reproducing the expected informal and experiential tone. Linguistically, AI-generated texts were more informational and less involved, less narrative, more explicit, less persuasive, and less abstract than student writing.

Jiang and Hyland (2025) focused on a more restricted linguistic phenomenon: the use of lexical bundles (Biber et al., 2004; Cortes, 2004, 2013) in student argumentative essays. Their aim was to identify differences in formulaic sequences between ChatGPT and human-produced essays as a way to detect AI-generated writing. The unit of analysis was the three-word bundle (a very frequent, non-overlapping three-word sequence), which was classified for grammatical structure and rhetorical function.

The authors found that ChatGPT essays exhibited greater formulaicity and predictability than student writing. Although AI-generated texts used fewer lexical bundles overall, those that were used showed a higher type/token ratio, indicating more rigid and repetitive phraseology. Structurally, AI essays favored simple noun/preposition bundles, especially of the "noun + of" type, whereas student essays used more clause-related and human-subject bundles. Functionally, AI bundles leaned toward descriptive expressions and discourse structuring devices, whereas student essays relied more on epistemic stance markers and expressions of personal perspective. These patterns reflect a key distinction: AI-generated texts lacked the rhetorical flexibility, argumentative depth, and subjective voice found in human-authored writing.

revistadelta
41.4
2025

Berber Sardinha (2024) used a register-diversified corpus that included conversations, academic writing, and news in addition to student essays. The study utilized an additive MD analysis to understand how closely AI-generated texts approximate human communication across the dimensions of linguistic variation from Biber (1988). The goal was to quantify the resemblance between AI and human production in different registers and to assess to what extent AI-generated texts could be considered human-like. Statistical tests, including ANOVA and linear discriminant analysis, were used to determine whether AI- and human-written texts differed significantly and whether such differences could predict authorship. The discriminant analysis offered a probabilistic model of classification.

The findings showed that AI-generated texts differed sharply from human-authored texts across all four registers. In the case of learner essays, the results showed that AI-generated texts were markedly different from those written by students: AI-produced texts were significantly more informational and less involved, with a greater density of content-heavy features and fewer signs of personal engagement. They were also less narrative and more expository, showed lower use of persuasive elements, and displayed reduced abstraction in academic contexts. Although student essays often integrated situation-dependent references and rhetorical variation, AI-generated texts tended toward surface-level coherence and predictability. A similar general trend was observed for all registers. Across all dimensions, AI-produced texts were consistently more informational, less narrative, and less persuasive. They lacked situation-dependent references and exhibited inconsistent abstraction patterns, such as an overuse of abstraction in conversations and a deficit in academic contexts. In general, AI failed to emulate human linguistic dimensional profiles, with a classification accuracy of 79% for AI-generated texts and 69% for human-written texts.

These studies share several key findings regarding AI-generated student writing with the sample study presented here (Shimazumi & Berber Sardinha, 2025). First, all studies point to a consistent informational density in AI-generated texts, which tend to be more expository and content-heavy than student-authored essays. This pattern is evident regardless of method, which strengthens this information-bearing tendency of AI writing.

Second, AI-written texts exhibit lower levels of involvement and personal engagement. Again, the previous studies reported a reduced use of

features associated with personal stance, such as private verbs or interactive pronouns, as well as an avoidance of bundles with explicit authorial presence.

Third, the studies converged on the observation that AI-produced texts are less persuasive and rhetorically varied. Both Berber Sardinha (2024) and Goulart et al. (2024) documented a lack of persuasive markers in AI-generated writing, especially in genres where persuasion is expected. Jiang and Hyland's (2025) findings reinforce this through the underuse of epistemic stance bundles by AI.

Finally, AI-generated writing was described as more rigid, predictable, and homogeneous. This finding emerged from Jiang and Hyland's (2025) study, where lexical bundles were more standardized and structurally simple. This rigidity contrasts with the rhetorical flexibility and contextual sensitivity found in the human-authored student essays in their study.

These findings point to interactive features such as persuasion and stance being harder for LLMs to emulate than the transmission of information. Interpersonal awareness is lacking in LLMs because they are software programs that have never actually taken part in real human conversations, debates, and so on. Interaction awareness requires the model to adopt a stance, present a point of view, anticipate counterarguments, gauge reactions, and so on. These elements are not directly captured through word sequence prediction, which forms the basis of AI language generation. In addition, the large corpora used for training AI do not include large quantities of interactive registers (especially spoken ones such as spontaneous conversation), and therefore the resulting models are not exposed to the linguistic features commonly used for interactive purposes.

6.2. Discourse Characteristics of AI-generated Language

The findings from the sample study presented here, which describe the discourses produced by LLMs in the generation of pop music lyrics, generally echo earlier research on AI discourse. Although these studies use different data and methods, they converge on a number of major characteristics of AI-generated discourse. The discussion in this subsection reviews selected studies in both corpus linguistics and natural language processing (NLP) and then considers how the findings from our sample study correspond to general trends in the literature.

revistadelta
41.4
2025

Breazu and Katsos (2024) explored how ChatGPT-4 reproduces journalistic discourse and whether its language reflects existing media ideologies or introduces a distinct form of reporting. They sought to determine whether AI-generated journalism adopts, modifies, or resists the dominant perspectives found in mainstream media. The study is based on a purpose-built corpus of AI-generated articles. The authors prompted ChatGPT-4 to write news reports about Roma migrants in the United Kingdom, using real headlines from two newspapers on the opposite ends of the British political spectrum: *The Guardian* (liberal) and the *Daily Mail* (conservative). The selected topics relate to the period surrounding the 2016 Brexit referendum, a context in which the Roma population was often framed through discriminatory or sensational narratives. The model was tasked to produce an article written "in the style of a journalist," without explicit instruction to imitate the ideological stance of either newspaper. The corpus analysis followed a critical discourse analysis by focusing on social actor representations (for example, whether Roma people are portrayed as active agents or passive victims), cause and responsibility assignment, and construction of sympathy or blame.

The results showed that ChatGPT-4 adopted a discourse style that differed from both newspapers, as its articles generally avoided sensationalism and moral judgment. The AI-generated articles displayed lexical moderation and tended to attribute responsibility in balanced terms, through background explanations that promoted empathy toward marginalized groups. In other words, ChatGPT adopted a cautious and formal register that minimized emotional polarization. However, this restraint does not mean ideological absence; rather, the model displayed an implicit liberal-humanitarian orientation, emphasizing fairness, social inclusion, and respect for diversity. When prompted with headlines that invited bias, such as those from the *Daily Mail*, the model generally refused to produce such content or reformulated it into a neutral account. This reveals a gatekeeping mechanism in the discourse that reflects the ethical guidelines built into model during training.

These results suggest that ChatGPT-4 filters journalistic content through value-based constraints designed to prevent harm. This filtering mechanism produces an appearance of neutrality that, in practice, corresponds to a specific moral stance, one that prioritizes tolerance and restraint over pluralism of voice. The resulting discourse may appear more balanced, but risks excluding legitimate expressions of emotion, conflict, and dissent that are part of authentic journalism. In practice, AI-generated

news discourse seems to produce a new, rule-governed style of journalism shaped by moderation protocols. As the model seeks to minimize offense and bias, it introduces another ideological layer: the institutional morality of the developers. AI journalism is therefore a controlled discourse form that is well-written, ethically filtered, and socially cautious, but limited in representing the diversity of human perspectives.

Szczepanik (2025) investigated whether AI-generated language reproduced the dominant ideology of *growthism* (i.e., the assumption that continuous economic expansion is inherently positive). To address this goal, the author compiled a purpose-built corpus of ChatGPT responses generated from prompts about sustainability, development, and ecological limits. To analyze the corpus, the study resorted to corpus linguistic methods used in corpus-assisted discourse studies (CADS), like keyword and concordance analysis, to identify lexical and semantic patterns that suggested underlying ideological orientations.

The analysis indicated that ChatGPT reproduced the discourse of *growthism* by associating economic growth with notions of progress, innovation, and prosperity and by treating environmental constraints as technical obstacles to be managed rather than as limits that question the growth paradigm. Even when addressing sustainability, the model tended to rephrase it within the logic of "green growth," suggesting that environmental goals can be achieved without altering economic priorities. The study finds little evidence of oppositional or post-growth perspectives; instead, ChatGPT seems to foster pro-growth ideology by presenting it as balanced and reasonable.

These findings provide evidence to suggest that ChatGPT mirrors the ideological formations embedded in its training data, which are dominated by pro-market and techno-optimistic perspectives. The LLM normalizes these ideologies by presenting them as common sense, thereby contributing to the success of this and other economic ideologies in public communication. The study demonstrates that using corpus-based methods to uncover AI discourse can shed light on how machine-generated texts reinforce existing dominant ideological positions.

Gallegos et al. (2024) conducted a literature survey of NLP studies that have evaluated discourse bias in AI-generated text using test suites, prompts, and benchmark datasets. The authors' goal was to measure how bias appears in AI-generated language in three ways: through embeddings,

revistadelta
41.4
2025

token probabilities, and generated text. The authors identified a pattern in the literature: The models repeat social stereotypes, display unequal treatment between social groups, and privilege Western cultural frames in their narratives. For example, gendered or racial assumptions surface in the adjectives or occupations attached to named entities. A system completes "The doctor said that…" with the pronoun "he," and "The nurse said that…" with "she." Likewise, it associates names like "John" or "Michael" with roles such as "engineer" or "CEO," and names like "Latisha" or "Jamal" with low-status occupations or criminal activities. Moreover, even when developers apply statistical "debiasing," the resulting texts continue to express preference through the choice of topic or the moral stance adopted in the generated story.

From this synthesis, the authors proposed several conclusions. First, bias detection must include the analysis of discourse rather than only embeddings or probabilities. It is in the generated language that the harm of bias becomes communicatively effective. Second, quantitative metrics should be complemented by linguistic inspection because many harmful patterns operate through lexical patterns rather than toxicity measures. Third, discourse analysis should be incorporated into model evaluation so that the social meaning of generated language can be assessed alongside performance metrics. In essence, the study underscores the notion that fairness in AI is a linguistic and communicative question by showing how LLMs produce discourse that carries social assumptions.

Buyl et al. (2024) assessed the assumption that AI-generated discourse is not neutral, but rather reflects the worldview of its creators and the environment in which it was trained. The authors argued that, as models are trained on human text and finetuned by human evaluators, their language inevitably reproduces political and moral patterns that exist in the societies where they were developed. The study created a corpus of LLM-generated texts by ChatGPT, Claude, Gemini, and others in both English and Chinese that consisted of short descriptive paragraphs about a set of politically charged figures and issues. These included international political leaders, economic topics, and moral concepts such as corruption or equality. The results revealed that English-language texts from Western-developed models tended to favor liberal democratic values and present political authority in terms of individual rights and transparency. In contrast, the same models, when prompted in Chinese, framed the same topics using collectivist or harmony-based language. The study also found that Chinese-developed

models differed in moral emphasis, displaying stronger acceptance of hierarchy and stability as positive values. For instance, models varied in whether they portrayed political leaders as "visionary" or "controversial" or whether they interpreted social order as "control" or "security."

Based on these findings, the authors concluded that the notion of ideological neutrality in AI-generated language is misleading. The discourse that models produce always reflects particular moral and political orientations, even when the prompt does not ask for opinion. In addition, ideology in AI discourse depends on both model origin and prompting language. Finally, the authors warned that, because LLMs are increasingly used to generate news summaries, classroom materials, and public information, their ideological patterns could shape public discourse.

Hirose and Uchida (2025) examined whether LLMs take consistent ideological positions and how flexible these positions are when confronted with opposing viewpoints by measuring the way models state opinions, accept or reject propositions, and modify their stance when faced with contradiction. The study created a bilingual corpus comprising 539 politically or morally debatable yes/no questions in both English and Japanese that ChatGPT-4o-mini and Gemini 1.5 Pro responded to. The questions covered topics such as immigration, gender equality, religion, environmental policy, and freedom of speech. The authors then performed a two-phase test: a baseline phase, where the models answered the questions without any additional context, and an opposition phase, where each model was told that the previous answer was wrong and was then asked to respond again. Three quantitative indicators were derived from this corpus. The first, bias, measured the proportion of questions for which the model consistently chose yes or no, reflecting an ideological leaning. The second, willingness, captured internal consistency across rounds of generation for the same prompt. The third, shift under opposition, evaluated how readily the model reversed its stance after being told that its previous answer was mistaken.

The results showed that ChatGPT tended to shift more readily when confronted with opposing views, often modifying its phrasing to accommodate the interlocutor. In contrast, Gemini was more resistant to change and maintained its earlier positions more frequently. The English responses also displayed stronger individualistic and rights-based reasoning, whereas the Japanese responses showed greater orientation toward harmony and consensus. These findings suggested three main conclusions. First,

revistadelta
41.4
2025

AI-generated language exhibits identifiable ideological patterns even when prompts are designed to be neutral. Second, flexibility in changing position shows that alignment mechanisms built into their training influence the discourse to favor agreement and politeness. Third, the presence of both stability and adaptability suggests that AI discourse operates according to a form of situational ideology: The model reproduces the stance most compatible with its communicative context rather than maintaining a fixed worldview. In summary, the study revealed that LLMs not only produce opinions, but also adapt those opinions in ways that resemble human conversational negotiation. The study considered this adaptive behavior as evidence that AI discourse functions as a reflection of both embedded ideological preferences and context-sensitive alignment toward the interlocutor's viewpoint.

Smith et al. (2025) addressed how LLMs reinforce standard language ideology through their generated texts, that is, discourse promoting a narrow linguistic norm and reproducing ideas of correctness and prestige that privilege standardized varieties such as Standard American English. By promoting this standard language ideology, AI-generated language may influence perceptions of what counts as "proper," "neutral," or "intelligent" speech. The authors relied on examples and findings from existing empirical research in which AI systems were prompted to produce text in different varieties of English. The authors identified a pattern of preference for formal registers, avoidance of regional or ethnic language features, and correction of perceived "errors" that are in fact dialectal forms. In other words, LLMs tend to treat standard varieties as the default, marking other varieties as deviant, ungrammatical, or inappropriate. When asked to produce text in a dialect or vernacular, models may simplify (or event caricature) it, using stereotypical vocabulary rather than authentic forms. Moreover, when the models attempt to imitate dialectal speech, they generally fall back on exaggerated or outdated features, turning representation into parody. This production in turn creates serious problems. First, speakers of minoritized varieties receive less accurate or respectful representation, as the model discourse implicitly labels their language as inferior. Second, when the model refuses to produce non-standard output in the name of avoiding offense, it effectively erases those varieties from the communicative space. The authors describe this as a form of "linguistic invisibility," where diversity is lost under the appearance of neutrality.

In light of these findings, the authors concluded that AI-generated discourse acts as an instrument of standard language ideology, not as a neutral medium of communication. Fairness in AI language use cannot be achieved through technical calibration alone, as it requires attention for understanding how discourse practices shape linguistic hierarchies. Importantly, linguistic diversity should be approached as a resource rather than as noise to be filtered out. Their argument situates AI discourse as a form of power relation: When a model presents one variety as normal and others as marginal, it reinforces inequality through everyday language generation. The authors call for future models to be trained and evaluated on corpora that reflect genuine variation rather than artificially uniform text.

Our sample investigation of AI discourse shares common findings with these studies. As in all of them, we treat bias and ideology as linguistic phenomena observable in output rather than as abstract properties of the model and consider lexical patterning as a reflection of cultural and ideological input that leaves traces in the text, which can be identified through linguistic analysis. At the same time, our study moves beyond a focus on propositional or evaluative discourse found in the four cited studies. Whereas Buyl et al. (2024), Szczepanik (2025), and Hirose and Uchida (2025) investigated political, economic and moral discourse, our work looks at artistic discourse. We see the creative domain as a valid site for ideological reproduction, demonstrating that bias can appear not only in argumentative or journalistic texts, but also in aesthetic genres. This approach expands the scope of AI discourse research from opinion expression to cultural expression.

Our study also detected an ethic of moderation and positivity in AI discourse, similar to Breazu and Katsos (2024). In our corpus, Gemini and ChatGPT produced an exaggerated discourse of self-empowerment, stripping rap of its social and political specificity in favor of moral uplift. In Breazu and Katsos's (2024) study, ChatGPT performed a comparable filtering function by rewriting sensational headlines into moralized neutrality. In both cases, AI discourse replaced confrontation with affirmation. However, whereas Breazu and Katsos (2024) interpreted this as cautious journalism, our results expose a reductionist aesthetic that turns collective struggle into an individual virtue narrative. In both cases, the findings reinforce the idea that LLM training protocols promote morally "safe" discourses in various domains (e.g., music lyrics and news reporting).

revistadelta
41.4
2025

Furthermore, as in Hirose and Uchida (2025), our analysis detected a form of contextual pliability. Their bilingual question corpus revealed that ChatGPT adjusted its answers to accommodate interlocutors, showing situational flexibility. In our case, the results demonstrated a related process in which AI adapts an existing genre (i.e., rap) by reinterpreting its conflictual discourse in a way that fits the moral and stylistic expectations encoded in its training and finetuning data. Both studies' results indicated that LLMs modulate their language toward normative, socially acceptable discourse, regardless of topic or register.

Finally, our study reflects the findings from Smith et al. (2025), who described the privileging of standard linguistic norms as a mechanism of complexity erasure. In our case, the findings showed the privileging of standard emotional and moral narratives, which are a form of linguistic and ideological conformity that leads to uniform and socially sanctioned expression.

In summary, our sample study confirmed patterns in AI discourse: (a) the expression of bias through discourse rather than form alone, (b) the moral filtering of content, and (c) the tendency of LLMs to reshape their language to match socially acceptable norms. We also argued for the utility of describing AI discourses in creative registers, where ideological simplification takes the form of aesthetic and moral normalization rather than explicit political bias.

7. Conclusion

We have argued that the distance between AI and human language use is best understood as a register problem that occurs because the probabilities that guide word-sequence generation in AI are not register-specific; consequently, the resulting texts combine probabilities from different registers, giving rise to distorted realizations of the intended registers. This finding highlights a symptom of what we termed a *register deficit*, or an unawareness of human register variation.

The analysis of EFL textbook texts suggests that artificial human language and LLM artificial language are markedly different. Human textbook writers deliberately design artificial texts, yet these texts maintain the communicative and pedagogical orientation expected in foreign language

education. Their simplification reflects the pedagogical constraints of EFL teaching. In contrast, LLM artificiality is not pedagogically motivated but mechanical; it lacks both situational grounding and the discursive conventions of language teaching. This should come as no surprise: LLMs lack register awareness in general, and educational registers are no exception.

The results from the lyric corpus analysis confirm that this deviation extends from grammar to discourse. When generating artistic texts, AI systems transform socially and culturally specific discourses into generalized moral narratives. What appears as creativity is instead a process of normalization, through which conflict and oppression are replaced by moral uplift and self-empowerment.

In sum, AI-generated texts are carriers of a confusing linguistic environment that distorts the diversity of human communication, replacing register variation with standardized language and social historic discourses with morally filtered expression. We have only begun to scratch the surface of the linguistic makeup of AI-generated texts. With its commitment to the identification, quantification, and interpretation of recurrent patterns of language use, corpus linguistics offers the means to describe and measure AI linguistic limitations.

Conflict of interests

I declare that there is no conflict of interest in this article.

Data available

The data used in this study are available in the article.

References

Ash, A. M. (2025). Three in 10 Teachers Use AI Weekly, Saving Six Weeks a Year. 2025/06/24. https://news.gallup.com/poll/691967/three-teachers-weekly-saving-six-weeks-year.aspx.

Baker, P., & McEnery, T. (2015). *Corpora and Discourse Studies: Integrating Discourse and Corpora*. Palgrave Macmillan.

Berber Sardinha, T. (2000a). Análise Multidimensional [Multi-Dimensional Analysis]. *DELTA*, *16*(1), 99-127.

Berber Sardinha, T. (2000b). Linguística de corpus: Histórico e problemática. *DELTA*, *16*(2), 323-367.

revistadelta
41.4
2025

Berber Sardinha, T. (2024). AI-generated vs human-authored texts: A Multidimensional comparison. *Applied Corpus Linguistics*, *4*(1), 100083. https://doi.org/10.1016/j.acorp.2023.100083

Berber Sardinha, T., & Fitzsimmons-Doolan, S. (2025). *Lexical Multidimensional Analysis: Identifying Discourses and Ideologies*. Cambridge University Press.

Biber, D. (1988). *Variation Across Speech and Writing*. Cambridge University Press.

Biber, D. (2012). Register as a predictor of linguistic variation. *Corpus Linguistics and Linguistic Theory*, *8*(1), 9-37.

Biber, D., & Conrad, S. (2009). *Register, Genre, and Style*. Cambridge University Press.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, *25*(3), 371-405.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics - Investigating Language Structure and Use*. Cambridge University Press.

Breazu, P., & Katsos, N. (2024). ChatGPT-4 as a journalist: Whose perspectives is it reproducing? *Discourse & Society*, *35*(6), 687–707. https://doi.org/10.1177/09579265241251479

Burr, V. (1995). *An Introduction to Social Constructionism*. Routledge.

Buyl, M., Rogiers, A., Noels, S., Dominguez-Catena, I., Heiter, E., Romero, R., Johary, I., Mara, A.-C., Lijffijt, J., & De Bie, T. (2024). Large language models reflect the ideology of their creators. Cornell University. https://doi.org/arXiv:2410.18417

Carter, R. (1998). Orders of reality: CANCODE, communication, and culture. *ELT Journal*, *52*(1), 43-56.

Collentine, J., & Berber Sardinha, T. (in press). Artificial Intelligence and Corpus Linguistics. In C. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics* (2nd ed.). Wiley.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, *23*(397-423).

Cortes, V. (2013). Lexical bundles and grammar. In C. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics* (pp. 3313–3319). Wiley.

Delfino, M. C. N. (2022). *More than words: Análise multimodal multidimensional da música popular em língua inglesa [More than words: Multimodal Multi-Dimensional Analysis of popular music in English]* [PhD Dissertation, PUCSP]. São Paulo.

Delfino, M. C. N., & Berber Sardinha, T. (2025). *Exploring AI creativity in songwriting: A lexical multi-dimensional analysis of human and machine-generated lyrics* [Paper]. Workshop "Corpus linguistics 2040: Which data, which methods, which models?", Leibniz-Institut für Deutsche Sprache (IDS), Mannheim, Germany.

Egbert, J., & Biber, D. (2019). Incorporating text dispersion into keyword analyses. *Corpora*, *14*(1), 77-104.

Egbert, J., & Staples, S. (2019). Doing Multi-Dimensional Analysis in SPSS, SAS, and R. In T. Berber Sardinha & M. Veirano Pinto (Eds.), *Multi-dimensional Analysis: Research Methods and Current Issues* (pp. 125-144). Bloomsbury.

Friginal, E., & Hardy, J. (2019). From factors to dimensions: Interpreting linguistic co-occurrence patterns. In T. Berber Sardinha & M. Veirano Pinto (Eds.), *Multi-Dimensional Analysis: Research Methods and Current Issues* (pp. 145-164). Bloomsbury Academic.

Gallegos, I. O., Rossi, R. A., Barrow, J., Tanjim, M. M., Kim, S., Dernoncourt, F., Yu, T., Zhang, R., & Ahmed, N. K. (2024). Bias and fairness in large language models: A survey. *Computational Linguistics*, *50*(3), 1097–1179.

Gilmore, A. (2004). A comparison of textbook and authentic interactions. *ELT Journal*, *58*(4), 363-374.

Goulart, L. (2024). *Variation in University student writing : A communicative text type approach*. John Benjamins.

Goulart, L., Matte, M. L., Mendoza, A., Alvarado, L., & Veloso, I. (2024). AI or student writing? Analyzing the situational and linguistic characteristics of undergraduate student writing and AI-generated assignments. *Journal of Second Language Writing*, *66*(101160), 1-19.

Hajer, M. (1993). Discourse coalitions and the institutionalization of practice. In *The argumentative turn in policy analysis and planning* (pp. 43-76). Duke University Press.

Hall, S. (1992). *Formations of Modernity*. Polity Press.

Halliday, M. A. K. (1991). Corpus studies and probabilistic grammar. In K. Aijmer & B. Altenberg (Eds.), *English Corpus Linguistics: Studies in Honour of Jan Svartvik* (pp. 30-43). Longman.

Hirose, M., & Uchida, M. (2025). Decoding the mind of large language models: A quantitative evaluation of ideology and biases. Cornell University. https://doi.org/arXiv:2505.12183

Jiang, F., & Hyland, K. (2025). Does ChatGPT argue like students? Bundles in argumentative essays. *Applied Linguistics*, *46*, 375-391.

Schmid, H. (1994). *Probabilistic Part-of-Speech Tagging Using Decision Trees* Proceedings of the International Conference on New Methods in Language Processing, Manchester, UK.

Scott, M. (1997). PC Analysis of key words - and key key words. *System*, *25*(2), 233-245.

Scott, M. (2000). Focusing on the text and its key words. In L. Burnard & T. McEnery (Eds.), *Rethinking Language Pedagogy from a Corpus Perspective - Papers from the Third International Conference on Teaching and Language Corpora* (pp. 103-122). Peter Lang.

revistadelta
41.4
2025

Shimazumi, M., & Berber Sardinha, T. (2023). *Variation in learner writing and a comparison with textbook models* [Paper]. 12th Corpus Linguistics Conference, Lancaster, UK.

Shimazumi, M., & Berber Sardinha, T. (2024). *Comparing EFL textbook and AI-generated writing: A multi-dimensional analysis* [Paper]. American Association for Corpus Linguistics (AACL 2024), Oregon University, Eugene, OR, USA.

Shimazumi, M., & Berber Sardinha, T. (2025). *Simulating human artificiality: How well does AI generate EFL textbook texts?* [Paper]. Corpus Linguistics 2025 International Conference (CL2025), Aston University, Birmingham, UK.

Sinclair, J. M. (1991). *Corpus, Concordance, Collocation*. Oxford University Press.

Smith, G., Fleisig, E., Bossi, M., Rustagi, I., & Yin, X. (2025). Standard language ideology in AI-generated language. Cornell University. https://doi.org/10.48550/arXiv.2406.08726

Szczepanik, R. J. (2025). The limits to growth(ism) in ChatGPT—corpus assisted discourse studies in AI-generated texts. *Discourse & Society*, *Online access*. https://doi.org/10.1177/09579265241308585

Recebido em: 15.07.2025
Aprovado em: 02.11.2025

EDITORA GERAL: Marcia Veirano Pinto