

INTRICACIA GRAMATICAL EM TEXTOS GERADOS PELO CHATGPT EM LÍNGUA ADICIONAL: UMA PERSPECTIVA SISTÊMICO-FUNCIONAL

COMPLEJIDAD GRAMATICAL EN TEXTOS GENERADOS POR CHATGPT EN LENGUAS ADICIONALES: UNA PERSPECTIVA SISTÊMICO-FUNCIONAL

GRAMMATICAL INTRICACY IN TEXTS GENERATED BY CHATGPT IN AN ADDITIONAL LANGUAGE: A SYSTEMIC-FUNCTIONAL PERSPECTIVE

COMPLEXITÉ GRAMMATICALE DANS LES TEXTES GÉNÉRÉS PAR CHATGPT DANS UNE LANGUE SUPPLÉMENTAIRE : UNE PERSPECTIVE SYSTÉMIQUE-FONCTIONNELLE

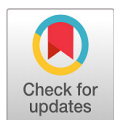
Antonio Marcio da Silva

Professor associado, University of Essex, Colchester, Essex, Reino Unido.
antonio.dasilva@essex.ac.uk
<https://orcid.org/0000-0002-4628-4091>

Lucia Rottava

Professora associada, Universidade Federal do Rio Grande do Sul (UFRGS), Brasil.
lucia.rottava@ufrgs.br
<https://orcid.org/0000-0003-3094-6270>

A coleta de dados para este estudo foi financiada pelo Executive Dean Fund da University of Essex, Reino Unido, e é parte do projeto "The Effect of ChatGPT on Student Writing in Multiple Languages: A Systemic Functional Linguistics Analysis". O projeto foi aprovado pelo Comitê de Ética da Universidade, sob o número ETH2223-0987.



RESUMO

A intricacia gramatical refere-se à complexidade da organização sintática dos complexos oracionais em um texto. Este estudo analisa essa intricacia em um corpus de seis tarefas de português como língua adicional produzidas pelo ChatGPT. Foram examinados os complexos oracionais, considerando-se a interdependência entre as orações – parataxe e hipotaxe – e o sistema lógico-semântico de expansão e projeção. A análise dos dados incluiu tratamento estatístico, utilizando Python e SPSS, além de uma avaliação qualitativa de uma amostra de textos. Os resultados indicaram que a correlação entre intricacia gramatical e densidade lexical geralmente segue o que Halliday (1985) propôs em termos de proporcionalidade. No entanto, a correlação entre intricacia gramatical e níveis de proficiência não se confirmou nos textos gerados pela inteligência artificial, indicando que o ChatGPT não segue de forma consistente as instruções referentes ao nível de proficiência. Quanto aos complexos oracionais, prevaleceram relações hipotáticas, e o desenvolvimento textual ocorreu principalmente por meio de expansões de extensão (+) e de intensificação (x). As contribuições deste estudo destacam a importância de que os professores que frequentemente utilizam inteligência artificial para criar materiais didáticos prestem atenção à complexidade e à organização lógico-semântica dos textos, o que também auxilia na identificação de produções que não seguem os princípios de integridade acadêmica.

Palavras-chave: ChatGPT, intricacia gramatical, sistema lógico-semântico, complexos oracionais, LSF, português língua adicional (PLA)

RESUMEN

La complejidad gramatical se refiere al grado de dificultad en la organización sintáctica de las oraciones complejas en un texto. Este estudio analiza dicha complejidad en un corpus de seis tareas de portugués como lengua adicional producidas

Recebido: 2024-08-21 / Aceito: 2025-06-12 / Publicado: 2026-02-20

<https://doi.org/10.17533/udea.ikala.358152>

Editora: Luanda Sito, Universidad de Antioquia, Medellín, Colômbia.

Direitos patrimoniais, Universidad de Antioquia, 2026. Este é um artigo em acesso aberto, distribuído nos termos da licença BY-NC-SA 4.0 Internacional da Creative Commons.



Íkala, Revista de Lenguaje y Cultura

MEDELLÍN, COLOMBIA, VOL. 31 NÚM. 1 (JANEIRO-ABRIL, 2026), PP. 1-21, ISSN 0123-3432
www.udea.edu.co/ikala

por ChatGPT. Se examinaron las oraciones complejas, teniendo en cuenta la interdependencia entre las oraciones (parataxis e hipotaxis) y el sistema lógico-semántico de expansión y proyección. El análisis de los datos incluyó un tratamiento estadístico, utilizando Python y SPSS, además de una evaluación cualitativa de una muestra de textos. Los resultados indicaron que la correlación entre la complejidad gramatical y la densidad léxica suele seguir lo que Halliday (1985) propuso en términos de proporcionalidad. Sin embargo, la correlación entre la complejidad gramatical y los niveles de competencia no se confirmó en los textos generados por la inteligencia artificial, lo que indica que ChatGPT no sigue de manera consistente las instrucciones relativas al nivel de competencia. En cuanto a las estructuras oracionales complejas, prevalecieron las relaciones hipotáticas, y el desarrollo textual se produjo principalmente mediante expansiones de extensión (+) e intensificación (x). Las contribuciones de este estudio destacan la importancia de que los profesores que utilizan con frecuencia la inteligencia artificial para crear materiales didácticos presten atención a la complejidad y la organización lógico-semántica de los textos, lo que también ayuda a identificar producciones que no siguen los principios de integridad académica.

Palabras clave: ChatGPT, complejidad gramatical, sistema lógico-semántico, oraciones complejas, LSF, portugués como lengua adicional (PLA)

ABSTRACT

Grammatical intricacy refers to the complexity of the syntactic organization of complex sentences in a text. This study analyzes this intricacy in a corpus of six Portuguese as an additional language tasks produced by ChatGPT. Sentence complexes were examined, considering the interdependence between sentences—parataxis and hypotaxis—and the logical-semantic system of expansion and projection. Data analysis included statistical treatment using Python and SPSS, as well as a qualitative evaluation of a sample of texts. The results indicated that the correlation between grammatical intricacy and lexical density generally follows what Halliday (1985) proposed in terms of proportionality. However, the correlation between grammatical intricacy and proficiency levels was not confirmed in the texts generated by artificial intelligence, indicating that ChatGPT does not consistently follow instructions regarding proficiency level. As for sentence complexity, hypotactic relations prevailed, and textual development occurred mainly through extension (+) and intensification (x) expansions. The contributions of this study highlight the importance of teachers who frequently use artificial intelligence to create teaching materials paying attention to the complexity and logical-semantic organization of texts, which also helps in identifying productions that do not follow the principles of academic integrity.

Keywords: ChatGPT, grammatical intricacy, logical-semantic system, complex sentences, SFL, Portuguese as an additional language (PAL)

RÉSUMÉ

La complexité grammaticale fait référence à la complexité de l'organisation syntaxique des phrases complexes dans un texte. Cette étude analyse cette complexité dans un corpus de six tâches en portugais comme langue supplémentaire produites par ChatGPT. Les phrases complexes ont été examinées en tenant compte de l'interdépendance entre les phrases – parataxe et hypotaxe – et du système logique-sémantique d'expansion et de projection. L'analyse des données a inclus un traitement statistique, à l'aide de Python et SPSS, ainsi qu'une évaluation qualitative d'un échantillon de textes. Les résultats ont indiqué que la corrélation entre la complexité grammaticale et la densité lexicale suit généralement ce que Halliday (1985) a proposé en termes de proportionnalité. Cependant, la corrélation entre la complexité grammaticale et les niveaux de compétence n'a pas été confirmée dans les textes générés par l'intelligence artificielle, ce qui indique que ChatGPT ne suit pas de manière cohérente les instructions relatives

au niveau de compétence. En ce qui concerne les phrases complexes, les relations hypotactiques ont prévalu, et le développement textuel s'est principalement fait par des expansions d'extension (+) et d'intensification (x). Les contributions de cette étude soulignent l'importance pour les enseignants qui utilisent fréquemment l'intelligence artificielle pour créer du matériel didactique de prêter attention à la complexité et à l'organisation logique et sémantique des textes, ce qui aide également à identifier les productions qui ne respectent pas les principes d'intégrité académique.

Palavras-chave: ChatGPT, complexité grammaticale, système logico-sémantique, phrases complexes, linguistique systémique fonctionnelle (LSF), portugais langue additionnelle (PLA)

Introdução

A intricacia gramatical é um conceito recorrente na linguística sistêmico-funcional (LSF), utilizado para explicitar a complexidade oracional na língua, definida pela razão entre o número total de orações e o número de períodos presentes em um texto. Trata-se de um conceito introduzido por Halliday (1985, 1987) para evidenciar que a língua é um sistema aberto e dinâmico, que representa e constrói os parâmetros semióticos. Por tanto, a complexidade é uma característica tanto da língua falada quanto da escrita, embora de natureza diversa. Para o autor, o uso da língua configura-se em um *continuum* que envolve baixa ou alta metáfora gramatical, baixa ou alta densidade lexical, baixa ou alta intricacia gramatical, ou ainda dinamicidade multimodal (Halliday, 1993), dependendo da modalidade da língua: falada ou escrita. Halliday (1985) já salientava que há organização e complexidade tanto na escrita quanto na fala, e assinalava que é preciso considerar três aspectos indissociáveis: “a natureza do meio, as funções desempenhadas e as proporções formais exibidas, ou seja, meio, função e forma” (p. 78)¹.

Com a expansão de ferramentas que auxiliam na produção textual por intermédio da inteligência artificial (IA) no século XXI, esses três aspectos continuam indissociáveis, especialmente a natureza do meio em que o texto se realiza (Halliday, 1985). Entretanto, a escrita produzida com o auxílio de ferramentas de IA, como os denominados *chatbots*, tem sido alvo de muita especulação quanto à qualidade dos textos gerados tecnologicamente, incluindo a questão de se essa produção apresenta maior ou menor complexidade sintática e/ou semântica em comparação à escrita convencional (produzida por humanos). A complexidade sintática diz respeito ao sistema de interdependência lógica (parataxe e hipotaxe), enquanto a complexidade semântica se refere às relações lógico-semânticas (Halliday & Matthiessen, 2014).

Pesquisas recentes (Lancaster, 2023; Ramos, 2023; Rottava & Da Silva, 2023) sugerem que textos gerados por *chatbots*, especialmente o denominado ChatGPT (transformador gerador pré-treinado), desenvolvido pela OpenAI, tendem a ser mais abstratos e aparentam ser mais complexos. Porém, não há estudos que demonstrem o quão complexos esses textos são em termos linguísticos, nem se possuem características mais alinhadas à modalidade falada ou escrita. Assim, este artigo tem como objetivo compreender a complexidade dos textos em português como língua adicional (PLA) gerados pelo ChatGPT, particularmente no que diz respeito à complexidade lógico-semântica. Para investigar as características desses textos, este artigo recorre ao conceito de intricacia gramatical (Halliday, 1985).

Além disso, é importante destacar que a complexidade da língua, conforme Halliday (2004), não advém apenas da intricacia gramatical, mas também da densidade lexical, uma vez que esta se concentra no aspecto semântico, enquanto a intricacia gramatical se refere aos elementos significativos do ponto de vista sintático.

Em um estudo recente, Da Silva e Rottava (2024) investigaram quantitativamente um extenso corpus de 2.991 textos escritos (706.401 palavras), produzidos pelo ChatGPT em diferentes línguas adicionais (alemão, espanhol, francês, italiano e português), com o objetivo de entender a complexidade dos textos, analisando-os sob o viés da densidade lexical. Os resultados indicaram que a média dessa densidade é de aproximadamente 46,72%, sugerindo uma correlação estatisticamente significativa entre a extensão dos textos e as escolhas lexicais com carga semântica. No entanto, essa complexidade seria mais bem compreendida se analisada em relação à configuração dos complexos oracionais recorrentes.

Assim, este artigo busca preencher essa lacuna, levando em consideração os resultados referentes à densidade lexical observados nos textos que compõem o corpus, a fim de compreender se há relação com a intricacia gramatical na amostra de textos analisados, desta vez, produzidos em PLA.

1 No original: “(...) the nature of the medium, the functions served, and the formal properties displayed – let us say function, medium, and form”.

Para tanto, o tratamento dispensado aos dados é quanti-qualitativo. Esse recorte permite uma análise quantitativa das relações de interdependência (parataxe e hipotaxe) em cada uma das orações presentes nos textos – realizadas gramaticalmente – e uma análise qualitativa das relações lógico-semânticas. Para orientar a análise, foram elencadas as seguintes perguntas de pesquisa:

- Qual é a correlação entre a intrincacia gramatical e os níveis de proficiência indicados pelo Quadro Comum Europeu de Referência para Línguas (Council of Europe, 2001) nos textos que compõem o corpus produzido pelo ChatGPT?
- Como a organização das orações e dos complexos oracionais nos textos revela a intrincacia gramatical?
- Quais relações lógico-semânticas recorrentes ilustram a intrincacia gramatical em uma amostra de textos deste mesmo corpus?

Espera-se que este estudo contribua para a compreensão de algumas implicações da complexidade textual resultante da IA na interpretação de textos produzidos pelo ChatGPT. Além disso, acredita-se que as respostas às perguntas deste estudo revelam características linguísticas que podem orientar encaminhamentos teórico-práticos no ensino da escrita em língua adicional na era digital.

Este artigo está organizado em cinco seções, além desta introdução. A primeira aborda o conceito de intrincacia gramatical na teoria da LSF; a segunda resume as categorizações relativas às relações de interdependência e às relações lógico-semânticas; a terceira detalha os procedimentos metodológicos utilizados para a constituição do corpus de pesquisa; a quarta apresenta os dados e discute os resultados; e, finalmente, a quinta apresenta as considerações finais.

Intrincacia gramatical como recurso de organização oracional

A preponderância de complexos oracionais em uma língua é uma indicação da complexidade linguística. A frequência com que uma oração se

Figura 1 Fórmula para calcular a intrincacia gramatical

$$\text{Intrincacia gramatical} = \frac{\text{Número total de orações no texto}}{\text{Número total de períodos no texto}}$$

Fonte: Eggins (2004, p. 97).

combina para formar um complexo oracional em um texto refere-se à intrincacia gramatical e está relacionada à organização das orações presentes nos textos. Para calcular a intrincacia gramatical, é necessário dividir o número total de orações no texto pelo número total de períodos (Figura 1).

No entanto, outros traços linguísticos podem indicar a complexidade da escrita, como a densidade lexical e a metáfora gramatical. As metáforas gramaticais são definidas por Halliday (2004) como construções que ocorrem no estrato léxico-gramatical da língua e permitem a variação na forma de expressão de um mesmo significado. Especificamente em relação à densidade lexical, Halliday (1985) e Eggins (2004) mostram que uma alta densidade lexical em um texto está associada a uma baixa intrincacia, enquanto uma baixa densidade lexical indica alta intrincacia gramatical. Nessa perspectiva, Eggins (2004) destaca que um texto altamente nominalizado, característico da escrita, contém um número significativamente maior de “palavras de conteúdo” (substantivos, verbos, adjetivos e alguns advérbios de modo), resultando em alta densidade lexical e baixa intrincacia gramatical.

Portanto, a densidade e a intrincacia estão relacionadas ao fato de um texto ser falado ou escrito: a densidade alta teria valores de entre 3 e 6 ou, em termos percentuais, acima de 40%, enquanto a densidade baixa de entre 1,5 e 2 ou abaixo de 40% (Halliday, 1985; Johansson, 2008). Da mesma forma, a alta intrincacia gramatical corresponde a valores acima de 2, enquanto a baixa intrincacia corresponde a valores abaixo de 2 (Eggins, 2004; Gulton & Pintubatu, 2022). A variação desses percentuais pode refletir um *continuum* entre fala e escrita, dependendo do texto e de seu propósito.

Estudar o complexo oracional sob a perspectiva da LSF significa situá-lo na metafunção ideacional e no estrato léxico-gramatical, tendo como premissa fundamental a inter-relação entre o linguístico e o social (Halliday, 1978, 1989, 2004). Por isso, os conceitos de contexto de cultura e contexto de situação (Malinowski, 1923, 1935; Firth, 1957) são essenciais para entender que os elementos linguísticos só adquirem sentido quando interpretados no(s) contexto(s) em que são usados (Halliday, 1978; Halliday & Hasan, 1989; Eggins, 2004). O contexto de cultura refere-se ao uso da língua em relação às convenções sociais, ideologias e instituições, enquanto o contexto de situação diz respeito à representação abstrata da situação e orienta as escolhas linguísticas do usuário ao produzir um texto (Ghio & Fernández, 2008).

Essas escolhas linguísticas variam quanto ao *campo* (o que está acontecendo), às *relações* (quem interage) e ao *modo* (como o texto está organizado). Essas variáveis situacionais estão interconectadas às metafunções, pois constroem realidades que representam a experiência (função ideacional), estabelecem interação entre as pessoas (função interpessoal) e criam um discurso coerente e articulado (função textual), manifestando-se no texto (Christie & Unsworth, 2000). Portanto, o contexto é um componente essencial na LSF e representa um sistema de opções estratificado em níveis funcionais, que mantém uma relação hierárquica entre si: o estrato contextual inclui categorias da situação social; o estrato semântico envolve os sistemas de significados; o estrato léxico-gramatical abrange os sistemas de expressão; e, o estrato fonológico compreende os sistemas de sons. Dependendo do propósito, a língua pode ser analisada a partir de diferentes perspectivas.

No estrato gramatical, a análise trinocular é realizada da seguinte forma: (I) “de baixo” (da manifestação para a organização gramatical), observando como os elementos constituintes da estrutura são realizados; (II) “ao redor” (nas relações funcionais), descrevendo as opções sistêmicas em termos de oposição, delicadeza e *valeur*; e (III) “de cima” (do significado para a organização

gramatical), partindo do contexto de situação, passando pela semântica e examinando os significados produzidos pelas funções gramaticais no desenvolvimento do texto (Figueredo, 2015).

As orações e/ou complexos oracionais são unidades que combinam três tipos de significados, cada uma delas expressando um tipo de organização semântica, de acordo com as metafunções (Halliday & Matthiessen, 2014). Essa organização semântica constrói orações com três dimensões: representação (a oração revela um processo da experiência humana no mundo), interação (a oração envolve alguma forma de relação entre falante e ouvinte ou interlocutor) e mensagem (comunica uma determinada quantidade de informação). Cada uma dessas dimensões expressa um tipo de organização semântica. Entre esses três tipos, o foco deste artigo é a metafunção ideacional, com ênfase no subcomponente lógico.

O subcomponente lógico e o subcomponente experiencial constituem a metafunção ideacional. O subcomponente lógico fornece recursos para expressar diversas complexidades (orações, grupos, palavras, etc.) e para formar complexos gramaticais, constituídos por meio dos sistemas de interdependência (parataxe e hipotaxe) e do sistema lógico-semântico. Por outro lado, o subcomponente experiencial permite que os seres humanos construam uma imagem mental da realidade que os rodeia – e de sua própria realidade interior – em termos de processos, participantes e circunstâncias.

Essa interação entre os subcomponentes lógico e experiencial é crucial na formação de complexos oracionais. Um complexo oracional sempre inclui uma proposição; e noções relacionadas a eventos, atributos, coisas, conceitos, participantes, processos ou circunstâncias estão presentes em uma oração e contribuem para criar um significado específico. Quanto mais proposições e elementos estiverem presentes em uma oração, maior será a complexidade sintática, o que impacta diretamente no processamento da informação. Essa complexidade é evidenciada no texto escrito pelas escolhas feitas

no sistema léxico-gramatical. Na perspectiva da LSF, o termo “sistema” refere-se ao ordenamento paradigmático da linguagem que engloba padrões ou regularidades. A linguagem é concebida como um recurso, ou uma rede de sistemas de opções, para a construção e interpretação do significado. Esse sistema é complexo e apresenta a propriedade formal de estar estratificado em níveis ou (sub)sistemas funcionais: contexto, semântica, léxico-gramática e fonologia/grafologia (Ghio & Fernández, 2008).

Intricacia gramatical e o subcomponente lógico

O subcomponente lógico da metafunção ideacional oferece recursos para formar diversos tipos de orações e complexos oracionais. As relações lógicas na gramática são expressas por meio de relações de interdependência – incluindo parataxe, hipotaxe e encaixamento –, além das relações lógico-semânticas.

O encaixamento é um mecanismo pelo qual uma oração funciona como constituinte dentro da estrutura de um grupo (nominal ou verbal), que, por sua vez, integra a estrutura de outra oração. Nesse caso, não há uma relação direta entre a oração encaixada e a oração principal. De acordo com Ghio e Fernández (2008), é fundamental distinguir as relações paratáticas e hipotáticas das orações encaixadas. Enquanto a parataxe e a hipotaxe estabelecem relações entre orações (ou entre elementos do mesmo nível estrutural), o encaixamento difere

por envolver uma mudança de nível: a oração deixa de funcionar como unidade independente e passa a atuar como um constituinte interno de um grupo que compõe outra oração. Portanto, conforme as autoras, o encaixamento não configura uma relação interoracional, mas sim um fenômeno estrutural interno à oração.

Para descrever o complexo oracional, dois sistemas funcionais são utilizados. O primeiro é o sistema tático ou de interdependência, que inclui: parataxe, representada pelo uso de números arábicos (1, 2), e hipotaxe, representada pelo uso de letras gregas (α , β), aplicáveis a todos os tipos de complexos. O segundo é o sistema lógico-semântico, que abrange expansão e projeção, estabelecendo relações entre as orações. A seguir, apresenta-se a Tabela 1, que ilustra o funcionamento desses sistemas no nível oracional.

Dentre os sistemas lógicos, este estudo enfatiza as relações observáveis nos complexos oracionais de hipotaxe, que envolvem orações que designam expansão e/ou projeção, bem como nos complexos oracionais formados por parataxe, também associados à expansão e/ou projeção (Eggins, 2004). Além dessa ênfase, também são analisadas as relações lógico-semânticas que constituem o texto.

Desenho do estudo

O corpus deste estudo foi gerado pelo *chatbot* ChatGPT e o tratamento analítico dos dados segue

Tabela 1 O complexo oracional

Sistemas lógicos	Subsistemas	Ocorrências	Definição
(i) Sistema de interdependência	Hipotaxe		Relação entre um elemento dependente (β) e um dominante (α). Os elementos que se relacionam não têm o mesmo status. Ex: <i>Ele disse que iríamos à praia amanhã.</i>
	Parataxe		Relação entre um iniciador (1) e um continuativo (2). Os elementos que se relacionam têm o mesmo status. Ex: <i>A chuva caiu e o frio logo começou.</i>

Tabela 1 O complexo oracional (continuação)

Sistemas lógicos	Subsistemas	Ocorrências	Definição
(ii) Sistema lógico-semântico	Expansão	Elaboração	(=) Uma oração reformula, exemplifica ou especifica a outra. <i>Ex: Comecei a ler o livro, o mesmo que ganhei de presente no ano passado.</i>
		Extensão	(+) Uma oração agrega, contrapõe ou oferece alternativa à outra. <i>Ex: Gosto de poesia, mas prefiro romances.</i>
	Projeção	Intensificação	(x) Uma oração qualifica a outra com circunstância de modo, tempo, lugar, causa, etc. <i>Ex: Ela saiu cedo porque precisava chegar ao trabalho antes das 8h.</i>
		Citação	(") Uma oração projetada por outra como algo dito literalmente. <i>Ex: José disse: "Chegarei atrasado na aula hoje."</i>
		Ideia	(') Uma oração apresenta outra como um pensamento ou crença. <i>Ex: João pensou que José chegaria atrasado.</i>

8

uma metodologia quanti-qualitativa (Dörnyei, 2007), priorizando uma amostra de textos de PLA. Essa amostra integra um conjunto mais amplo de dados em cinco línguas adicionais (alemão, espanhol, francês, italiano e português), organizados conforme os seis níveis de proficiência estabelecidos pelo Quadro Comum Europeu de Referência para Línguas (Council of Europe, 2001). Esses dados estão vinculados a um projeto de pesquisa desenvolvido em uma universidade britânica.

O corpus foi gerado seguindo uma metodologia estruturada composta por várias etapas. Após a aprovação do projeto de pesquisa pelo conselho de ética da universidade, a primeira etapa consistiu na divulgação do projeto por meio de páginas web e e-mails direcionados a todos os alunos regularmente matriculados na instituição, independentemente do curso ou nível de ensino. Em seguida, os interessados em participar do estudo preencheram um formulário eletrônico criado pelos pesquisadores utilizando a ferramenta *Microsoft Forms*, no qual puderam escolher

entre duas datas disponíveis para a realização das atividades em um laboratório de informática.

Após essas etapas iniciais, os participantes criaram uma conta pessoal no ChatGPT antes de comparecerem ao laboratório, na data escolhida pelos pesquisadores, para a execução das atividades propostas. As tarefas foram realizadas a partir de 100 contas distintas no ChatGPT, e cada sessão foi acompanhada por pelo menos um dos pesquisadores, que estiveram disponíveis para esclarecer dúvidas ou dificuldades relacionadas ao uso da ferramenta de IA. O tempo máximo estipulado para a conclusão das atividades foi de duas horas.

No dia agendado, no laboratório de informática da universidade, os participantes receberam seis tarefas elaboradas pelos pesquisadores, planejadas nas cinco línguas adicionais mencionadas anteriormente. As tarefas em PLA estão incluídas no Anexo como referência. O papel dos participantes consistiu em copiar as instruções das tarefas apresentadas em um documento de Word e inseri-las

Tabela 2 Número de textos, distribuídos de acordo com os níveis do CEFR

	Nível A1	Nível A2	Nível B1	Nível B2	Nível C1	Nível C2	Total
Português	100	100	100	98	100	100	598

Tabela 3 Número de palavras por nível

Nível A1	Nível A2	Nível B1	Nível B2	Nível C1	Nível C2	Total
13,314	16,806	23,409	27,599	27,795	25,228	134,151

em seus perfis pessoais na plataforma de IA, sem nenhuma intervenção ou *input* individual.

Importa destacar que, para cada tarefa, o ChatGPT foi instruído explicitamente em inglês a produzir um texto em português correspondente a um determinado nível de proficiência do CEFR. Por exemplo, o *prompt* incluía a instrução: “*write a text in Portuguese for the beginner level A1, following the instructions given below...*”, seguida da tarefa em português. Essa estrutura foi padronizada para todos os níveis e línguas utilizados no corpus. Os enunciados completos, conforme submetidos ao ChatGPT, estão disponíveis no Anexo, o que assegura a transparência dos procedimentos adotados e a comparabilidade entre os textos gerados. Essa padronização metodológica garante que o modelo foi orientado de forma direta quanto ao nível esperado, não deixando esse aspecto à interpretação da IA.

Posteriormente, cada um dos 30 textos gerados pelo ChatGPT (seis em cada um dos cinco idiomas) foi copiado e inserido pelos participantes em um formulário específico no formato *Microsoft Forms*, criado pelos pesquisadores para cada língua. De maneira geral, a plataforma operou de forma eficiente durante as atividades.

A partir do corpus resultante do procedimento descrito, este artigo foca na análise dos textos gerados em PLA, conforme detalhado na Tabela 2.

O corpus foi organizado conforme a Tabela 2, que identifica a quantidade de textos produzidos em cada nível de proficiência estabelecido pelo CEFR.

O corpus em PLA totaliza 134.151 palavras, conforme apresentado na Tabela 3.

Para a análise de dados, o corpus foi tratado seguindo os critérios e passos descritos a seguir: (1) Limpeza dos textos em planilhas do Excel: Preparação inicial dos textos para garantir a consistência e precisão dos dados; (2) Organização dos textos de acordo com os níveis do CEFR e cálculo da quantidade de palavras em cada nível; (3) Inserção dos textos em planilhas no formato CSV no ambiente *Google Colaboratory*: Importação dos textos para análise automatizada, na qual foi utilizado *Python* para identificar as “palavras de conteúdo” (substantivos, adjetivos, verbos e advérbios); (4) Verificação da densidade lexical: Uso de planilhas do Excel e do programa de análise estatística SPSS para calcular a densidade lexical, que serviu como base para a análise da intricacia gramatical, conforme o modelo de Halliday (2004); (5) Seleção de textos para análise: Escolha de dois textos de cada nível de proficiência, sendo um com menor densidade lexical (indicado pelo sinal –) e outro com maior densidade lexical (indicado pelo sinal +), totalizando 12 textos; (6) Análise manual das relações de interdependência das orações: Realização de uma análise quantitativa manual para mapear as relações de interdependência entre as orações nos 12 textos selecionados; (7) Verificação do percentual de intricacia gramatical: Cálculo da intricacia gramatical como a razão entre o número total de orações e o número de períodos em cada texto, seguindo a abordagem de Eggins (2004); (8) Análise e teste das variáveis: Aplicação do programa SPSS para análise estatística e teste das variáveis identificadas;

(9) Análise qualitativa das relações lógico-semânticas: Realização de uma análise detalhada em dois textos selecionados para examinar o funcionamento semântico das orações.

Os passos foram seguidos rigorosamente para garantir o adequado tratamento dos dados e permitir a análise quanti-qualitativa das relações gramaticais e semânticas no corpus gerado por IA, objeto de estudo nesta pesquisa.

Apresentação e análise dos dados

A análise dos dados segue as perguntas de pesquisa, utilizando critérios específicos para a organização, seleção e processamento do corpus. Os resultados são apresentados em três subseções, correspondentes às perguntas de pesquisa: as duas primeiras adotam uma abordagem quantitativa, enquanto a terceira se baseia em uma análise qualitativa.

Intricacia gramatical e níveis de proficiência

Conforme descrito na seção de metodologia, os dados foram gerados a partir de tarefas elaboradas segundo as diretrizes do CEFR. Para investigar se o ChatGPT produziu textos considerando essa variável, a análise dos dados foi orientada pela primeira pergunta de pesquisa: *Qual é a correlação entre a intricacia gramatical e os níveis de proficiência indicados pelo Quadro Comum Europeu de Referência para Línguas (CEFR) nos textos que compõem o corpus produzido pelo ChatGPT?*

Para responder a essa questão, os dados foram organizados com base no argumento de Halliday (1985, 2004), segundo o qual a intricacia gramatical é inversamente proporcional à densidade lexical, ou seja, à medida que uma aumenta, a outra tende a diminuir. Com base nesse parâmetro, e considerando as evidências empíricas de estudos focados na densidade lexical (Da Silva & Rottava, 2024), os resultados obtidos no cálculo da intricacia gramatical por texto e nível de proficiência estão apresentados na Tabela 4.

Na Tabela 4, os resultados foram agrupados de acordo com o nível de densidade lexical, sendo identificados com o sinal mais (+) os textos que apresentaram maior densidade lexical e com o sinal menos (–) aqueles com menor densidade. Os resultados, em geral, corroboram a hipótese proposta por Halliday em termos de proporcionalidade, uma vez que 91,7% dos casos confirmam essa relação inversa entre intricacia gramatical e densidade lexical. A única exceção foi observada no nível B2(–), no qual se verificou uma intricacia gramatical menor em comparação ao texto B2(+), além de uma diminuição simultânea tanto na densidade lexical quanto na intricacia gramatical. Esse comportamento discrepante, no entanto, não pôde ser amplamente explicado com base nos dados disponíveis. Uma hipótese possível é que as especificidades da tarefa atribuída ao nível B2 tenham influenciado a geração textual por parte da IA. Considera-se que uma análise comparativa mais detalhada entre os níveis de proficiência,

Tabela 4 Intricacia gramatical por nível

	Densidade (+)	Intricacia	Densidade (–)	Intricacia
Nível A1	5.4	2.9	3.9	3.3
Nível A2	5.7	2.1	4.8	3.1
Nível B1	5.1	2.1	4.6	2.4
Nível B2	5.2	2.4	4.7	1.7
Nível C1	4.8	2.0	4.3	2.6
Nível C2	5.1	2.0	4.1	2.1
Média	5.2	2.2	4.4	2.5

especialmente em relação às instruções fornecidas e ao tipo de tarefa, poderá elucidar essas diferenças em investigações futuras.

Ainda que não haja variações significativas entre os demais textos, os dados parecem confirmar a hipótese de que a intrincacia gramatical tende a ser maior nos níveis iniciais de proficiência (A1 e A2), uma vez que esses textos apresentam menor densidade lexical. Nos níveis mais avançados, observa-se uma densidade maior e uma intrincacia menor, o que reflete um maior domínio semântico nas escolhas linguísticas em uma língua adicional. No entanto, é importante destacar que, independentemente da densidade dos textos, os índices de intrincacia gramatical não indicam um domínio expressivo da escrita, mesmo quando há redução na densidade lexical. Em 83% dos textos, a intrincacia apresenta índice inferior a 3. Esse dado pode estar relacionado ao fato de que a densidade lexical

foi superior a 4 em 91,7% dos textos, inclusive naqueles classificados como de menor densidade. Além disso, esse resultado encontra respaldo no argumento de Halliday (1985, 2004), segundo o qual, em textos escritos, a densidade lexical geralmente varia entre 3 e 6, a depender do grau de formalidade, o que resulta, consequentemente, em menor intrincacia gramatical.

Relações de interdependência

Nesta subseção, a compreensão da intrincacia gramatical é analisada quantitativamente com o objetivo de verificar o número de orações que formam os complexos oracionais, considerando as relações de interdependência, e de visualizar como o texto está organizado estruturalmente. Para isso, a segunda pergunta de pesquisa orienta esta análise: *Como a organização das orações e dos complexos oracionais nos textos revela a intrincacia gramatical?*

Tabela 5 Complexos oracionais nos textos de densidade alta (D+)

	Oração simples (+)	N complexos oracionais (+)	Hipotaxe	Parataxe
Nível A1	1	6	19	0
Nível A2	4	6	16	2
Nível B1	4	7	12	8
Nível B2	3	10	18	10
Nível C1	6	11	25	8
Nível C2	8	7	21	2
Total	26	47	111	30

Tabela 6 Complexos oracionais nos textos de densidade baixa (D-)

	Oração simples (-)	N complexos oracionais (-)	Hipotaxe	Parataxe
Nível A1	1	11	32	12
Nível A2	3	7	28	4
Nível B1	2	7	18	2
Nível B2	6	5	12	2
Nível C1	1	15	32	12
Nível C2	6	7	17	8
Total	19	52	139	40

Tabela 7 Correlação entre complexo oracional e intrincacia gramatical

Correlações			
		Complexo oracional	Intrincacia gramatical
Complexo Oracional	Correlação de Pearson	1	.626*
	Sig. (2 extremidades)		.030
	N	12	12
Intrincacia Gramatical	Correlação de Pearson	.626*	1
	Sig. (2 extremidades)	.030	
	N	12	12

*. A correlacao é significativa no nível 0,05 (2 extremidades).

Tabela 8 Correlação entre complexo oracional e ocorrência de hipotaxe e parataxe

Correlações				
		Complexo Oracional	Hipotaxe	Parataxe
Complexo oracional	Correlação de Pearson	1	.914**	.788**
	Sig. (2 extremidades)		<.001	.002
	N	12	12	12
Hipotaxe	Correlação de Pearson	.914**	1	.510
	Sig. (2 extremidades)	<.001		.091
	N	12	12	12
Parataxe	Correlação de Pearson	.788**	.510	1
	Sig. (2 extremidades)	.002	.091	
	N	12	12	12

**. A correlação é significativa no nível 0,01 (2 extremidades).

As Tabelas 5 e 6 apresentam, respectivamente, o número de complexos oracionais, de alta e baixa densidade, identificados nos dados.

Com base nas informações das Tabelas 5 e 6, dois aspectos são fundamentais para responder à pergunta de pesquisa: (1) a existência de correlação entre o número de complexos oracionais e a intrincacia gramatical e (2) a relação entre o tipo de complexo oracional (hipotático ou paratático) e a ocorrência desses elementos nos textos. Para investigar o primeiro aspecto, foi realizada uma análise estatística bivariada no SPSS, correlacionando as variáveis “complexo oracional” e “intrincacia gramatical”, conforme ilustrado na Tabela 7.

Os coeficientes de correlação de Pearson (r) medem a relação linear entre variáveis, em que valores próximos de 1 indicam uma correlação positiva forte,

valores próximos de -1 denotam uma correlação negativa forte e valores próximos de 0 sugerem uma correlação fraca ou inexistente. No presente caso, $r = 0,626^*$ indica uma correlação positiva moderada, ou seja, à medida que o número de complexos oracionais aumenta, a intrincacia gramatical também tende a aumentar. A significância estatística da correlação entre as variáveis é verificada por meio do valor de significância (*Sig.*), que, neste caso, foi de 0,030. Esse valor indica que a correlação entre “complexo oracional” e “intrincacia gramatical” é estatisticamente significativa ao nível de 0,05 (bilateral). Portanto, os resultados revelam uma correlação positiva e estatisticamente significativa entre as variáveis mencionadas.

O mesmo teste foi aplicado para verificar a correlação entre as três variáveis: complexo oracional, hipotaxe e parataxe, conforme demonstrado na Tabela 8.

Como evidenciado na Tabela 8, a correlação entre “complexo oracional” e “hipotaxe” é altamente positiva ($r = 0,914^{**}$), indicando uma forte relação linear. Isso sugere que, à medida que o número de complexos oracionais aumenta, o uso de estruturas hipotáticas também tende a aumentar. De forma semelhante, a correlação entre “complexo oracional” e “parataxe” também é forte e positiva ($r = 0,788^{**}$), embora ligeiramente inferior à observada com a hipotaxe. Já a correlação entre “hipotaxe” e “parataxe” é positiva, porém mais fraca ($r = 0,510$), e não é estatisticamente significativa ao nível de 0,05, embora atinja significância ao nível de 0,10.

Os valores de significância (*Sig.*) são $<0,001$ para a correlação entre “complexo oracional” e “hipotaxe”, 0,002 para “complexo oracional” e “parataxe”, e, 0,091 para “hipotaxe” e “parataxe”. Isso indica que as correlações entre “complexo oracional” e “hipotaxe”, bem como entre “complexo oracional” e “parataxe”, são estatisticamente significativas ao nível de 0,01 (bilateral), enquanto a correlação entre “hipotaxe” e “parataxe” apresenta significância apenas ao nível de 0,10, o que limita sua robustez estatística.

Em síntese, os resultados revelam correlações positivas e estatisticamente significativas entre “complexo oracional” e “hipotaxe”, assim como entre “complexo oracional” e “parataxe”, sugerindo que essas variáveis estão diretamente relacionadas. A correlação entre “hipotaxe” e “parataxe” é mais fraca e não alcança o patamar de significância convencional de 0,05. Esses resultados indicam que os textos gerados pela IA tendem a apresentar maior complexidade gramatical, em contraste com textos mais próximos da oralidade, especialmente devido à predominância de complexos oracionais com relações hipotáticas. Tal resultado sugere que a construção do sentido textual, nesses casos, é realizada por meio de estruturas sintáticas características da escrita formal, conforme discutido por Halliday (1985). Para compreender de forma mais detalhada como essa complexidade se manifesta no sistema lógico-semântico, dois textos da amostra serão analisados a seguir.

Relações lógico-semânticas

A análise qualitativa dos complexos oracionais permite examinar as relações lógico-semânticas e compreender o funcionamento semântico dessas estruturas. Esta etapa da análise é orientada pela terceira pergunta de pesquisa: *Quais relações lógico-semânticas recorrentes ilustram a intrincacia gramatical em uma amostra de textos deste mesmo corpus?*

Para essa análise, foram selecionados dois textos: um com menor número de complexos oracionais (B2-) e outro com maior número (C1-). Esses textos representam dois níveis de proficiência distintos, sendo o texto B2(-) referente ao nível intermediário e o texto C1(-) ao nível avançado, conforme os critérios estabelecidos pelo CEFR.

Com o objetivo de compreender as escolhas lógico-semânticas realizadas pelo ChatGPT, ambos os textos são apresentados na íntegra. Todas as orações simples e os complexos oracionais foram identificados e classificados de acordo com as relações de interdependência. Além disso, as relações lógico-semânticas foram analisadas, identificando-se as ocorrências de expansão - elaboração (=), extensão (+), intensificação (x) e, no caso da projeção, os exemplos de citação (“) e ideia (‘)⁶.

Na Tabela 9, apresenta-se o texto B2(-). O texto B2(-) organiza-se em seis períodos simples e cinco complexos oracionais (equivalentes aos períodos compostos, segundo a gramática tradicional), totalizando 11 períodos. Os complexos oracionais são formados por um total de 13 orações. Para o cálculo da intrincacia gramatical, o número total de orações no texto (19) foi dividido pelo número de períodos, resultando em um índice de intrincacia gramatical de 1,7.

As relações lógico-semânticas no texto são realizadas exclusivamente por meio de recursos de expansão. Dentre esses recursos, predomina a intensificação (x), caracterizada pela presença de orações que qualificam a oração principal (dominante ou iniciadora) com uma circunstância,

Tabela 9 Exemplo B2(-): análise das relações lógico-semânticas (continuação)

α		(1i) As novas tecnologias têm transformado a sociedade humana de diversas maneiras,
$x\beta$		(1iii) oferecendo oportunidades e desafios em diferentes áreas.
Oração simples		(2i) No âmbito econômico, a tecnologia tem possibilitado a automatização de tarefas e a criação de novas formas de produção e consumo de bens e serviços.
Oração simples		(3i) Porém, essa mesma tecnologia também tem gerado preocupações quanto ao desemprego e à precarização do trabalho.
1	$x\beta$	(4i) No que se refere às relações sociais,
		(4ii) as novas tecnologias têm proporcionado novas formas de conexão entre as pessoas,
+2	α	(4iii) mas também têm gerado preocupações quanto à privacidade e à segurança dos dados pessoais.
Oração simples		(5i) Além disso, a relação entre as pessoas e as máquinas têm gerado discussões acerca da perda de empatia e da necessidade de um equilíbrio entre a tecnologia e a humanidade.
Oração simples		(6i) A educação e a cultura são fundamentais [[na preparação das pessoas para lidar com as novas tecnologias]].
α		(7i) É necessário um investimento em educação para a formação de profissionais qualificados e preparados
$x\beta$		(7iii) para lidar com as mudanças no mercado de trabalho.
Oração simples		(8i) Além disso, a cultura deve ser valorizada [[como um meio de desenvolver a criatividade e a capacidade de adaptação dos indivíduos]].
Oração simples		(9i) Em suma, o impacto das novas tecnologias para a sociedade humana é inegável.
$x\beta$		(10i) Embora ofereçam oportunidades,
α		(10iii) também apresentam desafios [[que devem ser enfrentados]].
$\alpha 1 \alpha$		(11i) É necessário
$=x \alpha 1 \beta$		(11ii) que haja um debate amplo e democrático acerca dessas mudanças,
$\beta \alpha$		(11iii) de forma a garantir
$\beta \beta$		(11iv) que a tecnologia esteja a serviço da sociedade e não o contrário.

14

especialmente de modo. Um exemplo é: “*para lidar com as mudanças no mercado de trabalho*”, presente no sétimo complexo oracional do texto.

As expansões observadas ocorrem tanto por extensão quanto por intensificação. A extensão é verificada no quarto complexo oracional, em que o uso da conjunção adversativa permite agregar um novo elemento ao oferecer uma alternativa: “*mas também têm gerado preocupações quanto à privacidade e à segurança dos dados pessoais*”, relacionado à afirmação de que as novas tecnologias têm gerado novas formas de interação. A intensificação manifesta-se, por exemplo, em “*oferecendo oportunidades e desafios em diferentes áreas*”, que amplia a oração dominante “*As novas tecnologias*”, ao acrescentar novas informações qualificadoras.

Nos cinco complexos oracionais que compõem o texto B2(-), predominam as relações lógico-semânticas de hipotaxe, sendo que um deles combina hipotaxe e parataxe. Um exemplo de complexo oracional com essa combinação é apresentado na Tabela 10.

Tabela 10 Combinação de relações de hipotaxe e parataxe no exemplo B2

		(4i) No que se refere às relações sociais,
	$x\beta$	(4ii) as novas tecnologias têm proporcionado
1		novas formas de conexão entre as pessoas,
+2	α	(4iii) mas também têm gerado preocupações
		quanto à privacidade e à segurança dos
		dados pessoais.

No exemplo, o complexo oracional é iniciado por uma oração dependente que apresenta um atributo (“*relações sociais*”), seguida pela oração dominante, cujo portador é “*as novas tecnologias*”, configurando uma estrutura hipotática. Essa oração dominante também funciona como a oração iniciadora do complexo paratático, sendo seguida por uma oração continuativa, que retoma o referente “*as novas tecnologias*”.

Ainda no âmbito desse complexo oracional, o ChatGPT estabelece uma relação lógico-semântica de expansão a partir da informação contida na oração dominante e iniciadora. Inicialmente, na oração dependente, a expansão ocorre por intensificação, por meio de uma circunstância causal. Em seguida, na parataxe, a oração continuativa expande a anterior por extensão, detalhando os efeitos gerados pelas novas tecnologias: “*preocupações quanto à privacidade e à segurança dos dados pessoais*”. Dessa forma, as relações lógico-semânticas giram em torno de um mesmo referente, modificando a ordem direta da oração e resultando em maior intrincacia gramatical.

Vale destacar que esse complexo oracional poderia, sem prejuízo semântico, ter sido reformulado como uma oração simples, o que reduziria a intrincacia e aumentaria a densidade lexical do texto. Por exemplo: “*As novas tecnologias têm proporcionado não só novas formas de conexão entre as pessoas, mas também preocupações quanto à privacidade e à segurança dos dados pessoais*”. Observa-se que a ideia de “*relações sociais*” já está implícita na expressão “*conexão entre as pessoas*”. Outra reformulação possível envolveria o uso de um termo resumidor (empacotamento), estabelecendo uma relação paratática no complexo oracional: “*as novas tecnologias têm proporcionado novas formas de conexão entre as pessoas e gerado preocupações quanto à privacidade e à segurança dos dados pessoais*”.

Em relação à hipotaxe, a Tabela 11 ilustra as relações geradas pelo ChatGPT no complexo oracional do texto B2(-).

Tabela 11 Relação de hipotaxe no texto B2(-)

$x\beta$	(10i)	Embora ofereçam oportunidades,
α	(10ii)	(Θ) também apresentam desafios [[que devem ser enfrentados]].

O complexo oracional ilustrado corresponde à décima oração (ou período, segundo a gramática tradicional) do texto. Essa construção resulta de uma relação hipotática, iniciada pela oração dependente “*Embora ofereçam oportunidades*”, seguida por uma oração dominante, na qual o participante elíptico ((Θ) -> *as novas tecnologias*) precisa ser recuperado da oração anterior, e finalizada com uma oração encaixada. Esse tipo de construção é recorrente na produção escrita do ChatGPT, conferindo aos textos maior intrincacia gramatical – uma característica que também pode ser observada, embora de forma distinta, na língua falada.

A expansão por intensificação é empregada na primeira oração dependente “*ofereçam oportunidades*”, que reformula e contrapõe a informação contida na oração dominante “*apresentam desafios*”. Esse complexo oracional poderia ser reestruturado por meio de um processo de nominalização, como em: “*o enfrentamento de desafios e o oferecimento de oportunidades (pelas novas tecnologias)*”, resultando em uma maior carga semântica e, consequentemente, em menor intrincacia gramatical.

O segundo texto, apresentado na íntegra a seguir, pertence ao nível de proficiência C1(-) do CEFR. Assim como no texto B2(-), as relações de interdependência e lógico-semânticas são indicadas, a fim de proporcionar uma visão clara de como a complexidade gramatical se manifesta nesse nível de proficiência (Tabela 12).

O texto C1(-) é composto por um período simples e 15 complexos oracionais (equivalentes aos períodos compostos, segundo a gramática tradicional), totalizando 40 orações. A razão entre o número total de orações (41) e o número de períodos (16) resulta em um índice de intrincacia gramatical de 2,6. Esse valor difere do texto B2(-),

Tabela 12 Relações de interdependência e lógico-semânticas no texto C1 (–)

α $+x\beta$	(1i) (1iit)	A vida é uma arte [[que exige equilíbrio, sabedoria e muita paciência]] para ser dominada.
1 =2 α $x\beta$	(2i) (2iit) (2iit)	A busca pela felicidade e pelo bem-estar é uma constante na vida humana, mas será que existe um segredo [[para viver bem]]
α $+ \beta\alpha$ $\beta\beta$	(3i) (3iit) (3iit)	Penso que a arte de viver bem está nas pequenas coisas da vida, [[aquelas que muitas vezes passam despercebidas diante dos nossos olhos]].
Oração simples	(4i)	Aprendi que a gratidão [[é um dos pilares para a felicidade.]]
2 1 +2 $x\beta\beta$ α $\beta\alpha$	(5i) (5iit) (5iit) (5iiv)	Quando somos gratos [[pelo que temos, por quem somos e pelas pessoas ao nosso redor]], nossa visão de mundo se transforma e passamos a valorizar o que realmente importa.
α $+ \beta$	(6i) (6i)	A gratidão nos ajuda a ser mais humildes, tolerantes e solidários.
$x\beta$ α	(7i) (7iit)	Outro ponto fundamental para viver bem é o equilíbrio entre o trabalho e o lazer.
1 =2 α $+B$	(8i) (8iit) (8iit)	Trabalhar é importante, mas é preciso saber dosar as responsabilidades com momentos de descanso e diversão.
α $x\beta$ 1 = β 2	(9i) (9iit) (9iit)	É nos momentos de lazer que recuperamos nossas energias e podemos desfrutar das coisas boas da vida.
α $+ \beta$	(10i) (10iit)	Por fim, mas não menos importante, a arte de viver bem está nas relações humanas.
1 +2	(11i) (11iit)	As pessoas ao nosso redor são nossa maior riqueza e precisamos cultivar boas relações [[para vivermos em harmonia]].
1 +2	(12i) (12iit)	Cada pessoa [[que passa pela nossa vida]] tem um propósito e pode nos ensinar algo valioso.
1 $x2$ α = β	(13i) (13iit) (13iit)	A vida é uma jornada [[que não tem fim]], e a arte de viver bem é um aprendizado diário.
α $+ \beta$ 1 $+ \beta$ 2 $+ \beta$ 3	(14i) (14iit) (14iit) (14iiv)	Precisamos estar abertos a novas experiências, aprender com nossos erros e valorizar cada momento de felicidade.
α $x\beta$	(15i) (15iit)	Viver bem é um estado de espírito [[que depende da forma como enxergamos a vida e as pessoas ao nosso redor]].
α $+ \beta$ 1 $+ \beta$ 2	(16i) (16iit) (16iiv)	Afinal, a verdadeira arte de viver bem está em ser feliz e fazer o bem [[a quem nos rodeia]].

no qual se observou uma redução simultânea da intrincacia gramatical e da densidade lexical, destoando da tendência observada nos demais níveis. No texto C1(+), a intrincacia foi de 2,0 e a densidade de 4,8; já no C1(-), a densidade é de 4,3, o que confirma, ainda que com variação mínima, a relação inversamente proporcional entre essas duas variáveis, conforme proposto por Halliday (1995, 2004).

Assim como no texto anterior, e apesar das diferenças temáticas e de nível de proficiência, as relações lógico-semânticas no texto C1(-) baseiam-se exclusivamente em recursos de expansão, com predominância das extensões (+), seguidas por intensificações (x) e elaborações (=). Essa tendência é evidente nos complexos oracionais do texto, dos quais sete combinam hipotaxe e parataxe, seis são formados somente por hipotaxe, dois apenas por parataxe e há apenas uma oração simples.

Para ilustrar a intrincacia gramatical produzida por meio de orações dependentes, apresenta-se a seguir o primeiro complexo oracional do texto (Tabela 13).

No exemplo da Tabela 13, o complexo oracional é formado por hipotaxe e a relação lógico-semântica corresponde a uma expansão por intensificação (x) da oração dominante, realizada por meio de uma circunstância de modo. O exemplo da Tabela 14 ilustra a predominância dos complexos oracionais no texto C1(-), que frequentemente combinam parataxe e hipotaxe.

Tabela 13 Intrincacia gramatical por meio de orações dependentes

α	(i) A vida é uma arte [[que exige equilíbrio, sabedoria e muita paciência]]
$+x\beta$	(iii) para ser dominada.

Tabela 14 Combinação de parataxe e hipotaxe

2	$x\beta\beta$	(si) Quando somos gratos [[pelo que temos, por quem somos e pelas pessoas ao nosso redor]],
1	α	(siii) nossa visão de mundo se transforma
$+2$	$\beta\alpha$	(siii) e passamos a valorizar
		(siv) o que realmente importa.

O quinto complexo oracional é composto por quatro orações, nas quais tanto a oração dominante quanto a iniciadora desempenham a função de oração principal. Vale destacar que a oração iniciadora e dominante não ocupa a posição inicial no complexo – uma escolha recorrente na produção do ChatGPT – o que resulta em textos gramaticalmente mais intrincados. No entanto, é importante observar que maior intrincacia gramatical não implica necessariamente maior formalidade, uma vez que esse aspecto é determinado pelos recursos de metáfora gramatical (Halliday, 2004; Eggins, 2004; Rottava & Santos, 2018).

Neste exemplo, todas as relações lógico-semânticas são de expansão, especificamente por extensão (+) e intensificação (x), referindo-se ao participante “*gratidão*”, mencionado na quarta oração. Esses recursos conferem ao texto um caráter mais abstrato, sem o fornecimento de exemplos ou parâmetros concretos para ilustrar esse referente. Essa abstração é reforçada pelos processos verbais predominantes no complexo oracional: processos relacionais (“*somos*” e “*temos*”), que contribuem para o caráter conceitual do texto; processos materiais (“*transformar*”, “*passar*”), que indicam mudanças no fluxo dos eventos e produzem efeitos diferentes dos demais; e, por fim, um processo mental (“*importar*”), que se relaciona à experiência consciente do leitor.

Considerações finais

O objetivo central deste estudo foi investigar como o ChatGPT gera textos em línguas adicionais, com base nas instruções correspondentes aos seis níveis de proficiência definidos pelo CEFR. A análise concentrou-se em variáveis relacionadas aos níveis de proficiência, à interdependência entre orações e às relações lógico-semânticas, com o intuito de compreender a organização textual, especialmente em termos de intrincacia gramatical.

Os resultados obtidos a partir dos textos selecionados revelaram uma correlação entre intrincacia gramatical e densidade lexical, coerente com o que já é discutido na literatura, particularmente nos

trabalhos de Halliday (1985, 2004). Esses dados indicam que o fato de os textos serem gerados por IA não afeta significativamente essa relação. Em média, os textos com baixa densidade lexical (D-) apresentaram índices mais elevados de intrinsecidade gramatical em comparação aos textos com alta densidade lexical (D+). No entanto, os textos D(-) demonstraram maior variabilidade, enquanto os textos D(+) exibiram índices mais baixos e estáveis. Esses resultados estão alinhados ao estudo de Syarif (2019), que analisou introduções de trabalhos acadêmicos redigidos por estudantes de pós-graduação em inglês como língua adicional, constatando que textos com menor densidade lexical tendem a apresentar maior intrinsecidade gramatical.

Os textos analisados continham um número superior de complexos oracionais (99) em relação a orações simples (45), o que sugere que o ChatGPT tende a gerar textos com maior complexidade gramatical do que seria esperado, especialmente nos níveis iniciais de proficiência. Esse padrão indica que a organização dos complexos oracionais pelo ChatGPT resulta em uma construção textual mais intrincada.

A predominância da intrinsecidade gramatical observada nos textos gerados pode, em alguns casos, facilitar a compreensão, conforme argumentam Gultom e Pintubatu (2022). Isso ocorre porque, embora apresentem maior complexidade estrutural, os textos com intrinsecidade gramatical elevada podem organizar o fluxo de informações de modo mais claro. Por outro lado, a alta densidade lexical pode dificultar a leitura ao condensar excessivamente a informação. Ressalta-se que a média de densidade lexical não variou entre os diferentes níveis de proficiência, conforme Tabela 15.

Em termos de complexos oracionais, observaram-se predominantemente relações de interdependência do tipo hipotática, seguidas por estruturas que combinam hipotaxe e parataxe, com menor ocorrência de parataxe isolada e de orações simples. A análise evidenciou uma correlação entre a presença de complexos oracionais e a intrinsecidade gramatical, indicando que a hipotaxe tende a dominar a

Tabela 15 Média de densidade lexical do corpus completo

Nível de proficiência	Média de densidade lexical
Nível A1	5,1
Nível A2	5,4
Nível B1	4,6
Nível B2	5,1
Nível C1	4,8
Nível C2	5,1

organização oracional, representando 78,72% dos casos nos textos D(+) e 77,65% em textos D(-). No entanto, esperava-se maior incidência de parataxe nos níveis iniciais (A1 e A2), padrão comum na escrita de aprendizes humanos de PLA, que frequentemente utilizam períodos simples e curtos.

Quanto às relações lógico-semânticas observadas na amostra de dois textos analisados qualitativamente, verificou-se a predominância de expansões, sem ocorrências de projeção. A análise revelou que o ChatGPT tende a agregar constantemente novos elementos aos seus referentes, o que se reflete na estrutura dos complexos oracionais. A alta frequência de intensificações sugere que a ferramenta frequentemente inclui circunstâncias adicionais aos participantes e processos descritos no texto.

Este estudo oferece uma contribuição relevante para o entendimento do impacto das ferramentas de IA no ensino de línguas adicionais. Esse impacto manifesta-se na organização lógico-semântica dos textos gerados, cuja análise considerou aspectos de complexidade gramatical e semântica. Sob esse viés, a LSF oferece recursos valiosos para uma análise detalhada do potencial textual, permitindo o desenvolvimento de estratégias práticas que auxiliem os aprendizes de línguas adicionais a aprimorar sua proficiência com base na estrutura dos textos produzidos. Assim, a compreensão da intrinsecidade gramatical nos textos gerados pelo ChatGPT pode fomentar reflexões importantes sobre a construção de complexos oracionais e a prática da escrita (Ma'mun, 2017).

Entretanto, reconhece-se que embora a LSF forneça uma estrutura robusta para descrever aspectos gramaticais e semânticos, apresenta limitações quando aplicada à análise de práticas pedagógicas em contextos reais de uso da IA, especialmente no ensino da escrita. Nesse sentido, a abordagem adotada neste estudo pode ser expandida em pesquisas futuras que considerem diferentes gêneros textuais e outras línguas adicionais. Do ponto de vista didático, uma possibilidade promissora consiste na elaboração de tarefas baseadas na pedagogia de gêneros, sob a perspectiva da LSF (Rose & Martin, 2012), permitindo a articulação entre os princípios teóricos, o uso de ferramentas de IA e metodologias de ensino da escrita em PLA.

Do ponto de vista pedagógico, os resultados deste estudo contribuem para o ensino de PLA ao evidenciar padrões de organização lógico-semântica nos textos gerados por IA. Professores podem utilizar esse conhecimento para promover maior consciência linguística entre os aprendizes, por meio da análise crítica da estrutura de complexos oracionais, da densidade lexical e da intrincacia gramatical presentes em textos gerados por ferramentas como o ChatGPT (Doyle, 2008). Em especial, o estudo alerta para o fato de que a IA pode produzir textos que não se alinham aos níveis de proficiência esperados, o que exige uma mediação docente criteriosa na seleção e no uso de tais materiais. Em estudos futuros, será fundamental considerar também os aspectos éticos envolvidos na utilização da IA na produção textual, como a autoria e a autenticidade. Embora não tenha sido o foco deste artigo, reconhece-se a relevância dessas questões no contexto educacional contemporâneo.

Referências

- Christie, F. & Unsworth, L. (2000). Developing socially responsible language research. In L. Unsworth (Ed.), *Researching language in schools and communities: Functional linguistic perspectives* (pp. 1–26). Cassell.
- Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.
- Da Silva, A. M. & Rottava, L. (2024). Densidade lexical em textos gerados pelo ChatGPT: Implicações da inteligência artificial para a escrita em línguas adicionais. *Texto Livre*, 17, 1–19. <https://doi.org/10.1590/1983-3652.2024.47836>
- Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford University Press.
- Doyle, H. (2008). Grammatical intricacy – The pedagogical significance of counting by word, clause or idea. In C. Wu, C. Matthiessen, & M. Herke (Eds.), *Proceedings of ISFC 35: Voices around the world*. Sydney.
- Eggins, S. (2004). *An introduction to systemic functional linguistics* (2nd Ed.). Pinter Publishers.
- Figueredo, G. (2015). Uma descrição sistêmico-funcional dos marcadores discursivos avaliativos em português brasileiro: a gramática das partículas modais. *Alfa*, 59(2), 281–307. <https://doi.org/10.1590/1981-5794-1504-3>
- Firth, J. R. (1957). A synopsis of linguistic theory 1930–1955. In *Studies in linguistic analysis* (pp. 1–32). Blackwell.
- Ghio, E. & Fernández, M. D. (2008). *Lingüística sistémico funcional. Aplicaciones a la lengua española* (2a. Ed.). Waldhuter Editorial.
- Gultom, J. J. & Pintubatu, E. (2022). Grammatical intricacy and lexical density of reading texts of English textbook for senior high school grade XII. *BAHAS*, 33(2), 173–182. <https://doi.org/10.24114/bhs.v33i2.35972>
- Halliday, M. A. K. (1985). *An introduction to functional grammar*. Edward Arnold.
- Halliday, M. A. K. (1993). Language and the order of nature. In M. A. K. Halliday & J. R. Martin (Eds.), *Writing science: Literacy and discursive power* (pp. 117–136). Routledge.
- Halliday, M. A. K. (1978). *Language as a social semiotic: The social interpretation of language and meaning*. Edward Arnold.
- Halliday, M. A. K. (1989). Part A. In M. A. K. Halliday & R. Hasan, *Language, context and text* (2nd ed.; pp. 3–49). Oxford University Press.
- Halliday, M. A. K. (1987/2002). Spoken and written modes of meaning. In M. A. K. Halliday & J. Webster (Eds.), *On grammar. Collected works of M. Halliday* (vol. 1; pp. 323–351). Edited by Jonathan Webster. Continuum.
- Halliday, M. A. K. (2004). *The language of science*. Continuum.
- Halliday, M. A. K. & Hasan, R. (1989). *Language, context and text* (2nd Ed.). Oxford University Press.

- Halliday, M. A. K. & Matthiessen, C. M. I. M. (2014). *An introduction to functional grammar* (4th Ed.). Edward Arnold.
- Johansson, V. (2008). Lexical diversity and lexical density in speech and writing: a develop-mental perspective. *Working Papers* 53 (pp. 61–79). Lund University.
- Lancaster, T. (2023). Artificial intelligence, text generation tools and ChatGPT – does digital watermarking offer a solution? *International Journal for Educational Integrity*, 19(10), 1–14. <https://doi.org/10.1007/s40979-023-00131-6>
- Ma'mun, N. (2017). Grammatical intricacy on students' writing. *Vision: Journal for Language and Foreign Language Learning*, 6(1), 89–98. <https://doi.org/10.21580/vjv6i11619>
- Malinowski, B. (1935). *Coral Gardens and their magic*, (vol. II): *The language of magic and gardening*. Allen and Unwin.
- Malinowski, B. (1923). The problem of meaning in primitive languages. In C. K. Ogden & I. A. Richards (Eds.), *The meaning of meaning* (8th Ed.; pp. 296–336). Harper Paperbacks.
- Ramos, A. S. M. (2023). Generative artificial intelligence based on large language models —Tools for use in academic research. *SciELO Preprints*. <https://doi.org/10.1590/SciELOPreprints.6105>
- Rose, D., Martin, J. R. (2012). *Learning to write, reading to learn: Genre, knowledge and pedagogy in the Sydney School*. Equinox Publishing Ltd.
- Rottava, L. & DaSilva, A. M. (2023). Sistematológico-semântico de expansão na reescrita de textos acadêmicos: escolhas linguísticas de uma estudante versus as escolhas do ChatGPT. *Diálogo das Letras*, 12, 1–18. <https://doi.org/10.22297/2316-17952023v12e02307>
- Rottava, L. & Santos, S. S. dos (2018). Os efeitos de construções metafóricas em textos produzidos em contexto acadêmico. *DELTA: Documentação e Estudos em Linguística Teórica e Aplicada*, 34(1), 55–79. <https://doi.org/10.1590/0102-445085262849162431>
- Santos, S. S.; Silva, L. B. M. G. & Bordim, C. (2024). O sistema de atitude em textos criados pelo ChatGPT – uma análise de textos em português como língua adicional. *Entretextos*, 24(2), 193–214. <https://doi.org/10.5433/1519-5392.2024v24n2p193-214>
- Syarif, H. (2019). Lexical density vs. grammatical intricacy: how are they related? *Proceedings of the Sixth International Conference on English Language and Teaching (ICOELT 2018)*, 276, 16–23. <https://doi.org/10.2991/icoelt-18.2019.3>

Anexo. Instruções das atividades realizadas pelos participantes

Write a text in Portuguese for the beginner level A1 following the instructions given below:	Write a text in Portuguese for the elementary level A2 following the instructions given below:
Tema: Amizade e fazer amigos Escreva um texto de 80 a 90 palavras sobre como fazer amigos. Fale sobre: • Como conhecer pessoas novas • Como ser um bom amigo • O que evitar em uma amizade • Como fez um amigo e o que fez para mantê-lo próximo.	Tema: Estresse Escreva um texto de 120 a 150 palavras sobre o estresse. Fale sobre os seguintes pontos: • O que é estresse • As causas do estresse • Como lidar com o estresse • A importância de reduzir o estresse.

Write a text in Portuguese for the intermediate level B1 following the instructions given below:

Tema: A educação no Brasil

Escreva um texto de 180 a 200 palavras sobre a educação no Brasil. Inclua os seguintes pontos nas informações:

- Sistema educacional no Brasil
- Desafios da educação no Brasil
- Programas governamentais para melhorar a educação no país
- O papel do professor.

Write a text in Portuguese for the upper-intermediate level B2 following the instructions given below:

Tema: O impacto das novas tecnologias para a sociedade humana

As novas tecnologias estão transformando rapidamente a nossa forma de viver, trabalhar e interagir com o mundo. Essas mudanças são tão profundas que podem afetar significativamente as relações sociais, a economia e a cultura. Nesse sentido, escreva um texto argumentativo que discuta o impacto das novas tecnologias para a sociedade humana, considerando os seguintes pontos:

- Oportunidades e desafios das novas tecnologias
- A relação entre as pessoas e as máquinas
- As mudanças na forma como produzimos e consumimos bens e serviços.
- Educação e cultura na preparação das pessoas para lidar com as novas tecnologias.

Ao escrever o seu texto, lembre-se de apresentar argumentos claros e coerentes, fundamentados em exemplos e evidências. O seu texto deve ter entre 200 e 250 palavras.

Write a text in Portuguese for the advanced level C1 following the instructions given below

Tema: A arte de viver bem

A busca pela felicidade e pelo bem-estar é uma constante na vida humana. Muitos procuram o segredo para viver bem, mas será que esse segredo existe? Nessa crônica, reflita sobre a arte de viver bem, considerando os seguintes pontos:

- O valor das pequenas coisas.
- A importância da gratidão.
- O equilíbrio entre trabalho e lazer.
- A importância das relações humanas.

Ao escrever a sua crônica, use sua experiência pessoal e observações do mundo ao seu redor para construir um texto pessoal e intimista. Lembre-se de usar recursos literários, metáforas e analogias, para tornar o seu texto mais expressivo e envolvente. O seu texto deve ter entre 250 e 300 palavras.

Write a text in Portuguese for the proficiency level C2 following the instructions given below

Tema: Pontes culturais

A interculturalidade é um tema cada vez mais presente no mundo globalizado em que vivemos. A diversidade cultural é uma riqueza que deve ser celebrada e valorizada, mas, muitas vezes, a falta de entendimento e comunicação entre culturas diferentes pode gerar conflitos e incompreensões. Nessa atividade de escrita criativa, escreva um texto explorando o tema da interculturalidade, considerando os seguintes pontos:

- A beleza da diversidade
- As barreiras da comunicação
- A importância do intercâmbio cultural
- A construção de pontes culturais.

Ao escrever o seu texto, use a sua criatividade e imaginação para explorar diferentes aspectos da interculturalidade. Utilize recursos literários, metáforas e analogias, para tornar o seu texto mais expressivo e envolvente. O seu texto deve ter entre 250 e 300 palavras.

Como citar este artigo: Silva, A. M. & Rottava, L. (2026). Intricacia gramatical em textos gerados pelo ChatGPT em língua adicional: uma perspectiva sistêmico-funcional. *Íkala, Revista de Lenguaje y Cultura*, 31(1), e358152. <https://doi.org/10.17533/udea.ikala.358152>