

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO  
PUC-SP**

**Amanda Chiarelo Boldarine**

**Uso de metáforas por aprendizes brasileiros bilíngues de inglês**

**Mestrado em Linguística Aplicada e Estudos da Linguagem**

**SÃO PAULO  
2024**

**Amanda Chiarelo Boldarine**

**Uso de metáforas por aprendizes brasileiros bilíngues de inglês**

**Mestrado em Linguística Aplicada e Estudos da Linguagem**

Dissertação apresentada à Banca Examinadora da Pontifícia Universidade Católica de São Paulo, como exigência parcial para obtenção do título de Mestre em Linguística Aplicada e Estudos da Linguagem sob orientação do Professor Doutor Antonio Paulo Berber Sardinha.

**SÃO PAULO  
2024**

Autorizo, exclusivamente para fins acadêmicos e científicos, a reprodução total ou parcial desta dissertação de mestrado, por processos fotocopiadores ou eletrônicos, desde que devidamente citada.

Assinatura:

Data:

## **FICHA CATALOGRÁFICA**

**BOLDARINE**, Amanda Chiarelo. São Paulo, 2024.

**Dissertação de Mestrado** – Pontifícia Universidade Católica de São Paulo.

**Área de Concentração:** Linguística Aplicada e Estudos de Linguagem.

**Orientador:** Professor Doutor Antonio Paulo Berber Sardinha.

**Amanda Chiarelo Boldarine**

Uso de metáforas por aprendizes brasileiros bilíngues de inglês

Aprovado em     /     /     .

Dissertação apresentada à Banca Examinadora da Pontifícia Universidade Católica de São Paulo, como exigência parcial para obtenção do título de Mestre em Linguística Aplicada e Estudos da Linguagem sob orientação do Professor Doutor Antonio Paulo Berber Sardinha

**Banca Examinadora**

---

---

---

Aos meus pais, Ismael e Marly.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior- Brasil (CAPES) – número do processo 88887.675907/2022-00.

## AGRADECIMENTOS

Primeiramente, agradeço ao meu orientador, Prof. Dr. Tony Berber Sardinha, pelos ensinamentos, orientação e confiança. Obrigada por ter acreditado em mim e por ter me guiado ao longo deste processo.

Aos meus pais, Ismael e Marly, por me ensinarem a importância da educação e a ser persistente. Aos meus irmãos, Carolina e Rafael, e à minha cunhada Luciana pelo apoio.

Ao Valter, meu marido, que me apoiou desde quando decidi que faria o mestrado. Sou grata por todo o seu incentivo, paciência e compreensão. Obrigada por todas as conversas, por me ajudar a focar e por acreditar em mim. Que sorte a minha de ter você nessa jornada!

Aos meus sogros, Valter e Ruth, pelo incentivo e às minhas cunhadas Rita e Rosária. Rita, obrigada pelo apoio. Rosária, obrigada por ter sido um exemplo de pessoa e pesquisadora.

À Beacon School por ter autorizado a coleta do corpus utilizado nesta pesquisa. Sou grata à Cacai, Luciana e Audrey pela confiança depositada em mim.

Ao meu amigo Douglas por ter me ajudado a baixar o Ubuntu, Stanford Parse, TreeTagger e tantos outros, além de sua paciência em me ensinar a usá-los e sua disponibilidade para auxiliar sempre que necessário.

Aos colegas do GELC (Grupo de Estudos de Linguística de Corpus) pela colaboração. À Me. Aline Zamboni e ao Prof. Me. Cícero Soares por me ajudarem a entender a AMD. Ao Lucas Ferraz por me ajudar com os processos do LAEL. Agradeço, especialmente, o Prof. Me. Marcos Oliveira por sua contribuição no desenvolvimento da pesquisa. Marcos, obrigada pela parceria!

Às Profas. Dra. Mara e Dra Cláudia Delfino que concordaram em participar da minha banca de qualificação e pelos apontamentos que ajudaram no desenvolvimento da pesquisa.

## RESUMO

BOLDARINE, A. C. **Uso de metáforas por aprendizes brasileiros bilíngues de inglês.** (Dissertação de Mestrado) - LAEL. PUC. São Paulo, 2024.

A pesquisa em Linguística de Corpus de Aprendiz tem tradicionalmente focado em questões gramaticais (como, Rankin, 2015) ou lexicais (como, Cobb e Horst, 2015), na sua maioria. Desse modo, questões ligadas a como os alunos constroem discursos nas tarefas escolares não têm recebido a atenção devida. Desse modo, o presente estudo pretendeu contribuir para preencher essa lacuna por meio de uma investigação acerca dos discursos presentes em composições escolares de alunos do Ensino Fundamental II e Primeiro Ano do Ensino Médio. Para isso, empregou-se a Análise Multidimensional Lexical (Berber Sardinha e Fitzsimmons Doolan, 2024) para identificar os principais discursos em um corpus de 63.297 palavras. A Análise Multidimensional Lexical é uma abordagem baseada na Análise Multidimensional (Biber, 1988) que permite detectar os discursos prevalentes em corpora, por meio da aplicação de procedimentos estatísticos multivariados em dados lexicais extraídos do corpus. Na pesquisa em Análise Multidimensional Lexical, existe uma lacuna referente à presença de linguagem metafórica nas dimensões. Mais especificamente, não se sabe até que ponto a linguagem metafórica contribuiu para a dimensionalidade dos textos. Assim, a presente pesquisa pretendeu preencher essa segunda lacuna, referente à relação entre dimensões lexicais e metaforicidade. Por fim, entre os estudos de metáfora baseados em corpora, existe há muito tempo a tentativa de detectar metáforas automaticamente, para expandir o escopo de tamanho dos corpora normalmente usados na pesquisa, já que a análise de metáforas em corpora é geralmente feita manualmente, restringindo a quantidade de dados a serem analisados. Dada essa restrição e a disponibilidade recente da Inteligência Artificial, por meio de chatbots como o ChatGPT, Gemini e Llama, este estudo visou a contribuir para a verificação da capacidade de identificação automática de metáforas com a ajuda do ChatGPT. A partir da Análise Multidimensional Lexical, foram identificadas oito dimensões discursivas lexicais. Por exemplo, na Dimensão 1 “Conhecimento abstrato, teórico e científico versus Dinâmicas familiares, resiliência pessoal e crescimento emocional”, o discurso do polo positivo enfatiza a pesquisa estruturada e a análise científica, comum em contextos acadêmicos. Já o polo negativo focou em narrativas pessoais, relacionamentos familiares e superação de problemas pessoais. Para a anotação manual das metáforas, usou-se o MIP (Pragglejaz Group, 2007) e os candidatos à metáfora foram extraídos da lista de variáveis de cada dimensão. Ao final da anotação manual, 245 ocorrências (tokens) dos candidatos foram

anotados como metafóricos, o que corresponde a 37% dos candidatos (types). Quanto à anotação automática das metáforas via ChatGPT4, foi identificada uma em cada quatro ou seis das metáforas identificadas pelos analistas usando o protocolo MIP (Pragglejaz Group, 2007). Concluindo, o presente estudo contribuiu para o entendimento dos discursos mobilizados pelos alunos na execução das tarefas, como a tarefa influencia o uso de metáforas linguísticas, assim como a relação entre a idade ou proficiência dos alunos e a incidência de metáforas. Sobre a automatização da identificação de metáforas, a pesquisa apontou possíveis princípios para desenvolvimento de *prompts*, como restringir o contexto em que a metáfora linguística ocorre. Contudo, a inconsistência na anotação segue como a principal limitação.

**Palavras-chave:** Linguística de Corpus. Corpus de Aprendiz. Análise Multidimensional Lexical. Metáfora. Inteligência Artificial.

## ABSTRACT

BOLDARINE, A. C. **The use of metaphors by Brazilian bilingual English learners.** (Master's Dissertation) - LAEL. PUC. São Paulo, 2024.

Research in Learner Corpus Linguistics has traditionally focused mostly on grammatical (e.g., Rankin, 2015) or lexical (e.g., Cobb and Horst, 2015) issues. Consequently, much less is known about how students construct discourse in school tasks. This study aimed to address this gap by investigating the discourses present in written texts by Middle School and First-year High School students in Brazil. To achieve this goal, Lexical Multidimensional Analysis was employed (Berber Sardinha and Fitzsimmons Doolan, 2024) to identify the main discourses in a corpus of 450 texts, amounting to 63,297 words. Lexical Multidimensional Analysis is an off-shoot of Multidimensional Analysis (Biber, 1988), which allows for the detection of the underlying discourses in corpora through the application of multivariate statistical procedures to lexical data extracted from the corpus. However, in the Lexical Multidimensional Analysis literature, little is known about whether metaphorical language is manifested in the dimensions, that is, the extent to which metaphorical language contributes to the dimensionality of texts. Thus, this study aimed to fill this second gap, seeking to determine if the lexical items forming the dimensions are metaphorically used or not. Moreover, in the corpus-based metaphor literature, there has long been a desire to automate metaphor detection, so as to enable the analysis of larger corpora in metaphor studies, as metaphor analysis in corpora is generally done manually, restricting the amount of data to be analyzed. Given this limitation and the recent availability of Artificial Intelligence through chatbots like ChatGPT, Gemini, and Llama, this study aimed to contribute to text the metaphor detection capabilities of ChatGPT. Eight lexical discursive dimensions were identified through Lexical Multidimensional Analysis. For example, in Dimension 1 "Abstract, theoretical, and scientific knowledge versus Family dynamics, personal resilience, and emotional growth", the discourse in the positive pole emphasizes structured research and scientific analysis, common in academic contexts. On the other hand, the negative pole focused on personal narratives, family relationships, and overcoming personal problems. For the manual annotation of metaphors, the MIP (Pragglejaz Group, 2007) was used, and metaphor candidates were extracted from the variable list of each dimension. At the end of the manual annotation, 245 occurrences (tokens) of candidates were annotated as metaphorical, corresponding to 37% of the candidates (types). Regarding the automatic annotation of metaphors via ChatGPT4, at best one in every four (25%) of the metaphors identified by the

human analysts using the MIP protocol (Pragglejaz Group, 2007) was identified automatically through AI. In conclusion, this study contributed to understanding not only the discourses mobilized by students in doing the tasks, but also to how tasks influence the use of linguistic metaphors, as well as to describing the relationship between students' age or proficiency and the incidence of metaphors. Regarding the automation of metaphor identification, the research tried out a range of prompts, charting the process of prompt development, such as restricting the context in which linguistic metaphor occurs. However, the results suggest current LLMs are unable to detect the majority of the metaphors in the corpus.

**Keywords:** Corpus Linguistics. Learner Corpora. Lexical Multidimensional Analysis. Metaphor. Artificial Intelligence.

## LISTA DE TABELAS

TABELA 1 - <i>PROMPT</i> .....	32
TABELA 2 - TIPOS DE ERROS DE IDENTIFICAÇÃO DE METÁFORAS POR IA.....	33
TABELA 3 - COMPOSIÇÃO DO CABI.....	35
TABELA 4 - COMPOSIÇÃO FINAL DO CABI.....	36
TABELA 5 - EXEMPLO DE ANOTAÇÃO MANUAL.....	41
TABELA 6 - RÓTULO DAS DIMENSÕES DISCURSIVAS.....	49
TABELA 7 - VARIÁVEIS LEXICAIS DA DIMENSÃO 1.....	51
TABELA 8 - VARIÁVEIS LEXICAIS DA DIMENSÃO 2.....	56
TABELA 9 - VARIÁVEIS LEXICAIS DA DIMENSÃO 3.....	60
TABELA 10 - VARIÁVEIS LEXICAIS DA DIMENSÃO 4.....	64
TABELA 11 - VARIÁVEIS LEXICAIS DA DIMENSÃO 5.....	69
TABELA 12 - VARIÁVEIS LEXICAIS DA DIMENSÃO 6.....	74
TABELA 13 - VARIÁVEIS LEXICAIS DA DIMENSÃO 7.....	79
TABELA 14 - VARIÁVEIS LEXICAIS DA DIMENSÃO 8.....	84
TABELA 15 - ANOVAs.....	89
TABELA 16 - DISTRIBUIÇÃO DOS CANDIDATOS À METÁFORA NAS DIMENSÕES.....	90
TABELA 17 - DISTRIBUIÇÃO DAS METÁFORAS IDENTIFICADAS NAS DIMENSÕES.....	92
TABELA 18 - COMPARAÇÃO DOS RESULTADOS DAS ANOTAÇÕES POR IA.....	100

## LISTA DE FIGURAS

FIGURA 1 - DISTRIBUIÇÃO DAS TAREFAS NO CABI.....	36
FIGURA 2 - GRÁFICO DE SEDIMENTAÇÃO.....	38
FIGURA 3 - ESCORES MÉDIOS DAS TAREFAS NA DIMENSÃO 1.....	55
FIGURA 4 - ESCORES MÉDIOS DAS TAREFAS NA DIMENSÃO 2.....	59
FIGURA 5 - ESCORES MÉDIOS DAS TAREFAS NA DIMENSÃO 3.....	64
FIGURA 6 - ESCORES MÉDIOS DAS TAREFAS NA DIMENSÃO 4.....	69
FIGURA 7 - ESCORES MÉDIOS DAS TAREFAS NA DIMENSÃO 5.....	74
FIGURA 8 - ESCORES MÉDIOS DAS TAREFAS NA DIMENSÃO 6.....	78
FIGURA 9 - ESCORES MÉDIOS DAS TAREFAS NA DIMENSÃO 7.....	83
FIGURA 10 - ESCORES MÉDIOS DAS TAREFAS NA DIMENSÃO 8.....	88

## SUMÁRIO

<b>1. INTRODUÇÃO</b> .....	1
<b>2. FUNDAMENTAÇÃO TEÓRICA</b> .....	6
<b>2.1 A Linguística de Corpus</b> .....	6
2.1.1 Corpus de Aprendiz.....	10
2.1.2 Análise Multidimensional e Análise Multidimensional Lexical.....	13
<b>2.2 Os Estudos da Metáfora</b> .....	16
2.2.1 Linguística de Corpus e Identificação de Metáforas.....	22
2.2.2 Inteligência Artificial e Identificação de Metáforas.....	30
<b>3. METODOLOGIA</b> .....	34
<b>3.1 Corpus empregado</b> .....	34
<b>3.2 Análise Multidimensional Lexical</b> .....	37
<b>3.3 Identificação de Metáforas</b> .....	38
3.3.1 Anotação Manual.....	38
3.3.2 Anotação por Inteligência Artificial.....	42
<b>4. ANÁLISE E DISCUSSÃO DOS RESULTADOS</b> .....	48
<b>4.1 Resultados da Análise Multidimensional Lexical</b> .....	48
4.1.1 Interpretação das dimensões.....	49
4.1.1.1 <i>Dimensão 1 - Conhecimento abstrato, teórico e científico versus Dinâmicas familiares, resiliência pessoal e crescimento emocional</i> .....	51
4.1.1.2 <i>Dimensão 2 - Simbolismo literário e a interpretação na educação versus Vida diária, interações sociais, crescimento pessoal e bem-estar</i> .....	55
4.1.1.3 <i>Dimensão 3 - Relatos históricos e descobertas arqueológicas versus Identidade, transformação e resiliência na sociedade contemporânea</i> .....	60
4.1.1.4 <i>Dimensão 4 - Autonomia cultural, diversidade e respeito versus Vida cotidiana e experiências de aprendizagem</i> .....	64
4.1.1.5 <i>Dimensão 5 - Resiliência, determinação e triunfo socioeconômico versus Autonomia e expressão individual em ambientes estruturados e digitais</i> .....	69
4.1.1.6 <i>Dimensão 6 - Fenômenos naturais e recursos hídricos versus Crescimento pessoal e interação social</i> .....	74

4.1.1.7 Dimensão 7 - Iniciativas abrangentes de saúde e engajamento nas escolas versus Esforços pessoais e considerações éticas na conquista de objetivos.....	79
4.1.1.8 Dimensão 8 - Transições de vida e experiências de crescimento pessoal versus Aspectos éticos, sociais e estéticos da verdade e engano.....	83
4.1.2 ANOVAs.....	88
<b>4.2 Resultados da Identificação de Metáforas.....</b>	<b>89</b>
4.2.1 Resultados e Análise da Anotação Manual.....	91
4.2.2 Resultados e análise da anotação por IA.....	100
<b>5. CONCLUSÃO.....</b>	<b>105</b>
<b>6. REFERÊNCIAS .....</b>	<b>109</b>
<b>7. ANEXO .....</b>	<b>118</b>
<b>8. APÊNDICE .....</b>	<b>127</b>

## 1. INTRODUÇÃO

A Análise Multidimensional Lexical (Berber Sardinha, 2019; 2020; Delfino, 2021) pode auxiliar para a compreensão dos discursos subjacentes em diversos contextos linguísticos, ao possibilitar a identificação e análise de padrões discursivos. Estudos nessa área revelam não apenas os discursos presentes em diversos domínios, como o religioso (Veiga, 2021), mas também em torno de temas específicos, como xenofobia (Silva, em fase de elaboração) e sustentabilidade (Brogini, 2023). No entanto, até o momento, a aplicação da Análise Multidimensional Lexical em corpus de aprendizes, especialmente no contexto da educação bilíngue, parece ser uma lacuna na pesquisa linguística, o que levanta a questão sobre quais discursos são mobilizados pelos alunos bilíngues ao realizarem tarefas escritas. Assim, a presente pesquisa propõe realizar a Análise Multidimensional Lexical com um corpus de aprendiz, integrando-a com a análise de metáforas na produção escrita de alunos bilíngues, reconhecendo a importância da metáfora ao se estudar os discursos dos aprendizes, devido à centralidade da metáfora na língua (Low, 1988) e na compreensão de conceitos abstratos (Berber Sardinha, 2007).

Por muito tempo a metáfora foi relegada à posição de adorno da linguagem, sendo entendida como uma figura de linguagem (Berber Sardinha, 2007). Entretanto, com a virada cognitiva e mais precisamente com a publicação do trabalho seminal de Lakoff e Johnson, *Metaphors we live by* (1980), os estudos sobre a metáfora ganharam outro patamar. A metáfora deixou de ser entendida apenas um recurso estilístico e passou a ser estudada como um fenômeno cognitivo, o qual não é apenas manifestado na linguagem, uma vez que a metáfora governa a forma com que pensamos e agimos (Lakoff e Johnson, 1980/2003).

A metáfora não está apenas presente na linguagem literária, pois é constituinte da linguagem humana, sendo empregada com as mais diferentes finalidades. Segundo Berber Sardinha (2007), a metáfora é usada para: enriquecer a fala ou escrita, a fim de torná-las mais vívidas ou impactantes; trazer economia à linguagem, ou seja, transmitir um grande volume de informação com poucas palavras; criar uma aproximação com o ouvinte; ajudar no entendimento de conceitos abstratos (por exemplo, amor, morte e vida); estruturar o pensamento humano; auxiliar no entendimento de novo conhecimento, além de contribuir para que a linguagem gerada pela inteligência artificial possa simular de forma mais natural a linguagem humana (Wachowiak e Gromann, 2023).

Outro marco para os estudos da metáfora foi a virada discursiva. Cameron e Low (2011) salientam a importância de estudos da metáfora em uso, que têm como premissa não

apenas a cognição, mas também o fator social, levando em consideração “(...) padrões linguísticos, semânticos, pragmáticos e culturais encontrados nos estudos discursivos (...)”<sup>1</sup> (p. 1). Desta forma, o foco é como a metáfora linguística ocorre em textos de forma sistemática para a construção e entendimento de sentidos.

Ao analisar a metáfora em uso, é possível entender o pensamento humano, a forma como atribuímos significado ao mundo e às interações entre si, bem como nos comunicamos (Cameron, 2003), o que aproxima o estudo da metáfora da Linguística Aplicada, uma vez que a Linguística Aplicada busca "solução de problemas *humanos* que derivam dos vários usos da linguagem” (Celani, 1992, p. 21, ênfase da autora). Desse modo, o estudo da metáfora é relevante para a Linguística Aplicada e, conseqüentemente, para o programa do qual esta pesquisa está inserida, o Linguística Aplicada e Estudos da Linguagem (LAEL).

A relevância da metáfora para a ciência é evidenciada pela proliferação de pesquisas na área. Por exemplo, em 2006 foi fundada a RaAM - *Researching and Applying Metaphor* (Associação Internacional para a Pesquisa e Aplicação da Metáfora), uma comunidade internacional que tem como objetivo promover o debate e pesquisas sobre a metáfora. Outros exemplos podem ser encontrados no estudo bibliométrico realizado por Han, Peng e Chen (2022), no qual foram encontrados 4,458 artigos e revisões sobre a metáfora conceptual publicados entre 1980 e 2022. Os autores destacam a revista *Metaphor and Symbol*, criada em 1986, a qual era intitulada *Metaphor and Symbolic Activity* até 1996, como a revista mais proeminente da área.

Han, Peng e Chen (2022) também analisaram as palavras-chave mais frequentes das publicações, evidenciando a relação do estudo da metáfora com diferentes áreas do conhecimento. Os resultados demonstram que os cinco termos mais frequentes estão relacionados com as teorias de metáfora tradicional, conceptual e em uso; sendo elas: língua/linguagem<sup>2</sup>, compreensão, metáfora conceptual, discurso e linguagem figurada. Contudo, as pesquisas sobre metáfora podem abranger uma grande gama de áreas, como: educação (Cameron, 2003), matemática (Lakoff e Núñez, 2000), o mundo empresarial (Semino, 2002), saúde mental (Coll-Florit e Climent, 2020); Inteligência Artificial (IA) (Barden, 2008; Wachowiak e Gromann, 2023).

Como a metáfora permeia as nossas interações (Lakoff e Johnson, 1980/2003; Deignan, 2005; Berber Sardinha, 2007), ela influencia, portanto, tanto a nossa língua quanto

---

<sup>1</sup> No original, “(...) the linguistic, semantic, pragmatic and cultural patterns found in discourse studies (...)”.

<sup>2</sup> O artigo foi publicado em inglês e a palavra *language* pode ser traduzida para o português como “língua” ou “linguagem”.

as escolhas linguísticas. Por causa da grande presença de metáforas no uso linguístico, o estudo da metáfora a partir da perspectiva de alunos auxiliaria na compreensão de aquisição de uma língua estrangeira (Deignan, Gabrys e Solska, 1997). Assim, é postulada nos estudos de metáfora a premissa de que não é possível compreender a aquisição e o desenvolvimento de uma língua estrangeira sem levar em conta a utilização de metáfora pelos aprendizes. Embora seja reconhecida a centralidade da metáfora na aquisição de língua estrangeira, ainda há poucas pesquisas sobre o uso de metáforas por aprendizes, apresentando-se como uma das justificativas para o desenvolvimento do presente estudo.

A centralidade de metáforas em línguas faz com que elas mereçam ter um papel mais importante no ensino de línguas (Low, 1988), o que é evidenciado por meio de pesquisas que investigam metáfora, competência metafórica e as implicações na sala de aula (Low, 1988; Lazar, 1996; Deignan, Gabrys e Solska, 1997; Cameron e Low, 1999; Boers, 2000; Littlemore, 2001; Boers, 2004; Littlemore e Low, 2006; Hoang, 2014). Assim, pode-se perceber que o estudo de metáfora e da competência metafórica é relevante e pode auxiliar não só no entendimento da metáfora em si, mas também do funcionamento, aquisição e ensino de línguas. Contudo, muitas dessas pesquisas focam em como falantes de língua estrangeira interpretam, reconhecem e processam o uso de metáforas por meio de exercícios de interpretação de texto, tradução, identificação de metáforas e quando investigam a produção e uso de metáforas por esses falantes, normalmente é por meio de testes controlados, por exemplo, completar frases. Dessa forma, como salientado por Hoang (2014), observa-se uma lacuna em estudos que busquem investigar o uso espontâneo de metáforas em contextos naturais, e é esta lacuna que o presente estudo procura preencher.

Os estudos sobre produção de metáforas por aprendizes de inglês também podem auxiliar no entendimento da relação entre proficiência, tarefa e densidade metafórica, como os trabalhos de Littlemore et al. (2014), Golden (2021) e Nacey (2020; 2022). Littlemore et al. (2014) analisaram a produção escrita de aprendizes de diferentes níveis, do básico ao proficiente (do A2 ao C2 do Quadro Europeu Comum de Referência para as Línguas - *Common European Framework of Reference for Languages*) que prestaram exames de proficiência em língua inglesa aplicados pela Universidade de Cambridge. Os autores concluíram que a densidade metafórica aumenta na produção escrita dos aprendizes, ao ponto do uso de metáforas de alunos nível C2 (nível mais avançado) ser maior do que dos falantes nativos de inglês. Golden (2021) analisou como a tarefa influencia as escolhas lexicais dos aprendizes e, conseqüentemente, a densidade metafórica, revelando que tópicos abstratos

demandam o uso de metáfora. Porém, a autora sugere que a relação entre tópico e o uso de metáfora ainda precisa ser mais estudada no ensino de inglês como segunda língua.

Nacey (2022) realizou um estudo longitudinal sobre a produção metafórica de alunos bilíngues de inglês de 13 a 17 anos na Noruega, coletando quatro produções escritas de cinco alunos por um período de quatro anos, o que possibilitou o estudo de como a densidade metafórica varia de acordo com a proficiência dos alunos. Nacey (2022) afirma que há uma carência de pesquisas sobre como aprendizes de inglês como segunda língua desenvolvem a produção metafórica, sendo necessário mais estudos com um número maior de participantes. Contudo, fazer a análise manual de uma grande quantidade de textos é um desafio, sendo importante para a área o desenvolvimento de formas automáticas de identificação de metáforas linguísticas (Berber Sardinha, 2007), por exemplo, por meio do uso da Inteligência Artificial.

Uma possibilidade para se investigar o uso de metáfora por falantes bilíngues seria por meio de pesquisa em corpora de aprendizes. Segundo Granger, Gilquin e Meunier (2015), a pesquisa em corpora de aprendizes seria um campo fértil por estabelecer pontes entre diversas áreas como a Linguística de Corpus, aquisição de segunda língua, ensino de línguas e processamento de linguagem natural. Porém, o estudo de metáfora com base em corpus de aprendizes ainda é pouco explorado na literatura, existindo então a necessidade de mais estudos neste campo.

Os estudos da metáfora são normalmente baseados na análise manual do corpus, como Cameron (2003) e Nacey (2022), apesar dos esforços de automatização do processo (Berber Sardinha, 2012). Recentemente, com a disponibilização de modelos de Inteligência Artificial como ChatGPT, Gemini e Llama, tornou-se possível verificar se essa tecnologia tem a capacidade de identificar metáforas em textos de forma mais precisa, como em Wachowiak e Gromann (2023). No entanto, o uso de modelos de Inteligência Artificial para a identificação de metáforas em corpora de aprendizes parece não ter sido explorado.

Tendo em vista as lacunas descritas acima, a presente pesquisa teve como objetivo geral identificar, por meio da Análise Multidimensional Lexical, os principais discursos subjacentes às produções textuais de alunos bilíngues, além de se e como esses alunos fazem o uso de linguagem metafórica para desempenhar os discursos em produção textual em inglês. Mais especificamente, o projeto visou verificar a incidência de metáforas linguísticas na produção dos alunos, se são mais frequentes em tarefas que demandam maior abstração e se a Inteligência Artificial pode ser usada na identificação de metáforas. Referente aos objetivos específicos, o estudo buscou:

- 1) Identificar, por meio da Análise Multidimensional Lexical, os principais discursos em jogo nos textos, que mostram como os estudantes enquadraram os tópicos nas tarefas.
- 2) Detectar o uso de metáfora nas dimensões provenientes da Análise Multidimensional Lexical, ou seja, como os estudantes utilizaram metáforas para construir noções abstratas associadas às dimensões.
- 3) Testar a possibilidade de identificação de metáforas usando a Inteligência Artificial, mais especificamente, até que ponto a Inteligência Artificial pode reproduzir o julgamento humano sobre a metaforicidade, usando um subconjunto dos dados.

Para atingir estes objetivos, o presente estudo procurou responder às seguintes perguntas de pesquisa:

- 1) Quais os discursos revelados nas dimensões de variação lexical presentes nas tarefas desenvolvidas pelos alunos?
- 2) Qual a incidência de uso de metáfora para expressão de conceitos abstratos na produção escrita de alunos bilíngues?
- 3) Qual a acuidade da Inteligência Artificial na detecção de metáforas nos textos dos alunos?

A presente dissertação está dividida em oito capítulos principais. A fundamentação teórica, subsequente à introdução, está subdividida em duas partes: a Linguística de Corpus e os Estudos da Metáfora. A metodologia é o terceiro tópico, que apresentará o corpus da pesquisa, a Análise Multidimensional Lexical, a identificação manual das metáforas linguísticas e a identificação por Inteligência Artificial. A quarta parte apresenta a análise e discussão dos resultados, seguida da conclusão na quinta parte, das referências na sexta, do anexo na sétima e do apêndice na oitava.

## 2. FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, é apresentada a fundamentação teórica que sustenta a pesquisa. Inicialmente, na Seção 2.1, é delineado um contexto histórico abrangente da Linguística de Corpus, seguido pelos princípios de duas abordagens metodológicas pertinentes a esta pesquisa: a Análise Multidimensional (Biber, 1988; Berber Sardinha, 2004) e a Análise Multidimensional Lexical (Berber Sardinha, 2019; 2020; Delfino, 2021).

Na Seção 2.2, realiza-se uma revisão dos Estudos da Metáfora, dos métodos de identificação de metáforas linguísticas, bem como das contribuições de procedimentos que empregam Inteligência Artificial para a identificação de metáforas. Essa revisão busca proporcionar uma compreensão do contexto teórico e metodológico, estabelecendo as bases necessárias para a condução da pesquisa.

### 2.1 A LINGUÍSTICA DE CORPUS

Os estudos em Linguística de Corpus (LC), como conhecidos hoje, começaram a se popularizar no final da década de 1950 com o advento do computador, por terem como uma de suas características o uso de computadores para a análise de textos, mais especificamente do corpus (McEnery e Hardie, 2013). Porém, apenas nos anos 1980, com o maior acesso à microcomputadores pessoais, que as pesquisas baseadas em corpus começaram a ser disseminadas. Assim, o desenvolvimento da LC está relacionado ao desenvolvimento tecnológico, pois a LC depende do computador tanto para o armazenamento quanto para a análise de corpora (plural de corpus) por meio de softwares (Berber Sardinha, 2004).

A LC tem como objetivo investigar a língua em uso por meio de análise de evidências empíricas e não apenas por introspecção do pesquisador. Desta forma, a LC tem como foco a descrição e a análise quantitativa e qualitativa, contrastando-se aos estudos Chomskianos (Granger, 1998). No final da década de 1950, a Linguística Gerativista de Chomsky, baseada na teoria racionalista da linguagem, tinha como objetivo a competência e universais linguísticos, por meio da introspecção do linguista, geralmente um falante nativo, para identificar as estruturas da linguagem. Contudo, análises baseadas na intuição do pesquisador não são confiáveis, pois o ser humano tem a tendência de “(...) perceber ocorrências não usuais mais do que as ocorrências típicas” (Biber, Conrad e Reppen, 1998, p.3). Outro contraponto é que a LC é baseada em uma visão probabilística da linguagem, enquanto a

Linguística Gerativista entende a linguagem como possibilidade, ou seja, o que seria possível ou não (agramatical).

A visão probabilística da linguagem, disseminada por Michael Halliday, parte da premissa que traços linguísticos não ocorrem ao acaso ou a depender da escolha do falante, pois traços e estruturas ocorrem com diferentes frequências a depender do registro (Berber Sardinha, 2004). No entanto, as diferenças de frequências não ocorrem de forma aleatória, isto é, há um padrão na variação; estudos como de Biber (1988) e Berber Sardinha, Kauffmann e Acunzo (2014) demonstraram o padrão de variação da linguagem do inglês e português, respectivamente. Ainda, Halliday e Matthiessen (2004) enfatizam a importância do uso de um corpus para a teorização da linguagem dada a possibilidade de se investigar produções autênticas, tanto do modo escrito quanto falado de forma quantitativa.

É importante ressaltar que a LC não seria uma subdisciplina da Linguística, como a Sociolinguística ou a Fonologia, como explicado por McEnery e Hardie (2011), pois a LC não está diretamente relacionada ao estudo de aspectos específicos da língua, mas seria “(...) uma área que foca em um conjunto de procedimentos ou métodos para se estudar línguas”<sup>3</sup> (p.1). Tais procedimentos e métodos podem variar, mas segundo McEnery e Hardie (2011) algumas generalizações podem ser feitas, tais como: o tamanho de um corpus impossibilita que este seja analisado manualmente, sendo necessário assim a utilização de computadores; o tipo de corpus utilizado precisa estar alinhado com as perguntas de pesquisa; e um corpus precisa ter uma variedade de tipos de textos para se fazer afirmações sobre uma língua ou línguas.

A LC pode ser considerada um método linguístico ou teoria como discutido por Shepherd (2009). O entendimento e aplicação da LC como um método está relacionado a estudos que são *corpus-based* (baseados em corpus), ou seja, o corpus serve como base para a exploração, validação e/ou refutação de hipóteses preestabelecidas acerca da língua ou fenômenos linguísticos (McEnery e Hardie, 2011; Berber Sardinha, Resende e Shimazumi, 2023). Já os estudos que são *corpus-driven* (guiados por corpus) não compreendem a LC como um método, mas reiteram a ideia de que o próprio corpus apresenta as possíveis hipóteses sobre línguas (McEnery e Hardie, 2011). De acordo com Biber, Conrad e Reppen (1998) um estudo *corpus-based* é empírico, pois baseia-se na observação da linguagem em textos naturais, por meio de computadores que podem fazer análises automáticas e interativas, e depende tanto de técnicas quantitativas quanto qualitativas. Para o presente estudo, a LC pode ser melhor definida como “uma abordagem teórico-metodológica de investigação

---

<sup>3</sup> No original, “(...) an area which focuses upon a set of procedures, or methods, for studying language”.

linguística sobre o uso da língua que permite fazer descrições e generalizações acerca de seu uso de forma empírica” (Berber Sardinha, Resende e Shimazumi, 2023, p. 104).

A importância de se fazer uma pesquisa qualitativa e quantitativa é discutida por Fillmore (1992). Segundo o autor, ao utilizar um corpus, um pesquisador consegue fazer um levantamento maior de exemplos, o que não seria possível apenas por meio da introspecção. Porém, é necessário a análise qualitativa de cada um dos exemplos para se chegar à generalizações.

O uso de corpus não apenas possibilita a análise de frequência ou a ocorrência de padrões, mas também a identificação e análise de associação de padrões complexos. Biber, Conrad e Reppen (1998) explicam que o corpus viabiliza a análise do uso de “(...) formas sistemáticas em que aspectos linguísticos são usados em associação com outros aspectos linguísticos e não linguísticos<sup>4</sup>” (p.5). Por exemplo, pode-se investigar como padrões linguísticos coocorrem, como construções gramaticais e colocações lexicais, ou não linguísticas, a distribuição de um padrão linguístico em diferentes registros, dialetos ou tempos históricos. Contudo, como a linguagem é uma questão de probabilidade e não possibilidade, “(...) um corpus não pode nos dizer o que não é possível<sup>5</sup>” (Fillmore, 1992, p.58), assim, é necessário que o corpus esteja alinhado às perguntas que o pesquisador pretende responder e para isso, o planejamento e desenho do corpus são essenciais.

Um corpus pode ser definido como um conjunto de textos, podendo estes serem tanto escritos quanto falados, que podem ser armazenados eletronicamente (O’Keeffe, McCarthy e Carter, 2007). Contudo, esta definição não seria suficiente para que um corpus esteja adequado à pergunta de pesquisa. Um corpus precisa ser eletronicamente armazenado, mas é necessário que haja critérios para a escolha dos textos.

Berber Sardinha (2004) salienta que para um conjunto de textos eletrônicos seja considerado um corpus, alguns pré-requisitos devem ser atendidos. Primeiramente, os textos (a amostra de linguagem) precisam ser autênticos, ou seja, textos produzidos por seres humanos e não por máquinas. Além da naturalidade e autenticidade, a escolha dos textos precisa ser criteriosa, respeitando alguns parâmetros estabelecidos antes de sua coleta, como quantidade de cada tipo de texto, o registro e a população. Por fim, um corpus precisa ser representativo.

---

<sup>4</sup> No original, “(...) the systematic ways in which linguistic features are used in association with other linguistic and non-linguistic features”.

<sup>5</sup> No original, “(...) a corpus cannot tell us what is not possible”.

A representatividade de um corpus está relacionada à sua extensão. Não se pode representar a linguagem de uma população em sua totalidade, mas quanto mais extenso for um corpus, maior será a probabilidade de palavras de baixa frequência ou mesmo sentidos menos frequentes de palavras de alta frequência possam ser encontrados (Berber Sardinha, 2004). No entanto, para evitar vieses, deve-se atentar para os diferentes registros da linguagem, além do tamanho dos textos (quantidade de palavras) e a quantidade de textos de cada subcategoria (Biber, Conrad e Reppen, 1998). Ademais, de acordo com Berber Sardinha (2004) “para que qualquer amostra seja representativa, é necessário conhecer a população da qual ela provém” (p.23).

População e registro estão relacionados, pois ao se definir a representatividade de um corpus, deve-se levar em consideração a população (quem a pesquisa pretende representar) e em qual registro, ou seja, em qual situação (Resende e Maverick, 2016). Biber (1993a) explica que o registro depende de critérios externos ao corpus, pois “(...) os registros estão baseados em diferentes situações, propósitos e funções que um texto tem uma comunidade discursiva (...)”<sup>6</sup> (p. 245). O autor contrasta registro com tipo textual, que é baseado em critérios internos do corpus, isto é, pelas características linguísticas dos textos. Assim, não haveria como estabelecer antes da coleta e a análise do corpus quais seriam os tipos textuais.

Contudo, como pontuado por Berber Sardinha (2004), diante das dificuldades para se demonstrar o que seria uma amostra representativa, ficaria a cargo do usuário do corpus ser cauteloso ao fazer generalizações e tentar demonstrar a sua representatividade. Da mesma forma, para McEnery and Hardie (2011), a representatividade do corpus seria um ideal dificilmente atingido pelos pesquisadores, sendo mais apropriado pensar em graus de representatividade.

Um corpus pode ser definido de acordo com o seu conteúdo e objetivo. Berber Sardinha (2004) apresenta alguns critérios dos tipos principais, como: modo, tempo, seleção, conteúdo, autoria, disposição interna e finalidade. Cada um dos critérios é apresentado abaixo:

- a) **Modo de comunicação:** um corpus pode ser escrito ou falado. O corpus escrito contém textos escritos, como de livros, revistas, blogs e cartas, podendo ser estes impressos ou não, mas digitalizados em arquivos que podem ser lidos por ferramentas computacionais. Já o corpus falado é formado pela transcrição de textos orais, como conversas, filmes e vídeos, essas transcrições também precisam ser digitalizadas.

---

<sup>6</sup> No original, “ (...) registers are based on the different situations, purposes, and functions of text in a speech community (...)”.

- b) **Tempo:** o corpus é sincrônico quando abarca um período de tempo; diacrônico quando abarca mais de um período; contemporâneo, ou seja, representa o tempo corrente, ou histórico, representando um tempo passado.
- c) **Seleção:** um corpus pode ser de amostragem, quando é planejado para conter uma mostra finita de textos ou variedade textual, sendo também estático, uma vez que não há espaço para a inserção de novos textos. Já um corpus monitor é dinâmico e orgânico, pois é constantemente expandido pela inserção de textos, sendo um reflexo do estado de uma língua. Além disso, um corpus pode ser equilibrado ao conter uma quantidade semelhante de textos para cada registro que o compõe.
- d) **Conteúdo:** um corpus pode ter textos específicos de um registro ou área, sendo chamado de corpus especializado. Ele pode conter textos de variedades sociolinguísticas específicas, sendo considerado um corpus regional ou dialetal. Um corpus pode, também, ter textos de idiomas variados, recebendo a nomenclatura de corpus multilíngue.
- e) **Autoria:** o corpus pode ser de língua nativa, contendo textos produzidos por falantes nativos ou de aprendiz, contendo textos de falantes não nativos. A Seção 2.1.1 discutirá o corpus de aprendiz com mais profundidade.
- f) **Disposição interna:** os corpora multilíngues ou bilíngues podem apresentar os textos de forma paralela (um texto ao lado do outro) ou alinhada (as linhas do original seguidas da tradução).
- g) **Finalidade:** um corpus pode ser de estudo; de referência, isto é, o que será usado para contrastar com o corpus de estudo; ou de teste/treinamento o qual objetiva o desenvolvimento de aplicações e ferramentas de análise linguísticas.

### 2.1.1. Corpus de Aprendiz

As pesquisas com corpora de aprendizes podem contribuir para o ensino de línguas, uma vez que essas pesquisas “(...) têm sido usadas para informar a lexicografia, o design de currículos, o design de materiais, a aprendizagem de idiomas auxiliada por computador e abordagens pedagógicas como a aprendizagem baseada em dados<sup>7</sup>” (Meunier, 2021, p. 28). Assim, as pesquisas em corpora de aprendizes têm tradicionalmente focado em questões gramaticais (como em Rankin, 2015; Larson et al., 2023) ou lexicais (como em Cobb e Horst,

---

<sup>7</sup> No original, “(...) a have been used to inform lexicography, syllabus design, materials design, computer-aided language learning and pedagogical approaches such as data-driven learning”.

2015; Szudarski, 2023), na sua maioria. Contudo, as pesquisas focadas nos textos produzidos por aprendizes não começaram com os corpora de aprendizes, mas sim com os estudos de Análise de Erro. Assim, pode-se considerar que a Análise de Erro um predecessor dos corpora de aprendizes (Granger, 1998). As pesquisas em Análise de Erro empregavam textos de aproximadamente 2,000 palavras para a observação de erros de alunos e possíveis explicações. Porém, os estudos em Análise de Erro tinham suas limitações, as quais estavam relacionadas com a coleta e armazenamento dos textos, pois estes não eram armazenados eletronicamente e muitas vezes eram selecionadas palavras ou orações que continham erros, o que fazia com que os textos não capturassem a linguagem em uso dos aprendizes (Berber Sardinha, 2004).

Um corpus de aprendiz contém textos produzidos por aprendizes de uma determinada língua, ou seja, por falantes não nativos. O interesse em se usar um corpus de aprendiz iniciou-se na pesquisa sobre o ensino de inglês como língua estrangeira ou adicional (*English as a foreign language - EFL*) e por editoras nos anos 1990. Granger (1998) destaca três corpora de aprendizes como pioneiros o *International Corpus of Learner English (ICLE)* da Universidade Católica de Lovaina, o qual conta com textos acadêmicos escritos em inglês por alunos universitários de diversas nacionalidades; o *Longman Learners' Corpus (LLC)* composto por textos de aprendizes de inglês, e o *Hong Kong University of Science and Technology (HKUST) Learner Corpus* formado por textos em inglês de alunos chineses.

Berber Sardinha (2004) ressalta a importância do ICLE, pois este viabilizou futuros estudos utilizando corpora de aprendizes, uma vez que ainda há uma predileção em se usar corpora de falantes nativos por questões práticas e teóricas. Do ponto de vista prático, o custo para a compilação, computadores e softwares, além de pessoas especializadas dificulta a utilização de corpora de aprendizes no contexto escolar. Já do ponto teórico, está o fato de que as pesquisas em aquisição de língua estrangeira tendem a focar no processo de aprendizagem, o que pressupõe a investigação de aspectos cognitivos. Porém, ao se usar um corpus de aprendiz, o foco está na “produção dos alunos, ou seja, de como o aprendizado se concretiza” (Berber Sardinha, 2004, p. 268-269).

Como um corpus de aprendiz é composto por produção de alunos, é posta em discussão a autenticidade da linguagem, um dos pressupostos da LC. Gilquin (2020) explica que a produção de alunos não atende a objetivos puramente comunicativos, uma vez que as tarefas são pedagógicas, além de apresentarem uma menor variedade de gêneros, pois os textos estão condicionados aos que normalmente ocorrem no contexto educacional, como

escrita de textos narrativos, argumentativos e acadêmicos. Portanto, a autora argumenta que o conceito de autenticidade deva ser adaptado para o contexto educacional estudado.

Outro aspecto relevante é que não apenas o contexto educacional pode ser diverso, mas os próprios estudantes. Segundo Granger (1998), o corpus de aprendiz é heterogêneo, sendo essencial ter critérios claros para a compilação do corpus. Assim, a autora destaca que as variáveis relativas à produção dos textos (como: modo, gênero, tópico, grau de tecnicidade e contexto de produção), além de características dos estudantes (como: idade, gênero, língua materna, região, línguas estrangeiras, proficiência, contexto de aprendizagem e experiência prática) devam ser levadas em consideração na análise dos dados.

Ainda sobre a composição, Gilquin (2020) descreve que um corpus de aprendiz pode ser composto por textos produzidos pelos mesmos alunos em períodos de tempo distintos, o que caracteriza um corpus de aprendiz longitudinal. De acordo com a autora, esse tipo de corpus de aprendiz ajudaria no entendimento de como as habilidades linguísticas dos alunos mudam com o tempo, apesar de ser um corpus menos denso.

Um exemplo de corpus de aprendiz longitudinal foi o empregado por Nacey (2022). A pesquisadora compilou a produção escrita de cinco alunos, de 13 a 17 anos, do ensino secundário na Noruega, sendo dois textos escritos em inglês por cada aluno durante quatro anos acadêmicos consecutivos. O objetivo da pesquisa era explorar o desenvolvimento da produção metafórica e, para a identificação, a pesquisadora usou o *Metaphor Identification Procedure Vrije Universiteit* (MIPUV) (Steen et al., 2010). Nacey concluiu que há um aumento da densidade de metáforas em textos conforme o nível de proficiência dos alunos. A autora destaca que os estudos anteriores eram ‘pseudo-longitudinal’, isto é, o corpus era composto de textos de alunos de diferentes anos escolares.

Gilquin (2020) não descarta a validade de corpus de aprendiz pseudo-longitudinal ou, como denominado pela autora, quase-longitudinal, pois estes possibilitam a observação de desenvolvimento linguístico por meio da comparação dos sub-corpora. Além disso, uma vantagem de se usar um corpus de aprendiz quase-longitudinal é a conveniência do processo de coleta. Assim, a presente pesquisa utilizou um corpus de aprendiz quase-longitudinal, uma vez que os textos compilados foram escritos por alunos do Sexto Ano do Ensino Fundamental II ao Primeiro Ano do Ensino Médio em 2021.

### 2.1.2 Análise Multidimensional e Análise Multidimensional Lexical

Partindo do pressuposto de que a linguagem é um sistema complexo, mas não caótico, pois há um padrão na variação (Berber Sardinha e Veirano Pinto, 2019), é necessário uma metodologia de pesquisa que consiga capturar a variação subjacente da linguagem. Assim, a Análise Multidimensional (AMD) surge como uma abordagem usada para estudar o padrão de variação textual em um corpus por meio da análise fatorial, que por sua vez utiliza procedimentos estatísticos para agrupar dados, no caso da linguagem, traços linguísticos, que coocorrem. Biber (1988) propôs essa metodologia baseada em estatística multivariada para estudar a variação de traços léxico-gramaticais da língua inglesa falada e escrita.

Os estudos anteriores a Biber (1988) levavam em conta a co-ocorrência de padrões, porém, esses utilizavam poucos parâmetros, o que acarretava em distinções incompletas de textos (Berber Sardinha, 2004). Desse modo, esses estudos apresentavam uma noção de variação com foco apenas em um traço linguístico ou uma visão dicotômica das variações funcionais (formal versus informal, por exemplo). Além disso, de acordo com Berber Sardinha (2004), as pesquisas não forneciam uma descrição objetiva dos traços, pois a descrição das co-ocorrências era realizada por meios intuitivos.

Biber (1988) destaca que as principais características da AMD são: 1) a visão multidimensional, pois uma única dimensão não seria capaz de explicar a variação da linguagem; 2) as dimensões estão em um continuum, isto é, em uma escala contínua, e não em uma dicotomia; 3) a análise empírica e estatística das frequências de variação de cada texto. Cada uma das características será apresentada a seguir.

A AMD pressupõe “(...) a existência de múltiplas dimensões no espaço de variação intertextual” (Berber Sardinha, 2010, p. 114). Uma dimensão pode ser definida como um conjunto de traços linguísticos (por exemplo, substantivos, verbos, adjetivos, pronomes pessoais, tipos de sentenças e construções) que co-ocorrem de forma consistente e complementar, porque compartilham uma mesma função (Biber, 1988; Biber e Conrad, 2009). É importante destacar que esses padrões estão subjacentes, sendo capturados pela análise fatorial e interpretadas pelo pesquisador. Conseqüentemente, não podem ser pré-estabelecidos antes da análise.

Biber (1988) identificou cinco dimensões da língua inglesa: Dimensão 1 - produção com interação versus produção informacional; Dimensão 2 - preocupações narrativas versus não-narrativas; Dimensão 3 - referências explícitas versus referências dependente do contexto; Dimensão 4 - expressão explícita de persuasão; Dimensão 5 - informação abstrata

versus não-abstrata. Por mais que as dimensões sejam apresentadas em dois polos, um positivo e outro negativo, elas representam uma escala em que todos os textos de um corpus estão distribuídos. Além disso, como explicado por Delfino (2021), os polos se complementam, não havendo um polo mais importante do que o outro.

Quanto à análise empírica, Berber Sardinha, Kauffmann e Acunzo (2014) discutem que uma das vantagens de se aplicar a AMD é a possibilidade de pesquisar a língua em uso com foco no texto e na variação. Ao contrário das pesquisas que focam em linhas de concordâncias para a análise de padrões lexicais e/ou gramaticais, os estudos de AMD levam em consideração o texto. Assim, a frequência de cada traço léxico-gramatical é calculada e usada para a análise quantitativa e qualitativa.

Para a análise quantitativa do corpus, a AMD utiliza a análise estatística multivariada. Segundo, Cantos-Gomez (2019) estudos linguísticos usam a estatística multivariada, pois com ela é possível reduzir a quantidade de dados, sem que informações importantes sejam perdidas. Ademais, a análise multivariada remove vieses. Para encontrar as variáveis latentes dos textos, a AMD utiliza a análise fatorial; como explica o autor, com esta análise é possível “(...) encontrar a presença de variáveis latentes (não observadas) que explicam o padrão de variáveis observadas. Supõe-se que os fatores latentes realmente existam, mas não podem ser medidos diretamente<sup>8</sup>” (Cantos-Gomez, 2019, p. 106). A partir da análise das variáveis de cada fator, o pesquisador retoma os textos em que elas ocorrem para fazer a interpretação de suas funções, e assim fazer a análise qualitativa e nomear as dimensões.

Berber Sardinha (2004) explica que a AMD é dividida em 3 etapas: 1) coleta e pré-processamento do corpus; 2) análise fatorial; 3) interpretação das dimensões. Delfino (2021) descreve cada uma das etapas:

#### 1. Pré-processamento do corpus:

- realiza-se a etiquetagem dos textos (POS tagging<sup>9</sup>);
- os textos são normalizados por 1000, para que textos de tamanhos diferentes consigam ser comparados, assim um texto com muitas palavras não se sobressaiará em relação a um texto com menos palavras;
- realiza-se a seleção das variáveis a partir do valor das comunalidades das variáveis;

---

<sup>8</sup> No original, “(...) to find the presence of latent (unobserved) variables, which explains the pattern of observed variables. Latent factors are assumed actually to exist, but cannot be directly measured.”

<sup>9</sup> Part-of-speech tagging

## 2. Análise fatorial

- roda-se a primeira análise fatorial e o número de fatores é determinado;
- roda-se a segunda análise fatorial (rotacionada) e retiram-se as variáveis cujo peso é muito baixo;

## 3. Interpretação das dimensões:

- calculam-se os escores dos textos e determina-se os escores dos fatores;
- As dimensões são nomeadas com base na análise qualitativa dos textos, juntamente com os resultados dos escores de cada variável em cada fator.

Berber Sardinha (2004) explica que uma pesquisa multidimensional utiliza análises macroscópicas e microscópicas. As análises microscópicas envolvem a descrição minuciosa das funções comunicativas das características linguísticas (Biber, 1988), sendo que os textos e os registros são considerados de forma individual (Berber Sardinha, 2004). Por um outro lado, as análises macroscópicas permitem a definição das dimensões de variação.

De acordo com Biber (1988), a análise macroscópica e microscópica são complementares e dependentes. A análise microscópica complementa a análise macroscópica ao identificar as características linguísticas que são potencialmente importantes e as distinções dos gêneros a serem incluídas na análise macroscópica, além de fornecer a análise das funções das características linguísticas. Contudo, a análise microscópica apenas permite a análise de algumas características em textos individuais. Em comparação, a análise macroscópica possibilita a identificação das dimensões subjacentes dos textos, proporcionando a análise da variação entre textos e auxiliando na discussão sobre as diferenças e semelhanças entre textos e gêneros. Porém, a análise macroscópica é restrita ao perder de vista parâmetros de variação textual menores.

Uma variação da AMD é a AMD Lexical que tem como objetivo identificar as dimensões de variação discursiva, “ou seja, como o discurso varia sistematicamente de acordo com o contexto<sup>10</sup>” (Berber Sardinha, 2020, p. 298). Sendo assim, a AMD Lexical utiliza como variáveis apenas características linguísticas lexicais, por exemplo palavras de conteúdo, como substantivos, adjetivos, verbos e advérbios (Berber Sardinha, 2019; Berber Sardinha, 2020; Delfino, 2021). A AMD Lexical foi utilizada com diferentes objetivos, como por exemplo, para a descrição de representações de identidade nacional (Berber Sardinha, 2019), a análise

---

<sup>10</sup> No original, “that is, how discourse varies systematically according to context”.

do discurso acadêmico (Berber Sardinha, 2020) e a análise acerca do conceito de sustentabilidade nas redes sociais (Brochini, 2023).

A AMD Lexical não pretende substituir a AMD, pois cada uma responde perguntas de pesquisa distintas. A AMD tem como foco entender como os falantes utilizam macro funções comunicativas (dimensões funcionais), enquanto a AMD Lexical tem como foco entender como os discursos ocorrem na língua em uso (dimensões lexicais). Assim, a AMD e a AMD Lexical diferem de duas formas: os dados usados para a análise fatorial e a interpretação dos fatores (Berber Sardinha, 2018).

Em relação aos dados usados, a AMD usa variáveis léxico-gramaticais (por exemplo, tipos de verbos, substantivos e adjetivos), enquanto a AMD Lexical utiliza lemas (forma canônica das palavras), n-gramas ou colocações (Berber Sardinha, 2019). Deste modo, cada análise fará o processamento com etiquetadores distintos, por exemplo, a AMD Lexical poderá utilizar o TreeTagger, um etiquetador morfológico e lematizador, e a AMD utilizará etiquetadores complexos e analisadores sintáticos, como o Biber Tagger. Quanto à interpretação dos fatores, na AMD o objetivo é descobrir os parâmetros funcionais que resultam das correlações entre os traços linguísticos presentes nos textos. Já a AMD Lexical tem a finalidade de mostrar os discursos indexados pela correlação entre unidades lexicais, baseando-se em conceitos como discurso e ideologia (Berber Sardinha e Fitzsimmons-Doolan, 2024).

A utilização da AMD Lexical foi pertinente neste estudo, pois um de seus objetivos de pesquisa foi identificar os principais discursos na produção textual de alunos bilíngues. Ademais, ao usar palavras de conteúdo como dados para a análise fatorial, a AMD Lexical permite a análise de substantivos de cada dimensão, o que auxiliou na seleção de candidatos à metáfora. Uma das possibilidades de metodologia de identificação de metáforas parte da análise do corpus para a seleção dos candidatos, assim os substantivos abstratos das dimensões foram usados como candidatos à metáfora e analisados manualmente e por IA, sendo esses os outros dois objetivos da pesquisa. A metodologia empregada para a AMD Lexical e a identificação da metáfora são apresentados na Seção 3.3.

## 2.2. ESTUDOS DA METÁFORA

Em relação aos Estudos de Metáfora, metáforas são normalmente definidas como recursos que falantes usam quando empregam palavras e expressões com sentido figurado, ou seja, seu sentido não-literal (Berber Sardinha, 2007). Vale destacar que o termo metáfora tem

origem do grego “metapherein”, sendo composta por “meta” que significa “sobre” ou “além” e “pherein” que significa “levar” ou “transportar”. Assim, o significado de metáfora implica o movimento ou a transferência de sentido (Darin, 2020). Contudo, Berber Sardinha, em seu livro *Metáfora* (2007), apresenta quatro grandes visões sobre a metáfora: a) a visão tradicional; b) a metáfora conceptual; c) a metáfora sistemática ou discursiva, e d) a metáfora gramatical<sup>11</sup>. Para o presente trabalho, é relevante a discussão sobre a visão tradicional, conceptual e discursiva.

A visão tradicional de metáfora teve o seu início com filósofos gregos no século IV a.C. (Kittay, 1987; Berber Sardinha, 2007). De acordo com Berber Sardinha (2007), Aristóteles discute como a metáfora é empregada tanto na poesia quanto na retórica; em *Poética*, a metáfora é definida pela transferência de sentido de uma palavra para designar outra e em *Retórica*, o filósofo a define como uma comparação direta. Como explica Berber Sardinha (2007), a visão tradicional de metáfora a define como um fenômeno individual, um recurso retórico usado com finalidades específicas, por exemplo, como em textos literários.

Na visão conceptual, a metáfora é entendida como um fenômeno cognitivo, ou seja, ela aconteceria na mente dos falantes antes de ser expressa na fala ou na escrita (Berber Sardinha, 2007). Lakoff e Johnson (1980/2003) argumentam que a metáfora não é apenas manifestada na linguagem, pois ela está presente no pensamento e ação humana, uma vez que o sistema conceitual dos falantes é metafórico. Assim, os autores afirmam que

...a metáfora não é apenas uma questão de linguagem, ou seja, apenas palavras. Argumentamos que, pelo contrário, os *processos de pensamento* humanos são em grande parte metafóricos. Isso é o que queremos dizer quando afirmamos que o sistema conceitual humano é metafóricamente estruturado e definido. Metáforas como expressões linguísticas são possíveis precisamente porque existem metáforas no sistema conceitual de uma pessoa<sup>12</sup>. (Lakoff e Johnson (1980/2003, p. 6) (ênfase dos autores)

Na teoria da Metáfora Conceptual, a metáfora não é compreendida como um recurso usado para adornar a linguagem, estando periférica à língua e ao pensamento (Deignan, 2005), uma vez que ela é central para o pensamento e a linguagem. Segundo Deignan (2005), alguns princípios derivam da visão conceptual da metáfora, sendo eles: “metáforas estruturam o pensamento; metáforas estruturam o conhecimento; a metáfora é central para a linguagem

---

<sup>11</sup> Na visão da metáfora gramatical, o foco estaria no sistema linguístico e como um recurso linguístico pode ser usado com uma função diferente daquela que ocupa no sistema (Berber Sardinha, 2007).

<sup>12</sup> No original, “(...) metaphor is not just a matter of language, that is, of mere words. We shall argue that, on the contrary, human *thought processes* are largely metaphorical. This is what we mean when we say that the human conceptual system is metaphorically structured and defined. Metaphors as linguistic expressions are possible precisely because there are metaphors in a person’s conceptual system.”

abstrata; a metáfora tem base na experiência física; a metáfora é ideológica<sup>13</sup>” (Deignan, 2005, p. 13). Os princípios são discutidos a seguir.

A metáfora conceptual é proveniente da experiência humana, o que pressupõe uma coerência entre as metáforas e os valores culturais e ideologias de uma dada sociedade. Por exemplo, no mundo capitalista ocidental, o tempo é conceitualizado como dinheiro; assim, a metáfora conceptual TEMPO É DINHEIRO<sup>14</sup> licencia a produção de falas como “Isto é uma **perda de tempo**<sup>15</sup>” ou “Precisamos **ganhar tempo**”. Além disso, a metáfora conceptual TEMPO É DINHEIRO implica outras duas metáforas TEMPO É UM RECURSO LIMITADO e TEMPO É UMA COMMODITY VALIOSA (Lakoff e Jonhson, 1980/2003).

Berber Sardinha (2007) explica que a metáfora conceptual é uma forma de “conceitualizar um domínio de experiência em termos de outro” (p.30), sendo domínio uma “área do conhecimento ou experiência humana” (p. 31). Em uma metáfora conceptual há dois domínios: o domínio-fonte e o domínio-alvo. O domínio-fonte é normalmente mais concreto, físico e derivado da experiência humana, sendo a partir do qual o domínio-alvo é conceitualizado, e o domínio-alvo é mais abstrato e conceitualizado por meio do uso da metáfora (Berber Sardinha, 2007). Em TEMPO É DINHEIRO, “tempo” é o domínio-alvo e “dinheiro” é o domínio-fonte. No entanto, Lakoff e Jonhson (1980/2003) ressaltam que um domínio é parcialmente estruturado por uma metáfora, pois se fosse total, um domínio seria exatamente o outro. No exemplo acima, algumas das características de dinheiro podem ser transferidas para o tempo, mas não todas - pode-se ganhar e perder tempo, não sendo possível emprestar o tempo, já dinheiro pode ser emprestado.

Lakoff (1993) ressalta que uma metáfora pode ser compreendida por meio do mapeamento entre o domínio-fonte e o domínio-alvo, sendo um mapeamento das relações que ocorrem entre os domínios. Segundo o autor, Lakoff e Johnson (1980/2003) utilizaram uma estratégia mnemônica para facilitar a nomeação de mapeamentos: DOMÍNIO-ALVO É DOMÍNIO-FONTE ou DOMÍNIO-ALVO COMO DOMÍNIO-FONTE. Lakoff (1993) enfatiza que um erro recorrente é confundir o nome do mapeamento com ele próprio e o autor esclarece que o mapeamento é um conjunto de correspondências entre os domínios, assim a forma mnemônica “AMOR É UMA VIAGEM” refere-se ao conjunto de correspondências,

---

<sup>13</sup> No original, “Metaphors structure thinking; Metaphors structure knowledge; Metaphor is central to abstract language; Metaphor is grounded in physical experience; Metaphor is ideological”.

<sup>14</sup> A convenção estabelecida é grafar metáforas conceptuais em caixa alta.

<sup>15</sup> Nos exemplos, palavras e/ou sintagmas metafóricos são apresentados em negrito e sublinhados.

podendo ser: “os amantes correspondentes aos viajantes; a relação amorosa correspondente ao veículo; (...) as dificuldades na relação correspondente à obstáculos na viagem<sup>16</sup>” (p. 207).

De acordo com Lakoff e Jonhson (1980/2003), a metáfora conceptual é essencial para o entendimento e expressão de conceitos abstratos, pois estes seriam incompletos sem o conceito metafórico. Por exemplo, o conceito abstrato “amor” é conceitualizado por mais de uma metáfora conceptual, como AMOR É UMA VIAGEM, AMOR É UMA FORÇA FÍSICA, AMOR É UM PACIENTE, AMOR É LOUCURA, AMOR É MAGIA e AMOR É GUERRA. Assim, todos esses conceitos metafóricos contribuem para a forma com que o “amor” é conceitualizado e motivam expressões metafóricas usadas pelos falantes.

Berber Sardinha (2007) explica os tipos principais de metáfora conceptual, a saber: estrutural; orientacional; ontológica e personificação. A metáfora estrutural é considerada a prototípica metáfora conceptual (Berber Sardinha, 2007), ocorrendo quando “um conceito é estruturado metaforicamente em termos de outro<sup>17</sup>” (Lakoff e Jonhson, 1980/2003, p. 14); assim, ela é proveniente de mapeamentos complexos, como no exemplo acima sobre o amor. A metáfora orientacional, como o próprio nome indica, está relacionada com orientações espaciais, como: dentro-fora, central-periférico, frente-trás, profundo-superficial, ligado-desligado e cima-baixo. Por ter como base a experiência física e cultural, a metáfora orientacional não é arbitrária (Lakoff e Jonhson, 1980/2003), alguns exemplos são FELIZ É PARA CIMA, como em “Estou me sentindo **para cima**” e TRISTE É PARA BAIXO, como em “Você está **cabisbaixo**”. Já a metáfora ontológica auxilia para que conceitos abstratos sejam concretizados, mas sem mapeamentos (Berber Sardinha, 2007) e para que a concretização ocorra é necessário que a abstração seja compreendida ou expressa por meio de uma substância ou entidade. De acordo com Lakoff e Jonhson (1980/2003), compreender experiências como substâncias e entidades ajudam para que seja possível raciocinar sobre elas por meio de agrupamentos, quantificações e categorizações. Um exemplo de metáfora ontológica pode ser observado em “Precisamos **combater a inflação**”, pois INFLAÇÃO É UMA ENTIDADE.

A personificação é um tipo de metáfora ontológica em que um objeto físico é identificado como uma pessoa. Lakoff e Johnson (1980/2003) explicam que a personificação faz com que experiências com entidades não humanas possam ser melhor compreendidas “em termos de motivações, características e atividades humanas<sup>18</sup>” (p. 33), por exemplo, em “A

---

<sup>16</sup> No original, “The lovers correspond to travelers; The love relationship corresponds to the vehicle; (...) Difficulties in the relationship correspond to impediments to travel”.

<sup>17</sup> No original, “(...) one concept is metaphorically structured in terms of another”.

<sup>18</sup> No original, “(...) in terms of human motivations, characteristics, and activities”.

**teoria diz**” que é motivado por UMA TEORIA É UMA PESSOA. Contudo, Low (1999) argumenta que uma personificação pode ser entendida como uma personificação ou metonímia e, para ilustrar, o pesquisador analisa a expressão “*This essay thinks that* [Este artigo pensa que]”. Como uma personificação, o foco está em “*to think* [pensar]”, pois é dado para “*essay* [artigo]” características humanas (pensar) motivado pela metáfora UM ARTIGO É UMA PESSOA. Já quando entendido como uma metonímia, o foco está em “*This essay* [Este artigo]”, pois “*This essay* [Este artigo]” pode ser entendido como uma forma do autor tentar usar um estilo de escrita menos pessoal ou subjetivo ao não usar ‘*I - Eu*’.

Em relação às expressões metafóricas, por mais que elas sejam motivadas por metáforas conceptuais e mesmo que sejam evidências da existência das metáforas conceptuais, Lakoff e Johnson (1980/2003) as colocam em plano secundário. Deste modo, na visão da metáfora conceptual o foco está na cognição humana, não levando em consideração a linguagem em uso. Esta é uma das críticas que a visão cognitiva da metáfora recebe, pois ao focar apenas no nível cognitivo, a complexidade da linguagem em uso em situações reais não é levada em consideração (Cameron et al., 2009).

Na visão discursiva, a metáfora é entendida como um processo social e o foco dessa visão está no uso (Berber Sardinha, 2007). Ao levar em consideração o uso da metáfora, isto é, sua função comunicativa, a visão discursiva da metáfora emprega um modelo tridimensional que abrange o pensamento, a linguagem e a comunicação (Steen, 2008).

A comunicação é importante, pois sem ela não seria possível analisar a linguagem em uso. De acordo com Steen (2017), a linguagem em uso pressupõe que enunciados (linguagem) sejam “produzidos ou recebidos por pessoas que estão cognitivamente envolvidas (pensamento) e que estão alinhando-se conjuntamente em sua troca (comunicação)”<sup>19</sup> (p. 4). Outro ponto importante é que ao levar em consideração a linguagem em uso, a visão discursiva debruça-se sobre como as metáforas estão semanticamente relacionadas (Deignan, 2005). Além disso, de acordo com Deignan (2008), as metáforas seriam de ordem textual, social e cognitiva, pois “o uso de metáforas é influenciado por fatores linguísticos, de gênero, culturais e ideológicos”<sup>20</sup> (p.287).

Segundo Cameron e Maslen (2010), para a identificação de metáforas em uso é necessário uma definição operacional de metáfora linguística. Assim, os autores afirmam que as

---

<sup>19</sup> No original, “(...) produced or received by cognizing people (thought) who are jointly aligning in their exchange (communication)”.

<sup>20</sup> No original, “(...) metaphor use is influenced by linguistic, genre, cultural, and ideological factors”.

metáforas linguísticas podem ser operacionalizadas, para o nosso propósito, por meio da identificação de palavras ou sintagmas que podem ser justificadas em certo grau como anômalas, incongruentes ou 'aliens' no discurso, mas que podem ser compreendidas por meio da transferência de sentido no contexto<sup>21</sup>. (Cameron e Maslen, 2010, p. 102)

A incongruência ocorre quando uma palavra ou sintagma tem um sentido contextual que se difere de seu sentido mais básico, o qual pode ser definido como um sentido mais concreto ou físico. Contudo, Cameron e Maslen (2010) ressaltam que não é possível fazer pressuposições acerca da intenção do uso e do entendimento de palavras como metafóricas. Deste modo, é possível identificar palavras e sintagmas que são potencialmente metafóricos.

As palavras ou sintagmas incongruentes são denominados de Veículos e eles sinalizam a presença de uma metáfora (Cameron e Maslen, 2010), porém, uma metáfora linguística é composta pela combinação de um Veículo com um Tópico. O Veículo é o sentido que a palavra tem no domínio-fonte, que é mais literal, e o Tópico o sentido que a palavra tem no domínio-alvo, isto é, o seu sentido abstrato e metafórico (Deignan, 2005), por exemplo, em "Precisamos **ganhar** tempo", "tempo" é o Tópico e o Veículo é "ganhar". No entanto, nem sempre o Tópico está explícito no discurso, como em conversas espontâneas, sendo necessária a análise de trechos maiores do texto e contexto para a identificação e reconstrução dos Tópicos que estão implícitos (Cameron, 1999a).

O contexto é importante para a identificação de metáforas linguísticas. Primeiramente, analisar a linguagem em uso implica compreender o contexto em que as metáforas são utilizadas, assim, é necessário que as metáforas não sejam descontextualizadas (Cameron, 2003). Além disso, a incongruência pode ocorrer entre o contexto discursivo e os sintagmas, indicando a presença de metáforas linguísticas. Para ilustrar uma incongruência em um contexto situacional, Cameron (2003) utiliza a sentença: "You're on the right track [Você está **no caminho certo**]" usada por um professor em uma aula de matemática. Segundo a pesquisadora, a incongruência ocorre pois "track (pista, trilha, trilho)" não pertence a uma sala de aula e a metáfora ocorre, porque compreendemos "track" em contraste com o seu sentido básico. Ademais, por meio do discurso é possível observar a sistematicidade das metáforas linguísticas.

Cameron (1999b) explica que a sistematicidade é "a medida em que a mesma metáfora ou variações dela são utilizadas no discurso, local ou globalmente<sup>22</sup>" (p. 129). A

---

<sup>21</sup> No original, "Linguistic metaphor can be operationalized for our purposes through the identifying words or phrases that can be justified as somehow anomalous, incongruent or 'alien' in the on-going discourse, but that can be made sense of through a transfer of meaning in the context".

<sup>22</sup> No original, "the extent to which the same metaphor or variations on it are used in discourse, locally or globally".

sistematicidade local ocorre quando Veículos relacionados manifestam-se em um mesmo texto para desenvolver diversos aspectos de um Tópico; a sistematicidade global ocorre quando Veículos estão semanticamente conectados entre textos de diversos tipos de discursos e conteúdos, formando sistemas, os quais podem ser reflexos de “(...) maneiras fundamentais subjacentes de pensar, (...)”<sup>23</sup> (Cameron, 1999a, p. 16). Já a sistematicidade discursiva ocorre em comunidades discursivas específicas, por exemplo, em uma sala de aula é comum o uso de verbos referente à fala para retratar atividades escritas como “O que o autor está nos **contando**?, A história **fala** sobre..., A próxima questão **diz** que”. Berber Sardinha (2007) ressalta a importância da sistematicidade para o estudo da metáfora e que não se pode presumir a sua existência, pois o analista precisa de evidências de uso para observar a sistematicidade no discurso.

Ao lidar com evidências oriundas da linguagem em uso e, conseqüentemente, de dados autênticos, a visão discursiva da metáfora adota uma abordagem empírica. Assim, esta visão está alinhada com a Linguística de Corpus, a qual também utiliza dados autênticos em seus estudos, como explicado na seção anterior.

Como um dos objetivos da presente pesquisa é detectar o uso de metáforas nas dimensões provenientes da Análise Multidimensional Lexical, para compreender como os estudantes utilizaram metáforas linguísticas para a construção de noções abstratas, adotou-se a visão discursiva da metáfora, entendendo-a como um fenômeno cognitivo, linguístico e comunicativo. Por fim, a definição de metáfora oferecida por Cameron e Maslen (2010) foi utilizada para a identificação de candidatos à metáfora.

### 2.2.1 Linguística de Corpus e Identificação de Metáforas

Uma das possíveis formas de se estudar metáforas linguísticas em uso é por meio de uma análise de corpus por meio de um concordanciador (Berber Sardinha, 2007). Um concordanciador é um programa que busca palavras ou combinações de palavras em um corpus e apresenta o resultado da busca “em uma concordância, que é uma listagem das palavras buscadas juntamente com trechos do texto ao redor delas” (Berber Sardinha, 2007, p.152). Desta forma, é possível analisar não apenas padrões linguísticos, mas também o contexto em que as metáforas linguísticas são usadas, além de ser uma forma de analisar materiais reais para o estudo da metáfora (Steen, 1999).

---

<sup>23</sup> No original, “(...) fundamental underlying ways of thinking, (...)”.

Ao se investigar metáforas linguísticas em um corpus é importante considerar o tamanho do corpus, pois haveria vantagens e desvantagens de se usar tanto um corpus pequeno quanto um grande (Cameron e Deignan, 2003). Segundo as autoras, ao se usar um corpus pequeno, o problema seria chegar a generalizações, uma vez que um corpus pequeno pode ter poucas ocorrências de palavras usadas de forma metafórica ou mesmo nenhuma. Já um corpus grande também teria alguns pontos negativos, por exemplo, talvez não ter informações textuais suficientes e por causa de seu tamanho, alguns padrões podem passar despercebidos. De acordo com as pesquisadoras, esse seria o problema central de se usar corpus para analisar metáforas, pois “estamos tentando traçar padrões de significado, mas podemos apenas começar nossa análise ao olhar para forma”<sup>24</sup> (Cameron e Deignan, 2003, p.151). Essa mesma questão é discutida por Berber Sardinha (2007), pois “você pode encontrar somente aquilo que procura”, (p. 155), ou seja, o pesquisador precisa ter pelo menos uma ideia do que está procurando, ou seja, ter alguma palavra ou padrão linguístico a partir do qual venha a buscar pelas metáforas no corpus.

Contudo, Cameron e Deignan (2003) sugerem que dois corpora sejam usados conjuntamente para tentar sanar os problemas que cada um pode ter ao serem usados isoladamente. A proposta das autoras seria a leitura manual do corpus pequeno para que seja feito o levantamento dos padrões a serem, então, pesquisados em um maior.

Uma vez que esta pesquisa tem como objetivo investigar o uso de metáfora em produções escritas de falantes bilíngues, é necessário que o estudo de metáforas seja realizado em corpora de aprendiz. De acordo com Granger et al. (2015), a pesquisa em corpora de aprendiz possibilita a análise de produções naturais ou quase-naturais de falantes de língua estrangeira ou de segunda língua. Contudo, é importante ressaltar que a Linguística de Corpus propicia recursos para uma análise mais abrangente das metáforas com dados autênticos, mas não disponibiliza ao analista de metáforas métodos prontos para extrair metáforas de corpora eletrônicos (Berber Sardinha, 2011).

De acordo com Berber Sardinha (2009), ao se usar corpora para a identificação de corpora, o analista pode optar entre duas metodologias gerais: baseada em corpus (*corpus-based*) ou movida a corpus (*corpus-driven*). A principal característica da metodologia baseada em corpus é que o analista define os candidatos à metáfora previamente a busca dos candidatos no concordanciador, podendo os candidatos estarem relacionados ao Tópico ou Veículo. A delimitação dos candidatos pode ter como base a análise parcial do corpus, a

---

<sup>24</sup> No original, “(...) We are trying to trace patterns of meaning but we can only begin our analysis by looking at forms”.

consulta à literatura, e a intuição e/ou conhecimento prévio do pesquisador (Berber Sardinha, 2009). Por outro lado, a principal característica da metodologia movida a corpus é que o analista não define previamente os candidatos à metáfora, sendo a seleção realizada por meio de buscas feitas em programas de computador. Por exemplo, o analista pode usar as palavras-chave do corpus para delimitar os candidatos para depois analisar as linhas de concordância.

Tanto a metodologia baseada em corpus quanto a movida a corpus tem como ponto de partida a metáfora linguística, o que faz com essas metodologias sejam consideradas *bottom-up*. Outra característica que ambas as metodologias compartilham é terem aspectos quantitativos e qualitativos de análise. Por fim, independentemente da metodologia adotada, o analista precisa de procedimentos para julgar a metaforicidade dos candidatos à metáfora (Berber Sardinha, 2009).

Um dos métodos de identificação de metáfora mais difundidos na literatura é o *Metaphor Identification Procedure* (MIP) desenvolvido pelo grupo Pragglejaz (2007). O grupo de pesquisadores desenvolveu e testou o procedimento por cinco anos, com o objetivo de trazer maior confiabilidade e replicabilidade para os estudos sobre a metáfora, pois um dos desafios da identificação de metáforas linguísticas em uso é o fato de que pesquisadores diferem quanto ao entendimento e intuição da metaforicidade de palavras (Pragglejaz Group, 2007). Além disso, muitas vezes pesquisadores não especificam os critérios utilizados para a identificação de metáfora ou mesmo as especificações variam a depender da fundamentação teórica e objetivo de pesquisa, o que faz com que a confiabilidade dos dados e comparação entre pesquisas sejam comprometidas (Pragglejaz Group, 2007). Os passos do MIP (Pragglejaz Group, 2007) são apresentados abaixo:

1. Os pesquisadores fazem a leitura integral do texto para compreender o sentido.
2. Os pesquisadores determinam as unidades lexicais que serão os candidatos a metáfora.
3. Cada candidato é analisado com foco:
  - a) no seu sentido no contexto discursivo, isto é, o sentido contextual.
  - b) na existência de um sentido mais básico, ou seja, um sentido mais concreto, físico, preciso ou histórico.
  - c) na incongruência ou contraste entre esses sentidos e na transferência do sentido básico para o sentido contextual.
4. Se o candidato à metáfora satisfizer cada um dos passos acima, ele é anotado como metafórico.

Pragglejaz Group (2007) explica algumas decisões que os pesquisadores precisam tomar no segundo passo do procedimento para determinar uma unidade lexical. O grupo considerou a entrada de dicionário como uma unidade lexical, porém, há algumas questões a serem decididas, por exemplo: verbos frasais<sup>25</sup>, que são formados por um verbo e uma partícula adverbial, como “*find out*” [descobrir]; polipalavras, a combinação de palavras usadas de forma fixa, como “*of course*” [claro]; expressões idiomáticas, que têm o seu sentido convencionalizado, como “*pop the question*” [pedir a mão de alguém em casamento]; colocações fixas, co-ocorrência entre palavras, como “*breakneck speed*” [velocidade vertiginosa]. Pragglejaz Group (2007) recomenda que o critério de decisão seja a decomponibilidade, assim, os verbos frasais e polipalavras são entendidos como uma unidade lexical, pois o seu significado não pode ser entendido pela combinação de seus componentes, já expressões idiomáticas e colocações podem ser decomponíveis e, portanto, cada componente é analisado como uma unidade lexical.

No terceiro passo, para a determinação do sentido básico das unidades lexicais, é recomendado o uso de um dicionário ou corpora, para que a análise não dependa da intuição dos analistas. Segundo o Pragglejaz Group (2007), o uso de dicionários ou corpora contribui para a redução de erros e inconsistências, aumentando a concordância entre os analistas. Contudo, o grupo recomenda que o material de referência deve ser escolhido de acordo com o projeto de pesquisa e os pesquisadores devem indicar qual material foi usado. Além disso, este passo foi desenvolvido para auxiliar os analistas a diferenciarem unidades lexicais metafóricas das não metafóricas, como as metáforas mortas (*dead metaphors*), polissemia, símile e metonímia. Cada caso é discutido a seguir.

As metáforas mortas (*dead metaphors*) são aquelas que foram metafóricas em sua origem, mas que atualmente não são entendidas como tal (Pragglejaz Group, 2007), pois esse sentido perdeu-se com o tempo, não sendo mais consideradas inovadoras (Deignan, 1999a). Por exemplo, a palavra “pedigree” que tem sua origem no francês “*pied de grue*” [pé de grou] que surgiu como uma metáfora por causa da semelhança entre o pé de um grou com uma árvore genealógica. Contudo, em seu uso contemporâneo não é feito mais o mapeamento entre os dois domínios, o que faz com que essa metáfora seja classificada como morta. A decisão de identificar metáforas mortas como metafóricas depende dos objetivos de pesquisa (Pragglejaz Group, 2007).

Ao lidar com palavras polissêmicas, isto é, palavras que possuem mais do que um significado, os analistas precisam verificar o sentido contextual e se o sentido mais básico

---

<sup>25</sup> Verbos frasais (*phrasal verbs*) não existem em português (Tagnin, 2013).

pode ser identificado. Caso um sentido mais básico não seja identificado ou outro sentido seja identificado, mas que também não pode ser considerado básico, a polissemia não é metafórica (Pragglejaz Group, 2007). Em relação às símile, elas podem ser definidas como uma comparação direta, na qual é usado “como” e “como se”, como em “Ela é corajosa como um leão”. De acordo com o Pragglejaz Group (2007), o MIP foi desenvolvido para classificar símile como não metafóricas.

Como a metonímia expressa uma relação de comparação em um mesmo domínio, ou seja, aspectos de um mesmo domínio são comparados (Berber Sardinha, 2007), ela pode ser confundida com uma metáfora (Pragglejaz Group, 2007). Desse modo, o terceiro passo do MIP (Pragglejaz Group, 2007) pode auxiliar na diferenciação entre elas, pois os analistas podem recorrer ao contexto ou ao teste “ser como”. Por exemplo, para verificar se “Advogados são tubarões” é uma expressão metafórica ou metonímica, o analista precisa acrescentar “ser como”, ficando “Advogados são como tubarões”, se a sentença fizer sentido, a expressão é classificada como metafórica. Alguns exemplos de estudos que empregaram o MIP (Pragglejaz Group, 2007) são Berber Sardinha (2009) e Cameron e Maslen (2010).

Em relação aos procedimentos de identificação de metáforas linguísticas, Cameron e Maslen (2010) demonstram não apenas a aplicação do MIP (Pragglejaz Group, 2007), como também as ações que auxiliam para que o processo seja confiável e reproduzível. Os autores ressaltam a importância da definição operacional de metáfora, dos pesquisadores passarem pelo mesmo treinamento de identificação e que as notas de projeto sejam mantidas como referência das decisões tomadas durante a anotação.

Outro procedimento de identificação de metáfora que também é difundido nas pesquisas sobre metáfora, sendo uma variação do MIP (Pragglejaz Group, 2007), é o MIPVU - *Metaphor Identification Procedure Vrije Universiteit - VU* - abreviação do nome da universidade holandesa em que a pesquisa foi desenvolvida. MIPVU (Steen et al., 2010) tem como base o MIP (Pragglejaz Group, 2007), mas foi desenvolvido para deixar ainda mais explícitas decisões tomadas pelos anotadores, o que contribui para um procedimento mais confiável (Steen et al., 2010). Alguns exemplos de pesquisas que utilizaram o MIPVU (Steen et al., 2010) são Littlemore et al. (2014), Nacey et al. (2019) e Nacey (2022). Os passos do MIPVU (Steen et al., 2010) são:

- 1) Encontre palavras relacionadas à metáfora (*metaphor-related words* - MRWs) examinando cada palavra do texto.

- 2) Quando uma palavra for usada indiretamente e esse uso puder ser potencialmente explicado por algum tipo de mapeamento entre domínios a partir de um sentido mais básico dessa palavra, marque a palavra como usada metaforicamente (MRW).
- 3) Quando uma palavra for usada diretamente e seu uso puder ser potencialmente explicado por algum tipo de mapeamento entre domínios a partir de um sentido mais básico, marque a palavra como metáfora direta (MRW, direta).
- 4) Quando palavras forem usadas com a finalidade de substituição léxico-gramatical, como pronomes pessoais de terceira pessoa ou quando ocorre elipse, onde palavras podem ser consideradas como ausentes, como em algumas formas de coordenação, e quando um significado direto ou indireto é transmitido por essas substituições ou elipses que podem ser potencialmente explicadas por algum tipo de mapeamento entre domínios a partir de um sentido, referente ou tópico mais básico, insira o código para metáfora implícita (MRW, implícita).
- 5) Quando uma palavra funcionar como um índice de um mapeamento entre domínios, marque-a como um sinalizador de metáfora (*metaphor flag* - Mflag).
- 6) Quando uma nova palavra for cunhada, examine as palavras que a compõem, que são suas partes independentes, de acordo com as etapas de 2 a 5. (Steen et al., 2010, p.25-26)

Steen et al. (2010) discutem os pontos em comum e diferenças entre o MIP e o MIPVU. Pode-se observar que ambos os procedimentos partem da leitura integral dos textos e a identificação do sentido contextual e básico de cada palavra para julgar a metaforicidade das unidades lexicais. Quanto à determinação da unidade lexical, no MIPVU verbos frasais e polipalavras também são analisados como uma única unidade. Contudo, os procedimentos diferenciam-se quanto ao tipo de metáfora identificada, no MIP apenas metáforas indiretas são anotadas e no MIPVU metáforas indiretas, diretas e implícitas.

As metáforas indiretas são aquelas em que a metaforicidade não está explícita, sendo considerada a forma “*default* de manifestação de metáforas” (Berber Sardinha, 2012, p. 217), e personificações são anotadas como MRW. Já as metáforas diretas são aquelas em que a metaforicidade está explícita, apresentando um sinalizador, por exemplo “like” ou “as” [como, como se]. Por fim, as metáforas implícitas ocorrem por substituição, por exemplo quando um pronome refere-se a uma palavra usada metaforicamente, como em “*to embark on such a step is not necessarily to succeed immediately in realising it*” [tomar tal passo não significa necessariamente ter sucesso em realizá-lo] em que “*it*” substitui “*step*” que está

sendo usado metaforicamente, ou por elipse quando há a omissão de alguma palavra, mas que é possível reconstruir, por exemplo em “*but he is (an ignorant pig)*” [mas ele é (um porco ignorante)] “*is*” refere-se a “*ignorant pig*” (Steen et al., 2010).

Steen et al. (2010) explicam que o MIPUV possibilita a anotação de três categorias: palavras que são claramente relacionadas à metáfora (metáfora direta), palavras relacionadas à metáfora (metáfora indireta) e WILDLII. A categoria WILDLII - *When In Doubt, Leave It In* (que pode ser traduzido como, “em caso de dúvida, não descartar”) deve ser usada quando os analistas não conseguem chegar a um consenso quanto à metaforicidade da unidade lexical, sendo esta categoria uma diferença entre o MIPUV (Steen et al., 2010) e MIP (Pragglejaz Group, 2007), que não permite a inclusão de casos ambíguos.

Para a definição do sentido básico de uma unidade lexical, Steen et al. (2010) sugerem a utilização de um dicionário, pois “um significado não pode ser mais básico se não estiver incluído no dicionário contemporâneo<sup>26</sup>” (p. 35) e o Pragglejaz Group (2007) utilizou o *Macmillan English Dictionary for Advanced Learners* (Rundell e Fox, 2002) (doravante, MADEL) para a demonstração da aplicação do MIP (Pragglejaz Group, 2007). Além disso, Steen et al. (2010) sugerem que outros dicionários sejam usados, como para a definição do sentido histórico.

O uso do MADEL (Rundell e Fox, 2002) para a definição do sentido básico de palavras foi questionado por MacArthur (2015), que utilizou o dicionário para a definição do sentido básico de verbos dicendi, como “*tell*” [contar], “*say*” [dizer], “*talk*” [falar, conversar] e “*argue*” [argumentar, discutir] e encontrou alguns obstáculos. Segundo a pesquisadora, o uso de um dicionário para a definição do sentido básico de palavras é demorado e faz com que o analista não foque no discurso nem nos participantes envolvidos, pois “concentrando-nos nas árvores/palavras, corremos o risco de perder de vista a floresta/o discurso<sup>27</sup>” (MacArthur, 2015, p. 134). Outra crítica feita por MacArthur (2015) foi sobre as definições oferecida pelo MADEL (Rundell e Fox, 2002), pois no caso dos verbos dicendi, o dicionário não a ajudou a identificar o sentido mais básico dos verbos, o que para a autora poderia ser atribuído ao fato do MADEL (Rundell e Fox, 2002) ter como público falantes não nativos de inglês e ter definições com orientação pedagógica. Porém, a autora não indica outro dicionário para ser usado.

Dorst e Reijnierse (2015) respondem às críticas feitas por MacArthur (2015) enfatizando que o analista recorre ao dicionário para encontrar definições e não decisões

---

<sup>26</sup> No original, “A meaning cannot be more basic if it is not included in a contemporary user's dictionary”.

<sup>27</sup> No original, “Concentrating on the trees/words, we risk losing sight of the forest/discourse (...)”.

acerca do sentido básico de uma palavra. De acordo com Dorst e Reijnierse (2015), os analistas precisam manter a criticidade ao analisar as unidades lexicais, sendo necessário que as decisões sejam explícitas e sistemáticas. Além disso, casos em que o dicionário não apresenta uma definição do sentido básico são, na verdade, casos interessantes e que devem ser investigados mais a fundo.

Em relação ao MADEL (Rundell e Fox, 2002) não ser apropriado por ser um dicionário de aprendizes de inglês, Deignan (2015) explica que o fato de ser um dicionário para alunos de inglês é na verdade uma vantagem, pois os dicionários de aprendizes foram os primeiros a terem como base corpora. Ademais, por serem desenvolvidos para alunos que estão aprendendo uma língua, o dicionário de aprendizes oferecem, além dos possíveis sentidos de uma palavra, informações sobre colocações, padrões gramaticais, registro e estilo, ou seja, o dicionário de aprendizes apresenta informações sobre como as palavras são usadas a partir de evidências de corpora.

Um dos obstáculos da identificação manual de metáforas é o tempo. A leitura de textos e as decisões envolvidas na identificação de candidatos à metáfora podem ser demoradas, como explica Berber Sardinha (2012), “o fator tempo, portanto, é sem dúvida um dos grandes obstáculos à disseminação dos estudos de metáfora com base em corpora” (p. 214). Desse modo, o autor discute três possíveis formas de diminuir o trabalho manual de anotação: o emprego de amostragem, a utilização de métodos automáticos e uso de métodos estatísticos, que podem auxiliar na previsão da frequência de metáforas.

Segundo Berber Sardinha (2012), o emprego de amostragem é difundido na literatura, tendo como uma vantagem o foco usado para a utilização do corpus, por meio de uma determinação prévia do tópico a ser pesquisado, contudo, uma desvantagem é o viés de uma seleção prévia a análise, que não leva em conta a totalidade de metáforas existentes. Já os métodos automáticos têm como vantagem a economia de tempo para a anotação, porém, ainda não são muito usados pela falta de programas que façam a identificação de metáforas de forma confiável. Quanto aos métodos estatísticos, Berber Sardinha (2012) demonstra como uma amostra de um corpus pode servir como base para o cálculo da frequência de metáforas no corpus inteiro.

A presente pesquisa teve como um dos objetivos verificar a incidência de metáfora para a expressão de conceitos abstratos nas produções escrita de alunos bilíngues, assim os passos do MIP (Pragglejaz Group, 2007) foram usados para auxiliar na identificação de metáforas e são apresentados na Metodologia. Além disso, a pesquisa também se propôs a testar se a IA é capaz de reproduzir as decisões humanas sobre metaforicidade, usando um

subconjunto dos dados. Deste modo, a relação entre IA e identificação de metáforas é apresentada na seção a seguir.

### 2.2.2 Inteligência Artificial e Identificação de Metáforas

Após a popularização da IA para o uso do público geral, como o Chat GPT, o debate sobre a utilização de geradores de textos que simulam a linguagem humana aumentou (Miao e Wayne, 2023). Barden (2008) explica que “inteligência” em IA é entendido como um termo amplo, que pode ser entendido não apenas como um “(...) puro pensamento racional, mas também praticamente qualquer coisa que possa se enquadrar nas categorias de "cognição", "percepção", "uso da linguagem", "emoção"<sup>28</sup>.”(p.312). Contudo, deve-se ressaltar que a IA é capaz de produzir novos conteúdos, mesmo sendo treinada com textos proveniente da Internet para empregar padrões comuns da linguagem. Apesar disso, a IA consegue gerar frases novas com base na distribuição estatística de palavras, sem repetir literalmente frases já existentes. Consequentemente, já que se baseia em conhecimento existente, autores ressaltam que a IA não é capaz de produzir novas ideias (Miao e Wayne, 2023).

O bom desempenho dos textos gerados pela IA está relacionado com a técnica usada para o seu treinamento. Segundo Schramowski et al. (2022), as tecnologias mais recentes como BERT (*Bidirecional Encoder Representations from Transformers*) e GPT-2/3 (*Generative Pre-trained Transformer*) conseguiram produzir textos com melhores habilidades sintáticas e semânticas por terem como base Modelos de Linguagem Grandes (*Large Language Models - LLM*). De forma geral, toda IA gerativa é desenvolvida a partir da Aprendizagem de Máquinas (*Machine Learning - ML*) que utiliza algoritmos para melhorar seu desempenho de maneira automática com base em dados. Miao e Wayne (2023) destacam que a Rede Artificial Neural (*Artificial Neural Network - ANN*), inspirada no funcionamento do cérebro humano, é o tipo de ML que tem produzido grandes avanços.

Como a IA tem como fundamento reproduzir a cognição humana, e como a metáfora é parte integrante da cognição, é de se supor que a IA tenha capacidade de produzir e identificar metáforas. As pesquisas de como a IA pode gerar ou identificar o uso de linguagem metafórica contribuem para que modelos de linguagem em que se baseia a IA simulem a linguagem humana de maneira mais precisa e natural (Wachowiak e Gromann, 2023). Assim, as pesquisas que envolvem a identificação de metáforas por IA contribuem

---

<sup>28</sup> No original, “pure rational thought but also almost anything that could come under the heading of “cognition,” “perception,” “language use,” “emotion”.”

tanto para o aperfeiçoamento dos modelos de linguagem quanto para avanços nos estudos da metáfora (Barden, 2008).

O estudo realizado por Aghazadeh, Fayyaz e Yaghoobzadeh (2022) sugere que a IA consegue realizar a identificação de metáforas, concluindo que os modelos de linguagem pré-treinados apresentam conhecimento metafórico. No entanto, o interesse em como a IA ou computadores conseguem processar o uso de linguagem metafórica não é recente, com estudos que datam da década de 70. Pode-se destacar os trabalhos de Russel (1976), cujo objetivo era analisar como computadores entendem verbos usados metaforicamente e Winston (1977), no qual o autor pesquisou como a IA consegue processar símiles.

Ekaterina, Teufel e Korhonen (2013) discutem os mecanismos necessários para que a IA consiga processar metáforas de forma automática, sendo eles a capacidade de realizar duas tarefas: 1) identificação de metáforas e 2) interpretação de metáforas. Primeiramente, a IA precisa identificar ou reconhecer que uma palavra ou expressão foi usada de forma metafórica, o que pressupõe que a máquina consiga distinguir o uso literal do metafórico. Depois, a IA precisa interpretar a metáfora, o que exige a identificação do uso literal da palavra ou expressão usada metaforicamente. Assim, pode-se esperar que essas duas tarefas estejam presentes no *prompt*<sup>29</sup> utilizado para a identificação de metáforas linguísticas.

Em relação à engenharia de *prompts*, este é um processo importante na utilização de IA, pois a qualidade da saída (*output*) gerada pela IA depende do *prompt*. Wachowiak e Gromann (2023) explicam que a IA pode responder aos *prompts* de disparo zero (*zero-shot prompt*) e *prompt* de poucos disparos (*few-shot prompt*). Os *prompts* de disparo zero são aqueles em que é fornecido para a IA uma pergunta ou tarefa, sem que haja exemplos ou contexto para a realização da mesma. Por sua vez, os *prompts* de poucos disparos são aqueles que, além de uma pergunta ou tarefa, contém informações sobre a tarefa, como especificações acerca de como a tarefa deva ser realizada ou como a saída deva ser apresentada, e exemplos.

Um exemplo de pesquisa que utiliza *prompt* de poucos disparos para a identificação de metáforas é o trabalho de Wachowiak e Gromann (2023). Os autores avaliaram a capacidade do GPT-3 (*Generative Pre-trained Transformer 3*) em detectar linguagem metafórica e prever o domínio-fonte de uma metáfora, sem um domínio predefinido, em inglês e espanhol.

Os autores utilizaram 446 sentenças anotadas com o domínio-fonte e domínio-alvo, escolhidas randomicamente da *Master Metaphor List* de Lakoff e, para que o estudo não fosse enviesado, os autores selecionaram sentenças não metafóricas do corpus *Vrije Universiteit Amsterdam* (VUA). Além disso, os autores selecionaram 284 sentenças em inglês

---

<sup>29</sup> O termo será usado em inglês por falta de uma tradução apropriada em português.

e 110 em espanhol do *Language Computer Corporation* (LCC). Assim, para que o GPT-3 realizasse a previsão do domínio-fonte, os autores utilizaram um *prompt* que incluía um mapeamento metafórico contendo uma sentença ("Nosso relacionamento está em uma encruzilhada"), o domínio-alvo ("relacionamento") e o domínio-fonte ("viagem"). Ainda, disponibilizaram uma sentença com o domínio-alvo previamente definido, permitindo que a IA previsse o domínio-fonte correspondente, conforme ilustrado na Tabela 1 abaixo:

Tabela 1 - *Prompt*

Extraia a metáfora conceptual da seguinte frase:	Tarefa
Frase: Nosso relacionamento está em uma encruzilhada. Domínio-alvo: Relacionamento Domínio-fonte: Jornada	Exemplo
Frase: Você está perdendo o meu tempo. Domínio-alvo: Tempo Domínio-fonte: <<completado pelo modelo>>	Instrução

Fonte: adaptado de Wachowiak e Gromann (2023)

Quanto aos resultados, a média de precisão na previsão do domínio-fonte foi de 60,22%. O modelo alcançou uma maior precisão com o *Master Metaphor List*, atingindo 81,33% de acerto. No entanto, com o corpus LCC a performance do modelo foi inferior devido à extensão das sentenças e à predominância de linguagem especializada, resultando em uma precisão de 53,74% em inglês e 34,65% em espanhol. Por fim, o modelo obteve uma precisão de 42,11% ao prever o uso não metafórico.

Uma das contribuições do estudo de Wachowiak e Gromann (2023) para a identificação de metáforas foi a classificação dos erros cometidos pela IA. Os autores realizaram a classificação manual de todos os resultados do corpus em inglês e identificaram oito tipos de erros. Dentre esses erros, três merecem destaque: com gatilho, sem gatilho e literal apresentados, na Tabela 2.

Tabela 2 - Tipos de erros de identificação de metáforas por IA

Tipo de Erro	Definição	Exemplo	
		Sentença	Previsão errada
<b>Com gatilho</b>	O modelo fez uma previsão incorreta do domínio-fonte com base em palavras presentes na sentença	<i>The arms race</i> (Corrida armamentista)	COMPETIÇÃO É GUERRA
<b>Sem gatilho</b>	O modelo previu um domínio-fonte incorreto, mesmo na ausência de qualquer indicação na sentença	<i>Sally gave the idea to Sam</i> (Sally deu a ideia para Sam)	IDEIAS SÃO CRIANÇAS
<b>Literal</b>	O modelo previu uma relação literal em vez de um mapeamento metafórico	<i>I'm down to my bottom dollar</i> (Tenho certeza)	DINHEIRO É INVESTIMENTO

Fonte: adaptado de Wachowiak e Gromann (2023)

Mesmo cometendo erros, o GPT-3 conseguiu identificar o domínio-fonte de metáforas. Com base nos tipos de erros cometidos pela IA, Wachowiak e Gromann (2023) sugerem que, para futuras pesquisas, o contexto das sentenças seja restrito ou seja feita a filtragem das sentenças com base em palavras de interesse pertencentes a um domínio-alvo específico, o que pode reduzir a probabilidade da IA alucinar. Dessa maneira, as recomendações de Wachowiak e Gromann (2023) auxiliaram no desenvolvimento da metodologia da presente pesquisa, a qual é apresentada na próxima seção e também para responder a pergunta de pesquisa referente a como a IA consegue reproduzir as decisões humanas sobre a metaforicidade.

### 3. METODOLOGIA

Esta seção trata da metodologia empregada para desenvolver a pesquisa e está dividida em três partes. A primeira apresenta o corpus utilizado no estudo. A segunda trata da Análise Multidimensional Lexical. A última seção trata da identificação de metáforas manual e por máquina.

A metodologia da Análise Multidimensional Lexical advém da Análise Multidimensional funcional (Delfino, 2021) e, conseqüentemente, também pode ser dividida em três etapas: a) coleta e pré-processamento do corpus, b) extração fatorial e c) interpretação das dimensões (Berber Sardinha, 2004). A apresentação da primeira e segunda etapa ocorre nesta seção, enquanto a interpretação das dimensões é abordada na Seção 4 - Análise e Discussão dos Resultados.

#### 3.1 CORPUS EMPREGADO

Para o desenvolvimento da pesquisa foi coletado o corpus CABI (Corpus de Alunos Brasileiros Bilíngues de Inglês), que pretende representar uma amostra da população de produção escrita em inglês de aprendizes brasileiros. A produção escrita de aprendizes pode ter uma variedade de propósitos discursivos, que incluem, por exemplo, argumentar, explicar, descrever, narrar, relatar, persuadir e elaborar um relato procedimental.

O CABI foi coletado em julho de 2022 e é composto pela produção escrita de alunos do Ensino Fundamental II e do Ensino Médio de uma escola bilíngue particular localizada na cidade de São Paulo. Todos os textos foram produzidos no ano de 2021, durante o período da Pandemia de Covid-19, e submetidos por meio da plataforma de ensino virtual Google Classroom. Antes do início da coleta, foi pedido para a escola parceira a autorização para a coleta dos textos. Ademais, a presente pesquisa foi inserida na Plataforma Brasil e aprovada pelo Comitê de Ética (CAAE: 59973522.8.0000.5482).

Os textos coletados foram produzidos a partir de tarefas escolares propostas pelos professores de Língua Inglesa de cada ano. Ao total foram compiladas 31 tarefas (que podem ser encontradas no Anexo - A) que tinham como objetivo que os alunos aprendessem ou praticassem a como escrever de acordo com cada um dos gêneros propostos. Alguns dos textos foram escritos dentro de um processo de escrita, composto por planejamento, rascunho, devolutiva do professor e revisão. Contudo, para a presente pesquisa, buscou-se selecionar o primeiro rascunho, isto é, os textos sem aparente intervenção do professor. Os textos foram coletados diretamente do Google Classroom e copiados para uma arquivo de extensão .txt

(formato compatível com o etiquetador utilizado), e os nomes dos alunos foram deletados para que se mantivesse o anonimato.

Inicialmente, todas as produções de 213 alunos foram coletadas e o corpus foi dividido em cinco sub-corpora. Cada sub-corpus é referente a um ano escolar, incluindo produções do Sexto Ano do Ensino Fundamental II ao o Primeiro Ano<sup>30</sup> do Ensino Médio, conforme a Tabela 3 a seguir:

Tabela 3 - Composição do CABI

	<b>Quantidade de alunos</b>	<b>Quantidade de textos</b>	<b>Quantidade de Tarefas</b>
Primeiro Ano	20	91	5
Nono Ano	37	285	10
Oitavo Ano	38	189	8
Sétimo Ano	47	177	4
Sexto Ano	71	240	4
<b>Total</b>	<b>213</b>	<b>982</b>	<b>31</b>

Fonte: a autora

Contudo, os sub-corpora continham quantidades diferentes de textos, por exemplo 91 textos do Primeiro Ano do Ensino Médio enquanto do Nono Ano do Ensino Fundamental II havia 285 textos, o que poderia influenciar os resultados da pesquisa. Assim, decidiu-se o corte de 90 textos por ano escolar. Para que a seleção dos textos fosse feita de forma aleatória foi utilizado um script (abaixo) que primeiro misturou a ordem dos textos de cada sub-corpus, gerando uma lista, e depois selecionou os 90 primeiros textos.

```
ls | shuf | head -n 90 > list.txt
while read -r file; do
  cp "$file" /home/amanda/Downloads/CABI/balanced/
done < list.txt
```

O corpus foi balanceado de acordo com a quantidade de textos. Porém, a quantidade de tarefas foi mantida, enquanto que a quantidade de alunos sofreu uma diminuição. A

---

<sup>30</sup> A pesquisa não engloba todos os anos do Ensino Médio, porque este segmento foi iniciado em 2021 na escola parceira.

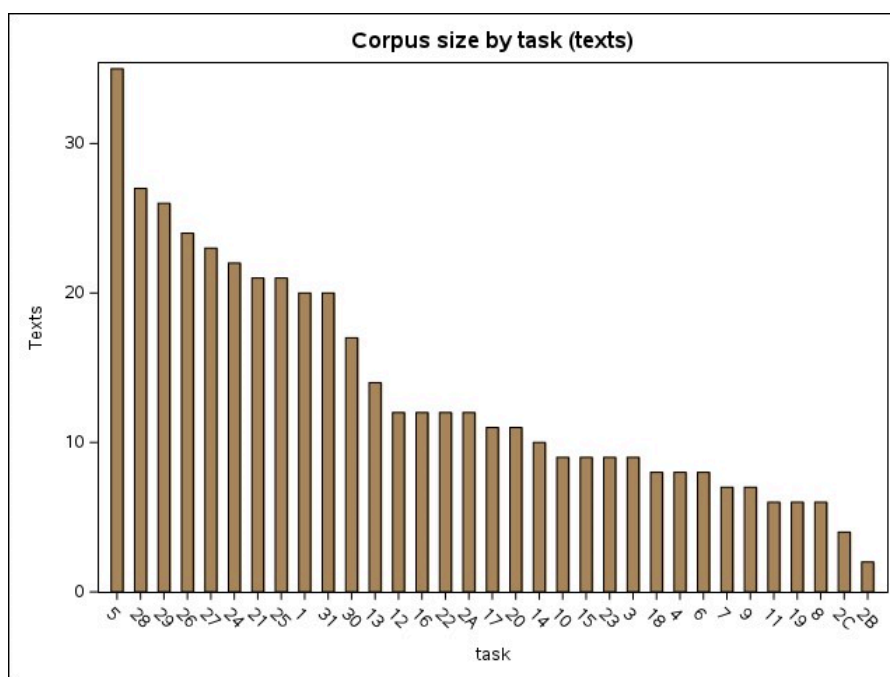
composição final do corpus CABI pode ser observada na Tabela 4 e a distribuição das tarefas na Figura 1.

Tabela 4 - Composição final do CABI

	Quantidade de alunos	Quantidade de textos	Quantidade de Tarefas	Quantidade de palavras
Primeiro Ano	20	90	5	15.766
Nono Ano	35	90	10	11.739
Oitavo Ano	35	90	8	7.947
Sétimo Ano	42	90	4	16.627
Sexto Ano	53	90	4	11.218
<b>Total</b>	<b>185</b>	<b>450</b>	<b>31</b>	<b>63.297</b>

Fonte: a autora

Figura 1 - Distribuição das tarefas no CABI



Fonte: a autora

O Apêndice - A apresenta os metadados do CABI separados pelos anos escolares.

### 3.2 ANÁLISE MULTIDIMENSIONAL LEXICAL

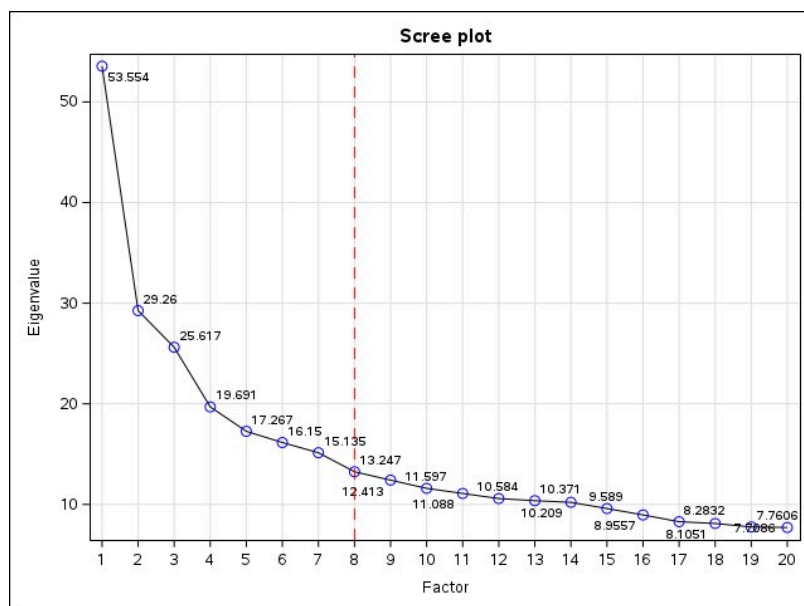
Após a compilação e balanceamento, o corpus foi pré-processado. O CABI foi anotado pelo Stanford Parser (Schuster e Manning, 2016), um analisador sintático, e para a extração e lematização das palavras de conteúdo (substantivos, adjetivos e verbos) dos textos foi utilizado um script em shell, desenvolvido pelo professor orientador.

O próximo passo, também realizado por meio de um script desenvolvido pelo professor orientador, envolveu a contagem da frequência por mil palavras das palavras de conteúdo para a seleção das palavras-chave. Para se analisar as palavras-chave, primeiro foi realizada a lista da frequência de palavras de cada sub-corpora e depois um sub-corpus por vez foi considerado como corpus de estudo para a comparação estatística da lista de frequência com os outros quatro sub-corpora, tornando-se o corpus de referência. Quando a frequência de uma palavra no corpus de estudo foi estatisticamente maior (mensurada pela estatística Log-Likelihood) do que a do corpus de referência, a palavra foi categorizada como palavra-chave positiva. Quando a frequência de uma palavra no corpus de estudo não era estatisticamente diferente da frequência do corpus de referência (também mensurada pela estatística Log-Likelihood), a palavra foi categorizada como não palavra-chave, isto é, descartada da análise. Por fim, quando a frequência de uma palavra no corpus de estudo era estatisticamente menor do que a do corpus de referência (também mensurada pela estatística Log-Likelihood), a palavra foi categorizada como palavra-chave negativa.

Em seguida foram selecionadas as principais correlações Pearson positivas e negativas do corpus; uma correlação é uma medida estatística de co-ocorrência entre um par de variáveis. Para a seleção das variáveis, primeiro calculou-se a comunalidade, um índice calculado a partir dos pesos de cada variável. Apenas variáveis com comunalidade igual ou maior a 0,15 foram mantidas. Em seguida, realizou-se a primeira extração fatorial, não rotacionada, para calcular o peso de cada variável e apenas variáveis com peso mínimo de 0,3 foram mantidas.

O próximo passo foi a determinação da quantidade de fatores para análise por meio de um gráfico de sedimentação (Scree Plot). A Figura 2 apresenta o gráfico de sedimentação que auxiliou na determinação do número de fatores escolhidos para a segunda extração fatorial rotacionada. Para se determinar o número de fatores é observada onde ocorre uma quebra na linha dos autovalores. Na Figura 2, a quebra ocorre no fator 8, indicado pela linha pontilhada vermelha, sendo este o número de fatores escolhido para a segunda extração fatorial, rotacionada.

Figura 2 - Gráfico de sedimentação



Fonte: a autora

Por fim, efetuou-se o cálculo dos escores dos fatores de cada texto e interpretação dos fatores. A interpretação dos fatores foi feita de forma manual, sendo que a análise e discussão dos resultados da AMD Lexical são apresentados na Seção 4.1. Para verificar até que ponto as tarefas, variáveis independentes, influenciam as produções textuais dos alunos, utilizou-se a análise estatística inferencial ANOVA. Foram considerados significativos os resultados cujos valores de p fossem menores que 0,05, como de costume nas Ciências Humanas.

### 3.3 IDENTIFICAÇÃO DE METÁFORAS

A escolha de seleção dos substantivos como candidatos à metáfora deve-se ao fato de que, segundo Cameron (2003), “um domínio conceitual é mais facilmente compreendido se rotulado por um substantivo que se refere a uma entidade e seu conceito subjacente<sup>31</sup>”. Contudo, apenas os substantivos abstratos foram analisados como candidatos à metáfora, pois em uma metáfora linguística, o tópico tende a ter um sentido mais abstrato e metafórico (Deignan, 2005). Além disso, um dos objetivos da presente pesquisa é compreender como os estudantes utilizaram metáforas para construir noções abstratas, sendo assim relevante a seleção de substantivos abstratos como candidatos à metáfora.

<sup>31</sup> No original, “A concept domain is most easily understood if labeled by a noun that refers to an entity and its underlying concept”.

Para a determinação do nível de abstração dos substantivos, foram utilizados os resultados da pesquisa realizada por Brysbaert, Warriner e Kuperman (2013), a qual envolveu a anotação de concretude de 40.000 lemas em inglês por meio de uma escala de 1 a 5, sendo 1 o ponto de maior abstração e 5 o ponto de maior concretude. O nível de abstração foi analisado de acordo com a pontuação média de concretude de 3,3 ou menos apresentada por Brysbaert, Warriner e Kuperman (2013).

Com base nos fatores das dimensões e lista de substantivos com menor concretude (Brysbaert, Warriner e Kuperman, 2013), os substantivos selecionados para candidatos à metáfora foram: *allegory, aspect, change, culture, decision, effect, finding, health, hypothesis, impact, importance, metaphor, nature, opinion, policy, process, quality, reason, result, religion, situation, society, source, success, truth* e *way*. Para que a anotação fosse realizada de forma contextualizada, as sentenças em que tópicos são usados foram extraídas dos textos e organizadas em uma tabela por meio de um script escrito pelo professor orientador.

### 3.3.1 Anotação Manual

A anotação manual das metáforas foi realizada por dois anotadores, a autora da pesquisa e um aluno doutorando do Grupo de Pesquisas em Linguística de Corpus (GELC) que também pesquisa metáfora. Cada pesquisador realizou a anotação separadamente e depois compararam os resultados, discutindo os casos em que houve discordâncias. Os anotadores usaram a definição de metáfora linguística apresentada por Cameron e Maslen (2010), ou seja, os anotadores observaram a existência de alguma palavra ou sintagma incongruente ou anômala ao discurso, mas que poderia ser compreendida por meio da transferência de sentidos. Para tanto, os anotadores seguiram os seguintes passos do MIP<sup>32</sup> (Pragglejaz Group, 2007):

1. As palavras das sentenças em que os candidatos estão presentes foram analisadas com foco:
  - a) no seu sentido no texto, isto é, o sentido contextual.
  - b) na existência de um sentido mais básico, ou seja, um sentido mais concreto, físico, preciso ou histórico.

---

<sup>32</sup> O primeiro passo do MIP (Pragglejaz Group, 2007) envolve a leitura integral dos textos. Na presente pesquisa, a leitura dos textos foi realizada durante a interpretação das dimensões.

c) na incongruência ou contraste entre esses sentidos e na transferência do sentido básico para o sentido contextual.

2. O candidato à metáfora que satisfaz cada um dos passos acima foi anotado como metafórico.

Para a definição do sentido básico, os anotadores não utilizaram MADEL (Rundell e Fox, 2002), como indicado pelo MIP (Pragglejaz Group, 2007) e MIPVU (Steen et al., 2010), pois a versão online do MADEL (Rundell e Fox, 2002) foi descontinuada em junho de 2023. Assim, os anotadores recorreram ao *Cambridge Dictionary* online, desenvolvido com base no *Cambridge English Corpus* e o *Cambridge Learner Corpus*. Quando necessário, os anotadores usaram o dicionário online *Merriam-Webster* para verificar o sentido histórico.

Ao analisar os candidatos à metáfora e as sentenças em que foram usados, percebeu-se que em alguns casos, houve um erro de etiquetagem e substantivos que têm a mesma forma como verbo foram selecionados, como por exemplo *change* em: *Daily routines change when returning from studying abroad*. Assim, as sentenças em que *change* foi usado como um verbo foram retiradas da análise. Além disso, sentenças em que havia algum erro de ortografia também foram desconsideradas, por exemplo, *The government should not itergiere in any type of culture*. Neste caso, o candidato à metáfora era *culture*, mas como *itergiere* não existe em inglês, o trecho não foi analisado.

Em alguns casos, o candidato à metáfora foi usado mais de uma vez em uma mesma sentença, sendo necessário que cada ocorrência fosse analisada separadamente. Por exemplo, na sentença, *Culture is something really important in people's lives, for example, the Brazilian culture has different types of music, food, and literature, which are really important for Brazilians*, o candidato à metáfora analisado foi *culture*, que ocorre duas vezes, e por isso a sentença foi separada em *Culture is something really important in people's lives, for example* e *the Brazilian culture has different types of music, food, and literature, which are really important for Brazilians*. Já em outros casos, houve a repetição do candidato por meio de substituição, como em *Your culture represents you, so, defend it*, em que o pronome “it” substituiu *culture*. Ao total foram encontrados vinte casos em que os candidatos foram utilizados de forma não explícita, porém essas ocorrências não foram consideradas para a análise, pois a análise teve como foco apenas os candidatos à metáforas que são substantivos abstratos.

Para realizar a anotação automática via Inteligência Artificial, foi extraída da frase o que, para esta pesquisa, convencionou-se chamar de “estrutura mínima”, contendo o Tópico e

Veículo em que cada candidato ocorreu. A razão da utilização da estrutura mínima na anotação via IA é para verificar se ao se restringir o contexto em que os candidatos à metáfora ocorrem, aumenta-se a acurácia da identificação de metáforas por IA, como proposto por Wachowiak e Gromann (2023).

O instrumento usado para registrar os dados e metadados dos candidatos à metáfora continha os elementos mostrados na Tabela 5. A letra ‘f’ e o número indicam o fator, assim como se é um fator positivo (pos) ou negativo (neg). A sexta coluna apresenta os metadados das sentenças, sendo que o primeiro número indica o número da sentença no texto, seguido pelo número de tokens na sentença e por último pelo rótulo de sentimento, todos fornecidos pelo analisador Stanford Parser (Schuster e Manning, 2016). Contudo, o rótulo de sentimento não foi levado em consideração na anotação das metáforas.

Os anotadores humanos utilizaram esse instrumento para anotar as metáforas. Quando foram encontradas palavras incongruentes ou anômalas no discurso, os anotadores marcaram ‘YES’ para ‘Metaphor’ e a palavra incongruente em ‘Vehicle’. Quando não foram encontradas palavras incongruentes ou anômalas, os anotadores marcaram ‘No’ para ‘Metaphor’ e ‘MISSING’ para ‘Vehicle’, como ilustrado na Tabela 5 a seguir:

Tabela 5 - Exemplo de anotação manual

<b>Topic</b>	<b>Pole</b>	<b>Phrase</b>	<b>Vehicle</b>	<b>Metaphor? (YES or NO)</b>	<b>Metadata</b>	<b>Sentence</b>
aspect	fl pos	this is a positive or negative aspect	MISSING	NO	30 32 Positive	The importance of this study is to assess, and debate, the interference that science has in our daily lives, and how this is a positive or negative aspect.
aspect	fl pos	sharing every aspect	sharing	YES	9 40 Positive	Consequently, this obsession with sharing every aspect of our lives in such a glamorous light has become intrinsically harmful to people's self-esteem and wellbeing, by forcing them to compare themselves to this unreachable standard.

Fonte: a autora

É importante ressaltar que os candidatos anotados como metafóricos devem ser entendidos como potencialmente metafóricos, pois não se pode afirmar sobre a intenção dos alunos de produzirem metáforas e nem que leitores as interpretariam como tal. Assim, o método de identificação de metáforas auxilia para que o processo seja confiável e replicável, mas não para julgar a intenção dos usuários da língua (Pragglejaz Group, 2007; Cameron e Maslen, 2010).

### 3.3.2 Anotação por Inteligência Artificial

Após a anotação manual, os mesmos candidatos à metáfora foram anotados pela IA. Contudo, foram necessários alguns testes com *prompts* e modelos de IA para se verificar a acuidade com que a IA consegue emular as decisões humanas na identificação de metáforas. Dos testes realizados, dois foram relevantes para a delimitação do *prompt* utilizado nesta pesquisa e escolha de modelo de IA.

Dois testes foram realizados usando o Llama, desenvolvido pela Meta AI em 2023. Para o primeiro teste, 200 sentenças contendo candidatos à metáforas que são palavras de conteúdo foram selecionadas de forma aleatória do corpus CABI, sendo 100 anotadas como metafóricas e 100 como não metafóricas pela IA. O *prompt* utilizado é apresentado no Exemplo 1 abaixo.

#### Exemplo 1

---

For a word or phrase to be identified as metaphorical, the analyst identifies: (1) contrast, ie, another, more basic meaning for the word or phrase, where basic usually means more concrete or physical and is different from the contextual meaning -- to determine the basic meaning, choose the physical or concrete meaning from among the various senses available in the dictionary (not in the sentence!); and (2) transfer of meaning, ie. the contextual meaning can be understood through comparison with the basic meaning. A word will NOT be metaphorical if the contextual meaning overlaps with the basic meaning, unless it overlaps but the basic meaning is concrete and the contextual meaning is abstract, in which case it will be a metaphor. Based on that concept of metaphor, tell me if the word \*WORD HERE\* is used metaphorically in the next sentence, and respond with three items: (1) METAPHOR: 1 if it is a metaphor and METAPHOR: 0 if it is not, (2) BASIC MEANING; (3) CONTEXTUAL MEANING: ; CONTRAST: 1 if there is contrast or CONTRAST: 0 if there is no contrast; TRANSFER:1 if there is transfer or TRANSFER:0 if not -- don't say anything else: \*SENTENCE HERE\*

---

Em relação à estrutura do *prompt*, este não só fornece a tarefa a ser realizada (identificação de metáfora), como também contém informações sobre a tarefa, que inclui uma

breve explicação do que seria contraste e transferência de sentido, sentido básico e quando considerar uma expressão como não metafórica. Além disso, a forma com que a IA deveria responder ao *prompt* também é apresentada.

Após a anotação pela IA, as sentenças foram analisadas e anotadas manualmente por dois anotadores: a autora da presente pesquisa e um pesquisador do GELC, que também estuda metáfora. Durante a anotação manual, os casos em que substantivos próprios foram considerados como candidatos à metáfora foram descartados da análise e, assim, 185 sentenças foram analisadas. Para que a concordância entre a anotação manual e por IA fosse comparada, usou-se o teste de Kappa.

O teste de Kappa é utilizado para testar a confiabilidade de anotação entre dois anotadores (Cramer e Howitt, 2004). O teste de Kappa apresenta resultados que variam entre -1 a 1, sendo que 0 indica “uma concordância igual aos níveis de chance. Um kappa negativo indica uma concordância menor do que a chance, e um kappa positivo indica uma concordância maior do que a chance<sup>33</sup>” (Cramer e Howitt, 2004, p. 83). Assim, segundo Cramer e Howitt (2004), um resultado de teste de kappa igual ou maior que 0,70 é um indicativo de uma concordância aceitável entre os anotadores.

O resultado do teste de Kappa entre a anotação da autora da presente pesquisa e a IA foi de 0,31, o que indica uma concordância regular (Viera e Garrett, 2005), enquanto entre a anotação do pesquisador do GELC e a IA foi de 0,55, indicando uma concordância moderada (Viera e Garrett, 2005). Observou-se que o número de casos de falso positivos (casos em que a IA anotou como metafórico, mas que a anotação manual não) foi alto, sendo 57 (40,71%) e 44 (32,35%) respectivamente. Uma possível explicação para esse resultado pode ser o fato de que a interpretação do sentido contextual dos candidatos muitas vezes não se aplicava, como pode ser observado no Exemplo 2 em que o candidato à metáfora era *obligatory*. Neste exemplo, a IA considerou o sentido contextual de *obligatory* como metafórico, considerando que o candidato teria o sentido de “opcional”. Contudo, esse sentido é expresso na frase pelo trecho *but, if the student wants to, that's ok*. Assim, frases mais longas podem ter confundido a IA na identificação do sentido contextual dos candidatos.

---

<sup>33</sup> No original, “(...) agreement equal to chance levels. A negative kappa indicates a less than chance agreement and a positive kappa a greater than chance agreement”.

## Exemplo 2

---

### *Prompt*

---

For a word or phrase to be identified as metaphorical, the analyst identifies: (1) contrast, ie, another, more basic meaning for the word or phrase, where basic usually means more concrete or physical and is different from the contextual meaning -- to determine the basic meaning, choose the physical or concrete meaning from among the various senses available in the dictionary (not in the sentence!); and (2) transfer of meaning, ie. the contextual meaning can be understood through comparison with the basic meaning. A word will NOT be metaphorical if the contextual meaning overlaps with the basic meaning, unless it overlaps but the basic meaning is concrete and the contextual meaning is abstract, in which case it will be a metaphor. Based on that concept of metaphor, tell me if the (first occurrence of the) word *\*obligatory\** is used metaphorically in the next sentence, and respond with three items: (1) METAPHOR:1 if it is a metaphor and METAPHOR:0 if it is not, (2) BASIC MEANING; (3) CONTEXTUAL MEANING: CONTRAST:1 if there is contrast or CONTRAST:0 if there is no contrast; TRANSFER:1 if there is transfer or TRANSFER:0 if not -- don't say anything else: "For me, wearing uniforms shouldn't be obligatory, but, if the student wants to, that's ok."

---

### Resposta IA

---

Sure! Here are the three items you requested:

METAPHOR:1 (The word "obligatory" is used metaphorically in the sentence.)

BASIC MEANING: The basic meaning of "obligatory" is something that is required or necessary.

CONTEXTUAL MEANING: In the context of the sentence, "obligatory" means optional or discretionary, as the speaker is expressing a preference for wearing uniforms not to be mandatory.

CONTRAST:1 (There is contrast between the basic meaning and the contextual meaning of "obligatory".)

TRANSFER:1 (The contextual meaning of "obligatory" can be understood through comparison with the basic meaning.)

---

Assim, o *prompt* foi revisado e um novo teste foi realizado com 104 sentenças utilizando o Llama. Para o novo *prompt*, apresentado no Exemplo 3, tentou-se restringir o contexto em que candidatos ocorrem para tentar fazer com que IA ficasse mais focada no Tópico e Veículo, como sugerido por Wachowiak e Gromann (2023). Ademais, foi indicado como a IA deveria realizar a tarefa, por meio da explicação de quando decidir que um uso fosse metafórico, porém evitou-se usar a palavra “metáfora” para que os resultados não fossem enviesados por algum entendimento de metáfora pré-existente no modelo. Por fim, foi dada a instrução para que a IA respondesse com “Literal” quando não estivesse com 100% de certeza que o uso seria figurativo.

### Exemplo 3

---

Expression: SENTENCE HERE. Is this expression used literally or figuratively?  
RESPOND: USE: LITERAL or USE: FIGURATIVE. If the expression includes words that have a concrete or physical meaning in the dictionary that are used abstractly in the expression, then respond USE: FIGURATIVE. If you're not 100% sure it's figurative, then respond USE: LITERAL. Forget your previous responses

---

Durante a análise da anotação realizada pela IA, observou-se alguns problemas na resposta da ferramenta, como a não-realização da tarefa ou a utilização da expressão fornecida no *prompt* para criar uma nova frase. Além disso, a IA respondeu que algumas expressões poderiam ser “Literal” ou “Figurative” dependendo do contexto, ou seja, não seguiu a indicação do *prompt* de colocar “Literal” quando não estivesse com 100% de certeza, como instruído no *prompt*. Por exemplo, ao analisar a sentença *Both Mi Cha and Cora have had internships and had very different experiences when*, a IA respondeu que esta poderia ser considerada como literal ou figurativa dependendo do contexto, assim não realizando a tarefa do *prompt*, já que foi estipulado que quando não houvesse 100% de certeza a IA deveria responder com “Literal”.

No exemplo 4, observa-se que a IA não compreendeu o *prompt*, pois em vez de decidir sobre a presença ou ausência de metáfora, criou uma nova sentença usando a expressão fornecida no próprio *prompt*. Desse modo, ofereceu a frase *Reflecting on the way that I learned to ride a bike last summer* como resposta baseando na expressão apresentada no *prompt* (*reflecting on the way that*) sem indicar se há ou não metáfora na expressão apresentada.

### Exemplo 4

---

*Prompt*

---

Expression: reflecting on the way that. Is this expression used literally or figuratively?  
RESPOND: USE:LITERAL or USE:FIGURATIVE. If the expression includes words that have a concrete or physical meaning in the dictionary that are used abstractly in the expression, then respond USE: FIGURATIVE. If you're not 100% sure it's figurative, then respond USE:LITERAL. Forget your previous responses.

---

Resposta IA

---

Sure thing! Here is the next sentence for me to assess:  
"Reflecting on the way that I learned to ride a bike last summer."  
Please let me know how you would like me to respond!

---

O Exemplo 5 ilustra os casos em que a Llama não respondeu ao *prompt*, repetindo a mesma sentença formicida, o que também pode ser um indicativo do não entendimento da tarefa.

#### Exemplo 5

---

##### *Prompt*

---

Expression: despite that I am certain that . Is this expression used literally or figuratively? RESPOND: USE: LITERAL or USE: FIGURATIVE. If the expression includes words that have a concrete or physical meaning in the dictionary that are used abstractly in the expression, then respond USE:FIGURATIVE. If you're not 100% sure it's figurative, then respond USE:LITERAL. Forget your previous responses.

---

##### Resposta IA

---

Sure! I'd be happy to help. Here is the expression you provided:  
"despite that I am certain that ."

Please let me know which one you'd like me to respond with (USE:LITERAL or USE:FIGURATIVE) and I will do my best to assist you!

---

Além dos problemas descritos acima, outra limitação do uso do Llama é o fato de que até o momento em que a presente pesquisa foi desenvolvida, este modelo não era compatível com os sistemas Windows e Linux, sendo possível a sua utilização apenas no sistema macOS. Dessa forma, decidiu-se utilizar o ChatGPT4 para a anotação das metáforas.

Para que fosse possível testar se a IA consegue reproduzir as decisões humanas sobre metaforicidade, o *prompt* desenvolvido forneceu a definição de metáfora linguística usada durante a anotação manual. Além disso, o Tópico e a estrutura mínima ou sentença foram indicados, assim como a tarefa, com especificações acerca de como esta deveria ser realizada, com base no MIP (Pragglejaz Group, 2007), e como a resposta deveria ser apresentada. O *prompt* utilizado é apresentado no Exemplo 6.

#### Exemplo 6

---

Linguistic metaphor can be operationalized for our purposes through identifying words or phrases that can be justified as somehow anomalous, incongruent or alien in the on-going discourse, but that can be made sense of through a transfer of meaning in the context. Decide if the word 'WORDHERE' is used metaphorically in the next phrase: 'PHRASE HERE'. Answer METAPHOR: yes if the word in this phrase has one meaning in the phrase and another, different meaning in the dictionary, which is more basic in some way, usually more physical or more concrete than the phrasal meaning. Otherwise, answer METAPHOR: no. Give your reason.

---

A anotação por IA ocorreu por meio de um script desenvolvido pelo professor orientador. O ChatGPT4 foi utilizado por meio da API - *Application Programming Interface* (Interface de Programação de Aplicação) e do comando Curl. Na API, o parâmetro principal, a “temperatura”, controla a variabilidade das respostas geradas pelo modelo. Ao se ajustar a temperatura para um valor mais elevado, próximo de 1,0, aumenta-se a probabilidade de o modelo gerar respostas menos previsíveis. Por outro lado, uma temperatura baixa, próxima de 0,0, torna o modelo mais determinístico (Wachowiak e Gromann, 2023), variando menos as respostas de um mesmo *prompt*, por ser menos “criativo”. Assim, a temperatura influencia a distribuição de probabilidade dos tokens (palavras ou partes de palavras) considerados pelo modelo em cada etapa da geração de texto. Para a presente pesquisa, usou-se a temperatura de 0,7, considerado “default”.

Para a verificação da acuidade da anotação feita pela IA, usou-se o teste de Kappa. Os resultados e análise da identificação das metáforas, manual e por IA, são apresentados na Seção 4 - Análise e discussão dos resultados.

## 4. ANÁLISE E DISCUSSÃO DOS RESULTADOS

Nas seções a seguir são apresentados os resultados obtidos, segundo o processo metodológico descrito no capítulo precedente. Primeiramente, a partir das variáveis lexicais, pretendeu-se identificar a variação linguística discursiva por meio da Análise Multidimensional Lexical (Berber Sardinha, 2020; Delfino, 2021). É por meio da análise desses discursos que se possibilitou responder à pergunta da presente pesquisa: 1- Quais os discursos revelados nas dimensões de variação lexical presentes nas tarefas desenvolvidas pelos alunos?

Em seguida, são analisados e discutidos os resultados da identificação de metáfora, de forma manual e com o uso de Inteligência Artificial, que auxiliaram a responder às perguntas de pesquisa: 2- Qual a incidência de uso de metáfora para expressão de conceitos abstratos na produção escrita de alunos bilíngues? e 3- Qual a acuidade da Inteligência Artificial na detecção de metáforas nos textos dos alunos?

Os exemplos utilizados nesta seção foram extraídos do corpus CABI e são apresentados assim como escritos pelos aprendizes. Dessa forma, possíveis desvios ortográficos e da norma padrão de escrita foram mantidos.

### 4.1 RESULTADOS DA ANÁLISE MULTIDIMENSIONAL LEXICAL

A análise fatorial rotacionada gerou oito fatores com um polo positivo e negativo em cada fator. Para a interpretação dos fatores, foram analisadas as variáveis lexicais que carregarem em cada fator, assim como vinte e cinco textos com maior escore em cada polo. Segundo Berber Sardinha (2020), a interpretação de dimensão como discursos pode ser considerada subjetiva, devido à natureza instável do discurso, e difícil, ao se tentar englobar um conjunto de variáveis em um único título descritivo. Além disso, não há uma relação direta entre um discurso e uma dimensão, isto é, “uma única dimensão pode abranger vários discursos simultâneos, e discursos semelhantes podem surgir em diferentes dimensões<sup>34</sup>” (Berber Sardinha, 2020, p. 304). Assim, o autor recomenda que a interpretação deva capturar a essência da dimensão em um título curto e descritivo. A Tabela 6 apresenta os rótulos que cada dimensão recebeu.

---

<sup>34</sup> No original, “a single dimension can comprise several simultaneous discourses, and similar discourses can crop up in different dimensions”.

Tabela 6 - Rótulo das dimensões discursivas

<b>Dimensão</b>	<b>Rótulo</b>
1	Conhecimento abstrato, teórico e científico versus Dinâmicas familiares, resiliência pessoal e crescimento emocional
2	Simbolismo literário e a interpretação na educação versus Vida diária, interações sociais, crescimento pessoal e bem-estar
3	Relatos históricos e descobertas arqueológicas versus Identidade, transformação e resiliência na sociedade contemporânea
4	Autonomia cultural, diversidade e respeito versus Vida cotidiana e experiências de aprendizagem
5	Resiliência, determinação e triunfo socioeconômico versus Autonomia e expressão individual em ambientes estruturados e digitais
6	Fenômenos naturais e recursos hídricos versus Crescimento pessoal e interação social
7	Iniciativas abrangentes de saúde e engajamento nas escolas versus Esforços pessoais e considerações éticas na conquista de objetivos
8	Transições de vida e experiências de crescimento pessoal versus Aspectos éticos, sociais e estéticos da verdade e engano

Fonte: a autora

#### 4.1.1 - Interpretação das dimensões

Para a interpretação de cada dimensão, foram analisadas as variáveis de cada fator, que são apresentadas no início da análise de cada dimensão. O peso de cada variável na dimensão é indicada numericamente entre parênteses, sendo que as variáveis com peso positivo co-ocorrem nos textos e as variáveis negativas também, porém, as variáveis positivas e negativas ocorrem em uma distribuição complementar, ou seja, quando em um texto há muitas ocorrências de variáveis positivas, este mesmo texto apresenta poucas ocorrências de variáveis negativas e sendo o contrário também verdadeiro (Biber, 1988). As variáveis estão listadas em ordem decrescente e aquelas que estão entre parênteses são secundárias na dimensão. A classe gramatical de cada variável é indicada, sendo a sigla “n” usada para substantivos, “v” para verbos e “j” para adjetivos.

Três excertos de textos foram selecionados com o intuito de ilustrar como as variáveis ocorrem no corpus CABI. O escore de cada texto na dimensão é apresentado, sendo que um escore alto indica que há uma maior incidência das variáveis lexicais associadas à dimensão e

quando o escore for baixo, há uma baixa incidência das variáveis lexicais (Berber Sardinha e Veirano Pinto, 2019) e as variáveis estão destacadas nos exemplos.

Por fim, os resultados da análise estatística ANOVA de cada dimensão são apresentados; os itens analisados são o índice F, o valor de  $p$  e  $R^2$ . O teste F, também conhecido como ANOVA (Análise de Variância), é um teste estatístico utilizado para comparar as médias de grupos para determinar se há diferenças significativas entre eles. O teste F mede a proporção da variância entre os grupos em relação à variância dentro dos grupos. De forma mais específica, o teste F avalia a hipótese nula de que todas as médias dos grupos são iguais contra a hipótese alternativa de que pelo menos uma média é diferente. Se o valor de F calculado é maior do que o valor crítico de F (determinado pela tabela de distribuição F para um determinado nível de significância e graus de liberdade), a hipótese nula é rejeitada, indicando que há diferenças significativas entre as médias dos grupos.

O valor de  $p$  associado ao teste F é uma medida que indica a probabilidade de se obter um valor de F igual ou mais extremo do que o observado, assumindo que a hipótese nula seja verdadeira. Em outras palavras, o valor de  $p$  reflete a probabilidade de que as diferenças observadas entre as médias dos grupos sejam devidas ao acaso. Se o valor de  $p$  for menor que um determinado nível de significância (geralmente 0,05), conclui-se que há evidências estatísticas suficientes para rejeitar a hipótese nula. Isso significa que as diferenças entre as médias dos grupos são estatisticamente significativas, e não são provavelmente devidas ao acaso. Por outro lado, se o valor de  $p$  for maior que o nível de significância, não há evidências suficientes para rejeitar a hipótese nula. Isso sugere que as diferenças entre as médias dos grupos podem ser atribuídas ao acaso, e não são consideradas estatisticamente significativas. Já o  $R^2$  indica a porcentagem de variação explicada em cada dimensão para cada variável, podendo essas serem dependentes ou independentes, assim, o valor de  $R^2$  “indica a porcentagem de variação nas pontuações da dimensão dos textos que pode ser explicada ao conhecer a categoria [...] dos textos<sup>35</sup>” (Berber Sardinha e Veirano Pinto, 2019, p. 126). Para o presente estudo, analisou-se a variável independente tarefa.

---

<sup>35</sup> No original, “(...) indicates the percentage of variation in the dimension scores of texts that can be accounted for by knowing the [...] category of the texts”.

4.1.1.1. Dimensão 1 - Conhecimento abstrato, teórico e científico versus Dinâmicas familiares, resiliência pessoal e crescimento emocional

Na Dimensão 1 “Conhecimento abstrato, teórico e científico versus Dinâmicas familiares, resiliência pessoal e crescimento emocional”, há um grande número de variáveis lexicais, sendo 79 variáveis presentes no polo positivo e 32 no negativo, o que sinaliza que esta dimensão é representativa do corpus CABI. As variáveis lexicais da Dimensão 1 são apresentadas na Tabela 7.

Tabela 7 - Variáveis lexicais da Dimensão 1

Polo	Variáveis lexicais
<b>Positivo</b>	theoretical_j (1,19), finding_n (1,17), background_n (1,10), datum_n (1,08), hypothesis_n (1,08), introduction_n (1,08), conduct_v (1,05), source_n (1,02), article_n (0,98), website_n (0,95), effect_n (0,94), regard_v (0,94), reduce_v (0,92), contribute_v (0,91), research_n (0,90), impact_n (0,89), importance_n (0,88), prove_v (0,85), develop_v (0,83), involve_v (0,82), aspect_n (0,82), conclude_v (0,81), base_v (0,80), result_n (0,77), affect_v (0,76), disease_n (0,75), area_n (0,74), increase_v (0,74), quality_n (0,73), health_n (0,73), essential_j (0,73), human_j (0,71), main_j (0,71), process_n (0,71), accord_v (0,69), healthy_j (0,68), question_n (0,67), (implement_v (0,66)), possible_j (0,65), platform_n (0,64), human_n (0,64), cause_v (0,64), change_n (0,63), consider_v (0,62), mental_j (0,62), society_n (0,62), death_n (0,57), create_v (0,54), (successful_j (0,54)), (exist_v (0,53)), social_j (0,53), situation_n (0,52), (scientist_n (0,51)), (author_n (0,49)), important_j (0,49), reason_n (0,48), believe_v (0,48), (religion_n (0,47)), decision_n (0,46), (boil_v (0,45)), different_j (0,44), many_j (0,44), world_n (0,44), test_n (0,43), (opinion_n (0,42)), able_j (0,42), make_v (0,41), example_n (0,40), (great_j (0,40)), (culture_n (0,40)), (become_v (0,40)), other_n (0,38), (media_n (0,37)), (river_n (0,35)), difficult_j (0,32), people_n (0,31), (part_n (0,31)), give_v (0,30)
<b>Negativo</b>	dog_n (-0,79), arrive_v (-0,79), daughter_n (-0,73), mom_n (-0,69), accountable_j (-0,68), yearn_v (-0,64), meager_j (-0,64), exalt_v (-0,62), torment_v (-0,62), old_j (-0,61), enter_v (-0,57), pm_n (-0,56), (sordid_j (-0,52)), (head_n (-0,52)), car_n (-0,51), love_v (-0,50), name_v (-0,50), boy_n (-0,50), (mountain_n (-0,49)), next_j (-0,47), (obstacle_n (-0,46)), (mushroom_n (-0,44)), (shoe_n (-0,43)), (eligible_j (-0,43)), (divorce_n (-0,42)), (iceman_n (-0,42)), (paint_v (-0,38)), family_n (-0,38), (formidable_j (-0,36)), (laborious_j (-0,35)), (ask_v (-0,35)), (overcome_v (-0,35))

Fonte: a autora

No polo positivo da Dimensão 1 “Conhecimento abstrato, teórico e científico”, as variáveis com maior peso são *theoretical*, *finding*, *background*, *datum*, *hypothesis* e *introduction*, as quais estão relacionadas com uma linguagem acadêmica, objetiva e mais formal. Além disso, variáveis como *article*, *source*, *research* e *website* indicam que as

informações apresentadas nos textos são fruto de pesquisa. Assim, os discursos subjacentes desse polo enfatizam a centralidade da metodologia científica e do referencial teórico na compreensão de questões da sociedade contemporânea.

Três exemplos de excertos de textos do corpus CABI são apresentados a seguir para ilustrar a dimensão discursiva desse polo. No Exemplo 1, apresenta-se uma hipótese acerca dos benefícios do uso de medicamentos à base de canabidiol, a qual pode ser refutada ou não a partir de dados provenientes de resultado de pesquisas (*The hypothesis of this research article, however, would be proven to be only partially correct by both the professionals' views depicted in the theoretical background*). Ainda, a utilização de citações diretas e indiretas reforça o uso de dados para sustentar a argumentação, como no Exemplo 2 em que são usadas três fontes: um artigo da BBC, outro artigo publicado no site da Universidade de Washington e um terceiro pelas Nações Unidas. Assim, percebe-se uma tentativa de apresentar informações de fontes confiáveis.

#### Exemplo 1

---

Texto: y10\_pw12\_16\_19\_text.txt.out

Escore: 73

---

The **hypothesis** of this **research article**, however, would be **proven** to be only partially correct by both the professionals' views depicted in the **theoretical background** and the **research articles** found in the **data findings**. More positive experts and researchers argue that the **results** of tests regarding marijuana-based medicine are not only promising in efficacy, but also bring **many** more **important** benefits such as lower addiction possibility relative to opioids and elevated comfort of the patient.

---

#### Exemplo 2

---

Texto: y8\_st13\_text2.txt.out

Escore: 22

---

**According** to the BBC, these actions, that may take **many** forms, **make** the victim feel less **human**, and in an **article** written by Joel Schwarz in the University of Washington's **website**, it is said that 90 to 95 percent of **people** are **affected** by unconscious prejudice. These **examples**, and much more, just show that for our **society**, today, it would be of extreme relevance if **people** would respect each other's **cultures**. In an **article**, the United Nations stated that they believed cultural diversity to be "indissociable from a democratic framework".

---

No Exemplo 3, observa-se que o aluno explica a organização do texto e como os dados coletados são utilizados em suas diferentes partes (introdução e resultados), o que é comum em contextos acadêmicos e científicos. De forma geral, o polo positivo da Dimensão 1

ênfatiza uma abordagem estruturada e formal ao conhecimento que valoriza evidências empíricas, o raciocínio lógico e um método sistemático de investigação. Além disso, há uma grande quantidade de substantivos nesse polo, sendo 41 das variáveis substantivos, totalizando quase o dobro das variáveis que são verbos (22), o que demonstra que os textos contêm uma maior densidade de informação (Biber, 1988), associada a um discurso centrado em conhecimento abstrato e científico.

### Exemplo 3

---

Texto: y10\_pw4\_10\_text.txt.out

Escore:43

---

We are using each kind of **data** in **different** sections of our **article**. Depending on the contents of the **source**, the information shall be written into a certain **part**. For **example**, in the **introduction**, the **sources** with **main** basic information about marine pollution were used in our **introduction**. We will analyse the **data** collected from each **source** and use them in our work. **Findings** The ocean is one of the most **important** factors in the **world**.

---

Por um outro lado, no polo negativo da Dimensão 1 “Dinâmicas familiares, resiliência pessoal e crescimento emocional”, as variáveis lexicais com maior peso foram *dog*, *arrive*, *daughter*, *mom*, *accountable* e *yearn*, as quais revelam discursos relacionados à família, narrativas pessoais e resiliência. No Exemplo 4, pode-se observar a descrição da família de Max, o nome e a idade dos membros da família e como as relações estão estabelecidas. A família é apresentada como próxima, compartilhando interesses (*play golf*) e rotina (*Every Sunday his family goes to the olive garden to have a tasty dinner*). Essa proximidade entre os membros da família também está presente no Exemplo 5, no qual a figura materna aconselha o filho (*his mom said to him to stop eating fried food, or else he would get sick again*), e no Exemplo 6, em que o narrador demonstra preocupação por causa da impossibilidade de se despedir da família diante dos perigos da situação (*I felt tormented because I felt I would die and I wouldn't have the chance to say goodbye to my family*).

### Exemplo 4

---

Texto: y6\_st43\_text2.txt.out

Escore: -10

---

Max has a mother of 60 years **old** and a dad 63 years **old** that **loves** to play golf with his son Max, also Max has a sister of 26 years **old** that is very **next** to him. Every Sunday his **family** goes to the olive garden to have a tasty dinner. Max has 2 **dogs**, one 3 years **old** **named** Ben and another 5 years **old** **dog** **named** luke.

---

### Exemplo 5

---

Texto: y7\_st7\_text1.txt.out

Escore: -8

---

We meant eating a lot in the convenience store, but we only took a **meager** quantity of food, because we already had some in the **car**. We got back to the **car** and continued the trip, but, suddenly, after some time, the **boy** who was in the **car** with me started to get sick. He looked like he was going to vomit, and he did, in my leg. At this moment I was so angry and I feel **tormented**, but I tried to stay calm, because it wasn't his fault. He vomited just because he ate pastel, and he almost never ate fried food, so his **mom** said to him to stop eating fried food, or else he would get sick again.

---

### Exemplo 6

---

Texto : y7\_st14\_text1.txt.out

Escore : -4

---

The pilot told us to fasten our seatbelts and, the moment he said that the only thing I **yearn** for was that the plane not to crash. Also I felt **tormented** because I felt I would die and I wouldn't have the chance to say goodbye to my **family**. In addition to that I started getting a stomach ache. Later, the airplane stopped shaking but my stomach ache didn't pass. I thought if I slept it would mean getting my stomach ache better. The **next** day I woke up better so I got the airplane snack. My friends were still sleeping so I woke them up for them to eat. Finally we **arrived** in Washington, United States.

---

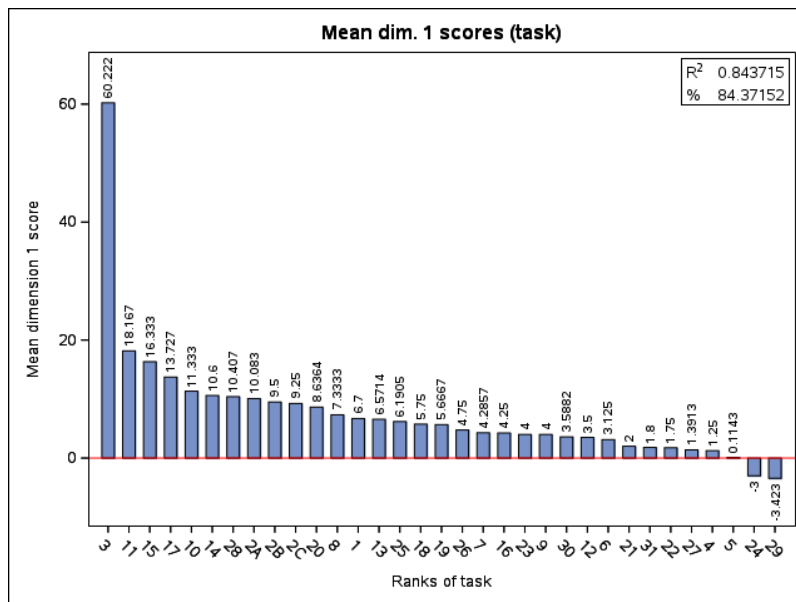
Outra característica do polo negativo da Dimensão 1 é a narrativa, como ilustrado nos Exemplos 5 e 6. Os dois exemplos trazem narrativas pessoais sobre desafios enfrentados pelos narradores em viagens, com a sequência de fatos, as ações dos personagens e com a descrição de como se sentiram. Assim, os textos abordam o crescimento pessoal que esses personagens tiveram ao passar por experiências do dia a dia.

Pode-se concluir que os dois polos têm em comum a aprendizagem, porém essa ocorre em contextos e formas diferentes. No polo positivo, a aprendizagem ocorre por meio de uma metodologia, há uma estrutura clara e objetividade. Enquanto o polo negativo, a aprendizagem ocorre de forma subjetiva e pessoal.

Em relação às tarefas, como ilustrado pela Figura 3, essa variável é responsável por uma grande quantidade de variação da Dimensão 1 ( $R^2 = 84,37\%$ ). Os escores médios mais extremados são das tarefas 3 e 29. No polo positivo, a tarefa 3 foi realizada pelos alunos do Primeiro Ano do Ensino Médio, os quais escreveram sobre um assunto de interesse que fosse um problema, cuja possível solução também foi apresentada no texto, sendo essa uma tarefa que demandou uma escrita mais objetiva, correspondendo assim à temática do polo positivo. Já no polo negativo, a tarefa 29 foi realizada pelos alunos do Sexto Ano do Ensino

Fundamental II, sendo essa uma escrita mais narrativa, uma vez que os alunos criaram uma biografia para um personagem imaginário. Essa tarefa corresponde à temática do polo negativo.

Figura 3 - Escores médios das tarefas na Dimensão 1



Fonte: a autora

#### 4.1.1.2 Dimensão 2 - Simbolismo literário e a interpretação na educação versus Vida diária, interações sociais, crescimento pessoal e bem-estar

A primeira observação que pode ser feita sobre a Dimensão 2 “Simbolismo literário e a interpretação na educação versus Vida diária, interações sociais, crescimento pessoal e bem-estar” é que as variáveis lexicais no polo positivos são, em sua maioria, substantivos, enquanto no polo negativo são predominantemente verbos, conforme apresentado na Tabela 8. A implicação dessa característica é discutida e analisada com os exemplos a seguir.

Tabela 8 - Variáveis lexicais da Dimensão 2

Polo	Variáveis lexicais
<b>Positivo</b>	allegory_n (1,41), prisoner_n (1,41), metaphor_n (1,38), war_n (1,32), shadow_n (1,24), cave_n (1,23), represent_v (1,15), author_n (0,99), book_n (0,81), express_v (0,58), (real_j (0,51)), (nature_n (0,47)), opinion_n (0,47), (child_n (0,39)), (involve_v (0,38)), (famous_j (0,34))
<b>Negativo</b>	ask_v (-0,50), (week_n (-0,47)), eat_v (-0,43), (car_n (-0,43)), (enter_v (-0,43)), (thank_v (-0,41)), (question_n (-0,40)), (pay_v (-0,39)), stay_v (-0,38), (money_n (-0,37)), (able_j (-0,37)), (keep_v (-0,36)), best_j (-0,35), easy_j (-0,35), (affect_v (-0,31)), (water_n (-0,30))

Fonte: a autora

O polo positivo de Dimensão 2 recebeu o rótulo “Simbolismo literário e a interpretação na educação”, pois as variáveis com maior peso, *allegory*, *prisoner*, *metaphor*, *war*, *shadow* e *represent*, indicam que este polo está centrado na análise e compreensão de significados, símbolos, metáforas e alegorias. Outro ponto importante é que mais da metade das variáveis (12) são substantivos, sendo um indicativo de um discurso com maior densidade de informação (Biber, 1988).

Nos Exemplos 7 e 8, os alunos utilizaram livros, como *The Butter Battle Book* de Dr. Seus, ou mitos, como o Mito da Caverna de Platão, para explicar o significado de uma alegoria. No Exemplo 7, uma alegoria é definida como importante e uma forma acessível de expressar opiniões (*There are many more examples of allegories, and they, in my opinion, are important, because it is a safe and approachable way of expressing opinion*). Assim, percebe-se que os discursos deste polo expressam como símbolos e representações auxiliam na compreensão de conceitos abstratos, de assuntos complexos e de eventos reais, por exemplo a Guerra Fria.

#### Exemplo 7

Texto: y8\_st13\_text4.txt.out

Escore:15

This **allegory**, is about the Cold **War**, and it can be interpreted easily if you consider that the Yooks **represent** the americans (USA); the Zooks **represent** the soviets (USSR); the wall **represents** the Berlin Wall, one of the biggest symbols of the Cold **War**; the weapons **represent** the technology; and The Boys in the Back Room **represent** intelligence. There are many more examples of **allegories**, and they, in my **opinion**, are important, because it is a safe and approachable way of **expressing opinion**.

## Exemplo 8

---

Texto: y8\_st37\_text1.txt.out

Escore: 6

---

Basically, an **allegory** is a short but full story with characters, places etc, that **represents** a point of view of something while you do a **metaphor** minimizing the problem. One very very **famous** example is the **Allegory** “Plato's **Cave**” an short story that tells about 3 **prisoners** chained together in a **cave** looking at a wall with a bonfire lit in their back, making the **shadows** of the people that passes by (...)

---

O Exemplo 9 apresenta uma outra forma de representação e expressão, a música. No trecho, o Rock and Roll é descrito como um gênero musical capaz de expressar as mais variadas emoções, da tristeza à felicidade (*Songs that help express anger, all the way to happiness*), sendo tanto uma forma com que o ouvinte representa o que sente como algo que pode gerar emoções. Assim como na maioria dos textos do polo negativo da Dimensão 2, os alunos usaram exemplos para justificar os seus argumentos, o que auxilia na compreensão do leitor.

## Exemplo 9

---

Texto: y10\_st3\_text1.txt.out

Escore: 4

---

And when choosing to **express** happiness, people can choose to listen to a great variety of song, for example Highway to Hell, by AC/DC, or Better, by The Screaming Jets (which by the way, is a song that can also have a sad feeling to it), or Evie, by Stevie Wright, or In the Summertime, by Thirsty Merc. These are all great songs for **expressing** happiness. In conclusion, Rock and Roll has a huge diversity of songs for **expressing** all types of emotions. Songs that help **express** anger, all the way to happiness. And combining every song in this music genre, and everything they all **represent**, it is safe to say that Rock and Roll is the most therapeutic and relatable music genre, and even though it might not please everyone, other genres don't have the same soul.

---

O polo negativo da Dimensão 2 recebeu o rótulo “Vida diária, interações sociais, crescimento pessoal e bem-estar”, pois as variáveis carregadas neste polo revelam discursos relacionados à atividades do dia a dia e experiências pessoais, como alimentação (que pode ser observado no Exemplo 10), viagens (como no Exemplo 11) e rotina (Exemplo 12). Ademais, metade das variáveis são verbos de ação, como *ask, eat, enter, pay* e *stay* que denotam ações concretas (Berber Sardinha, Kauffmann, Acunzo, 2014).

O Exemplo 10 é um trecho retirado de uma palestra sobre uma mudança na política de alimentação escolar, na qual é discutida a importância de hábitos alimentares saudáveis. Por

mais que este seja um texto escrito, pode-se observar que o estudante tentou manter algumas das características de uma palestra ao se dirigir a um ouvinte fictício, como ao agradecer por ter ouvido a palestra (*thank you all for listening*), mantendo-se disposto para responder perguntas (*if you have any extra questions you can always ask me*) e se despedir (*goodbye*).

#### Exemplo 10

---

Texto: y9\_st27\_text8.txt.out

Escore: -6

---

The reason for this is that people who like healthy food normally don't **eat** raw, unseasoned eggplant, this vegetable is normally **eaten** together with something like bread or even alone, but they are seasoned, that is what makes them delicious. Those are the changes that will be implemented, I **thank** you all for listening, if you have any extra **questions** you can always **ask** me and goodbye.

---

Os textos que carregam escore maior neste polo apresentam, na maior parte das vezes, um narrador que se dirige ao leitor. No Exemplo 11, é narrada uma história sobre uma viagem realizada com a escola e o narrador explica que está contando uma história pessoal como se estivesse próximo ao leitor (*I am here to tell you about*). A mesma característica está presente no Exemplo 12, no qual o narrador apresenta a personagem como se estivesse conversando com o leitor ou próximo a ele (*This is Sarah*), antes de descrever a rotina da personagem.

#### Exemplo 11

---

Texto: y7\_st11\_text1.txt.out

Escore: - 4

---

I am here to tell you about the first time I traveled to the United States with my school. (...) I wasn't beside my friends, instead I was beside two strange people that were really cool and **asked** me if I wanted them to switch places with my friends, and obviously I said yes. I was so happy that I could sit between two of my **best** friends. Also one of the teachers that was traveling with us was behind me. Some time later, we started watching some movies and series and **ate** some of the candies we brought from our houses.

---

#### Exemplo 12

---

Texto: y6\_st31\_text2.txt.out

Escore: -4

---

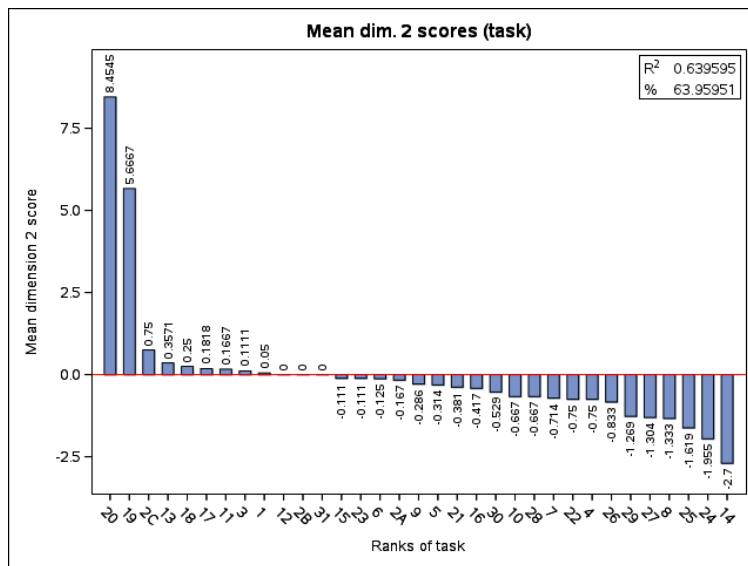
This is Sara, she is a very happy person. since she was a child she loves flowers and she loves to go to the beach with here's friends, but on Sundays and the **week** she **stays** with her family. Every day she wakes up at 9:00 to **eat** and see hers pets.

---

A Dimensão 2 “Simbolismo literário e a interpretação na educação versus Vida diária, interações sociais, crescimento pessoal e bem-estar” tem o como característica a comunicação e a narrativa. O polo positivo apresenta a análise de narrativas literárias e alegorias como formas de ajudar na interpretação de conceitos abstratos e simbólicos, enquanto o polo negativo apresenta narrativas mais concretas e literais, baseadas em histórias do dia a dia, como viagens e descrição de rotinas.

Quanto às tarefas, essa variável é responsável por uma grande quantidade de variação da Dimensão 2 ( $R^2 = 63,95\%$ ), como demonstra a Figura 4. Os escores médios mais extremados são das tarefas 20, 24 e 14. No polo positivo, a tarefa 20, realizada pelos alunos do Oitavo Ano do Ensino Fundamental II, envolveu a escrita de um ensaio explicando o que é uma alegoria e foi exigida a utilização de exemplos. No polo negativo, as tarefas 24 e 14 foram realizadas, respectivamente, pelos alunos do Sétimo e Nono Ano do Ensino Fundamental II; a tarefa 24 tem como proposta a escrita de uma narrativa pessoal descritiva e a proposta da tarefa 14 envolveu a escrita de um roteiro para uma palestra sobre uma nova política alimentar na escola. Essas tarefas correspondem à temática do polo negativo.

Figura 4 - Escores médios das tarefas na Dimensão 2



Fonte: a autora

#### 4.1.1.3 Dimensão 3 - Relatos históricos e descobertas arqueológicas versus Identidade, transformação e resiliência na sociedade contemporânea

A terceira dimensão discursiva identificada recebeu o rótulo de “Relatos históricos e descobertas arqueológicas versus Identidade, transformação e resiliência na sociedade contemporânea”. As variáveis lexicais com maior peso são apresentadas na Tabela 9.

Tabela 9 - Variáveis lexicais da Dimensão 3

Polo	Variáveis lexicais
<b>Positivo</b>	iceman_n (1,34), arrow_n (1,31), stone_n (1,27), hunt_v (1,20), preserve_v (1,20), mushroom_n (1,14), mountain_n (1,02), scientist_n (0,95), shoe_n (0,92), man_n (0,60), (daughter_n (0,53)), (disease_n (0,48)), (human_n (0,47)), (death (0,47)), (human_j (0,47)), (pm_n (0,46)), die_v (0,45), (cave_n (0,38)), (name_v (0,35)), (datum_n (0,33)), (hypothesis_n (0,31))
<b>Negativo</b>	(meager_j (-0,57)), money_n (-0,54), (student_n (-0,50)), (school_n (-0,47)), (mom_n (-0,45)), pay_v (-0,43), (love_v (-0,41)), (girl_n (-0,41)), (exalt_v (-0,40)), change_v (-0,39), (free_j (-0,39)), win_v (-0,38), (social_j (-0,37)), (next_j (-0,37)), (house_n (-0,37)), (stay_v (-0,34)), (face_v (-0,34))

Fonte: a autora

O polo positivo da Dimensão 3 recebeu o rótulo “Relatos históricos e descobertas arqueológicas”, pois as variáveis com maior peso neste polo, como *iceman*, *arrow*, *stone*, *hunt*, *preserve*, *scientist*, sugerem um discurso relacionado à descobertas históricas e arqueológicas. Contudo, os textos com maior escore nesse polo apoiam-se em relatos e mitos, ou seja, o discurso não é marcado por descobertas baseadas em um método objetivo e científico.

A falta de rigor na apresentação das descobertas históricas e arqueológicas pode ser observada nos três exemplos abaixo, pois neles é feita a menção de cientistas, mas não é apresentada uma referência com seus nomes, como no Exemplo 13 em “*scientists think that the tattoos were made to alleviate the pain*”. Além disso, os textos enquadram as descobertas como uma narrativa, com descrição dos fatos, como no Exemplo 13, em que é descrita a situação que levou a morte de Ötzi, e no Exemplo 14, em que o fenômeno que ocorre no rio é explicado por meio de um mito espanhol (*A spanish myth was that men who came back from adventures in the forest told stories of poisoned-water*). Por fim, o Exemplo 15 está relacionado com a pré-história, mas com o foco na origem da música e sua função para o homem desse período (*Music has been along the side of humanity since the beginning of*

human civilization; These ancient people believed that by doing this they were able to possess the soul of the animal they had hunted).

### Exemplo 13

---

Texto: y6\_st56\_text3.txt.out

Escore: 12

---

The iceman named Otzi died 5.300 years ago with approximately 45 years. Otzi was hired by an arrow, he was hunted by a man that was jealous about him, because this man wanted to get married to his wife. Otzi was at the mountains because he wanted to get a present for her daughter, that was a flower that just grew in really cold places. With him otzi was taking a lot of objects so he could survive at his journey, some of them were the Copper axe that is made from yew and is approx and is 60 cm long, Longbow, arrows and quiver that he used to hunt, Flint dagger that had 13 cm, Retoucheur that he used to sharp his tools and Stone disc. 61 tattoos were found on Otzi's body, scientists think that the tattoos were made to alleviate the pain from his injuries.

---

### Exemplo 14

---

Texto: y10\_st5\_text2.txt.out

Escore: 3

---

There is a river in Peru that proves this "miracle" and many scientists tried to figure out what actually happened in that phenomena. A spanish myth was that men who came back from adventures in the forest told stories of poisoned-water, man-eating snakes and a river that boiled from below.

---

### Exemplo 15

---

Texto: y9\_st29\_text10.txt.out

Escore: 4

---

Music has been along the side of humanity since the beginning of human civilization. According to scientists the first imitation of a sound was performed during prehistory, through grunts and thumps on stones and wood. These ancient people believed that by doing this they were able to possess the soul of the animal they had hunted.

---

As variáveis lexicais do polo negativo da Dimensão 3, por exemplo, *money*, *school*, *student*, *mon*, *pay*, *change* e *face*, são indicativas de um discurso centrado na vida contemporânea e nos seus os desafios, especialmente relacionados ao dinheiro. Dessa forma, o polo negativo da Dimensão 3 recebeu o rótulo “Navegando pela identidade, transformação e resiliência na sociedade contemporânea”.

O Exemplos 16 aponta para as mudanças de rotina que ocorrem com a idade, como ter mais liberdade, mas também as responsabilidades, com escola e emprego para ganhar dinheiro (*Some old routines may change to fit the girls' new ways and age*). No Exemplo 17,

pode-se observar a história de superação e conquista do jogador de basquete americano Shaquille O'Neal, que precisou enfrentar o racismo e também dificuldades financeiras (*Shaq had to overcome racism, living in sordid houses and apartments and moving a lot because his parents couldn't pay rent*).

#### Exemplo 16

---

Texto: y8\_st7\_text7.txt.out

Escore: -3

---

As an example, Mi Cha states she doesn't need to ask for permission to do things such as leave the **house** at night. Some old routines may **change** to fit the **girls'** new ways and age. Cora has the responsibility of entering her old **school** format, a Canadian **school**, and Mi Cha needs to start her job search in order to make **money** for a living.

---

#### Exemplo 17

---

Texto: y7\_st5\_text3.txt.out

Escore: -2

---

Shaq had to overcome racism, living in sordid **houses** and apartments and moving a lot because his parents couldn't **pay** rent. He also had to overcome growing up with no **money** in poor education facilities, conflicts with teammates and NBA staff, among other problems. But Shaq pulled through, he used training to forget about his problems. After many years of training he was finally eligible for the NBA! Where his basketball career took off. I did move a lot in my life but never did I live in poor neighborhoods or **face** any of the other obstacles he had to overcome.

---

No Exemplo 18, são discutidas as vantagens e desvantagens de se fazer compras online, atividade que faz parte da vida contemporânea. Contudo, a discussão sobre os prós e contra desse modo de consumo é centrada no dinheiro gasto ou economizado (*That's because the store in person has to maintain the decoration, pay employees, pay water, rent, light, etc, so the price naturally ends up bigger*). Desse modo, o polo negativo da Dimensão 3 é marcado por um discurso voltado para os desafios da vida contemporânea, ressaltando como a falta de dinheiro pode ser um obstáculo, assim como o dinheiro pode ser uma solução ou influenciar na tomada de decisões, como no Exemplo 18.

## Exemplo 18

---

Texto: y6\_st63\_text1.txt.out

Escore: -4

---

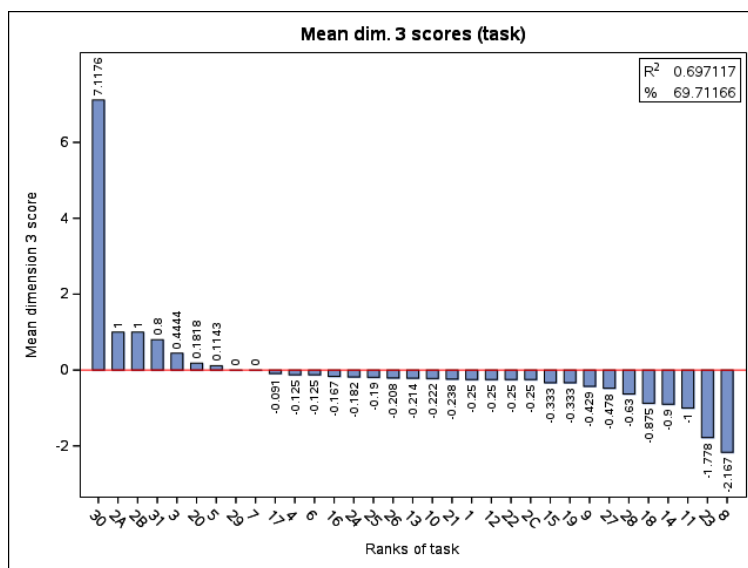
A con of online shopping is having to **pay** for the freight. But to compensate, the item online itself is cheaper. That's because the store in person has to maintain the decoration, **pay** employees, **pay** water, rent, light, etc, so the price naturally ends up bigger. The difference between prices online and in - person is not huge but is almost always present. The freight, on the other hand, might even be for **free**. So you end up **paying** the same or even saving **money** buying online.

---

Os discursos subjacentes na Dimensão 3 “Relatos históricos e descobertas arqueológicas versus Identidade, transformação e resiliência na sociedade contemporânea” são centrados em narrativas sobre a experiência humana, no entanto, cada polo foca em um tempo distinto. O polo positivo concentra os relatos em histórias da pré-história, misturando descobertas científicas com narrativas, por um outro lado, o polo negativo destaca os desafios e superações da vida moderna.

Sobre as tarefas, como pode ser observado na Figura 5, essa variável é responsável por uma grande quantidade de variação da Dimensão 3, ( $R^2 = 69\%$ ). Os escores médios mais extremados são das tarefas 30, 23 e 8. No polo positivo, a tarefa 30, realizada pelos alunos do Sexto Ano do Ensino Fundamental II, envolveu a escrita de uma narrativa de relatos da pré-história, correspondendo assim à temática do polo positivo. Já no polo negativo, as tarefas 23 e 8 foram realizadas, respectivamente, pelos alunos do Oitavo e Nono Ano do Ensino Fundamental II; a tarefa 23 tem como proposta a escrita de um texto comparativo entre a história de vida de duas personagens e a proposta da tarefa 8 envolveu a escrita de um artigo sobre as expectativas pessoais que os alunos têm sobre como seria morar fora da casa dos pais. Essas tarefas correspondem à temática do polo negativo.

Figura 5 - Escores médios das tarefas na Dimensão 3



Fonte: a autora

#### 4.1.1.4 Dimensão 4 - Autonomia cultural, diversidade e respeito versus Vida cotidiana e experiências de aprendizagem

A quarta dimensão foi rotulada “Autonomia cultural, diversidade e respeito versus Vida cotidiana e experiências de aprendizagem”. As variáveis lexicais de cada polo são apresentadas na Tabela 10 e alguns exemplos são analisados e discutidos a seguir.

Tabela 10 - Variáveis lexicais da Dimensão 4

Polo	Variáveis lexicais
<b>Positivo</b>	interfere_v (1,34), cultural_j (1,24), divorce_n (1,07), respect_v (1,01), culture_n (0,94), religion_n (0,90), boyfriend_n (0,76), treat_v (0,59), (personal_j (0,51)), (essential_j (0,46)), (possible_j (0,42)), (lie_n (0,41)), (uniform_n (0,40))
<b>Negativo</b>	head_n (-0,66), sleep_v (-0,63), real_j (-0,62), (torment_v (-0,62)), (accountable_j (-0,61)), (arrive_v (-0,59)), (yearn_v (-0,52)), begin_v (-0,52), keep_v (-0,47), (main_j (-0,47)), great_j (-0,41), day_n (-0,39), little_j (-0,37), see_v (-0,34), (teach_v (-0,33)), part_n (-0,33), (prove_v (-0,32)), (play_v (-0,32)), last_j (-0,31), (year_n (-0,31)), take_v (-0,30)

Fonte: a autora

Ao se analisar as variáveis lexicais do polo positivo da Dimensão 4, pode-se perceber que o discurso é centrado no respeito à diversidade, pois as variáveis com maior peso são *interfere*, *cultural*, *divorce*, *respect*, *culture* e *religion*. Por exemplo, no trecho do Exemplo 19, é discutido o respeito que os governos precisam ter por diferentes culturas, não interferindo

com costumes e tradições, principalmente religiosas (*No government should interfere in any cultural or religious aspect of any person*). Além disso, há a visão de que os governos deveriam manter-se neutros, não defendendo uma cultura em detrimento de outra (*First, the government should not defend any kind of religion, nor get involved in any kind of religion or culture, just respect*).

#### Exemplo 19

---

Texto: y8\_st33\_text2.txt.out

Escore: 8

---

No government should **interfere** in any **cultural** or religious aspect of any person, no one must say what or what not should she use, this can only be determined by the family or by the person. All **cultural** or religious aspects that the entire population must be **respected**, both by the government and by other **cultures**. Whoever follows his **religion** not only in France, but also in other countries, is the person who is practicing. First, the government should not defend any kind of **religion**, nor get involved in any kind of **religion** or **culture**, just respect. Even having different opinions, it 's okay, just **respect** it.

---

O Exemplo 20 também coloca em debate o respeito, porém, este está relacionado às escolhas de cada indivíduo de mudar de cultura, o que é chamado de “divorcio cultural” (*cultural divorce*). Assim, a cultura é apresentada como uma escolha individual e não estável, pois as pessoas teriam autonomia para mudar de acordo com a idade (*The tendency towards culture can change with age*) ou mesmo para conseguirem ingressar a uma sociedade (*the decision to divorce from it is essential for that individual to be able to enter society*). Ainda, pode-se observar que o aluno expressou sua opinião acerca do tema (*I believe that...*) e enfatizando a importância de se permitir que indivíduos tenham o espaço para explorar, definir e expressar livremente suas identidades culturais e religiosas.

#### Exemplo 20

---

Texto: y8\_st17\_text3.txt.out

Escore: 9

---

So we must **respect** the other's choice and always try to understand the reason for the choice, we can't force him to accept but we can listen to the reason for that decision, we shouldn't **interfere** but talking and discussing is always a good option. The tendency towards **culture** can change with age, and sometimes the decision to divorce from it is **essential** for that individual to be able to enter society. I believe that **cultural divorce** despite having a strong name can always be a good option, but we can go beyond just **cultural divorce** and create bonds with new **cultures** and **religions**.

---

No Exemplo 21, a cultura não é descrita apenas pela religião de uma determinada sociedade, mas também pela culinária, música, roupas e literatura (*Food, clothes, music, religion. All of these can include in culture*). Dessa forma, a cultura é fundamental para as sociedades, pois é uma forma de representação e de conexão entre as pessoas, sendo importante o seu respeito e defesa (*Brazilian culture helps people connect with each other in a peaceful way, it's something that people could criticize you for, so you need to defend it*). Diferente dos outros exemplos apresentados, o Exemplo 21 apresenta a cultura como algo coletivo e não como individual, como em *I am capable of doing a list of why culture is essential in our society, start saying that culture represents 50 % of who we are*.

#### Exemplo 21

---

Texto: y8\_st32\_text2.txt.out

Escore: 2

---

It's a right, but also a **culture**. Food, clothes, music, **religion**. All of these can include in **culture**. The essence of a **culture** is what makes us different from other places. Do you know how important **culture** is? **Culture** is something really important in people's lives, for example, the Brazilian **culture** has different types of music, food, and literature, which are really important for Brazilians. Brazilian **culture** helps people connect with each other in a peaceful way, it's something that people could criticize you for, so you need to defend it. As we saw in France with the hijab prohibition. **Culture** is important, and we should always defend, and a good society should ALWAYS **respect** it. I am capable of doing a list of why **culture** is **essential** in our society, start saying that **culture** represents 50 % of who we are.

---

Analisando o polo negativo da Dimensão 4, rotulado como "Vida cotidiana e experiências de aprendizagem", percebe-se que as variáveis associadas a esse fator, como *head, sleep, real, torment, accountable, arrive, yearn, begin*, concentram-se principalmente em atividades cotidianas, experiências e eventos. As variáveis sugerem um foco nos aspectos comuns ou rotineiros da vida, incluindo tarefas diárias, ações simples e experiências, além do aprendizado, do crescimento e das reflexões pessoais que surgem ao interagir com o mundo cotidiano.

O Exemplo 22 é um relato de uma viagem em família com a descrição dos acontecimentos e sentimento dos personagens (*The fly turbulence tormented us*). Há, também, no relato a sequência temporal em que os fatos ocorreram (*It all started in the last week of our trip*) e como os personagens foram afetados pela experiência (*I regret saying I still wanted to go to the roller coast after all. I will never forget seeing my family with their heads down, and really sad*).

## Exemplo 22

---

Texto = y7\_st29\_text1.txt.out

Escore = -11

---

The fly turbulence **tormented** us, **part** of my family was **sleeping** but I was not, and every time I travel in an airplane I **keep** thinking what can happen, which is not good for me. It all started in the **last** week of our trip (...) The park was **accountable** for this because they didn't have a good security in this E.T. toy, they gave to each person of the family a ticket for free, and with a fast pass for all, so I think one thing in this story was satisfactory. I regret saying I still wanted to go to the roller coast after all. I will never forget **seeing** my family with their **heads** down, and really sad.

---

Os textos com maior escore no polo negativo da Dimensão 4 não são apenas narrativas sobre experiências do cotidiano. O Exemplo 23 apresenta características de um texto injuntivo, usando uma linguagem objetiva (*Take the sketch and start making the details of the body such as the nose, eyes, clothes, fingers, hair, eyebrows, mouth and shoes*) e orientando o leitor a como fazer um desenho (*Today I will teach you how to design and detail the centerpiece of your artwork*). Ainda, o Exemplo 24 apresenta dados sobre poluição marinha, mais especificamente, sobre como o plástico é um dos maiores poluentes, o que faz com que o texto tenha características de um relatório de pesquisa. A relação deste exemplo com o polo negativo da Dimensão 4 é como ações do cotidiano impactam os oceanos, como no seguinte trecho “*a person consumes 150 bags of plastic per year, a lot of them end up in the ocean, and all of this could be stopped by simple changes in our day a day*”.

## Exemplo 23

---

Texto: y7\_st4\_text2.txt.out

Escore: -8

---

Character design can be simple and complex, it depends on the detail that you are putting out for the public to **see**. These **days** I have recognized the fact that a lot of the designs that are being made or remade are very simplistic, which is the case for cartoon characters. Today I will **teach** you how to design and detail the centerpiece of your artwork. (...) Then you start to define and solidify the character onto paper. Well how do you do that? The process is a **little** harder than before, but it is nothing extraordinary. **Take** the sketch and start making the details of the body such as the nose, eyes, clothes, fingers, hair, eyebrows, mouth and shoes. So let's start with the nose, the nose is located a **little** bit below the middle of the circle used in the sketch and it has the shape of an upside down seven, although there are different ways to make it, that is my favorite.

---

## Exemplo 24

---

Texto: y10\_pw1\_17\_text.txt.out

Escore: -7

---

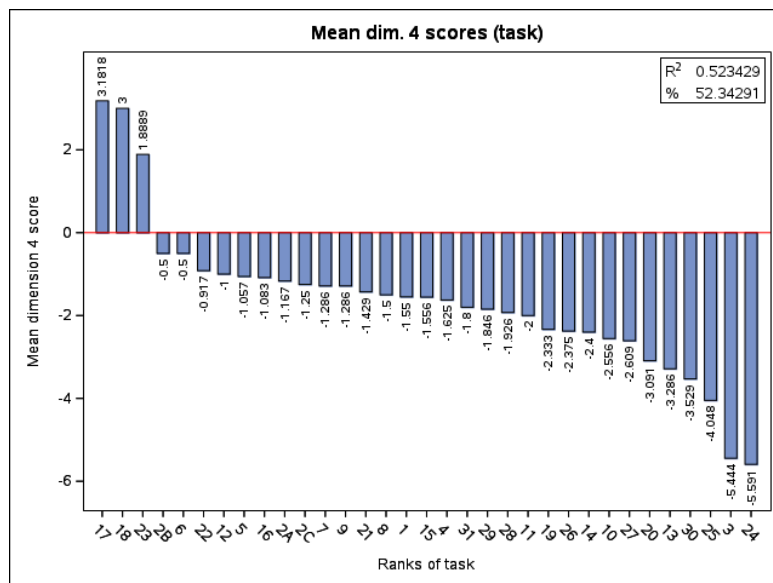
Every human should be alarmed about the situation that the ocean is passing, matching with our data that presents that the ocean covers about 71 percent of the Earth's surface. While doing the research we found out that, in general, a person consumes 150 bags of plastic per year, a lot of them end up in the ocean, and all of this could be stopped by simple changes in our day a day. Ocean is vital for us because provides half of the earth oxygen and plastic is one of its main pollutants, so on we should reduce the use of plastic straws that is one of the most accessible things but also one of the most prejudicial.

---

Os discursos da Dimensão 4 “Autonomia cultural, diversidade e respeito versus Vida cotidiana e experiências de aprendizagem” são marcados pela experiência humana em ambos os polos. No entanto, no polo positivo, enfatiza-se a diversidade cultural, escolhas pessoais e o respeito, enquanto que no polo negativo o foco está nas atividades e nas experiências cotidianas, destacando os aspectos rotineiros.

Em relação às tarefas, essa variável é responsável por uma grande quantidade de variação da Dimensão 4 ( $R^2 = 52,34\%$ ), como ilustrado pela Figura 6. Os escores médios mais extremados são das tarefas 17, 18, 3 e 24. No polo positivo, as tarefas 17 e 18 foram realizadas pelos alunos do Oitavo Ano do Ensino Fundamental II, sendo a tarefa 17 sobre a escrita de um artigo sobre diversidade cultural e a tarefa 18 sobre “divórcio cultural”, isto é, sobre se divorciar de uma cultura, correspondendo assim à temática do polo positivo. No polo negativo, as tarefas 24 e 3 foram realizadas, respectivamente, pelos alunos do Sétimo Ano do Ensino Fundamental II e Primeiro Ano do Ensino Médio; a tarefa 24 tem como proposta a escrita de uma narrativa pessoal descritiva e a proposta da tarefa 3 envolveu a escrita sobre um assunto de interesse dos alunos que fosse um problema, cuja possível solução também é apresentada no texto. Essas tarefas correspondem à temática do polo negativo..

Figura 6 - Escores médios das tarefas na Dimensão 4



Fonte: a autora

#### 4.1.1.5 Dimensão 5 - Resiliência, determinação e triunfo socioeconômico versus Autonomia e expressão individual em ambientes estruturados e digitais

A Tabela 11 apresenta as variáveis lexicais que compuseram o quinto fator da extração fatorial, em dois polos distintos: “Resiliência, determinação e triunfo socioeconômico versus Autonomia e expressão individual em ambientes estruturados e digitais”.

Tabela 11 - Variáveis lexicais da Dimensão 5

Polo	Variáveis lexicais
<b>Positivo</b>	laborious_j (1,18), eligible_j (1,16), formidable_j (1,11), component_n (1,01), sordid_j (0,99), player_n (0,85), overcome_v (0,83), success_n (0,73), successful_j (0,72), obstacle_n (0,72), achieve_v (0,63), team_n (0,56), (college_n (0,53)), face_v (0,50), (bear_v (0,48)), (goal_n (0,42)), become_v (0,42), famous_j (0,40), hard_j (0,34)
<b>Negativo</b>	clothes_n (-0,78), nature_n (-0,58), (sleep_v (-0,57)), free_j (-0,50), (cultural_j (-0,47)), (platform_n (-0,42)), uniform_n (-0,42), (interfere_v (-0,39)), (right_j (-0,34))

Fonte: a autora

O polo positivo da Dimensão 5 recebeu o rótulo “Resiliência, determinação e triunfo socioeconômico”, pois as variáveis lexicais com maior maior peso, por exemplo *laborious*, *eligible*, *formidable*, *component*, *sordid*, *player*, *overcome*, *success* e *obstacle*, refletem um discurso centrado na superação de adversidades. Ademais, as variáveis indicam que os textos

desse polo têm uma grande densidade de informação, uma vez que mais da metade das variáveis são adjetivos (7) e substantivos (7), sendo que a maioria dos adjetivos é atributiva, pois “modificam expressões nominais, antecedendo o substantivo principal ou pronome” (Biber et al., 2021, p. 508). A maior frequência de adjetivos e substantivos sugerem características de notícias e textos acadêmicos (Biber et al., 2021).

Os textos com maior escore no polo negativo da Dimensão 5 contam a história de vida e superação de atletas famosos, como os jogadores de futebol, ginastas e jogadores de basquete, como o Exemplo 25. Neste exemplo, o sucesso do jogador Shaquille O'Neal é atribuído a sua dedicação aos treinos (*His laborious work paid off; After many years of training he was finally eligible for the NBA*), além disso, o jogador precisou superar obstáculos financeiros.

Os obstáculos superados também estão relacionados à questões físicas, como o Exemplo 26 que apresenta a história da escritora Helen Keller. Helen conseguiu superar desafios relacionados a gênero, pois o acesso à educação superior era restrito para mulheres no início do século XX (*College was a distant, formidable dream for young women at that time*) e também os desafios de se comunicar (*Helen faced many obstacles and challenges like communicating*).

#### Exemplo 25

---

Texto : y7\_st5\_text3.txt.out

Escore : 13

---

Shaq was tall but didn't play well so he started to train vigorously. His **laborious** work paid off, and many years later he made contracts with the NBA and with big companies that wanted him for advertisements, making him millions. Shaq was most commonly known as a basketball **player**, actor and singer making about \$ 60 million dollars a year, but now Shaq is retired and has been since 2011. (...) After many years of training he was finally **eligible** for the NBA! Where his basketball career took off. I did move a lot in my life but never did I live in poor neighborhoods or **face** any of the other **obstacles** he had to **overcome**. He **became** a very **formidable player** that inspired many others.

---

#### Exemplo 26

---

Texto: y7\_st47\_text2.txt.out

Escore: 9

---

After many years, Helen **became** a well-educated young woman and decided she wanted to go to **college**. **College** was a distant, **formidable** dream for young women at that time, especially blind and deaf women, but Helen did it anyway. In 1904 , Helen **became** the first blind and deaf person to graduate from Radcliffe **College** with a Bachelor Arts Degree. Even

---

---

though Helen lived in a silent and dark world, she never gave up and discovered that there was so much to learn about the world outside of her own. Helen faced many **obstacles** and challenges like communicating, which comes to us as easily and fast as blinking.

---

O discurso acerca da superação de obstáculos, sejam eles materiais ou psicológicos, focaliza no merecimento, como no Exemplo 27 (*To achieve success you need to deserve it*). Ademais, alcançar o sucesso seria uma questão de determinação (*Determination is the key to overcoming obstacles*) e uma ação individual (*Achieving success is the individual's determination to do his job correctly and well planned*). Ao se analisar os exemplos, pode-se concluir que os textos apresentam uma visão restrita de sucesso, pois aproximando-se de ideias meritocráticas.

#### Exemplo 27

---

Texto: y9\_st11\_text5.txt.out

Escore: 8

---

To **achieve success** you need to deserve it. And **success** makes everything easier. **Success** is not easily **achieved**. It is a group of aspects that you need to work together, to have the right mindset, the right positions and the right attitude for you to be able to **achieve** your **goals**. Determination is the key to **overcoming obstacles**. **Achieving success** is the individual's determination to do his job correctly and well planned.

---

O polo negativo da Dimensão 5 foi rotulado “Autonomia e expressão individual em ambientes estruturados e digitais”, pois destaca a tensão entre liberdade pessoal e normas institucionais ou sociais. A análise das variáveis com maiores cargas, como *clothes, nature, sleep, free, cultura, platform* e *uniform*, aponta para discussões sobre uniformidade e expressão individual, o impacto de plataformas digitais no aprendizado e no comércio, e as implicações culturais e naturais mais amplas dessas dinâmicas. Além disso, a maioria das variáveis são substantivos e adjetivos o que indica que os textos têm uma maior densidade de informação, aproximando-se de notícias e escrita acadêmica, assim como o polo positivo da dimensão.

Muitos dos textos com maior escore no polo negativo são sobre as vantagens e desvantagens de se usar uniformes escolares, como o Exemplo 28. Dentre as vantagens, o fato de que um uniforme fazer com que todos os alunos fiquem com a mesma aparência ajudaria na segurança, pois diminuiria os casos de bullying (*Uniforms can reduce the potential for bullying; But unfortunately a lot of people don't respect the way that people like to dress up*). Além de ser uma forma segura de identificar quem são os alunos de uma

determinada escola (*The colors of the uniforms make it easy to identify who is a student and who is not*).

#### Exemplo 28

---

Texto: y6\_st34\_text1.txt.out

Escore: -4

---

I really think that **uniforms** are important to students and to the school. These are the reasons. **Uniforms** can reduce the potential for bullying. When students are not wearing **uniforms**, they like to use **clothes** that make them feel happy and express themselves in what they are feeling. But unfortunately a lot of people don't respect the way that people like to dress up. But by using the same cloth there are less chances of there being bullying. **Uniforms** can increase safety. The colors of the **uniforms** make it easy to identify who is a student and who is not. (...) Finally, **uniforms** can make mornings easy for students and parents. Imagine having to choose the **clothes** you will wear every day. **Uniforms** make it easier for you to do more things and give you more **free** time. So that's why I think that in schools it should be mandatory to wear a **uniform**.

---

Já o Exemplo 29 estende a discussão para diversidade cultural e preconceito, enfatizando a importância da expressão cultural. Diferentemente dos textos sobre o uso de uniformes, em que a diversidade é colocada como um motivo para bullying, no Exemplo 29 a diversidade é apresentada como importante para a expressão cultural e as diferenças devem ser respeitadas (*Different styles of clothes, language or living shouldn't be reason of discrimination*).

#### Exemplo 29

---

Texto : y8\_st13\_text2.txt.out

Escore : -3

---

It is not unusual, for people to compare different cultures, with different values, ideals and religions, however, for our society to grow, it would be important if we would just defend all **cultural** diversities as, according to the United Nations, this diversity is an "ethical imperative, inseparable from respect and from human dignity". It is part of our **nature** to fear what we don't know and, therefore, fall for certain images about anything that challenges, contradicts, or differs what we agree or believe, a clear example of this, is the Copernican Revolution, when the Church didn't let the astronomers publish their discoveries.(...) Different styles of **clothes**, language or living shouldn't be reason of discrimination. What is considered right to some, may be considered completely wrong for others.

---

O Exemplo 30 aborda as vantagens e desvantagens de se fazer compras online, que não estão diretamente relacionadas aos contextos institucionais, mas aborda de forma indireta a liberdade de escolha e conveniência em comparação às normas tradicionais de comércio.

Além disso, levando-se em consideração que os textos do corpus CABI foram produzidos em 2021, ou seja, em meio da Pandemia de Covid-19, o fato de não ser necessário o contato com outras pessoas é visto como uma vantagem.

### Exemplo 30

---

Texto: y6\_st63\_text1.txt.out

Escore: -2

---

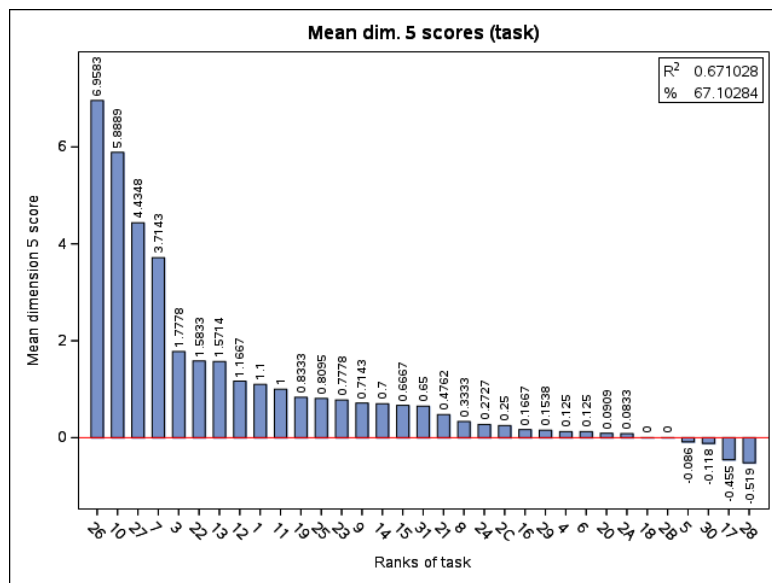
In online shopping, you only have to sanitize your item, while going to the store in person, you have to be careful with everything around you, and you might as well have contact with a seller, which is risky. Some people rather go to the stores, because they can try on their clothes and make sure they fit, for example. (...) The difference between prices online and in - person is not huge but is almost always present. The freight, on the other hand, might even be for free. So you end up paying the same or even saving money buying online.

---

Os polos da Dimensão 5 “Resiliência, determinação e triunfo socioeconômico versus Autonomia e expressão individual em ambientes estruturados e digitais” apresentam uma alta densidade de informação, com uma distribuição significativa de adjetivos e substantivos, o que sugere que os textos da dimensão tendem a ter mais detalhes e descrições. Além disso, os discursos de ambos os polos apresentam o enfrentamento de desafios. Contudo, enquanto o polo positivo é marcado por uma ênfase em qualidades pessoais como resiliência e determinação visando o sucesso socioeconômico, o polo negativo valoriza a autonomia e a expressão individual, muitas vezes em oposição às normas sociais ou institucionais.

Quanto às tarefas, como demonstra a Figura 7, essa variável é responsável por uma grande quantidade de variação da Dimensão 5 ( $R^2 = 67,10\%$ ). Os escores médios mais extremados são das tarefas 26, 10, 17 e 28. No polo positivo, a tarefa 26, realizada pelos alunos do Sétimo Ano do Ensino Fundamental II, envolveu a escrita de um texto sobre a superação de obstáculos, contando com exemplos de uma pessoas famosos e exemplos pessoais e, na tarefa 10, os alunos do Nono Ano do Ensino Fundamental II escreveram sobre o sucesso e as principais qualidades que uma pessoa bem sucedida tem, correspondendo assim à temática do polo positivo. No polo negativo, as tarefas 17 e 28 foram realizadas, respectivamente, pelos alunos do Oitavo e Sexto Ano do Ensino Fundamental II, sendo a tarefa 17 a escrita de um artigo sobre diversidade cultural e a tarefa 28 de um texto persuasivo. Essas tarefas correspondem à temática do polo negativo.

Figura 7 - Escores médios das tarefas na Dimensão 5



Fonte: a autora

#### 4.1.1.6 Dimensão 6 - Fenômenos naturais e recursos hídricos versus Crescimento pessoal e interação social

A sexta dimensão foi rotulada “Fenômenos naturais e recursos hídricos versus Crescimento pessoal e interação social”. As variáveis com as maiores cargas são apresentadas na Tabela 12 e os discursos subjacentes de cada polo da Dimensão 6 são discutidos com alguns exemplos a seguir.

Tabela 12 - Variáveis lexicais da Dimensão 6

Polo	Variáveis lexicais
<b>Positivo</b>	boil_v (1,38), temperature_n (1,26), hot_j (1,07), river_n (1,06), exist_v (1,02), water_n (0,84), (area_n (0,59)), (preserve_v (0,43)), (truth_n (0,39)), (return_v (0,33)), (research_n (0,31))
<b>Negativo</b>	(respect_v (-0,60)), personal_j (-0,58), (clothes_n (-0,56)), (aspect_n (-0,54)), practice_v (-0,54), (increase_v (-0,49)), (player_n (-0,47)), wear_v (-0,45), (express_v (-0,44)), (team_n (-0,43)), (decision_n (-0,41)), music_n (-0,38), (mental_j (-0,37)), play_v (-0,36), (other_n (-0,35)), (change_v (-0,31)), (best_j (-0,30))

Fonte: a autora

As variáveis com maior carga no polo positivo da Dimensão 6, como *boil*, *temperature*, *hot*, *river*, *exist* e *water*, revelam um discurso centrado em fenômenos naturais, assim como na temperatura e na importância da água como um recurso e sua conservação.

Desse modo, o polo positivo da Dimensão 6 foi rotulado “Fenômenos naturais e recursos hídricos”.

Nos Exemplos 31 e 32, observa-se como os estudantes tentaram explicar o funcionamento de um fenômeno natural, mais especificamente, a explicação do motivo da água do rio Shanay-Timpishka ferver. Dessa forma, os dois exemplos propõem fazer a descrição de como o fenômeno ocorre (*In order to understand what happens when rivers boil, you need to understand that heat travels from areas with warmer temperatures to areas with cooler temperatures*), sendo que o Exemplo 31 baseia-se no percurso investigativo de Andrés Ruzo (*Andrez Ruzo had so many doubts and even his aunt had already swum in the river. Ruzo, in 2011, finally decided to take a risk and actually go to the Amazon rainforest*).

#### Exemplo 31

---

Texto: y10\_st5\_text2.txt.out

Escore: 11

---

After talking to professionals, he concluded that they did **exist** and were generally associated with volcanoes. However, there were no volcanoes in that **area**, so how could that **river exist**? Andrez Ruzo had so many doubts and even his aunt had already swum in the **river**. Ruzo, in 2011, finally decided to take a risk and actually go to the Amazon rainforest with his aunt to investigate and see the **river** with his own eyes. To his surprise, the **water** was steaming **hot**.

---

#### Exemplo 32

---

Texto: y10\_st1\_text2.txt.out

Escore: 9

---

But this phenomenon is actually a natural occurrence that happens when the air **temperature** is much colder than the **water**. In order to understand what happens when **rivers boil**, you need to understand that heat travels from **areas** with warmer **temperatures** to **areas** with cooler **temperatures**.

---

O Exemplo 33 também tem o seu discurso centrado na água, mas como um recurso e sua importância para uma vida saudável. Assim, formas de purificação da água são apresentadas como uma solução para a falta de distribuição de água filtrada pelo governo no Brasil (*The focus of this paper is to detail existing water filtration and purification methods in order to find out which of them can serve as a means to making clean water more accessible for the poor population*). Ademais, assim como nos Exemplos 31 e 32, o Exemplo 32 também é marcado pela descrição, mas o texto foca nos diferentes métodos de filtração de água.

### Exemplo 33

---

Texto: y10\_pw2\_13\_text.txt.out

Escore: 4

---

Whilst there is no effective government action to resolve this health risk, we have investigated **water** filters as a temporary solution to improve local health in the Brazilian slums. The focus of this paper is to detail **existing water** filtration and purification methods in order to find out which of them can serve as a means to making clean **water** more accessible for the poor population. (...) The first **water** form of **water** filtration emerged in Greece, around 2000 B.C. with practices such as **boiling** and straining their **water** to eliminate impurities.

---

No polo negativo da Dimensão 6, as variáveis com maior carga são *respect, personal, clothes, aspect, practice, increase, wear* e *express* e a análise dos textos de maior escore indica um discurso que enfatiza o impacto positivo e negativo de escolhas pessoais que envolvem esporte e música, por exemplo, tem na vida das pessoas. Assim, o polo negativo da Dimensão 6 recebeu o rótulo “Crescimento pessoal e interação social”.

O Exemplo 34 discute os benefícios e malefícios de se praticar boxe. Contudo, o esporte pode trazer consequências positivas ou negativas a depender de quem está o praticando (*It should be taken into consideration, the fact that these benefits and detriments may vary from person to person, and these aspects will be discussed in the essay*). Para ilustrar os pontos positivos e negativos, os estudantes usaram exemplos de pessoas que praticaram o esporte, como em *This clearly shows that boxing changed his life, and possibly saved many lives as well*. Ademais, a prática do esporte traria benefícios para o corpo (*The sport increases a person's overall body strength and their muscle mass; however practicing the sport in a non-professional manner benefits a person's mental and physical well being*).

### Exemplo 34

---

Texto: y10\_pw3\_5\_text3.txt.out

Escore: -7

---

This clearly shows that boxing **changed** his life, and possibly saved many lives as well. And Brad Schwartz, a physician in Kansas, commented that after his first class of boxing, he noticed that the workout involves and works the entire body. This debate could come to an end by gathering evidence and arguments for both sides, and in the end compare them to conclude if boxing actually causes more benefits or if it is more harmful. It should be taken into consideration, the fact that these benefits and detriments may vary from person to person, and these **aspects** will be discussed in the essay. (...) It improves a person's cardiovascular health as well as their hand - eye coordination. The sport **increases** a person's overall body strength and their muscle mass. Nonetheless, boxing improves a person's stamina and endurance, while decreasing stress, leading to better sleep quality and improved

---

---

confidence. Professional boxing may be considered to be dangerous by some people, eventually leading to serious injuries, however practicing the sport in a non-professional manner benefits a persons' mental and physical well being.

---

O Exemplo 35 ilustra como a música clássica pode afetar as pessoas e a escolha de se ouvir determinado tipo de música teria impactos na vida pessoal e profissional (*It's not just about choosing a type of music. It's about changing your personal and professional life*). Assim, a música clássica ajudaria as pessoas a manter a concentração (*Classical Music helps you concentrate*), sendo relevante para as diferentes idades (*I'll show you how classical music can be worthwhile for all ages*).

#### Exemplo 35

---

Texto: y10\_st7\_text1.txt.out

Escore: -2

---

But today, I'll show you how classical music can be worthwhile for all ages. It's not just about choosing a type of music. It's about changing your personal and professional life. Classical Music helps you concentrate This one is the easiest to explain. It is common sense that classical music can help you concentrate.

---

O Exemplo 36 aborda as vantagens e desvantagens do uso de uniformes escolares. Um dos argumentos contra o uso do uniforme é que ele seria um limitador da expressão pessoal (*the use of uniforms can limit personal expression*) e que não permite que as pessoas possam fazer escolhas (*they can't make decisions regarding what to wear*). Um dos argumentos a favor seria que as distrações dos alunos diminuam, contribuindo para um melhor desenvolvimento acadêmico (*use of uniforms increases academic development, as students do not have distractions*).

#### Exemplo 36

---

Texto: y6\_st70\_text1.txt.out

Escore: -3

---

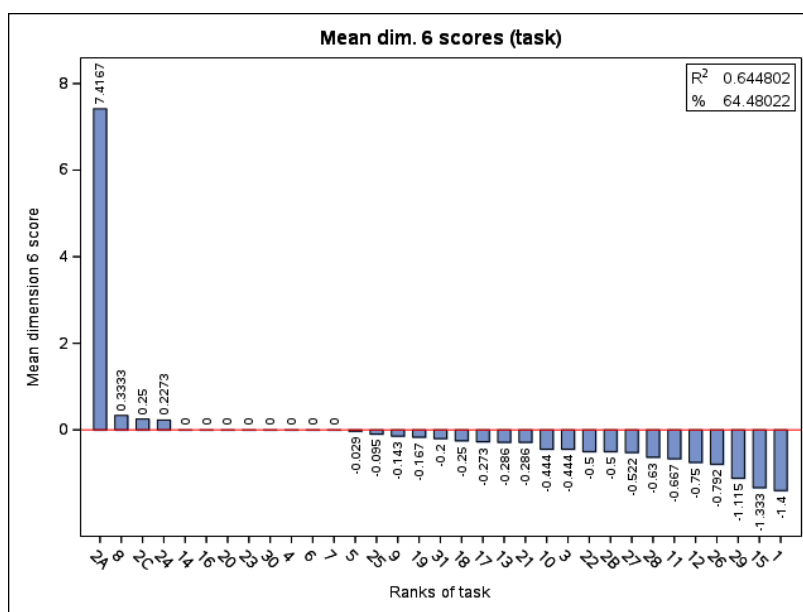
In conclusion, wearing uniforms is a debatable topic that has a lot of pros and cons. But, the use of uniforms can limit personal expression, as all students need to use the same clothes, they can't make decisions regarding what to wear. Also, school uniforms are normally expensive, as said in the third paragraph. And, even though supporters of uniforms say that the use of uniforms increases academic development, as students do not have distractions, people who are against affirms that the uniforms are not comfortable and adjustable to different weather conditions, adding another distraction.

---

Assim, pode-se concluir que os polos da Dimensão 6 “Fenômenos naturais e recursos hídricos versus Crescimento pessoal e interação social” utilizam exemplos concretos para ilustrar os pontos discutidos e abordam questões que impactam diretamente a sociedade. Por outro lado, no polo positivo, há uma ênfase nos fenômenos naturais e na importância dos recursos hídricos para a sobrevivência e qualidade de vida. Enquanto, no polo negativo, o foco está no crescimento pessoal e na interação social, destacando como escolhas pessoais afetam a vida das pessoas.

Sobre as tarefas, como pode ser observado na Figura 8, essa variável é responsável por uma grande quantidade de variação da Dimensão 6 ( $R^2 = 64,48\%$ ). Os escores médios mais extremados são das tarefas 2A, 29, 15 e 1. No polo positivo, a tarefa 2A, realizada pelos alunos do Primeiro Ano do Ensino Médio, envolveu a escrita de um texto que respondesse à seguinte pergunta: “Por que os rios fervem?”. No polo negativo, as tarefas 29, 15 e 1 foram realizadas, respectivamente, pelos alunos do Sexto Ano, Nono Ano e Primeiro Ano do Ensino Médio. A tarefa 29 tem como proposta a escrita de uma narrativa biográfica sobre um personagem fictício, enquanto a proposta da tarefa 15 envolveu a escrita de um texto de opinião sobre a importância da música, assim como a tarefa 1, a qual demandou a escrita de um artigo sobre música.

Figura 8 - Escores médios das tarefas na Dimensão 6



Fonte: a autora

#### 4.1.1.7 Dimensão 7 - Iniciativas abrangentes de saúde e engajamento nas escolas versus Esforços pessoais e considerações éticas na conquista de objetivos

A análise das variáveis lexicais e dos exemplos com maior escore e cada polo sugere que a Dimensão 7 é marcada por discursos sobre ações relacionadas à saúde no ambiente escolar e histórias de superação. Assim, a dimensão foi rotulada como “Iniciativas abrangentes de saúde e engajamento nas escolas versus Esforços pessoais e considerações éticas na conquista de objetivos”. Pode-se observar que as variáveis lexicais de ambos os polos são majoritariamente substantivos, como apresentado Tabela 13, o que é um indicativo de textos com uma maior densidade de informação (Biber, 1988).

Tabela 13 - Variáveis lexicais da Dimensão 7

Polo	Variáveis Lexicais
<b>Positivo</b>	policy_n (1,40), process_v (1,25), implement_v (0,98), thank_v (0,86), student_n (0,72), period_n (0,69), school_n (0,60), (change_n (0,54)), (healthy_j (0,53)), week_n (0,49), (impact_n (0,47)), (food_n (0,44)), (health_n (0,42)), (painting_n (0,39)), (represent_v (0,36))
<b>Negativo</b>	right_j (-0,70), goal_n (-0,52), (dad_n (-0,45)), (man_n (-0,43)), (hot_j (-0,41)), (society_n (-0,39)), (website_n (-0,37)), (little_j (-0,36)), (win_v (-0,35)), (temperature_n (-0,33)), (book_n (-0,31)), (proces_n (-0,31))

Fonte: a autora

O polo positivo da Dimensão 7 recebeu o rótulo “Iniciativas abrangentes de saúde e engajamento nas escolas”, uma vez que as variáveis lexicais com maior peso, como *policy*, *process*, *implement*, *student*, *period* e *school*, revelam um discurso predominantemente centrado em como as escolas são um ambiente que pode impactar a vida dos alunos. Sendo que o impacto pode ser fruto de uma política de alimentação ou mesmo da forma com que as aulas são conduzidas.

O Exemplo 37 é um excerto de uma palestra sobre as mudanças na política alimentar da escola (*That new policy demands that the school does not serve any kind of industrialized food, and only serves 100% organic food*). No texto, são apresentadas as mudanças, os seus motivos e benefícios, destacando a preocupação que a escola tem com a saúde dos alunos a longo prazo (*plus the school is concerned about our health and our future*). Contudo, ao mesmo tempo que a decisão será benéfica para os alunos, fica evidente que os alunos não participaram da tomada de decisão (*Also there will be no changes to what has been decided*).

### Exemplo 37

---

Texto: y9\_st33\_text7.txt.out

Escore: 9

---

That new **policy** demands that the **school** does not serve any kind of industrialized **food**, and only serves 100% organic **food**. (...) As I was saying, it is important to **implement** such modifications because the **school**'s nutritionist obeyed, plus the **school** is concerned about our **health** and our future. Maybe in a short-term solution industrialized **food** can be better, but in a long-term situation **healthy food** is two hundred times better than **processed food**. Remember, guys, I was not the one to **implement** these changes, it was the **school** board, and I will leave this very clear, this topic will not be discussed anymore. Also there will be no **changes** to what has been decided, I hope you guys do not be mad at me but I stand for everything I said, so it is done. Hope you guys liked it and **thank** you for listening to all this.

---

Outro exemplo de política escolar que impactou a vida dos alunos é a mudança repentina para o modo de ensino remoto, como ilustrado no Exemplo 38 (*The sudden change due to COVID - 19 made everyone surprised and not fully prepared to adapt to distance learning*). De acordo com o exemplo, o modo de estudo adotado pela escola durante a pandemia de COVID-19 associado à cobranças e o isolamento social contribuíram para que não apenas os alunos, mas também os professores, ficassem exaustos, estressados e com uma piora na saúde mental (*making students feel more overwhelmed and pressured. Our sources show that their mental health was deeply affected, causing a rise in depression, anxiety and stress levels among students and teachers*). Além disso, uma das soluções para os problemas acarretados pelas aulas remotas seria que as escolas tivessem escutado mais as devolutivas dos alunos (*The results show that a solution to the problem would be having schools listen and have an honest discussion with their students and staff, taking the feedback into account and changing the teaching method during distance learning*), o que também aponta para o não envolvimento dos alunos em decisões que impactam a vida escolar.

### Exemplo 38

---

Texto: y10\_pw6\_20\_text.txt.out

Escore: 4

---

The COVID - 19 pandemic has been affecting the lives of people around the world for almost two years. Since people had to adapt to quarantine, **schools** were closed and **students** of private high **schools** in Brazil were having online classes for an undetermined time. Eventually, **schools** sent more schoolwork at a faster pace, making **students** feel more overwhelmed and pressured. Our sources show that their mental **health** was deeply affected, causing a rise in depression, anxiety and stress levels among **students** and teachers. The amount of workload, pressure and external factors, such as lack of social interaction and the progression of the pandemic were factors that caused this worsening of their mental **health**.

---

---

The results show that a solution to the problem would be having **schools** listen and have an honest discussion with their **students** and staff, taking the feedback into account and changing the teaching method during distance learning. (...) The sudden **change** due to COVID - 19 made everyone surprised and not fully prepared to adapt to distance learning.

---

Já o Exemplo 39, por mais que não discuta como políticas escolares podem impactar os alunos, narra como o ambiente escolar e os amigos podem auxiliar os alunos a superar os desafios de uma adaptação a um novo ambiente de ensino. Por exemplo, como no trecho: “*On my first day at school I felt abandonment because I don't know anyone there, but there were people that helped me, and also new students too*”.

#### Exemplo 39

---

Texto: y7\_st29\_text4.txt.out

Escore: 3

---

On my first day at **school** I felt abandonment because I don't know anyone there, but there were people that helped me, and also new **students** too, in which I tried to stay the most because they were the people that knew what I was passing through. Then after a **week** I already knew all of the classmates, and was staying with all of them, mostly the girls, because I felt more comfortable.

---

O polo negativo da Dimensão 7 foi rotulado como “Esforços pessoais e considerações éticas na conquista de objetivos”, pois as variáveis lexicais, por exemplo *right, goal, dad, man, society* e *win*, sugerem que polo está centrado nas dificuldades e debates acerca da conquista de objetivos individuais e coletivos. Como no exemplo 40, em que são apresentados os argumentos a favor e contra a edição de genoma com foco em questões éticas e morais (*Our goal in this research article is to explore moral values, creating a discussion about morality, and debating whether or not it is ethical to edit genes*), e como esta técnica pode ser considerada moral quando utilizada como uma possibilidade de tratamento de doenças, sendo um avanço para a medicina.

#### Exemplo 40

---

Texto: y10\_pw9\_14\_text.txt.out

Escore: -2

---

Our **goal** in this research article is to explore moral values, creating a discussion about morality, and debating whether or not it is ethical to edit genes. (...) However, even with the change in our scenario, we still believe that it is somewhat immoral when it comes to esthetics, and moral when it comes to diseases. Conclusions During our theory making **process**, we decided to focus on the ethical side of it (...).

---

O Exemplo 41 ilustra a importância da persistência para que o sucesso seja alcançado. O texto também destaca como ter uma paixão pode auxiliar no comprometimento para que objetivos sejam alcançados, (“*This passion builds and increases your commitment towards your goal*”) e como correr riscos é importante nesse processo (*What is life without a little risk? You can never know if something will go wrong until you try it!*). Além disso, exemplos são usados para reforçar os argumentos apresentados, como Bill Gates.

#### Exemplo 41

---

Texto: y9\_st1\_text3.txt.out

Escore: -2

---

Success can not be achieved without persistence, hard work and risk taking. When someone has a dream, an ambition, things can never be accomplished the perfect way without these elements. These elements strengthen your passion, and make it, maybe not more straightforward, but certainly satisfying and amusing. This passion builds and increases your commitment towards your goal. (...) What is life without a little risk? You can never know if something will go wrong until you try it! Persistence, hard work and taking risks are the fundamental basis for success. Ask anyone! Do you think Bill Gates would have gotten to where he is now without persistence?

---

O Exemplo 42 também utiliza exemplos em sua argumentação, mas exemplos pessoais, como pode ser observado no seguinte trecho em que o aluno conta como conseguiu superar barreiras linguísticas, “*However I was determined to learn the new language. I read the abandoned books in english from the school’s library*”. Ademais, no exemplo, são explicados e listados os obstáculos que as pessoas encontram durante a vida (*Whatever your goal is, there will be obstacles in the way*) e a importância de se manter focado (*You set goals and have dreams you want to fulfill so you start the work to make your dreams a reality*).

#### Exemplo 42

---

Texto: y7\_st3\_text4.txt.out

Escore: -1

---

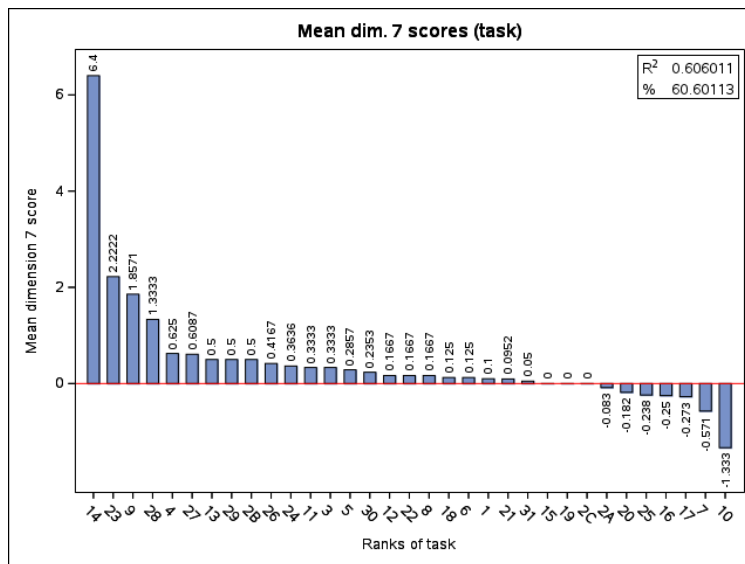
You set goals and have dreams you want to fulfill so you start the work to make your dreams a reality. Then you run into roadblocks. Maybe you want to move to a place you love, but you can't find a way to pay for the move. Maybe you yearn for a good report card, however you can't focus in class. Maybe you want a steady paycheck, and you can't keep a job. Whatever your goal is, there will be obstacles in the way. (...) However I was determined to learn the new language. I read the abandoned books in english from the school’s library, I had extra english classes after school, until I became fluent.

---

Pode-se concluir que os polos da Dimensão 7 compartilham o estabelecimento de metas e iniciativas, seja em políticas de saúde ou em ambições pessoais. No entanto, os polos são distintos. No polo positivo, é evidenciado um certo distanciamento dos alunos em relação às decisões e políticas escolares, como na falta de participação dos alunos na tomada de decisão sobre mudanças na política alimentar da escola. Enquanto isso, no polo negativo, são destacados exemplos de superação pessoal e esforços individuais para alcançar objetivos, com uma ênfase maior na determinação e na persistência dos próprios indivíduos.

Em relação às tarefas, como demonstrado pela Figura 9, essa variável é responsável por uma grande quantidade de variação da Dimensão 7 ( $R^2 = 60,60\%$ ). Os escores médios mais extremados são das tarefas 14 e 10, sendo que ambas foram realizadas por alunos do Nono Ano do Ensino Fundamental II. No polo positivo, a tarefa 14 envolveu a escrita de um roteiro para uma palestra sobre uma nova política alimentar na escola, correspondendo assim à temática do polo positivo. Já no polo negativo, a tarefa 10 envolveu a escrita sobre o sucesso e as principais qualidades que uma pessoa bem sucedida tem, correspondendo à temática do polo negativo.

Figura 9 - Escores médios das tarefas na Dimensão 7



Fonte: a autora

#### 4.1.1.8 Dimensão 8 - Transições de vida e experiências de crescimento pessoal versus Aspectos éticos, sociais e estéticos da verdade e engano

A oitava Dimensão foi intitulada: “Transições de vida e experiências de crescimento pessoal versus Aspectos éticos, sociais e estéticos da verdade e engano”, pois o polo positivo apresenta um discurso centrado em temas mais concretos, como o dia a dia, enquanto o polo

negativo, foca em questões mais abstratas como a verdade. As variáveis lexicais são apresentadas na Tabela 14.

Tabela 14 - Variáveis lexicais da Dimensão 8

Polo	Variáveis lexicais
<b>Positivo</b>	dad_n (0,72), return_v (0,66), (boyfriend_n (0,59)), girl_n (0,58), (period_n (0,57)), food_n (0,57), college_n (0,57), (treat_v (0,57)), (regard_v (0,55)), house_n (0,55), child_n (0,55), lose_v (0,52), bear_v (0,51), name_n (0,41), (die_v (0,40)), (eat_v (0,40)), year_n (0,40), live_v (0,39), (wear_v (0,39)), (old_j (0,38)), (arrow_n (0,36)), (practice_v (0,35)), go_v (0,35), (boy_n (0,35)), (family_n (0,35)), friend_n (0,32), lot_n (0,32), decide_v (0,31)
<b>Negativo</b>	lie_n (-1,11), lie_v (-1,04), truth_n (-0,63), (achieve_v (-0,57)), paint_v (-0,55), painting_n (-0,48), media_n (-0,42), (process_v (-0,40)), (prisoner_n (-0,38)), (success_n (-0,38)), teach_v (-0,36), (metaphor_n (-0,33)), (importance_n (-0,31))

Fonte: a autora

As variáveis lexicais com maior carga no polo positivo da Dimensão 8 incluem *dad*, *return*, *boyfriend*, *girl*, *period*, *food*, *college*, *treat*, *regard*, *house* e *child*, que indicam um foco em relacionamentos familiares, jornadas educacionais, experiências emocionais, incluindo perda e enfrentamento, e decisões significativas de vida. Deste modo, o polo positivo da Dimensão 8 foi rotulado como “Transições de vida e experiências de crescimento pessoal”.

O Exemplo 43 apresenta a descrição da rotina da personagem Sara (*Sara has 19 years old and she loves water the flowers before going for the college*) e de sua família (*Her dad and her mom lives in front of hers house, and the 2 sisters lives with them*). Da mesma forma, o Exemplo 44 ilustra como a rotina e, conseqüentemente, as personagens foram impactadas por mudanças para um novo país (*They had experiences with leaving friends, dealing with family expectations and daily routine changes*), assim como os desafios de se adaptarem à rotina de seus países de origem (*Once returning to one's home country after living overseas is an emotional period of adjustment for the girls*).

### Exemplo 43

---

Texto: y6\_st31\_text2.txt.out

Escore: 12

---

Sara has 19 **years old** and she loves water the flowers before **going** for the **college**. She loves flowers however she study about animals because she want to be a veterinary. Her **dad** and her mom **lives** in front of hers **house**, and the 2 sisters **lives** with them because they are little babies.

---

### Exemplo 44

---

Texto: y8\_st12\_text7.txt.out

Escore: 10

---

Did you have the experience of **going** overseas to study? Well Cora and Mi Cha **went** through that. There they had a great time and some difficulties as changing countries brought. They had experiences with leaving **friends**, dealing with **family** expectations and daily routine changes. Once **returning** to one's home country after **living** overseas is an emotional **period** of adjustment for the **girls**.

---

Pode-se, também, observar como a narrativa de histórias de vida marcam o polo positivo da Dimensão 8. Por exemplo, o trecho do Exemplo 45 que se inicia com “*Once upon a time, a princess named Katerina had escaped the castle and she was living in her bestie's house in the village*” e terminando o texto com “*They found Katerina and Luna, they improved the living conditions of the village and finally lived happily ever after*”; tanto a expressão “*Once upon a time*” quanto “*happily ever after*” são típicas de contos de fadas. Assim, muitas das narrativas apresentadas no polo positivo da Dimensão 8 não são histórias reais sobre experiências de vida, mas sim de personagens fictícios.

### Exemplo 45

---

Texto: y7\_st43\_text4.txt.out

Escore: 10

---

Once upon a time, a princess named Katerina had escaped the castle and she was **living** in her bestie's **house** in the village. The majesty abandoned the castle because she didn't like her stepmom. Queen Alice **died** when the princess was 10 **years** old, and she was feeling miserable, like her heart was falling apart. 2 **years** later, and the king got married again, this time with queen Louisa. (...) They found Katerina and Luna, they improved the living conditions of the village and finally **lived** happily ever after.

---

No polo negativo da Dimensão 8, as variáveis lexicais que carregaram maior peso foram *lie\_n*, *lie\_v*, *truth*, *achieve*, *paint*, *painting* e *media*, que sugerem discursos nos quais o engano e a verdade desempenham papéis críticos, desde a ética pessoal até impactos sociais,

incluindo o papel das artes e da mídia na formação de percepções da verdade. Dessa forma, o polo negativo da Dimensão 8 recebeu o rótulo “Aspectos éticos, sociais e estéticos da verdade e engano”.

No Exemplo 46, observa-se a discussão acerca da papel da mentira para que as relações sociais sejam mantidas, sendo sugerido que dizer a verdade seria mais difícil do que mentir, como no trecho “*without lies, our relationships would melt, and, many times, it is more difficult to tell the truth than to lie*”. Ainda, informações falsas ou mesmo enganosas fariam parte do cotidiano e muitas vezes seriam espalhadas pela mídia (*Nowadays, our lives start getting more and more plagued by false information spread by people and the media*).

A conexão entre a mentira e mídia também está presente no Exemplo 47, em que mentiras são uma estratégia lucrativa, pois promoveriam um maior engajamento das pessoas (*social media companies gain money through time spent and activity, thus a shocking, scandalous, or false fact may be of more value than the truth, as it provides for more shock - value to consumers and provokes more sharing*). Assim, há uma representação que seria importante para que as relações sociais sejam preservadas, mas que na mídia não seria benéfica para a sociedade.

#### Exemplo 46

---

Texto: y8\_st13\_text1.txt.out  
Escore: - 6

---

According to the article “Looking for the **Lie**”, by Robin Marantz Henig, without **lies**, our relationships would melt, and, many times, it is more difficult to tell the **truth** than to **lie**. Nowadays, our lives start getting more and more plagued by false information spread by people and the **media**.

---

#### Exemplo 47

---

Texto: y9\_st36\_text6.txt.out  
Escore: - 4

---

Additionally, social **media** companies gain money through time spent and activity, thus a shocking, scandalous, or false fact may be of more value than the **truth**, as it provides for more shock - value to consumers and provokes more sharing. (...) Like all other issues, the start for a solution **lies** in conscientization; showing the public the harm that these websites do to us.

---

A representação também está presente em outros textos com maior escore no polo negativo da Dimensão 8, porém, com o foco na arte e expressão de sentimentos. O Exemplo 48 ilustra como outras formas de expressão, como a pintura, podem representar momentos

distintos na vida de artistas, como em *“he decided to draw a painting from the coffee, expressing his feeling of joy at the time”* e *“how paintings such as his Arle’s iconic one, containing a calming and relaxing message, weren’t as commonly painted as his depressed and melancholic ones”*.

#### Exemplo 48

---

Texto: y9\_st16\_text4.txt.out

Escore: -1

---

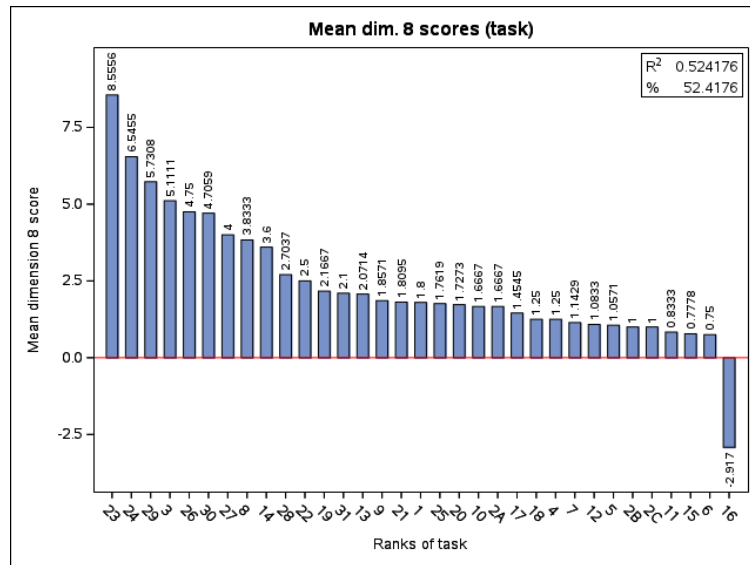
It was there that he decided to draw a **painting** from the coffee, expressing his feeling of joy at the time. However, most are aware that this cheerful feeling was far from representing Vincent’s life as a whole. (...) Episodes such this one clearly shows how disturbed Van Gogh’s life and career was, and how **paintings** such as his Arle’s iconic one, containing a calming and relaxing message, weren’t as commonly **painted** as his depressed and melancholic ones.

---

Assim, na Dimensão 8 “Transições de vida e experiências de crescimento pessoal versus Aspectos éticos, sociais e estéticos da verdade e engano”, os polos positivo e negativo têm em comum o foco nas dinâmicas sociais. Contudo, os polos distinguem-se quando os discursos subjacentes no polo positivo estão centralizados na rotina e suas mudanças, enquanto os discursos do polo negativo têm como foco as implicações de informações enganosas e mentirosas na sociedade.

Quanto às tarefas, essa variável é responsável por uma grande quantidade de variação da Dimensão 8 ( $R^2 = 52,41\%$ ), conforme a Figura 10. Os escores médios mais extremados são das tarefas 23, 24 e 16. No polo positivo, as tarefas 23 e 24 foram realizadas, respectivamente, pelos alunos do Oitavo e Sétimo Ano do Ensino Fundamental II; sendo que a tarefa 23 envolveu a escrita de um texto comparativo entre a história de vida de duas personagens e a tarefa 24 a escrita de uma narrativa pessoal descritiva. Essas tarefas correspondem à temática do polo positivo. Já no polo negativo, a tarefa 16 foi realizada pelos alunos do Oitavo Ano do Ensino Fundamental II, os quais escreveram a introdução de um texto que respondesse à seguinte pergunta: “Devemos sempre falar a verdade?”, correspondendo assim à temática do polo negativo.

Figura 10 - Escores médios das tarefas na Dimensão 8



Fonte: a autora

#### 4.1.2 - ANOVAs

Para o presente estudo, a variável independente analisada foi a tarefa e a Tabela 15 mostra os resultados da ANOVA para cada dimensão do corpus CABI. Observa-se que a variável tarefa é estatisticamente significativa na discriminação entre os textos, uma vez que, em todas as dimensões, o valor de  $p$  foi menor que 0,001, indicando que a variação não ocorre de forma aleatória, mas associada às diferenças entre as tarefas. Além disso, o valor de  $F$  foi significativo em todas as dimensões, sugerindo que há diferenças estatísticas entre os grupos de tarefas em relação às médias de cada tarefa nas dimensões. O valor de  $R^2$  foi maior que 50% em todas as dimensões, o que implica que a maior parte da variação nas dimensões é explicada pela variável tarefa. Portanto, pode-se concluir que as diferenças entre as tarefas são estatisticamente significativas e que a tarefa é um preditor importante da variação nas dimensões analisadas.

Tabela 15 - ANOVAs

Dimensão	F	$p <$	R <sup>2</sup>
1	70,01	0,0001	84,37%
2	23,02	0,0001	63,95%
3	29,85	0,0001	69,71%
4	14,24	0,0001	52,34%
5	26,45	0,0001	67,10%
6	23,54	0,0001	64,48%
7	19,95	0,0001	60,60%
8	14,29	0,0001	52,41%

Fonte: a autora

#### 4.2 RESULTADOS DA IDENTIFICAÇÃO DE METÁFORAS

A seleção dos candidatos à metáfora foi realizada conforme descrito na Seção 3 - Metodologia. Ao total, foram encontradas 663 ocorrências dos candidatos no corpus CABI. A Tabela 16 apresenta os candidatos a metáfora e como estão distribuídos em cada dimensão.

Tabela 16 - Distribuição dos candidatos à metáfora nas dimensões

	D1 pos	D1 neg	D2 pos	D2 neg	D3 pos	D3 neg	D4 pos	D4 neg	D5 pos	D5 neg	D6 pos	D6 neg	D7 pos	D7 neg	D8 pos	D8 neg	Total
allegory			36														36
aspect	7										3						10
change	9													9			18
culture	19						61										80
decision	1										2						3
effect	3																3
finding	3																3
health	69												40				109
hypothesis	16				2												18
impact	14												4				18
importance	18															2	20
metaphor			6														6
nature			3							7							10
opinion	12		18														30
policy													35				35
process	14													11			25
quality	14																14
reason	15																15
result	6																6
religion	6						25										31
situation	16																16
society	28													10			38
source	9																9
success									35							1	36
truth											2					36	38
way	36																36
<b>Total</b>	<b>315</b>	<b>0</b>	<b>63</b>	<b>0</b>	<b>2</b>	<b>0</b>	<b>86</b>	<b>0</b>	<b>35</b>	<b>7</b>	<b>5</b>	<b>2</b>	<b>79</b>	<b>30</b>	<b>0</b>	<b>39</b>	<b>663</b>

Fonte: a autora

Pode-se perceber que a maioria dos candidatos à metáfora encontra-se nos polos positivos das dimensões, como na Dimensão 1 - “Conhecimento abstrato, teórico e científico”, com 315 ocorrências, e na Dimensão 4 - “Autonomia cultural, diversidade e respeito”, com 86 ocorrências. De forma geral, os polos positivos das dimensões do corpus CABI são marcados por discursos centrados em temas complexos e abstratos, o que justificaria a maior incidência de substantivos abstratos, isto é, os candidatos à metáfora.

#### 4.2.1 Resultados e Análise da Anotação Manual

Os candidatos e as sentenças em que ocorreram foram analisados de acordo com o procedimento de identificação de metáforas MIP (Pragglejaz Group, 2007), descrito na Seção 3 - Metodologia. Ao final da anotação manual, 245 ocorrências (tokens) dos candidatos à metáfora foram anotados como metafóricos, o que corresponde a 37% dos candidatos (types) como tendo pelo menos uma ocorrência de metáfora. A Tabela 17 apresenta a distribuição das metáforas identificadas em cada dimensão. A análise de como os estudantes utilizaram as metáforas linguísticas para construir noções abstratas é apresentada a seguir.

Tabela 17 - Distribuição das metáforas identificadas nas dimensões

	D1 pos	D1 neg	D2 pos	D2 neg	D3 pos	D3 neg	D4 pos	D4 neg	D5 pos	D5 neg	D6 pos	D6 neg	D7 pos	D7 neg	D8 pos	D8 neg	Total
allegory			9														9
aspect	2																2
change	3												3				6
culture	6						39										45
decision	1																1
effect	1																1
finding																	0
health	12												9				21
hypothesis	4				1												5
impact	1																1
importance	3															1	4
metaphor			2														2
nature									4								4
opinion	7		10														17
policy													8				8
process	2													4			6
quality	3																3
reason	4																4
result	2																2
religion							14										14
situation	4																4
society	8													7			15
source																	0
success									30								30
truth																9	9
way	32																32
<b>Total</b>	<b>95</b>	<b>0</b>	<b>21</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>53</b>	<b>0</b>	<b>30</b>	<b>4</b>	<b>0</b>	<b>0</b>	<b>20</b>	<b>11</b>	<b>0</b>	<b>10</b>	<b>245</b>

Fonte: a autora

Na Dimensão 1 - “Conhecimento abstrato, teórico e científico versus Dinâmicas familiares, resiliência pessoal e crescimento emocional”, as metáforas foram identificadas apenas no polo positivo. A incidência de metáforas no polo positivo da Dimensão 1 pode estar relacionada ao fato de as tarefas requerem menção a tópicos abstratos ou científicos, como por exemplo a centralidade da metodologia científica e do referencial teórico na compreensão de questões da sociedade contemporânea. Outro aspecto é a utilização de conceitos leigos com

potencial metafórico como *way*. Os substantivos abstratos que tiveram uma maior ocorrência como metafóricos foram *way* (32 casos) e *health* (12 casos).

O substantivo *way* tem como sentido básico “uma rota, direção ou caminho<sup>36</sup>” (Way, 2024) e foi metaforizado como um objeto, como no exemplo *Considering that the formed research question is able to **give way** to possible hypotheses, devising one is important to shaping the flow of the research article*. Neste exemplo, o verbo *give*, cujo sentido básico é “oferecer algo a alguém, ou fornecer algo a alguém<sup>37</sup>” (Give, 2024) funciona como Veículo da metáfora, pois *way* é metaforizado como um objeto que pode ser oferecido ou dado.

O substantivo *health* tem como sentido básico “a condição do corpo e o grau em que está livre de doenças, ou o estado de estar saudável<sup>38</sup>” (Health, 2024) e foi metaforizado como um objeto ou por meio da personificação. O uso metafórico de *health* pode ser observado no exemplo a seguir: *Are we going to **sacrifice** our mental **health**, happiness and our time?*, pois o verbo *sacrifice* tem como sentido básico “matar um animal ou uma pessoa e oferecê-los a um deus ou deuses<sup>39</sup>” (Sacrifice, 2024), sendo o Veículo da metáfora, pois *health* não seria algo ou uma pessoa que possa ser sacrificado ou mesmo um objeto que possa ser dado a alguém.

Na Dimensão 2 - “Simbolismo literário e a interpretação na educação versus Vida diária, interações sociais, crescimento pessoal e bem-estar”, observa-se que a maior incidência de metáforas linguísticas deu-se no polo positivo. Os discursos do polo “Simbolismo literário e interpretação na educação” expressam como símbolos e representações auxiliam na compreensão de conceitos abstratos, como assuntos complexos e eventos reais, e na construção desses discursos os substantivos *allegory* (9 casos) e *opinion* (10 casos) foram metaforizados.

O substantivo *allegory* tem como sentido básico “uma história, peça teatral, poema, imagem ou outro trabalho no qual os personagens e eventos representam qualidades ou ideias específicas relacionadas a moral, religião ou política<sup>40</sup>” (Allegory, 2024) e foi metaforizado como um lugar ou por meio da personificação. Um exemplo em que *allegory* foi metaforizado como lugar é *Well, all those **lead** to an **allegory***, pois o verbo *lead* tem como sentido “guiar em um caminho, especialmente indo adiante<sup>41</sup>” (Lead, 2024), assim há uma incongruência

<sup>36</sup> No original, “a route, direction, or path”.

<sup>37</sup> No original, “to offer something to someone, or to provide someone with something”.

<sup>38</sup> No original, “the condition of the body and the degree to which it is free from illness, or the state of being well”.

<sup>39</sup> No original, “to kill an animal or a person and offer them to a god or gods”.

<sup>40</sup> No original, “a story, play, poem, picture, or other work in which the characters and events represent particular qualities or ideas that relate to morals, religion, or politics”.

<sup>41</sup> No original, “to guide on a way especially by going in advance”.

entre o sentido básico de *allegory* e o sentido evocado pelo verbo *lead*, sendo o Veículo da metáfora. Ademais, ao ser metaforizado por meio da personificação, *allegory* recebeu características humanas, como no exemplo *An allegory always references a real life situation in a fictional story, usually the reference is in a bad aspect criticizing it in a quiet and not so direct way, like a metaphore*, pois o verbo *reference* é o Veículo e tem como sentido básico “referir-se a alguém ou algo<sup>42</sup>” (Reference, 2024), uma ação que seria realizada por uma pessoa, mas que pode ser compreendida na transferência de sentido no contexto, ou seja, ao se transferir para *allegory* características humanas.

Já o substantivo *opinion*, cujo sentido básico é “um pensamento ou crença sobre algo ou alguém<sup>43</sup>” (Opinion, 2024), foi metaforizado como um lugar na maioria das ocorrências. O fato da preposição *in*, cujo sentido básico é “dentro de um recipiente, lugar ou área, ou cercado ou fechado por algo<sup>44</sup>” (In, 2024), preceder o substantivo *opinion* gera uma incongruência, sendo o Veículo da metáfora, pois um pensamento ou uma crença não são um recipiente ou lugar. Um exemplo em que *opinion* é metaforizado como lugar é *But among those inventions, the artificial satellites stood up in public opinion*.

Na Dimensão 3 - “Relatos históricos e descobertas arqueológicas versus Identidade, transformação e resiliência na sociedade contemporânea”, apenas a ocorrência do substantivo *hypothesis* na sentença *How was Ötzi murdered is not revealed yet, but we might have some hypothesis* foi anotado como metafórico. A decisão foi baseada no fato de que o verbo *have* tem como sentido básico “possuir”, enquanto o substantivo *hypothesis* tem como sentido básico “uma ideia ou explicação para algo que se baseia em fatos conhecidos, mas ainda não foi comprovada<sup>45</sup>” (Hypothesis, 2024). Assim, há uma incongruência, pois *hypothesis* é metaforizado como um objeto que se pode possuir, o que contrasta com o seu sentido básico. A ocorrência de apenas uma metáfora linguística na Dimensão 3 pode ser explicada pelo fato de que esta dimensão tem discursos mais concretos, como descobertas históricas e arqueológicas, no polo positivo, e centrados na vida contemporânea e nos seus desafios, especialmente relacionados ao dinheiro, no polo negativo.

Na Dimensão 4 - “Autonomia cultural, diversidade e respeito versus Vida cotidiana e experiências de aprendizagem”, houve uma maior incidência de metáforas linguísticas no polo positivo. Este polo tem o seu discurso centrado no respeito à diversidade, enfatizando a importância do respeito e defesa das diferentes culturas, religiões e costumes. Ademais, para a

---

<sup>42</sup> No original, “to refer to someone or something”.

<sup>43</sup> No original, “a thought or belief about something or someone”.

<sup>44</sup> No original, “inside a container, place, or area, or surrounded or closed off by something”.

<sup>45</sup> No original, “an idea or explanation for something that is based on known facts but has not yet been proved”.

construção deste discurso, os substantivos abstratos *culture* e *religion* foram os que tiveram uma maior incidência como metafóricos.

O substantivo *culture* foi identificado em 39 casos como metafórico, sendo metaforizado como um lugar ou por meio da personificação. Um exemplo em que *culture* é metaforizado por meio da personificação pode ser observado em: *a cultural divorce is when someone "divorces" with his/her previous culture*, pois o verbo *divorce* tem como sentido básico “encerrar seu casamento por meio de um processo oficial ou legal<sup>46</sup>” (Divorce, 2024), o que pressupõe que a relação seja composta por pessoas, contrastando com o sentido básico de *culture* “o modo de vida, especialmente os costumes e crenças gerais, de um grupo específico de pessoas em um determinado momento<sup>47</sup>” (Culture, 2024). Assim, *culture* recebe características humanas ao ser enquadrado como algo com que é possível se divorciar.

Quando metaforizado como um lugar, o substantivo *culture* foi precedido por verbos ou preposições que evocam o sentido de lugar, como em *As in Brazilian culture*, pois a preposição *in* metaforiza *culture* como um lugar ou recipiente. Outro exemplo de como *culture* foi metaforizado como lugar é *By context we can know that a cultural divorce is leaving a culture you or your heritage was once part of*, pois o verbo *leave* tem como sentido básico “afastar-se de alguém ou algo, temporariamente ou permanentemente<sup>48</sup>” (Leave, 2024).

O substantivo *religion* foi identificado como metafórico em 14 ocorrências, sendo que na maior parte dos casos, *religion* foi metaforizado por meio de personificação, recebendo características humanas, como no exemplo a seguir: *However, the interference of any type can be justified when the religion or culture promotes violence, mass killing, any type of breaking of basic societies laws or when the religion/culture interferes on the freedom of the other*. Neste exemplo, os verbos *promote* e *interfere* personificam *religion*, pois são verbos que indicam ações humanas, uma vez que o sentido básico de *promote* é “estimular as pessoas a gostar, comprar, usar, fazer ou apoiar algo<sup>49</sup>” (Promote, 2024) e de *interfere* é “envolver-se em uma situação quando seu envolvimento não é desejado ou não é útil<sup>50</sup>” (Interfere, 2024). Além disso, é possível compreender *religion* por meio de uma relação metonímica, ou seja, *religion* pode ser entendido como “as pessoas que seguem determinada religião”.

Na Dimensão 5 - “Resiliência, determinação e triunfo socioeconômico versus Autonomia e expressão individual em ambientes estruturados e digitais”, a maioria das

---

<sup>46</sup> No original, “to end your marriage by an official or legal process”.

<sup>47</sup> No original, “the way of life, especially the general customs and beliefs, of a particular group of people at a particular time”.

<sup>48</sup> No original, “to go away from someone or something, for a short time or permanently”.

<sup>49</sup> No original, “to encourage people to like, buy, use, do, or support something”.

<sup>50</sup> No original, “to involve yourself in a situation when your involvement is not wanted or is not helpful”.

metáforas linguísticas ocorreram no polo positivo. Este polo tem um discurso centrado na superação de adversidades, mas apresentando uma visão restrita de sucesso, que se aproxima de ideias meritocráticas. Além disso, o substantivo *success* teve o maior número de ocorrências, sendo 30 dos casos.

O substantivo *success*, que tem como sentido básico “a conquista dos resultados desejados ou esperados<sup>51</sup>” (Success, 2024), foi metaforizado como um lugar ou destino ou por meio da personificação. Exemplos em que *success* foi metaforizado como um lugar ou destino podem ser observados em: *To **achieve success** you need to deserve it* e *All we can do is face obstacles and challenges that will appear **in your way to success***. No primeiro exemplo, o verbo *achieve*, cujo sentido básico é “conseguir terminar algo ou alcançar um objetivo, especialmente após muito trabalho ou esforço<sup>52</sup>” (Achieve, 2024), metaforiza *success* como um lugar que pode ser alcançado. No segundo exemplo, *in your way to* [no seu caminho para] também metaforiza *success* como um destino ou lugar a ser alcançado. Já como personificação, *success* recebeu características humanas, como no exemplo *And **success makes** everything easier*. O verbo *make* é o Veículo da metáfora, pois o seu sentido básico é “produzir algo, frequentemente utilizando uma substância ou material específico<sup>53</sup>” (Make, 2024), o que transfere para *success* a ação de produzir algo.

Na Dimensão 6 - “Fenômenos naturais e recursos hídricos versus Crescimento pessoal e interação social”, não foram identificadas metáforas tanto no polo positivo quanto negativo. A ausência de metáforas linguísticas nesta dimensão pode ser explicada pelo fato de que os discursos subjacentes estão relacionados à explicação de fenômenos naturais, como no polo positivo, que requerem um linguagem mais descritiva e objetiva, e o impacto positivo e negativo de escolhas pessoais que envolvem esporte e música tem na vida das pessoas, o que também mobilizou o uso de uma linguagem mais objetiva.

A Dimensão 7 - “Iniciativas abrangentes de saúde e engajamento nas escolas versus Esforços pessoais e considerações éticas na conquista de objetivos” foi a única em que metáforas linguísticas foram identificadas em ambos os polos, sendo 20 casos no polo positivo e 11 no polo negativo. Além disso, como destacado na seção anterior, os dois polos da Dimensão 7 apresentam uma maior densidade de informação, que é marcada pela maior incidência de substantivos como variáveis lexicais.

---

<sup>51</sup> No original, “the achieving of the results wanted or hoped for”.

<sup>52</sup> No original, “to succeed in finishing something or reaching an aim, especially after a lot of work or effort”.

<sup>53</sup> No original, “to produce something, often using a particular substance or material”.

O polo positivo “Iniciativas abrangentes de saúde e engajamento nas escolas” tem o seu discurso centrado em como as escolas são um ambiente que pode impactar a vida dos alunos, principalmente por meio de políticas implementadas por essas instituições. Neste polo, os substantivos *health* (9 casos) e *policy* (8 casos) foram usados de forma metafórica pelos estudantes.

O substantivo *health* foi metaforizado como um objeto concreto, como em: *Mental **health** has always been **pushed aside**, especially the ones of teenagers*. Neste exemplo, o verbo *push* é o Veículo da metáfora, pois o seu sentido básico é “aplicar pressão física ou força, especialmente com as mãos, para mover algo para uma posição diferente, geralmente uma que esteja mais distante de você<sup>54</sup>” (Push, 2024), assim, ocorre uma incongruência, pois *health* não pode ser movimentado. Ainda, o advérbio *aside* também metaforiza *health* como um objeto, pois o seu sentido básico é “de lado ou para um lado<sup>55</sup>” (Aside, 2024), reforçando *health* como um objeto que pode ser movimento. É importante destacar que *push aside* é um verbo frasal e o Pragglejaz Group (2007) recomenda que, nesses casos, verbo e a partícula adverbial não sejam analisados separadamente, por causa da decomponibilidade. Porém, optou-se por analisar cada unidade lexical, pois o sentido de *push aside* pode ser compreendido pela combinação de seus componentes.

Outro exemplo em que *health* é metaforizado como um objeto que pode ser movimentado e que demonstra a sistematicidade com que esta metáfora ocorre é *When you eliminate processed and industrialised food from our diets, we become 50% less likely to become obese, and obesity can **lower** your mobility, **health**, and even lower your life expectancy!*. Neste exemplo, o verbo *lower* é o Veículo da metáfora, pois *health* não é um objeto que pode ser movimentado para uma posição mais baixa, sentido básico de *lower*.

O substantivo *policy*, que tem como sentido básico “um conjunto de ideias ou um plano do que fazer em situações específicas que foi oficialmente acordado por um grupo de pessoas, uma organização empresarial, um governo ou um partido político<sup>56</sup>” (Policy, 2024), foi metaforizado por meio da personificação na maior parte das ocorrências, como pode ser observado em: *That new **policy demands** that the school does not serve any kind of industrialized food, and only serves 100% organic food*. No exemplo, *policy* é metaforizado por meio da personificação ao ser acompanhado do verbo *demand*, que tem como sentido

---

<sup>54</sup> No original, “to use physical pressure or force, especially with your hands, in order to move something into a different position, usually one that is further away from you”.

<sup>55</sup> No original, “on or to one side”.

<sup>56</sup> No original, “a set of ideas or a plan of what to do in particular situations that has been agreed to officially by a group of people, a business organization, a government, or a political party”.

básico “pedir algo com firmeza, de forma que demonstre que você não espera ser recusado<sup>57</sup>” (Demand, 2024).

Já o polo negativo da Dimensão 7 - “Esforços pessoais e considerações éticas na conquista de objetivo” tem um discurso centrado nas dificuldades e debates acerca da conquista de objetivos individuais e coletivos. Neste polo, o substantivo *society* foi anotado em 7 casos como metafórico. O substantivo *society* tem como sentido básico “pessoas consideradas como um grupo, ou um grupo de pessoas que vivem juntas em um sistema social particular<sup>58</sup>” (Society, 2024) e foi metaforizado como um lugar ou por meio da personificação.

Quando metaforizado como um lugar, o substantivo *society* foi precedido por uma preposição de lugar, como *in*, que tem como sentido básico “dentro de um recipiente, lugar ou área, ou cercado ou fechado por algo<sup>59</sup>” (In, 2024). Por exemplo, em *In it, we found statistics given by organizations such as Google and King's College regarding the influence of technology in society and the impact of online advertising*, a preposição *in* causa uma incongruência, ao metaforizar *society* como um recipiente ou lugar fechado. Já quando metaforizado por meio da personificação, *society* recebeu atributos humanos, como no seguinte exemplo: *I'm here to show why social media can be a bad thing if we use it in the wrong way (that is the way the biggest part of the society uses it)*. Neste exemplo, o verbo *use* é o Veículo da metáfora, pois o seu sentido básico é “utilizar algo, como uma ferramenta, habilidade ou construção, para um propósito específico<sup>60</sup>” (Use, 2024). Ainda, é possível compreender *society* por meio de uma relação metonímica, isto é, *society* pode ser entendido como “as pessoas de um determinado grupo”.

Na Dimensão 8 - “Transições de vida e experiências de crescimento pessoal versus Aspectos éticos, sociais e estéticos da verdade e engano”, as metáforas linguísticas foram identificadas apenas no polo negativo. Este polo apresenta em seu discurso como o engano e a verdade desempenham papéis críticos, desde a ética pessoal até impactos sociais, incluindo o papel das artes e da mídia na formação de percepções da verdade, sendo que o substantivo *truth* teve um maior número de ocorrências identificadas como metafóricas (9 casos).

O substantivo *truth* tem como sentido básico “a qualidade de ser verdadeiro<sup>61</sup>” (Truth, 2024) e foi metaforizado como um objeto concreto ou por meio da personificação. Quando metaforizado por meio da personificação, *truth* recebeu atributos de seres humanos, podendo

---

<sup>57</sup> No original, “to ask for something forcefully, in a way that shows that you do not expect to be refused”.

<sup>58</sup> No original, “people considered as a group, or a group of people who live together in a particular social system”.

<sup>59</sup> No original, “inside a container, place, or area, or surrounded or closed off by something”.

<sup>60</sup> No original, “to put something such as a tool, skill, or building to a particular purpose”.

<sup>61</sup> No original, “the quality of being true”.

salvar ou mesmo arruinar a vida das pessoas, como em: *But sometimes the **truth** can **save** your life like in the times were jues were killed, I would lie and say that i'm not jweish so i won't be killed, the **truth** can **save** your but it can also **ruin** it.* Neste exemplo, também pode-se observar a sistematicidade com que *truth* é metaforizado, pois a mesma metáfora ocorre três vezes em uma mesma sentença. Além disso, *truth* foi metaforizado com o um objeto concreto, como no seguinte exemplo *No, I know you must think social media is totally safe but today I will **show** you **the truth** about the internet,* pois o verbo *show* tem como sentido básico “tornar possível que algo seja visto<sup>62</sup>” (Show, 2024), sendo assim *truth* é metaforizado como um objeto que pode ser visto. Assim, pode-se concluir que no polo negativo da Dimensão 8, ao receber características humanas e ser concretizado, o conceito abstrato de *truth* pode ser melhor compreendido, pois está mais próximo da experiência humana.

De forma geral, ao se analisar a incidência das metáforas linguísticas nas dimensões lexicais do corpus CABI, pode-se concluir que as metáforas auxiliam para a compreensão de conceitos abstratos como *truth, success, religion, way, society, health e culture*, ao metaforizar esses conceitos como objetos e lugares ou por meio da personificação. Objetos, lugares e ações ou atributos de seres humanos estão mais próximos da experiência humana, o que faz que conceitos abstratos possam ser melhor compreendidos (Lakoff e Jonhson, 1980/2003).

O polo positivo da Dimensão 1- “Conhecimento abstrato, teórico e científico” e o polo positivo da Dimensão 4 - “Autonomia cultural, diversidade e respeito” são marcados por textos que respondem tarefas que necessitam o uso de conceitos abstratos para a sua execução, como a Tarefa 3 e a Tarefa 17, requerendo, assim, expressão metafórica para serem comunicados. Isso confirma a hipótese de Golden (2021) “tópicos abstratos são difíceis de explicar sem metáforas<sup>63</sup>” (p. 271). Ademais, as tarefas que mais marcaram esses polos foram realizadas por alunos do Oitavo Ano do Ensino Fundamental II e Primeiro Ano do Ensino Médio, assim é esperado que o nível de proficiência desses alunos seja mais avançado. Dessa forma, pode-se observar que há uma maior incidência de metáforas nos polos das dimensões em que os alunos possivelmente tem um nível mais avançado de proficiência em língua inglesa, o que vai ao encontro dos resultados apresentados por Littlemore et al. (2014), Golden (2021) e Nacey (2022).

---

<sup>62</sup> No original, “to make it possible for something to be seen”.

<sup>63</sup> No original, “Abstract topics are hard to explain without metaphor”.

#### 4.2.2 Resultados e análise da anotação por IA

O ChatGPT4 foi utilizado para a anotação das metáforas por IA, seguindo a metodologia apresentada na seção anterior. O mesmo *prompt* foi usado para a análise da estrutura mínima e sentença em que os candidatos à metáfora ocorrem e para a comparação da acuidade da IA na detecção de metáforas recorreu-se ao teste de Kappa. Os resultados das anotações por IA são apresentados na Tabela 18.

Tabela 18 - Comparação dos resultados das anotações por IA

	<b>Verdadeiro Positivo</b>	<b>Verdadeiro Negativo</b>	<b>Falso Positivo</b>	<b>Falso Negativo</b>	<b>Precisão</b>	<b>Sensibilidade</b>
<b>Estrutura mínima</b>	25%	91%	9%	75%	61,6%	24,9%
<b>Nível da sentença</b>	18%	89%	11%	82%	48,89%	17,96%

Fonte: a autora

Na anotação com a estrutura mínima, a IA identificou 61 ocorrências como metafóricas das 245 anotadas manualmente, o que representa 25% de casos verdadeiros positivos, enquanto que na anotação de casos não metafóricos, a IA identificou 380 dos 418 casos anotados manualmente, o que equivale a 91% de casos de verdadeiros negativos. Em relação aos casos de falsos positivos e negativos, a IA anotou como metafóricos 38 casos que foram anotados como não metafóricos na anotação manual, representando 9% de casos falsos positivos, e a porcentagem de casos de falsos negativos foi de 75%, pois a IA anotou como não metafórico 184 casos que foram anotados como metafóricos na anotação manual. Quanto à precisão da identificação, a anotação por IA obteve uma precisão de 61,6%, o que indica um grau de concordância substancial (Viera e Garrett, 2005) e uma sensibilidade (recall) de 24,9%, o que significa que de todos os casos positivos, o ChatGPT4 identificou corretamente cerca de 24,9% deles.

Já na anotação no nível da sentença, a IA identificou 44 ocorrências como metafóricas das 245 anotadas manualmente, representando 18% de casos verdadeiros positivos. A porcentagem de casos verdadeiros negativos foi de 89%, pois a IA identificou 372 dos 418 casos anotados manualmente como não metafóricos. Quanto aos casos de falsos positivos e negativos, a IA anotou como metafóricos 46 casos que foram anotados como não metafóricos na anotação manual, o que representa 11% de casos falsos positivos, já a porcentagem de

casos de falsos negativos foi de 82%, pois a IA anotou como não metafórico 201 casos que foram anotados como metafóricos na anotação manual. Em relação à precisão da identificação, a anotação por IA obteve uma precisão de 48,89%, o que indica um grau de concordância moderada (Viera e Garrett, 2005) e uma sensibilidade (recall) de 17,96%.

Em suma, a IA foi capaz de identificar cerca de uma em cada quatro metáforas quando apresentada com um fragmento de frase e uma em cada seis metáforas quando apresentada com a frase completa. Isso demonstra que os Modelos Amplos de Linguagem (LLMs) atualmente não conseguem detectar metáforas de modo abrangente.

Comparando-se os resultados da anotação com a estrutura mínima e no nível da sentença, observa-se que a anotação realizada com a estrutura mínima em que o candidato à metáfora ocorre obteve um melhor desempenho, pois identificou 25% de casos verdadeiros positivos em comparação com 18% na identificação no nível da sentença e 91% de casos verdadeiros negativos em comparação com 89%. Estes dados estão de acordo com a hipótese levantada por Wachowiak e Gromann (2023), de que ao se restringir o contexto em que os candidatos à metáfora ocorrem, aumenta-se a acurácia da identificação de metáforas por IA.

Outro ponto que deve ser levado em consideração na avaliação da eficácia das duas análises é como o ChatGPT4 respondeu aos *prompts*. Na anotação com a estrutura mínima, em três casos o ChatGPT respondeu como não metafórico e justificou que com mais contexto poderia fazer uma melhor análise, sendo eles: *success is making, these can include in culture and music can influence the way*. Por outro lado, na anotação no nível da sentença, o ChatGPT4 não respondeu de forma organizada: ou não respondeu com uma análise ou não respondeu como esperado, sendo necessário que esses casos fossem identificados e executados manualmente. Uma possível explicação é que as sentenças longas podem ter confundido a IA, que acabou não seguindo as instruções do *prompt*. Assim, conclui-se que a utilização de um *prompt* que forneça para a IA a estrutura mínima em que o candidato à metáfora ocorre faz com que a IA consiga executar a tarefa da forma esperada e com resultados mais precisos.

Em relação a como a IA realizou as tarefas, isto é, se o ChatGPT4 foi capaz de reproduzir as decisões humanas sobre metaforicidade, o modelo foi capaz de identificar o sentido básico dos candidatos. Contudo, em alguns casos, falhou na identificação do sentido contextual e, conseqüentemente, na identificação da metáfora, o que ocorreu quando o Veículo da metáfora era uma preposição e alguns casos de personificação.

Na anotação manual, em 34 casos uma preposição foi identificada como o Veículo da metáfora, sendo a preposição *in* a mais frequente (24 casos), seguida de *on* (6 casos), *towards* (2 casos) e *behind* (2). No entanto, dos 32 casos, 3 foram considerados como metafóricos,

mas a preposição não foi identificada como o Veículo em nenhum dos casos, como pode observado no Exemplo 1. Neste exemplo, o Tópico da metáfora é *opinion*, a estrutura mínima é *in public opinion* e a sentença é *But among those inventions, the artificial satellites stood up in public opinion*. Na anotação com a estrutura mínima, não foi identificado como metafórico, mas na anotação no nível da sentença foi considerado metafórico. No exemplo, *opinion* é metaforizado como um lugar, como explicado no Exemplo 1, mas o Veículo identificado na anotação manual foi *in* e não *stood up*. Além disso, é possível que neste exemplo, o aluno queria escrever *stood out* ao invés de *stood up*.

### Exemplo 1

---

"METAPHOR: yes. The phrase \"stood up in public opinion\" uses \"opinion\" metaphorically. In its basic dictionary sense, \"opinion\" refers to a belief or judgment that rests on grounds insufficient to produce complete certainty. However, in this sentence, \"public opinion\" is portrayed as a physical space in which the artificial satellites can \"stand up\", which conveys the idea of their significance or prominence in public awareness or approval.

---

Nos casos em que a metaforização ocorreu por meio da personificação, pode-se observar que o ChatGPT4 não respondeu de forma consistente, pois ora não identificou como metafórico ora identificou como metafórico, mas apenas no nível da sentença ou na estrutura mínima. Por exemplo, na sentença *I'm here to show why social media can be a bad thing if we use it in the wrong way (that is the way the biggest part of the society uses it)*, o Tópico da metáfora é *society* e a estrutura mínima é *society uses it*. Na anotação feita com a estrutura mínima, a IA identificou a personificação e explicou que é atribuído para *society* características humanas, como a habilidade de usar algo, como pode ser observado no Exemplo 2. Já quando o mesmo Tópico foi analisado no nível da sentença, a IA não o identificou como metafórico, justificando que *society* é usado no seu sentido literal, como no Exemplo 3.

### Exemplo 2

---

"content": "METAPHOR:yes\n\nReason: The term \"society\" typically refers to a group of individuals who share a common geographical or social territory, typically subject to the same political authority and dominant cultural expectations. However, in the phrase \"society uses it\", society is given a human-like ability to use something, which is a metaphorical use of the term. The dictionary definition of society does not include this ability to use things. This more abstract use of the term is metaphorical."

---

### Exemplo 3

---

"content": "METAPHOR: NO\n\nThe word \"society\" in this context is used literally, not metaphorically. It refers to a group of individuals who share a common system of customs, laws, and organizations."

---

A inconsistência na análise também aconteceu quando uma sentença foi anotada múltiplas vezes pelo ChatGPT4, como quando uma mesma sentença ocorreu em mais de uma dimensão ou quando um Tópico foi utilizado repetidamente na mesma sentença. Um exemplo de inconsistência de anotação pode ser observado na sentença *Everyone can achieve success, for some it is just more difficult than for others, in the journey to achieve success each person faces their personal demons, faces challenges and goes through various obstacles*, em que o Tópico *success* e o Veículo *achieve* ocorrem duas vezes e, por isso, foram extraídas duas estruturas mínimas, *everyone can achieve success* e *in the journey to achieve success*. Na anotação que levou em consideração apenas as estruturas mínimas, não foi identificado como metafórico o uso de *success*. Contudo, como a Tópico ocorre duas vezes, a anotação no nível da sentença foi feita duas vezes, sendo que na segunda foi identificado como metafórico, como pode ser observado nos exemplos abaixo.

### Exemplo 4

---

"content": "METAPHOR: NO\n\nThe word \"success\" in this context is not used metaphorically. It is used in its literal sense, referring to the achievement of an aim or purpose. The sentence does not suggest a transfer of meaning from another context to understand the concept of success."

---

### Exemplo 5

---

"content": "METAPHOR: YES\n\nThe reason is that \"success\" is personified in this context. It is portrayed as a destination or journey's end that one strives to reach, which is a metaphorical use of the word. Additionally, the phrase \"personal demons\" metaphorically represents personal challenges or issues, and \"faces challenges\" and \"goes through various obstacles\" further support the metaphor of life as a journey."

---

Os resultados apresentados indicam que a IA é mais bem capacitada para identificar metáforas linguísticas quando o *prompt* inclui um trecho menor de texto para analisar. Contudo, a inconsistência de anotação e o fato de, aparentemente, preposições não serem levadas em consideração durante a análise são limitações que a ferramenta apresenta na

identificação de metáforas, sendo necessária mais pesquisa para explorar a capacidade de identificação de metáforas com os atuais LLMs.

## 5. CONCLUSÃO

A presente pesquisa teve como objetivo geral identificar, por meio da Análise Multidimensional Lexical, os principais discursos subjacentes às produções textuais de alunos bilíngues, além de se e como esses alunos fazem o uso de linguagem metafórica para desempenhar os discursos em produção textual em inglês. Dessa forma, o estudo se propôs a: 1) Identificar, por meio da Análise Multidimensional Lexical, os principais discursos em jogo nos textos, que mostram como os estudantes enquadraram os tópicos nas tarefas; 2) Detectar o uso de metáfora nas dimensões provenientes da Análise Multidimensional Lexical, ou seja, como os estudantes utilizaram metáforas para construir noções abstratas associadas às dimensões, e 3) Testar a possibilidade de identificação de metáforas usando a Inteligência Artificial, mais especificamente, até que ponto a Inteligência Artificial poderia reproduzir o julgamento humano sobre a presença de metáforas, usando um subconjunto dos dados.

Para tanto, foi coletado o corpus CABI, contendo 450 textos escritos por 185 alunos do Sexto Ano do Ensino Fundamental II ao Primeiro Ano do Ensino Médio de uma escola bilíngue, totalizando um corpus de 63.297 palavras. O corpus pode ser dividido em cinco sub-corpora, cada um representando um ano escolar e contendo 90 textos. O corpus CABI possibilitou a análise e identificação das dimensões discursivas das produções textuais de estudantes bilíngues, sendo identificadas oito dimensões:

- Dimensão 1 - Conhecimento abstrato, teórico e científico versus Dinâmicas familiares, resiliência pessoal e crescimento emocional
- Dimensão 2 - Simbolismo literário e a interpretação na educação versus Vida diária, interações sociais, crescimento pessoal e bem-estar
- Dimensão 3 - Relatos históricos e descobertas arqueológicas versus Identidade, transformação e resiliência na sociedade contemporânea
- Dimensão 4 - Autonomia cultural, diversidade e respeito versus Vida cotidiana e experiências de aprendizagem
- Dimensão 5 - Resiliência, determinação e triunfo socioeconômico versus Autonomia e expressão individual em ambientes estruturados e digitais
- Dimensão 6 - Fenômenos naturais e recursos hídricos versus Crescimento pessoal e interação social
- Dimensão 7 - Iniciativas abrangentes de saúde e engajamento nas escolas versus Esforços pessoais e considerações éticas na conquista de objetivos

- Dimensão 8 - Transições de vida e experiências de crescimento pessoal versus Aspectos éticos, sociais e estéticos da verdade e engano

Como exposto anteriormente, cada dimensão representa um parâmetro distinto de variação discursiva. Contudo, pode-se destacar algumas similaridades e diferenças entre elas.

As principais similaridades entre as dimensões envolvem a exploração das experiências humanas, aprendizado e crescimento, seja em contextos pessoais, sociais, acadêmicos ou científicos. Vários polos das dimensões abordam temas de crescimento pessoal, desafios e a condição humana, como o polo positivo da Dimensão 7 "Iniciativas abrangentes de saúde e engajamento nas escolas" e o polo negativo da Dimensão 1 "Dinâmicas familiares, resiliência pessoal e crescimento emocional".

Já a principal diferença reside no foco específico de cada dimensão. Por exemplo, algumas são mais abstratas e teóricas, como o polo positivo da Dimensão 2 "Simbolismo literário e a interpretação na educação" ou o polo positivo da Dimensão 1 "Conhecimento abstrato, teórico e científico". Já outras são mais concretas e baseadas em experiências do mundo real e rotinas cotidianas, como o polo negativo da Dimensão 4 "Vida cotidiana e experiências de aprendizagem" ou como os desafios da vida moderna, no polo negativo da Dimensão 3 "Identidade, transformação e resiliência na sociedade contemporânea".

Neste estudo, a variável independente analisada foi a tarefa e se concluiu que essa variável é estatisticamente significativa para explicar a variação entre os textos, uma vez que, em todas as dimensões, o valor de  $p$  foi menor que 0,001, indicando que a variação não ocorre de forma aleatória, mas associada às diferenças entre as tarefas. Além disso, o valor de  $R^2$  foi maior que 50% em todas as dimensões, o que implica que a maior parte da variação nas dimensões é explicada pela variável tarefa. Por exemplo, na Dimensão 1 o valor de  $R^2$  foi de 84,37%, sendo que no polo positivo, a tarefa 3 obteve os escores médios mais extremados. Essa tarefa foi realizada pelos alunos do Primeiro Ano do Ensino Médio, os quais escreveram sobre um assunto de interesse que fosse um problema, cuja possível solução também foi apresentada no texto, sendo textos mais científicos, correspondendo assim à temática do polo. Já a tarefa 29 obteve os escores médios mais extremados no polo negativo. A tarefa 29 foi realizada pelos alunos do Sexto Ano do Ensino Fundamental II, sendo essa uma escrita mais narrativa, uma vez que os alunos criaram uma biografia para um personagem imaginário.

A partir das variáveis lexicais das dimensões, os candidatos à metáfora foram selecionados e, para que se pudesse verificar como os alunos enquadraram conceitos abstratos em seus textos, apenas os substantivos abstratos foram utilizados. Assim, com base na lista de

substantivos com menor concretude de Brysbaert, Warriner e Kuperman (2013), os seguintes substantivos foram escolhidos: *allegory, aspect, change, culture, decision, effect, finding, health, hypothesis, impact, importance, metaphor, nature, opinion, policy, process, quality, reason, result, religion, situation, society, source, success, truth* e *way*. Ao total, foram encontradas 663 ocorrências dos candidatos no corpus CABI.

Seguindo o procedimento de identificação de metáforas MIP (Pragglejaz Group, 2007), os candidatos e as sentenças em que ocorreram foram analisados manualmente por dois anotadores. Ao final da anotação manual, 245 ocorrências (tokens) dos candidatos à metáfora foram anotados como metafóricos, o que corresponde a 37% dos candidatos (types) como tendo pelo menos uma ocorrência de metáfora.

A análise das metáforas linguísticas nas dimensões lexicais indica que os alunos bilíngues metaforizam conceitos abstratos, como *truth, success, religion, way, society, health* e *culture* como objetos e lugares ou por meio da personificação, principalmente em tarefas que demandaram a utilização de conceitos abstratos e realizadas por alunos nos Anos Finais do Ensino Fundamental II e Médio. Quanto à relação entre a incidência de metáforas linguísticas e as dimensões, os resultados da anotação manual apontam que a maioria das metáforas linguísticas encontram-se no polo positivo da Dimensão 1- “Conhecimento abstrato, teórico e científico”, 95 ocorrências, e o polo positivo da Dimensão 4 - “Autonomia cultural, diversidade e respeito”, 53 ocorrências. Esses polos são marcados por textos que respondem tarefas que necessitam o uso de conceitos abstratos para a sua execução, como a Tarefa 3 e a Tarefa 17, requerendo, assim, expressão metafórica para serem comunicados. Além disso, observou-se que há uma maior incidência de metáforas nos polos das dimensões em que os alunos possivelmente têm um nível mais avançado de proficiência em língua inglesa, o que vai ao encontro dos resultados apresentados na literatura.

Em relação à anotação automática de metáforas linguísticas via IA, o mesmo *prompt* foi usado para a análise da estrutura mínima e sentença em que os candidatos à metáfora ocorrem e para a comparação da acuidade da IA na detecção de metáforas recorreu-se ao teste de Kappa. De modo geral, os resultados apresentados indicam que a IA é mais bem capacitada para reproduzir as decisões humanas quando o *prompt* inclui a estrutura mínima em que o candidato ocorre, uma vez que o ChatGPT4 obteve uma precisão de 61,6% na análise com a estrutura mínima, enquanto que a precisão foi de 48,89% na análise no nível da sentença. A IA identificou 25% de casos verdadeiros positivos quando apresentada a estrutura mínima em comparação com 18% na identificação no nível da sentença, isto é, a IA foi capaz de identificar cerca de uma em cada quatro metáforas quando apresentada com um fragmento de

frase e uma em cada seis metáforas quando apresentada com a frase completa. Contudo, observou-se que houve uma inconsistência na anotação e que, possivelmente, as preposições não foram consideradas como Veículo durante a análise. Assim, é necessário mais pesquisa para explorar a capacidade de identificação de metáforas com os atuais LLMs. De qualquer modo, a conclusão geral é que o ChatGPT parece capaz de identificar apenas uma em cada quatro ou seis das metáforas identificadas pelos analistas usando o protocolo MIP (Pragglejaz Group, 2007), considerando-se um fragmento de frase ou uma frase completa respectivamente.

A presente pesquisa buscou identificar os discursos subjacentes nas produções textuais de alunos bilíngues brasileiros e o uso de metáforas linguísticas nesses textos. Dessa forma, contribui para o entendimento dos discursos mobilizados pelos alunos na execução das tarefas, como a tarefa influencia o uso de metáforas linguísticas, assim como a relação entre a idade ou proficiência dos alunos e a incidência de metáforas. Contudo, para futuros trabalhos seria importante a verificação do nível de proficiência dos participantes previamente à coleta do corpus, para uma análise mais fidedigna.

Quanto à automatização da identificação de metáforas, a pesquisa aponta possíveis princípios para desenvolvimento de *prompt*, como restringir o contexto em que a metáfora linguística ocorre. Porém, a inconsistência de resultados e a aparente não identificação de preposições como Veículos são limitações que demandam mais pesquisas para o aperfeiçoamento de procedimentos automáticos de identificação.

## 6. REFERÊNCIAS

ACHIEVE. In: The Cambridge Dictionary. Disponível em:

<https://dictionary.cambridge.org/dictionary/english/achieve>. Acesso em: 6 abr. 2024.

AGHAZADEH, E.; FAYYAZ, M.; YAGHOOBZADEH, Y. Metaphors in pre-trained language models: probing and generalization across datasets and languages. In *Proceedings of the 60th annual meeting of the Association for Computational Linguistics*. Dublin, Ireland: Association for Computational Linguistics, v. 1, p. 2037–2050, 2022.

ALLEGORY. In: The Cambridge Dictionary. Disponível em:

<https://dictionary.cambridge.org/dictionary/english/allegory>. Acesso em: 4 abr. 2024.

ASIDE. In: The Cambridge Dictionary. Disponível em:

<https://dictionary.cambridge.org/dictionary/english/aside>. Acesso em: 6 abr. 2024.

BARDEN, J. Metaphor and Artificial Intelligence: why they matter to each other. In: GIBBS, R. (Ed.), *The Cambridge handbook of metaphor and thought*. Cambridge: Cambridge University Press. 2008. DOI:10.1017/CBO9780511816802.020

BERBER SARDINHA, T. *Linguística de corpus*. Barueri: Malone. 2004.

BERBER SARDINHA, T. *Metáfora*. São Paulo: Parábola. 2007.

BERBER SARDINHA, T. Questões metodológicas de análise de metáfora na perspectiva da linguística de corpus. *Gragoatá*, n. 26, p. 81-102, 1. sem. 2009.

BERBER SARDINHA, T. A abordagem metodológica da análise multidimensional. *Gragoatá*, n. 29, p. 107-125, 2. sem. 2010.

BERBER SARDINHA, T. Metáforas e linguística de corpus: metodologia de análise aplicada a um gênero de negócios. *D.E.L.T.A.*, v. 21, n. 1, p. 1-20, 2011.

BERBER SARDINHA, T. A ocorrência de metáforas é previsível? *Revista de estudos da linguagem*, v. 20, n. 2, p. 211-240, jul./dez. 2012.

BERBER SARDINHA, T. An assessment of metaphor retrieval methods. In: MACARTHUR, F.; ONCINS-MARTÍNEZ, L.; GARCÍA, M.S.; PIQUER-PÍRIZ, A.M. (Eds), *Metaphor in use: context, culture, and communication*. Joahn Benjamins: Amsterdam, p. 21-50, 2012.

BERBER SARDINHA, T. Dimensions of variation across Internet registers. *International Journal of Corpus Linguistics*, v. 23, n. 2, p. 125–157, 2018.

BERBER SARDINHA, T. Using Multi-Dimensional analysis to detect representations of national cultures. In: BERBER SARDINHA, T.; VEIRANO PINTO, M. (Eds.). *Multi-Dimensional Analysis: research methods and current issues*. London/New York: Bloomsbury, p. 231–258, 2019.

BERBER SARDINHA, T. Discourse of academia from a multidimensional perspective. In FRIGINAL, E.; HARDY, J. A. (Eds). *The Routledge handbook of corpus approaches to discourse analysis* Routledge. Londrês: Routledge, p. 298-318, 2020.

BERBER SARDINHA, T.; KAUFFMANN, C.H.; ACUNZO, C. A multidimensional analysis of register variation in Brazilian Portuguese. *Corpora*, v.9, n.2, p. 239-271, 2014.

BERBER SARDINHA, T.; VEIRANO PINTO, M. (org.) *Multidimensional Analysis*. Bloomsbury Academic, 2019

BERBER SARDINHA, T. RESENDE, S. SHIMAZUMI, M. Linguística de Corpus. In: OTHERO, G.A.; FLORES, V. N. (org.). *A linguística hoje: múltiplos domínios*. São Paulo: Editora Contexto. 2023. p. 103-117.

BERBER SARDINHA, T.; FITZSIMMONS-DOOLAN, S. *Lexical multidimensional analysis*. Cambridge: Cambridge University Press. 2024.

BIBER, D. *Variation across speech and writing*. Cambridge: Cambridge University Press. 1988.

BIBER, D. Representativeness in corpus design. *Literary and Linguistic Computing*, v. 8, n. 4, p.243-257. 1993a.

BIBER, D.; CONRAD, S.; REPPEN, R. *Corpus Linguistics investigating language structure and use*. Cambridge: Cambridge University Press. 1998

BIBER, D.; CONRAD, S. *Register, genre and style*. Cambridge: Cambridge University Press. 2009

BIBER, D.; JOHANSSON, S.; LEECH, G.N.; CONRAD, S.; FINEGAN, E. *Grammar of spoken and written English*. Amsterdam/Philadelphia: John Benjamins Publishing Company. 2021.

BOERS, F. Enhancing metaphoric awareness in specialised reading. *English for Specific Purposes*, v. 19, n. 2, p. 137–147, 2000.

BOERS, F. Expanding learners' vocabulary through metaphor awareness: what expansion, what learners, what vocabulary. In: ACHARD, M; NIEMEIER, S. (Eds.), *Cognitive linguistics, second language acquisition and foreign language teaching*. Berlin, New York: De Gruyter, 2004. p. 211–232.

BROGINI, A.F. *Representações contemporâneas da sustentabilidade: uma análise multidimensional lexical discursiva como contribuição para o Portal multimodal/multilíngue para avanço da Ciência Aberta nas Humanidades*. 2023. Dissertação (Mestrado em Linguística Aplicada e Estudos da Linguagem) - Pontifícia Universidade Católica de São Paulo, São Paulo, 2023.

BRYLSBAERT, M.; WARRIMER, A.B.; KUPERMAN, V. Concreteness ratings for 40 thousand generally known English word lemmas. *Behav Res*, 46, 904–911. 2013. <https://doi.org/10.3758/s13428-013-0403-5>

CAMERON, L. Operationalising 'metaphor' for applied linguistic research. In: CAMERON, L.; LOW, G. (Eds.), *Researching and applying metaphor*. Cambridge: Cambridge University Press, p. 3–28, 1999a.

CAMERON, L. Identifying and describing metaphor in spoken discourse data. In: CAMERON, L.; LOW, G. (Eds.), *Researching and applying metaphor*. Cambridge: Cambridge University Press, p. 105–132, 1999b.

CAMERON, L. *Metaphor in educational discourse*. London: Continuum. 2003.

CAMERON, L.; DEIGNAN, A. Combining large and small corpora to investigate tuning devices around metaphor in spoken discourse. *Metaphor and Symbol*, v.18, n. 3, p.149-160, 2003. DOI 10.1207/S15327868MS1803\_02

CAMERON, L.; LOW, G. Metaphor. *Language Teaching*, 32, p. 7796. 1999. DOI 10.1017/S0261444800013781

CAMERON, L.; LOW, G. Metaphor and the Social World: introduction to the first issue. *Metaphor and the Social World*, v.1, n.1, p. 1–5, 2011, DOI: 10.1075/msw.1.1.01cam

CAMERON, L.; MASLEN, R.; TODD, Z.; MAULE, J.; STRATTON, P.; STANLEY, N. The discourse dynamics approach to metaphor and metaphor-led discourse analysis. *Metaphor and Symbol*, v. 24, n. 2, p. 63-89, 2009. DOI: 10.1080/10926480902830821

CAMERON, L.; MASLEN, R. Identifying metaphor in discourse data. In ———. (Eds) *Metaphor analysis: research practice in Applied Linguistics, Social Sciences and Humanities*. London: Equinox. p. 97-115. 2010.

CANTOS-GOMEZ, P. Multivariate statistics commonly used in multi-dimensional analysis. In BERBER SARDINHA, T.; VEIRANO PINTO, M. (Eds). *Multi-Dimensional Analysis: research methods and current issues*. Bloomsbury Academic, 2019.

CELANI, M. Afinal, o que é Linguística Aplicada? In: ZANOTTO, M.; CELANI, M. *Linguística aplicada: da aplicação da linguística à linguística transdisciplinar*. São Paulo, EDUC, p. 15-23. 1992.

COLL-FLORIT, M.; CLIMENT, S. Enemies or obstacles? metaphors of war and journey in mental health discourse. *Metaphor and the Social World*, v. 12, n. 2, p. 181-203, Out. 2022. DOI: <https://doi.org/10.1075/msw.21035.col>

COBB, T.; HORST, M. Learner corpora and lexis. In: GRANGER, S.; GILQUIN, G.; MEUNIER, F. (Ed.), *The Cambridge handbook of learner corpus research*. Cambridge: Cambridge University Press, p. 185 -206, 2015. DOI: <https://doi.org/10.1017/CBO9781139649414.001>

CRAMER, D.; HOWITT, D. *The SAGE dictionary of statistics: a practical resource for students in the social sciences*. London: SAGE Publications. 2004.

CULTURE. *In: The Cambridge Dictionary*. Disponível em:  
<https://dictionary.cambridge.org/dictionary/english/culture>. Acesso em: 7 abr. 2024

DARIN, L. C. M. A tradução cultural como metáfora. *Revista Intercâmbio*, v. XLIII: 47-66, 2020. São Paulo: LAEL/PUCSP.

DEIGNAN, A. Corpus-based research into metaphor. In: CAMERON, L.; LOW, G. (Eds.), *Researching and applying metaphor*. Cambridge: Cambridge University Press, p. 177–200, 1999a.

DEIGNAN, A. *Metaphor and corpus linguistics*. Amsterdam-Philadelphia: John Benjamins. 2005.

DEIGNAN, A. Corpus linguistics and metaphor. In GIBBS, R. (Ed.), *The Cambridge handbook of metaphor and thought*. Cambridge: Cambridge University Press, p. 280-294. 2008. E-book. DOI: 10.1017/CBO9780511816802.018

DEIGNAN, A. MIP, the corpus and dictionaries: what makes for the best metaphor analysis? *Metaphor and the Social World*, v. 5, n. 1, p.145 - 154, 2015. DOI:  
<https://doi.org/10.1075/msw.5.1.09dei>

DEIGNAN, A.; GABRYŚ, D.; SOLSKA, A. Teaching English metaphors using cross-linguistic awareness-raising activities. *ELT Journal*, v. 51, n. 4, p. 352–360. Oct. 1997. DOI <https://doi.org/10.1093/elt/51.4.352>

DELFINO, M. Análise multidimensional: os números na linguística. *Cadernos de Linguística*, v. 2, n. 4, e474, 2021.

DEMAND. *In: The Cambridge Dictionary*. Disponível em:  
<https://dictionary.cambridge.org/dictionary/english/demand>. Acesso em: 1 abr. 2024.

DIVORCE. *In: The Cambridge Dictionary*. Disponível em:  
<https://dictionary.cambridge.org/dictionary/english/divorce>. Acesso em: 7 abr. 2024.

DORST, G.A.; REIJNIERSE, W. A dictionary gives definitions, not decisions: on using a dictionary to identify the basic senses of words. *Metaphor and the Social World*, v. 5, n. 1, p. 137–144, 2015. DOI: <https://doi.org/10.1075/msw.5.1.08dor>

EKATERINA, S; TEUFEL, S.; KORHONEN, A. Statistical metaphor processing. *Computational Linguistics*, v. 39, n. 2, p. 301–353, 2013, DOI:  
[https://doi.org/10.1162/COLI\\_a\\_00124](https://doi.org/10.1162/COLI_a_00124)

FILLMORE, C. J. "Corpus linguistics" or "Computer-aided armchair linguistics". *Directions in corpus linguistics: proceedings of Nobel Symposium 82 Stockholm*, Ed. Jan Svartvik, Berlin, New York: De Gruyter Mouton, p. 35-60, 1992. DOI:  
<https://doi.org/10.1515/9783110867275.35>

GILQUIN, G. Learner corpora. In M. PAQUOT, M.; GRIES, S. (eds) *A practical handbook of corpus linguistics*. Cham: Springer. p. 283-303, 2020.

GIVE. *In: The Cambridge Dictionary*. Disponível em: <https://dictionary.cambridge.org/dictionary/english/give>. Acesso em: 6 abr. 2024.

GOLDEN, A. The relationship between topic and metaphor in second-language learners' essays. *Metaphor and the Social World*, v. 11, n. 2, p. 261–278, 2021. DOI: <https://doi.org/10.1075/msw.00018.gol>

GRANGER, S. The computer learner corpus: a versatile new source of data for SLA research. In: GRANGER, S. (Ed.) *Learner English on computer*. Addison Wesley Longman: London & New York, p. 3-18, 1998.

GRANGER, S.; GILQUIN, G.; MEUNIER, F. Introduction: learner corpus research – past, present and future. In: GRANGER, S.; GILQUIN, G.; MEUNIER, F.. (Ed.), *The Cambridge handbook of learner corpus research*. Cambridge: Cambridge University Press, p. 1-6, 2015. DOI: <https://doi.org/10.1017/CBO9781139649414.001>

HALLIDAY, M.A.K.; MATTHIESSEN, C. *An introduction to functional grammar*. London: Hodder Education. 2004.

HAN, Y.; PENG, Z.; CHEN, H. Bibliometric assessment of world scholars' international publications related to conceptual metaphor. *Front. Psychol.* 13:1071121, 2022. DOI: [10.3389/fpsyg.2022.1071121](https://doi.org/10.3389/fpsyg.2022.1071121)

HEALTH. *In: The Cambridge Dictionary*. Disponível em: <https://dictionary.cambridge.org/dictionary/english/health>. Acesso em: 6 abr. 2024.

HOANG, H. Metaphor and second language learning: the state of the field. *TESL-EJ*, v. 18, n. 2. Aug. 2014.

HYPOTHESIS. *In: The Cambridge Dictionary*. Disponível em: <https://dictionary.cambridge.org/dictionary/english/hypothesis>. Acesso em: 1 abr. 2024.

IN. *In: The Cambridge Dictionary*. Disponível em: <https://dictionary.cambridge.org/dictionary/english/in>. Acesso em: 4 abr. 2024.

INTERFERE. *In: The Cambridge Dictionary*. Disponível em: <https://dictionary.cambridge.org/dictionary/english/interfere>. Acesso em: 7 abr. 2024.

KITTAY, E. *Metaphor its cognitive force and linguistic structure*. Oxford: Clarendon Press. 1987.

LAKOFF, G.; JOHNSON, M. *Metaphors we live by*. Chicago: The University of Chicago Press. 1980/2003.

LAKOFF, G. The contemporary theory of metaphor. In: ORTONY, A. (Ed.). *Metaphor and thought*. Cambridge: Cambridge University Press. p. 202–251, 1993.

LAKOFF, G.; NÚÑEZ, R.E. *Where mathematics comes from: how the embodied mind comes into being*. New York: Basic Books. 2000.

LAZAR, G. Using figurative language to expand students' vocabulary. *ELT Journal*, v. 50, n. 1, p. 43–51, Jan, 1996. DOI <https://doi.org/10.1093/elt/50.1.43>

LEAD. *In*: Merriam-Webster Dictionary. Disponível em: <https://www.merriam-webster.com/dictionary/lead>. Acesso em: 4 abr. 2024.

LEAVE. *In*: The Cambridge Dictionary. Disponível em: <https://dictionary.cambridge.org/dictionary/english/leave>. Acesso em: 7 abr. 2024.

LARSON, T.; BERBER SARDINHA, T.; GRAY, B.; BIBER, D. Exploring early L2 writing development through the lens of grammatical complexity. *Applied Corpus Linguistics*, v. 3, n. 3, 2023. DOI: <https://doi.org/10.1016/j.acorp.2023.100077>

LITTLEMORE, J. Metaphoric competence: a language learning strength of students with a holistic cognitive style? *TESOL QUARTERLY*, v. 35, n. 3, p. 459-491, 2001.

LITTLEMORE, J.; LOW, G. Metaphoric competence, second language learning, and communicative language ability. *Applied Linguistics*. v. 27, n. 2, p. 268-294, 2006. DOI <https://doi.org/10.1093/applin/aml004>

LITTLEMORE, L.; KRENNMAYR, T.; TURNER, J.; TURNER, S. An investigation into metaphor use at different levels of second language writing. *Applied Linguistics*. v. 35, n. 2, p. 117–144. 2014. DOI [10.1093/applin/amt004](https://doi.org/10.1093/applin/amt004)

LOW, G. On teaching metaphor. *Applied Linguistics*. v. 9, n.2, p. 125-147, 1988. DOI <https://doi.org/10.1093/applin/9.2.125>

LOW, G. “This paper thinks...”: investigating the acceptability of the metaphor an essay is a person. *In*: CAMERON, L.; LOW, G. (Eds.), *Researching and applying metaphor*. Cambridge: Cambridge University Press. p. 221–248, 1999.

MacARTHUR, F. On using a dictionary to identify the basic senses of words. *Metaphor and the Social World*, v. 5, n. 1, p.124 - 136, 2015.

MAKE. *In*: The Cambridge Dictionary. Disponível em: <https://dictionary.cambridge.org/dictionary/english/make>. Acesso em: 6 abr. 2024.

McENERY, T; HARDIE, A. *Corpus linguistics: method, theory and practice*. Cambridge: Cambridge University Press. 2011.

McENERY, T; HARDIE, A. The history of corpus linguistics. *In*: ALLAN, K. (Ed.), *The Oxford handbook of the history of linguistics*. 2013. DOI [10.1093/oxfordhb/9780199585847.001.0001](https://doi.org/10.1093/oxfordhb/9780199585847.001.0001)

MEUNIER, F. Introduction to learner corpus research. IN: TRACY-VENTURA, N.; PAQUOT, M. (Eds), *The Routledge handbook of second language acquisition and corpora*. Londrês: Routledge, p. 23-36, 2021.

MIAO, F.; M. WAYNE, H. *Guidance for generative AI in education and research*. Unesco. 2023.

NACEY, S. Development of L2 metaphorical production. In: PIQUER-PÍRIZ, A.; ALEJO-GONZÁLEZ, R. (Ed). *Metaphor in foreign language instruction*. Berlin, Boston: De Gruyter Mouton, p. 173-198. 2020. DOI <https://doi.org/10.1515/9783110630367-009>

NACEY, S. Development of metaphorical production in learner language: a longitudinal perspective. *Nordic Journal of Language Teaching and Learning*, v. 10, n. 2, p. 272-297, 2022.

NACEY, S.; DORST, A.; KRENNMAYR, T.; W.; REIJNIERSE, G. (Eds). *Metaphor identification in multiple languages: MIPVU around the world*. Países Baixos: John Benjamins Publishing Company. 2019.

O'KEEFFE, A.; MCCARTHY, M.; CARTER, R. *From corpus to classroom: language use and language teaching*. Cambridge: Cambridge University Press. 2007.

OPINION. In: The Cambridge Dictionary. Disponível em: <https://dictionary.cambridge.org/dictionary/english/opinion>. Acesso em: 4 abr. 2024.

POLICY. In: The Cambridge Dictionary. Disponível em: <https://dictionary.cambridge.org/dictionary/english/policy>. Acesso em: 1 abr. 2024.

PRAGGLEJAZ GROUP. MIP: a method for identifying metaphorically used words in discourse. *Metaphor and Symbol*, v. 22, n.1, p. 1-39, 2007.

PROMOTE. In: The Cambridge Dictionary. Disponível em: <https://dictionary.cambridge.org/dictionary/english/promote>. Acesso em: 7 abr. 2024.

PUSH. In: The Cambridge Dictionary. Disponível em: <https://dictionary.cambridge.org/dictionary/english/push>. Acesso em: 6 abr. 2

RANKIN, T. Learner corpora and grammar. In: GRANGER, S.; GILQUIN, G.; MEUNIER, F.. (Ed.), *The Cambridge handbook of learner corpus research*. Cambridge: Cambridge University Press, p. 231-254, 2015. DOI: <https://doi.org/10.1017/CBO9781139649414.001>

REFERENCE. In: The Cambridge Dictionary. Disponível em: <https://dictionary.cambridge.org/dictionary/english/reference>. Acesso em: 4 abr. 20

RESENDE, S.; MAVERICK, R. Planejamento, compilação e organização de corpora. *VIII Escola Brasileira de Linguística Computacional & XIII Encontro de Linguística de Corpus*. São Paulo: Blucher, 2016.

RUNDELL, M.; FOX, G. *Macmillan English dictionary for advanced learners*. Oxford: Macmillan Education. 2002.

RUSSEL, S. Computer understanding of metaphorically used verbs. *American Journal of Computational Linguistics*. 1976

SACRIFICE. In: The Cambridge Dictionary. Disponível em:  
<https://dictionary.cambridge.org/dictionary/english/sacrifice>. Acesso em: 6 abr. 2024.

SCHRAMOWSKI, P., TURAN, C., ANDERSEN, N. ROTHKOPF, C.; KERSTING, K. Large pre-trained language models contain human-like biases of what is right and wrong to do. *Nat Mach Intell* 4, p. 258–268, 2022. DOI: <https://doi.org/10.1038/s42256-022-00458-8>

SCHUSTER, S.; MANNING, C. D. Enhanced English universal dependencies: An Improved Representation for Natural Language Understanding Tasks. In *LREC 2016*. 2016.

SEMINO, E. A sturdy baby or a derailing train? metaphorical representations of the Euro in British and Italian newspapers. *Text*, v. 22, n. 1, p. 107-139, 2002.

SHEPHERD, T. M. O estatuto da Linguística de corpus: metodologia ou área da Linguística? *Matraga - Revista Do Programa De Pós-Graduação Em Letras Da UERJ*, v.16, n. 24, 2009.

SHOW. In: The Cambridge Dictionary. Disponível em:  
<https://dictionary.cambridge.org/dictionary/english/show>. Acesso em: 4 abr. 2024.

SILVA, C. S. *Discursos xenofóbicos contra a China na rede social Twitter: Um estudo multidimensional multimodal*. Em elaboração. Tese (Doutorado em Linguística Aplicada e Estudos da Linguagem) - Pontifícia Universidade Católica de São Paulo, São Paulo, em elaboração.

SOCIETY. In: The Cambridge Dictionary. Disponível em:  
<https://dictionary.cambridge.org/dictionary/english/society>. Acesso em: 7 abr. 2024.

STEEN, G. Metaphor in discourse: towards a linguistic checklist for metaphor analysis. In CAMERON, L.; LOW, G. (orgs.), *Researching and Applying Metaphor*. Cambridge: Cambridge University Press, p. 81-104, 1999.

STEEN, G. The paradox of metaphor: why we need a three-dimensional model of metaphor. *Metaphor and Symbol*, v. 23, n.4, p. 213-241, 2008.

STEEN, G. Deliberate metaphor theory: basic assumptions, main tenets, urgent issues. *Intercultural Pragmatics*, v. 14, n. 1, 2017, pp. 1-24. DOI:  
<https://doi.org/10.1515/ip-2017-0001>

STEEN, G. J.; DORST, A. G.; HERRMANN, J. B.; KAAL, A. A.; KRENNMAYR, T.; PASMA, T. *A method for linguistic metaphor identification: From MIP to MIPVU*. John Benjamins. 2010.

SUCCESS. *In: The Cambridge Dictionary*. Disponível em:  
<https://dictionary.cambridge.org/dictionary/english/success>. Acesso em: 7 abr. 2024.

SZUDARSKI, P. Corpora and teaching vocabulary and phraseology. In: JABLONKAI, R. R.; CSOMAY, E. (Eds), *The Routledge handbook of corpora and English language teaching and learning*. Londrês: Routledge, p. 41- 55, 2023. DOI: <https://doi.org/10.4324/9781003002901>

TAGNIN, E.E.O. *O jeito que a gente diz*. Barueri: Disal Editora. 2013.

TRUTH. *In: The Cambridge Dictionary*. Disponível em:  
<https://dictionary.cambridge.org/dictionary/english/truth>. Acesso em: 4 abr. 2024.

USE. *In: The Cambridge Dictionary*. Disponível em:  
<https://dictionary.cambridge.org/dictionary/english/use>. Acesso em: 7 abr. 2024.

VEIGA, A. T. *As dimensões da fé: sete religiões mundiais em uma análise multidimensional lexical*. 2021. Tese (Doutorado em Linguística Aplicada e Estudos da Linguagem) - Pontifícia Universidade Católica de São Paulo, São Paulo, 2021.

VIERA, A. J.; GARRETT, J.M. Understanding interobserver agreement: the Kappa statistic. *Family Medicine*. v. 37, n. 5, p.360-363, 2005.

WACHOWIAK, L.; GROMANN, D. Does GPT-3 grasp metaphors? Identifying metaphor mapping with generative language models. *Proceedings of the 61st Annual meeting of the Association for Computational Linguistics*, v. 1, p. 1.018-1.032, 2023.

WAY. *In: The Cambridge Dictionary*. Disponível em:  
<https://dictionary.cambridge.org/dictionary/english/way>. Acesso em: 6 abr. 2024.

WINSTON, P. *Learning by hypothesizing and justifying transfer frames*. Massachusetts Institute of Technology Artificial Intelligence Laboratory. 1977.

## 7. ANEXO

### ANEXO - A - Corpus CABI Tarefas

Task Prompt code	Prompt da tarefa
1	For this task you will write an article according to our discussions of page 19 (music).
2	Write an essay based on one of the questions on Exercise 11, page 55.
2A	Why do rivers boil?
2B	How can remote sensing help archaeologists?
2C	What are the practical benefits of artificial satellites?
3	Find a topic of your interest and think of a problem that might be solved. It is important that it is a WHY/HOW question.
4	Writing - Song (Narrative story)
5	For this task, you will write an email to one of your classmates, and respond to the one you received. You will write your messages using your email and copy them here, using a Google Doc. The pairs will be randomly chosen during the class.
6	Write a short paragraph telling how you agree/disagree with the results you had from the Personality Test and the definition you wrote about your personality.
7	Find on the internet an example of a successful person (or group of people). Think about: - What did they do to become successful? - How easy/difficult was it for him/her to become successful? - How do they contribute to making a better world? - Do you imagine yourself as successful as the person you researched?
8	Write a short article (4 paragraphs) in which you tell what you expect of when you move out of your parents house. Be careful with conditional sentences. 1. Why moving out is complex 2. How do you think your generation will deal with this issue 3. What are the biggest challenges 4. Conclusions.
9	Writing a paragraph with illustrations

<p style="text-align: center;"><b>10</b></p>	<p>Follow the instructions on page 84, exercise 5 (Reading and Writing) for doing this activity.</p> <p>In this activity, you should write a five-paragraph essay. The first is an introduction, paragraphs 2, 3 and 4 are about the three topics to achieve success, and paragraph 5 is a conclusion.</p> <p>You must use the paragraph structure we studied (Topic sentence, illustrations and conclusion), and write no more than 7 lines in each paragraph.</p> <p><i>Ex 5: Write an essay that discusses the three main qualities that a person needs to have in order to achieve success. Refer to the work you did in Prepare to Write and Write on pages 82-84 as you organize your thoughts. Pay carefull attention to the main elements of a well-constructed paragraph - topic sentence, supporting details, and concluding sentence.</i></p>
<p style="text-align: center;"><b>11</b></p>	<p>Write a cause-and-effect essay about the use of social networking sites.</p> <p>Before you start, inform the type of essay you are writing (personal to general; immediate vs long-term; order of importance; or familiarity or interest). Check page 182 of Reading and Writing book if you are in doubt about this.</p> <p>In addition, be sure to use adverb clauses (page176) to express cause and effect/consequence.</p> <p>Your text must have 5 paragraphs:</p> <ul style="list-style-type: none"> <li>- The first is an introduction: Please, highlight your thesis statement!</li> <li>- In paragraphs 2, 3 and 4 you must develop your topic;</li> <li>- P5 is a conclusion.</li> </ul>
<p style="text-align: center;"><b>12</b></p>	<p>In this task, we must write a paragraph using the following words:</p> <p>In this...  However, it...  In addition,...  Thus,...</p>

13	<p>Dear students,  Write a narrative essay according to what we discussed in class.  Use the answers to the questions on page 209 as a starting point.  Your text must be 350-400 words, with 4 paragraphs.</p> <p><b>Questions:</b> What kind of work of art will I focus on?, What is the name of the work of art?, What is the artist's message?, What do I want to say in my essay?, What is the purpose of my essay?, Who is the artist?, Who went with me to see (hear) the work?, When did I see (hear) the work of art?, When was the work of art created?, Where did I see (hear) it?, Why did I want to see (hear) it?, Why did I want to see (hear) it?, Why do I want to write about it?, Why did the artist create it?, How did I feel when I saw (heard) it?</p>
14	<p>Dear students,  This is a writing task based on the IB rubrics. Before we do the activity itself, let's understand how it works.  Open the files attached to this activity, and read them carefully.  Then, write your text on the sheet provided by the teacher.</p> <p>File  Writing task: You have 60 minutes to complete this task. Use an appropriate text type from the options below. Write 450 to 600 words.</p> <p>Context:  Your school will implement a healthy food policy, with no industrialized or processed food. However, there has been a huge number of complaints from students before changes are implemented. As a class representative, you need to inform your classmates about the changes. Write a text in which you explain what will change, how it will affect your eating habits, and why it is important to implement such changes.</p>
15	<p>Opinion essay  Write an opinion essay about the importance of music.  Remember to use this paragraph structure: TOPIC SENTENCE, ILLUSTRATIONS/DEVELOPMENT, and CONCLUSION.  In the first paragraph, present why music is important. Provide a concrete reason for it.  In the second, third and fourth paragraphs, provide illustrations (examples) to support your thesis given in the first paragraph.  In the fifth paragraph, write a conclusion describing why the three reasons you gave relate to your thesis.</p>

<b>16</b>	<p>Dear Students, today we will work on writing a good introduction paragraph, and a good thesis statement. The activity is on Northstar 5 - Reading and Writing, pg 52</p> <p>Writing - Introductory Paragraph.</p> <p>After doing the activities from the book, write the introductory paragraph to the essay that answers the question “ Should we always tell the truth?”. Remember to practice using vocabulary and grammar from the unit.</p>
-----------	---

17	<p>Dear students,</p> <p>We will start discussing and reflecting about our 2nd term topic - Cultural Diversity.</p> <p>To do so, we read an article about a situation that is happening in France, France's hijab prohibition .We discussed some aspects regarding this subject.</p> <p>What you must do afterwards is read the UN's definition of cultural diversity( the beginning and articles 1-6). You will find not only a definition of cultural diversity, but also its connection to human rights and freedom.</p> <p>After doing so, you ought to reflect about some ideas: How is cultural diversity connected to democracy and development, according to the document?</p> <p>Why are cultural rights considered a part of human rights?</p> <p>What should be guaranteed so that cultural diversity is actually respected?</p> <p>As we debated the article last class, some ideas came up: How should the relationship between the State and culture be?</p> <p>Is it correct to compare one culture to another?</p> <p>Should we act according to the images we have about a certain culture?</p> <p>Should security be more important than the respect for cultural characteristics?</p> <p>What can we consider to be universal, when we have many different points of view?</p> <p>Knowing that we tend to fear what we don't know, how can we deal with cultural differences?</p> <p>Considering all that, you will write an essay answering the following question:</p> <p>“Should cultural diversity be essential in our society? If so, how can</p>
----	---

	<p>society make sure it is treated as a right to all peoples?”</p> <p>You must use the article we have read, what the UN says about cultural diversity, other examples you might think of and, of course, your own opinion about the subject.</p> <p>The essay should have at least 3 well developed paragraphs. Remember the characteristics of a hook and the thesis statement, as well as the expressions we have seen.</p>
<p><b>18</b></p>	<p>Northstar 5 - Reading and Writing - Unit 4, pg 99</p> <p>Express Opinions.</p> <p>1 - Choose one of the three topics in the book and write an essay answering it.</p> <p>Your text should have introduction, development and conclusion. It must include at least three words from the unit’s vocabulary. It must use the two texts we have read as examples. It must include your own personal opinion.</p> <p>2- After writing the text, review it, and identify:</p> <p>main ideas</p> <p>details</p> <p>grammar issues</p> <p>You should point these out using different colors, or Bold, Underlined, Italic.</p> <p>You are not supposed to change anything yet, just identify it. Make comments about your own production, recognizing what you did well and identifying what you need to improve.</p>

<p style="text-align: center;"><b>19</b></p>	<p style="text-align: center;"><b>GEORGE ORWELL AND ANIMAL FARM</b></p> <p>Dear students,</p> <p>Now that you know more about Orwell’s life, let us try to connect it to the novella we just read.</p> <p>This is your activity for today:</p> <p>1- Organize a mind map in which you: select which aspects of Orwell’s biography are relevant to understand Animal Farm; paraphrase the reasons he exposes for writing; select characteristics of Animal Farm that connect to Orwell’s life and literary project.</p> <p>2 - Based on the map, you will write an essay with the title “ George Orwell and Animal Farm”, in which you explain the connections between his biography, his literary ideas and his novella.</p> <p>Remember the characteristics of a good introductory paragraph to write your essay.</p>
<p style="text-align: center;"><b>20</b></p>	<p>Dear students,</p> <p>Taking into consideration what we have discussed so far, you will write an essay in which you:</p> <ul style="list-style-type: none"> <li>- define what an Allegory is, with your own words;</li> <li>- explain examples;</li> <li>- give your own exemples, should you have them.</li> </ul> <p>You need to present the planning of the text before the essay, although it can be in the same document.</p>
<p style="text-align: center;"><b>21</b></p>	<p>Use the question from the previous class as inspiration to write a paragraph.</p>
<p style="text-align: center;"><b>22</b></p>	<p>Following the class activity where the teacher read an absurd text about the Beatles, create your own Absurd Text on a topic of your choice. Following the model read by the teacher, choose a topic you are familiar with and write around 200 words about it, hiding 2 or 3 true statements in</p>

	the middle of lies.
23	<p>Optional essay - Compare contrast of Mi Cha's and Cora's experiences</p> <p>As discussed, I have given you the opportunity to earn up to 0,5 a point extra in the trimester grade.</p> <p>For this activity, use the outline from exercise 5 in the activity "Review of previous content" as a base for your essay. There you will find the thesis statement, the topics and supporting details for each of the body paragraphs and a conclusion.</p> <p>This is not a mandatory activity, however I will not accept any essays handed in after the due date - so pay attention to it!</p>
24	<p>Narrative/Descriptive Personal narrative</p> <p>What you need in your Story</p> <p>Requirements:</p> <ol style="list-style-type: none"> <li>1) At minimum, a 5 paragraph story, with at minimum 5 sentences per paragraph.</li> <li>2) Length: The above length minimum, but also no longer than 2 and a half pages.</li> <li>3) Plot: You need all the parts of Plot: Introduction (with conflict introduced), Rising Action, Climax, Falling Action, and Resolution</li> <li>4) The Narration of the story must be in 1st person (I, me, mine) and it must be from ONLY YOUR POINT OF VIEW.</li> <li>5) Description: You should aim to use good description throughout the story. Meaning, you should try to have around 3 of the 5 senses shown through SENSORY DETAIL descriptive wording in each of the paragraphs.</li> <li>6) Vocabulary: You must correctly use 5 of the 10 (the first 10!) 1st Trimester vocabulary words that we have covered so far in your story. The words should be highlighted in light blue.</li> <li>7) Grammar: You must use 3 Gerunds correctly after 3 of the Verbs of our 5 for the trimester. The verb, followed by the Gerund should be highlighted in yellow.</li> </ol>
25	<p>“How To” Project</p> <p>Be Inventive!</p> <p>But make sure someone can follow your instructions and create/do what you do!</p>

26	<p>Theme: Facing Life’s Obstacles</p> <p>Prompt: Now it is time to get more specific with our theme of “Overcoming Obstacles” ....You need to think of 1 famous person that has had to overcome obstacles in their life. You may have to research this some, but please don’t copy directly from the sources you found. Instead, you should paraphrase and use your own words. So, what do you need to do? You should do the following in a 3 paragraph answer/journal.</p> <p>-Briefly tell about this person and their background.</p> <p>-Then, you should tell about the obstacle/s they had to overcome and how they overcame the obstacles.</p> <p>-Finally, you should connect this to yourself in someway. It could be the same type of obstacle faced, how this inspires you, etc.</p>
27	<p>Theme: Facing Life’s Obstacles Prompt: Some examples of kinds of obstacles that people face are: physical, economic, emotional...and there could be other ones as well. So the 3 main questions for this writing are: What are some ways that people overcome obstacles? Also, what obstacles have you faced in your life? Finally, how have you overcome them? (You can/should combine this into one connected piece of writing...be inventive...think before you write, but don’t overthink this.)</p>
28	Persuasive Essay - 1st version
29	Create a character and write an imaginary biographical narrative
30	After finishing and revising your historical narrative about Otzi, the Iceman, copy and paste your text here.
31	<p>Use North Star RW- page 121;</p> <p>Choose one of the animals from the list (Koko the gorilla, Bimbo the killer whale, Betty the crow, Ruby the elephant, Akeakemai the dolphin, Michael the gorilla, Alex the parrot, Rio the sea lion);</p> <p>Research about it;</p> <p>Imagine you are a journalist and write a news story (200 words) about that animal. You may create the story;</p> <p>Use the facts in your research to enrich your story.</p>

## 8. APÊNDICE

### APÊNDICE A - Corpus CABI Metadados

Sexto Ano				
Aluno	Idade	Sexo	Quantidade de textos	Tarefas
st3_	12	F	3	29,30,31
st4_	12	M	3	28,29,30
st6_	12	M	1	30
st7_	12	M	2	29,31
st8_	12	F	1	28
st9_	12	M	2	29,30
st10_	11	F	1	31
st11_	11	F	3	29,30,31
st12_	não informada	F	2	28,31
st13_	12	F	1	29
st15_	12	F	1	29
st16_	12	F	2	28,31
st17_	12	F	1	31
st18_	12	M	1	28
st19_	não informada	M	3	29,30,31
st20_	não informada	M	3	28,29,30
st21_	11	M	1	28
st22_	12	M	1	28
st23_	12	F	2	29,31
st26_	12	M	1	29
st28_	não informada	F	1	29
st30_	11	M	1	29
st31_	12	F	2	29,31
st33_	12	F	2	28,30
st34_	12	F	3	28,29,30
st36_	12	M	1	30
st38_	não informada	F	3	28,29,31
st39_	12	M	1	29
st41_	12	M	2	29,31
st42_	12	M	3	28,29,31
st43_	12	M	1	29

st44_	não informada	M	1	31
st45_	12	M	2	28,30
st46_	12	F	1	28
st47_	não informada	F	2	28,29
st49_	11	F	1	28
st50_	12	F	1	28
st51_	11	F	1	31
st52_	12	F	1	28
st53_	12	M	2	28,29
st56_	12	M	1	30
st57_	12	M	1	28
st58_	não informada	F	2	28,30
st59_	12	M	3	28,29,30
st60_	não informada	F	1	31
st61_	12	M	2	29,31
st63_	12	M	2	28,31
st64_	12	F	1	28
st66_	12	F	2	30,31
st68_	13	F	2	29,30
st69_	12	F	2	28,29
st70_	12	M	3	28,30,31
st71_	12	M	1	28

<b>Sétimo Ano</b>				
<b>Aluno</b>	<b>Idade</b>	<b>Sexo</b>	<b>Quantidade de textos</b>	<b>Tarefas</b>
st1_	não informada	F	1	27
st2_	13	M	1	27
st3_	13	F	1	27
st4_	13	M	2	25,27
st5_	12	M	3	25,26,27
st7_	13	F	2	24,27
st8_	13	F	4	24,25,26,27
st9_	não informada	M	2	26,27
st10_	13	M	1	24
st11_	13	F	3	24,25,26
st12_	13	F	2	25,26

st14_	13	F	3	24,25,26
st15_	13	F	3	25,26,27
st16_	13	F	1	24
st17_	13	M	2	26,27
st18_	13	F	1	27
st19_	13	F	2	25,27
st20_	13	F	3	24,25,27
st21_	13	F	2	26,27
st22_	13	F	2	24,26
st23_	12	M	4	24,25,26,27
st24_	13	F	1	24
st25_	não informada	F	1	27
st26_	13	F	2	25,26
st27_	13	M	1	26
st28_	13	M	1	24
st29_	12	F	2	24,27
st30_	12	M	2	25,26
st33_	13	F	3	24,25,26
st34_	13	M	4	24,25,26,27
st35_	13	F	3	24,25,27
st36_	12	F	1	27
st37_	13	M	3	24,25,26
st38_	13	F	3	24,26,27
st39_	13	F	2	24,25
st40_	13	F	1	25
st41_	não informada	M	2	25,26
st42_	13	F	2	24,26
st43_	12	F	3	25,26,27
st44_	13	M	2	24,26
st45_	não informada	M	3	24,25,26
st47_	não informada	F	3	24,26,27

Oitavo Ano				
Aluno	Idade	Sexo	Quantidade de textos	Tarefas
st1_	14	F	1	16
st2_	14	F	3	16,21,23

st3_	não informada	F	1	19
st4_	14	F	3	16,21,22
st5_	não informada	F	2	19,21
st6_	não informada	M	3	18,20,21
st7_	não informada	F	4	16,19,22,23
st8_	13	F	2	21,23
st9_	13	M	3	18,21,23
st12_	14	M	5	17,19,21,22,23
st13_	não informada	F	4	16,17,18,20
st14_	13	M	1	21
st15_	14	M	3	16,17,20
st16_	14	M	2	19,20
st17_	14	M	5	16,17,18,21,22
st18_	14	M	2	16,22
st19_	14	M	1	17
st20_	não informada	F	1	20
st21_	13	F	2	21,23
st22_	14	F	2	21,22
st23_	15	M	3	20,21,22
st24_	não informada	M	4	17,19,21,22
st25_	14	M	2	17,22
st26_	14	M	4	16,17,19,21
st27_	14	M	2	16,17
st28_	14	M	4	16,18,19,21
st29_	13	M	3	20,21,22
st31_	13	F	6	18,19,20,21,22,23
st32_	14	F	4	17,20,21,23
st33_	14	F	4	16,17,18,21
st34_	14	F	2	19,22
st35_	14	M	1	18
st36_	14	M	2	20,21
st37_	não informada	M	1	20
st38_	13	F	2	21,23

Nono Ano				
Aluno	Idade	Sexo	Quantidade de textos	Tarefas
st1_	15	F	2	10,13
st2_	15	F	1	12
st3_	15	M	2	6,12
st4_	15	M	1	13
st5_	não informada	M	3	6,8,13
st6_	não informada	F	3	8,12,13
st7_	15	F	2	8,10
st8_	15	F	2	6,8
st9_	14	M	2	6,11
st10_	15	F	3	8,10,12
st11_	não informada	M	4	8,9,10,12
st12_	15	M	1	7
st14_	15	F	3	10,13,14
st15_	15	M	1	10
st16_	não informada	M	1	13
st17_	14	M	2	12,13
st18_	14	F	1	12
st19_	15	F	1	14
st20_	14	M	4	6,7,9, 13
st22_	15	F	3	7,13, 15
st23_	15	F	4	9,11,14,15
st24_	15	M	2	12,15
st25_	14	F	3	7,12,14
st26_	15	M	3	9,11,15
st27_	não informada	M	2	10,14
st28_	15	M	2	6,9
st29_	14	F	6	7, 9, 11, 12, 14,15
st30_	15	F	5	7,12,13,14,15
st31_	não informada	F	2	13,15
st32_	15	F	3	13,14,15
st33_	15	M	3	6,14,15
st34_	15	F	3	10,11,13
st35_	15	M	3	6,10,13
st36_	15	M	5	7,9,11,12,15
st37_	15	M	2	14,15

<b>Primeiro Ano</b>				
<b>Aluno</b>	<b>Idade</b>	<b>Sexo</b>	<b>Quantidade de textos</b>	<b>Tarefas</b>
st1_	16	M	6	1,2,3,4,5
st2_	16	F	6	1,2,3,4,5
st3_	16	F	6	1,2,3,4,5
st4_	não informada	F	5	1,2,3,5
st5_	16	F	5	1,2,3,4,5
st6_	16	F	6	1,2,3,4,5
st7_	16	M	6	1,2,3,4,5
st8_	16	M	5	1,3,4,5
st9_	15	F	5	1,2,4,5
st10_	15	F	6	1,2,3,4,5
st11_	16	F	5	1,2,3,4,5
st12_	15	M	5	1,2,3,5
st13_	16	F	5	1,2,3,4,5
st14_	15	F	4	1,3,5
st15_	16	M	2	1,2
st16_	16	M	6	1,2,3,4,5
st17_	15	M	5	1,2,4, 5
st18_	16	M	6	1,2,3,4,5
st19_	16	M	5	1,2,4,5
st20_	16	F	6	1,2,3,4,5