

UNIVERSIDADE ESTADUAL DE GOIÁS
CÂMPUS CORA CORALINA – SEDE CIDADE DE GOIÁS
PROGRAMA DE PÓS-GRADUAÇÃO *STRICTO SENSU* EM LÍNGUA LITERATURA E
INTERCULTURALIDADE

IASMIN KNUIVERS FRANCO MACIEL

**O USO DE *LEXICAL BUNDLES* DA ORALIDADE PARA O ENSINO DE INGLÊS:
UMA SISTEMATIZAÇÃO BASEADA EM CÓRPUS PARA CRIANÇAS NÃO
ALFABETIZADAS**

GOIÁS-GO

2024

IASMIN KNUIVERS FRANCO MACIEL

**O USO DE *LEXICAL BUNDLES* DA ORALIDADE PARA O ENSINO DE INGLÊS:
UMA SISTEMATIZAÇÃO BASEADA EM CÓRPUS PARA CRIANÇAS NÃO
ALFABETIZADAS**

Trabalho apresentado à Universidade Estadual de Goiás,
Câmpus Cora Coralina, Programa de Pós-Graduação *Stricto
Sensu* em Língua, Literatura e Interculturalidade; para
conclusão do programa, sob orientação do Prof. D.r Eduardo
Batista da Silva.

GOIÁS-GO

2024



TERMO DE AUTORIZAÇÃO PARA PUBLICAÇÃO DE TESES E DISSERTAÇÕES NA BIBLIOTECA DIGITAL (BDTD)

Na qualidade de titular dos direitos de autor, autorizo a Universidade Estadual de Goiás a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UEG), regulamentada pela Resolução, CsA nº 1.087/2019 sem ressarcimento dos direitos autorais, de acordo com a Lei nº 9.610/1998, para fins de leitura, impressão e/ou *download*, a título de divulgação da produção científica brasileira, a partir desta data¹. Estando ciente que o conteúdo disponibilizado é de inteira responsabilidade do(a) autor(a).

Dados da autora

Nome completo: IASMIN KNUIVERS FRANCO MACIEL

E-mail: iasminfknuivers@gmail.com

Dados do trabalho

Título: O USO DE *LEXICAL BUNDLES* DA ORALIDADE PARA O ENSINO DE INGLÊS: UMA SISTEMATIZAÇÃO BASEADA EM CÓRPUS PARA CRIANÇAS NÃO ALFABETIZADAS

Tipo:

Tese Dissertação

Curso/Programa: POSLLI

Concorda com a liberação documento

SIM NÃO

¹ Período de embargo é de até **um ano** a partir da data de defesa.

Goiás, 15 de março de 2024

Assinatura autora

Assinatura do orientador

CATALOGAÇÃO NA FONTE

Biblioteca Frei Simão Dorvi – UEG Câmpus Cora Coralina

M152u	<p>Maciel, Iasmin Knuiuers Franco.</p> <p>O uso de “lexical bundles” da oralidade para o ensino de inglês : uma sistematização baseada em córpis para crianças não alfabetizadas [manuscrito] / Iasmin Knuiuers Franco Maciel. – Goiás, GO, 2024.</p> <p>90 f. ; il.</p> <p>Orientador: Prof. Dr. Eduardo Batista da Silva.</p> <p>Dissertação (Mestrado em Língua, Literatura e Interculturalidade) – Câmpus Cora Coralina, Universidade Estadual de Goiás, 2024.</p> <p>1. Linguística córpis. 1.1. Ensino de inglês para crianças. 1.2. Lexical bundles. 1.3. Registro oral. I. Título. II. Universidade Estadual de Goiás, Câmpus Cora Coralina.</p> <p>CDU: 811.111:37</p>
-------	---

Bibliotecária responsável: Marília Linhares Dias – CRB 1/2971

UNIVERSIDADE ESTADUAL DE GOIÁS

(Criada pela lei nº 13.456 de Abril de 1999, publicada no DOE-GO de 20 de Abril de 1999)

Pró-Reitoria de Pesquisa e Pós-Graduação Coordenação
de Pós-Graduação Stricto Sensu

UEG CÂMPUS CORA CORALINA

Av. Dr. Deusdeth Ferreira de Moura Centro - GOIÁS CEP: 76600000
Telefones: (62)3936-2161 / 3371-4971 Fax: (62) 3936-2160 CNPJ: 01.112.580/0001-71

ATA DE EXAME DE DEFESA 01/2024

Aos vinte e três dias do mês de fevereiro de dois mil e vinte e quatro às dezesseis horas, realizou-se o Exame de Defesa da dissertação do(a) mestrando(a) Iasmin Knuiuers Franco Maciel, intitulado **“UMA SISTEMATIZAÇÃO DE *LEXICAL BUNDLES* PARA CRIANÇAS NÃO ALFABETIZADAS”**. A banca examinadora foi composta pelos seguintes professores: Dr. Eduardo Batista da Silva – Presidente – (POSLLI/UEG), Dra. Talita Serpa (UNESP), Dra. Cristiane Rosa Lopes (POSLLI/UEG). Os membros da banca fizeram suas observações e sugestões, as quais deverão ser consideradas pelo(a) mestrando(a) e seu/sua orientador(a). Em seguida, a banca examinadora reuniu-se para proceder a avaliação do exame de defesa. Reaberta a sessão, o(a) presidente da banca examinadora, proclamou o resultado, segundo o qual a dissertação foi (x) aprovada, () aprovada com ressalvas, () reprovada com as seguintes exigências (se houver): _____

Cumpridas as formalidades de pauta, às 18h, a presidência da mesa encerrou esta sessão do Exame de Defesa e lavrou a presente ata que, após lida e aprovada, será assinada pelos membros da banca examinadora.

Goiás-GO, 23 de fevereiro de 2024.



Prof. Dr. Eduardo Batista da Silva (POSLLI/UEG)



Profa. Dra. Talita Serpa (UNESP)



Prof. Dra. Cristiane Rosa Lopes (POSLLI/UEG)

“É realmente verdade que gratidão gera gratidão e lamúria atrai lamúria. Isto ocorre porque o coração agradecido comunica-se com as forças divinas, e o queixoso, com as forças malignas. Assim, quem vive agradecendo, torna-se naturalmente feliz e quem vive lamuriando, torna-se infeliz. [...] Alegrem-se que virão coisas alegres [...]”. (MOKITI OKADA, 2019, p. 53).

AGRADECIMENTOS

Neste momento, gostaria de expressar minha profunda gratidão àqueles que contribuíram para a realização deste trabalho acadêmico.

Primeiramente, quero expressar minha gratidão a Deus, fonte de força e inspiração. Sua orientação e bênçãos têm sido fundamentais em minha jornada acadêmica e pessoal.

Aos meus pais, Carlos e Morgana Franco, dedico minha profunda gratidão. Seus ensinamentos e amor moldaram quem sou hoje. A cada desafio superado, refletia a sabedoria e valores que me transmitiram, mostrando-me o caminho da verdade e do bem. Seu apoio constante é o alicerce dos meus sonhos. Obrigada, do fundo do coração.

Aos meus irmãos, Iúri e Ingrid Franco, agradeço pela fonte inesgotável de apoio, incentivo e inspiração que representam. Nossa ligação é meu tesouro mais precioso. Cada vitória é compartilhada em meu coração. Obrigada por serem parte essencial da minha jornada.

A meu marido Vinícius Maciel, meu companheiro de vida e meu amor, expresso gratidão especial. Sua presença é minha âncora e seu amor incondicional é a base dos meus sonhos. Este trabalho é dedicado a você, que me inspira a ser a melhor versão de mim mesma. Obrigada, meu amor.

Por fim, gostaria de estender meus agradecimentos à minha escola IKF Easy, English e a todos os professores, pais e colegas que contribuíram para o meu crescimento acadêmico.

Este trabalho acadêmico é o resultado de um esforço coletivo, e sou profundamente grata a todos que desempenharam um papel importante em minha jornada. Obrigada.

RESUMO

Esta pesquisa explora *lexical bundles* no ensino de inglês para crianças não alfabetizadas segundo a expectativa de alfabetização proposta pela Base Nacional Curricular (BNCC). *Lexical bundles* são sequências lexicais fixas que ocorrem frequentemente em um determinado registro linguístico. Nossa fundamentação teórica recorre aos trabalhos vinculados à Linguística de Córpus, incluindo Biber, Conrad, Cortes (2003), Romer (2008), Biber (2009), Berber Sardinha (2010) e Egbert et al. (2022), e à Interculturalidade, como abordado por Kramsch (2021). Nosso objetivo geral é criar uma lista abrangente de *lexical bundles* relevantes e aplicáveis na sala de aula de língua inglesa, com uma preocupação não apenas descritiva, mas também voltada ao ensino. O objetivo específico do presente estudo é identificar e analisar os padrões de *lexical bundles* presentes na língua falada direcionada para crianças em vídeos do *YouTube*. Para isso, compilamos um cörper composto por 1.940 textos, totalizando 1.722.722 milhões de palavras. A transcrição do conteúdo textual dos vídeos foi realizada por meio da plataforma online *YouTube Scribe*. Com o auxílio do software *Wordsmith Tools*, versão 5.0, 2021, processamos o cörper e elencamos os *lexical bundles* compostos por três e quatro itens. Como resultado, identificamos 135 *lexical bundles* típicos da comunicação destinada à criança, ou seja, padrões linguísticos orais em diferentes contextos e produzidos por diferentes interlocutores, como por exemplo "*what are you*", "*go to the*", "*what do you*", "*one, two, three*" e "*where are you*". Essas evidências linguístico-estatísticas têm o potencial de impactar a compreensão da linguagem destinada às crianças não alfabetizadas, fornecendo insumos didático-pedagógicos para o ensino de inglês na educação infantil, expondo as crianças a padrões linguísticos autênticos e recorrentes na língua.

Palavras-Chave: *Ensino de Inglês para crianças. Lexical bundles. Cörper. Registro oral.*

ABSTRACT

This research explores lexical bundles in English teaching for preliterate children according to the literacy expectations proposed by the National Curricular Base (BNCC). Lexical bundles are fixed lexical sequences that occur frequently in a specific linguistic register. Our theoretical framework draws on works related to Corpus Linguistics, including Biber, Conrad, Cortes (2003), Romer (2008), Biber (2009), Berber Sardinha (2010), and Egbert et al. (2022), and on Interculturality, as approached by Kramsch (2021). Our general objective is to create a comprehensive list of relevant and applicable lexical bundles in the English language classroom, with a concern not only descriptive but also focused on teaching. The specific objective of this study is to identify and analyze the patterns of lexical bundles present in spoken language directed towards children in YouTube videos. To do this, we compiled a corpus consisting of 1,940 texts, totaling 1,722,722 million words. The transcription of the textual content of the videos was carried out through the online platform YouTube Scribe. With the assistance of the software Wordsmith Tools, version 5.0, 2021, we processed the corpus and listed the lexical bundles composed of three and four items. As a result, we identified 135 lexical bundles typical of communication aimed at children, i.e., oral linguistic patterns in different contexts and produced by different interlocutors, such as "what are you", "go to the", "what do you", "one, two, three", and "where are you". These linguistic-statistical evidences have the potential to impact the understanding of language aimed at preliterate children, providing didactic-pedagogical inputs for the teaching of English in early childhood education, exposing children to authentic and recurring linguistic patterns in the language.

Key words: *English Language teaching for children. Lexical bundles. Corpus. Oral register.*

LISTA DE FIGURAS

Figura 1 - A relação entre Linguística de Córpus e ensino de línguas.....	36
Figura 2 - O uso de córpus no ensino e aprendizado de línguas.....	37
Figura 3 – Características da fala X da escrita.....	59
Figura 4 – <i>Screen shot</i> da página inicial do <i>YouTube</i>	62
Figura 5 - <i>Screen shot</i> da página inicial do <i>Ytscribe</i>	62
Figura 6 - <i>Screen shot</i> da página inicial do <i>Word Smith Tools</i>	63
Figura 7 - <i>YouTube transcript extension</i> – passo 1.....	68
Figura 8 - <i>YouTube transcript extension</i> – passo 2.....	68
Figura 9 - <i>YouTube transcript extension</i> – passo 3.....	69
Figura 10 - <i>YouTube transcript extension</i> – passo 4.....	69
Figura 11 - <i>YouTube transcript extension</i> – passo 5.....	70
Figura 12 – seleção dos vídeos – passo 1.....	70 71
Figura 13 – seleção dos vídeos – passo 2.....	71 72
Figura 14 – seleção dos vídeos – passo 3.....	73 73
Figura 15 – seleção dos vídeos – passo 4.....	74 74
Figura 16 – seleção dos vídeos – passo 5.....	75 75
Figura 17 – seleção dos vídeos – passo 6.....	passo 6.....
Figura 18 – seleção dos vídeos – passo 7.....	passo 7.....
Figura 19 – <i>Word Smith Tools</i> – passo 1.....	passo 1.....
Figura 20 – <i>Word Smith Tools</i> – passo 2.....	passo 2.....
Figura 21 – <i>Word Smith Tools</i> – passo	passo

3.....	
Figura 22 – <i>Word Smith Tools</i> – passo 4.....	76
Figura 23 – <i>Word Smith Tools</i> – passo 5.....	76

LISTA DE QUADROS

Quadro 1 – Professor como “filtro”.....	26
Quadro 2 - Aplicações pedagógicas diretas de cópús específico e geral	38
Quadro 3 - Aplicações pedagógicas indiretas de cópús específico e geral	39
Quadro 4 - Conceituação de <i>lexical bundles</i>	40
Quadro 5 - <i>Lexical bundles</i> específicos.....	77
Quadro 6 - <i>Lexical bundles</i> em comum.....	78
Quadro 7 – Comparação entre a presente pesquisa e Biber, Conrad e Cortes (2003) quanto à <i>lexical bundles</i> presentes/ausentes.....	81
Quadro 8 – Comparação de <i>lexical bundles</i> presentes/ausentes em Biber, Conrad e Cortes (2003)	81
Quadro 9 – Comparação entre a presente pesquisa e Biber (2009) quanto à <i>lexical bundles</i> presentes/ausentes.....	83
Quadro 10 – Comparação de <i>lexical bundles</i> presentes/ausentes em Biber (2009).....	

SUMÁRIO

1 CONTEXTUALIZAÇÃO SOCIOCULTURAL.....	10
2 INTRODUÇÃO.....	14
3 FUNDAMENTAÇÃO TEÓRICA.....	26
3.1 Linguística de Córpus.....	27
3.2 Linguística de Córpus no ensino.....	28
3.3 <i>Lexical Bundles</i>	40
3.4 Interculturalidade.....	50
3.5 Fala x Escrita.....	56
4 MATERIAL.....	62
4.1 <i>YouTube</i>	63
4.2 <i>YouTube Scribe</i>	64
4.3 Word Smith Tools.....	64
4.4 Córpus.....	65

5 MÉTODO.....	68
6 RESULTADOS E ANÁLISE DOS DADOS	77
7 CONSIDERAÇÕES FINAIS	85
REFERÊNCIAS	

1 CONTEXTUALIZAÇÃO SOCIOCULTURAL

O termo "sociocultural" refere-se à interseção entre dois aspectos fundamentais da vida humana: o social e o cultural. Esse termo é utilizado em diversos campos do conhecimento, incluindo antropologia, sociologia, psicologia, educação e outros, para descrever a influência e interconexão entre os fatores sociais e culturais na formação e compreensão de diferentes aspectos da sociedade e da experiência humana.

O "aspecto social" abrange a estrutura, interações e dinâmicas de uma sociedade, incluindo normas, instituições, hierarquias, relações entre grupos e indivíduos, bem como sistemas políticos e econômicos. Por outro lado, o "aspecto cultural" engloba os valores, crenças, língua, tradições, expressões artísticas e formas simbólicas compartilhadas por uma comunidade ou grupo. A cultura molda identidades, comportamentos e perspectivas individuais, enquanto o aspecto social define os contextos e estruturas nos quais essas culturas se desenvolvem e se manifestam. Juntos, esses aspectos influenciam significativamente a maneira como as pessoas percebem, interagem e dão sentido ao mundo ao seu redor, fornecendo uma base para compreendermos a complexidade e a riqueza das interações humanas.

"Contextualização sociocultural" refere-se ao processo de situar um fenômeno, evento, conceito ou objeto dentro de seu contexto social e cultural mais amplo. Isso envolve considerar as influências sociais e culturais que moldam e influenciam o significado, a interpretação e as práticas relacionadas a esse fenômeno.

A contextualização sociocultural no contexto deste trabalho envolveria considerar não apenas os padrões linguísticos específicos dos *lexical bundles* destinados a crianças não alfabetizadas, mas também os contextos sociais e culturais em que esses padrões surgem e são utilizados. Isso pode incluir aspectos como o ambiente familiar e escolar das crianças, as interações sociais cotidianas em que a linguagem é empregada, as influências culturais e sociais que moldam as formas de comunicação infantil, e até mesmo as tecnologias e mídias contemporâneas, como vídeos do YouTube, que podem influenciar a exposição linguística das crianças.

Portanto, ao explorar *lexical bundles* para crianças não alfabetizadas, é essencial considerar não apenas os aspectos linguísticos, mas também as dimensões sociais e culturais que permeiam o uso e desenvolvimento desses padrões linguísticos, proporcionando uma compreensão mais abrangente e contextualizada do fenômeno em questão.

Em resumo, contextualização sociocultural refere-se à compreensão e interpretação de algo dentro do contexto mais amplo de sua sociedade e cultura, reconhecendo as complexas interações entre o indivíduo e seu ambiente social e cultural.

A abordagem sociocultural reconhece que tanto o contexto social quanto o cultural desempenham um papel fundamental na formação das experiências individuais e coletivas, nas percepções, nos processos cognitivos, na aprendizagem, no desenvolvimento humano e na compreensão das interações humanas. Por exemplo, na educação, essa abordagem reconhece a importância do contexto cultural e social na forma como os alunos aprendem e se desenvolvem.

Em resumo, "sociocultural" refere-se à compreensão de que aspectos sociais e culturais estão intrinsecamente interligados e influenciam a forma como as pessoas vivem, se comportam, aprendem, se relacionam e dão significado ao mundo ao seu redor.

Em relação aos temas de "aspecto social" e "aspecto cultural", deparamo-nos com instituições ao redor do mundo que desempenham um papel importante na promoção e no entendimento desses conceitos. Abordaremos sobre o British Council, uma instituição internacional originada do Reino Unido, que possui uma missão clara: promover a compreensão intercultural e reforçar os laços entre o Reino Unido e outras nações ao redor do mundo. Fundado em 1934, o British Council é uma organização sem fins lucrativos ativa em mais de 100 países, colaborando estreitamente com governos, instituições educacionais, entidades culturais e a sociedade civil.

No ensino da língua inglesa, o British Council oferece cursos variados, programas de capacitação para professores do idioma e certificações de proficiência. Além disso, trabalha em parceria com escolas e instituições para aprimorar os métodos de ensino. A organização promove programas de intercâmbio estudantil, oportunidades para artistas e profissionais culturais, bem como eventos que celebram e compartilham a riqueza da cultura britânica e internacional. Outra área de atuação crucial é o apoio à educação e desenvolvimento. O British Council estabelece parcerias sólidas com instituições educacionais para fortalecer os sistemas de ensino, compartilhar as melhores práticas e apoiar o desenvolvimento educativo em diversas comunidades.

O British Council publicou relatórios sobre o índice de proficiência em inglês no Brasil através do EF English Proficiency Index (EF EPI). O EF EPI é um estudo anual que avalia e classifica a proficiência em inglês de diferentes países com base em testes de proficiência de idiomas. Em geral, o Brasil historicamente tem sido classificado como tendo um nível de proficiência em inglês considerado intermediário ou intermediário-baixo. Isso

significa que muitos brasileiros têm conhecimentos básicos ou moderados no idioma, mas podem enfrentar dificuldades para se comunicar fluentemente ou compreender aspectos mais complexos da língua.

No entanto, é essencial considerar que a proficiência em inglês pode variar significativamente em diferentes regiões do Brasil e entre diferentes grupos demográficos, como idade, nível socioeconômico, educação e acesso a oportunidades de aprendizado de idiomas. Além disso, iniciativas privadas, como escolas de idiomas e cursos online, também desempenham um papel fundamental na oferta de oportunidades de aprendizado de inglês para os brasileiros. É fundamental continuar investindo em educação e programas que promovam o aprendizado de idiomas, como o inglês, para aumentar a proficiência da população brasileira, o que pode trazer benefícios significativos em termos de oportunidades de emprego, educação e intercâmbio cultural com o mundo globalizado.

Para os dados que serão apresentados a seguir, foram utilizadas múltiplas fontes de dados. Inicialmente, foram examinados dados públicos do IBGE, Ministério do Trabalho e informações do Data Popular. Em seguida, realizou-se uma pesquisa qualitativa, incluindo entrevistas detalhadas com profissionais de RH, especialistas e agentes governamentais, juntamente com grupos de discussões envolvendo indivíduos de 25 a 35 anos da classe média. Complementando, houve uma fase quantitativa, ouvindo 720 pessoas de 18 a 55 anos, representando diferentes estratos sociais em todo o país, que têm interesse ou estão realizando cursos de inglês. As etapas qualitativa e quantitativa ocorreram em 2013.

É importante citar que “no Brasil, 5,1% da população de 16 anos ou mais afirma possuir algum conhecimento do idioma inglês. Existem, porém, diferenças entre as gerações. Entre os mais jovens, de 18 a 24 anos, o percentual dos que afirmam falar inglês dobra, chegando a 10,3% das pessoas nessa faixa etária” (COUNCIL, 2014, p. 7).

É importante notar que esse levantamento se concentra nas faixas etárias a partir dos 16 anos, e não há menção específica sobre estudos voltados para crianças mais novas. Isso ressalta uma lacuna significativa na compreensão do nível de conhecimento ou exposição à língua inglesa entre as crianças brasileiras antes da faixa etária de 16 anos. Frente à ausência de trabalhos anteriores, procuramos contribuir e lançar luz para orientar um procedimento teórico metodológico.

A falta de dados ou estudos dedicados ao conhecimento de inglês entre crianças no Brasil é uma questão que merece atenção, pois a exposição precoce a idiomas estrangeiros pode ser benéfica para o desenvolvimento linguístico e cognitivo. Estudos sugerem que a

aprendizagem de idiomas desde a infância pode ter impactos positivos no desenvolvimento global das crianças, além de facilitar a aquisição de habilidades linguísticas.

Portanto, encontramos-nos defasados quanto a obrigatoriedade da língua inglesa antes do Ensino Fundamental, como citado em Brasil (2020, p. 19):

A Base Nacional Comum Curricular (BNCC) estipulou a língua inglesa como a única língua adicional obrigatória a partir do 6º ano do Ensino Fundamental. A BNCC estabeleceu campos de experiência para a Educação Infantil e campos de atuação para o Ensino Fundamental e o Ensino Médio.

A Base Nacional Comum Curricular (BNCC) estabelece os direitos de aprendizagem e desenvolvimento para todos os estudantes da educação básica, mas, até o momento, não definiu objetivos específicos de proficiência na língua inglesa para os diferentes níveis de ensino.

Além disso, a ausência de referências específicas sobre a proficiência dos professores de inglês pode ser problemática. A formação e a capacitação dos docentes são componentes essenciais para garantir a qualidade do ensino de línguas estrangeiras. A falta de diretrizes claras sobre as habilidades e competências que os professores devem possuir para lecionar inglês pode dificultar o desenvolvimento profissional e aprimoramento pedagógico desses profissionais. Assim como mencionado no Parecer CNE/CEB Nº: 2/2020, p. 20: “Entretanto, a BNCC não estipulou objetivos de aprendizagem para a língua inglesa em termos de proficiência a ser atingida na educação básica. E ainda não temos referências específicas sobre a proficiência dos professores”.

Nesse contexto, é importante que sejam estabelecidos critérios e referências que apoiem tanto os estudantes na busca por proficiência na língua inglesa quanto os professores na busca por qualificação e aprimoramento de suas práticas pedagógicas. A definição de objetivos de aprendizagem e referências de proficiência na BNCC para a língua inglesa poderia contribuir significativamente para a melhoria do ensino e da aprendizagem desse idioma nas escolas brasileiras.

Portanto, a ausência de dados específicos sobre o conhecimento de inglês entre crianças no Brasil destaca a importância de investigações e pesquisas direcionadas a esse grupo etário, permitindo uma compreensão mais abrangente e um melhor planejamento de políticas educacionais voltadas para a aprendizagem de idiomas desde os primeiros anos de vida.

2 INTRODUÇÃO

Mais do que nunca, aprender inglês é uma habilidade necessária, tendo em vista o mundo globalizado moderno, e apesar da notória necessidade de estudar língua inglesa, segundo uma pesquisa realizada pelo Instituto de Pesquisa Data Popular para o British Council, observa-se que no Brasil, 5,1% da população de 16 anos ou mais afirma possuir algum conhecimento do idioma inglês. Levando em consideração crianças da educação infantil, não há registros de estudos sobre o conhecimento da língua para crianças não alfabetizadas.

Conforme citado em Cronquist, K; Fiszbein, A., 2017 apud Brasil, 2020, p. 17:

Um estudo sobre a qualidade do aprendizado de língua inglesa na América Latina identificou avanços e desafios nas escolhas feitas em países como Argentina, Brasil, Chile, Colômbia, Costa Rica, Equador, México, Panamá, Peru e Uruguai. Esses países concentravam 84% da população e do 87% do Produto Interno Bruto (PIB) da região em 2015. Mesmo sendo marcadamente hispanófono, 68% dos executivos apontam o inglês como a principal língua dos negócios, contra 6% do espanhol e 8% do Mandarim.

Os números apresentados sobre a preferência do inglês pelos executivos para assuntos comerciais (68%), em contraste com o espanhol (6%) e o Mandarim (8%), ressaltam a importância estratégica do domínio do inglês no ambiente empresarial na região latino-americana. Essa constatação aponta para a necessidade de investimentos e políticas públicas que promovam o ensino e aprofundamento do inglês nas escolas e nas áreas profissionais, visando preparar indivíduos para competir em um mercado de trabalho cada vez mais globalizado.

Não obstante da explícita necessidade dos estudos atualmente, observa-se que o estudo da língua inglesa é realizado por apenas uma taxa da população de 1%. Quanto ao ensino de inglês especificamente para crianças, encontramos muitos profissionais não qualificados exclusivamente para tal faixa etária. Entende-se que não pode ser professor aquele que apenas possui proficiência na língua, pois o ensino de língua estrangeira depende de diversos outros fatores, e enfrentamos diversas dificuldades, tais como: “pouca estrutura para um ensino adequado da língua e turmas com número elevado de alunos. Somam-se a isso a carga horária insuficiente e a dificuldade de encontrar professores com formação adequada” (COUNCIL, 2014, p. 12).

Atualmente, encontramos em sala de aula de língua inglesa na educação infantil profissionais que ensinam o que livros didáticos propõe, e não se utilizam da práxis para unir a teoria à prática. Editoras produzem livros a partir do que acham ser útil ensinar, e não possuem estudos prévios com seleção de *lexical bundles* mais recorrentes no dia a dia infantil.

Lidamos com uma grande defasagem do aprendizado de língua inglesa para crianças, o que nos leva a graves problemas educacionais. Contudo, o presente trabalho visa contrapor tal situação por meio dos *lexical bundles* agregados ao ensino para crianças não alfabetizadas. A partir da presente pesquisa criaremos um portfólio de vocabulário visual de língua inglesa para crianças não alfabetizadas baseado em evidências por meio dos *lexical bundles*.

A citação "nos dois primeiros anos do Ensino Fundamental, a ação pedagógica deve ter como foco a alfabetização, a fim de garantir amplas oportunidades para que os alunos se apropriem do sistema de escrita alfabética" (BRASIL, 2017, p. 59) destaca a importância da alfabetização durante os primeiros anos da educação formal. Isso implica que as crianças não alfabetizadas mencionadas em nossa pesquisa estão dentro da faixa etária recomendada e considerada pelo documento da Base Nacional Comum Curricular (BNCC). De acordo com a BNCC, nessa fase inicial do Ensino Fundamental, é crucial fornecer um ambiente propício para que os alunos desenvolvam habilidades de leitura e escrita, possibilitando-lhes a apropriação do sistema de escrita alfabética. Portanto, ao nos referirmos a crianças não alfabetizadas, estamos direcionando nosso foco para aquelas que se encontram nessa fase inicial da educação formal, conforme preconizado pela BNCC, onde a ênfase pedagógica está na alfabetização como base fundamental para seu desenvolvimento educacional.

Levemos em consideração a citação: "embora, desde que nasce e na Educação Infantil, a criança esteja cercada e participe de diferentes práticas letradas, é nos anos iniciais (1º e 2º anos) do Ensino Fundamental que se espera que ela se alfabetize" (BRASIL, 2017, p. 89). Tal citação enfatiza que, embora as crianças sejam expostas a diversas práticas relacionadas à linguagem escrita desde a primeira infância e durante a Educação Infantil, é nos primeiros anos do Ensino Fundamental que se espera que elas efetivamente se alfabetizem. Isso ressalta a importância dos primeiros anos da escolaridade formal como um período crucial para o desenvolvimento da alfabetização, onde há uma transição de práticas letradas mais informais para um ensino mais estruturado e sistemático da leitura e escrita. Portanto, essa passagem sublinha a relevância do trabalho pedagógico nesses anos iniciais do Ensino Fundamental, destacando o papel fundamental desses anos no processo de alfabetização das crianças.

Este trabalho pretende inovar a abordagem no ensino de língua inglesa para crianças não alfabetizadas, abordando a defasagem identificada no aprendizado. A proposta de incorporar *lexical bundles* ao currículo destina-se a proporcionar uma base sólida e acessível para o desenvolvimento linguístico precoce. Os *lexical bundles*, por meio de expressões e padrões linguísticos frequentemente utilizados, visam facilitar a compreensão e retenção do vocabulário essencial. A construção de um portfólio de vocabulário visual fundamentado em evidências derivadas desses bundles busca não apenas enriquecer o repertório linguístico das crianças, mas também criar um ambiente de aprendizado interativo e lúdico. Ao adotar essa abordagem, a intenção é não apenas superar desafios educacionais, mas também promover um engajamento mais efetivo e duradouro no processo de aprendizagem da língua inglesa desde a infância.

Lexical bundles são combinações de três ou mais palavras que são identificadas empiricamente em um *corpus*¹ de linguagem natural. Biber et al. (1999) define *lexical bundles* como combinações de palavras que de fato se repetem mais comumente em um dado registro, e como um tipo muito particular de combinação de palavras. No trabalho de Biber et al. (1999), os *lexical bundles* foram descritos de maneira simples como sendo as sequências de palavras recorrentes mais comuns. No estudo de Biber, Conrad e Cortes (2003), os *lexical bundles* são conjuntos de palavras que são altamente comuns e amplamente usadas em textos de um determinado registro. Segundo Conrad e Biber (2004), *lexical bundles* são sequências recorrentes de palavras ou frases que são frequentemente encontradas juntas na linguagem natural. Observa-se a similaridade entre os conceitos de diferentes autores.

O avanço tecnológico, marcado pelo aumento no uso de computadores e pela disponibilidade de grandes *corpus* linguísticos, desempenhou um papel significativo nas transformações observadas na pesquisa de *lexical bundles* em língua inglesa. A coleta extensiva de dados linguísticos por meio de *corpus* oferece uma riqueza de informações que possibilita uma análise mais abrangente e detalhada dos padrões linguísticos. A utilização de ferramentas computacionais facilita a identificação de *lexical bundles* específicos,

1 A discussão sobre a forma plural de "corpus" é um tópico linguístico interessante e, em certos contextos, pode gerar divergências. Tradicionalmente, "corpus" é uma palavra em latim que, ao passar para o inglês, tem seu plural como "corpora". No entanto, no uso coloquial e em certos círculos acadêmicos, alguns autores adotam "corpus" tanto para o singular quanto para o plural, inspirando-se na prática de certas palavras latinas que não mudam quando se referem a mais de uma instância.

Neste trabalho, utilizaremos "corpus" tanto para o singular quanto para o plural. Essa abordagem busca simplificar a linguagem e evitar a distinção tradicional entre "corpus" (singular) e "corpora" (plural). Optando por essa forma mais simplificada, reconhecemos a diversidade de práticas linguísticas existentes e adotaremos consistentemente "corpus" ao longo deste trabalho, seja para se referir a uma única coleção de dados ou a várias delas.

contribuindo para um entendimento mais preciso de como essas unidades linguísticas são empregadas em diferentes contextos. Essa interseção entre avanços tecnológicos e pesquisa linguística abre novas perspectivas para o desenvolvimento de estratégias pedagógicas mais eficazes, alinhadas com as demandas contemporâneas do ensino de língua inglesa para crianças não alfabetizadas.

Na introdução deste estudo, é crucial destacar nossa opção por preservar a integridade do termo original em inglês, "*lexical bundles*", em detrimento da utilização da tradução para "pacotes lexicais". Essa escolha visa resguardar a singularidade e a profundidade semântica do conceito, evitando possíveis perdas de sutilezas na tradução. Ao adotar o vocábulo em inglês, buscamos alinhar-nos de maneira mais precisa com a literatura especializada e fomentar uma abordagem uniforme na análise das unidades *multiword*. Essa decisão linguística fortalece a clareza conceitual e contribui para uma comunicação mais eficiente no âmbito acadêmico e científico.

A delimitação do escopo desta pesquisa é uma decisão fundamental, considerando a vastidão do tema em questão. O reconhecimento da ausência de trabalhos anteriores que ofereçam um repertório específico destinado a crianças não alfabetizadas revela uma lacuna significativa no conhecimento existente. A falta de recursos específicos para este grupo demográfico destaca a necessidade urgente de investigações nesse campo. Esta pesquisa se propõe a preencher essa lacuna, contribuindo para a literatura acadêmica ao desenvolver e apresentar um repertório adaptado às características e necessidades específicas das crianças não alfabetizadas. Ao fazê-lo, espera-se fornecer uma base sólida para práticas pedagógicas inovadoras e eficazes no ensino de língua inglesa a esse público-alvo, que são professores em formação, professores em atuação na rede básica, pública e privada, professores de escolas de idiomas, profissionais liberais e pesquisadores e/ou pais interessados pelo tema.

Biber (2009) utilizou uma abordagem baseada em *cópus* para identificar os padrões de *bundles* mais comuns em conversação e redação acadêmica. Os padrões típicos da fala são fundamentalmente diferentes daqueles típicos da escrita acadêmica. O estudo também conclui que os padrões na escrita acadêmica são estruturas estereotipadas com um espaço variável intermediário que é preenchido por "content words". Neste sentido, Conrad, Biber (2004) exploram a ocorrência e a utilidade de *lexical bundles*, também, em contextos de conversação e de escrita acadêmica. No estudo de Biber, Conrad e Cortes em 2003, eles categorizaram os agrupamentos em grupos que se alinham com funções discursivas específicas em relação ao emprego desses grupos funcionais tanto em conversas cotidianas quanto em textos acadêmicos.

O trecho realça o uso de uma abordagem baseada em *córpus* por parte de Biber (2009) para identificar padrões de *lexical bundles* em conversas e escritas acadêmicas. Essa análise revela discrepâncias fundamentais nos padrões de linguagem entre fala e escrita acadêmica. Além disso, o estudo indica que os padrões presentes na escrita acadêmica consistem em estruturas convencionais com espaços intermediários que são preenchidos por "palavras de conteúdo". Conrad e Biber (2004) também investigaram a ocorrência e a utilidade de *lexical bundles* em contextos de conversação e escrita acadêmica.

É evidente que as pesquisas anteriores sobre o tema têm suas raízes em fontes derivadas de momentos de registros orais e escritos, mas predominantemente em ambientes acadêmicos. No entanto, esta pesquisa distingue-se ao fornecer evidências que indicam a viabilidade do ensino de inglês como língua estrangeira na educação infantil. Essa abordagem diverge de trabalhos anteriores ao considerar o contexto de aplicação, direcionando-se para a educação infantil e, ao mesmo tempo, explorando formas inovadoras de registro. A pesquisa busca, assim, estender as fronteiras do conhecimento ao adaptar-se a um cenário educacional específico, oferecendo uma perspectiva única e relevante para a introdução de estratégias de ensino de línguas estrangeiras desde a infância.

O aspecto distintivo da presente pesquisa em relação a trabalhos anteriores reside no seu enfoque no ensino de inglês como língua estrangeira na educação infantil. A pesquisa oferece evidências sólidas e indica a viabilidade dessa abordagem em um novo contexto, introduzindo formas inovadoras de registro. Desta maneira, contribui significativamente para a expansão do entendimento sobre a aplicação de *lexical bundles* não apenas em contextos acadêmicos, mas também em um cenário educacional mais amplo, particularmente no que diz respeito ao ensino de inglês para crianças na educação infantil.

Assim sendo, optou-se pelo estudo de *lexical bundles* pela importância de suas contribuições para o ensino de língua inglesa para crianças não alfabetizadas pois refletem a língua falada, norteiam o ensino e é baseado em evidências e uso, e não em subjetividade. Como aponta Biber (2009), “o estudo de *lexical bundles* é importante em vários campos, incluindo linguística, processamento de linguagem natural e ensino de línguas, pois ajuda pesquisadores e alunos de línguas a compreender como as palavras funcionam em determinado contexto”. Discutindo sobre *lexical bundles* nas palavras de Biber, Conrad e Cortes (2003): “É evidente que desempenham funções discursivas de grande importância”. Esta pesquisa é o primeiro passo em direção ao objetivo de conscientização e apresentação da necessidade de um ensino de língua inglesa para crianças baseado em evidências.

Como indicado por Biber (2009), a análise de *lexical bundles* desempenha um papel crucial em várias áreas, abrangendo a linguística, o processamento de linguagem natural e o ensino de idiomas. Isso ocorre porque fornece valiosas informações para pesquisadores e estudantes de línguas, permitindo uma compreensão mais profunda de como as palavras funcionam em contextos específicos. Abordar os *lexical bundles*, nas palavras de Biber, Conrad e Cortes (2003), revela sua importância nas funções discursivas, destacando seu papel significativo na organização do discurso.

Este estudo representa um passo inicial em direção à conscientização sobre a necessidade de um método de ensino de língua inglesa para crianças que seja fundamentado em evidências. Isso sugere a importância de basear a educação de crianças em pesquisas sólidas e compreensão científica, a fim de promover uma abordagem eficaz no ensino do idioma desde a infância.

A criança que tem contato com a língua inglesa o quanto antes tende a ter mais possibilidades no futuro, como vagas de empregos, qualificação profissional, desenvolvimento pessoal e intelectual. Como comentado por Camargo, Jessyca (2022, p. 18):

Dessa maneira, as justificativas para a busca do ensino da LI mais precocemente deveriam pautar também o fato de que o aprendiz de menor idade dispõe de mais horas por dia para que possa dedicar-se a esse ensino e isso o levaria, por consequência, a um maior período de exposição à língua. Isso contribui para que ele consiga efetivar de maneira mais fácil esse conhecimento do que o aprendiz que inicia posteriormente.

A presente pesquisa auxiliará professores de língua inglesa da educação infantil a terem um material de referência para subsidiar o ensino. Estudantes de letras poderão se basear no material para se qualificarem para o mercado de trabalho cada vez mais exigente. As crianças serão expostas a produções autênticas e recorrentes da língua inglesa. Ao tomar conhecimento sobre o tipo de *bundle* utilizado na comunicação para/com crianças, o professor poderá reproduzir em suas aulas, tornando-as mais objetivas e eficazes.

Conforme mencionado por Egbert et al. (2022, p. 57): "Quando um córpus está sendo criado para um novo projeto, os objetivos de pesquisa linguística irão orientar todos os aspectos do design e compilação do córpus".²

Tal afirmação ressalta a importância crucial de alinhar os objetivos de pesquisa linguística com o processo de criação de um córpus. Ao iniciar um novo projeto de córpus, é essencial ter clareza sobre os objetivos específicos da pesquisa, pois esses objetivos orientarão todas as etapas do design e compilação do córpus.

² When a corpus is being created for a new project, the linguistic research goals will guide all aspects of corpus design and compilation.

Os objetivos de pesquisa linguística têm um papel fundamental na determinação do escopo do *corpus*, na seleção dos textos a serem incluídos, na definição das características linguísticas a serem analisadas e nos métodos de coleta e organização dos dados. Eles influenciam diretamente as decisões sobre a extensão do *corpus*, a diversidade de gêneros textuais a serem representados, o tamanho da amostra, entre outros aspectos.

Essa abordagem orientada por objetivos garante que o *corpus* seja construído de forma a atender às necessidades específicas da pesquisa linguística em questão. Portanto, a definição clara dos objetivos de pesquisa desde o início é crucial para garantir a relevância, a representatividade e a utilidade do *corpus* na investigação linguística.

Frente ao papel da necessidade de destaque do ensino de inglês na educação infantil, a pesquisa tem o objetivo geral de criar uma lista abrangente de *lexical bundles* relevantes e aplicáveis na sala de aula de língua inglesa. Foram analisadas comunicações reais que não são feitas por crianças, mas sim expostas para crianças.

Levando em consideração essa conexão essencial entre os objetivos de pesquisa e a construção do *corpus*, o objetivo específico de pesquisa para este estudo é identificar e analisar os padrões de *lexical bundles* presentes na língua falada voltada para crianças em vídeos do *YouTube*. Esse objetivo é fundamental para direcionar o estudo, fornecendo uma estrutura clara para a análise e aplicação dos *lexical bundles* nos contextos específicos selecionados, visando contribuir de maneira significativa para o ensino e aprendizagem da língua inglesa.

Com base em um processamento linguístico robusto – um *corpus* grande e com diferentes sub-registros –, destacamos a precisão na seleção dos *lexical bundles*. O *corpus* criado especialmente para a pesquisa possui 1940 textos, 135 *lexical bundles* e 1.722.722 milhões de palavras. A fundamentação teórica recorre à Linguística de *Corpus* (BIBER, 2014, 2022), e à Interculturalidade (KRAMSCH, 2021). O objeto de estudo da pesquisa são os *lexical bundles*, intimamente vinculados à léxico gramática da língua inglesa, pois não se trata apenas de léxico, nem apenas de gramática, e sim ambas nutrindo-se mutuamente.

Ao serem analisados, os *lexical bundles* revelam como palavras específicas estão intrinsecamente associadas a determinadas construções gramaticais, demonstrando a maneira pela qual a gramática e o léxico se entrelaçam e se complementam no uso autêntico da língua inglesa. Essa pesquisa visa explorar a função e a frequência dessas unidades linguísticas, fornecendo uma compreensão mais profunda da natureza complexa e interdependente do léxico e da gramática no idioma inglês. Os materiais utilizados na realização da presente pesquisa foram 4, a saber: 1) *YouTube*; 2) *Ytscribe*; 3) *Word Smith Tools* e 4) *O corpus*. A

metodologia proposta para esta pesquisa dar-se-á por meio da seleção de *córpus* presente em vídeos do *YouTube* destinados às crianças. Logo após essa seleção, o *córpus* passará pelo *ytscribe*, onde transcreveremos todo o material oral, feito isso, extrairemos os *lexical bundles* mais comuns, utilizando a plataforma *Word Smith Tools* (SCOTT, 2021) para o processamento dos textos.

Na condução desta pesquisa, é fundamental destacar que empregamos a versão 5.0 do software *WordSmith Tools*, datada de 2021. Essa especificação é de suma importância para fornecer clareza e transparência aos próximos pesquisadores que possam se interessar pelo tema ou realizar estudos semelhantes. Ao mencionar a versão específica utilizada, garantimos que haja consistência e precisão na compreensão das ferramentas e metodologias empregadas ao longo do processo de análise de dados. Dessa forma, ao explicitar o uso da versão 5.0 de 2021 do *WordSmith Tools*, buscamos contribuir para a replicabilidade e a validade dos resultados obtidos, permitindo que outros pesquisadores possam contextualizar e comparar seus próprios achados com os desta pesquisa. Essa prática de transparência e documentação detalhada das ferramentas e recursos utilizados fortalece a integridade e a confiabilidade da pesquisa acadêmica, promovendo um ambiente propício para o avanço do conhecimento e o desenvolvimento contínuo da investigação científica.

Com esse trabalho, pretendemos entregar constatações linguístico-estatísticas para que professores tenham subsídios, e para aprendizes não alfabetizados, com base em evidências.

Romer (2008, p. 124) comenta que “uma certeza que temos é que o campo da Linguística de *Córpus* e do ensino de línguas tem um futuro emocionante, pelo qual tanto os pesquisadores quanto os professores podem e devem aguardar com expectativa”.³ Essa afirmação ressalta a perspectiva otimista em relação ao desenvolvimento futuro da interseção entre a linguística de *córpus* e o ensino de idiomas. Ela sugere que há um potencial empolgante para avanços, inovações e descobertas nesses campos, o que oferece oportunidades tanto para os pesquisadores quanto para os professores. Essa visão positiva destaca a importância de estar aberto às mudanças, ao progresso e às novas abordagens no uso de *córpus* no ensino, antecipando uma evolução promissora que pode trazer benefícios significativos para o aprimoramento das práticas de ensino e aprendizagem de idiomas.

Elaboramos dois problemas para a pesquisa, a saber: 1) quais os *bundles* específicos dos quatro sub-registros contidos em vídeos do *YouTube*: literatura infantil, vídeos

³ One thing that we can be sure of, however, is that the field of corpus linguistics and language teaching has an exciting future that both researchers and teachers can, and should, look forward to.

educacionais, filmes e músicas infantis, com base em um processamento linguístico robusto? e 2) quais os *bundles* compartilhados pelos quatro sub-registros?

A abordagem para a análise do primeiro problema de pesquisa envolverá o uso de um processamento linguístico robusto, o que significa que se empregará uma técnica de análise de linguagem detalhada e abrangente para identificar esses agrupamentos específicos em cada tipo de registro.

O segundo problema de pesquisa busca determinar quais são os *lexical bundles* que são compartilhados entre os registros mencionados anteriormente. Isso implica em identificar os *lexical bundles* que ocorrem em comum nos cinco tipos de registros. Essa análise poderá fornecer informações valiosas sobre os elementos linguísticos que são frequentemente encontrados em diferentes tipos de materiais voltados para crianças, ajudando a entender possíveis semelhanças ou padrões na linguagem usada nesses contextos.

No geral, esses dois problemas de pesquisa visam aprofundar a compreensão da linguagem presente em materiais direcionados para crianças, utilizando uma abordagem de análise linguística detalhada e comparativa para identificar tanto os elementos específicos de cada registro quanto os elementos compartilhados entre eles.

Atualmente, a língua inglesa é a língua mais falada do mundo, e conta com 1.268 milhões de falantes, sendo 379 milhões falantes nativos e 753 milhões falantes não nativos. David, P. D. (2005, p. 210) comenta:

Sendo falado como língua mãe por cerca de 400 milhões de pessoas, o inglês tornou-se a língua franca, o Latim dos tempos modernos, falado em todos os continentes por cerca de 800 milhões de pessoas. Estimativas, incluindo falantes com níveis de menor percepção e fluência, sugerem a existência atualmente de um total superior a um bilhão.

É considerada a língua universal, portanto, é de extrema importância que as pessoas tenham sucesso ao aprenderem essa segunda língua. Como citado por Figueiredo; Marzari (2012, p. 01) apud Camargo, Jessyca (2022, p. 14):

Na atualidade, é inquestionável a importância e a necessidade de comunicação entre os povos linguisticamente distintos. Com o advento da globalização, bem como com os avanços tecnológicos, tornamo-nos cidadãos sedentos de informação. Para isso, usufruímos do inglês, hoje considerado a mais importante ferramenta de interação tanto acadêmica como profissional para saciarmos nossa necessidade. Mais do que uma questão política e social, o inglês é um facilitador e formador de pessoas bilíngues aptas a contribuir para o crescimento e sucesso da sociedade global.

Tendo isso em mente, sublinha-se a importância dos *lexical bundles* para atingir tal objetivo, já que a criança pode assimilar a língua desde tenra idade, sem perceber que está adquirindo algo tão valioso para seu desenvolvimento pessoal e carreira futura.

Ademais, Berber Sardinha (2010, p. 22) apresenta as dificuldades da Linguística de Córpus:

- Escassez de praticantes;
- Falta de doutores na área;
- Dificuldade em conseguir financiamento para projetos de uso de corpora no ensino;
- Pouca ou nenhuma integração com a escola pública;
- Pouca presença nos cursos de formação de professor na universidade;
- Pouca integração com educadores e especialistas em educação, assim como no mundo inteiro;
- Falta de materiais de ensino feitos com (base em) corpora e livros de apoio ao professor.

A pesquisadora responsável pelo presente estudo se identifica profundamente com as dificuldades destacadas por Berber Sardinha (2010) no âmbito da Linguística de Córpus. As questões levantadas por ele ressoam fortemente, refletindo desafios significativos enfrentados por aqueles envolvidos na pesquisa e prática desta área especializada.

Durante a trajetória da autora como estudante na educação básica e no ensino médio, não foi dada a oportunidade de ter contato com o tema da Linguística de Córpus ou com professores que possuíam conhecimento substancial sobre esse campo específico da linguística. Infelizmente, esse é um cenário comum para muitos estudantes, já que a abordagem de disciplinas tão especializadas pode ser limitada no currículo escolar.

A ausência de discussões ou abordagens sobre Linguística de Córpus pode refletir a falta de incorporação de disciplinas mais específicas e contemporâneas nos programas educacionais. Muitas vezes, os currículos escolares são moldados por diretrizes mais gerais, o que pode deixar de fora temas mais especializados ou emergentes, como é o caso da análise de córpus na linguística.

Além disso, os materiais didáticos não eram construídos com base em córpus. Geralmente, os materiais utilizados nas salas de aula costumavam seguir abordagens mais tradicionais, sem incorporar a riqueza e a autenticidade que o uso de córpus poderia proporcionar.

Essa ausência de materiais didáticos embasados em córpus limitou a experiência de aprendizado, já que tais recursos poderiam oferecer uma visão mais próxima da língua em seu uso real, apresentando exemplos autênticos e contextualizados para compreensão e análise linguística.

A falta de materiais didáticos baseados em córpus também pode ter impacto na maneira como os estudantes interagem com a língua, pois a exposição a exemplos reais da linguagem utilizada em contextos autênticos pode enriquecer a compreensão e a prática linguística de forma mais significativa.

A falta de exposição a esse assunto ao longo da jornada escolar evidencia a lacuna entre os avanços acadêmicos e o ensino convencional, destacando a necessidade de atualização e inclusão de tópicos mais atuais e relevantes nos programas educacionais. É essencial promover uma abordagem mais ampla e inclusiva no ensino, proporcionando aos estudantes a oportunidade de explorar uma gama diversificada de áreas do conhecimento, incluindo campos especializados como a Linguística de Córpus.

Como aluna de Letras, a autora experienciou a lacuna no currículo acadêmico em relação aos estudos específicos sobre Linguística de Córpus. É notável que, apesar da amplitude do campo de estudo de Letras, muitos programas de cursos de formação de professores ainda não incluem de forma significativa ou aprofundada disciplinas voltadas para a análise de córpus na linguística.

O curso de Letras é reconhecido por sua abrangência no estudo da linguagem, literatura e cultura, preparando os estudantes para atuarem como professores, pesquisadores, tradutores, entre outras áreas. No entanto, a ausência de disciplinas ou abordagens dedicadas à Linguística de Córpus durante a formação como estudante de Letras reflete uma limitação no currículo, especialmente considerando que o curso tem o propósito de formar professores.

A formação de professores é uma das bases fundamentais do curso de Letras, e a inclusão de disciplinas mais especializadas, como a análise de córpus na Linguística, poderia enriquecer a preparação dos futuros educadores. Ao proporcionar conhecimentos mais específicos e atualizados, os cursos de Letras poderiam capacitar os professores para adotar novas abordagens pedagógicas e ferramentas inovadoras, como o uso de córpus no ensino da língua e literatura.

A falta de ênfase nos estudos de Linguística de Córpus dentro do curso de Letras evidencia uma oportunidade de aprimoramento do currículo, permitindo que os futuros profissionais da área possam explorar e integrar conhecimentos, promovendo, assim, uma educação mais completa e alinhada às demandas contemporâneas do ensino de línguas e literaturas.

A autora só teve contato com a Linguística de Córpus ao ingressar no PIBID (Programa Institucional de Bolsa de Iniciação à Docência). Essa experiência representou um ponto de virada significativo, pois foi através desse programa que a autora teve a oportunidade de se envolver mais profundamente com a Linguística de Córpus e suas aplicações no ensino.

O PIBID proporcionou um ambiente propício para a descoberta e exploração dessa área da linguística, oferecendo oportunidades práticas de aplicar os conhecimentos adquiridos

em situações reais de sala de aula. Essa imersão prática possibilitou à autora compreender como os *corpuses* podem ser utilizados de maneira eficaz no ensino da língua e literatura, bem como na análise linguística.

Essa experiência destacou a importância de programas acadêmicos que promovem a interação direta com áreas específicas do conhecimento, como a Linguística de *Corpus*, expandindo os horizontes de aprendizado para além das disciplinas tradicionais e permitindo o desenvolvimento de habilidades práticas relevantes para a atuação profissional.

Após o contato inicial com a Linguística de *Corpus* durante sua participação no PIBID, a autora se sentiu especialmente motivada a explorar mais profundamente essa área no âmbito acadêmico. Essa motivação resultou na escolha de utilizar *corpuses* como base para seu Trabalho de Conclusão de Curso (TCC) na faculdade de Letras.

Ao realizar seu TCC com base em *corpuses*, a autora teve a oportunidade de aprofundar seus conhecimentos na área e aplicar na prática os conceitos aprendidos, desenvolvendo uma pesquisa que explorava o uso e a aplicação de *corpuses* em um contexto linguístico específico.

A experiência e o interesse adquiridos durante a realização do TCC impulsionaram a autora a seguir adiante com essa linha de pesquisa. Assim, ela elaborou seu pré-projeto de mestrado no mesmo campo, buscando expandir e aprofundar seu conhecimento sobre a utilização de *corpuses* na análise linguística.

Essa continuidade nos estudos reflete o comprometimento da autora em explorar e contribuir com o campo da Linguística de *Corpus*, demonstrando seu interesse em se aprofundar e desenvolver pesquisas mais avançadas nessa área específica da linguística.

Portanto, a autora desta pesquisa reconhece e se identifica profundamente com as dificuldades apontadas por Sardinha (2010), reconhecendo a importância de superar tais obstáculos para promover o avanço e a implementação eficaz da Linguística de *Corpus* no contexto educacional.

O trabalho está organizado em 06 partes, sendo elas: 1) contextualização sociocultural; 2) introdução; 3) fundamentação teórica; 4) material e método; 5) resultados e análise de dados; e 6) considerações finais. No que se refere à produção final, nosso objetivo principal de criar uma lista abrangente de *lexical bundles* relevantes e aplicáveis na sala de aula de língua inglesa foi alcançado com sucesso, assim como o objetivo específico de identificar e analisar os padrões de *lexical bundles* presentes na língua falada voltada para crianças em vídeos do *YouTube*.

A análise abrangente de *lexical bundles* provenientes de vídeos do *YouTube*, literatura infantil, vídeos educacionais, filmes e músicas infantis proporcionou uma compreensão

comparativa dos padrões linguísticos em diferentes formas de comunicação para crianças. A pesquisa não apenas contribui para a teoria linguística, mas também oferece recursos práticos e estratégias pedagógicas valiosas.

A lista abrangente de lexical bundles derivada da análise do *cópus* é um recurso útil para professores e alunos, enriquecendo o ensino de inglês na educação infantil. A proposta de uma referência didático-pedagógica que integra os lexical bundles identificados fortalece ainda mais a utilidade prática da pesquisa, oferecendo estratégias de ensino eficazes alinhadas às necessidades específicas de crianças não alfabetizadas. Em resumo, a pesquisa visa enriquecer o ensino de inglês na educação infantil e aprimorar significativamente a competência linguística dos alunos, contribuindo tanto para a teoria linguística quanto para a prática educacional.

3 FUNDAMENTAÇÃO TEÓRICA

Quadro 1 – Professor como “filtro”



Fonte: a autora

A preocupação central desta pesquisa reside na análise e identificação de padrões linguísticos presentes em vídeos do YouTube direcionados para crianças não alfabetizadas, visando fornecer insights valiosos para professores no contexto do ensino de inglês na educação infantil. No entanto, é relevante considerar que, em um segundo momento, as próprias crianças também podem se beneficiar indiretamente caso o professor opte por recomendar vídeos que julgue pertinentes e de qualidade para o desenvolvimento linguístico e educacional delas. Nesse sentido, o professor desempenha um papel fundamental como um "filtro" dos conteúdos disponíveis, selecionando e compartilhando com as crianças aquilo que considera importante e adequado para seu aprendizado e desenvolvimento. Essa mediação do professor não apenas contribui para garantir que as crianças tenham acesso a materiais relevantes e de qualidade, mas também promove uma experiência educacional mais enriquecedora e orientada para os objetivos pedagógicos estabelecidos. Assim, ao atuar como um intermediário entre os recursos digitais e as crianças, o professor desempenha um papel crucial na orientação e no direcionamento do processo de aprendizado, facilitando a

assimilação de conteúdos significativos e apropriados para o desenvolvimento linguístico e educacional das crianças não alfabetizadas.

Nesta seção deste trabalho, embarcaremos em uma exploração abrangente da Linguística de Córpus, Linguística de Córpus no ensino, *lexical bundles*, Interculturalidade e Fala x Escrita. A Linguística de Córpus, como base metodológica, oferece uma abordagem inovadora para a análise linguística, permitindo-nos examinar padrões recorrentes em grandes conjuntos de dados textuais autênticos. A Linguística de Corpus no ensino fornece uma abordagem fundamentada em dados reais que pode melhorar a compreensão, o aprendizado e a aplicação prática de uma língua, tornando o ensino mais eficaz e significativo para os estudantes. Ao nos aprofundarmos nos *lexical bundles*, exploraremos as unidades *multiword* que desempenham papéis cruciais na expressão de significado e na construção discursiva. Esta análise minuciosa dos padrões lexicais contribuirá para a compreensão mais profunda das nuances linguísticas. Além disso, a dimensão intercultural acrescentará uma camada vital, considerando como essas estruturas linguísticas se manifestam em diferentes contextos culturais, ampliando assim a compreensão da interação entre linguagem e cultura. Discutindo sobre a relação entre a fala e a escrita entendemos que são formas complementares de comunicação que compartilham a mesma base linguística, mas apresentam diferenças fundamentais em termos de modalidade, produção, características linguísticas e contexto de uso. Estudar a relação entre fala e escrita é crucial para compreender a linguagem em sua totalidade e para desenvolver habilidades comunicativas eficazes em diferentes contextos.

3.1 Linguística de Córpus

A Linguística de Corpus é um campo da linguística que se concentra na análise de grandes coleções de textos escritos ou falados, chamados *córpus*, para extrair informações linguísticas. Esses *córpus* podem ser constituídos por uma grande variedade de materiais, como livros, artigos, transcrições de conversas, sites da internet, entre outros.

O objetivo principal da Linguística de Córpus é investigar e descrever padrões linguísticos reais, revelando como palavras, frases e estruturas são usadas em contextos autênticos. Isso permite aos linguistas e pesquisadores entenderem melhor a linguagem natural, identificar tendências, variações, usos coloquiais, gramaticais e semânticos.

As técnicas da Linguística de Córpus envolvem o uso de softwares específicos para análise de dados linguísticos. Estas ferramentas permitem pesquisas detalhadas, como identificação de frequência de palavras, colocações (combinações frequentes de palavras),

análise de concordância (contextos em que palavras ou expressões ocorrem) e estudo de padrões gramaticais.

As aplicações da Linguística de Corpus são amplas e podem ser encontradas em várias áreas, como por exemplo:

- I. Linguística Aplicada: no ensino de línguas, para criar materiais didáticos baseados em uso real da língua; na tradução, para entender o uso de termos e expressões em diferentes contextos.
- II. Processamento de Linguagem Natural: para desenvolver sistemas de inteligência artificial, tradução automática, análise de sentimento e assistentes de voz.
- III. Análise de Discurso e Comunicação: para investigar padrões de comunicação em diferentes contextos sociais, políticos e culturais.
- IV. Estudos Literários e Culturais: para analisar estilos literários, discursos políticos, mídia e aspectos culturais por meio da linguagem.

A Linguística de Corpus oferece uma abordagem baseada em evidências empíricas e tem contribuído significativamente para o entendimento da linguagem humana, fornecendo *insights* valiosos sobre seu funcionamento e uso nas mais diversas situações.

3.2 Linguística de Córpus no ensino

Os autores Biber, McCarthy, Carter, Brown (2001, p. 332) ressaltam importantes descobertas sobre análises quantitativas de córpus, a saber:

Ao contrário da aparência de que análises quantitativas de córpus consistem em uma contagem elaborada de elementos, nossas investigações apontam para duas generalizações principais que são cruciais para o ensino de inglês como segunda língua (ESL) / inglês como língua estrangeira (EFL):

1. A centralidade do registro para estudos de uso da linguagem: padrões fortes de uso em um registro frequentemente representam padrões fracos em outros registros. Se os linguistas pretendem realizar uma análise completa de padrões gramaticais, eles devem considerar os padrões de uso em diferentes registros, e os aprendizes muitas vezes podem se beneficiar dessas informações;

2. A falta de confiabilidade das intuições sobre o uso: professores, autores e profissionais de avaliação frequentemente confiam em suas intuições para escolher as palavras e estruturas mais importantes para se concentrar. No entanto, estudos de córpus mostram que tais intuições sobre o uso muitas vezes estão incorretas.⁴

⁴Contrary to the appearance that quantitative corpus analyses consist of elaborate bean counting, our investigations point to two major generalizations that are crucial for ESL/ EFL teaching: 1. the centrality of register for studies of language use. Strong patterns of use in one register often represent only weak patterns in other registers. If linguists are to undertake a complete analysis of grammatical patterns, they must consider the patterns of use across registers, and learners can often benefit from this information. 2. the unreliability of intuitions about use. Teachers, authors, and testing professionals constantly rely on their intuitions to choose the most important words and structures to focus on. However, corpus studies show that such intuitions about use are often incorrect.

A citação em questão salienta duas ideias cruciais para o ensino de inglês como segunda língua (ESL) ou inglês como língua estrangeira (EFL), com base em descobertas obtidas por meio de análises quantitativas de *cópus*. Em primeiro lugar, é ressaltada a importância do contexto linguístico ao analisar os padrões gramaticais. Ela aponta que padrões fortes observados em um determinado contexto podem não ser igualmente predominantes em outros. Isso implica que os linguistas, ao realizarem uma análise abrangente dos padrões gramaticais, devem considerar e ponderar os diferentes padrões de uso em diversos ambientes linguísticos. Além disso, destaca-se que os alunos aprendendo uma língua estrangeira podem se beneficiar ao conhecerem essas distintas formas de uso da língua, o que pode influenciar positivamente o processo de ensino e aprendizagem.

Em segundo lugar, os autores questionam a confiabilidade das intuições individuais sobre o uso da linguagem. Muitas vezes, professores, autores e profissionais de avaliação confiam em suas próprias intuições ao selecionar palavras e estruturas linguísticas que consideram mais relevantes. No entanto, a pesquisa indica que os estudos de *cópus* frequentemente contradizem essas intuições, sugerindo que as percepções individuais podem não refletir com precisão os padrões reais de uso da linguagem em situações reais de comunicação.

Essas generalizações ressaltam a importância de basear o ensino de idiomas em evidências empíricas provenientes de estudos de *cópus*, em vez de depender exclusivamente de pressentimentos individuais. Esse método pode resultar em um ensino mais embasado e eficaz, fundamentado nos padrões reais de uso da linguagem em uma variedade de contextos, oferecendo benefícios valiosos aos alunos de ESL/EFL.

Este trabalho reforça a premissa expressa em Biber, McCarthy, Carter, Brown (2001) de que o estudo e a análise baseados em evidências são essenciais para um entendimento mais aprofundado e preciso. Concordamos plenamente com a ideia de que confiar unicamente em pressentimentos individuais, sem considerar as evidências obtidas por meio de análises e estudos concretos, pode limitar nossa compreensão real do uso da linguagem.

Ao destacar a importância de examinar dados empíricos provenientes de estudos de *cópus*, estamos alinhados com a noção de que as percepções individuais sobre o uso da linguagem podem não ser sempre precisas. Essa abordagem fundamentada em evidências oferece uma compreensão mais sólida dos padrões reais de uso da linguagem em diferentes contextos, o que é crucial no desenvolvimento de estratégias de ensino mais eficazes para estudantes de ESL/EFL.

Portanto, concordamos plenamente com a ideia de que o estudo baseado em evidências, como proposto na citação, é fundamental para um ensino informado e eficaz de línguas estrangeiras, contribuindo significativamente para o progresso e o aprimoramento no campo do ensino de idiomas.

O estudo linguístico baseado em *cópus* para crianças não alfabetizadas demanda atenção especial a duas considerações fundamentais, como apontado por Egbert et al. (2022, p. 11):

No caso de um estudo linguístico baseado em *cópus*, existem duas considerações principais que devem ser abordadas para avaliar o grau de representatividade de um *cópus*: (1) Em que medida a coleção de textos no *cópus* representa a gama de textos e tipos de texto no domínio de interesse? Estas são as considerações de domínio. (2) Em que medida o *cópus* representa com precisão a distribuição quantitativa das características linguísticas de interesse? Estas são as considerações de distribuição.⁵

Essas considerações são de extrema relevância ao avaliar o grau de representatividade do *cópus* utilizado:

1. Considerações de domínio: refere-se à medida em que a coleção de textos no *cópus* representa de forma abrangente a variedade de textos e tipos de texto presentes no contexto específico das crianças não alfabetizadas. Aqui, a diversidade dos materiais de leitura e o conteúdo devem refletir com precisão os contextos linguísticos, culturais e sociais em que essas crianças estão imersas. Dessa forma, garantir uma ampla representatividade do domínio de interesse é essencial para compreender e atender às necessidades linguísticas desse grupo.

2. Considerações de distribuição: envolve a precisão com que o *cópus* reflete a distribuição quantitativa das características linguísticas relevantes para o estudo das crianças não alfabetizadas. Isso significa que a ocorrência e a frequência das palavras, estruturas gramaticais e outros elementos linguísticos devem espelhar fielmente a realidade linguística desse público-alvo. Uma distribuição representativa desses aspectos linguísticos contribuirá para uma compreensão mais profunda das particularidades linguísticas das crianças em fase de aprendizado da linguagem.

Ao considerar esses dois aspectos - a abrangência de textos e a representação precisa das características linguísticas - o estudo baseado em *cópus* para crianças não alfabetizadas pode obter percepções valiosas sobre suas necessidades linguísticas específicas, contribuindo

⁵ In the case of a corpus-based linguistic study, there are two major considerations that must be addressed to evaluate the degree of representativeness of a corpus: (1) To what extent does the collection of texts in the corpus represent the range of texts and text types in the domain of interest? These are the domain considerations. (2) To what extent does the corpus accurately represent the quantitative distribution of the linguistic features of interest? These are the distribution considerations.

assim para o desenvolvimento de estratégias mais eficazes no processo de aprendizado da linguagem nessa fase crucial do desenvolvimento infantil.

Tenhamos em mente a seguinte citação: "a meta final da análise de córpus é uma descrição empírica generalizável do uso da linguagem em um domínio discursivo específico." (Egbert et al., 2022, p. 5).⁶ Os autores destacam o objetivo fundamental da análise de córpus, que é a obtenção de uma descrição empiricamente embasada e generalizável do uso da linguagem em um contexto discursivo específico. Quebrando a citação em partes para analisá-la, temos:

- "A meta final da análise de córpus": refere-se ao objetivo final ou principal da análise de córpus, que é um método linguístico que envolve o estudo de grandes coleções de textos ou amostras de linguagem.
- "é uma descrição empírica generalizável": essa análise visa criar uma descrição baseada em evidências empíricas, ou seja, dados reais encontrados nos textos analisados. Essa descrição não é apenas específica para um texto, mas é generalizável, ou seja, pode ser aplicada de forma mais ampla ao domínio discursivo em questão.
- "do uso da linguagem em um domínio discursivo específico": foca no uso da linguagem dentro de um contexto particular ou domínio discursivo, que pode ser um campo específico como letras, medicina, direito, jornalismo, entre outros. Essa análise busca compreender como a linguagem é utilizada dentro desse contexto específico.

Se fossemos transcrever a citação de Egbert et al. (2022) incluindo o assunto proposto neste trabalho, seria: "a meta final da análise de córpus de '*Lexical bundles* para crianças não alfabetizadas' é alcançar uma descrição empírica generalizável do uso da linguagem em um contexto discursivo específico (os sub-registros apresentados), voltado para as particularidades e padrões de expressões linguísticas agrupadas para crianças nessa fase pré-alfabetização."

No entanto, as abordagens fundamentadas em córpus operam de maneira empírica, examinando os padrões reais de uso em textos naturais; empregam uma extensa coleção de textos naturais seguindo princípios específicos; fazem uso intensivo de computadores para análise, incorporando tanto técnicas automáticas quanto interativas; e dependem de métodos analíticos quantitativos e qualitativos. É crucial abordar a pesquisa em córpus como uma oportunidade multifacetada, consolidando seu espaço no meio acadêmico e promovendo um

⁶ The ultimate goal of corpus analysis is a generalizable empirical description of language use in a target discourse domain.

crescimento abrangente na compreensão da linguagem. Os *córpus* não seguem uma fórmula predefinida, podendo abranger qualquer tipo de escrita, desde instruções de uso de um produto até obras completas de Shakespeare.

A Linguística de *Córpus* (LC) é a investigação da linguagem baseada em exemplos autênticos de diversos registros, e ela se complementa mutuamente. Berber Sardinha (2010) aborda a LC educacional no Brasil, destacando que, mesmo após 10 anos, ela ainda não alcançou uma visibilidade considerável. Para reverter esse cenário, os linguistas de *córpus* poderiam: proporcionar uma maior oferta de cursos de extensão, integrar a Linguística de *Córpus* aos programas de graduação em Letras, enfatizar a importância da Linguística de *Córpus* na rede pública, dar prioridade ao ensino a distância (EAD), complementar os materiais didáticos com a utilização da Linguística de *Córpus*, aumentar a disponibilidade de livros didáticos baseados em *córpus* e promover a publicação de materiais de apoio destinados aos professores de inglês no Brasil.

A presente pesquisa baseada em *córpus* pode contemplar absolutamente todos os tópicos propostos por Berber Sardinha (2010), pois:

1. Oferecer mais cursos de extensão: esta pesquisa pode contribuir para a criação de cursos de extensão que se baseiem na metodologia e nos resultados obtidos no estudo de *córpus*. Esses cursos poderiam capacitar professores, pesquisadores e profissionais interessados em aprofundar seu conhecimento sobre o uso da Linguística de *Córpus* no ensino de língua inglesa para crianças não alfabetizadas.

2. Incluir a Linguística de *Córpus* na graduação em Letras: os resultados deste trabalho podem ser fundamentais para justificar a inclusão de disciplinas ou módulos sobre Linguística de *Córpus* nos currículos de cursos de Letras. Essa inclusão pode ampliar o conhecimento dos futuros profissionais da área sobre a importância e aplicabilidade dessa abordagem no ensino de línguas.

3. Priorizar a Linguística de *Córpus* na rede pública: a pesquisa pode servir como uma base teórica e prática para promover discussões e políticas que priorizem o uso da Linguística de *Córpus* no ambiente educacional público, destacando sua relevância para melhorar o ensino-aprendizagem de línguas estrangeiras nas escolas.

4. Priorizar o ensino a distância (EAD): os resultados podem ser adaptados para o ambiente de ensino a distância, permitindo que plataformas online e cursos remotos integrem a abordagem de *córpus* no ensino de crianças não alfabetizadas, oferecendo materiais e recursos eficazes.

5. Complementar livros didáticos com a Linguística de Córpus: a pesquisa pode influenciar a criação de materiais complementares para livros didáticos que explorem os resultados do córpus, fornecendo exemplos autênticos e relevantes para o ensino de línguas para crianças em fase de alfabetização.

6. Publicar mais livros didáticos feitos com córpus: os achados da pesquisa podem contribuir para a criação de livros didáticos que sejam construídos com base em córpus, oferecendo aos professores materiais de ensino que reflitam com mais precisão os padrões reais de uso da linguagem.

7. Publicar livros de apoio a professores de inglês no Brasil: os resultados podem ser a base para a produção de livros de apoio que ofereçam orientações práticas para professores de inglês no Brasil, destacando como a Linguística de Córpus pode ser aplicada de maneira eficaz no contexto educacional brasileiro.

Este trabalho baseado em córpus para crianças não alfabetizadas pode ter um impacto significativo na maneira como a Linguística de Córpus é percebida, aplicada e integrada no ensino de línguas, tanto em nível acadêmico quanto em contextos educacionais reais.

Berber Sardinha (2010) também destaca uma série de desafios enfrentados pela Linguística de Córpus no contexto educacional brasileiro. A escassez de praticantes, a falta de doutores especializados na área e as dificuldades em obter financiamento para projetos que visam incorporar o uso de córpus no ensino são obstáculos significativos que podem impactar o desenvolvimento e a disseminação dessa abordagem. A falta de integração efetiva com a escola pública e a presença limitada nos cursos de formação de professores nas universidades brasileiras destacam a necessidade de uma maior inserção da Linguística de Córpus no ambiente educacional.

Outro desafio apontado é a escassa integração com educadores e especialistas em educação, tanto no Brasil quanto internacionalmente. Essa lacuna pode limitar as oportunidades de troca de conhecimento e experiências entre os praticantes da Linguística de Córpus e profissionais da educação. A falta de materiais de ensino desenvolvidos com base em córpus e de livros de apoio ao professor é outro fator crítico que pode impactar a eficácia da implementação dessa abordagem no contexto educacional. Superar esses desafios exigirá esforços coordenados para promover a visibilidade, a formação e a colaboração entre os praticantes da Linguística de Córpus e os educadores, contribuindo para um avanço mais robusto e integrado dessa disciplina no cenário educacional brasileiro.

A observação de Sinclair (2004) sobre a disponibilidade de coleções substanciais de textos linguísticos em formato eletrônico, ou córpus, por mais de 40 anos destaca a riqueza e a

longa tradição desse recurso na pesquisa linguística. O autor enfatiza que o *córpus* proporciona uma visão da estrutura da linguagem que anteriormente não estava acessível. No contexto brasileiro, a reflexão sobre a demora de mais de 25 anos para a incorporação do *córpus* no meio acadêmico levanta questões intrigantes. A caracterização do *córpus* como algo revolucionário e inovador para a época sugere que, apesar da sua disponibilidade há décadas, o seu reconhecimento pleno e sua adoção no meio acadêmico brasileiro ocorreram de maneira gradual. Mesmo após 41 anos desde o surgimento do primeiro *córpus*, as dúvidas persistentes sobre sua utilidade e benefícios indicam que ainda há desafios a serem superados para uma compreensão mais abrangente e uma aceitação consolidada dessa ferramenta na comunidade acadêmica brasileira. Essa ponderação ressalta a importância de promover uma compreensão mais ampla do *córpus* e explorar suas potencialidades para avançar no conhecimento linguístico no contexto nacional.

Ressalta-se que o *córpus* eletrônico tem sido considerado um valioso recurso para o estudo linguístico e sua contribuição ao ensino de línguas assim como à linguística é amplamente reconhecida. Porém, o uso de *córpus* no país, para o ensino de línguas, iniciou-se somente nos últimos anos da década de 1990, ou seja, aproximadamente 26 anos depois do lançamento do *Córpus Brown*, que foi o primeiro *córpus* lançado.

O *Corpus Brown*, desenvolvido por Francis; Kucera (1979), representa uma significativa contribuição para a linguística de *córpus*. Baseado em textos escritos americanos, esse *córpus* abrange uma ampla variedade de gêneros e contextos, proporcionando uma visão abrangente do uso da língua inglesa. Sua extensa cobertura temporal e diversidade de fontes tornam o *Corpus Brown* uma ferramenta valiosa para investigações linguísticas, permitindo a análise de padrões linguísticos ao longo do tempo e a compreensão das nuances da expressão verbal em diferentes domínios. Sua influência perdura, servindo como referência para pesquisadores e estudiosos interessados na análise empírica da linguagem.

O conhecimento e entendimento do *córpus* eletrônico é reservado principalmente a professores universitários e linguistas com foco em pesquisa baseada em *córpus*. Embora o *córpus* ofereça benefícios significativos para o ensino de línguas, a constatação de que a maioria dos professores desconhece essa ferramenta ressalta a necessidade de uma maior integração do *córpus* na formação acadêmica. A sugestão de incluir disciplinas relacionadas ao *córpus* na matriz curricular do curso de Letras destaca uma oportunidade crucial para capacitar estudantes a utilizarem efetivamente essa ferramenta em sua futura prática como professores ou pesquisadores. Ao oferecer conhecimento e habilidades práticas sobre o *córpus* durante a formação acadêmica, os estudantes podem maximizar seu potencial como

educadores, aproveitando os benefícios dessa valiosa ferramenta na elaboração de aulas, exercícios e materiais didáticos.

São inúmeros os benefícios que o *córpus* traz à pesquisa e ensino. Cobb; Boulton (2015, p. 480) comentam:

Do ponto de vista do professor, os *córpus* podem ajudar a decidir o que ensinar. Frequentemente, os *córpus* usados para esta finalidade não são grandes e modernos como o BNC ou COCA, mas sim *córpus* bastante pequenos como o Brown (Kuc̃era e Francis 1979), ou então coleções de texto construídas para o propósito e às vezes adequadas ao nível não necessariamente representativas de um idioma em sua totalidade. Esses *córpus* podem ser particularmente úteis no ensino de línguas para fins específicos, onde os materiais publicados são difíceis de encontrar.⁷

É possível criar materiais online para o ensino de forma prática e eficiente. “Aplicações pedagógicas de estudos de *córpus* podem ir além das descrições linguísticas, tendo impacto direto no planejamento de currículos e nas práticas pedagógicas ligadas ao ensino de línguas” Oliveira (2009, p. 54). Portanto, os *córpus* podem ser usados de várias maneiras, indiretamente para informar materiais de ensino e obras de referência, ou diretamente como ferramentas de aprendizagem de línguas e repositórios para a concepção de atividades de ensino. Porém, para que os professores e os alunos tenham o apoio necessário, é preciso melhorar consideravelmente a comunicação entre pesquisadores de *córpus* e profissionais. Deve-se construir uma ponte entre a pesquisa de *córpus* e a prática e deve ser explorado o papel da linguística de *córpus* (e linguistas de *córpus*) com a intenção de auxiliar melhor os professores em seu trabalho.

Romer (2008) comenta que para os pesquisadores de *córpus*, uma grande e importante tarefa será “espalhar a palavra” sobre *córpus* e informar os professores sobre o que já está disponível, por exemplo: *córpus* online, materiais baseados em *córpus*, trabalhos e coleções de exercícios de aprendizagem baseados em dados. Com isso, deve-se convencer mais professores que eles (e seus alunos) poderiam se beneficiar muito com a consultoria de *córpus*. As universidades deveriam, então, promover palestras e *workshops* sobre questões que se relacionam diretamente com os professores, seus problemas e necessidades, facilitando o contato e explanação sobre o tema. O trabalho do docente já é, naturalmente, difícil e

⁷ From the teacher’s perspective, corpora can help in deciding what to teach. Often the corpora used for this purpose are not large modern corpora like the BNC or COCA but rather smallish corpora like the Brown (Kuc̃era and Francis 1979), or else purpose-built and sometimes level-appropriate text collections not necessarily meant to be representative of a language in its entirety. Such corpora can be particularly useful in teaching languages for specific purposes where published materials are difficult to come by.

estressante, então se os linguistas de *córpus* podem fazer qualquer coisa para oferecer suporte, eles devem fazê-lo.

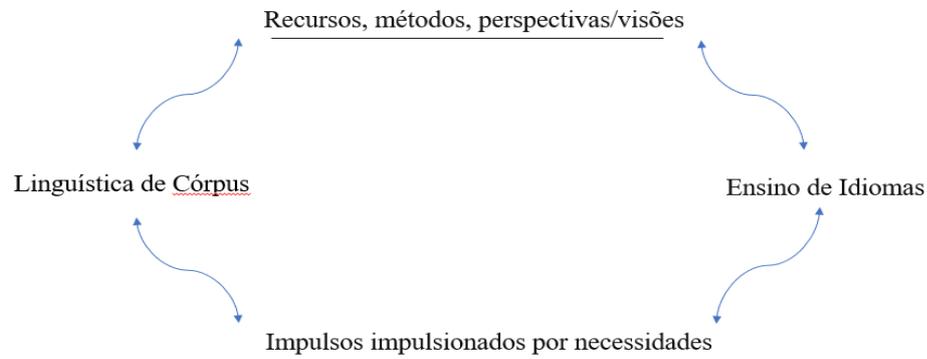
Neste momento, apresentaremos um esquema baseado na proposta de Romer (2008, p. 113) para ilustrar e explicar a interconexão entre a linguística de *córpus* e o ensino de línguas. Este esquema visa destacar os pontos-chave que demonstram como a utilização dos *córpus* na prática educacional oferece representações autênticas da língua, orienta os alunos na compreensão de padrões linguísticos, possibilita a manipulação ativa dos dados linguísticos, incentiva a exploração significativa e, por fim, promove a reflexão crítica sobre o uso e a estrutura da língua-alvo. Através dessa abordagem, busca-se proporcionar aos alunos uma compreensão mais funcional do idioma, tornando o aprendizado mais eficaz e contextualizado. Como comentado por Romer (2008, p. 112):

Essa relação é dinâmica, na qual os dois campos se influenciam significativamente. Enquanto o Ensino de Idiomas (LT) se beneficia dos recursos, métodos e perspectivas fornecidos pela Linguística de *Córpus* (CL), também fornece impulsos importantes que são incorporados na pesquisa linguística de corpora. As exigências do LT têm, portanto, um impacto nos projetos de pesquisa em CL e no desenvolvimento de recursos e ferramentas adequados. Este artigo investigará qual influência a CL teve até o momento no LT e de que maneiras os corpora têm sido utilizados para aprimorar a prática pedagógica. Também discutirá possíveis efeitos adicionais da CL no LT e do LT na CL, e destacará algumas tarefas futuras para pesquisadores e praticantes no campo.⁸

Em resumo, a relação entre a linguística de *córpus* e o ensino de línguas oferece oportunidades significativas para aprimorar a eficácia do ensino-aprendizagem de idiomas, fornecendo uma abordagem mais prática, autêntica e contextualizada para os estudantes.

Figura 1 - A relação entre Linguística de *Córpus* e ensino de línguas

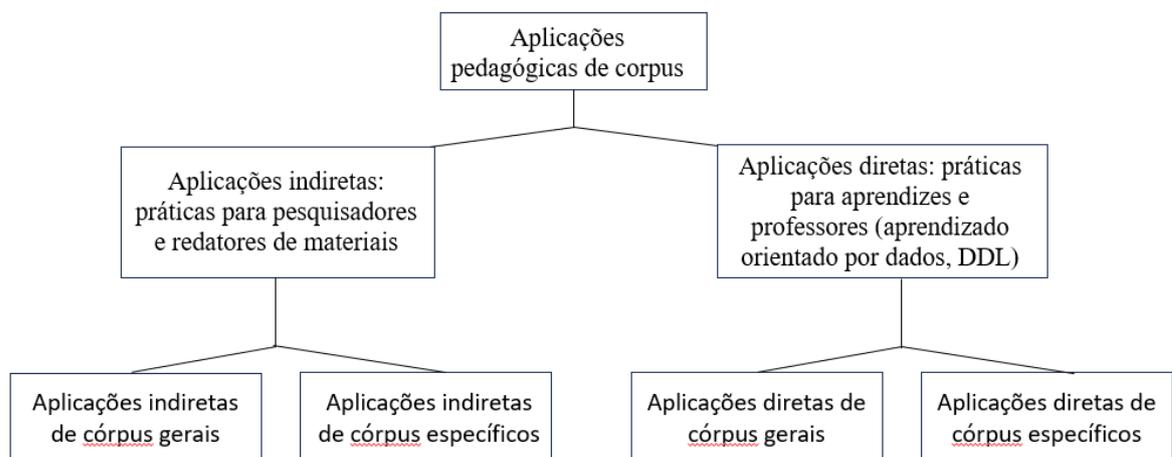
⁸ this relationship is a dynamic one in which the two fields greatly influence each other. While LT profits from the resources, methods, and insights provided by CL, it also provides important impulses that are taken up in corpus linguistic research his relationship is a dynamic one in which the two fields greatly influence each other. While LT profits from the resources, methods, and insights provided by CL, it also provides important impulses that are taken up in corpus linguistic research. The requirements of LT hence have an pact on research projects in CL and on the development of suitable resources and tools. The present article will investigate what influence CL has had on LT so far, and in what ways corpora have been used to improve pedagogical practice. It will also discuss further possible effects of CL on LT and of LT on CL and highlight some future tasks for researchers and practitioners in the field.



Fonte: traduzido de Romer (2008, p. 113) pela autora

A relação entre linguística de cópus (CL, do inglês *Corpus Linguistics*) e ensino de línguas (LT, do inglês *Language Teaching*) é de grande relevância e interesse no campo da educação linguística. No ensino de línguas, a aplicação da linguística de cópus tem demonstrado ser extremamente valiosa. Ela oferece aos educadores e alunos acesso a exemplos autênticos e contextualizados da língua-alvo, permitindo um aprendizado mais prático e relevante. Os cópus fornecem uma base de dados rica e diversificada que pode ser explorada para melhorar diversas habilidades linguísticas.

Figura 2 – O uso de cópus no ensino e aprendizado de línguas



Fonte: traduzido de Romer (2008, p. 19) pela autora

As aplicações pedagógicas de cópus podem ser divididas em diretas e indiretas, específicas e gerais, todas desempenham papéis significativos no processo de ensino e

aprendizagem de línguas. A seguir, apresentaremos onde o *córpus* pode ser utilizado e como ele pode ser aplicado.

Levando em consideração o texto de Romer (2008) sobre a relação entre *córpus* e o ensino de idiomas, a autora apresenta duas tabelas que especificam locais e métodos específicos nos quais os *córpus* podem ser utilizados de maneira benéfica para o ensino de línguas. Esses quadros detalham como os *córpus* podem ser integrados de forma eficaz no processo educacional, identificando áreas como análise gramatical, construção de vocabulário, compreensão auditiva, expressão escrita e oral, além de evidenciar a relevância dos *córpus* na elaboração de materiais didáticos e no planejamento de aulas mais dinâmicas e contextualizadas. Esses recursos gráficos fornecem um roteiro prático para a aplicação dos *córpus* no contexto do ensino de idiomas, destacando seu potencial para enriquecer a experiência de aprendizagem dos alunos.

As aplicações diretas de um *córpus* específico e geral na educação podem ser visualizadas, conforme apresentado no quadro 2, abaixo:

Quadro 2 – Aplicações pedagógicas diretas de *córpus* específico e geral

Onde o <i>córpus</i> pode ser utilizado	Como o <i>córpus</i> pode ser utilizado
Produção de Materiais Educacionais Específicos: Elaboração de Exercícios Contextualizados	Aplicação de um <i>córpus</i> particular para criar atividades linguísticas que sejam contextualizadas e relevantes para os interesses e demandas dos estudantes.
Análise de Uso de Vocabulário e Estruturas Gramaticais: Identificação de Colocações e Expressões Idiomáticas	Investigação minuciosa do emprego de termos e expressões específicos em variados contextos, viabilizando uma compreensão mais acurada do léxico e das construções gramaticais.
Suporte à Compreensão e Produção de Textos Específicos: Análise de Estilos e Gêneros Textuais Particulares	Utilização do <i>córpus</i> específico para estudar e compreender melhor estilos de escrita ou gêneros textuais específicos, como textos acadêmicos, técnicos ou literários.
Apoio ao Ensino de Línguas Estrangeiras: Identificação de Erros Recorrentes de Aprendizes	Análise dos erros comuns cometidos pelos aprendizes de uma língua estrangeira com base em um <i>córpus</i> específico para adaptar estratégias de ensino que abordem essas dificuldades de forma direcionada.
Pesquisa Linguística Focada: Estudo de Variações Linguísticas em um Contexto Determinado	Investigação de variações regionais, sociais ou situacionais específicas de um <i>córpus</i> , o que pode fornecer insights valiosos para a compreensão de particularidades linguísticas em um contexto determinado.

Análise de Textos para Fins Específicos: Avaliação de Textos Específicos para Determinada Finalidade:	Utilização do cópús para avaliar textos específicos em relação a uma finalidade particular, como adequação de textos para ensino, divulgação científica, entre outros.
---	--

Fonte: a autora

As aplicações diretas se relacionam ao emprego específico dos cópús como instrumentos educacionais. Isso compreende atividades em salas de aula que consistem na análise de textos autênticos com o intuito de instruir em gramática, vocabulário e nas habilidades de leitura, audição, escrita e fala. Os cópús oferecem aos estudantes exemplos reais e contextualizados do idioma estudado, contribuindo para uma compreensão mais aprofundada das nuances linguísticas e de seu uso prático.

O uso de um cópús específico e/ou geral permite uma análise mais detalhada de aspectos linguísticos que podem ser aplicados diretamente no ensino e aprendizagem, adaptando-se às necessidades particulares de um grupo de alunos, área de estudo ou contexto educacional específico.

As aplicações indiretas de um cópús específico e geral na educação podem ser visualizadas, conforme apresentado no quadro 3:

Quadro 3 - Aplicações pedagógicas indiretas de cópús específico e geral

Onde o cópús pode ser utilizado	Como o cópús pode ser utilizado
---------------------------------	---------------------------------

Desenvolvimento Profissional de Professores: Formação Continuada	Utilização do <i>corp</i> us para capacitar educadores no aprimoramento de métodos de ensino mais eficazes e adaptados.
Desenvolvimento Profissional de Professores: Formação Continuada	Utilização do <i>corp</i> us para capacitar educadores no aprimoramento de métodos de ensino mais eficazes e adaptados.
Avaliação e Aprimoramento de Currículos: Análise de Materiais Didáticos	Uso do <i>corp</i> us para avaliar a efetividade e adequação de materiais de ensino em relação à autenticidade linguística.
Pesquisa Linguística e Pedagógica: Investigação de Tendências Educacionais	Utilização do <i>corp</i> us para pesquisas sobre tendências linguísticas e pedagógicas, auxiliando na compreensão de métodos de ensino mais eficazes.
Desenvolvimento de Estratégias Institucionais: Planejamento Curricular	Análise do <i>corp</i> us para orientar a criação de diretrizes educacionais e planejamento curricular de modo mais alinhado com a linguagem autêntica.
Aprimoramento de Programas de Ensino: Revisão de Programas Acadêmicos	Uso do <i>corp</i> us para aprimorar e adaptar programas educacionais visando à melhoria do ensino linguístico e da compreensão textual.
Contribuição para Estudos Linguísticos e Educacionais: Auxílio em Pesquisas Acadêmicas	Utilização do <i>corp</i> us como fonte de dados para estudos linguísticos e educacionais mais amplos, contribuindo para o avanço desses campos de conhecimento.

Fonte: a autora

As aplicações indiretas referem-se aos efeitos mais amplos que o uso de *corp*us pode ter no ensino de idiomas. Isso envolve a influência no desenvolvimento de materiais didáticos, currículos e metodologias de ensino. Os *corp*us fornecem informações valiosas sobre a frequência de uso de certas estruturas linguísticas, colocações e padrões, o que pode orientar a seleção de conteúdo e abordagens de ensino mais eficazes e alinhadas com o uso real da língua.

Essas aplicações, diretas e indiretas, destacam a importância dos *corp*us no ensino de idiomas, não apenas como uma ferramenta prática para a sala de aula, mas também como um recurso que influencia a maneira como os idiomas são ensinados, estruturados e compreendidos em um contexto mais amplo.

3.3 Lexical Bundles

No próximo quadro será apresentado um mergulho aprofundado nas definições de *lexical bundles* por meio da perspectiva de cinco trabalhos distintos. Cada definição oferece uma lente única para compreender essas unidades linguísticas, destacando nuances e abordagens específicas. A análise comparativa dessas perspectivas enriquecerá nossa compreensão sobre a natureza e o papel dos *lexical bundles*, fornecendo um panorama abrangente que servirá como base sólida para as investigações subseqüentes neste trabalho. Este quadro não apenas evidenciará as convergências e divergências nas concepções dos autores, mas também oferecerá uma plataforma para a reflexão crítica sobre o entendimento atual dessas estruturas linguísticas fundamentais.

Quadro 4 - Conceituação de *lexical bundles*

AUTORES	DEFINIÇÃO
Biber, Conrad, Cortes (2003, p. 73)	Pacotes lexicais são definidos como seqüências de palavras múltiplas que ocorrem pelo menos 20 vezes por milhão de palavras, embora muitos deles sejam muito mais comuns (ocorrendo mais de 200 vezes por milhão de palavras). ⁹
Conrad, Biber (2004, p. 59)	Definimos pacotes lexicais como as seqüências lexicais fixas mais frequentemente recorrentes em um registro. Quanto mais comum um pacote lexical, mais útil ele parece ser na construção do discurso, mas exatamente onde definir um limite de frequência é algo um tanto arbitrário. ¹⁰
Biber (2009, p. 17)	Lexical bundles são seqüências recorrentes de palavras que são altamente resistentes à substituição e eliminação, representando fragmentos pré-fabricados de múltiplas palavras que são relativamente difíceis de analisar gramaticalmente. ¹¹
Cortes (2002, p. 1)	Pacotes lexicais podem ser descritos como colocações estendidas, grupos de três ou mais palavras que ocorrem juntas em um registro, independentemente de sua estrutura ou idiomatidade. ¹²

9 Lexical bundles are defined as the multi-word sequences that occur at least 20 times per million words, although many of them are much more common (occurring more than 200 times per million words).

10 We define lexical bundles as the most frequent recurring fixed lexical sequences in a register. The more common a lexical bundle, the more useful it would appear to be in building discourse, but precisely where to set a frequency cut-off is somewhat arbitrary.

11 Lexical bundles are recurring sequences of words that are highly resistant to substitution and deletion, representing multi-word prefabricated chunks that are relatively difficult to analyze grammatically.

12 Lexical bundles can be described as extended collocations, groups of three or more words that occur together in a register, regardless of their structure or idiomatity.

Cortes (2015, p. 200)

Os pacotes lexicais são definidos como sequências formulaicas recorrentes, mas frequentemente não são unidades estruturais completas. Embora alguns pacotes lexicais sejam considerados como colocações estendidas, eles diferem desse tipo de expressão principalmente na classe de palavras dos componentes que compõem os pacotes lexicais.¹³

Fonte: traduzido pela autora

As citações têm em comum o tema "*lexical bundles*" e oferecem uma explicação abrangente sobre esse conceito. Elas fornecem informações sobre como os *lexical bundles* são definidos, destacam sua importância na construção do discurso e mencionam a variação na frequência com que essas sequências de palavras ocorrem. Além disso, as citações também abordam a relação entre *lexical bundles* e colocações, incluindo as colocações estendidas, o que fornece uma visão mais abrangente sobre como essas construções linguísticas são usadas na comunicação.

A citação de Conrad, Biber (2004) é a mais abrangente entre as mencionadas. O motivo para isso reside no fato de que essa definição não somente explica o que constitui os *lexical bundles*, mas também aborda a importância dessas sequências na construção do discurso e investiga a problemática de estabelecer um limite de frequência, reconhecendo a natureza um tanto subjetiva dessa decisão.

Enquanto as outras definições se concentram principalmente na descrição dos *lexical bundles* e em como eles se manifestam, a segunda definição vai mais além, contextualizando a relevância dessas sequências na linguagem e enfatizando as nuances envolvidas na determinação dos limites de frequência. Isso faz com que essa definição seja mais completa e esclarecedora para aqueles que buscam uma compreensão mais abrangente sobre o tema.

Tomando como referência as conceituações apresentadas no quadro acima, definimos *lexical bundles*, também conhecido como pacote de palavras, como sequências de múltiplas palavras que aparecem frequentemente em textos ou registros de linguagem. Essas sequências podem consistir de três palavras ou mais e são suficientemente recorrentes para serem consideradas expressões fixas na língua. A frequência dos *lexical bundles* pode variar, mas geralmente são agrupamentos de palavras que ocorrem juntas com relevância estatística. Eles desempenham um papel fundamental na compreensão e produção de textos, pois

¹³ Lexical bundles are defined as recurrent formulaic sequences, but they are frequently not complete structural units. Although some lexical bundles are regarded as extended collocations, they differ from this type of expressions mostly in the word class of the components that make up lexical bundles.

frequentemente representam unidades de expressão prontamente acessíveis na língua, tornando a comunicação mais eficaz.

Biber, Conrad e Cortes (2003) defendem a visão de que nenhum enfoque único pode oferecer todas as informações necessárias. Pelo contrário, a compreensão abrangente da importância global das combinações de várias palavras na linguagem só pode ser alcançada através da síntese de resultados provenientes de pesquisas empíricas em várias perspectivas. Este estudo em particular explora a utilidade de uma dessas perspectivas, adotando uma abordagem orientada para a frequência e conduzindo uma análise comparativa abrangente das combinações de múltiplas palavras em registros de orais e escritos. Apoiamos a posição dos autores em foco, entendemos que a exploração léxico gramatical interdependente apresenta potencial educacional satisfatório.

A perspectiva de Biber, Conrad e Cortes (2003), ao argumentar que nenhum único método pode oferecer todas as informações necessárias acerca de combinações de palavras na linguagem, ressalta a complexidade e a riqueza desse domínio de estudo. Eles enfatizam a necessidade de integrar descobertas de pesquisas empíricas de diversas perspectivas para obter uma compreensão global.

O estudo atual adere a essa visão e adota uma abordagem focada na frequência, conduzindo uma análise comparativa ampla de combinações de palavras em registros orais e escritos. Isso fortalece a ideia de que explorar elementos léxico-gramaticais interligados oferece um potencial educacional considerável. Em outras palavras, compreender a relação entre palavras e suas combinações pode ser uma ferramenta valiosa para o ensino e para um aprofundamento na compreensão da linguagem. Assim, a pesquisa respalda a posição dos autores e reconhece a utilidade desse enfoque no contexto educacional.

Biber, Conrad e Cortes (2003) exploram a visão tradicional de que a linguagem segue um modelo composicional, no qual as unidades menores se combinam para formar estruturas maiores, como fonemas formando morfemas e morfemas formando palavras, e assim por diante. No entanto, nos últimos 20 anos, um número crescente de linguistas questionou essa perspectiva, argumentando que uma parte considerável da nossa comunicação cotidiana é baseada em expressões pré-existentes. Há evidências de que algumas combinações de várias palavras são armazenadas em nossos repertórios mentais como unidades distintas. Em outras palavras, falantes nativos empregam tanto estruturas compostas quanto expressões de várias palavras que funcionam como unidades predefinidas.

Essa abordagem destaca a ideia de que a língua é mais flexível e complexa do que apenas seguir um modelo rígido de construção de palavras e frases, e que o uso de unidades predefinidas desempenha um papel importante em nossa comunicação diária.

Biber, Johansson, Leech, Conrad e Finegan (1999) destacam que as sequências de palavras, conhecidas como "*lexical bundles*" ou *lexical bundles*, não precisam necessariamente ser estruturas linguísticas completas ou expressões fixas. Ao examinar esses *lexical bundles*, eles oferecem uma nova perspectiva para entender a língua. Anteriormente, o ensino de idiomas costumava focar exclusivamente em regras gramaticais e vocabulário, com essas duas áreas sendo tratadas separadamente. No entanto, a análise de *lexical bundles* permite uma integração mais eficaz entre gramática e vocabulário, proporcionando uma visão de como ela é realmente usada na comunicação cotidiana. Isso representa uma mudança significativa na maneira como a língua é ensinada e compreendida.

A pesquisa realizada por Biber, Conrad e Cortes (2003) foi distinta em sua abordagem, pois empregou análises empíricas em grandes conjuntos de dados linguísticos, cada um contendo cerca de 5 milhões de palavras. O que a tornou notável foi a sua ênfase na identificação de combinações de palavras com base em critérios de alta frequência. Além disso, o estudo se destacou da maioria das pesquisas anteriores, uma vez que concentrou sua atenção em sequências de palavras mais longas, abrangendo conjuntos de 4, 5 e 6 palavras. Isso resultou em uma compreensão mais profunda das combinações de palavras usadas na linguagem, indo além das análises tradicionais de pares de palavras ou tríades, e oferecendo informações valiosas sobre as estruturas linguísticas mais complexas e como são usadas na comunicação.

O trabalho de Biber, Conrad e Cortes (2003) é fundamental para a compreensão das diferenças entre a linguagem usada na escrita acadêmica e na conversação. Ao classificar *lexical bundles* com base em seus contextos de discurso e permitir comparações entre registros, eles destacam como a linguagem é adaptada para diferentes finalidades comunicativas. As notáveis disparidades entre a prosa acadêmica, com sua ênfase em marcadores de referência e organização do texto, e a conversação, que se concentra em expressar atitudes e interações, ressaltam a flexibilidade e a complexidade da linguagem. Ambos os registros dependem de combinações de palavras frequentes, mas suas estruturas e funções variam consideravelmente, enfatizando a importância de considerar o contexto ao analisar a linguagem. Essas descobertas têm implicações significativas para a pesquisa linguística e o ensino da linguagem, destacando a necessidade de reconhecer as diferenças entre diferentes tipos de discurso e como a linguagem se adapta a esses contextos.

O estudo de Chen e Baker (2010) emprega um método automatizado que se baseia na frequência para identificar combinações de palavras comuns, chamadas *lexical bundles*, na escrita acadêmica. Foram analisados *lexical bundles* provenientes de um cópulo de textos acadêmicos publicados e de dois cópulos de escrita acadêmica de estudantes (um em L1 e o outro em L2) de forma abrangente, tanto quantitativa quanto qualitativamente. Os resultados revelaram que a escrita acadêmica publicada apresenta uma ampla variedade de *lexical bundles*, enquanto a escrita dos estudantes em L2 demonstra uma variedade mais limitada. Além disso, algumas expressões frequentes em textos publicados, como “*in the context of*” não foram usadas com frequência em nenhum dos cópulos de estudantes, enquanto os escritores em L2 faziam um uso excessivo de certas expressões, por exemplo: “*all over the world*”, que eram raramente empregadas por acadêmicos nativos. As conclusões provenientes das análises estruturais e funcionais dos *lexical bundles* também têm importantes implicações pedagógicas.

O estudo comparativo de Chen e Baker (2010) evidenciou as diferenças e semelhanças essenciais entre a produção acadêmica de escritores nativos e aprendizes. Através de comparações estruturais e funcionais, foi constatado que o uso de *lexical bundles* em ensaios de estudantes nativos e não nativos é notavelmente similar. No entanto, uma análise qualitativa mais aprofundada revelou que a escrita de estudantes nativos compartilha algumas características distintivas da escrita acadêmica, como o domínio do uso de linguagem cuidadosa na escrita profissional nativa. Por outro lado, a escrita não nativa exibe uma tendência que parece ser exclusiva da escrita em L2, incluindo a tendência à generalização excessiva e à preferência por certas expressões idiomáticas e conectores.

Conrad e Biber (2004) propõem que a análise da frequência e emprego em diversos cenários linguísticos, tais como interações cotidianas e produção acadêmica, pode conceder perspicácia a respeito de diversidade linguística e modelos de comunicação. Portanto, mais uma vez os *lexical bundles* se fazem importantes e necessários, pois podem demonstrar diferentes tipos de padrões na fala (seja ela escrita ou oral).

Os autores Conrad e Biber (2004) enfatizam a importância dos *lexical bundles* como uma parte significativa do discurso falado e escrito, que não ocorre por acaso. Eles demonstram que diferentes tipos de comunicação dependem de diferentes conjuntos de *lexical bundles*, sugerindo eles desempenham um papel estrutural e funcional vital nos discursos de seus respectivos registros. A ideia de que seu uso não é acidental, mas sim propositado e adaptado ao contexto, sublinha a importância desses bundles no entendimento do discurso e na análise linguística.

A ênfase dada por Conrad e Biber (2004) à importância dos *lexical bundles* no discurso oral e escrito destaca a compreensão de que essas sequências de palavras não são o resultado do acaso, mas sim uma parte significativa e intencional da linguagem. Os autores apontam que diferentes tipos de comunicação dependem de conjuntos específicos de *lexical bundles*, o que sugere que essas combinações desempenham papéis essenciais na estrutura e função dos discursos em diferentes contextos e registros. A ideia de que o uso dessas sequências não é aleatório, mas sim propositado e adaptado ao contexto em que são empregados, enfatiza a importância de compreender os *bundles* para analisar a linguagem e o discurso de forma mais profunda. Em suma, essa passagem destaca o valor dos *lexical bundles* na análise linguística e na interpretação do discurso, evidenciando que eles são elementos essenciais para compreender como a linguagem é usada de maneira eficaz em diversos contextos comunicativos.

No estudo de Cortes (2015), o foco inicial foi traçar uma perspectiva histórica da evolução das pesquisas sobre *lexical bundles*, com ênfase posterior na identificação e categorização dessas expressões, sobretudo em investigações que se debruçaram sobre a utilização de *lexical bundles* em contextos de escrita acadêmica. É inegável que as pesquisas pioneiras destacaram a frequência como o traço mais proeminente dessas expressões. A frequência, sem dúvida, é um componente crucial, porém estudos posteriores também demonstraram que as funções desempenhadas por esses *lexical bundles* desempenham um papel significativo em sua importância e relevância no discurso e na escrita acadêmica.

A extensa utilização das estruturas taxonômicas já definidas para a classificação de *lexical bundles* terá um impacto positivo na consolidação das categorias e subcategorias dentro desses sistemas, permitindo a inclusão de novas quando relevante. Isso proporciona uma estrutura sólida e flexível para o estudo dessas unidades linguísticas recorrentes.

Adicionalmente, as mais recentes inovações que estabelecem conexões entre *lexical bundles* e funções comunicativas, evidenciadas por meio de movimentos retóricos, realçam a importância de pesquisas que analisem essas expressões no contexto do discurso em que estão inseridas. Isso implica que entender como os *lexical bundles* são usados para cumprir funções específicas na comunicação é fundamental para uma compreensão completa de sua relevância e impacto nos textos e discursos. Portanto, o estudo da relação entre essas expressões e o contexto discursivo torna-se essencial para uma análise abrangente e aprofundada dessas unidades linguísticas.

Apesar de parecer simples aprender e usar essas expressões devido à sua alta ocorrência, é essencial notar que pesquisas abrangentes indicam que escritores iniciantes não

as incorporam com a mesma frequência nem com os mesmos propósitos que escritores experientes. Isso destaca que a utilização eficaz de *lexical bundles* envolve mais do que simplesmente reconhecê-los; requer uma compreensão aprofundada de como essas expressões se encaixam no contexto da escrita.

Além disso, é crucial avaliar como o uso de *lexical bundles* se relaciona com o nível de habilidade na escrita e a experiência na produção de textos. Esse exame é fundamental para identificar abordagens mais eficazes destinadas a orientar escritores iniciantes na utilização apropriada dessas expressões em contextos acadêmicos. Compreender como a proficiência na escrita influencia o emprego de *lexical bundles* é um passo importante para aprimorar a escrita acadêmica e auxiliar escritores em seu desenvolvimento e aprimoramento na expressão escrita.

Biber (2009) explorou a aplicação de uma abordagem radicalmente orientada por córpus para o estudo da linguagem formulaica em inglês. As análises realizadas neste estudo revelaram tendências surpreendentes, com destaque para a preferência por sequências linguísticas fixas e contínuas em contextos de conversação, em contraposição à inclinação por estruturas formulaicas que possibilitam variações internas em contextos de escrita. Além disso, a análise identificou as diferentes funções desempenhadas pelas palavras de conteúdo nas partes estáticas em comparação com aquelas nas partes flexíveis.

Essas descobertas têm implicações significativas para nossa compreensão de como o discurso é produzido e compreendido na fala em contraste com a escrita, bem como para o processo de aprendizado e ensino de discursos orais em comparação com discursos escritos. Elas fornecem insights valiosos sobre as características distintas da linguagem utilizada em diferentes contextos de comunicação, o que pode ser relevante para o desenvolvimento de estratégias pedagógicas eficazes.

Este comentário destaca a consistência das conclusões deste estudo com pesquisas anteriores baseadas em córpus, enfatizando a distinção intrínseca entre a gramática da fala e a gramática da escrita. O estudo ressalta que, durante conversas, as orações, incluindo as orações subordinadas, desempenham um papel fundamental no discurso, enquanto na escrita informativa, as frases, especialmente as frases nominais e preposicionais, estabelecem a base do discurso.

Essa distinção entre a fala e a escrita é respaldada pela longa trajetória de pesquisa, com Halliday sendo um dos pioneiros nesse entendimento desde os anos 1970, com foco especial na importância da nominalização no discurso acadêmico escrito.

Essas constatações ressaltam a complexidade da linguagem e a necessidade de compreender como ela varia e se adapta em diferentes contextos de comunicação, seja na fala ou na escrita, o que tem implicações significativas para o ensino e a aprendizagem de línguas, bem como para a análise linguística.

Biber, Egbert, Keller e Wizner (2021) conduziram uma pesquisa baseada em um corpus de 2,3 milhões de palavras de interações conversacionais do British National Corpus (2014). Seus resultados revelam que a maior parte do discurso em conversas é composta por sequências de unidades de discurso coerentes que têm objetivos de comunicação identificáveis. Essas unidades de discurso podem ser classificadas em diferentes tipos de discurso conversacional, cada um com propósitos de comunicação específicos, mesmo que não sigam estruturas de gênero convencionais, como narrativas pessoais ou piadas. A pesquisa identificou 16 tipos de discurso conversacional associados a combinações de 9 propósitos de comunicação gerais. É interessante observar que os tipos de discurso focados na expressão de opiniões pessoais, sentimentos e avaliações são comuns em conversas. Além disso, o estudo enfatiza que os tipos de discurso destinados à transmissão de informações são igualmente frequentes nas conversações.

Pesquisadores de diversas correntes acadêmicas têm investigado como os interlocutores, por vezes, utilizam unidades de discurso mais unificadas em suas conversações. Para a maioria dos estudiosos, a conversa geralmente segue uma estrutura de sequências locais de intervenções, com unidades de discurso mais abrangentes sendo uma ocorrência incomum, reservada principalmente para gêneros de fala culturalmente reconhecidos que se encaixam no curso normal das conversas.

Neste processo, Biber, Egbert, Keller e Wizner (2021) questionaram duas suposições teóricas fundamentais de pesquisas anteriores: a ideia de que a conversação é predominantemente organizada como sequências de turnos (com unidades de discurso de nível superior sendo consideradas especiais e não comuns), e a concepção de que unidades de discurso de nível superior - quando ocorrem - são manifestações de gêneros culturais reconhecidos de forma especial (como piadas ou narrativas pessoais) com estruturas convencionais. Pelo contrário, demonstraram que: 1) a maior parte do discurso conversacional é composta por sequências de unidades de discurso coesas que têm objetivos comunicativos identificáveis, e 2) essas unidades de discurso podem ser classificadas em tipos de discurso conversacional que têm propósitos comunicativos distintos, embora geralmente não sigam uma estrutura de gênero convencional.

Biber, Conrad e Cortes (2003) esclareceram que as pesquisas iniciais sobre unidades de várias palavras que se repetem adotaram uma abordagem subjetiva, mas os avanços tecnológicos e a coleta de grandes corpú linguísticos resultaram em alterações nas abordagens de pesquisa. Durante as décadas de 1990 e 2000, houve um aumento significativo em estudos voltados para a definição e classificação de combinações de palavras fixas. Essas pesquisas apresentaram diferenças notáveis, incluindo a metodologia empregada na identificação e descrição de diversos tipos de unidades de várias palavras, o enfoque na expressão fixa investigada e os corpú utilizados para análise.

A identificação de *lexical bundles*, como exemplo, envolveu a avaliação de grupos de n palavras em um corpú por meio de um software de computador. Apenas as expressões que atendiam a um limite predefinido eram consideradas *lexical bundles*. Biber et al. (1999) inicialmente propuseram um limite de frequência de dez vezes em um milhão de palavras para combinações de quatro palavras, mas pesquisas posteriores adotaram critérios mais rigorosos, elevando o limite para vinte ou quarenta vezes em um milhão de palavras.

Em resumo, as pesquisas pioneiras sobre combinações de palavras recorrentes passaram por uma transição significativa, abandonando uma abordagem subjetiva em favor de métodos mais sistemáticos e avançados na análise linguística. É fundamental destacar as contribuições de outros pesquisadores que dedicaram ao estudo de *lexical bundles* por meio de corpú.

A pesquisa de Biber, Conrad e Cortes em 2003 identificou uma série de "*lexical bundles*" em textos, demonstrando como essas sequências de palavras são frequentemente usadas em linguagem escrita e falada. Alguns exemplos de "*lexical bundles*" mencionados em sua pesquisa podem incluir:

1. "*at the same time*"
2. "*in the case of*"
3. "*as a result of*"
4. "*in terms of*"
5. "*on the other hand*"
6. "*in order to*"
7. "*in the context of*"
8. "*in this paper*"
9. "*on the basis of*"
10. "*in the following*"

Esses exemplos representam *lexical bundles* comuns que foram identificados na pesquisa e que ocorrem frequentemente em textos. Eles são sequências de palavras pré-fabricadas que desempenham um papel importante na construção do discurso e na comunicação eficaz. A pesquisa de Biber, Conrad e Cortes (2003) ajudou a destacar a recorrência dessas unidades predefinidas na linguagem.

A pesquisa conduzida por Biber, Conrad e Cortes em 2003 revelou descobertas relevantes no estudo dos *lexical bundles*, sequências de palavras recorrentes na linguagem escrita e falada. Essas descobertas incluíram a observação de que certas combinações de palavras ocorrem com alta frequência em textos, indicando que não são meramente sequências aleatórias, mas unidades predefinidas na linguagem.

Um aspecto importante identificado na pesquisa foi a variação na frequência e composição dos *lexical bundles* entre diferentes gêneros de texto e registros linguísticos. Isso sugeriu que o uso dessas sequências está fortemente relacionado ao contexto de comunicação em que são encontradas.

Além disso, a pesquisa enfatizou o papel crítico desempenhado pelos *lexical bundles* na construção de significado e na estruturação do discurso. Essas sequências de palavras contribuem para a coesão e a organização do discurso, facilitando uma comunicação eficaz.

No geral, a pesquisa de Biber, Conrad e Cortes (2003) ampliou nossa compreensão da linguagem e da importância das unidades predefinidas na comunicação. Suas descobertas têm tido um impacto significativo no campo da linguística e na análise da linguagem.

A pesquisa de Biber (2009) sobre *lexical bundles* apresenta vários exemplos dessas sequências. Lembre-se de que os *lexical bundles* são sequências de palavras frequentemente usadas que tendem a ocorrer juntas em textos, representando unidades predefinidas na linguagem. Alguns exemplos de *lexical bundles* citados em sua pesquisa podem incluir:

1. "at the end of the"
2. "in terms of the"
3. "it is important to"
4. "as a result of"
5. "in order to"
6. "on the other hand"
7. "in the context of"
8. "the nature of the"

9. “*you know what I mean*”

10. “*I don’t want to*”

Esses são apenas alguns exemplos, e na pesquisa de Biber (2009), ele explora uma variedade de sequências de palavras que são consideradas *lexical bundles* com base em sua recorrência em corpúscos linguísticos. Cada exemplo representa uma combinação comum de palavras que é frequentemente encontrada em textos. A pesquisa e análise de corpúscos linguísticos realizadas por Douglas Biber trouxeram contribuições significativas para o campo da linguística, resultando em várias descobertas e insights relevantes. Biber identificou uma variedade de sequências de palavras comuns que são frequentemente usadas em textos, destacando que a linguagem muitas vezes se baseia em unidades predefinidas ou colocações de palavras. Além disso, sua pesquisa revelou que a frequência e a composição desses *lexical bundles* variam consideravelmente entre diferentes gêneros e registros linguísticos, sugerindo que a escolha de palavras predefinidas está fortemente relacionada ao contexto comunicativo.

A importância dos *lexical bundles* na análise do discurso foi enfatizada por Biber, uma vez que essas sequências de palavras desempenham um papel fundamental na construção de significado e na expressão de ideias. Além disso, sua pesquisa lançou luz sobre como os aprendizes de línguas adquirem e utilizam essas sequências de palavras predefinidas em seus discursos.

O trabalho de Biber (2009) também contribuiu para o avanço da linguística de corpúscos, uma abordagem que enfatiza a análise de extensos conjuntos de dados linguísticos para compreender padrões e estruturas na linguagem. Em resumo, as descobertas de Biber expandiram nossa compreensão da estrutura da linguagem e do papel das unidades predefinidas na comunicação e na construção de significado em diversos contextos linguísticos. Seu impacto continua a ser relevante na pesquisa linguística e na análise da linguagem.

3.4 Interculturalidade

A interculturalidade refere-se à interação e convivência entre diferentes culturas, promovendo a troca de valores, ideias, práticas e conhecimentos entre pessoas de origens diversas. É um conceito que reconhece e valoriza a diversidade cultural, buscando a compreensão mútua e o diálogo respeitoso entre indivíduos ou grupos que representam diferentes origens étnicas, linguísticas, religiosas, sociais e históricas.

Este conceito vai além da simples coexistência de culturas, visando à criação de ambientes onde as diferenças são reconhecidas, respeitadas e utilizadas como oportunidades para aprendizado e enriquecimento mútuo. A interculturalidade promove a tolerância, a aceitação e a valorização da diversidade como um ativo para o desenvolvimento de sociedades mais inclusivas e harmoniosas.

Neste contexto, a interculturalidade é fundamental em várias esferas da vida contemporânea, incluindo a educação, o ambiente de trabalho, as relações sociais e o cenário global. Ela desempenha um papel crucial na formação de cidadãos mais conscientes, capazes de compreender e interagir de forma construtiva em sociedades multiculturalmente diversas.

Promover a interculturalidade implica em reconhecer a igualdade de direitos, a justiça social e a valorização das diferentes perspectivas culturais. Ela é essencial para o estabelecimento de pontes entre culturas, contribuindo para a construção de um mundo mais solidário, onde a diversidade é reconhecida como um elemento enriquecedor da experiência humana.

A associação entre proficiência linguística e competência sociocultural torna-se essencial ao tratar-se do entendimento de uma outra língua, pois a cultura influencia decisivamente nos valores da comunidade, interação, normas, fazendo com que a compreensão da língua, de fato, dependa da prévia compreensão de mundo acerca da língua estudada.

Na década de 60, no Brasil, a abordagem predominante no ensino da língua portuguesa estava fortemente fundamentada na visão de linguagem como expressão do pensamento. Os professores direcionavam seus esforços para o ensino de conceitos normativos, enfatizando o domínio da metalinguagem. Esse período foi marcado por uma prática educacional preocupada principalmente com a transmissão de conhecimentos.

Conforme apontado por Zanini (1999), o ensino de português nessa época concebia a língua como um sistema. A instrução de leitura e escrita tinha seu foco em textos literários, enquanto os alunos eram instruídos sobre as normas e regras de funcionamento da língua. Notavelmente, os estudantes desse período pertenciam majoritariamente a classes sociais privilegiadas, muitos dos quais já chegavam à escola com certo domínio da "norma culta".

Esse cenário educacional refletia uma ênfase na estrutura normativa da língua, dando menos destaque à comunicação efetiva e à compreensão contextualizada da linguagem. Barbosa (2009) aponta que ao definir-se como prática e produto sócio-histórico, a língua é o melhor meio de acesso à cultura, já que é, simultaneamente, veículo, produto e produtor de todas as culturas.

A cultura propriamente dita é disseminada por meio do léxico de uma língua. Sarmiento (2001), que o papel do professor não é o de fazer com que os alunos se comportem como as pessoas da cultura alvo, mas de dar espaço para que os alunos interpretem os significados dessa cultura. Portanto, o docente precisa tomar conhecimento da cultura presente na língua ensinada, para que assim os alunos sejam capazes de realmente interpretar e entender, de fato, a língua.

Léxico é o conjunto dos vocábulos pertencentes a uma língua. Cultura é compreendida pelos comportamentos e tradições de um determinado povo. Ao unir esses dois conceitos, encontramos léxicos presentes em diferentes culturas, e que devem ser levados em consideração. Se a cultura do povo determina a língua, é necessário conhecer a cultura para compreender a língua.

A escola de qualidade, por meio dos conteúdos e das relações interpessoais que seu ambiente proporciona, contribui para o desenvolvimento humano de forma integral, respeitando e valorizando as diferenças que são encontradas ali (BRASIL, 2013). Com o passar dos anos e avanços na educação fez-se necessário, cada vez mais, educarmos seres críticos e pensantes, para que possam compreender as diferentes culturas e escolher como agir frente à elas, e assim resultará na “construção de um cidadão crítico perante as fortes investidas internacionais em divulgar e conseguir adeptos as suas práticas sem que esses tenham a possibilidade de decidir o que de fato eles querem assimilar ou não” (PAULINO, 2009, p. 53). A proposta ideal é formar cidadãos que consigam distinguir as diversas culturas e saibam interpretá-las nos momentos necessários para estarem preparados, efetivamente, para o mundo do trabalho e as práticas sociais.

É de suma importância que o docente foque nas potencialidades de cada aluno para aprimorá-las, mas que também reconheça as dificuldades, para desenvolvê-las. O aluno precisa se sentir confiante e capaz de exercer as funções que lhe for atribuído. As línguas exercem um papel de ponte entre o indivíduo falante e as tradições, crenças e manifestações. Por meio da língua é possível entender um povo e sua cultura.

Segundo Silva (2017) cultura é o conjunto de realizações humanas, materiais ou imateriais leva-nos a caracterizá-la como um fundamento básico da História, que por sua vez pode ser definida como o estudo das realizações humanas ao longo do tempo. Portanto, o que poderia ser melhor para materializar essa cultura e mudança das pessoas, se não a língua?

De acordo com Geertz (1989), cultura é a própria condição de existência dos seres humanos, produto das ações por um processo contínuo, através do qual, os indivíduos dão sentido à suas ações, sendo que ocorre na mediação das relações dos indivíduos entre si, na

produção de sentidos e significados. Não obstante de tal pensamento, é de exímia importância que o profissional educador se adeque à culturalidade do ensino proposto, uma vez que infringir tal premissa pode resultar no não entendimento do tema que venha a ser discutido em aula.

Terry Eagleton (2005), aborda o conceito de cultura de maneira abrangente e multifacetada. Em suas obras, Eagleton oferece uma visão complexa e dinâmica da cultura, que vai além de uma definição estática. Em suas análises, ele destaca que a cultura não se limita apenas às manifestações artísticas ou intelectuais de uma sociedade, mas é um conjunto amplo e em constante evolução de práticas, valores, crenças, instituições e produções simbólicas que caracterizam uma comunidade ou grupo social.

Para Eagleton (2005), a cultura é um fenômeno profundamente enraizado na vida cotidiana, influenciado por fatores políticos, sociais, econômicos e históricos. Ele enfatiza a natureza dinâmica e mutável da cultura, que está em constante interação e transformação, moldada por diferentes contextos e relações de poder.

Portanto, a definição de cultura segundo Eagleton é mais ampla e inclusiva, considerando-a como um conjunto complexo de elementos que permeiam todas as esferas da vida e que estão sujeitos a interpretações e mudanças constantes.

A compreensão da importância do léxico cultural estar presente nas aulas de língua inglesa insere o falante na comunidade linguística, além de possibilitar o entendimento e a produção de enunciados. A compreensão do ensino não deve ser realizada somente a partir da ótica teórica do conteúdo, e deve ter como base a junção do léxico, cultura e ensino. Há habilidades linguísticas essenciais para o aprendizado de uma língua, mas que ao caminhar em conjunto com a compreensão e absorção da cultura, faz-se mais eficaz.

Paulino (2009) comenta que o professor pode tornar-se um catalizador no que se refere à percepção cultural, estimulando a pesquisa e multiplicando o respeito às diferenças de língua. Espera-se que o docente tenha o conhecimento necessário da cultura anteriormente, para assim poder repassar para seus alunos o verdadeiro sentido da língua em foco. O professor intercultural consegue atingir a dimensão cultural da língua. O ensino da língua inglesa contribui com o desenvolvimento não somente intelectual do aluno, mas também do desenvolvimento físico, social e psicológico.

Trabalhar a competência intercultural nas aulas de língua estrangeira é crucial nos dias de hoje. A habilidade de compreender e interagir de forma eficaz com pessoas de diferentes

culturas vai além do mero domínio linguístico. Envolve uma compreensão mais profunda dos contextos sociais, históricos e culturais que moldam a língua em questão.

Ao integrar elementos interculturais no ensino de línguas estrangeiras, os alunos têm a oportunidade de não apenas aprender vocabulário e gramática, mas também de explorar e apreciar as nuances culturais que influenciam a língua. Isso permite uma melhor compreensão do uso da linguagem em diferentes situações e contextos sociais, além de promover a empatia, o respeito e a tolerância em relação a outras culturas.

A inclusão da competência intercultural nas aulas de língua estrangeira não apenas enriquece a experiência de aprendizagem, mas também prepara os alunos para interagir de maneira mais eficaz e respeitosa em um mundo globalizado, onde a comunicação intercultural é cada vez mais essencial.

Segundo Juvino; De Assunção Barbosa (2013, p. 01):

Aprender uma língua estrangeira compreende, além do domínio gramatical da língua, conhecimento cultural, haja vista que todas as palavras acarretam, além de seu sentido literal, outras acepções que foram incorporadas paulatinamente através dos hábitos culturais de determinada comunidade. Ao estudo dessa dimensão cultural da palavra, Robert Galisson deu nome de lexicultura e ao “viés cultural” agregado ao léxico, carga cultural compartilhada.

A citação acima ressalta a grande importância do aprendizado de línguas estrangeiras para além da gramática, destacando a relevância do conhecimento cultural. Galisson introduz o termo 'lexicultura', referindo-se à carga cultural intrínseca às palavras, além de seus significados literais. Essa noção enfatiza como as palavras são moldadas pelos hábitos culturais de uma comunidade, ampliando a compreensão linguística para além das fronteiras gramaticais. Nesse contexto, a compreensão da 'lexicultura' torna-se essencial no processo de domínio de uma língua estrangeira, enriquecendo a compreensão cultural e contextual do idioma estudado.

Mendes (2004, p.177) apud Santos; Alvarez (2010, p.177) afirma que:

[...] os materiais didáticos para o ensino de LE/L2, de maneira geral, não incentivam as relações interculturais, uma vez que os conteúdos veiculados quase sempre representam aspectos estanques da cultura da língua-alvo, expostos através de amostras descontextualizadas, que servem apenas para “praticar” os aspectos da língua.

A observação sobre os materiais didáticos para o ensino de línguas estrangeiras é pertinente e levanta uma questão fundamental. De fato, muitos desses materiais costumam focar primariamente na prática linguística, negligenciando as nuances culturais e as relações interculturais que são essenciais para um aprendizado integral e contextualizado.

É comum encontrar nos livros e recursos para o ensino de línguas estrangeiras exemplos descontextualizados, representando aspectos isolados da cultura associada à língua-alvo. Essa abordagem reducionista não apenas limita a compreensão cultural dos alunos, mas também pode reforçar estereótipos ou visões superficiais sobre a cultura em questão.

Os materiais didáticos têm um papel crucial no processo de aprendizagem de idiomas, e sua abordagem deveria ser mais abrangente, integrando não apenas a prática linguística, mas também elementos culturais autênticos e contextualizados. A inclusão de conteúdos interculturais mais ricos e diversificados pode não só enriquecer o aprendizado linguístico, mas também promover uma compreensão mais profunda e respeitosa das diferentes culturas associadas à língua em estudo.

Portanto, repensar a maneira como os materiais didáticos abordam a cultura associada à língua-alvo é crucial para criar experiências de aprendizado mais enriquecedoras, contextualizadas e culturalmente sensíveis para os alunos de línguas estrangeiras.

Ao longo do tempo, as estratégias para ensinar o componente cultural em aulas de línguas estrangeiras têm evoluído, abrangendo diversas abordagens. Inicialmente, as perspectivas se concentravam em conceitos de assimilação cultural ou aculturação, enfatizando a absorção da cultura do país de destino como parte do processo de aprendizagem de uma língua estrangeira. Essas abordagens muitas vezes promoviam a ideia de que o aprendiz deveria assimilar e se adaptar à cultura alvo.

Posteriormente, houve uma transição em direção à interculturalidade, reconhecendo a importância do diálogo entre culturas, destacando a necessidade de compreensão mútua e troca de perspectivas entre a cultura do aluno e a cultura estudada. Isso permitiu uma abordagem mais equilibrada, na qual as diferenças culturais eram valorizadas e vistas como um meio enriquecedor de aprendizagem.

Além disso, surgiram abordagens mais abrangentes, como o multi e o pluriculturalismo, que reconhecem não apenas a interação entre duas culturas, mas a presença e a inter-relação de múltiplas culturas dentro de uma sociedade ou contexto. Essas perspectivas ampliaram a compreensão da diversidade cultural, enfatizando a coexistência e o entrelaçamento de diferentes identidades culturais.

Contudo, é importante notar que essas abordagens muitas vezes estiveram centradas no estudo da "cultura do outro", o que pode limitar a compreensão da complexidade cultural do próprio aluno, deixando de lado a valorização da cultura do aprendiz como parte integrante do processo de aprendizagem de línguas estrangeiras.

Portanto, as mudanças nas abordagens para o ensino do componente cultural em aulas de línguas estrangeiras refletem uma evolução em direção a uma compreensão mais ampla e inclusiva da diversidade cultural, mas ainda há espaço para uma maior valorização e integração da cultura do aprendiz no contexto educacional.

Como comentado por Salomão (2015, p. 384): “Parece-nos pertinente desafiar o ensino do componente cultural como fatos ou prescrições, assim como desfazer a equação “uma língua = uma cultura”, principalmente no ensino de inglês como língua internacional”. Desafiar a abordagem do ensino do componente cultural como fatos estáticos ou prescrições fixas é crucial para promover uma compreensão mais ampla e dinâmica da inter-relação entre língua e cultura, especialmente no contexto do ensino de inglês como língua internacional.

A ideia de que uma língua está intrinsecamente ligada a uma única cultura é uma simplificação que pode limitar a compreensão dos estudantes sobre a diversidade cultural existente dentro da própria língua. No caso do inglês, como língua internacional, é especialmente importante desfazer essa equação "uma língua = uma cultura", pois o inglês é utilizado em contextos diversos ao redor do mundo, com diferentes variantes, influências culturais e usos sociais.

Ao desafiar essa equação, os educadores podem incentivar os alunos a explorar a diversidade cultural presente na língua inglesa, reconhecendo que ela é moldada por múltiplos contextos culturais e experiências. Isso permite uma compreensão mais ampla e crítica da língua, além de promover uma mentalidade aberta e inclusiva em relação à diversidade cultural.

Além disso, ao invés de apresentar a cultura como um conjunto fixo de fatos ou normas a serem memorizados, é mais benéfico estimular os estudantes a analisar e refletir sobre as dinâmicas culturais, valores, crenças e contextos sociais associados à língua em estudo. Isso ajuda a desenvolver habilidades críticas, promove a compreensão da complexidade cultural e encoraja uma abordagem mais reflexiva e contextualizada em relação à língua e à cultura.

Portanto, desafiar a ideia de uma conexão unívoca entre língua e cultura e evitar a apresentação da cultura como um conjunto estático de fatos contribui para uma abordagem mais inclusiva, dinâmica e enriquecedora no ensino de línguas estrangeiras, especialmente no contexto do inglês como língua internacional.

Para Kramsch (2021), a interculturalidade vai além da ideia de apenas aprender sobre outra cultura; trata-se de um processo dinâmico e complexo de negociar significados, valores e identidades entre falantes de diferentes línguas e culturas. Ela destaca que a linguagem

desempenha um papel crucial nesse processo, não apenas como um meio de comunicação, mas como um reflexo e um veículo das diferenças culturais.

Kramersch (2021) argumenta que a interculturalidade se manifesta no espaço entre as culturas, onde ocorre o encontro e a negociação de significados. Nesse espaço, os indivíduos confrontam suas próprias perspectivas culturais com as dos outros, e é nesse confronto que ocorre o verdadeiro processo de aprendizado intercultural.

3.5 Fala x Escrita

Esta seção aborda o processo de comparação entre a fala e a escrita, destacando a transição dos registros orais presentes nos vídeos infantis do YouTube para a forma escrita. Essa transformação foi crucial para a análise crítica e aprofundada do conteúdo apresentado, proporcionando *insights* valiosos sobre as nuances da comunicação dirigida a crianças nesse meio digital.

Na presente pesquisa, utilizamo-nos de registros orais, porém foram transformados em registros escritos para que a análise do *cópus* pudesse acontecer, pois entende-se que um registro oral não pode ser estudado sem que haja uma transcrição.

I. Transcrição de Registros Oraais:

A base fundamental para esta pesquisa foi constituída pelos registros orais presentes nos vídeos infantis analisados. A linguagem falada nessas produções, repleta de entonação, expressividade e elementos não verbais, constitui um aspecto vital para compreender a dinâmica comunicativa destinada ao público infantil. A captura desses elementos foi alcançada por meio da transcrição integral dos diálogos presentes nos vídeos.

II. Processo de Transcrição:

A transcrição, neste contexto, foi realizada por meio da ferramenta automatizada e especializada (*YouTube Scribe*) de reconhecimento de fala. Este método proporcionou uma abordagem eficiente para lidar com grandes volumes de dados de áudio, permitindo a rápida conversão de registros orais em texto escrito. Contudo, é crucial destacar que a automação dessa etapa pode suscitar desafios significativos, principalmente no que diz respeito à precisão da transcrição.

III. Comparação entre Fala e Transcrição:

A comparação entre o conteúdo falado nos vídeos e sua representação transcrita revelou discrepâncias notáveis. A natureza dinâmica da linguagem falada, repleta de entonações específicas e variações tonais, muitas vezes desafiou a capacidade da plataforma de transcrição em capturar nuances sutis. Além disso, expressões idiomáticas, gírias infantis e

elementos culturais específicos podem ter sido interpretados de maneira incompleta ou equivocada.

IV. Margem de Erro na Transcrição:

É essencial reconhecer a existência de uma margem de erro inerente ao processo de transcrição automatizada. A dependência de algoritmos e modelos de linguagem pode resultar em imprecisões, especialmente em situações onde a fala é acelerada, possui sobreposição de vozes ou apresenta sotaques regionais específicos. Essa margem de erro deve ser considerada ao interpretar os resultados da análise, reconhecendo que algumas nuances linguísticas podem ter escapado à precisão da transcrição.

Como comentado por Batista-Santos; Batista Pereira (2017, p. 168):

A fala e a escrita possuem elementos diferentes, próprios e, por utilizar o mesmo sistema linguístico, não devem ser discutidas de forma dicotômica pois, como afirma Marcuschi (2010, p. 21), “a passagem da fala para a escrita não é a passagem do caos para a ordem: é a passagem de uma ordem para outra ordem”. Assim, a oralidade e a escrita não podem ser entendidas no nível de superioridade ou inferioridade em relação à outra.

O comentário apresentado destaca uma perspectiva valiosa sobre a relação entre fala e escrita, enfatizando a complexidade e a riqueza inerentes a ambas as formas de expressão linguística. Ao considerar que a fala e a escrita são elementos diferentes, mas intrinsecamente ligados pelo mesmo sistema linguístico, evita-se uma abordagem dicotômica que poderia reduzir a compreensão da riqueza da linguagem humana.

A citação de Marcuschi (2010) acrescenta uma dimensão importante à discussão, ressaltando que a transição da fala para a escrita não deve ser vista como uma mudança do caos para a ordem, mas sim como a evolução de uma ordem para outra ordem. Isso sugere que ambas as formas de comunicação possuem suas próprias complexidades e estruturas, desafiando a ideia de superioridade ou inferioridade entre oralidade e escrita.

A noção de que a oralidade e a escrita não devem ser entendidas em termos de superioridade ou inferioridade ressalta a importância de reconhecer as diversas funções e contextos em que cada uma é empregada. A fala, com sua espontaneidade e nuances expressivas, e a escrita, com sua formalidade e capacidade de registro, desempenham papéis complementares no vasto espectro da comunicação humana.

Segundo Marcuschi (2010, p. 17) apud Batista-Santos; Batista Pereira (2017, p. 169):

[...] sob o ponto de vista mais central da realidade humana, seria possível definir o homem como um ser que fala e não como um ser que escreve. Entretanto, isto não significa que a oralidade seja superior à escrita, nem traduz a convicção, hoje tão

generalizada quanto equivocada, de que a escrita é derivada e a fala é primária. A escrita não pode ser tida como uma representação da fala.

O ponto de vista apresentado destaca uma perspectiva central na compreensão da realidade humana, ao sugerir que a capacidade de fala é intrínseca à natureza do ser humano. Essa visão enfatiza o papel fundamental da oralidade como uma característica essencial do homem. No entanto, Marcuschi (2010) adverte contra a interpretação simplista de que a oralidade é superior à escrita ou que a escrita é uma forma derivada e secundária de comunicação.

A ideia de que o homem pode ser definido primariamente como um ser que fala ressoa com a experiência cotidiana, uma vez que a comunicação verbal é uma parte integral da vida humana desde tempos imemoriais. No entanto, a ressalva de que isso não implica superioridade da oralidade sobre a escrita é crucial. Ambas as formas de comunicação têm suas próprias características, funções e relevâncias em contextos específicos.

Figura 3 – Características da fala X da escrita

FALA	ESCRITA
Contextualizada	Descontextualizada
Implícita	Explícita
Redundante	Condensada
Não-planejada	Planejada
Predominância do "modus pragmático"	Predominância do "modus sintático"
Fragmentada	Não fragmentada
Incompleta	Completa
Pouco elaborada	Elaborada
Pouca densidade informacional	Densidade Informacional
Predominância de frases curtas, simples ou coordenadas	Predominância de frases completas subordinação abundante
Pequena frequência de passivas	Emprego frequente de passivas
Pouca nominalização	Abundância de nominalizações
Menor densidade lexical	Maior densidade lexical

Fonte: Fonte: Koch (2005, p. 78) apud Batista-Santos; Batista Pereira (2017, p. 171)

A afirmação de que a escrita não é uma mera representação da fala destaca a autonomia e complexidade inerentes à linguagem escrita. A escrita não apenas preserva informações, mas também oferece possibilidades únicas de expressão, organização e reflexão. Ela possui suas próprias regras e nuances, contribuindo para a diversidade e riqueza da comunicação humana.

Assim, o entendimento proposto evita dicotomias simplistas entre oralidade e escrita, reconhecendo que ambas desempenham papéis distintos e valiosos na experiência humana. Essa abordagem enriquece a compreensão da complexidade da linguagem, destacando que a fala e a escrita são expressões complementares e igualmente essenciais da capacidade comunicativa do ser humano. Analisemos a figura a seguir:

A fala e a escrita são duas formas distintas de comunicação, cada uma com suas características e nuances específicas. Aqui estão algumas das principais diferenças entre fala e escrita:

1. Tempo de Processamento:

- Fala: Normalmente, a fala é mais rápida no processo de produção e compreensão. As informações podem ser transmitidas mais rapidamente por meio da fala devido à velocidade natural da linguagem falada.

- Escrita: A escrita geralmente é mais lenta, pois envolve a formulação cuidadosa de palavras e frases. A leitura também pode ser um processo mais demorado do que a audição.

2. Formalidade:

- Fala: A linguagem falada tende a ser mais informal, com o uso de gírias, contrações e entonação expressiva. A comunicação oral muitas vezes é adaptada ao contexto e à audiência.

- Escrita: A escrita geralmente segue regras mais formais e é estruturada de maneira mais cuidadosa. A linguagem escrita pode ser mais padronizada e menos flexível em termos de expressão informal.

3. Registros:

- Fala: Pode ser mais efêmera, já que as palavras não ficam registradas de forma permanente, a menos que sejam gravadas.

- Escrita: Permanece como um registro fixo do que foi comunicado. Textos escritos podem ser revisados, armazenados e referenciados ao longo do tempo.

4. Recursos Não Verbais:

- Fala: Inclui entonação, ritmo, expressões faciais e gestos, o que pode adicionar camadas adicionais de significado e nuance.

- Escrita: Dependendo do contexto (como em mensagens de texto ou e-mails), pode haver emojis ou outros recursos para tentar transmitir nuances emocionais.

5. Público-Alvo:

- Fala: Pode ser mais adaptada ao público imediato, aproveitando a capacidade de feedback em tempo real para ajustar a comunicação conforme necessário.

- Escrita: Geralmente é dirigida a um público mais amplo e pode ser revisada para atender a diferentes audiências.

6. Correções e Edições:

- Fala: Erros de fala podem ocorrer e serem corrigidos no momento, mas não há uma oportunidade de editar ou revisar as palavras depois de pronunciadas.

- Escrita: Permite revisões e edições antes da comunicação ser compartilhada, resultando em uma expressão mais cuidadosa e precisa.

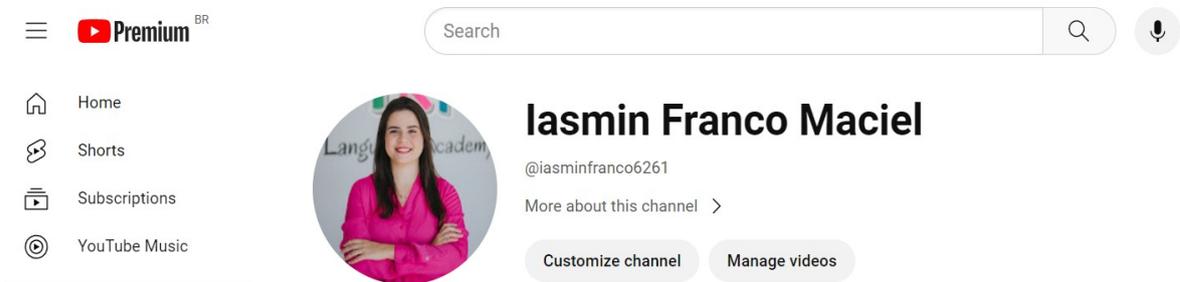
Ambas a fala e a escrita são formas valiosas de comunicação, cada uma com suas próprias vantagens e limitações. A escolha entre fala e escrita muitas vezes depende do contexto, do propósito da comunicação e da preferência do comunicador.

4 MATERIAL

Utilizamos quatro materiais para a realização satisfatória da presente pesquisa, a saber: *YouTube*, *Ytscribe*, *Word Smith Tools* e o *cópus da pesquisa*.

Os *screen shot* de cada página inicial das fontes dos três primeiros materiais citados pode ser visualizado a partir das figuras 4, 5 e 6, respectivamente, a saber:

Figura 4 - *Screen shot* da página inicial do *YouTube*



Fonte: YouTube (2023)

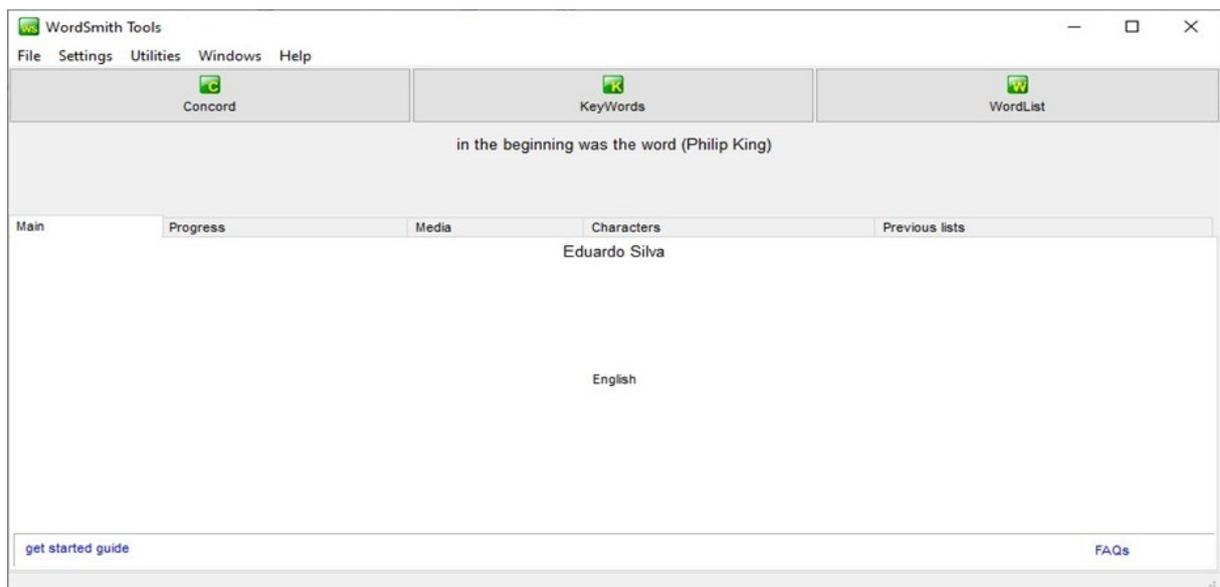
Figura 5 – *Screen shot* da página inicial do *Ytscribe*

YouTube Scribe

The best way to get free transcripts of any YouTube Video

Fonte: YouTube Scribe (2023)

Figura 6 – *Screen shot* da página inicial do *Word Smith Tools*.



Fonte: Scott (2021)

Os materiais expostos nas imagens anteriores e que foram utilizados para a presente pesquisa serão descritos com alto nível de detalhamento nos sub títulos nas páginas subsequentes.

4.1 *YouTube*

O YouTube é uma plataforma online de compartilhamento de vídeos, criada em 2005 e posteriormente adquirida pelo Google, que se tornou uma das maiores e mais influentes plataformas de vídeo do mundo. Ela oferece uma vasta gama de conteúdo em diversas categorias, como entretenimento, música, educação, notícias, tutoriais, jogos e vlogs.

Qualquer pessoa pode criar um canal no YouTube e fazer upload de vídeos, o que resultou numa comunidade diversificada de criadores de conteúdo, desde indivíduos até empresas e organizações. A plataforma oferece a possibilidade de monetização através de anúncios exibidos nos vídeos, inscrições pagas, doações e *merchandising*.

O YouTube utiliza um algoritmo de recomendação que sugere vídeos com base nas preferências do usuário, histórico de visualizações e comportamento na plataforma, proporcionando uma experiência personalizada. Os espectadores podem interagir com os vídeos por meio de curtidas, comentários e compartilhamentos, além de se inscreverem nos canais para receberem atualizações.

Além disso, o YouTube oferece o serviço premium, YouTube Premium, que remove anúncios, permite o download de vídeos para visualização offline e oferece acesso ao YouTube Music, entre outros benefícios. A plataforma também permite a transmissão ao vivo de eventos, shows, jogos e podcasts.

O YouTube se tornou uma parte integral da cultura online, proporcionando entretenimento, informação e educação para milhões de pessoas ao redor do mundo, transformando a maneira como o conteúdo audiovisual é consumido na internet.

4.2 YouTube Scribe

O YouTube Scribe, uma extensão gratuita do Google, representa uma ferramenta potencialmente valiosa para usuários que buscam transcrições de vídeos no YouTube. Essa extensão pode simplificar o processo de transformar conteúdo de áudio em texto, proporcionando uma experiência mais acessível e inclusiva para aqueles que preferem ou necessitam de informações textuais. Ao integrar-se diretamente ao navegador, essa extensão pode oferecer uma maneira intuitiva e eficiente de obter transcrições precisas e confiáveis, aprimorando a acessibilidade e utilidade dos vídeos na plataforma.

A funcionalidade do YouTube Scribe abrange uma ampla gama de benefícios, incluindo a capacidade de personalizar o formato da transcrição, destacar trechos específicos ou mesmo traduzir automaticamente o texto para diversos idiomas. Esses recursos adicionais tornam a extensão uma ferramenta versátil para usuários com diferentes necessidades e preferências linguísticas.

4.3 Word Smith Tools

O WordSmith Tools é um conjunto de programas de software amplamente utilizado no campo da linguística computacional e análise de textos. Desenvolvido por Mike Scott, o *software* foi lançado pela primeira vez em 1996. Ao longo dos anos, passou por atualizações e novas versões foram lançadas para aprimorar suas funcionalidades e oferecer recursos mais avançados para análise de texto e córpus linguísticos.

O WordSmith Tools oferece várias ferramentas para análise de córpus, incluindo:

1. Concordancer: permite a pesquisa de palavras em um corpus, exibindo contextos em que essas palavras ocorrem, ajudando na análise de uso e frequência.

2. KeyWords: ajuda na identificação e análise de palavras-chave ou termos frequentes em um corpus, o que pode ser útil para estudos de frequência de palavras.

3. Collocates: facilita a identificação de palavras que ocorrem frequentemente em conjunto com uma palavra específica, ajudando a entender associações linguísticas.

4. Clusters: agrupa palavras com base em sua coocorrência frequente no texto, mostrando padrões e associações semânticas.

5. WordList Tool: gera listas de palavras únicas em um corpus de texto, acompanhadas de informações estatísticas como frequência de ocorrência e distribuição.

6. File Tool: permite a leitura e análise de arquivos de texto em diferentes formatos, facilitando o processamento de grandes conjuntos de dados textuais.

7. Lexical Diversity Tool: calcula medidas de diversidade lexical, como o índice de diversidade lexical de Herdan ou o índice de honra de Honore, para avaliar a riqueza vocabular de um texto.

8. Word Frequency Tool: analisa a frequência de ocorrência de palavras em um corpus de texto, ajudando a identificar padrões e termos mais comuns.

9. N-grams Tool: identifica sequências de palavras consecutivas de comprimento específico (n-gramas), úteis para análise de estilo e padrões linguísticos.

Essas são apenas algumas das ferramentas disponíveis no Word Smith Tools 5.0. Cada uma delas pode ser útil para diferentes tipos de análise textual, desde estudos de linguística até análises de conteúdo em áreas como pesquisa acadêmica, marketing ou ciências sociais.

4.4 Córpus

Para iniciar nossa apresentação, partimos da definição do termo *cópus*, que se refere a uma “coletânea de exemplos naturais de linguagem, que consistem desde algumas frases até conjuntos de textos escritos ou gravações orais que foram coletados para serem usados como base para pesquisa linguística” (HUNSTON, 2002, p. 2 apud SARMENTO, 2010, p. 89-90), ou como “um conjunto de textos escritos ou falados numa língua, disponível para análise”. (TRASK, 2004 apud ALUISIO; ALMEIDA, 2006, p. 157). A amplitude da definição de “*cópus*”, que abarca desde pequenos trechos até coleções abrangentes de textos e gravações, reflete a diversidade intrínseca desse recurso. Egbert (2022, p. 7) define *cópus* como “uma amostra grande e criteriosa de textos projetada para representar um domínio alvo de uso da linguagem (por exemplo, uma língua, dialeto ou registro)”.¹⁴ A coletânea de exemplos autênticos de linguagem, conforme delineado por Sarmiento (2010), ressalta a importância de capturar a linguagem em seu contexto genuíno, permitindo uma compreensão mais profunda dos padrões linguísticos. A mencionada coletânea não apenas serve como uma ferramenta vital para a pesquisa linguística, mas também representa uma fonte valiosa para explorar nuances e variações na expressão linguística. Além disso, a noção de *cópus* como um conjunto disponível para análise, conforme enfatizado por Trask (2004), destaca sua utilidade como uma base de dados linguísticos acessível para pesquisadores, professores e estudantes, contribuindo assim para avanços significativos na compreensão da linguagem em suas diversas manifestações. Essa abordagem abrangente destaca não apenas a riqueza de informações linguísticas contidas nos *cópus*, mas também a sua relevância como uma ferramenta fundamental para a pesquisa e a análise linguística.

Neste trabalho, buscamos contemplar todos os elementos citados por Viana (2011, p. 24), quando cita que um *cópus* deve:

- (a) ser compreendido como um conjunto de textos;
- (b) contempla textos (orais ou escritos) que tenham sido efetivamente produzidos por falantes de determinada língua;
- (c) consiste numa forma de representar empiricamente o uso que se faz de uma língua em seu sentido geral ou específico;
- (d) é uma reprodução da produção linguística de toda a população que se quer investigar ou uma amostra representativa dessa população, com base em princípios claros e bem definidos;
- (e) assume a forma eletrônica com vistas a ser investigado pelo computador;
- (f) é concebido com o objetivo de possibilitar a realização de uma pesquisa Linguística

¹⁴ A large and principled sample of texts designed to represent a target domain of language use (e.g., a language, dialect, or register).

Analisando os tópicos separadamente, explicaremos como contemplamos cada um deles com nossa pesquisa:

- a) deve ser compreendido como um conjunto de textos: nosso cópuz é um compilado de vídeos infantis, como já mencionado anteriormente, ou seja, construímos um conjunto de textos orais;
- b) contempla textos (orais ou escritos) que tenham sido efetivamente produzidos por falantes de determinada língua: os sub-registros utilizados foram produzidos por falantes de língua inglesa;
- c) consiste numa forma de representar empiricamente o uso que se faz de uma língua em seu sentido geral ou específico: a língua inglesa voltada para crianças;
- d) é uma reprodução da produção linguística de toda a população que se quer investigar ou uma amostra representativa dessa população, com base em princípios claros e bem definidos: crianças não alfabetizadas;
- e) assume a forma eletrônica com vistas a ser investigado pelo computador: utilizamos plataformas e *softwares* em computador;
- f) é concebido com o objetivo de possibilitar a realização de uma pesquisa Linguística: a presente pesquisa.

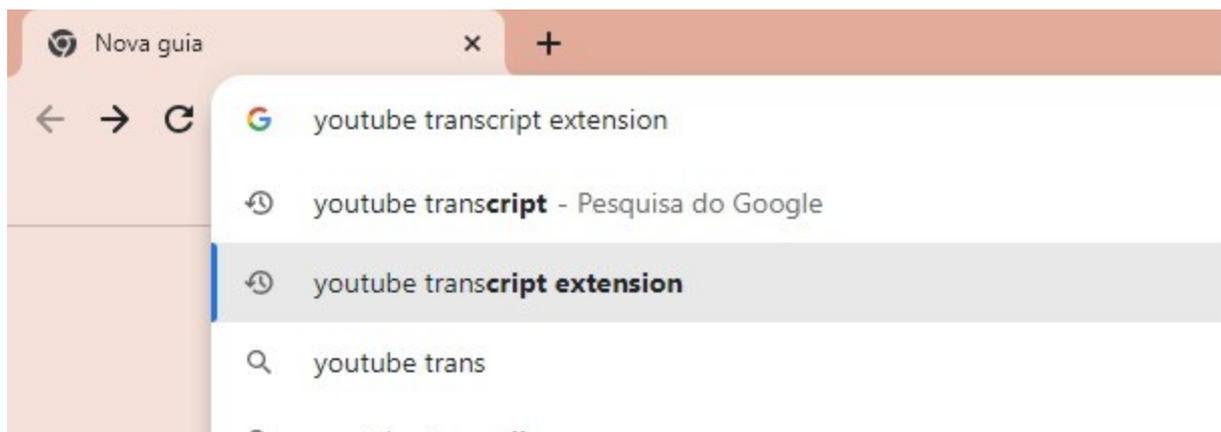
Os tópicos mencionados por Viana (2011), relacionados aos critérios que definem o que deve compor um cópuz, foram cuidadosamente considerados pela autora ao conceber sua pesquisa. Isto inclui aspectos como representatividade dos dados, volume suficiente para análise estatística, diversidade de fontes textuais, boa organização dos dados, entre outros requisitos fundamentais para que um conjunto de textos seja reconhecido como cópuz na área da Linguística de Cópuz.

Ao contemplar todos esses elementos em sua pesquisa, a autora demonstra um cuidado metodológico e rigor acadêmico, assegurando que o cópuz desenvolvido ou utilizado atenda aos critérios estabelecidos para análise linguística, garantindo assim a validade e a solidez dos resultados obtidos em sua investigação.

5 MÉTODO

Para iniciarmos o processo, a primeira etapa consistiu em acessar a internet pelo *Google Chrome*, digitamos “*YouTube transcript extension*” para obter a extensão do *ytscribe* no *Chrome*, conforme figura a seguir:

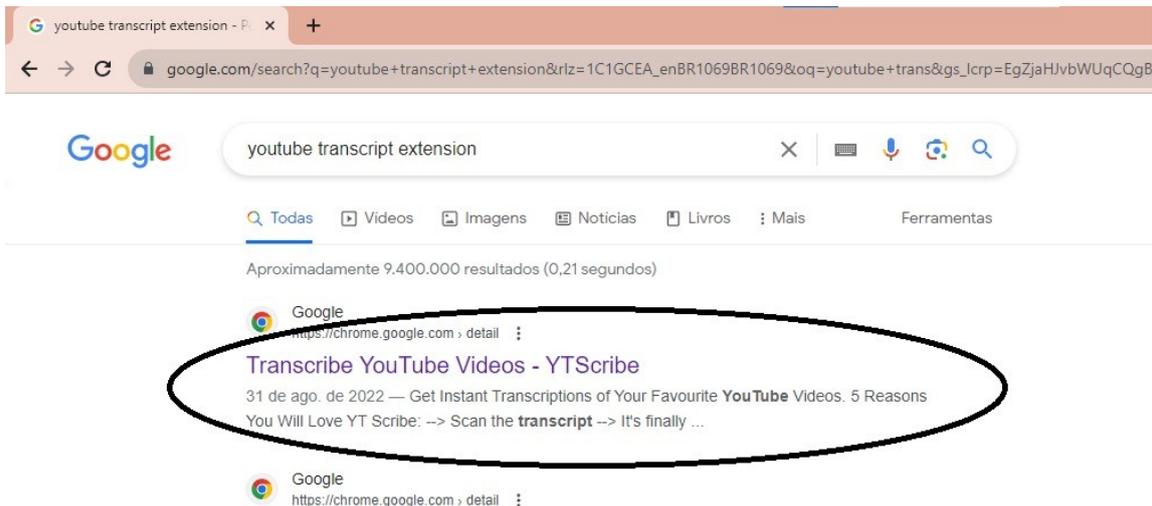
Figura 7 - *YouTube transcript extension* – passo 1



Fonte: *YouTube* (2023)

No passo número 2, selecionamos e clicamos no primeiro *link* disponível na página de pesquisas, conforme figura a seguir:

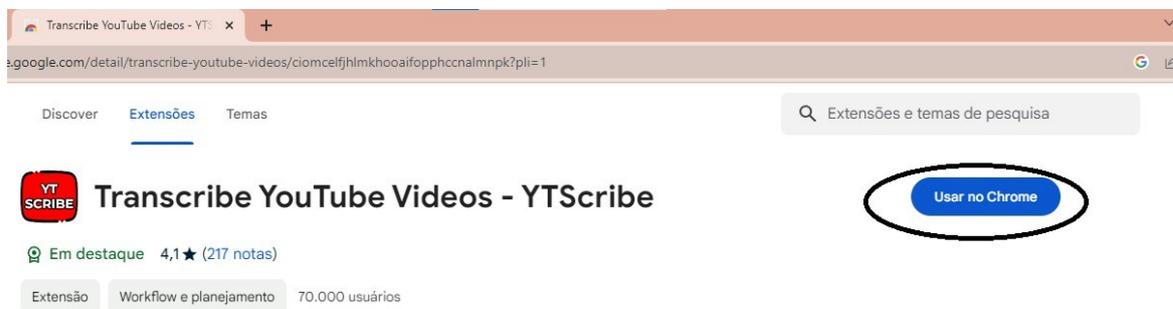
Figura 8 - *YouTube transcript extension* – passo 2



Fonte: *YouTube* (2023)

Após esse procedimento, clicamos em “Usar no *Chrome*” para que pudéssemos ter acesso à plataforma diretamente do *YouTube*, como mostrado na figura abaixo:

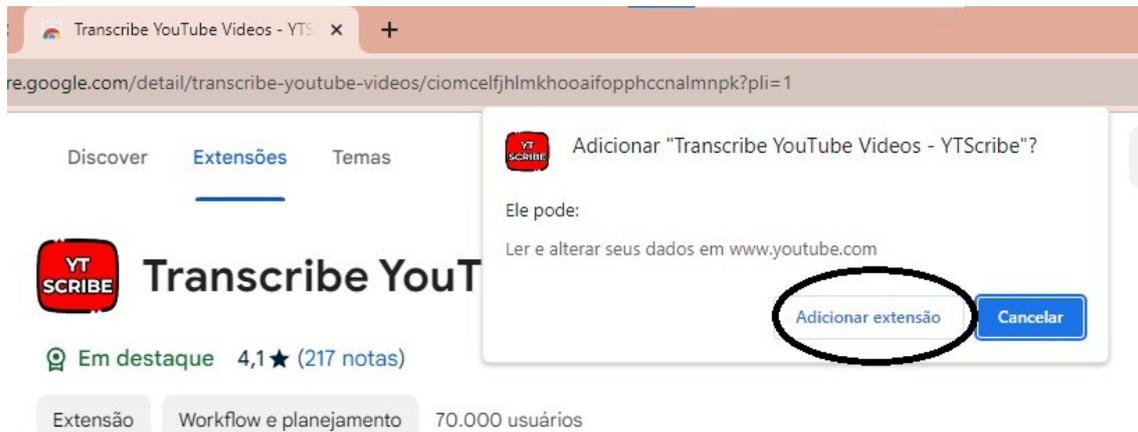
Figura 9 - *YouTube transcript extension* – passo 3



Fonte: *YouTube* (2023)

Logo a seguir, confirmamos a adição da extensão clicando na caixa de texto “adicionar extensão” para que pudéssemos ter acesso à plataforma diretamente do *YouTube*, a saber:

Figura 10 - *YouTube transcript extension* – passo 4



Fonte: *YouTube* (2023)

Após esse passo a passo, tivemos confirmação da inclusão da extensão do *YouTube* Scribe ao *Chrome*, como indica a figura 11 abaixo:

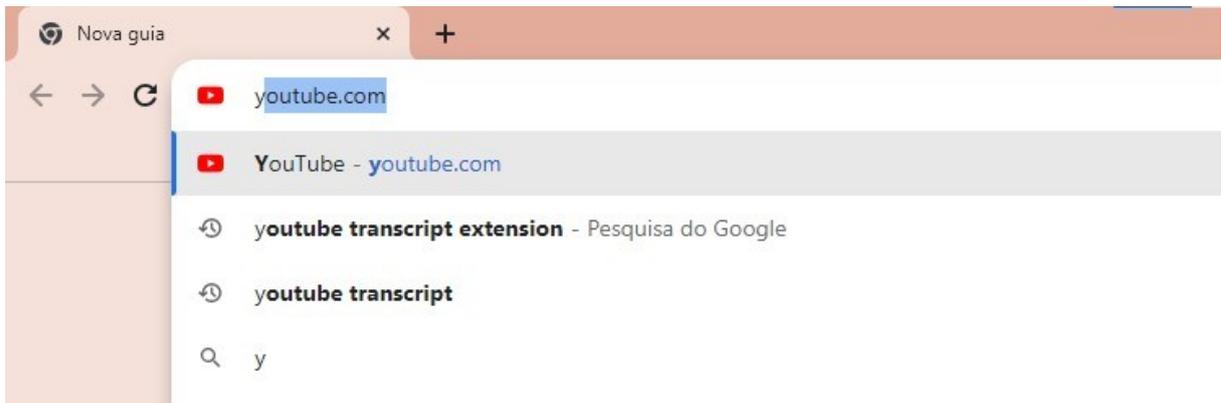
Figura 11 - *YouTube transcript extension* – passo 5



Fonte: *YouTube* (2023)

Em seguida, digitamos “*youtube.com*” na barra de pesquisa para prosseguir, como apresentado na figura 12 abaixo:

Figura 12 – seleção dos vídeos – passo 1

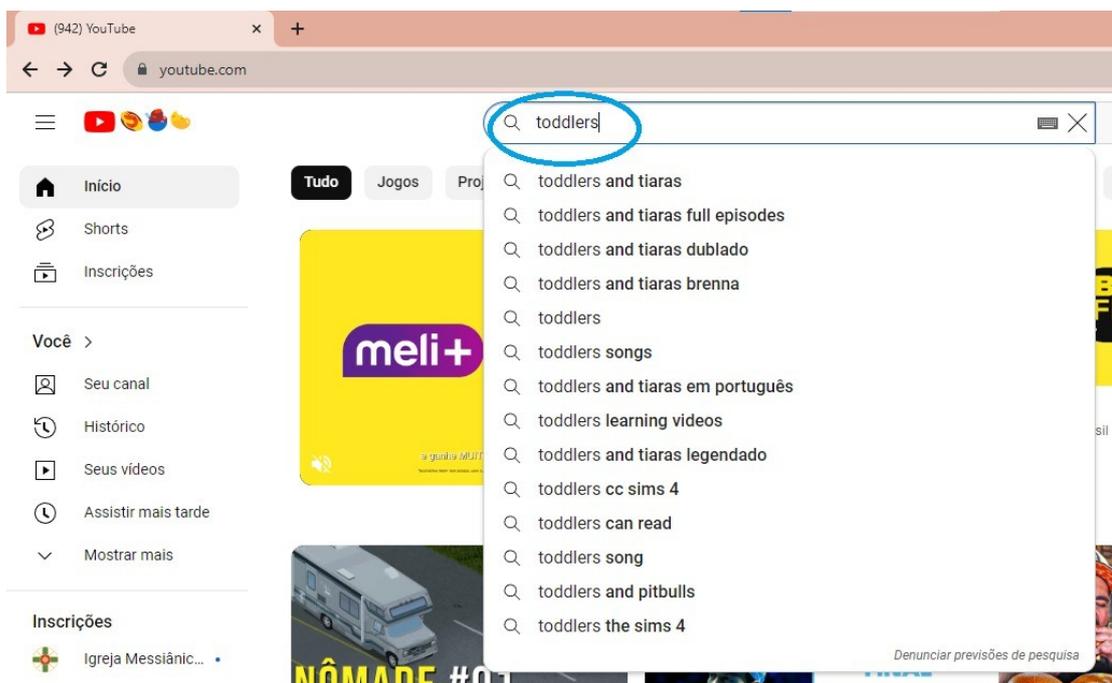


Fonte: *YouTube* (2023)

Palavras sementes

Ao acessar o site em questão, pesquisamos por palavras chaves que nos levariam a vídeos de conteúdo infantil, como por exemplo: inglês para crianças, inglês infantil, histórias infantis em inglês, música infantil em inglês, filmes infantis em inglês, músicas infantis em inglês ou digitamos em inglês: *"videos for toddlers"*, *"english for kids"*, *"english stories"*, entre outras. Como apresentado na figura 13 abaixo:

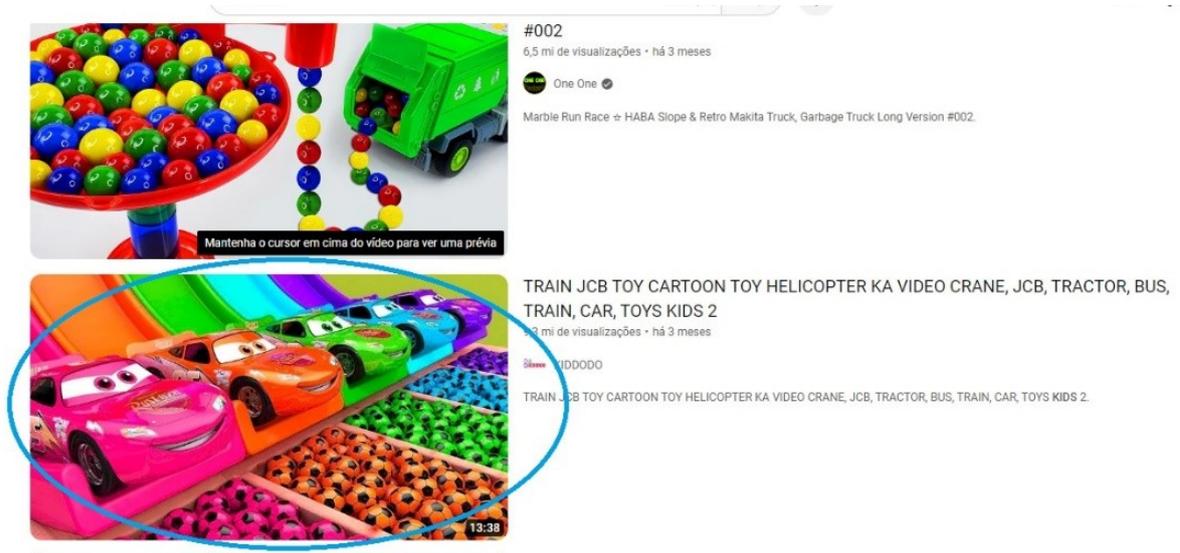
Figura 13 – seleção dos vídeos – passo 2



Fonte: *YouTube* (2023)

Posteriormente, um vídeo de interesse era selecionado, independentemente de seu comprimento ou duração.

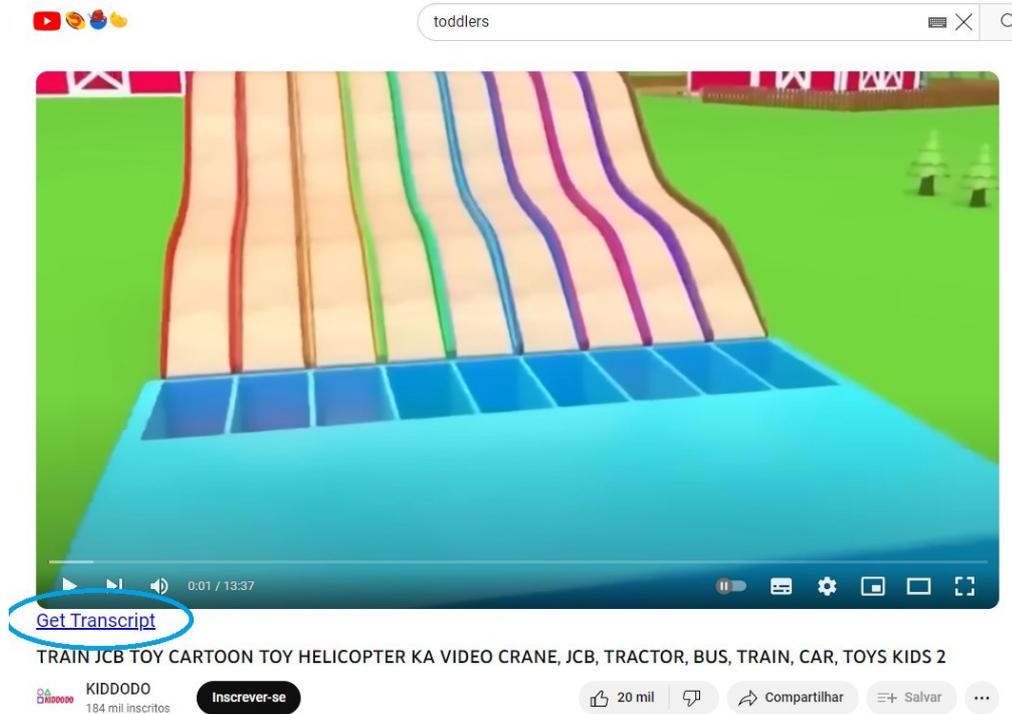
Figura 14 – seleção dos vídeos – passo 3



Fonte: *YouTube* (2023)

Após selecionar o vídeo, procedemos clicando sobre o mesmo e, então, selecionamos a opção "*get transcript*" localizada logo acima do título, conforme ilustrado na figura 15:

Figura 15 – seleção dos vídeos – passo 4



Fonte: *YouTube* (2023)

Ao clicar em "*get transcript*", automaticamente foi aberta outra plataforma, o "*YouTube Scribe*". Clicamos em "*download txt file*". Vale lembrar que os arquivos precisam estar salvos em txt para posteriormente serem lançados no software "*Word Smith Tools*".

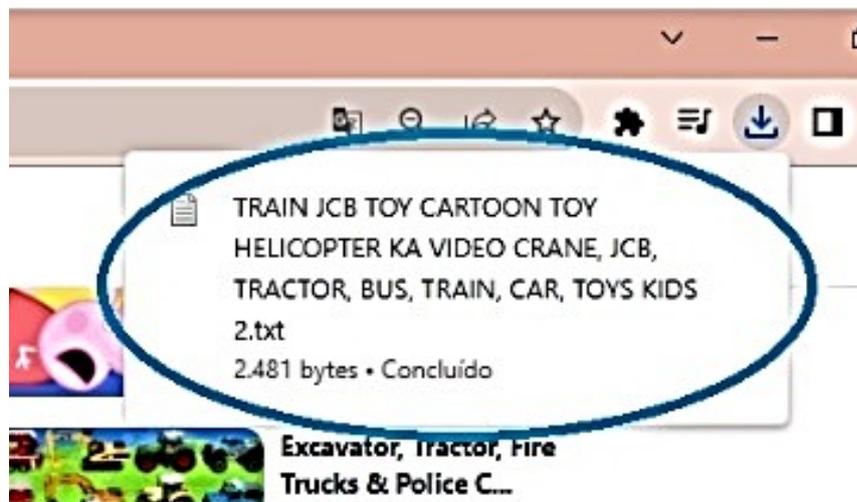
Figura 16 – seleção dos vídeos – passo 5



Fonte: *YouTube Scribe* (2023)

Após esse procedimento, o documento no formato TXT será salvo automaticamente no dispositivo utilizado para acessar a plataforma *YouTube Scribe*, como podemos ver na figura a seguir:

Figura 17 – seleção dos vídeos – passo 6



Fonte: *YouTube Scribe* (2023)

E então acontecerá a transcrição do vídeo selecionado:

Figura 18 – seleção dos vídeos – passo 7

TRAIN JCB TOY CARTOON TOY HELICOPTER KA VIDEO CRANE, JCB, TRACTOR, BUS, TRAIN, CAR, TOYS KIDS 2 - Bloco de Notas
 Arquivo Editar Formatar Exibir Ajuda
 Now it's amazing [Music] what's your name [Music] Wow Let's play together [Applause] I slide first are you okay let's go oh [Applause] my gosh oh we can do it
 I'm so funny what's happening it's my turn oh who's next [Applause] it's water tank truck yes where are you going the world is colorful [Music] what is it surprise
 I got a bowl oh [Music] I like it what's it in hi there very nice watch next [Music] awesome you got a ball that's right [Music] I'm curious it's exciting I'm so happy
 There are many cars here one two [Music] three four five six seven eight nine ten eleven [Music] look at that [Music] wow it's nice slide
 [Music] yum yummy I go first go Yahoo [Music] that's amazing thank you I can do it wow I'm so funny what's next is it my turn I'm ready yeah good job
 [Music] follow me [Music] hey home [Music] Full Speed Ahead [Music] I got it fantastic who's in next Hill it is [Music] yeah let's go

Fonte: *YouTube* Scribe (2023)

Após a coleta e arquivamento dos 1.940 vídeos voltados para crianças, fizemos *download* do software “*Word Smith Tools*”. Após o término do *download*, o acesso ao software exibe a seguinte tela inicial, e após acessar a imagem inicial do software, o próximo passo consistiu em clicar na letra “W” maiúscula, que corresponde à opção “*Word list*”

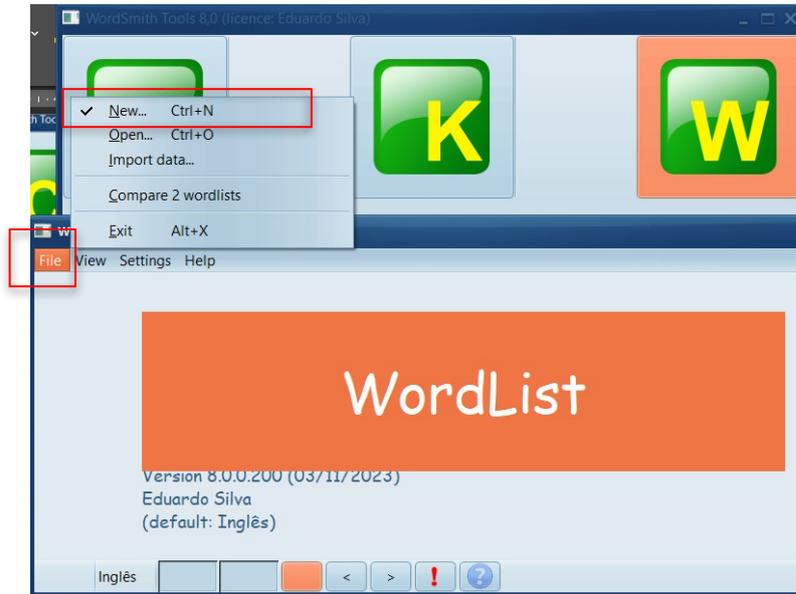
Figura 19 – *Word Smith Tools* – passo 1



Fonte: *Scott* (2021)

Então, selecionamos “*File*” e depois, “*New...*” para iniciarmos uma análise nova, conforme apresentado a seguir:

Figura 20 – *Word Smith Tools* – passo 2



Fonte: Scott (2021)

Então, clicamos em “choose texts now” para a seleção dos arquivos em TXT.

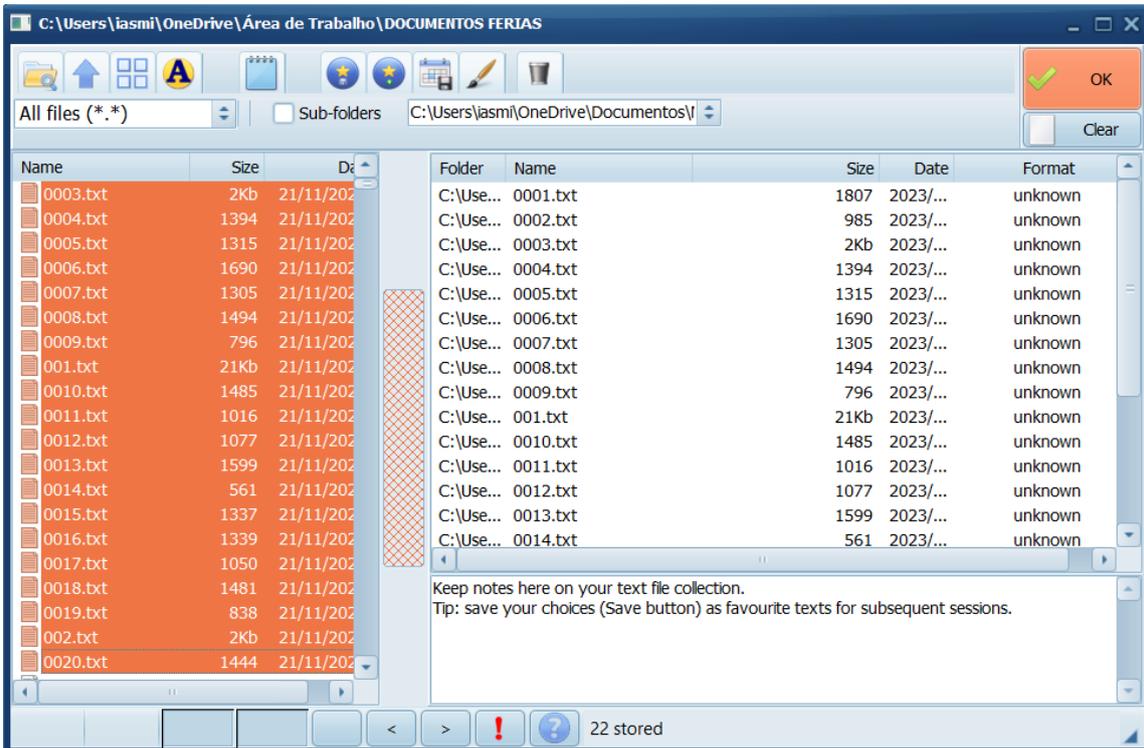
Figura 21 – *Word Smith Tools* – passo 3



Fonte: Scott (2021)

Selecionamos os arquivos para análise, conforme figura 22 abaixo:

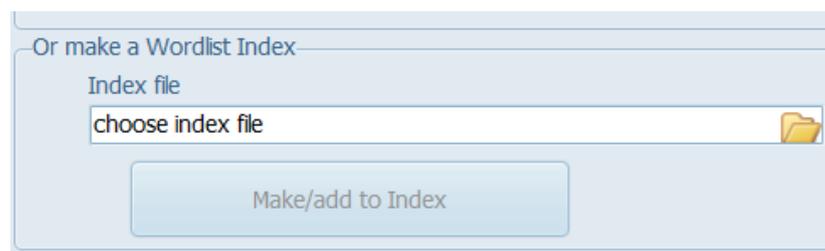
Figura 22 – *Word Smith Tools* – passo 4



Fonte: Scott (2021)

Clicamos em “choose index file” e foi gerado um documento em excel com a lista estatística contendo informações como: a) tamanho do arquivo; b) tipos de palavras diferentes; c) comprimento das palavras; d) sentenças; e) significados; f) parágrafos; g) palavras de 1 letra; h) palavras de 2 letras; i) palavras de 3 letras; entre outras várias informações. Também foi gerado um outro documento *excel* apenas com os *lexical bundles* presentes no *corpus*, sua frequência e em quantos textos diferentes estava presente cada *lexical bundle* em questão.

Figura 23 – Word Smith Tools – passo 5



Fonte: Scott (2021)

Utilizando o segundo arquivo gerado é que foi possível analisar os 86 *lexical bundles* que discurremos ao longo deste trabalho.

6 RESULTADOS E ANÁLISE DOS DADOS

A seguir, avançaremos para a apresentação dos resultados obtidos por meio da análise dos dados coletados. Nesta seção, examinaremos os padrões identificados nos *lexical bundles* para crianças não alfabetizadas, fornecendo uma análise detalhada das estruturas linguísticas recorrentes encontradas em diferentes sub-registros. Além disso, exploraremos as implicações desses resultados, destacando a relevância dos *lexical bundles* na comunicação infantil e discutindo seu potencial impacto no ensino de inglês para crianças em idade pré-alfabética. Vamos agora mergulhar nos dados e nas conclusões que emergem dessa análise. Abaixo apresentaremos os 60 primeiros *lexical bundles* específicos mais frequentes selecionados por meio do *software Word Smith Tools*:

Quadro 5 - *Lexical bundles* específicos

1. <i>What are you</i>	2. <i>Go to the</i>	3. <i>What do you</i>	4. <i>One two three</i>	5. <i>Where are you</i>
6. <i>We have to</i>	7. <i>Are you doing</i>	8. <i>This is the</i>	9. <i>What are you doing</i>	10. <i>Upon a time</i>
11. <i>Once upon a time</i>	12. <i>Once upon a</i>	13. <i>Do you want</i>	14. <i>Out of the</i>	15. <i>Do you think</i>
16. <i>Do you know</i>	17. <i>Take care of</i>	18. <i>So much fun</i>	19. <i>We need to</i>	20. <i>Be able to</i>
21. <i>Don't want to</i>	22. <i>It's time for</i>	23. <i>There was a</i>	24. <i>I'm going to</i>	25. <i>Here we go</i>
26. <i>Two three four</i>	27. <i>One two three four</i>	28. <i>It's time to</i>	29. <i>This is a</i>	30. <i>Why are you</i>
31. <i>Are you ready</i>	32. <i>Have an idea</i>	33. <i>What to do</i>	34. <i>For your favorite</i>	35. <i>Do you want to</i>
36. <i>Thank you for</i>	37. <i>In the morning</i>	38. <i>Get out of</i>	39. <i>Oh my god</i>	40. <i>How do you</i>
41. <i>Went to the</i>	42. <i>Up and down</i>	43. <i>Want to play</i>	44. <i>Back to the</i>	45. <i>One two three four five</i>

- | | | | | |
|-----------------------------------|------------------------|--------------------------|-------------------------|------------------------------|
| 46. <i>How are you</i> | 47. <i>In front of</i> | 48. <i>Why don't you</i> | 49. <i>Have to go</i> | 50. <i>Upon a time there</i> |
| 51. <i>Once upon a time there</i> | 52. <i>All the way</i> | 53. <i>Want to be</i> | 54. <i>Let's go to</i> | 55. <i>As soon as</i> |
| 56. <i>In the sky</i> | 57. <i>Come on</i> | 58. <i>Are you going</i> | 59. <i>Go to school</i> | 60. <i>This is my</i> |

Fonte: a autora

Agora, apresentaremos os 60 primeiros *lexical bundles* em comum mais frequentes selecionados por meio do *software Word Smith Tools*:

Quadro 6 – *Lexical bundles* em comum

1. <i>What are you</i>	2. <i>Go to the</i>	3. <i>What do you</i>	4. <i>One two three</i>	5. <i>Where are you</i>
6. <i>We have to</i>	7. <i>Are you doing</i>	8. <i>This is the</i>	9. <i>What are you doing</i>	10. <i>Upon a time</i>
11. <i>Once upon a time</i>	12. <i>Once upon a</i>	13. <i>Do you want</i>	14. <i>Out of the</i>	15. <i>Do you think</i>
16. <i>Do you know</i>	17. <i>Take care of</i>	18. <i>So much fun</i>	19. <i>We need to</i>	20. <i>Be able to</i>
21. <i>Don't want to</i>	22. <i>It's time for</i>	23. <i>There was a</i>	24. <i>I'm going to</i>	25. <i>Here we go</i>
26. <i>Two three four</i>	27. <i>One two three four</i>	28. <i>It's time to</i>	29. <i>This is a</i>	30. <i>Why are you</i>
31. <i>Are you ready</i>	32. <i>Have an idea</i>	33. <i>What to do</i>	34. <i>For your favorite</i>	35. <i>Do you want to</i>
36. <i>Thank you for</i>	37. <i>In the morning</i>	38. <i>Get out of</i>	39. <i>Oh my god</i>	40. <i>How do you</i>
41. <i>Went to</i>	42. <i>Up and</i>	43. <i>Want to</i>	44. <i>Back to</i>	45. <i>One two</i>

<i>the</i>	<i>down</i>	<i>play</i>	<i>the</i>	<i>three four five</i>
46. <i>How are you</i>	47. <i>In front of</i>	48. <i>Why don't you</i>	49. <i>Have to go</i>	50. <i>Upon a time there</i>
51. <i>Once upon a time there</i>	52. <i>All the way</i>	53. <i>Want to be</i>	54. <i>Let's go to</i>	55. <i>As soon as</i>
56. <i>In the sky</i>	57. <i>Come on</i>	58. <i>Are you going</i>	59. <i>Go to school</i>	60. <i>This is my</i>

Fonte: a autora

Conforme citado no Quadro 4, a definição de lexical bundles por Conrad e Biber (2004, p. 59, grifo nosso): “Definimos pacotes lexicais como as sequências lexicais fixas mais **frequentemente recorrentes** em um registro. Quanto mais **comum** um pacote lexical, mais **útil** ele parece ser na construção do discurso, mas exatamente onde definir um limite de frequência é algo um tanto arbitrário”, os autores enfatizam que quanto mais comum um pacote lexical, mais útil ele parece ser na construção do discurso, ressaltando, porém, que definir um limite de frequência para esses pacotes é algo um tanto arbitrário.

Analisando os resultados apresentados nos quadros 5 e 6 de nosso estudo, observamos que os 60 lexical bundles listados são os mesmos. Isso evidencia uma congruência notável com a citação dos autores. As palavras-chave "frequentemente recorrentes", "quanto mais comum" e "mais útil" mencionadas na definição de Conrad e Biber estão em perfeita sintonia com os resultados encontrados em nossa pesquisa.

A constatação de que os mesmos lexical bundles foram identificados como sendo frequentes, comuns e úteis corrobora diretamente com a definição dos autores. Isso ressalta a validade e a relevância desses padrões linguísticos identificados em nosso estudo, reforçando a importância desses lexical bundles na construção e compreensão do discurso direcionado a crianças não alfabetizadas. Essa congruência entre a definição teórica e os resultados empíricos fortalece ainda mais a fundamentação de nosso trabalho e sua contribuição para o entendimento e o ensino da linguagem dirigida a esse público-alvo.

A escolha do foco em registros orais foi motivada pela consideração cuidadosa do público-alvo: crianças não alfabetizadas. Reconhecendo que essas crianças ainda estão em

processo de desenvolvimento da leitura e escrita, optamos por explorar o registro oral como uma janela valiosa para compreender a linguagem que permeia seu ambiente comunicativo.

Entende-se como registro: “uma variedade de textos associados a características situacionais, linguísticas e funcionais específicas.”¹⁵ (EGBERT et al., 2022, p. 222).

Essa definição de registro fornecida por Egbert, Jesse, et al. destaca a compreensão do termo no contexto linguístico. Eles explicam que o registro refere-se a uma diversidade de textos que estão conectados a características específicas de situação, linguagem e função.

Esta explicação ressalta que o conceito de registro não se restringe apenas aos textos em si, mas também considera o contexto em que são utilizados. Isso inclui a situação em que os textos são produzidos ou circulam, as características linguísticas distintivas presentes nesse contexto e a função comunicativa ou propósito que esses textos atendem.

Ao delinear esses elementos, essa definição de registro amplia a compreensão sobre a diversidade textual, destacando a importância de considerar não apenas o texto em si, mas também o ambiente e o propósito comunicativo ao analisar e interpretar diferentes registros linguísticos.

É importante considerar que dentro do conceito de registro linguístico, existem não apenas registros gerais, mas também sub-registros específicos, cada um com suas próprias características distintas. No contexto deste trabalho, visando obter uma compreensão mais aprofundada e contextualizada, o registro principal é o oral, e dentro desse registro, estão os sub-registros representados por vídeos do *YouTube*, literatura infantil, filmes e músicas voltadas para crianças, bem como vídeos educativos.

Esses sub-registros do registro oral foram escolhidos estrategicamente para capturar as nuances da linguagem oral utilizada em diferentes contextos e abrangem uma variedade de contextos linguísticos que refletem diferentes situações, estilos de linguagem e finalidades comunicativas dentro do âmbito da linguagem oral. Cada um desses sub-registros possui características particulares que influenciam a forma como a linguagem é usada, estruturada e compreendida, seja na comunicação informal, como nos vídeos do *YouTube* e filmes, ou na educação formal, como na literatura infantil e vídeos educativos.

Ao considerar esses sub-registros dentro do registro oral, é possível compreender e analisar de forma mais detalhada as nuances linguísticas e os padrões de uso da linguagem presentes em cada contexto específico, contribuindo assim para uma compreensão mais ampla e aprofundada dos *lexical bundles* nesses diferentes cenários de uso da língua.

15 A variety of texts associated with particular situational, linguistic, and functional characteristics.

A seguir, encontraremos um quadro comparativo entre 10 *lexical bundles* identificados por Biber, Conrad e Cortes (2003), Biber (2009) e aqueles que emergem a partir deste estudo. Ao confrontar os padrões identificados por Biber, Conrad e Cortes (2003) e Biber (2009) com os que foram identificados na presente pesquisa, buscaremos destacar semelhanças e disparidades, lançando luz sobre possíveis áreas de convergência ou divergência na linguagem utilizada nos três contextos. A análise dos *lexical bundles* proposta visa não apenas destacar as similaridades e diferenças, mas também oferecer insights sobre como a linguagem evolui ao longo do tempo, refletindo mudanças no discurso e na expressão linguística.

Quadro 7 – Comparação entre a presente pesquisa e Biber, Conrad e Cortes (2003) quanto à *lexical bundles* presentes/ausentes

Biber, Conrad e Cortes (2003)	Presença/Ausência nos registros orais analisados nesta pesquisa
1. "at the same time"	X
2. "in the case of"	X
3. "as a result of"	X
4. "in terms of"	X
5. "on the other hand"	X
6. "in order to"	X
7. "in the context of"	X
8. "in this paper"	X
9. "on the basis of"	X
10. "in the following"	X

Quadro 8 – Comparação de *lexical bundles* presentes/ausentes em Biber, Conrad e Cortes (2003)

Presente pesquisa	Presença/Ausência nos registros orais analisados em Biber, Conrad e Cortes (2003)
1. "so much fun"	X
2. "we need to"	X
3. "don't want to"	✓
4. "are you ready"	X
5. "it's time for"	X
6. "have an idea"	X
7. "went to the"	X
8. "in the morning"	X
9. "get out of"	X
10. "want to play"	X

Fonte: a autora

Quadro 9 – Comparação entre a presente pesquisa e Biber (2009) quanto à *lexical bundles* presentes/ausentes

Biber (2009)	Presença/Ausência nos registros orais analisados nesta pesquisa
1. "at the end of the"	X
2. "what are you doing"	✓

3. "it is important to"	X
4. "as a result of"	X
5. " do you want me "	✓
6. "on the other hand"	X
7. "in the context of"	X
8. "what are you doing"	X
9. "you know what I mean"	X
10. "I don't want to"	X

Fonte: a autora

Quadro 10 – Comparação de *lexical bundles* presentes/ausentes em Biber (2009)

Presente pesquisa	Presença/Ausência nos registros orais analisados em Biber (2009)
1. "I'm going to"	X
2. "here we go"	X
3. "this is a"	X
4. "for your favorite"	X
5. "why are you"	X
6. " how do you "	✓
7. "back to the"	X
8. "what are you"	✓
9. "go to the"	✓

10. "what do you"

✓

Fonte: a autora

O sinal representado por X significa a ausência dos *lexical bundles* em questão no trabalho descrito e ✓ significa a presença do *lexical bundles* em questão no trabalho descrito.

Obs: No processo de elaboração dos quadros que descrevem os *lexical bundles* presentes ou ausentes nos trabalhos, o critério de inclusão foi cuidadosamente considerado, levando em conta não apenas os elementos iniciais ou finais, e sim qualquer parte do *lexical bundle*, com complemento ou sem. Essa abordagem criteriosa assegurou uma representação ampla e diversificada das possibilidades de utilização dos *córpus* no ensino de línguas, englobando uma variedade de contextos e formas de emprego nos processos de aprendizagem linguística.

A pesquisa em questão e as pesquisas de Biber, Conrad e Cortes (2003) e Biber (2009) adotaram abordagens distintas na análise de *lexical bundles*, destacando diferenças na extensão de suas investigações. Enquanto este trabalho se limitou à base do *lexical bundle*, considerando exemplos como "what are you", os outros autores adotaram uma abordagem mais abrangente ao incluir a expansão dessas unidades linguísticas. Eles analisaram construções mais elaboradas, como "what are you doing". Essa diferença no escopo da análise reflete nuances importantes na compreensão das sequências de palavras, proporcionando uma visão mais detalhada das variações e complexidades presentes no uso de *lexical bundles*. Enquanto focamos na forma mais básica do *bundle*, eles exploraram as potenciais extensões dessas sequências, enriquecendo assim a compreensão do fenômeno linguístico em determinados contextos.

Ao analisar os quatro quadros que descrevem a presença ou ausência de *lexical bundles* nas pesquisas apresentadas, observa-se uma distinção marcante. O trabalho atual evidencia a inclusão de *lexical bundles* direcionados para o público infantil, uma área que anteriormente não havia sido estudada. Em contraste, as pesquisas anteriores delineadas nos quadros anteriores enfocam principalmente os *lexical bundles* presentes em registros acadêmicos e textos mais formais. Essa diferenciação indica uma ampliação do escopo de estudo, explorando *lexical bundles* específicos para um público diferente e revelando a variedade de contextos e usos linguísticos que podem ser abordados através da análise de *lexical bundles*.

7 CONSIDERAÇÕES FINAIS

Ao final desta pesquisa, é evidente que alcançamos o objetivo geral de criar uma lista abrangente de *lexical bundles* relevantes e aplicáveis na sala de aula de língua inglesa. O empenho em compreender os padrões linguísticos em diferentes contextos resultou em uma contribuição valiosa para a compreensão da linguagem destinada às crianças não alfabetizadas. O objetivo específico delineado de identificação e análise dos *lexical bundles* presentes na língua falada direcionada para crianças em vídeos do *Youtube* foi plenamente atingido, destacando-se o sucesso alcançado na meta proposta.

A apresentação de uma lista abrangente de *lexical bundles*, derivada da análise de *corp*us de diferentes sub-registros, representa um recurso prático valioso para professores e alunos, enriquecendo o ensino de inglês na educação infantil. Ao fornecer uma base sólida fundamentada em estudos e evidências, em vez de suposições, tanto os professores quanto os alunos se beneficiarão enormemente. Essa abordagem baseada em dados promove uma prática pedagógica mais informada e eficaz, na qual o ensino e a aprendizagem são orientados por princípios empiricamente validados, aumentando assim a qualidade e a eficácia do processo educacional.

Quanto aos problemas de pesquisa, a resposta completa é evidente. Conforme apresentado ao longo do trabalho, sistematizamos uma lista abrangente de *lexical bundles* presentes nos sub-registros e identificamos quais eram comuns entre eles. Notavelmente, observamos que praticamente todos os *lexical bundles* identificados como mais comuns em cada sub-registro separadamente também eram os mesmos quando comparamos os quatro sub-registros. Isso demonstra uma compreensão abrangente e comparativa dos padrões linguísticos presentes em diferentes formas de comunicação destinadas a crianças.

A presente pesquisa se destaca por sua inegável importância para públicos envolvidos no cenário educacional, trazendo contribuições significativas para professores, alunos, pais e interessados pelo assunto. Abaixo, destacamos como cada um desses grupos pode se beneficiar diretamente do trabalho desenvolvido:

1. Professores:

- Aprimoramento pedagógico: ao abordar os principais *lexical bundles* nos vídeos infantis, a pesquisa oferece *insights* valiosos para o aprimoramento das práticas pedagógicas. Professores encontrarão recursos e estratégias inovadoras que poderão ser incorporados em sala de aula, proporcionando um ambiente mais dinâmico e eficaz de aprendizado.

- Desenvolvimento Profissional: as descobertas desta pesquisa também contribuem para o desenvolvimento profissional dos educadores, oferecendo novas perspectivas e ferramentas que podem ser integradas ao seu repertório, enriquecendo sua abordagem no processo de ensino.

- Desenvolvimento de materiais didáticos: com a lista abrangente de *lexical bundles* apresentada, os professores podem criar materiais didáticos mais eficazes e direcionados, integrando esses padrões linguísticos autênticos em atividades de ensino de vocabulário, gramática e comunicação oral, por exemplo, além de diversas dinâmicas e brincadeiras.

- Planejamento de aulas: os *lexical bundles* identificados podem orientar o planejamento de aulas, permitindo que os professores integrem esses padrões linguísticos de forma sistemática ao longo do currículo, garantindo uma progressão natural no desenvolvimento da linguagem dos alunos.

- Aprimoramento da prática pedagógica: ao adotar uma abordagem baseada em evidências e estudos linguísticos, os professores podem aprimorar sua prática pedagógica, garantindo que o ensino da língua inglesa seja fundamentado em princípios empiricamente validados, em vez de depender apenas de intuições ou suposições.

- Promoção da compreensão linguística: o uso dos *lexical bundles* identificados em seu estudo pode ajudar os alunos a desenvolver uma compreensão mais profunda e intuitiva da língua inglesa, facilitando a internalização de padrões linguísticos comuns e promovendo a fluência e a expressão natural.

- Fomento da competência comunicativa: ao incorporar *lexical bundles* autênticos em atividades de ensino, os professores podem ajudar os alunos a melhorar sua competência comunicativa, capacitando-os a se expressarem de forma mais eficaz e fluente em situações de comunicação reais.

Em suma, este trabalho de lexical bundles oferece aos professores uma valiosa ferramenta para aprimorar o ensino de inglês, promovendo uma abordagem mais fundamentada, eficaz e orientada por evidências na sala de aula.

2. Alunos:

- Melhoria da experiência de aprendizagem: Os resultados desta pesquisa impactam diretamente a experiência de aprendizagem dos alunos, proporcionando um ambiente mais estimulante e envolvente, pois ao se tratar de aulas de idioma para crianças, um dos maiores desafios é cativar a atenção dos aprendizes durante todo o período reservado para a aula.

- Estímulo ao interesse acadêmico: Ao explorar os *lexical bundles*, os alunos são expostos a conceitos desafiadores e instigantes, promovendo o desenvolvimento do pensamento crítico e estimulando o interesse por áreas relacionadas ao estudo.

3. Pais:

- Parceria escola-família: A pesquisa fortalece a parceria entre escola (ou professor autônomo) e família, fornecendo aos pais uma compreensão mais profunda dos métodos de ensino e abordagens pedagógicas. Essa transparência contribui para uma participação mais ativa na educação de seus filhos.

- Auxílio na orientação educacional: Os pais encontrarão na pesquisa informações valiosas que os auxiliarão na orientação educacional de seus filhos, possibilitando um suporte mais eficaz no desenvolvimento acadêmico e pessoal.

Em resumo, esta pesquisa não apenas aborda questões relevantes na área educacional, mas também oferece soluções tangíveis e aplicáveis que impactam positivamente a experiência de professores, alunos e pais. O conhecimento gerado contribui para a construção de um ambiente educacional mais rico, colaborativo e voltado para o sucesso acadêmico e pessoal de todos os envolvidos.

Apesar das descobertas esclarecedoras apresentadas, é crucial reconhecer certas limitações nesta pesquisa. Uma limitação importante diz respeito à quantidade exata de cada sub-registro extraído do *YouTube*. Embora tenhamos compilado meticulosamente um corpus composto por 1.940 textos, totalizando 1.722.722 milhões de palavras, a distribuição precisa desses textos em diferentes sub-registros permanece incerta. Devido à natureza dinâmica do conteúdo online, determinar a representação exata de cada contexto linguístico extraído do *YouTube* é desafiador. Assim, enquanto nosso estudo oferece percepções valiosas sobre os padrões de *lexical bundles* na linguagem falada direcionada a crianças no *YouTube*, a falta de dados precisos sobre a distribuição dos sub-registros constitui uma limitação que requer investigação adicional em futuras pesquisas.

A pesquisa deixa margens para o desenvolvimento de atividades didático-pedagógicas cuja criação está prevista para pesquisas de doutoramento da autora. Possível sequência didática, material didático, e metodologia de ensino.

REFERÊNCIAS

ALUÍSIO, Sandra Maria; DE BARCELLOS ALMEIDA, Gladis Maria. O que é e como se constrói um corpus? Lições aprendidas na compilação de vários corpora para pesquisa lingüística. **Calidoscópico**, v. 4, n. 3, p. 156-178, 2006.

BARBOSA, Ana Mae; GALVÃO, Rejane. **Arte educação como mediação cultural e social**. Unesp, 2008.

BATISTA-SANTOS, Dalve Oliveira; BATISTA PEREIRA, Elemária. A influência da oralidade na escrita: reflexões e desafios no ensino de língua materna. **Revista Desafios**, v. 4, n. 2, p. 167-184, 2017.

BERBER SARDINHA, Tony et al. Como usar a Linguística de Corpus no Ensino de Língua Estrangeira—por uma Linguística de Corpus Educacional brasileira. **Corpora no ensino de línguas estrangeiras**. São Paulo: Hub Editorial, p. 293-348, 2010.

BIBER, Douglas et al. Grammar of spoken and written English. **(No Title)**, 1999.

BIBER, Douglas et al. Towards a taxonomy of conversational discourse types: An empirical corpus-based analysis. **Journal of Pragmatics**, v. 171, p. 20-35, 2021.

BIBER, Douglas. A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. **International journal of corpus linguistics**, v. 14, n. 3, p. 275-311, 2009.

BIBER, Douglas; CONRAD, Susan. **Variation in English: Multi-dimensional studies**. Routledge, 2014.

BIBER, Douglas; CONRAD, Susan; CORTES, Viviana. If you look at...: Lexical bundles in university teaching and textbooks. **Applied linguistics**, v. 25, n. 3, p. 371-405, 2004.

BIBER, Douglas; CONRAD, Susan; CORTES, Viviana. **Lexical bundles in speech and writing: An initial taxonomy**. na, 2003.

BIBER_ *Corpus linguistics is for text-lovers_2021*. 2022. Kindle file.

BIBER_McCARTHY_CARTER_BROWN_Quantitative Corpus-Based Research_Much More than Bean Counting_2001. 2022. Kindle file.

BNC CONSORTIUM et al. British national corpus. **Oxford Text Archive Core Collection**, 2007.

BRASIL, Ministério da Educação. **A Base Nacional Comum Curricular – A Etapa do Ensino Fundamental**. Brasília: MEC. 2017 g.

CAMARGO, Jessyca, C. M. H. REFLEXÕES SOBRE A FORMAÇÃO REQUERIDA AOS PROFESSORES DE LÍNGUA INGLESA PARA CRIANÇAS 2022. Dissertação (Mestrado em Letras) – Programa de Pós Graduação em Letras, Universidade Tecnológica Federal do Paraná. Pato Branco, 2022.

CHEN, Yu-Hua; BAKER, Paul. Lexical bundles in L1 and L2 academic writing. 2010.

COBB, Tom; BOULTON, Alex. Classroom applications of corpus analysis. 2015.

CONRAD, Susan M.; BIBER, Douglas. The frequency and use of lexical bundles in conversation and academic prose. **Lexicographica**, 2004.

CORTES, V. **Situating lexical bundles in the formulaic language spectrum: origins and functional analysis developments**. In: CSOMAY, E.; CORTES, V. (ed.). *Corpus-based research in Applied Linguistics: studies in honor of Doug Biber*. Amsterdam: John Benjamins Publishing Company, 2015. p. 197-216.

CORTES, Viviana. Lexical bundles in freshman composition. **Using corpora to explore linguistic variation**, v. 9, p. 131-145, 2002.

COUNCIL, British. Demandas de aprendizagem de inglês no Brasil: elaborado com exclusividade para o British Council pelo Instituto de Pesquisa Data Popular. **São Paulo**, 2014.

DAVID, Patricia Denicoló. O inglês no mundo: língua de prestígio. **Trama**, v. 1, n. 2, p. 209-215, 2005.

DE OLIVEIRA, Lúcia Pacheco. Linguística de Corpus: teoria, interfaces e aplicações. **Matraga-Revista do Programa de Pós-Graduação em Letras da UERJ**, v. 16, n. 24, 2009.

EAGLETON, T. (2005). A ideia de cultura. Trad. Sandra Castello Branco; revisão técnica Cezar Mortari. São Paulo: Editora UNESP.

EGBERT, et al. Designing and Evaluating Language Corpora: A Practical Framework for Corpus Representativeness. Cambridge University Press, 2022. Kindle file.

Ensinos de Meishu-Sama/Organização e tradução IMMB, 6 ed. - São Paulo: Fundação Mokiti Okada, 2019, p. 53

FERREIRA PAULINO, Suzana. **Interculturalismo e ensino de língua inglesa: é na tua língua que falo, mas é na minha língua que te compreendo**. 2009. Dissertação de Mestrado. Universidade Federal de Pernambuco.

FRANCIS, W. Nelson; KUCERA, Henry. Brown corpus manual. **Letters to the Editor**, v. 5, n. 2, p. 7, 1979.

GEERTZ, Clifford. A Interpretação das Culturas. São Paulo: LTC, 1989.

GOVERNO FEDERAL. Diretrizes Curriculares Nacionais para a oferta de Educação Plurilíngue -Parecer CNE/CEB Nº: 2/2020 aprovado em 09/07/2020.

JUVINO, Drielle Caroline Izaias; DE ASSUNÇÃO BARBOSA, Lúcia Maria. O LÉXICO E A CULTURA EM CANÇÕES BRASILEIRAS. **Revista Desempenho**, n. 19, 2013.

KRAMSCH, Claire. Learning and Teaching Pragmatics in the Globalized World: Commentary. **The Modern Language Journal**, v. 105, n. 3, p. 760-763, 2021.

MARCUSCHI, Beth. Escrevendo na escola para a vida. **Coleção explorando o ensino: língua portuguesa. Brasília: Ministério da Educação, Secretaria de Educação Básica**, p. 65-84, 2010.

SALOMÃO, Ana Cristina Biondo. O componente cultural no ensino e aprendizagem de línguas: desenvolvimento histórico e perspectivas na contemporaneidade. **Trabalhos em Linguística Aplicada**, v. 54, p. 361-392, 2015.

SANTOS, P., ALVAREZ, M. L. O. Língua e Cultura no Contexto de Português Língua Estrangeira. Campinas, SP: Pontes Editores, 2010.

SARDINHA, Tony Berber. Corpus brasileiro. **Informática**, v. 708, p. 0-1.2010, 2010.

SARMENTO, Simone. Linguística de Corpus: histórico, metodologia, campos de aplicação. **Trama. Marechal Cândido Rondon, PR. Vol. 6, n. 12 (2010), p. 89-112**, 2010.

SARMENTO, Simone. **O ensino de cultura na aula de língua estrangeira: o discurso e a prática do professor**. 2001.

SCOTT, M. WordSmith Tools. Versão 5.0. Oxford: Oxford University Press, 2021.

SILVA, E. B. **Léxico, cultura e formação crítica na aula de língua inglesa**. In: LUTERMAN, L. A.; POZZOBON, M. M.; SILVA, V. R.; THEREZA JÚNIOR, A. H. (org.). Educação linguística e formação docente: diferentes olhares epistemológicos. Campinas: Pontes. 2017. p. 231-243.

SINCLAIR, John. Intuition and annotation—the discussion continues. In: **Advances in corpus linguistics**. Brill, 2004. p. 39-59.

TRASK, Robert Lawrence. **Dicionário de linguagem e lingüística**. Contexto, 2004.

VIANA, Vander; TAGNIN, Stella EO. **Corpora no ensino de línguas estrangeiras**. Hub Editorial, 2011.

ZANINI, M. Uma visão panorâmica da teoria e da prática do ensino de língua materna. *Acta Scientiarum*, v. 21, n. 1, 1999.