

ARTIGO ORIGINAL

# Data Driven Learning para aprendizes iniciantes de inglês: uma experiência em um curso de estratégias de leitura

*Data Driven Learning for beginning English learners: an experience in a reading strategies course*

Luana Viana dos Santos<sup>1</sup>, Paula Tavares Pinto<sup>2</sup>

1 PPGEL-Universidade Estadual Paulista "Júlio de Mesquita Filho" - luana.viana-santos@unesp.br

2 Universidade Estadual Paulista "Júlio de Mesquita Filho" - paula.pinto@unesp.br

## Como citar o artigo

SANTOS, L. V. dos; PINTO, P. T. Data Driven Learning para aprendizes iniciantes de inglês: uma experiência em um curso de estratégias de leitura. *Revista Horizontes de Linguística Aplicada*, v. 23, n. 1, AG3, 2024.

## Resumo

Este artigo apresenta as etapas de elaboração de atividades direcionadas por dados voltadas aos aprendizes de nível básico de proficiência em inglês de um curso de estratégias de leitura oferecido em uma universidade multicampi. Também é discutida a percepção dos alunos acerca das atividades por meio da análise de suas respostas ao questionário aplicado no final do curso. A elaboração do material didático se deu a partir do levantamento dos termos multipalavras do *corpus* de estudo, que guiou a extração de cinco artigos utilizados para o desenvolvimento de atividades do tipo *hands on* e *hands off* com o auxílio das ferramentas de *corpora* voltadas ao ensino *CorpusMate* (BAISA; CROSTHWAITE, 2023), *Skell* (BAISA; SUCHOMEL, 2014) e *VersaText* (THOMAS, 2020). O estudo baseia-se nos trabalhos da Linguística de *Corpus* (LC) e da abordagem Aprendizagem Direcionada por Dados, *Data Driven Learning* (DDL) (BERBER SARDINHA, 2000a; 2000b; 2004; 2005; 2011; BOULTON; COBB, 2017; BAISA; CROSTHWAITE, 2021; JOHNS, 1988; MARK; O'KEEFFE, 2022; PINTO *et al.*, no prelo; TAGNIN; VIANA, 2011). Os resultados apontam como a DDL pode ser proveitosa para estudantes de níveis A1 e A2 desde que sejam feitas as adaptações necessárias.

**Palavras-chave:** Atividades direcionadas por dados. *Data Driven Learning*. Estratégias de Leitura. Nível básico.

## Abstract

This paper presents the stages of developing data-driven activities aimed at learners at a basic level of English proficiency in a reading strategies course offered at a multi-campus university. Students' perception of the Skell activities is also discussed through the analysis of their responses to the questionnaire administered at the end of the course. The preparation of the teaching material was based on the survey of multiword terms in the study *corpus*, which guided the extraction of five articles used to develop hands-on and hands-off activities with the help of the corpora tools *CorpusMate* (BAISA; CROSTHWAITE, 2023), *Skell* (BAISA; SUCHOMEL, 2014) and *VersaText* (THOMAS, 2020). The study is based on the works of *Corpus Linguistics* (LC) and *Data Driven Learning* (DDL) approach (BERBER SARDINHA,

Fonte de financiamento: Capes.

Conflito de interesse: As autoras declaram não haver.

Data de recebido: 18 Mar 2024. Data de aprovado: 21 Maio 2024.



Este é um artigo publicado em acesso aberto (Open Access) sob a licença Creative Commons Attribution Non-Commercial No Derivative, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que sem fins comerciais, sem alterações e que o trabalho original seja corretamente citado.

2000a; 2000b; 2004; 2005; 2011; BOULTON; COBB, 2017; BAISA; CROSTHWAITE, 2021; JOHNS, 1988; MARK; O'KEEFFE 2022; PINTO *et al.*, in press; TAGNIN; VIANA, 2011). The results show how DDL can be beneficial for students at levels A1 and A2 as long as the necessary adaptations are made.

**Keywords:** Data driven tasks. Data Driven Learning. Reading Strategies. Basic level.

## 1 INTRODUÇÃO

A Língua Inglesa (doravante LI) passou a ser considerada indispensável para o sucesso acadêmico de graduandos, pós-graduandos e demais pesquisadores (HYLAND, 2008). Diante do atual cenário de crescente incentivo à internacionalização das instituições de ensino superior, as universidades têm investido no desenvolvimento da proficiência linguística de suas comunidades, uma vez que as instituições que oferecem oportunidades de intercâmbio e mobilidade têm como língua franca o inglês (BAFFI-BONVINO; KANEKO-MARQUES, 2020; GARCIA *et al.*, 2021). Com isso, o oferecimento de cursos de curta duração com propósitos específicos tem ocorrido sob a supervisão dos professores dos cursos de Letras e com o apoio de graduandos e pós-graduandos das instituições (BAFFI-BONVINO; KANEKO-MARQUES, 2020; GARCIA *et al.*, 2021), que passaram a desenvolver material didático especializado para seu público-alvo.

Compreendido sob a perspectiva dos pressupostos teóricos e metodológicos das abordagens de Línguas Para Fins Específicos (LinFE), o ensino de LI caracteriza-se por colocar as necessidades específicas dos alunos no centro da aprendizagem (HUTCHINSON; WATERS, 1987). A abordagem *English for Specific Purposes* (ESP), ou Inglês Para Fins Específicos (IFE), baseia-se no ensino de habilidades, gêneros e aspectos da língua dos quais os alunos mais necessitam, com o intuito de desenvolver capacidades linguísticas específicas para a comunicação em diversas atividades, sejam elas acadêmicas ou profissionais (HEEMANN, 2009). No contexto acadêmico, com base no aporte teórico do *English for Academic Purposes* (EAP), ou Inglês Para Fins Acadêmicos (IFA), uma ramificação do ESP (O'BOYLE; VIANA, 2022), dá-se ênfase para as habilidades comunicativas próprias das atividades realizadas no âmbito universitário, nos cursos de graduação e pós-graduação.

Neste artigo, discutiremos o processo de elaboração das atividades didáticas desenvolvidas para o curso Estratégias de Leitura em Língua Inglesa – Nível Básico, oferecido em 2023 no formato de disciplina optativa *on-line* para os alunos da pós-graduação de todos os câmpus da Universidade Estadual Paulista (Unesp), compondo o Portifólio de Integração de Disciplinas da Pós-Graduação da Unesp (PROPG)<sup>1</sup>, conjunto de disciplinas transversais e específicas voltadas para o corpo discente da instituição, no formato de ensino híbrido, com o objetivo de gerar a integração entre os Programas de Pós-Graduação (PPGs) (UNESP, 2023).

As atividades foram desenvolvidas conforme a abordagem *Data Driven Learning* (DDL), ou em português, Aprendizagem Direcionada por Dados, resultado da associação entre a Linguística de *Corpus* (doravante LC) e o ensino de línguas cuja máxima é o desenvolvimento da autonomia do aprendiz por meio do livre acesso aos dados da língua-alvo e de sua investigação, mediada pelo professor (BERBER SARDINHA, 2011; BOULTON; COBB, 2017; JOHNS, 1988). Os dados linguísticos são apresentados sob forma de um *corpus*, coletânea de textos autênticos coletados conforme critérios pré-estabelecidos para servir a uma pesquisa (BERBER SARDINHA, 2004).

Conforme Nesi (2016), a união entre a LC e o EAP tornou-se comum uma vez que os *corpora* podem fornecer informações quantitativas sobre o discurso acadêmico, embasando informações qualitativas, além de possibilitarem oportunidades de aprendizagem baseadas em dados provindos do uso real da língua. O material que será apresentado neste trabalho é resultado de uma pesquisa maior, em nível de Mestrado, assim como os dados de pesquisa que serão analisados.

---

<sup>1</sup> Portfólio de integração de disciplinas da pós-graduação. Disponível em: <https://www2.unesp.br/portal#!/propg/disciplinas-portfolio>.

O público-alvo do curso de estratégias de leitura são aprendizes iniciantes de inglês, de níveis A1 e A2, conforme o Quadro Comum Europeu de Referência para Línguas (QCER)<sup>2</sup> (BRITISH COUNCIL, 2023); assim, nos propusemos a desenvolver atividades orientadas pela DDL para esse perfil de estudantes, ainda que as pesquisas em torno dos resultados da abordagem evidenciem melhores resultados com alunos intermediários e avançados (BOULTON; COBB, 2017). Apesar de serem alunos iniciantes, os participantes da pesquisa estão inseridos no contexto acadêmico e realizam atividades com a LI requeridas a esse perfil de estudante, como a leitura de artigos científicos.

Diante do cenário ilustrado, buscamos responder à seguinte questão norteadora: Em que medida atividades didáticas direcionadas por dados podem influenciar a aprendizagem de leitura em LI de estudantes de nível básico? A investigação se apoiou no Modelo de Aquisição, complexidade e formulaicidade para o *design* de DDL proposto por Mark e O’Keeffe (2022).

## 2 PRESSUPOSTOS TEÓRICOS

### 2.1 O ensino de inglês no contexto acadêmico: ESP e EAP

O IFE é uma das abordagens mais bem estabelecidas no ensino de LI (ANTHONY, 2018), que coloca as necessidades dos alunos no centro da aprendizagem (HUTCHINSON; WATERS, 1987). Conforme Anthony (2018, p. 10-11, tradução nossa),

Inglês para Fins Específicos (IFE) é uma abordagem de ensino de línguas que visa às necessidades acadêmicas ou ocupacionais atuais e/ou futuras dos alunos, concentra-se no idioma, nos gêneros e nas habilidades necessários para atender a essas necessidades e auxilia os alunos a atenderem a essas demandas através do uso de materiais e métodos de ensino gerais e/ou específicos de disciplinas<sup>3</sup>.

Trata-se de uma abordagem que se vale da análise prévia de necessidades de aprendizagem, da análise de gêneros textuais, dos métodos comunicativos de ensino e do construtivismo (HYLAND, 2022), diferenciando-se do Inglês para Fins Gerais (IFG), que está pautado no ensino das quatro habilidades linguísticas e na formação de falantes proficientes sem propósitos acadêmicos ou profissionais previamente definidos. Por dar ênfase às necessidades de aprendizes de contextos acadêmicos e profissionais variados, o IFE desdobrou-se em várias outras abordagens consideradas ramificações. O EAP, Inglês para Fins Acadêmicos (IFA), em português, é talvez a ramificação mais influente do IFE (Anthony 2018). Segundo O’Boyle *et al.* (2022, p. 2, tradução nossa), o IFA

[...] refere-se ao aprendizado, ensino e pesquisa de práticas de linguagem e discurso em inglês que as pessoas usam para estudar e/ou trabalhar no ensino superior. Ele se baseia em uma variedade de áreas temáticas, como Linguística Aplicada, Educação e Linguística<sup>4</sup>.

De acordo com Hyland (2006), o IFA pode ainda ser dividido em mais duas ramificações, o *English for General Academic Purposes* (EGAP), ou Inglês para Fins Acadêmicos Gerais (IFAG), e o *English for Specific Academic Purposes* (ESAP), Inglês para Fins Acadêmicos Específicos (IFAE). O primeiro compreende o ensino focado em habilidades, conteúdos linguísticos e atividades comuns a todas as disciplinas de forma geral, enquanto o segundo compreende o ensino da

<sup>2</sup> *Common European Framework of Reference for Languages*: documento europeu desenvolvido para a descrição de proficiência linguística de aprendizes (BRITISH COUNCIL, 2023).

<sup>3</sup> No original: “*English for Specific Purposes (ESP) is an approach to language teaching that targets the current and/or future academic or occupational needs of learners, focuses on the necessary language, genres, and skills to address these needs, and assists learners in meeting these needs through the use of general and/or discipline specific teaching materials and method*” (ANTHONY, 2018, p. 10-11).

<sup>4</sup> No original: “[...] *refers to the learning, teaching and research of language and discourse practices in English that people use to study and/or work in higher education. It draws on a range of subject areas such as Applied Linguistics, Education and Linguistics*” (O’BOYLE *et al.*, 2022, p. 2).

língua e das habilidades relacionadas às demandas particulares de uma disciplina (HYLAND, 2006).

No Brasil, o IFE foi introduzido por meio do Projeto “Ensino de Inglês Instrumental em Universidades Brasileiras” a partir de 1970 na Pontifícia Universidade Católica (PUC-SP) sob a orientação e supervisão da Prof.<sup>a</sup> Dr.<sup>a</sup> Maria Antonieta Alba Celani (CELANI, 2005). A denominação “Inglês Instrumental” se justificava pelo caráter funcional da língua: o de usá-la como instrumento para a execução de tarefas dentro de determinada área de especialidade.

Conforme Celani (1983, p. 4), o inglês já era uma necessidade naquele momento para os estudantes brasileiros e a habilidade mais requerida era a leitura: “a necessidade de inglês, particularmente para a finalidade de leitura, é indiscutível, quer em relação a um curso determinado, quer para complementação dos estudos”. Por ter chegado ao Brasil como uma disciplina com foco para a Compreensão Escrita (CE), o IFE acabou sendo associado ao ensino de leitura instrumental; no entanto, já é de amplo conhecimento que a abordagem visa ao ensino de quaisquer habilidades, desde que sejam elencadas como prioritárias a determinado grupo de aprendizes.

## 2.2 Linguística de *Corpus* e ensino: a abordagem *Data Driven Learning*

Desde as décadas de 1980 e 1990 a LC, área da Linguística voltada para a “exploração da linguagem por meio de evidências empíricas, extraídas por computador” (BERBER SARDINHA, 2004, p. 3) tem sido usada no ensino de LE, especialmente nos contextos de ensino de línguas para fins específicos, primeiramente no ensino de inglês (TAGNIN; VIANA, 2011), por meio do *classroom concordancing*, ensino baseado no uso de linhas de concordâncias de *corpora*, e do *Data Driven Learning* (DDL), aprendizagem direcionada por dados introduzida pelo linguista Tim Johns (BERBER SARDINHA, 2011).

O aluno, dentro do escopo da DDL, é visto como um pesquisador ou mesmo um detetive da língua (BERBER SARDINHA, 2011). Nas atividades baseadas em *corpora*, os estudantes exploram os dados fornecidos pelo *corpus*, amostras da linguagem que contêm as informações morfológicas, gramaticais, fraseológicas e discursivas tomadas como objeto de ensino pelos professores, a fim de entrar em contato com o léxico e com a gramática de determinados gêneros discursivos.

Berber Sardinha (2000b, p. 61) considera que a DDL gera impacto no aluno, no professor e no ensino como um todo:

(1) Aluno: Assume um papel mais central e independente, tornando-se um pesquisador dotado de habilidades que lhe permitem obter respostas para suas dúvidas e verificação de suas hipóteses. (2) Professor: Deixa de ser a fonte única de informação na sala de aula. Por conseguinte diminui a pressão sobre ele para saber todas as respostas. O computador, enquanto informante, assume um papel de destaque como provedor de informação. (3) Ensino: Centra-se mais no aluno e permite que se fundamente pelo princípio de descoberta. Deixa, portanto, de estar baseado na transmissão de regras prontas.

A DDL possui um caráter transformador, pois redimensiona os papéis tradicionalmente estabelecidos para professores e alunos. Essa característica também pode ser muito desafiadora, uma vez que, diante das informações contidas em um *corpus*, os alunos podem se deparar com dados que podem ser tão inusitados para eles como para os professores. A riqueza linguística contida em amostras reais de uso da língua não cabe nos livros didáticos ou nas gramáticas, por isso, o aluno, assim como professor descobrirão novas informações sobre a língua. Segundo Johns (1988, p. 3, tradução nossa):

O que há de distintivo na abordagem DDL para o ensino indutivo de línguas é o princípio de que os dados são primários e o professor não consegue antecipar exatamente quais regras ou padrões os alunos descobrirão: na verdade, eles muitas vezes perceberão coisas que são desconhecidas não apenas pelo professor, mas também pelas obras padrão de

referência sobre o idioma. É este elemento de desafio e de descoberta que dá ao DDL o seu sabor especial e estimulante<sup>5</sup>.

De acordo com Boulton e Cobb (2017), a DDL concentra seus melhores resultados no contexto de ensino superior, em que se encontram estudantes entre os níveis intermediário e avançado de proficiência, do B1 ao C2, embora ressaltem que já existem pesquisas voltadas para experiência com DDL no ensino básico e nos cursos livres de idiomas. Mark e O’Keeffe (2022) reforçam que os princípios da abordagem, como a frequência de exposição à linguagem natural no processo de aprendizagem de línguas, podem ser amplamente aproveitados por todos os aprendizes de LI, desde que haja uma curadoria do material retirado dos *corpora* de acordo com o que se espera dos alunos em cada nível proposto pelo QCER.

Conforme Mark e O’Keeffe (2022, p. 2, tradução nossa), “[...] o ethos tradicional da DDL é promover a descoberta indutiva independente por meio da exploração livre de *corpus*, no entanto, isso requer pelo menos um nível intermediário de competência”<sup>6</sup>. As autoras, então, argumentam a favor de orientações para o ensino dirigido por *corpora*, conforme as quais os professores desempenham um papel mais influente na seleção e mediação das tarefas, diferenciadas por níveis, para que os estudantes possam de fato se beneficiar dos recursos fornecidos pela LC.

Mark e O’Keeffe (2022) associam descobertas provindas das pesquisas em Aquisição de Segunda Língua (ASL), *Second Language Acquisition* (SLA), especialmente entre os modelos de aquisição *usage-based* (UB), que concebem a aquisição como um resultado do uso e da exposição do aprendiz à língua (MARK; O’KEEFFE, 2022). Os modelos UB mostram que “os aprendizes de segunda língua normalmente passam de um repertório de sequências holofrásticas fixas em níveis baixos para um sistema expandido de *slots* e *frames* para padrões totalmente abstratos (geralmente figurativos)”<sup>7</sup> (MARK; O’KEEFFE, 2022, p. 2, tradução nossa), ou seja, inicialmente, descobrem padrões da língua, como combinações de palavras, para então se apropriarem de sequências mais complexas que requerem do falante um conhecimento maior do idioma.

As autoras propõem um modelo para aprimorar a curadoria de DDL e o *design* de atividades baseado nos princípios da aquisição, da complexidade e da formulaicidade (MARK; O’KEEFFE, 2022), isso porque: i. a aquisição nos níveis mais baixos se difere da aquisição nos níveis mais avançados (MARK; O’KEEFFE, 2022), ii. conseqüentemente, nos níveis mais baixos, a complexidade varia e iii. a formulaicidade é um marcador de um aluno avançado, pois o uso de linguagem formulaica, como colocações e expressões idiomáticas, requer maior conhecimento e consciência da língua (MARK; O’KEEFFE, 2022).

O esquema contido na imagem a seguir ilustra a organização do processo de aquisição de LE com base na variabilidade da complexidade e da formulaicidade inerente a cada nível do QCER conforme o modelo para *design* de materiais orientados pela DDL proposto por Mark e O’Keeffe (2022) (fig. 1).

<sup>5</sup> No original: “What is distinctive about the DDL approach to inductive language teaching is the principle that the data is primary, and the teacher does not know in advance exactly what rules or patterns the learners will discover: indeed, they will often notice things that are unknown not only to the teacher, but also to the standard works of reference on the language. It is this element of challenge and of discovery that gives DDL its special flavour and stimulus” (JOHNS 1988, p. 3).

<sup>6</sup> No original: “[...] the traditional ethos of DDL is to promote independent inductive discovery through free corpus foraging, however, this requires at least an intermediate level of competency” (MARK; O’KEEFFE, 2022, p. 2).

<sup>7</sup> No original: “[...] second language learners typically move from a repertoire of fixed holophrasal sequences at low levels to an expanded slot and frame system to fully abstracted (often figurative) patterns [...]” (MARK; O’KEEFFE, 2022, p. 2).

	A1	A2	B1	B2	C1	C2
<b>Princípio</b>	→					
<b>Aquisição</b>	combinação de palavras		identificação de slots e frames		saber formulaico: refinamento de padrões abstratos para a percepção do registro de associação de palavras	
<b>Complexidade</b>	foco em padrões verbais		mudança de foco em padrões verbais para padrões nominais		foco na complexidade nominal	
<b>Formulaicidade</b>	foco no sentido literal		ampliação do foco no sentido literal para foco no sentido figurado		foco no sentido figurado	

Figura 1: Modelo de aquisição, complexidade e formulaicidade para o *design* de DDL  
 Fonte: Elaborado com base em Mark e O’Keeffe (2022, p. 10).

De acordo com as autoras,

Esta estrutura se afasta do ethos original do DDL, no qual o aluno se envolve em um processo de descoberta de aprendizagem indutiva. Embora este ainda seja um ideal alcançável, argumentamos que, para que o DDL funcione em todos os níveis, há necessidade de uma mediação baseada em princípios para diferenciar dados e tarefas para diferentes níveis<sup>8</sup>. (MARTIN; O’KEEFFE, 2022, p. 9, tradução nossa).

As atividades de leitura devem facilitar o reconhecimento dos recursos facilitadores da compreensão, e para o caso do público-alvo desta pesquisa, devem trabalhar com os itens lexicais relacionados às estratégias de leitura contidas no quadro de conteúdos previstos no curso. Como nos propomos a desenvolver material didático com auxílio da LC, nos pautamos no modelo proposto por Mark e O’Keeffe (2022) e nos critérios obrigatórios para atividades baseadas em *corpus* propostos por Delfino (2016) (quadro 1).

**Quadro 1.** Critérios obrigatórios para elaboração de atividades com base em *corpora*

- O1 – O exercício faz uso de *corpus*;
- O2 – O exercício precisa ter enunciados claros;
- O3 – O exercício tem como foco principal o padrão léxico-gramatical;
- O4 – O exercício é ético;
- O5 – O exercício é replicável;
- O6 – O exercício é motivador;
- O7 – O exercício não simplifica a língua usada nos textos/concordâncias/lista de palavras, etc.;
- O8 – O exercício deve apresentar nível de dificuldade adequado;
- O9 – O exercício contém conteúdo relevante para o aluno e para a construção do conhecimento em inglês;
- O10 – O professor é facilitador e não distribuidor de conhecimento;
- O11 – O aluno é descobridor, pesquisador e não recipiente de conhecimento.

Fonte: Delfino (2016, p. 56).

<sup>8</sup> No original: “This framework moves away from the original ethos of DDL in which the student engaged in a discovery process of inductive learning. While this is still an attainable ideal, we argue that if DDL is to work across levels, there is need for principled mediation to differentiate data and tasks for different levels” (MARTIN; O’KEEFFE, 2022, p. 9).

Valendo-nos dos dados linguísticos obtidos pelo *corpus* de estudo e da concepção de leitura como processo interativo de produção de sentidos, elaboramos as atividades de CE em LI conforme os critérios acima apresentados. Buscamos também investigar os recursos fornecidos pelas ferramentas *on-line CorpusMate* (CROSTHWAITE; BAISA, 2023), *Skell* (BAISA; SUCHOMEL, 2014) e *VersaText* (THOMAS, 2020) com o propósito de desenvolver exercícios para que os alunos leiam trechos selecionados dos artigos e entrem em contato com ferramentas simples e de fácil manuseio.

### 2.3 A habilidade de compreensão escrita e as estratégias de leitura

Para a elaboração de atividades didáticas voltadas para o aprimoramento de CE, cabe entender quais são os processos de leitura. Leffa (1996) considera a existência de duas acepções do que seja a leitura, consideradas definições restritas e antagônicas, a de que “(a) ler é extrair significado do texto” (p. 11) e a de que “(b) ler é atribuir significado ao texto” (p. 11). Na primeira, a leitura é tomada como um processo ascendente, uma vez que o significado parte do texto para o leitor (LEFFA, 1996). Já na segunda, a leitura é tomada como um processo descendente, em que os significados relacionados às hipóteses do leitor são confirmados ou não pela leitura do texto (LEFFA, 1996). Ambos os processos possuem suas limitações, já que “A complexidade do processo da leitura não permite que se fixe em apenas um de seus polos” (LEFFA, 1996, p.17) e que “Para compreender o ato da leitura temos que considerar então (a) o papel do leitor, (b) o papel do texto e (c) o processo de interação entre o leitor e o texto” (LEFFA, 1996, p. 17). No contexto de IFE, o ensino de CE volta-se para as estratégias de leitura, cuja função é:

[...] desenvolver habilidades tais como deduzir o significado e uso de itens lexicais não familiares; entender a informação explícita e a implícita dentro da sentença; entender o valor comunicativo das sentenças; extrair seletivamente pontos relevantes do texto; distinguir ideias principais das secundárias, etc. [...] (HEEMANN, 2009, p. 144).

Conforme Baldo (2009), a leitura eficiente em LE está relacionada ao processo de aquisição de vocabulário, principalmente contextual, de modo que o conhecimento de estruturas sintáticas teria menor influência na proficiência leitora uma vez que essa é influenciada pelo repertório de informações garantido pelo acervo lexical dos leitores. Segundo a autora, leitores mais proficientes na LE conseguem mobilizar as estratégias de leitura com que dispõem na sua Língua Materna (LM). Nesse sentido, os recursos facilitadores da compreensão são eficientes quando se é capaz de reconhecê-los, porque não importa a quantidade de vezes que se lance mão deles sem que se consiga enxergar sua presença nos textos (BALDO, 2009).

Assim, é possível afirmar que no processo de leitura “o leitor passa a ser visto como um sujeito ativo, um bom usuário de estratégias e um aprendiz cognitivo” (HEEMANN, 2009, p. 138), que é capaz de relacionar seu conhecimento de mundo, ou conhecimento prévio, com as informações do texto. A leitura concebida como processo de produção de sentidos, resultado de uma atividade interativa entre o texto, o autor, o leitor e o contexto mobiliza tanto o processo ascendente, do inglês *bottom up*, quanto o descendente, *top down* (LEFFA, 1996; HEEMANN, 2009).

Os recursos oferecidos pelas ferramentas citadas favorecem a produção de atividades para três momentos importantes de uma aula de CE em LE, segundo Tomitch (2009): a pré-leitura, o momento anterior à leitura de textos na língua alvo em que o professor faz uma sondagem a respeito do conhecimento prévio dos alunos acerca do tema da aula/do texto e em que se faz o acionamento de esquemas relevantes para a construção do significado; a leitura, em que o aluno recebe o texto e o lê efetivamente; e a pós-leitura, fase em que faz-se “a ligação entre o texto e a realidade do aluno” (TOMITCH, 2009, p. 197). A pós-leitura é um momento muito importante, pois é a fase de consolidação do que foi aprendido (TOMITCH, 2009).

Conforme sugere Tomitch (2009), visamos elaborar as atividades levando em consideração o objetivo do curso em que serão aplicadas, o interesse e o nível de proficiência dos alunos, a adequação ao seu nível linguístico em inglês, o gênero escolhido para leitura e o tamanho dos textos levados para as aulas. Destacamos que não objetivamos testar o nível de proficiência leitora dos alunos, uma vez que entendemos que as atividades propostas devem apresentar as estratégias de leitura como ferramentas para alcançar a compreensão leitora, ou seja, não as tomar como o objetivo final do curso.

Assim, entendendo que as atividades serão voltadas para aprendizes de nível básico, voltamo-nos para o ensino das estratégias de leitura como ferramentas “que levem o aluno a otimizar o conhecimento que já possui na língua materna e na estrangeira, seja esse conhecimento linguístico, metalinguístico ou de mundo [...]” (TOMITCH, 2009, p. 195).

### 3 METODOLOGIA

#### 3.1 O contexto de aplicação das atividades

O curso Estratégias de Leitura em Língua Inglesa – Nível Básico é fruto da parceria público-privada entre a Unesp e o banco Santander que deu origem ao Programa de Língua Inglesa da Unesp (PLIU), hoje intitulado Programa de Língua Estrangeira da Unesp (PLEU). Em 2020, o curso teve como público-alvo 200 graduandos dos campus da universidade (PINTO; SANTOS, 2022) e foi novamente oferecido no segundo semestre de 2023 como disciplina optativa do Portifólio de Integração de Disciplinas da PROPG da Unesp, como mencionado na seção introdutória. Foram oferecidas quatro turmas com 30 vagas cada uma, intituladas Turmas 1, 2, 3 e 4, com carga horária de 30 horas e com aulas ofertadas nos períodos da manhã e da tarde.

As atividades desenvolvidas foram aplicadas para as Turmas 3 e 4, em que as autoras puderam ter contato com os alunos. As aulas aconteciam às segundas e às quartas das 16h30 às 18h30, por meio de aplicativo de videoconferências, durante o período de 7 de agosto a 29 de setembro de 2023. A Turma 3 era composta por 26 alunos e a Turma 4 por 29 alunos.

As atividades desenvolvidas foram aplicadas juntamente com outras já previstas no material didático do curso, então, os alunos fizeram exercícios baseados em *corpus* e exercícios de outras naturezas, o que diversificou as metodologias empregadas nas aulas. Antes de aplicar o material preparado com base em *corpus*, a tutora, juntamente com a professora responsável pela disciplina, deu uma aula com tarefas simples de investigação das ferramentas *CorpusMate*, *Skell* e *VesaText* para que os alunos pudessem se familiarizar com elas.

Todas as atividades direcionadas por dados pautavam-se nos conteúdos previstos no cronograma, a saber: reconhecimento do *layout* de diferentes gêneros, processo de inferência e de identificação de cognatos e falsos cognatos, técnicas de *skimming* e de *scanning*, processos de formação de palavras em inglês, análise das características de resumos acadêmicos, estabelecimento de coesão e coerência nos textos e o uso e a função de marcadores do discurso e de elementos de referência.

#### 3.2 A elaboração das atividades direcionadas por dados

O *corpus* utilizado na pesquisa é o *SDG-Plos* (Pinto *et al.*, no prelo), uma coletânea de 400 artigos científicos de pesquisadores internacionais que abordaram em seus trabalhos os 17 Objetivos de Desenvolvimento Sustentável (doravante ODS) da Organização das Nações Unidas (ONU). O *corpus* foi criado a partir do *software AntCorGen* (ANTHONY, 2022) e foi compilado por meio da plataforma *Sketch Engine* (KILGARRIFF *et al.*, 2014).

Para a elaboração das atividades foram extraídas as palavras-chave do *corpus* de estudo, que foram então tratadas e analisadas como termos multipalavras dos ODS. As palavras-chaves são “[...] aquelas cujas frequências são estatisticamente diferentes no *corpus* de estudo e no *corpus* de referência” (BERBER SARDINHA, 2005, p. 238) e são geradas a partir de uma

comparação entre a lista de palavras do *corpus* de estudo e a lista de palavras do *corpus* de referência (BERBER SARDINHA, 2005).

Ao falar das palavras-chave abordamos o conceito de “chavicidade”, “[...] uma qualidade possuída por palavras, grupos de palavras, frases etc.”<sup>9</sup> (BONDI; SCOTT, 2010, p. 43, tradução nossa), índice que aponta itens lexicais proeminentes que caracterizam um texto (BONDI; SCOTT, 2010). Esses itens lexicais são representativos da amostra linguística na qual estão contidos, e quando entendidos como termos multipalavras passam a ser considerados como “uma locução ou expressão multipalavra que possui estatuto terminológico” (RAMISCH; VILLAVICENCIO, 2010, p. 41).

Como a lista de palavras-chave gerada pela ferramenta *Keywords* do *Sketch Engine* é extensa, uma das etapas da pesquisa foi a seleção daquelas que seriam analisadas como termos multipalavras e poderiam servir ao desenvolvimento de material didático. O foco deste trabalho é apresentar o processo de elaboração das atividades movidas por dados e os resultados coletados a partir do *feedback* dado pelos estudantes de nível básico, por isso, não nos debruçaremos sobre a etapa de extração de dados do *corpus*.

A figura 2 ilustra as etapas de elaboração das atividades.

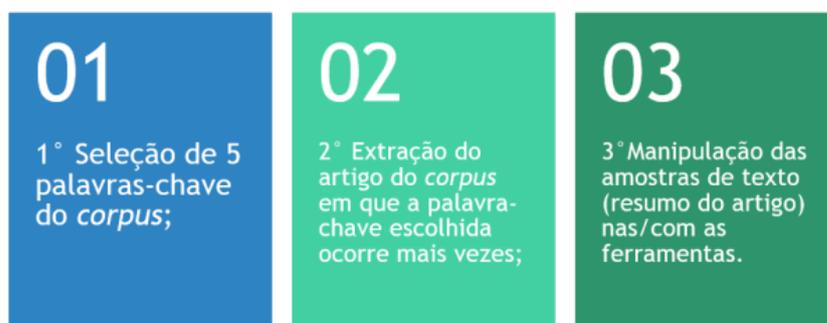


Figura 2. Etapas da elaboração das atividades

Fonte: Elaboração própria.

Foram selecionados cinco termos multipalavras dos ODS a partir dos quais seriam escolhidos os textos a serem lidos e trabalhados nas aulas. As cinco palavras-chave selecionadas para a extração dos textos foram: *ecosystem service* (com chavicidade igual a 89.6), *protected area* (com chavicidade igual a 39.7), *ocean health* (com chavicidade igual a 38.7) e *waterbird habitat* (com chavicidade igual a 35.7). Os artigos escolhidos foram aqueles em que as palavras-chave citadas ocorriam mais vezes.

Selecionados os textos, optamos por trabalhar com a seção de *abstracts*, ou seja, os resumos dos artigos, que são de extensão curta e contêm de forma sintetizada os objetivos, o referencial teórico, a metodologia e os resultados dos trabalhos. Essa organização do material, que seria lido e utilizado nas atividades, foi fundamental, dado o tempo de duração do curso, que não permitia a leitura detalhada de todas as seções do artigo.

Os resumos foram manipulados no *VersaText*, e a partir dos recursos contidos na ferramenta, bem como dos recursos do *CorpusMate* e do *Skell*, elaboramos os exercícios de cada atividade. Foram desenvolvidas cinco atividades, mas foi possível aplicar somente quatro em razão da carga horária do curso.

O *CorpusMate* é uma plataforma desenvolvida pelos pesquisadores Peter Crosthwaite e Vít Baisa (2023) pensada para o ensino de línguas baseado e dirigido por *corpora* a alunos de ensino básico (BAISA; CROSTHWAITE, 2023). Seu *corpus* foi compilado a partir de cinco recursos, de acordo com os autores, o *corpus* BAWE, o *corpus* de TED Talks, o *Simple English*

<sup>9</sup> No original: “*Keyness is a quality possessed by words, word-clusters, phrases etc. [...]*” (BONDI; SCOTT, 2010, p. 43).

Wikipédia, o corpus BBC Tech, o Elsevier Corpus e o BNC 2014 Spoken, o que fornece uma gama de dados da modalidade escrita e falada da língua.

O *Skell* (*Sketch Engine for Language Learners*) é uma vertente simplificada do *Sketch Engine* com *layout* amigável e com recursos que são aproveitados dos *corpora* de seu *software* de origem. Os recursos do *Skell* são o *Examples*, que apresenta frases retiradas dos *corpora* do *Sketch Engine* com as palavras de busca para que o aluno as veja em contexto; o *Word sketch*, que oferece ao aluno as combinações da palavra de busca com outras que podem atuar como seu sujeito, seu objeto, seu modificador, etc.; e o *Similar Words*, que traz sinônimos da palavra de busca.

A terceira ferramenta, *VersaText*, que “oferece ferramentas para a exploração linguística de um único texto”<sup>10</sup> (THOMAS, 2020, p. 9, tradução nossa). Nele, o aluno pode colar amostras de textos e investigá-las quanto ao seu vocabulário por meio da criação de nuvem de palavras através do *Wordcloud*, da análise estatística e morfológica do léxico por meio do *Profiler* ou pela leitura detalhada das linhas de concordância, com os termos coloridos, por meio do *Concordance*.

Cada uma das quatro atividades está estruturada em exercícios *pre-reading*, *reading* e *post-reading*. Para ilustrar sua organização, iremos descrever a atividade número 2. A figura 3 apresenta o primeiro exercício da seção *pre-reading*, em que o aluno deve relacionar as palavras do quadro com as imagens dadas.

### 1- Relacione os termos do quadro com as imagens a seguir:

**Climate - Urbanization - Water management - Waterbird habitat**









Figura 3. Exercício 1 da seção *pre-reading* da atividade 2

Fonte: Elaboração própria.

<sup>10</sup> No original: “[...] offers tools for exploring the language of a single text” (THOMAS, 2020, p. 9).



### Reading

- Click here to read text:  
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0169780>;
- Pay attention to the way the information is organized (Are there any pictures? Graphs? Videos? How is the text organized?);
- Keep your focus on the abstract;
- You should skim the text in 2 or 3 minutes.

Figura 6. Momento *Reading* da atividade 2  
Fonte: Elaboração própria.

Os alunos puderam ler o resumo no próprio *site* em que foi publicado, seu suporte de veiculação original. Nas instruções, pede-se que os estudantes observem o *layout* do texto e façam uma leitura rápida, *skimming*, para mobilizar as estratégias já vistas no início do curso.

O exercício 2 da seção *post-reading* focaliza o conteúdo ministrado naquela semana de aplicação das atividades, os marcadores do discurso, ou *linking words*:

2- Assinale a alternativa em que a *linking word* destacada exerça a mesma função que "despite" no trecho a seguir: *The Central Valley of California is one of the most important regions for wintering waterbirds in North America **despite** extensive anthropogenic landscape modification and decline of historical wetlands there.*

- a) The Central Valley of California is one of the most important regions for wintering waterbirds in North America **because of the** extensive anthropogenic landscape modification and decline of historical wetlands there.
- b) The Central Valley of California is one of the most important regions for wintering waterbirds in North America **as well as** extensive anthropogenic landscape modification and decline of historical wetlands there.
- c) The Central Valley of California is one of the most important regions for wintering waterbirds in North America **such as** extensive anthropogenic landscape modification and decline of historical wetlands there.
- d) The Central Valley of California is one of the most important regions for wintering waterbirds in North America **even with** extensive anthropogenic landscape modification and decline of historical wetlands there.

Figura 7. Exercício 2 da seção *post-reading* da atividade 2  
Fonte: Elaboração própria.

Para ilustrar um exercício em que o aluno deveria utilizar a ferramenta *Skell*, apresentamos a seguir o exercício 5 da seção *pre-reading* da atividade 4, elaborada a partir da palavra-chave *human well-being*:



Os itens de 1 a 4 os alunos deveriam pontuar de 1, considerado *weak* ou fraco, a 5, considerado *great* ou "ótimo". Essas avaliações dadas pelos alunos, assim como as sugestões oferecidas serão analisadas na seção seguinte.

#### 4 RESULTADOS ENCONTRADOS

Analizamos a percepção dos alunos acerca das atividades com base em suas respostas dadas no *Final Survey*, questionário aplicado ao final do curso. Serão analisadas primeiramente as respostas dadas pelos alunos da Turma 3, e, em seguida, as respostas dadas pelos alunos da Turma 4.

Como o questionário já era um instrumento do curso, não havia em sua estrutura perguntas que direcionassem o olhar dos alunos para a avaliação das atividades direcionadas por dados especificamente, por isso, entendemos que o formulário fornece informações sobre a percepção das atividades de forma geral, e não com especificidade sobre o uso de cada uma das ferramentas, por exemplo. A avaliação dos alunos sobre o curso será discutida, mas as perguntas 1 e 2, voltadas à qualidade do material e à adequação ao nível básico, serão priorizadas porque podem ajudar a responder a pergunta norteadora do estudo com mais detalhes.

Na Turma 3, para o primeiro item, "Qualidade das atividades propostas durante o curso", foram marcadas 19 respostas "5" (*great*/"ótimo"), e duas respostas "4", um nível abaixo de "*great*" ou "ótimo" (fig. 9).

1. Qualidade das atividades propostas durante o curso  
22 respostas

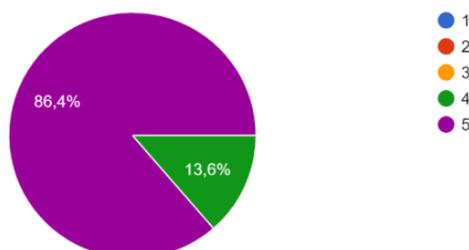


Figura 9. Gráfico com as respostas dadas na questão 1 do *Final Survey* (Turma 3)  
Fonte: *Final Survey* aplicado no final do curso.

Para o item 2, "Adequação das atividades para o nível básico", na Turma 3, 18 alunos marcaram "5", 3 (três) marcaram "4", e 1 (um) marcou "3" (fig. 10).

2. Adequação das atividades para o nível básico  
22 respostas

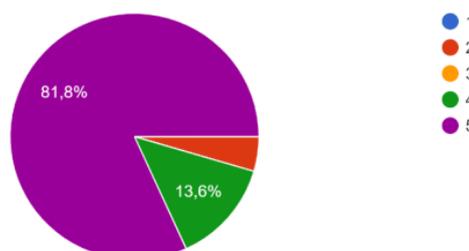


Figura 10. Gráfico com as respostas dadas na questão 2 do *Final Survey* (Turma 3)  
Fonte: *Final Survey* aplicado no final do curso.

Para o item 1, “Qualidade das atividades propostas durante o curso”, na Turma 4 tivemos 15 respostas com “5”, 3 (três) respostas com “4” e 1 (uma) resposta com “3” (fig. 11).

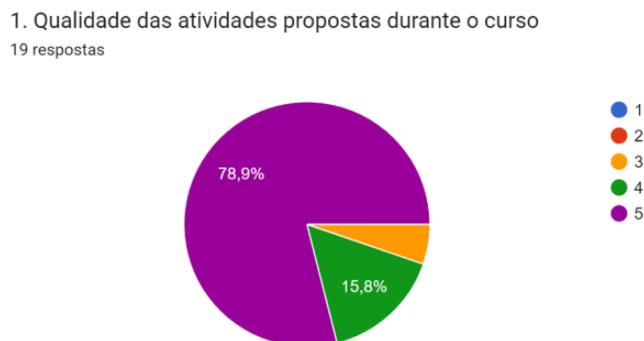


Figura 11. Gráfico com as respostas dadas na questão 1 do *Final Survey* (Turma 4)  
Fonte: *Final Survey* aplicado no final do curso.

No item 2 “Adequação das atividades para o nível básico”, 14 alunos responderam “5”, 3 (três) alunos responderam “4”, e 2 (dois) responderam “3” (fig. 12).



Figura 12. Gráfico com as respostas dadas na questão 2 do *Final Survey* (Turma 4)  
Fonte: *Final Survey* aplicado no final do curso.

Ao todo, obtivemos 41 respostas, o que significa que, somando os alunos das Turmas 3 e 4 (55), 14 alunos deixaram de responder ao questionário. Por conta disso, não podemos afirmar categoricamente que todos os alunos aprovaram as atividades; contudo, consideramos que em sua maioria os estudantes demonstraram estar satisfeitos com os materiais empregados, pois mesmo sem contar com as respostas dos que não responderam ao formulário, do total de respostas dadas, nenhuma pode ser considerada negativa.

Poucos alunos revelaram dificuldade em acompanhar o curso em razão do nível de proficiência. Como os gráficos apontam, cerca de 82% dos alunos da Turma 3 e cerca de 74% dos alunos da Turma 4 marcaram “5 (*great/ótimo*)” para esse quesito, logo, consideraram que as atividades estavam adequadas ao nível básico. As respostas dadas às questões dissertativas 5 e 6 reforçam o percentual de satisfação demonstrado nas respostas às perguntas objetivas.

Da mesma forma, podemos afirmar que os alunos consideraram as atividades de alta qualidade, como demonstrou a maioria das respostas dadas à pergunta 2: 86% dos alunos da Turma 3 e cerca de 80% dos alunos da Turma 4 avaliaram esse quesito com “5”.

## 5 CONSIDERAÇÕES FINAIS

Neste artigo, apresentamos as atividades direcionadas por dados com foco para as estratégias de leitura para alunos de níveis A1 e A2 aplicadas em um curso de estratégias de leitura; desse modo, foi necessária a investigação de quais itens lexicais poderiam servir para tal finalidade, os termos multipalavras. A criação das atividades levou tempo, tanto para a análise do *corpus* quanto para a elaboração dos exercícios em si, no entanto, conseguimos levar para alunos iniciantes os princípios da descoberta da língua e da autonomia no aprendizado, habilidades que podem ser fortalecidas a longo prazo, de um nível de proficiência para o outro.

No que tange à pergunta de pesquisa, “Em que medida atividades didáticas direcionadas por dados podem influenciar a aprendizagem de leitura em LI de estudantes de nível básico?”, pode-se dizer que as ferramentas utilizadas, *CorpusMate*, *Skell* e *VersaText* oferecem recursos de investigação da língua e de seus padrões, o que possibilita que os alunos ampliem seu vocabulário sem ter que lançar mão de uma ferramenta tradutória, por exemplo.

Além disso, a DDL proporciona ao aprendiz o contato com materiais autênticos de uso da língua. Com a curadoria adequada do material, as ferramentas potencializaram as estratégias de leitura ensinadas aos alunos que, mesmo com conhecimentos básicos de inglês, foram capazes de realizar os exercícios propostos e, de maneira guiada, exploraram amostras reais da linguagem. As atividades direcionadas por dados podem ainda estimular a autonomia na aprendizagem, característica requisitada aos estudantes de contexto acadêmico.

A eficácia da DDL no ensino superior já é amplamente conhecida (BOULTON; COBB, 2017; NESI, 2016). O que foi possível demonstrar neste artigo é que a DDL pode, sim, render bons resultados com alunos de nível básico, como observado na análise do questionário feito com os alunos, desde que sejam feitas as adaptações necessárias.

As atividades baseiam-se nos princípios de Mark e O’Keeffe (2022), pois foram elaboradas a partir de combinações de palavras, entendidas e tratadas enquanto termos multipalavras, padrões relevantes para esta pesquisa por permitirem a associação com o ensino das estratégias de leitura. O nível de proficiência dos estudantes foi respeitado por meio do uso de ferramentas de *corpora* próprias para o ensino e por meio da estruturação das tarefas a serem realizadas em etapas com instruções e direcionamentos da professora e da tutora.

O material desenvolvido está também fortemente ancorado nos critérios obrigatórios para elaboração de exercícios baseados em *corpora* de Delfino (2016) porque respeita a integridade do aluno e foi pensado para o seu nível de dificuldade, portanto é ético; baseia-se no padrão léxico-gramatical pois foi criado a partir dos termos multipalavras do *SDG-Plas*; e é reaplicável por outros professores por utilizar ferramentas de *corpora* para o ensino consideradas abertas, ou seja, gratuitas e de fácil acesso.

## REFERÊNCIAS

- ANTHONY, L. *Antconc (Version 3.4.4) [Computer Software]*. Tokyo: Waseda University, 2011.
- ANTHONY, L. *Introducing English for Specific Purposes*. London/Nova York: Routledge, 2018.
- ANTHONY, L. *AntCorGen (Version 1.2.0) [Computer Software]*. Tokyo: Waseda University, 2022.
- ANTHONY, L.; THOMPSON, P. *Automatic Creation and Discourse-Level Annotation of Individualized Discipline-Specific Corpora for the Data-Driven Learning (DDL) Classroom*. University of Birmingham, UK, 2017.
- BAFFI-BONVINO, M. A.; KANEKO-MARQUES, S. M. O processo de internacionalização e o ensino e aprendizagem de língua inglesa no contexto de uma universidade estadual. *Estudos Linguísticos*, São Paulo, v. 49, n. 1, p. 510-527, 2020.
- BAISA, V.; CROSTHWAIT, P. *Corpus Mate*, 2023. Disponível em: <https://corpusmate.com/>. Acesso em: 29 jun. 2023.
- BAISA, V.; SUCHOMEL, V. SKELL – Web Interface for English Language Learning. In: *Eighth Workshop on Recent Advances in Slavonic Natural Language Processing*. Brno: Tribun EU, 2014. p. 63-70.

- BALDO, A. Recursos de inferência lexical em L2. *Letras de Hoje*, Porto Alegre, v. 44, n. 3, p. 60-69, jul./set. 2009.
- BERBER SARDINHA, A. P. Linguística de *Corpus*: histórico e problemática. *Revista D.E.L.T.A.*, São Paulo, v. 16, n. 2, p. 323-367, 2000a.
- BERBER SARDINHA, A. P. Computador, *corpus* e concordância no ensino da léxico: gramática de língua estrangeira. In: LEFFA, V. (Org.). *As palavras e sua companhia: o léxico na aprendizagem*. Pelotas, RS: EDUCAT, 2000b. p. 45-75.
- BERBER SARDINHA, A. P. *Linguística de Corpus*. Barueri, SP: Manole, 2004.
- BERBER SARDINHA, A. P. Como encontrar as palavras-chave mais importantes de um *corpus* com Word Smith Tools. *Revista D.E.L.T.A.*, São Paulo, v. 21, n. 2, p. 237-250, 2005.
- BERBER SARDINHA, A. P. Como usar a Linguística de *Corpus* no ensino de língua estrangeira: por uma linguística de *corpus* educacional brasileira. In: VIANA, V.; TAGNIN, S. (Org.). *Corpora no ensino de línguas estrangeiras*. São Paulo: Hub Editorial, 2011. p. 301-356.
- BONDI, M.; SCOTT, M. *Keyness in Texts*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2010.
- BOULTON, A.; COBB, T. *Corpus* use in Language Learning: a meta-analysis. *Language Learning. Ann Arbor*, v. 67, n. 2, p. 348-396, 2017.
- BRITISH COUNCIL. British Council Brasil, 2023. *Quadro Comum Europeu de Referência para Línguas (QCER)*. Disponível em: <https://www.britishcouncil.org.br/quadro-comumeuropeu-de-referencia-para-linguas-cefr>. Acesso em: 25 set. 2023.
- CELANI, M. A. A. Considerações sobre a pesquisa "A necessidade e eficiência do ensino de inglês instrumental em Universidades brasileiras". *The ESPECIALIST*, v. 6, p. 2-9, 1983.
- CELANI, M. A. A. Introduction. In: CELANI, M. A. A. *et al.* (Org.). *ESP in Brazil: 25 years of evolution and reflection*. Campinas, SP: Mercado de Letras, 2005. p. 13-26.
- DELFINO, M. C. N. *O uso de música para o ensino de inglês como língua estrangeira em um ambiente baseado em corpus*. 2016. 159f. Dissertação (Mestrado em Linguística Aplicada e Estudos da Linguagem) – Pontifícia Universidade Católica de São Paulo, São Paulo, 2016.
- GARCIA, D. N. M.; PINTO, P. T.; SANTOS, D. C. Ações de internacionalização para o ensino e a pesquisa na área de línguas. *Fórum Linguístico*, Florianópolis, v. 18, n. 1, p. 5618-5630, jan./mar. 2021.
- HEEMANN, C. Inglês instrumental (ESP): o uso de estratégias de leitura em um curso *on-line* de leitura. *Signo*, Santa Cruz do Sul-RS, v. 34, n. 56, p. 137-156, jan. 2009.
- HUTCHINSON, T.; WATERS, A. *English for Specific Purposes: a learning – centred approach*. Cambridge, UK: Cambridge University Press, 1987.
- HYLAND, K. *English for academic purposes: an advanced resource book*. New York: Routledge, 2006.
- HYLAND, K. As can be seen: lexical bundles and disciplinary variation. *English for Specific Purposes*, v. 27, p. 4-21, 2008.
- HYLAND, K. English for Specific Purposes: What is it and where is it taking us?. *ESP Today*, v. 10, n. 2, p. 202-2020, 2022.
- JOHNS, T. F. Should you be persuaded: two samples of data-driven learning materials. *ELR Journal*, v. 4, p.1-16, 1988.
- KILGARRIFF, A. *et al.* Ten years on. *Lexicography*, v. 1, n. 1, p. 7-36, 2014.
- LEFFA, V. *Aspectos da leitura: uma perspectiva psicolinguística*. Porto Alegre: Sagra - DC. Luzzatto Editores, 1996.
- MARK, G.; O'KEEFFE, A. Principled pattern curation to guide data-driven learning design. *Applied Corpus Linguistics*, p. 1-15, 2022.
- NESI, H. *Corpus* studies in EAP. In: HYLAND, K.; SHAW, P. (Ed.). *The Routledge Handbook of English for Academic Purposes*. Abingdon, Oxfordshire, UK: Routledge, 2016. p. 206-217.
- O'BOYLE, A.; VIANA, V. *Corpus Linguistics for English for Academic Purposes*. London/New York: Routledge, 2022.
- PINTO, P. T. *et al.* Unveiling the academic vocabulary of sustainable development goal-research papers: a comprehensive lexical multidimensional analysis. *Revista Estudos da Linguagem* (no prelo).

- PINTO, P. T.; SANTOS, L. V. Avaliação de um curso on-line de estratégias de leitura em língua inglesa ministrado no início da pandemia. *EntreLinguas*, Araraquara-SP, v. 8, n. 00, p. 1-22, 2022.
- RAMISCH, C.; VILLAVICENCIO, A. Chutando o balde ou batendo as botas? Processamento de linguagem natural e expressões multipalavras na linguagem cotidiana e científica. In: LOPES PERNA, C. B.; DELGADO, H. O. K.; FINATTO, M. J. B. (Org.). *Linguagens especializadas em corpora: modos de dizer e interfaces de pesquisa*. Porto Alegre: ediPUCRS, 2010. p. 29-49.
- TAGNIN, S. E. O.; VIANA, V. *Corpora no ensino de línguas estrangeiras*. São Paulo: Hub Editorial, 2011.
- THOMAS, J. Two web-based tools for learning language from language. *Versatile*: 2020. Disponível em: [https://www.versatile.pub/uploads/8/1/6/3/81634112/two\\_webbased\\_tools\\_for\\_learning\\_language\\_from\\_language\\_thomas\\_2020\\_v2.pdf](https://www.versatile.pub/uploads/8/1/6/3/81634112/two_webbased_tools_for_learning_language_from_language_thomas_2020_v2.pdf). Acesso em: 15 set. 2023.
- TOMITCH, L. M. B. Aquisição de leitura em língua inglesa. In: LIMA, D. C. de. (Org.). *Ensino e aprendizagem de língua inglesa: conversas com especialistas*. São Paulo: Parábola Editorial, 2009.
- UNESP. Centro de Línguas e Desenvolvimento de Professores, 2021. Programa de Língua Inglesa da Unesp. Disponível em: <https://www.ibilce.unesp.br/#!/centro-de-linguas1/idiomas-sem-fronteiras/>. Acesso em: 27 set. 2023.
- UNESP. *Portifólio de Integração de Disciplinas da Pós-Graduação*, 2023. Disponível em: <https://www2.unesp.br/portal#!/propg/disciplinas-portfolio>. Acesso em: 18 nov. 2023.
- UNITED NATIONS. *The 17 goals*. Disponível em: <https://sdgs.un.org/goals>. Acesso em: 09 abr. 2023.
- VIANA, V. Linguística de *Corpus*: conceitos, técnicas & análises. In: VIANA, V.; TAGNIN, S. E. O. (Org.). *Corpora no ensino de línguas estrangeiras*. São Paulo: HUB Editorial, 2011. p. 25-95.

### **Contribuição dos autores.**

Luana Viana dos Santos coletou e descreveu os dados apresentados e Paula Tavares Pinto revisou o texto, assim, as autoras trabalharam colaborativamente na escrita do artigo.