

Grupo de pesquisa En-Corpora: ensino baseado e dirigido por *corpora***En-Corpora research group: *corpus*-based and *corpus*-driven teaching**

Silmara Moscatelli¹
Fatec/ Presidente Prudente

Paula Tavares Pinto²
UNESP/São José do Rio Preto

Carolina Tavares de Carvalho³
UNESP/Ibilce

Luciano Franco da Silva⁴
Instituto Federal do Paraná

Resumo

O objetivo deste artigo é apresentar alguns resultados sobre as discussões e estudos realizados pelo Grupo de Pesquisa En-*Corpora*: Estudos Baseados e Dirigidos por *Corpora* (CNPq/UNESP), que conta com o apoio financeiro da Coordenação de Aperfeiçoamento de Pessoal e Nível Superior (CAPES). Como fundamentação teórica, tomamos como base a abordagem da Linguística de *Corpus* e da Aprendizagem Moviada por Dados (DDL), pautados nos trabalhos de autores como Berber Sardinha (2004), Tognini-Bonelli (2001) e Gilquin e Granger (2010). Os resultados obtidos fornecem possibilidades para que professores e pesquisadores da área da Linguística de *Corpus*, com a abordagem em DDL, possam produzir suas atividades didáticas de maneira mais adequada a contextos de ensino-aprendizagem de línguas.

¹ Docente da Faculdade de Tecnologia do Estado de São Paulo (Fatec), campus Presidente Prudente, doutorado em Estudos Linguísticos pela Universidade Estadual Paulista "Júlio de Mesquita Filho" (UNESP-Ibilce) (2023), e mestre (2015) pelo Proletras, câmpus de Assis, área de Conhecimento: Linguagens e Letramentos. Membro do Grupo de Pesquisa En-Corpora (UNESP-Ibilce). Orcid: <https://orcid.org/0000-0002-3899-7614>.

² Mestre e Doutora em Estudos Linguísticos pela Universidade Estadual Paulista Júlio de Mesquita Filho, com estágio na Universidade de Manchester e pós-doutorado na Universidade de Surrey, Inglaterra. Formada em Tradução (1992) e em Letras. Docente da UNESP, Câmpus de São José do Rio Preto. Coordenadora geral do Programa Idiomas sem Fronteiras e do projeto de English Teaching Assistants (CAPES/Fulbright) na UNESP. Atua no Programa de Pós-Graduação em Estudos Linguísticos. ORCID: <https://orcid.org/0000-0001-9783-2724>

³ Doutoranda em linguística aplicada pela Universidade Estadual Paulista Júlio de Mesquita Filho - UNESP/Ibilce. Mestre na mesma instituição em Pedagogia. Membro do Grupo de Pesquisa Tradução, Terminologia e Corpora e En-Corpora: Ensino Baseado e Orientado por Corpora (UNESP-Ibilce), e do projeto: Os Objetivos de Desenvolvimento Sustentável (ODS) das Nações Unidas em a pesquisa brasileira: uma abordagem baseada em corpus para apoiar a redação e tradução de artigos de pesquisa.

⁴ Docente EBT/T do Instituto Federal do Paraná (IFPR), câmpus Goioerê e doutorando em Estudos Linguísticos pela Universidade Estadual Paulista "Júlio de Mesquita Filho" (UNESP-Ibilce), e mestre (2018) pela mesma instituição, na linha de pesquisa Pedagogia do Léxico e da Tradução Baseada em Corpora. É membro do Grupo de Pesquisa En-Corpora (UNESP-Ibilce).

Palavras-chave: Grupo de pesquisa. Ensino de línguas. Linguística de *Corpus*. Aprendizagem Movida por Dados

Abstract

This article aims to present some results from the discussions and studies conducted by the Research Group En-Corpora: Corpus-Based and Corpus-Driven Studies (CNPq/UNESP), with financial support from the by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior– Brasil (CAPES). The theoretical framework is based on the approach of Corpus Linguistics and Data-Driven Learning (DDL), drawing on the works of authors such as Berber Sardinha (2004), Tognini-Bonelli (2001), and Gilquin and Granger (2010). The results highlight the role of the DDL approach in improving didactic activities, providing opportunities for teachers and researchers in Corpus Linguistics to develop activities better suited to language teaching and learning contexts.

Keywords: Research group. Language teaching. *Corpus* linguistics. Data-driven learning

Introdução

O ensino de línguas estrangeiras está em contínua transformação, com diversas abordagens e metodologias sendo discutidas. Refletir sobre mecanismos de apoio ao ensino de línguas estrangeiras e a incorporação de novas tecnologias são pontos essenciais para promover a autonomia do estudante no processo de ensino-aprendizagem. Dessa forma, a Linguística de *Corpus* (LC) se apresenta como uma alternativa viável, uma vez que trabalha amostras autênticas de registros linguísticos que estudantes precisam aprender para uma comunicação efetiva, seja para o mercado de trabalho, seja para o ensino de línguas para fins gerais. Além disso, a LC utiliza ferramentas computacionais capazes de gerar dados detalhados sobre padrões lexicogramaticais em diferentes registros, a partir de linhas de concordâncias, frequências de determinadas classes gramaticais, entre outros.

Berber Sardinha *et al.* (2012) enfatiza a importância de acessar e interagir com os textos que circulam nas mídias digitais, bem como integrá-los ao contexto de sala de aula (SA) de língua estrangeira. Para os autores, a junção entre o computador e a Internet é imprescindível para os professores de língua estrangeira, pois:

finalmente, o computador e a Internet são de inestimável valor porque a maioria dos professores de língua estrangeira teve sua formação nas Humanidades. É natural que esses profissionais escolham atividades para a sala de aula que coloquem para trabalhar o lado direito do cérebro, o lado responsável pela criatividade e pelos aspectos estéticos da vida. Entretanto, o aprendizado de língua se dá no lado esquerdo – o lado da lógica, da resolução de problemas, o lado que faz correlações entre as partes e o todo. Este é o lado que, em 96% dos seres humanos, controla a aquisição da língua materna e de outras línguas. Parece haver, portanto, um descompasso entre o que acreditamos ser a “receita” certa para uma boa prática pedagógica (o criativo e o lúdico) e o que realmente vai ajudar nossos aprendizes a aprenderem e reterem a língua estrangeira (o lógico) [...] é a

utilização das novas tecnologias na sala de aula de língua estrangeira. É justamente aí que os padrões típicos de aprendizagem (a resolução de problemas, a independência do aprendiz) se associam às formas típicas de comunicação do terceiro milênio para implementar as rotinas de sala de aula – a união do criativo e do lúdico ao lógico, do agradável e motivador ao eficaz e duradouro (Berber Sardinha *et al.*, 2012, p. 7).

Observa-se, portanto, a necessidade de trabalhar com textos autênticos⁵ para facilitar a compreensão do idioma por parte dos alunos. Ademais, por meio da abordagem orientada por dados, os discentes podem desenvolver papel de pesquisadores, uma vez que as atividades estimulam a busca de novas fontes de conhecimento, promovem o desenvolvimento da autonomia e o do pensamento crítico, além de oferecerem acesso a um vasto banco de textos oriundos de diversas fontes (Berber Sardinha, 2011).

A LC pode ser aplicada tanto no ensino de línguas gerais quanto para fins específicos. No grupo de pesquisa En-Corpora: Ensino Baseado e Dirigido por *Corpora*, os estudos se concentram em discutir o ensino de línguas com o suporte das ferramentas da LC (Carvalho, 2021; Garcia, 2020; Moscatelli, 2023; Nazzi-Laranja, 2020; Santos e Pinto, no prelo). Adicionalmente, destaca-se a publicação do Caderno de Atividades de Aprendizagem Movida por Dados e sua versão traduzida para o inglês (Pinto *et al.* 2022; 2023), elaboradas pelo grupo de pesquisa, que oferecem propostas didáticas para o ensino de inglês e espanhol. Esse material foi concebido para ser utilizado como recurso digital em sala de aula, tanto com acesso à internet (*hands on*) quanto sem ele (*hands off*). Nesse contexto, os alunos são incentivados a assumir o papel de “detetives” e protagonistas de sua própria aprendizagem, ao explorarem coletâneas de textos autênticos, os *corpora* digitais, que permitem a investigação aprofundada da língua em uso. Portanto, o presente artigo tem como objetivo detalhar a teoria e a metodologia adotadas pelo grupo de estudos na elaboração de atividades de Aprendizagem Movida por Dados ou *Data Driven Learning* (DDL).

Linguística de *Corpus*

A LC é uma área da Linguística centrada na coleta e análise de dados a partir de uma abordagem empirista. Baseia-se em dados concretos, em experiências reais, observando a frequência com que determinadas construções linguísticas são empregadas em contextos específicos (Berber Sardinha, 2000).

Para compreendermos as contribuições da LC, tanto para a análise linguística quanto para o ensino de línguas, é essencial, em primeiro lugar, entender o significado do termo *corpus* dentro do amplo campo da Linguística. Em termos abrangentes, um *corpus* é uma coletânea de textos, compilados em formato eletrônico, a partir de critérios específicos, para servirem como objeto de estudos linguísticos (Tagnin, 2011). Todavia, é preciso destacar que nem toda coleção de textos pode ser considerada um *corpus*, visto que diversos critérios devem ser respeitados em sua compilação (Friginal, 2018). Nesse artigo,

⁵ Neste artigo entendemos que textos autênticos são aqueles provenientes de comunicações genuínas no mundo real, ou seja, textos que não foram criados com a intenção de serem utilizados para o ensino.

utilizaremos o conceito de *corpus* com base em Sinclair (2005), cuja definição delimita bem esse conceito. Segundo o autor:

um *corpus* é uma coleção de porções de linguagem em formato de textos eletrônicos, selecionados de acordo com critérios externos para representar, o máximo possível, uma língua ou variedades de uma língua como fonte de dados para a pesquisa linguística. (Sinclair, 2005, n.p., tradução nossa)⁶

É importante destacar que as porções de linguagem mencionadas pelo autor devem ser extraídas de textos autênticos, ou seja, textos que não foram produzidos com o intuito de integrarem o *corpus*. Em relação aos critérios para a escolha dos textos, Berber Sardinha (2004) e Sinclair (2005) elencam alguns dos mais comuns utilizados nas pesquisas, são eles: i) modo (oral ou escrito); ii) composição (jornais, revistas, palestras, artigos científicos); iii) origem e data; iv) extensão e representatividade⁷.

Complementando as observações anteriores, McEnery e Xiao (2011) caracterizam a LC como uma abordagem empirista, que analisa a língua como um sistema probabilístico, a partir da exploração sistemática de um *corpus*. Dado que essas pesquisas buscam descrever a língua em uso, muitos pesquisadores também utilizam os dados extraídos dos *corpora* para conduzir análises críticas de currículos e materiais didáticos de línguas estrangeiras.

Com base nessas questões, acreditamos ser importante uma breve explicação sobre dois tipos de investigações linguísticas no campo da LC, as chamadas investigações “baseadas em *corpus*”⁸ e “dirigidas por *corpus*”⁹. Tognini-Bonelli (2001) explica que as investigações baseadas em *corpus* são utilizadas para testar ou exemplificar uma teoria, formulada previamente à consulta dos dados, ou seja, tais estudos são realizados para validarem as intuições dos pesquisadores acerca do uso da língua. Nesta linha de pensamento, “o papel do *corpus* é o de ser um depósito de exemplos para ilustrar uma teoria ou conceitos previamente estabelecidos” (Berber Sardinha, 2002, p. 33). Portanto, podemos entender que as pesquisas baseadas em *corpus* têm um caráter descritivo da língua, ao invés de prescritivo.

Por outro lado, Tognini-Bonelli (2001) esclarece que, nas investigações dirigidas por *corpus*, os pesquisadores analisam primeiro os dados linguísticos disponíveis para depois proporem teorias e explicações acerca dos padrões linguísticos que surgirem. Nessa abordagem, o status teórico da LC é caracterizado por sua natureza indutiva, no qual o pesquisador é instigado a fazer o mínimo de suposições acerca do funcionamento da língua antes do levantamento de dados empíricos. Logo, os padrões linguísticos mais recorrentes

⁶ Original: “a corpus is a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as source of data for linguistic research”.

⁷ Berber Sardinha (2004) afirma que a representatividade de um *corpus* comporta três dimensões: i) número de palavras (quanto maior o número de palavras, mais chances ele tem de possuir palavras de baixa frequência, que são a maioria das palavras de uma língua); ii) número de textos da qual o *corpus* foi compilado (quanto maior o número de textos, melhor um *corpus* é representado); e iii) números de gêneros, registros ou tipos textuais presentes no *corpus*, em especial para a compilação de *corpus* que visam representar a língua como um todo.

⁸ *Corpus-based investigations*

⁹ *Corpus-driven investigations*

em um *corpus* devem ser a base para a elaboração de uma descrição linguística que reflita e explique tais evidências.

Para a área de línguas para fins acadêmicos (LinFE) e profissionais, abordagens baseadas em *corpora* têm ganhado grande destaque (Frankenberg-Garcia *et al.*, 2022; McEnery, Xiao, 2011), uma vez que essas pesquisas proporcionam a descrição e a análise do discurso acadêmico. Conforme Flowerdew (1994), os *corpora* acadêmicos podem direcionar professores e pesquisadores sobre as características linguísticas e discursivas com as quais os alunos precisam estar familiarizados para otimizar a compreensão de gêneros específicos da área, portanto, podem auxiliar a evidenciar os conteúdos que devem ser incorporados ao curso de LinFE.

Em vista disso, Charles (2012) destaca três usos de *corpora* para o ensino Inglês para fins acadêmicos (IFA) e Inglês para fins profissionais (IFP), a saber: i) repetição de padrões linguísticos dentro de uma quantidade massiva de dados, revelando, assim, evidências objetivas, e não mais subjetivas, na análise linguística; descrição de características específicas do discurso acadêmico, assim como a fraseologia de diferentes disciplinas e gêneros acadêmicos; e, por último, iii) compilação de listas de palavras acadêmicas¹⁰, tais como o trabalho de Coxhead e Byrd (2007). Complementando os usos de *corpora* mencionados acima, Cheng (2010) e Friginal (2018) apresentam o uso de *corpora* na produção de materiais voltados ao IFA, visto que, para o contexto acadêmico, eles são mais especializados em termos de tópico e registro. Além disso, pesquisas com base em *corpus* podem oferecer recursos mais realistas, ilustrativos e atuais para a criação de materiais de ensino (Cheng, 2010), ideia que também é defendida por McCarthy (2001, *apud* O'Keeffe; McCarthy; Carter, 2007, p. 21, tradução nossa)¹¹ o afirmar que:

a linguística de *corpus* representa uma mudança pioneira nas técnicas e métodos científicos e, provavelmente, antecipa mudanças tecnológicas que mudarão nossa noção de educação, papel do professor e contexto cultural dos serviços de educação e mediação de teorias e técnicas.

Pelo fato de a LC ser uma abordagem empirista, sua principal contribuição para o ensino de LE é o de demonstrar, por meio de amostragens de usos autênticos da língua, características e padrões que não são percebidos somente pela intuição do pesquisador. Nessa mesma linha de pensamento, O'Keeffe, McCarthy e Carter (2007) e Walsh (2010) apresentam uma contradição em relação à preparação de materiais didáticos de LE, afirmando que, em sua maioria, eles são baseados na intuição de como a língua é utilizada, ao invés de apresentar evidências reais de seu uso. Como solução para esse problema, os pesquisadores defendem a ideia de que materiais com base em *corpora* devem colocar o aluno em contato direto com a língua autêntica (McEnery; Xiao, 2011).

¹⁰ Informações disponíveis em: <http://www.victoria.ac.nz/lals/resources/academicwordlist/>. Acesso em: 19 ago. 2023.

¹¹Original: “Corpus linguistics represents cutting-edge change in terms of scientific techniques and methods and probably foreshadows even more profound technological shifts that will “impinge upon our long-held notions of education, roles of teachers, the cultural context of the delivery of educational services and the mediation of theory and technique.

A partir desta introdução sobre o conceito de *corpus* e sua aplicabilidade no ensino de línguas, apresentaremos, a seguir, algumas das principais abordagens de ensino de LE que utilizam *corpora* em SA.

Aprendizagem direcionada por dados

A DDL¹², descrita no trabalho de Johns (1991), é uma abordagem que consiste no uso das ferramentas da LC para propósitos pedagógicos. Segundo Chambers (2010), um dos diferenciais no uso da DDL é o alto grau de interação entre alunos, professores e *corpora*, algo que não acontece em outras abordagens de ensino de línguas.

A aprendizagem indutiva é a essência da DDL (Chambers 2010; Gilquin; Granger, 2010), ou seja, o processo de descoberta de padrões linguísticos por parte do aluno faz com que o aprendizado seja motivador e interessante. Nesse contexto investigativo de ensino, espera-se que o aluno assuma o papel de um " Sherlock Holmes", atuando como um investigador, semelhante ao famoso detetive, conforme argumentam Gilquin e Granger (2010) e Chambers (2010). Na abordagem DDL, os alunos são incentivados a observar os dados, formular hipóteses e estabelecer regras sobre os padrões linguísticos identificados no *corpus* em estudo (abordagem indutiva). Além disso, eles podem verificar a validade das regras gramaticais presentes nos livros didáticos (abordagem dedutiva). Dessa forma, eles se tornam mais envolvidos, ativos e autônomos no processo de aprendizagem. Com a DDL, Johns (1994) concorda, também, com a mudança na função do professor, que passaria a ser um mediador, assumindo um papel secundário na sala de aula. O educador, portanto, passa a ser considerado um orientador de evidências que auxilia os alunos a responderem suas próprias questões. Todavia, a abordagem DDL, assim como qualquer outra, também recebeu suas críticas, em especial, Gilquin e Granger (2010) destacam quatro aspectos da DDL que podem atrapalhar a sua utilização em SA: a logística; as crenças do professor; as crenças do aluno; e o conteúdo apresentado.

Sobre a crença dos professores não utilizarem a DDL em suas aulas, Gilquin e Granger (2010) explicam que tal situação decorre muitas vezes da falta de familiaridade com os recursos que a LC oferece, pois, segundo Mauranen (2004), antes de os alunos serem apresentados aos possíveis usos de *corpora*, é importante que o professor os conheça primeiro. Entretanto, Boulton (2009) e Gilquin e Granger (2010) argumentam que o problema pode ser mais complexo do que a simples falta de conhecimento técnico, pois na abordagem DDL a principal fonte de conhecimento tem origem na análise do *corpus*, o que pode fazer com que o professor a "perca o controle" sobre os dados e informações que alunos irão encontrar durante as aulas, podendo causar assim uma impressão de falta de conhecimento ou falta de preparação por parte do professor.

Já em relação à crença dos alunos sobre a DDL, Gilquin e Granger (2010) explicam que há relatos de que essa abordagem pode ser desanimadora, pois atividades envolvendo *corpora* nem sempre são diretas e necessitam de treino e prática para um uso efetivo. Além disso, dentre os principais problemas reportados em pesquisas empíricas envolvendo alunos e *corpora*, é possível citar: dificuldade de enxergar padrões linguísticos, dificuldade

¹² *Data-Driven Learning* (Aprendizagem dirigida por dados)

para realizar buscas dentro do *corpus* e a possibilidade de fazer inferências erradas a partir dos dados linguísticos encontrados.

O último problema apresentado por Gilquin e Granger (2010) é a forma como o conteúdo é apresentado. Segundo as autoras, essa discussão se dá a partir de críticas lançadas em relação à autenticidade de dados linguísticos provenientes de *corpora*, além disso, as autoras alertam que é preciso muito cuidado para que os alunos não recebam informações demais ou informação nenhuma ao pesquisarem em *corpora*, problema que também foi levantado por outros estudos envolvendo a DDL. Em relação a isso, Kennedy e Miceli (2001) alertam para a importância de os alunos aprenderem a distinguir se uma determinada ocorrência não existe ou se ela simplesmente não é representativa o suficiente dentro do *corpus* de estudo.

Apesar de todos os desafios apresentados, Gilquin e Granger (2010, p. 367, tradução nossa)¹³ defendem a implementação da DDL no ensino de línguas. Para as autoras, a abordagem é promissora e

coloca o aluno em contato com a língua (potencialmente) autêntica, motiva os alunos por meio de descobertas, desenvolve habilidades cognitivas importantes e, de maneira geral, fornece benefícios que vão além de conhecimento linguístico do item estudado.

Em relação à logística, as autoras explicam que esse é o problema mais comum na DDL, pois é esperado que os alunos tenham contato direto com as ferramentas e softwares para a análise de *corpora*. Embora existam hoje alguns *softwares* de análise linguística e *corpora* gratuitos online, ainda é necessário o uso de computadores, preferencialmente um ou dois alunos por computador, e tudo isso gera despesas com equipamentos, mão de obra e manutenção que nem sempre as instituições de ensino estão dispostas ou têm condições de manter.

Como solução para esse problema, muitos professores optam por preparar e levar os próprios materiais e/ ou atividades impressas para a SA. Entretanto, Chambers (2010) argumenta que tudo isso é muito trabalhoso e demanda muito tempo tanto do professor quanto do aluno. Gilquin e Granger (2010) também apresentam o consumo de tempo como um obstáculo na implementação da DDL, explicando que, além do período gasto na preparação de materiais, é necessário tempo para treinar o aluno a lidar com *corpora*, e, mais importante ainda, é necessário que o aluno tenha tempo hábil em SA para a realização das tarefas propostas.

Apesar do risco mencionado, Chambers (2010) argumenta que o professor em uma SA não deve ser a principal fonte de conhecimento linguístico. Em vez disso, sua função é atuar como facilitador do processo de aprendizagem, auxiliando os alunos na interpretação dos dados, orientando-os sobre como realizar buscas no *corpus* de maneira eficiente e analisando melhor os resultados obtidos.

¹³ Original: “brings learners into contact with (potentially) authentic language, motivates them by introducing an element of discovery, develops important cognitive skills and, more generally, provides benefits which go well beyond the knowledge of the item under study”.

Complementando o argumento acima, Kennedy e Miceli (2001) sugerem quatro passos que o professor deve seguir ao colocar o aluno em contato com *corpora*, são eles: 1) Formulação das perguntas; 2) Desenvolvimento de estratégias de busca; 3) Observação dos exemplos mais relevantes e 4) Criação de hipóteses. No Quadro 1 a seguir, apresentamos algumas dicas de como os *corpora* podem ser utilizados de maneira mais eficaz na SA, de acordo com Kennedy e Miceli (2001, p. 82-87).

Quadro 1: Sugestões e cuidados no uso de *corpora* em sala de aula

| Passos na investigação com <i>corpora</i> | Dicas |
|--|---|
| #1 - Formulação das perguntas | Tente ser o mais preciso possível no enunciado da questão; Certifique-se de que a questão está de acordo com o contexto da sala/aluno; Considere utilizar questões abertas, por exemplo, ao invés de perguntar “A preposição x aparece depois do verbo y?”, formule a questão da seguinte maneira: “O que aparece depois do verbo y?”; Atente-se tanto para possíveis problemas gramaticais quanto lexicais. |
| #2 - Desenvolvimento de estratégias de busca | Refleta sobre a eficiência da estratégia, escolha por aquela que irá gerar resultados com ocorrências mais relevantes para os alunos; Certifique-se de que o item de busca seja uma parte fixa da palavra/expressão, visto que há muitas colocações formadas por uma parte fixa e outra variável. |
| #3 - Observação dos exemplos mais relevantes | Lembre-se de checar os significados das palavras que serão usadas como evidências, e escolha aquelas que mais se aproximam do significado esperado; Tente não ser influenciado por suposições, observe tanto à esquerda quanto à direita da palavra-chave e veja quais palavras estão mais associadas a ela; Apesar do número de ocorrências ser um aspecto importante, é preciso lembrar que às vezes o melhor resultado pode não ser aquele mais frequente. |
| #4 - Criação de hipóteses | Mesmo que apareça apenas uma ocorrência daquilo que está sendo procurado, pode ser o suficiente para a criação de uma hipótese. Lembre-se de que o importante é a qualidade das evidências, e não a quantidade delas; Caso os exemplos apresentados pelo <i>corpus</i> não sejam os esperados, é necessário refletir sobre o porquê de isso ter acontecido e se isso pode afetar os resultados esperados; Caso você não encontre ocorrências do que está sendo procurado, reflita com cuidado antes de tirar suas conclusões, certifique-se de que sua resposta esteja de acordo com o que foi proposto pela atividade. |

Fonte: Adaptado de Kennedy e Miceli (2001, p. 82-87).

As autoras explicam que as sugestões apresentadas acima visam a ajudar alunos e professores a lidarem melhor com problemas que não são oriundos de conhecimento linguístico, mas sim de uma habilidade de investigação com *corpora* inadequado. Conforme explicado anteriormente, parte das críticas em relação à DDL está mais associadas a fatores extralinguísticos do que no nível de proficiência¹⁴ que o aluno precisa ter antes de utilizar *corpora* em SA.

Para Johns (1994), observar e analisar as linhas de concordância têm um impacto positivo sobre a autonomia do aluno no processo de ensino e aprendizagem de uma segunda língua. Ao investigar essas concordâncias, os aprendizes têm acesso a um *corpus* de língua autêntica, e não só a exemplos de usos pré-fabricados pelo professor e/ou pelo material didático. Assim, a fonte de conhecimento vem do *corpus*, e não do educador. Nos últimos anos, temos visto alguns autores questionando se o uso da DDL em SA ainda é relevante, como indagado por Crosthwaite e Boulton (2022), a resposta é afirmativa e evidenciada no volume organizado por Crosthwaite (2024) com mais de vinte autores discutindo o uso da LC em SA e a importância do DDL para a motivação da investigação linguística por parte dos alunos. De acordo com Berber Sardinha (2004), a DDL busca tornar o aprendiz um pesquisador. O autor também reforça o aspecto indutivo da vertente, pois o estudante inicia a aprendizagem a partir da observação de dados de uso autêntico da língua. Essa análise dos dados pode ser realizada com base nas linhas de concordância, ou seja, linhas do texto que apresentam a palavra pesquisada em seu centro, bem como parte do texto à sua volta.

Critérios para a preparação das atividades

Antes de elaborarmos as atividades, pensamos em relação à tipologia do *corpus*. Berber Sardinha (2004) proporciona um esquema com os principais critérios quanto à definição e objetivo do *corpus* para elaborarmos as atividades compostas na SD. Nossa compilação apresenta as seguintes definições:

Quadro 2: Critérios quanto à definição e objetivo do *corpus*

| | | |
|---|------------|---|
| 1 | Modo | Escrito |
| 2 | Tempo | Contemporâneo |
| 3 | Seleção | Amostragem |
| 4 | Conteúdo | Especializado |
| 5 | Autoria | Língua nativa |
| 6 | Finalidade | Atividades de estudo com foco na compreensão leitora e escritora. |

Fonte: Berber Sardinha (2004, p. 20).

¹⁴ É importante ressaltar que Kennedy e Miceli (2001) reconhecem sim a importância de um nível adequado de proficiência linguística por parte do aluno, em especial nos passos #1 e #3. Entretanto, as autoras defendem que falhas na preparação e aplicação de atividades DDL podem causar a impressão de que a investigação em *corpora* é muito mais complicada do que realmente deveria ser.

Além de definir a tipologia do *corpus*, utilizamos os critérios obrigatórios (O) e não-obrigatórios (NO), estabelecidos por Delfino (2016), como métodos para a preparação de atividades baseadas em *corpus*. Vale ressaltar que as atividades com base em *corpora* devem estimular o aluno a descobrir evidências linguísticas com base em padrões e em frequência dos dados. O professor é facilitador, e não distribuidor do conhecimento. Para cumprir esse critério, é necessário que o professor compartilhe a responsabilidade de aprendizagem com seus alunos, apresentando a eles caminhos para que façam suas próprias conclusões sobre o uso da língua. Delfino (2016) lembra que a atenção do professor não deve estar somente no ensino do conteúdo, mas, sim, em estimular o aluno a buscar novas fontes de conhecimento. Os critérios obrigatórios dessa preparação e suas características gerais estão reunidos no quadro a seguir.

Quadro 3: Atividades baseadas em *corpus*

| |
|--|
| 01 - O exercício faz uso de <i>corpus</i> . |
| 02 - O exercício precisa ter enunciados claros. |
| 03 - O exercício tem como foco principal o padrão lexicogramatical. |
| 04 - O exercício é ético. |
| 05 - O exercício é replicável. |
| 06 - O exercício é motivador. |
| 07 - O exercício não simplifica a língua usada nos textos/concordâncias/listas de palavras, etc. |
| 08 - O exercício deve apresentar nível de dificuldade adequado. |
| 09 - O exercício contém conteúdo relevante para o aluno e para a construção do conhecimento na língua estrangeira. |
| 10 - O professor é facilitador e não distribuidor de conhecimento. |
| 11 - O aluno é descobridor, pesquisador e não recipiente de conhecimento. |

Fonte: Delfino (2016, p. 56).

A partir dos critérios obrigatórios para a criação de atividades com base em *corpora*, descritos acima, apresentamos os critérios não obrigatórios (NO), isto é, aqueles que não precisam ser necessariamente seguidos pelo professor no momento da confecção de atividades com *corpora* (Delfino, 2016). Entretanto, muitos desses critérios são relevantes para a criação da atividade, como, por exemplo, pensar se o exercício é participativo. Isso significa que a atividade em questão deve estimular, sempre que possível, a participação do aluno. As atividades desenvolvidas proporcionam que a aula seja dialogada, debatida e discutida, tanto por meio da valorização dos conhecimentos que os alunos já trazem de suas vivências, seus conhecimentos prévios, quanto à medida que vão identificando a frequência, as linhas de concordâncias e os contextos em que eles aparecem.

Se o exercício é de fácil adaptação, a atividade pode ser reproduzida para outras turmas e adaptada para alunos com diferentes níveis de proficiência linguística, facilitando, assim, o trabalho do professor. O quadro a seguir apresenta as características gerais para tais critérios.

Quadro 4: Lista dos critérios não obrigatórios para a preparação de atividades

| |
|------------------------------------|
| NO1 – O exercício é participativo; |
| NO2 – O exercício é colaborativo; |

| |
|---|
| NO3 – O exercício exige pouco tempo de preparação por parte do professor; |
| NO4 – O exercício trabalha com o conceito de padronização; |
| NO5 – O exercício lida com o conceito de frequência; |
| NO6 – O exercício trabalha com o conceito de variação; |
| NO7 – O exercício trabalha com o conceito de variedade textual; |
| NO8 – O exercício incorpora mídias diferentes; |
| NO9 – O exercício incorpora linhas de concordância; |
| NO10 – O exercício incorpora textos; |
| NO11 – O exercício incorpora lista de frequência de palavras; |
| NO12 – O exercício incorpora lista de palavras-chave; |
| NO13 – O exercício faz com que os alunos trabalhem diretamente com <i>corpora</i> ; |
| NO14 – O exercício incorpora diagramas e diferentes formas de visualização; |
| NO15 – O exercício é de fácil adaptação; |
| NO16 – O exercício desenvolve a autonomia; |
| NO17 – O exercício ensina e não testa. |

Fonte: Delfino (2016, p. 56).

Análise e discussões de atividades com LC

Nesta seção apresentaremos uma atividade que foi proposta no livro do grupo de pesquisa Caderno de Atividades de Aprendizagem Movida por Dados (Pinto; Garcia; Serpa, 2022). A lição intitulada *¿Cuándo me falte fuerza, resistiré!* foi elaborada para o ensino de língua espanhola a nível do Ensino Médio – 1º, 2º e 3º anos.

O objetivo dessa atividade didática é o de observar o uso das orações temporais com a seguinte estrutura: CUÁNDO + PRESENTE DO SUBJUNTIVO/FUTURO DO INDICATIVO, são ações que não sabemos se de fato ocorrerão, por isso o uso desses modos e tempos verbais. Essa lição está disponível em dois modelos: um *hands on* e outro *hands off*. A diferença entre essas duas versões é que a primeira exige que os estudantes acessem a ferramenta *Voyant Tools* para descobrir quais os termos mais frequentes na música e, a partir disso, realizem os exercícios; enquanto a segunda não requer esse acesso, uma vez que os dados já foram previamente selecionados e disponibilizados na própria lição.

Com relação à ferramenta de análise de *corpus Voyant Tools*, espera-se que os alunos a usem para que consigam determinar as combinações que podem ser feitas com o pronome complemento ME e a utilização dos verbos no PRESENTE DO SUBJUNTIVO + FUTURO DO INDICATIVO. Dessa forma, apresentamos abaixo a sequência de atividade, que teve como ponto gerador a música: *Resistiré* como nosso *corpus*.

Quadro 5: Lição: *¿Cuándo me falte fuerza, resistiré!*

Atividades 1: Em pequenos grupos, converse a respeito da música que tem o título “Resistiré”. Vocês conhecem? Qual o tema que irá tratar? Em que época foi feita? Anote abaixo seus palpites.

Atividade 2: Assista ao vídeo da música, regravada em 2020, acompanhando a leitura:
Cuando pierda todas las partidas
Cuando duerma con la soledad

Cuando se me cierran las salidas
Y la noche no me deje en paz
Cuando sienta miedo del silencio
Cuando cueste mantenerse en pie
Cuando se rebelen los recuerdos
Y me pongan contra la pared
Resistiré, erguido frente a todo
Me volveré de hierro para endurecer la piel
Y aunque los vientos de la vida soplen fuerte
Soy como el junco que se dobla,
Pero siempre sigue en pie
Resistiré, para seguir viviendo
Soportaré los golpes y jamás me rendiré
Y aunque los sueños se me rompan en pedazos
Resistiré, resistiré.
Cuando el mundo pierda toda magia
Cuando mi enemigo sea yo
Cuando me apuñale la nostalgia
Y no reconozca ni mi voz
Cuando me amenace la locura
Cuando en mi moneda salga cruz
Cuando el diablo pase la factura
O Si alguna vez me faltas tu
Resistiré erguido frente a todo,
Me volveré de hierro para endurecer la piel,
Y aunque los vientos de la vida soplen fuerte,
Soy como el junco que se dobla
Pero siempre sigue en pie'
Resistiré para seguir viviendo,
Soportare los golpes
Y jamás me rendiré,
Y aunque los sueños se me rompan en pedazos'
Resistiré, resistirééé'
O que mais chamou sua atenção? Qual mensagem a música passa?

Fonte: Moscatelli (2023).

A lição inicia com o levantamento do conhecimento prévio dos alunos sobre a temática da música *Resistiré*. Em seguida, é realizada a projeção do vídeo da música, contextualizando sua gravação e regravações, permitindo que os alunos compreendam os diferentes contextos históricos e sociais em que a canção foi utilizada como símbolo de resistência. Após essa introdução, os alunos são divididos em duplas para anotarem suas conclusões sobre a temática e os contextos da canção, compartilhando posteriormente suas observações com a turma.

A próxima etapa envolve a introdução das ferramentas digitais *Voyant Tools*, especificamente as funcionalidades *Word Tree*, *Mandala*, *Cirus*, *Terms Berry* e *Contexts*. Cada uma dessas ferramentas oferece uma abordagem distinta para a análise do texto da música,

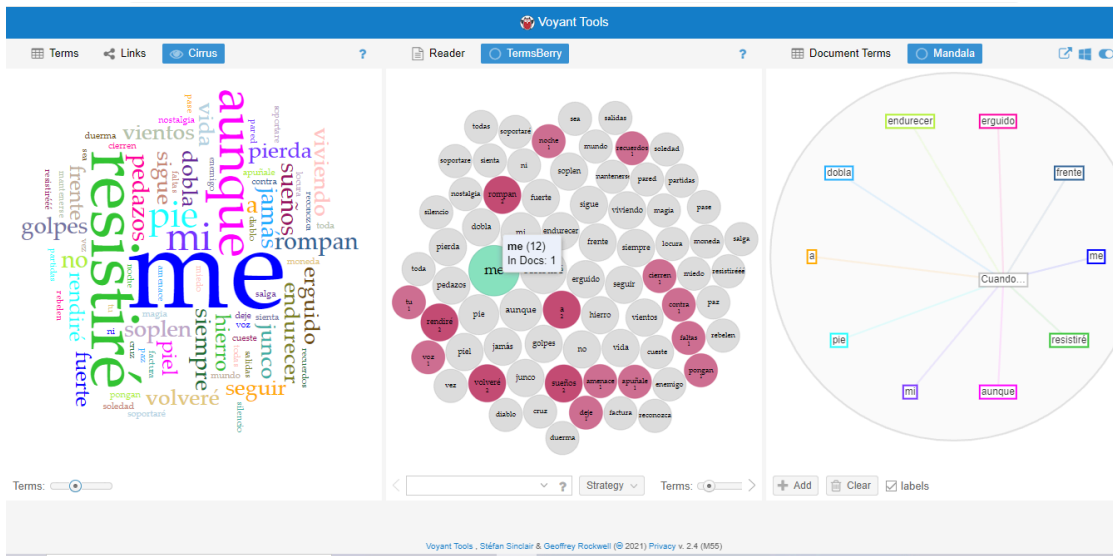
proporcionando uma compreensão mais aprofundada dos padrões linguísticos e do uso da língua.

Na fase de desenvolvimento da lição, os alunos utilizam essas ferramentas em grupos para investigar os tempos e modos verbais presentes na música, focando especialmente no uso do presente do subjuntivo e do futuro, além do uso do pronome complemento com os verbos. Eles discutem como esses modos verbais são utilizados para expressar necessidades e mensagens de resistência na música. Para aplicar o conhecimento adquirido, os alunos elaboram cartazes com frases utilizando a estrutura gramatical aprendida, transmitindo mensagens de resiliência com o tema “5 cosas que haré cuándo los planes no salgan bien”. A aula é finalizada com a apresentação dos cartazes e uma avaliação formativa contínua, analisando a participação dos alunos e os dados levantados durante as atividades.

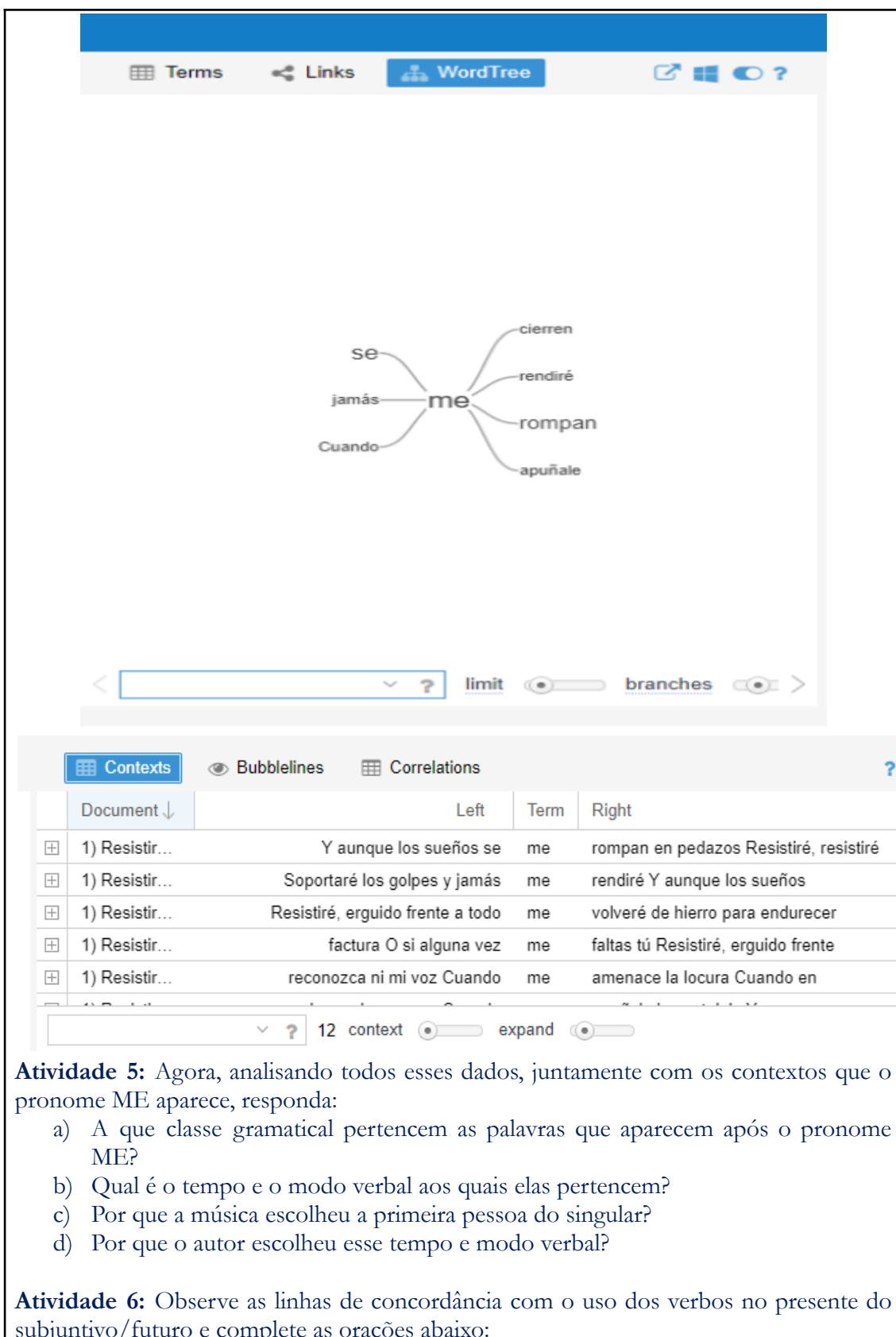
A ferramenta proporciona um olhar mais direcionado e reflexivo diante das combinações, uma vez que evidencia como é importante refletir com os alunos sobre as características da língua, porém, sempre em contextos significativos, ou seja, realizando a análise linguística a partir da leitura e da produção textual, porque não basta só estudar o modo verbal específico, por exemplo, mas é necessário também, observar como ele atua na construção do sentido do texto. Na Figura 1, podemos observar o comando da atividade e o layout das ferramentas utilizadas e o que elas oferecem de dados para a análise linguística.

Figura 01: Layout das atividades de DDL com a ferramenta *Voyant*.

Atividade 3: Utilize a ferramenta *Voyant Tools* para encontrar quais palavras são mais frequentes:



Atividade 4: Analise o pronome “ME”, por meio das ferramentas *Corpus Tools Word Tree*:



The screenshot displays a software interface for analyzing text. At the top, there are navigation buttons: 'Terms', 'Links', and 'WordTree'. The main area shows a word tree for the term 'me', with branches leading to 'se', 'jamás', 'Cuando', 'cierren', 'rendiré', 'rompan', and 'apuñale'. Below the tree is a search bar and a 'limit' slider. At the bottom, there are tabs for 'Contexts', 'Bubblelines', and 'Correlations'. The 'Contexts' tab is active, showing a table with columns for 'Document', 'Left', 'Term', and 'Right'. The table contains five rows of context data for the term 'me'.

| Document | Left | Term | Right |
|----------------|----------------------------------|------|--|
| 1) Resistir... | Y aunque los sueños se | me | rompan en pedazos Resistiré, resistiré |
| 1) Resistir... | Soportaré los golpes y jamás | me | rendiré Y aunque los sueños |
| 1) Resistir... | Resistiré, erguido frente a todo | me | volveré de hierro para endurecer |
| 1) Resistir... | factura O si alguna vez | me | faltas tú Resistiré, erguido frente |
| 1) Resistir... | reconozca ni mi voz Cuando | me | amenace la locura Cuando en |

Atividade 5: Agora, analisando todos esses dados, juntamente com os contextos que o pronome ME aparece, responda:

- A que classe gramatical pertencem as palavras que aparecem após o pronome ME?
- Qual é o tempo e o modo verbal aos quais elas pertencem?
- Por que a música escolheu a primeira pessoa do singular?
- Por que o autor escolheu esse tempo e modo verbal?

Atividade 6: Observe as linhas de concordância com o uso dos verbos no presente do subjuntivo/futuro e complete as orações abaixo:

| Document | Left | Term | Right |
|------------|---------------------------------|------|--|
| 1) Cuan... | con la soledad Cuando se | me | cierren las salidas Y la |
| 1) Cuan... | salidas Y la noche no | me | deje en paz Cuando sienta |
| 1) Cuan... | se rebelen los recuerdos Y | me | pongan contra la pared Resistiré |
| 1) Cuan... | Resistiré Erguido frente a todo | me | volveré de hierro para endurecer |
| 1) Cuan... | Soportaré los golpes y jamás | me | rendiré Y aunque los sueños |
| 1) Cuan... | Y aunque los sueños se | me | rompan en pedazos Resistiré, resistiré |
| 1) Cuan... | mi enemigo sea yo Cuando | me | apuñale la nostalgia Y no |
| 1) Cuan... | reconozca ni mi voz Cuando | me | amenace la locura Cuando en |
| 1) Cuan... | factura O si alguna vez | me | faltas tú Resistiré Erguido frente |
| 1) Cuan... | Resistiré Erguido frente a todo | me | volveré de hierro para endurecer |
| 1) Cuan... | Soportaré los golpes y jamás | me | rendiré Y aunque los sueños |
| 1) Cuan... | Y aunque los sueños se | me | rompan en pedazos Resistiré, resistiré |

12 context expand

Voyant Tools · Stéfán Sinclair & Geoffrey Rockwell (© 2022) Privacy v. 2.5.3 [2021/12/16] Não há notificações no

a) Cuándo _____ (terminar, yo) el curso _____ (trabajar) en un laboratorio.

b) Cuándo _____ (ser, yo) rico, me _____ (compraré) mi finca.

c) Cuándo _____ (tener, nosotros) diez años más, _____ (construir) nuestra propia casa.

d) Cuándo _____ (acabar) la pandemia, _____ (abrazar) muchas personas.

e) Cuándo me _____ (faltar) alegría, _____ (llamar) a los amigos verdaderos.

f)

Atividade 7: Após a reflexão da letra da música e do uso das orações temporais presente do subjuntivo/futuro, elabore um cartaz com frases, usando essa estrutura gramatical, transmitindo uma mensagem de resiliência com o seguinte tema: 5 cosas que haré cuándo los planes no salgan bien.

Fonte: Moscatelli (2023).

À medida que os alunos trabalham com a ferramenta, observam a padronização da estrutura, eles sentem-se mais confiantes e mais familiarizados para usarem esse vocabulário. Por isso, a atividade final é a produção do cartaz utilizando a estrutura gramatical trabalhada na sequência de atividades.

Esta abordagem integradora, que combina a análise de *corpus* com ferramentas digitais, promove não apenas a aprendizagem da língua espanhola, mas também o protagonismo e a autonomia dos alunos, ajudando-os a pensar criticamente sobre o uso da língua.

Considerações finais

Os conhecimentos produzidos a partir do grupo de estudo representam alguns casos em um dos campos que mais tem crescido na área da LC, o de *corpora* como recursos na criação de atividades didáticas para o ensino de língua estrangeira. No decorrer do

artigo, ficou evidente que é possível desenvolver atividades didáticas, provenientes de estudos com *corpora*.

Também evidenciamos que os dados são um caminho promissor tanto para a área de preparação de atividades didáticas quanto para o ensino de línguas estrangeiras, pois são importantes ferramentas para o ensino do uso autêntico da língua. Corroboramos ainda o que Flowerdew (1994) declarou sobre utilização da LC poder direcionar professores e pesquisadores quanto às características linguísticas e discursivas que os alunos precisam estar familiarizados, a fim de otimizar tanto a compreensão de textos específicos da área de atuação como ensino de línguas para fins gerais.

Os resultados obtidos no grupo de estudo com o livro Caderno de Atividades de Aprendizagem Movidas por Dados (Pinto; Garcia; Serpa, 2022) podem, por exemplo, fornecer possibilidades para que professores e pesquisadores da área da LC, com a abordagem em DDL, produzam suas atividades de maneira mais adequada aos contextos de ensino-aprendizagem de línguas estrangeiras. Além disso, essa abordagem permite o protagonismo e a autonomia dos alunos, ajudando-os a pensar criticamente sobre o uso da língua.

Referências

- BERBER SARDINHA, T. Computador, *corpus* e concordância no ensino de léxico-gramática de língua estrangeira. In: LEFFA, V. J. (Org.). **As palavras e sua companhia: o léxico na aprendizagem**. Pelotas, RS: EDUCAT / ALAB, 2000. p. 47- 75.
- BERBER SARDINHA, T. *Corpora* eletrônicos na pesquisa em tradução. **Cadernos de Tradução**, Florianópolis, v. 1, n. 9. jan. 2002. DOI: <https://doi.org/10.5007/%025x>. Disponível em: <https://periodicos.ufsc.br/index.php/traducao/article/view/5980>. Acesso em: 6 ago. 2024
- BERBER SARDINHA, T. **Linguística de Corpus**. São Paulo: Manole: 2004.
- BERBER SARDINHA, T. Como usar a Linguística de *Corpus* no ensino de língua estrangeira: por uma linguística de *corpus* educacional brasileira. In: VIANA, V.; TAGNIN, S (Orgs.). **Corpora no ensino de línguas estrangeiras**. São Paulo: Hub, 2011. p. 293-348.
- BERBER SARDINHA, T.; SHEPHERD, T. M. G.; DELEGÁ-LÚCIO, D. FERREIRA, T. L. S. B. **Tecnologias & mídias no ensino de Inglês: o corpus nas “receitas”**. 1. ed. São Paulo: Macmillan, 2012.
- BOULTON, A. Data-driven learning: reasonable fears and rational reassurance. **Indian Journal of Applied Linguistics**, v. 35, n. 1, p. 81-106, jan.-jun. 2009. DOI: <https://www.learntechlib.org/p/105180/>. Disponível em: <https://hal.science/hal-00326990v2>. Acesso em: 10 ago. 2024.
- CARVALHO, C. T. **Linguística de Corpus e suas contribuições para a elaboração de atividades de Inglês nos níveis A2 e B1 (QCER)**. 2021. 176 f. Dissertação (Mestrado

em Estudos Linguísticos) - Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, São José do Rio Preto, 2021.

CHARLES, M. Proper vocabulary and juicy collocations: EAP students evaluate do it-yourself *corpus*-building. **English for Specific Purposes**, v. 31, n. 2, p. 93-102, abr. 2012. DOI: <https://doi.org/10.1016/j.esp.2011.12.003>. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0889490611000780>. Acesso em: 10 ago. 2024.

CHENG, W. What can a *corpus* tell us about language teaching? In: O'KEEFFE, A.; MCCARTHY, M. (Eds.). **The Routledge handbook of corpus linguistics**. Londres / Nova York: Routledge, 2010. p. 319-332.

COXHEAD, A.; BYRD, P. Preparing writing teachers to teach the vocabulary and grammar of academic prose. **Journal of Second Language Writing**, v. 16, n. 3, p. 129- 147, set. 2007. DOI: <https://doi.org/10.1016/j.jslw.2007.07.002>. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S1060374307000471>. Acesso em: 12 jul. 2024.

CROSTHWAITE, P.; BOULTON, A. DDL is dead? Long live DDL! Expanding the boundaries of data-driven learning. In: TYNE, H; BILGER, M.; LUSCAIL, L.; LERAY, M.; CURRY N.; PÉREZ-SABATER, C. **Discovering language: Learning and affordance**. Frankfurt: Peter Lang, 2022. No prelo. Disponível em: <https://hal.science/hal-03506624>. Acesso em: 20 jul. 2024.

CROSTHWAITE, P. **Corpora for Language Learning: Bridging the Research-Practice Divide**. Londres: Routledge, 2024. DOI: <https://doi.org/10.4324/9781003413301>. Disponível em: <https://espace.library.uq.edu.au/view/UQ:0173875>. Acesso em: 10 ago. 2024.

DELFINO, M. C. N. **Uso de Música para o ensino de Inglês como língua estrangeira em um ambiente baseado em corpus**. 2016. 159 f. Dissertação (Mestrado em Linguística Aplicada e Estudos da Linguagem) – LAEL, Pontifícia Universidade Católica de São Paulo, São Paulo, 2016.

FLOWERDEW, J. **Academic Listening Research Perspectives**. Cambridge: Cambridge University Press, 1994.

FRANKENBERG-GARCIA, A.; PINTO, P. T.; BOCORNY, A. E. P.; SARMENTO, S. *Corpus*-aided EAP writing workshops to support international scholarly publication. **Applied Corpus Linguistics**, v. 2, n. 3, dez. 2022. DOI: <https://doi.org/10.1016/j.acorp.2022.100029>. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S2666799122000144>. Acesso em: 23 jul. 2024.

FRIGINAL, E. **Corpus linguistics for English teachers: new tools, online resources, and classroom activities**. 1. ed. Nova York: Routledge, 2018. 367 p.

GARCIA, W. D. **Fanfictions, Linguística de Corpus e Aprendizagem Direcionada por Dados: tarefas de produção escrita com foco no uso autêntico de língua e**

atividades que visam à autonomia dos alunos de Letras em analisar preposições. 2020. 145 f. Dissertação (Mestrado em Estudos da Linguagem) - Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, São José do Rio Preto, 2020.

GILQUIN, G.; GRANGER, S. How can data-driven learning be used in language teaching?. *In: O'KEEFFE, A.; MCCARTHY, M. (Eds.). **The Routledge handbook of corpus linguistics.*** Londres / Nova York: Routledge, 2010. p. 387-398.

JOHNS, T. Should you be persuaded: two samples of data-driven learning materials. *In: JOHNS, T.; KING, P. (Eds.). **Classroom concordancing:*** English Language Research Journal. Birmingham: University of Birmingham, 1991. p. 1-16.

JOHNS, T. From printout to handout: grammar and vocabulary teaching in the context of data-driven learning. *In: ODLIN, T. (Ed.). **Perspectives on pedagogical grammar.*** New York: Cambridge University Press, 1994.

KENNEDY, C.; MICELI, T. An evaluation of intermediate students approaches to *corpus* Investigation. **Language Learning & Technology**, v. 5, n. 3, p. 77-90, set. 2001. DOI: <https://dx.doi.org/10125/44567>. Disponível em: <https://www.lltjournal.org/item/10125-44567/>. Acesso em: 2 de jun. 2024.

NAZZI-LARANJA, L. A. **O uso de agrupamentos lexicais em atividades de língua inglesa para o vestibular: uma metodologia baseada em corpora.** 2020. 148 f. Dissertação (Mestrado em Estudos Linguísticos) - Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, São José do Rio Preto, 2020.

MAURANEN, A. Investigating English as a lingua franca with a spoken *corpus*. *In: CAMPOY, M. C; LUZÓN, M. J. (Orgs.). **Spoken corpora in applied linguistics.*** Bern: Peter Lang, 2004. p. 33-56.

MCENERY, T.; XIAO, R. What *corpora* can offer in language teaching and learning. *In: HINKEL, E. (Orgs.). **Handbook of research in second language teaching and learning.*** Nova York: Routledge, 2011. p. 364-380.

MOSCATELLI, S. Atividade 15: ¡cuándo me falte fuerza, resistiré! *In: PINTO, P. T.; GARCIA, W. D.; SERPA, T. (Orgs.) **Caderno de atividades de aprendizagem movida por dados.*** 1. ed. – Campinas, SP: Pontes Editores, 2022. p. 201-213. *E-book*: 222 p. Disponível em: https://ponteseditores.com.br/loja3/pontes-editores-home-2_trashed/ensino-de-linguas/caderno-de-atividades-de-aprendizagem-movida-por-dados/. Acesso em: 14 jul. 2024.

MOSCATELLI, S. **Ensino de Língua Espanhola em Curso de Eventos: o uso de corpora para desenvolver atividades didáticas com fins específicos.** 2023. Tese (Doutorado em Estudos Linguísticos) - Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, São José do Rio Preto, 2023.

O'KEEFFE, A.; MCCARTHY, M.; CARTER, R. **From corpus to classroom: language use and language teaching.** Cambridge: Cambridge University Press, 2007. 333 p.

PINTO, P. T.; CROSTHWAITE, P.; CARVALHO, C.; SPINELLI, F.; SERPA, T.; ORENHA-OTTAIANO, A. **Using language data to learn about language: a teachers' guide to classroom *corpus* use**. Brisbane: The University of Queensland, 2023. *E-book*: DOI: <https://doi.org/10.14264/3bbe92d>. Disponível em: <https://uq.pressbooks.pub/using-language-data/>. Acesso em: 3 jul. 2024.

PINTO, P. T.; GARCIA, W. D.; SERPA, T. **Caderno de atividades de aprendizagem movida por dados**. Campinas: Pontes Editores, 2022. *E-book*: 222 p. Disponível em: https://ponteseditores.com.br/loja3/pontes-editores-home-2_trashed/ensino-de-linguas/caderno-de-atividades-de-aprendizagem-movida-por-dados/. Acesso em: 14 jul. 2024.

SANTOS, L. V.; PINTO, P. T. Data driven learning para aprendizes iniciantes de inglês: uma experiência em um curso de estratégias de leitura. **Horizontes de Linguística Aplicada**. (no prelo).

SINCLAIR, J. M. *Corpus and Text - Basic Principles*. In: WYNNE, M. (Org.). **Developing Linguistic Corpora: a guide to good practice**. Oxford: Oxbow Books, 2005. 96 p. cap. 1. Disponível em: <https://users.ox.ac.uk/~martinw/dlc/index.htm>. Acesso em: 25 jun. 2024.

TAGNIN, S. E. O. Glossário de linguística de *corpus*. In: VIANA, V.; TAGNIN, S. (Orgs.) **Corpora no ensino de línguas estrangeiras**. São Paulo: HUB Editorial, 2011. p. 357-361.

TOGNINI-BONELLI, E. **Corpus linguistics at work**. Philadelphia, Amsterdam: John Benjamins, 2001.

WALSH, S. What features of spoken and written *corpora* can be exploited in creating language teaching materials and syllabuses. In: O'KEEFFE, A.; MCCARTHY, M. (Orgs.). **The Routledge handbook of corpus linguistics**. Nova York: Routledge, 2010. p. 319-332.

Recebido em 15 de agosto de 2024
Aceito em 8 de setembro de 2024