



UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
Câmpus de São José do Rio Preto

William Danilo Garcia

***DATA-DRIVEN LEARNING: PROPOSTAS E PRINCÍPIOS
PEDAGÓGICOS PARA A IMPLEMENTAÇÃO DESSA ABORDAGEM
NO CONTEXTO DA EDUCAÇÃO BÁSICA***

São José do Rio Preto

2024

William Danilo Garcia

***DATA-DRIVEN LEARNING: PROPOSTAS E PRINCÍPIOS
PEDAGÓGICOS PARA A IMPLEMENTAÇÃO DESSA ABORDAGEM
NO CONTEXTO DA EDUCAÇÃO BÁSICA***

Tese apresentada como parte dos requisitos para obtenção do título de Doutor em Estudos Linguísticos, junto ao Programa de Pós-Graduação em Estudos Linguísticos, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Câmpus de São José do Rio Preto.

Orientadora: Profa. Dra. Paula Tavares Pinto

São José do Rio Preto

2024

G216d

Garcia, William Danilo

Data-Driven Learning: propostas e princípios pedagógicos para a implementação dessa abordagem no contexto da educação básica / William Danilo Garcia. -- São José do Rio Preto, 2024

161 p.

Tese (doutorado) - Universidade Estadual Paulista (UNESP), Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto

Orientador: Paula Tavares Pinto

1. Linguística de Corpus. 2. Aprendizagem Movida por Dados. 3. Base Nacional Comum Curricular. 4. Ensino de Língua Inglesa. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Universidade Estadual Paulista (UNESP), Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

WILLIAM DANILO GARCIA

***DATA-DRIVEN LEARNING: PROPOSTAS E PRINCÍPIOS
PEDAGÓGICOS PARA A IMPLEMENTAÇÃO DESSA ABORDAGEM
NO CONTEXTO DA EDUCAÇÃO BÁSICA***

Tese apresentada como parte dos requisitos para obtenção do título de Doutor em Estudos Linguísticos, junto ao Programa de Pós-Graduação em Estudos Linguísticos, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Câmpus de São José do Rio Preto.

COMISSÃO EXAMINADORA

TITULARES

Profa. Dra. Paula Tavares Pinto
UNESP – Câmpus de São José do Rio Preto
Orientadora

Profa. Dra. Deise Prina Dutra
Universidade Federal de Minas Gerais (UFMG)

Prof. Dr. Eduardo Batista da Silva
Universidade Estadual de Goiás (UEG)

Profa. Dra. Cibele Cecilio de Faria Rozenfeld
UNESP – Câmpus de Araraquara

São José do Rio Preto

30 de agosto de 2024

A todos aqueles que me apoiaram e me incentivaram a continuar trilhando os caminhos acadêmicos. Vocês fazem parte do meu desenvolvimento e da minha formação.

AGRADECIMENTOS

A Deus e à espiritualidade, pela dádiva da vida, pela oportunidade de crescimento e de aprendizagem e pelas inúmeras bênçãos.

Aos meus pais, Mara e Rolemberg, pela vida e ancestralidade, por todo o amor, por todas as lições e os ensinamentos. Sem a educação que vocês me deram, eu alcançaria tamanha conquista.

À Profa. Dra. Paula Tavares Pinto, pela brilhante e atenciosa orientação, pelos aconselhamentos e conversas e por toda a confiança depositada em mim. Sem o seu apoio e incentivo, minha jornada acadêmica não renderia tantos frutos.

À Profa. Dra. Cibele Cecilio de Faria Rozenfeld, pelas leituras tão enriquecedoras do texto e por todos os apontamentos e considerações feitas, desde a sessão de debate do SeLin até a defesa.

Ao Prof. Dr. Eduardo Batista da Silva, por toda a colaboração, pelas valiosas contribuições trazidas no exame de qualificação e pelo aceite em compor a banca examinadora.

À Profa. Dra. Deise Prina Dutra, por aceitar compor a comissão examinadora.

A todos os docentes que ministraram workshops no curso *Práticas pedagógicas inovadoras para o ensino de língua inglesa: o aluno como detetive da língua e protagonista da aprendizagem*, Profa. Dra. Paula Tavares Pinto, Profa. Dra. Adriane Orenha Ottaiano, Profa. Dra. Talita Serpa, Prof. Dr. Eduardo Batista da Silva, Prof. Dr. Peter Crosthwaite, Profa. Dra. Ariane Donizete Delgado Ribeiro Caldas, Profa. Me. Carolina Tavares de Carvalho e Prof. Me. Emanuel Henrique Alves. Sem a colaboração de vocês, esta pesquisa não se concretizaria.

À Carolina Tavares de Carvalho, pela parceria na organização do curso *Práticas pedagógicas inovadoras para o ensino de língua inglesa: o aluno como detetive da língua e protagonista da aprendizagem*.

A todos os professores que aceitaram realizar o curso *Práticas pedagógicas inovadoras para o ensino de língua inglesa: o aluno como detetive da língua e protagonista da aprendizagem* e desenvolver atividades didáticas tão valiosas.

A todos os envolvidos na publicação do livro *Caderno de Atividades de Aprendizagem Movida por Dados*.

Ao Prof. Dr. Peter Crosthwaite, da Univesidade de Queensland, pela receptividade e parceria com o grupo de pesquisa *En-Corpora*.

Ao corpo docente do PPGEI, em especial, à Profa. Dra. Fabiana Komesu, ao Prof. Dr. Lourenço Chacon Jurado Filho, ao Prof. Dr. Manoel Luiz Gonçalves Corrêa, ao Prof. Dr. Lauro Maia Amorim e à Profa. Dra. Maria Angélica Deângeli, pelas aulas de altíssimo nível, pelos ensinamentos e por todas as orientações e direcionamentos.

Aos funcionários do PPGEL, da Seção Técnica da Pós-Graduação, da Biblioteca e aos representantes discentes, pela assistência e pela prontidão nos serviços.

À minha tia, Aparecida, por toda a ajuda em momentos tão difíceis e angustiantes. Sem o seu cuidado, jamais realizaria esse sonho.

Ao meu irmão, Jefferson, pelo carinho, incentivo e apoio em momentos tão desafiadores.

Ao meu parceiro, Danilo, por todo o carinho e incentivo. Suas palavras me inspiraram a seguir em frente e a concretizar esse passo tão importante na minha formação acadêmica.

Aos amigos e colegas do grupo de pesquisa *En-Corpora*, por todas as ideias compartilhadas, projetos realizados juntos e por toda a ajuda e motivação.

Aos amigos, aos integrantes do time de vôlei Melhores na Rede, e aos meus colegas de trabalho, pela torcida, por todo o apoio e motivação, e por me ajudarem a seguir firme no propósito acadêmico.

Aos meus alunos, pela paciência, colaboração e oração.

RESUMO

A relação da Linguística de *Corpus* com o ensino é comumente atrelada à década de 90, quando Tim Johns cunhou a Aprendizagem Movida por Dados. Desde então, inúmeros estudos são realizados no âmbito de defender a consulta em *corpora* como uma abordagem de aprendizagem de línguas que desperta no estudante um papel ativo, autônomo e investigativo com relação à análise do funcionamento da língua. No entanto, há certas críticas sobre a Aprendizagem Movida por Dados, no sentido de que ela ainda está atrelada ao Ensino Superior e a estudantes com níveis mais altos de proficiência, e de que ela ainda não adentrou as teorias mais convencionais de ensino. Desse modo, esta pesquisa objetiva correlacionar os pressupostos pedagógicos presentes nos documentos norteadores da Educação Básica à Aprendizagem Movida por Dados, a fim de desenvolver lições e atividades didáticas que visem à implementação de materiais didáticos. Além disso, esta tese tem a finalidade de promover formação continuada a professores, para que eles também desenvolvam atividades didáticas, e propor princípios pedagógicos que orientem a elaboração de lições que se pautem na Aprendizagem Movida por Dados. Como fundamentação teórica, este trabalho se embasa na Linguística de *Corpus* e o ensino, na Aprendizagem Movida por Dados, na Base Nacional Comum Curricular e no Currículo Paulista. Com relação aos materiais e métodos, destacam-se as ferramentas computacionais da Linguística de *Corpus*, assim como o trabalho colaborativo entre pesquisadores e professores durante o curso de formação continuada em serviço e no decorrer do desenvolvimento de atividades didáticas. Como resultados alcançados, pontua-se a criação de duas lições e de dez atividades didáticas pautadas na Aprendizagem Movida por Dados, assim como a proposta de quinze princípios pedagógicos que orientem a utilização da Aprendizagem Movida por Dados na implementação de materiais didáticos ou na elaboração de lições e/ou atividades didáticas. Todos esses resultados evidenciaram o quanto a Aprendizagem Movida por Dados é aplicável a diferentes contextos educacionais e pode contribuir com a implementação de materiais didáticos já conhecidos e utilizados pelos professores.

Palavras-chave: Linguística de *Corpus*. Aprendizagem Movida por Dados. Base Nacional Comum Curricular. Ensino de Língua Inglesa.

ABSTRACT

The relationship between Corpus Linguistics and teaching is commonly linked to the 90s, when Tim Johns elaborated Data-Driven Learning. Since then, numerous studies have been carried out to defend corpora consultation as a language learning approach that encourages students to take an active, autonomous and investigative role in analyzing how the language works. However, there are certain criticisms about Data-Driven Learning, in the sense that it is still linked to Higher Education and to students with higher proficiency levels, and that it has not yet entered mainstream classroom pedagogy. Therefore, this research aims to correlate the pedagogical assumptions present in the guiding documents of Basic Education to Data-Driven Learning, in order to develop lessons and teaching activities focused on implementing teaching materials. Furthermore, this thesis aims to promote continued training for teachers, so that they can also develop their own teaching activities, and propose pedagogical principles that guide the development of lessons based on Data-Driven Learning. As theoretical basis, this work is based on Corpus Linguistics and teaching, Data-Driven Learning, the National Common Curricular Base and the Paulista Curriculum. Regarding materials and methods, the computational tools of Corpus Linguistics stand out, as well as the collaborative work between researchers and teachers during the in-service teacher training course and during the development of teaching activities. The results achieved include the creation of two lessons and ten teaching activities based on Data-Driven Learning, as well as the proposal of fifteen pedagogical principles that can guide the use of Data-Driven Learning in the implementation of teaching materials or in the preparation of lessons and /or teaching activities. These results highlighted how Data-Driven Learning is applicable to different educational contexts and can contribute to the implementation of teaching materials already known and used by teachers.

Keywords: Corpus Linguistics. Data-Driven Learning. National Common Curricular Base. English Language Teaching.

LISTA DE ILUSTRAÇÕES

Figura 1. A Educação Básica	22
Figura 2. Ensino Médio e as áreas do conhecimento	23
Figura 3. O uso de <i>corpora</i> na aprendizagem e no ensino de línguas	42
Figura 4. Atividade 1 da lição <i>So many things in a day</i>	52
Figura 5. Atividade 2 da lição <i>So many things in a day</i>	53
Figura 6. Atividade 3 da lição <i>So many things in a day</i>	53
Figura 7. Atividade 3 da lição <i>So many things in a day</i> , modelo <i>hands-off</i>	54
Figura 8. Atividade 4 da lição <i>So many things in a day</i>	55
Figura 9. Atividade 1 da lição <i>It depends on the situation</i>	58
Figura 10. Atividade 2 da lição <i>It depends on the situation</i>	58
Figura 11. Atividade 3 da lição <i>It depends on the situation</i> , modelo <i>hands-on</i>	59
Figura 12. Atividade 3 da lição <i>It depends on the situation</i> , modelo <i>hands-off</i>	61
Figura 13. Atividade 4 da lição <i>It depends on the situation</i>	62
Figura 14. Representação da aula de línguas estrangeiras	64
Figura 15. O curso “Práticas pedagógicas inovadoras para o ensino de língua inglesa: o aluno como detetive da língua e protagonista da aprendizagem	69
Figura 16. Capa da apresentação de slides utilizada na oficina do dia 04/05/2022	71
Figura 17. Roteiro dos pontos abordados na oficina do dia 04/05/2022	72
Figura 18. Capa da apresentação de slides da oficina do dia 25/05/2022	74
Figura 19. Capa da apresentação de slides da oficina do dia 01/06/2022	75
Figura 20. Página inicial do Skell	82
Figura 21. Consultando uma palavra no Skell	83
Figura 22. Resultado de busca do verbo <i>take</i> na função <i>Examples</i>	84
Figura 23. Resultado da busca do verbo <i>take</i> na função <i>Word sketch</i>	85
Figura 24. Resultado de busca do verbo <i>take</i> na função <i>Similar words</i>	86
Figura 25. Página inicial do Linggle	87
Figura 26. Os comandos de busca do Linggle	88
Figura 27. Resultados da busca <i>go _</i> no Linggle	89
Figura 28. Resultados da busca <i>go *</i> no Linggle	90
Figura 29. Resultados da busca <i>love n.</i> no Linggle	91
Figura 30. Página inicial do Versatext	92

Figura 31. Inserção do texto no Versatext	93
Figura 32. A função <i>Word cloud</i>	94
Figura 33. A função Profiler	95
Figura 34. Ilustração do material didático <i>Super Seek and Find</i>	98
Figura 35. Exemplificação do procedimento de busca com a palavra <i>throw</i>	99
Figura 36. Texto <i>Are girls smarter than boys?</i>	101
Figura 37. Proposta de identificação das <i>noun phrases</i> no texto	102
Figura 38. Procedimento de busca de <i>gray matter</i> no Skell	103
Figura 39. Resultados obtidos para modificadores de <i>matter</i> , por meio do <i>Work Sketch</i>	104
Figura 40. Atividade de leitura com o tema <i>fashion</i>	106
Figura 41. Atividade de nuvem de palavras com o termo <i>fashion</i>	107
Figura 42. Pesquisa por <i>fashion</i> no Skell	108
Figura 43. Atividade de combinações de duas palavras da área médica	109
Figura 44. Consulta de <i>anaesthetic</i> no Skell	110
Figura 45. Exercício de correlacionar as combinações aos respectivos sentidos	111
Figura 46. Consulta de <i>anaesthetic</i> no Skell por meio da opção <i>Word Sketch</i>	112
Figura 47. Busca pelo quantificador <i>many</i> no Linggle	115
Figura 48. Busca pelo quantificador <i>many</i> e os substantivos que o acompanham	116
Figura 49. Atividade de análise do funcionamento dos quantificadores a partir do uso	117
Figura 50. Atividade sobre as <i>dependent prepositions</i>	119
Figura 51. Exemplo de busca com o verbo <i>arrive</i>	120
Figura 52. Exemplos de <i>arrive in</i>	121
Figura 53. Atividade de leitura com tema relacionado à beleza	122
Figura 54. Atividade de nuvem de palavras com o substantivo <i>beauty</i>	123
Figura 55. Exemplificação da busca por <i>beauty</i> no Linggle	124
Figura 56. Exemplificação de busca com <i>beauty</i> *	124
Figura 57. Exemplo de busca com o comando <i>~beauty</i>	125
Figura 58. Atividade com foco nos sinônimos de <i>voting</i>	126
Figura 59. Exemplo da busca <i>~voting</i> na ferramenta Linggle	127
Figura 60. Atividade com foco nas palavras derivadas de <i>vote</i>	128
Figura 61. Resultado da busca <i>*vote</i>	128

Figura 62. Exemplos de uso da palavra <i>devote</i>	129
Figura 63. Atividade de <i>listening</i> com a música <i>I want to break free</i>	130
Figura 64. Atividade com foco em gerar nuvem de palavras com a letra da música	131
Figura 65. Nuvem de palavras gerada a partir da canção <i>I want to break free</i>	131
Figura 66. Atividade com foco nas classes gramaticais	132
Figura 67. Nuvem de palavras do <i>abstract</i> do artigo	134
Figura 68. Atividade de busca das palavras desconhecidas no Linggle	135
Figura 69. Exemplo de pesquisa de <i>intake</i> no Linggle	135
Figura 70. Atividade com foco nas palavras-chaves em contexto	136
Figura 71. Procedimento de elaboração da atividade com foco nas palavras-chaves no Versatext	137
Gráfico 1. Facilidade de aplicação do Skell e do Linggle	77
Gráfico 2. Probabilidade de incorporar o Skell e o Linggle no contexto de ensino	77
Gráfico 3. Facilidade de aplicação do Versatext	78
Gráfico 4. Probabilidade de incorporar o Versatext no contexto de ensino	78
Quadro 1. Competências Gerais da Educação Básica	25
Quadro 2. Taxonomia de <i>corpora</i>	40
Quadro 3. Os módulos 1 e 2 do curso do Prof. Dr. Peter Crosthwaite	73
Quadro 4. Os dados que alimentam o Skell	81
Quadro 5. Lista das atividades didáticas elaboradas pelos cursistas	97
Quadro 6. Atividade para análise de vocabulários da música	99
Quadro 7. Atividade proposta para pesquisa de <i>noun phrases</i> no Skell	103
Quadro 8. Exercício de formar combinações	112
Quadro 9. Exercício dos quantificadores	114

SUMÁRIO

1 INTRODUÇÃO	14
2 FUNDAMENTAÇÃO TEÓRICA	20
2.1 A Base Nacional Comum Curricular (BNCC)	20
2.1.1 A BNCC e a relação com as metodologias ativas	27
2.1.2 BNCC: um olhar reflexivo	29
2.2 O Currículo Paulista – etapa do Ensino Médio	32
2.3 A Linguística de <i>Corpus</i>	34
2.4 A Linguística de <i>Corpus</i> e o ensino de línguas	41
2.5 A Aprendizagem Movida por Dados (<i>Data-Driven Learning</i>)	44
3 DA TEORIA À PRÁTICA: EXEMPLIFICAÇÃO DE PROPOSTAS DE ATIVIDADES DIDÁTICAS	51
3.1 <i>So many things in a day</i>: proposta de atividade didática elaborada para o Ensino Médio	52
3.2 <i>It depends on the situation</i>: proposta de atividade didática elaborada para o contexto de curso livre de inglês	57
3.3 Da organização da aula	63
4 MATERIAIS E MÉTODOS	66
4.1 Dos métodos e do curso de formação continuada em serviço	66
4.1.1 O curso “Práticas pedagógicas inovadoras para o ensino de língua inglesa: o aluno como detetive da língua e protagonista da aprendizagem.”	68
4.2 As ferramentas	79
4.2.1 A ferramenta Skell	80
4.2.2 A ferramenta Linggle	86
4.2.3 A ferramenta Versatext	91
5 PROPOSTAS DE ATIVIDADES DIDÁTICAS: IMPLEMENTAÇÕES POSSÍVEIS AOS MATERIAIS DIDÁTICOS	96

5.1 Propostas de atividades didáticas desenvolvidas pelos professores durante o curso de formação continuada em serviço	96
5.1.1 Atividade com a canção <i>Kangaroo Groove</i>	97
5.1.2 Atividade com foco em <i>noun phrases</i> a partir da leitura de texto	100
5.1.3 Atividade com foco em ampliar os conhecimentos lexicais que acompanham a palavra <i>fashion</i>	105
5.1.4 Atividade com foco em vocabulários específicos da área da Medicina	109
5.1.5 Atividade com foco nos quantificadores	113
5.1.6 Atividade com foco nas preposições	118
5.1.7 Atividade com foco em ampliar os conhecimentos lexicais que acompanham a palavra <i>beauty</i>	122
5.1.8 Atividade com foco em ampliar os conhecimentos lexicais que acompanham a palavra <i>voting</i>	126
5.1.9 Atividade com a música <i>I want to break free</i> , da banda <i>Queen</i>	130
5.1.10 Atividade com foco em estratégias de leitura	133
5.2 Princípios pedagógicos para a elaboração de lições que recorrem à Aprendizagem Movida por Dados	138
6 CONSIDERAÇÕES FINAIS E ENCAMINHAMENTOS	146
REFERÊNCIAS	151
APÊNDICE A – EMENTA DO CURSO “PRÁTICAS PEDAGÓGICAS INOVADORAS PARA O ENSINO DE LÍNGUA INGLESA: O ALUNO COMO DETETIVE DA LÍNGUA E PROTAGONISTA DA APRENDIZAGEM”	156
ANEXO A – CERTIFICADO DE TRADUÇÃO DO CURSO “IMPROVING WRITING THROUGH CORPORA: DATA-DRIVEN LEARNING?” PARA O PORTUGUÊS	161

1 INTRODUÇÃO

A delimitação deste estudo está estreitamente relacionada à dissertação de Mestrado desenvolvida pelo pesquisador (Garcia, 2020), que enfatizou a utilização da Linguística de *Corpus* (LC) e de ferramentas computacionais (os concordanciadores) para o ensino de preposições aos estudantes do curso de Letras. Isso foi feito de modo a contrastar a forma como os graduandos utilizam esses itens gramaticais em relação ao uso convencional e padrão na língua. Assim, ao comparar os *corpora* de aprendizagens (a produção escrita dos estudantes) ao *corpus* de referência (representativo do funcionamento geral da língua), observou-se uma melhora significativa na escrita e no emprego das preposições pelos envolvidos na pesquisa. Com base nisso, estabeleceu-se, no pesquisador, o interesse em seguir investigando como a LC pode implementar o ensino.

Além disso, há de se considerar que esse desejo de seguir realizando pesquisas que se voltem ao contexto educacional se faz presente não apenas pelas conclusões obtidas no Mestrado, como também pela jornada profissional do autor desta tese, que atua há mais de dez anos na Educação, tanto em contextos de cursos livres de idiomas quanto no ensino regular, mais precisamente no Ensino Fundamental II e Médio da rede pública estadual de São Paulo.

Devido a isso, definiu-se o aprofundamento científico na abordagem da Aprendizagem Movida por Dados em contextos que vão além do Ensino Superior, como o desenvolvido no Mestrado. Desta vez, portanto, o foco foi expandir as etapas educacionais a fim de inserir, nesta pesquisa, a Educação Básica. Nesse sentido, definiu-se aprofundar os estudos na Aprendizagem Movida por Dados, do inglês *Data-Driven Learning* (DDL).

Desse modo, no que tange ao estado da arte em relação à DDL, pontua-se, por meio de diversos estudos e pesquisadores a serem devidamente abordados na fundamentação teórica desta tese, que essa metodologia desperta o papel ativo, autônomo e investigativo no estudante. Tal fato possibilita que ele não só lance questionamentos e levante hipóteses com relação ao funcionamento da língua, mas analise e compreenda, a partir de exemplos de uso da língua, seus padrões e como ela funciona.

Entretanto, por mais que haja um consenso em relação à relevância da DDL, há também algumas considerações sobre como ela pode ser ainda mais explorada. Dentre essas ponderações, pontua-se que, conforme Crosthwaite (2024), embora haja muitos trabalhos científicos sobre a Aprendizagem Movida por Dados, ela ainda não conseguiu adentrar totalmente as teorias de ensino estabelecidas como as mais convencionais. Somado a isso, Boulton (2024) afirma que os estudos em DDL se restringem ao contexto universitário e do Ensino Superior e carecem de colaboração entre pesquisadores e professores.

Diante disso, esta pesquisa identificou a necessidade de desenvolver uma série de ações que visassem à superação das lacunas em relação à Aprendizagem Movida por Dados. A primeira delas, portanto, é a de focar a pesquisa na Educação Básica e nas teorias de ensino que embasam os documentos norteadores dessa etapa educacional. Para isso, recorreu-se à Base Nacional Comum Curricular e ao Currículo Paulista, no intuito de compreender os pressupostos pedagógicos que os permeiam.

A partir dessa investigação, foram observados certos pontos em comum entre os pressupostos pedagógicos contidos nesses documentos e as conceitualizações acerca da DDL. Esse olhar atento a como essas teorias todas podem ser atreladas de modo a potencializar a elaboração de lições, que recorram à Aprendizagem Movida por Dados, contribuiu com a publicação do livro *Caderno de Atividades de Aprendizagem Movida por Dados* em parceria aos integrantes do grupo de pesquisa *En-Corpora*¹ e em sua versão para o inglês, intitulada *Using Language Data to Learn about Language: a Teacher's Guide to Classroom Corpus Use*, que contou com o apoio do Prof. Dr. Peter Crosthwaite, da Universidade de Queensland.

Ademais, como forma de superar a carência de estudos que contem com a colaboração de pesquisadores e professores (Boulton, 2024), outra ação desenvolvida por esta pesquisa foi a elaboração de um curso de formação continuada em serviço para professores atuantes na Educação Básica. Salienta-se que esse projeto teve um caráter bastante colaborativo, tanto no sentido de reunir pesquisadores da Universidade Estadual Paulista (UNESP), da Universidade Estadual de Goiás (UEG) e da Universidade de Queensland, quanto por levar formação sobre a DDL e as ferramentas da LC a professores em serviço, sendo muitos deles atuantes no Ensino Médio da rede pública estadual de São Paulo.

¹ Link do Grupo de Pesquisa: <https://dgp.cnpq.br/dgp/espelhogrupo/619726>

Ressalta-se que, como resultado da participação nesse curso, os professores cursistas puderam não só ver exemplos de lições pautadas na DDL, como também elaborar, com o apoio do autor desta tese, suas próprias atividades didáticas, de modo a implementarem os materiais didáticos com os quais trabalham com a metodologia da Aprendizagem Movida por Dados. Todas essas atividades, além das lições desenvolvidas pelo pesquisador para compor o livro *Caderno de Atividades de Aprendizagem Movida por Dados*, foram devidamente analisadas com o intuito de se estabelecer quais são os pontos cruciais no desenvolvimento e/ou implementação de materiais didáticos que recorram à Aprendizagem Movida por Dados. Eles são, portanto, propostos como princípios pedagógicos para que a abordagem da Aprendizagem Movida por Dados seja utilizada de forma eficaz na implementação de materiais didáticos ou na elaboração de lições e/ou atividades.

Visto tudo isso, as **perguntas de pesquisa** que norteiam este estudo são:

- Em que medida a Aprendizagem Movida por Dados e os pressupostos pedagógicos abordados nos documentos orientadores da Educação Básica podem se complementar e contribuir para a disseminação dessa abordagem em contextos educacionais que vão além daqueles convencionalmente pesquisados?

- De quais formas a elaboração de um curso de formação continuada em serviço para professores pode promover a implementação de materiais didáticos e/ou elaboração de atividades didáticas que se utilizam da Aprendizagem Movida por Dados?

- Quais são os princípios pedagógicos que podem orientar a utilização da Aprendizagem Movida por Dados na implementação de materiais didáticos ou na elaboração de lições e/ou atividades didáticas?

À vista dessas perguntas de pesquisa, reitera-se que o **objetivo geral** desta tese é:

- Propor princípios pedagógicos que possam orientar a utilização da Aprendizagem Movidada por Dados na implementação de materiais didáticos ou na elaboração de lições e/ou atividades didáticas.

Já no que tange aos **objetivos específicos**, são eles:

- Desenvolver lições que utilizam da abordagem da Aprendizagem Movidada por Dados e que se amparam nos pressupostos pedagógicos dos documentos norteadores da Educação Básica, com o intuito de demonstrar que a DDL também se aplica a outros contextos educacionais (não apenas ao Ensino Superior), assim como a estudantes com níveis mais baixos de proficiência (não apenas a alunos avançados).
- Oferecer um curso de formação continuada em serviço de curta duração para professores que atuam na Educação Básica, a fim de divulgar as ferramentas da LC e a Aprendizagem Movidada por Dados.
- Atrelar, por meio desse curso, pesquisadores da área da LC de diferentes universidades aos professores atuantes na Educação Básica, com o objetivo de que estes utilizem da Aprendizagem Movidada por Dados para produzirem atividades didáticas que implementem os materiais didáticos com os quais trabalham.

Compreendidas as perguntas de pesquisa e os objetivos, são descritos os procedimentos metodológicos envolvidos nesta tese. Inicialmente, realizou-se a revisão bibliográfica, que está organizada na fundamentação teórica, capítulo dois, de modo a considerar a BNCC, o Currículo Paulista, a Linguística de *Corpus* e, em seguida, sua relação com o ensino de línguas para, então, lidar com a Aprendizagem Movidada por Dados.

Sobre o capítulo três, ele envolve a demonstração das duas lições desenvolvidas pelo autor desta tese e publicadas no livro *Caderno de Atividades de Aprendizagem Movidada por Dados*, como forma de se estabelecer uma relação entre as teorias recorridas e a prática. Desse modo, compreende-se que esse capítulo já

apresenta resultados relevantes em relação à elaboração de lições que recorrem à DDL.

Com relação ao capítulo quatro, ele engloba uma descrição da natureza da pesquisa para, na sequência, descrever o curso de formação continuada em serviço de curta duração, que foi desenvolvido pelo pesquisador em parceria com outra discente do Programa, a Carolina Tavares de Carvalho, com o objetivo de divulgar a Aprendizagem Movida por Dados e suas possibilidades de aplicação em sala de aula com o auxílio das ferramentas da LC. Por último, apresentam-se as ferramentas da LC utilizadas nesta pesquisa, que são: Skell, Linggle e Versatext.

Já o capítulo cinco contempla os resultados alcançados, ou melhor, as atividades didáticas propostas pelos participantes do curso em colaboração com o autor desta tese. Além disso, apresentam-se, nesse capítulo, os princípios pedagógicos para que a abordagem da Aprendizagem Movida por Dados seja utilizada de forma eficaz na implementação de materiais didáticos ou na elaboração de lições e/ou atividades didáticas.

Finalmente, esclarece-se que esta tese de Doutorado se classifica, segundo Paiva (2019), como uma pesquisa de natureza aplicada, do gênero prático e com abordagem mista (utilizando-se de métodos qualitativos e quantitativos). Ela foi desenvolvida junto ao Programa de Pós-Graduação em Estudos Linguísticos da Universidade Estadual Paulista (UNESP), câmpus de São José do Rio Preto, e está relacionada à área de concentração da Linguística Aplicada, mais precisamente atrelada à linha de pesquisa da Pedagogia do Léxico e da Tradução Baseada em *Corpora*.

Essa linha de pesquisa abarca “[...] não apenas os estudos da tradução e do léxico, mas também aplicações pedagógicas a partir destas investigações” (Orenha-Ottaiano; Pinto, 2018, p. 127). As autoras enfatizam ainda que um dos objetivos dessa linha de pesquisa envolve a elaboração de materiais didáticos baseados em *corpora* (Orenha-Ottaiano; Pinto, 2018). Esse enfoque, inclusive, passou a ter mais notoriedade com os trabalhos mais recentes do grupo *En-Corpora*, que têm enfatizado a Aprendizagem Movida por Dados. É válido citar, por exemplo, o estudo da DDL no contexto de Inglês para Fins Acadêmicos (Pinto, 2024) e na melhoria da leitura e escrita acadêmica (Pinto e Crosthwaite, no prelo). Desse modo é no desenvolvimento

de atividades didáticas que se amparam na Aprendizagem Movida por Dados para o ensino de língua inglesa que se enquadra este estudo.

Por fim, ainda no âmbito do delineamento desta tese, vale destacar que, justamente por ter esse caráter aplicável ao ensino e à etapa da Educação Básica, é que ela se ampara no projeto de pesquisa da Profa. Dra. Paula Tavares Pinto intitulado *The United Nations Sustainable Development Goals (SDGs) in the Brazilian research: a corpus-based approach to support research-paper writing and translation*. Isso porque essa preocupação com a educação está em consonância com o Objetivo de Desenvolvimento Sustentável de número quatro, que diz respeito à educação de qualidade.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo propõe-se a revisar os fundamentos teóricos que embasam esta pesquisa. Para tanto, inicia-se com uma descrição dos documentos norteadores da Educação Básica no Brasil e no estado de São Paulo. Em seguida, traz-se luz à Linguística de *Corpus* por meio de conceitualizações mais gerais dessa vertente dos estudos da linguagem. Feito isso, parte-se para a exploração da relação da LC com o ensino de línguas e, finalmente, encaminha-se para a aplicação mais direta de *corpus* ao ensino mediante a revisão dos conceitos sobre a Aprendizagem Movida por Dados.

2.1 A Base Nacional Comum Curricular (BNCC)

Prévia à conceitualização da Base Nacional Comum Curricular, considera-se apropriado pontuar que esse documento educacional e o Currículo Paulista compõem a seção da fundamentação teórica, pois há neles princípios pedagógicos que se alinham à Aprendizagem Movida por Dados e que serão, por conseguinte, recorridos como argumentos que defendem a utilização da DDL no contexto do Ensino Médio da rede pública estadual de São Paulo. Nesse sentido, entende-se que a discussão sobre esses documentos é pertinente à área da Educação e imprescindível no sentido de atrelar a Aprendizagem Movida por Dados às teorias de ensino mais convencionais e em vigor, suprimindo, assim, um dos desafios ainda enfrentados pela DDL: o de ainda não ter adentrado as práticas pedagógicas mais convencionais (Crosthwaite, 2024).

Homologado o texto referente a etapa do Ensino Médio em 14 de dezembro de 2018 pelo ministro da Educação da época, Rossieli Soares, a Base Nacional Comum Curricular (BNCC)² passou a valer como “documento de caráter normativo que define o conjunto orgânico e progressivo de aprendizagens essenciais que todos os alunos devem desenvolver ao longo das etapas e modalidades da Educação Básica [...]” (Brasil, 2018, p. 7).

Entende-se que a Educação Básica, de acordo com (Brasil, 2018), compreende as seguintes etapas:

² Disponível em: <http://basenacionalcomum.mec.gov.br/>. Acesso em: 15 de julho de 2022.

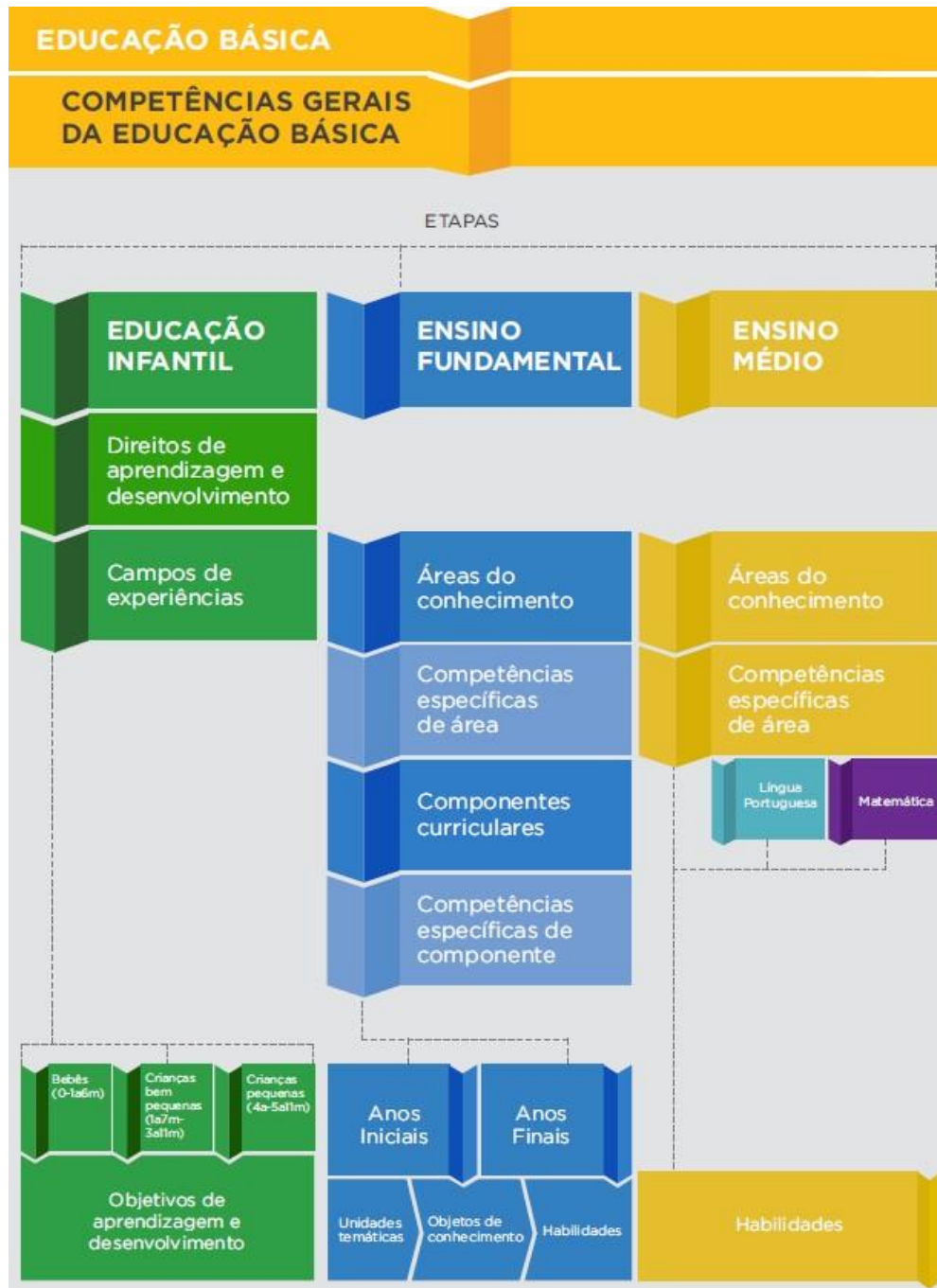
- Educação Infantil, subdividida em bebês (entre 0 e 1 ano e 6 meses), crianças bem pequenas (entre 1 ano e 7 meses e 3 anos e 11 meses) e crianças pequenas (entre 4 anos e 5 anos e 11 meses);

- Ensino Fundamental, desdobrado em Anos Iniciais (do 1º ao 5º ano) e Anos Finais (do 6º ao 9º ano);

- Ensino Médio, correspondente a 1ª, 2ª e 3ª série.

A figura a seguir contempla bem toda essa divisão.

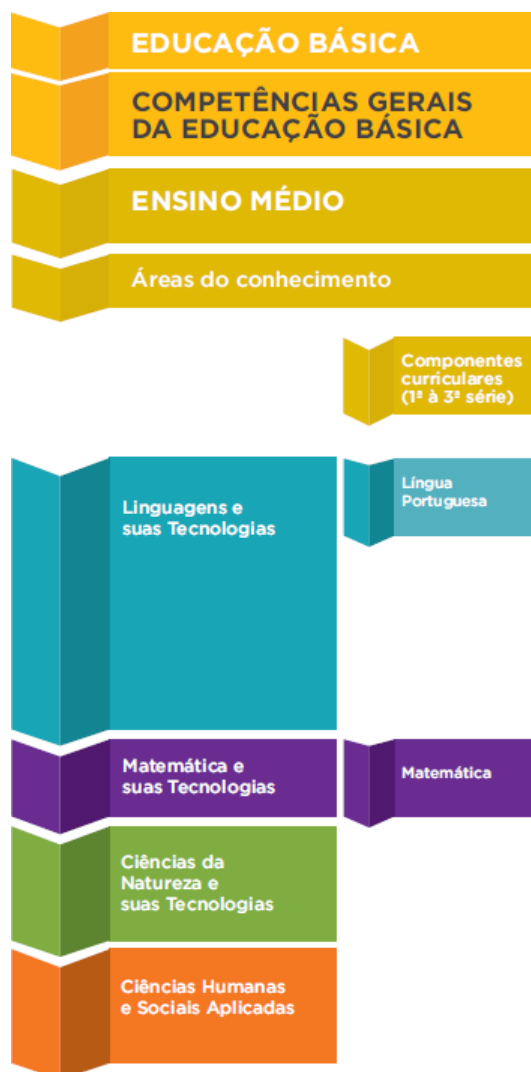
Figura 1. A Educação Básica



Fonte: Brasil (2018, p. 24).

Tanto para o Ensino Fundamental quanto para o Ensino Médio, os componentes curriculares, ou melhor, as disciplinas, tais como Língua Portuguesa, Matemática, História e assim por diante, são categorizados por áreas do conhecimento, conforme a figura subsequente:

Figura 2. Ensino Médio e as áreas do conhecimento



Fonte: Brasil (2018, p. 32).

Nesse sentido, a BNCC compreende, em consonância com a LDB, que, para a etapa do Ensino Médio (foco desta pesquisa), a área de Linguagens e suas Tecnologias integra os componentes curriculares de Arte, Educação Física, Língua Inglesa e Língua Portuguesa. Matemática e suas Tecnologias está atrelada apenas à Matemática. Ciências da Natureza e suas Tecnologias engloba a Biologia, a Química e a Física. Já Ciências Humanas e Sociais Aplicadas abrange a Filosofia, a Geografia, a História e a Sociologia. Destaca-se, ainda, que, para além da Formação Geral Básica, que envolve esses componentes curriculares aqui descritos, a BNCC também prevê que o Ensino Médio possui certa flexibilização em sua organização curricular, de forma a proporcionar aos alunos uma maior escolha em relação àquilo que

gostariam de estudar, o que a Base Nacional Comum Curricular denomina de Itinerários Formativos.

Compreendidas, então, essas questões de caráter organizacional das etapas de ensino e das áreas de conhecimento, julga-se necessário pontuar que, no que tange à delimitação das aprendizagens essenciais da Educação Básica no Brasil pela BNCC, observa-se que essa é uma demanda existente desde a implementação da Lei de Diretrizes e Bases da Educação Nacional (LDB, Lei nº 9.394/1996) e que se intensificou a partir do Plano Nacional de Educação (PNE)³, homologado em 2014 pela Lei nº 13.005, que estabelece vinte metas para a educação nacional a serem cumpridas no prazo de dez anos.

O intuito dessas aprendizagens essenciais, propostas pela Base Nacional Comum Curricular, é visto como um aspecto positivo por Fidalgo, Sprenger e Ramos (2021), pois “pode evitar que cada estado, região ou escola de uma cidade ofereçam currículos e propostas completamente diferentes.” Portanto, essa tentativa de padronizar, de certa maneira, o mínimo que precisa ser ensinado em todo o território nacional visa a assegurar aos estudantes o desenvolvimento de dez competências gerais que perpassam toda a Educação Básica nacional. São elas:

³ Disponível em: <https://pne.mec.gov.br/>. Acesso em: 15 de julho de 2022.

Quadro 1. Competências Gerais da Educação Básica

01	Valorizar e utilizar os conhecimentos historicamente construídos sobre o mundo físico, social, cultural e digital para entender e explicar a realidade, continuar aprendendo e colaborar para a construção de uma sociedade justa, democrática e inclusiva.
02	Exercitar a curiosidade intelectual e recorrer à abordagem própria das ciências, incluindo a investigação, a reflexão, a análise crítica, a imaginação e a criatividade, para investigar causas, elaborar e testar hipóteses, formular e resolver problemas e criar soluções (inclusive tecnológicas) com base nos conhecimentos das diferentes áreas.
03	Valorizar e fruir as diversas manifestações artísticas e culturais, das locais às mundiais, e também participar de práticas diversificadas da produção artístico-cultural.
04	Utilizar diferentes linguagens – verbal (oral ou visual-motora, como Libras, e escrita), corporal, visual, sonora e digital –, bem como conhecimentos das linguagens artística, matemática e científica, para se expressar e partilhar informações, experiências, ideias e sentimentos em diferentes contextos e produzir sentidos que levem ao entendimento mútuo.
05	Compreender, utilizar e criar tecnologias digitais de informação e comunicação de forma crítica, significativa, reflexiva e ética nas diversas práticas sociais (incluindo as escolares) para se comunicar, acessar e disseminar informações, produzir conhecimentos, resolver problemas e exercer protagonismo e autoria na vida pessoal e coletiva.
06	Valorizar a diversidade de saberes e vivências culturais e apropriar-se de conhecimentos e experiências que lhe possibilitem entender as relações próprias do mundo do trabalho e fazer escolhas alinhadas ao exercício da cidadania e ao seu projeto de vida, com liberdade, autonomia, consciência crítica e responsabilidade.
07	Argumentar com base em fatos, dados e informações confiáveis, para formular, negociar e defender ideias, pontos de vista e decisões comuns que respeitem e promovam os direitos humanos, a consciência socioambiental e o consumo responsável em âmbito local, regional e global, com posicionamento ético em relação ao cuidado de si mesmo, dos outros e do planeta.
08	Conhecer-se, apreciar-se e cuidar de sua saúde física e emocional, compreendendo-se na diversidade humana e reconhecendo suas emoções e as dos outros, com autocrítica e capacidade para lidar com elas.
09	Exercitar a empatia, o diálogo, a resolução de conflitos e a cooperação, fazendo-se respeitar e promovendo o respeito ao outro e aos direitos humanos, com acolhimento e valorização da diversidade de indivíduos e de grupos sociais, seus saberes, identidades, culturas e potencialidades, sem preconceitos de qualquer natureza.
10	Agir pessoal e coletivamente com autonomia, responsabilidade, flexibilidade, resiliência e determinação, tomando decisões com base em princípios éticos, democráticos, inclusivos, sustentáveis e solidários.

Fonte: Brasil (2018, p. 9-10)

De certo modo, compreende-se que essas dez competências dizem respeito:

1. ao conhecimento;
2. ao pensamento científico, crítico e criativo;
3. ao repertório cultural;
4. à comunicação;
5. à cultura digital;
6. ao trabalho e projeto de vida;
7. à argumentação;
8. ao autoconhecimento e autocuidado
9. à empatia e cooperação;
10. à responsabilidade e cidadania⁴.

⁴ Conforme apresenta a Revista Educação. Disponível em: <https://revistaeducacao.com.br/2018/10/05/bncc-competenciasgerais/>. Acesso em: 15 de julho de 2022.

Ao se analisarem essas dez competências, nota-se que aquelas relacionadas ao conhecimento (1), ao pensamento científico, crítico e criativo (2), à comunicação (4) e à cultura digital (5) são as que mais vão ao encontro desta pesquisa, cujo embasamento teórico (relacionado à Linguística de *Corpus*) prevê que a língua seja investigada pelo viés científico do empirismo e com o auxílio de ferramentas tecnológicas e *softwares* computacionais. Esse ponto central da LC não apenas é congruente com as competências supracitadas, como também se associa ao enfoque de exploração de novas perspectivas de pesquisa e obtenção de informações dado ao componente curricular de Língua Inglesa pela BNCC.

Ainda no que tange às competências gerais, a BNCC prevê que elas tratem da “mobilização de conhecimentos (conceitos e procedimentos), habilidades (práticas cognitivas e socioemocionais), atitudes e valores para resolver demandas complexas da vida cotidiana, do pleno exercício da cidadania e do mundo do trabalho” (Brasil, 2018, p. 8).

Percebe-se, assim, que a visão de educação é consideravelmente abrangente, pois se preocupa não só com os saberes e objetos do conhecimento, mas também com as práticas e ações necessárias para se atuar de forma crítica na sociedade, sem mencionar o autoconhecimento no que compete às emoções, por exemplo. Essas considerações alinham-se com o conceito de educação integral. Conforme Brasil (2018, p. 14),

[...] a BNCC afirma, de maneira explícita, o seu compromisso com a educação integral. Reconhece-se, assim, que a Educação Básica deve visar à formação e ao desenvolvimento humano global, o que implica compreender a complexidade e a não linearidade desse desenvolvimento, rompendo com visões reducionistas que privilegiam ou a dimensão intelectual (cognitiva) ou a dimensão afetiva. Significa, ainda, assumir uma visão plural, singular e integral da criança, do adolescente, do jovem e do adulto – considerando-os como sujeitos de aprendizagem – e promover uma educação voltada ao seu acolhimento, reconhecimento e desenvolvimento pleno, nas suas singularidades e diversidades.

Conclui-se, assim, que a educação integral expande o foco de desenvolvimento do aluno para muito além do âmbito acadêmico (dos conteúdos escolares). Dessa maneira, perspectivas, tais como o aprimoramento de competências e habilidades socioemocionais, a reflexão sobre a sociedade e o mundo do trabalho, e a participação no mundo digital etc., passam a ser enfatizadas. Assim sendo, na educação integral os alunos são concebidos como atuentes ativos no processo de aprendizagem e que

precisam ser acolhidos, reconhecidos e respeitados em suas totalidades. Isso tudo no intuito de se desenvolver no aluno o protagonismo e o seu projeto de vida.

2.1.1 A BNCC e a relação com as metodologias ativas

Em vista de os fundamentos pedagógicos da BNCC enfatizarem o desenvolvimento de competências – que envolvem a mobilização do conhecimento, de habilidades, de atitudes e de valores para se resolverem demandas da vida cotidiana – e a educação integral do estudante, que os percebe como sujeitos da aprendizagem, compreende-se que a Base fomenta as práticas pedagógicas pautadas nas metodologias ativas. Por esse motivo, elas são mobilizadas nesta seção.

As metodologias ativas, apesar de configurarem estratégias inovadoras em relação ao processo de ensino e aprendizagem, possuem bases epistemológicas que não são, necessariamente, tão recentes. Segundo Lovato *et al.* (2018), desde o século XVIII, quando houve as revoluções liberais na Europa e a independência dos Estados Unidos, vem se questionando os métodos mais tradicionais de ensino. Essa mudança de foco a favor de uma aprendizagem em que a busca pelo conhecimento parta do aluno se consolida, no entanto, com o movimento Escola Nova (Lovato *et al.*, 2018). Entre os defensores dessa metodologia centrada na aprendizagem pela experiência, destaca-se John Dewey, que enfatizava a relevância de se aprender fazendo (*learning by doing*) (Almeida, 2018). Somadas a isso, as contribuições de John Dewey enfatizaram a demanda de se relacionar teoria à prática, considerando o contexto diário do aluno (Daros, 2018).

Além de a Escola Nova trazer à tona a questão da contextualização com a vida do aluno, Daros (2018) aponta outro ponto positivo no que tange às bases teóricas das metodologias ativas, que é a teoria defendida por Ausubel. O filósofo corrobora com as teorias de John Dewey e de outros precursores do escolanovismo ao sustentar que os conhecimentos prévios dos estudantes precisam ser levados em consideração para que a aprendizagem seja significativa (Daros, 2018).

Considerados, então, os aspectos principais do movimento Escola Nova (a aprendizagem pela experiência, o aprender fazendo, a relação entre teoria e prática, o contexto diário do aluno e a aprendizagem significativa), passa-se à

conceitualização das metodologias ativas, que englobam, de certo modo, todos esses pressupostos.

Embora pesquisas em Neurociência apontam que toda aprendizagem é, de certa forma, ativa (Moran 2019), por metodologias ativas, compreende-se as “alternativas pedagógicas que colocam o foco do processo de ensino e aprendizagem nos aprendizes [...]” (Moran, 2019, p. 7). Desse modo, entende-se que, diferentemente de uma aprendizagem por transmissão, do professor para o aluno, que também tem seu devido valor, como aponta Moran (2018), por intermédio das metodologias ativas, o estudante assume o papel de protagonista central desse processo de aprendizagem, que é ativo, como asseguram Lovato *et al.* (2018), Moran (2018) e Camargo (2018).

Enquanto protagonista da aprendizagem, o aluno, por meio das metodologias ativas, adquire o conhecimento a partir de questionamentos, investigações, descobertas, experimentações, compartilhamentos e resoluções de problemas (Moran, 2018, 2019). Isso tudo é feito de modo a considerar as situações reais vivenciadas pelos alunos (Moran, 2015), o que corrobora inclusive com a contextualização e, por conseguinte, com a aprendizagem significativa.

Todo esse percurso da aprendizagem de forma ativa é desenvolvido por meio de atividades interativas entre os alunos, sendo o conhecimento constituído de modo colaborativo (Camargo, 2018; Moran, 2019). Além disso, a construção do saber, que é feita de forma interligada, crítica, reflexiva e híbrida (Moran, 2018, 2019), torna-se ainda mais profunda quando feita de modo multissensorial, valorizando-se dos conhecimentos prévios dos alunos (Moran, 2018). Ademais, esse processo de aprendizagem ativa avança, segundo Moran (2018), em espiral, ou seja, parte daquilo que o aluno já sabe (os conhecimentos prévios) para desenvolver novos saberes, que progridem, gradativamente, dos níveis mais simples para os mais complexos (Moran, 2018).

Posto tudo isso, percebe-se que, de certa forma, as metodologias ativas promovem uma inversão no foco, que passa do ensinar para o aprender. Isso ocorre, pois, “ao contrário do método tradicional, que primeiro apresenta a teoria e dela parte, o método ativo busca a prática e dela parte para a teoria” (Moran, 2019, p. 9). Com isso, a ênfase no aluno demanda que ele, junto ao professor, seja corresponsável pelo processo (Moran, 2019).

Já com relação aos professores, Lovato *et al.* (2018) afirmam que eles passam a assumir o papel de mediadores e facilitadores da aprendizagem. Moran (2015) vai além, acrescentando que os docentes atuam como orientadores e curadores. Isso porque eles ajudam na seleção das informações pertinentes, dentre tantas disponíveis, especialmente no mundo virtual, além de acolherem, valorizarem, incentivarem e orientarem os estudantes.

Diante do exposto, acredita-se que “ensinar e aprender tornam-se fascinantes quando se convertem em processos de pesquisa constantes, de questionamento, de criação, de experimentação, de reflexão e de compartilhamento crescentes [...]” (Moran, 2018, p. 37).

Em conclusão, a BNCC, certamente, representa uma série de mudanças em relação à Educação Básica no Brasil. Embora tenham sido discutidos até este momento apenas alguns pontos mais centrais, acredita-se que fica evidente que esse documento estabelece uma série de demandas novas que precisam ser atendidas em nível nacional. Por esse motivo, julga-se pertinente estabelecer algumas ponderações e/ou visões crítico-reflexivas em relação a esse documento.

2.1.2 BNCC: um olhar reflexivo

Considerados os aspectos mais globais da Base Nacional Comum Curricular e a relação com as metodologias ativas, faz-se pertinente estabelecer um olhar reflexivo em relação a esse documento. Entre tantos pontos a serem ponderados, destacam-se, nesta seção, os que podem ser correlacionados, de algum modo, com o escopo de estudo abarcado nesta tese. Estudos sobre como a BNCC impacta a educação nacional foram aprofundados nesta pesquisa, pois se enquadram melhor no escopo das pesquisas voltadas à área da educação e das políticas públicas educacionais.

Entretanto, algumas colocações enriqueceram este trabalho. A primeira delas diz respeito à implementação das aprendizagens essenciais a todo o território nacional. Embora isso seja visto de forma positiva por Fidalgo, Sprenger e Ramos (2021), que apontam que essa iniciativa já foi utilizada, inclusive, em outros países que apresentam resultados de aprendizagem melhores pelos estudantes. As autoras defendem que essa implementação do mínimo a ser ensinado não pode acontecer desacompanhada da formação docente. Isso porque a efetiva aplicação dessas

aprendizagens mínimas demandará das redes de ensino e dos professores a reformulação dos currículos, com o desafio de adequação aos seus devidos contextos econômicos, culturais e educacionais. Nesse sentido, elas pontuam o quanto isso tudo se torna desafiador aos professores e afirmam que a “BNCC de Língua Inglesa pode trazer contribuições e avanços para o ensino de língua inglesa no Brasil desde que a formação docente não seja negligenciada” (Fidalgo; Sprenger; Ramos, 2021, p. 261).

Ainda em relação às aprendizagens essenciais, Dornelles e Irala (2021, p. 205) criticam que um documento como a BNCC, que “compõe uma política educacional nacional para a educação básica, deve-se constituir um orientador e não um normatizador de políticas curriculares locais”. Desse modo, elas não defendem a normatização, mas sim o caráter formativo nas práticas de ensino e aprendizagem. Vê-se, portanto, que a formação continuada em serviço dos professores é um dos pontos a serem pensados quando se reflete a BNCC, tanto no sentido de como implementá-la quanto de refleti-la na direção do que se deve precisamente desenvolver, a depender do contexto, e como isso precisa acontecer. Desse modo, espera-se que as propostas realizadas nesta tese possam, em alguma medida, contribuir com a formação docente, pois se propõe, além das atividades didáticas voltadas à língua inglesa que podem viabilizar algumas das colocações trazidas na Base, um curso de formação continuada em serviço que capacite os docentes a utilizarem as ferramentas necessárias para tais atividades.

Outra crítica das autoras à BNCC está centrada no conceito de educação linguística. Esse ponto abordado pela Base remete, segundo Dornelles e Irala (2021), a outro conceito, o da análise linguística. Para elas, “a BNCC se beneficiaria de uma breve exposição sobre esse conceito [...]” (Dornelles; Irala, 2021, p. 209). Elas prosseguem essa crítica argumentando que a falta, de certo modo, da explicitação desses conceitos enfraquece a reflexão docente sobre a prática pedagógica. Para as autoras, portanto, é preciso que a BNCC torne “mais explícitos os elementos teóricos-chave da nova proposição” (Dornelles; Irala, 2021, p. 212). Isso não apenas no que tange à análise linguística, como também à abordagem dela por parte dos estudantes, que, segundo a BNCC, precisa ser feita de modo indutivo, como alegam Dornelles e Irala (2021).

Dessa maneira, entende-se que a proposta trazida nesta tese pode oferecer uma forma de se alcançar essa análise linguística a ser feita de modo indutivo pelos

estudantes, que é justamente pela utilização da Aprendizagem Movida por Dados (a ser tratada mais devidamente em seção posterior) e com o auxílio das ferramentas da LC. Esse atrelamento da DDL com essas demandas da BNCC se mostra relevante porque o cerne dessa metodologia é tornar o estudante um pesquisador/investigador da língua, que levantará, a partir do contato com os dados da língua, formulações e hipóteses de como o idioma funciona.

Esse estímulo a atividades em que o estudante descubra a língua de modo indutivo, ou ainda, que ele atue como pesquisador da língua é também mencionado por Ribas (2021), que apresenta, em seu capítulo, reflexões acerca do papel do letramento e das tecnologias digitais na BNCC. Para a autora, em relação ao letramento digital e ao ensino de língua inglesa, é “necessário propor atividades e o desenvolvimento de habilidades que estimulem uma postura de pesquisador nos aprendizes [...] (Ribas, 2021, p. 237).

Além dessa consideração, a autora reflete ainda que, sobre o letramento digital e os gêneros digitais, é importante que a BNCC não perpetue “um modelo de ensino fechado e antigo, que perdura até hoje nas escolas [...]” (Ribas, 2021, p. 237). Com isso, entende-se, mais uma vez, a relevância da Aprendizagem Movida por Dados como uma forma de diversificação dos modelos de ensino, de modo a proporcionar um modelo não apenas moderno, por meio das ferramentas e *softwares* disponibilizados a partir dos avanços dos estudos em Linguística de *Corpus*, como também indutivo, no que tange ao papel do estudante. Afinal, ele passa a ser um pesquisador da língua quando se emprega tais ferramentas por meio da Aprendizagem Movida por Dados.

Em suma, a BNCC é um documento que possibilita muita discussão e reflexão crítica. Certamente, as análises e revisões da BNCC vão para muito além dos pontos levantados nesta seção. Porém, como se observou, alguns deles acabam abrindo espaço para a implementação das propostas trazidas nesta tese, tais como a formação docente e a utilização de metodologia de ensino em que o estudante assuma o papel de pesquisador e descubra informações sobre a língua de forma indutiva. Além disso, outro aspecto a ser considerado é que a BNCC ainda abre espaço para a utilização das tecnologias digitais de informação e comunicação, o que é visto de forma positiva por Fidalgo, Sprenger e Ramos (2021) ao realizarem suas

reflexões sobre esse documento. Isso também é um fator que beneficia os argumentos e propostas defendidas neste estudo.

Já para a seção seguinte, lidar-se-á com a questão da adequação curricular da qual o estado de São Paulo, assim como todas as outras redes de ensino, precisou realizar, no intuito de responder às demandas articuladas pela BNCC, atendendo às competências e às habilidades essenciais normatizadas por esse documento.

2.2 O Currículo Paulista – etapa do Ensino Médio

Como já mencionado, esta seção aborda o Currículo Paulista na etapa do Ensino Médio apenas, pois ela é o foco desta pesquisa. No entanto, a publicação da BNCC fez com que os currículos de todas as etapas da Educação Básica fossem reelaborados. Isso é, inclusive, o que pontuam Fidalgo, Sprenger e Ramos (2021, p. 241) ao afirmarem que, com a publicação da BNCC, “caberá às redes de ensino, públicas e privadas, a tarefa de construir currículos com base nas orientações dessa referência”, ou seja, das normativas implementadas pelo documento que definem as aprendizagens essenciais da Educação Básica.

Essa reestruturação curricular não isentou o Estado de São Paulo da necessidade de adequação. Dessarte, apesar de ter sido concluído em 2020, o Currículo Paulista – etapa do Ensino Médio começou a ser desenvolvido em 2018 e contou com o apoio das universidades estaduais USP, UNESP e UNICAMP, dentre outros órgãos, com o intuito de garantir a melhoria da qualidade da educação (São Paulo, 2020).

De modo geral, pontua-se que o Currículo Paulista foi redigido a partir das novas demandas trazidas com a publicação da Base Nacional Comum Curricular. Uma delas, por sinal, relaciona-se às dez competências gerais. Nesse sentido, o Currículo Paulista reforça a necessidade de atendimento dessas competências gerais como forma de garantia de que os estudantes atuem na sociedade contemporânea. Em se tratando dos saberes e das competências, o Currículo Paulista

“indica claramente o que o estudante deve ‘saber’ (em termos de conhecimentos, habilidades, atitudes e valores) e, sobretudo, o que deve ‘saber fazer’, considerando a mobilização desses conhecimentos, habilidades, atitudes e valores para resolver demandas complexas da vida cotidiana, do pleno exercício da cidadania e do mundo do trabalho” (São Paulo, 2020, p. 30).

Nota-se, assim, o caráter prescritivo em relação à aprendizagem comum a todo currículo educacional e ressalta-se que, no caso do documento paulista, prega-se que o desenvolvimento das competências englobe tanto o que os estudantes precisam saber quanto o que eles precisam saber fazer para que possam atuar ativa e criticamente na sociedade. Essa combinação dos objetos de conhecimento e dos desdobramentos deles de forma prática na vida dos estudantes evidencia ainda o quanto questões como a interdisciplinaridade, o papel central do aluno (protagonismo), a educação integral e o projeto de vida se fazem presente nesse documento (São Paulo, 2020).

Em relação à interdisciplinaridade, o Currículo Paulista frisa que a aprendizagem supere “a fragmentação disciplinar do conhecimento, estimulando sua contextualização e aplicação na vida real, para dar sentido ao que se aprende.” (São Paulo, 2020, p. 32). Nesse sentido, então, valorizam-se abordagens pedagógicas mais práticas e interativas, que superem, de certo modo, a segmentação por componentes curriculares (as disciplinas). Além disso, o documento preza ainda por ações pedagógicas que possibilitem aos alunos assumirem um papel ativo e participativo na construção da aprendizagem, exercendo o chamado protagonismo (São Paulo, 2020).

No tocante à formação global do aluno, é que se dispõe a educação integral, a qual o Currículo Paulista também adota em conformidade com a BNCC. Esse viés da educação integral implica no desenvolvimento dos estudantes em suas dimensões intelectual, física, socioemocional e cultural (São Paulo, 2020).

Em detrimento dessa educação integral, contribui-se ainda para que os estudantes percebam quais são os seus objetivos de vida, o que é denominado como projeto de vida (São Paulo, 2020). Desse modo, o Currículo Paulista prescreve a importância da escola ser um espaço em que se discuta e se reflita sobre os sonhos dos alunos e como eles precisam agir tanto na escola quanto na sociedade para alcançar essas aspirações (São Paulo, 2020).

Em síntese do que se discutiu até então, julga-se pertinente acrescentar que, no Currículo Paulista, afirma-se que

A principal característica da etapa final da Educação Básica é a flexibilização curricular, tendo como objetivo consolidar, aprofundar e ampliar a formação integral do estudante. A finalidade é desenvolver o conjunto de competências e habilidades, propiciando protagonismo ao jovem e maior autonomia e

assertividade nas suas escolhas, por meio do desenvolvimento do projeto de vida em consonância com os princípios da justiça da ética e da cidadania (São Paulo, 2020, p. 46).

Já com relação à estruturação e/ou organização do Ensino Médio, o Currículo Paulista, assim como a Base Nacional Comum Curricular, prevê que essa etapa seja composta por uma Formação Geral Básica comum a todos os estudantes (envolvendo as áreas de Linguagens e suas Tecnologias, Matemática e suas Tecnologias, Ciências da Natureza e suas Tecnologias e Ciências Humanas e Sociais Aplicadas) com carga horária máxima de 1.800 horas e por Itinerários Formativos (parte diversificada e flexível) com carga mínima de 1.200 horas (São Paulo, 2020).

Sobre a área de Linguagens e suas Tecnologias, na qual esta pesquisa se enquadra, o Currículo Paulista determina cinco campos de atuação social para a proposta de práticas de linguagem. São eles: o campo da vida pessoal, o campo das práticas de estudo e pesquisa (ao qual essa pesquisa se mostra fortemente atrelado, uma vez que assume a Linguística de *Corpus* e a Aprendizagem Movida por Dados como escopo de trabalho), o campo jornalístico-midiático, o campo de atuação na vida pública e o campo artístico-literário (São Paulo, 2020).

Por fim, no que diz respeito ao componente curricular de Língua Inglesa (objeto de pesquisa deste estudo), o Currículo Paulista defende que ele se centre na ampliação da autonomia e do protagonismo dos alunos, proporcionando-lhes o pensamento crítico em relação à língua e seus usos e intencionalidades, o que inclui a relação com a tecnologia (São Paulo, 2020). Acredita-se, portanto, que esses preceitos estejam de acordo com o escopo teórico e metodológico adotado nesta pesquisa, que é o da Linguística de *Corpus* e sua relação com o ensino de línguas e Aprendizagem Movida por Dados, que são abordados na sequência.

2.3 A Linguística de *Corpus*

Exploradas as temáticas referentes aos documentos que descrevem o regimento da Educação Básica a nível nacional e estadual (São Paulo), respectivamente a BNCC e o Currículo Paulista, iniciam-se, nesta seção, as considerações a respeito da Linguística de *Corpus* (LC), apontada por McEnery e Hardie (2012, p. 1) como uma abordagem heterogênea e possível de ser aplicada a diferentes vertentes da Linguística.

Antes de defini-la, porém, compete analisar que, apesar de a Linguística de *Corpus* estar fortemente atrelada à tecnologia, como afirma Berber Sardinha (2004), já havia estudos em *corpora* mesmo antes de o computador existir. Eles datam do século XX e enfatizavam o ensino de línguas. Um exemplo de *corpus* não computadorizado mencionado por Berber Sardinha (2004) é o SEU (*Survey of English Usage*), que serviu também como inspiração para os *corpora* que foram desenvolvidos posteriormente e de forma eletrônica.

No entanto, esse aspecto tecnológico está tão arraigado na LC a ponto de se fazer presente em muitas das definições dadas a essa área dos estudos da linguagem. Veja a conceitualização de *corpus* a seguir:

um conjunto de dados linguísticos (pertencentes ao uso oral ou escrito da língua, ou ambos), sistematizados segundo determinados critérios, suficientemente extensos em amplitude e profundidade, de maneira que sejam representativos da totalidade do uso linguístico ou de algum de seus âmbitos, dispostos de tal modo que possam ser **processados por computador**, com a finalidade de propiciar resultados vários e úteis para a descrição e análise (Sanchez, 1995, p. 8-9 *apud* Berber Sardinha, 2004, p. 338, grifo meu).

Destaca-se que o *corpus*, ou seja, o conjunto de dados linguísticos, precisa ser organizado de forma a ser lido pelos computadores para que os resultados gerados sejam analisados e descritos de forma criteriosa. Outra definição amplamente difundida, especialmente no Brasil, é a do pesquisador Berber Sardinha, que alega:

A linguística de *corpus* ocupa-se da coleta e da exploração de *corpora*, ou conjuntos de dados linguísticos textuais coletados criteriosamente, com o propósito de servirem para a pesquisa de uma língua ou variedade linguística. Como tal, dedica-se à exploração da linguagem por meio de evidências empíricas, **extraídas por computador** (Berber Sardinha, 2000, p. 325, grifo meu)

Novamente, observa-se, a preocupação do autor em expressar que a exploração dos dados precisa ocorrer por intermédio do computador. Na verdade, convém considerar que os avanços na tecnologia levaram a uma série de vantagens aos linguistas de *corpus*, tais como a agilidade e a eficiência no processamento das informações, fatores que certamente contribuíram para a popularização desse campo de pesquisas linguísticas (Reppen; Simpson-Vlach, 2020).

A LC, segundo Biber *et al.* (1998, p. 4), por compreender análises delimitadas por critérios que envolvem o computador, mescla técnicas qualitativas e quantitativas

dos padrões encontrados em textos naturais. Essas considerações do autor, além de corroborarem com a necessidade do viés tecnológico, ressalta outro ponto congruente nas definições estudadas até então, que é a questão de a LC lidar com dados naturais. Por sinal, essa ressalva aparece na própria definição do autor, que afirma que o *corpus* se trata de “uma ampla e criteriosa seleção de textos naturais”⁵ (Biber *et al.*, 1998, p. 12, tradução minha). Outro teórico que demonstra concordar com a especificidade dos textos naturais é Sinclair (1991, p. 171, tradução minha), que afirma que “um *corpus* é uma coleção de textos naturais escolhidos para caracterizar um estado ou uma variação da língua”⁶. Mas afinal o que são esses textos naturais? Eles se referem ao fato de “os textos presentes nos *corpora* necessitarem de terem sido produzidos com um objetivo comunicativo autêntico”⁷ (Frankenberg-Garcia, 2016, p. 383, tradução minha) ou então de serem “textos coletados de fontes que ocorrem naturalmente, ao invés daqueles redigidos para uma pesquisa ou questionário”⁸ (Reppen; Simpson-Vlach, 2020, p. 92, tradução minha).

Entretanto, não se pode deixar de se considerarem aquelas situações cujos objetivos de pesquisa são bem específicos, o que Reppen e Simpson-Vlach (2020) denominam de *corpus* especializado. Um exemplo bastante comum de dados que não condizem com essa demanda de um *corpus* se pautar em textos naturais é o caso dos *corpora* de aprendizes – “coleções eletrônicas de dados linguísticos produzidos por aprendizes de uma segunda língua, sejam eles aprendizes de segunda língua ou de língua estrangeira”⁹ (Granger, 2013, p. 1, tradução minha).

Em se tratando disso, essa ressalva de um *corpus* envolver características específicas que atendam a uma finalidade em particular daquilo que se deseja analisar na língua também se mostra como um consenso entre os pesquisadores. Nota-se isso tanto nas definições vistas anteriormente (em “dados linguísticos [...] sistematizados segundo determinados critérios” e em “dados linguísticos textuais coletados criteriosamente”) quanto no que prescreve Frankenberg-Garcia (2016, p. 383,

⁵ “[...] a *corpus* is a large and principled collection of natural texts.”

⁶ “A *corpus* is a collection of naturally-occurring language text, chosen to characterize a state or variety of a language”.

⁷ “Many scholars also believe that the texts in *corpora* need to have been produced for authentic communicative goals”.

⁸ “The use of natural texts means that language has been collected from naturally occurring sources rather than from surveys or questionnaires”.

⁹ “Learner *corpora* are electronic collections of language data produced by L2 learners, that is, second or foreign-language learners”.

tradução minha) quando afirma que “os textos compilados em um *corpus* são normalmente selecionados de forma a se enquadrarem a um propósito específico¹⁰”. Reppen e Simpson-Vlach (2020) também advogam a favor de o *corpus* precisar ser compilado a partir de critérios específicos para que se alcance a devida representatividade. Sendo assim, “as características linguísticas ou questões disparadoras da pesquisa moldarão a coletânea de textos usada na compilação do *corpus*¹¹” (Reppen; Simpson-Vlach, 2020, p. 92, tradução minha).

Outro aspecto relevante e comum entre os teóricos que definem a LC é a questão da representatividade, haja vista que um *corpus* “é tido como representativo da linguagem, de um idioma, ou de uma variedade dele” (Berber Sardinha, 2004, p. 22). Com relação a isso, é pertinente destacar que, embora estudos mais incipientes considerassem que, quanto mais vasta a extensão de um *corpus*, mais chances ele teria de representar uma língua ou uma variedade dela. Considerações mais recentes parecem concordar que “o tamanho ideal de um *corpus* depende de para que ele será utilizado¹²” (Frankenberg-Garcia, 2016, p. 384, tradução minha). Assim sendo, a autora acredita que o bom senso vai ditar o quão extenso um *corpus* precisa ser para que possa ser representativo (Frankenberg-Garcia, 2016).

Além disso, ao se pensar na LC de forma geral, deve-se considerar que ela surgiu como um rompimento à forma tradicional de investigação linguística, que enfatizava somente a estrutura. Desse modo, essa vertente dos estudos linguísticos salienta, conforme Biber *et al.* (1998, p. 1), o uso da língua e não as realizações dela, que seriam possíveis de serem previstas com base em uma linha teórica (oposição ao Gerativismo). Isso evidencia, portanto, o caráter empírico da LC, uma vez que ela se baseia em evidências e padrões encontrados em dados linguísticos compilados a partir de textos naturais (escritos ou orais).

Nesse sentido, inclusive, Viana (2011, p. 26) descreve que, ao invés de estudar o funcionamento da língua por um viés intuitivo, a LC observa os fenômenos linguísticos a partir da forma como a língua é usada pelos usuários. Nas palavras do próprio autor: “[...] para compreender a questão linguística é necessário basear o estudo em um alicerce empírico no qual os resultados advêm da observação de

¹⁰ “[...] *the texts compiled into a corpus are normally selected so as to be fit for a particular purpose.*”

¹¹ “*The linguistic features or research questions being investigated will shape the collection of texts used in creating the corpus.*”

¹² “*The ideal size of a corpus will ultimately depend on what the corpus is for.*”

exemplos reais” (Viana, 2011, p. 26). Nota-se, assim, que o autor reforça esse caráter inovador da LC no sentido de analisar a língua a partir do uso, como mencionou Biber *et al.* (1998).

Em síntese, portanto, das colocações levantadas nesta seção, julga-se pertinente trazer uma reunião dos pontos-chaves no que se refere às características principais da LC. São eles:

- Utilização de análise computacional;
- Análises quantitativas e qualitativas;
- Criteriosidade e objetivos específicos na seleção dos textos e na definição da extensão do *corpus*;
- Empirismo ao analisar padrões em textos naturais (Reppen; Simpson-Vlach, 2020).

Em relação ao quesito da classificação dos *corpora*, ambos Berber Sardinha (2004) e Viana (2011) apresentam uma organização e estruturação das diferentes formas de se classificar um *corpus*.

Segundo Viana (2011), um aspecto a se levar em consideração é a abrangência. Sendo assim, pode-se classificar um *corpus* em geral ou especializado. Como exemplo de um *corpus* geral, Viana (2011) diz que se pode pensar no *British National Corpus* (BNC) e o *American National Corpus* (ANC). Relacionado ao especializado, o autor menciona o COMPARA, que, de acordo com a definição do próprio site¹³, serve para estudar as traduções entre o português e o inglês.

Ao se pensar no meio, o autor aponta que um *corpus* pode ser oral ou escrito. McEnery e Hardie (2012, p. 4) destacam que, em geral, um *corpus* oral consome mais tempo no processo de compilação quando comparado a um *corpus* escrito.

Viana (2011) ressalta também o quesito tempo, o que faz com que um *corpus* seja considerado sincrônico (envolvendo um determinado espaço de tempo), diacrônico (envolvendo diferentes períodos), contemporâneo (como o caso do COCA – *Corpus of Contemporary American English*, que possui relação com a atualidade) ou histórico.

Ao se julgar se um *corpus* se renova ou não, Viana (2011) descreve que ele pode ser tipificado entre dinâmico (em que sempre são feitas alterações com certa frequência) e estático (em que alterações não são permitidas). Nesse sentido da

¹³ Endereço eletrônico: <https://www.linguateca.pt/COMPARA/>.

dinamicidade ou reciclabilidade dos *corpora*, enfatiza-se também o *corpus* que se classifica como *do-it-yourself (DIY) corpora*. Carvalho, Laranja e Pinto (2021, p. 66, tradução minha) o definem como “coleções de textos que podem ser usadas para responder questões atuais ou para comprovar hipóteses, além de poderem ser descartadas ou ampliadas com novos textos em um momento posterior¹⁴.”

No que se refere às línguas, Viana (2011) diz que um *corpus* pode ser monolíngue ou multilíngue, podendo este último ser paralelo (alinhado ou não alinhado) e/ou não paralelo. No que envolve os produtores do *corpus*, o autor propõe uma classificação distinta em que se considere *corpus* de primeira língua, de segunda língua ou de língua estrangeira. Isso engloba os *corpora* de aprendizes descritos anteriormente (Granger, 2013).

Viana (2011) considera ainda que um *corpus* pode ser classificado como *corpus* de estudo quando seu uso se concretizar no intuito de explicar uma determinada variação da língua ou como *corpus* de referência quando seu uso representar o padrão de comparação para o *corpus* de estudo. O quadro a seguir resume essa discussão sobre as classificações dos *corpora*:

¹⁴ “In this paper we argue for the use of do-it-yourself (DIY) corpora, or recyclable corpora, as you wish, which are collections of texts that can be used to answer current questions or test hypotheses and, afterwards, be disposed of or enlarged with new texts.”

Quadro 2. Taxonomia de *corpora*

Critérios		Corpus		
Abrangência		Geral Especializado		
Meio		Oral Escrito		
Tempo	Número de períodos históricos	Sincrônico Diacrônico		
	Relação com a atualidade	Contemporâneo Histórico		
Renovação		Dinâmico Estático		
Línguas	Número	Monolíngue Multilíngue	Paralelos Não paralelos	Alinhados
	Produtores	De primeira língua (L1) De segunda língua (L2) De língua estrangeira (LE)		
Emprego	Estudo Referência			

Fonte: Viana (2011, p. 30)

Nesta seção, buscou-se apresentar a LC, destacando seu caráter heterogêneo e aplicável às diferentes pesquisas em linguagem e enfatizando tanto suas características gerais quanto suas possíveis classificações.

2.4 A Linguística de *Corpus* e o ensino de línguas

Como menciona Boulton (2010, p. 534), “as potenciais aplicações de *corpora* eletrônicos no ensino e aprendizagem de línguas vem recebendo atenção considerável nos últimos anos¹⁵.” No entanto, esse cenário nem sempre foi assim.

Como visto na seção anterior, estudos que envolviam *corpora* de alguma forma já existiam mesmo antes do advento do computador e tinham, desde então, relação com o ensino. Nesse sentido, Berber Sardinha (2000) destaca o trabalho de Thorndike em 1921, que envolveu o levantamento manual de um *corpus* e que acarretou mudanças no ensino de línguas.

Ainda sob a perspectiva histórica, Berber Sardinha (2000) pontua a relevância das décadas de 80 e 90 no que tange à relação da LC com o ensino de línguas. No entanto, diferentemente dos trabalhos do início do século XX (como o de Thorndike em 1921), nessas décadas já se observava a presença maior da tecnologia computacional. Um marco que ambos Berber Sardinha (2000) e Romer (2010) apontam para esse período foi o projeto COBUILD – inovador não só para a LC como também para o ensino de idiomas.

Mais precisamente, essa referida proposta tratou-se do “primeiro dicionário a ser compilado a partir de um *corpus* computadorizado” (Berber Sardinha, 2000, p. 332). Além do lançamento desse dicionário voltado ao ensino de inglês, o projeto COBUILD envolveu a elaboração de gramáticas e livros didáticos que também eram voltados ao ensino, como destaca Berber Sardinha (2000, p. 329).

Há de se ressaltar, entretanto, que o projeto COBUILD se pautou na compilação de textos produzidos por falantes nativos e, portanto, corrobora com as considerações de Nesselhauf (2004, p. 125) no âmbito de o *corpus* nativo ser responsável pela melhoria no ensino de línguas. No entanto, a autora acrescenta que o *corpus* de aprendizes é também pertinente ao ensino, pois ele expõe as dificuldades típicas dos aprendizes de determinada língua (Nesselhauf, 2004, p. 126). Granger (1998, 2002) também reconhece que a utilização de *corpus* de aprendizes oferece vantagens para o ensino, justamente porque evidencia as dificuldades linguísticas encontradas pelos alunos.

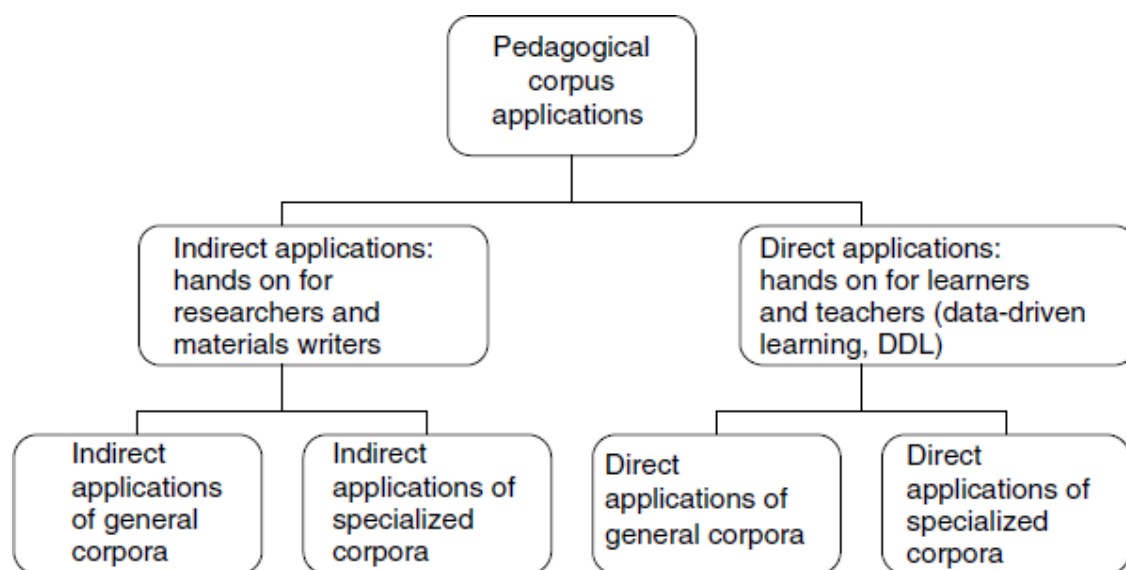
¹⁵ “The potential applications of electronic corpora in language teaching and learning have received considerable attention in recent years”.

Nesselhauf (2004, p. 126) salienta ainda que o *corpus* de aprendizes, além de informar as dificuldades dos alunos, fornece informações ao processo de aquisição de uma segunda língua. Nesse âmbito, destacam-se as pesquisas de Granger (1998; 2002) que não só definiram essa vertente da LC, como também investigaram o processo de aquisição de uma segunda língua por meio dos *corpora* de aprendizes.

Associados ao projeto COBUILD da década de 80, esses estudos e pesquisas relacionados aos *corpora* de aprendizes constituem também o motivo de autores como Berber Sardinha (2000) e Romer (2010) destacarem esse período (anos 80 e 90) como centrais à relação da LC com o ensino de línguas.

Compreendida essa relação entre a LC e o ensino de idiomas por uma perspectiva histórica, passa-se então para um viés analítico de como essas aplicações de *corpora* se dão em sala de aula. Frankenberg-Garcia (2016) e Romer (2010) parecem concordar em apresentar que essa relação se constrói tanto de forma indireta quanto direta. Considere a figura a seguir:

Figura 3. O uso de *corpora* na aprendizagem e no ensino de línguas



Fonte: Romer (2010, p. 19).

Como sugere a figura, nota-se que, por um viés indireto das aplicações pedagógicas do *corpus*, entende-se, por um lado, o desenvolvimento de materiais (livros didáticos, dicionários etc.), a partir de informações levantadas em estudos baseados em *corpora*, sejam eles gerais ou especializados. Por outro lado, o aspecto

direto das aplicações pedagógicas do *corpus* diz respeito ao seu manuseio por professores e estudantes.

Em relação ao sentido indireto da aplicação de *corpus* ao ensino, Frankenberg-Garcia (2016, p. 390, tradução minha) alega que “dicionários de aprendizes modernos fornecem exemplos baseados em *corpus* que podem ajudar os estudantes a entenderem como as palavras são usadas em contextos, além de utilizarem dados de *corpora* de aprendizes para chamar a atenção aos erros mais recorrentes¹⁶.” Ela ainda destaca que, além dos dicionários, há uma crescente em compêndios gramaticais e materiais didáticos de ensino de línguas que se baseiam em *corpus* (Frankenberg-Garcia, 2016, p. 390).

Ainda segundo a autora, as informações proporcionadas por análises com base em *corpora* (tais como a frequência de palavras, colocações e/ou aspectos gramaticais) corroboram para a tomada de decisões do que se deve ou não incluir em materiais didáticos, dicionários e gramáticas. Essa relevância da frequência é explicada por Leech (2011, p. 12, tradução minha) pela simples relação de, “quanto mais frequente a palavra for em relação ao uso da língua, maiores as chances de ela ser útil ao aprendiz¹⁷.”

Já com relação às aplicações mais diretas do *corpus* ao ensino, Frankenberg-Garcia (2016, p. 391) acredita que “os *corpora* podem oferecer mais recursos da língua e apresentar soluções para questões da linguagem que não foram tratadas em materiais e em recursos linguisticamente editados, o que impulsiona a autonomia dos alunos¹⁸.” Isso se dá justamente pelo *corpus* ser, entre outras coisas, uma compilação de textos naturais, como visto anteriormente e proporcionar, como sugere Reppen (2010), materiais e textos autênticos, ou seja, em que a língua tenha sido usada com uma finalidade comunicativa e significativa, ao invés de um propósito particular de pesquisa e/ou estudo.

Contudo, Frankenberg-Garcia (2016, p. 391) reconhece que os professores de idiomas nem sempre têm tempo hábil de compilar um *corpus* ou de conduzir uma

¹⁶ “*Modern learners’ dictionaries also provide corpus-based examples that can help learners see how words are used in context and utilise data from learner corpora to draw attention to recurrent errors.*”

¹⁷ “[...] *the more frequent a word is in language use, the more likely it is to be useful to the learner.*”

¹⁸ “*Corpora can provide more language and can disclose solutions to language queries that have not been dealt with in edited language resources, propelling language users to completely new levels of learner autonomy.*”

pesquisa com base em *corpus* e, por isso, sugere a utilização de *corpora* e recursos que já estejam disponíveis e sejam, assim, mais facilmente acessados.

No que diz respeito aos percalços do efetivo uso de *corpus* por professores e estudantes, Romer (2010, p. 18, tradução minha) apresenta uma ponderação pertinente de que “poucos professores e estudantes sabem da utilidade dos recursos e põem a mão nos computadores e nos *corpora*¹⁹.” Friginal (2018, p. 7) pontua que, dentre os principais motivos para esse receio em manusear as ferramentas oferecidas pela LC, estão a falta de confiança, a escassez de tempo por parte dos professores e a dificuldade deles em orientarem seus alunos. Sem mencionar aqueles docentes que ainda não se familiarizaram com a abordagem que a LC proporciona.

Sendo assim, Romer (2010, p. 21) enfatiza a valia de pesquisadores e linguistas de *corpus* desenvolverem projetos que auxiliem os professores e os alunos a se apropriarem das ferramentas disponibilizadas por essa vertente dos estudos linguísticos.

Para finalizar esta parte, julga-se válida a síntese de que a LC e a área do ensino de idiomas, apesar de também já contarem com estudos anteriores aos computadores, foram fortemente impulsionadas devido à tecnologia. Destacou-se, nesse âmbito, o quanto o projeto COBUILD em meados dos anos 80 teve um papel preponderante para esse avanço, tal como as pesquisas direcionadas aos *corpora* de aprendizes. Por fim, esta seção discutiu as aplicações pedagógicas da LC ao ensino de línguas, pontuando as formas indiretas e diretas segundo as quais elas acontecem.

2.5 A Aprendizagem Movida por Dados (*Data-Driven Learning*)

Viu-se na seção anterior que as décadas de 80 e 90 se mostraram de grande valia para a intersecção da LC com o ensino de idiomas pelo fato de o computador estar presente em pesquisas como o projeto COBUILD e por haver estudos em *corpora* de aprendizes. Entretanto, Berber Sardinha (2011), ao estabelecer um panorama sobre o uso da LC no contexto do ensino, indica que, dentre as três formas diferentes de se elaborarem atividades didáticas pautadas na LC, a primeira delas se originou, mais precisamente, nos anos 90. Ela ficou conhecida como *classroom*

¹⁹ “[...] a few teachers and learners know about the availability of useful resources and get their hands on corpus computers or concordancers themselves.”

concordancing, pois era evidente o uso das concordâncias em sala de aula nesse período. Berber Sardinha (2011, p. 302) define essas concordâncias como “listagens de palavras de um *corpus*, ladeadas por porções dos textos em que ocorrem, feitas por computador.” Já Tagnin (2011, p. 358) as define como a “relação de todas as ocorrências de uma palavra de busca em um *corpus* junto com o seu contexto. Em geral, apresenta a palavra de busca em posição central, cuja formatação é denominada de KWIC.”

Evidentemente, essa forma de utilização de *corpora* em sala de aula surgida nos anos 90 da qual se refere Berber Sardinha (2011) diz respeito à Aprendizagem Movida por Dados, comumente conhecida em inglês como *Data-Driven Learning* (DDL). Porém, antes de defini-la apropriadamente, ressalta-se, ainda pelo viés histórico, que foi o pesquisador Tim Johns quem cunhou essa proposta, ao introduzir, com o auxílio do *software Microconcord*, as concordâncias em seu trabalho docente voltado ao ensino de inglês para fins específicos (Chambers, 2010). A autora prossegue descrevendo ainda que os alunos com os quais Tim Johns aplicou essa abordagem pontuaram que “trabalhar com as linhas de concordância foi uma forma mais eficaz de estudar o uso de preposições²⁰” (Chambers, 2010, p. 346, tradução minha).

Como descreve Johns (1991, p. 1), essa abordagem evidencia o uso do computador, mais precisamente, do concordanciador – “programa que extrai todas as ocorrências de uma palavra de busca em um *corpus* juntamente com o seu contexto [...]” (Tagnin, 2011, p. 358) – como uma forma de prover informações sobre a língua. Apesar de esse informante ser, *a priori*, passivo, o processo se torna ativo a partir do momento em que o aprendiz lança um questionamento a ser investigado com o uso da ferramenta. Ainda no sentido de uma definição para essa abordagem, Johns (1991, p 2, tradução minha) alega que “a Aprendizagem Movida por Dados fornece ao aluno o acesso a fatos relacionados ao uso da língua²¹.”

Johns (1991, p. 2) prossegue acrescentando o fato do concordanciador não ser uma novidade, haja vista que já era empregado por pesquisadores com o intuito de descobrir padrões da língua. Isso sem mencionar que também já havia indícios do

²⁰ “His learners considered working with a concordance printout to be a much more effective way of studying the use of common prepositions.”

²¹ “[...] and that is a data-driven approach which gives the learner access to the facts of linguistic ‘performance’.”

uso dessa ferramenta para o ensino em 1969 em Birmingham (Chambers, 2010). Todavia, o que se mostrou inovador à época foi a descentralização do acesso à ferramenta (feito apenas por pesquisadores até então), tornando-o disponível aos alunos e os colocando em uma posição em que um perfil de pesquisador se mostre necessário. Nos dizeres do próprio autor, “o aprendiz é também um pesquisador e a aprendizagem precisa ser movida pelo acesso aos dados linguísticos, portanto o termo Aprendizagem Movida por Dados para descrever essa abordagem²²” (Johns, 1991, p. 2, tradução minha).

Nesse sentido, o fato de o aluno atuar como um pesquisador que acessa o *corpus* para responder suas possíveis dúvidas sobre o funcionamento da língua é que se insere a metáfora do Sherlock Holmes (Johns, 1997, p. 101 *apud* Crosthwaite, 2020, p. 169), que sugere o papel investigativo do aprendiz, como um detetive da língua. Desse modo, enfatizam-se as ressalvas de Chambers (2010), que apontam que a Aprendizagem Movida por Dados, além de proporcionar um acesso mais amplo aos dados pautados em *corpora*, estabelece uma abordagem que implica em uma participação mais central e ativa dos alunos no processo da aprendizagem, caracterizado por ela como indutivo.

Desse modo, Johns (1994, p. 297) prega ainda que a DDL busca afastar a necessidade de um intermediador e aproximar o aprendiz ao dado linguístico por um acesso direto. Como resultado dessa aproximação, Johns (1991, p. 2-3) apresenta que a Aprendizagem Movida por Dados se postula como um processo de aprendizagem de línguas em que o aluno faz especulações sobre a língua, que o ajuda a ver os padrões linguísticos e até mesmo a fazer generalizações sobre o funcionamento da língua. Além dessas vantagens, evidenciam-se ainda processos como observação, análise, reflexão, racionalização e exploração da língua (O’Sullivan, 2007, p. 277 *apud* Boulton, 2010, p. 535).

Certamente, esse processo todo envolve uma mudança no papel do aluno, que se torna mais ativo e autônomo. Contudo, há de se considerar que o professor também passa a assumir uma posição diferente. Segundo Johns (1991, p. 3), o professor “tem que se tornar um diretor e um coordenador de pesquisas iniciadas pelos alunos²³.”

²² “[...] *language-learner is also, essentially, a research worker whose learning needs to be driven by access to linguistic data – hence the term “data-driven learning (DDL)” does describe the approach.*”

²³ “*The second main effect of DDL is on the role of the teacher, who has to learn to become a director and a coordinator of student-initiated research.*”

Johns (1994, p. 297) descreve que o professor tem que deixar a posição de detentor do conhecimento para assumir essa posição de alguém que organiza a pesquisa. Essa é uma ressalva pertinente, pois é comum que os estudantes, ao pesquisarem uma dúvida sobre a língua nos *corpora*, acabem deparando-se com linhas de concordâncias que contenham palavras, expressões ou até mesmo estruturas linguísticas as quais eles, possivelmente, desconhecem. Por isso, é importante que o professor seja um mediador de todo esse processo.

Consideradas as colocações trazidas até então, convém mencionar a definição de Aprendizagem Movida por Dados feita por Crosthwaite (2020, p. 169, tradução minha), que aponta que

a DDL envolve o engajamento direto do aprendiz com os dados do *corpus*, seja na forma de materiais impressos ou na consulta mais direta mesmo, o que possibilita que os alunos visualizem e internalizem informações contextuais e estatísticas sobre a língua em uso²⁴.

Apesar de essa conceitualização compilar, de certa forma, os pontos levantados anteriormente, ela abre escopo para as formas de aplicação dessa abordagem. Em relação a isso, Chambers (2010) sugere que o trabalho com a DDL em sala de aula pode acontecer de forma indireta – quando o professor prepara as linhas de concordância previamente – ou direta – quando o aluno acessa o *corpus*. Essas opções de proporcionar ou não ao aluno o acesso direto ao concordanciador (sem preparação prévia do professor) são também abordadas por Gabrielatos (2005), que denomina a forma indireta como *soft* (branda, suave) e a direta como *hard* (rígida, difícil). Já Boulton (2012) cunha as nomenclaturas ***hands-on*** para o acesso mais direto aos dados linguísticos e ***hands-off*** para quando essa consulta é feita por intermediação prévia do professor.

Ainda a esse respeito, mais precisamente em relação à aplicação *hands-off*, Boulton (2010) defende que a opção de operacionalização de atividades DDL em que o contato do aprendiz com o dado linguístico seja feito por meio de materiais preparados pelo professor e/ou pesquisador. Embora alguns teóricos acreditem que a não utilização dos computadores para acessar os concordanciadores possa ferir, de

²⁴ “DDL typically involves direct learner engagement with corpus data, in the form of printed materials or in the form of learner-led hands-on corpus consultation, allowing students to visualize and internalize statistical and contextual information about language in use [...]”

certo modo, a essência da Aprendizagem Movida por Dados, Boulton (2010, p. 559) aponta que esses materiais preparados pelos professores oferecem também benefícios e se mostram possíveis de serem utilizados por aprendizes com baixo nível linguístico, sem mencionar que podem ser preparados por professores que não detêm vasto conhecimento da LC. Nos dizeres do próprio autor, “eliminar o computador não representa menosprezar a base conceitual da DDL, mas sim tornar a tarefa do aprendiz mais fácil²⁵” (Boulton, 2010, p. 559, tradução minha).

Visto que o contato com o dado linguístico pode ocorrer por meio de materiais preparados por professores e ainda sim representar uma postura de pesquisador nos alunos, como defende Boulton (2010), torna-se necessário apontar algumas formas de acordo com as quais as atividades que exploram a concordância podem ser elaboradas. Há, conforme Berber Sardinha (2011, p. 310), as que são integrais (com todas as ocorrências, as selecionadas (com apenas algumas ocorrências), as editadas (em que as linhas são modificadas), as preenchidas (com exibição da palavra de busca) e as lacunadas (em que a palavra de busca não é disponibilizada).

Observadas as ressalvas acerca das aplicações da Aprendizagem Movida por Dados, segue-se para uma sucinta discussão das limitações e das vantagens de se empregar essa abordagem no contexto educacional. Com relação a isso, Boulton (2010) aborda três pontos que restringem, de certa forma, a adoção da DDL. São eles: (i) a falta de convencimento de que essa aprendizagem vale a pena, (ii) o fato de ela ser comumente associada a contextos de ensino-aprendizagem de cunho mais acadêmico e/ou que envolvem um nível de proficiência linguística mais avançada – ponto esse também levantado por Friginal (2018) – (iii) e pelo fato de a aplicação *hands-on* ser preponderante.

No que tange ao primeiro ponto, Chambers (2010) acrescenta que a DDL pode consumir muito tempo e ser trabalhosa, o que acaba desmotivando alguns professores e/ou pesquisadores e dificultando o entendimento do quão pertinente ela é. Nesse sentido, a própria autora sugere o emprego de linhas de concordância preparadas antecipadamente pelos professores (Chambers, 2010). Outra solução a essa problemática dada pela teórica é a utilização de atividades elaboradas pelos professores (*hands-off*) em um primeiro momento para que, de forma processual e

²⁵ “We have argued that eliminating the computer from the equation, far from fatally undermining the conceptual basis of DDL, can in fact make the learner’s task considerably easier.”

gradual, sejam inseridas atividades que visem ao acesso mais direto ao *corpus* (*hands-on*) (Chambers, 2010).

Compreende-se, inclusive, que essa estratégia proposta por Chambers (2010) também se mostra válida aos outros dois pontos de atenção mencionados por Boulton (2010). Iniciar a implementação da Aprendizagem Movida por Dados com atividades *hands-off* possibilita a aplicabilidade dessa abordagem em contextos cujos níveis de proficiência da língua sejam básicos e/ou intermediários. Além disso, essa sugestão de inserir atividades *hands-on* de forma gradual e processual representa uma forma de desmistificar a percepção de que só se faz jus à DDL atividades didáticas pautadas no acesso direto aos dados do *corpus*.

Essa gradação do acesso menos direto ao *corpus* ao mais direto pode, inclusive, ser interpretada em consonância com a observação feita por Crosthwaite (2017, p. 448) ao retratar que a DDL promove interações condizentes com a teoria sociocultural vygotskiana. Conforme Paiva (2014), essa perspectiva afirma que existe uma zona próxima ao desenvolvimento, ou seja, o que é possível de ser aprendido por intermédio de mediação, metaforicamente relacionada a figura do andaime. Nesse sentido de se construir a aprendizagem aos poucos, a inserção, em doses homeopáticas, do acesso mais direto ao *corpus* pode contribuir, portanto, para o desenvolvimento dos saberes necessários para que o aluno consiga, em determinado ponto, acessar o *corpus* autonomamente e sem grandes dificuldades.

Em defesa ainda dessa utilização de materiais previamente preparados pelos professores, Boulton (2012) destaca que eles selecionam quantidades menores de dados linguísticos, limpando, em certa medida, o excesso de informações que podem ser desnecessárias e/ou disruptivas aos alunos quando está buscando por alguma informação sobre a língua.

Por fim, outra crítica à DDL descrita por Chambers (2010) é a de que alguns pesquisadores afirmam que essa abordagem é descontextualizada. Isso porque as linhas de concordância lidam com fragmentos curtos de textos. Entretanto, a própria autora rebate essa percepção, ressaltando que os concordanciadores, muitas vezes, disponibilizam o acesso ao texto todo. Chambers (2010) vai além e pontua que a Aprendizagem Movida por Dados disponibiliza informações sobre o uso da língua feito por um número substancial de falantes nativos. Boulton (2010) mostra-se a favor dessa colocação ao alegar que os usos da língua disponíveis para investigação no

corpus estão evidenciados em seus contextos de produção e ainda organizados de maneira que o aluno possa facilmente acessar várias linhas de concordância perceber, portanto, as múltiplas possibilidades de uso da língua.

Em conclusão, esta seção visou abordar a Aprendizagem Movida por Dados sob um olhar reflexivo que trouxe, em primeira instância, o panorama histórico para, em seguida, discorrer sobre as definições mais gerais dessa abordagem. Isso sem deixar de mencionar as formas segundo as quais a DDL pode ser aplicada e algumas das limitações e dos benefícios apontados por certos pesquisadores a seu respeito.

3 DA TEORIA À PRÁTICA: EXEMPLIFICAÇÃO DE PROPOSTAS DE ATIVIDADES DIDÁTICAS

Concluída a seção referente à fundamentação teórica, segue-se para um capítulo que visa a apontar possíveis desdobramentos das teorias na prática. Para isso, recorre-se à demonstração de duas lições, publicadas no livro *Caderno de Atividades de Aprendizagem Movida por Dados* (Pinto; Garcia; Serpa, 2022). O intuito, pois, desta apresentação é o de exemplificar como os pressupostos pedagógicos e os documentos norteadores da Educação Básica, assim como as conceitualizações acerca da Linguística de *Corpus* e da Aprendizagem Movida por Dados, se convergem de modo que podem ser desdobrados em atividades didáticas reais e aplicáveis à sala de aula.

Reitera-se ainda que, além de servir como forma de atrelar a teoria à prática nesta tese, as atividades didáticas contidas neste capítulo foram utilizadas também pelos professores que participaram do curso de formação continuada em serviço (vide capítulo 4) como forma de terem exemplos práticos de como a Aprendizagem Movida por Dados pode oferecer atividades didáticas a serem levadas às salas de aulas de diferentes contextos de ensino.

Desse modo, as atividades trazidas neste capítulo, que possibilitam que os estudantes investiguem como a língua funciona por meio da observação das ocorrências dela, estão categorizadas conforme dispõem o acesso aos dados linguísticos, ou melhor, aos *corpora*. Dessa forma, elas podem ser do tipo **hands-on**, em que os estudantes acessam diretamente as ferramentas e/ou softwares computacionais para consultarem as informações da língua. No entanto, também podem ser do tipo **hands-off**, em que esse procedimento de consulta é previamente delimitado pelo professor que elabora a atividade, evidenciando, então, um aspecto indireto de acesso às informações da língua. Independentemente, porém, de como essas atividades se dispõem, ambos os tipos viabilizam uma postura investigativa (de detetive) por parte dos estudantes para aprenderem, indutivamente, sobre a língua.

3.1 *So many things in a day*: proposta de atividade didática elaborada para o Ensino Médio

A primeira proposta apresentada neste capítulo foi publicada no livro *Caderno de Atividades de Aprendizagem Movida por Dados* (Pinto; Garcia; Serpa, 2022). Ela foi desenvolvida pelo pesquisador para o contexto do Ensino Médio e explora, em especial, a função *WordSketch*, da ferramenta Skell (vide capítulo 4), para a pesquisa por expressões coloquiais em língua inglesa com os verbos *be*, *have*, *get*, *put* e *take*. Essa proposta de atividade didática, elaborada em forma de uma lição completa, foi pensada em dois modelos: um *hands-on* e outro *hands-off*.

So many things in a day se trata de uma lição planejada para a primeira, segunda e/ou terceira séries do Ensino Médio, a ser aplicada em duas aulas, com duração entre 45-50 minutos cada. O objetivo principal é usar expressões coloquiais com os verbos *be*, *have*, *get*, *put* e *take* no contexto de atividades diárias, ou melhor, da rotina. Já os específicos envolvem (i) discutir atividades comuns à rotina; (ii) utilizar o Skell para descobrir possíveis combinações com os referidos verbos e (iii) redigir frases que descrevam a rotina diária.

Essa lição, projetada para ser aplicada com os estudantes do Ensino Médio, subdivide-se em estágios. O primeiro deles, de introdução, propõe que os alunos usem a língua inglesa para conversarem entre si sobre suas atividades diárias e suas rotinas, conforme ilustração a seguir:

Figura 4. Atividade 1 da lição *So many things in a day*

Atividade 1:



Em pequenos grupos, converse a respeito da rotina de cada um. O que há em comum entre vocês? O que há de diferente?

SEMELHANÇAS	DIFERENÇAS

Fonte: Pinto, Garcia e Serpa (2022)

A partir disso, visa-se ao levantamento dos conhecimentos prévios dos estudantes em relação ao conteúdo proposto na lição. Para isso, a atividade 2 sugere que os alunos pensem a respeito de quais combinações com os verbos *be*, *get*, *have*, *put* e *take* eles já conhecem.

Figura 5. Atividade 2 da lição *So many things in a day*

Atividade 2:

Quais atividades diárias você consegue formar com os verbos *be*, *get*, *have*, *put* e *take*? Consegue pensar em possíveis combinações com esses verbos?



BE

GET

HAVE

PUT

TAKE

Fonte: Pinto, Garcia e Serpa (2022)

O estágio seguinte, de investigação em *corpus*, está relacionado ao momento em que os estudantes passam a conhecer a ferramenta Skell e aprendem a utilizá-la. Isso no caso da versão *hands-on* da lição.

Figura 6. Atividade 3 da lição *So many things in a day*

Atividade 3:

Utilize a ferramenta SKELL para encontrar combinações possíveis na língua inglesa envolvendo os verbos *be*, *get*, *have*, *put* e *take*.

BE	GET	HAVE	PUT	TAKE



Fonte: Pinto, Garcia e Serpa (2022)

Nota-se que a atividade 3 demanda dos estudantes que eles acessem a ferramenta Skell para encontrarem ao menos 4 combinações com cada um dos verbos.

Já o modelo *hands-off* traz um recorte prévio (feito pelo próprio pesquisador) das informações disponibilizadas pela ferramenta Skell. O intuito, no entanto, permanece o de os estudantes observarem os exemplos reais de uso da língua para levantarem possíveis combinações com os verbos abordados na lição. Veja a seguir:

Figura 7. Atividade 3 da lição *So many things in a day*, modelo *hands-off*

Atividade 3:

Observe algumas frases retiradas de uma ferramenta que apresenta exemplos de usos reais da língua.

We **have** extensive **experience** in application migration.

This **puts** extra **money** in your pocket.

So he **gets** **home** by 3 am.

These boots are easy to **put on** and **off**.

It means the volleyball season **is over**.

By early 1894 the war **was over**.

He married twice and **has** three **children**.

People **are able to** access information online.

A few customers **were able to** resolve problems.

It can **take** four **photos** per second while recording video.

More skills needed to **get** a **job**.

Who is **taking care of** your forest?

I **put** so much **effort into** it.

The group **gets** some **money** in corporate contracts.

I started to **take pictures** every few minutes.

Many students **have problems** recalling simple facts.

Identifique nas frases anteriores possíveis combinações na língua inglesa envolvendo os verbos be, get, have, put e take.

BE	GET	HAVE	PUT	TAKE

Na sequência, o terceiro estágio, de aplicação, envolve a aplicação das combinações encontradas pelos estudantes, seja por meio do acesso direto à ferramenta Skell ou por meio da consulta ao recorte prévio desses dados feito pelo pesquisador. Para isso, a atividade 4 propõe que os alunos redijam frases sobre as suas rotinas.

Figura 8. Atividade 4 da lição *So many things in a day*

Atividade 4:

Descreva a sua rotina. Forme frases empregando as combinações levantadas com auxílio da ferramenta SKELL.



Fonte: Pinto, Garcia e Serpa (2022)

Finalmente, o último estágio, de conclusão, diz respeito à discussão, por parte dos estudantes envolvidos na possível aplicação dessa lição e do professor que ministrou essa aula. Isso com o objetivo de se obter o *feedback* dos alunos em relação à ferramenta Skell.

Em conclusão, a lição *So many things in a day*, descrita nesta subseção, visa a proporcionar aos estudantes a utilização de combinações coloquiais da língua com certos verbos do inglês. Isso é organizado, na lição, de modo que busca a atender os pressupostos pedagógicos contidos na BNCC e no Currículo Paulista, assim como as conceitualizações acerca da LC e o ensino, e da Aprendizagem Movida por Dados.

De início, na atividade 1 da lição, notou-se o enfoque do material na contextualização em relação ao tema e à vida dos estudantes ao direcionar a discussão sobre a rotina diária. Esse é um ponto importante em relação às metodologias ativas mencionadas na BNCC e no Currículo Paulista, pois favorece o engajamento dos estudantes, contribuindo com o processo de aprendizagem. Além disso, a contextualização com a vida e com a temática da aula nesse estágio inicial da lição é uma forma de refutar o que Chambers (2010) lista como uma das críticas à Aprendizagem Movida por Dados, que é a descontextualização das linhas de concordância. Isso porque, ao se estabelecer o contexto já de pronto, durante o

momento da busca e da consulta no *corpus*, já haverá um contexto pré-estabelecido do qual o estudante se atentará no momento da pesquisa.

Na atividade 2, há outro fator relevante de se destacar, que é o do levantamento do conhecimento prévio. Isso também se mostra condizente com a aprendizagem ativa e, logo, com os fundamentos pedagógicos dos documentos norteadores da Educação Básica. Além disso, esse estágio já se configura também como uma preparação para a pesquisa no *corpus* ou nas linhas de concordâncias previamente selecionadas pelo professor (modelo *hands-off*). Isso porque, nessa etapa em que se direciona os estudantes a pensarem em combinações possíveis com os verbos abordados na lição, já se promove uma instigação de encontrar combinações possíveis. Esse estabelecimento de uma necessidade de se buscar por outras formas combinatórias com os verbos consegue relacionar a atividade 2 com o exercício seguinte.

Dessa maneira, a atividade 3, que pode ser desenvolvida de forma *hands-off*, trazendo, conforme Berber Sardinha (2011), linhas de concordância selecionadas, ou então de maneira *hands-on*, promovendo o acesso a linhas de concordância integrais (Berber Sardinha, 2011), configura-se como a instância em que os estudantes fazem, com a mediação do professor, as especulações com base no que já haviam pensado, analisam os padrões da língua e estabelecem generalizações. Compreende-se, pois, que é nessa atividade em que a Aprendizagem Movidada por Dados acontece efetivamente. Outro aspecto de grande relevância desse momento da lição é que há um propósito de busca estabelecido, que é o de encontrar possíveis combinações com os verbos, de modo a preencher uma tabela com os resultados da pesquisa. Por mais que isso pareça algo simples, essa mera tabela auxilia no estabelecimento do propósito de se consultar o *corpus*, o que é visto também como uma maneira de contextualizar o procedimento da busca.

Por fim, a atividade 4 propõe que os estudantes redijam frases com base nas descobertas feitas no exercício anterior. Esse também é um estágio relevante, pois objetiva contextualizar as informações encontradas no *corpus* com a vida dos estudantes. Logo, isso rebate as críticas da falta de contextualização normalmente feita à Aprendizagem Movidada por Dados, como aponta Chambers (2010), e está também em consonância com a aprendizagem ativa, que enfatiza a relevância de o estudante ser o sujeito central, o protagonista.

3.2 *It depends on the situation*: proposta de atividade didática elaborada para o contexto de curso livre de inglês

A segunda proposta apresentada neste capítulo também foi publicada no livro *Caderno de Atividades de Aprendizagem Movida por Dados* (Pinto; Garcia; Serpa, 2022). Porém, diferentemente da proposta anterior, esta foi desenvolvida pelo autor desta tese para o contexto de curso livre de inglês, ou seja, aquele que existe em escolas de idiomas. Com relação à ferramenta, ela aborda a utilização do Linggle (vide capítulo 4), para a pesquisa por algumas *dependent prepositions* (sintagmas preposicionais) da língua inglesa. Essa proposta de atividade didática, elaborada em forma de uma lição completa, também foi concebida em dois modelos: um *hands-on* e outro *hands-off*.

It depends on the situation se trata de uma lição planejada para aprendizes de nível intermediário e/ou avançado, a ser aplicada em duas aulas, com duração de aproximadamente 45 minutos cada. O objetivo principal é usar as *dependent prepositions* que acompanham determinados verbos, substantivos e adjetivos. Já os específicos envolvem (i) discutir e dividir opiniões sobre frases de temas variados; (ii) utilizar o Linggle para descobrir as *dependent prepositions* que acompanham certos verbos, substantivos e adjetivos e (iii) empregá-las em frases de temas diversos.

Assim como a lição vista anteriormente, esta também se subdivide em estágios. O primeiro deles, de introdução, propõe uma contextualização do conteúdo principal (as *dependent prepositions*). Isso é feito por meio da atividade 1, cujo foco está na produção oral, uma vez que é sugerido que os aprendizes conversem e discutam, entre si, se concordam ou não com temas presentes em determinadas frases.

Figura 9. Atividade 1 da lição *It depends on the situation*

Atividade 1:

Observe as frases a seguir. Você concorda com elas? Converse com um colega e descubra se ele(a) também concorda ou não com cada frase.

	ME		STUDENT A	
I've always tried to provide <u>my friends</u> with positive thoughts.	Yes	No	Yes	No
Non-native speakers have difficulty in <u>understanding</u> some linking sounds	Yes	No	Yes	No
I'm never anxious about my personal plans or professional goals.	Yes	No	Yes	No

Fonte: Pinto, Garcia e Serpa (2022)

A partir disso, visa-se, ainda como parte da introdução, à etapa em que o foco recai sobre o fato de os aprendizes despertarem o interesse e a atenção em relação às preposições. Essa atividade, ainda que inicial, auxiliará os estudantes a lidarem com as informações às quais eles terão acesso no exercício subsequente, instância em que eles, de fato, acessarão a ferramenta Linggle para investigar a língua. Veja a atividade 2 a seguir:

Figura 10. Atividade 2 da lição *It depends on the situation*

Atividade 2:

Considere as frases do exercício anterior para responder as questões que se seguem.

a) A primeira frase traz um exemplo de uma preposição precedida de:

☐ um adjetivo; ☐ um substantivo; ☐ um verbo.

b) O que fica entre **provide** e a preposição **with**?

c) A segunda frase traz um exemplo de uma preposição precedida de:

☐ um adjetivo; ☐ um substantivo; ☐ um verbo.

d) Como é usado o verbo após a combinação **difficulty in**?

e) A terceira frase traz um exemplo de uma preposição precedida de:

☐ um adjetivo; ☐ um substantivo; ☐ um verbo.



Fonte: Pinto, Garcia e Serpa (2022)

O estágio seguinte, de investigação em *corpus*, está relacionado ao momento em que os estudantes passam a conhecer a ferramenta Linggle e aprendem a utilizá-la, ou seja, a aplicar os procedimentos de busca descritos no capítulo *Materiais e métodos*, mais especificamente, na seção das ferramentas.

Figura 11. Atividade 3 da lição *It depends on the situation*, modelo *hands-on*

Atividade 3:

Selecione 3 itens de cada uma das categorias (verbo, substantivo e adjetivo) da tabela abaixo. Utilize o *LINGGLE* para descobrir quais preposições acompanham cada uma das palavras escolhidas.

VERBS	NOUNS	ADJECTIVES
Blame / Blame (sb/sth)	Attention	Available
Focus	Belief	Delighted
Object	Capacity	Different
Persist	Information	Good / Bad
Suffer	Problem	Prejudiced
Warn (sb)	Reputation	Suited

Use a tabela a seguir para anotar e esquematizar algumas das suas descobertas.

Verb	Preposition	Observation / Example

Noun	Preposition	Observation / Example

Adjective	Preposition	Observation / Example

Fonte: Pinto, Garcia e Serpa (2022)

Como se pode notar, a atividade 3, no modelo *hands-on*, direciona os estudantes a escolherem três verbos, três substantivos e três adjetivos para

pesquisarem, de forma direta no Linggle, quais preposições são empregadas juntamente a cada um deles.

Entretanto, no modelo *hands-off*, é trazido um recorte prévio (feito pelo próprio pesquisador) das informações disponibilizadas pela ferramenta Linggle. O intuito, no entanto, permanece o de os estudantes observarem os exemplos reais de uso da língua para levantarem as possíveis combinações preposicionais feitas com verbos, substantivos e adjetivos. Veja a seguir:

Figura 12. Atividade 3 da lição *It depends on the situation*, modelo *hands-off*

Atividade 3:

Observe os exemplos extraídos do *LINGGLE* e identifique as combinações verbo-preposição, substantivo-preposição e adjetivo-preposição.

- * Santayana argued that these disciplines , the religious narratives in which they were presented in exemplary ways, @@ narratives , were not inevitably to blame for superstition
- * It is this phenomenon of self - suppression that I want to focus on today.
- * No one could object to the richness of choice if it served a clear and coherent educational purpose , but , if it does , who can say what that is ?
- * The problem is that it is not clear that such a pattern of rights could persist in anything but name only .
- * Furthermore , most other books we wxamined suffer from more or less of the same problem.
- * Bit I want to warn you against letting this longed - for child become the embodiment of all you hope.
- * The students I knew did n't pay much attention to Gide 's homosexuality ; that was just a personal quirk that he was entitled to indulge in , if he felt so inclined.
- * Nevertheless , it gradually does grow dear that logocentrism designates a supposedly mistaken belief in the stability of language and phenomena.
- * They took place only whaen the ego 's usual capacity for control had been overcome either through the use of a psychoactive substance or therapeutic @@.
- * I sought out the information on every munitions factory and storage center and began to dismantle them.
- * The problem with the concept of the quality of a performance , with respect to reward in a world of limited good (i.e.
- * The daughter , who had a reputation for dishonesty , added details each time she gave her story .
- * If only inevitably decreases what is aviable for original appropriation by others , which is not the same thing at all .
- * Some of the most successful career track women label themselves failures ; some of the least professionally successful are delighted with their lives and believe they have realized their
- * If we look at the pictures in " The Urban Villagers , " we find units not very different from Little Italy in New York or West Village tenements .

Use a tabela a seguir para anotar e esquematizar algumas das suas descobertas.

Verb	Preposition	Observation / Example

Noun	Preposition	Observation / Example

Adjective	Preposition	Observation / Example



Fonte: Pinto, Garcia e Serpa (2022)

Na sequência, o terceiro estágio, o de aplicação, envolve a aplicação das *dependent prepositions* encontradas pelos estudantes, seja por meio do acesso direto à ferramenta Linggle ou por meio da consulta ao recorte prévio desses dados feito pelo pesquisador. Para isso, a atividade 4 propõe que os alunos redijam frases contendo informações verdadeiras de suas vidas e que empreguem as preposições e as combinações descobertas durante a aula. Em seguida, a atividade sugere que os aprendizes conversem entre si e discutam se as frases dos colegas são verdadeiras para eles também ou não.

Figura 13. Atividade 4 da lição *It depends on the situation*

Atividade 4:	STUDENT A	
	Yes	No
Use as suas descobertas para formar frases que sejam significativas e verdadeiras para você. Em seguida, pergunte a um colega de sala se ele(a) concorda com a sua afirmação.		

Fonte: Pinto, Garcia e Serpa (2022)

Finalmente, o último estágio, o de conclusão, diz respeito à discussão, por parte dos estudantes envolvidos, na possível aplicação dessa lição e do professor que ministrou essa aula. Isso com o objetivo de se obter o *feedback* dos alunos em relação à ferramenta Linggle.

Em suma, a lição *It depends on the situation*, descrita nesta subseção, visa a proporcionar aos estudantes a utilização de algumas *dependent prepositions*. Observa-se que, nessa lição, assim como naquela descrita anteriormente, o contexto de ensino não é acadêmico (Ensino Superior), como rotulam algumas das críticas à Aprendizagem Movida por Dados, que alegam que essa abordagem é mais comumente disseminada nesse contexto. Pelo contrário, ambas as lições foram desenvolvidas para cenários educacionais diferentes, sendo a primeira voltada ao Ensino Médio e a segunda às escolas e/ou centros de idiomas.

Ainda em comparação à lição anterior, esta também apresenta um estágio introdutório em que se contextualiza o tema e o atrela à vida do estudante. Em seguida, há o levantamento do conhecimento prévio. Essa parte inicial é

especialmente importante, pois já prepara, de certa maneira, o aluno para as noções de funcionamento da língua no que se refere às preposições, levando, assim, o estudante a compreender os padrões da língua com os quais ele se deparará no estágio subsequente, o da busca no *corpus*. Toda essa preparação para o ato da pesquisa em si se mostra em conformidade com os pressupostos pedagógicos da BNCC e do Currículo Paulista no sentido de a aprendizagem ser ativa, sendo o estudante o sujeito central desse processo.

Na atividade 3, o aluno consulta as linhas de concordância, sejam elas as selecionadas ou as integrais (Berber Sardinha, 2011), ou seja, pelo modelo *hands-off* ou *hands-on*. Destaca-se que há um propósito para a realização da busca, que é o de encontrar padrões de uso das preposições de modo a preencher a tabela com verbos, substantivos e adjetivos. Nota-se, assim, que esse objetivo estabelecido para a consulta no *corpus* é uma forma de tornar as especulações e generalizações a serem feitas pelos estudantes a partir dos resultados alcançados mais contextualizadas, afinal há uma razão para a realização da pesquisa.

Por fim, a atividade 4 proporciona ao estudante empregar os padrões encontrados por meio da busca na ferramenta de maneira personalizada. Isso também contribui para a contextualização de toda a lição e do processo da aprendizagem, que é ativa e que atende às competências relacionadas ao conhecimento, ao pensamento científico, crítico e criativo, à comunicação e à cultura digital, presentes na Base Nacional Comum Curricular.

3.3 Da organização da aula

Como se descreveu nas duas últimas subseções, ambas as lições apresentaram uma sequenciação das atividades que favoreceram tanto os pressupostos pedagógicos das metodologias ativas e, por conseguinte, da BNCC e do Currículo Paulista, quanto as conceitualizações acerca da LC e o ensino e da Aprendizagem Movida por Dados. Dessa maneira, segue-se, então, para uma análise desses estágios e uma revisão das fases da aula propostas por Rozenfeld e Viana (2020), com o objetivo de traçar pontos congruentes.

Em ambas as lições, foi possível identificar uma primeira atividade, cujo foco recai sobre o estabelecimento do tema e a contextualização à vida dos estudantes,

promovendo também interação entre eles. Em seguida, a atividade busca levantar o conhecimento prévio dos estudantes e prepará-los com relação aos conteúdos linguísticos que eles possivelmente encontrarão durante a consulta às linhas de concordância. Sequencialmente, estabelece-se um propósito claro para a realização das pesquisas por meio de atividades que visam ao preenchimento de tabelas com as informações levantadas pelas linhas de concordância. Feito isso, os estudantes realizam as buscas nas ferramentas da LC para responderem a essa questão contida nas lições. Por fim, após resolução e devolutivas dos professores, os estudantes aplicam os padrões linguísticos identificados no *corpus* de modo personalizado, ou melhor, contextualizado à realidade deles.

Visto isso, observa-se uma organização sequencial que estrutura ambas as lições. Nesse sentido, é válido ressaltar as considerações de Rozenfeld e Viana (2020), que estabelecem as fases que perpassam as aulas de línguas. Segundo os autores, elas são cinco, como demonstra a figura a seguir:

Figura 14. Representação da aula de línguas estrangeiras



Fonte: Rozenfeld e Viana (2020)

Como ilustra a imagem, há uma fase prévia à aula propriamente dita, cujo objetivo é o de “selecionar o conteúdo e o material a serem apresentados aos alunos” (Rozenfeld; Viana, 2020). Esse é, portanto, o momento do planejamento em que o professor analisa o que será utilizado e reflete, inclusive, sobre as dificuldades da aula.

Já a fase 2, como apresentam Rozenfeld e Viana (2020), subdivide-se em três momentos, sendo eles: (a) o da discussão das questões organizacionais, que lida com recados e assuntos institucionais, por exemplo; (b) o do aquecimento, que envolve a retomada de conteúdos, e (c) o da preparação do insumo, em que busca dar apoio aos estudantes com relação ao objetivo principal da aula.

Nesse sentido, entende-se que a fase 3 é, justamente, “o momento de o(a) professor(a) apresentar o material selecionada na fase 1” (Rozenfeld; Viana, 2020). Assim sendo, esse é o momento cujo objetivo central da aula é desenvolvido. Finalmente, a fase 4 é quando o aluno “faz o uso da língua-alvo em situações significativas de comunicação” (Rozenfeld; Viana, 2020). Por esse motivo, ela é descrita como a fase prática, do uso livre, do fechamento da aula.

Visto isso, visualiza-se uma possível relação da sequenciação das lições com as fases da aula postuladas por Rozenfeld e Viana (2020). Assim, a elaboração e o planejamento da lição em relação a qual conteúdo abordar e a qual ferramenta da LC recorrer se relaciona à fase 1. Já o momento de estabelecimento do tema e levantamento dos conhecimentos prévios está em consonância com a fase 2-b (aquecimento); enquanto a preparação dos estudantes em relação aos conteúdos linguísticos que eles possivelmente precisarão acionar durante a consulta no *corpus* e as formas de busca nas ferramentas da LC se atrelam à fase 2-c (preparação do insumo). Seguidamente, a pesquisa e a resolução da atividade que direciona os estudantes aos conteúdos linguísticos que eles precisam analisar nas linhas de concordância configuram a fase 3 e o uso personalizado e contextualizado dos padrões linguísticos identificados por intermédio dos concordanciadores corresponde à fase 4.

Em conclusão, esta subseção correlacionou as fases da aula (Rozenfeld e Viana, 2020) à organização sequencial das lições que se utilizam da Aprendizagem Moviada por Dados. Isso com o intuito de estabelecer, ainda que previamente, critérios que perpassam a implementação de materiais didáticos ou o desenvolvimento de lições e/ou atividades didáticas amparadas pela abordagem da DDL.

4 MATERIAIS E MÉTODOS

Este capítulo discorre sobre a natureza desta pesquisa de doutoramento, de modo a contemplar os materiais e/ou ferramentas por ela utilizados, além da metodologia envolvida em seu desenvolvimento.

De forma clara e objetiva, esta tese, desenvolvida junto ao Programa de Pós-Graduação em Estudos Linguísticos da Unesp de São José do Rio Preto, situa-se na área de concentração da Linguística Aplicada, mais precisamente, na linha de pesquisa da Pedagogia do Léxico e da Tradução Baseada em *Corpora*.

Ela classifica-se, conforme Paiva (2019), em uma pesquisa de natureza aplicada, do gênero prático, haja vista que visa à implementação de novas atividades didáticas de língua inglesa voltadas à etapa do Ensino Médio. Com relação à abordagem, este estudo se trata de uma pesquisa mista, pois “se utiliza de métodos qualitativos e quantitativos para a coleta de dados” (Paiva, 2019, p. 13).

Delimitados esses aspectos mais gerais, prossegue-se para os métodos e a descrição do curso elaborado para professores de língua inglesa com o intuito de apoiá-los na utilização e na aplicação da Aprendizagem Movida por Dados em sala de aula, para, em seguida, abordar as ferramentas recorridas para a elaboração das atividades. Enfatiza-se que esse curso se classifica como uma ação de formação continuada em serviço, pois considerou os docentes e suas experiências e seus conhecimentos técnicos em todo o processo, um dos postos-chaves para esse tipo de formação, conforme Salles (2004, p. 5), que pontua que “a formação continuada em serviço diz respeito à valorização da prática docente como a única possibilidade de êxito do professor aplicar criativamente a racionalidade técnica obtida no processo de aquisição de competências escolares.”

4.1 Dos métodos e do curso de formação continuada em serviço

De modo geral, este estudo contou, em um primeiro momento, com uma pesquisa de levantamento bibliográfico acerca dos aportes teóricos adotados. Dessarte, pesquisou-se sobre os documentos oficiais relacionados à Educação Básica (a Base Nacional Comum Curricular (BNCC) e o Currículo Paulista – etapa do

Ensino Médio), sobre a LC e sua relação com o ensino de línguas, e sobre a Aprendizagem Movida por Dados.

Posterior à revisão bibliográfica, o pesquisador, junto à orientadora e aos demais membros do grupo de pesquisa *En-Corpora*, trabalhou na elaboração das lições didáticas apresentadas no capítulo anterior e na organização do livro *Caderno de Atividades de Aprendizagem Movida por Dados* (Pinto; Garcia; Serpa, 2022). Como já mencionado, o intuito da elaboração das lições que compõem essa obra foi o de atrelar os pressupostos da aprendizagem ativa presente nos documentos norteadores da Educação Básica à Aprendizagem Movida por Dados, de modo a rebater algumas das críticas feitas a essa abordagem.

Além disso, destaca-se que essas lições também foram utilizadas como exemplificação da aplicação prática da Aprendizagem Movida por Dados em um curso de formação continuada em serviço intitulado “Práticas pedagógicas inovadoras para o ensino de língua inglesa: o aluno como detetive da língua e protagonista da aprendizagem”. Ele foi ministrado pelo pesquisador, em parceria com a orientadora, a Profa. Dra. Paula Tavares Pinto, além de outros professores e discentes do Programa de Pós-Graduação em Estudos Linguísticos (vide apêndice A). Essa ação formativa aos docentes se justifica pelo fato de a DDL ainda não ter alcançado as teorias de ensino mais convencionais (Crosthwaite, 2024) e por precisar envolver mais a relação entre pesquisadores e professores (Boulton, 2024).

Destaca-se ainda que parte desse curso envolveu dois módulos do curso do Prof. Dr. Peter Crosthwaite, da Universidade de Queensland, intitulado *Improving writing through corpora: Data-driven learning?*, que foi traduzido para o português pelo pesquisador (vide Anexo A). Essa parceria com o professor, que atua em uma universidade australiana, foi de grande relevância para a elaboração do curso de formação continuada em serviço proposto nesta pesquisa.

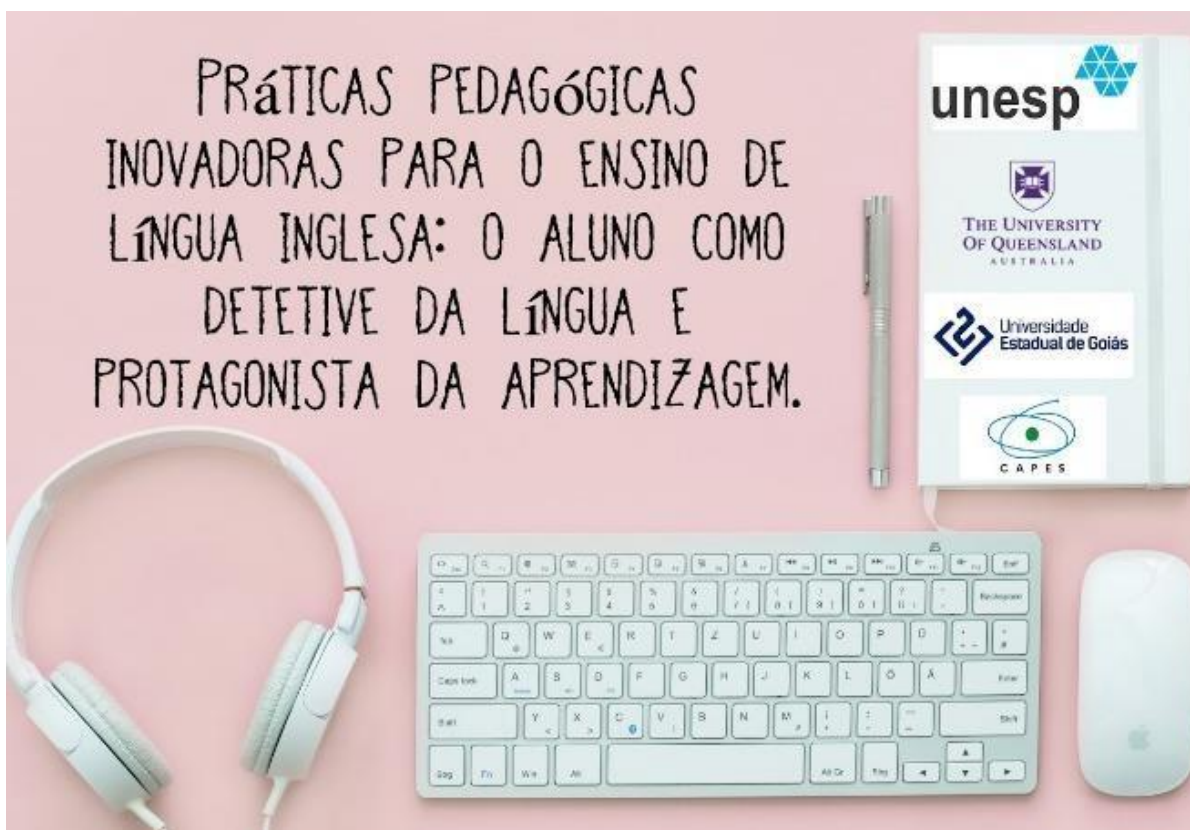
Após a aplicação do curso, que visou a apoiar os professores da Educação Básica na implementação da Aprendizagem Movida por Dados em sala de aula, foram coletadas as propostas de atividades didáticas pautadas nessa abordagem como forma de complementação das habilidades e/ou conteúdos previstos no Currículo Paulista. Essa coleta aconteceu por meio das atividades avaliativas propostas durante todo o curso e da avaliação final, requisitos para a obtenção do certificado de conclusão do curso.

Dessarte, somadas às lições descritas no capítulo 3 e que foram usadas também no curso como uma referência aos professores que participaram da formação continuada em serviço, são apresentadas, no próximo capítulo, as propostas de atividades didáticas que os próprios cursistas desenvolveram com o auxílio do pesquisador. Portanto, compreendidos os procedimentos metodológicos utilizados, passa-se, então, ao aprofundamento da descrição desse curso de formação continuada em serviço oferecido aos professores da Educação Básica.

4.1.1 O curso “Práticas pedagógicas inovadoras para o ensino de língua inglesa: o aluno como detetive da língua e protagonista da aprendizagem.”

O curso intitulado “Práticas pedagógicas inovadoras para o ensino de língua inglesa: o aluno como detetive da língua e protagonista da aprendizagem”, organizado pelo pesquisador e por outra doutoranda do Programa, a discente Carolina Tavares, contou com módulos ministrados por docentes da Unesp de São José do Rio Preto – a Profa. Dra. Paula Tavares Pinto, a Profa. Dra. Adriane Orenha Ottaiano e a Profa. Dra. Talita Serpa –, da Universidade Estadual de Goiás – o Prof. Dr. Eduardo Batista da Silva –, da Universidade de Queensland – o Prof. Dr. Peter Crosthwaite –, além do Prof. Me. Emanuel Henrique Alves e da Profa. Me. Arianne Donizete Delgado Ribeiro Caldas (na época, ambos doutorandos do Programa no qual esta pesquisa se insere).

Figura 15. O curso “Práticas pedagógicas inovadoras para o ensino de língua inglesa: o aluno como detetive da língua e protagonista da aprendizagem



Fonte: elaborada pela coorganizadora do curso

Por se tratar de um curso de formação continuada em serviço, o público-alvo foram os professores de língua inglesa que atuam na Educação Básica, em especial aqueles que lecionam na rede pública estadual de São Paulo. No entanto, houve inscrições de alguns profissionais que atuam na vertente de ensino de idiomas também.

Com carga horária de 30 horas, o curso, oferecido na modalidade on-line, se caracteriza pelo formato de curta duração. Por essa limitação, necessitou-se fazer uma seleção das ferramentas a serem abordadas. Optou-se, então, por expor o

Versatext, o Sketch Engine²⁶, o Skell, o Linggle, o Vocabprofile²⁷, o Platcol²⁸, o Online English Collocations Platform²⁹ e o Tradcorpus³⁰.

Enfatiza-se que todas essas ferramentas foram apresentadas aos cursistas a partir de uma estruturação didática organizada em forma de oficinas síncronas, que proporcionaram aos professores participantes a reflexão e a análise dos possíveis desdobramentos dessas tecnologias no contexto de ensino onde atuam. Salienta-se, entretanto, que duas dessas oficinas foram oferecidas de maneira assíncrona, pois envolveram a realização, por parte dos cursistas, de dois dos módulos do curso *Improving writing through corpora: Data-driven learning?*, ministrado pelo Prof. Dr. Peter Crosthwaite, da Universidade de Queensland, que foi traduzido para o português pelo pesquisador.

Essas oficinas aconteceram por meio de encontros síncronos semanais, realizadas às quartas-feiras, das 19h às 21h. Elas foram ministradas com o auxílio do Google Meet (para as aulas síncronas) e do Google Sala de Aula, onde os materiais e os recursos apresentados no curso eram disponibilizados aos participantes.

A primeira dessas oficinas, ministrada pela Profa. Dra. Paula Tavares Pinto, a doutoranda Carolina Tavares de Carvalho e o autor desta tese, teve o intuito de apresentar e elucidar alguns pontos centrais da LC e da Aprendizagem Movidada por Dados.

²⁶ Disponível em: <https://www.sketchengine.eu/>. Acesso em: 19 de julho de 2022.

²⁷ Disponível em: <https://www.lexutor.ca/vp/eng/>. Acesso em: 19 de julho de 2022.

²⁸ Disponível em: <http://www.institucional.grupogbd.com/dicionario/sobre>. Acesso em: 19 de julho de 2022.

²⁹ Disponível em: <http://www.institucional.grupogbd.com/workbook/index>. Acesso em: 19 de julho de 2022.

³⁰ Disponível em: <https://sites.google.com/unesp.br/tradcorpus/p%C3%A1gina-inicial?authuser=0>. Acesso em: 19 de julho de 2022.

Figura 16. Capa da apresentação de slides utilizada na oficina do dia 04/05/2022



Aprendizagem Movidada por Dados no Ensino de Língua Inglesa *(Data-Driven Learning)*

Profa. Dra. Paula Tavares Pinto
Profa. Me. Carolina Tavares de Carvalho
Prof. Me. William Danilo Garcia

Práticas Pedagógicas Inovadoras para o Ensino de Língua Inglesa

Fonte: elaborado pelos professores responsáveis pela oficina

A partir da imagem, faz-se necessário reiterar o fato de que esse curso de formação continuada em serviço foi desenvolvido a partir de uma parceria entre três universidades, além de contar com professores e pesquisadores que atuam, com afinco, na área da LC.

Outro aspecto também enfatizado nessa primeira oficina foi a relação que a Aprendizagem Movidada por Dados pode estabelecer com algumas das diretrizes tanto da BNCC quanto do Currículo Paulista. Alguns desses pontos são: a autonomia do estudante em pesquisar sobre a língua, a utilização de ferramentas digitais para essa busca, além da aprendizagem indutiva, já que o estudante, a partir das informações encontradas nos *corpora*, levanta hipóteses sobre a língua até se chegar ao entendimento de como ela, de fato, funciona. A imagem a seguir ilustra bem o roteiro dessa oficina:

Figura 17. Roteiro dos pontos abordados na oficina do dia 04/05/2022

Roteiro



**1. BNCC/Novo Currículo Paulista/Currículo de Goiás
+
Aprendizagem Movidada por Dados (DDL)**

**2. Aprendizagem Movidada por Dados (DDL)
Habilidades de Corpus (HACs)**

3. Hands-on Activities!!

Fonte: elaborado pelos professores responsáveis pela oficina

Posterior a essa primeira oficina, de caráter mais introdutório e elucidativo em relação às vantagens da LC e da Aprendizagem Movidada por Dados, as duas semanas subsequentes envolveram a participação dos cursistas em dois módulos do curso do Prof. Dr. Peter Crosthwaite. Esses dois módulos foram os que trataram, mais especificamente, da ferramenta Sketch Engine. Considere os conteúdos deles no quadro exposto a seguir.

Quadro 3. Os módulos 1 e 2 do curso do Prof. Dr. Peter Crosthwaite

Módulo	Título	Conteúdo
01	Substituindo o dicionário ou site de tradução por <i>corpora</i> ?	Comparação entre os resultados de um dicionário ou google tradutor e os resultados de <i>corpus</i> . Isso de forma a focar particularmente na utilidade relativa para expressões com mais de uma palavra.
02	Compreendendo as funções básicas de <i>corpus</i> e os resultados obtidos.	Treinamento das concordâncias no Google Acadêmico e no SketchEngine for Language Learning (SKELL). Treinamento inicial no <i>Open SketchEngine</i> , usando o BAWE. Leitura e compreensão das linhas de concordância, buscas simples e pautadas no POS / lema / <i>wildcards</i> . Compreensão da frequência / listas de colocações usando a função <i>Word Sketch</i> .

Fonte: Crosthwaite (2020, p. 172, tradução minha)

Percebe-se que, no módulo 1, o tema de interesse é o esclarecimento das vantagens e dos benefícios de se consultar os *corpora* como forma de se pesquisarem aspectos da língua que podem causar dúvidas. Assim, a ênfase dessa etapa recaiu sobre a comparação entre as buscas feitas em ferramentas como dicionários *online* e/o Google Tradutor e aquelas realizadas com o auxílio das ferramentas da LC. Esse momento, portanto, foi importante para ajudar os cursistas a compreenderem ainda melhor o que, de fato, são *corpora* e como eles podem auxiliar a sanar dúvidas que se pode ter sobre a língua.

Já o módulo 2 do curso do Prof. Dr. Peter Crosthwaite enfatiza a utilização do Sketch Engine, a partir do manuseio da versão gratuita da ferramenta³¹, em que a fonte de dados consultada é o *British Academic Written English Corpus* (BAWE)³². Logo, essa foi a etapa em que os cursistas praticaram os comandos de busca mais comumente utilizados ao se pesquisar nos *corpora*. Destaca-se que, apesar de assíncrono, o curso do Prof. Dr. Peter Crosthwaite é organizado de modo didático e

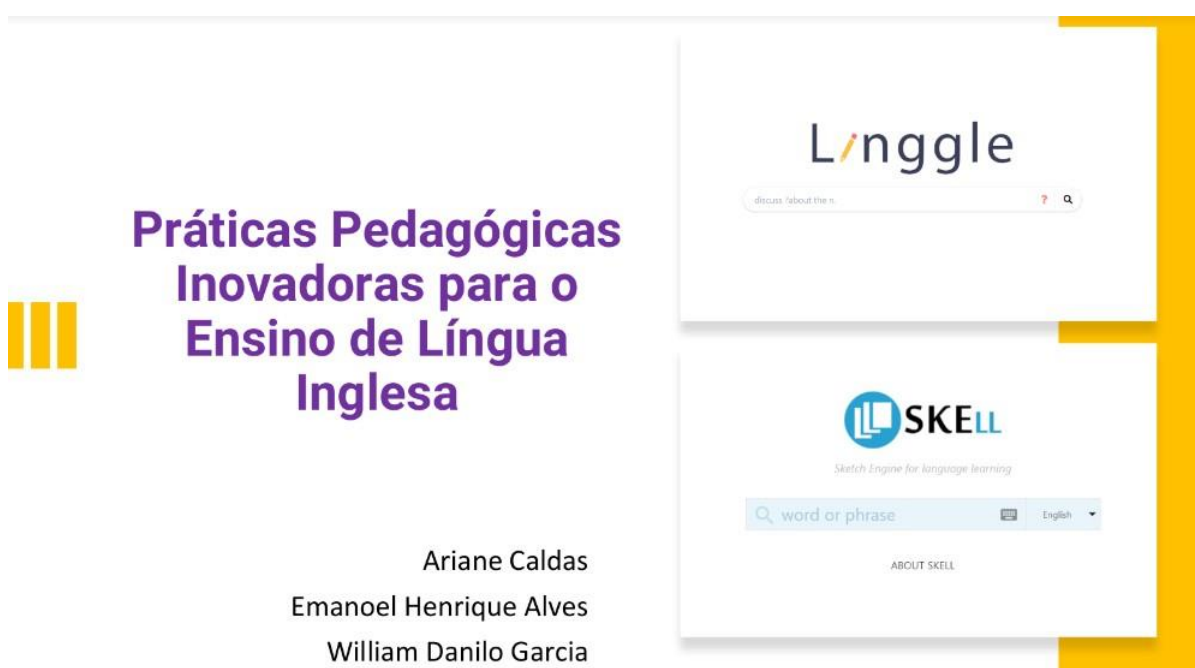
³¹ Disponível em: <https://app.sketchengine.eu/#dashboard?corpname=preloaded%2Fbawe2>. Acesso em: 19 de julho de 2022.

³² Mais informações disponíveis em: <https://www.coventry.ac.uk/research/research-directories/current-projects/2015/british-academic-written-english-corpus-bawe/>. Acesso em: 19 de julho de 2022.

autoexplicativo, para que esses procedimentos de busca abordados no curso possam ser executados pelos cursistas.

Após as duas semanas que envolveram a realização dos dois módulos do curso do Prof. Dr. Peter Crosthwaite, a próxima oficina, ministrada no dia 25/05/2022 pela Profa. Dra. Ariane Caldas, pelo Prof. Me. Emanuel Henrique Alves e pelo autor desta tese, teve como enfoque a apresentação e a instrução de como utilizar as ferramentas Linggle e Skell para a realização de pesquisas sobre a língua.

Figura 18. Capa da apresentação de slides da oficina do dia 25/05/2022



Fonte: elaborado pelos professores responsáveis pela oficina

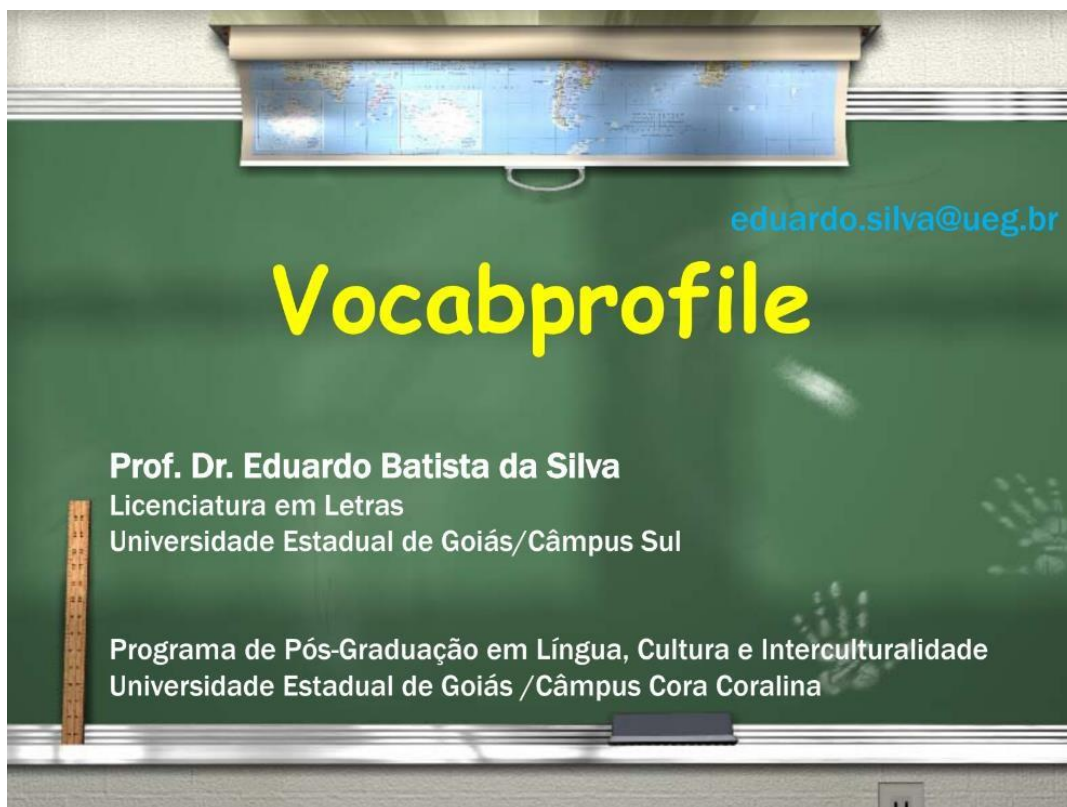
Relevante ainda mencionar que essa oficina também apresentou modelos de atividades didáticas de língua inglesa já desenvolvidas pelos pesquisadores envolvidos no curso. Isso como forma de exemplificar aos cursistas como que eles mesmos podem elaborar suas próprias atividades didáticas. Essas exemplificações foram pautadas no livro *Caderno de Atividades de Aprendizagem Movida por Dados*³³, cujo pesquisador é um dos organizadores. Ademais, duas dessas propostas de exemplificação são trazidas no capítulo 4 desta tese.

³³ Disponível em:

https://www.ponteseditores.com.br/loja/index.php?route=product/product&product_id=1646. Acesso em: 12 de julho de 2023.

A próxima oficina, realizada em 01/06/2022 pelo Prof. Dr. Eduardo Batista da Silva, da Universidade de Goiás, enfatizou a utilização da ferramenta Vocabprofile e lidou com suas possíveis aplicações à sala de aula.

Figura 19. Capa da apresentação de slides da oficina do dia 01/06/2022



Fonte: elaborado pelo professor responsável pela oficina

As duas últimas oficinas, oferecidas em 08/06/2022 e 22/06/2022, lidaram, respectivamente, com as ferramentas Platcol e Online English Collocations Platform, desenvolvidas pela pesquisadora Profa. Dra. Adriane Orenha Ottaiano. Além do Tradcorpus, um repositório de dados de *corpora* on-line que disponibiliza uma série de recursos que podem ser utilizados por professores e/ou pesquisadores para o ensino de línguas estrangeiras e de tradução (Serpa; Pinto; Sbrogio; Camargo, 2023).

Ademais, vale destacar que todas as oficinas ofereceram aos cursistas uma atividade de caráter mais prático-aplicado, no sentido de que eles pudessem experimentar e utilizar, de fato, as ferramentas abordadas em cada uma das oficinas. Além disso, houve uma atividade de conclusão de curso que envolveu o planejamento de uma aula que empregasse alguma das ferramentas estudadas no curso. Foi exatamente com base nessas atividades que o pesquisador recortou algumas das

propostas de atividades didáticas feitas pelos professores participantes e as disponibiliza no capítulo 5.

Compreendidas, então, essas características mais gerais do curso, segue-se para uma análise dos participantes inscritos. De modo breve, atesta-se que houve um total de 42 inscritos no curso, sendo 25 deles do estado de São Paulo, seis do Paraná, cinco de Goiás, dois do Rio de Janeiro e um da Bahia.

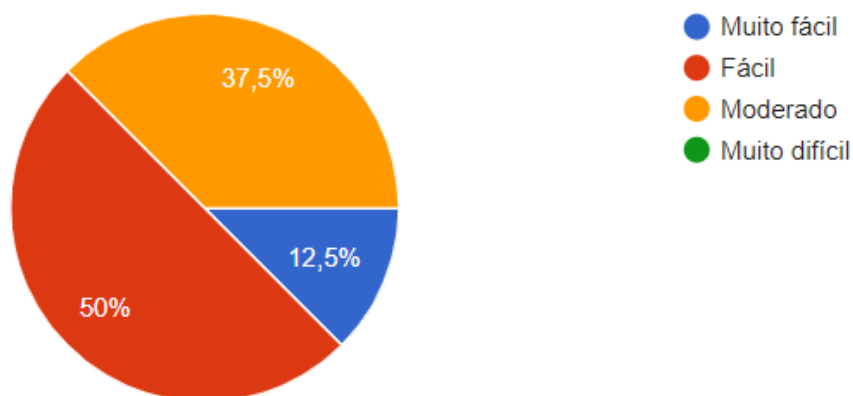
Em relação aos níveis de escolaridade, um cursista indicou o nível superior incompleto enquanto quinze indicaram o nível superior completo. Um acusou o nível de pós-graduação *lato sensu* incompleto e quatorze o nível de pós-graduação *lato sensu* completo. Dois respondentes afirmaram possuir nível de pós-graduação *stricto sensu* incompleto enquanto nove afirmaram ter o nível de pós-graduação *stricto sensu* completo.

No tocante aos contextos de ensino nos quais atuam profissionalmente, 37 participantes do curso são professores das etapas do Ensino Fundamental e do Ensino Médio (seja na rede pública e/ou particular). Desses 37, um deles ainda leciona no Ensino Superior (público), cinco dividem a jornada de trabalho com escolas de idiomas e um com aulas particulares. Para além desses profissionais que se subdividem em mais de um segmento, houve três inscritos que trabalham somente em escolas de idiomas, um que atua apenas como professor do Ensino Superior em uma universidade pública e outro como coordenador.

Apesar de o número de inscritos ser relativamente alto, há de se ponderar que nem todos chegaram a realizar plenamente o curso de formação continuada. Sendo assim, esse fator impactou não só o alcance de formação docente, que acabou ficando reduzido, como também os dados relacionados às respostas deles em relação às ferramentas apresentadas no decorrer das oficinas.

Em relação à análise das opiniões dos professores que participaram do curso sobre as ferramentas da LC, aponta-se, sobre o Skell e o Linggle, o que 8 alunos do curso opinaram sobre essas ferramentas. Das respostas obtidas em relação à facilidade de aplicação desses recursos, não houve nenhum respondente que assinalou a opção muito difícil, o que pode ser visto como um fator positivo. 37,5% deles optaram pelo nível de facilidade moderada, 50% deles sinalizaram o nível fácil e 12,5% deles o nível muito fácil.

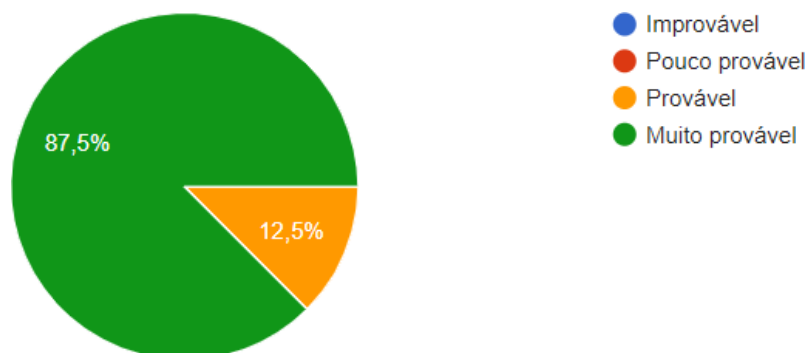
Gráfico 1. Facilidade de aplicação do Skell e do Linggle



Fonte: elaborado pelo autor

Já no que diz respeito à probabilidade de incorporá-los nos contextos de ensino onde atuam, os respondentes assinalaram, majoritariamente, a opção muito provável (87,5%) e 12,5% deles marcaram a opção provável. Considere o gráfico em seguida.

Gráfico 2. Probabilidade de incorporar o Skell e o Linggle no contexto de ensino

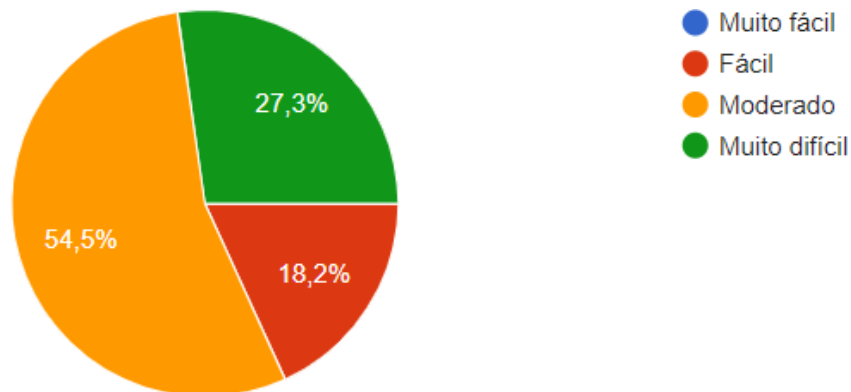


Fonte: elaborado pelo autor

Nota-se, portanto, que ambos o Skell e o Linggle foram considerados não apenas facilmente aplicáveis, como também muito prováveis de serem inseridos nos contextos de ensino onde os respondentes atuam.

Com relação ao Versatext, obteve-se um total de onze cursistas que responderam aos questionários opinando sobre esse recurso. Em relação à facilidade de aplicação da ferramenta, os resultados apontam que 54,5% dos respondentes consideram-na com um nível de facilidade moderado, 18,2% deles fácil e 27,3% deles muito difícil. Nenhum respondente marcou a opção muito fácil.

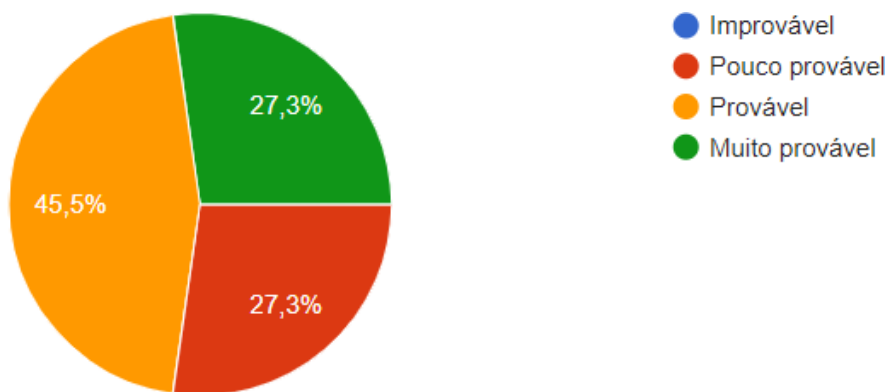
Gráfico 3. Facilidade de aplicação do Versatext



Fonte: elaborado pelo autor

Quando questionados acerca da probabilidade de incorporar o Versatext no contexto de ensino onde atuam, 45,5% dos respondentes consideram-no provável, 27,3% deles muito provável e 27,3% deles pouco provável. Nenhum cursista assinalou a opção improvável.

Gráfico 4. Probabilidade de incorporar o Versatext no contexto de ensino



Fonte: elaborado pelo autor

Esses dados sugerem que, embora nenhum dos cursistas que se propôs a opinar sobre o Versatext o tenha achado uma ferramenta muito fácil de ser aplicada, nenhum deles o considerou improvável de ser incorporado no contexto de ensino onde atuam. Certamente que os dados apontam a necessidade de os alunos do curso compreenderem melhor a aplicação do Versatext haja vista que 27,3% deles o

acharam muito difícil. Todavia, pode-se analisar positivamente o fato de que o maior percentual de respostas acerca da probabilidade de incorporação ao ensino se mostrou favorável (cerca de metade deles sinalizaram a opção provável e cerca de ¼, a opção muito provável).

Com base nas análises das respostas dos cursistas, o pesquisador optou por filtrar, dentre todas as ferramentas abordadas no curso de formação continuada em serviço, essas três, o Linggle, o Skell e o Versatext, para lhes conferir um maior enfoque. Isso tanto nas propostas de exemplificação, trazidas no capítulo 4, quanto nas propostas de atividades didáticas de língua inglesa que foram elaboradas pelos participantes do curso e pelo próprio pesquisador, como forma de contemplar as habilidades e/ou conteúdos estipulados para o Ensino Médio, conforme o Currículo Paulista, e que são apresentadas no capítulo 5 desta tese.

Em conclusão, viu-se, neste capítulo, a natureza desta pesquisa e a metodologia que estrutura e organiza esta tese de doutoramento. Na sequência, explicou-se sobre o curso de formação continuada em serviço de curta duração, organizado pelo autor desta tese, em parceria com outra doutoranda da mesma linha de pesquisa e com a colaboração de outros pesquisadores da LC. Por fim, foram apresentadas as análises das opiniões de alguns dos cursistas, em relação a três das ferramentas abordadas no curso, o Skell, o Linggle e o Versatext, justamente pelo fato de eles as considerarem mais facilmente utilizáveis e mais prováveis de serem empregadas nos contextos educacionais onde atuam.

4.2 As ferramentas

Prévia à apresentação das ferramentas, é preciso esclarecer que elas só são devidamente pontuadas, nesta tese, após a descrição do curso de formação continuada em serviço porque foi, justamente, durante a formação docente que se obtiveram dados que apontaram os *softwares* considerados mais fáceis de serem manuseados e utilizados na preparação de atividades didáticas. Assim, por mais que haja um número considerável de ferramentas de *corpora* disponíveis atualmente e que podem ser incorporadas em sala de aula para o ensino de língua inglesa, como o caso de Garcia (2020), por exemplo, que lidou com a utilização do *Corpus of Contemporary*

*American English (COCA)*³⁴ para o ensino de preposições, este estudo delimita um recorte com base nos apontamentos feitos no curso de formação docente, lidando, portanto, com apenas três dessas ferramentas: o Skell, o Linggle e o Versatext.

4.2.1 A ferramenta Skell

O *Sketch Engine for Language Learners* (SKELL) é uma ferramenta que permite que estudantes e professores de língua inglesa consultem como uma palavra é usada por falantes reais da língua inglesa³⁵. Além de ser gratuito, o Skell é disponibilizado sem a necessidade de se registrar e/ou criar uma conta. Nesse sentido, isso facilita o seu acesso em sala de aula e ainda preserva os dados dos usuários.

Por se tratar de uma ferramenta com base em *corpora*, faz-se necessário descrever que seus dados são pautados em artigos da Wikipedia, em partes do *English Web 2013 corpus*³⁶, no *Timestamped web corpus*³⁷, no *British National Corpus*³⁸ e em sites em inglês explorados pela ferramenta *WebBootCat*³⁹, como demonstra o quadro disponibilizado a seguir:

³⁴ *Corpus* disponível em: <https://www.english-corpora.org/coca/>. Acesso em: 19 de julho de 2022.

³⁵ Conforme definição do site institucional do Skell: <https://www.sketchengine.eu/skell/>. Acesso em: 19 de julho de 2022.

³⁶ Mais informações em: <https://www.sketchengine.eu/entente-english-corpus/>. Acesso em: 19 de julho de 2022.

³⁷ Mais informações em: <https://www.sketchengine.eu/feed-corpus-project/>. Acesso em: 19 de julho de 2022.

³⁸ Mais informações em: <http://www.natcorp.ox.ac.uk/>. Acesso em: 19 de julho de 2022.

³⁹ Mais informações em: <https://www.sketchengine.eu/build-a-corpus-from-the-web-lesson/>. Acesso em: 19 de julho de 2022.

Quadro 4. Os dados que alimentam o Skell

Source	no. of words	percentage
Wikipedia	~ 403,653,953	~ 38.88%
English Web 2013	~ 320,892,066	~ 30.91%
Timestamped web corpus	~ 146,082,464	~ 14,07%
British National Corpus	~ 90,379,212	~ 8,71%
WebBootcat	~ 77,192,480	~ 7,44%
Total	1,038,200,313	100%

Fonte: site institucional da ferramenta⁴⁰

Observa-se, portanto, que o Skell conta com uma base de dados que totaliza mais de 1 bilhão de palavras, o que indica o quanto as informações são representativas da língua inglesa. Outra consideração importante a ser feita é a de que essa ferramenta se baseia em bons exemplos de uso da língua, o que facilita ainda mais sua utilização para fins didáticos e pedagógicos. Isso se deve a um sistema denominado *Good Dictionary Examples* (GDEX)⁴¹, que se responsabiliza por filtrar, de certa maneira, frases que sejam muito longas ou que contenham vocabulários mais complicados da língua estrangeira.

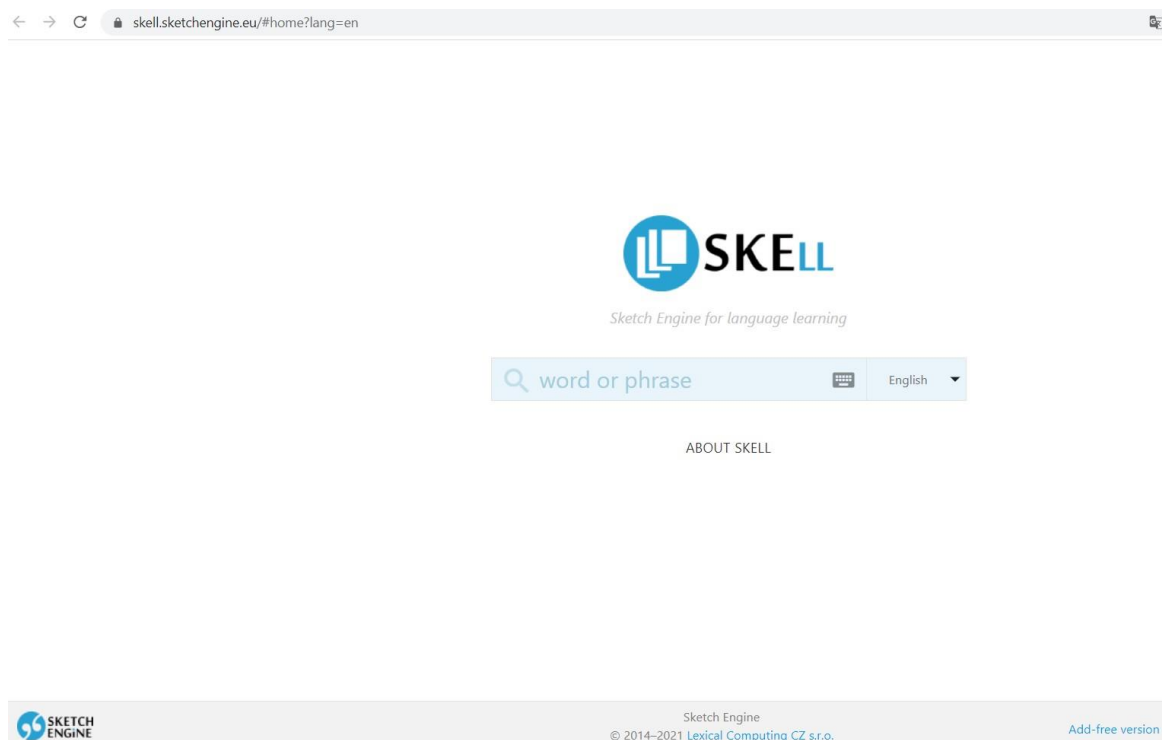
Compreendido isso, vale destacar que o Skell disponibiliza, basicamente, três funções. A primeira delas é denominada *Examples* e diz respeito às linhas de concordância. A segunda é responsável por apresentar colocações, ou seja, combinações de duas palavras e é chamada de *Word Sketch*. A terceira é a *Synonyms* e ela apresenta uma lista de palavras – também dispostas no formato de nuvem de palavras (*word cloud*) – que revela uma relação de sinonímia com a palavra buscada.

Para melhor entendimento dessas funções, mostra-se pertinente a exemplificação de como utilizá-las. Primeiramente, é necessário acessar a página inicial do Skell, conforme a imagem a seguir.

⁴⁰ Disponível em: <https://www.sketchengine.eu/english-skell-corpus/>. Acesso em: 19 de julho de 2022.

⁴¹ Mais informações em: <https://www.sketchengine.eu/guide/gdex/>. Acesso em: 19 de julho de 2022.

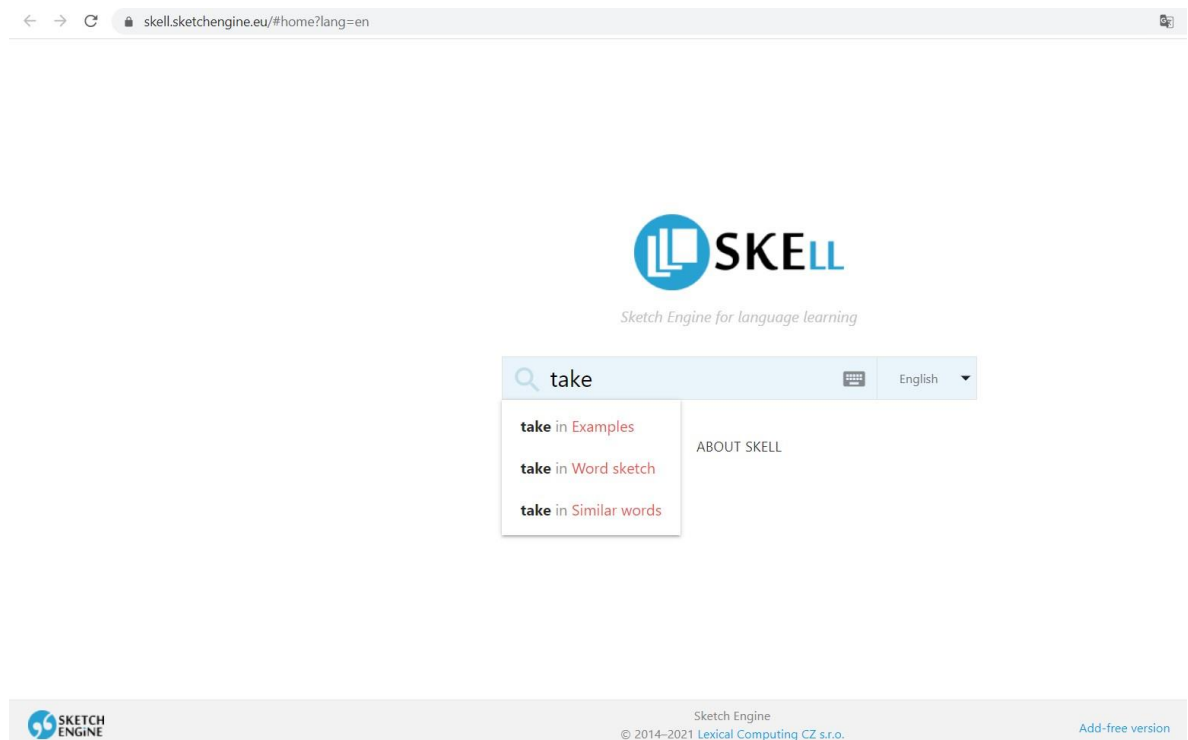
Figura 20. Página inicial do Skell



Fonte: site do Skell

Feito isso, insere-se um termo de busca do qual se queira obter mais informações. Suponha-se que o interesse seja compreender melhor o verbo *take*. Só o que precisa ser feito é digitá-lo no campo de busca. A ferramenta automaticamente lhe dará as três funções disponíveis, como ilustra a figura a seguir.

Figura 21. Consultando uma palavra no Skell



Fonte: site do Skell

Caso se queira analisar os diferentes usos desse verbo, basta selecionar a opção *Examples* e serão disponibilizadas as linhas de concordância, como se pode observar na figura apresentada a seguir.

Figura 22. Resultado de busca do verbo *take* na função *Examples*

The screenshot shows the Skell sketch engine interface. The browser address bar displays the URL: `skell.sketchengine.eu/#result?lang=en&query=take&f=concordance`. The page header includes the Skell logo and a search bar with the query "take". Below the search bar, there are tabs for "Examples", "Word sketch", and "Similar words". The "Examples" tab is selected, showing a list of 17 concordance lines for the verb "take". The word "take" is highlighted in red in each line. The list includes various grammatical forms: "taken", "taking", "took", and "Take".

take 1,293.84 hits per million ⓘ

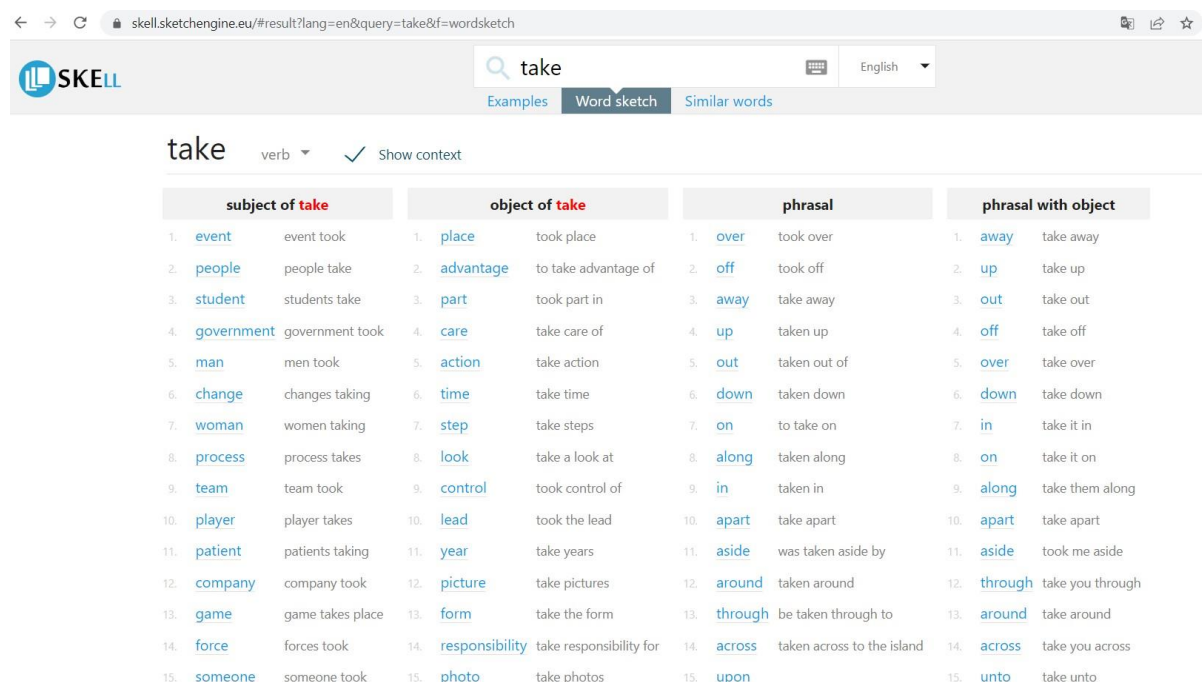
1. These natural body substances are usually **taken** together.
2. Many conflicts are **taking** place amid general indifference.
3. The whole procedure had **taken** almost three hours.
4. Individual clinicians are constantly **taking** management decisions.
5. The traumatized cell phone conversations **took** place again.
6. Todd began **taking** drum lessons 8 years ago.
7. This trail **takes** approximately $2\frac{1}{2}$ hours.
8. Can **taking** vitamins prevent daily drinking damage?
9. The conversation **took** place many years ago.
10. There are more systemic changes **taking** place.
11. This intangible evil however presently **took** shape.
12. **Take** every 30 minutes until relief is maintained.
13. The government **takes** private property through condemnation proceedings.
14. She **took** communion without realising anything yet.
15. It was **taken** before oral medication was administered.
16. This contact was soon **taken** under fire.
17. The team decision making is **taken** real seriously.

Fonte: site do Skell

Percebe-se, com base nas linhas de concordância, que os resultados obtidos levam em consideração todas as possíveis formas do verbo *take*. Isso é devido ao fato de o Skell proporcionar resultados lematizados do termo de busca. O aspecto positivo disso é que os estudantes já conseguem visualizar as diferentes formas de uso da palavra que estiverem buscando.

Já no caso de a busca ter o intuito de consultar as combinações possíveis de palavras (colocações e/ou coligações) envolvendo o verbo *take*, basta selecionar a opção *Word sketch*. Como exemplo, considere a próxima imagem.

Figura 23. Resultado da busca do verbo *take* na função *Word sketch*



The screenshot shows the Skell website interface. At the top, there's a search bar with 'take' entered. Below the search bar, there are tabs for 'Examples', 'Word sketch' (which is selected), and 'Similar words'. The main content area displays the word 'take' with a dropdown menu set to 'verb' and a 'Show context' link. Below this, there are four columns of results, each with a header: 'subject of take', 'object of take', 'phrasal', and 'phrasal with object'. Each column contains a list of 15 items, each consisting of a numbered word and a short example sentence.

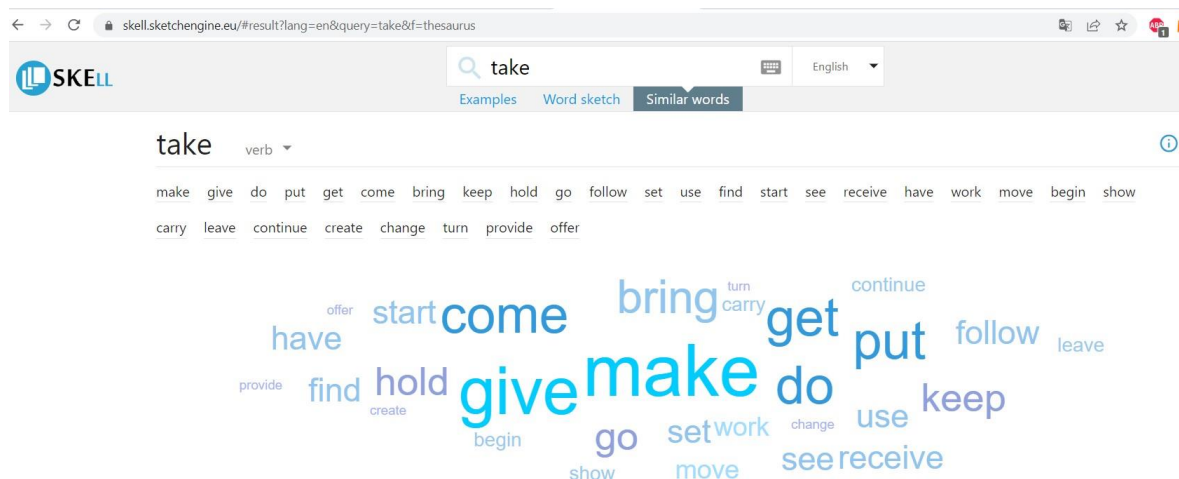
subject of take		object of take		phrasal		phrasal with object	
1. event	event took	1. place	took place	1. over	took over	1. away	take away
2. people	people take	2. advantage	to take advantage of	2. off	took off	2. up	take up
3. student	students take	3. part	took part in	3. away	take away	3. out	take out
4. government	government took	4. care	take care of	4. up	taken up	4. off	take off
5. man	men took	5. action	take action	5. out	taken out of	5. over	take over
6. change	changes taking	6. time	take time	6. down	taken down	6. down	take down
7. woman	women taking	7. step	take steps	7. on	to take on	7. in	take it in
8. process	process takes	8. look	take a look at	8. along	taken along	8. on	take it on
9. team	team took	9. control	took control of	9. in	taken in	9. along	take them along
10. player	player takes	10. lead	took the lead	10. apart	take apart	10. apart	take apart
11. patient	patients taking	11. year	take years	11. aside	was taken aside by	11. aside	took me aside
12. company	company took	12. picture	take pictures	12. around	taken around	12. through	take you through
13. game	game takes place	13. form	take the form	13. through	be taken through to	13. around	take around
14. force	forces took	14. responsibility	take responsibility for	14. across	taken across to the island	14. across	take you across
15. someone	someone took	15. photo	take photos	15. upon		15. unto	take unto

Fonte: site do Skell

Nota-se que há um botão denominado *show context*. Ao selecioná-lo, obtém-se uma visualização mais ampla das colocações envolvendo o termo de busca. Além disso, é válido mencionar que essa função traz diferentes combinações com *take*, tais como possíveis sujeitos e objetos para esse verbo, além de combinações com preposições etc. Caso haja interesse em ver exemplos dessas combinações, só o que precisa ser feito é clicar nelas que o próprio Skell o direcionará para linhas de concordâncias que demonstrem os usos possíveis da colocação selecionada.

Por fim, se o interesse da busca for checar palavras que possuam certo tipo de relação sinônímica com o verbo *take*, escolha a opção *Similar words*. Considere a imagem a seguir.

Figura 24. Resultado de busca do verbo *take* na função *Similar words*



Fonte: site do Skell

Observa-se que o resultado obtido é dado tanto na forma de lista de palavras quanto na forma de nuvem de palavra (*word cloud*). Caso haja interesse em analisar alguma das palavras trazidas pela ferramenta semelhante, em algum sentido, ao termo buscado, só o que precisa ser feito é clicar nessa palavra que o Skell o direcionará às linhas de concordância ou às combinações possíveis desse item.

Enfim, viu-se nesta subseção a ferramenta Skell e uma breve descrição de como ela pode ser utilizada para consultar diferentes informações sobre uma palavra da língua inglesa. Reitera-se que a facilidade em manuseá-la é um fator bastante positivo e que facilita sua aplicação em sala de aula.

4.2.2 A ferramenta Linggle

O Linggle⁴² é uma ferramenta de busca linguística e gramatical em língua inglesa desenvolvida pela Universidade Nacional Tsing Hua, no Taiwan. A base de dados abarca *corpora* como o *Google Web 1T 5-gram*⁴³ (lançado em 2006 em contribuição com o Google), o *British Nacional Corpus* e o *New York Times Annotated Corpus*⁴⁴ (lançado em 2008 e contendo mais de 1.8 milhões de artigos publicas pelo *The New York Times*).

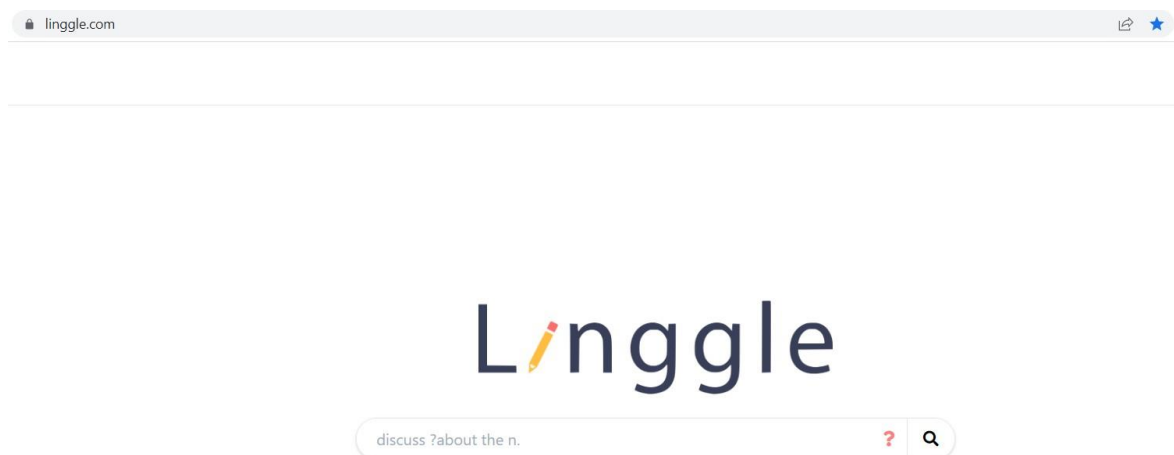
⁴² Mais informações em: <https://libguides.lb.polyu.edu.hk/online-tools-for-assignment/linggle>. Acesso em: 19 de julho de 2022.

⁴³ Mais informações em: <https://catalog ldc.upenn.edu/LDC2006T13>. Acesso em: 19 de julho de 2022.

⁴⁴ Mais informações em: <https://catalog ldc.upenn.edu/LDC2008T19>. Acesso em: 19 de julho de 2022.

O Linggle⁴⁵ também pode ser considerado uma ferramenta de fácil manuseio para os alunos e os professores e que, portanto, pode ser mais facilmente incorporada à sala de aula. A seguir, tem-se a página inicial dele.

Figura 25. Página inicial do Linggle

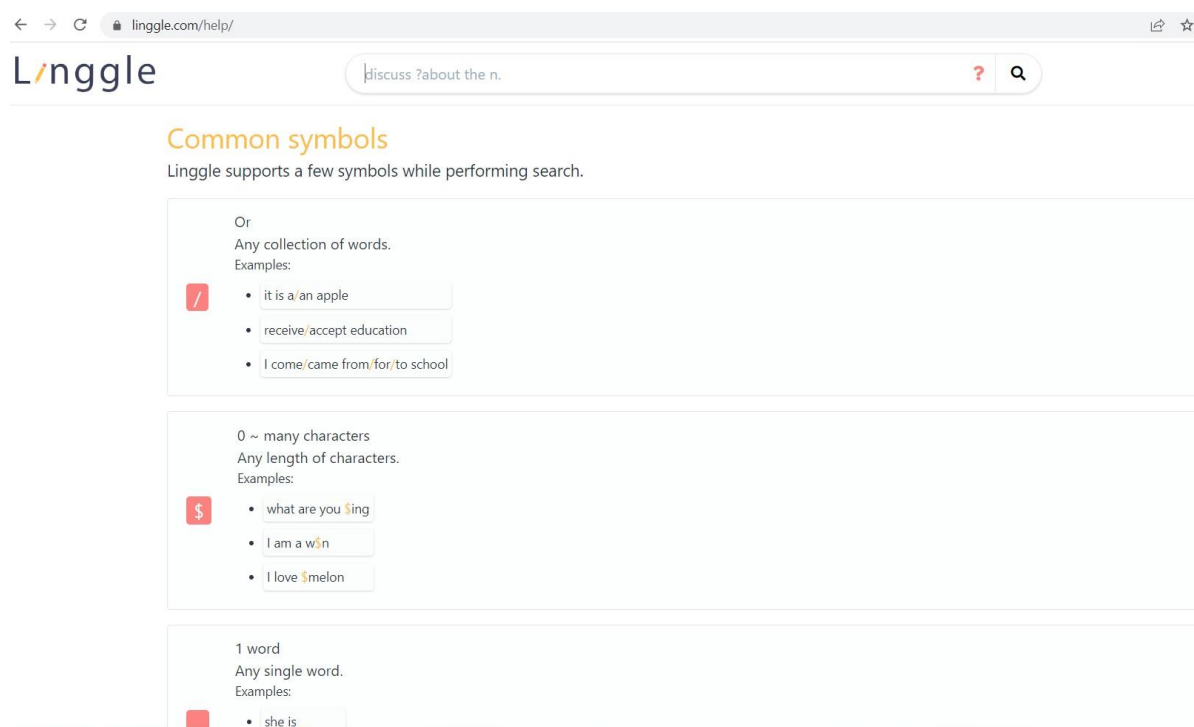


Fonte: site do Linggle

Ao observar a imagem, percebe-se que, ao lado da barra de busca, há um botão no formato de um ponto de interrogação. Trata-se de uma forma de botão de ajuda em que, ao clicá-lo, encontram-se as formas com as quais se pode realizar as buscas na ferramenta. A próxima figura retrata esses comandos.

⁴⁵ Link para acesso à ferramenta: <https://linggle.com/>. Acesso em: 19 de julho de 2022.

Figura 26. Os comandos de busca do Linggle



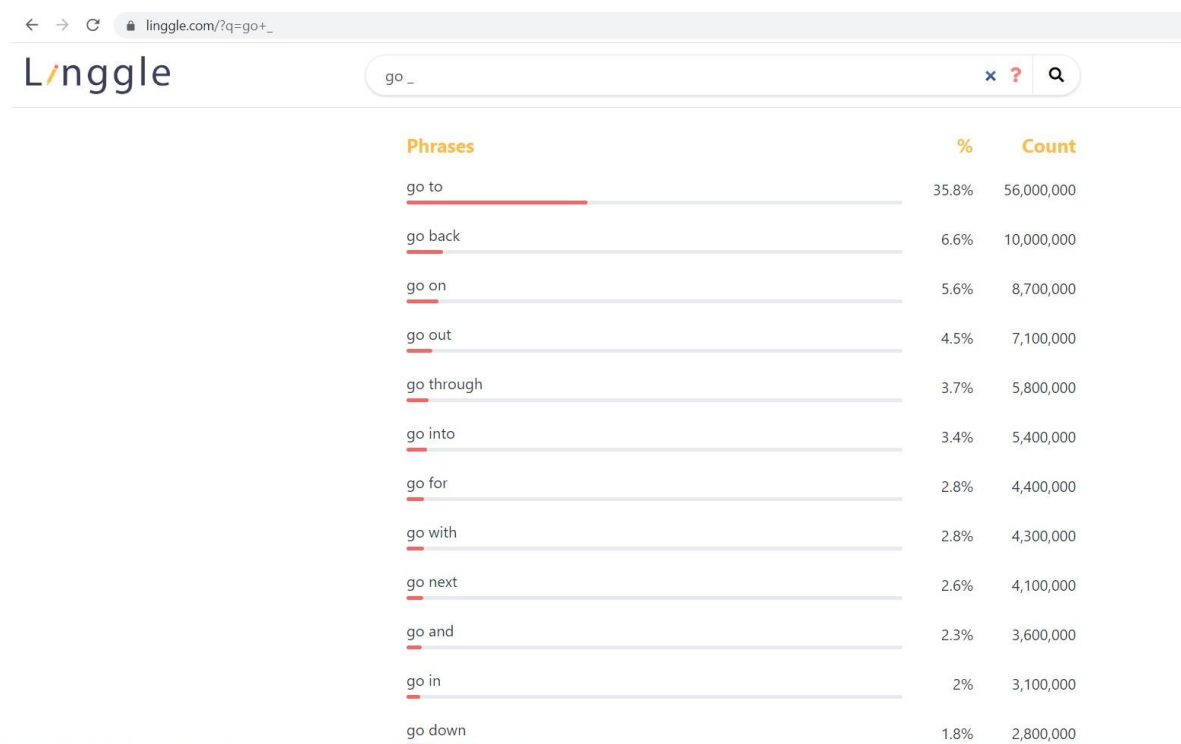
Fonte: site do Linggle

Como se pode ver, as opções de comando de busca são variadas, porém muito bem explicadas pelo próprio site, que traz, inclusive, exemplos de como realizar cada uma das formas de pesquisa. Por esse motivo, não serão descritas aqui todas as formas de busca disponíveis, mas aquelas mais comumente utilizadas, conforme observou Crosthwaite, Wong e Cheung (2019) ao analisar e rastrear as formas a partir das quais estudantes em contexto de pós-graduação realizaram suas consultas no *corpus*. Dentre as descobertas dos pesquisadores, destacam-se o fato de que o uso de *wildcards* (elementos curingas) como o asterisco (*) ou o símbolo do subtraço (□) se mostrou recorrente, assim como o das etiquetas morfossintáticas (as *POS tags*) (Crosthwaite; Wong; Cheung, 2019).

Dessa maneira, julga-se pertinente explicar que o subtraço (□) no Linggle é usado como representação de qualquer palavra. Assim, caso se queira saber qual palavra combina com o verbo *go*, por exemplo, basta utilizar o seguinte comando: go

—.

Figura 27. Resultados da busca go _ no Linggle



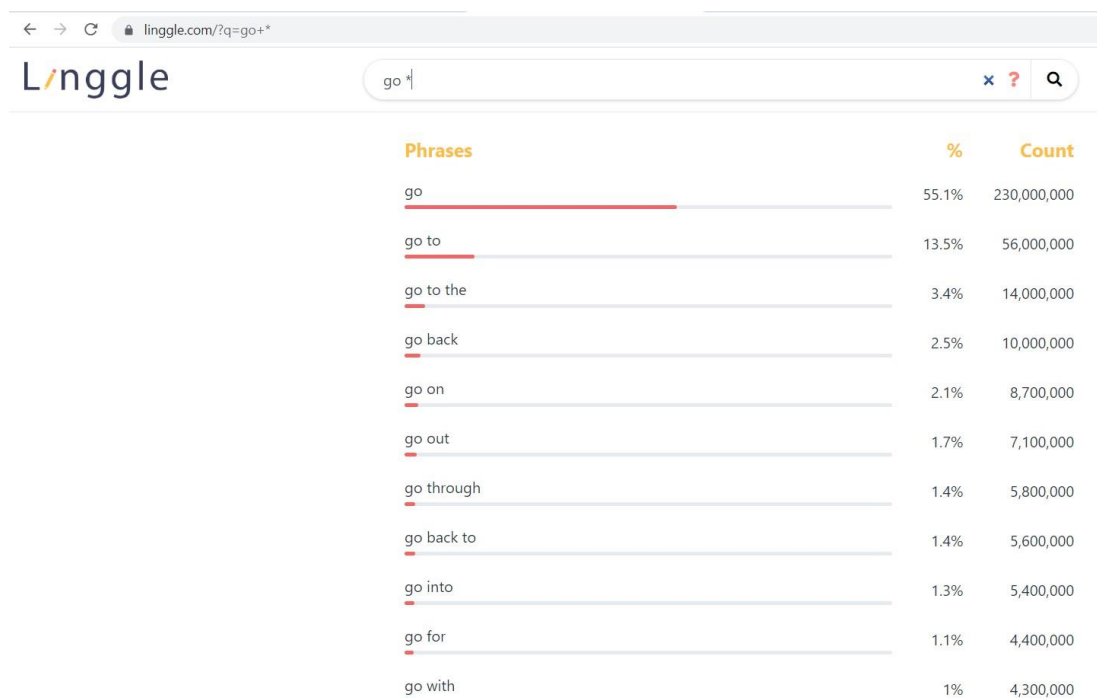
Phrases	%	Count
go to	35.8%	56,000,000
go back	6.6%	10,000,000
go on	5.6%	8,700,000
go out	4.5%	7,100,000
go through	3.7%	5,800,000
go into	3.4%	5,400,000
go for	2.8%	4,400,000
go with	2.8%	4,300,000
go next	2.6%	4,100,000
go and	2.3%	3,600,000
go in	2%	3,100,000
go down	1.8%	2,800,000

Fonte: site do Linggle

Logo se vê que os resultados obtidos indicam que a palavra que mais ocorre com esse verbo é a preposição *to*, haja vista que essa combinação aparece em primeiro lugar. Caso haja a necessidade de ver como esse verbo e essa preposição são usados juntos, basta clicar em cima dessa opção que a ferramenta trará linhas de concordância contendo exemplos dessa combinação.

Já o uso do asterisco (*) se refere a busca por qualquer número de palavras. Sendo assim, o procedimento de busca *go ** traz os seguintes resultados:

Figura 28. Resultados da busca go * no Linggle



Phrases	%	Count
go	55.1%	230,000,000
go to	13.5%	56,000,000
go to the	3.4%	14,000,000
go back	2.5%	10,000,000
go on	2.1%	8,700,000
go out	1.7%	7,100,000
go through	1.4%	5,800,000
go back to	1.4%	5,600,000
go into	1.3%	5,400,000
go for	1.1%	4,400,000
go with	1%	4,300,000

Fonte: site do Linggle

Vê-se que, diferentemente do procedimento de busca anterior, a pesquisa com o asterisco (*) mostra combinações com mais de uma palavra, tais como *go to the* e *go back to*.

No que tange às etiquetas morfossintáticas (as *POS tags*), é preciso que o consulente leve em consideração as legendas oferecidas na aba ajuda, representada pelo símbolo do ponto de interrogação, conforme descrição prévia. De modo que, caso o que se queira saber seja quais substantivos são recorrentemente usados após o verbo *love*, o comando de busca deverá incluir a abreviação *n.* (referente a *noun*, ou seja, substantivo). Considere a figura a seguir como exemplo.

Figura 29. Resultados da busca *love n.* no Linggle

The screenshot shows the Linggle website interface. At the top, there's a browser address bar with 'linggle.com/?q=love+n.' and the Linggle logo. A search bar contains 'love n.' with a magnifying glass icon. Below the search bar, a table displays the results. The table has three columns: 'Phrases', '%', and 'Count'. The phrases are listed with a red underline on the word 'love'. The percentages and counts are shown to the right of each phrase.

Phrases	%	Count
love story	9.8%	710,000
love affair	5.8%	420,000
love life	5.6%	400,000
love song	4%	290,000
love being	3.9%	280,000
love will	3.7%	270,000
love one	3.6%	260,000
love songs	3.5%	250,000
love love	3.1%	220,000
love music	2.9%	210,000
love triangle	2.8%	200,000

Fonte: site do Linggle

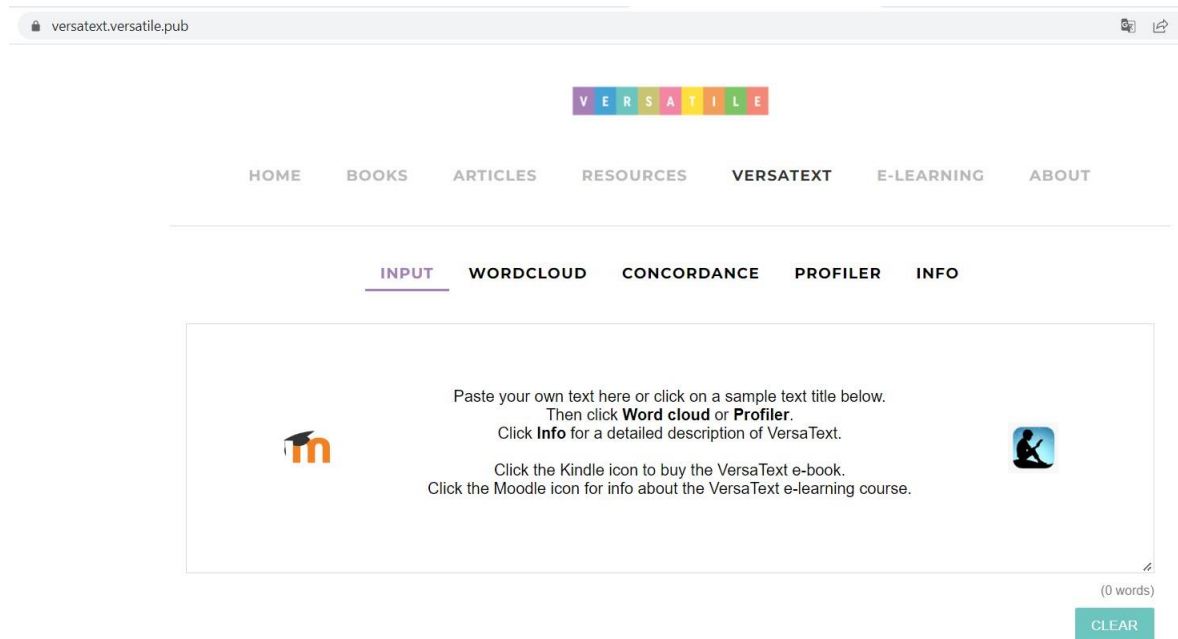
Espera-se que por meio desses exemplos das diferentes formas de pesquisa no Linggle possa se ter notado o quão facilmente utilizável é essa ferramenta. Nesse sentido, acredita-se que seu uso em sala de aula não cause tanto estranhamento nos alunos e nem nos professores, tal como o Skell.

4.2.3 A ferramenta Versatext

Diferentemente das outras duas ferramentas descritas até então, que representaram formas mais fáceis de consulta ao *corpus*, o Versatext⁴⁶ é uma ferramenta que possibilita a compilação rápida de textos.

⁴⁶ Link para acessar a ferramenta: <https://versatext.versatile.pub/>. Acesso em 19 de julho de 2022.

Figura 30. Página inicial do Versatext



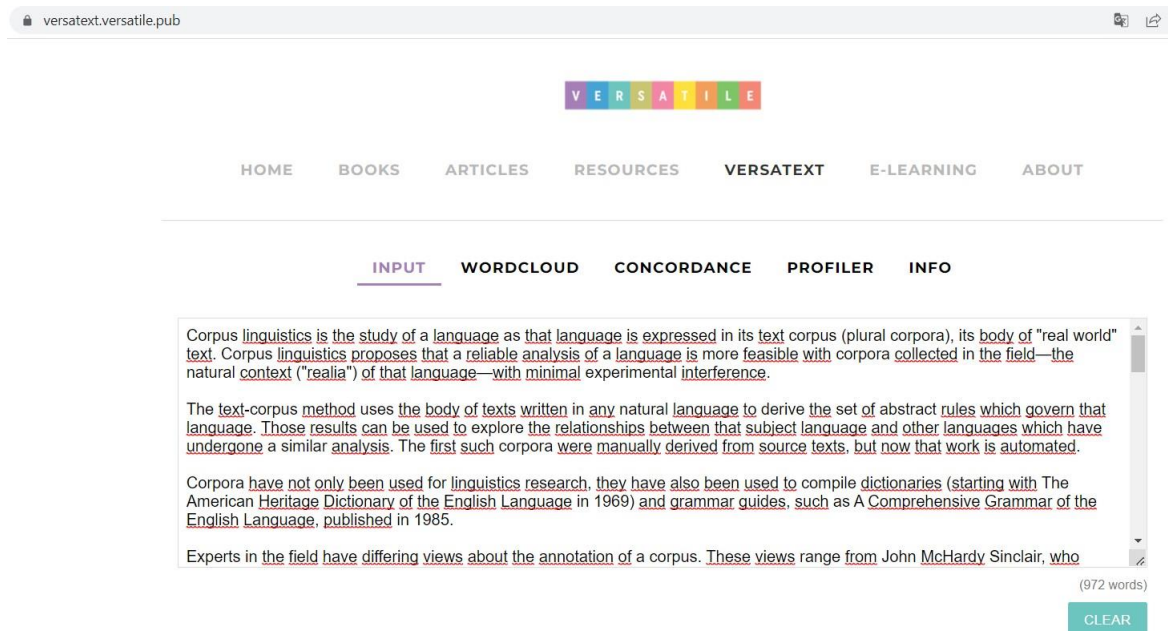
Fonte: site do Versatext

Como sugere a figura, basta utilizar a aba *input* para colar um texto ou mais dentro do quadro indicado e, então, gerar a análise (por meio da opção *Word cloud* ou *Profiler*) do material textual inserido. Uma observação relevante é que os professores podem utilizar textos que serão abordados nas aulas de língua inglesa ou até mesmo as produções que foram escritas pelos próprios alunos.

Para fins de exemplificação do funcionamento da ferramenta, utilizou-se aqui um excerto do texto da Wikipedia que aborda a definição de LC⁴⁷. Considere a demonstração por meio da figura que se segue.

⁴⁷ Texto disponível em: https://en.wikipedia.org/wiki/Corpus_linguistics. Acesso em: 19 de julho de 2022.

Figura 31. Inserção do texto no Versatext



Fonte: site do Versatext

Em seguida, pode-se clicar na opção *Word cloud* para se gerar a nuvem de palavras desse texto, que consta com um total de 972 palavras, como ilustrado na imagem anterior. Observa-se que essa função, além de gerar a nuvem de palavras, disponibiliza uma série de opções de filtro dos resultados alcançados.

Nesse sentido, é possível delimitar o número de palavras a serem exibidas na nuvem, assim como optar se elas deverão ser consideradas como palavras ou em suas formas lematizadas ou ainda em suas classificações morfossintáticas. Além disso, pode-se filtrar quais palavras lexicais ou de conteúdo (as *content words*) e quais palavras funcionais (as *function words*) se deseja visualizar nos resultados, tal como é retratado na imagem a seguir.

versatext.versatile.pub

INPUT WORDCLOUD CONCORDANCE PROFILER INFO

Words to show

10 20 30 40 50 60 70 80 90 100

WORD ☒ LEMMA ☐ PART OF SPEECH ☐ TAG ☐

Colour scheme

DEFAULT ☒ BW ☐ ACCESSIBLE ☐

Include content words

NOUN ☒ VERB ☒ ADJECTIVE ☒ ADVERB ☒

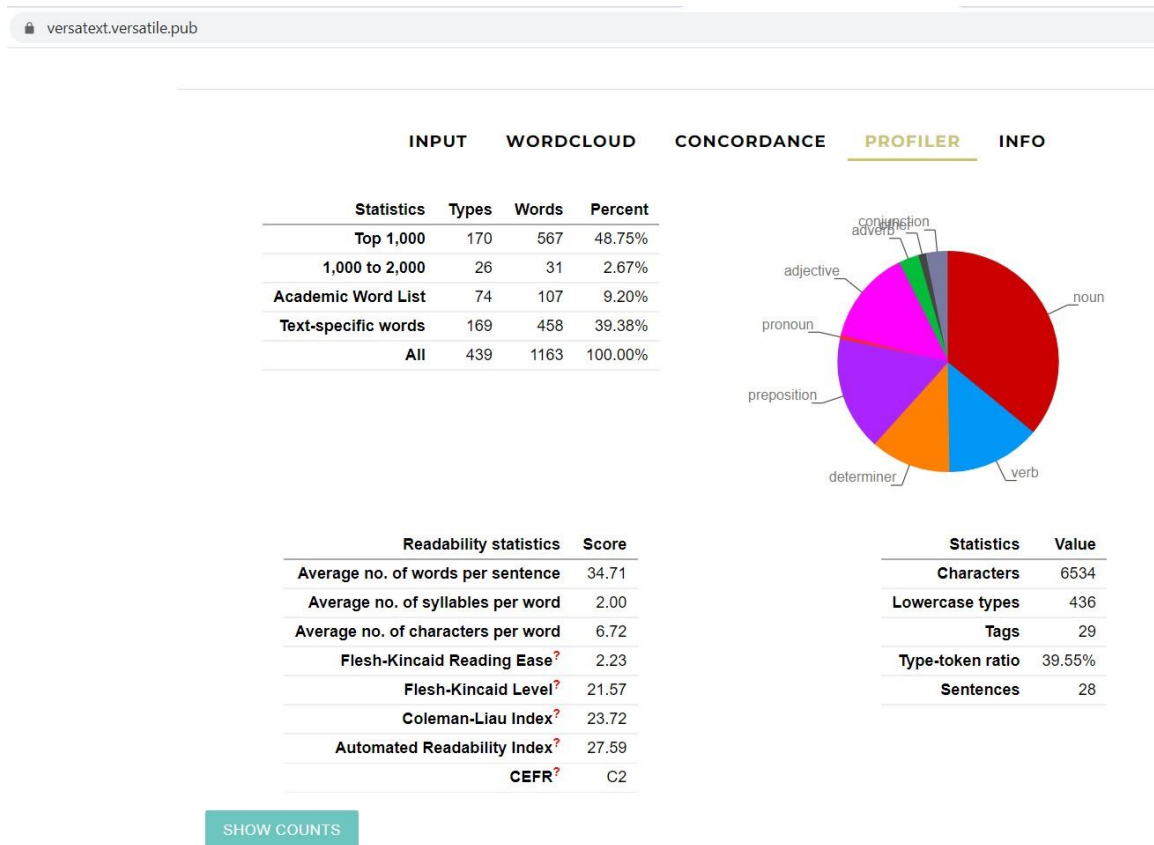
BE+HAVE+DO ☐

Include function words

PREPOSITION ☐ CONJUNCTION ☐ PRONOUN ☐

Por fim, a função *Profiler* diz respeito à demonstração de dados com caráter mais estatístico em relação ao texto que foi inserido na ferramenta. Desse modo, ela informa, por exemplo, a quantidade de itens (total do número de palavras) e de formas (total de palavras diferentes) presentes no material textual analisado. Considere a imagem a seguir para exemplificação dessa função.

Figura 33. A função Profiler



Fonte: site do Versatext

Para concluir, foram exploradas, nesta subseção, as aplicabilidades da ferramenta Versatext e o quanto ela oferece possibilidades de uso em atividades didáticas que visam ao ensino de língua inglesa, assim como as outras duas abordadas anteriormente – o Skell e o Linggle.

5 PROPOSTAS DE ATIVIDADES DIDÁTICAS: IMPLEMENTAÇÕES POSSÍVEIS AOS MATERIAIS DIDÁTICOS

Após a descrição dos materiais e métodos, torna-se pertinente abordar as atividades didáticas desenvolvidas pelos professores que participaram do curso de formação continuada em serviço. Dentre as propostas elaboradas pelos cursistas, serão descritas, neste capítulo, aquelas que recorreram à utilização das ferramentas tratadas nesta tese (Skell, Linggle e Versatext) e que contaram com a revisão e o apoio técnico do pesquisador para que pudessem melhor atender às especificidades da Aprendizagem Movida por Dados. Ademais, destaca-se que essas atividades didáticas se mostram como formas de implementação de materiais didáticos de língua inglesa.

5.1 Propostas de atividades didáticas desenvolvidas pelos professores durante o curso de formação continuada em serviço

No decorrer dos módulos do curso de formação continuada em serviço, os professores precisaram realizar algumas tarefas, que envolveram a elaboração de atividades didáticas amparadas pela Aprendizagem Movida por Dados. Para tanto, utilizaram ferramentas da LC apresentadas nas oficinas, tais como Skell, Linggle e Versatext. Assim sendo, listam-se, na sequência, as propostas desenvolvidas pelos cursistas. Veja, a seguir, um quadro que resume as informações dessas atividades.

Quadro 5. Lista das atividades didáticas elaboradas pelos cursistas

Proposta	Conteúdo/foco	Contexto de ensino	Ferramenta da LC
1.	Verbos	Curso de idiomas - crianças	Skell
2.	<i>Noun phrases</i>	Ensino Médio	Skell
3.	<i>Collocations</i>	Ensino Médio	Skell
4.	Collocations	Ensino Superior	Skell
5.	Quantificadores	Ensino Médio	Linggle
6.	Preposições	Escola de idiomas	Linggle
7.	<i>Collocations</i>	Ensino Médio	Linggle
8.	Vocabulário	Ensino Médio	Linggle
9.	Vocabulário	Ensino Médio	Versatext
10.	Leitura	Ensino Superior	Versatext e Linggle

Fonte: elaborado pelo autor

5.1.1 Atividade com a canção *Kangaroo Groove*

Inicialmente, enfatiza-se que esta proposta foi planejada para o contexto de ensino de língua inglesa para o público infantil. A atividade visa a implementar o material didático intitulado *Super Seek and Find*, da editora *Macmillan*, conforme se observa na imagem a seguir.

Figura 34. Ilustração do material didático *Super Seek and Find*



Fonte: *Super Seek and Find level 4*⁴⁸

Como se pode notar, trata-se de uma atividade com música, cuja letra envolve um canguru que ensina às crianças alguns movimentos. O professor que desenvolveu essa proposta sugeriu que, após trabalhada a canção com os estudantes, uma forma de explorar os vocabulários nela contidos pode ser concebida com o auxílio da ferramenta Skell.

Para isso, o professor seleciona uma lista de palavras retiradas da letra da música e propõe que os estudantes utilizem a opção *Similar Words*, no Skell, para identificar possíveis sinônimos. Além disso, o exercício envolve a identificação, dentre

⁴⁸ Mais informações em: <https://www.macmillanenglish.com/br/catalogue/courses/young-learners/super-seek-and-find/levels-samples>. Acesso em: 12 de janeiro de 2024.

as opções listadas, de palavras que não expressam o sentido de movimento. O quadro seguinte demonstra a sugestão do cursista.

Quadro 6. Atividade para análise de vocabulários da música

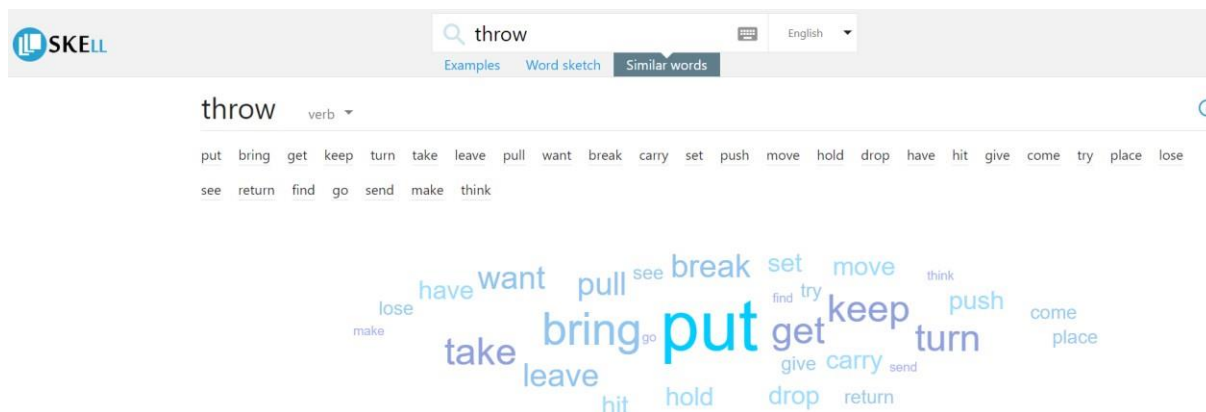
Atividade: Utilizando a opção Similar Words, na ferramenta SKELL, encontre palavras parecidas com as listadas abaixo. Em seguida, diga qual delas não indica movimento?

		Palavras parecidas	Indica movimento?
1.	Throw		() Sim () Não
2.	Surf		() Sim () Não
3.	Stamp		() Sim () Não
4.	Gallop		() Sim () Não
5.	Stomp		() Sim () Não
6.	Slide		() Sim () Não
7.	Heaps		() Sim () Não
8.	Hop		() Sim () Não
9.	Wiggle		() Sim () Não

Fonte: pesquisador a partir da proposta do cursista

Para responderem à atividade, os estudantes precisam realizar a busca no Skell por meio da opção *Similar Words*. Desse modo, espera-se que seja feita a consulta conforme o exposto na sequência.

Figura 35. Exemplificação do procedimento de busca com a palavra *throw*



Fonte: captura de tela da ferramenta Skell

Nota-se, então, que as possíveis respostas, nesse caso, para palavras parecidas com *throw* seriam *put*, *bring*, *get*, *keep*, *turn* etc. Além disso, pelas opções dadas pelo Skell, os estudantes, possivelmente, podem perceber que *throw* se trata de um vocábulo com o sentido de movimento.

Sob uma perspectiva analítica, afirma-se que esta proposta de atividade didática é exequível e de simples implementação por parte do docente, que simplesmente elaborou uma lista de palavras a partir dos vocabulários presentes na música contida no material didático e um enunciado claro e conciso, que direciona o estudante para o que precisa ser feito. Pelo viés dos estudantes, uma observação a ser feita é a de que o professor precisa ensiná-los a consultar o Skell, mais precisamente a opção *Similar Words*, o que também representa uma tarefa descomplicada.

No entanto, apesar de ser uma tarefa de simples execução, destaca-se que ela demanda do estudante o papel ativo e investigativo durante a sua realização. Ao observarem e interpretarem os resultados das buscas, os estudantes precisam levantar hipóteses e chegar a conclusões, pelos exemplos, dos significados das palavras. Esse papel certamente é o que se preza no âmbito da Aprendizagem Movida por Dados.

Ademais, é crucial destacar que a proposta envolve contextualização, pois parte da canção para explorar determinados vocábulos. Ela também estabelece um propósito claro para a pesquisa no *corpus*, que é o de compreender se os verbos possuem sentido relacionado a movimento ou não e ainda de encontrar sinônimos dessas palavras.

Por fim, no que tange às críticas à Aprendizagem Movida por Dados, essa proposta supera algumas delas: afinal, ela é voltada às crianças e não a estudantes do contexto universitário e/ou com alto nível de proficiência mais alto. Tal fato demonstra, portanto, as possíveis aplicações da abordagem a realidades distintas.

5.1.2 Atividade com foco em *noun phrases* a partir da leitura de texto

A segunda proposta foi planejada para o contexto de ensino de língua inglesa no Ensino Médio. A atividade objetiva implementar o material didático intitulado *Inglês – Série Nove Ensino Médio*⁴⁹. Ela envolve a utilização da ferramenta Skell como forma de complementar a própria atividade fornecida pelo livro didático, cujo foco é encontrar algumas *noun phrases* presentes no texto contido no material.

⁴⁹ Marques, A. *Inglês – Série Novo Ensino Médio. Volume único*. 7 ed. São Paulo: Ática, 2008.

Em vista disso, os primeiros passos sugestionados são os referentes à leitura e à compreensão textual do artigo existente no livro didático, conforme demonstrado na sequência.

Figura 36. Texto *Are girls smarter than boys?*



Fonte: Inglês – Série Novo Ensino Médio.

Realizada a interpretação do texto, a atividade subsequente do material foca em algumas *noun phrases* contidas nesse artigo. Desse modo, nessa tarefa os

estudantes precisam reler o texto para encontrar as combinações apontadas no exercício, como direciona a tarefa do material didático.

Figura 37. Proposta de identificação das *noun phrases* no texto

Noun Phrases

Encontre no texto os adjetivos (ou substantivos com função de adjetivo) que completam as locuções a seguir:

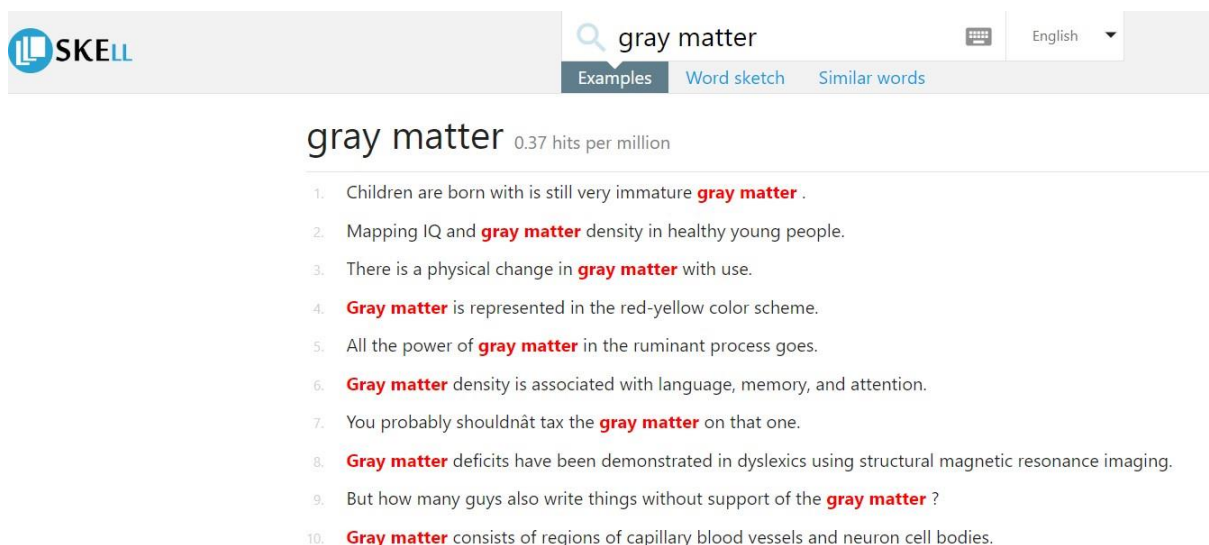
1 ▲ matter gray	massa cinzenta
2 ▲ tests intelligence	testes de inteligência
3 ▲ cells brain	células do cérebro
4 ▲ times modern	tempos modernos
5 ▲ news scientific	notícia científica
6 ▲ size head	tamanho da cabeça
7 ▲ abilities language	habilidades de linguagem
8 ▲ size body	tamanho do corpo
9 ▲ tasks cognitive	tarefas cognitivas
10 ▲ matter white	substância branca
11 ▲ skills spatial	habilidades espaciais
12 ▲ size brain	tamanho do cérebro

Inset photo: Three scientists in a laboratory setting, one holding a camera or microscope.

Fonte: Inglês – Série Novo Ensino Médio.

Após identificarem as *noun phrases* no texto, a professora sugere uma atividade que implemente o material didático e que envolve a utilização da ferramenta Skell. Dessa maneira, a proposta é a de que os estudantes investiguem essas *noun phrases* levantadas a partir do texto e encontrem exemplos de uso delas por meio da opção *Example*. Isso com o intuito de confirmarem o entendimento do sentido dessas combinações. Para tanto, os alunos precisam, então, realizar o procedimento de busca como o evidenciado a seguir.

Figura 38. Procedimento de busca de *gray matter* no Skell



The screenshot shows the Skell search interface. At the top, there is a search bar with the text 'gray matter' and a magnifying glass icon. To the left of the search bar is the Skell logo. Below the search bar, there are three tabs: 'Examples', 'Word sketch', and 'Similar words'. The 'Examples' tab is selected. Below the tabs, the search results are displayed. The first result is 'gray matter' with a score of '0.37 hits per million'. Below this, there is a list of 10 example sentences, each with the phrase 'gray matter' highlighted in red. The sentences are numbered 1 through 10.

gray matter 0.37 hits per million

1. Children are born with is still very immature **gray matter** .
2. Mapping IQ and **gray matter** density in healthy young people.
3. There is a physical change in **gray matter** with use.
4. **Gray matter** is represented in the red-yellow color scheme.
5. All the power of **gray matter** in the ruminant process goes.
6. **Gray matter** density is associated with language, memory, and attention.
7. You probably shouldnât tax the **gray matter** on that one.
8. **Gray matter** deficits have been demonstrated in dyslexics using structural magnetic resonance imaging.
9. But how many guys also write things without support of the **gray matter** ?
10. **Gray matter** consists of regions of capillary blood vessels and neuron cell bodies.

Fonte: captura de tela da ferramenta Skell

Com base nos resultados obtidos, os estudantes podem analisar os exemplos e confirmar o sentido da *noun phrase* utilizada no texto, no caso *gray matter*, que se refere à massa cinzenta (cérebro). No entanto, além de perceber o sentido, a proposta elaborada pela cursista também envolve a busca por outros modificadores que podem acompanhar os termos listados no material didático (vide última coluna da atividade).

Quadro 7. Atividade proposta para pesquisa de *noun phrases* no Skell

Noun phrase	Outros exemplos confirmam o sentido?	Outros modificadores que acompanham
Gray matter		
Intelligence tests		
Brain cells		
Modern times		
Scientific news		
Head size		
Language abilities		
Body size		
Cognitive tasks		
White matter		
Spatial skills		
Brain size		

Fonte: pesquisador a partir da proposta da cursista

Já para realizar essa segunda etapa os alunos têm que utilizar a opção *Word Sketch*, consultando, por exemplo, a palavra *matter* e observando os resultados obtidos, tal como a seguir.

Figura 39. Resultados obtidos para modificadores de *matter*, por meio do *Work Sketch*

modifiers of matter		
1.	<u>subject</u>	the subject matter
2.	<u>organic</u>	organic matter
3.	<u>particulate</u>	particulate matter
4.	<u>dark</u>	dark matter
5.	<u>practical</u>	As a practical matter ,
6.	<u>such</u>	such matters
7.	<u>simple</u>	a simple matter of
8.	<u>criminal</u>	criminal matters
9.	<u>related</u>	related matters .
10.	<u>legal</u>	legal matters
11.	<u>serious</u>	a serious matter
12.	<u>gray</u>	gray matter
13.	<u>fecal</u>	fecal matter
14.	<u>important</u>	important matters
15.	<u>plant</u>	plant matter

Fonte: captura de tela da ferramenta Skell

Como se pode perceber, assim como a atividade anterior, essa proposta também envolveu a adaptação de um material didático utilizado pelo professor e refuta algumas das críticas à Aprendizagem Movida por Dados, pois foi pensada para o contexto educacional do Ensino Médio, possivelmente para estudantes com nível de proficiência baixo. Outro ponto relevante em relação à superação das críticas à DDL é que a atividade didática não demandou do professor um processo de elaboração complexo, uma vez que o que se fez foi adaptar e implementar o material didático que já faz parte da rotina de trabalho do professor.

Desse modo, além de se mostrar prática, funcional e de simples planejamento, essa proposta possibilita que o aluno se posicione como um pesquisador ao consultar o Skell tanto no momento de confirmar o sentido quanto no de encontrar outras possibilidades de combinações. Ademais, ela envolveu uma contextualização, por meio da realização da leitura do texto, além do levantamento do conhecimento prévio dos estudantes no exercício de encontrar as *noun phrases* do texto e da preparação de como esses itens linguísticos se relacionam previamente à consulta no *corpus*. Esses são procedimentos importantes e que precisam perpassar as atividades e/ou lições que recorrem à Aprendizagem Movida por Dados.

5.1.3 Atividade com foco em ampliar os conhecimentos lexicais que acompanham a palavra *fashion*

A terceira atividade trazida neste capítulo foi delineada por uma cursista que trabalha na rede estadual de ensino de São Paulo. Por esse motivo, o material didático por ela utilizado foi a apostila oferecida pela própria rede na época, chamada Currículo em Ação. Com relação ao público, esta é uma atividade didática voltada aos estudantes do Ensino Médio.

Novamente, trata-se de uma atividade elaborada com o intuito de implementar um material didático já existente. A proposta também parte de uma leitura e compreensão textual, como se verifica a seguir.

Figura 40. Atividade de leitura com o tema *fashion*

FASHION

Throughout history, fashion has greatly influenced the “fabric” of societies all over the world. What people wear often characterizes who they are and what they do for a living. As Mark Twain once wrote, “Clothes make the man. Naked people have little or no influence on society.”

The fashion industry is a global industry, where fashion designers, manufacturers, merchandisers, and retailers from all over the world collaborate to design, manufacture, and sell clothing, shoes, and accessories. The industry is characterized by short product life cycles, erratic consumer demand, an abundance of product variety, and complex supply chains.

Source: <https://www.bls.gov/spotlight/2012/fashion/home.htm>. Accessed on October 2, 2020.

c) Search in a dictionary for the words you circled in the text.

d) Answer the questions on your notebook.

1. What is fashion to you?
2. Do you agree to Mark Twain's quote present in the text? Why?
3. Who dictates fashion?

Fonte: Currículo em Ação – Linguagens e suas tecnologias – 1ª série do Ensino Médio⁵⁰

Nota-se que, em um primeiro momento, a proposta é a de que os estudantes leiam o texto, cuja temática é a moda, e o compreenda. Na sequência, a própria apostila sugere que os estudantes façam uma espécie de nuvem de palavras (*word cloud*) com palavras que se relacionem ao vocábulo *fashion*, como estabelecido em seguida.

⁵⁰ Currículo em Ação – Linguagens e suas tecnologias – 1ª série do Ensino Médio, Caderno do Estudante, Volume 1, 2022. Disponível em: <https://efape.educacao.sp.gov.br/curriculopaulista/ensino-medio/materiais-de-apoio-2/>. Acesso em 13 de janeiro de 2024.

Figura 41. Atividade de nuvem de palavras com o termo *fashion*

e) Complete the graphic organizer with words you can relate to “fashion”.



Fonte: Currículo em Ação – Linguagens e suas tecnologias – 1ª série do Ensino Médio

Como se pode perceber, a atividade propõe a elaboração de uma nuvem de palavras que se relacionam com o termo *fashion*, mas não especifica nenhum recurso que assista o estudante a desempenhar tal tarefa. Dessa forma, a professora estabelece que os alunos acessem a plataforma Skell e utilizem a opção *Word Sketch* para encontrar verbos, modificadores e quaisquer outros termos que acompanhem a palavra *fashion*. Assim sendo, presume-se que os estudantes realizem um procedimento de busca como o demonstrado na sequência.

Figura 42. Pesquisa por *fashion* no Skell

The screenshot shows the Skell dictionary website. At the top, there is a search bar with the word 'fashion' entered. Below the search bar, there are tabs for 'Examples', 'Word sketch', and 'Similar words'. The 'Word sketch' tab is selected. Below the tabs, the word 'fashion' is displayed with a dropdown menu showing 'noun' and a checkmark. To the right of the word, there is a link 'Show context'. Below this, there are three columns of results:

verbs with fashion as subject	verbs with fashion as object	adjectives with fashion
1. model fashion modelling	1. relax in a relaxed fashion	1. accessory
2. dictate dictated by fashion	2. coordinate in a coordinated fashion .	2. fickle
3. change fashions changed	3. humiliate in humiliating fashion .	3. chic
4. influence influenced by fashion	4. mock in a mocking fashion	4. savvy
5. inspire inspired by the fashion	5. showcase showcased fashions	5. forward
6. mind	6. alternate in an alternating fashion	6. similar in a fashion similar to the
7. crowd	7. understate	7. conscious fashion conscious and
8. brand	8. dictate fashion dictated	8. galore
9. obsess	9. love love fashion	9. groundbreaking
10. design fashion designing	10. co-ordinate	10. trendy

Fonte: captura de tela da ferramenta Skell

Conforme os resultados alcançados, verifica-se que palavras como *model*, *influence*, *brand*, *design*, *trendy*, *accessory*, dentre outras, são algumas das que estabelecem relação com *fashion*.

Logo, considera-se que essa proposta da utilização da ferramenta Skell também representou uma forma simples de implementar uma atividade didática já disponível em um material didático já existente – o Currículo em Ação. Dessa maneira, esse é mais um exemplo de como a Aprendizagem Movidada por Dados pode ser introduzida sem muitos esforços no cotidiano escolar, inclusive no caso do ensino público. Vale ressaltar ainda que o contexto educacional envolvido não é o do Ensino Superior, mas o do Ensino Médio e com alunos com nível de proficiência baixo, o que mais uma vez demonstra a versatilidade da Aprendizagem Movidada por Dados.

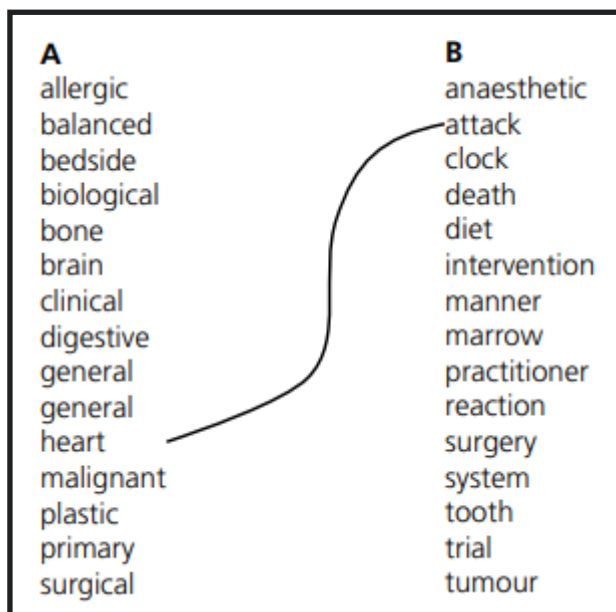
Ademais, é pertinente pontuar que, assim como nas propostas anteriores, há a apresentação do tema e a contextualização, que são feitas por meio do texto e da atividade com foco na interpretação. Há também o momento de levantamento do conhecimento prévio no instante em que os alunos precisam preencher as possíveis combinações com *fashion*, além do estabelecimento de um propósito claro para a realização da pesquisa, que é o de encontrar combinações que vão além daquelas que eles já conhecem ou puderam encontrar no texto.

5.1.4 Atividade com foco em vocabulários específicos da área da Medicina

A quarta proposta, também criada por uma cursista, foi planejada para o contexto do Ensino Superior, mais precisamente para estudantes de língua inglesa do curso de Medicina. A atividade se pauta em um material didático específico para essa área médica, focando em vocabulários e combinações de palavras desse segmento.

A atividade inicial proposta é a do próprio material didático, intitulado *Check Your English Vocabulary for Medicine: All You Need to Improve Your Vocabulary*⁵¹. Ela demanda dos estudantes que eles encontrem combinações de duas palavras tipicamente usadas na medicina.

Figura 43. Atividade de combinações de duas palavras da área médica



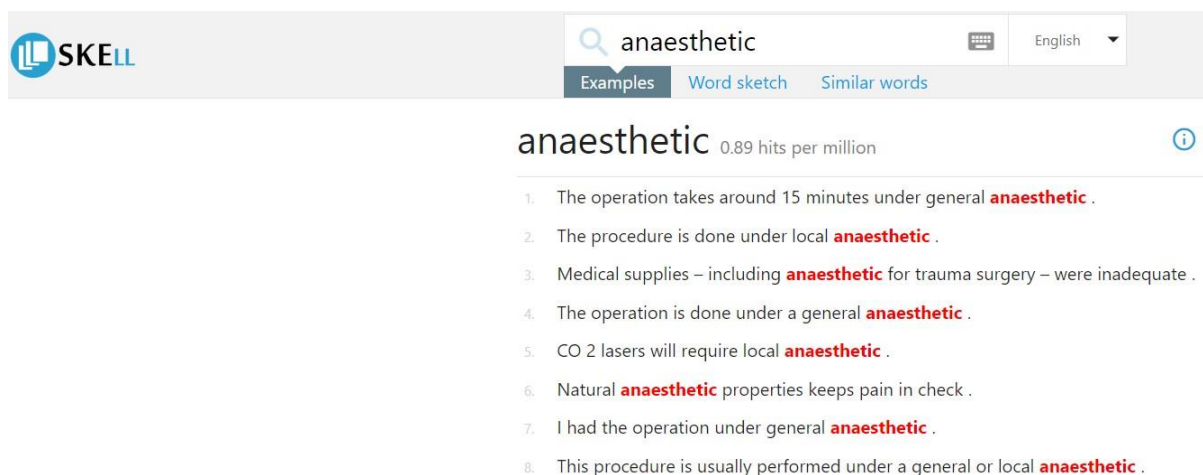
Fonte: Wyatt, 2006

Vê-se que o intuito é o de que os estudantes liguem a coluna A com a B, formando as combinações de duas palavras, com base em seus conhecimentos prévios. Já para a correção desse exercício, a cursista propõe que os estudantes chequem se suas hipóteses estão corretas ou não por meio da consulta no Skell. Ao fazerem isso, eles assumem um papel central na atividade, ao invés de apenas esperarem o gabarito das respostas pelo professor no momento da correção. Para

⁵¹ Wyatt, R. *Check Your English Vocabulary for Medicine: All You Need to Improve Your Vocabulary*. A&C Black, 2006.

isso, os graduandos em medicina precisam realizar buscas como a que se mostra na sequência.

Figura 44. Consulta de anaesthetic no Skell



The screenshot shows the Skell dictionary interface. At the top, there is a search bar with the word 'anaesthetic' entered. To the left of the search bar is the Skell logo, and to the right is a language dropdown menu set to 'English'. Below the search bar are three tabs: 'Examples', 'Word sketch', and 'Similar words'. The 'Examples' tab is selected, displaying a list of eight sentences. The word 'anaesthetic' is highlighted in red in each sentence. The search results also show '0.89 hits per million' and an information icon.

anaesthetic 0.89 hits per million

1. The operation takes around 15 minutes under general **anaesthetic** .
2. The procedure is done under local **anaesthetic** .
3. Medical supplies – including **anaesthetic** for trauma surgery – were inadequate .
4. The operation is done under a general **anaesthetic** .
5. CO 2 lasers will require local **anaesthetic** .
6. Natural **anaesthetic** properties keeps pain in check .
7. I had the operation under general **anaesthetic** .
8. This procedure is usually performed under a general or local **anaesthetic** .

Fonte: captura de tela da ferramenta Skell

É notório que as linhas de concordância 1, 4, 7 e 8 trazem a combinação *general anaesthetic*, o que pode evidenciar aos alunos que essa seja uma combinação usual na medicina.

Em seguida, o material didático sugere que os estudantes correlacionem as combinações formadas com seus respectivos sentidos, conforme demonstrado a seguir.

Figura 45. Exercício de correlacionar as combinações aos respectivos sentidos

1. A condition in which the heart has a reduced blood supply because one of the arteries becomes blocked by a blood clot, causing myocardial ischemia and myocardial infarction

heart attack

2. A substance given to make someone lose consciousness so that a major surgical operation can be carried out

3. Soft tissue in cancellous bone.

4. The treatment of disease or other condition by surgery.

5. Any one of the first twenty teeth which develop in children between about six months and two-and-a-half years of age and are replaced by the permanent teeth at around the age of six.

6. Surgery to repair damaged or malformed parts of the body.

7. A condition in which the nerves in the brain stem have died, and the person can be certified as dead, although the heart may not have stopped beating.

8. The way in which a doctor behaves towards a patient, especially a patient who is in bed.

9. An effect produced by a substance to which a person has an allergy, such as sneezing or a skin rash.

10. A trial carried out in a medical laboratory on a person or on tissue from a person.

11. A tumour which is cancerous and can grow again or spread into other parts of the body, even if removed surgically.

12. A doctor who provides first-line medical care for all types of illness to people who live locally, refers them to hospital if necessary and encourages health promotion.

13. The rhythm of daily activities and bodily processes such as eating, defecating or sleeping, frequently controlled by hormones, which repeats every twenty-four hours.

14. The set of organs such as the stomach, liver and pancreas which are associated with the digestion of food.

15. A diet that provides all the nutrients needed in the correct proportions.

Fonte: Wyatt, 2006

Para essa tarefa, a cursista também sugere a consulta feita pela ferramenta Skell (vide figura 43), mas, dessa vez, com o intuito de comparar se o sentido trazido pelo material didático condiz com outros exemplos trazidos nas linhas de concordância.

Em um terceiro momento, como forma de ampliação vocabular, a professora sugere um exercício em que os estudantes precisam listar outras combinações possíveis com os termos estudados. Veja.

Quadro 8. Exercício de formar combinações

Termo	Outras combinações
Anaesthetic	
Attack	
Clock	
Death	
Diet	
Intervention	
Manner	
Marrow	
Practitioner	
Reaction	
Surgery	
System	
Tooth	
Trial	
Tumour	

Fonte: pesquisador a partir da proposta da cursista

Para a realização dessa atividade em que se deve listar outras combinações possíveis para os termos estudados previamente, projeta-se novamente a utilização do Skell, utilizando-se a opção *Word Sketch*, para que se obtenham as expressões de duas palavras. Dessa forma, espera-se que buscas como a apresentada a seguir sejam feitas pelos estudantes.

Figura 46. Consulta de *anaesthetic* no Skell por meio da opção *Word Sketch*

The screenshot shows the Skell Word Sketch interface for the word "anaesthetic". The search bar at the top contains "anaesthetic" and the language is set to "English". Below the search bar, there are tabs for "Examples", "Word sketch", and "Similar words". The "Word sketch" tab is selected, showing a list of words related to "anaesthetic".

The results are organized into four columns:

- verbs with anaesthetic as subject**: numb, paralyse, inject, extract, induce, wear, fear, block, predict, combine.
- verbs with anaesthetic as object**: palpate, inject (anaesthetic is injected), administer (anaesthetic being administered), inhale, rub, spray, squeeze, picture, smell, infiltrate.
- modifiers of anaesthetic**: epidural, dissociative, inhalational, inhalation, longacting, lobal, longacting, short-acting, long-acting, lipophilic.
- nouns modified by anaesthetic**: M.O.A, xylocaine, benzocaine, Novocaine, lignocaine, laparotomy, phenobarbital, chloroform, anticonvulsant, adrenaline.

Fonte: captura de tela do Skell

A partir dos resultados gerados, os estudantes podem concluir, por exemplo, que *epidural anaesthetic*, *dissociative anaesthetic* e *lobal anaesthetic* são outras combinações possíveis. Essas conclusões podem ser feitas ao se observar o quadro que apresenta os modificadores de *anaesthetic*.

Em conclusão, a proposta de ensino de combinações típicas da área médica para graduandos do curso de Medicina visa à implementação de exercícios já elaborados em um material didático. Mais uma vez, as ferramentas e os pressupostos da Aprendizagem Moviada por Dados puderam implementar e oferecer novas aplicações didáticas. Destaca-se também que, como nos outros casos, essa proposta requer o acesso à internet para a consulta à ferramenta Skell e demanda pouco esforço do professor no tocante à preparação e à elaboração da aula. Esses podem ser vistos como pontos favoráveis a utilização da DDL.

Além disso, os exercícios propostos visaram a levantar os conhecimentos prévios dos estudantes. Assim, eles precisam tentar estabelecer as combinações de palavras para, na sequência, verificarem se de fato pensaram em combinações possíveis de serem feitas na língua. Dessa forma, a atividade incentiva a formulação de hipóteses sobre a língua para, então, despertar nos estudantes o interesse em descobrir se as combinações que eles pensaram são realmente possíveis na língua. Isso faz com que os exercícios sejam contextualizados e promove um propósito sobre o que deve ser feito no momento da realização da consulta ao *corpus*. Esses são pontos estruturantes da proposta didática que fazem a diferença e enriquecem a Aprendizagem Moviada por Dados.

5.1.5 Atividade com foco nos quantificadores

A quinta atividade, elaborada por outra cursista, enfoca a primeira série do Ensino Médio. Ela foi esboçada com vistas ao ensino dos quantificadores *many*, *much*, *few* e *little* e se baseia em uma das estratégias do ensino híbrido – a sala de aula invertida⁵². Por isso, o objetivo é fazer com que o aluno pesquise, levante hipóteses e tire conclusões sobre o uso dos quantificadores na língua inglesa, de forma prévia ao trabalho com o livro didático e as explicações do professor no decorrer das aulas.

⁵² Segundo Moran (2019, p. 29), “a aprendizagem invertida transfere para o digital uma parte do que era explicado em aula pelo professor.”

Isto posto, a concepção é a de que, com o auxílio da ferramenta Linggle, os estudantes busquem por exemplos de uso dos quantificadores, de modo a preencherem o quadro exposto a seguir.

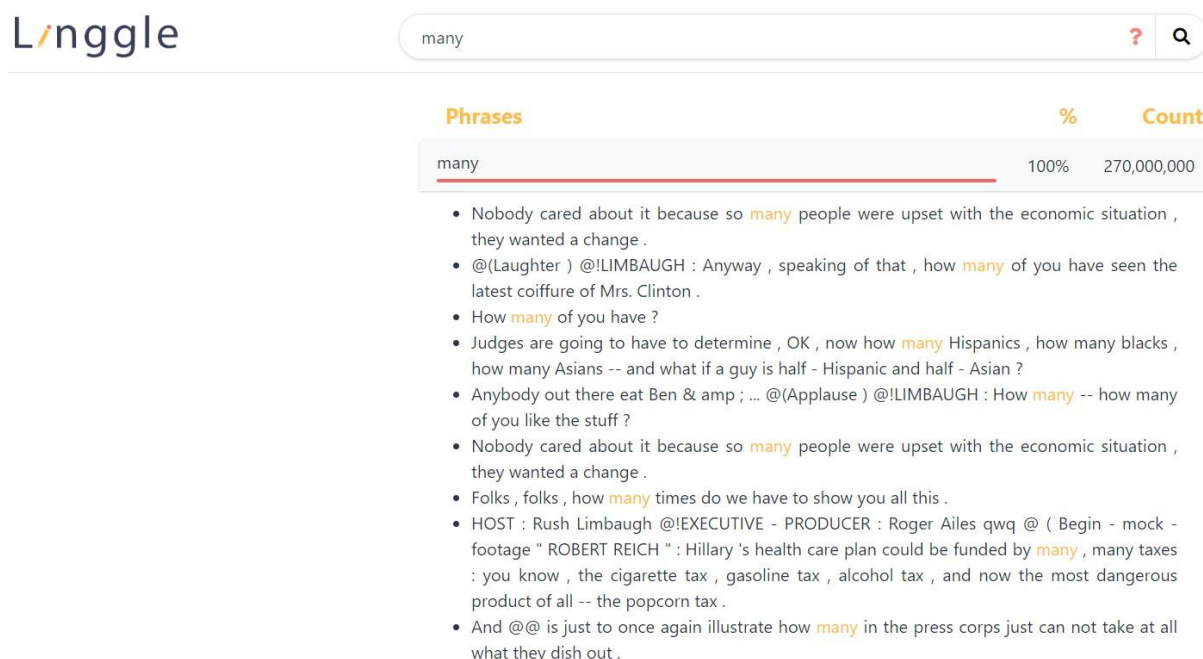
Quadro 9. Exercício dos quantificadores

Termo	Exemplos	
MANY	1. 2. 3. 4.	5. 6. 7. 8.
MUCH	1. 2. 3. 4.	5. 6. 7. 8.
FEW	1. 2. 3. 4.	5. 6. 7. 8.
LITTLE	1. 2. 3. 4.	5. 6. 7. 8.

Fonte: pesquisador a partir da proposta da cursista

A fim de preencher a atividade com exemplos reais de uso dos quantificadores, pressupõe-se que os estudantes façam pesquisas simples no Linggle, inserindo apenas a palavra na barra de busca, como demonstra a imagem em seguida.

Figura 47. Busca pelo quantificador *many* no Linggle



Fonte: captura de tela do Linggle

No entanto, os estudantes podem ainda optar por uma consulta mais direcionada, inserindo a palavra a ser consultada somada ao comando *n.*, que representa os substantivos. Assim, a busca filtrará por substantivos que acompanham o termo pesquisado. Veja:

Figura 48. Busca pelo quantificador *many* e os substantivos que o acompanham

The screenshot shows the Linggle search interface. At the top left is the Linggle logo. To its right is a search bar containing the text 'many n.' with a search icon (magnifying glass) and a question mark icon. Below the search bar is a table with three columns: 'Phrases', '%', and 'Count'. The table lists various phrases starting with 'many' followed by a noun, along with their percentage of occurrence and their count.

Phrases	%	Count
many people	14.5%	11,000,000
many other	13.8%	10,000,000
many years	11.6%	8,500,000
many times	8%	5,900,000
many others	4.7%	3,500,000
many ways	4.7%	3,400,000
many cases	4.4%	3,200,000
many things	3.6%	2,600,000
many countries	1.9%	1,400,000
many areas	1.6%	1,200,000
many hours	1.3%	980,000

Fonte: captura de tela do Linggle

Dessa maneira, evidencia-se mais claramente as palavras que sucedem o quantificador. Entretanto, ambas as formas de busca, uma que é mais simples (demonstrada pela figura 46) e outra que filtra os substantivos (ilustrada pela figura 47), fornecem informações a respeito da natureza dos substantivos que acompanham o quantificador *many*. Sendo assim, compreende-se que, com as informações coletadas, eles conseguem perceber que *many* é seguido por substantivos contáveis.

Essas análises sobre o funcionamento dos quantificadores com base nos usos obtidos pelo Linggle são ainda mais bem direcionadas quando os estudantes se deparam com o próximo exercício elaborado e proposto pela cursista em parceria com o pesquisador. Veja:

Figura 49. Atividade de análise do funcionamento dos quantificadores a partir do uso

A partir da análise dos resultados encontrados, completem as regras abaixo.

- a) Usamos MANY para substantivos: () contáveis () incontáveis.
- b) Usamos MUCH para substantivos: () contáveis () incontáveis.
- c) Usamos FEW para substantivos: () contáveis () incontáveis.
- d) Usamos LITTLE para substantivos: () contáveis () incontáveis.

Fonte: pesquisador a partir da proposta da cursista

Contempla-se que a própria atividade, ao oferecer sempre duas alternativas de resposta (substantivos contáveis ou incontáveis), direciona o estudante para a informação que ele precisa checar quando estiver analisando o exemplo de uso dos quantificadores no Linggle. Esta é, portanto, uma atividade didática elaborada para ser realizada pelos estudantes antes da aula em que o professor elucidará essas questões gramaticais. Por isso a importância de que ela facilite e oriente o olhar dos estudantes em relação àquilo que eles precisam compreender.

Ao se analisar esta proposta, algumas considerações se fazem necessárias. Por se tratar de uma metodologia ativa (Sala de Aula Invertida), compreende-se que o nível de complexidade se torna um pouco maior, pensando na zona próxima ao desenvolvimento vygotskyano (Paiva, 2014). Por esse motivo, a mediação em relação à compreensão do funcionamento dos itens linguísticos (quantificadores) se fez de modo a trazer alternativas as quais o estudante assinalará com base na observação das linhas de concordância. Essa foi uma estratégia sugerida pelo pesquisador e que o cursista também achou relevante.

Assim, o que se nota é que a proposta, apesar de não envolver tanto uma apresentação e nem um levantamento do conhecimento prévio dos estudantes, estabelece propósitos claros para a pesquisa no *corpus*. Em um primeiro momento, deve-se encontrar combinações com os quantificadores para, a partir desses exemplos, formular hipóteses sobre como esses itens linguísticos funcionam. Feito isso, há a atividade que estabelece esses funcionamentos em forma de perguntas com duas opções de alternativas para que eles respondam conforme o que haviam observado nas linhas de concordância.

Dessa forma, conclui-se que esta proposta enfatizou bastante o papel ativo do estudante no processo da aprendizagem, de modo a atuarem como pesquisadores/detetives da língua. Isso tudo elaborado de forma descomplicada para que o estudante consiga analisar as ocorrências da língua e compreender os padrões linguísticos. Esses fatores contribuem, portanto, para a defesa da aplicabilidade da Aprendizagem Movida por Dados como uma abordagem que contribui para a aprendizagem ativa.

5.1.6 Atividade com foco nas preposições

A sexta proposta de atividade didática foi desenvolvida por uma cursista que também se baseou na utilização de um material didático, mas voltado para o contexto de ensino de língua inglesa em escolas de idiomas. O tema selecionado foi as preposições que obrigatoriamente sucedem determinados verbos, as *dependent prepositions*.

A atividade do material didático fornece uma lista de preposições e pede que os estudantes as associem com algumas frases. Dessa forma, o intuito do exercício é que se estabeleça a relação entre o verbo e a preposição que o sucede. A imagem a seguir ilustra a atividade.

Figura 50. Atividade sobre as *dependent prepositions*

2 VERBS + PREPOSITIONS

a Complete the **Prepositions** column with a word from the list.

about at for in of on to with



- 1 I arrived New York City on Friday night.
- 2 I was very tired when I arrived the hotel.
- 3 I hate waiting people who are late.
- 4 A What are you going to do the weekend?
B I don't know. It depends the weather.
- 5 I'm sorry, but I really don't agree you.
- 6 I asked a chicken sandwich, but this is tuna!
- 7 Let's invite Debbie and Tim the party.
- 8 Who's going to pay the meal?
- 9 I need to speak Martin the meeting.
- 10 I don't spend much money food.
- 11 Are you going to write him soon?
- 12 Don't worry the exam. It isn't very hard.
- 13 She fell love a man she met on the Internet.
- 14 You're not listening! What are you thinking ?
- 15 A What do you think Shakira?
B I really like her. I think she's great.

Prepositions

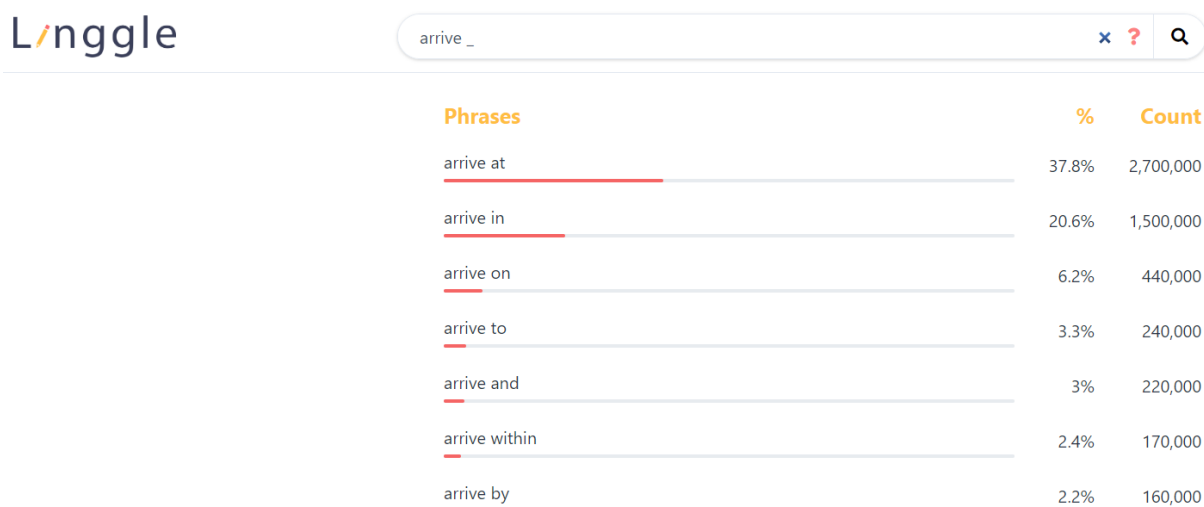
in _____

Fonte: *American English File 2*⁵³

Para a realização dessa atividade, ao invés de os estudantes completarem as preposições com base apenas no que eles acreditam estar correto ou consultando o professor, a cursista propõe a utilização da ferramenta Linggle. Para isso, recomenda-se que seja utilizado o verbo contido na frase seguido do comando _ (subtração), como retrata a imagem a seguir.

⁵³ Latham-Koenig, C. et al. *American English File 2: American English File 2 Student Book with Online Practice*. 2. ed. Oxford: Oxford University Press, 2013.

Figura 51. Exemplo de busca com o verbo *arrive*



The screenshot shows the Linggle search interface. At the top left is the Linggle logo. To its right is a search bar containing the text 'arrive _'. To the right of the search bar are three icons: a blue 'x', a red '?', and a magnifying glass. Below the search bar is a table with three columns: 'Phrases', '%', and 'Count'. The table lists eight phrases with their corresponding percentages and counts. Each phrase in the 'Phrases' column has a red horizontal bar of varying length to its left, representing its relative frequency.

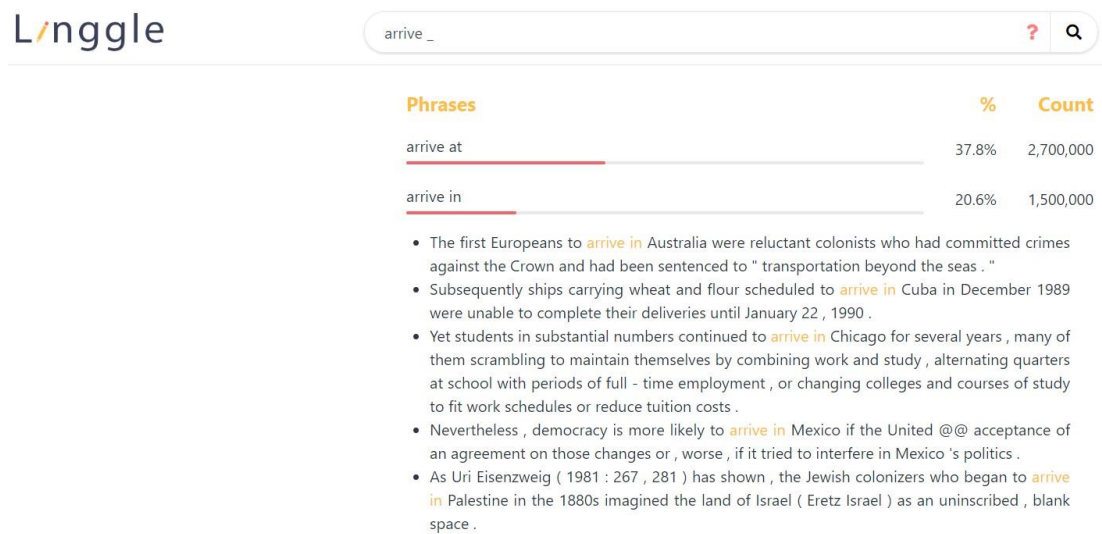
Phrases	%	Count
arrive at	37.8%	2,700,000
arrive in	20.6%	1,500,000
arrive on	6.2%	440,000
arrive to	3.3%	240,000
arrive and	3%	220,000
arrive within	2.4%	170,000
arrive by	2.2%	160,000

Fonte: captura de tela da ferramenta Linggle

Percebe-se que o procedimento de busca *arrive _* apresentou várias possibilidades preposições. No entanto, as combinações *arrive at* e *arrive in* foram as mais expressivas. Consequentemente, sugere-se, nesse caso, que o professor, no decorrer das buscas feitas pelos estudantes, oriente-os a clicar em cada um desses resultados para que, a partir dos exemplos, eles possam levantar hipóteses de quando se usa uma preposição e quando se usa outra.

Por exemplo, ao se clicar no resultado *arrive in*, compreende-se facilmente que essa opção de preposição é utilizada ao se referir a países, como se pode notar na figura a seguir:

Figura 52. Exemplos de *arrive in*



Fonte: captura de tela da ferramenta Linggle

Posto isso, conclui-se que a proposta de atividade didática apresentada representa uma forma de descentralizar a figura do professor como fonte de consulta dos estudantes, pois oferece a ferramenta Linggle como meio de pesquisa para a realização do exercício contido no material didático. No entanto, ressalta-se a importância de o professor atuar no papel de mediador desses procedimentos de busca, orientando os aprendizes a clicarem também nos exemplos e a terem um olhar investigativo em relação a quando se usa cada preposição. Essa mediação do docente é um dos pontos que as teorias acerca da Aprendizagem Movida por Dados frisam bastante.

Somado a isso, faz-se pertinente considerar que o exercício também se propõe a levantar os conhecimentos prévios dos estudantes, pois eles precisam tentar completar as preposições antes de as consultarem por meio da ferramenta. Nota-se, ainda, que essa tentativa inicial de eles resolverem o exercício para, na sequência, realizarem a busca contribui, inclusive, para o levantamento de hipóteses sobre a língua, desperta o interesse em se fazer a busca no Linggle e, ainda, estabelece um propósito claro para a investigação no *corpus*.

Logo, essa atividade consegue conceber o aluno como detetive da língua, de modo a evidenciar a aplicabilidade da Aprendizagem Movida por Dados a contextos para além do Ensino Superior, pois se trata do contexto de escolas de idiomas. Ademais, utiliza-se de recursos que contextualizam e dão propósito à pesquisa no *corpus*, sem mencionar que representa uma forma simples, no quesito preparação da

aula, de se implementar um material didático de língua inglesa já disponível no mercado.

5.1.7 Atividade com foco em ampliar os conhecimentos lexicais que acompanham a palavra *beauty*

A sétima atividade é semelhante à apresentada na seção 5.1.3. Logo, trata-se também de uma proposta para a primeira série do Ensino Médio da rede pública estadual de São Paulo e utiliza o material Currículo em Ação como base. Inicialmente, a tarefa do material envolve a leitura e a compreensão de um texto relacionado com o tema beleza. Considere a imagem a seguir.

Figura 53. Atividade de leitura com tema relacionado à beleza

b) Read the text. Circle the cognates, and look up for the words you don't know in a dictionary.

WHY DO WOMEN WANT TO BE BEAUTIFUL?

Beauty is only skin-deep, but the perceived absence of beauty may lead to damaging social bias. Compared to men, women may suffer more from social anxiety, prejudice, and inequality based on their appearance. To comprehend beauty-related socio-cultural phenomena, many studies have examined the pursuit of beauty related to body image, body perception, a body-related self-concept, and body satisfaction. These studies have clarified factors that impact and result in the pursuit of beauty. However, although numerous previous studies on these aspects of beauty have been conducted, there is still a considerable controversy about why women, in particular, are focused on with a beautiful appearance.

Source: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6075765/>. Accessed on July 15, 2020.

c) Answer the following questions on your notebook.

1. According to the text, do women and men react the same way to beauty?
2. What does the absence of beauty may cause? Do you agree with that?
3. In your opinion, what is beauty?
4. Do you think the media affects the way people understand beauty? Why?

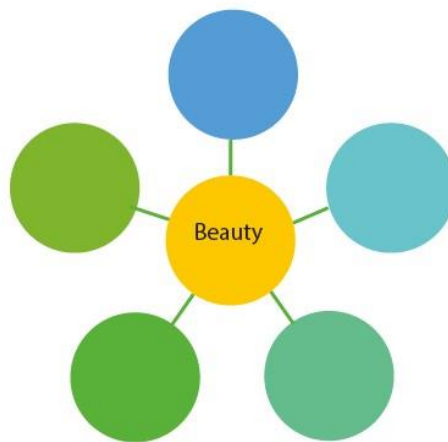
Fonte: Currículo em Ação – Linguagens e suas tecnologias – 1ª série do Ensino Médio⁵⁴.

Logo após a leitura e a compreensão do texto, o material aborda um exercício de elaboração de uma nuvem de palavras relacionadas ao substantivo *beauty*. Observe:

⁵⁴ Currículo em Ação – Linguagens e suas tecnologias – 1ª série do Ensino Médio, Caderno do Estudante, Volume 1, 2022. Disponível em: <https://efape.educacao.sp.gov.br/curriculopaulista/ensino-medio/materiais-de-apoio-2/>. Acesso em: 23 de janeiro de 2024.

Figura 54. Atividade de nuvem de palavras com o substantivo *beauty*

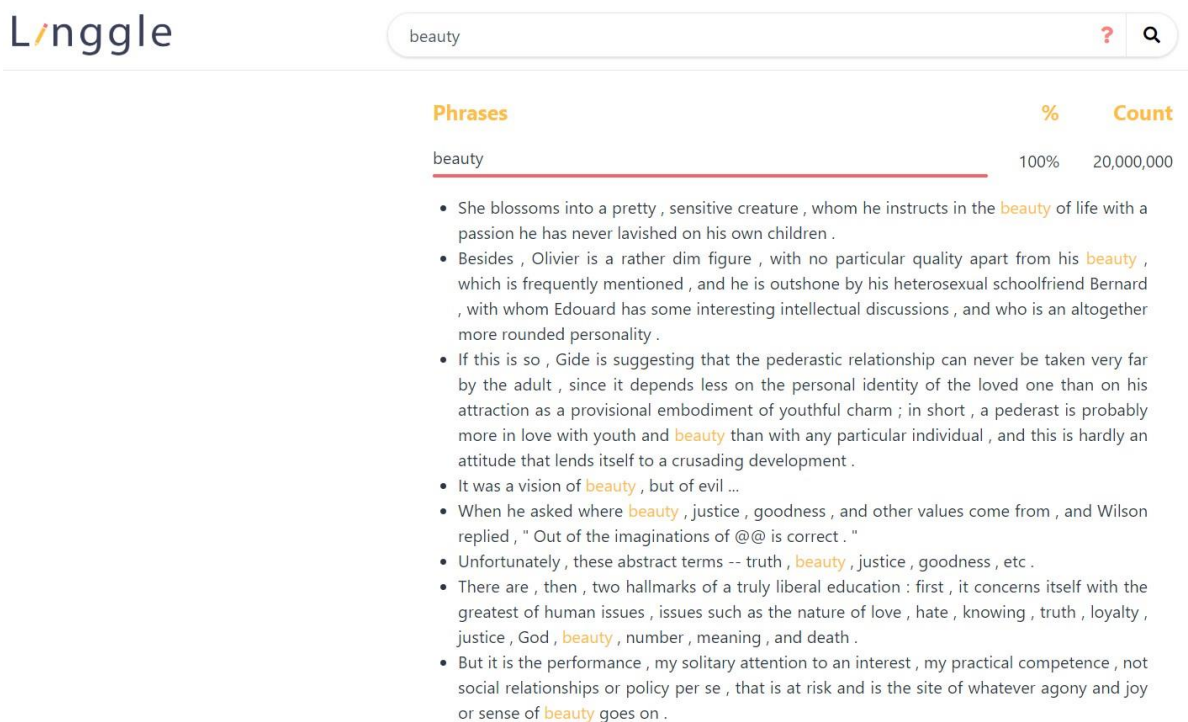
d) Complete the graphic organizer with words you can relate to “**beauty**”.



Fonte: Currículo em Ação – Linguagens e suas tecnologias – 1ª série do Ensino Médio

O comando do exercício direciona os estudantes a completarem a nuvem de palavras com vocabulários que se relacionem a *beauty*. Entende-se que isso seja feito com base nos conhecimentos prévios dos estudantes e no texto lido anteriormente. Entretanto, a proposta da cursista visa a ampliar as possibilidades didáticas desse exercício a partir da consulta à ferramenta Linggle. Para isso, os estudantes podem simplesmente pesquisar por *beauty* e analisar, com base nos exemplos obtidos, quais outras palavras se relacionam, de alguma maneira, com esse substantivo. Esse procedimento é exemplificado a seguir.

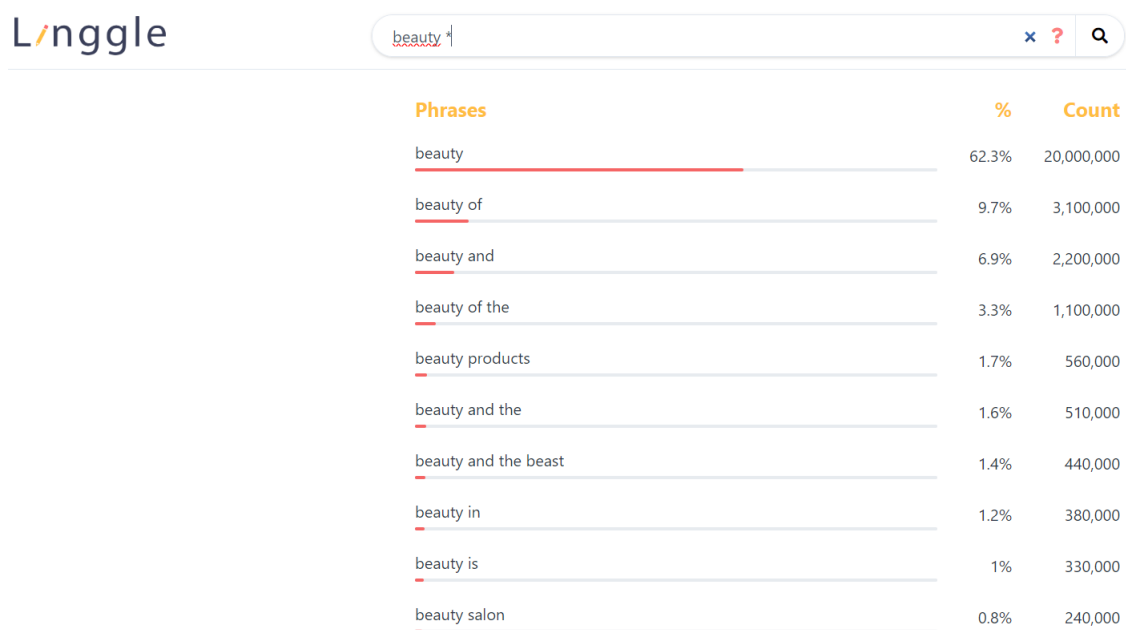
Figura 55. Exemplificação da busca por *beauty* no Linggle



Fonte: captura de tela da ferramenta Linggle

Somada a essa possibilidade, os estudantes dispõem de outras maneiras de buscas, utilizando, por exemplo, o asterisco, que demonstrará quais palavras acompanham o termo, tais como *products*, *salon* e assim por diante. Veja o exemplo:

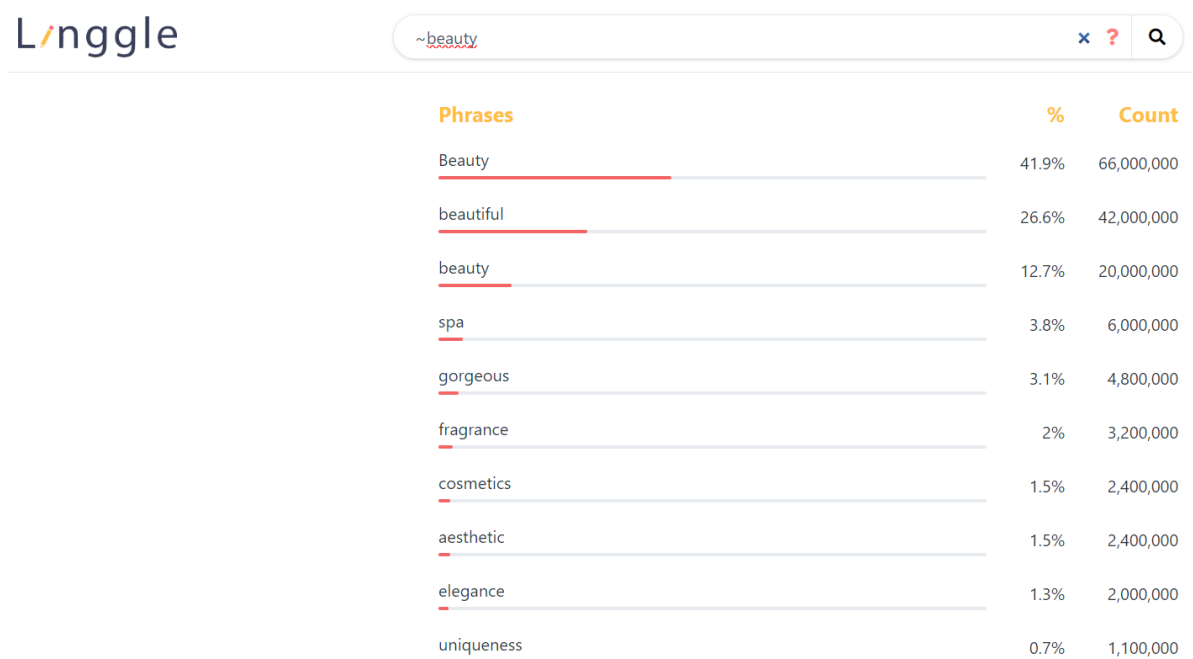
Figura 56. Exemplificação de busca com *beauty* *



Fonte: captura de tela da ferramenta Linggle

Além disso, pode-se ainda buscar por palavras sinônimas por meio do comando ~, o que resulta em opções como *gorgeous*, *cosmetics*, dentre outras. Verifique o exemplo trazido a seguir:

Figura 57. Exemplo de busca com o comando ~*beauty*



The screenshot shows the Linggle search interface. At the top, the Linggle logo is on the left, and a search bar contains the query '~beauty'. To the right of the search bar are icons for clearing the search (x), help (?), and searching (Q). Below the search bar, a table displays the results. The table has three columns: 'Phrases', '%', and 'Count'. The results are as follows:

Phrases	%	Count
Beauty	41.9%	66,000,000
beautiful	26.6%	42,000,000
beauty	12.7%	20,000,000
spa	3.8%	6,000,000
gorgeous	3.1%	4,800,000
fragrance	2%	3,200,000
cosmetics	1.5%	2,400,000
aesthetic	1.5%	2,400,000
elegance	1.3%	2,000,000
uniqueness	0.7%	1,100,000

Fonte: captura de tela da ferramenta Linggle

Percebe-se que esses procedimentos de busca auxiliam os estudantes a encontrarem os vocabulários que se relacionam com o substantivo *beauty*. Já sob um olhar de análise da proposta, destaca-se que ela também parte da apresentação de um tema e da contextualização à vida dos alunos por meio dos exercícios que enfatizam a leitura e a compreensão textual.

A atividade didática também levanta o conhecimento prévio dos estudantes quando sugere que eles pensem em possíveis combinações com o substantivo. Vê-se que esse estágio, além de considerar aquilo que os alunos já sabem, desperta-lhes o interesse em realizar a busca, no sentido de que ela tem o propósito de encontrar outras formas combinatórias com *beauty*, para além daquelas pensadas por eles em um primeiro instante. Essa sequenciação se mostra favorável à aprendizagem ativa e aos documentos norteadores da Educação Básica e desempenham um papel relevante para a organização da Aprendizagem Movida por Dados.

Enfim, presume-se que a proposta da cursista visa a implementar o material didático já disponível na rede estadual de São Paulo. Por esse motivo, é possível afirmar que os esforços demandados do professor para a elaboração dessa proposta não apresentaram muita complexidade. Isso mostra, mais uma vez, um aspecto em defesa da utilização da Aprendizagem Movida por Dados em contextos que vão além do Ensino Superior.

5.1.8 Atividade com foco em ampliar os conhecimentos lexicais que acompanham a palavra *voting*

Assim como na atividade anterior, a oitava proposta, também desenvolvida por um cursista, direciona-se ao Ensino Médio. O material didático não foi disponibilizado pelo idealizador. Entende-se que, nesse caso, o objetivo é o de complementar uma tarefa de leitura, cujo tema é a votação nos Estados Unidos, presente nesse material. Por esse motivo, após os exercícios de leitura e compreensão textual, o cursista propõe algumas tarefas aos estudantes como forma de ampliarem os conhecimentos lexicais que acompanham a palavra *voting*.

O primeiro exercício tem o foco em identificar possíveis sinônimos para *voting*. Analise:

Figura 58. Atividade com foco nos sinônimos de *voting*

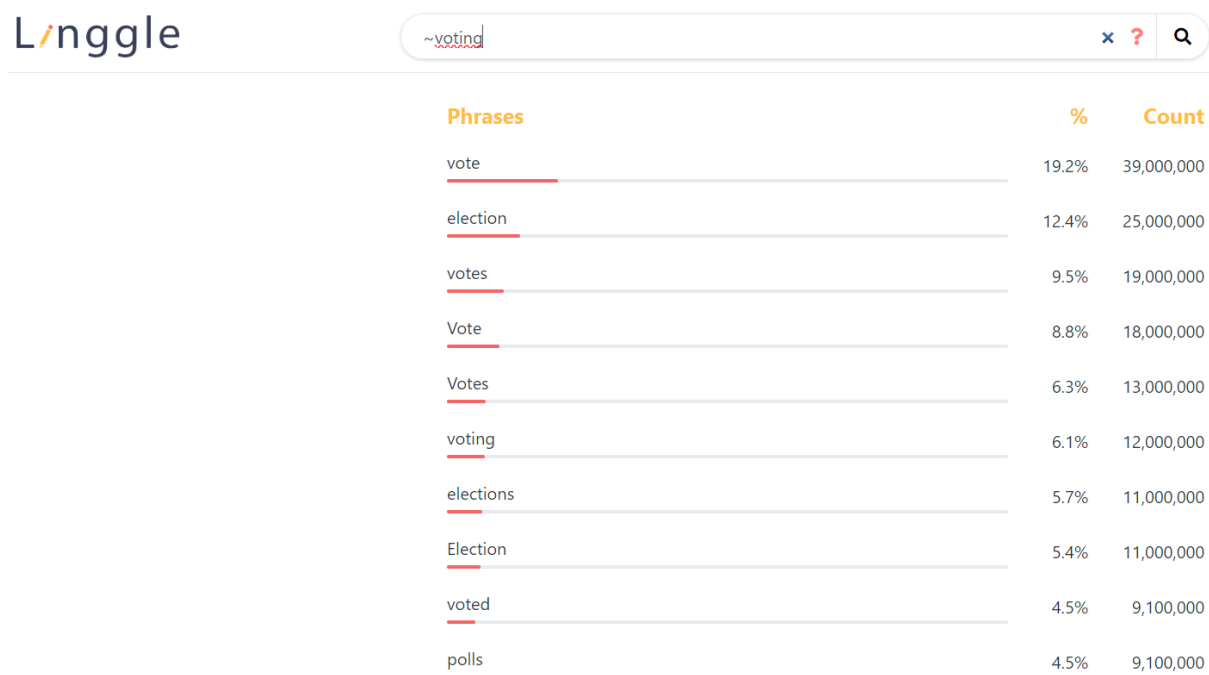
A. Using Linggle, find 5 different SYNONYMS for **VOTING (don't forget to use ~ before Voting)**

1.
2.
3.
4.
5.

Fonte: pesquisador a partir da proposta do cursista

Como o próprio enunciado da questão sugere, é preciso que os estudantes acessem a ferramenta Linggle e utilizem ~ na busca para encontrarem os sinônimos do vocábulo *voting*. Logo, pressupõe-se que eles realizem a seguinte pesquisa.

Figura 59. Exemplo da busca ~voting na ferramenta Linggle



The screenshot shows the Linggle search interface. The search bar contains '~voting'. Below the search bar, there is a table with three columns: 'Phrases', '%', and 'Count'. The table lists various phrases related to voting, along with their percentage of usage and their count.

Phrases	%	Count
vote	19.2%	39,000,000
election	12.4%	25,000,000
votes	9.5%	19,000,000
Vote	8.8%	18,000,000
Votes	6.3%	13,000,000
voting	6.1%	12,000,000
elections	5.7%	11,000,000
Election	5.4%	11,000,000
voted	4.5%	9,100,000
polls	4.5%	9,100,000

Fonte: captura de tela da ferramenta Linggle

Facilmente, os estudantes conseguem identificar palavras como *election*, *polls*, dentre outras. Ao descer a barra de rolagem para visualizar outros resultados alcançados, eles seguramente encontrarão termos semelhantes e conseguirão identificar ao menos cinco opções válidas, conforme demandado pelo enunciado do exercício.

Na sequência, é desenvolvido um exercício que visa ao processo de formação de palavras (prefixação) com o termo *vote*. Para isso, projeta-se a tarefa a seguir.

Figura 60. Atividade com foco nas palavras derivadas de *vote*

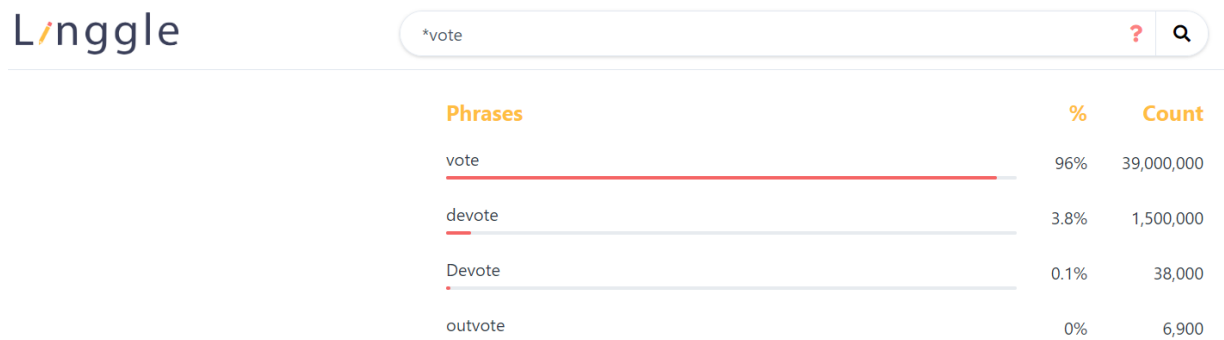
B. Use Linggle to find two derived words for *vote*. To identify prefixes that go along with *vote*, search for **vote*. Then analyze what these words mean by clicking on them and seeing examples in real sentences.

Word	Meaning
1.	
2.	

Fonte: pesquisador a partir da proposta do cursista

Ao analisar o enunciado da questão, compreende-se que o intuito é o de que os estudantes levistem palavras derivadas de *vote* ao buscarem por prefixos que acompanham o termo. Dessa maneira, os estudantes precisam realizar a busca da seguinte forma para responderem à primeira parte do exercício:

Figura 61. Resultado da busca **vote*

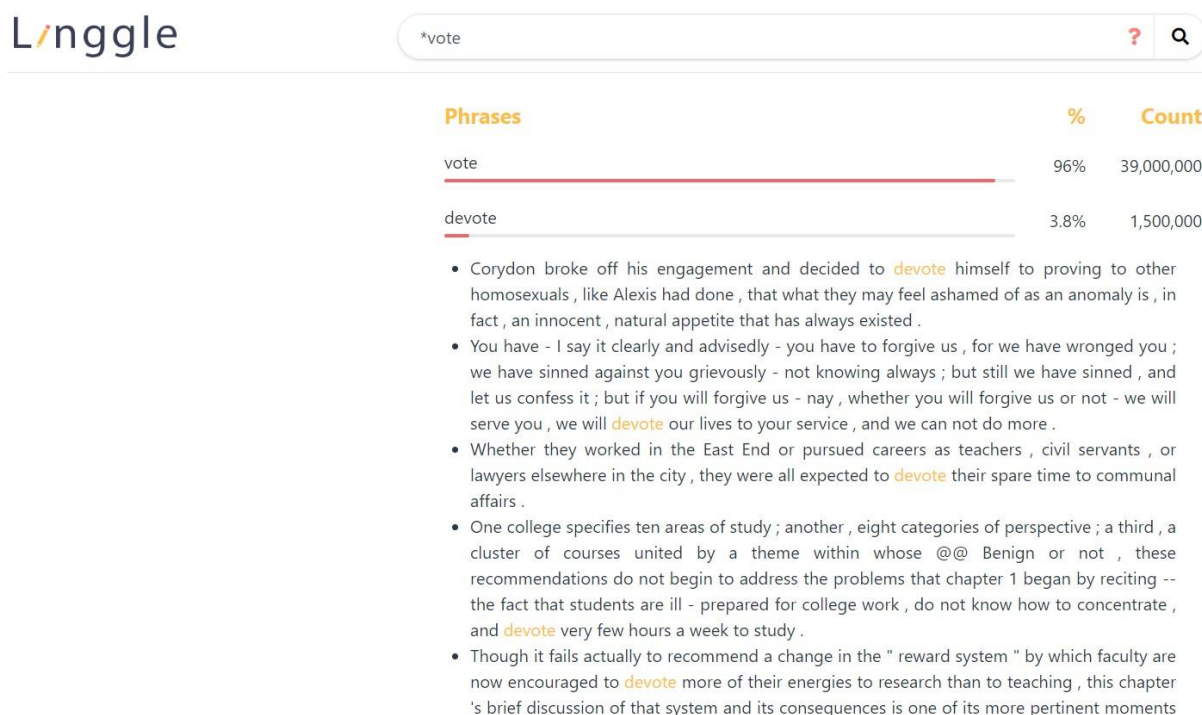


Phrases	%	Count
vote	96%	39,000,000
devote	3.8%	1,500,000
Devote	0.1%	38,000
outvote	0%	6,900

Fonte: captura de tela da ferramenta Linggle

Como se percebe na imagem, a pesquisa resulta em *devote* e *outvote*. Entretanto, para responder a segunda parte da atividade, que se refere aos sentidos dessas palavras, os estudantes deverão clicar nelas para que vejam exemplos reais de uso em frases e, conseqüentemente, levistem as hipóteses em relação aos significados. Esse procedimento é demonstrado a seguir.

Figura 62. Exemplos de uso da palavra *devote*



Fonte: captura de tela da ferramenta Linggle

Desse modo, entende-se que, ao observar os exemplos, os alunos presumirão os significados dessas palavras e perceberão, inclusive, que elas consistem em verbos. Logo, essa proposta de se complementar a atividade de leitura do material didático com foco tanto em explorar palavras sinônimas quanto nos processos de formação de palavras coloca os estudantes como detetives da língua a partir do acesso direto à ferramenta Linggle. Além disso, pontua-se que essa pesquisa é feita de forma contextualizada, pois envolve a leitura em um momento anterior à busca. Há ainda um propósito claro estabelecido para que os estudantes façam a pesquisa no *corpus*. Por esse motivo, afirma-se que esta atividade didática, voltada para o Ensino Médio, representa mais uma maneira de contribuir com a implementação de materiais didáticos já disponíveis. Isso sem mencionar a simplicidade de preparação e desenvolvimento da proposta, o que demonstra o quanto a Aprendizagem Movida por Dados pode ser utilizada de forma descomplicada.

5.1.9 Atividade com a música *I want to break free*, da banda *Queen*

A nona atividade, proposta por uma cursista, considera o contexto do Ensino Médio, mais especificamente, a primeira série. Trata-se de uma atividade com música que visa à exploração das classes gramáticas das palavras contidas na letra da canção.

Em primeiro lugar, a professora sugere que os estudantes escutem a música e tomem nota das palavras que eles conseguirem ouvir na letra, conforme demonstrado na imagem a seguir.

Figura 63. Atividade de *listening* com a música *I want to break free*

Atividade 1. Ouça a música da banda Queen e anote, abaixo, as palavras da letra que você conseguir ouvir/compreender.

_____	_____
_____	_____
_____	_____
_____	_____
_____	_____

Fonte: pesquisador a partir da proposta da cursista

Após esse primeiro contato com a música, propõe-se que os estudantes identifiquem, com o auxílio do Versatext, por meio de função *Wordcloud* especificamente, as palavras mais relevantes na letra da música e comparem se as opções dadas na nuvem de palavras coincidem com os vocabulários que eles ouviram na música. Esse exercício é ilustrado na sequência.

Figura 64. Atividade com foco em gerar nuvem de palavras com a letra da música

Atividade 2. No Versatext, insira a letra da música e utilize o botão Wordcloud para gerar as palavras mais relevantes. Alguma dessas opções coincidem com as palavras que você anotou na atividade 1? Quais?

_____	_____
_____	_____
_____	_____

Fonte: pesquisador a partir da proposta da cursista

Dessa maneira, para compararem o que eles ouvirem com as palavras indicadas como mais relevantes na nuvem de palavras feita no Versatext, pontua-se que os aprendizes, inicialmente, insiram a letra da música na ferramenta e, sequencialmente, cliquem no botão *Wordcloud*. Essa ação gerará a nuvem de palavras, tal como expõe a figura a seguir.

Figura 65. Nuvem de palavras gerada a partir da canção *I want to break free*



Fonte: captura de tela do Versatext

Por fim, ainda com base nessa nuvem de palavras gerada no Versatext pelos aprendizes, recomenda-se que os estudantes observem as informações oferecidas pela ferramenta para responderem ao exercício seguinte.

Figura 66. Atividade com foco nas classes gramaticais

Atividade 3. Observe a nuvem de palavras para preencher a tabela.

Exemplos de substantivos	Exemplos de verbos	Exemplos de adjetivos	Exemplos de advérbios.

Agora, redija algumas frases em inglês com essas palavras.

Fonte: pesquisador a partir da proposta da cursista

Percebe-se que a atividade demanda que os estudantes verifiquem as cores das palavras contidas na nuvem de palavras e se atentem às informações contidas na legenda. Desse modo, eles poderão compreender que as palavras em vermelho representam os substantivos, as azuis os verbos, as rosas os adjetivos e as verdes os advérbios. Feito isso, a tarefa final determina que, utilizando desses exemplos encontrados na letra da música, os estudantes produzam frases.

Em conclusão, esta proposta, por um lado, difere-se das demais, pois não parte de um material didático. Dessa maneira, isso evidencia o quanto o leque de possibilidades de criação de atividades didáticas a partir das ferramentas da LC é vasto. Por outro lado, o que ela tem em comum com as outras propostas é que ela

também foi projetada para ser aplicada de modo direto (*hands-on*) na sala de aula. Verifica-se também a praticidade e exequibilidade desta atividade que, assim como as demais, não exige muito esforço e/ou tempo do professor em relação à preparação.

Além disso, considera-se relevante o fato de a proposta ser contextualizada, mesmo que sem o apoio de um material didático. Isso porque ela se pauta na música mencionada e no que os estudantes conseguem compreender da letra dessa canção. Somado a isso, há também um propósito definido para a utilização da ferramenta, que é a comparação das palavras que os estudantes compreenderam com aquelas listadas como sendo as mais relevantes pelo Versatext. Essa definição do que deve ser feito desperta, portanto, o interesse nos alunos em contrastarem se o que eles puderam entender são, de fato, palavras relevantes na letra da música.

Outro aspecto de grande relevância é o fato de a proposta dar uma oportunidade para que os estudantes utilizem os resultados obtidos na busca de forma livre, produzindo frases na língua inglesa. Isso se mostra em conformidade com a lições apresentadas no capítulo 3 e com a aprendizagem ativa prevista na BNCC e no Currículo Paulista, além de contemplar a fase 4 (de prática, uso livre e fechamento) dos estágios da aula propostos por Rozenfeld e Viana (2020).

5.1.10 Atividade com foco em estratégias de leitura

A décima atividade apresentada foi elaborada por uma cursista com foco no desenvolvimento de estratégias de leitura no contexto de Inglês Para Fins Acadêmicos. O objetivo da atividade recai sobre as palavras-chaves. Para isso, a proposta se baseia em um artigo científico intitulado *Climate change mitigation opportunities based on carbon footprint estimates of dietary patterns in Peru*⁵⁵.

Com o intuito de engajar os estudantes e de levantar os conhecimentos prévios, propõe-se que se apresente a seguinte nuvem de palavras, gerada pelo Versatext, a partir do resumo (*abstract*) do artigo.

⁵⁵ Vázquez-Rowe, I. *et al.* Climate change mitigation opportunities based on carbon footprint estimates of dietary patterns in Peru. *PLoS ONE*, 2017. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0188182>. Acesso em: 27 janeiro 2024.

Figura 68. Atividade de busca das palavras desconhecidas no Linggle

Exercise: Go to Linggle, write the word you don't know on the search bar and click on it. Write down two sentences that helped you understand the meaning of the word.

Fonte: pesquisador a partir da proposta da cursista

Conforme estabelece o comando do próprio exercício, os estudantes pesquisarão no Linggle as palavras das quais eles não conhecem o sentido. Assim sendo, presume-se que eles façam um procedimento de busca semelhante ao apresentado na sequência.

Figura 69. Exemplo de pesquisa de *intake* no Linggle

The screenshot shows the Linggle website interface. At the top, the Linggle logo is on the left, and a search bar with the word 'intake' is in the center. To the right of the search bar are icons for help (?) and search (Q). Below the search bar, there is a table with three columns: 'Phrases', '%', and 'Count'. The first row of the table shows 'intake' with a percentage of 100% and a count of 5,500,000. Below the table, there is a list of example sentences where the word 'intake' is highlighted in orange.

Phrases	%	Count
intake	100%	5,500,000

- Body weight , food **intake** , and water intake were recorded daily during both this experimental period and the habituation period .
- For those animals @@ differences were noted among the three groups with respect to increase or decrease in body weight , foot **intake** , or water intake .
- All participants had completed three or more years of high school at the time of testing , resided in New York City , and were described on **intake** forms as Caucasian .
- But with the food comes a message of proscription and suppression of what he experiences as a very arousing sensation accompanying his food **intake** .
- The " good " breast - mother who provides the food can not be at the same time this vile polluting entity , so the mytho - empirical projection solves this dilemma for the simplistic mouth - ego by attributing it to the seduction of the breast - mother by an extraneous badness that lured her to **intake** the polluting sexual desire , which she in turn transmits to the mouth - ego , while at the same time suppressing it .
- Prior to the first **intake** , the sponsor described the students ' English language proficiency as high intermediate to advanced .
- The right to food is the requirement that everyone has enough caloric **intake** to maintain healthy growth of mind and body .

Fonte: captura de tela do Linggle

Pelos exemplos obtidos, julga-se que os estudantes levantarão hipóteses acerca do significado da palavra *intake*, que é ingestão ou consumo. A recomendação é a de que eles façam esse tipo de busca para consultarem as palavras desconhecidas. Percebe-se que essas atividades são prévias ao ato da leitura e da compreensão do texto em si.

Após realizados esses exercícios de levantamento do conhecimento prévio, propõe-se a leitura do resumo do artigo, acompanhada de questões que enfatizam a compreensão textual. Feito isso, sugestiona-se uma atividade que envolve, novamente, a análise da nuvem de palavras mostrada aos estudantes no momento inicial da proposta. Observe o exercício a seguir.

Figura 70. Atividade com foco nas palavras-chaves em contexto
Complete the sentences below with one word from the wordcloud.

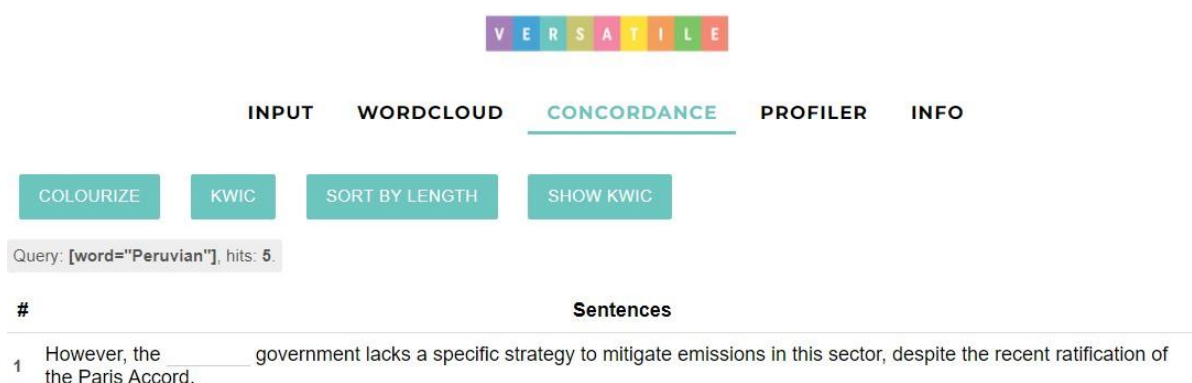
- a) However, the _____ government lacks a specific strategy to mitigate emissions in this sector, despite the recent ratification of the Paris Accord.
- b) Previous studies have delved into the nature of dietary _____, showing that GHG reductions can be achieved in diets if certain foods are consumed rather than other, more GHG intensive products.
- c) In view of this, the main objective of this study is to analyze the environmental impacts of a set of 47 Peruvian food diet profiles, including _____ and socioeconomic scenarios.
- d) Previous studies have delved into the nature of dietary patterns, showing that _____ reductions can be achieved in diets if certain foods are consumed rather than other, more GHG intensive products.
- e) For instance, vegetarian and low-meat diets have proved to be less carbon intensive than _____ that are based on ruminant meat.

Fonte: proposta da cursista

Verifica-se que essa atividade demanda do aluno observar a nuvem de palavras apresentada no início da aula para identificar qual palavra completa cada uma das

lacunas. A elaboração desse exercício envolveu o *Concordance*, no Versatext, que disponibiliza as palavras-chaves no contexto, do inglês *Keyword in Context* (KWIC). No entanto, ao invés de utilizar o botão KWIC, a cursista optou pelo botão *Sentence*, pois, assim, o contexto em volta da palavra-chave envolve toda a frase (período) e não apenas um recorte dela. Por se tratar de uma atividade didática, a opção do período todo (botão *Sentence*) ao invés do recorte feito pelo *Keyword in Context* se mostra mais pedagógico. Veja a ilustração desse procedimento a seguir.

Figura 71. Procedimento de elaboração da atividade com foco nas palavras-chaves no Versatext



Fonte: captura de tela do Versatext

Resumidamente, esta proposta de atividade didática enfatizou, por meio da nuvem de palavras geradas com o Versatext, a apresentação do tema do texto e o levantamento dos conhecimentos prévios dos estudantes. Esse estágio visou contribuir com a leitura e a interpretação contextualizadas do resumo do artigo para que, na sequência, enfocasse nas palavras-chaves presentes no resumo do artigo. Esse exercício final demandou dos alunos observarem novamente a nuvem de palavras para completarem as lacunas com as palavras-chaves correspondentes. Desse modo, estabelece-se um propósito também para a observação da nuvem de palavras. Ademais, destaca-se que a elaboração dessa atividade final ocorreu por meio da utilização do *Concordance*, no Versatext (vide figura 70).

Em conclusão, esta atividade didática promoveu procedimentos de acesso às informações linguísticas de forma indireta (*hand-off*), ao oferecer a nuvem de palavras e ao proporcionar o exercício com foco nas palavras-chaves. No entanto, também envolveu o acesso direto aos dados da língua por meio da busca das palavras desconhecidas no Linggle (*hands-on*). Dessa forma, essa atividade se mostrou um

pouco mais complexa no sentido de preparação por parte do professor. Entretanto, com relação à aplicação aos alunos, ela não é complexa. Vê-se, portanto, que, além de adaptar propostas já existentes em materiais didáticos, as ferramentas da LC possibilitam ainda que os professores planejem atividades completas e/ou lições.

Por fim, ressalta-se que esta subseção apresentou uma série de propostas de atividades didáticas envolvendo os princípios da Aprendizagem Movidada por Dados. A maioria delas se basearam em atividades contidas em materiais didáticos, porém as duas últimas propostas foram diferentes. Elas foram desenvolvidas inteiramente pelos cursistas com o apoio do pesquisador. Nesse sentido, pontua-se que as formas de implementação da Aprendizagem Movidada por Dados vão desde a adaptação de materiais didáticos já existentes até a elaboração completa de atividades didáticas e/ou lições.

Com relação ao acesso ao *corpus*, a maior parte das propostas aqui trazidas valeram-se do acesso direto aos dados linguísticos, ou seja, da opção *hands-on*. Isso, apesar de aparentemente ser um procedimento mais complexo de ser levado à sala de aula, provou-se ser, com base nas propostas aqui apresentadas, uma opção simples de ser executada pelos alunos que requer pouca preparação do professor em relação à preparação e/ou elaboração das aulas. Logo, esses são fatores importantes de serem pontuados em defesa da efetiva aplicação da Aprendizagem Movidada por Dados nas aulas de Língua Inglesa por meio das ferramentas da LC, mais precisamente do Skell, do Linggle e do Versatext.

5.2 Princípios pedagógicos para a elaboração de lições que recorrem à Aprendizagem Movidada por Dados

Concluída a exposição das dez propostas de atividades didáticas produzidas pelos professores que participaram do curso de formação continuada em serviço e com o suporte do pesquisador, prossegue-se para um levantamento dos pontos em comum dessas atividades. Isso com o intuito de se estabelecer, com base no que os cursistas propuseram e em todo o referencial teórico recorrido nesta tese, uma série de princípios pedagógicos que guiem a elaboração de lições voltadas à Aprendizagem Movidada por Dados.

Dessa forma, ao se analisar os aspectos mais relevantes que perpassaram as dez propostas descritas na seção anterior, conclui-se que muitas delas apresentaram o tema e o contextualizaram. Além disso, elas também levantaram os conhecimentos prévios dos estudantes, preparando-os e os motivando, em certa medida, para a realização da consulta no *corpus*, por meio da ferramenta da LC. Outro ponto crucial é que todas estabeleceram um propósito claro em relação a que os estudantes precisavam buscar ao acessarem as linhas de concordância. Tais procedimentos, portanto, conceberam o estudante como sujeito ativo da aprendizagem, de modo a atuarem como pesquisadores/detetives da língua.

Nesse sentido, julga-se que esse passo a passo esteve em consonância com as duas lições de exemplificação trazidas no capítulo 3 desta tese. Isso porque os cursistas tiveram acesso a elas no decorrer da formação e as utilizaram como inspiração para elaborarem suas próprias atividades didáticas. Ademais, nota-se também uma sequenciação que, de certo modo, atrela-se às fases da aula propostas por Rozenfeld e Viana (2020) e que também atendem a aspectos importantes das metodologias ativas previstas nos documentos orientadores da Educação Básica, a Base Nacional Comum Curricular e o Currículo Paulista.

Entretanto, diferentemente das lições exemplificadas no capítulo 3, as dez propostas dos cursistas, por configurarem mais atividades didáticas e não a elaboração de uma lição toda, não abordaram, na maioria das vezes, certos critérios prescritos tanto nas fases da aula de língua (Rozenfeld; Viana, 2020) quanto nos documentos da Educação Básica.

O primeiro exemplo disso é que as propostas dos professores não contemplaram, por exemplo, a fase 1, relacionada ao planejamento. Porém, por mais que esse estágio não tenha aparecido explicitamente nas propostas dos cursistas, é notável que eles planejaram suas atividades didáticas e as discutiram com o pesquisador, no intuito de que atendessem às demandas das conceitualizações acerca da Aprendizagem Movidada por Dados.

Outro exemplo disso é que apenas uma proposta promoveu atividade voltada ao uso livre e prático dos itens linguísticos encontrados no decorrer da análise das linhas de concordância como forma de fechamento das atividades. No entanto, é válido pontuar que ambas as lições elaboradas pelo pesquisador contam com atividades de encerramento e utilização das descobertas feitas pela pesquisa no

corpus. Essa fase de fechamento e uso prático, como sugere Rozenfeld e Viana (2020), é vista como importante parte do processo de aprendizagem ativa e movida por dados, pois oferece ao estudante a chance de aplicar, de modo personalizado e contextualizado, o que foi aprendido no decorrer da lição.

Com base, portanto, em todas essas colocações, ou melhor, no que se pôde observar nas duas lições produzidas pelo pesquisador e nas dez atividades didáticas propostas pelos cursistas (apoiados pelo pesquisador), além do que se notou em relação à aprendizagem ativa prevista na BNCC e no Currículo Paulista (norteadores da Educação Básica), à Aprendizagem Movida por Dados e às fases da aula de Rozenfeld e Viana (2020), propõem-se os princípios pedagógicos para que a abordagem da Aprendizagem Movida por Dados seja utilizada de forma eficaz na implementação de materiais didáticos ou na elaboração de lições e/ou atividades didáticas.

Desse modo, compreende-se que as lições e/ou atividades que se utilizam da Aprendizagem Movida por dados precisam:

1. Contar com um planejamento prévio em que se considere o objetivo da aula e da consulta ao *corpus*, assim como as ferramentas da LC a serem utilizadas.

Esse primeiro princípio diz respeito à relevância do planejamento prévio da lição e/ou do material didático e está relacionado ao que Rozenfeld e Viana (2020) apontam como a fase 1 da aula. Assim, é nesse estágio que se precisa considerar os objetivos da aprendizagem, a organização da(s) aula(s), a(s) ferramenta(s) da LC e os demais recursos didáticos que serão utilizados. Inclusive, essa é a etapa de pensar nas possíveis dificuldades que serão encontradas no decorrer do desenvolvimento da lição e/ou material didático.

2. Apresentar um tema que perpassasse por todos os estágios, subsidiando e dando suporte à pesquisa a ser feito pelos estudantes no *corpus*;

3. Contextualizar esse tema à realidade dos estudantes;

Com relação aos princípios 2 e 3, eles são especialmente relevantes, pois

estabelecem a importância de a lição e/ou material didático perpassar uma temática a ser desenvolvida nas atividades. Além disso, enfatizam a relevância de esse tema estar contextualizado à vida dos estudantes. Essa é uma preocupação das metodologias ativas que são reiteradas pela BNCC. Em certa medida, esses princípios 2 e 3 ainda estão no nível do planejamento da aula. Já os efetivos desdobramentos do tema e dessa preocupação com ele estar contextualizado aos estudantes podem ser vistos nos próximos princípios.

4. Promover interação entre os estudantes, correlacionando o tema ao objeto do conhecimento;

5. Levantar os conhecimentos prévios dos estudantes;

Com relação ao quarto e ao quinto princípio, eles se destacam, pois visam a promover a interação comunicativa entre os estudantes, o que é endossado pelas metodologias ativas e pelos documentos orientadores da Educação Básica (BNCC e o Currículo Paulista). Ademais, correlacionam essa interação com o tema a ser efetivamente trabalhado na(s) aula(s), além de já oferecer uma oportunidade de os estudantes resgatarem os conhecimentos prévios, tais como estruturas e/ou léxico comumente relacionados ao tema da aula. Um exemplo da efetiva aplicação desses princípios é a atividade 1 da lição *So many things in a day* (vide capítulo 3.1).

Atividade 1:



Em pequenos grupos, converse a respeito da rotina de cada um. O que há em comum entre vocês? O que há de diferente?

SEMELHANÇAS	DIFERENÇAS

Nota-se que o intuito dessa atividade é o de contextualizar o tema da lição (rotina), promover a interação entre os estudantes e o levantamento de

conhecimentos prévios, tais como os vocabulários relacionados à rotina, o tempo verbal presente e assim por diante.

6. Incentivar a formulação prévia de hipóteses sobre o funcionamento da língua;

7. Preparar os estudantes em relação aos conteúdos linguísticos que eles possivelmente precisarão acionar durante a consulta ao *corpus*;

Apresentado o contexto e dado a chance de interação e levantamento dos conhecimentos prévios, é relevante que, antes de consultar o *corpus* por meio da ferramenta da LC, os estudantes comecem a racionalizar e levantar hipóteses sobre o funcionamento da língua, em especial no que se volta aos conteúdos linguísticos a serem abordados na(s) aula(s). Um bom exemplo disso é a atividade 2 da lição *It depends on the situation* (vide seção 3.2).

Atividade 2:

Considere as frases do exercício anterior para responder as questões que se seguem.

a) A primeira frase traz um exemplo de uma preposição precedida de:

☐ um adjetivo; ☐ um substantivo; ☐ um verbo.

b) O que fica entre *provide* e a preposição *with*?

c) A segunda frase traz um exemplo de uma preposição precedida de:

☐ um adjetivo; ☐ um substantivo; ☐ um verbo.

d) Como é usado o verbo após a combinação *difficulty in*:

e) A terceira frase traz um exemplo de uma preposição precedida de:

☐ um adjetivo; ☐ um substantivo; ☐ um verbo.



Percebe-se que essa atividade já focaliza o conteúdo abordado na aula (as preposições) e dá a chance de o aluno fazer algumas hipóteses primárias com relação a como o uso das preposições acontece na língua. Esse é um passo inicial no sentido de despertar o interesse do estudante em compreender melhor como a língua funciona (no que se trata do objeto do conhecimento abordado). Com foco nisso é que

se estabelece o próximo princípio.

8. Despertar o interesse dos estudantes em realizar a pesquisa no *corpus*;

Ao se considerar a enumeração dos princípios pedagógicos, nota-se que as transições dos estágios da lição e/ou material didático acontecem de maneira sequencial e lógica. Assim, esse levantamento prévio e essa motivação de os alunos realizarem mais pesquisas sobre a língua se alinham às competências gerais 1 e 2 da BNNC, relacionadas ao conhecimento e ao pensamento científico. Além disso, fazem com que as premissas de aluno ser protagonista, conforme as metodologias ativas, e detetive da língua, como preza a Aprendizagem Movida por Dados, aconteçam com engajamento por parte dos estudantes.

9. Ensinar aos estudantes a utilização das ferramentas da LC usadas para a consulta ao *corpus*;

10. Posicionar o professor como mediador do processo de busca e de análise dos resultados obtidos (as linhas de concordância);

No que tange aos professores, uma vez que os estudantes se mostram engajados para a realização das pesquisas, esses profissionais precisam ensiná-los os procedimentos de buscas nas ferramentas da LC. Além disso, têm que agir como mediadores de todo esse processo. Afinal, durante a realização das pesquisas os estudantes se depararão com muitas informações e elas levantarão inúmeras dúvidas. Nesse sentido, o professor precisa auxiliá-los a filtrar as informações e a compreendê-las. Uma forma de se fazer isso é também lembrar os estudantes do que eles precisam encontrar com relação à língua e a que conclusão eles precisam chegar com relação a como ela funciona. Para isso o princípio a seguir se mostra bastante relevante.

11. Estabelecer um propósito claro que norteie a realização da pesquisa e a análise dos dados obtidos (as linhas de concordância);

12. Conceber o estudante como sujeito ativo no processo de aprendizagem, de

modo que o aluno seja pesquisador/detetive da língua;

13. Ter uma progressão linear em relação ao nível de complexidade, partindo, preferencialmente, do *hands-off* para o *hands-on*⁵⁶;

14. Proporcionar colaboração entre os estudantes e o professor durante o processo de especulação e generalização acerca dos padrões da língua observados nos resultados das buscas (as linhas de concordância);

Com relação à pesquisa e às conclusões que os estudantes vão fazer sobre como a língua funciona, é necessário destacar que esse processo deve ser ativo, sendo o aprendiz o responsável por levantar hipóteses, fazer investigações e descobrir os padrões da língua. Esses são pressupostos tanto da Aprendizagem Movida por Dados quanto dos documentos orientadores da Educação Básica. No entanto, esse processo pode ser mediado de modo a se considerar, em um primeiro momento, um acesso *hands-off* para, progressivamente, se se pautar no procedimento *hands-on*.

Além disso, é aconselhável que haja o trabalho colaborativo entre os estudantes, de modo que eles discutam e compartilhem suas hipóteses e suas descobertas. Essa colaboração está em conformidade com as metodologias ativas e a BNCC. Já com relação a figura do professor, ele é também muito importante nesse processo colaborativo, pois também auxiliará no levantamento dos padrões e do funcionamento da língua. Assim, após serem feitas as devidas descobertas sobre a língua – considerando os objetivos pretendidos para a lição e/ou material didático –, é de suma importância que os estudantes pratiquem o que foi aprendido. Para isso, estabelece-se o próximo princípio.

15. Oportunizar aos estudantes a utilização personalizada e contextualizada dos padrões linguísticos descobertos pela pesquisa em *corpus*.

Em conclusão, salienta-se que esses princípios foram elaborados a partir da

⁵⁶ Sugere-se essa progressão linear do nível de complexidade, partindo do *hands-off* para o *hands-on*, pelo fato de entender-se que a quantidade de informação a qual o estudante estará exposto no modelo *hands-off* ser menor do que no *hands-on*.

revisão teórica de teorias de ensino convencionais da Educação Básica e que se atrelam aos pressupostos da Aprendizagem Movidada por Dados, além da aplicação prática, que ocorreu na elaboração das lições, na formulação do curso de formação continuada em serviço aos professores e no apoio aos cursistas em relação ao desenvolvimento das atividades didáticas. Dessa maneira, evidencia-se que esses princípios são essenciais na implementação de materiais didáticos e na elaboração de lições e/ou atividades didáticas que recorram à Aprendizagem Movidada por Dados.

6 CONSIDERAÇÕES FINAIS E ENCAMINHAMENTOS

Este último capítulo retoma as perguntas de pesquisa mencionadas na introdução, como forma de articulá-las às discussões e às considerações finais. Na sequência, são apresentados os encaminhamentos para aqueles, pesquisadores e/ou professores, que tenham interesse em se aprofundar na utilização da abordagem da Aprendizagem Movidada por Dados na implementação de materiais didáticos ou na elaboração de lições e/ou atividades didáticas.

Sendo assim, relembra-se que a primeira pergunta de pesquisa é:

- Em que medida a Aprendizagem Movidada por Dados e os pressupostos pedagógicos abordados nos documentos orientadores da Educação Básica podem se complementar e contribuir para a disseminação dessa abordagem em contextos educacionais que vão além daqueles convencionalmente pesquisados?

Nesse sentido, é apropriado mencionar que, ao se correlacionar os documentos norteadores da Educação Básica, a Base Nacional Comum Curricular e o Currículo Paulista, à Aprendizagem Movidada por Dados, pode-se estabelecer uma série de pontos complementares. A BNCC, por exemplo, estabelece o desenvolvimento de competências gerais, sendo a 1 (voltada ao conhecimento), a 2 (relacionada ao pensamento científico, crítico e criativo), a 4 (concernente à comunicação) e a 5 (associada à cultura digital) as que mais se alinham à utilização das ferramentas computacionais para acessar o *corpus* e, conseqüentemente, analisar o funcionamento da língua.

Além disso, a BNCC toma por base as metodologias ativas, que, no geral, enfatizam:

- O *learn by doing* (aprender fazendo), que atrela a teoria à prática;
- A aprendizagem contextualizada à vida do aluno e às suas necessidades;
- O papel protagonista do estudante, sujeito ativo no processo da aprendizagem;
- A aprendizagem a partir de questionamentos, investigações e descobertas;

- A necessidade de as atividades serem interativas e colaborativas e considerarem os conhecimentos prévios dos estudantes.

Todas essas considerações, portanto, alinham-se às conceitualizações acerca da Aprendizagem Movida por Dados, o que favoreceu a defesa de correlacionar a DDL às teorias mais convencionais de ensino do contexto da Educação Básica. Isso contribuiu, inclusive, para que o escopo de aplicabilidade da Aprendizagem Movida por Dados se ampliasse ao Ensino Médio, aos cursos livres de idiomas e àqueles estudantes cujos níveis de proficiência ainda não são altos. Isso evidencia a possibilidade, portanto, de a DDL ser empregada em contextos que vão além daqueles convencionalmente considerados na maior parte das pesquisas.

Entretanto, não só os pressupostos teóricos presentes na BNCC podem complementar a DDL. É adequado salientar também o quanto a Aprendizagem Movida por Dados, assim como as ações desenvolvidas por este estudo, pode suplementar a Base Nacional Comum Curricular, em especial, no que se trata das críticas que são feitas a esse documento, tanto no sentido da necessidade de haver formações docentes (Dornelles; Irala, 2021; Fidalgo; Sprenger; Ramos, 2021), quanto no que diz respeito à explicitação do que exatamente se trata a perspectiva da análise linguística mencionada no documento (Dornelles; Irala, 2021). Nessa perspectiva, compreende-se que promover a formação continuada em serviço aos professores foi uma forma de superar a primeira crítica. Já o desenvolvimento das lições e das atividades didáticas amparadas pela Aprendizagem Movida por Dados, assim como o estabelecimento dos princípios pedagógicos, foi uma maneira de vencer a segunda. Essas colocações respondem, de certa forma, à segunda pergunta desta pesquisa:

- De quais formas a elaboração de um curso de formação continuada em serviço para professores pode promover a implementação de materiais didáticos e/ou elaboração de atividades didáticas que se utilizam da Aprendizagem Movida por Dados?

A elaboração desse curso promoveu um trabalho colaborativo entre os pesquisadores da área da LC, envolvendo a participação de estudiosos de três universidades (Universidade Estadual Paulista, Universidade Estadual de Goiás e a Universidade de Queensland), além de proporcionar que eles interagissem com

professores da Educação Básica. Tal colaboração contribuiu para uma superação à ponderação feita por Boulton (2024), no sentido de que a DDL ainda carece de pesquisas que envolvam professores e pesquisadores.

Ademais, esse curso pode divulgar a Aprendizagem Movidada por Dados a professores em serviço, de modo a estabelecer conexões com as teorias de ensino com as quais trabalham em seus contextos de ensino. Isso se mostra de grande relevância ao se considerar a colocação de Crosthwaite (2024), no que tange ao fato de a DDL ainda não estar amplamente disseminada ou correlacionada às teorias mais convencionais de ensino.

Somado a isso tudo, esse processo de formação dos professores ainda gerou uma série de atividades didáticas que utilizaram a abordagem da Aprendizagem Movidada por Dados para implementar os materiais didáticos. Todas as propostas desenvolvidas pelos cursistas com o apoio do pesquisador comprovaram o quanto a DDL não é necessariamente trabalhosa ou complexa, mas sim possível de ser incorporada aos recursos didáticos e/ou materiais que o professor já tenha disponíveis. Dois fatores atribuem-se a essa afirmação, que é (i) a disponibilização de ferramentas de fácil utilização, como o Skell, o Linggle e o Versatext; e (ii) a flexibilização em se poder usar tanto o modelo de acesso direto ao *corpus* (*hands on*) quanto o de acesso indireto (*hands-off*).

Todas essas reflexões culminam na última pergunta desta pesquisa:

- Quais são os princípios pedagógicos que podem orientar a utilização da Aprendizagem Movidada por Dados na implementação de materiais didáticos ou na elaboração de lições e/ou atividades didáticas?

Eles são quinze e versam que as lições e/ou atividades que se utilizam da Aprendizagem Movidada por dados precisam:

1. Contar com um planejamento prévio em que se considere o objetivo da aula e da consulta ao *corpus*, assim como as ferramentas da LC a serem utilizadas;
2. Apresentar um tema que perpassasse por todos os estágios, subsidiando e dando suporte à pesquisa a ser feita pelos estudantes no *corpus*;
3. Contextualizar esse tema à realidade dos estudantes;

4. Promover interação entre os estudantes, correlacionando o tema ao objeto do conhecimento;
5. Levantar os conhecimentos prévios dos estudantes;
6. Incentivar a formulação prévia de hipóteses sobre o funcionamento da língua;
7. Preparar os estudantes em relação aos conteúdos linguísticos que eles possivelmente precisarão acionar durante a consulta ao *corpus*;
8. Despertar o interesse dos estudantes em realizar a pesquisa no *corpus*;
9. Ensinar aos estudantes a utilização das ferramentas da LC usadas para a consulta ao *corpus*;
10. Posicionar o professor como mediador do processo de busca e de análise dos resultados obtidos (as linhas de concordância);
11. Estabelecer um propósito claro que norteie a realização da pesquisa e a análise dos dados obtidos (as linhas de concordância);
12. Conceber o estudante como sujeito ativo no processo de aprendizagem, de modo que o aluno seja pesquisador/detetive da língua;
13. Ter uma progressão linear em relação ao nível de complexidade, partindo, preferencialmente, do *hands-off* para o *hands-on*;
14. Proporcionar colaboração entre os estudantes e o professor durante o processo de especulação e generalização acerca dos padrões da língua observados nos resultados das buscas (as linhas de concordância);
15. Oportunizar aos estudantes a utilização personalizada e contextualizada dos padrões linguísticos descobertos pela pesquisa em *corpus*.

Enfatiza-se que eles foram cunhados após um árduo processo científico, que envolveu uma revisão bibliográfica vasta e interdisciplinar, haja vista que se apoia em teorias ora voltadas à Educação, ora à LC. No entanto, o procedimento não se findou na leitura e na análise dessas teorias, uma vez que elas foram mobilizadas na elaboração das lições que compõem o livro *Caderno de Atividades de Aprendizagem Movidada por Dados*, que relacionou teoria à prática, no oferecimento do curso de formação continuada em serviço para os professores e no apoio dado pelo pesquisador aos cursistas em relação ao desenvolvimento das atividades didáticas que utilizam da DDL para a implementação dos materiais didáticos.

Por fim, no tocante aos encaminhamentos, espera-se que esta pesquisa contribua com futuros estudos que correlacionem a LC e a Aprendizagem Movida por Dados ao ensino, em especial, à etapa da Educação Básica, que se subdivide em Educação Infantil, Ensino Fundamental e Ensino Médio. Além disso, almeja-se ainda que professores também usufruam das lições, das atividades didáticas e dos princípios pedagógicos para o planejamento de suas aulas de língua inglesa, de modo que levem a DDL às suas salas de aulas.

REFERÊNCIAS

- ALMEIDA, M. E. B. Apresentação. In: BACICH, L.; MORAN, J. (Org.). *Metodologias ativas para uma educação inovadora: uma abordagem teórico-prática*. Porto Alegre: Penso, 2018.
- BERBER SARDINHA, T. Como usar a Linguística de Corpus no ensino de língua estrangeira. In: TAGNIN, S. E. O.; VIANA, V. (Org.). *Corpora no ensino de línguas estrangeiras*. São Paulo: Hub Editorial, 2011. p. 301-356.
- BERBER SARDINHA, T. *Linguística de Corpus*. Barueri: Manole, 2004.
- BERBER SARDINHA, T. Linguística de Corpus: histórico e problemática. *Delta*, São Paulo, v. 16, n. 2, p. 323-367, 2000.
- BIBER, D. et al. *Corpus linguistics investigating language structure and use*. Cambridge: Cambridge University Press, 1998.
- BOULTON, A. Data-drive learning: in conversation with Alex Boulton. In: CROSTHWAITE, P. (Org.). *Corpora for language learning: bridging the research-practice divide*. New York: Routledge, 2024. p. 43-58.
- BOULTON, A. Hands-on / hands-off: alternative approaches to data-driven learning. In: THOMAS, J.; BOULTON, A. (Org.). *Input, Process and Product: Developments in Teaching and Language Corpora*. Brun: Masaryk University Press, 2012. p. 152-168.
- BOULTON, A. Data-driven learning: taking the computer out of the equation. *Language Learning*, v. 60, n. 3, p. 534-572, 2010.
- BRASIL, Ministério da Educação. *Base Nacional Comum Curricular: educação é a base*. Brasília, 2018. Disponível em: http://basenacionalcomum.mec.gov.br/images/BNCC_EI_EF_110518_versaofinal_sit_e.pdf. Acesso em: 20 julho 2022.
- BRASIL. *Lei nº 9.394*, de 20 de dezembro de 1996. Estabelece as diretrizes e bases da educação nacional. Diário Oficial da União, Brasília, DF, 1996. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/l9394.htm. Acesso em: 20 julho 2022.
- BRASIL. *Lei nº 13.005*, de 25 de junho de 2014. Aprova o Plano Nacional de Educação – PNE e dá outras providências. Diário Oficial da União, Brasília, DF, 2014. Disponível em: http://www.planalto.gov.br/ccivil_03/ato2011-2014/2014/lei/l13005.htm. Acesso em 20 de julho 2022.
- CAMARGO, F. Por que usar metodologias ativas de aprendizagem. In: CAMARGO, F.; DAROS, T. (Org.). *A sala de aula inovadora: estratégias pedagógicas para fomentar o aprendizado ativo*. Porto Alegre: Penso, 2018.
- CARVALHO, C.T.; LARANJA, L. A. N.; PINTO, P. T. Do-it-Yourself Corpora: what are they and who are they for? *TradTerm*, São Paulo, v. 37, n. 1, p. 67-84, 2021.

CHAMBERS, A. What is data-driven learning? In: O'KEEFFE, A; MCCARTHY, M. (Org.). *The Routledge Handbook of Corpus Linguistics*. Abingdon: Routledge, 2010. p. 345-358.

CROSTHWAITE, P. Bridging the research-practice divide for data-driven learning. In: CROSTHWAITE, P. (Org.). *Corpora for language learning: bridging the research-practice divide*. New York: Routledge, 2024. p. 1-8.

CROSTHWAITE, P. Taking DDL online: Designing, implementing and evaluating a SPOC on data-driven learning for tertiary L2 writing. *Australian Review of Applied Linguistics*, v. 43, n. 2, p. 169-195, 2020.

CROSTHWAITE, P.; WONG, L.; CHEUNG, J. Characterising postgraduate students' corpus query and usage patterns for disciplinary data-driven learning. *ReCALL*, v. 31 n. 3, p. 255-275, 2019.

CROSTHWAITE, P. Retesting the limits of data-driven learning: feedback and error correction. *Computer Assisted Language Learning*, v. 30, p. 447-473, 2017.

DAROS, T. Metodologias ativas: aspectos históricos e desafios atuais. In: CAMARGO, F.; DAROS, T. (Org.). *A sala de aula inovadora: estratégias pedagógicas para fomentar o aprendizado ativo*. Porto Alegre: Penso, 2018.

DORNELLES, C; IRALA, V. B. Dos apagamentos na BNCC de Língua Inglesa: entre novas e velhas roupagens. In: SILVA, K. A.; XAVIER, R. P. (Org.). *Múltiplos olhares para a Base Nacional Comum Curricular: Língua Portuguesa e Língua Inglesa*. Campinas: Pontes Editores, 2021. p. 205-220.

FIDALGO, S. S.; SPRENGER, T. M.; RAMOS, R. C. G. Formação de professores e a BNCC de Língua Inglesa: expectativas e desafios. In: SILVA, K. A.; XAVIER, R. P. (Org.). *Múltiplos olhares para a Base Nacional Comum Curricular: Língua Portuguesa e Língua Inglesa*. Campinas: Pontes Editores, 2021. p. 241-264.

FRANKENBERG-GARCIA, A. Corpora in ELT. In: HALL, G. (Org.). *The routledge handbook of English language teaching*. Abingdon: Routledge, 2016. p. 383-398.

FRIGINAL, E. *Corpus Linguistics for English Teachers: New Tools, Online Resources and Classroom Activities*. Nova York: Routledge, 2018.

GABRIELATOS, C. Corpora and Language Teaching: Just a Fling or Wedding Bells? *Tesl-Ej*. v. 8, n. 4, p. 4, 2005.

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. 145f. Dissertação (Mestrado em Estudos Linguísticos) – Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista (UNESP), São José do Rio Preto, 2020.

GRANGER, S. Learner Corpora. In: CHAPPELLE, C. A. (Org.). *The encyclopedia of Applied Linguistics*. Oxford: Blackwell Publishing, 2013.

GRANGER, S. A bird's-eye view of learner corpus research. In: GRANGER, S.; HUNG, J.; PETCH-TYSON, S. (Org.). *Computer learner corpora, second language acquisition and foreign language teaching*. Amsterdã: John Benjamins, 2002. p. 03-33.

GRANGER, S. The computer learner corpus: a versatile new source of data for SLA research. In: GRANGER, S. (Org.). *Learner English on computer*. New York: Longman, 1998. p. 03-18.

JOHNS, T. From printout to handout: grammar and vocabulary teaching in the context of data-driven learning. In: ODLIN, T. (Org.). *Perspectives on pedagogical grammar*. Cambridge: Cambridge University Press, 1994. p. 293-313.

JOHNS, T. Should you be persuaded: two samples of data-driven learning materials. *English language research journal*, Birmingham, v. 4, p. 1-16, 1991.

LEECH, G. Frequency, corpora and language learning. In: MEUNIER, F. et al. (Org.). *A Taste for Corpora: in honour of Sylviane Granger*. Amsterdã: John Benjamins, 2011. p. 7-31

LOVATO, L. F. et al. Metodologias ativas de aprendizagem: uma breve revisão. *Acta Scientiae*, Canoas, v. 20, n. 2, p. 154-171, 2018.

MCENERY, T.; HARDIE, A. *Corpus linguistics: method, theory and practice*. Cambridge: Cambridge University Press, 2012.

MORAN, J. *Metodologias ativas de bolso: como os alunos podem aprender de forma ativa, simplificada e profunda*. São Paulo: Editora do Brasil, 2019.

MORAN, J. Metodologias ativas para uma aprendizagem mais profunda. In: BACICH, L.; MORAN, J. (Org.). *Metodologias ativas para uma educação inovadora: uma abordagem teórico-prática*. Porto Alegre: Penso, 2018.

MORAN, J. Mudando a educação com metodologias ativas. In: SOUZA, C. A.; MORALES, O. E. T. (Org.). *Convergências midiáticas, educação e cidadania: aproximações jovens*. PG: Foca Foto-PROEX/UEPG, 2015. p. 15-33.

NESSSELHAUF, N. Learner corpora and their potential for language teaching. In: SINCLAIR, J. M. (Org.). *How to use corpora in language teaching*. Amsterdã: John Benjamins, 2004. p. 125-152.

ORENHA-OTTAIANO, A.; PINTO, P. T. Pedagogia do Léxico e da Tradução: Novas Práticas em Pesquisa. In: ROCHA, N.; RODRIGHUES, A.; CAVALARI, S. (Org.). *Novas práticas em pesquisa sobre a linguagem: rompendo fronteiras*. São Paulo: Cultura Acadêmica, 2018. p. 127-144.

PAIVA, V. L. M. O. *Manual de pesquisa em estudos linguísticos*. São Paulo: Parábola, 2019.

PAIVA, V. L. M. O. *Aquisição de segunda língua*. São Paulo: Parábola Editorial, 2014.

PINTO, P. T. Using a Data-Driven Learning Approach to Teach English for Academic Purposes and Improve Higher Education Internationalisation. *Corpora for Language Learning*. 01ed.: Routledge, 2024, p. 227-239.

PINTO, P. et al. (Org.). *Using language data to learn about language: a teachers' guide to classroom corpus use*. Brisbane, QLD Australia: The University of Queensland. <https://doi.org/10.14264/3bbe92d>

PINTO, P. T.; CROSTHWAITE, P. Exploring new approaches to concordancing: for enhancing academic reading and writing across disciplines. In: Komesu, F.; Assis, J. A.; Lamrani, Z.; El Houdna, Y. (Org). *Academic-Scientific Literacies*. Belo Horizonte: Editora PUC Minas, (in press).

PINTO, P. T.; GARCIA, W. D.; SERPA, T. S. (Org.). *Caderno de Atividades de Aprendizagem Movida por Dados*. Campinas, SP: Pontes Editores, 2022.

REPPEN, R.; SIMPSON-VLACH, R. Corpus Linguistics. In: SCHMITT, N.; RODGERS, M. P. H. (Org.). *An Introduction to Applied Linguistics*. Abingdon: Routledge, 2020. p. 91-108.

REPPEN, R. *Using Corpora in The Language Classroom*. Cambridge: Cambridge University Press, 2010.

RIBAS, F. C. Política de ensino de Língua Inglesa: a BNCC versus as pesquisas sobre letramento digital e uso de tecnologias digitais. In: SILVA, K. A.; XAVIER, R. P. (Org.). *Múltiplos olhares para a Base Nacional Comum Curricular: Língua Portuguesa e Língua Inglesa*. Campinas: Pontes Editores, 2021. p. 221-240.

ROMER, U. Using General and Specialized Corpora in English Language Teaching: Past, Present and Future. In: CARMEN, M. et al. (Org.). *Corpus-Based Approaches to English Language Teaching*. Londres: Continuum, 2010. p. 18-35.

ROZENFELD, C. C. de F.; VIANA, N. Reflexões teóricas sobre a aula de língua estrangeira: organizando materiais, analisando contextos, definindo percursos. *DELTA: Documentação e Estudos em Linguística Teórica e Aplicada*, v. 35, n. 4, 2020. Disponível em: <https://revistas.pucsp.br/index.php/delta/article/view/47112>. Acesso em: 11 fev. 2024.

SALLES, F. C. A formação continuada em serviço. *Revista Iberoamericana de Educación*, v. 34, p. 1-8, 2004

SÃO PAULO, Secretaria da Educação do Estado de São Paulo. *Currículo Paulista Etapa Ensino Médio*. São Paulo: SEDUC, 2020. Disponível em: <https://efape.educacao.sp.gov.br/curriculopaulista/wp-content/uploads/2020/08/CURR%C3%8DCULO%20PAULISTA%20etapa%20Ensino%20M%C3%A9dio.pdf>. Acesso em 20 julho 2022.

SERPA, T; PINTO, P. T.; SBROGIO, R. de O.; CAMARGO, D. C. de. O ambiente digital Tradcorpus: proposta de repositório de dados de corpora on-line para o ensino de línguas estrangeiras e de tradução. *Trabalhos em Linguística Aplicada*, Campinas, SP, v. 62, n. 1, p. 121-139, 2023. Disponível em:

<https://periodicos.sbu.unicamp.br/ojs/index.php/tla/article/view/8666006>. Acesso em: 12 julho 2023.

SINCLAIR, J. *Corpus, concordance, collocation*. Oxford: Oxford University Press, 1991.

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

VIANA, V. Linguística de Corpus: conceitos, técnicas & analyses. In: VIANA, V; TAGNIN, S. E. O. (Org.). *Corpora no ensino de línguas estrangeiras*. São Paulo: Hub Editorial, 2011. p. 25-95.

APÊNDICE A – EMENTA DO CURSO “PRÁTICAS PEDAGÓGICAS INOVADORAS PARA O ENSINO DE LÍNGUA INGLESA: O ALUNO COMO DETETIVE DA LÍNGUA E PROTAGONISTA DA APRENDIZAGEM”

EMENTA DO CURSO

Práticas pedagógicas inovadoras para o ensino de língua inglesa: o aluno como detetive da língua e protagonista da aprendizagem.

DOCENTES

Prof. Me. William Danilo Garcia, Profa. Dra. Paula Tavares Pinto, Profa. Dra. Adriane Orenha Ottaiano, Profa. Dra. Talita Serpa, Prof. Dr. Eduardo Batista da Silva, Prof. Dr. Peter Crosthwaite, Profa. Me. Arianne Donizete Delgado Ribeiro Caldas, Profa. Me. Carolina Tavares de Carvalho e Prof. Me. Emanuel Henrique Alves.

MODALIDADE

Este curso será ministrado de forma *online*.

APRESENTAÇÃO

O curso “*Práticas pedagógicas inovadoras para o ensino de língua inglesa: o aluno como detetive da língua e protagonista da aprendizagem*” introduz ferramentas digitais comuns à Linguística de *Corpus* com o intuito de que os professores se familiarizem com esses recursos e os implementem em suas práticas pedagógicas. Essa formação se faz necessária, pois fornece aos docentes possibilidades de aprimorarem seus métodos de ensino, trabalhando de forma inovadora e visando ao desenvolvimento da autonomia e do protagonismo dos alunos, que, por meio dessas tecnologias, assumem o papel ativo de investigador e/ou detetive da língua inglesa. Reconhecemos, inclusive, que esses preceitos são alguns dos norteadores tanto da BNCC quanto do Currículo Paulista.

OBJETIVOS

Oferecer formação continuada aos professores de inglês da rede pública do estado de São Paulo e do Goiás. Para isso, o foco será:

- Conhecer algumas das ferramentas digitais comuns à Linguística de *Corpus* e compreender como elas podem contribuir para a pesquisa e, conseqüentemente, a construção de novos conhecimentos acerca da língua inglesa;

- Refletir as possibilidades de implementação desses recursos em práticas pedagógicas inovadoras e que promovam o protagonismo dos alunos ao colocarem os estudantes em uma posição ativa e de pesquisadores da língua.

A QUEM SE DESTINA?

Aos professores de língua inglesa da rede pública estadual de São Paulo e do Goiás.

COMO SERÁ O CURSO?

O conteúdo programático do curso será desenvolvido por meio da realização de dois módulos do curso “*Improving writing through corpora*”, desenvolvido pelo professor Peter Crosthwaite da Universidade de Queensland da Austrália e de oficinas que apresentarão algumas das ferramentas digitais comuns à Linguística de *Corpus* que estejam disponíveis gratuitamente online. Para isso, a cada *workshop*, os cursistas serão, primeiramente, expostos às atividades didáticas que utilizam desses recursos para que, em seguida, elas sejam devidamente apresentadas aos professores. Essa forma de aplicação das oficinas se faz necessária, pois desse modo os docentes compreendem exatamente como seus alunos se sentirão ao serem expostos a essas tecnologias digitais de informação e comunicação. Sendo assim, acreditamos que essa experiência prática de atuarem como pesquisadores/detetives da língua promoverá nos cursistas engajamento e interesse em conhecer os recursos próprios da Linguística de *Corpus*, assim como utilizá-los em sala de aula e na elaboração de atividades.

PROGRAMAÇÃO

Oficina	Docente	Data
<i>Introdução: aplicações pedagógicas do DDL e introdução aos módulos 1 e 2</i>	Profa. Dra. Paula Tavares Pinto e a Profa. Me. Carolina Tavares de Carvalho (Unesp)	04/05/2022 às 19h
<i>Improving writing through corpora – módulo 1</i>	Prof. Dr. Peter Crosthwaite (Universidade de Queensland)	11/05/2022
<i>Improving writing through corpora – módulo 2</i>	Prof. Dr. Peter Crosthwaite (Universidade de Queensland)	18/05/2022
<i>Linggle e Skell: aplicações pedagógicas</i>	Prof. Me. William Danilo Garcia, Prof. Me. Emanuel Henrique Alves e Profa. Me. Ariane Donizete Delgado Ribeiro Caldas (doutorandos pela Unesp)	25/05/2022 às 19h
<i>VocabProfile: aplicações pedagógicas</i>	Prof. Dr. Eduardo Batista da Silva (Universidade Estadual de Goiás)	01/06/2022 às 19h
<i>PlatCol e Online English Collocations Platform: aplicações pedagógicas</i>	Profa. Dra. Adriane Orenha Ottaiano e Prof. Me. Emanuel Henrique Alves (Unesp)	08/06/2022 às 19h
<i>Atividades hands on/off (Classroom) e TradCorpus (elaboração de atividades didáticas)</i>	Prof. Me. William Danilo Garcia e Profa. Dra. Talita Serpa (Unesp)	22/06/2022 às 19h

AVALIAÇÃO

A avaliação do curso possui um caráter formativo e envolverá, portanto, a participação e o engajamento dos cursistas durante todas as oficinas. Ela também se

pautará na atividade didática envolvendo uma das ferramentas digitais, que será elaborada pelos participantes do curso. Além disso, utilizaremos questionários com o objetivo de identificar o público-alvo e suas necessidades específicas, assim como obter dados relacionados ao aproveitamento do curso pelos professores.

PRÉ-REQUISITOS

Esperamos que os cursistas possuam boa conexão de internet, visto que o curso será ministrado na modalidade online. No que diz respeito à proficiência no idioma, é desejável um nível intermediário na habilidade de compreensão escrita.

RESPONSABILIDADES COMO CURSISTA

Os cursistas precisam estar atentos às datas e aos prazos estabelecidos na programação. Esperamos que os participantes tenham uma postura engajada e participativa durante as oficinas e uma predisposição com relação à elaboração de uma atividade didática que faça uso de uma das ferramentas digitais abordadas no curso.

CERTIFICAÇÃO

Será gerado um certificado de 30 horas aos participantes que obtiverem presença superior a 75% das oficinas ministradas e apresentarem uma proposta de atividade didática que utilize uma das ferramentas digitais comuns à Linguística de *Corpus* abordadas no curso.

TIRE SUAS DÚVIDAS

Em caso de dúvidas, entre em contato pelo e-mail: william.danilo@unesp.br e/ou carolina.tavares@unesp.br

LINKS

Sketch Engine <https://www.sketchengine.eu/>

Skell: <https://skell.sketchengine.eu/#home?lang=en>

Dicionário

de

Colocações

Platcol:

<http://www.institucional.grupogbd.com/dicionario/sobre>

Dicionário Linguee: <https://www.linguee.com.br/>

Vocab Profile: <https://www.lex tutor.ca/vp/eng/>

Curso Improving writing through corpora - University of Queensland:
[https://edge.edx.org/courses/course-](https://edge.edx.org/courses/course-v1:UQx+HASS0653N_7120_00002+2021_S1/about)
[v1:UQx+HASS0653N_7120_00002+2021_S1/about](https://edge.edx.org/courses/course-v1:UQx+HASS0653N_7120_00002+2021_S1/about)

ANEXO A – CERTIFICADO DE TRADUÇÃO DO CURSO “IMPROVING WRITING THROUGH CORPORA: DATA-DRIVEN LEARNING?” PARA O PORTUGUÊS



School of Languages and Cultures

CRICOS PROVIDER NUMBER 000258

CERTIFICATE OF TRANSLATION

This is to certify that

William Danilo Garcia

has satisfactorily translated the course SLATx: *Improving your writing through corpora* into Portuguese on **11 July 2021**.



Dr. Peter Crosthwaite

Fellow: Higher Education Academy: <https://www.heacademy.ac.uk/>

UQ researcher's profile: <http://researchers.uq.edu.au/researcher/18380>

See publications at https://www.researchgate.net/profile/Peter_Crosthwaite

Editorial Board member *Applied Corpus Linguistics* (<https://www.journals.elsevier.com/applied-corpus-linguistics>)

Section editor for corpus linguistics at *Open Linguistics* (<https://www.de Gruyter.com/view/i/opli>)

orcid.org/0000-0002-1482-8381, CRICOS: 000258