

ABORDAGEM ORIENTADA POR DADOS: DESAFIOS E APLICAÇÕES NO ENSINO DE LE

Data-Driven Learning Approach: challenges and applications in the teaching of foreign languages

Jean Michel Pimentel ROCHA¹

Resumo: Diante da influência das tecnologias da informação e comunicação (TICs) e da aprendizagem de línguas assistida por computador (CALL - *Computer Assisted Language Learning*) no ensino de inglês como Língua Estrangeira (LE), as quais possibilitam novos olhares para a prática didático-pedagógica, objetivamos, neste artigo, refletir acerca das aplicações de *corpora* no ensino, principalmente no que concerne à Abordagem Orientada por Dados (*Data-Driven Learning Approach*). Trata-se de uma abordagem centrada no aprendiz, o qual, mediado pelo professor, busca entender o funcionamento da língua observando-a em um *corpus*. Nesse contexto, fundamentados em Johns (1991b), em Boulton (2009b, 2010) e em Boulton e Tyne (2013), examinamos os principais desafios para sua implementação, bem como problematizamos sua relação com a noção de autonomia na aprendizagem. Ademais, partindo dessa reflexão e do instrumental teórico-metodológico da Linguística de Corpus, exploramos didaticamente um *corpus* de resumos científicos da área de linguística a fim de auxiliar a escrita desse gênero textual quanto ao reconhecimento de padrões lexicogramaticais, por licenciandos em Letras, especialmente os iniciantes na pesquisa acadêmica. Ao propor este trabalho, esperamos despertar nesse público o interesse pela utilização de recursos tecnológicos que visam a processos de aprendizagem autônomos a partir de *corpora*, de modo que possam conscientizar-se da existência desses recursos e possam aplicá-los em suas práticas acadêmicas cotidianas.

Palavras-chave: *Corpora* no ensino. Abordagem Orientada por Dados. Autonomia na aprendizagem. Escrita acadêmica. Resumo científico.

Abstract: Considering the influence of information and communication technologies (ICTs) and of computer-assisted language learning (CALL) in the teaching of English as a foreign language, which make possible new perspectives for didactic-pedagogical practice, in this paper, we reflect upon the application of corpora in language teaching, especially regarding Data-Driven Learning Approach. It is a learner-centered approach, which, mediated by the teacher, seeks to understand the functioning of the language by observing it in a corpus. In this context, founded on the works of Johns (1991a), Boulton (2009b, 2010) and Boulton and Tyne (2013), we examine the main challenges for its implementation, as well as we problematize its relation with the notion of learning autonomy. Furthermore, based on this reflection and on the theoretical framework of Corpus Linguistics, we explore a corpus made of abstracts from the field of linguistics in order to assist Portuguese language undergraduate students' writing process, particularly those beginners in academic research, in relation to lexicogrammatical patterns. In proposing this work, we hope to arouse in this public the interest in the use of technological resources that aim at autonomous learning processes from corpora, so that they can become aware of its existence and can apply them in their daily academic practices.

Keywords: Corpora and language teaching. Data-driven learning approach. Learning autonomy. Academic writing. Scientific abstract.

¹ Rocha. Unesp. Endereço eletrônico: jeanpimentel_sp@yahoo.com.br

- | Abordagem orientada por dados: desafios e aplicações no ensino de LE

Apresentação

Na era digital, a utilização de recursos possibilitados pelas TICs e *CALL* faz-se cada vez mais presente no ensino e na aprendizagem de LE, acarretando profundas transformações na cultura de aprendizagem, em especial na relação professor/aprendiz, que têm seus papéis reavaliados ante a emergência de (novos) métodos (FITZPATRICK, 2004). Em meio aos recursos empregados, destacamos, neste artigo, a aplicação direta de *corpora*, evidenciada na chamada Abordagem Orientada por Dados (Abordagem DDL), a qual centra-se no aprendiz e no uso de programas de concordância. Nesse sentido, examinamos os principais desafios para implementá-la, bem como refletimos sobre a sua relação com a noção de autonomia, tida como um de seus diferenciais.

A partir dessa discussão, a fim de ilustrar uma possível aplicação da abordagem, trazemos uma proposta de ensino voltada para a escrita de resumos (*abstracts*) em língua inglesa, uma prática letrada exigida no meio acadêmico. O foco da proposta está no reconhecimento de padrões lexicogramaticais (colocações e coligações) comumente empregados em resumos científicos da área de linguística nas direções tradutórias português > inglês. Nosso público-alvo são licenciandos em Letras, especialmente os iniciantes na pesquisa acadêmica. Partimos do pressuposto de que, sendo capazes de explorar e reconhecer essas combinatórias em um *corpus*, os estudantes ampliarão e enriquecerão seu repertório lexical na área enfocada, podendo empregar as combinações lexicais aprendidas na redação de seus próprios resumos, aproximando-se, dessa maneira, dos usos que seus pares fazem dessas estruturas. Ressaltamos ainda que, apesar de enfocarmos uma temática específica, os princípios da abordagem DDL e as formas de exploração de um *corpus* por meio dela podem se estender a qualquer temática e a outros níveis de ensino. Nosso público-alvo não se restringe, assim, ao estudante pesquisador, mas se estende também ao estudante que se interessa pelo ensino, seja de língua materna ou de língua estrangeira.

Abordagem Orientada por Dados: vantagens e desafios no ensino de LE

Idealizada por Johns (1991), a Abordagem DDL desenvolveu-se no final da década de 80, na área de ensino de inglês como LE. Surge como revolucionária ao propor mudanças na relação professor/aluno e no emprego de recursos tecnológicos em sala de aula.

Em se tratando do primeiro aspecto, em contraposição a um ensino tradicional, caracterizado pelo modelo IRF (*Initiation, Response e Feedback*) – em que o professor inicia uma discussão, o aluno responde e o professor oferece um *feedback* – essa abordagem reorganiza a relação professor/aluno. Desse modo, o modelo engessado é substituído por um mais dinâmico, no qual o aprendiz torna-se um pesquisador e sua tarefa é “descobrir a língua estrangeira”, e a [...] do professor é providenciar um contexto em que o aprendiz possa desenvolver estratégias para a descoberta – estratégias pelas quais ele ou ela possa ‘aprender a aprender’” (JOHNS, 1991b, p. 1, tradução nossa²). Quanto ao segundo aspecto, a ênfase é dada à ferramenta de concordância *KWIC* (*keywords in context*), principal recurso para a aplicação da abordagem. Com essa ferramenta, o aprendiz é levado a buscar e a observar em um *corpus* as regularidades dos fenômenos linguísticos e chegar a generalizações nem sempre possíveis pela utilização de materiais didáticos, que, em geral, trazem prontas as regras a serem ensinadas (JOHNS, 1991b).

Como precursor da abordagem, o autor enfatiza três principais mudanças no processo de ensino e aprendizagem: (i) **a primazia do dado** – ao fazer as observações em um *corpus*, tanto aluno quanto professor não sabem, a princípio, quais padrões irão encontrar; (ii) **o papel do professor** – este torna-se diretor e coordenador do aluno iniciante na pesquisa. Além disso, precisa refletir sobre sua própria prática de modo a se adaptar e adaptar a abordagem a sua realidade pedagógica; e (iii) **o papel da gramática** – regras e exemplos, por vezes intuitivos, dão lugar a evidências de uso autênticas, as quais contribuem para uma conscientização gramatical por parte do aprendiz, que, por sua vez, passa a fazer suas próprias descobertas.

Passados quase 30 anos de seu surgimento, tais mudanças, que, do nosso ponto de vista, mostram-se como principais vantagens, colocam-se nos dias de hoje, paradoxalmente, como os principais desafios para sua execução. Assim sendo, no ensino de LE, adotar uma abordagem que privilegie a primazia do dado e as formas de examiná-lo por meio de um *corpus* requer uma reflexão sobre como fazê-la chegar à sala de aula e sobre como será sua recepção por parte de professores e aprendizes. Com isso em mente, respaldados pelos estudos conduzidos por Boulton e Tyne (2013) e Boulton (2010, 2009b, 2009a), pode-se afirmar, resumindo as transformações mencionadas, que os desafios para se colocar em prática a abordagem DDL recaem, sobretudo, na relação estabelecida entre recursos, professores e alunos.

2 [...] the task of the learner is to “discover” the foreign language, and that the task of the language teacher is to provide a context in which the learner can develop strategies for discovery – strategies through which he or she can “learn how to learn”.

- | Abordagem orientada por dados: desafios e aplicações no ensino de LE

Ao fazer menção aos recursos (computador, *corpora*, concordanciador, plataformas *online*, entre outros), entendemo-los como vantagens, pois, ao lado de instrumentos de ensino tradicionais (a exemplo da lousa, papel e caderno), tais oferecem novos caminhos para o processo de ensino e aprendizagem. Entretanto, tornam-se um problema quando se tem dificuldades quanto ao acesso e à operacionalização (BOULTON; TYNE, 2013).

O problema do acesso seria, talvez, mais em relação à estrutura física das instituições de ensino do que em relação às ferramentas diretas para a utilização da abordagem, como plataformas *online* de gerenciamento de *corpora* (*Sketch Engine*³ e *BYU Corpora*⁴), ferramentas de compilação automática de *corpora* (*BootCat Front End*⁵) e programas de concordância (*AntConc*⁶ e *WordSmith Tools*⁷). Muitos desses recursos são gratuitos e até mesmo os que não são permitem acesso temporário ou não têm limitações quanto às funções. Observa-se, então, que o acesso direto a elas não seria em si um grande desafio. Além disso, há na *web* ou nos próprios *sites* que as hospedam tutoriais que auxiliam a operacionalizá-las.

Mesmo assim, segundo Boulton e Tyne (2013), as dificuldades poderiam permanecer, ainda mais quando se sabe que muitos professores têm pouca experiência com o manuseio do próprio computador, quadro que se complica ao se exigir o domínio de *softwares* mais complexos idealizados para a pesquisa científica. Uma solução para exploração pedagógica desse tipo de ferramenta, apontada pelos autores a partir de McCarthy (2008), estaria na participação direta dos professores no desenvolvimento delas, de modo que, baseando-se em suas próprias experiências, juntamente com editoras e acadêmicos, fossem criados *softwares* idealizados para a sala de aula, com uma interface mais “amigável”, por exemplo.

No que diz respeito à reavaliação da relação professor e aluno, conforme Johns (1991), ambos assumem uma nova postura: o aluno deixa de ser apenas um objeto receptor do conhecimento e o professor passa a ser um mediador da descoberta. Essa mudança radical, também vantajosa, traz consequências para o processo de ensino e aprendizagem. No entanto, ao assumir o papel de orientador do aluno, o professor poderia acreditar na redução do seu próprio papel. Essa seria, de acordo com Boulton (2009a), uma postura

3 <https://www.sketchengine.co.uk/>

4 <http://corpus.byu.edu/>

5 <http://bootcat.dipintra.it/>

6 <http://www.laurenceanthony.net/software/antconc/>

7 <http://www.lexically.net/wordsmith/>

ingênuo, visto que o computador não substituiria o professor, muito pelo contrário, a tarefa de guiar o aluno por um caminho, a princípio, desconhecido, exigiria mais cuidado. Nessa perspectiva, a responsabilidade do professor é ainda maior, pois tira o aluno de sua zona de conforto, auxiliando-o a perceber a língua por meio de um outro olhar e a chegar a generalizações acerca dos fenômenos linguísticos que a intuição por si só não daria conta de explicar.

Ao agir sobre os dados, os aprendizes tomariam consciência dos padrões de uso da língua, em especial, os lexicogramaticais, como as colocações e as coligações, além de outros aspectos sintagmáticos. Notariam a frequência de uso desses padrões, em geral, não abarcados nos livros didáticos, particularmente aqueles que não se baseiam em evidências empíricas (BOULTON, 2009a). Essa nova forma de conceber o ensino da gramática oportunizaria ao aluno ganhos cognitivos, com melhoras na capacidade de observar, interpretar, explorar, inferir, comparar, hipotetizar, etc. (O'SULLIVAN, 2007 *apud* BOULTON, 2009a).

Entretanto, uma questão anterior à mudança na postura professor/aluno e também desafiadora para a inserção da abordagem está em sua recepção em diferentes níveis de ensino. Essa é uma das principais críticas de Boulton e Tyne (2013) e Boulton (2009a). Assim, um obstáculo está em fazê-la transpor os limites da pesquisa e do ambiente acadêmico.

O distanciamento da abordagem de outras esferas de ensino evidencia-se, por exemplo, em Cassany (2016). Ao pesquisar o uso de recursos tecnológicos disponíveis na *web* por estudantes pré-universitários espanhóis, o autor revela que os mais utilizados são as ferramentas de tradução e os dicionários. Dos alunos entrevistados nesse estudo, nenhum fez referência ao emprego de *corpora*, o que aponta para o desconhecimento desse recurso pelos alunos e pelos próprios professores (CASSANY, 2016). No Brasil, não temos conhecimento de pesquisas similares. Mas, no que concerne a trabalhos voltados à temática DDL, embora não tenhamos localizado estudos que a explorassem de modo mais aprofundado, identificamos alguns que, de certa forma, problematizam sua aplicabilidade (MONZÓN; FADANELLI, 2016; MARTINS, 2016; ORENHA-OTTAIANO, ROCHA, 2015; SANTOS, 2011; BERBER SARDINHA, 2000).

Diante desse cenário, segundo Boulton (2009a), a lógica da Abordagem DDL, e mesmo das TICs, está em desacordo com os paradigmas tradicionais de ensino, orientados para o professor. Para superar esse entrave, uma saída seria a proposta de Conrad (2000), ao defender a necessidade de uma nova geração de professores de LE para dar conta dos desafios impostos à prática pedagógica pelas novas tecnologias. Isso exigiria, por exemplo, mudanças político-pedagógicas que só poderiam ser contempladas a longo

- | Abordagem orientada por dados: desafios e aplicações no ensino de LE

prazo. No entanto, Boulton (2009a) inclina-se para uma solução mais imediata. Em resposta a tais questões e a possíveis apontamentos negativos (dificuldade de manuseio do *corpus*, leitura das linhas de concordância, erros de interpretação) para viabilização da abordagem, argumenta em favor de uma opção integrativa, adaptando-a ao já existente em sala de aula – a exemplo de materiais impressos, como fizera o precursor da abordagem – e menos dependente de recursos tecnológicos. Para o autor, a melhor forma de expandi-la para além da comunidade científica é não promovendo-a como revolucionária: “[...] por um lado, DDL *não* é uma prática ordinária; por outro, DDL tem que se *tornar* uma prática ordinária (ou pelo menos não ser vista como uma prática *extraordinária* quando encontrada pela primeira vez)” (BOULTON, 2009b, p. 20, tradução e grifo nossos⁸).

Tal posicionamento distancia-se, nos termos de Gabrielatos (2005), de uma versão *hard* da abordagem – em que o aluno pesquisador se envolve em todas as etapas do processo – e aproxima-se de uma versão *soft*, a qual se justificaria pela praticidade, uma vez que incorporando-a a procedimentos comuns no ensino evitar-se-ia um possível ‘trauma’ para os alunos e também para os professores, em decorrência da inclusão, de uma só vez, de uma nova abordagem (*DDL approach*), de novos materiais (*corpora*) e de novos recursos tecnológicos (computador/ferramenta de concordância) (BOULTON, 2009b).

Além disso, a flexibilização da abordagem abriria perspectivas para alcançar um público maior – e, mais importante, não eliminaria sua característica primordial, ou seja, o procedimento indutivo, que possibilitaria ao aluno explorar os dados, buscar por padrões e formular hipóteses para explicá-los (BOULTON, 2009a). O que não pode acontecer, nas palavras dos autores, é a privação de alunos e professores de terem contato com o potencial pedagógico da abordagem.

A noção de autonomia e a abordagem DDL

Na literatura acerca da Abordagem DDL (LEŃKO-SZYMAŃSKA; BOULTON, 2015; BOULTON; TYNE, 2013; ASTON, 2001; MUKHERJEE, 2006; JOHNS, 1991a, 1991b), um dos ganhos a ela atribuídos diz respeito à autonomia na aprendizagem/do aprendiz. Em tais trabalhos, não há, todavia, uma formulação teórica mais aprofundada que permita-nos compreender o que fundamenta o conceito de autonomia empregado. Depreende-se,

8 [...] on the one hand, DDL is not ordinary practice; on the other, DDL has to become ordinary practice (or at least, not to be seen as extraordinary when first encountered).

de maneira geral, que seja o desenvolvimento por parte do aprendiz da capacidade de tomar decisões e tornar-se mais independente em relação ao seu próprio aprendizado. Frente a isso, faz-se necessária uma discussão, ainda que breve, sobre seu emprego no ensino de LE e que nos possibilite relacioná-la à própria abordagem DDL.

De acordo com Little (2007) e Benson (2006), a conceituação de autonomia nesse âmbito não é consensual e apresenta-se de várias maneiras. Apesar dessa problemática, uma das definições precursoras e difundidas é a de Holec (1981, p. 3, tradução nossa⁹), que a concebe como “a habilidade de responsabilizar-se pela própria aprendizagem”. Tal posição centra-se na individualidade do aprendiz e representa a mudança de um ensino direcionado a um ensino autoinstruído, cujo conhecimento objetivo e universal substitui-se pelo conhecimento subjetivo e individual do aluno, que se responsabiliza pelos objetivos e conteúdos a serem aprendidos, lançando mão de capacidades cognitivas e metacognitivas (LITTLE, 2007).

Para Benson (2011), a associação autonomia e individualização marcou os estudos da área nas décadas de 70 e 80 e foi bastante criticada, pois passava-se a impressão de que, nesse modelo, a aprendizagem ocorreria de forma isolada, sem a necessidade de professores ou de outros aprendizes. Em anos mais recentes, a ênfase recai na colaboração e na interdependência:

As decisões pessoais são necessariamente feitas em relação às normas morais e sociais, tradições e expectativas. Autonomia então inclui a noção de interdependência, o que é ser responsável pela própria conduta em um contexto social: ser capaz de cooperar com outros e resolver conflitos em maneiras construtivas. (KOHONEN, 1992, p. 19 *apud* BENSON, 2011, p. 15, tradução nossa¹⁰)

Essa relação de interdependência também está presente em Little (1996), ao defender que o desenvolvimento da autonomia envolve a participação crítica nas interações sociais (BENSON, 2011). Nesse contexto, Little (2007) propõe que a autonomia na aprendizagem de uma segunda língua abrange três princípios interativos: (i) envolvimento do aprendiz; (ii) reflexão do aprendiz e (iii) uso da língua-alvo. O primeiro

9 [...] the ability to take charge of one's own learning [...]

10 Personal decisions are necessarily made with respect to social and moral norms, traditions and expectations. Autonomy thus includes the notion of interdependence, that is being responsible for one's own conduct in the social context: being able to cooperate with others and solve conflicts in constructive ways.

- | Abordagem orientada por dados: desafios e aplicações no ensino de LE

princípio requer que o professor chame a atenção do aluno para o próprio processo de aprendizagem, isso envolve tomar decisões concernentes aos conteúdos, seleção de atividades e materiais da aprendizagem, além do gerenciamento e da avaliação da interação na sala de aula. Esse envolvimento leva ao segundo princípio, o da reflexão, que, além de abarcar o primeiro, visto que para alcançá-lo exige-se uma reflexão, mesmo incidental, abarca ainda uma reflexão explícita do processo de aprendizagem. O princípio do uso da língua-alvo, por sua vez, coloca-se como canal pelo qual o envolvimento e os processos reflexivos, e, conseqüentemente, a interação se estabelece (LITTLE, 2007). Segundo Little, entende-se que é por meio de interdependência desses princípios que se pode alcançar a autonomia na aprendizagem de LE. Para o autor, a autonomia, além de ser fruto de processos cognitivos e metacognitivos, é também produto de um processo interativo.

Ao traçar um paralelo com a Abordagem DDL, assumindo uma posição próxima das ideias de Little (2007) e de uma perspectiva construtivista no ensino e aprendizagem de LE, não concebemos que a autonomia esteja relacionada apenas à individualidade do aluno. Dessa forma, também entendemos que é na interação que ela se funda, isto é, a partir do compartilhamento de experiências, das trocas de conhecimentos e das reflexões promovidas entre aqueles que constituem o espaço de aprendizagem e aqueles que mesmo indiretamente se fazem presentes. A esse respeito, vale destacar as considerações de Leffa (2003), ao sustentar – mobilizando os conceitos vigotskianos de mediação e artefato cultural – que a autonomia do sujeito, em tese, pode ser uma ilusão ou uma pretensão, pois, mesmo na ausência de um professor, estamos sempre submetidos a outras vozes e discursos.

Alinhando-nos ainda ao pensamento de Freire (1996), podemos afirmar que, com a Abordagem DDL, especialmente pela reavaliação do papel do professor e do aprendiz, este também torna-se sujeito da produção do saber, não sendo apenas objeto de um conhecimento transferido, mas atuante em sua construção. Expostos aos dados, professores e alunos discutem e, em conjunto, problematizam possíveis caminhos para a produção do conhecimento. Nesse sentido, estamos em consonância com Freire (1996, p. 96, grifo nosso) quando assevera: “fundamental é que professor e alunos saibam que a postura deles [...] é *dialógica*, aberta, curiosa, indagadora e não apassivada [...]. O que importa é que professor e aluno se assumam *epistemologicamente curiosos*.” Na abordagem DDL, então, não se trata simplesmente de “*cut out the middleman*” – eliminar o professor –, mas de reavaliar seu posicionamento, bem como o do aluno para que juntos, enquanto seres sociais e históricos, compartilhem e construam conhecimento.

A capacidade de poder decidir sobre a própria aprendizagem é, desse modo, algo construído em conjunto e não um processo acabado. Pensar na existência de uma autonomia absoluta é assumir a postura radical de que somos seres acabados, quando, na verdade, tomar consciência do nosso inacabamento é que nos faz seres em constante busca do conhecimento, já que a educação é um processo de busca permanente (FREIRE, 1996). Assim sendo, “inacabados e conscientes do inacabamento, abertos à procura, curiosos, ‘programados’, mas, para aprender, exercitaremos tanto mais e melhor a nossa capacidade de aprender e de ensinar quanto mais sujeitos e não puros objetos do processo nos façamos” (FREIRE, 1996, p. 65).

Explorando um *corpus* de resumos científicos

Pelo que foi exposto, observa-se que o grande desafio para implementar a abordagem é fazê-la chegar à sala de aula. Vimos, além de outros aspectos mencionados, que um impedimento está no próprio desconhecimento de muitos educadores, os quais sequer ouviram falar da utilização de *corpora* no ensino. À vista disso, tendo em mente a formação de licenciandos em Letras, também aprendizes de LE, especialmente aqueles em processo de iniciação na pesquisa acadêmica, é que objetivamos, com esta proposta, oferecer subsídio para aplicação de uma versão *soft* da abordagem DDL, visando o enriquecimento da competência lexical do público em questão.

Apesar de o foco ser o estudante iniciante na pesquisa, como afirmado no início deste trabalho, salientamos que a proposta ora apresentada a ele não se limita. Desse modo, volta-se, de forma geral, à capacitação de professores em formação, os quais passam, respaldados pela abordagem DDL, a conhecer meios de exploração de um *corpus* em sala de aula, permitindo, independente da temática abordada, a replicação desse conhecimento. De forma mais específica, nossa proposta destina-se a um licenciando em Letras que, paralelamente ao interesse pelo ensino, desenvolve pesquisa acadêmica e vê-se diante da necessidade de escrever resumos científicos em língua inglesa – uma prática letrada exigida do estudante que começa seu percurso como pesquisador. Nesse contexto, o reconhecimento e a utilização de padrões lexicogramaticais mais recorrentes/convencionais e comumente empregados na redação acadêmica podem contribuir para o enriquecimento do repertório lexical de aprendizes de LE e, por conseguinte, na obtenção de melhorias ao redigirem esse gênero.

- | Abordagem orientada por dados: desafios e aplicações no ensino de LE

Nosso *corpus* constitui-se de resumos científicos da área de linguística, recolhidos de periódicos da biblioteca eletrônica *Scielo*¹¹. Foram 259 resumos em língua portuguesa e suas traduções em língua inglesa. Trata-se de um *corpus* paralelo, isto é, composto de textos originais e de suas respectivas traduções. Após fazer o *download* dos textos, os separamos em pastas distintas para cada língua e fizemos uma limpeza, retirando, por exemplo, a identificação de autoria. Feito isso, salvamos em formato .txt para que pudessem ser lidos pelo *Software WordSmith Tools* (SCOTT, 2012). Apesar de ser uma versão paga, pode-se ter acesso livre as suas ferramentas básicas (*Concord*, *KeyWords* e *WordList*).

Para averiguar as características do *corpus*, utilizamos a ferramenta *WordList*, que gera uma lista de palavras do *corpus*. Tanto a lista do *corpus* em inglês quanto a do *corpus* em português apresentam pouco mais de 37.000 palavras (*tokens*). Salvando-as em formato .lst e contrastando-as com as listas de palavras de um *corpus* de referência, obtém-se uma outra lista, a de palavras-chave, gerada pela ferramenta *KeyWords*, a qual apresenta as palavras que são mais características dos *corpora* sob análise. Para obtê-la, em inglês e em português, respectivamente, empregamos a lista dos *corpora* de referência BNC (*British National Corpus*)¹² e Corpus da Folha¹³. De posse desse vocabulário chave, é possível, por exemplo, observar vários pontos a serem problematizados, como mostraremos a seguir.

Levantamento de padrões lexicogramaticais

Inicialmente, de posse das listas de palavras-chave, o professor convidaria seus alunos para um breve exame da lista em português, conduzindo-os a refletir sobre as palavras que encabeçam a listagem (organizada por ordem de frequência). Os aprendizes logo perceberiam que a maioria são palavras lexicais típicas de resumos científicos, já que a ideia dessa lista é filtrar a lista geral de palavras, deixando justamente aquelas que são chave no *corpus*.

Notariam que um dos primeiros vocábulos, em português, é *artigo*, com 168 ocorrências, seguido, um pouco mais abaixo da listagem, por vocábulos do mesmo campo semântico, tais como *pesquisa*, *trabalho* e *estudo*. Em seguida, o professor solicitaria aos alunos que observassem, utilizando a ferramenta *concord*, os principais padrões lexicogramaticais com os quais tais vocábulos coocorrem. Os aprendizes, após algumas

¹¹ <http://www.scielo.org/php/index.php?lang=pt>

¹² <http://www.natcorp.ox.ac.uk/>

¹³ http://www.linguateca.pt/cetenfolha/index_info.html

observações, diriam que tais vocábulos colocam-se ao lado de dêiticos (*este, neste, o/a presente*), de verbos (*apresentar, discutir, analisar, examinar, buscar, pretender, descrever*) e antes de substantivo (*objetivo*). Além disso, concluiriam que aparecem na introdução do resumo, no qual, em geral, os objetivos do artigo costumam ser inseridos, conforme ilustra a figura:

N	Concordance
1	mação inicial do professor. O objetivo deste artigo é apresentar o Projeto Variação
2	objetos, promovendo novas identidades. Este artigo discute a noção de verdade
3	terialidade é lingüística e histórica. Neste artigo ilustramos o mérito da aplicação
4	segundas metades do século precedente. O artigo revisita, entre outras, hipóteses
5	internacionais do português. No presente artigo , discutimos alguns dados coletados

Figura 1 – Recorte da ferramenta *concord*

A partir dessa observação inicial, os aprendizes já teriam, minimamente, ideia de quais combinações lexicais costumam estruturar as partes iniciais de artigos científicos na área de linguística, o que já seria uma informação bastante útil aos que não costumam redigi-los. Mas, como precisam escrevê-los em língua inglesa, caberia averiguar como a comunidade de pesquisadores dessa área vem fazendo isso. Daí a necessidade do *corpus* em língua inglesa.

Dando continuidade à orientação, o professor pediria aos alunos que verificassem, no *corpus* em inglês, os padrões correspondentes daqueles até então levantados. Ao comparar as listas, não teriam dificuldades em encontrar vocábulos como *research, paper, article* e *study*. Novamente, uma busca com auxílio das ferramentas *concord, clusters* e *collocates* seria necessária. Os aprendizes constatariam que a unidade lexical mais utilizada para traduzir *artigo* é *paper*, mas que outras palavras como *article, study* e *research* também são empregadas. Entre os padrões mais recorrentes destacariam, outra vez, expressões dêiticas (*this, the present, in this*) seguidas de substantivo (*paper, article, study, research*) e de verbos (*aims to/at, study, intend, discuss, focus, show, analyze, examine, present, investigate, revisit, address, propose, describe, explain, seek*, entre outros). Examinando as palavras-chave, tanto da lista em português quanto da lista em inglês, o aprendiz poderia eleger vocábulos como verbos e substantivos característicos da área e consultar seu cotexto

- | Abordagem orientada por dados: desafios e aplicações no ensino de LE

nas linhas de concordância. Com essas informações, tomariam notas em seus cadernos, criando um pequeno banco de dados com as colocações mais recorrentes que possam auxiliá-los na escrita do gênero textual sob enfoque.

Usos de verbos seguidos de preposição e verbos frasais

Partindo das observações anteriores, um aprendiz poderia notar, por exemplo, diferenças de uso do verbo *to aim*, ora seguido da preposição *at*, ora da preposição *to*. Para esclarecer a dúvida, recorreriam às linhas de concordância na tentativa de chegar a uma resposta. Ao procurar por *aim** (*aims, aimed, aiming*) *at*, encontrariam 38 ocorrências e verificariam que em todas elas o verbo que segue a preposição estaria na forma infinitiva. Ao procurar por *aim** (*aims, aimed, aiming*) *to*, encontrariam 38 ocorrências, no entanto, para esses casos, o verbo seguido da preposição *at* estaria no gerúndio (*-ing*).

N	Concordance – AIMS TO	N	Concordance - AIMS AT
1	ncrypt. The present article aims to analyze how	1	TRACT The present article aims at analyzing reco
2	Abstract This article aims to present the path of	2	Abstract The present essay aims at delineating a the
3	in history. This paper aims to demonstrate	3	practices. This paper aims at presenting partial res
4	Linguistics, this paper aims to discuss the notion	4	The present paper aims at discussing the questio
5	ironment. This paper aims to revisit the interge	5	in the novel. This paper aims at proposing a reflect

Figura 2 - Linhas de concordância para *aims to* e *aims at*

Ao interpretar as linhas, alunos e professores tentariam descobrir se há alguma diferença semântica entre os distintos usos ou se é apenas uma diferença de forma. Veriam que, no *corpus* em português, as correspondências para tais coocorrências são

o artigo/o trabalho objetiva ou *tem por objetivo*, com cerca de 132 ocorrências. Mapeariam ainda construções similares como em *o objetivo deste artigo/do trabalho*, cujas opções tradutórias em inglês mais frequentes são as estruturas coligacionais: *the purpose of, the aim of, the objective of* e *the focus of*.

A problematização do uso de *aims in/at* pode ser o gancho para discussão dos verbos que geralmente vêm seguidos de preposição. Recorrendo às linhas de concordância e pesquisando por preposições comuns, os aprendizes podem se deparar com tais verbos (*agree with, deal with, arise from, able to, access to, resort to, think about* etc). Esse pode ser o momento ideal para tratar também dos verbos frasais mais recorrentes nesses textos (*based on, draw on, focus on, look at, rely on, bring about, carry out, point out, find out*).

Um exercício envolvendo verbos frasais e verbos seguidos por preposição, que pode ser desenvolvido a partir do recorte das linhas de concordância, é o exercício “*fill in the blank*”. Editando as linhas de concordância, o professor pode omitir a preposição, criar um *handout* impresso e pedir aos alunos que completem as lacunas com as preposições que julgarem adequadas, como ilustrado a seguir:

Concordance – Verbos Frasais	Concordance - Verbos seguidos por preposição
1 Corpus based study was carried * for the	1 I conclude, agreeing * Ortega (2009) that a
2 mentioned and found *	2 how public translators deal
the relevance of this acad	* these conflicting
3 The results point	3 field of study that arises *
* the necessity of continued	an interface between
4 This paper, based * the theory of Systemic	4 metaphor that emerges * frequency based
5 Its draws * semiotic and Marxian tradit	studies
6 The article relies * the theoretical	5 the results obtained * the analysis lead us to
framework	6 we proceed to some claims * the current

Figura 3 - Linhas de concordância: verbos frasais e verbos seguidos por preposição

- | Abordagem orientada por dados: desafios e aplicações no ensino de LE

Usos de conectores

Ainda por meio da lista de palavras-chave, alunos podem fazer o levantamento de preposições e conjunções mais frequentes (*in, on, by, at, over* etc.) e examinar, pela ferramenta de concordância, quais são os conectores mais comuns em que tais preposições/conjunções aparecem. Certamente, encontrarão vários (**in**: *in the light of, in order to, in addition to, in conclusion, in terms of, in turn*; **on**: *one the one hand, on the other, on this field*; **by**: *by way of, by the use of, by means of, by this view*; **at**: *at least, at the same time*; **over**: *over the past*; **to**: *used to, due to, according to*; **as**: *as well as, as a sign of, as far as, as part of, such as, as a result of*). Além do levantamento e da busca por correspondentes, os alunos podem ser levados a discutir e agrupar os conectores encontrados conforme as relações (causa, decorrência, tempo, adição, exemplificação, concordância, comparação, entre outros) que estabelecem no texto.

Considerações finais

Com este artigo, refletimos acerca da Abordagem DDL, enfatizando, especialmente, os desafios para sua inserção no ensino de LE. Em um primeiro momento, vimos que os princípios essenciais da abordagem – primazia do dado, o papel do professor (e também do aluno) e o papel do ensino gramatical – são os principais desafios para implementá-la, haja vista que implicam mudanças na prática pedagógica que vão desde a instrumentalização de uma ferramenta nova à reavaliação da dinâmica professor/aprendiz em sala de aula.

Por meio de uma revisão teórica dos trabalhos de Boulton, verificamos que uma alternativa para sanar o problema estaria na integração da abordagem às atividades didáticas cotidianas. Apesar de concordar com essa visão, acreditamos tratar-se de uma medida paliativa e que exigiria, mesmo diante de uma versão mais *soft*, conhecimentos mínimos para aplicá-la. Dessa forma, posicionamos ainda em favor de mudanças a longo prazo, ou seja, mudanças de caráter político-pedagógicas concentradas na formação de professores, pela inserção de disciplinas nos currículos das instituições formadoras que enfatizem as TICs e CALL e, conseqüentemente, abordagens como a DDL. A nosso ver, apesar de esse ser o caminho mais árduo, é o que pode trazer a solução mais acertada, pois daria oportunidade para que os educadores pelo menos conheçam a gama de recursos disponíveis para que possam, posteriormente, avaliar sua eficácia e considerar a aplicação. Iniciativas a curto prazo também podem ser desenvolvidas na tentativa de criar uma ponte entre a abordagem, limitada aos ambientes institucionais de pesquisa, e outros níveis de

ensino. Essa ponte poderia ser construída a partir de cursos universitários de extensão, de aperfeiçoamento ou disciplinas optativas oferecidas à comunidade interna e externa, de modo que se estabeleça uma troca de conhecimentos e estratégias em conjunto, que possam ser pensadas para a inserção desse modelo em práticas educativas distintas. A difusão poderia, ainda, ser concretizada pela publicação de materiais didáticos e de artigos científicos resultantes desses cursos formativos.

Em um segundo momento de nossa discussão, relacionamos a noção de autonomia com a abordagem DDL, dado que aquela é considerada como uma das principais vantagens desta. Argumentamos que, embora haja por parte do aluno uma responsabilidade maior em relação ao seu aprendizado, não compactuamos com a ideia de que há uma autonomia absoluta, em razão de a busca do conhecimento, conforme Freire (1996), ser algo constante e que se dá sempre na dependência do outro, isto é, na relação dialógica estabelecida entre professores, alunos e demais sujeitos sócio-historicamente constituídos.

Finalmente, com propósito de ilustrar uma aplicação prática da abordagem, propusemos algumas sugestões de como trabalhar em sala de aula com um *corpus* de resumos científicos. Nossa exploração, mesmo que de forma breve, privilegiou padrões lexicogramaticais, no entanto, destacamos que essa investigação não se esgota aqui, muito mais poderia ser explorado, como tempos e vozes verbais mais recorrentes, usos de advérbios e de pronomes, estratégias de referenciação, características tradutórias, entre outros aspectos linguísticos que possam ajudar professores em formação e iniciantes na pesquisa acadêmica na escrita de um resumo científico em outra língua. Estudos contrastivos também poderiam ser conduzidos a partir da compilação de outros *corpora* constituídos de resumos de outras áreas do conhecimento e de outras línguas, a fim de se evidenciar as similaridades e diferenças em relação aos aspectos aqui abordados.

Vale a pena ressaltar que apenas sugerimos uma forma de exploração da abordagem, evidenciamos, assim, a necessidade de realmente colocá-la em prática no intuito de validar a sua aplicabilidade. Salientamos que as questões tratadas no *corpus* deste trabalho refletem o uso que a comunidade científica brasileira da área de linguística faz da língua inglesa na sua condição de língua franca, isto é, uma língua de comunicação global amplamente empregada por falantes que possuem línguas maternas distintas (SEIDLHOFER, 2005). Dessa forma, as características linguísticas da produção oral e escrita de falantes não-nativos nem sempre podem corresponder às características empregadas por falantes nativos. Caberia aos professores levar essa reflexão aos seus alunos, considerando que eles podem questionar se as construções com as quais se depararam são, de fato, empregadas por falantes/pesquisadores nativos. Essa seria uma outra tarefa a ser desenvolvida, cuja resposta poderia ser buscada via *corpora*.

- | Abordagem orientada por dados: desafios e aplicações no ensino de LE

Partindo da problematização do papel da língua inglesa enquanto língua franca, embora não tenha sido o objetivo desse artigo, pode-se problematizar a escrita acadêmica para além de um modelo de letramento – *modelo de habilidades de estudo* – que privilegie apenas o reconhecimento de aspectos formais da língua (LEA; STREET, 2014). Nesse sentido, as distintas manifestações mundiais da língua inglesa abrem perspectivas para se pensar a escrita acadêmica a partir de um *modelo de letramentos acadêmicos*, como defendido por Lea e Street (2014), em que é dada ênfase às relações identitárias, de produção de sentido e de poder (a exemplo das imposições linguísticas) que envolvem a escrita em contextos acadêmicos.

Concluimos nossa reflexão reiterando que, com este trabalho, mesmo que minimamente, almejamos chamar a atenção de professores em formação para as aplicações de *corpora* no ensino, de maneira que lhes seja possível conhecê-las, refletir sobre elas e poder direcioná-las para suas práticas pedagógicas futuras, sabendo-se que, no contexto curricular dos cursos de Licenciatura em Letras no Brasil, pouca atenção é dada a tais questões. Sendo assim, apresentar a Abordagem já é um passo inicial para que tais educadores possam conscientizar-se de sua existência e avaliarem sua aplicabilidade.

Referências

ALLAN, R. Data-driven learning and vocabulary: investigating the use of concordances with advanced learners of English. **Centre for Language and Communication Studies Occasional Paper**, n. 66. Dublin: Trinity College Dublin, 2006.

ASTON, G. **Learning with Corpora**. Houston: Athelstan, 2001.

BENSON, P. Autonomy in language teaching and learning: state of the art article. **Language Teaching**, v. 40, p. 21-40, 2006.

_____. **Teaching and researching autonomy**. London: Pearson, 2011.

BERBER SARDINHA, A. P. Computador, corpus e concordância. In: LEFFA, V. J. **As palavras e sua companhia: o léxico e a aprendizagem das línguas**. Pelotas: EDUCAT, 2000. p. 47-75.

BOULTON, A. Data-driven learning: reasonable fears and rational reassurance. **Indian Journal of Applied Linguistics**, v. 35, n. 1, p. 81-106, 2009a.

_____. Data-driven learning: on paper, in practice. In: HARRIS, T.; MORENO JAÉN, M. **Corpus linguistics in language teaching**. Bern: Peter Lang, 2009b.

_____. Data-driven learning: taking the computer out of the equation. **Language Learning**, v. 60, n. 3, p. 534-572, 2010.

BOULTON, A.; TYNE, H. Corpus linguistics and data-driven learning: a critical overview. **Bulletin Suisse de linguistique appliquée**, v. 97, p. 97-118, 2013.

CASSANY, D. Recursos lingüísticos en línea: contextos, prácticas y retos. **Revista Signos: Estudios de Lingüística**, v. 49, p. 7-29, 2016.

CONRAD, S. Will corpus linguistics revolutionize grammar teaching in the 21st century? **TESOL Quarterly**, v. 34, p. 548-560, 2000.

FITZPATRICK, A. Information and communication technology in foreign language teaching and learning: an overview. In: _____. **Information and communication technologies in the teaching and learning of foreign languages: state of the art, needs and perspectives**. Moscow: Institute for Information Technologies in Education (IITE), 2004. p. 10-26.

FREIRE, P. **Pedagogia da autonomia: saberes necessários à prática educativa**. São Paulo: Paz e Terra, 1996.

GABRIELATOS, C. Corpora and language teaching: just a fling or wedding bells? **Teaching english as a second language – electronic journal**, v. 8, n. 4, p. 1-35, 2005.

HOLEC, H. **Autonomy in foreign language learning**. Oxford: Pergamon, 1981.

JOHNS, T. From printout to handout: grammar and vocabulary teaching in the context of data-driven learning. In: JOHNS, T.; KING, P. Classroom concordancing. **English Language Research Journal**, v. 4, p. 27-45, 1991a.

_____. Should you be persuaded: two samples of data-driven learning materials. In: JOHNS, T.; KING, P. Classroom concordancing. **English Language Research Journal**, v. 4, p. 1-16, 1991b.

KOHONEN, V. Experiential language learning: second language learning as cooperative learner education. In: NUNAN, D. **Collaborative language learning and teaching**. Cambridge: Cambridge University Press, 1992. p. 14-39.

LEA, M. R.; STREET, B. V. O modelo de “letramentos acadêmicos”: teoria e aplicações. Traduzido por Fabiana Komesu e Adriana Fischer. **Revista Filol. Linguíst. Port.**, São Paulo, v. 16, n. 2, p. 477-493, 2014.

LEFFA, V. J. Quando menos é mais: a autonomia na aprendizagem de línguas. In: NICOLAIDES, C. *et al.* **O desenvolvimento da autonomia no ambiente de aprendizagem de línguas estrangeiras**. Pelotas: UFPEL, 2003. p. 33-49.

- | Abordagem orientada por dados: desafios e aplicações no ensino de LE

LEŃKO-SZYMAŃSKA, A.; BOULTON, A. **Multiple affordances of language corpora for data-driven learning**. Philadelphia: John Benjamins, 2015.

LITTLE, D. Freedom to learn and compulsion to interact: Promoting learner autonomy through the use of information systems and information technologies. In: PEMBERTON, R. *et al.* **Taking control: autonomy in language learning**. Hong Kong: Hong Kong University Press, 1996. p. 203- 218.

_____. Language learner autonomy: some fundamental considerations revisited. **Innovation in language learning and teaching**, v. 1, p. 14-29, 2007.

MARTINS, M. O aprendizado movido por dados na sala de aula de línguas adicionais. In: MARTINS, M.; ZONI, M. **Experiências e reflexões sobre a educação de línguas adicionais**. Macapá: UNIFAP, 2016. p. 107-121.

MCCARTHY, M. Accessing and interpreting corpus information in the teacher education context. **Language Teaching**, v. 41, n. 4, p. 563-574, 2008.

MONZÓN, A. J. B.; FADANELLI, S. B. Leitura de textos especializados anglófonos no ensino técnico: idiosincrasias terminológicas e pedagógicas sob a perspectiva da linguística de corpus. **LínguaTec**, Bento Gonçalves, v. 1, n. 1, p. 1-28, 2016.

MUKHERJEE, J. Corpus linguistics and language pedagogy: the state of the art – and beyond. In: BRAUN, S.; KOHN, K.; MUKHERJEE, J. **Corpus technology and language pedagogy: new resources, new tools, new methods**. Frankfurt: Peter Lang, 2006. p. 5-24.

ORENHA-OTTAIANO, A.; ROCHA, J. M. P. Ensino de inglês como LE e contribuições pedagógicas de um glossário bilíngue de colocações. **Revista Signótica**, v. 27, n. 2, p. 485-509, 2015.

O’SULLIVAN, I. Enhancing a process-oriented approach to literacy and language learning: the role of corpus consultation literacy. **ReCALL**, v. 19, n. 3, p. 269-286, 2007.

SANTOS, A. G. dos. **Working closely with corpora**: proposta de ensino de colocações adverbiais em inglês para negócios, sob a luz da Linguística de Corpus. 2011. 166 f. Dissertação (Mestrado em Estudos Lingüísticos e Literários em Inglês) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2011.

SCOTT, M. **WordSmith Tools**: version 7. Stroud: Lexical Analysis Software, 2016.

SEIDLHOFER, B. Key concepts in ELT: english as a lingua franca. **ELT Journal**, v. 59, n. 4, p. 339-341, 2005.

COMO CITAR ESTE ARTIGO: ROCHA, Jean Michel Pimentel. Data-Driven Learning Approach: challenges and applications in the teaching of foreign languages. **Revista do GEL**, v. 14, n. 3, p. 135-153, 2017. Disponível em: <https://revistadogel.gel.org.br/>.

DOI: <http://dx.doi.org/10.21165/gel.v14i3.1897>

Submetido em: 31/07/2017. | **Aceito em:** 09/11/2017.
