

## O USO DE *CORPUS* ESPECIALIZADO NO ENSINO PÚBLICO DE LÍNGUA INGLESA

Maria Carolina Zuppari (PUCSP)

**Resumo:** O objetivo deste artigo é apresentar a Linguística de *Corpus* como abordagem pedagógica adicional no processo ensino-aprendizagem dos estudantes da rede estadual de ensino de São Paulo. Para isso, foi coletado um *corpus* especializado voltado para o ensino de língua inglesa e contendo os gêneros contemplados na Proposta Curricular de São Paulo para Língua Estrangeira Moderna (LEM) no 4º bimestre do 6º ano do Ensino Fundamental – ciclo II. As informações extraídas do *corpus* foram usadas como subsídio para a criação de uma sequência didática baseada nas abordagens *data-driven learning* (Aprendizado Movido por Dados) e multimídia/multigênero. O objetivo da atividade é proporcionar a interação entre estudantes e os gêneros que permeiam as atividades sociais previstas na Proposta Curricular.

**Palavras-chave:** Linguística de *Corpus*; Língua Inglesa; Abordagem Multimídia/multigênero; Ensino público estadual.

**Abstract:** This paper aims to present Corpus Linguistics as an additional pedagogical approach in the teaching and learning process of students in public schools run by the state of Sao Paulo. To this end, a specialized corpus was created to be used in the teaching of English, containing the genres present on the Course Program Proposal of Sao Paulo for Contemporary Foreign Language in sixth grade, specifically the fourth quarter. The information extracted from the corpus served as subsidy for the development of a lesson plan based on the data-driven learning and multimedia/multigenre approaches. The goal of the lesson plan is to foster the interaction between students and the genres that underlie the social activities described in the Course Program Proposal.

**Keywords:** Corpus Linguistics; English; Multimedia/multigenre approach; Public school learning.

## INTRODUÇÃO

Ao longo dos anos, o processo de ensino-aprendizagem que estabelece o professor de língua estrangeira como a figura de autoridade na sala de aula, segundo Oxford (1989, p.10), tem se modificado para a visão do professor colaborador, consultor, e mediador do conhecimento, capaz de reduzir suas tarefas pedagógicas e gerenciais para dar mais autonomia aos estudantes, tornando-os mais independentes. Oxford (1989, p.8) descreve as estratégias de aprendizado desenvolvidas por ela como ferramentas essenciais para o desenvolvimento da competência comunicativa por meio, por exemplo, de planejamentos e ações conscientes em direção a um objetivo. Desse modo, com a utilização de uma ou mais estratégias, ou uma combinação de várias, o estudante passa a ser mais rápido e mais eficiente. Para a autora, a identificação e o uso apropriado da estratégia que melhor atende as necessidades do estudante são fundamentais para o seu desenvolvimento na sala de aula (1989, p.12). Com o advento das Novas Tecnologias de Informação e Comunicação (doravante TICs) estratégias, ferramentas e pedagogias têm sido introduzidas no processo ensino-aprendizagem, alternando também o panorama educacional. Segundo Perrenoud, o que acontece no mundo não pode ser ignorado pelas escolas: “as novas

tecnologias da informação e da comunicação (TIC ou NTIC) transformam espetacularmente não só nossas maneiras de comunicar, mas também de trabalhar, de decidir, de pensar” (2000, p.125).

Como se sabe, no ensino público, existe uma falta de vocabulário de língua inglesa contextualizado com base na Proposta Curricular disponível eletronicamente nas escolas para usufruto dos estudantes e professores. Apesar de a Proposta Curricular do estado de São Paulo prever a inserção da tecnologia no currículo da educação básica, em escolas que possuem laboratório de informática, em geral os professores de línguas não os utilizam para suplementar o aprendizado dos estudantes porque não dispõem do conhecimento técnico ou de ferramentas didáticas que tornem o ensino por meio de novas tecnologias de informação possível (PERRENOUD, 2000, p.131).

Com o crescimento da participação do Brasil no cenário econômico mundial, a demanda por estudantes e profissionais que dominam a língua inglesa tem aumentado exponencialmente. Ao mesmo tempo, o ensino de língua inglesa na rede estadual, apesar dos esforços, não tem acompanhado esta demanda. O desenvolvimento de uma ferramenta adicional de ensino capaz de integrar a sala de

aula às novas tecnologias da informação a fim de relacionar o currículo escolar à construção do conhecimento de mundo é de grande importância científica para os professores e estudantes que fazem parte do ensino público estadual. A criação de um *corpus* de língua inglesa baseado nas disposições da Proposta Curricular do Estado de São Paulo, ou seja, com léxico especializado nas áreas temáticas relacionadas ao propósito de ensino de cada ano escolar visa suprir essa falta de vocabulário contextualizado e de material didático baseado na Proposta Curricular do estado para ser utilizado nos laboratórios de informática. Além disso, almejamos a inclusão digital e social dos estudantes da rede estadual de ensino por meio da mediação de uma nova abordagem de ensino, a Linguística de *Corpus* (doravante LC).

De acordo com Berber Sardinha (2004, p.19), a construção de um *corpus* pressupõe a utilização de linguagem autêntica e a rigorosa seleção dos textos de acordo com critérios de autenticidade, naturalidade e representatividade, tornando possível derivar informações de frequência de uso dessas palavras, suas co-ocorrências e colocações. Portanto, a LC permite ao professor trabalhar com a linguagem autêntica, real, na sala de aula. A criação de um *corpus* especializado permite que tanto professores quanto estudantes tenham

acesso a fraseologias, colocações e co-ocorrências presentes em uma área específica, ou temática, da língua inglesa.

Para esse estudo, foram selecionados e coletados textos para a composição de um *corpus* com base nas estruturas e léxico previstos pela Proposta Curricular do estado de São Paulo para o ensino de LEM-Inglês no 4º bimestre do 6º ano do ciclo II do Ensino Fundamental. Escolheu-se o 6º ano por ser o primeiro ano do ciclo II do Ensino Fundamental no qual a língua inglesa passa a ser disciplina do currículo escolar estadual. A escolha do 4º bimestre foi devida a metodologia imposta pela LC. Os bimestres 1 e 2 tratam do aspecto oral da língua inglesa, enquanto o 3º bimestre introduz o léxico da língua escrita. Portanto, no 4º bimestre o estudante começa a ser exposto às estruturas léxico-gramaticais da língua escrita. A extração de termos, listas de frequência e concordância foi feita no software WordSmith Tools Versão 5.0 (SCOTT, 1999). A partir das informações extraídas do *corpus* coletado, uma aula-piloto foi elaborada com base nas abordagens direta por meio do *data-driven learning* (Aprendizado Movido por Dados – doravante DDL) e multimídia/multigênero, e indireta (conteúdos procedimentais e atitudinais), com a finalidade de ser aplicada futuramente em estudantes do 6º ano do Ensino Fundamental II nas escolas da rede estadual

de ensino de São Paulo. Espera-se que os estudantes se beneficiarão do vocabulário contextualizado oferecido pelo *corpus* nas atividades a serem realizadas, levando à autonomia dos estudantes e, segundo Perrenoud (2000, p.131), à construção da competência lógica, epistemológica e didática dos professores.

### **O CONTEXTO ESCOLAR E A PROPOSTA CURRICULAR DO ESTADO DE SÃO PAULO**

A Proposta Curricular do Estado de São Paulo contém os princípios norteadores para a prática educativa nas escolas da rede estadual de ensino. Além disso, conta com o documento *Orientações para a Gestão do Currículo*, o qual propõe a interdisciplinaridade, estimula a interação cultural na escola e o fortalecimento da interação escola-comunidade, e a elaboração do Projeto Pedagógico de cada escola da rede estadual de ensino. A Proposta Curricular conta também com os *Cadernos do Professor*, que são organizados por bimestre e por disciplina. São nesses cadernos que as situações de aprendizagem são apresentadas como orientação no ensino dos conteúdos disciplinares.

A Proposta defende a ideia de que universalizar a escola não é suficiente, a relevância da aprendizagem deve ser universal. Desse modo, é necessário que a educação esteja

voltada para o desenvolvimento pessoal dos estudantes, juntamente com a construção de ideais de identidade, autonomia e liberdade. Essa autonomia está relacionada à capacidade dos estudantes de gerenciar a própria aprendizagem, ou seja, ser capaz de aprender a aprender. Agindo com autonomia, estabelecendo uma identidade e aceitando as diversidades formamos a base para a formação do estudante como cidadão.

A presença da tecnologia na educação está prevista na Lei de Diretrizes e Bases da Educação Nacional (Lei nº 9.394/96). A Proposta Curricular mostra a inserção da tecnologia de dois modos complementares: como educação tecnológica básica e como compreensão dos fundamentos científicos e tecnológicos da produção.

A inserção da tecnologia como educação tecnológica básica refere-se à alfabetização tecnológica, a qual irá preparar os estudantes para o mundo no qual a tecnologia está presente em todos os lugares. A compreensão dos fundamentos científicos e tecnológicos da produção implica em “relacionar o currículo ao mundo da produção de bens e serviços, ou seja, aos processos pelos quais a humanidade – e cada um de nós – produz os bens e serviços de que necessita para viver” (SECRETARIA DA EDUCAÇÃO, p.23).

A área de Linguagens, Códigos e suas Tecnologias abrange as seguintes disciplinas, tanto no Ensino Fundamental quanto no Ensino Médio: Língua Portuguesa, Língua Estrangeira Moderna (LEM), Arte e Educação Física. A definição de linguagem para essa área, de acordo com os Parâmetros Curriculares Nacionais (PCN 2006) “é a capacidade humana de articular significados coletivos em sistemas arbitrários de representação, que são compartilhados e que variam de acordo com as necessidades e experiências da vida em sociedade”. Isto é, “a principal razão de qualquer ato de linguagem é a produção de sentido” (PCN 2006).

Para a disciplina de LEM, é importante criar a consciência linguística e a consciência crítica dos usos da língua estrangeira para que os estudantes possam acessar os bens culturais da humanidade (SECRETARIA DA EDUCAÇÃO, p.38). Para isso, deve-se focar a organização textual e o uso da linguagem em diversas situações de comunicação. A absorção do conhecimento é atingida mais facilmente quando ele é contextualizado. Essa contextualização pode acontecer em três níveis: (a) contextualização sincrônica, a qual ocorre num mesmo tempo e avalia algo em relação à época e à sociedade que o gerou; (b) contextualização diacrônica, a qual ocorre ao longo do tempo considerando

um objeto cultural no eixo do tempo; e a contextualização interativa, que permite a relação do texto com o universo específico do leitor.

A disciplina de Língua Estrangeira Moderna (doravante LEM) é decisiva para que o estudante possa conhecer diferentes meios de sentir, viver e se expressar. Portanto, a finalidade de LEM é construir a competência discursiva do indivíduo a partir de uma perspectiva pluricêntrica, voltada para a diversidade linguística dos diferentes povos. Assim, por meio de experiências com diversos grupos sociais e diferentes comunidades discursivas, torna-se possível promover a definição da identidade linguística e cultural dos estudantes.

A orientação metodológica para o trabalho com LEM na atualidade segue os princípios dos letramentos múltiplos, o qual se baseia nas relações entre o saber e o fazer em múltiplas linguagens e gêneros discursivos. Esse método de trabalho muda significativamente o conceito de conteúdo no ensino de LEM, eliminando o foco na gramática e funções comunicativas para privilegiar “o conhecimento e o reconhecimento de si e do outro, traduzido em diferentes formas de interpretação do mundo, concretizadas nas atividades de produção oral e escrita, desenvolvidas em cada uma das etapas da escolarização” (SECRETARIA DA EDUCAÇÃO, p. 43).

## **A TECNOLOGIA NA SALA DE AULA E A LINGUÍSTICA DE CORPUS**

Segundo Masetto (2000, p.139), não se pode falar em tecnologia e educação sem falar em processo de aprendizagem, pois a tecnologia se faz presente como um instrumento no desenvolvimento desse processo. O autor (2000, p.139-142) descreve o processo de aprendizagem como uma ação diretamente ligada ao aprendiz, pela qual ele é capaz de crescer e se desenvolver. Nesse processo, o professor pode assumir o papel de facilitador ou mediador da aprendizagem do estudante, aquele que incentiva e motiva o aprendiz em sua aprendizagem. Desse modo, o estudante participa ativamente do processo, levando a transformações em seu comportamento. Há também uma mudança de atitude no professor, que, apesar de assumir o papel de especialista em algumas ocasiões, na maioria das vezes trabalhará como facilitador da aprendizagem e orientador ou motivador das atividades do estudante, resultando na verdadeira mediação pedagógica. Nesse caso, o conceito de aprendizagem é relacionado ao “conceito de desenvolvimento do ser humano como um todo, em suas diferentes áreas: área de conhecimento, de sensibilidade, de competências e de atitudes ou valores” (MASETTO, 2000, p.140).

O foco não está mais em utilizar alguns recursos audiovisuais para dar a mesma aula expositiva, mas escolher tecnologias adequadas e variadas que compreendam o desenvolvimento intelectual e afetivo dos estudantes, além de competências e atitudes. Nesse sentido, para Masetto (2000, p.144) enfatizar o processo de aprendizagem significa encontrar diferentes estratégias para motivar o estudante, respeitando os ritmos e formas diferentes de aprendizagem, pois os estudantes aprendem a partir de ritmos e de forma diferentes. As TICs também deverão ser utilizadas para dar valor à autoaprendizagem e no desenvolvimento da interaprendizagem, isto é, “a aprendizagem como produto das inter-relações entre as pessoas” (MASETTO, 2000, p.153-4), nesse caso, os estudantes.

Nesse sentido, a LC aparece como uma abordagem que pode auxiliar na inclusão digital e social dos estudantes da rede estadual de ensino ao permitir que professores e estudantes trabalhem com a linguagem real, autêntica, na sala de aula, por meio do estudo e análise da léxico-gramática contextualizada em determinadas situações sociais.

A LC, de acordo com Berber Sardinha, “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” (2004, p.3). Portanto, a LC enquadra-se dentro do conceito formado pela abordagem empirista com a visão da linguagem como sistema probabilístico. Nesse caso, utilizar a abordagem empirista significa priorizar os dados oriundos da observação da linguagem, em contrapartida a posição racionalista da linguagem. Biber (2010) define a LC como uma descrição da linguagem em uso generalizável, quantitativa e que pode ser individualizada. De acordo com o autor (1998, p.4), um *corpus* é uma coletânea grande e criteriosa (*principled*) de textos de linguagem natural. Isso significa que uma coleção de textos qualquer não perfaz um *corpus*.

Para Berber Sardinha, “a visão da linguagem como sistema probabilístico pressupõe que, embora muitos traços linguísticos sejam possíveis teoricamente, não ocorrem com a mesma frequência” (2004, p.30). Isso significa que embora muitas combinações e características da estrutura da língua sejam possíveis, não ocorrem todas com a mesma frequência. Além disso, essas combinações não são aleatórias, pois há uma correlação entre características linguísticas e contextos de ocorrência da linguagem, ou seja, grupos de traços linguísticos podem variar de acordo com

a situação comunicativa (BERBER SARDINHA, 2004). Por não ser aleatória, essa variação indica que há uma padronização da linguagem (*patterned language*), evidenciada pela recorrência, isto é, colocações, coligações ou estruturas que se repetem de modo significativo parecem ser na realidade padrões lexicais ou léxico-gramaticais. Nesse caso, o ensino da língua por meio de padrões léxico-gramaticais (colocações, coligações e agrupamentos de palavras) de acordo com situações comunicativas específicas traz aos estudantes a possibilidade de construir o conhecimento da língua que permitirá que a interação bem-sucedida em as diversas situações comunicativas e sociais. Portanto, partimos do pressuposto que a inserção da LC como base para uma abordagem de ensino de língua estrangeira beneficiará a formação dos estudantes do 6º ano do ciclo II do Ensino Fundamental.

## **METODOLOGIA**

Conforme supracitado, um *corpus* especializado foi compilado para esse estudo, baseado nas estruturas e no léxico previstos pela Proposta Curricular do estado de São Paulo para o ensino de LEM-Inglês no 4º bimestre do 6º ano do ciclo II do Ensino Fundamental, ano em que a língua inglesa passa a ser disciplina do currículo escolar

estadual. A criação de listas de palavras por frequência e a extração de agrupamentos de palavras, bem como linhas de concordância foram feitas com o auxílio do software WordSmith Tools Versão 5.0 (SCOTT, 1999). A partir das informações extraídas do *corpus* coletado, um plano de aula foi elaborado com base nas abordagens direta por meio do *data-driven learning* (Aprendizado Movido por Dados – doravante DDL) e multimídia/multigênero. Espera-se que o plano de aula possa ser utilizado com estudantes do 6º ano do Ensino Fundamental II nas escolas da rede estadual de ensino de São Paulo.

### **COMPILAÇÃO DO *CORPUS* E EXTRAÇÃO DE DADOS**

O planejamento e a compilação do *corpus* especializado utilizado nesse trabalho seguem critérios linguísticos específicos. Conforme descrito anteriormente, o conteúdo do *corpus* reflete o tema da Proposta Curricular do estado de São Paulo para LEM – Língua Inglesa, referente ao 6º ano do Ensino Fundamental – ciclo II, que estabelece a “Aquisição de repertório lexical e reconhecimento de estruturas”. Em específico, compilamos um *corpus* especializado para a linguagem relativa ao 4º bimestre do 6º ano, sob o tema “Diferentes moradias”, que aborda os seguintes aspectos, incluindo gêneros e produção:

- Denominação de diferentes tipos de moradia
- Relação entre ilustração e descrição de diferentes tipos de moradia
- Denominação de espaços de uma casa e dos itens de mobília mais comuns
- Adjetivos usados para descrever casas e seus espaços

### **Gêneros para leitura e escrita**

- Descrições de diferentes moradias, de plantas baixas de empreendimentos imobiliários etc.

**Produção:** planta baixa de uma casa contendo itens de mobília, com os cômodos e móveis identificados (SECRETARIA DA EDUCAÇÃO, p.45)

De acordo com O’Keeffe (2010), um *corpus* especializado permite um equilíbrio entre as análises quantitativa e qualitativa, sendo possível encontrar sutilezas que não necessariamente apareceriam em um *corpus* maior. Isto é, em um *corpus* especializado, de dimensões menores, pode-se encontrar maior concentração de padrões da língua, padrões de uso localizados e especialização pragmática. Nesse caso, há maior concentração de cobertura do registro de linguagem sendo estudado e, maior consciência do contexto de uso por parte do usuário do *corpus*.

De acordo com Atkins (1992, p.3), o que difere o *corpus* sendo compilado de qualquer outra coletânea de textos computadorizados é justamente o interesse linguístico em particular. Nesse trabalho, o objetivo linguístico é compilar textos contendo léxico especializado em relação ao tema “diferentes moradias”.

Além disso, Atkins (1992, p.4) argumenta que a simples existência de um *corpus* não é suficiente para levantar dados linguísticos. Para isso, é necessária a utilização de um grupo de ferramentas de análise para o seu processamento. Algumas dessas ferramentas básicas são: concordâncias (ou KWIC – *key word in context*), listas de frequência de palavras e listas de colocados extraídas com a utilização de softwares específicos para esse fim como o WordSmith Tools<sup>1</sup> (SCOTT, p.1999), o qual foi usado nesse trabalho na versão 5.0. Para a extração das listas de frequência e concordâncias foram utilizadas as ferramentas *WordList* e *Concord*.

O *corpus* especializado compilado para esse trabalho é monolíngue e composto por dois gêneros: anúncios de imóveis e plantas baixas, os quais atendem ao disposto pela Proposta Curricular para o tema “Diferentes moradias”. Foram selecionados quatro jornais americanos disponíveis na

---

1 O software WordSmith Tools foi desenvolvido por Mike Scott e teve a primeira versão publicada em 1996. Atualmente está na versão 5.0.

Internet para a captura dos anúncios de imóveis: *San Francisco Chronicle*, *The Dallas Morning News*, *The New York Times* e *The Seattle Times*. Os jornais foram escolhidos por representarem diferentes partes geográficas dos Estados Unidos (Sudoeste, Sul, Nordeste e Noroeste), e, portanto, serem passíveis de refletir diferentes registros de linguagem. Além disso, foram selecionados jornais em que os anúncios aparecessem na página da Internet do jornal em si, e não fossem direcionados a páginas de imobiliárias individuais. Para o conteúdo de plantas baixas, foi selecionado o website *Architectural Designs* (<http://www.architecturaldesigns.com>) por se encaixar no perfil necessário e devido à grande quantidade e variedade de plantas baixas com descrição disponíveis. Os textos foram todos copiados e colados para o formato .txt, exigido para o processamento do *corpus* pelo software WordSmith Tools.

O *corpus* contém 100 (cem) anúncios de imóveis de cada jornal selecionado, totalizando 400 (quatrocentos) anúncios, e 240 (duzentos e quarenta) descrições de plantas baixas, divididas em quatro categorias contendo 60 (sessenta) descrições cada uma: *Beach House Plans*, *Canadian House Plans*, *Cape Cod House Plans* e *Carriage House Plans*. No total foram compilados 640 textos contendo 67.172<sup>2</sup>

---

2 Esse é o número de tokens utilizados para a construção da wordlist (lista de palavras), ou seja, exclui todas as palavras incluídas no cabeçalho do corpus. O número total de tokens (running words) é 69.564.

*tokens* (palavras) e 5.259 *types* (formas). De acordo com Sarmiento (2009, p.269), um *corpus* especializado traz mais palavras de conteúdo entre as mais frequentes. Esse fato foi confirmado em nosso *corpus* de estudo, que apresenta dezesseis palavras lexicais entre as trinta mais frequentes. Além da lista de palavras, foram extraídas listas dos *clusters* (pacotes lexicais) mais frequentes no *corpus*, divididas em *clusters* com 2 palavras e *clusters* com 3 palavras. Linhas de concordância foram geradas para as primeiras dez palavras de conteúdo mais frequentes no *corpus* e para os primeiros 10 *clusters* com 2 e 3 palavras de conteúdo mais frequentes com o intuito de produzir uma base de dados para a construção dos exercícios a serem utilizados no plano de aula. Para a escolha dos *clusters*, foram consideradas apenas combinações contendo palavras de conteúdo. A lista de palavras e *clusters* de conteúdo para as quais foram geradas linhas de concordância estão descritas na Figura 1.

Linhas de concordâncias foram feitas em conjunto para as palavras *bedroom/bedrooms* e *walk in closet/walk in closets*. Os *clusters* formados por nomes próprios não foram considerados para a compilação da lista. Utilizamos a frequência de palavras de conteúdo e *clusters* como base estatística para a confecção da sequência didática. Outra opção seria utilizar os resultados

da análise estatística de associação, como a Informação Mútua Específica, que mede a força de atração entre um nóculo e um colocado por meio de probabilidade logarítmica. Porém, por questão de tempo hábil para a realização da análise de dados, optamos por utilizar a lista de *clusters* mais frequentes.

### **APLICAÇÃO DO CORPUS: DATA-DRIVEN LEARNING E ABORDAGEM MULTIMÍDIA/MULTIGÊNERO**

Um exemplo de abordagem de Informática Aplicada ao Ensino de Línguas é o *data-driven learning* (Aprendizado Movido por Dados), ou DDL, que inclui o uso de *corpora* nas tarefas aplicadas no processo de ensino-aprendizagem. O’Keeffe (2007, p.24) explica que em DDL, o estudante utiliza o *corpus* em primeira mão por meio de tarefas guiadas ou materiais didáticos baseados em *corpus*, como, por exemplo, exercícios contendo linhas de concordância. Chambers e Kelly mostram que o contexto pedagógico de DDL traz, de uma só vez, as teorias construtivistas de aprendizado, a abordagem de ensino comunicativa e o desenvolvimento de autonomia no estudante (O’KEEFFE, 2007, p.24), conceito este que se encaixa harmoniosamente com o do Letramento Múltiplo da Proposta Curricular do estado de São Paulo.

A principal vantagem da utilização de DDL na sala de aula, de acordo com Granger (2010, p.359), é a autenticidade do

material exposto aos estudantes de forma compacta, o que os auxilia na expansão de vocabulário e na identificação de padrões da língua. De acordo com Granger, “[...] os estudantes são encorajados a observar os dados do *corpus*, construir hipóteses e formular regras a fim de ter uma visão da língua (abordagem indutiva) ou verificar a validade das regras da gramática que conhecem ou de seus livros (abordagem dedutiva)” (2010, p.359, tradução do autor). Desse modo, o processo ensino-aprendizagem se torna mais autônomo, com a participação ativa e engajada por parte dos estudantes. De acordo com Berber Sardinha, as vantagens da DDL causam impacto nos estudantes, nos professores e no processo de ensino-aprendizagem da seguinte forma (2004, p.292-293):

1. Os estudantes se tornam pesquisadores e assumem papel central e independente na sua aprendizagem ao serem capazes de identificar regularidades da língua (padronizações) e fazer generalizações para explicar essas regularidades;
2. O professor se torna um orientador ou facilitador do ensino e o computador passa a ser fonte de informação;
3. O ensino passa a ser centrado no estudante, com a mudança de foco da transmissão de regras para o princípio da descoberta.

Granger (2010, p.366-367) cita algumas das dificuldades de se trabalhar com a abordagem DDL na sala de aula em relação à logística do método, aos estudantes e aos professores. O uso de *corpora* no ensino exige computadores, acesso a um *corpus* e concordanceadores, o que pode ser uma barreira no desenvolvimento de materiais didáticos. Geralmente os professores precisarão compilar seu próprio material de DDL, e para tanto, devem estar aptos a essa atividade, com conhecimento de *corpora* e suas aplicações no ensino. Outra desvantagem é que tanto a criação quanto a aplicação de material didático de DDL levam tempo para serem feitas. Apesar dessas dificuldades, Berber Sardinha acredita que

“[...] a DDL pode, de fato, contribuir para o ensino de línguas. O conjunto de pressupostos em que a abordagem se baseia e a maneira pela qual é colocada em prática podem enriquecer o processo de ensino-aprendizagem e desenvolver habilidades fundamentais no estudante”. (2004, p.295).

Segundo Berber Berber Sardinha (2010, p.309-310), a preparação de atividades na abordagem DDL precisa considerar diversas variáveis, dentre elas:

a. A escolha da palavra de busca ou palavra nódulo, que pode ser individual ou um colocado;

- b. O modo de apresentação da aula, se no laboratório de informática ou na sala de aula convencional;
- c. O tipo de concordância a ser produzida;
- d. A quantidade de concordâncias a serem apresentadas ao estudante; e
- e. Se haverá ou não exercícios para acompanhar as concordâncias.

Dois recursos são indispensáveis para a construção de exercícios baseados na abordagem DDL: um *corpus* e um programa concordanceador para explorar esse *corpus* (GRANGER, 2010). No nosso caso, o *corpus* é o compilado de acordo com as características descritas nesse artigo e o programa concordanceador utilizado foi o WordSmith Tools versão 5.0. Este é um software pago, no entanto, há a opção de se usar o software gratuito AntConc (<http://www.laurenceanthony.net/software/antconc/>).

Antes de utilizar atividades de DDL na sala de aula, é necessário explicar aos estudantes como ler e interpretar uma concordância. Essa leitura não deve ser realizada como a leitura de um texto corrido, ao contrário, deve-se focar a área central da concordância, marcada pela palavra nóculo, e as palavras à direita e esquerda da palavra nóculo. Em seguida, deve-se focar como extrair os padrões lexicogramaticais das

concordâncias observadas. Para isso, o professor ajuda os estudantes a identificar o que é um padrão e como ele pode ser encontrado (BERBER SARDINHA, 2010).

Além das atividades de DDL, utilizamos uma nova abordagem para a aplicação da LC no processo ensino-aprendizagem proposta por Berber Sardinha (2010), as atividades multimídia/multigênero. De acordo com o autor, existem inúmeros gêneros, subgêneros e mídias atualmente e esse número tende a crescer cada vez mais. Além disso, há a tendência de se tornarem mais complexos com a transição das mídias analógicas para as digitais e com o surgimento e desaparecimento de novas tecnologias. É nesse contexto que Berber Sardinha sugere essa nova abordagem para materiais de ensino baseados em *corpus*:

[...] os estudantes e professores estarão cada vez mais expostos a esses gêneros (seja como produtores ou receptores), ao mesmo tempo em que tais gêneros se tornam parte de seu dia a dia, no lazer, em casa e no trabalho. A sala de aula precisa, então, capacitar os estudantes a terem bom desempenho no contato com esses gêneros, sob pena de ficar ultrapassada e, como já sabemos, acusada de ser um lugar anacrônico e desestimulante; sem falar que pode ser questionada em relação a constituir um local de formação de cidadãos capazes de enfrentar os desafios de uma sociedade cada vez mais globalizada, digital e em transformação. (2010, p.326-327)

Segundo Berber Sardinha (2010), ainda não há disponibilização de *corpora* de grande parte dos gêneros que surgem com a era digital, e sempre haverá significativo espaço de tempo entre o surgimento de um gênero, sua utilização em massa e a disponibilidade de um *corpus* que o contenha. No entanto, a quantidade de arquivos em formato digital em rede facilita a criação de *corpora* por parte dos praticantes da LC no ensino para seus próprios propósitos, como é o nosso caso.

Para o autor (2010), o objetivo final da atividade multimídia/multigênero não é tornar os estudantes especialistas nos gêneros, mas sim permitir que possa haver interação, mesmo que leve, entre o estudante e os gêneros que permeiam a atividade social em foco, por meio de elementos linguísticos que possibilitem desempenhar a atividade. Isto é, essa proposta de abordagem de ensino “visa capacitar os estudantes a lidar com (produzir e/ou receber) a linguagem envolvida em determinada atividade social” (2010, p.329).

### **SEQUÊNCIA DIDÁTICA: PLANO DE AULA**

Conforme descrito acima, a Proposta Curricular do estado de São Paulo para LEM – Inglês foca a construção da competência discursiva do estudante a partir de uma perspectiva pluricêntrica, seguindo os princípios dos

letramentos múltiplos e da pluralidade de gêneros discursivos. Por isso, utilizaremos a abordagem multimídia/multigênero e DDL para a construção das atividades didáticas que poderão ser realizadas em sala de aula. O objetivo educacional deste plano de aula é o estudante e desenvolver nele a consciência crítica por meio da prática transformadora, respeitando os preceitos da tendência histórico crítica e a integração entre estudante e sociedade.

A sequência didática a ser utilizada foi criada seguindo os critérios estabelecidos por Berber Sardinha (2010) para a abordagem multimídia/multigênero. A abordagem DDL foi incorporada nos exercícios com concordância e excertos do *corpus*. Os critérios da abordagem multimídia/multigênero são: (a) Modularidade: o material é dividido em partes distintas que podem ser preparadas e aplicadas separadamente; (b) Replicabilidade: os enunciados são gerais e podem ser aplicados a outros materiais ou a outras atividades sociais. O objetivo é reutilizar o material com poucas adaptações, diminuindo o tempo gasto na sua preparação; (c) Curso: o material foi criado para ser usado no 4º bimestre do 6º ano do Ensino Fundamental – ciclo II, nas escolas estaduais de São Paulo; e (d) Conhecimento Prévio: os estudantes devem ter conhecimento prévio do que é um gênero textual, atividade

social, e dos conceitos da LC, como *corpus*, frequência de palavras, *clusters* ou agrupamentos de palavras, linhas de concordância, etc. Por isso, é fundamental iniciar o plano de aula com uma explicação sobre esses conceitos.

Optamos por manter o enunciado de todos os exercícios em língua portuguesa, pois se supõe que o estudante terá menos dificuldades para completar as atividades ao se concentrar no conteúdo da aula e não nas instruções.

A sequência didática do plano de aula apresenta uma introdução, seguida de quatro seções: 1. Foco na atividade social (*focus on the social activity*); 2. Foco no gênero textual (*focus on the genre*); 3. Foco no texto (*focus on the text*); e 4. Foco no *corpus* (*focus on the corpus*). Em cada parte, foram criadas atividades relacionadas entre si a fim de criar um continuum entre as partes. Optamos por manter o enunciado de todos os exercícios em língua portuguesa, pois se supõe que o estudante terá menos dificuldades para completar as atividades ao se concentrar no conteúdo da aula e não nas instruções.

Na introdução da sequência didática, é discutido o que é um *corpus* e qual é sua finalidade e são apresentados os conceitos de linguagem autêntica disponível em *corpora*, leitura de linhas de concordância, extração de padrões. É importante explicar

aos estudantes como ler as linhas de concordância: primeiro, verticalmente, prestando atenção à palavra nóculo e, em seguida, observando as palavras ao redor do nóculo para achar padrões léxico-gramaticais presentes. O slide abaixo mostra como linhas de concordância podem ser mostradas aos estudantes (Figura 2).

O foco na atividade social compreende o uso de diferentes gêneros associados a uma atividade específica. Essa seção traz atividades voltadas para a identificação de ações e situações sociais e comunicativas relacionadas ao tema da moradia (Figura 3). Espera-se que os estudantes consigam identificar alguns dos gêneros associados aos diferentes tipos de moradia e empreendimentos imobiliários, bem como descrever ambientes e mobílias.

No foco no gênero textual, o estudante é guiado ao contexto de uso de cada gênero que está sendo estudado, sobretudo o conjunto físico e social de produção e recepção dos gêneros discursivos que circulam socialmente (Figuras 4 e 5). Já o foco no texto traz excertos de textos que exemplificam o gênero estudado, principalmente os elementos linguísticos e visuais presentes em cada excerto (Figura 6).

O objetivo da parte foco no *corpus* é explorar o *corpus* especializado criado para a confecção dos exercícios desse

plano de aula. Na realização dos exercícios para essa parte, foi enfatizado o trabalho com os padrões léxico-gramaticais existentes nos gêneros textuais sendo analisados por meio de linhas de concordância e listas de frequência de palavras, como pode ser visto nas Figuras 7 e 8.

## CONSIDERAÇÕES FINAIS

Partindo da premissa de que a Linguística de *Corpus* como base para uma abordagem de ensino de língua estrangeira diferenciada seria benéfica para a formação dos estudantes da rede estadual de ensino de São Paulo, buscamos criar um *corpus* especializado baseado nos temas da Proposta Curricular para o 4º bimestre do 6º ano do Ensino Fundamental – ciclo II. O objetivo da criação desse *corpus* foi o de possibilitar a inserção das TICs no processo ensino-aprendizagem dos estudantes da rede estadual de ensino de São Paulo por meio do acesso a um conteúdo lexical especializado para LEM – Inglês, tornando-os mais críticos e autônomos em relação ao seu aprendizado. A sequência didática elaborada a partir do *corpus* coletado pode ser aplicada tanto no laboratório de informática quanto na sala de aula presencial, por meio de exercícios baseados nas abordagens DDL e multimídia/multigênero. As abordagens DDL e multimídia/multigênero permitem

que, com a quantidade de dados disponíveis no *corpus*, os professores possam gerar exercícios moldados pela temática da Proposta Curricular para o trabalho completo de docência ao longo do 4º bimestre do 6º ano.

Uma das limitações desse estudo é ter somente dois gêneros contemplados no *corpus*. Outra limitação, devido ao tempo disponível para pesquisa, foi a escolha por utilizar a frequência observada como base estatística para seleção dos pacotes lexicais a serem estudados. Seria benéfico realizar estudos adicionais com estatísticas de associação para verificação da co-ocorrência entre a palavra nódulo e seus colocados, como, por exemplo, a Informação Mútua Específica ou T-score, permitindo assim o trabalho com outras listas de colocações.

Além disso, o *corpus* especializado criado é específico para uso no 4º bimestre do 6º ano. Possibilidades de pesquisas futuras seriam primeiramente aumentar o número de palavras do *corpus* compilado para esse estudo de modo a torná-lo mais representativo, e em seguida, a criação de um *corpus* de estudo contendo os temas e gêneros textuais para utilização nos outros anos do Ensino Fundamental – Ciclo II.

## REFERÊNCIAS

- ATKINS, Sue; CLEAR, Jeremy; OSTLER, Nicholas (1992). “Corpus Design Criteria”. In: *Literary and Linguistic Computing*, 7(1): 1-16.
- BERBER SARDINHA, Tony (2011). “Como usar a linguística de corpus no

ensino de língua estrangeira por uma linguística de corpus educacional brasileira”. In: VIANA, Vander; TAGNIN, Stella E.O. (Orgs.). *Corpora no ensino de línguas estrangeiras*. São Paulo: HUB Editorial, 2010, p. 301-356.

\_\_\_\_\_ (2004). *Linguística de Corpus*. São Paulo: Manole.

BERBER SARDINHA, Tony; ALMEIDA, Gladys Maria de B. (2008). “A Linguística de *Corpus* no Brasil”. In: TAGNIN, Stella E.O.; ARAÚJO VALE, Oto (Orgs.). *Avanços da Linguística de Corpus no Brasil*. São Paulo: Humanitas, p. 17-40.

BIBER, Douglas (2010). *Merging corpus linguistics and discourse analytic research goals: discourse units in biology research articles*. Trabalho apresentado em palestra na Universidade de São Paulo, São Paulo.

DAVIES, Mark (2008-). *The Corpus of Contemporary American English (COCA): 520+ million words, 1990-present*. In <http://corpus.byu.edu/coca/> Acesso em 20.Jul.2017.

GRANGER, Sylviane; GILQUIN, Gaëtanelle (2010). “How can DDL be used in language teaching?” In: O’KEEFFE, Anne; MCCARTHY, Michael (Orgs.), *The Routledge Handbook of Corpus Linguistics*. London: Routledge, p. 359-370.

GRANGER, Sylviane (2002). “A Bird’s-eye view of learner *corpus* research.” In: GRANGER, Sylviane; HUNG, Joseph; PETCH-TYSON, Stephanie (Orgs.), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam and Philadelphia: Benjamins, p.3-33 (Language Learning & Language Teaching, v.6).

MASETTO, Marcos T.; MORAN, José Manuel; BEHRENS, Marilda Aparecida (2000). *Novas tecnologias e mediação pedagógica*. 15.ed. Campinas: Papirus, p.133-173.

O’KEEFFE, Anne; MCCARTHY, Michael; CARTER, Ronald. (2007). *From Corpus to Classroom: language use and language teaching*. Cambridge: Cambridge University Press.

O’KEEFFE, Anne (2010). *Corpora, Pragmatics and Language Teaching*. Trabalho apresentado no IX Encontro de Linguística de Corpus, Porto Alegre.

OXFORD, Rebecca L. (1989). *Language Learning Strategies: What Every Teacher Should Know*. Wadsworth Publishing Company.

PERRENOUD, Philippe (2000). *10 novas Competências para ensinar*. Porto Alegre: Artes Médicas.

\_\_\_\_\_ (2002). “A formação dos professores no século XXI”. In: PERRENOUD, Philippe; THURLER, Monica Gather et al. (Orgs.). *As competências para ensinar no século XXI: a formação dos professores e o desafio da avaliação*. Porto Alegre: Artmed, p. 11-34.

São Paulo (Estado). SECRETARIA DA EDUCAÇÃO. *Proposta Curricular do Estado de São Paulo – LEM: Língua Estrangeira Moderna – Inglês: Ensino Fundamental – Ciclos I e II e Ensino Médio*, 2011. In <http://www.educacao.sp.gov.br/a2sitebox/arquivos/documentos/237.pdf> Acesso em 20.Jul.2017.

SARMENTO, Simone (2009). “Linguística de *corpus* e o desenvolvimento de material didático para Inglês”. In: SARMENTO, Simone; FREITAS, Ana Luiza P. de (Orgs.). *O ensino do Inglês como língua estrangeira: estudos e reflexões II*. Porto Alegre: ediPUCRS.

SCOTT, Mike (1999). *WordSmith Tools version 5*, Oxford: Oxford University Press.

SCOTT, Mike; TRIBBLE, Christopher (2006). *Textual Patterns: Key words and corpus analysis in language education*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

**Maria Carolina Zuppardi**. Mestre em Linguística Aplicada e Estudos da Linguagem pela Pontifícia Universidade Católica de São Paulo (PUCSP). Doutoranda em Linguística Aplicada e Estudos da Linguagem pela Pontifícia Universidade Católica de São Paulo (PUCSP), com bolsa CAPES. Atua como Especialista em Ensino Digital do Learning & Innovation Office da Rede Internacional de Universidades Laureate, na área de desenvolvimento de disciplinas para cursos de ensino superior da rede no mundo todo e capacitação docente para o ensino digital. Membro do Grupo de Pesquisa GELC – Grupo de Estudo em Linguística de Corpus (CNPq). E-mail: [carolzupp@gmail.com](mailto:carolzupp@gmail.com) / [carol@corpuslg.org](mailto:carol@corpuslg.org)

## ANEXOS

N	Palavras individuais	Frequência	clusters com 2 palavras	Frequência	clusters com 3 palavras	Frequência
1	ROOM	1005	LIVING ROOM	223	WALK IN CLOSET(S)	105(37)
2	HOME	551	DINING ROOM	192	SINGLE FAMILY HOUSE	86
3	BATH	523	MASTER SUITE	162	FORMAL DINING ROOM	65
4	LIVING	518	SINGLE FAMILY	149	OPEN FLOOR PLAN	42
5	KITCHEN	511	FULL BATH	115	EAT IN KITCHEN	32
6	BEDROOM(S)	445(375)	FAMILY ROOM	110	STAINLESS STEEL APPLIANCES	31
7	FLOOR	439	HOUSE PLAN	99	FLOOR TO CEILING	26
8	HAS	385	MASTER BEDROOM	98	WOOD BURNING FIREPLACE	22
9	FAMILY	372	FAMILY HOUSE	87	TWO CAR GARAGE	21
10	MASTER	361	CAR GARAGE	86	WRAP AROUND PORCH	17

Figura 1. Lista de palavras e *clusters* de conteúdo para os quais foram geradas linhas de concordância.

**Trabalhando com *corpus***

- Esse é um modelo de pesquisa com *corpus*. Buscamos a palavra *bedroom*:
  - 1 Then came their bathroom, and then a **bedroom** that, judging by the photographs on the
  - 2 Avoid attic placement directly above a **bedroom**. ## MOST RECREATION WORK calls
  - 3 (kitchen, dining room, living room, and **bedroom**), with air-conditioning and new modern
  - 4 to be put in. The wallpaper in the back **bedroom** is literally crumbling off "h". (Theresa
  - 5 e blue chair which still faced the closed **bedroom** door. At last, sitting there, in the
  - 6 floor sleeping porch and turn the front **bedroom** into a playroom where it's nice and
  - 7 Mrs& Coolidge spent more time in her **bedroom** among her doll collection. She kept the
  - 8 went up stairs and did the bath and her **bedroom**- way I always do in the morning".
  - 9 beside the door and went upstairs to her **bedroom**. On her bureau lay a small, brass
  - 10 at once and forcibly hauled him into his **bedroom**. Lewis warned him never to lay a hand
- Esse é um exemplo de **linhas de concordância**
- Elas são lidas primeiro, verticalmente, prestando atenção à palavra nóculo. Em seguida, devemos observar as palavras ao redor do nóculo para achar padrões.

81

Figura 2. Slide de introdução: Trabalhando com *corpus*.

### Describing a House

- Em grupos, complete o diagrama abaixo com as atividades e gêneros relacionados à atividade de descrever uma casa. A primeira caixa foi preenchida para você como exemplo. Ao redor, escreva outras atividades sociais e seus respectivos gêneros textuais associados à atividade principal.

### Buying a house or an apartment

- Compartilhe suas ideias com a classe:
  - Em quais situações você tem de descrever uma casa ou um apartamento?
  - Normalmente, quem compra uma casa ou um apartamento?
  - Por que motivo alguém compraria uma casa ou um apartamento?
  - Por que comprar uma casa é importante?
  - Como procuramos uma casa para comprar?

Figura 3. Slides contendo atividades da seção de foco na atividade social (*focus on the social activity*).

### Genre 1: Real Estate Ad

### Genre 1: Real Estate Ad

- Agora vamos focar as questões associadas ao gênero textual *Real Estate Ad*, associado à atividade social. Novamente, trabalhe em pares. Responda essas perguntas:
  - O texto/figura no slide anterior é um exemplo desse gênero textual. Você o reconhece? Você já usou esse tipo de gênero? Com que finalidade?
  - Analise o texto/figura acima e responda essas perguntas:
    - Faz parte de que mídia?
    - Quem tem acesso a ele?
    - Onde foi produzido? Com que finalidade?
    - Quem produz esse tipo de texto/figura?
    - Para quem ele é produzido?
    - Como você pode acessá-lo?
    - Você já produziu algo parecido? O quê?

### Genre 1: Real Estate Ad

- Analise a figura acima e identifique algumas partes, para que servem e qual a informação que elas transmitem:

Parte/Segmento/Ígura	Propósito	Informações transmitidas

Figura 4. Slides contendo atividades da seção de foco no gênero textual - gênero anúncio de imóveis (*focus on the genre - Real estate ad*)

### Genre 2: House Plan

Ground Floor

Second Floor

### Genre 2: House Plan

- Agora vamos focar as questões associadas ao gênero textual *House Plan*, associado à atividade social. Novamente, trabalhe em pares. Responda essas perguntas:
  - O texto/figura acima é um exemplo desse gênero textual. Você o reconhece? Você já usou esse tipo de gênero? Com que finalidade? Que tipo de texto é?
  - Análise o texto/figuras acima e responda essas perguntas:
    - Faz parte de que mídia?
    - Quem tem acesso a ele?
    - Onde foi produzido? Com que finalidade?
    - Quem produz esse tipo de texto/figura?
    - Para quem ele é produzido?
    - Como você pode acessá-lo?
    - Você já produziu algo parecido? O quê?

Adaptado de <http://comistas.ig.com.br/allesinara/taz/arquitetura/>

Figura 5. Slides contendo atividades da seção de foco no gênero textual - gênero planta baixa (*focus on the genre - House plans*)

### Texto: Real Estate Ad

- Acesse o texto online
  - <http://realestate.nytimes.com/sales/detail/2836-4230905/Forest-Hills-Queens-NY-11375>
- Quais são as dicas visuais que o ajudam a entender o texto?
- Há palavras parecidas com a língua portuguesa no texto? Quais são elas e o que elas significam?

### Foco no texto

- House Plan**

Perfect for the growing Family  
Style: Cape Cod, Traditional  
Perfect for the growing family with a smaller building budget, this home plan allows four bedrooms in a narrow lot size. The floor plan is efficient with the living room connecting to the dining area and L-shaped kitchen. Two bedrooms share a full bath on this floor.  
The second floor, which may be finished at a later date, holds two additional bedrooms and a full bath as well. One of these bedrooms has a convenient dressing room, and each contains ample closet space.

  - Quais palavras que fazem parte da figura estão presentes no texto acima?
  - Há palavras parecidas com a língua portuguesa no texto? Quais são elas e o que elas significam?

Figura 6. Slides contendo atividades da seção de foco no texto (*focus on the text*)

