

## CONSCIENTIZAÇÃO LINGUÍSTICA COM BASE EM CORPORA *ONLINE*

Deise P. DUTRA  
(Universidade Federal de Minas Gerais)  
deisepdutra@gmail.com

*RESUMO: Este artigo discute resultados de uma pesquisa desenvolvida em um curso de Inglês intermediário e avançado, que visou a utilizar corpora online, discutir sobre descrição linguística à luz da Linguística de Corpus e promover o desenvolvimento linguístico dos alunos. Os dados apresentados provêm de tarefas para a utilização de corpora online e para a conscientização da língua em uso e questionários, aplicados no início e ao final do curso. Os resultados revelam que os participantes perceberam a importância da descrição de dados reais para a compreensão da língua em uso, aprenderam a usar concordâncias e aprofundaram seus conhecimentos da língua inglesa.*

*PALAVRAS-CHAVE: corpora online; língua em uso; conscientização linguística.*

*ABSTRACT: This article discusses results from a research work developed in an intermediate and advanced English course that aimed at using online corpora, discussing about linguistic description in the light of Corpus Linguistics, and promoting students' linguistic development. The data presented derive from tasks for the use of online corpora and for language use awareness, and questionnaires answered in the beginning and end of the course. The results reveal that the participants have noticed the importance of real data description for the understanding of language use, have learned to use concordance data and have broaden their knowledge of English.*

*KEYWORDS: online corpora; language in use; linguistic consciousness*

### 0. Introdução

O ensino/aprendizagem de língua estrangeira (LE), no caso deste artigo, língua inglesa, pode ter foco na fluência e/ou na acuidade linguística, dependendo do objetivo do curso e de seus alunos. O século XX presenciou um grande número de métodos (DUTRA e MELLO 2004) que oscilavam entre o foco na língua para o uso e o conhecimento de formas linguísticas isoladas (CELCE-MURCIA 1991). A popularização da abordagem comunicativa nos últimos trinta anos ajudou a divulgar a idéia de que o ensino de gramática não deveria ser enfatizado. Todavia, mais recentemente, houve uma

releitura do trabalho com a acuidade dos aprendizes de LEs (LONG 1991; DOUGHTY e WILLIAMS 1998) e, assim, desde o artigo seminal de Canale e Swain (1980), discute-se quais seriam os componentes da competência comunicativa. Para Canale e Swain (1980) e Canale (1983), ela é constituída das competências linguística, discursiva, sociolinguística e estratégica. Pesquisadores como Swain (1991) observaram que mesmo os aprendizes expostos a altos níveis de *input* não atingiam um alto nível geral de proficiência. Como certos itens linguísticos (gramaticais ou sociolinguísticos) podem não ser percebidos com facilidade por aprendizes, a noção de que é preciso promover oportunidades no ensino de LEs para o desenvolvimento da percepção consciente<sup>1</sup> (SCHIMDT e FROTA 1986) foi desenvolvida.

Em contextos de ensino de LE parece haver uma tendência ao foco no ensino direto de gramática (NEVES 1996; CELCE-MURCIA, DORNYEI e THURRELL 1997; DUTRA 2004; DUTRA E MELLO 2004), motivada pelas influências que os professores têm de suas experiências atuais e prévias (como alunos e professores), pelos elementos pedagógicos e pelos fatores contextuais (DUTRA 2004). Mesmo com o aparente predomínio da abordagem comunicativa em nosso país, em certos contextos ela é adotada somente no discurso de professores<sup>2</sup>.

Pesquisas recentes têm procurado investigar como tem sido o ensino de gramática no Brasil (ALMEIDA 2003; ALMEIDA e DUTRA 2005; CARAZZAI 2002; GIL 2004; NÉRI 2006; OLIVEIRA 2006; SOUZA; 2008). O foco exacerbado nas formas parece não contribuir para o desenvolvimento da competência comunicativa dos aprendizes, sendo assim, o ensino de inglês como LE, hoje, tende a caminhar para um equilíbrio entre as atividades que enfatizam o desenvolvimento da fluência e a instrução com foco gramatical (RICHARDS 2002).

Tendo em vista os participantes desta pesquisa (alunos da graduação em Letras da habilitação em Inglês), procurou-se criar alternativas de ensino destinadas ao desenvolvimento da conscientização linguística, como uma habilidade necessária aos bons aprendizes de língua, principalmente, àqueles que serão professores dessa LE. Nesse sentido, a pesquisa ora descrita trata das ações pedagógicas desenvolvidas, com a utilização de corpora eletrônicos, que visavam à melhoria da proficiência linguística desses alunos, futuros profissionais da área de língua inglesa<sup>3</sup>. Alguns dos aspectos tratados foram escolhidos devido às necessidades linguísticas dos alunos.

---

<sup>1</sup> Tradução de Marques (2007) para *noticing* (SCHMIDT e FROTA, 1986).

<sup>2</sup> Ver Dutra (no prelo) para uma discussão sobre o ensino de inglês no Brasil.

<sup>3</sup> É comum no contexto pesquisado que alunos de Letras atuem como professores ou tradutores antes mesmo de se graduarem. No caso desta pesquisa a maior parte dos participantes já trabalhava como professor de inglês.

Como tornar possível uma conscientização linguística que englobe o conhecimento procedimental e declarativo da língua, com uma visão da língua para o uso e, também, como sistema que possibilita a compreensão entre seus usuários, que dê segurança ao professor em suas intervenções na sala de aula e que proporcione a seus alunos uma aprendizagem efetiva? Diferentes ações pedagógicas poderiam ter conduzido a esse objetivo, mas, devido ao perfil dos participantes deste estudo, aprendizes de inglês com domínio intermediário ou avançado, para os quais fluência e acuidade devem caminhar paralelamente, optei por pesquisar os resultados de intervenções pedagógicas com base em corpora eletrônicos no desenvolvimento linguístico dos participantes. Esse foco só foi possível devido aos avanços atuais da Linguística de Corpus.

A Linguística de Corpus é a área que trata dos estudos que utilizam corpora eletrônicos em suas análises. Essa área “é hoje uma das áreas mais vibrantes voltadas ao estudo da linguagem” (BERBER SARDINHA 2004: 42), devido à sua capacidade de descrição das línguas, baseada em amplas quantidades de dados, e que tem levado linguistas a repensarem o uso língua em diversos gêneros e com funções discursivas específicas. Berber Sardinha (2004: 3) aponta que a utilização de corpus para os estudos linguísticos ocorre há muito tempo: “Na Grécia Antiga, Alexandre, o Grande definiu o Corpus Helenístico. Na antiguidade e na Idade Média, produziam-se corpora de citações da Bíblia.”, todavia, o advento do computador trouxe capacidade de armazenamento e exploração de dados não possíveis anteriormente. Um corpus seria, então, uma coletânea de textos escritos e transcrições de fala armazenados em formato eletrônico, que possa ser investigada do ponto de vista linguístico. O corpus serve como registro do uso linguístico de um grupo de falantes, devido ao fato de que a escolha dos textos a serem incluídos segue princípios, tornando possível a identificação do gênero e, em certos casos, do grupo ao qual o participante pertence.

No Brasil é crescente o número de estudos baseados em corpora e ligados ao ensino de línguas. Há, por exemplo, estudos que fazem uso de corpora para aprendizagem (c.f. BEBER SARDINHA 1999; DUTRA, P. B. 2002; VICENTINI 2006), que analisam corpora de aprendizes (c.f. BERBER SARDINHA 2001; LÚCIO 2006; PINTO 2008), que propõem instrumentos de análise de corpus de aprendizes (c.f. BERBER SARDINHA e SHEPHERD 2008), que desenvolvem e analisam corpus de linguagem profissional (c.f. BERBER SARDINHA 2000; PARDO 2004).

O objetivo deste artigo, que une a Linguística de Corpus e a Linguística Aplicada ao ensino de línguas estrangeiras, é descrever e analisar os resultados das tarefas propostas no curso “Estudos Temáticos de Língua e Linguística Inglesa: Corpora e Aprendizagem” que visou familiarizar os alunos com recursos de corpora *online*, discutir sobre descrição linguística à luz da Linguística de Corpus e promover o desenvolvimento linguístico dos participantes.

## 1. Fundamentação teórica

Nesta seção apresentamos a fundamentação teórica deste artigo que se baseia nas seguintes discussões: a centralidade do domínio linguístico para professores de línguas estrangeiras e ações pedagógicas influenciadas pela LC com destaque para o Aprendizado Movido por Dados (JOHNS 1994).

### 1.1 Domínio lingüístico para professores de línguas estrangeiras

O domínio de língua inglesa está no cerne da profissão de professores de inglês e, por isso, cabe avaliar o papel da conscientização linguística na formação desses profissionais. Como ressaltado em Dutra e Mello (no prelo), em seu trabalho sobre o que as pesquisas em linguagem podem mostrar sobre um projeto de educação continuada,

A linguagem, necessária para a organização de nosso pensamento, é crucial para a compreensão do mundo onde estamos inseridos, fazendo-nos entender quem são os professores com os quais interagimos, como atuam e quais caminhos desejam seguir. Além disso, a língua estrangeira, no nosso caso o inglês, é, em sala de aula, o meio de comunicação entre professores e alunos e, também, seu objeto de estudo.

Em muitos contextos, a preocupação com o domínio da língua a ser ensinada por professores de línguas estrangeiras foi colocada em segundo plano em sua formação, devido ao foco central no desenvolvimento de habilidades pedagógicas. Por isso, Edge (1988: 9)<sup>4</sup> parece praticamente pedir desculpas ao argumentar sobre a importância do conhecimento linguístico do professor de línguas estrangeiras:

Minha posição pode parecer muito conservadora em alguns círculos... mas eu quero argumentar que o conhecimento sobre a língua e sobre aprendizagem de língua tem um papel central no treinamento de professores de língua inglesa para falantes de outras línguas.

Como Edge, outros autores passaram a defender abertamente a importância do domínio linguístico na formação de professores de

---

<sup>4</sup> Tradução minha para: *My position on this may seem over-conservative in some circles ... but I want to argue that knowledge about language and language learning still has a central role to play in English language teacher training for speakers of other languages.* (EDGE 1988: 9)

inglês, com reflexos importantes para as ações desenvolvidas na sala de aula. Um dos autores que se dedica a essa questão é Andrews (2001, 2007), ao discutir os efeitos da conscientização linguística do professor (CLP) (*teacher language awareness – TLA*) na prática pedagógica. Segundo Andrews (2007), a CLP estaria diretamente ligada à habilidade comunicativa do professor e ao seu conhecimento pedagógico, uma vez que ela é constituída das competências estratégica e linguística e do conhecimento do conteúdo da disciplina (nesse caso, a própria língua). Os estudos desse autor apontam para o fato de a conscientização linguística do professor afeta o ensino e, conseqüentemente, o tipo de aprendizagem favorecido em sala de aula. Os argumentos do autor defendem que, se ser professor de inglês como língua estrangeira é realmente uma profissão, então certos requisitos devem ser esperados desse profissional e, no cerne desses, está a conscientização linguística.

## 1.2 Aprendizado Movido por Dados

Berber Sardinha (2004) aponta as propostas de ensino de línguas baseadas em conceitos da LC e seus instrumentos, destacando o Currículo Lexical, a Abordagem Lexical e o Aprendizado Movido por Dados. Todas essas propostas foram inovadoras à sua maneira. As duas primeiras por trazerem o léxico para seu eixo central e a última por colocar em evidência a importância do desenvolvimento da habilidade de descoberta do aluno, que o induz a encontrar padrões e regularidades na língua. Segundo Berber Sardinha (2004: 288), há uma diferença crucial entre as duas primeiras propostas:

A diferença (...) é que segundo Lewis, o Currículo Lexical entende o léxico como palavra isolada. Já a Abordagem Lexical enfatiza o caráter composto do léxico por meio das noções de porção (*chunks*), polipalavras (*polywords*) e colocação.

Já o Aprendizado Movido por Dados é “uma das propostas mais sólidas para a utilização de material de corpus na sala de aula” (BERBER SARDINHA 2004: 290). Essa proposta se vale de concordâncias<sup>5</sup>, selecionadas pelo professor, que passam a fazer parte de exercícios indutivos para guiar os aprendizes a descobrirem como certos aspectos linguísticos funcionam. Esses exercícios podem trazer a concordância KWIC (*Key Word in Context*) “que mostra a palavra de busca no centro da listagem ladeada pelas palavras que ocorreram no texto junto a ela” (BERBER SARDINHA 2004: 272), ou

---

<sup>5</sup> “Uma concordância consiste de uma listagem dos cotextos (palavras ao redor) dos quais um dado item (palavra isolada, composta, estrutura, pontuação) ocorre.” (BERBER SARDINHA 2004: 272)

uma concordância sentencial que apresenta uma sentença completa, podendo ser acompanhada de orientações para completar lacunas, para juntar duas partes de uma sentença e diversas perguntas, capazes de levar o aprendiz a notar os padrões da língua. Os exercícios são baseados em concordâncias, mas não apresentam todas as concordâncias que resultam de uma busca em um corpus. Segundo Johns (1994: 297):

A suposição que fundamenta esta abordagem é que a aprendizagem efetiva de uma língua é uma forma de pesquisa linguística e que concordâncias impressas oferecem uma fonte única para a estimulação de estratégias de aprendizagem indutivas – principalmente as estratégias de perceber similaridades e diferenças e de formação e teste de hipótese (Johns 1988).<sup>6</sup>

A língua a que o aprendiz é exposto é autêntica e, ao fazer a construção de campos de sentidos e uso, ele passa a assumir o papel de pesquisador lingüístico. O professor é, nesse caso, acima de tudo, o coordenador da pesquisa. Por isso, ele deve conhecer bem seus alunos e prepará-los para as atividades compostas de perguntas com formato mais restrito ou mais amplo. Berber Sardinha (2004: 277) ressalta que "...os alunos devem ser ensinados a analisar as concordâncias. Isso é sem dúvida o que vai determinar o sucesso da aplicação das concordâncias, pois consiste em permitir que os alunos desenvolvam estratégias de análise de dados."

Devido ao perfil dos participantes desta pesquisa, alunos da graduação em Letras, pareceu propícia a adoção da abordagem do Aprendizado Movido por Dados. Para esses aprendizes, o desenvolvimento de habilidade de descoberta linguística é essencial ao garantir sua capacidade de desenvolvimento autônomo. Dessa maneira, espera-se que a conscientização linguística seja aumentada e colabore para sua aprendizagem da língua inglesa. A seguir, apresento a metodologia do trabalho.

## 2. Metodologia

O contexto da pesquisa é o curso semi-presencial optativo intitulado "Estudos Temáticos de Língua e Linguística Inglesa: Corpora e Aprendizagem" com carga horária de 60 horas. No curso,

---

<sup>6</sup> Minha tradução para "The assumption that underlies this approach is that effective language learning is itself a form of linguistic research, and that the concordance printout offers a unique resource for the stimulation of inductive learning strategies – in particular the strategies of perceiving similarities and differences and of hypothesis formation and testing (Johns 1988)."

foi utilizada a plataforma TelEduc para as atividades não-presenciais. Os primeiros encontros introduziram os alunos a vários corpora *online* e ocorreram no laboratório de informática, disponível para o grupo semanalmente, com a supervisão da professora<sup>7</sup> da turma.

Os participantes eram alunos da graduação em Letras (licenciatura e bacharelado) da habilitação em Inglês, da Faculdade de Letras da Universidade Federal de Minas Gerais. O pré-requisito estabelecido para cursar a disciplina foi a aprovação em Habilidades Integradas I, II e III<sup>8</sup>, o que significava que os alunos cursavam, aproximadamente, o quarto período de Letras. Todavia, devido ao caráter flexível da composição do currículo dessa faculdade, o grupo era formado por alunos de diversos períodos. A maior parte deles estava no sexto período, ou adiante, e tinha nível intermediário ou avançado de inglês. Houve 19 inscritos no curso e 14 alunos o completaram.

O curso, com atividades presenciais e não-presenciais, foi planejado com base em 3 eixos, que visavam à instrumentalização dos participantes para a utilização de corpora *online*, à compreensão de conceitos tais como língua, aprendizagem, uso, registro, gramática descritiva e prescritiva, além de proporcionar oportunidades para o desenvolvimento linguístico. Dentre as atividades presenciais houve aulas expositivas, discussões em grupo sobre os textos e seminário em grupo. As atividades não-presenciais, individuais ou em grupo, eram anunciadas na plataforma TelEduc (ver Ilustração 1), na qual os alunos as postavam em seus respectivos portfólios (ver Ilustração 2).

---

<sup>7</sup> A professora da disciplina descrita e é, também, pesquisadora autora deste artigo.

<sup>8</sup> As disciplinas Habilidades Integradas I, II e III são, normalmente, as primeiras disciplinas de Língua Inglesa cursadas por alunos da graduação com habilitação em Inglês. Caso o aluno tenha conhecimento prévio do conteúdo das disciplinas, pode fazer uma prova de Aproveitamento de Estudos, que o dispensa de cursar as disciplinas.

**Corpora e Aprendizagem**  
**Agendas Anteriores** - Sept. 30th to Oct 28th

Busca Ajuda

Histórico Voltar para as Agendas Anteriores

Sept. 30<sup>th</sup> to Oct 28<sup>th</sup> Agenda

Dear students,

Task 8 is due Oct. 2<sup>nd</sup> since it is a preparation for our class discussion on that day. Please compare the focuses of chapters 5 and 6. How does chapter 6 complement the language discussion presented in chapter 5? Please post it in your individual portfolio.

On Oct. 2<sup>nd</sup> you should also bring the summary of chapter 6.

The reading for Oct. 9<sup>th</sup> is from *Longman Grammar of Spoken and Written English* (Biber et al. 1999) – Section 14: The grammar of conversation. It is a long section, so get the copy soon and start reading it. We'll decide in class how you'll present its content.

For Oct. 16<sup>th</sup> you should read the text called Ten criteria for a spoken grammar (McCarthy and Carter 2002) and bring a summary of it. The text is also available at the copy center.

Prof. Douglas Biber will be here for SEVFALE and you can't miss his talk.

It'll be on Oct. 21<sup>st</sup> at 7 pm and the title is "Corpus linguistics and language teaching".

Your Task 11 is to summarize his talk and relate it to our course content (Due Oct. 28<sup>th</sup>).

**Ilustração 1 – Exemplo de uma página de abertura da disciplina na plataforma TelEduc com a agenda das atividades.**

**Corpora e Aprendizagem**

**Portfólio - Portfólio Individual** [Busca](#) [Ajuda](#)

 Rejane  
 Raiz

| Meus Portfólios           | Portfólios Individuais | Portfólios de Grupos | Portfólios Encerrados |
|---------------------------|------------------------|----------------------|-----------------------|
| <a href="#">Atualizar</a> |                        |                      |                       |

| Itens   | Data       | Compartilhamento         | Comentários | Avaliação |
|---|------------|--------------------------|-------------|-----------|
|  <a href="#">First Task</a>      | 22/08/2008 | Totalmente Compartilhado |             | Não       |
|  <a href="#">Second Task</a>     | 26/08/2008 | Totalmente Compartilhado |             | Não       |
|  <a href="#">Third Task</a>      | 02/09/2008 | Totalmente Compartilhado |             | Não       |
|  <a href="#">Fourth Task</a>     | 06/09/2008 | Totalmente Compartilhado |             | Não       |
|  <a href="#">Fifth Task</a>      | 09/09/2008 | Totalmente Compartilhado |             | Não       |
|  <a href="#">Sixth Task</a>      | 16/09/2008 | Totalmente Compartilhado |             | Não       |
|  <a href="#">Seventh Task</a>    | 20/09/2008 | Totalmente Compartilhado |             | Não       |
|  <a href="#">Eighth Task</a>     | 29/09/2008 | Totalmente Compartilhado |             | Não       |
|  <a href="#">Ninth Task</a>    | 06/10/2008 | Totalmente Compartilhado |             | Não       |
|  <a href="#">Eleventh Task</a> | 24/10/2008 | Totalmente Compartilhado |             | Não       |

 Comentário de Aluno    Comentário de Formador    Comentários enviados por mim    Item Avaliado

**Ilustração 2 – Exemplo de Portfólio de uma das participantes do curso.**

Os dados apresentados neste artigo foram coletados em três etapas: no início, no meio e ao final do curso oferecido. Primeiramente, foram coletadas as respostas ao questionário sobre a experiência de aprendizagem com corpora, sobre a expectativa em relação ao curso e sobre as dificuldades linguísticas dos alunos. Posteriormente, os participantes desenvolveram 10 tarefas baseadas nos princípios do Aprendizado Movido por Dados (JOHNS 1994). A abordagem para a preparação dessas tarefas foi tanto reativa quanto proativa. A primeira abordagem implicou a análise das redações produzidas pelos alunos<sup>9</sup> e do questionário, ambos realizados no início de curso. A abordagem proativa foi utilizada quando os tópicos linguísticos discutidos em O´Keeffe, McCarthy e Carter (2007) e Biber

<sup>9</sup> As redações deviam ser argumentativas e seguiram as orientações do *International Corpus of Learner English* (ICLE) apresentadas no *site* <http://cecl.fltr.ucl.ac.be/Cecl-Projects/Icle/icle.htm>

et al. (1999)<sup>10</sup> suscitavam o interesse dos alunos por aprofundamento. As tarefas criadas pela professora/pesquisadora utilizaram corpora gerais<sup>11</sup> *online* *Brown Corpus*, *British National Corpus* (BNC) e *Contemporary Corpus of American English* (COCA) que, por meio dos sites que os disponibilizam, geram concordâncias e gráficos a serem analisados.

### 3. Resultados

Nesta seção, apresentamos a discussão de tarefas de abordagem reativa e proativa, segundo os princípios do Aprendizado Movido por Dados (JOHNS 1994). A primeira é um exemplo de uma tarefa desenvolvida em grupo e apresentada como seminário, e a segunda foi feita individualmente. Conforme mencionado anteriormente, todas as tarefas foram postadas na plataforma TelEduc, o que facilitou o acesso de todos os participantes ao trabalho desenvolvido.

Durante a análise das redações produzidas pelos aprendizes chamou-nos a atenção a utilização de adjetivos modais (por exemplo, *It is possible*; *It is probable*). Por causa do número reduzido de redações<sup>12</sup>, preferimos falar que esses itens eram “salientes” e não “de alta frequência” nas redações. À luz dos conceitos de frequência e saliência, podemos discutir a relevância de tratar itens, como adjetivos modais, por serem menos frequentes do que verbos modais, por exemplo. Dentre os linguistas que trabalham com corpus e têm interesse no ensino, a frequência de um item é considerada condição necessária para torná-lo prioritário à aprendizagem (i.e. O’KEEFFE, MCCARTHY e CARTER 2007 sobre o ensino de porções - *chunks*). De outra feita, a importância de um determinado item para um grupo cultural específico é também suficiente para fundamentar a decisão de ensinar o item menos frequente em detrimento a outro mais frequente (HUNSTON 2002). No caso dos participantes deste

---

<sup>10</sup> O’Keeffe, McCarthy e Carter (2007) foi o livro texto e Biber et al. (1999) a gramática de referência do curso. O programa inicial previa algumas leituras e outras foram escolhidas durante o semestre, tendo em vista o interesse do grupo de participantes.

<sup>11</sup> Há vários tipos de corpora, por exemplo, especializados (i.e. artigos acadêmicos), gerais, que incluem muitos tipos de textos (i.e. *British National Corpus*); comparáveis que englobam dois ou mais corpora de língua diferentes ou de variantes de uma mesma língua; paralelos, que são formados por textos originais e traduzidos, ou “produzidos simultaneamente em duas ou mais línguas” (HUNSTON 2002, p. 15); de aprendizes, que são formados por textos produzidos por aprendizes de uma língua; pedagógicos, que consistem de materiais pedagógicos a que o aprendiz tenha sido exposto; históricos ou diacrônicos e monitores, que visam acompanhar mudanças na língua (HUNSTON 2002).

<sup>12</sup> Devido ao fato de ser uma disciplina optativa e de haver pré-requisito, o número de matriculados foi 19 alunos.

estudo, alunos da graduação de nível intermediário e avançado de inglês, que serão profissionais na área de inglês (e.g. tradutores, professores), a aprendizagem do uso apropriado de todas as estruturas modais pode ser muito benéfica. Além disso, o caminho escolhido para propiciar a descoberta do funcionamento da língua inglesa com base em dados de corpora, em nosso caso, disponíveis *online*, parece adequado ao grupo em questão, que precisa ter domínio do sistema de modalização em inglês para sua atuação profissional. A seguir, descrevemos os passos da tarefa de abordagem reativa, que visou ao aprofundamento do conhecimento dos aprendizes sobre as estruturas modais, seguida da discussão dos resultados apresentados pelos alunos nos seminários.

O primeiro passo foi uma atividade de leitura de material autêntico (*cartoon*) no qual apareciam verbos modais. Os alunos responderam perguntas de compreensão do texto e, em seguida, de descoberta do uso das estruturas linguísticas em foco. Essa atividade introdutória teve o objetivo de ativar o conhecimento prévio dos aprendizes sobre estruturas modais e, também, de despertar sua curiosidade quanto às possibilidades de expressão modal em inglês.

O segundo passo foi a organização dos grupos para os seminários sobre como expressar **possibilidade** (sentido epistêmico<sup>13</sup>) em inglês. A atividade objetivou a descoberta do perfil lexico-gramatical dos seguintes itens linguísticos: verbos modais (e.g. *may, might*), advérbios modais (e.g. *possibly, probably, definitely*), adjetivos modais (e.g. *possible, probable*), verbos modais lexicais (e.g. *seem, appear*). Os grupos receberam perguntas para guiar a descoberta de dados de corpora *online*<sup>14</sup> a fim de que pudessem pesquisar os itens linguísticos em foco quanto à sua frequência, à correlação entre estrutura e registro, às colocações mais frequentes (palavras que co-ocorrem com a palavra investigada), às restrições

---

<sup>13</sup> Os modais podem ser divididos em epistêmicos e deônticos sendo que os primeiros expressam sentidos de possibilidade e probabilidade e os outros denotam sentidos ligados ao mundo real, tais como obrigação, permissão e habilidade. (c.f. DUTRA 1998 para discussão sobre modais)

<sup>14</sup> As perguntas abaixo guiaram a investigação e continham os termos discutidos em sala que são comumente usados na Linguística de Corpus. Uma diferença entre as tarefas do curso ora descrito e o sugerido por Johns (1994) é que não houve seleção feita pela professora dos exemplos do corpus a serem analisados. Isso não foi feito, pois um dos objetivos do curso era ensinar os alunos a fazerem suas buscas nos corpora e a selecionarem os exemplos a serem discutidos.

1. How many examples did you get and in which corpus?
2. Is there a correlation with the use of the structure and a specific register? (include a chart);
3. Make a lexico-grammatical profile of the category your group was assigned to:
  - a) What are the most common words it collocates with?
  - b) Are there any syntactic restrictions?
  - c) What can you say about its semantic prosody?

sintáticas e à prosódia semântica<sup>15</sup>. Os seminários foram apresentados presencialmente. A seguir, discuto os resultados em relação à aprendizagem dos alunos. Ilustro, também, o que foi descoberto por eles com base nos exemplos discutidos pelos grupos.

Todos os grupos puderam investigar a frequência dos itens linguísticos em questão em vários corpora quanto: a) ao registro(s) no(s) qual(is) os itens ocorrem mais e b) à classe gramatical usada para expressar o sentido de possibilidade. Primeiramente, eles buscaram verificar se o registro era determinante na escolha do item linguístico. A Ilustração 3 apresenta um exemplo de um dos grupos sobre a alta frequência do modal *may* em prosa acadêmica no sentido epistêmico. Isso levou os alunos a procurarem dados sobre o uso de *may* no sentido deontico e a descobrirem a baixa frequência desse modal com esse sentido. Chegar a essa conclusão pela análise de dados de corpora foi gratificante para os alunos, principalmente quando a pesquisadora (e professora) mencionou que resultados parecidos constavam na gramática de Biber et al. (1999), baseada na análise de um corpus de aproximadamente 40 milhões de palavras de textos orais e escritos. Além disso, como cada grupo de alunos ficou responsável pela investigação de uma classe gramatical utilizada para expressar o sentido de possibilidade, ao final de todas as apresentações foi possível descobrir as classes mais produtivas. Dentre as quatro classes investigadas, ficou evidente que os verbos modais são muito mais frequentes do que as outras três classes. Devido ao trabalho específico de cada grupo também foi possível perceber quais itens, em cada classe gramatical, eram mais comumente usados em diferentes registros.

---

<sup>15</sup> Louw (1993) apud Berber Sardinha (2004: 40) define o termo prosódia semântica como “associação entre itens lexicais de conotação (negativa, positiva ou neutra) ou instância avaliativa”.

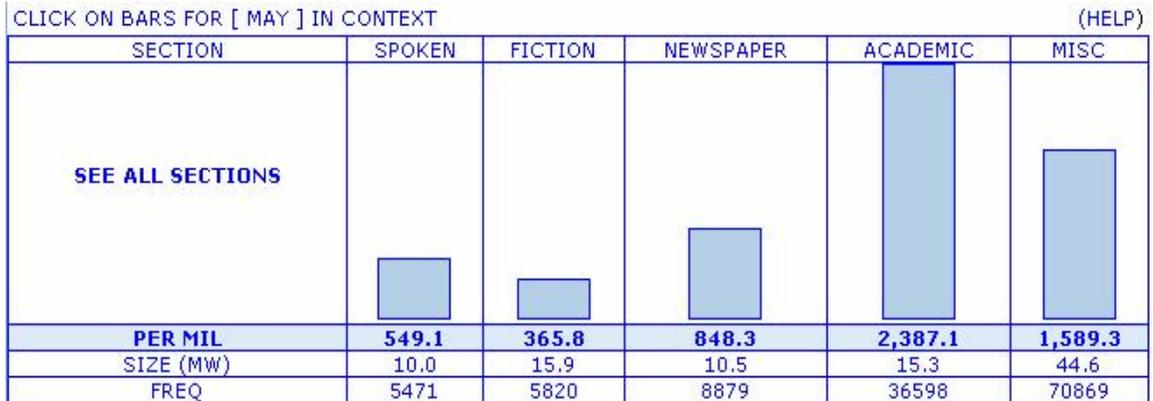
- **May**

1. **How many examples did you get and in which corpus?**

5471 examples were taken from the British National Spoken Corpus

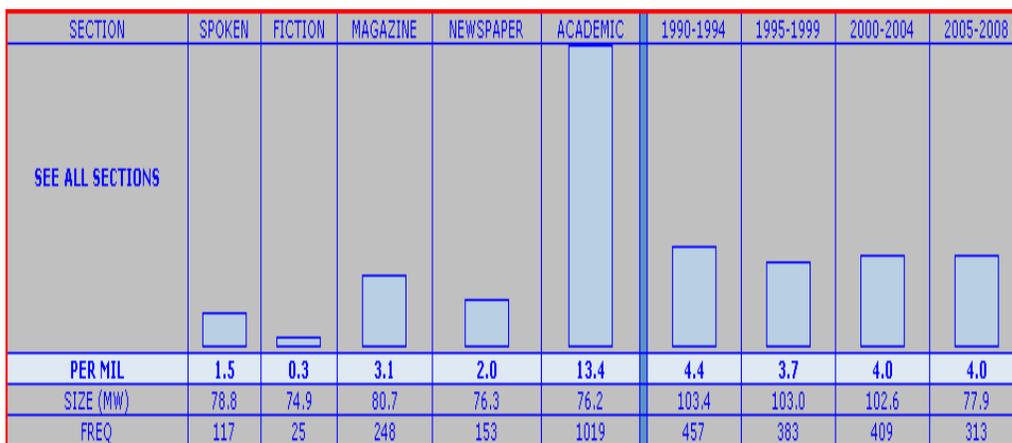
2. **Is there a correlation with the use of the structure and a specific register? (include a chart) .**

According to this corpus, *may* is a lot more frequent in academic context than in the other ones. Actually, it has a very low frequency in spoken discourse, fiction and newspaper.



**Ilustração 3 – Exemplo de *slide* projetado em seminário sobre os verbos modais**

Outro aspecto apreendido pelos participantes foi o uso específico dos itens linguísticos em questão. No caso dos adjetivos modais, a sua análise no contexto de uso, principalmente na escrita acadêmica, levou à hipótese de que estratégias de atenuação seriam responsáveis pela alta frequência de, por exemplo, *likely* nesse registro (Ilustração 4). Os autores desse tipo de texto utilizam uma linguagem cuidadosa para fazer suas asserções, já que é comum preferirem se resguardar quanto a resultados ou opiniões divergentes (Ilustração 5).



**Ilustração 4 – Exemplo de *slide* referente à construção modal *It's likely* – Corpus of Contemporary American English (COCA) [www.americancorpus.org](http://www.americancorpus.org).**

### Source information:

|                         |   |
|-------------------------|---|
| Date                    | 2008  |
| Publication information | Mar/Apr2008, Vol. 87 Issue 2, p49-62, 13p, 1 bw |
| Title                   | An Empty Revolution.                            |
| Author                  | Rodriguez, Francisco <sup>1</sup>               |
| Source                  | Foreign Affairs                                 |

### Expanded context:

the very high costs it will pay if it loses power. Unless a framework is forged through which the government and the opposition can reach a settlement, there is a significant risk that one or both sides will resort to force. Looking back, one persistent question (in itself worthy of a potentially fascinating study in international political economy) will be how the Venezuelan government has been able to convince so many people of the success of its antipoverty efforts despite the complete absence of real evidence of their effectiveness. When such a study is written, **it is likely** that the Chavez administration's strategy of actively lobbying foreign governments and launching a high-profile public relations campaign--spearheaded by the Washington-based Venezuela Information Office--will be found to have played a vital role. The generous disbursement of loans to cash-strapped Latin American and Caribbean nations, the sale of cheap oil and heating gas to support political allies in the developed and developing worlds, and the covert use of political contributions to buy the loyalty of politicians in neighboring countries must surely form part of the explanation as well. But perhaps an even

### Ilustração 5 – Contexto expandido com ocorrência da construção modal *It's likely* – Corpus of Contemporary American English (COCA) [www.americancorpus.org](http://www.americancorpus.org)

A segunda tarefa, que passo a descrever a seguir, foi planejada de acordo com a abordagem proativa e realizada pelos participantes individualmente. O objetivo era proporcionar oportunidades de aprofundamento quanto ao uso de porções (*chunks*)<sup>16</sup>, frequentes em inglês. Porções são "seqüências de duas ou mais palavras que ocorrem regularmente e parecem possuir unidade de significados ou funções" (O'KEEFFE, MCCARTHY e CARTER 2007: 46). Esses autores argumentam que as porções mais frequentes nem sempre são salientes<sup>17</sup> e, por isso, sugerem o aumento da conscientização sobre

<sup>16</sup> Minha tradução para ...."regularly occurring strings of two or more words which seem to possess unitary meanings or functions" (O'KEEFFE, MCCARTHY e CARTER 2007: 46).

<sup>17</sup> O'Keeffe, McCarthy e Carter (2007) apresentam uma interessante discussão sobre a saliência psicológica de algumas porções, como expressões idiomáticas e verbos frasais, que eram muitas vezes apresentados no ensino de inglês como curiosidades da língua, principalmente antes dos estudos de corpus. Atualmente,

a utilidade dessas sequências de palavras. Segundo (O´Keeffe, McCarthy, Carter 2007: 75)<sup>18</sup>:

Ao explorar os usos de porções no corpus oral, fica aparente que, entre os mais frequentes (os 20 mais comuns em cada caso), há um número considerável que tem funções pragmáticas claras e comuns na organização e manejo conversacional, e na relação entre o falante e o ouvinte.

Primeiramente, os participantes do curso leram o capítulo "*Lessons from the analysis of chunks*" de O´Keeffe, McCarthy, Carter (2007) e fizeram um resumo do texto. Posteriormente, houve discussão em sala de aula sobre os conceitos e dados apresentados. Em seguida, os alunos receberam a tarefa de investigar porções frequentes como *you know*, contrastando os resultados dos corpora *online* com os resultados apresentados no capítulo lido, baseado no corpus de inglês oral CANCODE (*Cambridge and Nottingham Corpus of Discourse in English*). Finalmente, cada aluno escolheu uma porção, que poderia ser formada de duas a seis palavras, e fez uma investigação quanto à sua frequência, ao seu perfil léxico-gramatical e à sua função.

Os resultados apontam para um aumento da compreensão dos aprendizes quanto às características linguísticas da porção escolhida, resultante da análise de concordâncias. Um dos participantes do curso escolheu analisar *sort of* e, como o capítulo estudado trazia dados de corpus oral, sua primeira busca foi no *BNC* oral, tendo analisado, posteriormente, como essa sequência de palavras aparecia em outros registros, ao utilizar o COCA. Foi confirmada a alta frequência no registro oral e, após a análise das palavras que a antecedem, verificou-se a incidência de hesitações conversacionais como *er er*, *erm*, *mm mm*, características desse tipo de registro (Ilustração 6). Outra constatação, em relação à função de *sort of*, foi a de que, ao expressar imprecisão, a porção pode ser também utilizada como um mecanismo para "salvar a face" do interlocutor, conforme o exemplo extraído do COCA: "I agree with you that my medical education is **sort of**, you might say, a little blotchy here and there and it wasn't in the United States. So what?", ou de diminuir a importância de um fato ou situação, também coletado do COCA "I felt **sort of** dizzy, like something's coming up the top of my head, you

---

com esses estudos, podemos avaliar a relevância do ensino de certas sequências de palavras menos pelas suas peculiaridades e mais pelo seu uso efetivo.

<sup>18</sup> By exploring the uses of the chunks in the spoken corpus, it is apparent that amongst the most frequent (the top 20 in each case), there are a considerable number which have clear, common pragmatic functions in the organization and management of conversation and the speaker-listener relationship." (O´KEEFFE, MCCARTHY e CARTER 2007: 75)

know, and it's just a funny feeling, it'll wear off." A análise das concordâncias proporcionaram uma maior conscientização do uso de *sort of* para uma aluna de nível avançado que, apesar de ter conhecimento prévio dessa porção, não sabia sobre as sutilezas pragmáticas identificadas durante as atividades do curso.

I can't understand why erm i i it's not been do , SORT OF not done ba by now because, oh yes. Perhaps I arrive, piano er score for it, and er it's a SORT OF fairy tale lyric you know, I've Very Paul than his shoulders so that makes him like like a, SORT OF normal person. A wimp! Ah! Did you see that aper, that paper Yeah. Mm mm. so that will give a SORT OF a background imprint you know? The way I've ere. We got a greenhouse. And some veggies! And a SORT OF Are these leeks? things that I don't know dark red or erm or white. It's sort of a is it a SORT OF bluey green? Or is it a racing green? It's it has got those little Unfortunately it's got a SORT OF general grubbiness to it as well, hasn't i ou now have to get, in er er, because there was a SORT OF erm gradation from being off ill, which wa a great deal of problem, others really do have a SORT OF off work for two or three days every month to do now erm twentieth centu no it isn't it's a SORT OF nineteen eighties version of wanting to be

### Ilustração 6 – Concordâncias extraídas do corpus BNC oral

A tarefa posterior a essa análise de porções envolveu a leitura e discussão de um capítulo de Biber et al. (1999) sobre *lexical bundles*, que são sequências de palavras que co-ocorrem com frequência, sem necessariamente terem uma unidade sintática ou função. Dessa maneira, as porções são *bundles*, mas nem todos os *bundles* são porções. Os alunos puderam escolher quais *bundles* queriam investigar e utilizaram as concordâncias para verificar seu perfil léxico-gramatical e sua função. Um dos alunos analisou *I tell you what* e constatou sua alta frequência no corpus oral. A análise do contexto linguístico das ocorrências levou-o a identificar que, em alguns casos, o *bundle* era um fragmento de sentença com uma oração encaixada, como em *I tell you **what I did***. Em outros casos, *I tell you what* aparecia como uma unidade com uma função, por exemplo, de fazer uma sugestão, como em *I guess that you could -- **I tell you what**, Harold, you can take the first one*.

Como detectado no início do curso, poucos alunos tinham conhecimento dos instrumentos disponíveis da Linguística de Corpus para análise linguística. Ao final do curso, um questionário sobre as tarefas propostas foi aplicado aos alunos e seus resultados revelaram um saldo positivo quanto às atividades desenvolvidas (Quadro 1).

| <b>As atividades o ajudaram a</b>     | <b>Porcentagem das respostas</b> |
|---------------------------------------|----------------------------------|
| aprender a usar recursos corpora      | 85                               |
| entender a área de estudos de corpora | 100                              |
| aprender mais sobre a língua inglesa  | 100                              |

**Quadro 1 – Resultado dos questionários de final de curso**

Segundo os participantes, as atividades de busca nos corpora eletrônicos são produtivas porque despertam o interesse pela pesquisa em LE, enfatizam a língua em uso, por serem fonte de material autêntico, e auxiliam na compreensão da relação entre aspectos contextuais e significado.

#### 4. Conclusão

A pesquisa ora relatada sobre um curso semi-presencial com o foco em corpora mostrou que, apesar de os alunos estarem na fase final da graduação, a maioria deles tinha pouca familiaridade com recursos de corpora *online*. Esses aprendizes passaram a ter uma melhor compreensão de conceitos linguísticos, além de terem incorporado novos conceitos sobre o que são colocações, porções (*chunks*), expressões idiomáticas, materiais autênticos, entre outros. Houve, dessa maneira, um aumento da conscientização de aspectos linguísticos.

Quanto aos instrumentos utilizados para o levantamento das dificuldades linguísticas dos participantes, detectamos que as redações foram mais produtivas para esse fim do que o questionário de início do curso, que constava de perguntas sobre o que eles acreditavam ser necessário melhorar em sua proficiência em inglês. Os itens linguísticos escolhidos para destaque nas tarefas suscitaram interesse dos alunos e os levaram a aprender aspectos dos quais não tinham conhecimento anterior. Todavia, não foi possível verificar o desenvolvimento linguístico dos aprendizes além do que foi relatado por eles, pois as redações foram coletadas somente no início do semestre. Acreditamos, por outro lado, que, mesmo que as redações tivessem sido coletadas em outros momentos do curso, apenas um semestre teria sido, de qualquer forma, um período muito curto para afirmar algo sobre o desenvolvimento linguístico dos alunos nos aspectos em foco durante as pesquisas.

De todo modo, as abordagens indutivas (reativa e proativa), baseadas nos princípios do Aprendizado Movido por Dados, mostraram-se eficazes ao propiciar a conscientização linguística desejável em aprendizes que vislumbram atingir um alto nível de

proficiência. Esse foi um aspecto relevante para os participantes deste estudo, alunos de graduação em Letras e futuros profissionais da área, como professores, tradutores, revisores de textos, dentre outros.

## REFERÊNCIAS BIBLIOGRÁFICAS

- ALMEIDA, Márcia Cristina Fontes. O uso da metalinguagem no discurso em sala de aula do professor de língua estrangeira. Dissertação (Mestrado em Estudos Linguísticos) - Universidade Federal de Minas Gerais, 2003.
- ALMEIDA, M. C. F.; DUTRA, D. P. O uso da abordagem proativa vs. a abordagem reativa no discurso metalingüístico do professor de língua estrangeira. In: VII Congresso Brasileiro de Linguística Aplicada, 2005, São Paulo. *Anais do VII CBLA*. São Paulo: ALAB, 2004.
- ANDREWS, S. The language awareness of the L2 teacher: its impact upon pedagogical practice. *Language awareness*, v. 10, n. 2&3, p. 75-90, 2001.
- \_\_\_\_\_. *Teacher language awareness*. Cambridge: Cambridge, 2007.
- BERBER SARDINHA, A. P. Beginning Portuguese corpus linguistics: exploring a corpus to teach Portuguese as a foreign language. *DELTA* [online] v. 15, n. 2: 289-299, 1999.
- \_\_\_\_\_. Compilação e anotação de um corpus de português de linguagem profissional. *The Specialist*, v. 21, n. 1: 111-147. São Paulo: EDUC, 2000.
- \_\_\_\_\_. corpus de aprendiz Br-ICLE. *Intercâmbio*, v. 10: 227-39, 2001.
- \_\_\_\_\_. *Linguística de corpus*. Barueri: Manole, 2004.
- BERBER SARDINHA, A. P.; SHEPERD, T. An online system for error identification in Brazilian learner English. *Anais do 8th Teaching and Language Corpora Conference*. Lisboa: Associação de Estudos e de Investigação Científica do ISLA-Lisboa. p. 257-262, 2008.
- CANALE, M. From communicative competence to communicative language pedagogy. In: RICHARDS, J.; SCHMIDT, R. *Language and communication*. London: Logman, 1983.
- CANALE, M.; SWAIN, M. Theoretical bases of communicative approaches to second language testing and teaching. *Applied Linguistics*, v. 1. n. 1, p. 1-47, 1980.
- CARAZZAI, M. Grammar and grammar teaching: a qualitative study of EFL teachers' beliefs and behaviors. 2002. Dissertação (Mestrado em Língua Inglesa). Universidade Federal de Santa Catarina.
- CELCE-MURCIA, M. *Teaching English as a second or foreign language*. Boston: Heinle & Heinle, 1991.
- CELCE-MURCIA, M., DÖRNYEI, Z., & THURRELL, S. Direct approaches in L2 instruction: A turning point in communicative language teaching? *TESOL Quarterly*, v. 31, p.141-152, 1997.

DOUGHTY, C.; WILLIAMS, J. (Ed.) *Focus on form in classroom second language acquisition*. Cambridge: Cambridge University Press, 1998.

DUTRA, D. P. The acquisition of English root modality by non-native speakers. 1998. Tese (Doutorado em Linguística) - Programa em Linguística, University of Florida, Gainesville.

\_\_\_\_\_ A metalinguagem no sistema pedagógico do professor. In: DUTRA, D. P.; MELLO, H. (org.) *A gramática e o vocabulário no ensino de inglês: novas perspectivas*. Belo Horizonte: POSLIN/FALE/UFMG, 2004. p.17-39.

\_\_\_\_\_ O ensino de L2 no Brasil. No prelo.

DUTRA, D. P.; MELLO, H. Os caminhos do ensino de gramática em línguas estrangeiras. In: DUTRA, D. P.; MELLO, H. (org.) *A gramática e o vocabulário no ensino de inglês: novas perspectivas*. Belo Horizonte: POSLIN/FALE/UFMG, 2004 (a). p. 9-15.

\_\_\_\_\_ Pesquisas em linguagem: o que elas revelam sobre um projeto de educação continuada. IN: TELLES, J. (Org.) No prelo.

DUTRA, P. B. Explorando a Linguística de Corpus e letras de música na produção de atividades pedagógicas. 2002. 127 f. Dissertação (Mestrado em Linguística Aplicada) – Programa de Pós-Graduação em Linguística Aplicada ao Ensino de Línguas, PUC-SP, São Paulo.

EDGE, J. Applying linguistics in English language teacher training for speakers of other languages. *ELT Journal*, v. 42, n.1, p. 9-13, 1988.

GIL, G. Foco-na-forma e foco-na-comunicação: dois focos complementares. In: DUTRA, D. P.; MELLO, H. (org.) *A gramática e o vocabulário no ensino de inglês: novas perspectivas*. Belo Horizonte: POSLIN/FALE/UFMG, 2004. p. 41-53.

HUSTON, S. *Corpora in Applied Linguistics*. Cambridge: Cambridge, 2002.

JOHNS, T. From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. In ODLIN, R. (ed.). *Perspectives on Pedagogical Grammar*. Cambridge: Cambridge, 1994. p. 293-313.

LONG, M. Focus on form: a design feature in language teaching methodology. In: de BOT, K.; GINSBERG, R.; KRAMSCH, C. (Ed.). *Foreign language research in cross-cultural perspectives*. Amsterdam: John Benjamins, 1991. p.39-52.

LÚCIO, D. D. A relexicalização de adjetivos nas redações de alunos de inglês: um estudo baseado em Corpus de aprendiz. 2006. 105 f. Dissertação (Mestrado em Linguística Aplicada) – Programa de Pós-Graduação em Linguística Aplicada ao Ensino de Línguas, PUC-SP, São Paulo.

MARQUES, A. L. S. P. O papel da percepção consciente de conectivos concessivos na produção escrita e na reformulação da interlíngua. 2007. Dissertação (Mestrado em Linguística Aplicada) - Programa de Pós-Graduação em Estudos Linguísticos, Universidade Federal de Minas Gerais, Belo Horizonte.

NÉRI, P. A. De foco nas formaS a foco na forma: relato de experiência de uma professora da escola regular. 2006. Dissertação

(Mestrado em Linguística Aplicada) - Programa de Pós-Graduação em Estudos Linguísticos, Universidade Federal de Minas Gerais, Belo Horizonte.

NEVES, M. S. Os mitos de abordagens tradicionais e estruturais ainda interferem na prática em sala de aula. In: PAIVA, V. L. M. O. (Org.). *Ensino de língua inglesa: reflexões e experiências*. Campinas: Pontes; Belo Horizonte: Departamento de Letras Anglo-Germânicas – UFMG, 1996. p. 69-80.

OLIVEIRA, A. L. A. M. Eu até hoje não sei se devo ensinar gramática ou não: a ação-colaborativa e sua influência na práxis docente. *Revista Intercâmbio*, XV, 2006.

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

PARDO, R. M. Critérios de Construção e Organização de um Corpus de Especialidade: o corpus técnico-científico de ortodontia. 2004. Dissertação (Mestrado em Estudos Linguísticos e Literários em Inglês) - Universidade de São Paulo.

PINTO, M. V. O uso de *things, anything, something* e *everything* em corpora de aprendiz. 2008. 221 f. Dissertação (Mestrado em Linguística Aplicada) – Programa de Pós-Graduação em Linguística Aplicada ao Ensino de Línguas, PUC-SP, São Paulo.

RICHARDS, J., Accuracy and fluency revisited. IN: HINKEL, E. FOTOS, S. *New perspectives on grammar teaching in second language classrooms*. Mahwah: Lawrence Erlbaum, 2002. p. 35-50.

SCHMIDT, R.; FROTA, S. Developing basic conversational ability in a second language: a case study of an adult learner of Portuguese. IN: DAY, R. (Ed.) *Talking to learn: conversation in second language acquisition*. Rowley: Newbury House. p. 237-326.

SOUZA, P. R. As conceptualizações sobre o ensino de gramática de duas professoras em formação inicial, suas práticas de sala de aula e seus processos reflexivos. 2008. Dissertação (Mestrado em Linguística Aplicada) – Programa de Pós-Graduação em Estudos Linguísticos, Universidade Federal de Minas Gerais, Belo Horizonte.

SWAIN, M. French immersion and its offshoots; getting two for one. IN: FREED, B. (Ed.). *Foreign Language Acquisition: research and the classroom*. Lexington: Heath, 1991.p.91-103.

VICENTINI, G. P. M. Aluno como pesquisador em e-classes: elaboração de atividades baseadas em um corpus do seriado *Friends*. *Revista Intercâmbio*, volume XV. São Paulo: LAEL/PUC-SP, 2006.

Recebido em agosto de 2009  
Aprovado em fevereiro de 2010