

# Descobertas linguísticas para pesquisadores e aprendizes: a Linguística de *Corpus* e o ensino de gramática<sup>1</sup>

## *Linguistic discoveries for researchers and learners: Corpus Linguistics and the teaching of grammar*

---

Deise Prina Dutra\*  
Universidade Federal de Minas Gerais - UFMG  
Belo Horizonte - Minas Gerais/Brasil

Rejane Protzner Silero\*\*  
Universidade Federal de Minas Gerais - UFMG  
Belo Horizonte - Minas Gerais/Brasil

RESUMO: Este artigo discute a ocorrência de agrupamentos com a palavra *for* em textos argumentativos produzidos por aprendizes brasileiros de inglês de nível universitário e sugere atividades pedagógicas de descoberta de padrões linguísticos. Utilizamos ferramentas da Linguística de Corpus (LC) para análise e sugestões de ensino de gramática. Este artigo baseia-se em dados do Corpus de Aprendizes Brasileiros do Inglês (CABrI) em comparação com outros *corpora*. Os resultados, que revelam o uso adequado ou não adequado de agrupamentos com verbo + *for* e substantivo + *for*, são seguidos de atividades que visam à integração da gramática com o léxico e a uma conscientização dos aprendizes sobre a frequência do uso de certos agrupamentos linguísticos.

PALAVRAS-CHAVE: *corpus* de aprendizes, frequência, ensino de gramática e vocabulário

---

<sup>1</sup> Este trabalho é um dos frutos do pós-doutorado da primeira autora na Pontifícia Universidade Católica de São Paulo sob a supervisão de Tony Berber Sardinha. Uma versão preliminar dessa pesquisa foi apresentada no VIII Encontro de Linguística de Corpus no Rio de Janeiro em 2009.

\* deisepdutra@gmail.com

\*\* rejaneprotzner@gmail.com

ABSTRACT: This article discusses the occurrences of clusters with the word *for*, in argumentative essays written by undergraduate Brazilian learners of English and suggests pedagogical activities to encourage the discovery of linguistic patterns. We use Corpus Linguistics (CL) tools in our analysis and pedagogical suggestions. Our research is based on data collected from the Brazilian Corpus of English Learners (*Corpus de Aprendizes Brasileiros do Inglês – CABrI*) in comparison with other corpora. The results, which pointed to appropriate or inappropriate use of bundles such as *verb + for* and *noun + for*, are followed by activities that aim to integrate grammar and lexis and to promote learners' awareness of frequency of use of certain linguistic bundles.

KEYWORDS: learner corpus, frequency, teaching grammar and vocabulary

## Introdução

O tema ensino de gramática de uma segunda língua (SL) ou língua estrangeira (LE), no nosso caso, o inglês, suscita diferentes opiniões de aprendizes e professores e / ou pesquisadores. Se por um lado há aqueles que acreditam na importância do ensino da gramática para o domínio efetivo de uma língua, há outros que preferem valorizar outros aspectos da língua, visando mais ao desenvolvimento da fluência do que da acuidade. Ainda há os que advogam por um equilíbrio entre as atividades que enfatizam o desenvolvimento da fluência e da acuidade, sem a necessidade de retorno à instrução com foco puramente gramatical (RICHARDS, 2002). Antes de apresentar nossa proposta de ensino de gramática, cabe discutir em qual vertente da Linguística nos inserimos. Isso nos leva a conceituar gramática e, posteriormente, a apresentar a abordagem empirista da linguagem, utilizada pela Linguística de *Corpus* (LC), base para este artigo, e sua frutífera ligação com a Linguística Aplicada, especialmente com o segmento que se ocupa das questões relativas ao ensino de línguas estrangeiras.

Optamos, neste trabalho, por nos filiar ao conceito de Linguística como uma ciência descritiva, que leva em consideração a linguagem em uso e não, necessariamente, o padrão tradicionalmente reconhecido como correto:

[...] o linguista tenta descobrir e registrar as regras que os membros de uma comunidade de linguagem seguem e não busca impor a eles outras regras (externas), ou normas do que seja correto. [...] A razão pela qual os linguistas atuais são tão insistentes sobre a distinção entre as regras descritivas e prescritivas é simplesmente porque a gramática tradicional tinha um forte caráter normativo. Os gramáticos viam

como sua tarefa formular os padrões corretos e impô-los, se necessário, aos falantes da língua. (LYONS, 1981, p. 47-48).<sup>2</sup>

Muito embora saibamos que a área de ensino de línguas não se baseie somente na Linguística para tomar suas decisões pedagógicas, nossa opção por favorecer o caráter descritivo da Linguística pode ser explicada pelo fato de que a gramática normativa teve e, em certos contextos, ainda tem forte influência nessas decisões. Em uma das tentativas de se discutir essa influência, Westney (1994, p. 73) parte de “uma definição ampla de gramática que considera gramática como qualquer aspecto da língua possível de ser sistematicamente descrito, envolvendo não somente morfologia e sintaxe, mas também incluindo fonologia e semântica”. Esse autor, baseando-se em Leech (1988),<sup>3</sup> afirma que a teoria gramatical e a prática pedagógica têm princípios e objetivos distintos; contudo, a gramática descritiva pode prover informações à teoria ou à prática. É nesse sentido que a Linguística de *Corpus* (LC) pode colaborar com a Linguística Aplicada (LA), pois, além de fornecer descrições da língua em uso, dispõe de ferramentas que podem ser utilizadas no ensino de línguas. Antes de nos atermos a esses dois aspectos relevantes para este artigo, cabe explicitar melhor alguns princípios da LC.

A LC segue uma abordagem empirista da linguagem, compreendida como um sistema probabilístico. Essa abordagem “significa primazia aos dados provenientes da observação da linguagem, em geral reunidos sob a forma de um *corpus*” (BERBER SARDINHA, 2004, p. 30). Segundo esse mesmo autor, a tradição empirista, impulsionada por Halliday, contrapõe-se ao racionalismo na linguística, defendido por Chomsky. Para os empiristas, deve-se analisar a língua partindo-se da frequência das ocorrências reveladoras das escolhas dos

---

<sup>2</sup> Nossa tradução para “[...] *the linguist tries to discover and record the rules to which the members of a language-community actually conform and does not seek to impose upon them other (i.e. extraneous) rules, or norms of correctness. [...] The reason why present-day linguists are so insistent about the distinction between descriptive and prescriptive rules is simply that traditional grammar was very strongly normative in character. The grammarian saw it as his task to formulate the standards of correctness and to impose these, if necessary, upon the speakers of a language.*” (LYONS, 1981, p. 47-48)

<sup>3</sup> Leech (1988) *apud* Westney (1994): LEECH, G. Varieties of English grammar: the state of the art from the grammarian’s point of view. In: BALD, W-D. (Ed.). *Kernprobleme der englischen Grammatik: Sprachliche Fakten und ihre Vermittlung*. München: Langescheidt-Longman, 1988. p. 5-17.

falantes em determinados contextos de uso. Enquanto isso, os racionalistas trabalham com as possibilidades que o sistema de uma língua coloca à disposição de seus falantes nativos. Para fins pedagógicos, é muito relevante considerar que as frequências de itens linguísticos não são aleatórias, mas, sim, dependentes do contexto (BIBER, 1988). Pesquisas recentes têm investigado, por exemplo, quais são os pacotes lexicais (*lexical bundles*) mais frequentes na escrita acadêmica (SIMPSON-VLACH; ELLIS, 2010), o que pode colaborar para um ensino mais efetivo àqueles que precisam ter sucesso no contexto acadêmico.

Delineamos brevemente os objetivos deste artigo, que visa exemplificar como as descrições da língua em uso e as ferramentas da LC<sup>4</sup> podem ser utilizadas no ensino da gramática de língua inglesa. Mais especificamente, nosso primeiro objetivo é descrever alguns aspectos da interlíngua de aprendizes de inglês, mostrando como os erros de aprendizes dessa língua, aparentemente restritos ao uso da preposição *for*, podem ser tratados de maneira inovadora na abordagem baseada em *corpus*. Nesse sentido, essa descrição aponta para o fato de que o ensino de gramática poderá ser mais integrado ao ensino de vocabulário (CONRAD, 2000). Para atingir esse primeiro objetivo descreveremos nosso *corpus* de estudo (Corpus de Aprendizes de Inglês - CABrI), formado por redações de alunos da graduação em inglês da Universidade Federal de Minas Gerais (UFMG), em comparação com outros *corpora* (*Contemporary Corpus of American English* – COCA, *Louvain Corpus of Native English Essays* – LOCNESS, e o subcorpus do ICLE coletado em outras instituições que não a nossa – Br-ICLE). Nosso segundo objetivo é mostrar como o foco nas condições de uso das construções linguísticas

---

<sup>4</sup> Esclarecemos brevemente que a LC parte da perspectiva da língua em uso, diferenciando-se da linguística tradicional que estuda questões estruturais das línguas (e.g. morfemas, palavras, frases etc.) (BIBER, CONRAD; REPPEN, 1998). Há autores (i.e. AARTS, 2002) que advogam que, tendo a LC o foco na língua em uso, que é distinto da ênfase da linguística teórica (que se ocupa da descrição da língua em geral), ela deve ser considerada uma disciplina. Aarts (2002) argumenta que a LC tem duas metodologias: baseada em *corpus* (*corpus-based*) e dirigida por *corpus* (*corpus-driven*). Outros autores (i.e. SCOTT; TRIBBLE, 2006, p. 4) não tratam a LC como um novo ramo de conhecimento e sim afirmam que os “métodos baseados em *corpus* são meramente um grupo de ferramentas e quadros”. Nossa tradução para “*corpus-based methods are merely a set of tools and frameworks*”. Neste artigo seguimos Scott e Tribble (2006), bem como Biber, Conrad e Reppen (1998), que filiam seus estudos linguísticos à abordagem baseada em *corpus* (*corpus-based approach*).

(CONRAD, 2000) pode favorecer o desenvolvimento da conscientização linguística dos aprendizes. Dessa maneira, o ensino de gramática para aprendizes que necessitem atingir boa acuidade passa a ser vinculado ao verdadeiro uso da língua. Antes de nos atermos aos resultados deste estudo, apresentamos a fundamentação teórica que norteia nosso trabalho, bem como a metodologia utilizada.

### **Fundamentação teórica**

O ensino de línguas baseado nas necessidades dos aprendizes, em alguns casos, tem ressaltado a ligação entre a forma, significado e uso (LARSEN-FREEMAN, 2003), considerados indissociáveis. Um dos exemplos desse enfoque seria a abordagem foco-na-forma, proposta por Long (1991), que pode ser tanto reativa quanto proativa (DOUGHTY; WILLIAMS, 1998). A abordagem reativa pressupõe que os aprendizes, ao tentarem utilizar certas estruturas linguísticas, tenham demonstrado dificuldades no seu uso apropriado, cabendo, então, por exemplo, um *feedback* corretivo por parte do professor (DOUGHTY; VARELLA, 1998). Já o tipo proativo da abordagem foco-na-forma implica que o professor, conhecendo seu grupo de alunos, possa prever dificuldades de aquisição de alguma estrutura e proponha atividades que ressaltem seu uso, ou seja, demonstrando como a forma é efetivamente usada em um contexto específico. Forma, significado e uso estão inequivocamente interligados, tanto na abordagem foco-na-forma reativa quanto na proativa.

Um mesmo curso pode implementar atividades reativas e proativas com sucesso, baseadas na LC, mais especificamente na proposta “Aprendizado Movido por Dados” (AMD – JOHNS, 1994), a qual se ancora na ideia de que os aprendizes de língua estrangeira devem ter a oportunidade de descobrir padrões na língua alvo a partir de análises de exemplos autênticos dessa língua (DUTRA, 2009). Esse estudo tratou das dificuldades dos aprendizes, alunos da graduação em inglês, propondo atividades preparadas com *corpora online*. A abordagem reativa envolveu a análise das redações produzidas pelos alunos. Esses dados mostraram, por exemplo, a necessidade dos aprendizes de desenvolverem seus conhecimentos sobre o uso apropriado do sistema de modalização do inglês. Já a abordagem proativa foi utilizada quando os alunos demonstravam interesse em aprofundar em algum tópico linguístico discutido nos textos do curso, por exemplo, o uso de porções (*chunks*) frequentes em inglês que têm “unidade de significados ou funções” (O’KEEFFE, MCCARTHY; CARTER, 2007, p. 46), como *you know*, *I mean* e *sort of*. Em ambos os

exemplos de atividades, a gramática não era compreendida somente em relação às questões das formas, e os aprendizes foram incentivados a fazer relações entre forma, significado e uso, ressaltando os padrões léxico-gramaticais<sup>5</sup> que emergiam das análises feitas. Para detectar esses padrões, os aprendizes eram expostos à concordância que é “uma listagem das ocorrências de um item específico, dispostas de tal modo que a palavra de busca (aquela que se tem interesse em investigar) aparece centralizada na página (ou tela do computador)” (BERBER SARDINHA, 2004, p. 187) (ver, por exemplo, FIG. 2 deste artigo). Eles recebiam a orientação de se concentrarem nas palavras que acompanham o item linguístico em questão, seja a sua direita ou esquerda a fim de que investigassem o cotexto<sup>6</sup> imediato, possibilitando a compreensão dos padrões léxico-gramaticais. Com a análise da concordância, destaca-se que “a palavra de busca é acompanhada do seu cotexto original, isto é, das palavras que ocorreram junto com ela no *corpus*.” (BERBER-SARDINHA, 2004, p. 187).

Os tipos de atividades descritos acima podem propiciar o desenvolvimento linguístico e o aumento da acuidade dos aprendizes, principalmente, quando o foco no item linguístico advém das necessidades desses aprendizes. É por isso que estudos que enfocam em *corpora* de aprendizes (c.f BERBER SARDINHA, 2001; GRANGER, 1998; GRIES; WULFE, 2005; LÚCIO, 2006; PINTO, 2008) podem colaborar para que tenhamos melhor conhecimento sobre as dificuldades dos aprendizes, e isso pode informar a preparação de atividades pedagógicas.

Mesmo não sendo fácil compilar dados de aprendizes (ver seção de metodologia sobre a formação do subcorpus do International *Corpus* of

---

<sup>5</sup> A análise de textos autênticos possibilita a observação de que algumas estruturas tendem a ocorrer frequentemente com certas palavras, permitindo a constatação de padrões léxico-gramaticais. “Esses padrões não são associações meramente arbitrárias; mais propriamente, estruturas gramaticais específicas frequentemente ocorrem com classes lexicais restritas porque tanto as estruturas quanto as classes lexicais servem às mesmas tarefas ou funções comunicativas.” (BIBER *et al.*, 1999, p. 13). Nossa tradução para “These patterns are not merely arbitrary associations; rather, particular grammatical structures often occur with restricted lexical classes because both the structures and the lexical classes serve the same underlying communicative tasks and functions.”

<sup>6</sup> Baseamo-nos em Partington (1998), Fox (1998), entre outros autores, que utilizam o termo cotexto para se referirem ao ambiente linguístico que acompanha o item linguístico em investigação. O termo contexto seria, então, utilizado para se referir ao ambiente situacional no qual o item é utilizado (FOX, 1998).

Learner English (ICLE) com redações de aprendizes brasileiros – Br-ICLE), o número de trabalhos na área tem crescido no Brasil e em diversos países. Ponderamos, entretanto, que estudos sobre *corpus* de aprendizes são recentes – da década de 90 em diante. O grande destaque entre as publicações nessa área é a coletânea de Granger (1998), que abrange discussões sobre metodologia, descrição e aplicação pedagógica desse tipo de *corpus*. Dentre as questões abordadas, podemos exemplificar as ferramentas computacionais para a análise de corpora de aprendizes (MEUNIER, 1998), o uso de conectivos adverbiais (ALTERNBERG; TAPPER, 1998) e a instrução com foco-na-forma, bem como o aprendizado movido por dados (GRANGER; TRIBBLE, 1998). Esses estudos têm o objetivo de descrever o uso da língua-alvo pelos aprendizes (HATZITHEODOROU; MATTHEOUDAKIS, 2008), além de, em alguns casos, visar também a fazer um diagnóstico da proficiência dos aprendizes, a fim de servir como subsídio de ações pedagógicas que utilizem os recursos da LC para propostas pedagógicas (c.f. JOHNS, 1994; MILTON, 1998).

Além disso, a compilação de *corpus* de aprendizes permite que conheçamos a interlíngua desses aprendizes e que façamos (a) um uso direto em sala de aula (SEIDLHOFER, 2002) com a análise das produções feitas pelos próprios alunos ou (b) um uso indireto. Nosso artigo enquadra-se na descrição da interlíngua de aprendizes de inglês para, posteriormente, propor atividades de conscientização linguística que podem fazer uso direto do *corpus* de aprendizes (CABrI). A área de descrição da interlíngua tem se mostrado bastante proveitosa, especialmente por contribuir com a compreensão sobre o que é a “produção de aprendizes” (*learner output*), que ainda se encontra um tanto quanto deficiente nos campos de pesquisa sobre o ensino de língua estrangeira. Dentre os elementos a serem analisados em um *corpus* de aprendizes estão seus erros. O erro, nessa perspectiva, é visto como parte do processo de aprendizagem, conforme também defendem vários autores (c.f. BERBER SARDINHA; SHEPHERD, 2008) que consideram os erros sistemáticos e passíveis de serem alvo de reflexão a respeito do desenvolvimento da interlíngua. Essa nova vertente da análise contrastiva com foco na interlíngua (GRANGER, 1998) é diferente da análise contrastiva tradicional, que comparava duas línguas, objetivando identificar as diferenças e, assim, prever erros. Os estudos que visam a investigar erros de alunos, embasados na LC, têm foco e método diferentes da Análise de Erro tradicional, priorizando-se o uso de construções linguísticas para que sobreuso e subuso sejam entendidos dentro de uma perspectiva não normativa, mas, sim, descritiva da

interlíngua. Nesse ponto, a comparação com o uso de estruturas com significados correspondentes produzidas por falantes nativos se faz necessária. Essa comparação é feita entre *corpora* com um grande número de dados, diferentemente do que ocorria nos estudos vinculados à Análise de Erro, cujo objeto de exame eram

coletâneas de fragmentos descontextualizados, selecionados de antemão, que exibiam certas características definidas *a priori*, no caso, erros, podendo se constituir em um apanhado de frases soltas ou meramente em uma lista de palavras e orações. (BERBER-SARDINHA, 2004, p. 267).

Como colocado anteriormente, a análise contrastiva da interlíngua (ACI) não é uma volta à análise contrastiva tradicional. A ACI propicia o levantamento de quais aspectos linguísticos merecem ser tratados pedagogicamente. Além disso, argumentamos que os estudos baseados em *corpus* podem revolucionar o ensino de gramática em três aspectos:

1. descrições monolíticas da gramática do inglês serão substituídas por descrições com base em registros específicos.
2. o ensino de gramática ficará mais integrado com o ensino de vocabulário.
3. a ênfase mudará da acuidade estrutural para as condições apropriadas de uso de construções gramaticais alternativas. (CONRAD, 2000, p. 549)<sup>7</sup>

De acordo com os objetivos deste artigo, propomo-nos a mostrar como Conrad pode ter razão quanto à potencialidade da LC para mudar os caminhos do ensino de gramática. Nosso estudo não prevê a descrição da gramática do inglês, mas se vale do conceito de registro para delimitar a formação de seu *corpus* de estudo (*corpus* de aprendizes constituído por redações) e também faz comparações com o registro acadêmico de um *corpus* de referência. Este artigo pretende mostrar como o ensino de gramática, aqui representado pelo ensino

---

<sup>7</sup> Nossa tradução para “1. Monolithic descriptions of English grammar will be replaced by register-specific descriptions. 2. The teaching of grammar will become more integrated with the teaching of vocabulary. 3. Emphasis will shift from structural accuracy to the appropriate conditions of use for alternative grammatical constructions” (CONRAD, 2000, p. 549).

do item linguístico *for*, lucra com a integração do ensino de vocabulário e, acima de tudo, o quanto o foco do ensino de gramática tem a ganhar ao promover no aluno a percepção sobre o uso de estruturas linguísticas.

Assim sendo, na próxima seção discorreremos acerca das metodologias utilizadas para a compilação de nosso *corpus* (CABrI) e para sua análise. Na seção seguinte, apresentamos alguns resultados que permitirão a ligação entre a LC e o ensino de línguas a partir da análise de agrupamentos com a palavra *for*. Discutiremos sobre possíveis atividades voltadas para o ensino da gramática integrado ao ensino de vocabulário – e que reforce a ideia de adequação das construções linguísticas para o registro e cotexto em questão.

## Metodologia

Os dados analisados são provenientes do *Corpus* de Aprendizes Brasileiros de Inglês (CABrI), composto de redações de alunos da graduação em Letras da UFMG (habilitação em inglês). Como este é um *corpus* ainda em formação, vale a pena ressaltar que, na época da análise dos dados, o *corpus* continha 36.187 palavras.

A maior parte dos *corpora* de aprendizes é de linguagem escrita, e o nosso *corpus*, CABrI, está sendo formado parcialmente de acordo com os princípios e padrões do *International Corpus of Learner English* (ICLE), cuja segunda versão (GRANGER; DAGNEAUX; MEUNIER; PAQUOT, 2009) contém 3.7 milhões de palavras provenientes de 16 subcorpora (Bulgária, China, Finlândia, Suécia, Itália, entre outros). Parte do CABrI, a correspondente às redações escritas por aprendizes de nível avançado, poderá contribuir com o subcorpus de produções de aprendizes brasileiros de inglês (Br-ICLE) que fará parte do ICLE. Berber Sardinha (2001) afirmou que o Br-ICLE tinha em 2001 20.259 palavras (cerca de 10% do total planejado – 200 mil palavras). Até maio de 2009, o *corpus* tinha dobrado (40.834 palavras). Isso demonstra a dificuldade dos pesquisadores em coletar os dados necessários. No Brasil, outra grande iniciativa para a formação de *corpora* de aprendizes é o *Corpus* Multilíngue de Aprendizes (CoMAprend), constituído de redações dos alunos da graduação e dos cursos de extensão das áreas do Departamento de Letras Modernas da Universidade de São Paulo: alemão, espanhol, francês, inglês e italiano (TAGNIN; FROMM, 2008).

Os procedimentos de coleta seguiram as indicações do ICLE. Primeiramente, os textos foram coletados de alunos do curso de Letras. Segundo, as redações foram escritas individualmente e fora da sala de aula.

Terceiro, os alunos podiam consultar dicionários e gramáticas, entre outros, indicando, em um formulário sobre seu histórico escolar, quais materiais tinham sido utilizados como apoio durante a escrita da redação. Quarto, cada redação deveria ter entre 500 e 1.000 palavras, e devia ser escrita por alunos de nível avançado ou intermediário alto. Parte das redações coletadas não atendeu a esse último requisito, por terem menos de 500 palavras e por terem sido escritas por alunos de níveis iniciante e intermediário. Sendo assim, 20 redações (12.082 palavras) poderão ser integradas ao Br-ICLE, e o restante apenas continuará como parte do CABrI, que tem como um de seus objetivos conhecer mais a fundo o nível linguístico dos alunos da graduação em inglês de nossa instituição, para que sejam realizadas intervenções pedagógicas compatíveis com as suas necessidades linguísticas.

Os dados foram analisados segundo quatro procedimentos. Em primeiro lugar, foi feita uma análise manual para a identificação de erros, entendidos como construções linguísticas que se distanciam dos padrões lexico-gramaticais mais frequentes no inglês. Segundo, utilizamos o *Error Tagger 1.0*, disponível no site <[www.corpuslg.org](http://www.corpuslg.org)>, para a identificação das palavras com maior probabilidade de serem usadas erroneamente no *corpus*. Apesar de o *Error Tagger 1.0* não possuir alta precisão (BERBER SARDINHA; SHEPHERD, 2008), resolvemos utilizá-lo, pois, por meio dele, foi possível comparar os resultados com aqueles obtidos pela análise manual e pelo programa *WordSmith Tools*. Terceiro, utilizamos também, em diversos momentos da análise, o programa *WordSmith Tools* para extrair a frequência de palavras, colocações e coligações, e analisar as linhas de concordâncias nas quais determinadas construções linguísticas ocorriam. Quarto, comparamos nossos resultados com o *corpus* geral COCA (*Corpus of Contemporary American English*) e com os *corpora* de aprendizes LOCNESS (*Louvain Corpus of Native English Essays*) e Br-ICLE (subcorpus do ICLE coletado em outras instituições que não a UFMG). Essas comparações também foram feitas em momentos distintos da pesquisa, a fim de verificar se havia sobreuso ou subuso de determinadas construções linguísticas.

O trabalho ora descrito faz comparações entre *corpora* produzidos em situações similares. Quando analisamos as escolhas feitas pelos aprendizes de inglês, comparamo-las às produções de falantes nativos de inglês que escreveram redações argumentativas (LOCNESS) e que escreveram textos acadêmicos (e.g. ao fazermos buscas no COCA, levamos em consideração apenas as ocorrências no registro acadêmico, e não em todos os registros desse

*corpus* geral de inglês). Quanto às comparações entre nosso *corpus* (CABrI) e o Br-ICLE, elas são justas também, uma vez que ambos os *corpora* são compostos de redações argumentativas produzidas por alunos de graduação com habilitação em inglês e de faixa etária semelhante. A única diferença reside no nível de proficiência de alguns aprendizes.

A partir de nossos resultados, que serão explicitados a seguir, verificamos que, principalmente para alunos de graduação em Letras – Inglês (futuros professores ou pesquisadores), a utilização de linhas de concordância provenientes das suas próprias produções textuais como material de atividades com foco-na-forma pode se mostrar bem proveitosa (SEIDLHOFER, 2002). Os alunos podem ter a oportunidade de reexaminar sua escrita, atentando-se aos padrões de colocações por eles utilizados e comparando com os padrões que são produzidos por falantes nativos em contextos semelhantes. Dessa forma, eles são encorajados a desenvolver uma postura mais autônoma diante do seu processo de aprendizagem, pois saberão realizar análises, fazendo uso de recursos da LC, a fim de refletir o quão próxima ou distante está sua interlíngua do que os nativos produzem.

### **Resultados obtidos e possíveis intervenções pedagógicas**

Nesta seção, discutiremos os resultados obtidos, ressaltando o percurso de nossas tentativas de análise até chegarmos aos agrupamentos com a palavra *for*, quais sejam: verbo + *for* e substantivo + *for*.<sup>8</sup> As sugestões de atividades pedagógicas baseadas em *corpus* seguirão a discussão dos resultados de cada agrupamento.

A primeira análise das redações foi manual e buscou identificar erros produzidos pelos aprendizes brasileiros de inglês quando da escrita das redações argumentativas. Detectamos diversos erros e, entre eles, chamou-nos a atenção o uso inapropriado de conjunções subordinadas como *after*, *before*, *since*, *while*, *whereas*, e também da palavra *for*. Ao compararmos a frequência dessas mesmas palavras no CABrI e no *corpus* de redações argumentativas escritas por falantes nativos (LOCNESS), constatamos que não havia diferença na frequência de uso. Esse resultado fez-nos perceber como nossa escolha dos itens linguísticos a serem analisados foi preponderantemente guiada pelo fator

---

<sup>8</sup> Outros agrupamentos com a palavra *for* constam de nosso *corpus* de estudo (CABrI), porém, por limite de espaço, optamos por apresentar e discutir somente verbo + *for* e substantivo + *for*.

saliência, em detrimento do fator frequência. A fim de tentarmos encontrar um objeto mais definido de análise, decidimos utilizar o etiquetador de erros – *Error Tagger*, – versão 1.0. *For* apareceu entre as dez mais fortes candidatas (mais especificamente na nona posição – ver TAB. 1) a ser usada inadequadamente no *corpus* (CABrI). Ainda assim, o fator saliência continuou prevalecendo, uma vez que, conforme dito, já havíamos nos atentado para o uso inadequado de *for* em nossa análise manual.

TABELA 1  
As 10 mais fortes candidatas ao erro segundo o *Error Tagger 1.0*

	Word	Tag	Word (prob)	Left bundle	Right bundle	Framework	Range	Word Class
1	the	0.6040	0.1431	1.00	1.00	1.00	0.3974	0.0837
2	it	0.5321	0.0451	1.00	1.00	1.00	0.0641	0.0837
3	to	0.5280	0.0354	1.00	1.00	1.00	0.0897	0.0431
4	a	0.5255	0.0310	1.00	1.00	1.00	0.0384	0.0837
5	that	0.5220	0.0228	1.00	1.00	1.00	0.0256	0.0837
6	are	0.5176	0.166	1.00	1.00	1.00	0.0384	0.0507
7	with	0.5142	0.0295	1.00	1.00	1.00	0.0128	0.0431
8	technology	0.5126	0.0238	1.00	1.00	1.00	0.0128	0.0392
9	for	0.5122	0.0176	1.00	1.00	1.00	0.0128	0.0431
10	as	0.5100	0.0045	1.00	1.00	1.00	0.0128	0.0431

Em nossa explicitação, propomos maneiras de se lançar mão dos recursos da LC no ensino da gramática a partir da descrição dos nossos resultados, os quais se baseiam em análises feitas com o programa *WordSmith Tools 5.0*, e em comparações com os *corpora* de aprendizes LOCNESS; Br-ICLE, e com a seção de textos acadêmicos que compõe o *corpus* geral COCA.

O terceiro passo da análise foi verificar no programa *WordSmith Tools 5.0* o número de linhas de concordância nos quais *for* aparecia nos *corpora* CABrI; Br-ICLE; LOCNESS. Os números da TAB. 2 poderiam significar a ocorrência de um sobreuso de *for* pelos aprendizes brasileiros de inglês que contribuíram para a formação do CABrI? A resposta parece ser negativa, haja vista que é, no mínimo, arriscado fazer tal afirmação, partindo de números demasiado gerais. Nosso próximo passo foi analisar as diversas construções linguísticas compostas pela palavra *for*. Com base nesses resultados, concentramo-nos nas atividades de ensino de gramática que podem ser desenvolvidas com aprendizes que almejam ter acuidade na escrita acadêmica.

TABELA 2  
Corpora e quantidade de linhas de concordância com *for*

<i>Corpus</i>	Total de palavras no <i>corpus</i>	Linhas de concordância com <i>for</i>	Porcentagem das linhas de concordância
CABrI	36.187	413	1,1412%
Br-ICLE	40.834	335	0,8203%
LOCNESS	324.072	3145	0,9705%

### Agrupamentos verbo + *for*

Os agrupamentos verbais com *for* presentes no CABrI foram: *work for*, *wait for*, *search for*, *live for*, *prepare + substantivo + for*, *consider + substantivo + for*, *contribute for*, *spread for* e *stand for*. Atestamos que o uso dos cinco primeiros agrupamentos verbais foram apropriados. Decidimos discutir como, com recursos da LC, podemos levar os alunos a descobrirem os padrões que acompanham *prepare + substantivo + for*, *prepare to* e *consider + substantivo + for*.

Como o verbo *prepare* ocorre também com a preposição *to*, comparamos suas ocorrências com as ocorrências do agrupamento *prepare + substantivo + for*. Aparentemente, poderíamos argumentar que houve sobreuso de *prepare + substantivo + for* no CABrI, porém esse fato é possivelmente explicado por uma das propostas de tema de redação oferecida aos aprendizes (“Most university degrees are theoretical and do not **prepare students for** the real world. They are therefore of very little value.”), que inclui o agrupamento em questão. A ocorrência de ambos os agrupamentos nos outros *corpora* de aprendizes (Br-ICLE; LOCNESS) é muito baixa ou inexistente (TAB. 3). Quanto à comparação com as ocorrências no registro acadêmico do COCA, há 585 ocorrências de *prepare for* e apenas 115 de *prepare to*, demonstrando o uso apropriado pelos aprendizes brasileiros que, de fato, utilizaram *prepare for* em maior número.

TABELA 3  
Porcentagem de ocorrência em cada *corpus*

<i>Corpus</i>	<i>prepare + substantivo + for</i>	<i>Prepare + to</i>
CABrI	0,0967%	0,0359%
Br-ICLE	0,0171%	0,0171%
LOCNESS	0,0022%	--

Por mais que os alunos tenham sido bem-sucedidos no uso desse agrupamento (*prepare + substantivo + for*), uma atividade possível de ser realizada teria como foco o contraste entre os tipos de colocados à direita de *prepare + for* e *prepare + to*. Exemplos do COCA (ver FIG. 1) seriam mostrados e os alunos seriam guiados a perceber que *prepare for* é geralmente acompanhado por um substantivo, enquanto *prepare to* coloca-se com um verbo. Tal atividade, aparentemente voltada apenas para a forma, pode também estender sua função e se mostrar útil para explorar o ensino de vocabulário e, até mesmo, o uso de sufixos em inglês. Na primeira linha de concordância abaixo, a palavra *learning* segue o agrupamento *prepare for*. Seria, portanto, interessante investigar se os alunos estão cientes de que o sufixo *-ing* é também formador de substantivos, e não só do gerúndio no inglês.

#### **prepare for**

- \* By following this approach children can progress and **prepare for** *learning*.
- \* When you **prepare for** *trial*, don't write out your questions. They'll sound stiff and...
- \* ...contacted the chief of police and the military to alert them to **prepare for** possible civil unrest;
- \* "...on a cold morning in March 1994 dozens of Hualapais woke up at 6:00 a.m. to **prepare for** a run that would take them nearly two hundred miles along the Colorado River";
- \* "As Dent and elections officials across the nation **prepare for** this November's presidential vote, they also increasingly find themselves..."

#### **prepare to**

- \* With the largest of Etowah's mounds in the background, archaeologists **prepare to** *survey* a section of the site
- \* As I **prepare to** *descend*, the three young women are already safely down
- \* ...the senior year experience represents a critical developmental period as undergraduate students **prepare to** make the difficult transition...
- \* ...and became convinced that city employees should **prepare to** take over the sales program.
- \* ...are some of the most influential issues facing community college leaders as they **prepare to** enter the next millennium.

FIGURA 1 – Dados originários do registro acadêmico do COCA

Quanto ao segundo agrupamento verbal a ser discutido (*consider* + substantivo + *for*), parece ter havido sobreuso dele no CABrI (0,0193% de ocorrência no *corpus*), tendo em vista que não houve nenhuma ocorrência desse agrupamento nos outros *corpora* analisados.

NConcordance

1 about this issue. This essay will **consider** arguments for the importance of  
2 about this issue. This essay will **consider** arguments for the supporting the  
3 about this issue. This essay will **consider** arguments for regarding the  
4 about the issue. This essay will **consider** arguments for the supporting the  
5 involving teaching degrees, should **consider**, for instance, amid many other  
6 about this issue. This essay will **consider** arguments for theoretial  
7 about this issue. This essay will **consider** arguments for the importance of  
8 people's questions. This essay will **consider** arguments for controlling the

**FIGURA 2 – Linhas de concordância com *consider* + substantivo + *for* no CABrI**

Pode ser que o sobreuso de *consider* + substantivo + *for* tenha origem na orientação que os aprendizes recebem na disciplina de escrita acadêmica da UFMG que eles cursavam na época da coleta de dados. Nesse curso, os alunos são incentivados a utilizarem porções (*chunks*) ou pacotes lexicais (*lexical bundles*) que facilitem a escrita de textos acadêmicos para que, ao realizarem um exame internacional, por exemplo, eles se sintam mais seguros e aptos a escreverem de maneira ágil. Todavia, conforme atestado, muito embora gramaticalmente correto e também compatível com o registro acadêmico, tal agrupamento não parece ter grande relevância em termos de uso. Seria o caso de se elaborar uma atividade para que os alunos investigassem a frequência e o modo de uso de outros agrupamentos frequentes no discurso acadêmico, tais como *be considered as*, *be argued that*, incluídos na Lista de Fórmulas Acadêmicas – baseada em *corpora* de textos acadêmicos– e que propõe a seguinte categorização funcional dessas fórmulas ou pacotes lexicais: expressões referenciais, expressões de posicionamento e expressões de organização discursiva (SIMPSON-VLACH; ELLIS, 2010). Para isso, eles teriam de ler um determinado número de textos acadêmicos, publicados em revistas de suas áreas de estudos, a fim de observar possíveis recorrências de expressões típicas desse registro com a função pragmática que eles necessitam para, posteriormente, verificar em *corpora online* de registro acadêmico a frequência (e.g. COCA) e os padrões de uso desses agrupamentos.

#### 4Agrupamento substantivo + for

Os agrupamentos substantivo + *for* ocorreram com baixa frequência no CABrI. Contudo, essas ocorrências ficaram salientes em nossos dados, devido, em certos casos, à sua clara adequação ou inadequação. O primeiro agrupamento é *opportunity + for*, que foi usado pelos participantes do CABrI, e também ocorreu no LOCNESS (*corpus* de redações produzidas falantes nativos de inglês – FIG. 3), porém não apareceu no Br-ICLE.

NConcordance

1 families are together. There is the most **opportunity** for frustration and therefore  
2 new jobs. It would be a great a great **opportunity** for even more teenagers and  
3 how much one is willing to give for the **opportunity** for something better.  
4 de Gaulle's emergency powers with no **opportunity** for parliamentary debate.  
5 At this stage we can already see the **opportunity** for a gap between primary &  
6 to learn more. This is because the **opportunity** for non-skilled employment is  
7 This renaissance period was an **opportunity** for Da Vinci and others to

FIGURA 3 – Linhas de concordância de *opportunity + for* extraídas do LOCNESS

O uso de *opportunity + for* pelos aprendizes brasileiros se mostrou apropriado. Consideramos interessante o fato de que as linhas de concordância do CABrI (FIG. 4) apresentaram as colocações *provide + opportunity + for* e *offer + opportunity + for*. No Br-ICLE e no LOCNESS, no entanto, não há ocorrências dessas colocações. Decidimos, então, investigar o uso dessas ocorrências na seção do registro acadêmico do COCA, e constatamos que *opportunity + for* é um agrupamento muito utilizado (2.060 ocorrências). Ainda investigando o COCA, chegamos a outras constatações: (a) enquanto a colocação *offer opportunity for* obteve apenas 1 ocorrência, *offer opportunities for* totalizou 31 ocorrências; (b) da mesma forma, *provide opportunity for* obteve 13 ocorrências, enquanto *provide opportunities for* obteve o total de 234 ocorrências.

NConcordance

1 difficult to get. My university offers this **opportunity** for most of its students. We  
2 are also not able to provide this **opportunity** for the students to put in

FIGURA 4 – Linhas de concordância de *opportunity for* extraídas do CABrI

Os alunos que contribuíram para o CABrI utilizaram os agrupamentos corretamente, porém seria de grande valia que eles descobrissem que *opportunities* (no plural) é mais usado para ambos os agrupamentos, e que a frequência de *provide opportunities for* é bem maior que *offer opportunities*. Logo, o professor poderia, novamente, atrelar o ensino da gramática ao ensino de vocabulário, mostrando linhas de concordância de um *corpus*, por exemplo, do COCA, para que os alunos sejam levados a descobrir a frequência os colocados de *opportunity for* e o porquê do maior uso da palavra em questão (*opportunity*) no plural nesses agrupamentos.

Outro agrupamento formado por um substantivo + *for* que chamou nossa atenção foi *necessity + for*, não por ser este um agrupamento não adequado na língua inglesa, mas pelo fato de o seu colocado à direita ter sido um verbo no gerúndio (*accomplishing*). Tal estrutura (*necessity + preposição + verbo no gerúndio*), portanto, configura-se como um ambiente linguístico que favorece o uso de *of*, e não de *for*. Isso poderia ser explorado em uma atividade que induziria os alunos a fazerem uma consulta no registro acadêmico do COCA, com o objetivo de testar diferentes preposições até perceberem quais são as coligações adequadas e quais são suas frequências de uso, como visto brevemente na FIG. 5.

*necessity for* = 369 ocorrências

(e.g. “...including the necessity for adequate corporate governance structures”; “This is the “keystone” for lifelong learning, an absolute necessity for success in the 21st century”; “emphasizing that Judicious Discipline is a necessity for students to feel safe, to be protected against dangers and to show respect.”; “All reports agreed on the necessity for prompt removal on detection of such migration in view of the possible injury to...”)

*necessity of* = 924 ocorrências

(e.g. “...the fact that the necessity of developing student’s critical thinking is proposed by academics...”; “Some in positions of responsibility spoke out for the necessity of change. Sister Maureen Fiedler (who now hosts the radio program Interfaith Voices...”; “For example, a spirited debate is occurring about the necessity of emphasizing evidence-based practices in behavioral health...”; “...their teachers can build on the pillars and extend them rather than face the necessity of starting over from scratch.”)

FIGURA 5 – Dados extraídos do registro acadêmico do COCA

O último agrupamento formado por um substantivo + *for* que comentaremos é *key + for*. Foi registrada uma inadequação no uso de *key + for* pelos participantes do CABrI, tendo em vista que nesse *corpus* não há ocorrências de *key + to*, enquanto no Br-ICLE e no LOCNESS só há ocorrências de *key + to*, e nenhuma de *key + for*. Além disso, analisando os colocados à direita (*key + to + determinante + substantivo; key + to + verbo no gerúndio*) em alguns exemplos aleatórios do registro acadêmico do COCA – FIG. 6, percebemos que os participantes do CABrI deveriam ter optado pela preposição *to*, em vez de *for*.

*key for* = 72 ocorrências

(e.g. The key for diagnosis of lupus vulgaris is the awareness of tuberculous etiology in any chronic ulcers; of these words and phrases and knowing when and how to structure this vocabulary is key for students who are wishing to understand the concepts)

*key to* = 1.795 ocorrências

(e.g. and that trust among the top officials is a key to effective policymaking; In arguing for his approach Florida also contends that tolerance is key to attracting and retaining the creative class.)

#### FIGURA 6 – Dados extraídos do registro acadêmico do COCA

Uma atividade similar àquela proposta para *necessity for* poderia ser também utilizada para *key for*, a fim de que os aprendizes pudessem ser induzidos a verificar quais colocados são realmente usados com esse agrupamento e com qual frequência.

### Conclusão

Concluimos que a palavra *for*, classificada como a nona palavra a ser uma forte candidata ao erro, pelo *Error Tagger*, nos textos do CABrI, pode tanto contribuir para formações de itens linguísticos que são usados adequada e não adequadamente pelos aprendizes brasileiros de inglês que forneceram suas redações argumentativas para esse *corpus*. Utilizando outras ferramentas da LC, como o programa *WordSmith Tools*, analisamos o CABrI, o Br-ICLE e o LOCNESS, além de acessar *online* o COCA, para melhor compreendermos o uso efetivo das construções analisadas.

Dentre os agrupamentos cujo uso se aproximou daquele dos falantes nativos, estão os agrupamentos verbais preposicionados: *prepare* + substantivo + *for*, *work for*, *wait for*, *search for*, *live for* e o substantivo preposicionado *opportunity for*. Por outro lado, foi constatado o sobreuso de *consider* + substantivo + *for* que pode ter sua origem nas orientações sobre escrita acadêmica do curso que os participantes frequentavam na época da coleta de dados. Houve também a constatação do uso inadequado de *for* com os verbos *contribute* e *spread*, cujos agrupamentos com a preposição *to* (*contribute to* e *spread to*) são muito mais frequentes nos *corpora* de falantes nativos.

As sugestões de atividades pedagógicas com a utilização de recursos da LC visam à descoberta dos padrões léxico-gramaticais que ficam evidentes por meio de linhas de concordância. Primeiramente, cabe ao professor preparar atividades que induzam seus alunos a comparar suas escolhas léxico-gramaticais com aquelas feitas por falantes nativos em contexto de produção similar ao seu. Segundo, o professor deve guiá-los a descobrir os colocados mais frequentes da palavra ou agrupamento lexical em questão. Terceiro, o professor deve ressaltar que a gramática não está dissociada do léxico e que o uso de certos agrupamentos é determinado pelo contexto e cotexto. O acesso a *corpora* gratuito *online* é uma realidade, porém cabe ao professor preparar seus alunos utilizar esse tipo de recurso. A princípio, atividades com linhas de concordância podem ser feitas em sala de aula com material impresso, até que os aprendizes se sintam familiarizados com as descobertas que podem fazer. Posteriormente, os aprendizes podem desfrutar das ferramentas que acompanham *corpora online* ou até mesmo de programas como *WordSmith Tools*. Ressaltamos que o trabalho de conscientização linguística sugerido aqui está baseado na abordagem Aprendizado Movido por Dados (JOHNS, 1994), que pode ser de grande valia aos aprendizes que necessitem atingir um alto nível de proficiência linguística, o que é o caso daqueles que participaram da coleta de dados para nosso *corpus* (CABrI).

## Referências bibliográficas

- AARTS, J. Does corpus linguistics exist? Soem old and new issues. In: BREIVIK, L.E.; HASSELGREN, A. (Ed.). *From the COLT's mouth ... and others: Language corpora studies in honour of Anna-Brita Stenström*, Amsterdam: Rodopi, 2002. p.1-16. (Republicado. In: TEUBERT, W.; KRISHNAMURTHY, R. (Ed.). *Corpus Linguistics: critical concepts in Linguistics*. New York: Routledge, 2007.)
- ALTENBERG, B.; TAPPER, M. The use of adverbial connectors in advanced Swedish learners' written English. In: GRANGER, S. (Ed.). *Learner English on Computer*. London & New York: Addison Wesley Longman, 1998. p. 80-93.
- BERBER SARDINHA, A. P. O *corpus* de aprendiz Br-ICLE. *Intercâmbio*, v. 10, p. 227-39, 2001.
- BERBER SARDINHA, A. P. *Linguística de corpus*. Barueri: Manole, 2004.
- BERBER 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, 2008. p.257-262.
- BIBER, D. *Variation across speech and writing*. Cambridge: Cambridge, 1988.
- BIBER, D; CONRAD, S.; REPPEN, R. *Corpus Linguistics: Investigating language structure and use*. Cambridge: Cambridge, 1998.
- BIBER, D.; JOHANSSON, S.; LEECH, G.; CONRAD, S.; FINEGAN, E. *Longman grammar of spoken and written English*. Essex:Longman, 1999.
- CARTER, R.; MCCARTHY, M. *Cambridge Grammar of English*. Cambridge: Cambridge, 2006.
- CONRAD, S. Will Corpus Linguistics revolutionize grammar teaching in the 21<sup>st</sup> century? *TESOL Quarterly*, v. 34, n. 3. p. 548-560, 2000.
- DOUGHTY, C.; VARELA, E. Communicative focus on form. In: DOUGHTY, C.; WILLIAMS, J. (Ed.). *Focus on form in classroom second language acquisition*. Cambridge: Cambridge, 1998. p. 114-138.
- DOUGHTY, C.; WILLIAMS, J. (Ed.). *Focus on form in classroom second language acquisition*. Cambridge: Cambridge, 1998.
- DUTRA, D. P. Conscientização linguística com base em corpora *online*. *Revista Intercâmbio*, v. XX, p. 79-98, 2009.
- FOX, G. Using *corpus* data in the classroom. In: TOMLINSON, B. *Materials development in language teaching*. Cambridge: Cambridge, 1998. p. 25-43.

- GRANGER, S. The computer learner *corpus*: a versatile new source of data for SLA research. In: GRANGER, S. (Ed.). *Learner English on Computer*. London & New York: Addison Wesley Longman, 1998. p. 3-18.
- GRANGER, S.; TRIBBLE, C. Learner corpus data in the foreign language classroom: form-focused instruction and data-driven learning. In: GRANGER, S. (Ed.). *Learner English on Computer*. London & New York: Addison Wesley Longman. 1998. p. 199-210
- GRANGER, S.; DAGNEAUX, E.; MEUNIER, F.; PAQUOT, M. *International Corpus of Learner English – Version 2*. Louvain-la-Neuve: UCL Presses Universitaires de Louvain, 2009
- GRIES, S.; WULFF, S. Do foreign language learners also have constructions? Evidence from priming, sorting, and *corpora*. *Annual Review of Cognitive Linguistics*. v.3, p. 182-200, 2005.
- HATZITHEODOROU, A.; MATTHEODAKIS, M. Exploring the marking of stance in argumentative essays written by EFL learners and native speakers of English. *Anais do 8th Teaching and Language Corpora Conference*. Lisboa: Associação de Estudos e de Investigação Científica do ISLA-Lisboa. 2008. p.133-141.
- 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.
- LARSEN-FREEMAN, D. *Teaching language: from grammar to grammaring*. Boston: Thomson/Heinle, 2003.
- LEECH, G. Varieties of English grammar: the state of the art from the grammarian's point of view. In: BALD, W-D. (Ed.). *Kernprobleme der englischen Grammatik: Sprachliche Fakten und ihre Vermittlung*. München: Langenscheidt-Longman, 1988. p. 5-17.
- 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 Lingüística Aplicada) – Programa de Pós-Graduação em Lingüística Aplicada ao Ensino de Línguas, PUC-SP, São Paulo, 2006.
- LYONS, J. *Language and Linguistics: an introduction*. Cambridge: Cambridge, 1981.
- 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.

- MEUNIER, F. *Computer tools for the analysis of learner corpora*. In: GRANGER, S. (Ed.). *Learner English on Computer*. London & New York: Addison Wesley Longman, 1998. p. 19-38.
- MILTON, J. Exploiting L1 and interlanguage *corpora* in the design of an electronic language learning and production environment. In: GRANGER, S. (Ed.). *Learner English on Computer*. London & New York: Addison Wesley Longman, 1998. p.186-198.
- O'KEEFFE, A.; MCCARTHY, M.; CARTER, R. *From corpus to classroom: language use and language teaching*. Cambridge: Cambridge, 2007.
- PARTINGTON, A. *Patterns and meanings: using corpora for English language research and teaching*. Amsterdam: John Benjamins. 1998.
- 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, 2008.
- RICHARDS, J. Addressing the grammar gap in task work. In: *Methodology in language teaching*. Cambridge: Cambridge, 2002. p. 153-166.
- SEIDLHOFER, B. Pedagogy and local learner *corpora*: Working with learningdriven data. In: GRANGER, S.; HUNG, J.; PETCH-TYSON, S. (Ed.). *Computer learner corpora, second language acquisition and foreign language teaching*. Amsterdam: Benjamins, 2002. p. 213–234.
- SCOTT, M.; TRIBBLE, C. *Textual patterns: key words and corpus analysis in language education*. Amsterdam: John Benjamins, 2006.
- SIMPSON-VLACH, R; ELLIS, N. *Applied Linguistics*, p. 1-26, 2010. (Publicação com acesso prévio online doi:10.1093/applin/amp058)
- TAGNIN, S.; FROMM, G. CoMAprend a experiência da construção de um *corpus* de aprendiz para estudos. *Domínio de Linguagem*, 2008.
- WESTNEY, P. Rules and pedagogical grammar. In: *Perspectives on pedagogical grammar*. ODLIN, T. (Ed.). Cambridge: Cambridge, 1994. p. 72-96.

Recebido em 4 de abril de 2010. Aprovado em 17 de maio de 2010.