

USO DE VERBOS MODAIS POR ALUNOS AVANÇADOS DE INGLÊS COMO LÍNGUA ESTRANGEIRA: UMA INVESTIGAÇÃO BASEADA EM *CORPUS*¹

Vander VIANA² (PUC-Rio)

RESUMO: A Lingüística de *Corpus* tornou-se uma grande tendência em Lingüística Aplicada desde a segunda metade do século XX devido às facilidades computacionais. Atualmente professores podem pesquisar e avaliar a produção de seus alunos por meio da compilação de *corpora* de aprendiz. Este artigo descreve como esta técnica foi utilizada para investigar o uso de verbos modais na escrita de alunos avançados de inglês como língua estrangeira oriundos de cursos livres na Cidade do Rio de Janeiro. Quando o *corpus* de pesquisa é comparado ao registro prosa acadêmica estudado por Biber et alii (1999), a divergência se torna aparente. Os resultados sugerem que os sujeitos escrevem de uma forma improficiente, contrariando as expectativas iniciais.

ABSTRACT: Corpus Linguistics has become a major trend in Applied Linguistics since the second half of the 20th century due to computing facilities. Nowadays teachers can research and assess their students' production by means of compiling learner corpora. This article describes how this technique was used to investigate the usage of modals in the writing of advanced EFL learners studying at private language schools in Rio de Janeiro. When the research corpus is compared to the academic prose register studied by Biber et alii (1999), the divergence becomes apparent. The findings, thus, suggest that subjects write in a non-proficient way, which runs counter to previous expectations.

1. Introdução

Investigações baseadas em *corpora* já não são novidade na área de Lingüística Aplicada. No entanto, apesar de coleções de dados reais poderem ser utilizadas na área de ensino/aprendizado de línguas (cf. SANTOS, 2001), grande parte dos professores de línguas não está ciente de seu potencial. Surge, então, a necessidade de pesquisadores pensarem formas de como *corpora* podem contribuir para o ensino de línguas. A bem da verdade, dados de uso da língua obtidos a partir de grandes conjuntos de textos já fazem parte do cotidiano, por exemplo, de professores de língua inglesa. A título de ilustração, pode-se citar a recém-lançada coleção intitulada *Top Notch*. Na metodologia proposta para o curso, Saslow et alii (2006, p. Txiii) explicam que a série oferece informação concisa e útil a respeito de frequência, colocações e emprego típico de certas expressões por falantes de inglês como primeira língua a partir da análise de *corpus*. Materiais como este podem ajudar professores a entenderem um pouco mais as vantagens de utilização dos princípios da Lingüística de *Corpus* na sala de aula.

Uma outra possível aplicação destes princípios concerne o mapeamento da escrita de alunos em qualquer estágio do processo de ensino/aprendizagem conforme comentado por Leech (1998, p. xiv) no prefácio ao livro *Learner English on Computer*:

suponhamos que uma professora X, em um país cuja primeira língua não é o inglês, ensine inglês aos seus alunos toda semana e regularmente solicite que eles escrevam redações ou realizem outras tarefas escritas em inglês. Agora, em vez de devolver as redações aos alunos com comentários e um suspiro de alívio, ela armazena as redações (com a permissão dos alunos, é claro) em seu computador, compilando gradualmente, semana por semana, uma coleção maior e mais representativa do trabalho de seus alunos. Auxiliada por ferramentas computacionais como um concordanciador, ela pode extrair dados e frequência deste 'corpus', e pode analisar a progressão de seus alunos como um grupo em certa profundidade.³

¹ Esta é uma versão do trabalho final apresentado à Profa. Dra. Maria Carmelita Pádua Dias como requisito parcial para aprovação na disciplina Introdução à Lingüística Computacional, cursada no 1º semestre de 2006 na Pontifícia Universidade Católica do Rio de Janeiro. Agradeço às Profas. Dras. Maria Carmelita Pádua Dias e Sonia Zyngier por seus comentários críticos às versões anteriores deste texto.

² Endereço de e-mail: vander.viana@terra.com.br.

³ Tradução livre do seguinte fragmento: "let us suppose that higher education teacher X, in a non-English speaking country, teaches English to her students every week, and every so often sets them essays to write, or other written tasks in English. Now, instead of

Neste fragmento, Leech revela a possibilidade que professores têm de poderem fazer pesquisas em Lingüística de *Corpus*, o que talvez possa diminuir o hiato existente entre ensino e pesquisa. Além disto, os resultados obtidos são de duas naturezas distintas: preenchem uma lacuna no panorama científico ao mesmo tempo em que auxiliam professores a estruturarem suas práticas pedagógicas.

O presente estudo faz uso da Lingüística de *Corpus* para investigar uma área específica da gramática da língua inglesa, a saber, verbos modais. O objetivo aqui é analisar o modo como estudantes brasileiros avançados de inglês como língua estrangeira de cursos livres utilizam verbos modais em suas composições e contrastar os resultados com os obtidos por Biber et alii (1999) no mapeamento da produção oral e escrita de falantes de inglês como primeira língua.

As questões principais que guiam o presente estudo são as seguintes:

- (a) Os estudantes brasileiros avançados de inglês como língua estrangeira utilizam verbos modais em sua escrita de forma semelhante àquela de falantes de inglês como primeira língua?
- (b) Caso negativo, quais são as diferenças entre estes dois grupos?

O presente artigo estrutura-se em quatro partes principais. Inicialmente, faz-se uma revisão da literatura com vistas à explicitação de conceitos como Lingüística de *Corpus* e verbos modais. Posteriormente, discute-se o procedimento metodológico empregado tanto na coleta quanto no tratamento de dados. Em um terceiro momento, são apresentados os resultados encontrados nesta pesquisa. Finalmente, algumas considerações finais são tecidas com base no estudo realizado.

2. Revisão da literatura

Esta seção compreende dois principais aspectos enfocados neste trabalho: a Lingüística de *Corpus* e os verbos modais em língua inglesa. A primeira subseção explica alguns dos principais termos utilizados neste artigo. Na segunda parte, os verbos modais são apresentados a partir da perspectiva de três gramáticas diferentes. Ainda nesta subseção, faz-se a revisão de dois estudos baseados em *corpora* que lidam, direta ou indiretamente, com o uso de verbos modais por aprendizes de inglês como língua estrangeira.

2.1 Lingüística de *Corpus*

A noção de *corpus* como uma “coleção de textos orais ou escritos”⁴ (WEHMEIER, 2000, p. 295) já existe há muito tempo, mas foi somente no século XX que esta palavra assumiu um novo significado na área de Lingüística. Como observam McEnery e Wilson (1996, p. 21),

em princípio, qualquer coleção de mais de um texto pode ser chamada de corpus: o termo ‘corpus’ é simplesmente a palavra latina para ‘corpo’; um corpus, portanto, pode ser definido como qualquer corpo de texto. [...] Porém, quando utilizado no contexto da Lingüística moderna, o termo ‘corpus’ tende freqüentemente a ter conotações mais específicas do que esta definição oferece.⁵

Um *corpus*, de acordo com Tognini-Bonelli (2001, p. 2), é “uma coleção de textos presumidamente representativa de uma dada língua que é compilada para que possa ser utilizada para a análise lingüística”⁶.

A Lingüística de *Corpus* se insere na área de Lingüística Computacional assim como o Processamento de Linguagem Natural (doravante PLN). Porém, enquanto o PLN compreende “o estudo da linguagem voltada para a construção de *softwares* e sistemas computacionais específicos, como tradutores automáticos, *chatbots*, *parsers*, reconhecedores automáticos de voz, geradores automáticos de resumos etc.” (OTHERO e MENUZZI, 2005, p. 23-24), a Lingüística de *Corpus* pode ser descrita como “o estudo da linguagem baseada em exemplos ‘reais’ de uso da mesma”⁷ (MCENERY e WILSON, 1996, p. 1).

returning those essays to students with comments and a sigh of relief, she stores the essays (of course with the students’ permission) in her computer, and is gradually building up, week by week, a larger and more representative collection of her students’ work. Helped by computer tools such as a concordance package, she can extract data and frequency information from this ‘corpus’, and can analyse her students’ progress as a group in some depth”.

⁴ Tradução livre do seguinte fragmento: “collection of written or spoken texts”.

⁵ Tradução livre do seguinte fragmento: “in principle, any collection of more than one text can be called a corpus: the term ‘corpus’ is simply the Latin for ‘body’, hence a corpus may be defined as any body of text. [...] But the term ‘corpus’ when used in the context of modern linguistics tends most frequently to have more specific connotations than this simple definition provides for”.

⁶ Tradução livre do seguinte fragmento: “a collection of texts assumed to be representative of a given language put together so that it can be used for linguistic analysis”.

⁷ Tradução livre do seguinte fragmento: “as the study of language based on examples of ‘real life’ language use”

É exatamente esta característica específica do estudo de linguagem natural que distingue a Lingüística de *Corpus* dos estudos lingüísticos baseados na tradição Chomskyana. Enquanto a primeira área se ocupa da linguagem em uso, a segunda se apóia em exemplos artificiais e inventados. Segundo Sinclair (2003, p. ix), “antes da disponibilização de grandes quantidades de dados, muitas das generalizações tiveram que ser feitas por suposições intuitivas; não era possível verificar tais noções na era anterior à Lingüística de *Corpus*”⁸. Em outras palavras, a Lingüística de *Corpus* lida com a probabilidade do uso lingüístico enquanto a visão racionalista está interessada em abstrações lingüísticas. Até então estas áreas permanecem com duas formas diferentes de olhar a linguagem e ainda não se chegou a um denominador comum.

O presente estudo parte da análise da produção escrita de aprendizes. Ao utilizar tais dados, professores podem desenvolver suas próprias pesquisas e descobrir, por exemplo, que aspectos devem ser focalizados quando do ensino de uma língua. Este objetivo está de acordo com o que Granger (2004, p. 291) argumenta:

Corpora de aprendiz servem a dois principais propósitos: (1) ao oferecer uma melhor descrição da interlíngua e uma melhor compreensão dos fatores que a influenciam, eles contribuem à teoria de aquisição de segunda língua; e (2) eles contribuem ao desenvolvimento de ferramentas pedagógicas e práticas de sala de aula que mais exatamente alvejam as necessidades do aprendiz.⁹

Assim sendo, estudos baseados em *corpus* são de grande auxílio a professores de línguas. Uma vez que professores estejam cientes das necessidades de seus alunos, eles poderão mais facilmente atingir seus objetivos educacionais.

2.2 Verbos modais

Os verbos modais sempre tiveram um lugar de destaque em gramáticas da língua inglesa. Apresenta-se, a título de exemplificação, a descrição de Swan (1998, p. 333):

Os verbos *can, could, may, might, will, would, shall* (principalmente no inglês britânico), *should, must* e *ought* são chamados de ‘verbos modais auxiliares’. Eles são usados antes da forma infinitiva de outros verbos e acrescentam certos tipos de significados conectados à certeza ou à obrigação e liberdade de agir [...]. *Need* [...] e *dare* [...] podem às vezes ser utilizados como verbos auxiliares modais e a expressão *had better* [...] também é utilizada como um verbo modal auxiliar.¹⁰

Swan considera ‘ought’ como um tipo de “verbo modal auxiliar”. Ele também sustenta a posição de que ‘need’, ‘dare’ e ‘had better’ podem ser utilizados como “verbos modais auxiliares” sem distingui-los dos verbos listados no início da citação. Isto talvez se deva ao fato de que a gramática de Swan (1998) ofereça apresentações práticas de tópicos gramaticais. Apesar de ser indicado na introdução que “o livro é voltado para alunos de níveis intermediário e avançado, e para *professores de inglês*”¹¹ (SWAN, 1998, p. xi – grifo meu), ele argumenta que ele não escreve para especialistas. Desta forma, presume-se que “onde foi necessária a utilização de terminologia gramatical, eu [ele] preferi [preferiu] utilizar termos tradicionais que são amplamente conhecidos e fáceis de compreender”¹² (SWAN, 1998, p. xi). Apesar de clara, a explicação de Swan não se enquadra no propósito deste estudo, que requer uma descrição mais detalhada de verbos modais e seu emprego.

⁸ Tradução livre do seguinte fragmento: “before large amounts of data were easily available, most of the generalisation had to be done by intuitive guesswork; pre-Corpus Linguistics were not able to check their notions”.

⁹ Tradução livre do seguinte fragmento: “Computer learner corpora (CLC) allegedly serve two main purposes: (1) by providing a better description of interlanguage and a better understanding of the factors that influence it, they contribute to Second Language Acquisition theory; and (2) they contribute to the development of pedagogical tools and classroom practices that more accurately target the needs of the learner”.

¹⁰ Tradução livre do seguinte fragmento: “The verbs *can, could, may, might, will, would, shall* (mainly British English), *should, must* and *ought* are called ‘modal auxiliary verbs’. They are used before the infinitives of other verbs, and add certain kinds of meaning connected with certainty or with obligation and freedom to act [...]. *Need* [...] and *dare* [...] can sometimes be used like modal auxiliary verbs, and the expression *had better* [...] is also used like a modal auxiliary.”

¹¹ Tradução livre do seguinte fragmento: “the book is intended for intermediate and advanced students, and for teachers of English”.

¹² Tradução livre do seguinte fragmento: “where it has been necessary to use grammatical terminology, I have generally preferred to use traditional terms that are well known and easy to understand”.

Uma diferente abordagem à análise da língua inglesa é oferecida pela *Collins Cobuild English Grammar* (SINCLAIR, 1990), que oferece dados retirados de amostras reais de linguagem em uso a partir da análise do *corpus* de Birmingham. Nela, os verbos modais são descritos como “uma categoria especial de verbos auxiliares”¹³ (SINCLAIR, 1990, p. 217), abrangendo ‘can’, ‘could’, ‘may’, ‘might’, ‘must’, ‘ought to’, ‘shall’, ‘should’, ‘will’ e ‘would’. Outros verbos como ‘dare’, ‘need’ e ‘used to’ são agrupados em um subtipo denominado ‘verbos semi-modais’. Esta gramática, contudo, apresenta uma descrição dos verbos modais na língua inglesa sem distinguir o emprego dos mesmos em diferentes gêneros.

Uma outra perspectiva é oferecida pela *Longman Grammar of Spoken and Written English* (BIBER et alii, 1999), uma publicação mais recente baseada em *corpus*. Segundo a mesma, os verbos modais se dividem em três grupos, a saber, verbos modais, verbos auxiliares marginais e verbos semi-modais. O primeiro grupo abrange ‘can’, ‘could’, ‘may’, ‘might’, ‘shall’, ‘should’, ‘will’, ‘would’ e ‘must’. Estes verbos modais (BIBER et alii, 1999, p. 483) têm um número de características específicas como (a) serem formas invariáveis, (b) precederem o sujeito em perguntas que exigem resposta sim ou não e (c) serem seguidos por um verbo na forma infinitiva sem ‘to’.

Os verbos auxiliares marginais correspondem a ‘need (to)’, ‘ought to’, ‘dare (to)’ e ‘used to’. De acordo com Biber et alii (1999, p. 484), estes verbos são raros e estão presentes praticamente somente no inglês britânico.

Expressões idiomáticas fixas como ‘(had) better’, ‘have to’, ‘(have) got to’, ‘be supposed to’ e ‘be going to’ são chamados de verbo semi-modais por Biber et alii (1999, p. 484). Eles diferem dos outros verbos modais porque podem ter marcação tanto de tempo como de pessoa. Ademais, eles podem também ocorrer como formas infinitas.

Para o escopo deste estudo, considera-se a descrição de verbos modais fornecida por Biber et alii (1999). A análise destes pesquisadores parece ser mais precisa uma vez que as características gramaticais são consideradas dentro de cada registro analisado na gramática em questão, a saber, conversa, ficção, linguagem jornalística e prosa acadêmica.

Os nove verbos modais que são agrupados por Biber et alii (1999) na primeira categoria por eles descrita também são chamados de “verbos modais centrais” no estudo de Wilson (2005). O autor afirma que estes verbos modais têm recebido grande atenção de estudiosos devido à sua alta complexidade semântica (WILSON, 1995, p. 151).

Um exemplo de estudo que trata tangencialmente de tal questão é o de Mindt (1996), intitulado “English Corpus Linguistics and the foreign language teaching syllabus”. Neste artigo, Mindt (1996) argumenta que a Linguística de *Corpus* tem influenciado dicionários e gramáticas, mas os materiais didáticos para o ensino de inglês como língua estrangeira continuam inalterados. Uma das seções do artigo compreende o tópico de verbos modais. Usando parte do *London-Lund Corpus*, ele argumenta que ‘would’, ‘can’ e ‘will’ são os verbos modais mais comuns em seu *corpus* de pesquisa. Considerando que “as formas presentes ocorrem mais freqüentemente em orações principais do que as formas passadas”¹⁴, (MINDT, 1996, p. 234) e que ‘will’ é um verbo modal extremamente freqüente em conversas em inglês, ele propõe que livros didáticos alemães de inglês como língua estrangeira devam introduzir tal verbo modal no primeiro ano de estudo em vez de fazê-lo no segundo. Em outras palavras, a apresentação de ‘will’ não deve ser adiada em favor de ‘must’ e ‘may’, que são verbos modais menos freqüentes.

Em sua compilação de freqüências vocabulares, Ringbom (1998) também contempla alguns verbos modais na escrita de estudantes de inglês de sete nacionalidades diferentes (holandesa, fino-sueca, finlandesa, francesa, alemã, espanhola e sueca). Os sete *subcorpora*, partes integrantes do projeto denominado *International Corpus of Learner English* (ICLE), foram comparados ao *Louvain Corpus of Native English Essays* (LOCNESS), que compreende redações argumentativas escritas por estudantes americanos e britânicos. Apesar de não haver uma explicação completa a respeito das 110 palavras mais freqüentes que são apresentadas no artigo, pode-se perceber algumas diferenças entre os grupos analisados. Parece que todos os grupos de estudantes utilizam ‘can’ em demasia e subutilizam ‘would’ e ‘will’. Uma única exceção concerne o grupo francês que utiliza ‘will’ freqüentemente. Em relação ao verbo modal ‘should’, franceses, finlandeses e alemães tendem a utilizá-lo mais do que americanos e britânicos enquanto espanhóis, fino-suecos, suecos e holandeses geralmente o subutilizam. Quanto ao verbo modal ‘could’, há três resultados distintos: (a) os estudantes finlandeses utilizam-no da mesma forma que seus pares americanos e britânicos; (b) os alunos espanhóis utilizam-no em maior escala; e (c) todos os outros cinco

¹³ Tradução livre do seguinte fragmento: “a special kind of **auxiliary verb**”.

¹⁴ Tradução livre do seguinte fragmento: “the present forms occur more frequently in main clauses than the past forms”.

grupos nacionais subutilizam-no. Em vez de oferecer uma interpretação completa a respeito dos dados apresentados no artigo, Ringbom (1998, p. 51) conclui que

[o] capítulo tentou mostrar como uma aparentemente simples contagem de frequência de palavras pode oferecer um ponto de partida útil para muitos e interessantes projetos de escala menor nos quais as características gerais da linguagem do aluno avançado assim como a relativa importância da transferência e das características universais podem ser exploradas.¹⁵

Apesar de os verbos modais já terem sido estudados por um grande número de lingüistas de *corpus* (cf. WILSON, 2005), parece haver uma lacuna de estudos que enfoquem a produção escrita de estudantes brasileiros de inglês como língua estrangeira.

3 Metodologia

Como este estudo trata da utilização de verbos modais em composições de alunos brasileiros avançados de inglês como língua estrangeira, foi necessária a compilação de um *corpus* que representasse tal produção. Para este propósito, composições escritas em língua inglesa foram coletadas em três cursos livres localizados em seis diferentes áreas da cidade do Rio de Janeiro. Decidiu-se coletar somente composições que fizessem parte de provas ou testes de forma a assegurar que os sujeitos de pesquisa não tivessem nenhum tipo de auxílio de terceiros durante o processo de escrita ou que eles colassem e/ou copiassem partes de suas composições. Os alunos podiam, contudo, utilizar dicionários e/ou gramáticas se isto fosse a eles permitido pelas regras de cada curso livre.

Dado que o foco da pesquisa é a escrita de alunos avançados, a coleta de dados foi realizada com alunos dos dois últimos períodos de cada curso livre. Em outras palavras, somente alunos que estavam prestes a se graduar foram solicitados a contribuir para com a pesquisa. Houve, no entanto, uma exceção: um dos cursos livres oferecia um programa específico de escrita. Os alunos deste programa opcional também foram convidados a participar.

O tópico das composições foi ditado pelo material didático adotado por cada curso livre ou pelo próprio professor. Desta forma, limitou-se a liberdade de escolha. Os alunos poderiam geralmente escolher um tópico a partir de uma lista tríplice, mas em alguns casos, eles tinham um único tópico obrigatório.

As composições variaram muito em termos de tamanho. A mais curta contém 112 palavras e a mais longa, 478. Em média, elas têm 288 palavras cada.

Após a coleta de dados, todas as composições foram digitadas de forma que pudessem ser investigadas com o auxílio de um programa computacional. As versões digitais correspondem ao que foi escrito pelos alunos. Os erros foram mantidos porque são representativos da escrita de aprendizes. Somente os erros ortográficos foram corrigidos. Caso contrário, o computador leria, por exemplo, 'should' e 'shuld' como duas palavras distintas, o que dificultaria a análise.

Apesar de ter havido um esforço para que os alunos produzissem composições por si mesmos, alguns fragmentos repetidos foram encontrados em certas composições como, por exemplo, títulos. Se estas seqüências tivessem sido mantidas, elas teriam constituído um problema na contagem final de itens lexicais. Assim sendo, tais seqüências repetidas foram excluídas porque elas são, na verdade, somente cópias da tarefa dada pelo professor.

No atual estágio, o *corpus* de pesquisa contém 105 composições escritas por alunos brasileiros avançados de inglês como língua estrangeira oriundos de três cursos livres em seis áreas da cidade do Rio de Janeiro. O *corpus* totaliza 30.261 itens e 2.870 formas.

A análise foi realizada com o programa *WordSmith Tools* (SCOTT, 1999), mais especificamente, com uma de suas ferramentas denominada de *WordList*. Assim, foi possível obter uma lista das palavras mais frequentes no *corpus*. Esta lista permitiu a identificação dos nove verbos modais que foram analisados

¹⁵ Tradução livre do seguinte fragmento: "chapter has tried to show that a seemingly simple word frequency count may provide a useful starting point for many interesting small-scale projects where the general characteristics of advanced learner language as well as the relative importance of transfer and universal features can be further explored."

(‘can’, ‘could’, ‘may’, ‘might’, ‘must’, ‘shall’, ‘should’, ‘will’ e ‘would’) e suas respectivas freqüências¹⁶. Em um segundo estágio, a ferramenta *Concord* foi utilizada para analisar o cotexto¹⁷ destes verbos modais.

O *corpus* de referência é o *Longman Spoken and Written English Corpus* no qual a *Longman Grammar of Spoken and Written English* (BIBER et alii, 1999) é baseada. Esta obra de referência “descreve o uso real de características gramaticais em diferentes variedades do inglês: principalmente em conversa, ficção, linguagem jornalística e prosa acadêmica”¹⁸ (BIBER et alii, 1999, p. 4). Na presente pesquisa, decidiu-se comparar os resultados aqui reportados àqueles obtidos por Biber et alii (1999) no mapeamento do registro prosa acadêmica.

4. Análise de dados e discussão

Oito dos nove verbos modais analisados podem ser agrupados em duas categorias: aqueles que não se referem ao passado e aqueles que podem se referir ao passado (cf. BIBER et alii, 1999, p. 484-485). No primeiro grupo, estão ‘may’, ‘can’, ‘will’, and ‘shall’; e no segundo, encontram-se, respectivamente, ‘might’, ‘could’, ‘would’ and ‘should’. A diferença na utilização entre modais que não se referem ao passado e aqueles que podem se referir ao passado é notável. A Tabela 1 resume este contraste.

Referência outra que não ao passado		Possível referência ao passado		Total
Verbo modal	Porcentagem	Verbo modal	Porcentagem	
<i>May</i>	86.96%	<i>Might</i>	13.04%	100%
<i>Can</i>	83.90%	<i>Could</i>	16.10%	100%
<i>Will</i>	74.83%	<i>Would</i>	25.17%	100%
<i>Shall</i>	0.00%	<i>Should</i>	100%	100%

Tabela 1: Distribuição de verbos modais no corpus de pesquisa (referência temporal)

‘May’, ‘can’ e ‘will’ são, no mínimo, três vezes mais comuns do que seus pares, a saber: ‘might’, ‘could’ e ‘would’. A única exceção é o par ‘shall’ e ‘should’, em que o último é muito mais freqüente do que o primeiro, não havendo nenhuma instância de ‘shall’ no *corpus* de pesquisa. Estes resultados são similares aos que foram encontrados por Biber et alii (1999, p. 486): “considerando os pares de verbos modais centrais, o verbo que pode expressar tempo passado é menos freqüente do que o seu par em todos os casos com exceção de ‘shall’/‘should’”¹⁹. A diferença entre os resultados do estudo destes pesquisadores e o presente é que no *corpus* de pesquisa não há nenhuma instância de ‘shall’ enquanto no *corpus* de referência este verbo modal está presente apesar de ser raro.

Verbos modais também podem ser agrupados em três categorias de acordo com as idéias que eles transmitem (cf. BIBER et alii, 1999, p. 489). Eles podem expressar (a) permissão, possibilidade ou habilidade (‘can’, ‘could’, ‘may’ e ‘might’), (b) volição ou predição (‘will’, ‘would’ e ‘shall’), e (c) necessidade ou obrigação (‘should’ e ‘must’). A Tabela 2 mostra o número de vezes que cada verbo modal ocorre no *corpus* de pesquisa.

¹⁶ Neste estudo, optou-se pela lematização dos verbos modais em questão. Em outras palavras, as formas negativas reduzidas destes verbos também foram consideradas quando do levantamento das freqüências de uso.

¹⁷ Sinclair (2003, p. 174) define cotexto como um grupo de palavras que ocorre em ambos os lados de uma determinada palavra de busca em um texto.

¹⁸ Tradução livre do seguinte fragmento: “describes the actual use of grammatical features in different varieties of English: mainly conversation, fiction, newspaper language, and academic prose”

¹⁹ Tradução livre do seguinte fragmento: “Considering the pairs of central modals, the tentative/past time member is less frequent than its partner in all cases except *shall/should*.”

Idéias	Verbos modais	Ocorrências
Permissão	'can'	271
	'could'	52
Possibilidade	'may'	20
	'might'	3
Volição	'will'	220
	'would'	74
Predição	'shall'	0
Necessidade	'should'	59
	'must'	31

Tabela 2: Distribuição de verbos modais no corpus de pesquisa (idéias)

Pode-se, então, comparar os resultados do presente estudo ao mapeamento de modais no registro prosa acadêmica realizado por Biber et alii (1999, p. 489).

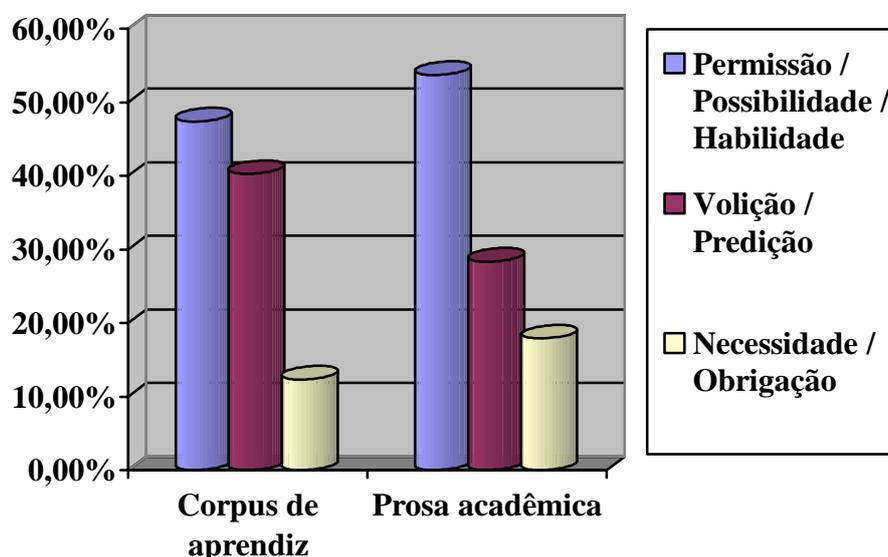


Gráfico 1: Idéias transmitidas por verbos modais

O Gráfico 1 indica que os participantes deste estudo tendem a subutilizar verbos modais que marcam tanto permissão, possibilidade e habilidade como necessidade ou obrigação. Por outro lado, eles mostram uma tendência à utilização em larga escala de verbos modais que indicam volição ou predição, especialmente com o uso de 'will', como ilustrado na Figura 1 abaixo.

Volição	Predição
But, even so, I <i>will</i> try to reach happiness and success, no matter what this may cost. ²⁰	“Christmas trees” <i>will</i> also be decorated with lights and Christmas ornaments.
After finishing my studies, I <i>will</i> start to work with my sister, who is a doctor, and has a little emergency hospital.	I’m sure that this <i>will</i> be a great experience and we will never forget it!
I <i>would</i> try to help them anyway.	because the most powerful country <i>would</i> impose its culture and, consequently, its language.

Figura 1: Exemplos de verbos modais que expressam volição e predição

²⁰ Este exemplo e todos os seguintes foram retirados do *corpus* de pesquisa e não foram corrigidos conforme descrito na seção 3. A única exceção diz respeito a correção de erros ortográficos.

De acordo com Biber et alii (1999, p. 489), ‘will’ e ‘would’ são menos freqüentes na prosa acadêmica. O registro que contém as maiores freqüências de tais verbos modais é o conversacional. Desta forma, a presença dos mesmos no *corpus* analisado pode sugerir que os sujeitos de pesquisa escrevem em uma forma que é similar à forma que os falantes de inglês como primeira língua se expressam oralmente.

Outra característica apontada por Biber et alii (1999) para a caracterização da prosa acadêmica é a utilização de sintagmas verbais que incorporem verbos modais na voz passiva. Como eles afirmam, “a voz passiva com verbos modais é rara em conversa e ficção, mas é relativamente comum para alguns verbos modais na prosa acadêmica”²¹ (BIBER et alii, 1999, p. 499). Os pesquisadores também concluem que “com a passiva, ‘can’ e ‘should’ são particularmente comuns, ‘could’ e ‘must’ também são razoavelmente comuns”²².

No *corpus* de aprendiz, no entanto, os resultados são bastante diferentes. A maior parte dos sintagmas verbais que incorporam verbos modais está na voz ativa como pode ser visto no Gráfico 2.

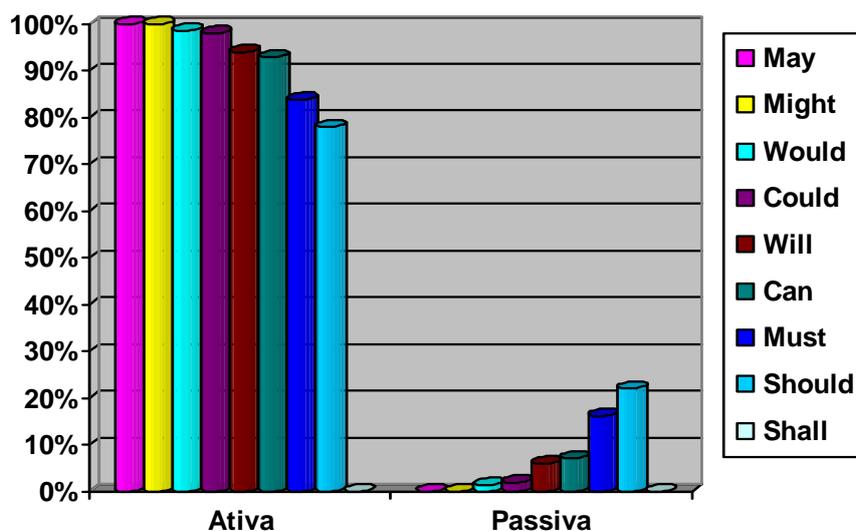


Gráfico 2: Distribuição de verbos modais (vozes ativa e passiva)

Há poucas instâncias de sintagmas verbais na voz passiva como ilustrado nos seguintes exemplos:

That is why death punishment it is not a good idea, this *should not be accepted* in any country, in any constitution.
 When a baby becomes a child, his or her growing *must be accompanied* attentively by his parents
 There are a lot of things that *can be done* to better your health
 If happiness brings health, friends and positive views of life, nothing *could* really *be considered* more important than it.

Ao todo, há somente 52 instâncias de verbos modais sendo utilizados na voz passiva, o que representa 7,12% dos sintagmas verbais contendo verbos modais. Mais uma vez pode-se observar a diferença existente entre o *corpus* de aprendiz e o registro prosa acadêmica investigado por Biber et alii (1999).

5. Conclusão

Os resultados do presente estudo revelam que os participantes da pesquisa e os falantes de inglês como primeira língua utilizam verbos modais de formas diferentes. Ao escreverem composições, os participantes desta pesquisa tendem a utilizar estruturas que caracterizam a produção oral de falantes de inglês como

²¹ Tradução livre do seguinte fragmento: “passive voice with modals is rare in conversation and fiction, but relatively common for some modals in academic prose”.

²² Tradução livre do seguinte fragmento: “with the passive, *can* and *should* are particularly common, *could* and *must* are also fairly common”.

primeira língua. Uma indicação de tal resultado é o uso freqüente de verbos modais que sinalizam volição ou predição, especialmente o verbo modal 'will', na escrita destes estudantes. Um outro indício que foi reportado neste artigo é a rara freqüência de voz passiva nos sintagmas verbais que contêm verbos modais, uma característica da prosa acadêmica estudada por Biber et alii (1999).

Além da descrição gramatical, este estudo também tem implicações pedagógicas. Os professores de inglês podem conscientizar seus alunos sobre o emprego em larga escala de 'will' e a subutilização da voz passiva. Ao fazer isto, estes alunos estarão aptos a escrever de forma mais proficiente e a comunicar suas idéias mais fluentemente.

A realização de estudos baseados em *corpus* é de grande importância para professores de línguas. Tal tipo de investigação permite que o professor/pesquisador identifique as áreas mais problemáticas de uma determinada língua para grupos específicos de alunos. Como afirmam Tribble e Jones (1990, p. 23), "mesmo com pequenos estudos baseados em sala de aula é possível chegar a algumas conclusões muito interessantes a respeito da forma como os alunos lidam com a língua inglesa"²³. Faz-se necessário ressaltar que estes estudos destacam a produção de alunos, ou seja, eles são baseados no que estes verdadeiramente escrevem ou falam em vez de considerar modelos abstratos de linguagem. Isto só pode ser alcançado por meio da Lingüística de *Corpus*.

Como comentário final, podem-se citar as palavras de Granger (2004, p. 299) a respeito do potencial de *corpora* de aprendiz:

Corpora de aprendiz talvez não tenham dado origem a um grande número de aplicações na área de ensino/aprendizagem, mas a atividade crescente nesta área e as ferramentas de ensino e de referência com base em corpora de aprendiz que já foram produzidas são evidências concretas de uma tendência em curso que deve resultar em aplicações pedagógicas inovadoras nos próximos anos.²⁴

6. Referências bibliográficas

BIBER, D. et alii *Longman grammar of spoken and written English*. London: Longman, 1999.

GRANGER, S. Practical applications of learner corpora. In: LEWANDOWSKA-TOMASZCZYK, B. (Ed.). *Practical applications in language and computers: PALC 2003*. Frankfurt am Main: Peter Lang, 2004. p. 291-301.

_____. (Ed.). *Learner English on computer*. London / New York: Longman, 1998.

LEECH, G. Preface. In: GRANGER, S. (Ed.). *Learner English on computer*. London / New York: Longman, 1998. p. xiv-xx.

MCENERY, T.; WILSON, A. *Corpus Linguistics*. Edinburgh: Edinburgh University Press, 1996.

MINDT, D. English Corpus Linguistics and the foreign language teaching syllabus. In: THOMAS, J.; SHORT, M. (Eds.). *Using corpora for language research*. London / New York: Longman, 1996. p. 232-247.

OTHERO, G. de Á.; MENUZZI, S. de M. *Lingüística computacional: teoria & prática*. São Paulo: Parábola, 2005.

RINGBOM, H. Vocabulary frequencies in advanced learner English: a cross-linguistic approach. In: GRANGER, S. (Ed.). *Learner English on computer*. London / New York: Longman, 1998. p. 41- 52.

²³ Tradução livre do seguinte fragmento: "even with very small classroom-based studies it is possible to come to some very interesting conclusions about the way students are dealing with English".

²⁴ Tradução livre do seguinte fragmento: "Learner corpora may not yet have given rise to a large number of teaching and learning applications, but the buzzing activity in the field and the CLC-informed reference and teaching tools that have already been produced are concrete evidence of an ongoing trend which should result in highly innovative pedagogical application in the years to come".

- SANTOS, D. Introdução ao processamento de linguagem natural através das aplicações. In: RANCHHOD, E. (Ed.). *Tratamento das línguas por computador: uma introdução à lingüística computacional e suas aplicações*. Lisboa: Caminho, 2001. p.229-259.
- SARDINHA, A. P. B. *Lingüística de Corpus*. São Paulo: Manole, 2004.
- SASLOW, J. et alii *Top Notch fundamentals: teacher's edition and lesson planner*. New York: Pearson Longman, 2006.
- SCOTT, M. *WordSmith tools 3.0*. Oxford: Oxford University Press, 1999.
- SINCLAIR, J. (Ed.). *Collins Cobuild English grammar*. London / Glasgow: Collins, 1990.
- _____. *Corpus, concordance, collocation*. Oxford: Oxford University Press, 1991.
- _____. *Reading concordances: an introduction*. London: Longman, 2003.
- SWAN, M. *Practical English usage*. Oxford: Oxford University Press, 1998.
- TOGNINI-BONELLI, E. *Corpus Linguistics at work*. Amsterdam / Philadelphia: John Benjamins Company, 2001.
- TRIBBLE, C.; JONES, G. *Concordances in the classroom: a resource book for teachers*. Harlow: Longman, 1990.
- WEHMEIER, S. (Ed.). *Oxford advanced learner's dictionary of current English*. Oxford: Oxford University Press, 2000.
- WILSON, A. Modal verbs in written Indian English: a quantitative and comparative analysis of the Kolhapur corpus using correspondence analysis. *ICAME Journal*, v. 29, p. 151-169, abr. 2005. Disponível em: <<http://gandalf.aksis.uib.no/icame/ij29/ij29-page151-170.pdf>>. Acesso em: 30 jun. 2006.