

1

Introdução

Modality is just one example of an area of learner language which has not previously been well described and where computer learner corpora can make a contribution²

(Aijmer, 2002, p. 56)

Como indica Aijmer (2002) na epígrafe que inicia este capítulo, a modalidade é ainda uma área pouco explorada em investigações acerca da produção lingüística de aprendizes. Não há muitos estudos com base na lingüística de corpus que enfoquem este traço lingüístico a partir da análise da escrita de estudantes, especialmente de falantes brasileiros de inglês como língua estrangeira (ILE).

Além da lacuna identificada, os verbos modais revelam-se recursos importantes em redações universitárias. Em uma comparação – realizada pelo pesquisador – da produção da escrita de universitários americanos e britânicos³ de inglês como primeira língua (IL1) com a produção oral e escrita de britânicos como um todo, notou-se a recorrência de verbos modais na lista de palavras-chave compilada pelo *WordSmith Tools* (Scott, 1999). Esta lista forneceu aquelas palavras que são características da prosa escrita dos universitários anteriormente mencionados, representada no *Louvain Corpus of Native English Essays* – LOCNESS, quando comparada a um amplo corpus contendo mais de 100 milhões

² Tradução livre para o português: “A modalidade é somente um exemplo de uma área da linguagem de aprendiz que não foi previamente bem descrita e onde corpora de aprendiz podem fazer uma contribuição”.

³ Nesta dissertação, emprega-se o termo ‘americano’ para indicar os cidadãos nacionais dos Estados Unidos. Por sua vez, ‘britânico’ refere-se aos nacionais da Grã-Bretanha apesar da mesma englobar países distintos. Ambas as escolhas foram feitas por questão de economia de linguagem.

de palavras da língua inglesa, como ocorre no *British National Corpus* – BNC (ver Seção 6.1.2 para mais detalhes acerca deste procedimento). O resultado indicou a existência de várias formas correspondentes aos verbos modais na lista de palavras-chave: ‘cannot’, ‘can’t’, ‘would’, ‘should’, ‘won’t’, ‘couldn’t’, ‘will’, ‘wouldn’t’, ‘shouldn’t’, ‘can’, ‘must’ e ‘shall’, em ordem decrescente de chavicidade.

Ao todo, sete dos nove verbos modais centrais (cf. Biber et al., 1999 – ver mais detalhes na Seção 4.1.4) aparecem como palavras-chave. Dentro deste grupo, a maioria é característica da prosa escrita do LOCNESS, sendo considerada palavra-chave positiva. O único verbo modal central mais comumente encontrado no BNC do que no LOCNESS, classificado como palavra-chave negativa, corresponde a ‘shall’.

Por estes motivos, optou-se pelo estudo de verbos modais na escrita de universitários de língua inglesa nesta dissertação. Os modais são recursos lingüísticos comumente empregados em redações de alunos universitários, que por sua vez são recorrentemente utilizadas como forma de avaliação dos alunos em foco (cf. Capítulo 3).

A presente investigação baseia-se nos preceitos da lingüística de corpus visto que esta possibilita a observação de padrões em situações efetivas de uso. Opta-se pela investigação de textos naturalmente produzidos pelos participantes da pesquisa nos quais o uso que eles fazem da língua inglesa torna-se visível. O auxílio da ferramenta computacional permite que grandes quantidades de textos possam ser processadas de forma mais rápida e confiável, facilitando o pesquisador em sua tarefa analítica.

Em termos de corpora existentes, a investigação já realizada no âmbito do projeto *International Corpus of Learner English* (ICLE), coordenado por Sylviane Granger na Universidade de Louvain na Bélgica, mostrou-se apropriada para o objetivo da pesquisa. Isto ocorre porque o projeto contempla a compilação de redações escritas por universitários em diversos países com base em critérios preexistentes e pela adoção de um corpus de referência já coletado e disponibilizado. Assim sendo, optou-se por investigar a parte brasileira do ICLE – *The Brazilian Portuguese Sub-corpus of ICLE* (Br-ICLE) – sediada na Pontifícia Universidade Católica de São Paulo (PUC-SP). Mais especificamente, foram consideradas somente as redações coletadas no Estado do Rio de Janeiro

(cf. Seção 5.1.1). Em relação ao corpus de referência, este estudo se alinha a outras investigações semelhantes (cf. Granger, 1998a) e adota o LOCNESS, que é um corpus comparável contendo redações produzidas por estudantes americanos e britânicos, falantes de IL1.

De forma resumida, o objetivo desta dissertação é verificar em que medida a escolha e o uso de verbos modais diferem ou se aproximam nas produções escritas de falantes brasileiros de ILE e de falantes americanos e britânicos de IL1. De forma mais ampla, o estudo preenche um nicho de pesquisa, a saber, a investigação constrastiva de verbos modais baseada no Br-ICLE e no LOCNESS.

As perguntas de pesquisa que norteiam esta dissertação são quatro, conforme explicitadas abaixo.

- (1) Qual é a frequência de uso de verbos modais na escrita de universitários brasileiros em ILE, representada no Br-ICLE?
- (2) Qual é a frequência de uso de verbos modais na escrita de universitários americanos e britânicos em IL1, representada no LOCNESS?
- (3) Quais são as diferenças observadas no emprego de verbos modais realizado por estes dois grupos de falantes?
- (4) Quais são os padrões de uso dos verbos modais em relação aos verbos lexicais nos dois corpora analisados (Br-ICLE e LOCNESS)?

Com o intuito de responder a estas perguntas, a presente dissertação se estrutura em dez capítulos. Os três capítulos que seguem esta introdução referem-se à fundamentação teórica da pesquisa. No Capítulo 2, é enfocada a lingüística de corpus, área de investigação que privilegia, com um foco empírico, a análise de dados reais. São discutidos, neste capítulo, o conceito de corpus e suas características definidoras assim como os diferentes tipos de corpora. São também introduzidos nesta parte os conceitos de colocação, coligação, preferência semântica e prosódia semântica além dos princípios idiomático e de livre escolha (cf. Sinclair, 1991).

O capítulo seguinte ocupa-se da atividade de escrita. Inicialmente, são discutidos modelos de abordagem à escrita (cf. Grabe & Kaplan, 1996; Hyland, 2002). Em um segundo momento, enfoca-se a escrita de universitários, questão central da presente dissertação. Ressaltam-se, especialmente na parte final do capítulo, estudos da escrita universitária com base na lingüística de corpus (cf. Capítulo 2) e através da abordagem textual (cf. Hyland, 2002).

O último capítulo dedicado à fundamentação teórica concentra-se na especificidade dos verbos modais. Faz-se um apanhado da descrição de tal recurso lingüístico em gramáticas de diferentes naturezas antes de delinear a opção feita neste estudo. Posteriormente, são discutidos estudos que tenham adotado os verbos modais como unidades analíticas, dando atenção especial às pesquisas baseadas em corpora escritos.

A metodologia empregada na presente pesquisa encontra-se reportada em dois capítulos por motivos de organização textual. O Capítulo 5 explicita os procedimentos adotados para a coleta de dados. Desta forma, são descritas as condições de compilação, os corpora e os participantes envolvidos.

O Capítulo 6 contempla outro aspecto metodológico, a saber, o tratamento dos dados. São comentados os papéis desempenhados por três programas computacionais na pesquisa em tela. Em um primeiro momento, são identificados os recursos do *WordSmith Tools* (Scott, 1999) para a investigação dos corpora de pesquisa e de referência utilizados. Posteriormente, indica-se de que forma o *Statistical Package for the Social Sciences* (SPSS) e o *Excel* foram empregados nesta dissertação para o cálculo de testes estatísticos.

A análise de dados encontra-se reportada e discutida à luz da fundamentação teórica no Capítulo 7. Nesta parte, há duas grandes subdivisões. A primeira enfoca uma análise freqüencial dos verbos modais no Br-ICLE e no LOCNESS tanto de forma isolada como de forma contrastiva, indicando-se a existência de diferenças significativas em alguns casos. Na segunda subdivisão, o capítulo centra-se na análise dos padrões de uso que são observados quando do emprego dos verbos modais no LOCNESS e em que medida eles também são verificados no Br-ICLE.

No Capítulo 8, são tecidos alguns comentários finais a respeito do estudo. Indica-se também de que forma esta dissertação aponta para estudos futuros, registrando encaminhamentos que poderão responder a novas perguntas de pesquisa originadas com esta investigação.

Finalmente, nos últimos dois capítulos podem ser encontradas as referências bibliográficas citadas no decorrer desta dissertação e os anexos que permitem uma melhor compreensão dos dados reportados.