

**PROBLEMAS NA MODALIZAÇÃO EM L2:  
ANÁLISE CONTRASTIVA DO MODAL WOULD  
EM CORPORA DIGITAL**

Renata Aparecida R. A. dos Santos (UERJ)  
[rarash@click21.com.br](mailto:rarash@click21.com.br)

**CONSIDERAÇÕES INICIAIS**

Uma das maiores contribuições da Linguística de Corpus, especialmente quando relacionada a corpus de aprendizes é a possibilidade de investigação de aspectos da aprendizagem que ela oferece. Sem a utilização de tais ferramentas e tecnologias, tais investigações seriam praticamente impossíveis. Muitos pesquisadores já estudaram o comportamento de aprendizes da língua inglesa em relação a certos aspectos da língua na produção escrita através de corpora, como Karin Aijmer (2002). Ela estudou a aplicação da modalização por alunos suecos, aprendizes de inglês em nível avançado. Com a constatação de um uso excessivo dos modais, concluiu que a expressão de atitude em geral e o uso de modais em especial por alunos de inglês como L2 são via de regra problemáticos.

Partindo dessa premissa, o presente trabalho tem como objetivo analisar, tendo como base os preceitos da Linguística de corpus, especialmente o corpus de aprendizes, o comportamento e a presença do modal *would* por aprendizes brasileiros.

Para isso, comparamos a utilização do modal *would* em dois corpora: um corpus de nativos e um de aprendizes, ambos exclusivamente de produção escrita. Através de ferramentas digitais, verificou-se a frequência do uso e os significados atribuídos.

**1. Orientação teórica**

Para que a pesquisa fosse feita, baseamo-nos nos conceitos de frequência e coocorrência lexicais da Linguística de Corpus. Nela, corpora digitais são analisados com o auxílio de um computador. Um corpus digital pode ser descrito por características básicas: deve ser uma grande coleção de textos produzidos por indivíduos diferentes,

## ***Cadernos do CNLF, Vol. XIII, N° 04***

passível de leitura por um programa de computador. O computador auxilia na distinção e listagem de concordâncias e padrões linguísticos.

Uma das áreas da Linguística de Corpus é a que trata dos corpora de aprendizes. Como já foi bastante discutido na literatura (LE-ECH, 1998; GRANGER, 1998C; HORVATH, 2001), a investigação de aspectos do aprendizado da linguagem é uma das maiores contribuições dos corpora de aprendiz computadorizados. Esta investigação seria, sem auxílio da tecnologia, praticamente impossível de ser realizada. Corpora são fontes de evidências, que podem resolver problemas de uso da língua, pois através deles pode-se observar padrões de uso da linguagem.

O objetivo da utilização de corpora de aprendizes é melhorar a descrição e teoria da linguagem, assim como sua relevância na prática e descrição do uso da linguagem. Tal prática tem, portanto, aplicações no ensino de línguas. Pode-se estudar como uma palavra ou construção gramatical é usada em textos autênticos. Em muitos casos, as investigações feitas são comparativas. Elas comparam os corpora de aprendizes com corpora de nativos de tipo semelhante.

Comparando-se língua nativa e interlíngua pode-se estabelecer distinções, diagnosticar se itens linguísticos estão sendo aplicados de forma excessiva ou com menor frequência em relação aos nativos, revelando diferenças na distribuição de padrões, que podem ajudar a explicar porque um texto que não contém erros gramaticais ou lexicais podem, mesmo assim, parecer que não foram escritos por um nativo.

Neste trabalho usamos corpora de aprendizes computadorizados para comparar a abrangência e a frequência do modal *would* em redações de alunos universitários, nativos e não nativos.

### ***2. Perguntas da pesquisa***

A partir do estudo da literatura e da pesquisa efetuada, o presente trabalho se propõe a responder as seguintes perguntas: analisando e comparando os corpora, há diferenças no uso do modal *would* entre nativos e não nativos? Se sim, que tipo de diferenças existem? Seriam elas apenas quantitativas ou também qualitativas?

## **Cadernos do CNLF, Vol. XIII, Nº 04**

Como podemos categorizá-las e o que elas dizem a respeito dos aprendizes?

### **3. Descrição dos Materiais Utilizados**

Para a que este estudo fosse possível, foram analisados dois corpora, o corpus de estudo e o comparável.

O corpus de estudo, o BrICLE (*The Brazilian Portuguese Sub-corpus of International Corpus of Learner English*), é na verdade um sub-corpus integrante dos ICLE, que é formado por textos argumentativos de não nativos de 14 nacionalidades diferentes. Classificado como pequeno e ainda em coleta, na época da pesquisa era constituído por mais de 65.000 palavras em 127 textos argumentativos. Este corpus de aprendizes é formado de redações de alunos universitários brasileiros, com inglês de nível avançado.

O corpus comparável, o LOCNESS (*Louvain Corpus of Native English Essays*) tem porte médio, com 288.177 palavras em 408 textos argumentativos. É um corpus comercial, coletado entre 1991 e 1995, composto por *essays* de mais de 500 palavras de autoria de universitários e vestibulandos cuja língua materna é o inglês. Tais *essays* foram escritos por estudantes universitários americanos e britânicos, além de vestibulandos britânicos, com tópicos variados.

Os corpora estudados possuem algumas características semelhantes. Por isso podem ser comparados. Ambos consistem exclusivamente de produção escrita. Contemporâneos, especializados e de amostragem, pois tem como finalidade ser uma amostra do total de estudantes por eles representados. São compostos de textos autênticos, com conteúdo escolhido criteriosamente, seguindo as regras estabelecidas por seus criadores.

### **4. Modalização**

Já foi constatado que a modalização é uma área problemática para aprendizes de inglês. Segundo a gramática de língua inglesa Collins Cobuild (1990), a modalização é a forma utilizada para indicar a atitude do falante em relação ao que é dito ou quando está pre-

## ***Cadernos do CNLF, Vol. XIII, N° 04***

ocupado com o efeito provocado no interlocutor. Pode ser expressa por advérbios (*probably, possibly*), verbos lexicais com significado modal (I think, I feel) ou verbos modais (*can, could, may, might, must, shall, should, will e would*). Cada um deles teria uma utilização diferente e especial.

Vários estudos mostraram que não nativos aparentam ter dificuldade ao utilizar os modais de forma apropriada (AIJMER, 2002) e que eles supervalorizam ou não seus possíveis significados. Assim, eles acabam usando muito mais ou muito menos os modais do que os nativos.

### **5. *O modal would***

*Would* faz parte do grupo de modais que são principalmente epistêmicos, ou seja, que expressam várias formas de expressar dúvida ou certeza. Pode-se atribuir inúmeras funções ao *would*.

A gramática *Collins Cobuild* sugere, entre outras, as seguintes categorias:

- Ação que acontecia com regularidade no passado;
- No discurso indireto, substituindo *will*;
- Pensamentos sobre o futuro em histórias;
- Suposição;
- Certeza de um fato sob certas circunstâncias;
- Possibilidade no passado (quando seguido de *have*);
- Instruções ou apelos;
- Ofertas ou convites;
- Expressar volição ou desejo;
- Expressar preferência, quando seguido de *rather* ou *sooner*;
- Expressar arrependimento, quando seguido de *that*.

A gramática *Longman Grammar os Spoken and Written English* sugere as seguintes categorias:

## ***Cadernos do CNLF, Vol. XIII, Nº 04***

- Volição - intrínseca
- Predição - extrínseca
- Hipótese no passado

O presente estudo compara o modal *would*, seus significados e uso em redações produzidas por nativos e não nativos. Para facilitar a análise, categorias foram criadas a partir das provenientes das gramáticas supracitadas. Para cada categoria, uma letra foi atribuída, para futura classificação e etiquetagem. São elas:

- Suposição (A)
- Certeza (C)
- Hipótese no passado (H)
- Ordens, Instruções, Convites, Ofertas (O)
- Volição (V)
- Discurso indireto (R)

### ***6. Método de coleta e tratamento de dados***

Os corpora foram analisados por várias ferramentas pelas quais a frequência pode ser determinada. A manipulação dos dados se dá através do programa WordSmith Tools 3.0 (SCOTT, 1999), através do qual se extraem de ambos os corpora as ocorrências de *would*, que são posteriormente analisadas. A vantagem da utilização deste software é que ele possui uma ferramenta de concordância que investiga a frequência dos verbos nos corpora. Para limitar a análise, utiliza-se o método de Sardinha, que consiste de uma seleção aleatória das ocorrências para que um ponto de corte seja estabelecido e que as ocorrências observadas sejam selecionadas aleatoriamente. Tais ocorrências são analisadas manualmente e etiquetadas de acordo com as categorias previamente analisadas.

## **Cadernos do CNLF, Vol. XIII, N° 04**

### **7. Análise dos dados**

O modal *would* é o mais utilizado no corpus de nativos e o segundo mais usado no de aprendizes brasileiros (**Tabela 1**). Um ponto de corte de 225 ocorrências<sup>35</sup> foi adotado para a classificação por categorias. Utilizando as definições e categorizações para o modal em questão extraídas da gramática pedagógica *Collins Cobuild* e da gramática baseada em corpus *Longman Grammar of Spoken and Written English*, chegou-se a conclusão que ambos nativos e não nativos utilizam o modal *would* prioritariamente como suposição, seguido por certeza (**Tabela 2**). Os outros sentidos atribuídos ao modal não teriam tanta utilização.

BriCLE		Locness	
Modal	%	Modal	%
Can	22,07	Would	24,15
Would	20,11	Can	19,38
Will	17,07	Will	18,42
Should	13,67	Should	12,72
May	9,47	Could	10,48
Could	9,47	May	8,04
Must	4,83	Must	5,31
Might	2,77	Might	1,40
Shall	0,54	Shall	0,10

**Tabela 1**

	A	C	H	V	R	O
BriCLE	44,1	36	10	7	4	0,4
Locness	47,6	27,6	8	9,3	2,7	0,9

**Tabela 2**

A partir desta análise, pode-se dizer que os textos de nativos e não nativos são suficientemente similares entre si, mesmo com a diferença percentual. Entretanto, tal conclusão seria superficial, pois não podemos analisar apenas um termo linguístico isoladamente. Assim, foram analisados os verbos com os quais o modal *would* foram combinados, de acordo com suas categorias (**Tabelas 3 e 4**). Pode-se notar que os aprendizes brasileiros utilizam o verbo *be* de maneira

---

<sup>35</sup> O ponto de corte (225) corresponde ao número total de ocorrências no BriCLE. No Locness, o número total de ocorrências foi 1460.

## **Cadernos do CNLF, Vol. XIII, N° 04**

excessiva, enquanto não se utilizam de outras formas verbais que poderiam ser utilizadas em seu lugar, o que enriqueceria o texto.

Categoria A (suposição)

BrICLE		Locness	
BE	40	BE	32
HAVE	8	HAVE	12
BECOME	5	MAKE	4
IMPROVE	3	DECREASE	5
WONDER	3	ALLOW	4
OUTROS	38	OUTROS	51

**Tabela 3**

Categoria C (certeza)

BrICLE		Locness	
BE	28	BE	17
HAVE	15	HAVE	8
OUTROS	28	HELP	3
		OUTROS	33

**Tabela 4**

### **8. Considerações finais**

Com base na Linguística de Corpus, observamos e comparamos dois corpora digitalizados para analisar como aprendizes brasileiros de língua inglesa em nível avançado utilizam o modal *would* em suas redações.

Os resultados parciais obtidos até este ponto sublinham que o modal em questão é usado de forma diferenciada entre nativos e não nativos, em termos de frequência, de coocorrência e distribuição. Além disso, nota-se no corpus brasileiro um uso menor dos 'colocados' do modal supracitado, pois os aprendizes brasileiros utilizam menor número de verbos acompanhando o modal estudado. Os verbos utilizados pelos nativos apresentam uma carga pragmática maior que os utilizados pelos aprendizes brasileiros.

A pesquisa pretende se aprofundar agora nos verbos que são utilizados para acompanhar o modal *would*, analisando quais verbos poderiam substituir o verbo *be*, que é usado em forma excessiva por

## ***Cadernos do CNLF, Vol. XIII, N° 04***

aprendizes brasileiros. Pretende-se que os resultados tenham importantes implicações pedagógicas.

### REFERÊNCIAS BIBLIOGRÁFICAS

BERBER, Tony Sardinha. *Linguística de corpus*. São Paulo: Manole, 2004.

BIBER, D.; JOHANSSON, S.; LEECH, G.; CONRAD, S. e FINAGAN, E. *Longman Grammar of Spoken and Written English*. Inglaterra: Pearson Education Limited, 1999.

COLLINS, William & sons. *Collins Cobuild English Grammar*. Grã-Bretanha: Collins Publishers, 1990.

DAVIS, Alan & ELDER, Catherine. *The handbook of Applied Linguistics*. Reino Unido: Blackwell Publishing, 2004.