

ANALISANDO A FALA DE PROFESSORES DE INGLÊS 'NATIVOS' E 'NÃO-NATIVOS' EM UM FÓRUM DE DISCUSSÃO ONLINE

Luciana S. Ferreira Simioni¹

Marcio Chrisostimo

Resumo: Este trabalho tem foco em uma investigação de corpora eletrônicos com o objetivo de extrair e analisar padrões lexicais no discurso dos professores “nativos” e “não-nativos” de língua inglesa em fóruns de discussão online de forma a conscientizar os professores “não-nativos” sobre sua própria performance linguística e sobre os padrões do gênero em questão, quer dizer, a comunicação mediada pelo computador. O presente estudo é comparativo e analisa, a Linguística de Corpus foi aplicada como metodologia, as escolhas lexicais de ambos os grupos em grandes quantidades de dados.

1) Introdução

O presente artigo tem por objetivo analisar o discurso de professores de língua inglesa cuja língua materna é o próprio inglês (doravante chamados de PN ou professores “nativos”), assim como o discurso de professores de língua inglesa cuja língua materna é o português do Brasil (doravante chamados de PNN ou professores “não-nativos”). O trabalho usa os fóruns de discussão profissional online e visa a encontrar e comparar possíveis padrões lexicais que possam caracterizar cada um dos grupos .

O estudo analisa grandes quantidades de dados e, para tanto, encontra suporte teórico na Linguística de Corpus. A Linguística de Corpus pode ser definida como a área responsável pelo estudo da língua por meio da observação de grandes quantidades de dados linguísticos reais, isto é, textos orais ou escritos autênticos, com o auxílio de ferramentas computacionais. Tais textos orais ou escritos, produzidos no mundo real, bem como coletados, digitalizados e analisados pelos estudiosos da área em questão, são conhecidos simplesmente como *corpora* (singular: *corpus*).

O trabalho aqui descrito fundamenta-se nesta área de pesquisa baseada em corpus digital e se preocupou com a descrição de uma determinada variedade de texto, neste caso específico, a interação em fóruns de discussão online. O estudo é iluminador pois, embora haja uma série de publicações e trabalhos que utilizem a Linguística de Corpus para a análise de escolhas lexicais de um grupo específico, não há estudos comparativos que utilizem fóruns de discussão profissional online como objeto de pesquisa, assim como não há um número significativo de pesquisa com foco no discurso de PN e PNN de língua inglesa em ambiente digital.

¹ Orientadora da pesquisa: Tania Shepherd.

Além dos fatores mencionados acima, outro fator também foi de suma importância e relevância para a escolha das questões a serem estudadas: a constatação de que a interação dos PNN em fóruns de discussão online é visivelmente diferente daquela percebida entre os PN, seja por uma necessidade de usar a língua corretamente, de acordo com as regras ditadas pela gramática normativa, seja por falta de conhecimento do gênero em questão.

Coffin *et al.* [2005]² destacam que:

A última década tem tido um interesse considerável pela essência da comunicação mediada pelo computador (CMC) e seu papel na educação, inclusive no ensino superior. [...] mas até hoje poucos são os estudos lingüísticos que focaram especificamente na maneira com que a argumentação ocorre em discussões online escritas.

Assim sendo, nota-se que poucos são os estudos lingüísticos com foco em CMC feitos até o momento e é justamente este fator que faz com que este meio de comunicação seja de extrema relevância para a Lingüística de Corpus, área cujo interesse central está no uso autêntico de uma determinada língua coletada em forma digital.

Dado o objetivo do trabalho, as questões de pesquisa investigadas foram as seguintes:

Os grupos de PN e PNN interagem de forma similar ou diferente em fóruns de discussão, que possam ser identificadas em termos lexicais?

Se as formas de interação são diferentes, existem padrões lexicais que os caracterizem como grupos?

Como essas diferenças, caso ocorram, podem ser explicadas?

2) Metodologia

Conforme mencionado anteriormente, o trabalho aqui proposto teve como fundamentação teórica a Linguística de Corpus, área dentro da Linguística que se ocupa da extração e análise de corpora a fim de estudar uma determinada língua ou variedade lingüística, por esta ser apropriada para investigar a linguagem em larga escala como ela ocorre em situações reais, bem como ordenar, contrastar e comparar textos digitalizados.

O uso do computador nesta, que alguns consideram como área da Lingüística e outros como somente uma metodologia, é de suma importância, pois este facilita a extração e reor-

² Tradução nossa para: “Over the last decade there has been considerable interest in the nature of computer-mediated communication (CMC) and its role in education, including higher education. (...) but to date there have been few linguistic studies which have specifically focused on the way in which argumentation operates in text-based electronic conferences.” [Coffin *et al.*, 2005, p.41].

ganização das informações contidas em um *corpus*, para uma futura análise e observação dos dados lingüísticos. As ferramentas computacionais mais comuns são:

- ✓ Listas de palavras: permitem fazer a contagem das palavras em um determinado corpus.
- ✓ Concordanciadores: estes fornecem listas de ocorrências de palavras específicas ou grupos de palavras em um determinado contexto, de acordo com a necessidade do usuário.
- ✓ Palavras-chave: estas fornecem as palavras-chave em um determinado texto.

Conforme Scott [2001] aponta,

O computador e seu software são apenas ferramentas. Fica a nosso critério usar a nossa criatividade para decidir o que fazer com eles. O fato de não ser possível prever o que cada um de nós possivelmente escolherá para descobrir desenvolve uma fascinação na nossa observação das novas descobertas [...].³

Esta diversidade nas possibilidades de análise é justamente o que faz com que os trabalhos feitos através da Linguística de Corpus tornem-se extremamente ricos e desafiadores, pois cabe ao pesquisador utilizar esta diversidade da forma mais eficaz e conveniente para a pesquisa a ser desenvolvida.

No que concerne aos dados utilizados na pesquisa em questão, estes foram extraídos de dois corpora distintos. O primeiro é composto de 35.531 palavras, extraídas de um fórum de discussão online profissional de uma determinada escola de idiomas brasileira, cujos participantes são todos professores brasileiros de língua inglesa. O segundo é composto de 58.784 palavras, extraídas de três diferentes fóruns de discussão online profissionais.

O corpus composto pelo discurso dos PNN de língua inglesa foi considerado o corpus de estudo e procurou-se um fórum de discussão online profissional como garantia de que todos os participantes fossem profissionais da língua inglesa. Após a escolha do fórum ter sido feita, a “fala” dos participantes começou a ser coletada manualmente, isto é, copiada e salva em bloco de notas, para uma futura análise através da ferramenta computacional escolhida para o presente trabalho, o *WordSmith Tool*, versão 3.0.

O corpus composto pelo discurso dos PN de língua inglesa foi considerado o corpus de referência e a compilação do corpus em questão foi feita da mesma forma que a compilação do corpus de estudo, isto é, manualmente. Contudo três fóruns de discussão online foram uti-

³ Tradução nossa para: “The computer and its software are only tools. It is up to us to use our creativity in deciding what to do with them. The fact that it is not possible to predict what we may each choose to find out, lends a fascination to our observation of the new findings [...]” [Scott, 2001, p.64].

lizados, basicamente por três motivos, conforme explicação a seguir. Além do fato de o corpus de referência ter que ser pelo menos duas vezes maior que o corpus de estudo [Sardinha, 2004: 343]⁴, observou-se uma tendência entre os PN de língua inglesa de expressarem suas idéias e opiniões com poucas palavras. Para dificultar ainda mais tal compilação, poucas são as listas de discussão online que divulgam a nacionalidade dos participantes, fazendo com que, muitas das vezes, as conversas entre eles tivessem que ser lidas na íntegra, para que suas respectivas origens fossem descobertas.

Vale a pena ressaltar que optamos por trabalhar com o mesmo gênero (listas de discussão), sujeitos que tivessem as mesmas ocupações (professores de língua inglesa). Nossa variável foi a nacionalidade: brasileiros e falantes nativos de língua inglesa.

Para a pesquisa desenvolvida, apenas os quadrigramas (conjunto de quatro palavras) foram analisados devido ao fato de esses feixes lexicais oferecerem uma ideia clara de suas respectivas estruturas e funções, além de serem menos comuns [Biber et al. 2006: 992; Scott & Tribble, 2006: 141].

Como não há uma metodologia pré-estabelecida para a análise comparativa de n-gramas (sejam eles bigramas, trigramas, quadrigramas, etc), optamos por um ponto de corte. Apenas os quadrigramas que ocorreram pelo menos quatro vezes foram analisados, uma vez que ocorrências abaixo de quatro repetições podem ser atribuídas ao acaso ou a idiosincrasias. Para este estudo, foram aplicadas as três categorias desenvolvidas por Hyland [2008] para dar conta da natureza de quadrigramas. Essas, são: 1- padrões focados no texto (padrões argumentativos, discursivos e avaliativos de um texto), 2- padrões focados no tópico (padrões focados na experiência e atividades do mundo real) e 3- padrões focados no participante (padrões focados nos interlocutores e suas respectivas formas de interação e engajamento). Para efeito de identificação, atribuímos a cada categoria, respectivamente, as seguintes letras Te, To e P.

3) Análise

Após o ponto de corte ter sido estabelecido (4 repetições), 59 quadrigramas foram encontrados no corpus de referência, bem como 38 no corpus de estudo, respectivamente.

Os quadrigramas foram classificados conforme as categorias acima e o seguinte quadro foi obtido:

⁴ Devido à dificuldade para se coletar o discurso dos PN de língua inglesa e ao fato de a pesquisa ter sido em menos de um ano (como pós-graduação Lato Sensu, não houve tempo hábil de formar um corpus de referência 2 vezes maior que o de estudo. Contudo uma nova fase de compilação para enriquecer tal corpus já foi iniciada.

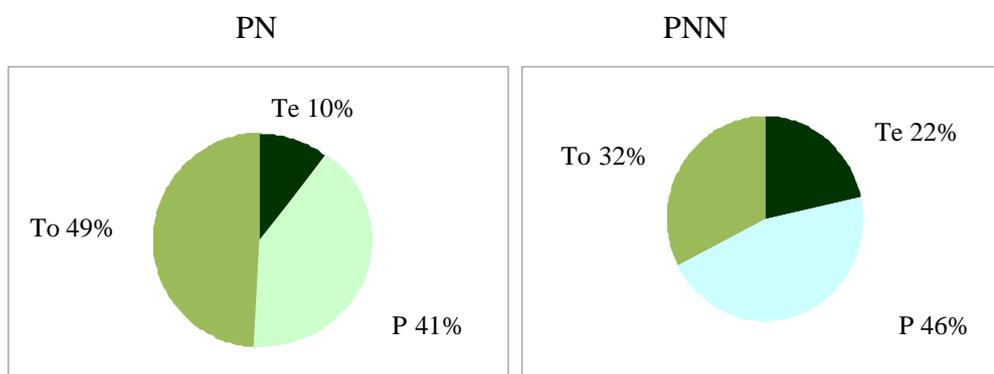
Quadro 1: Distribuição de categorias analíticas entre professores nativos e não nativos

1	WHAT KINDS OF ASSESSMENTS	To	1	OF THE LEARNING PROCESS	To
2	VIEW OF LANGUAGE LEARNING	To	2	PART OF THE LEARNING	To
3	TO BE A TEACHER	To	3	WITH THE HELP OF	To
4	TO BE A GOOD	To	4	ATMOSPHERE IN THE CLASSROOM	To
5	THERE ARE VERY FEW	To	5	CLOSER TO THE PROCESS	To
6	THERE ARE SO MANY	To	6	END OF THE CLASS	To
7	THE REST OF THE	To	7	HAS TO DO WITH	To
8	THE MAGICAL REALIST VIEW	To	8	ICT IN THE CLASSROOM	To
9	THE GRAMMAR TRANSLATION METHOD	To	9	ONE OF MY GROUPS	To
10	THE CULTURE OF THE	To	10	THE PROCESS OF LEARNING	To
11	THE CONTEXT OF ELT	To	11	TO THE PROCESS OF	To
12	THE AGE OF #	To	12	WRITING ON THE BOARD	To
13	OF THE TARGET LANGUAGE	To	13	WHEN IT COMES TO	Te
14	LANGUAGE LEARNERS NEED TO	To	14	ON THE OTHER HAND	Te
15	IN THE SAME CLASS	To	15	THE END OF THE	Te
16	IN THE CONTEXT OF	To	16	AS MUCH AS POSSIBLE	Te
17	HOW TO USE THE	To	17	TO THINK ABOUT #	Te
18	HERE IN THE UK	To	18	AT THE END OF	Te
19	ETHICS IN LANGUAGE TEACHING	To	19	AT THE SAME TIME	Te
20	DEFAULT THREAD SUBSCRIPTION MO- DE	To	20	TO THE FACT THAT	Te
21	CULTURE OF THE TARGET COOPERATIVE AND COLLABORATIVE	To	21	TO THINK ABOUT HOW	Te
22	LEARNING	To	22	AS I SEE IT	P
23	AN ENGLISH LANGUAGE TEACHER	To	23	FORWARD TO HEARING FROM	P
24	A CUP OF COFFEE	To	24	HOW CAN WE MAKE	P
25	THE END OF THE	Te	25	LOOKING FORWARD TO HEARING	P
26	ON THE OTHER HAND	Te	26	LOOKING FORWARD TO READING	P
27	MAKE A CUP OF	Te	27	WHAT DO YOU THINK	P
28	IS ONE OF THE	Te	28	I TOTALLY AGREE WITH	P
29	IN THE SAME WAY	Te	29	IT SEEMS TO ME	P
30	AT THE END OF	Te	30	FORWARD TO READING FROM	P
31	YOU SHOULD SAY SO	P	31	I HAVEN'T HAD ANY	P
32	YOU MIGHT WANT TO	P	32	IT IS IMPORTANT TO	P
33	WHAT YOU MEAN BY	P	33	SEEMS TO ME THAT	P
34	WHAT DO YOU WANT	P	34	THANK YOU FOR YOUR	P
35	WHAT DO YOU DO	P	35	THANKS FOR YOUR CONTRIBUTIONS	P
36	TO BE ABLE TO	P	36	THERE MUST BE A	P
37	THAT YOU SHOULD SAY	P	37	TO HEARING FROM YOU	P
38	SHOULD SAY SO SURPRISES	P	38	TO READING FROM YOU	P
39	SHOULD BE ABLE TO	P			
40	SEEMS TO ME THAT	P			
41	SAY SO SURPRISES ME	P			
42	QUERIES FROM OTHER MEMBERS	P			
43	OF ASSESSMENTS DO YOU	P			
44	NEED TO BE ABLE	P			
45	KNOW IF YOU HAVE	P			
46	KINDS OF ASSESSMENTS DO	P			
47	IT SEEMS TO ME	P			
48	I WOULD LIKE TO	P			
49	I WOULD BE INTERESTED	P			
50	I THINK YOU ARE	P			
51	I THINK THERE ARE	P			
52	I THINK IT IS	P			
53	I DON'T KNOW IF	P			
54	I AGREE WITH YOU	P			
55	FROM OTHER MEMBERS AND	P			
56	DO YOU WANT TO	P			
57	DO YOU GIVE FOR	P			
58	ASSESSMENTS DO YOU GIVE	P			
59	AGREE WITH THE IDEA	P			

Ao categorizar os quadrigramas extraídos de acordo com o estudo feito por Hyland [2008] e transformar a tabela acima em um gráfico, as diferenças entre ambos os grupos tor-

naram-se mais evidentes.

FIGURA 1: Distribuição de categorias analíticas entre professores nativos e não nativos



Primeiramente, percebe-se que os PN tendem a expressar suas idéias e pontos de vista mencionando situações reais de seu próprio mundo (49% de quadrigramas focados no tópico) e demonstrando suas opiniões e avaliações sobre o assunto discutido (41% de quadrigramas cujos padrões estavam focados no participante). Apenas 10% dos quadrigramas extraídos eram marcadores textuais, o que sugere que este grupo tende a tratar suas interações no ambiente em questão, isto é, o fórum de discussão online, como uma conversa.

Conforme exposto nos gráficos acima, PNN usam marcadores textuais frequentemente (22% dos quadrigramas encontrados têm seus padrões focados no texto), o que pode sugerir que este grupo trata o fórum de discussão como texto escrito (e-mails ou cartas). Além disso, poderia ser depreendido do diagrama que este grupo menciona experiências de vida e interage com os outros participantes. Contudo, um olhar mais criterioso para a tabela acima mostra que alguns dos quadrigramas categorizados como “padrões com foco no participante” (46%) tendem a se referir a expressões de despedida comumente usadas em cartas ou e-mails (Ex.: “looking forward to”, isto é, “guardo ansioso para”). Por último, pode ser inferido que os PNN tendem a focar em um tópico específico (32%) coletivamente e repetidamente, formando, assim, quadrigramas.

4) Conclusões

Este artigo analisou as interações de PN e PNN em fóruns de discussão profissionais. Teve como objetivo identificar padrões no discurso dos participantes nessas discussões. Dessa forma, a presente pesquisa não teve como objeto de estudo o discurso de professores isoladamente, mas os padrões lexicais que os caracterizam como um grupo, no ambiente de fóruns

online.

Nesta parte final, faz-se necessário o retorno às perguntas de pesquisa postas anteriormente:

- Os grupos de PN e PNN interagem de forma similar ou diferente em fóruns de discussão, que possam ser identificadas em termos lexicais?
- Como essas diferenças, caso ocorram, podem ser explicadas?

Em termos da questão 1, os corpora analisados mostram evidência de várias diferenças no discurso de PN e PNN. A primeira diferença refere-se à organização do texto. Os PNN parecem mais preocupados com a organização textual do que PN. Isto torna-se mais evidente quando o número de quadrigramas focados no texto dos dois grupos é comparado. Uma outra diferença marcante está relacionada à forma que os grupos trocam experiências e práticas profissionais. Os corpora analisados mostram que PN tendem a usar mais quadrigramas focados em assuntos relativos à prática profissional, demonstrando uma tendência maior a dividir suas visões e práticas diárias.

Em relação à segunda questão, os resultados sugerem que os dois grupos analisados usam quadrigramas similares, mas os PNN tendem a usar repetidamente os mesmos quadrigramas nas diversas discussões estudadas. Este fato indica que os PNN podem ter um repertório mais restrito dessas formas recorrentes.

Nos corpora estudados, com base nas categorias de Hyland [2008], observa-se que os PNN usam um grande número de quadrigramas com foco no texto. Em uma análise mais detalhada, verifica-se que diversos quadrigramas usados pelos PNN são característicos de cartas e e-mails, indicando que os PNN tendem a tratar as listas de discussão como textos escritos.

Os resultados aqui apresentados devem ser tratados com precaução, pois o corpus de estudo é relativamente pequeno. Aliado a este fato, deve-se considerar que, apesar de haver inúmeros estudos sobre o uso de agrupamentos ou feixes de palavras, este campo de listas de discussão profissionais de professores de inglês ainda é relativamente pouco explorado.

Considerando as possíveis limitações desta pesquisa, o corpus de estudo representa apenas as interações que acontecem na instituição onde o mesmo foi coletado e não é necessariamente representativo do discurso da maioria dos PNN no Brasil. Desta forma, as conclusões aqui apresentadas são apenas iniciais e podem ser usadas para estudos futuros. A riqueza do corpus de estudo pode ser objeto de pesquisa de diversos outros estudos relacionados, por e-

xemplo, à argumentação e ao posicionamento dos participantes nas listas de discussão profissionais.

Referências

- BIBER, D. Stance in spoken and written university registers. *Journal of English for Academic Purposes*, vol. 5, 2006. p. 97–116.
- COFFIN, C; HEWINGS, A. Engaging Electronically. Using CMC to develop students' argumentation skills in Higher Education. United Kingdom. *Language and Education*, 2005. p. 32- 49.
- HYLAND, Ken. As can be seen: lexical bundles and disciplinary variation. *Science Direct*, 2008. p. 4-21.
- SCOTT, M. *Small corpus studies and ELT: theory and practice*. In: GHADESSY, M.; HENRY, A.; Roseberry, R. (Org). Amsterdam, Holanda: John Benjamins, 2001.
- SCOTT, M; TRIBBLE, C. *Textual Patterns: key words and corpus analysis in language education*. Amsterdam, Holanda: John Benjamins, 2006.
- SARDINHA, T. B. *Linguística de Corpus*. Barueri, SP: Manole, 2004.