

**BLOCOS DE PALAVRAS NA FALA DE CARIOCAS:
ANÁLISE DE QUADRIGRAMAS NO CORPUS NURC**

Gustavo Estef Lino da Silveira (UERJ)
gutolino@oi.com.br

RESUMO

Análise de blocos de quatro palavras dentro de um *corpus* formado por falantes cariocas da língua portuguesa (NURC). Como base teórica, foi utilizada a classificação de Ken Hyland (2008) acerca de blocos lexicais, que divide as palavras quanto ao tamanho (denominado feixe) e quanto ao uso (de acordo com o gênero utilizado ou a modalidade: escrita ou oral). O *corpus* foi analisado com o auxílio de um programa de computador e os cinquenta blocos de quatro palavras (quadrigramas) mais presentes foram selecionados. Os resultados demonstraram que os falantes utilizavam, em grande parte, blocos voltados para a interação verbal com seu interlocutor e, por isso, assemelham-se a outros *corpora* de literatura infantil. Sendo assim, o estudo abre espaço para uma futura investigação contrastiva entre *corpora* escritos e orais.

Palavras-chave: Linguística textual. Feixes lexicais. Quadrigramas.

1. Introdução

Muitos são os estudos acerca do conceito de que a linguagem consiste não de palavras isoladas, mas sim de blocos de palavras que carregam sentido em si e unem-se a outros blocos para trazer significação a uma mensagem. Tais trabalhos têm estudado estes blocos de palavras que podem ser denominados de diversas maneiras, tais como: *chunks*, *bundles*, *lexical chunks*, *lexical units*, *multi-word units*, *ready-mades*, *prefabricated language*, *holophrases*, *clusters*, entre outros. Ken Hyland (2008) faz uso da denominação *lexical bundles* – aqui traduzida por feixes lexicais – para denominar os blocos de palavras que carregam um sentido único, sendo vistas como inseparáveis. Haja vista que este artigo tratará da visão de Ken Hyland sobre os agrupamentos de palavras usaremos a denominação dada por este autor.

Ken Hyland (2008) afirma que os feixes lexicais são um componente importante da fluência da produção linguística e um fator-chave na mensuração do sucesso da aprendizagem de uma língua. Com isso, o autor propõe que analisemos os feixes lexicais em n-gramas, que são feixes lexicais contendo “n” palavras que se repetem em um número considerável de vezes ao longo de textos escritos. Sendo assim, Ken Hyland avança na classificação dos n-gramas ao sugerir a subdivisão dos mesmos em:

trigramas, quadrigramas e pentagramas. No entanto, o autor parece priorizar a análise dos quadrigramas em suas pesquisas, pois, segundo o autor, estes carregariam maior significação e poderiam vir a ser encontrados com maior frequência.

Pode-se afirmar também que um n-grama vem a ser um feixe lexical que tem um número “n” de palavras que completam seu sentido e que não podem ser analisadas separadamente, mas apenas dentro do feixe. Assim sendo, um quadrigrama seria um feixe lexical de quatro palavras que forma um único sentido. Tais palavras não se agrupam por acaso, mas andam juntas no discurso como um bloco de significado único. Segundo John Sinclair (1991, p. 108), a maioria das palavras não tem um significado independente, ou seja, fazem apenas sentido dentro de um feixe lexical.

A análise de feixes lexicais pode vir a permitir com que possamos reconhecer se um texto pertence a um determinado gênero. Todavia, a ausência de feixes lexicais em um texto poderia demonstrar que um autor ainda não tem o domínio necessário tanto da língua da qual faz uso quanto do gênero em que está escrevendo.

Em textos de aprendizes iniciantes de uma língua estrangeira, por exemplo, pode-se notar pouco uso de expressões idiomáticas ou de feixes lexicais, enquanto que, em falantes que dominam mais a língua, o uso de tais feixes pode ser mais comum de ser encontrado.

O uso de alguns feixes lexicais específicos, especialmente, no início de um texto, antecipa ao leitor que tipo de texto o mesmo tem em mãos, facilitando assim sua leitura e compreensão. Por exemplo: ao ler um texto que se inicia com um pentagrama “o objetivo deste relatório é”, já é antecipado ao leitor o gênero do texto em questão (um relatório) fazendo com que o mesmo ative seu conhecimento prévio de outros exemplares prototípicos desse gênero. Assim também, ao nos depararmos com o feixe lexical “era uma vez”, sabemos que estamos prestes a iniciar a leitura de uma história infantil. Assim sendo, os feixes auxiliam o processo de leitura. Também é mister ressaltar que um feixe lexical usado inadequadamente em um determinado texto pode vir a causar estranheza ao leitor, pois é possível relacioná-lo a outros exemplares de textos do mesmo gênero e perceber que aquela escolha linguística não foi apropriada.

2. Fundamentação teórica

Ken Hyland afirma que seu modelo teórico de classificação dos feixes lexicais originou-se do sistema criado por Douglas Biber (2006). O estudo do primeiro autor agrupa os feixes lexicais em três tipos de orientação, sendo eles: os feixes orientados para o tópico, os orientados para o texto e os orientados para a interação. As categorias em questão também estão calcadas nas três meta-funções da linguística sistêmico-funcional *hallidayana* (meta-funções textual, ideacional e interpessoal).

Passamos agora a descrever cada uma dessas categorias de classificação de feixes lexicais. Primeiro, tratemos dos feixes orientados para o tópico, ou pesquisa. Esses, ajudam o escritor a estruturar a sua argumentação em termos de localização de tempo e espaço, os procedimentos e propósitos do texto, quantificações, descrições de algo e os relacionados ao campo de pesquisa.

Em segundo lugar, os feixes lexicais orientados para a organização textual. Estes ampliam o significado do texto auxiliando na coesão e coerência do mesmo. Estabelecem relações de contraste ou de adição de ideias entre elementos, marcam relações de causa e consequência, organizam e estruturam sequências do discurso e situam argumentos especificando condições específicas.

Por último, temos os feixes lexicais orientados para a interação. Estes estão focados na relação entre leitor e escritor. Podem indicar a atitude ou avaliação do autor e também podem dirigir-se diretamente ao leitor.

Tendo apresentado os preceitos teóricos que norteiam a análise de dados deste trabalho, passamos a seguir a detalhar os passos da metodologia da presente pesquisa.

3. Metodologia

Nesta seção, trataremos tanto da descrição da metodologia usada no estudo de quadrigramas feito por Ken Hyland (2008), quanto da metodologia de pesquisa deste estudo.

Ken Hyland compilou um *corpus* próprio de textos acadêmicos formado por artigos de pesquisa, teses de doutorado e dissertações de mestrado, sendo essas últimas das áreas humanas e exatas. O autor deci-

diu analisar quadrigramas porque esses são muito mais comuns que pentagramas e oferecem uma gama de estruturas e funções bem mais claras que os trigramas.

Já para o estudo que originou este artigo foi utilizado o *corpus* NURC. O *corpus* faz parte do Projeto da Norma Urbana Oral Culta do Rio de Janeiro, um projeto da Faculdade de Letras da Universidade Federal do Rio de Janeiro. Esse *corpus* é formado por entrevistas gravadas nas décadas de 1970 e 90 do século XX, num total de 350 horas, com informantes com nível superior completo, nascidos no Rio de Janeiro e filhos de pais, preferencialmente, cariocas.

O *corpus* foi analisado através do programa de computador Wordsmith Tools 5.0 e uma lista dos cinquenta quadrigramas mais frequentes foi criada e, conseqüentemente, classificada de acordo com o arcabouço teórico proposto por Ken Hyland. Conseqüentemente, os quadrigramas foram separados de acordo com as três categorias de feixes lexicais previamente mencionadas: tópico, interação e organização.

Trataremos da análise dos dados obtidos no *corpus* NURC na próxima seção.

4. Análise dos dados

A partir do *corpus* NURC, foi-se em busca dos cinquenta quadrigramas mais frequentes com o auxílio do programa de computador. Também foi feita a classificação dos quadrigramas de acordo com os três tipos de orientação mencionados por Ken Hyland.

Conforme podemos ver na tabela que segue abaixo (Tabela 1), 90% dos quadrigramas encontrados no *corpus* (um total de 45 quadrigramas) são de orientação interacional, ou seja, estabelecem uma relação entre autor e leitor.

Apenas 10% dos quadrigramas mais frequentes (5 exemplares) são orientados para o tópico, ajudando o leitor a se localizar em relação ao tempo, local, quantidade etc. Por fim, não encontramos feixes lexicais orientados para a organização textual entre os cinquenta quadrigramas mais frequentes.

Círculo Fluminense de Estudos Filológicos e Linguísticos

	Quadrigramas – NURC	Categorias de Ken Hyland		
		Tópico	Interação	Organização
1	COMO É QUE É		X	
2	QUE EU ACHO QUE		X	
3	EU ACHO QUE É		X	
4	EU ACHO QUE O		X	
5	EU ACHO QUE EU		X	
6	O QUE QUE É		X	
7	EU ACHO QUE A		X	
8	MAS EU ACHO QUE		X	
9	O QUE É QUE		X	
10	EU VER QUE MAIS		X	
11	LOC EU ACHO QUE		X	
12	EU ACHO QUE TEM		X	
13	COMO É QUE VOCÊ		X	
14	EU ACHO QUE NÃO		X	
15	DEIXA EU VER QUE		X	
16	COMO É QUE SE		X	
17	QUE MAIS QUE TEM		X	
18	VER QUE MAIS QUE		X	
19	EU ME LEMBRO QUE		X	
20	O QUE QUE VOCÊ		X	
21	O RIO DE JANEIRO	X		
22	NÉ EU ACHO QUE		X	
23	DOC E O QUE		X	
24	QUE MAIS QUE EU		X	
25	EU ACHO QUE ISSO		X	
26	O QUE A GENTE		X	
27	É EU ACHO QUE		X	
28	QUE HOJE EM DIA	X		
29	QUE O RIO DE	X		
30	QUE É UMA COISA	X		
31	COMO É QUE O		X	
32	QUE A GENTE TEM		X	
33	ENTÃO EU ACHO QUE		X	
34	A GENTE TEM QUE		X	
35	O QUE QUE TEM		X	
36	QUE MAIS QUE TINHA	X		
37	DOC O QUE QUE		X	
38	É QUE O SENHOR		X	
39	E EU ACHO QUE		X	
40	E COMO É QUE		X	
41	E O QUE QUE		X	
42	EU VER O QUE		X	
43	ACHO QUE TEM QUE		X	
44	O QUE QUE EU		X	

45	NÃO EU ACHO QUE		X	
46	QUE VOCÊ ACHA QUE		X	
47	COMO É QUE CHAMA		X	
48	DOC COMO É QUE		X	
49	NÃO SEI O QUE		X	
50	DEIXA EU VER O		X	

Tabela 1: Tabela de quadrigramas extraídos e classificados do NURC.

O gráfico abaixo (**Gráfico 1**) nos auxilia a visualizar a frequência dos quadrigramas encontrado no *corpus* NURC:

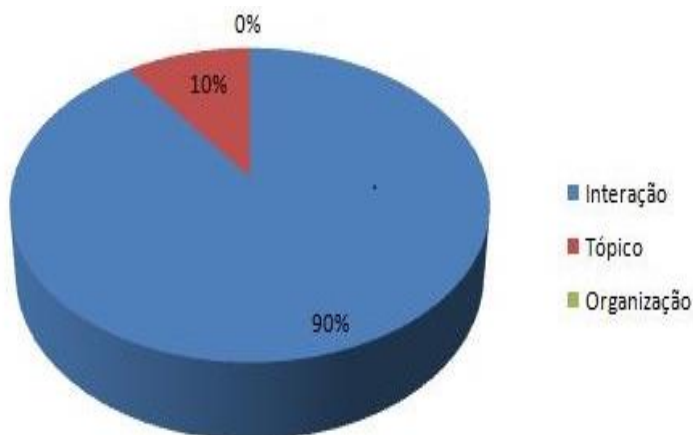


Gráfico 1: Gráfico com a divisão dos feixes encontrados no NURC

5. Conclusão

Ken Hyland, em seu estudo, afirma que mais da metade dos feixes mais comuns de cada tipo de texto dentro de seu *corpus* de estudo não era comum a outros tipos de texto. Isso quer dizer, que os feixes mais comuns no *subcorpus* de linguística aplicada, por exemplo, pouco se assemelhavam ao de ciências exatas. Isso nos leva a constatar que o gênero influencia as escolhas léxico-gramaticais e fala por si mesmo. Com isso, cada tipo de texto e cada área de estudo terá sua própria léxico-gramática e, por conseguinte, seus feixes lexicais próprios.

Em relação ao *corpus* de estudo do *Corpus Digitalizado do Projeto da Norma Urbana Oral Culta do Rio de Janeiro*, devemos levar em consideração sua especificidade, pois, por se tratar de um *corpus* de lin-

guagem oral, onde entrevistas foram gravadas e depois transcritas para análise, há a prevalência de quadrigramas orientacionais de interação.

Esse *corpus* de discurso oral se assemelha a *corpora* de literatura infantil, por exemplo, onde também são encontrados muitos feixes lexicais interacionais, pelos quais o escritor estabelece uma conversa com seus pequenos leitores.

Também podemos afirmar que não encontramos quadrigramas de orientação organizacional possivelmente por não se tratar de um *corpus* de textos escritos.

Com isso, poderíamos dizer que esse estudo nos abre a perspectiva para um futuro estudo comparativo entre uma análise baseada na comparação entre *corpora* de textos escritos e falados a fim de se traçar um perfil das escolhas de tipos de quadrigramas orientacionais e suas frequências entre os dois *corpora*.

REFERÊNCIAS BIBLIOGRÁFICAS

BIBER, Douglas. *University language: a corpus-based study of spoken and written genres*. Amsterdam: Benjamin, 2006.

HYLAND, Ken. As can be seen: lexical bundles and disciplinary variation. In: _____. *English for Specific Purposes*. London: Elsevier, 2008, vol. 27.

NURC. *Corpus digitalizado do Projeto da Norma Urbana Oral Culta do Rio de Janeiro*. Disponível em: <<http://www.lettras.ufrj.br/nurc-rj>>. Acesso em: 16-04-2016.

SINCLAIR, John. *Corpus, concordance, collocation*. Oxford: OUP, 1991.