

**O USO DE MARCADORES DE DISCURSO
EM TEXTOS DE APRENDIZES DE INGLÊS COMO LE:
UM ESTUDO BASEADO EM CORPUS**

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

RESUMO

Ao estudarem uma língua estrangeira, muitos aprendizes são ensinados a fazer uso de marcadores de discurso em suas composições escritas. No entanto, durante a prática docente tem-se visto que nem sempre as redações contêm tais expressões de organização de texto tornando sua leitura muitas vezes com ideias truncadas. Para tal investigação foi compilado um *corpus* de redações de alunos considerados avançados contendo quase 400 textos e cerca de 100.000 palavras coletadas em seis filiais de um curso de inglês da cidade do Rio de Janeiro. Os aprendizes tiveram cerca de sete anos de estudos na instituição, são adolescentes terminando o ensino médio e/ou ingressando na carreira universitária. O estudo faz uso de preceitos da linguística de *corpus* através do auxílio de programas de computador para compilar e extrair grandes quantidades de dados. Também faz uso das categorias de Hyland (2008) sobre a classificação de blocos de palavras para se entender como os aprendizes organizam suas ideias. O resultado do presente estudo demonstrou que mesmo tendo havido grande ênfase no ensino de marcadores discursivos voltados para a organização do texto os aprendizes não fizeram tanto uso dos mesmos. Foram encontrados mais blocos de palavras voltados para o tópico e para interação autor-leitor. Tal achado pode sugerir que a ênfase dada no ensino dos blocos pesquisados não necessariamente resultou em aprendizagem.

Palavras-chave: Linguística de *corpus*. Marcadores discursivos.
Escrita em inglês. Linguística aplicada

1. Introdução

Muitos estudos têm demonstrado a necessidade de se ensinar blocos léxico-gramaticais e conseqüentemente blocos que tenham a função de organizar ideias na produção escrita em língua estrangeira. Outras pesquisas têm ressaltado a importância de analisarmos esses blocos como forma de avaliarmos idiomaticidade e, por conseguinte, o nível de proficiência em um idioma (HYLAND, 2008; LEWIS, 2009). Entretanto, estudos dessa natureza envolvendo populações de aprendizes brasileiros de língua inglesa ainda são escassos. Este estudo vem, portanto, acrescentar aos estudos existentes, não só por ter como sujeitos de pesquisa aprendizes adolescentes de inglês, mas também por ter um grupo com características bem homogêneas, isto é, todos os aprendizes estudavam em um

mesmo nível de uma mesma instituição particular de ensino de língua inglesa, de uma única cidade (Rio de Janeiro), e todos começaram a aprender a língua inglesa ainda crianças. Além disso, o presente artigo não apenas se deteve à análise de quaisquer blocos léxico-gramaticais utilizados por aprendizes de inglês como língua estrangeira, mas tem como foco a descrição do uso de blocos léxico-gramaticais voltados para a organização textual, os chamados marcadores discursivos.

2. Justificativa do estudo

Um estudo sobre o uso de marcadores discursivos seria demasiado extenso para esta pesquisa e segundo Hyland (2008) os blocos léxico-gramaticais poderiam ser classificados de acordo com a quantidade de palavras que os mesmos se anelam. Sendo assim, se um bloco de palavras é formado de dois elementos o autor sugere a nomenclatura bigrama, se for com três elementos trigramas, seguindo assim uma ordem crescente onde seriam classificados como quadrigramas, pentagramas etc.

No entanto, o autor ressalta que os trigramas tem seu sentido um tanto quanto opacos ao serem analisados trazendo pouco sentido senão dentro de um contexto de análise maior da frase. Por outro lado, os pentagramas, seriam bastante escassos de serem encontrados em um *corpus*. Logo, aliando significado e frequência, a escolha por quadrigramas parece ser um recorte considerável para esse estudo de análise de marcadores discursivos. Em outras palavras, esse estudo se limita a analisar o uso de marcadores discursivos formados por quatro palavras, ou seja, os quadrigramas.

A pergunta que norteia esta pesquisa foi baseada em um estudo de Hyland (2008), onde o autor classifica blocos léxico-gramaticais de quatro palavras de acordo com três tipos diferentes de uso (conforme serão definidos na seção de fundamentação teórica). Sendo assim, o presente estudo visa analisar se ao fazermos uso do sistema para categorização de blocos léxico-gramaticais desenvolvido por Hyland como poderíamos identificar que tipos de quadrigramas são mais usados pelos aprendizes de inglês como língua estrangeira?

3. Metodologia de pesquisa

Este trabalho faz uso de ferramentas digitais para compilar, formatar, classificar e auxiliar a análise de dados oriundos de um grande

número de textos escritos por aprendizes de inglês como língua estrangeira. Dentro da linguística aplicada, a disciplina que se ocupa de compilação e tratamento de dados digitalizados é a linguística de *corpus*. Pode-se vir a questionar se a linguística de *corpus*, é uma área da linguística ou apenas um conjunto de métodos em si. A linguística de *corpus* trabalha com dados empíricos, oriundos da linguagem em uso e com esses dados é possível fazer deduções sobre a língua.

Alguns linguistas consideram a linguística de *corpus* uma metodologia e outros a veem como teoria. Em um dos primeiros volumes dedicados à linguística de *corpus*, McEnery e Wilson (1996, p. 2) perguntam e respondem textualmente “A linguística de *corpus* é um ramo da linguística? A resposta a essa pergunta é tanto *sim* como *não*”. Os autores alegam que a linguística de *corpus* não tem o mesmo estatuto da semântica, sintaxe ou sociolinguística visto que essas disciplinas têm um objeto de investigação definido. Ao mesmo tempo, alegam que o termo *corpus* pode ser atrelado a cada uma das áreas da linguística, gerando portanto a “semântica de *corpus*”, a “sintaxe de *corpus*”, por exemplo, em oposto à semântica ou sintaxe não baseadas em *corpora*. López (2009) afirma que se um dos princípios da pesquisa linguística é a observação e descrição de fenômenos, então, a linguística de *corpus* é mais do que uma metodologia, pois permite ao analista extrair dados, quantificá-los, observá-los e interpretá-los. Dessa forma, a natureza da pesquisa baseada em *corpus* não pode ser classificada nem como quantitativa e nem como qualitativa, pois já nasce híbrida em sua natureza. Sendo assim, López argumenta que a partir da extração de dados de grandes quantidades de textos através do auxílio de ferramentas de computador a pesquisa com base em preceitos da linguística de *corpus* permite não apenas a extração desses dados, mas também dá ao linguista evidências irrefutáveis que podem ser analisadas com base no *corpus*. A linguística de *corpus* permite que o pesquisador trabalhe com grandes quantidades de dados, sendo estudados tanto em termos de frequência quanto abrangência. Como resultado, o que deveria levar meses de trabalho através da análise manual, passa a levar segundos através do auxílio de programas de computador.

4. Fundamentação teórica

Em um estudo de 2008, no qual analisou textos acadêmicos, Hyland fez uma proposta enxuta para lidar com a categorização de quadrigramas: sugeriu que os blocos léxico-gramaticais extraídos de seu

corpus fossem divididos de acordo com as três macrofunções que porventura poderiam estar desempenhando no discurso. Haveria então: blocos orientados para o tópico, para o texto e os orientados para o leitor.

Em primeiro lugar temos os quadrigramas orientados para o tópico. Estes ajudariam os escritores a estruturar suas atividades e experiências do mundo real. Esses quadrigramas poderiam ser ainda subdivididos em categorias como lugar, procedimento, quantificação, descrição e tópico propriamente dito, conforme listados no **Quadro 1** abaixo.

	Propósito	Exemplos
Localização	Indicam tempo e lugar	at the beginning of, at the same time, in the present study
Procedimento	Sinalizam o procedimento ou papel de algo/alguém	the use of the, the role of the, the purpose of the, the operation of the
Quantificação	Relacionados à mensuração e valores	the magnitude of the, a wide range of, one of the most
Descrição	Possuem caráter descritivo	the structure of the, the size of the, the surface of the
Tópico	Voltados intrinsecamente para o tópico/ local	In the Hong Kong, the currency board system

Quadro 1: Quadrigramas orientados para o tópico

Em segundo lugar Hyland verificou a presença de uma categoria de blocos voltados para o texto. Esses estariam relacionados à organização textual e evidenciariam como seu significado pode atrelar valor às ideias e argumentos escritos relacionando, adicionando ou contrastando fatos, conforme pode ser visto no **Quadro 2** a seguir. Esses quadrigramas são o ponto nodal desta análise pois tratam-se dos marcadores discursivos que espera-se que os aprendizes façam uso em seus textos, pois assim foram ensinados a fazê-lo nas aulas de composição escrita na instituição de ensino analisada.

	Propósito	Exemplos
Marcador de transição	Estabelecem relações de adição ou contraste entre elementos e/ou ideias	On the other hand, in addition to the, in contrast to the
Marcador de resultados	Marcam relações de inferências ou causa entre elementos	As a result of it, it was found that, these results suggest that
Marcador estrutural	Marcadores reflexivos do texto que organizam pedaços do discurso ou direcionam o leitor a algum lugar no texto	In the present study, in the next section, as shown in figure
Marcador de suporte	Situam os argumentos especificando condições limitantes	In the case of, with respect to the, on the basis of, in the presence of, with the exception of

Quadro 2: Quadrigramas orientados para o texto.

A terceira e última categoria descreve os quadrigramas orientados para o leitor, cujo objetivo é estabelecer diálogo, uma espécie de direcionamento mais específico ao leitor do texto. Os quadrigramas orientados para o leitor estão subdivididos em dois tipos: os marcadores de postura e os de engajamento, conforme Quadro 5 a seguir.

	Propósito	Exemplos
Marcador de postura	Retratam a atitude a avaliação do autor	Are likely to be, may be due to, it is possible that
Marcador de engajamento	Dirige-se diretamente ao leitor	It should be noted that, as can be seen

Quadro 3: Quadrigramas voltados para o leitor.

Hyland (2008, p. 20) sugeriu com seu estudo que uma análise detalhada de n-gramas pode oferecer *insights* a uma dimensão crucial da análise de gênero muitas vezes ignorada. Com isso, tal análise pode vir a proporcionar um melhor entendimento das várias formas com que escritores fazem uso de recursos linguísticos em inglês em diferentes contextos acadêmicos.

5. O corpus de análise

O *corpus* deste estudo consiste de 394 redações coletadas entre os anos de 2009 e 2012 em seis filiais do Rio de Janeiro de um dos maiores cursos livres de inglês do Brasil abrangendo as filiais da Zona Norte, Zona Oeste e Grande Rio. Cada aprendiz escreveu um dos três tipos de textos trabalhados em sala ao longo do semestre durante uma avaliação escrita de final de período: história pessoal, resenha de filme e artigo.

O *corpus* compilado para este estudo é um *corpus* sincrônico, pois faz um recorte da língua dos aprendizes em um momento “X”. O tamanho do *corpus* contabiliza 91.333 *tokens*. Sendo assim, o *corpus* poderia ser classificado como de tamanho pequeno-médio, conforme afirma Berber-Sardinha (2004, p. 26), que estabelece essa categorização para *corpora* que têm entre 80 e 250 mil palavras. Também é importante mencionar que cada redação contém aproximadamente 250 palavras.

6. Análise dos dados

Na tentativa de entender os quadrigramas mais usados no *corpus* e a função desempenhada pelos mesmos, recorreremos às categorizações maiores de Hyland (tópico, texto e leitor). Os quadrigramas mais fre-

quentes encontrados, com 73% dos casos, foram aqueles voltados para o tópico, conforme pode ser visto na **Tabela 1** abaixo.

Núm ⁶⁷	Quadrigrama	Frequência
1	Tells the story of	38
2	The end of the	24
3	In the middle of	23
4	The story of a	22
5	One of the most	18
6	The most popular sport	17
7	The special effects are	17
8	One of the best	16
9	It tells the story	15
11	An unforgettable afternoon it	14
12	At the same time	14
14	Is a film that	13
15	Is one of the	13
16	The middle of the	13
17	To go to the	13
19	Have a lot of	12
20	I was going to	12
21	Unforgettable afternoon it was	12
23	In Rio de Janeiro	11
24	Life when I was	11
25	Most popular sport in	11
26	My life when I	11
28	A lot of things	10
30	Falls in love with	10
33	The film is set	10
34	The name of the	10
35	The plot is about	10
36	Film tells the story	9
37	For the first time	9
39	Plot is about a	9
40	The film tells the	9
42	With a lot of	9
43	Film is set in	8
45	Is a very good	8
46	Of the film is	8
48	Tells the story about	8
49	When I was a	8
50	Years ago I was	8
51	Are a lot of	7

⁶⁷ O número em questão refere-se à ordem numérica de extração dos quadrigramas pelo programa de computador e depois reagrupados manualmente por mim de acordo com a classificação proposta por Hyland em três diferentes tabelas.

52	Had a lot of	7
53	I was a child	7
56	It is a very	7
57	Sport in my country	7
59	The ball with the	7
60	The movie tells the	7
61	The resto f the	7
63	To go to a	7
64	Was a child I	7
65	Win the Oscar but	7
66	A lot of fun	6
67	A lot of people	6
68	A walk to remember	6
69	And in the end	6
70	And one of them	6
71	And tells the story	6
72	As the story unfolds	6
74	Beginning to the end	6
75	Changed my life because	6
76	Didn't win an Oscar	6
77	End of the film	6
78	Event changed my life	6
79	Fall in love with	6
81	Harry Potter and the	6
85	I was in the	6
86	I went to the	6
87	In front of the	6
91	Movie tells the story	6
92	Name of the film	6
93	Of the most famous	6
96	Story of a man	6
97	The beginning to the	6
98	The main character is	6
99	The Oscar but should	6
101	There are a lot	6
103	Went to the beach	6
104	A child I used	5
105	A story about a	5
106	Afternoon it was a	5
107	Afternoon when I was	5
108	An unforgettable afternoon I	5
109	An unforgettable afternoon when	5
112	And it is a	5
113	And the most important	5
114	And you have to	5
115	Ball with the hands	5
116	Based on a book	5
117	Because it is a	5

118	Changed my life was	5
119	Child I used to	5
120	Decided to go to	5
122	Did not win the	5
123	Film is directed by	5
128	In love with a	5
129	In love with her	5
130	In my country is	5
132	In this day I	5
133	Is a movie that	5
134	is based on a	5
136	It was an unforgettable	5
138	Lot of special effects	5
141	My mother and my	5
142	Not win the Oscar	5
143	Of a group of	5
144	Of a man who	5
145	Of Rio de Janeiro	5
146	Of the movie is	5
147	Out of the blue	5
148	Players in each team	5
149	Plays the role of	5
150	Popular sport in Brazil	5

Tabela 1: Quadrigramas voltados para o tópicos

Em segundo lugar com 23% dos quadrigramas estão os voltados para a interação entre escritor e leitor, conforme demonstrados na **Tabela 2** abaixo:

Núm	Quadrigrama	Frequência
10	Should have won the	15
13	Have won the Oscar	14
18	Changed my life when	12
22	I highly recommend it	11
27	Really changed my life	11
29	Changed my life I	10
31	Have won an Oscar	10
32	Should have won an	10
38	My point of view	9
41	What strikes you most	9
44	I would like to	8
54	I will never forget	7
58	Strikes you most is	7
62	This film should have	7
73	Asked me if I	6
80	Film should have won	6
82	I didn't want to	6
83	I don't know why	6

84	I have ever seen	6
88	Is a must see	6
89	Is the most popular	6
90	Me to go to	6
94	Recommend this film because	6
95	So I decided to	6
110	And I decided to	5
111	And I had to	5
121	Did a great job	5
124	I didn't know what	5
125	I had to do	5
126	I had to go	5
127	I highly recommend this	5
135	It should have won	5
137	It's a must see	5
139	My life because I	5
140	My life I was	5

Tabela 2: Quadrigramas voltados para o leitor

E por último com apenas 4% os quadrigramas organizacionais, conforme podem ser vistos na **Tabela 3** a seguir:

Núm	Quadrigrama	Frequência
47	On the other hand	8
55	In the end of	7
100	The other hand the	6
102	To sum up the	6
131	In the end he	5

Tabela 3: Quadrigramas voltados para a organização do texto

Decidi elaborar um **Gráfico 1** com base nas três categorias de classificação dos quadrigramas proposta por Hyland (2008), que contempla os quadrigramas voltados para o tópico do texto, para a organização textual e a interação.

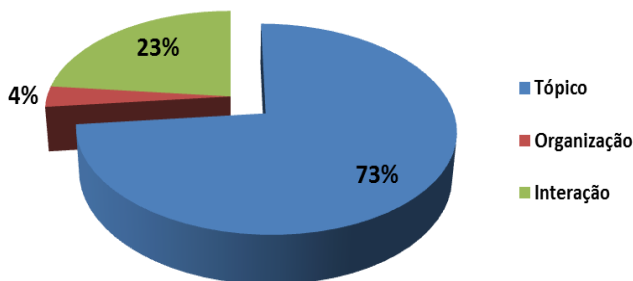


Gráfico 1: Divisão dos quadrigramas de acordo com a classificação de Hyland

As tabelas e gráficos parecem sugerir que os aprendizes estudados

preferiram os quadrigramas voltados para o tópico, ainda que grande parte do percentual a mais desses quadrigramas possa ser explicado porque alguns quadrigramas se superpõem a outros, como por exemplo, *the film tells the, film tells the story, tells the story about*.

7. Conclusão

De acordo com as três categorias propostas por Hyland os quadrigramas usados em sua grande maioria pelos aprendizes são aqueles voltados para o tópico do texto.

A falta de blocos léxico-gramaticais voltados para o texto causa truncamento de ideias e menos suavidade na leitura, o que provavelmente não ocorreria em um texto de falantes nativos, conforme estudo mostrado por Lewis (2009).

Era de se esperar que os aprendizes utilizassem mais quadrigramas voltados para a organização do texto. Por se tratarem de alunos avançados supostamente teriam ciência de que um texto coeso necessita de marcadores de discurso e conectivos. Mas foram encontrados apenas 4% de quadrigramas voltados para a organização. No entanto, os aprendizes utilizaram em maior quantidade os quadrigramas voltados para o tópico (73%). Isso demonstra que os aprendizes ainda não dominam a técnica de se escrever um texto com coerência e coesão, resultando muitas vezes em um texto repetitivo e com ideias truncadas. Parece muitas vezes ao leitor que falta fluência na leitura de muitos exemplares produzidos. Já com o uso de quadrigramas organizacionais o leitor conseguiria fazer uma leitura mais fluida, antecipando informações por vir, conectando ideias e contrastando-as.

O resultado do presente estudo demonstrou que mesmo tendo havido grande ênfase no ensino de marcadores discursivos voltados para a organização do texto os aprendizes não fizeram tanto uso dos mesmos. Foram encontrados mais blocos de palavras voltados para o tópico e para interação autor-leitor. Tal achado pode sugerir que a ênfase dada no ensino dos blocos pesquisados não necessariamente resultou em aprendizagem.

No entanto, deve ser levado em consideração que apenas quadrigramas foram analisados e que futuros estudos envolvendo outros agrupamentos lexicais com quantidades variadas de palavras devem ser feitos assim como estudos envolvendo outros gêneros textuais que não os utili-

zados nesta pesquisa.

REFERÊNCIAS BIBLIOGRÁFICAS

BERBER-SARDINHA, Tony. *Linguística de corpus*. Barueri: Manole, 2004.

HYLAND, Ken. As can be seen: lexical bundles and disciplinary variation. *ESP.*, vol. 27, p. 4-21, 2008.

LEWIS, Margareta. *The idiom principle in L2 english*. Stockholm: Stockholm University, 2009.

McENERY, Tony; WILSON, Andrew. *Corpus Linguistics*. Edinburgh: Edinburgh University Press, 1996.