

**A LINGUÍSTICA DE *CORPUS* E O ENSINO DE INGLÊS – UM
ESTUDO DA PRODUÇÃO ESCRITA DE APRENDIZES
BRASILEIROS**

Gustavo Estef Lino da Silveira (Mestrando, UERJ)

gutolino@oi.com.br

RESUMO: Segundo Bakhtin (2010) o gênero influencia as escolhas léxico-gramaticais. No entanto, estas escolhas significam que um texto foi bem escrito, de forma coesa e coerente? O que caracteriza um texto como um exemplar prototípico de um determinado gênero? Seria o uso de blocos lexicais característicos de um gênero o melhor método a ensinar a escrita de gêneros textuais? O objetivo deste trabalho é analisar os feixes lexicais produzidos por aprendizes brasileiros de inglês como língua estrangeira em um curso livre de idiomas e contrastar a produção linguística dos alunos em textos do gênero resenha de filme com o que se espera que os alunos produzam de acordo com o programa curricular da instituição de ensino. Para isso, lancei mão de preceitos da Linguística de *Corpus* para extração de dados de forma digital e analisar coletivamente o perfil do grupo de alunos. Os resultados demonstraram que os aprendizes não fazem uso dos blocos lexicais ensinados e que, teoricamente, deveriam fazer parte do gênero resenha de filme.

Palavras-chave: Linguística de *Corpus*, blocos lexicais, ensino de escrita, gêneros textuais

1. Introdução

O presente trabalho tem como objetivo principal analisar os feixes lexicais encontrados em um *corpus* de redações de aprendizes brasileiros de inglês como língua estrangeira. Para tal finalidade, farei uso de preceitos da Linguística de *Corpus* e de um programa de computador com o fim de me auxiliar a extrair os agrupamentos lexicais mais frequentes dentro deste *corpus*.

O programa a ser usado para extração dos feixes é o *WordSmith Tools* versão 5.0. Dentre as funções do programa estão a extração de listas de palavras mais frequentes assim como os agrupamentos de palavras mais comuns encontrados em um *corpus*. Assim sendo, o programa faz a extração por meios digitais, cabendo ao pesquisador a tarefa manual de analisar as escolhas dos aprendizes e tentar interpretar o que tais feixes lexicais podem vir a dizer sobre a aprendizagem dos alunos em questão.

A motivação deste trabalho teve como pontapé inicial as aulas de composição escrita ministradas em um curso livre de idiomas em que leciono. Tenho percebido durante minha prática docente que alguns alunos não fazem uso em suas redações de muitos dos blocos lexicais (*lexical chunks*) que são ensinados durante as aulas de produção escrita, os quais, teoricamente, deveriam ser reproduzidos.

Logo, meu objetivo é averiguar a que ponto os alunos aprenderam os blocos lexicais ensinados como parte inerente de um

gênero textual fazendo uso dos mesmos dentro de um tipo de texto. Neste estudo em específico analisarei a resenha de filme. Para o objetivo a que me proponho, farei uso do programa curricular prescrito pelo curso livre de idiomas, mencionado acima, que institui um programa a ser trabalhado pelos professores durante as aulas e farei uma análise contrastiva do que está especificamente no programa a ser ensinado em uma das aulas de produção textual, a que aborda o ensino da produção escrita do gênero resenha de filme, e o que os alunos realmente produziram em seus textos.

Na próxima seção mencionarei os fundamentos teóricos que nortearão este artigo com base na Linguística de Corpus e estudos de Berber Sardinha (2004) e Hyland (2008), assim como estudos anteriores que fizeram estudo da produção de *corpora* de aprendizes de língua estrangeira.

2. Fundamentação Teórica

Primeiramente, defino o conceito de Linguística de *Corpus*, a visão de linguagem a que esta se atrela e descrevo as duas abordagens de análise de um *corpus*. Em seguida, passo a distinguir o conceito de blocos e feixes lexicais.

É mister ressaltar que decidi lançar mão de preceitos da Linguística de *Corpus* pois o objetivo deste trabalho é olhar as redações dos aprendizes de inglês coletivamente, analisando a instituição de ensino e os aprendizes como um todo ao atingir

determinado nível de proficiência na língua. Logo, a análise manual apenas seria indicada caso eu quisesse priorizar a investigação do aluno como indivíduo, não sendo esse o objetivo desta pesquisa. Assim sendo, o uso de ferramentas digitais, permite-me obter um espectro mais amplo com relação ao meu grupo de estudo.

É fundamental definir o que vem a ser a Linguística de *Corpus*, doravante LC.

A Linguística de *Corpus* ocupa-se da coleta e da exploração de *corpora*, ou conjuntos de dados linguísticos textuais coletados criteriosamente, com o propósito de servirem para a pesquisa de uma língua ou variedade linguística. Como tal, dedica-se à exploração da linguagem por meio de evidências empíricas, extraídas por computador. (BERBER SARDINHA, 2004, p.3)

Com isso, vemos que a LC analisa dados de forma empirista, lado a lado com a visão de linguagem proposta por Halliday para quem, a linguagem é um sistema de probabilidades (BERBER SARDINHA, 2004, p.30-1). Por outra vertente, temos o racionalismo de Chomsky que estuda a linguagem por meio da introspecção, vendo a linguagem como um sistema de possibilidades.

Isso significa dizer que Chomsky vê a escrita do aluno como um sistema de possibilidades infinitas dada a criatividade do aluno em produzir sentenças. Porém, nem sempre “linguagem possível”

pode vir a significar “linguagem provável”. Ou seja, o que é gramaticalmente correto pode vir a não estar mais em voga ou não ser usado em determinado contexto. Concomitantemente, a linguística hallidayana, vê a linguagem sob outra perspectiva, a da probabilidade de determinada sentença ser produzida em um certo contexto ou não.

“Dentro da visão racionalista da língua, Chomsky sugeria que um *corpus* não poderia ser uma ferramenta útil para um linguista uma vez que o *corpus* deveria descrever a língua através da competência e não da performance”. (ALMEIDA, 2010, p. 26). Este estudo acredita que acima da competência do aprendiz está a performance do mesmo em atuar em determinadas situações, pois o aprendiz pode possuir certa competência e não vir a performar bem em um contexto específico.

Logo, o presente estudo está atrelado à visão hallidayana da linguagem, pois “a visão da linguagem como sistema probabilístico pressupõe que, embora muitos traços linguísticos sejam possíveis teoricamente, (esses traços) não ocorrem com a mesma frequência”. (BERBER SARDINHA, 2004, p. 30-1)

Este estudo considera que a Linguística de Corpus, ou análise eletrônica de textos segundo alguns autores é uma metodologia em si, pois ela ditará o caminho de análise a ser seguido. “A LC traz consigo algo mais do que simplesmente o instrumental

computacional. Aqui se encaixam as investigações do comportamento do léxico”. (BERBER SARDINHA, 2004, p. 36)

Existem duas abordagens de estudo dentro da LC a “abordagem dirigida pelo *corpus*” e a “baseada no *corpus*”. Tratarei de definir brevemente o que vem a ser ambas as abordagens.

Alguns linguistas de *corpus* como Douglas Biber preferem a terminologia “abordagem baseada em corpus”, ao referirem-se a primeira abordagem. Outros a denominam como *bottom-up theory*, *corpus guided theory* ou *corpus driven theory*: esta abordagem sugere que as hipóteses e teorias devem surgir a partir do *corpus*, sendo este que direcionará o estudo.

Por outro lado, temos a “abordagem baseada no *corpus*” ou *top-down theory*. Nesta abordagem a teoria existe de antemão à análise do *corpus*. O pesquisador parte de um princípio teórico e apenas depois analisa o *corpus* para comprovar ou refutar sua hipótese.

Este estudo está atrelado à primeira abordagem – abordagem dirigida pelo *corpus* – que visa analisar primeiro o *corpus* sem atrelar-se a categorias linguísticas pré-estabelecidas para, então, interpretar os dados obtidos.

Passo agora a tratar dos conceitos e da distinção entre blocos e feixes lexicais segundo alguns autores, visto que estas são as

unidades de trabalho com as quais estudarei o *corpus* de aprendizes compilado para este trabalho.

SHEPHERD (2012, p.15) afirma que a Linguística de *Corpus* enquanto método e a habilidade dos programas concordanciadores de extraírem e organizarem grandes quantidades de dados tornaram evidentes a inseparabilidade do léxico e da gramática, materializando hipóteses como a colocação (a propriedade de as palavras ‘preferirem a companhia’ de determinadas palavras). Muito antes de o computador poder mostrar tal fato, o linguista Firth já dizia que “devemos julgar uma palavra pela companhia que ela tem. (FIRTH, 1951, *apud* SHEPHERD, 2012, p. 15)ⁱ” (tradução nossa)ⁱⁱ.

Se a velha dicotomia léxico e gramática não se aplica mais nos estudos linguísticos de hoje em dia, podemos então dizer que o foco deste estudo é estudar a léxico-gramática do falante de língua estrangeira. Pois a mesma é constituída de blocos lexicais compostos de partes lexicais e gramaticais trazendo uma unidade de sentido única em si.

Muitos estudos foram feitos sobre colocação de palavras e diferentes nomenclaturas foram dadas a estas colocações, tais como “chunks” (Scrivener, 2005) e “feixes lexicais” (Hyland, 2008). SHEPHERD (2006, p. 128) menciona também as seguintes nomenclaturas dadas aos blocos lexicais “unidades de significado”

(Sinclair 1991), “sequências formulaicas” (Wray, 2002), “padrões gramaticais” (Hunston e Francis 2000), “fórmulas” (Wray, 2002), “rotinas”, “lexemas frasais” (Moon, 1998), “molduras colocacionais” (Renouf e Sinclair, 1991), ‘n-gramas’ (Sinclair, 2004), “padrões pré-fabricados” (Granger, 1998), entre outros.

HYLAND (2008) faz uso da denominação “lexical bundles” – aqui traduzida por “feixes lexicais”. Haja vista que este artigo tratará da visão de deste autor sobre os feixes de palavras usaremos a denominação dada por este autor para definir os blocos de palavras que são retirados de um *corpus*. Mas qual seria a definição exata de Hyland para estes feixes lexicais?

Para HYLAND (2008, p. 4-5) feixes seriam sequências de palavras que ocorrem com frequência, palavras que acompanham umas as outras com mais frequência do que o esperado. (tradução nossa)ⁱⁱⁱ

Hyland 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. Hyland (2008) menciona o uso de trigramas, quadrigramas e pentagramas

em seu estudo, dando prioridade aos quadrigramas por carregarem maior significação e serem encontrados com maior frequência.

Decidi focar-me nos quadrigramas porque estes são bem mais comuns do que pentagramas e oferecem uma gama de estruturas e funções mais claras que os trigramas. Feixes lexicais são essencialmente colocações estendidas definidas por sua frequência de ocorrência e extensão. Feixes são essencialmente colocações estendidas definidas por suas frequências de ocorrência e extensão de uso. (HYLAND, 2008: 8) (tradução nossa)^{iv}

Podemos afirmar também que um n-grama seria um feixe lexical que tem um número “n” de palavras que completam seu sentido e que não podem ser analisadas separadamente, mas sim dentro do feixe. Assim sendo, um quadrigrama seria um feixe lexical de quatro palavras formando um único sentido. Tais palavras não se agrupam por acaso, mas andam juntas no discurso como um bloco de significado. Segundo SINCLAIR *apud* HYLAND (2008, p.6) a maioria das palavras não tem um significado independente. Ou seja, fazem sentido dentro de um feixe lexical.

A análise dos feixes lexicais pode auxiliar no reconhecimento do gênero de um texto. Todavia, a ausência de feixes lexicais em um texto pode demonstrar que o autor ainda não tem o domínio necessário tanto da língua que faz uso quanto do gênero discursivo que está usando para escrever seu texto.

Em textos de aprendizes iniciantes de 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.

A presença de feixes lexicais ou *clusters* na escrita de um aprendiz pode revelar muito sobre seu nível de fluência. Pois, segundo HYLAND (2008, p. 5) “a ausência de tais *clusters* pode vir a revelar a ausência de fluência de um novato ou que aquele falante seria um recém-chegado àquela comunidade”. (tradução nossa)

SINCLAIR *apud* HYLAND (2008, p. 6) afirma que são “os componentes de um rico repertório de padrões de multi-palavras que fazem um texto”. (tradução nossa). Ou seja, os blocos lexicais é que trazem significação ao texto e não apenas palavras isoladas. Um texto com um vasto repertório lexical pode revelar a fluência do aprendiz de língua estrangeira ou também sua deficiência na língua alvo.

O uso de certos blocos lexicais, especialmente, no início de um texto, antecipa ao leitor que tipo de texto o mesmo tem em mãos, facilitando assim a leitura e compreensão do mesmo. 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 fazendo com que o mesmo ative seu conhecimento prévio de outros exemplares daquele gênero a fim de auxiliá-lo em sua leitura.

Também é mister ressaltar que um bloco lexical mal usado naquele exemplar pode vir a causar estranheza ao leitor pois ele poderá vir a relacionar aquele bloco a outros exemplares de textos daquele gênero e perceber que aquela escolha linguística não fora apropriada.

Embora a ideia de feixes e blocos lexicais pareça similar é necessário fazer a distinção entre os dois conceitos. Chamarei de feixes lexicais ou *clusters* os blocos de palavras extraídos do *corpus* através de um programa de computador e blocos lexicais ou *chunks* os agrupamentos de palavras, ou seja, as expressões que são ensinadas como um “tijolo lexical” que formam uma expressão de sentido único. Em outras palavras, feixe lexical é uma expressão fixa de n-palavras que ocorre com frequência e pode ser extraída de um *corpus* através de um programa de computador e bloco lexical é uma expressão de n-palavras, também de sentido único, que é ensinada e/ou produzida.

Este estudo alinha-se à visão de Hyland e seu estudo dos quadrigramas, a fim de verificar os feixes lexicais de quatro palavras mais frequentes dentro de resenhas de filmes produzidos por aprendizes brasileiros de inglês.

Na próxima seção, passo a descrever em detalhe a metodologia que fora usada para desenvolver este estudo assim como o grupo de aprendizes de inglês escolhidos.

3. Metodologia

Começo a seção com algumas considerações sobre a metodologia adotada para a análise de meu *corpus* assim como detalho os procedimentos de coleta e escolha do *corpus*.

O *corpus* consiste de 44 redações coletadas entre os anos de 2009 e 2012 em um curso de inglês da cidade do Rio de Janeiro. Os textos coletados fazem parte da avaliação final que os alunos escrevem ao término do semestre. Aos alunos são dadas as opções de escreverem três tipos de texto diferentes durante a avaliação, o quais já foram trabalhados em sala de aula e escritos em casa pelos mesmos como parte de tarefa extra-classe. São eles: relato pessoal, resenha de filme e artigo esportivo.

Ao escreverem os textos durante a avaliação os alunos dispõem de 1 hora e meia, tanto para a parte de conhecimentos linguísticos quanto para a redação, não consultam dicionário e já escreveram um texto semelhante em casa, corrigido pelo professor e devolvido ao aluno com o devido *feedback*.

As redações colhidas foram digitadas manualmente e os erros de ortografia tiveram de ser corrigidos a fim de auxiliar a busca por palavras e *clusters*, caso contrário o estudo seria prejudicado.

Os alunos escolhidos estudam na instituição de ensino há cerca de sete anos. Estão concluindo o curso com o nível que se equivaleria ao intermediário-avançado. São alunos que têm, em

média, 17 anos de idade e estão na fase do pré-vestibular ou início da vida universitária. Ao término do curso os alunos com bom desempenho estão equiparados ao nível B2 do Quadro Comum Europeu (*Common European Framework / Cambridge FCE*).

Ao longo dos anos de estudo de inglês os alunos aprendem a escrever diversos tipos de texto, sempre recebendo o *input* em sala de aula e produzindo-os em casa ou durante as provas. Escrevem textos como cartas (formais e informais), *e-mails*, relatórios, histórias, artigos, resenhas de filme, entre outros.

É importante dizer que os alunos são treinados a produzir os tipos de texto em sala, em uma aula de composição escrita. Eles lêem um texto prototípico daquele exemplar do gênero que é parte do livro do aluno, sugerido pelo programa curricular da instituição, respondem a perguntas de interpretação sobre o mesmo, analisam que tipo de texto se trata e quais características constituem aquele gênero, como por exemplo, registro (formal, neutro ou informal), blocos lexicais mais comuns, marcadores discursivos presentes, parágrafos e suas distribuições e conteúdo dos textos.

Passo a tratar, agora, da abordagem dada ao ensino da produção de resenha de filme em sala de aula. Os alunos são expostos aos movimentos retóricos característicos deste gênero. Os aprendizes lêem um exemplar prototípico do gênero e são guiados pelo professor através de perguntas a fim de identificar as características do gênero. Eles aprendem a identificar os cinco

parágrafos típicos que são inerentes ao gênero e os blocos lexicais que se adequam a cada um destes parágrafos, sendo eles: introdução (informação técnica do filme), roteiro (um pequeno resumo sobre a história do filme), pontos positivos, pontos negativos e recomendações.

As composições a serem analisadas foram compiladas de todas as turmas de último nível de uma mesma filial da instituição ao longo dos anos de 2009 a 2012. A partir daí, o material foi digitalizado e os erros de ortografia foram corrigidos para dar precisão ao estudo. O *corpus* foi compilado e analisado com o auxílio do programa *WordSmith Tools 5.0*.

Embora os aprendizes tenham a possibilidade de escrever três tipos diferentes de textos para a avaliação, foi decidido que o foco da minha análise seria a resenha de filme por ser um dos tipos de texto em que trabalhamos o uso de blocos lexicais para a construção de um exemplar típico do gênero. Durante as aulas os professores apresentam alguns tipos de resenha de filmes e trabalham com os alunos os tipos de blocos de palavras mais frequentes a serem usados em determinados parágrafos do texto.

Diferentemente dos outros tipos de texto, temos na resenha de filmes embasamento para comparar que blocos foram ensinados e quais foram realmente reproduzidos pelos aprendizes em seus textos, pois nesta aula exclusivamente, consta no programa curricular da instituição o ensino de blocos lexicais característicos

do gênero a fim de que os alunos possam utilizá-los na produção de seus textos.

No entanto, admito que a escolha apenas pela resenha de filmes reduz em bastante o tamanho do *corpus* a ser analisado, haja vista que durante esses anos consegui coletar apenas 44 amostras o que contabiliza um *corpus* com aproximadamente 10.000 palavras. Sendo assim, seria este um *corpus* classificado como de tamanho pequeno, com menos de 80 mil palavras, conforme afirma BERBER SARDINHA (2004, p. 26).

Entretanto, defensores de *corpora* de aprendizes como SHEPHERD (2012, p. 16) afirmam que é possível formular interessantes hipóteses com *corpora* pequenos, como é geralmente o caso dos *corpora* de aprendizes.

Como parte do estudo, irei contrastar o *corpus* obtido com o que se espera que o aluno produza através do programa curricular que o professor recebe para preparar sua aula (originariamente chamado de *scheme of work*). Neste programa, existem as indicações das atividades a serem feitas em todas as aulas, assim como os blocos lexicais a serem trabalhados no ensino da produção deste tipo de texto.

Passemos agora a analisar os quadrigramas encontrados no *corpus* assim como compará-los com o programa curricular do professor a ser trabalhado em sala de aula com os alunos.

4. Análise

Esta seção tem como objetivo descrever os dados extraídos no *corpus* e contrasta-los com o programa curricular do professor (*scheme of work*) a ser utilizado nas aulas. Também analisarei os dados que surgiram do *corpus* através de uma abordagem qualitativa de pesquisa.

A seguir reproduzo uma tabela (tabela 1) adaptada do programa curricular da instituição (*scheme of work* – semana 6 aula 1, p. 6 no original) utilizado pelos professores na aula de produção escrita de resenha de filme. A mesma foi resumida tendo sido mantidos apenas os blocos lexicais de quatro e cinco palavras a serem ensinados durante a aula (ou seja, os feixes lexicais de três, seis ou mais palavras foram omitidos por não serem estes objetos deste estudo). Foi mantido o léxico a ser ensinado e incluída a tarefa proposta durante a prova, que nada mais é, que uma tarefa similar à que se pede que os alunos façam como parte do dever de casa.

Tabela 1:

Habilidade	Tipo de texto	Tarefa	Léxico	Blocos lexicais ensinados
Writing: A film review	Resenha de filme	Write a film review about a film that should have won an Oscar but hasn't. (or) Write a review you'll send to a video shop. You may want to mention: the plot, the cast, the special effects (if appropriate), why you liked the film, etc.	visual imagery, atmosphere, acting, plot, performance, ending, cast, scene, slow, special effects, set, soundtrack, screenplay, stunning, first-rate, convincing, predictable, adapted, reveal, suspend, strike, direct, it stars.	is the latest film by,
				the Oscar-winning film ,
				adapted from a novel by,
				stars ... in the title role,
				The film is set in,
				It tells the story of,
				The story takes place,
				As the story unfolds ,
				what strikes you most is,
				create an atmosphere of,
				acting is first-rate,
				I was impressed by,
				not a gripping film,
				the plot is quite
				the ending is predictable,
what the film really lacks,				
characters fail to move,				
To sum up, ... is a ...				
It's well-worth seeing,				
I highly recommend it,				
It doesn't quite deliver the great				

Quadro: programa curricular simplificado da instituição de ensino.

Mesmo sendo o foco deste estudo a análise de quadrigramas, acredito que a extração de determinadas listas de palavras podem nos revelar algumas informações interessantes sobre esse grupo de alunos.

Decidi, então, extrair do *corpus* uma lista com as dez palavras de conteúdo mais frequentes (tabela 2), aqui leia-se por palavras de conteúdo aquelas que carregam em si uma significação, podendo ser substantivos, adjetivos, pronomes ou verbos; diferentemente das palavras gramaticais que tem a função de sustentação do texto, de ligar ideias, ou seja, de unirem-se a outras palavras criando significados maiores.

Tabela 2:

As dez palavras de "conteúdo" mais frequentes no corpus		
N	Palavra	%
1	FILM	1.93
2	I	0.78
3	MOVIE	0.78
4	HE	0.57
5	SHE	0.48
6	LOVE	0.44
7	LIKE	0.35
8	PEOPLE	0.35
9	OSCAR	0.30
10	STORY	0.30

Quadro: lista das dez palavras de “conteúdo” mais frequentes no corpus.

Como podemos ver acima, dentre as dez palavras de conteúdo mais frequentes do corpus, quatro delas encontram-se presente nos blocos lexicais contidos no programa curricular da aula, são elas: *film*, *I*, *Oscar* e *story*.

A seguir, faço uma compilação de todos os blocos lexicais de quatro palavras contidos no programa curricular (*scheme of work*) que foram encontrados no *corpus* de estudo. Além disso, menciono o número de vezes que cada aprendiz usou cada quadrigrama, os aprendizes foram representados por letras do alfabeto de “A” a “N”, conforme pode ser visto na tabela 3.

Tabela 3:

Os quadrigramas encontrados no corpus de acordo com o scheme of work			
#	Quadrigrama	Frequência	Aprendiz
1	tells the story of	7	A, D, G, H, J, M, N
2	To sum up	3	E, I e N
3	what strikes you most	2	I e N
4	the ending is predictable	2	M e N
5	I highly recommend it	2	G e N
6	it was set in	1	K
7	the movie is set in	1	I
8	the film set in...	1	N
9	is set in 2002	1	M
10	the latest film by	1	M
11	as the story unfolds	1	N
12	in the title role	1	N
13	scenery creates an atmosphere	1	L
14	I was impressed by	1	N
15	adapted from a true story	1	I
16	adapted for a movie show	1	C
17	adapted to the cinema	1	L
18	worth seeing and na	1	H
19	worth to be seen	1	F
20	It really worth.	1	B
21	is totally worth seeing	1	A
22	I highly recommend this...	1	C

Quadro: os blocos lexicais de quatro palavras que foram ensinados e encontrados no *corpus*.

Conforme a tabela 3 nos mostra, de um total de 21 quadrigramas ensinados na aula de composição escrita, os alunos utilizaram 22 quadrigramas, ou seja, um a mais. Isto ocorreu devido ao fato de os alunos terem adaptado um bloco lexical original causando variações do quadrigrama ensinado, como por exemplo: “set in” tendo sido usado de 4 formas diferentes “*it was set in*” (número 6), “*the movie is set in*” (número 7), “*the film set in*” (número 8) e “*is set in 2002*” (número 9). Outros exemplos seriam o quadrigrama original “*It’s well-worth seeing*” que fora usado de 4

formas diferentes (exemplos 18, 19, 20 e 21) e “*adapted from a ... by*” usado de 3 formas diferentes (exemplos 15, 16 e 17 no gráfico), aqui omitidos por questões de espaço.

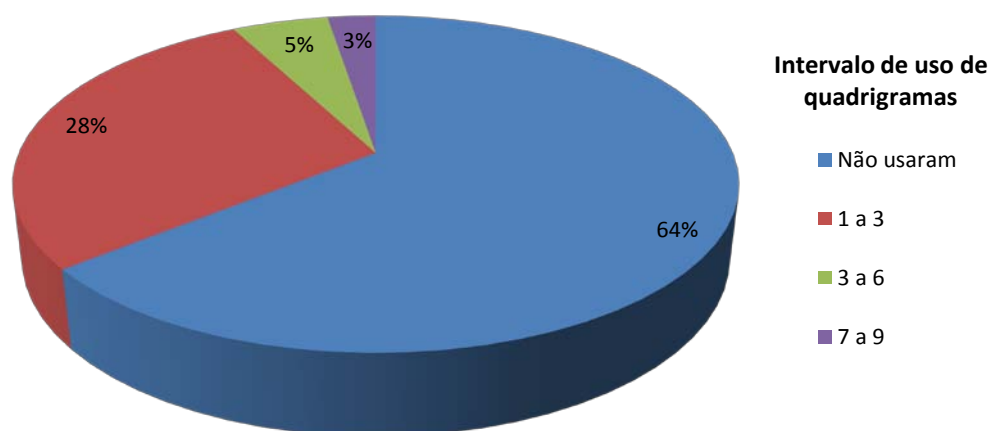
O quadrigrama “*tells the story of*” foi o *chunk* mais usado, com sete ocorrências no total. Este quadrigrama foi encontrado no início dos textos dos aprendizes. Enquanto que a segunda ocorrência mais encontrada foi na verdade, um trigrama, “*to sum up*” aqui mencionado, excepcionalmente, por ter sido usado três vezes e fazer parte de pentagramas como “*to sum up, the film*”, “*to sum up, it is...*”.

A grande maioria dos quadrigramas (17) foi usada apenas uma vez e também, em grande parte, por um único aprendiz.

No gráfico 1, é demonstrada a porcentagem de aprendizes em relação ao número de quadrigramas que cada um utilizou em seus textos.

Gráfico 1:

Uso de quadrigramas



Total de alunos: 44

Quadro: O uso de quadrigramas dividido por aprendiz.

De um total de 44 alunos, apenas 3% dos aprendizes usou 7 ou mais quadrigramas, o que significa, em termos práticos, que apenas 1 aprendiz usou 9 feixes lexicais de quatro palavras. 5% dos aprendizes, um total de 2 alunos, usaram entre 3 a 6 quadrigramas. 28% das redações continham de 1 a 3 quadrigramas.

Dos 44 alunos, apenas 14 deles usaram pelo menos um quadrigrama. Isto significa dizer que 64 % dos alunos formados pela instituição entre os anos de 2009 e 2012 não utilizou sequer um dos blocos lexicais ensinados em aula para escrever resenhas de filme durante a avaliação escrita.

5. Discussão

Nesta seção passo a discutir os dados que surgiram do *corpus* e foram analisados no capítulo anterior.

A análise dos dados, gráficos e tabelas extraídos do *corpus* apontam para o fato de que algumas vezes os aprendizes de inglês da instituição de ensino em questão parecem não ter aprendido os quadrigramas que, teoricamente, deveriam constituir o gênero resenha de filme, fazendo uso de estruturas livres e/ou linguagem aquém do nível que espera-se que os alunos deste nível tivessem.

Em suma, os dados sugerem que os aprendizes fazem pouco uso dos blocos lexicais de quatro palavras supostamente ensinados em sala de aula conforme nos mostra o gráfico 1. Como demonstrado anteriormente, um único aprendiz usou 9 feixes lexicais, enquanto que a maioria dos alunos que os utilizou usou apenas um do total ensinado. Ainda é importante ressaltar que dentro do grupo de alunos estudados, 64 % não usou nenhum quadrigrama sequer. Este dado pode sugerir, que os quadrigramas não foram internalizados pelos alunos ou também que o professor não deu a devida importância ao ensino dos blocos lexicais nas aulas de composição escrita, mesmo estando estes quadrigramas no programa curricular (*scheme of work*) da instituição e distribuídos ao professor como parte obrigatória do material a ser usado na preparação das aulas. Os feixes lexicais mencionados nesta

pesquisa foram marcados no programa curricular como léxico ativo a ser trabalhado em sala.

A análise deste estudo levou em consideração o uso adequado pelos aprendizes das expressões contidas no *scheme of work* da instituição. No entanto, foram encontrados alguns exemplos de erros de colocação “a propriedade de as palavras ‘preferirem a companhia’ de determinadas palavras” (SHEPHERD, 2012, p. 15), ou seja, palavras que foram colocadas com outras equivocadamente ocasionando um mau uso do bloco lexical. Como exemplos encontrados no *corpus*, temos o caso de palavras com preposições errôneas ou até mesmo expressões incomuns na língua.

Tais casos de mau uso da língua tiveram de ser descartados pois não faz parte do objetivo deste estudo a análise de erros, podendo, então, serem utilizados em um futuro estudo sobre erros de aprendizes e mau uso da língua.

No processo de ensino e aprendizagem de uma língua é preciso expor, primeiramente, o aprendiz ao léxico que se deseja ensinar. Com isso, os alunos começam a construir seu *priming* lexical, ou seja, a habilidade de unir o significado da palavra ao seu contexto de uso e as possíveis colocações que envolvem as palavras. Pode se dizer que muitos dos erros encontrados no *corpus* foram erros ocasionados porque os alunos ainda não tem seu léxico mental totalmente formado. Muitos erros de aprendizes de língua estrangeira são ocasionados por tentativas frustradas de uso da

língua em contexto, registro ou com colocações erradas, pois embora os aprendizes tentem fazer uso do léxico, e nesta pesquisa dos blocos lexicais, ainda não conhecem os campos lexicais apropriados ao fazer a escolha certa do léxico (quadrigrama) naquele contexto^v.

6. Considerações finais

Tendo analisado e discutido os dados que emergiram do *corpus* de estudo desta pesquisa, passo neste capítulo a tecer algumas considerações finais sobre este estudo, abordar lacunas deixadas pela pesquisa e sugerir estudos futuros na área de análise de *corpus* eletrônico de língua de aprendiz.

O presente estudo teve como objetivo analisar os quadrigramas mais frequentes usados por aprendizes de inglês no gênero resenha de filme. Com isso, pôde ser visto que os alunos não utilizam os blocos lexicais ensinados, como supostamente deveriam. A pesquisa também emergiu evidências de erros linguísticos na produção escrita destes alunos, que supostamente, seriam alunos de nível intermediário-avançado e não deveriam cometer erros simples de língua. Com isso, existiu uma lacuna a ser mencionada neste estudo que seriam os quadrigramas errôneos produzidos pelos aprendizes e os exemplos de mau uso e sobre uso da língua que não foram amplamente abordados. Em um estudo

futuro gostaria de voltar a este *corpus* sob o viés da análise de erros da língua de aprendiz.

Embora tenha mencionado alguns quadrigramas vagamente ao longo da pesquisa, gostaria de ressaltar a visão a que novos estudos de análise de erros vêm se atrelando. Há, hoje em dia, uma mudança na conceituação de erros de aprendizes conforme menciona GRANGER *apud* SHEPHERD (2012, p. 23):

A produção (escrita e oral) do aprendiz de língua estrangeira não inclui somente o erro, ou seja, o que o aprendiz simplesmente usa erradamente ao experimentar na língua estrangeira. ‘Errôneo’ evidencia também a frequência maior ou menor, bem como a distribuição de recursos léxico-gramaticais e discursivos utilizados pelos aprendizes de determinada língua.

Passo a tratar aqui das limitações do presente trabalho. Em um estudo futuro seria interessante aumentar o tamanho do *corpus* tornando-o de tamanho médio e usar redações provenientes de outras filiais da mesma instituição, sendo assim poderia-se contrastar que feixes lexicais são mais usados por aprendizes de diferentes áreas provenientes da mesma instituição de ensino ou em que filiais os aprendizes produzem mais blocos lexicais errôneos.

Este estudo tem entre suas limitações o fato de não ter analisado as notas obtidas nas redações dos alunos, não sabemos se as redações com mais blocos lexicais utilizados foram as que

obtiveram a maior nota. Logo, pergunto-me, o que caracteriza escrever uma boa resenha de filme? Haja vista que os textos não foram todos analisados integralmente e não temos um parâmetro do que seja escrever uma boa resenha de filme além da resenha proposta pelo livro didático. Seria escrever bem uma resenha, emular o texto sugerido pelo livro didático ou usar muitos blocos lexicais? As redações que não contem blocos lexical algum sugerido pelo programa curricular da instituição obtiveram notas maiores que as que fizeram uso de muitos *chunks*?

Este estudo aponta para o fato de que os alunos concluintes desta instituição de ensino, em sua maioria, não utilizam os clusters de quatro palavras sugeridos em seu programa curricular.

Mas por que isto ocorre? Entre as limitações deste estudo, também pode-se dizer que uma maior análise precisaria ser feita para explicar este motivo. O aluno precisaria de mais tempo para a aquisição do léxico ensinado, já que os mesmos são apresentados em apenas uma aula? Os aprendizes precisariam de mais oportunidades de prática dadas pelo professor? Ou será que a própria forma como os *chunks* são ensinados e sua importância precisam ser revistos?

Em suma, algumas aplicações deste estudo sugerem que tanto professores quanto coordenadores de treinamento possam vir a refletir sobre as práticas didático-pedagógicas em relação ao ensino da produção escrita deste tipo de texto (resenha de filme) e

como trabalhá-lo em sala de aula. Coordenadores de treinamento são os responsáveis em prover as ferramentas necessárias aos professores e treiná-los a ensinar os alunos a aprender melhor. Seria oportuno trabalhar com os professores em *workshops* como abordar convenções de gêneros textuais, como tornar a prática de leitura e escrita na sala de aula de língua estrangeira uma atividade mais motivadora e produtiva para os alunos. Professores precisam ter ciência do que realmente deve ser ensinado e esperado que seus alunos produzam. Tal prática poderia vir a ser alcançada através de maior conscientização do processo de escrita do aprendiz e mais *feedback* dado ao aluno por parte do professor, assim como maior conhecimento dos gêneros textuais e programas curriculares por parte dos docentes e maior treinamento da instituição aos professores no tocante a ministração de aulas de produção escrita. Por fim, os alunos precisam ser ensinados a pôr em prática o uso de determinadas expressões que são características de determinados gêneros textuais e também condizentes com o nível em que o aluno está estudando.

Estes são apenas alguns entre muitos levantamentos que poderiam ser feitos para o fato de os aprendizes não utilizarem os blocos lexicais de quatro palavras conforme esperado. No entanto, uma investigação mais de perto junto a professores e alunos e até mesmo possíveis observações de aulas poderiam auxiliar a um

futuro estudo a elucidar os questionamentos elencados neste estudo.

ABSTRACT: According to Bakhtin (2010) the genre influences the lexical-grammatical choices. However, do these choices mean that the text was well-written, in a cohesive and coherent way? What characterizes a text as a prototypical example of a particular genre? Would it be the use of lexical chunks that are particular of a genre the best way to teach the writing of textual genres? The aim of this article is to analyze the lexical bundles produced by Brazilian learners of English as a foreign language in a private English language school and contrast their linguistic production while writing film reviews with what is contained in the institution's syllabus for the teaching of writing of this genre. Therefore, I have made use of principles of *Corpus* Linguistics for digital data extraction and the collective analysis of the learners' profile. The results have shown that the learners do not make use of the lexical chunks taught and that, theoretically, should be part of the genre film review.

Key-words: *Corpus* linguistics, lexical chunks, teaching of writing, textual genres

REFERÊNCIAS

ALMEIDA, Maria Izabel. Prosa argumentativa em língua inglesa: um estudo contrastivo sobre advérbios em *corpora* digitais. Rio de Janeiro: UERJ, 2010, Dissertação de mestrado.

BAKHTIN, Mikhail. *Estética da criação verbal*. São Paulo, SP: WMF Martins Fontes, 2010.

HALIDAY, Michael A. K. *An introduction to functional grammar*. London, UK: OUP, 2004.

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

SARDINHA, Tony Berber. *Linguística de Corpus*. Barueri, SP: Manole, 2004.

SCRIVENER, JIM. *Learning teaching*. Oxford, UK: Macmillan, 2005.

SHEPHERD, Tania Maria Granja. “Panorama da Linguística de *Corpus*”. In: ---, SARDINHA, Tony Berber, PINTO, Marcia Veirano (org.). *Caminhos da Linguística de Corpus*. Campinas, SP: Mercado de Letras, 2012.

---. ZYNGIER, Sonia. VIANA, Vander. “Feixes lexicais e visões de mundo: um estudo sobre *corpus*.” In: UNIVERSIDADE do Estado do Rio de Janeiro. *Matraga*. Rio de Janeiro, RJ: Caetés, ano 13, n. 19, 2006, p. 125-140.

ⁱ Tradução de “You shall know a word by the company it keeps.”

ⁱⁱ Todas as citações em inglês deste artigo são de responsabilidade do autor.

ⁱⁱⁱ Tradução de “Essentially, these are words which follow each other more frequently than expected by chance.”

^{iv} Tradução de “I decided to focus on 4-word bundles because they are far more common than 5-word strings and offer a clearer range of structures and functions than 3-word bundles. Bundles are essentially extended collocations defined by their frequency of occurrence and breadth of use”.

^v Shepherd (2012) comunicação pessoal.