

# Avaliando a escrita em inglês com a Linguística de *Corpus*

Gustavo Estef Lino da Silveira (UERJ)

Resumo: O objetivo deste estudo é investigar o uso de blocos lexicais de quatro palavras por aprendizes de inglês como língua estrangeira ao escreverem textos do tipo resenha de filme. Avaliação é um conceito muito amplo e subjetivo. Logo, decidi focar o estudo na avaliação da produção escrita como forma de mensurar a aprendizagem dos alunos de blocos lexicais. Para tal, farei uso de preceitos da Linguística de *Corpus* e o auxílio de ferramentas computacionais como listas de palavras e um concordanciador. A escrita pode ser trabalhada e avaliada de diversas formas em sala de aula, mas há de se ressaltar que a Linguística de *Corpus* preenche um vão na análise linguística da produção dos alunos a partir do momento em que esta auxilia o professor a traçar um panorama da interlíngua dos seus aprendizes. Se, antes, o professor se perdia em textos isolados, agora o mesmo consegue ver todo um grupo de alunos mapeados, como uma espécie de exame de raios X. Com isso, o estudo investiga o uso de blocos lexicais de quatro palavras, chamados quadrigramas, usados pelos alunos. Foram coletados 117 textos em seis escolas de idiomas do RJ e a análise dos dados mostrou que apenas um quarto dos alunos estudados utilizou quadrigramas em seus textos. O estudo questiona não apenas o ensino e uso de blocos lexicais como forma de se avaliar a aprendizagem da escrita em inglês como também a própria dinâmica de ensino-aprendizagem da instituição.

## 1. Introdução

O objetivo deste trabalho é discutir a avaliação da aprendizagem de um grupo de alunos de um curso de inglês do Rio de Janeiro. O conceito de avaliação e o de aprendizagem são muito amplos. Além do mais, em uma sala de aula de língua estrangeira trabalham-se quatro habilidades como um todo: a produção oral, a escrita, a compreensão auditiva e a leitura. Logo, seria preciso estreitar a análise e ater-me somente a uma das quatro habilidades trabalhadas, nesse caso, a prática escrita.

E por que a decisão em focar na habilidade de *writing*? Primeiro porque avaliação é um conceito muito amplo e senti a necessidade de direcionar minha investigação a uma habilidade específica. Caso decidisse analisar a avaliação oral, por exemplo, um dos problemas que enfrentaria seria que esta é uma produção momentânea e possui um caráter mais efêmero que a escrita; pois, a não ser que a produção oral do aluno seja gravada e transcrita, os dados se perdem ao longo do tempo.

Não obstante, a avaliação escrita possui um caráter mais duradouro, pois esta permanece registrada para fins de correção posterior, reflexão e *feedback* do professor.

Outro motivo para tal pesquisa é que ao avaliar a escrita dá-se mais tempo para o aluno processar a língua, já que ele utilizará um processo mental diferente para a produção do conteúdo linguístico. Isto pode vir a contribuir para uma escrita com menos erros que a produção oral em si. No entanto, fatores como a interlíngua do aprendiz podem vir a resultar maior ou menor presença de erros no texto.

Avaliar a produção escrita é uma tarefa um tanto quanto trabalhosa e cansativa. Por um lado, temos os professores que têm pilhas de textos a corrigir e ficam ansiosos para terminar a

correção e se verem livres da tarefa. Além disso, a análise se perde em textos isolados, em que o professor se vê marcando erros e dando *feedback* para um aluno específico. Isto pode ocasionar que o professor perca a visão global da turma e que erros seus alunos mais cometem. Por fim, o professor não consegue estabelecer um panorama maior do seu grupo vendo que expressões e vocabulário sua turma como um todo faz uso ou não, ou que erros são mais recorrentes.

Já pelo lado do aluno, o mesmo lê o texto corrigido pelo professor e muitas vezes apenas vê a nota que tirou e põe o texto de volta na gaveta e o esquece. Este não faz uma autocorreção, não reescreve o texto ou reflete sobre possíveis alternativas de acertos. Ou seja, o aluno muitas vezes não volta com o texto para o professor, encerrando-se assim o processo de escrita em língua estrangeira.

Resumindo, pode-se dizer que os textos se perdem, o professor não consegue ter uma visão panorâmica da interlíngua do seu grupo e as redações são vistas isoladamente.

## 2. Fundamentação teórica

### 2.1. Interlíngua

Dentro do processo de aquisição de uma segunda língua, Ellis (2000) afirma que existe um sistema linguístico abstrato com suas próprias características, erros e estilos que cada aprendiz ou grupo de aprendizes carrega em si, chamado interlíngua.

Este sistema mental seria dotado de regras linguísticas abstratas transitórias. Isto quer dizer, que o falante é dotado de uma “gramática mental” e pode modificá-la sempre que desejar adicionando ou apagando regras e reestruturando todo o sistema.

O termo *interlíngua* foi criado pelo linguista americano Larry Selinker na década de 1970. Para ele, a interlíngua pode ser definida como um sistema linguístico que difere tanto da língua materna quanto da língua-alvo.

A interlíngua do aprendiz deve ser vista como um processo vital e natural para se compreender certas construções léxico-gramaticais da produção linguística de aprendizes de língua estrangeira.

Mesmo sendo a interlíngua um processo contínuo e em desenvolvimento, com características de ambas as línguas mãe e estrangeira, Selinker (apud Ellis, 2000) afirma que somente cerca de 5% dos aprendizes irão construir uma gramática mental semelhante à de um falante nativo.

Conhecer a interlíngua de um aprendiz pode vir a auxiliar o professor a compreender as limitações linguísticas de seu aluno e até mesmo intervir na sua prática social de ensino da escrita em uma segunda língua.

## 2.2. Formas de avaliação da produção escrita

A escrita pode ser trabalhada e avaliada de diversas formas: a primeira delas é a escrita como processo. Nesta abordagem, o aluno pré-escreve, toma notas, rascunha, revisa, edita. Aqui todo o processo de escrita é levado em consideração pelo professor e aluno.

A escrita também pode ser vista como um produto. Nesta visão, o texto é analisado como um produto final. Todo o *background* do texto final, anotações não são considerados, apenas o texto redigido em si.

A escrita também pode ser analisada como um fator cognitivo. Quando o professor quer checar se o aluno já possui não apenas as habilidades de ler e escrever, mas também de compreensão e interpretação de diferentes tipos de texto.

Ao fazer uso da escrita como uma prova de aprendizagem o professor quer saber se o aluno aprendeu determinada estrutura léxico-gramatical ensinada em sala de aula. O contexto comunicativo não importa muito, mas sim a estrutura da língua que foi ensinada.

E por último, a escrita como uma mensuração de aquisição da linguagem. O professor a utiliza como avaliação se a linguagem foi adquirida em um contexto mais amplo e se o aluno sabe fazer uso da língua em diferentes situações, espontâneas, livres e sem controle. Os alunos são estimulados a produzir diferentes tipos de textos como um processo natural de escrita, sem focar em itens léxico-gramaticais específicos.

Apesar de todas essas formas mencionadas acima, decidi usar a escrita para avaliar a produção dos alunos sob outro viés, o da Linguística de *Corpus*. Como o uso de ferramentas computacionais pode auxiliar o professor nessa avaliação? A seguinte citação de Leech (1998) ilustra claramente tal possibilidade:

Suponhamos que uma professora X, num país que não tenha inglês como primeira língua, ensine inglês a seus alunos todas as semanas, e de vez em quando lhes peça para escrever composições ou outros trabalhos naquela língua. Ora, ao invés de devolver os trabalhos aos alunos com comentários e um suspiro de alívio, ela guarde as composições em seu computador e construa, gradualmente, semana após semana, uma coletânea maior e mais representativa dos trabalhos de seus alunos. Ajudada por ferramentas computacionais como um concordanciador, ela poderá extrair dados e informações sobre as frequências lexicais desse ‘*corpus*’ e poderá analisar o progresso de seus alunos enquanto grupo, com alguma profundidade. As questões de pesquisa que se abrem são mais significativas quando se compila um *corpus*. (Leech, 1998, p. 14) [minha tradução].

Temos, então, um caminho onde a Linguística de *Corpus* se insere a partir do momento em que possibilita que grandes quantidades de textos sejam analisados como um todo. Com isso, o professor pode fazer um exame de raios X de sua turma vendo-a de modo abrangente. Se antes ele tinha uma visão dos galhos e folhas; agora ele consegue enxergar a árvore como um todo.

A Linguística de *Corpus* nos auxilia a estabelecer padrões em um *corpus* de estudo a partir da análise de todo um grupo de alunos e não de um aluno em si, ou até mesmo de uma instituição de ensino inteira.

## 2.3. Os *chunks* e a léxico-gramática

Programas concordanciadores que auxiliam na extração e organização de grandes quantidades de dados tornaram evidente a inseparabilidade do léxico e da gramática. Portanto, hoje, sendo chamada de léxico-gramática. Este fato materializou hipóteses como a *colocação*, ou seja, a propriedade de algumas palavras ‘preferirem a companhia’ de outras.

Segundo Michael Lewis (1993), a língua consiste da colocação de uma palavra com outra formando um bloco lexical (*chunk*), ou seja, palavras que andam juntas com outras como em um bloco de tijolos formando uma unidade única de sentido.

Logo, se a linguagem é formada de blocos lexicais, o aluno deveria aprender, ao invés de palavras isoladas, blocos de palavras e como unir os blocos lexicais uns aos outros, agrupando-os e reagrupando-os com outros blocos, formando a sua fala.

Outro teórico que falou sobre o conceito de colocações de palavras foi Ken Hyland (2008). Ele 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, 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, quadrigamas e pentagramas em seu estudo, dando prioridade aos quadrigamas por carregarem maior significação e serem encontrados com maior frequência.

Decidi focar-me nos quadrigamas 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 de uso. (Hyland, 2008, p. 8) [minha tradução].

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, tais feixes podem ser mais encontrados.

A presença de feixes ou *chunks* dentro da fala de um aprendiz pode revelar muito sobre seu nível de fluência. 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”. (minha tradução).

#### 2.4. Abordagens de análise da Linguística de *Corpus*

Existem duas formas de entrada ao se analisar um *corpus*. A primeira, a abordagem dirigida pelo *corpus*, não se atrela a nenhuma categoria linguística pré-estabelecida para analisar os dados. As hipóteses e teorias surgem a partir das evidências que emergem do texto.

Já na abordagem baseada no *corpus*, a teoria existe de antemão a análise do *corpus*. O pesquisador parte de um princípio teórico e apenas depois analisa o *corpus* para comprovar ou refutar sua hipótese.

Este estudo atrela-se a primeira abordagem – dirigida pelo *corpus*, onde os dados foram extraídos e analisados puramente sem a influência de alguma teoria linguística.

### 3. Metodologia

O *corpus* de estudo é composto de 117 redações, cada uma com cerca de 250 palavras. Os textos foram produzidos durante uma avaliação escrita de fim de semestre, cronometrada e sem consulta a material de referência. É mister ressaltar, que os alunos já produziram um texto semelhante como parte da tarefa de casa algumas semanas antes. Tal texto fora corrigido pelo professor, que deu *feedback* e o devolveu ao aluno para que o mesmo tenha que reescrevê-lo na avaliação, agora sem consulta.

O *corpus* é formado por textos do tipo resenha de filme, que foi trabalhado em uma aula de produção escrita com os alunos ao longo do semestre. Os alunos tiveram de ler um exemplar de resenha do livro didático, responder algumas perguntas de interpretação sobre o texto e produzir um exemplar semelhante em casa sobre um filme que eles quisessem.

E por que a escolha da resenha de filme como objeto de análise deste estudo? Pois nessa aula os alunos aprendem quais blocos lexicais de três, quatro e cinco palavras são característicos do gênero resenha de filme. Diferentemente das aulas em que aprendem a escrever um relato pessoal ou artigo de revista, onde não há ênfase no ensino de blocos lexicais como característicos de produção textual de um gênero.

Os alunos em questão seriam equivalentes ao nível intermediário alto. De acordo com o quadro comum europeu, comparados ao nível B2, nível FCE de Cambridge.

O *corpus* é de tamanho pequeno, com 25.000 palavras e foi coletado em seis filiais da instituição de ensino do Grande Rio: Bangu, Campo Grande, Méier, Tijuca, Freguesia e Nova Iguaçu. A quantidade de redações provenientes de cada escola encontra-se listada na tabela 1 abaixo:

Filial	Quantidade de textos
Bangu	7
Campo Grande	13
Méier	13
Tijuca	9
Freguesia	26
Nova Iguaçu	49

Tabela 1: quantidade de textos das filiais.

A seguir, na tabela 2, está o programa curricular adaptado da aula de produção escrita do gênero resenha de filme. Este programa é distribuído a todos os professores do nível final do curso com a tarefa a ser trabalhada em sala nesta aula. Fazem parte do programa o léxico ativo e os blocos lexicais a serem ensinados. Como se pode ver, temos aqui blocos lexicais de quatro palavras apenas (à esquerda) um total de 12 quadrigramas e a frequência de uso encontrada no *corpus* de estudo pelos aprendizes, totalizando 44 casos.

Programa curricular do professor	
Quadrigramas	Frequência
...in the title role...	1
the story takes place	1
create an atmosphere of	2
I was impressed by	2
not a gripping film	2
the ending is predictable	2
tells the story of	21
as the story unfolds,	2
what strikes you most	5
the plot is quite	0
What the film lacks	0
I highly recommend it	6
12	44

Tabela 2: programa curricular do professor.

#### 4. Discussão dos dados

A análise do *corpus* revelou que 75% dos alunos estudados não utilizaram nenhum quadrigrama sequer em seus textos, ou seja, apenas 25% dos aprendizes fizeram uso de blocos lexicais de quatro palavras, conforme mostra o gráfico 1 abaixo.

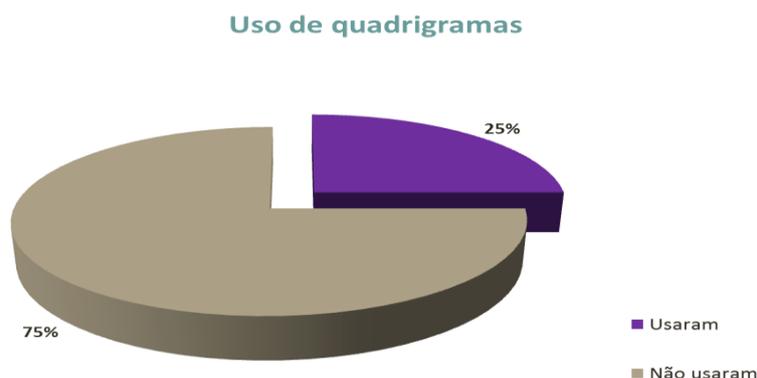


Gráfico 1: uso total de quadrigramas.

Também é possível que se faça uma análise por escola. Com isso, pode-se traçar um perfil de onde a produção é maior ou menor dentro do objetivo da busca. Foram extraídos os feixes lexicais e agrupados de acordo com a escola coletada conforme mostra a tabela 3. Como se pode ver, os alunos de Campo Grande utilizam mais blocos lexicais de quatro palavras, com 46% de alunos os empregando, que os de Nova Iguaçu, com apenas 20% das resenhas de filmes contendo quadrigramas.

Uso de quadrigramas	
Filial	Uso
Campo Grande	46%
Bangu	29%
Freguesia	23%
Méier	23%
Tijuca	22%
Nova Iguaçu	20%

Tabela 3: uso de quadrigramas por filial.

A análise dos dados sugere que os alunos parecem não ter internalizado os blocos lexicais ensinados por completo. 75% dos alunos não utilizaram um quadrigrama sequer. Este dado pode indicar, tanto que os quadrigramas não foram aprendidos pelos alunos como também que o professor não dera a devida importância ao ensino dos mesmos nas aulas de composição escrita. Há de se ressaltar que os quadrigramas em questão estão presentes no programa curricular da instituição e distribuídos ao professor como parte obrigatória do material a ser usado na preparação das aulas.

Foram ensinados 12 quadrigramas e desconsideradas as suas variantes apenas 44 casos foram encontrados. Este número pode ser considerado baixo diante do tamanho do *corpus*. Mas, em relação a esses 25% de alunos que utilizaram os quadrigramas, seria interessante questionar se os mesmos, de fato, internalizaram o conteúdo ensinado ou apenas reproduziram o conteúdo do livro didático a fim de tirar uma boa nota na avaliação. No entanto, tais motivos precisariam ser investigados mais a fundo para se entender o porquê.

É importante salientar, também, que houve a presença de quadrigramas errôneos (com erros de ortografia, concordância e colocação) que foram corrigidos. Os erros de ortografia foram alterados para a forma correta e os de colocação e concordância descartados por se tratarem do âmbito de estudo da Análise de Erros, não sendo esse o foco do estudo.

A pesquisa fez emergirem evidências de erros linguísticos na produção escrita destes alunos, que supostamente, seriam alunos intermediário-avançados e não deveriam cometer erros simples da língua. Tais casos de mau uso do idioma tiveram de ser descartados, pois também estão além das intenções deste trabalho.

## 5. Considerações finais

Para finalizar, é preciso tecer algumas considerações sobre o estudo que se fazem importantes. Em primeiro lugar, pode-se perguntar se o gênero resenha de filme faz parte da prática social desses alunos. Até que ponto os alunos estudados têm contato com resenhas de filmes no seu dia a dia? Quicá em inglês?

Em uma situação mais adiante, poderia se questionar também em que contexto além da sala de aula de inglês os alunos teriam que escrever resenhas de filme? Ou seja, o quão realístico e significativo seria avaliar a aprendizagem do ensino de um idioma através da produção de um tipo de texto apenas? Seria importante que, em próximos estudos, mais exemplares de gêneros diversos fossem coletados a fim de dar maior significância aos dados obtidos.

Também se pôde perceber que os alunos parecem não dominar muito bem o gênero resenha de filme. Inclusive utilizam quadrigramas de cunho avaliativo em parágrafos que teoricamente deveriam descrever o filme, por exemplo.

Outro fato que merece atenção é que o aprendiz ainda está fazendo tentativas com a língua. Por isso a presença de blocos lexicais errôneos ou em movimentos retóricos não apropriados no texto. A interlíngua destes aprendizes precisa ser levada em consideração ao avaliar seus textos, pois esta ainda está em processo de formação.

Há de se questionar se o aprendiz precisaria de mais tempo para aquisição e uso do léxico, haja vista que os alunos aprendem a escrever uma resenha de filme em uma aula, praticam o conteúdo de forma bem simples em sala e levam a tarefa para casa para depois reproduzirem o conteúdo de novo na prova.

Um fato a ser considerado é que a instituição de ensino prioriza a produção oral dando pouca ênfase à escrita. Este fato pode contribuir para uma deficiência da habilidade escrita em prioridade à prática oral.

Por fim, questiono: o que significa escrever uma boa resenha de filme? Usar os blocos lexicais ensinados pela instituição? Tirar uma boa nota ou reproduzir exemplares prototípicos deste gênero? De onde os feixes lexicais são extraídos para ensiná-los como exemplo de uma boa resenha de filme? De exemplares originais do gênero ou de um texto adaptado no livro didático? Será que analisar resenhas de filmes e a produção de blocos lexicais seria a melhor maneira de avaliar a aprendizagem de escrita em língua estrangeira?

A Linguística de *Corpus* permite que o estudo seja aplicado na avaliação do ensino e aprendizagem de outras línguas. Apesar de este estudo ter sido feito em língua inglesa, o mesmo pode ser replicado para a avaliação da aprendizagem de qualquer língua estrangeira ou produção textual em língua materna.

## Referências

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

ELLIS, Rod. *Second language acquisition*. Oxford: Oxford University Press, 2000.

HARMER, Jeremy. *How to teach writing*. Essex, UK: Pearson Longman, 2009.

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

LEECH, G. Teaching and Language Corpora: a convergence. In.: WICHMAN, A.; FLIGELSTONE, S; MCENERY, T.; KNOWLES, G. (Ed.). *Teaching and Language Corpora*. Harlow: Addison Wesley, 1997.

LEWIS, Michael. *The lexical approach*. Hove: Language Teaching Publications, 1993.

SELINKER, Larry. Interlanguage. *IRAL*, p. 209-231, 1972.