

A ESCRITA EM LÍNGUA INGLESA E O ERRO: UM BREVE ESTUDO SOB A LUZ DA LINGUÍSTICA DE CORPUS

Writing in the English language and error: a concise study under the light of Corpus Linguistics

Gustavo Estef Lino da Silveira¹

Resumo: O presente estudo teve como objetivo analisar quais elementos léxico-gramaticais de quatro palavras ou quadrigramas (HYLAND, 2008, p.5) ensinados em aulas de inglês como língua estrangeira seriam produzidos erroneamente pelos aprendizes em composições. Para tal investigação foram coletadas 394 redações de alunos concluintes de um curso livre de inglês da cidade do Rio de Janeiro que estudavam a língua há cerca de 7 anos. O *corpus* foi manipulado usando-se programas de computador para extração de grandes quantidades de dados digitais seguindo preceitos de análise da Linguística de *Corpus* como uma metodologia de pesquisa (SHEPHERD, 2009, p. 153). O resultado do trabalho demonstrou que os aprendizes produziram blocos de palavras errôneos sugerindo uma possível não assimilação do conteúdo ensinado, uma aprendizagem parcial ou até mesmo uma criatividade do aprendiz influenciada por sua língua materna.

Palavras-chave: análise de erros; linguística de *corpus*; linguística aplicada; inglês como língua estrangeira.

Abstract: *The present study aimed at analyzing which 4-word lexicogrammatical elements or 4-grams (HYLAND, 2008, p. 5) that had been taught in English as a foreign language classes would be produced wrongly by the learners in compositions. For this investigation, 394 compositions were collected from graduating learners at an English language school in Rio de Janeiro who had been studying the language for 7 years. The corpus was manipulated by using computer programs for extraction of a large number of digital data following the principles of Corpus Linguistics as a research methodology (SHEPHERD, 2009, p. 153). The result of the work demonstrated that the learners produced erroneous blocks of words suggesting that they may not have acquired the content taught, a partial learning or even creativity of the learner influenced by his mother tongue.*

Key-words: *error analysis; corpus linguistics; applied linguistics; EFL.*

¹ Mestre em Linguística pela UERJ e doutorando pelo mesmo programa e instituição. Professor assistente de língua inglesa do Colégio de Aplicação Instituto Fernando Rodrigues da Silveira (Cap-UERJ). E-mail: gutolino@oi.com.br

1 Introdução

O presente estudo discute, através de dados empíricos, aspectos léxico-gramaticais da produção escrita de aprendizes de inglês como língua estrangeira (L2). Ensinou-se a esses aprendizes a produzir um certo conjunto de blocos léxico-gramaticais de quatro palavras doravante chamados de quadrigramas (Hyland, 2008, p. 5), associados a determinados gêneros textuais. Os quadrigramas estavam presentes no livro didático dos alunos e guia do professor, os mesmos foram retirados de um texto em L2 a fim de realçar o uso dos mesmos por falantes-nativos. A pesquisa analisou se os alunos produziriam, em condições de exame, os quadrigramas conforme foram ensinados, e mais especificamente quantos dos quadrigramas foram produzidos erroneamente.

O trabalho usa como metodologia a extração de quadrigramas a partir de um *corpus* digitalizado de redações de aprendizes de inglês como L2, com o auxílio de um programa de computador. A fim de localizar quadrigramas errôneos, isto é, um número de quadrigramas que não possam ser considerados ‘bem formados’. É de suma importância, portanto, que se revise a tipologia do erro para incorporar-se a uma nova forma de entendimento do mesmo, que é aquela auxiliada por métodos computacionais, a qual lida com frequência e abrangência de padrões linguísticos. Logo, é mister ressaltar que a justificativa do presente estudo dá-se no fato de que: 1) são recentes e poucos os estudos envolvendo *corpora* de aprendizes de inglês como L2 no Brasil e, 2) o auxílio de programas de computador, para tratamento de grandes quantidades de textos, permite ao linguista observar uma miríade de informações como se fosse sob a lente de um “microscópio”, podendo-se, assim, traçar padrões de erros em um grupo de aprendizes de uma L2.

2 Tipos de erros

As tendências mais recentes de métodos comunicativos focam principalmente em desenvolver a habilidade oral do aprendiz, com os erros tendo menor importância; contanto que os mesmos não impeçam a comunicação e sejam produzidos em atividades cuja ênfase seja a fluência, os erros podem vir a ser negligenciados (UR, 1996, p. 246). Se, por um lado, temos no passado uma abordagem considerando o erro um demônio a ser exorcizado, por outro, no presente, o mesmo já é visto como um eterno e inseparável companheiro (LÓPEZ, 2009, p. 3). A verdade é que um estudo de erros oferece grandes vantagens para melhorar as práticas pedagógicas da língua. Isso significa

que os resultados de uma possível análise baseada em erros evidenciam aquelas áreas da língua que os professores precisam dar ênfase, tais como gramática, vocabulário, discurso etc. Ou seja, analisando-se os erros, importantes sugestões para o desenvolvimento de métodos de ensino podem vir a ser criados, e isso envolve todas as áreas do planejamento pedagógico, desde o programa curricular até os materiais de ensino.

No entanto, a avaliação da competência de um aprendiz vai além da análise de erros analisados isoladamente, mas também de outros aspectos tais como o não uso de estruturas difíceis da língua, o que poderia ser um indicador de pouco progresso na aprendizagem de uma língua estrangeira (YIP, 1995, p. 5).

A atitude em relação ao erro sempre variou em um contínuo de total rejeição ao erro a uma certa leniência. No Quadro abaixo (1), identificamos quatro momentos dentro da Linguística Aplicada com autores e suas visões de erros.

Autor	Ano	Tipos de erro
Corder	1973	Omissão, adição, seleção e ordem.
Wilkins	1976	Generalização da língua.
Ferris	2002	Global, local, tratável e não-tratável.
Thornbury	2008	Lexical, gramatical e discursivo. Transferência positiva e negativa.

Quadro 1- Classificação dos tipos de erros, segundo alguns autores.

Em 1973, Corder (*apud* ERDOGAN, 2005, p. 264) classificava os erros em quatro tipos: omissão, adição, seleção e ordem. Os erros de omissão seriam quando o aprendiz omitia uma palavra ou um morfema. Os de adição quando o aprendiz adicionava um vocábulo desnecessário à frase ocasionando um erro. Os erros de seleção seriam quando há uma escolha verbal errônea, um pronome errado ou verbo, ou seja, o aprendiz teria feito uma escolha errada. E por último, o erro de ordem, que poderia ser um erro de ortografia, ordem errada no eixo sintagmático da frase ou ordem de adjetivo e substantivo, por exemplo. Em outras palavras, Corder via o erro em L2 de uma forma isolada ignorando o sentido e o discurso.

De forma semelhante, Wilkins (1976, p. 203) via o erro em L2 dissociado do sentido: assim que os aprendizes tivessem formado regras mentais sobre um fato linguístico, estenderiam essas regras mentais a outros fatos, generalizando e, por vezes, errando. A terminação do passado simples em inglês estendida ao passado irregular é um típico exemplo de generalização.

Ferris (2002, p. 22) faz a distinção entre erros globais e locais e também entre erros tratáveis e não-tratáveis. Os erros globais seriam aqueles que afetam o entendimento de um texto, impedindo a comunicação e a compreensão da mensagem e erros locais seriam pequenos erros que não afetam a compreensão do texto; a mensagem pode ser perfeitamente entendida.

Erros tratáveis são aqueles relacionados a uma estrutura linguística que está relacionada à quebra de uma regra. São chamados de tratáveis porque o aprendiz pode ser direcionado a um livro de gramática e por meio de um conjunto de regras pode resolver o problema. Como exemplo poderíamos citar os tempos e formas verbais, o uso de artigo, erros de ortografia e pontuação etc. Um erro não-tratável pode ser idiossincrático e o aluno precisará utilizar conhecimento adquirido da língua para fazer a autocorreção. Como exemplo poderíamos citar: a escolha errada de palavras, uma estrutura frasal inexistente na língua, seja por adição ou subtração de palavras.

Ferris faz sua contribuição aos estudos sobre o erro, mas a grande questão de seu trabalho é quando a autora fala dos erros tratáveis e não-tratáveis. Entretanto, este trabalho não está interessado nos erros individuais, pois estes não são tratáveis por serem idiossincráticos. O presente estudo não lida com erros individuais, mas sim com os coletivos.

Thornbury (2008, p. 114) diz que existem duas razões principais pelas quais os erros podem ocorrer. A primeira delas seria pela influência da língua materna do aprendiz, também chamada de interferência ou transferência e, dessa forma, o autor fala de dois tipos de transferência: positiva e negativa. Para o autor, ambas sofrem influência da língua materna do aprendiz. No entanto, a transferência positiva não resulta em um erro em si.

O autor afirma que há três tipos de erros que o aprendiz de uma segunda língua comete: erros lexicais, gramaticais e discursivos. Para o autor, a primeira categoria envolve os erros de escolhas de palavras errôneas pelo aprendiz para o significado que se queria transmitir. Como, por exemplo, a escolha da forma errada da palavra (o uso do verbo *think* ao invés do substantivo *thing*) pela semelhança fonológica e gráfica, e erros na ordem natural das palavras, como em português temos primeiro o substantivo e depois o adjetivo e em inglês o inverso (ex.: *girl beautiful*).

O segundo tipo de erro seria o erro gramatical. Esses estão atrelados à forma e tempo do verbo, na estrutura da frase (*the doorbell rangs; we speaked*) e também em casos onde o sujeito da frase é omitido (*it*).

O último tipo de erro listado por Thornbury seria o erro de discurso. Esses seriam erros macro, de construção e coesão textual onde palavras usadas erroneamente, como um conectivo *eventually*, pode gerar no leitor a falsa impressão de que o texto está perto de sua conclusão.

Na visão do autor, o erro pode ser micro ou macro. O macro seria um erro discursivo e o erro micro o erro da estrutura linguística em si.

O fato de eu não ter encontrado uma literatura específica que classifique erros nas décadas de 1980 e 1990 parece convergir com o estudo de Truscott (*apud* FERRIS, 2002, p. 12), onde o autor afirma que houve um declínio na correção, nos estudos e até mesmo na importância dada à análise de erros como um reflexo do surgimento do *process-approach* naquelas décadas. No entanto, o Quadro (1) visto anteriormente sugere que parece ter havido um divisor de águas entre as eras pré e pós *corpus*. Estudos pré-*corpus* envolviam erros nas décadas de 1970; mais à frente parece haver um intervalo onde os erros foram esquecidos e mais adiante no período pós-*corpus* os mesmos voltaram a estar em voga com novas classificações e estudos nos anos 2000, agora com o auxílio de ferramentas de programas de computador e princípios da Linguística de *Corpus*.

Em 2000, Ellis (2000, p.52) notou que havia um tipo de erro que escapava a classificação clássica de erros em gramaticais ou lexicais, imperante até aquela época: erros também podem ter sua origem no uso excessivo de determinadas formas. Ellis notou que alguns aprendizes chineses usavam em excesso expressões de arrependimento ao se desculpar em inglês, e isso espelhava as regras pragmáticas de sua língua materna.

Mais recentemente com o auxílio de métodos computacionais aliados à Linguística Aplicada tais fatores tendem a incluir erros que envolvam a frequência de certas estruturas no texto, que vão desde o não uso da estrutura, até o uso excessivo, conforme proposto por Granger (2003, p.2).

3 O *corpus* de estudo

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 (aqui denominadas M e T), Zona Oeste (aqui denominadas F, B e CG) e Grande Rio (que recebeu as iniciais NI)².

² Os bairros das filiais foram omitidos para resguardar a confidencialidade dos dados. Com isso, siglas foram atribuídas para a identificação das mesmas.

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* de análise 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.

A metodologia de pesquisa utilizada é a Linguística de *Corpus* em si como um conjunto de programas de computador que permite manipular grandes quantidades de dados e através deles analisá-los, como defende Shepherd (2009, p. 150).

Os textos coletados fazem parte da avaliação final que os alunos escrevem ao término do último semestre na instituição. Aos alunos são dadas as opções de escreverem textos que já foram trabalhados em sala de aula e escritos em casa pelos mesmos como parte de tarefa extraclasse.

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

Os alunos estudados tiveram três aulas de composição escrita ao longo do semestre. O tema é trabalhado seguindo orientações do livro didático e com grande ênfase na prática oral do tópico a ser escrito em casa como tarefa extraclasse. Em termos práticos, isto significa dizer que os alunos tiveram pouca ou nenhuma oportunidade de praticar a habilidade escrita em sala, apenas ensaiando oralmente o conteúdo a ser reproduzido em casa ou recebendo o *input* acerca do conteúdo.

Os sujeitos desta pesquisa, à época da coleta, tinham estudado na instituição de ensino por cerca de sete anos exceto alguma possível exceção de aluno de outra instituição que tenha sido submetido a um teste de nivelamento. Logo, foram selecionados todos os alunos concluintes de sete filiais diferentes. São alunos que têm, em média, 17 anos de idade e estão na fase do pré-vestibular ou início da vida universitária; estão concluindo o curso com o nível que equivaleria ao intermediário-avançado, o que os equipararia ao nível do Quadro Comum Europeu B2 / Cambridge FCE, conforme a Figura 1 a seguir e as descrições para cada nível de proficiência.

CEFR	Levels	English
Common European Framework		Cambridge certificates
C2	Very advanced	CPE Certificate of Proficiency in English
C1	Advanced	CAE Certificate in Advanced English BEC Higher
B2	Upper intermediate	FCE First Certificate in English BEC Vantage
B1	Intermediate	PET Preliminary English Test BEC Preliminary
A2	Elementary	KET Key English Test
A1	Beginner	

Figura 1: Quadro Comum Europeu para referência de línguas³.

De acordo com o Quadro Comum Europeu para referência de línguas (CEF), os alunos que estão no nível B2 são classificados como “intermediário avançado” e aqueles que atingem esse nível são capazes de entender as ideias principais de textos complexos em tópicos concretos e abstratos, incluindo discussões técnicas em sua área de especialidade. Também podem interagir com um grau de fluência e espontaneidade, o que torna a interação regular com falantes-nativos bem possível sem dificuldades de ambas as partes. E, por último, podem produzir textos claros e detalhados em uma ampla variedade de assuntos e explicando pontos de vista em um determinado tópico dando as vantagens e desvantagens de várias opções (COUNCIL OF EUROPE, 2011).

A instituição de ensino afirma sobre as habilidades adquiridas pelos alunos ao longo do curso avançado, conforme consta no certificado de conclusão de curso, que os alunos tiveram 100 horas-aula em um período de 1 ano. No certificado também consta a seguinte afirmação acerca da habilidade linguística do aprendiz: “O aluno tem excelente domínio da língua inglesa falada e escrita. Interage com sucesso em situações sociais e acadêmicas, mesmo as mais complexas, como negociações, apresentações e debates. Tem habilidade para redigir textos de naturezas diversas tais como cartas formais e informais, ensaios e artigos”.

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 e mais adiante durante as

³Fonte da Figura 1: <http://www.cambridgeenglish.org/images/126130-cefr-diagram.pdf>

provas. É mister ressaltar, que este estudo não fez o acompanhamento das aulas dadas pelos diversos professores ao longo dos anos, podendo cada professor ter trabalhado o conteúdo direcionado pelo programa curricular da instituição a seu próprio modo. No entanto, baseio meus comentários nos princípios metodológicos da instituição e o programa curricular que cada professor tem em mãos para preparar suas aulas.

A grosso modo, em uma aula de composição escrita, os alunos leem um texto típico do exemplar do gênero a ser trabalhado que é parte integrante de uma lição do livro didático do aluno, sugerido pelo programa curricular da instituição. Logo em seguida, respondem a perguntas de interpretação sobre o mesmo, oralmente, analisam que tipo de texto se trata e quais características constituem aquele gênero, como, por exemplo, registro (formal, neutro ou informal), blocos léxico-gramaticais mais comuns, marcadores discursivos presentes, parágrafos e suas distribuições além do conteúdo dos textos. Embora algumas vezes haja espaço para um momento de prática individual escrita do conteúdo, grande ênfase é dada ao trabalho oral em pares ou em pequenos grupos de alunos. Resumindo, dessa prática didática foram extraídos os quadrigramas ensinados.

4 Análise e discussão dos dados

Antes de avançar na análise dos quadrigramas errôneos, seria importante destacar que todos os quadrigramas considerados fora do padrão foram analisados. Não importou a esse estudo a origem do erro, seja ele por influência da língua materna do aprendiz, da sua interlíngua ou mesmo um erro de tentativa); os quadrigramas foram analisados e classificados sob dois vieses. O primeiro deles sob uma perspectiva pedagógica, voltada para a prática didática e de fácil assimilação pelo professor de inglês e o aprendiz; e o segundo, sob uma classificação voltada para os princípios da Linguística de *Corpus*, conforme mencionarei mais adiante.

Analisando-se todo o *corpus*, foram encontrados 25 quadrigramas errôneos. O critério de busca foi que o quadrigrama tivesse sido produzido por, no mínimo, três aprendizes diferentes sendo da mesma filial ou não. Esse número pode ser considerado baixo, se levado em conta o total de quadrigramas produzidos por três aprendizes extraídos do *corpus*, um número bruto de 537. Ou seja, do total de quadrigramas extraídos, 4,7% são errôneos. Esse é um número aparentemente pequeno, mas que chama a atenção a partir do fato de que o mesmo erro ocorre não apenas entre aprendizes mas também filiais diferentes.

Passo a seguir a analisar a lista de quadrigramas errôneos, como pode-se ver na Tabela 1 abaixo. O quadrigrama errôneo mais frequente foi “have a lot of”, que foi produzido 12 vezes, por 9 alunos diferentes de NI e 3 de M.

Os quadrigramas de maior dispersão nas filiais foram os de número 3 e 5, tendo sido produzidos por cinco filiais diferentes. Ou seja, não importa a origem do erro, o fato é que ele ultrapassa os muros da filial e parece ser compartilhado por grupos de aprendizes.

Núm	Quadrigrama	Frequência	Textos	Filiais Individuais					
				B	CG	F	M	NI	F
1	Have a lot of ⁴	12	12	0	0	0	3	9	0
2	Tells the story about	8	8	0	1	0	2	2	3
3	Had a lot of	7	7	1	1	0	2	2	1
4	In this day I	4	4	0	0	0	0	1	3
5	Tells a story about	5	5	1	1	1	1	0	1
6	the history of a	5	5	1	0	0	0	1	3
7	A Oscar but should	4	4	0	1	1	0	2	0
8	Have won a Oscar	4	4	0	2	0	0	1	1
9	Not win a Oscar	4	4	0	1	1	0	2	0
10	Should have won a	4	4	0	2	0	0	1	1
11	Win a Oscar but	4	4	0	1	1	0	2	0
12	To enter in the	5	3	0	0	1	0	4	0
13	The final of the	4	3	0	1	0	1	2	0
14	A film that not	3	3	0	0	1	0	2	0
15	Cast is composed by	3	3	1	2	0	0	0	0
16	Film talk about a	3	3	0	0	0	1	2	0
17	Film tells the history	3	3	1	0	0	0	1	1
18	Film that not win	3	3	0	0	1	0	2	0

⁴ O quadrigrama em questão fora expandido e analisada toda a linha de concordância para, então, concluir-se que trata-se de um uso errôneo.

19	In love for a	3	3	0	0	0	1	2	0
20	In the other hand	3	3	0	0	1	0	0	2
21	In this moment I	3	3	1	0	0	0	0	2
22	The film talk about	3	3	0	0	0	1	2	0
23	The history and the	3	3	1	0	0	0	2	0
24	To insert an idea	3	3	0	0	1	0	0	2
25	In my point of (view)	4	4	0	0	0	0	3	1

Tabela 1: Lista de quadrigramas errôneos e filiais onde ocorreram.

A classificação para erros sugerida por Granger (2006), demasiado extensa (54 categorias), não se aplicaria aos itens desta pesquisa, visto que a autora trabalha com itens gramaticais individuais, isto é, as categorias são anotadas em cada item lexical individualmente.

A classificação dos erros nas Tabelas 2 a 8 a seguir é uma versão simplificada da categorização de Granger e dá conta dos blocos errôneos encontrados no *corpus* de estudo. Ela foi feita segundo 7 tipos diferentes de erros encontrados nos quadrigramas, sendo eles: preposição, artigo, 3ª pessoa do singular/escolha lexical errônea, falta de verbo auxiliar, escolha lexical errônea, erro de pronome demonstrativo e “there” (existencial). A seguir, explico cada tipo de erro mencionado acima e exemplifico com casos retirados do *corpus*.

O primeiro deles é o que classifico como erro de preposição. Muitos alunos parecem ter dificuldade em assimilar qual preposição ou verbo vem junto a uma determinada palavra. Mas na verdade, pode se dizer que tais erros seriam de colocação. Por colocação entende-se o grau de uma palavra de atrair uma outra tão fortemente que ambas parecem ser inseparáveis. Esses blocos léxico-gramaticais sempre atrairão certas preposições não havendo outra possibilidade de alteração, adição ou exclusão. Conforme mostra a Tabela 2 adiante, foram encontrados sete erros classificados como erros de preposição. O primeiro quadrigrama errôneo foi encontrado em oito textos diferentes, já a forma correta ocorreu 38 vezes, um número quase cinco vezes maior. Isso quer dizer que a cada seis tentativas de uso desses quadrigramas, uma é errônea.

No erro de número 6, o aprendiz utilizou a preposição errônea “in” ao invés de “on/from”, no entanto, esta também poderia ter sido omitida não sendo necessária na frase, podendo haver,

assim, uma exclusão da preposição. O exemplo de número 7 mostra um caso de sobreuso de preposição, sendo esta desnecessária na frase.

Núm	Erro	Quantid. de quadrigramas errôneos	Forma correta
1	Tells the story about	8	Tells the story of
2	Tells a story about	5	Tells a story of
3	In love for a	3	In love with a
4	In the other hand	3	On the other hand
5	In my point of (view)	4	From my point of
6	In this day I	4	(On) this day I
7	To enter in the (university)	5	To enter university

Tabela 2: Erros de proposição.

Embora os erros elencados a seguir possam ser considerados básicos para o nível estudado - o nível avançado -- parece haver um descuido quanto ao uso do artigo indefinido “a/an” por alguns alunos, já que foram encontrados cinco quadrigramas errôneos com tal ocorrência (Tabela 3).

Núm	Erro	Quant. de quadrigramas errôneos	Forma correta
1	A Oscar but should	4	An Oscar but should
2	Have won a Oscar	4	Have won an Oscar
3	Not win a Oscar	4	Not win an Oscar
4	Should have won a	4	Should have won an
5	Win a Oscar but	4	Win an Oscar but

Tabela 3: Erros de artigo indefinido.

À primeira vista, o erro descrito a seguir parece estar no verbo da terceira pessoa do singular que tem a terminação “s” faltando. No entanto, também houve uma escolha lexical errônea porque o livro não “fala sobre” e sim, em inglês, “é sobre” algo. Logo, a forma correta do pentagrama seria: “the film is about a”.

Núm	Erro	Quant. de quadrigramas errôneos	Forma correta
1	Film talk about a	3	Film is about a
2	The film talk about a	3	The film is about a

Tabela 4: Terceira pessoa do singular/escolha lexical errônea.

O erro listado adiante, como muitos outros, também foi extraído duplamente por ter sua origem em um pentagrama. No entanto, decidi também manter a divisão em dois quadrigramas a fim de manter o critério inicial de análise defendido por este estudo. Como pode se ver na Tabela 5, o erro é devido a uma má construção da forma negativa. Três aprendizes não internalizaram o uso do verbo auxiliar “did” necessário para expressar o passado na negativa.

Núm	Erro	Quant. de quadrigramas errôneos	Forma correta
1	A film that not	3	A film that did not win
2	Film that not win	3	Film that did not win

Tabela 5: Verbo auxiliar faltando.

Com seis ocorrências, as escolhas lexicais errôneas não variam muito. Em primeiro lugar pode-se observar o mau uso do substantivo “history” ao invés de “story”; logo depois, a possível influência da língua mãe ocasionando um erro na escolha do léxico “final” em detrimento de “end” e também um outro possível erro de interlíngua, no qual esse aluno demonstra não ter aprendido a expressão idiomática “put an idea into somebody’s head” fazendo uma tentativa errônea com base em sua língua materna “to insert an idea” (Tabela 6).

Núm	Erro	Quant. de quadrigramas errôneos	Forma correta
1	The history of a	5	The story of a
2	The final of the (game)	4	The end of the
3	Film tells the history	3	Film tells the story
4	The history and the	3	The story and the
5	To insert an idea	3	Put an idea into somebody’s head

6	Cast is composed by	3	Cast includes
---	---------------------	---	---------------

Tabela 6: Escolha lexical errônea.

Os quadrigramas que passo a analisar abaixo demonstram que houve um erro na escolha do pronome demonstrativo para referência anafórica e também da preposição que o acompanha, pois os aprendizes estavam se referindo a um momento anterior que havia mudado sua vida (Tabela 7).

Núm	Erro	Quant. de quadrigramas errôneos	Forma correta
1	In this moment I	3	At that moment I
2	At this moment I	3	At that moment I
3	In that moment I	3	At that moment I
4	On that moment I	2	At that moment I

Tabela 7: Erro de pronome demonstrativo e/ou preposição.

Em último lugar, um erro muito frequente entre os aprendizes brasileiros de inglês como língua estrangeira é o “there” com sentido existencial; muitas vezes confundido com o verbo “ter” em língua portuguesa, os alunos fazem mau uso do mesmo ao evitarem a expressão “there is/are” pelo verbo “have”, tanto no presente, passado, singular e plural (Tabela 8)⁵. O primeiro quadrigrama errôneo foi encontrado 12 vezes e o segundo teve 7 ocorrências.

Núm	Erro	Quant. de quadrigramas errôneos	Forma correta
1	Have a lot of	12	There is a lot of
2	Had a lot of	7	There were lots of

Tabela 8: Erro de “There” existencial.

Após listar cada uma das categorias que lancei mão para classificar os quadrigramas errôneos encontrados no *corpus*, decidi, então, agrupá-los em um gráfico que permitisse melhor visualização

⁵ Para classificar o quadrigrama como errôneo, foi preciso expandir a linha de concordância e analisar todo o seu entorno, para, então, considerá-lo como erro.

dos erros. De acordo com o Gráfico 1 a seguir, o erro mais frequente é o de preposição com 27% dos casos. Em segundo lugar, com 19% foram as escolhas lexicais errôneas. O erro na escolha do artigo e o uso do “there” com sentido existencial tiveram 17% e 16% de ocorrências, respectivamente. Em 5º lugar dentre os erros mais frequentes foram os quadrigramas contendo erros tanto de pronome demonstrativo quanto de preposição com 10% dos casos. E por último, com 6% e 7% estão os erros de 3ª pessoa do singular e omissão do verbo auxiliar.

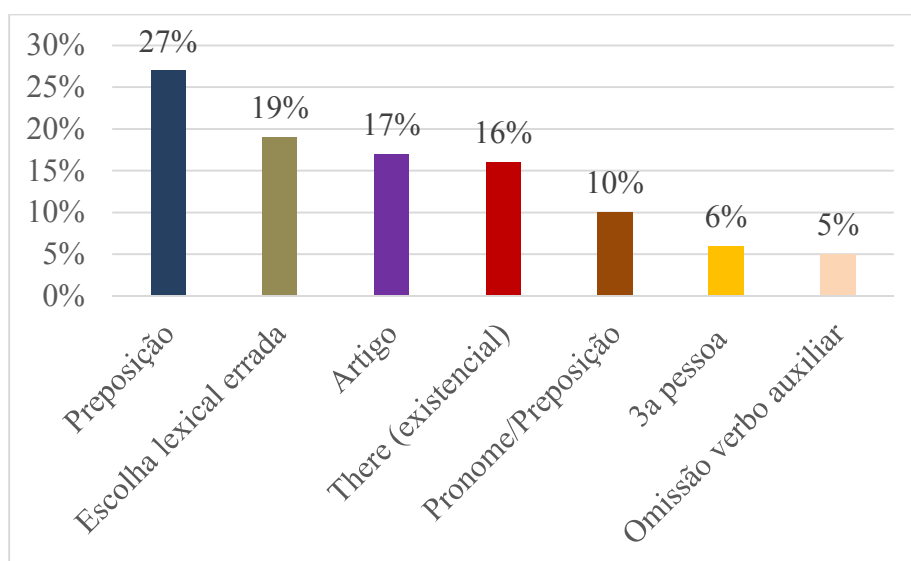


Gráfico 1: Tipos de erros encontrados no *corpus*.

O mapeamento dos erros acima pode ser muito útil para um trabalho de correção e tratamento de erros, auxiliando o professor a lidar com alunos que possuem erros fossilizados e ampliando seu entendimento da interlíngua do aprendiz brasileiro. Com isso, os erros encontrados, embora relativamente poucos em relação ao tamanho do *corpus* podem contribuir para uma melhor prática didática do professor e uma melhor produção do aprendiz de L2.

No entanto, para o linguista de *corpus*, a classificação descrita anteriormente parece fazer pouco sentido, sendo necessária uma análise que dê conta dos princípios que a Linguística de *Corpus* aborda como colocação, coligação, entre outros.

Antes de prosseguir com a análise dos dados, é importante que se abra um parêntese para que tais conceitos sejam elucidados. Por ‘colocação’, entende-se como sendo combinações frequentes de unidades lexicais fixadas na norma. Tais unidades, destacam-se por sua alta frequência

de uso, como por exemplo a expressão, *redondamente enganado*, o que motiva que esse tipo de construção, seja às vezes identificado como sendo uma unidade semi-fixa ou semi-idiomática.

Já por ‘coligação’, entende-se como a propriedade de as palavras preferirem determinadas posições sintáticas, na frase ou mesmo no texto (BERBER-SARDINHA e SHEPHERD, 2008).

Logo, o que pode se depreender das definições mencionadas anteriormente, é que a colocação está para o léxico assim como a coligação está para a gramática. Os colocados parecem, com isso, trazer sentido ao texto enquanto que os coligados teriam a função de unir ideias, de estruturar o texto.

Sendo assim, decidi organizar os erros previamente mencionados segundo tais princípios e sugiro a classificação dos mesmos conforme erros de colocação e coligação apenas, demonstrados nas Tabelas 9 e 10, a seguir.

Erros de colocação	Quant.	Forma correta
Film talk about a	3	Film is about a
The film talk about	3	The film is about
Cast is composed by	3	Cast includes
The history of a	5	The story of a
The final of the (game)	4	The end of the
Film tells the history	3	Film tells the story
The history and the	3	The story and the
To insert an idea	3	Put an idea into somebody’s head

Tabela 9: Erros de colocação.

Conforme demonstrado na Tabela 9 anteriormente, pôde ser extraído do *corpus* um total de 27 ocorrências de erros de colocação contra 83 erros de coligação, a seguir na Tabela 10; sendo assim temos um total de 110 quadrigramas errôneos dentro das duas tabelas. No entanto, como a Linguística de *Corpus* não enfatiza números brutos, mas sim percentuais, o que deve ser analisado é a proporção de 24% da primeira tabela contra 75% da segunda. Tal fato, sugere que os aprendizes estudados, cometem três vezes mais erros acerca da estrutura da frase, da gramática de um quadrigrama, os chamados erros de coligação, e menos erros lexicais, os colocacionais.

Erros de coligação	Quant.	Forma correta
Tells the story about	8	Tells the story of
In love for a	3	In love with a

In the other hand	3	On the other hand
In my point of (view)	4	From my point of
In this day I	4	(On) this day I
In this moment I	3	At that moment I
At this moment I	3	At that moment I
In that moment I	3	At that moment I
On that moment I	2	At that moment I
To enter in the (university)	5	To enter university
A Oscar but should	4	An Oscar but should
Should have won a	4	Should have won na
Win a Oscar but	4	Win an Oscar but
Have won a oscar	4	Have won an Oscar
Not win a oscar	4	Not win an Oscar
A film that not	3	A film that did not
Film that not win	3	Film that did not
Have a lot of	12	There was a lot of
Had a lot of	7	There were lots of

Tabela 10: Erros de coligação.

No Gráfico a seguir (2), pode-se ver a divisão de erros segundo conceitos provenientes da Linguística de *Corpus*. Há uma predominância de erros de coligação com 76% dos casos sobre os erros de colocação com 24%. Isso significa que os aprendizes produzem mais erros referentes à estrutura gramatical dos blocos e em menor quantidade erros de colocação.

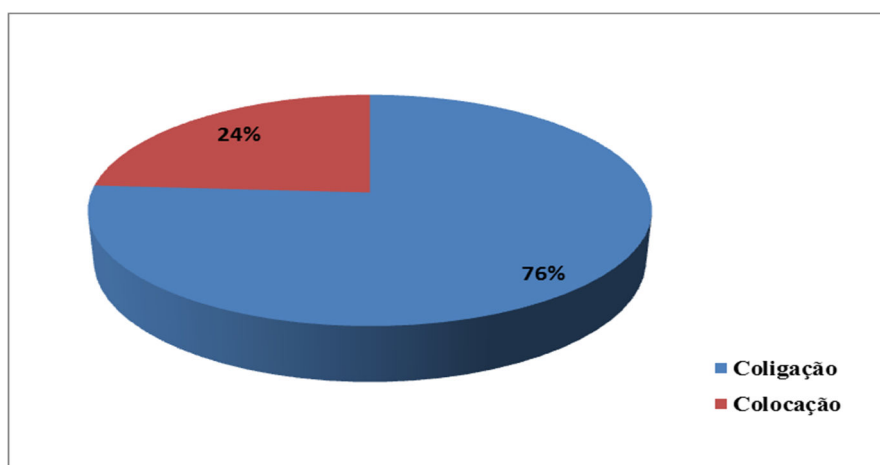


Gráfico 2: Erros de coligação e colocação.

Se comparada a proporção de produção de quadrigamas corretos com errôneos, pode-se fazer uma analogia que quanto maior a quantidade de textos e palavras maior a probabilidade de se encontrarem erros, pois parece haver uma relação entre quantidade de textos e número de

quadrigramas certos e errados. Os alunos erram muitas vezes porque estão tentando escrever certo. Dessa forma, seus erros são ocasionados por tentativas, isto é, quanto maior a tentativa maior a probabilidade do erro ou acerto. No caso deste estudo, pode-se dizer que os alunos acertam bastante, e que os mesmos escrevem razoavelmente bem, se analisados pelo viés da frequência de erros encontrados. Todavia, ainda há espaço para se analisar outras questões que dão conta do nível da escrita deste grupo de aprendizes como a densidade lexical dos textos, o nível de sofisticação léxico-gramatical das redações, entre outros filtros que poderiam ser aplicados em outras análises.

5 Conclusão

A conclusão desta pesquisa deve levar em consideração que, avaliar o nível de proficiência de um aprendiz de L2 apenas pela produção escrita de quadrigramas pode não ser inteiramente justo, já que aprender uma língua envolve a aquisição de quatro habilidades (escrever, falar, compreensão leitora e auditiva). Os resultados mostraram que os aprendizes produziram duas classes de quadrigramas errôneos: erros de colocação e de coligação. Tendo sido, o segundo tipo, três vezes mais frequente que o primeiro. Tal fato, pode vir a sugerir que maior ênfase à gramática necessite ser dada durante as aulas já que em termos lexicais os aprendizes erraram menos.

Em contrapartida, parece ser possível afirmar que muitos aprendizes ainda estavam cometendo erros de tentativa ao errarem as palavras que se justapunham umas a outras formando o quadrigrama. O estudo parece sugerir ainda que os 7 anos de estudo de uma L2 na instituição analisada não foram suficientes para a aquisição de uma segunda língua, entre a população estudada, isso, se levado em consideração os erros encontrados, que foram principalmente erros gramaticais (coligacionais). Ainda assim, alguns aprendizes foram criativos ao transformarem os quadrigramas ensinados e utilizando-os seguindo a influência da língua materna ou, com outras palavras, que não foram pertencentes ao quadrigrama original.

Durante a análise dos dados, foi encontrado sobreuso de alguns dos quadrigramas ensinados e pouco uso de outros pelos aprendizes. Isso pode vir a sugerir uma certa preferência idiossincrática dos aprendizes por certas expressões em detrimento de outras. Os resultados também demonstram que houve muitos erros de transferência entre a L1 e a L2, pois a língua materna exerceu grande influência na aprendizagem de uma segunda língua nos aprendizes em questão.

Embora estudos recentes sobre análise de erros sejam escassos tanto em Linguística Aplicada quanto em Linguística de *Corpus*, um novo viés pode ser dado a essa área através do uso de princípios da LC e de programas de computador a fim de auxiliar a investigação da interlíngua do aprendiz brasileiro de L2. Pesquisas futuras analisando diferentes tipos de n-gramas (maiores ou menores) e outras habilidades senão a escrita podem vir a contribuir a mapear de forma mais abrangente a pequena contribuição que o presente estudo pretendeu dar à análise de erros auxiliada por computador.

REFERÊNCIAS

- BERBER-SARDINHA, T. *Linguística de corpus*. Barueri, SP: Manole, 2004.
- _____; SHEPHERD, T. M. G. An online system for error identification in Brazilian learner English. In: TEACHING AND LANGUAGE CORPORA CONFERENCE, 8., 2008, Lisboa. *TaLC8*. Lisboa: Associação de estudos e de investigação do ISLA, 2008. p. 257-263.
- COUNCIL OF EUROPE. Common European framework of reference for languages: learning, teaching, assessment. [S.l.]: Council of Europe, 2011.
- ELLIS, R. *Second language acquisition*. Oxford: Oxford University Press, 2000.
- ERDOGAN, V. Contribution of error analysis to foreign language teaching. *Mersin University Journal of Faculty Education*. Turquia, v. 1, ed. 2, p. 261-270, dez. 2005.
- FERRIS, D. R. *Treatment of error in second language student writing*. Michigan: The University of Michigan Press, 2002.
- GRANGER, S. Error-tagged learner corpora and CALL: a promising synergy. *CALICO Journal*. San Marcos, TX, v. 20, n. 3, p. 1-16, 2003.
- _____. Using error-tagged learner corpora to create English-specific CEF descriptors. In: CONFERENCE OF EALTA, 3., 2006. [S.l.: s.n., 2006].
- HYLAND, K. As can be seen: lexical bundles and disciplinary variation. *ESP*, v. 27, p. 4-21, 2008.
- LEWIS, M. *The idiom principle in L2 english*. Stockholm: Stockholm University, 2009.
- LÓPEZ, W. C. Error analysis in a learner corpus: what are the learners' strategies? In: A survey on corpus-based research. Panorama de investigaciones basadas en *corpus*. AELINCO 2009. Disponível em: < <http://www.um.es/lacell/aelinco/contenido/index> >. Acesso em: 05 jul. 2013.

SHEPHERD, Tania. O estatuto da linguística de *corpus*: metodologia ou área da linguística? In: Matraca. Rio de Janeiro, v. 16, n. 24, jan./jun. 2009.

THORNBURY, S. *How to teach grammar*. Essex, UK: Pearson, 2008.

UR, P. *A course in language teaching*. Cambridge: Cambridge University Press, 1996.

WILKINS, D. A. *Second language learning and teaching*. London: Edward Arnolds, 1976.

YIP, V. *Interlanguage and learnability / from Chinese to English*. USA: John Benjamins Publishing, 1995.