UNIVERSIDADE DE SÃO PAULO
FACULDADE DE FILOSOFIA, LETRAS E CIÊNCIAS HUMANAS
DEPARTAMENTO DE LETRAS MODERNAS

DANILO SUZUKI MURAKAMI

Colocações verbais em um *corpus* de aprendizes brasileiros de inglês

Versão corrigida

São Paulo
2016

DANILO SUZUKI MURAKAMI

Colocações verbais em um *corpus* de aprendizes brasileiros de inglês

Dissertação apresentada à Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo para obtenção do título de mestre em Letras

Área de concentração:
Estudos Linguísticos e Literários em Inglês

Orientadora: Profa. Dra. Stella Esther Ortweiler Tagnin

De acordo.

Versão corrigida

São Paulo
2016

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Catalogação na Publicação
Serviço de Biblioteca e Documentação
Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo

M972c

Murakami, Danilo Suzuki
Colocações verbais em um corpus de aprendizes brasileiros de inglês / Danilo Suzuki Murakami ; orientadora Stella Esther Ortweiler Tagnin. - São Paulo, 2015.
110 f.

Dissertação (Mestrado)- Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo. Departamento de Letras Modernas. Área de concentração: Estudos Linguísticos e Literários em Inglês.

1. Linguístics de Corpus. 2. Ensino de língua estrangeira. 3. Língua inglesa. I. Tagnin, Stella Esther Ortweiler, orient. II. Título.

Nome: MURAKAMI, Danilo Suzuki.

Título: Colocações verbais em um *corpus* de aprendizes brasileiros de inglês

Dissertação apresentada à Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo para obtenção do título de mestre em Letras

Aprovado em:

Banca Examinadora

Prof. Dr. _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. Dr. _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. Dr. _____ Instituição: _____

Julgamento: _____ Assinatura: _____

AGRADECIMENTOS

Aos meus pais e à toda minha família pelo apoio durante minha jornada acadêmica.

À Stella Tagnin por ter me orientado com muita paciência desde a graduação e pelas "trocas de receitas".

À Daniele Pechuti Kowalewski por me inspirar a seguir estudando, aprendendo e criticando.

Aos colegas do Projeto CoMET (Alva, Andréa, Carmen, Elisa, Helmara, Joacyr, Katarinna, Luciana C., Luciana G., Malila, Mara, Marlene, Nilson, Rozane, Sabrina, Sandra e Sérgio) pela troca de ideias e momentos de descontração.

À Universidade de Simon Fraser (Canadá) por ter me recebido no seu Departamento de Linguística. Em especial, à Profa. Dra. Maite Taboada e ao Prof. Dr. Murray Munro. Ao governo canadense pelo financiamento de minha visita.

À Universidade de Cambridge (Reino Unido), especialmente à Profa. Dra. Dora Alexopoulou e ao Dr. Jeroen Geertzen, pela oportunidade de aprofundar a pesquisa. Ao programa *Language Education* (USP) que possibilitou tal oportunidade.

À Education First pelo apoio à pesquisa, principalmente à Dra. Rachel Baker.

Aos falantes nativos pela sua valiosa contribuição para a avaliação dos dados.

Aos meus amigos André Sanchez, Andressa Soreira, Bruna Polachini, Bruno Nogueira, Felipe Ramirez, Filipe Ferreira, Joyce Kitamura, Mariana Agoíno e Moara Cappellari.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de mestrado e pelo apoio financeiro para a realização desta pesquisa.

RESUMO

MURAKAMI, D. S. **Colocações verbais em um *corpus* de aprendizes brasileiros de inglês**. 2015. 110 f. Dissertação (Mestrado) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2015.

Muitas pesquisas reconhecem a importância das colocações para o aprendizado da língua inglesa. Contudo, poucos estudos investigaram o tema na escrita de aprendizes brasileiros de inglês. Esta pesquisa examina o papel das colocações verbais em um *subcorpus* do *EF-Cambridge Open Language Database* (EFCAMDAT) composto por redações de aprendizes brasileiros de inglês de nível avançado. A abordagem metodológica adotada neste estudo é baseada em técnicas da Linguística de *Corpus*. Para essa investigação, foi elaborada uma classificação semiautomática de todos os verbos com o auxílio de um programa de anotação de *corpora*. Em geral, os resultados mostram que praticamente uma em cada cinco combinações entre um verbo e um substantivo é uma colocação. No entanto, os aprendizes não empregam colocações verbais com sucesso mesmo sendo de nível avançado de aprendizado. As colocações verbais apresentaram desvios em 25% dos casos. O principal tipo de inadequação é o uso de um verbo inapropriado causado pela influência do português. Um pequeno número de estruturas sintáticas também pode ser responsável por desvios colocacionais. Mais pesquisas sobre esse tópico precisam ser conduzidas para a total compreensão dos fatores que determinam a taxa de sucesso. Os achados devem contribuir para a área de aprendizagem de inglês por brasileiros.

Palavras-chave: Linguística de *Corpus*, ensino de língua estrangeira, língua inglesa, *corpus* de aprendizes, colocação.

ABSTRACT

MURAKAMI, D. S. **Verbal collocations in a corpus of Brazilian learners of English**. 2015. 110f. Dissertação (Mestrado) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2015.

There is a growing body of literature that recognizes the importance of collocations in English language learning. However, few studies have investigated the use of collocations in the writing of Brazilian learners of English. This research examines the role of verbal collocations in a subcorpus of the EF-Cambridge Open Language Database (EFCAMDAT). The subcorpus comprises writings by advanced learners of English from Brazil. The methodological approach taken in this study is based on Corpus Linguistics. For this investigation, a semi-automatic classification of all verbs was applied with the aid of a computer program for annotation of text. Overall, the results indicate that nearly one out of every five combinations between a verb and a noun is a collocation and that learners are not completely successful in the use of verbal collocations despite their advanced level of learning. The use of verbal collocations was found to be deviant in 25% of the cases. The main type of inadequacy was the use of an inappropriate verb caused by the influence of Portuguese. A small number of syntactic patterns may also have been responsible for collocational deviations. More research on this topic needs to be undertaken before full comprehension of the factors that determine success rate. The findings should make a contribution to the field of English learning by Brazilians.

Keywords: Corpus Linguistics, language teaching, English studies, learner corpus, collocation.

LISTA DE FIGURAS

Figura 1 – Concordância de *make* em nosso *corpus* de estudo.....20
Figura 2 – Captura de tela da ferramenta *splitter* do *WordSmith Tools 6.0*.....34
Figura 3 – Captura de tela de parte de modelo de redação no *Englishtown* 37
Figura 4 – Captura de tela da ferramenta *AutoCode* do *UAM CorpusTool*: anotação automática de verbos.....42
Figura 5 – Captura de tela da ferramenta *AutoCode* do *UAM CorpusTool*: verbos anotados atuomaticamente 43
Figura 6 – Captura de tela da ferramenta *Project* do *UAM CorpusTool*: segmentos anotados automaticamente 43
Figura 7 - Captura de tela da ferramenta *Project* do *UAM CorpusTool*: segmentos corrigidos manualmente.....44
Figura 8 – Esquema de anotação adotado na pesquisa 47
Figura 9 – Esquema de anotação: tipo de combinação 48
Figura 10 – Esquema de anotação: não-combinação 49
Figura 11 – Esquema de anotação: colocação 51
Figura 12 – Esquema de anotação: estrutura sintáticas 55
Figura 13 – Captura de tela do *UAM CorpusTool* 60
Figura 14 – Captura de tela dos verbos com *thought* como objeto na ferramenta *Word Sketch* (SkELL).....93

LISTA DE GRÁFICOS

Gráfico 1 – Frequência de tipos de combinação 61
Gráfico 2 – Frequência de três combinações: combinação livre, colocação e expressão idiomática 63
Gráfico 3 – Distribuição de colocações por aprendiz 65
Gráfico 4 – Frequência de colocações adequadas e inadequadas 68
Gráfico 5 - Distribuição de colocações por uso 69
Gráfico 6 - Distribuição de colocações por aceitabilidade 70

LISTA DE QUADROS

Quadro 1 – Modelo de cabeçalho dos textos no EFCAMDAT 31
Quadro 2 – Exemplo de texto do EFCAMDAT com cabeçalho preenchido.....32
Quadro 3 – Dois exemplos de textos do EFCAMDAT com apenas alguns caracteres 35
Quadro 4 – Dois exemplos de textos do EFCAMDAT com pequenas mensagens.....36
Quadro 5 – Exemplo de texto do EFCAMDAT com cópia de modelo de redação.....37
Quadro 6 – Classificação de aceitabilidade 70

LISTA DE TABELAS

Tabela 1 – Comparação entre os níveis do *Englishtown* e do QECR 29
Tabela 2 – Tópicos das redações do nível 16 (C2) do *Englishtown* 30
Tabela 3 – Descrição quantitativa do *corpus* de estudo 38
Tabela 4 – Quantidade de não-combinações 62
Tabela 5 – Quantidade de colocações e textos por aprendiz 64
Tabela 6 – Quantidade de colocações por verbo lematizado 66
Tabela 7 – Quantidade de colocações por estrutura sintática 67
Tabela 8 – Distribuição de desvios combinatórios 71
Tabela 9 – Distribuição de desvio combinatório I: verbo inadequado 72
Tabela 10 – Desvio I: verbo inadequado (i. influência do português) 73
Tabela 11 – Desvio I: verbo inadequado (ii. verbo similar) 75
Tabela 12 - Desvio I: verbo inadequado (iii. *do* e *make*) 77
Tabela 13 – Desvio I: verbo inadequado (iv. verbo diverso) 79
Tabela 14 – Desvio I: verbo inadequado (v. *take*) 80
Tabela 15 – Desvio II: substantivo inadequado 81
Tabela 16 – Desvio III: verbo e substantivo adequados com estrutura inadequada 82
Tabela 17 – Desvio IV: verbo e substantivo inadequados 83
Tabela 18 – Desvio V: verbo como substantivo 84
Tabela 19 – Distribuição de estruturas sintáticas e aceitabilidade 86
Tabela 20 – Colocações empregadas e desvios em relação a estruturas sintáticas 87

LISTA DE ABREVIATURAS E SIGLAS

COCA	*Corpus of Contemporary American English*
CoMAprend	*Corpus* Multilíngue de Aprendizes
CoMET	*Corpus* Multilíngue para Ensino e Tradução
DDL	*Data-driven Learning*
EFCAMDAT	Education First-Cambridge Open Language Database
HTML	*HyperText Markup Language*
ICLE	*International Corpus of Learner English*
LC	Linguística de *Corpus*
QECR	Quadro Europeu Comum de Referência para as Línguas
SkELL	*Sketch Engine for Language Learning*
TXT	*Text File*

SUMÁRIO

INTRODUÇÃO **15**
CAPÍTULO 1. FUNDAMENTAÇÃO TEÓRICA **18**
1.1 Linguística de *corpus* 18
1.2 *Corpora* no ensino 19
1.3 *Corpus* de aprendizes 22
1.4 Colocação 23
1.4.1 Colocação verbal 24
CAPÍTULO 2. METODOLOGIA **26**
2.1 Início da pesquisa 26
2.2 O *corpus* de estudo: EFCAMDAT 27
2.2.1 *Englishtown* 28
2.2.2 Cabeçalho 30
2.2.3 Seleção de textos 32
2.3 UAM CorpusTool **39**
2.3.1 Esquema de anotação 40
2.3.1.1 Tipo de combinação 48
2.3.1.1.1 Não-combinação 48
2.3.1.1.2 Combinação livre 50
2.3.1.1.3 Colocação 51
2.3.1.1.4. Expressão idiomática 54
2.3.1.2 Estruturas sintáticas 54
2.3.2 Anotação 59
CAPÍTULO 3. RESULTADOS **61**
3.1 Resultados quantitativos 61
3.2 Análise das colocações 71
3.2.1 Desvio I: verbo inadequado 71
3.2.2 Desvio II: substantivo inadequado 80
3.2.3 Desvio III: verbo e substantivo adequados com estrutura inadequada 81
3.2.4 Desvio IV: verbo e substantivo inadequados 83
3.2.5 Desvio V: verbo como substantivo 84
3.2.6 Desvio VI: substantivo como verbo 84
3.2.7 Desvio VII: inclassificável 85

3.3 Colocações e estruturas sintáticas 85
CAPÍTULO 4. CONSIDERAÇÕES FINAIS 90
REFERÊNCIAS 96
APÊNDICE A - LISTA DE TEXTOS E CÓDIGOS DO *SUBCORPUS* 99
APÊNDICE B – INSTRUÇÕES PARA OS FALANTES NATIVOS REALIZAREM O TESTE DE ACEITABILIDADE 100
APÊNDICE C – PRÉ-TESTE DE ACEITABILIDADE COM FALANTES NATIVOS DE INGLÊS 101
APÊNDICE D – COLOCAÇÕES CLASSIFICADAS COMO ACEITÁVEIS .102

INTRODUÇÃO

A experiência de trabalho com o ensino de inglês para brasileiros motivou o início da pesquisa no sentido de tentar compreender as dificuldades desse público na tarefa de aprender uma outra língua. Devido à formação acadêmica em estudos linguísticos, o caminho mais natural do trabalho foi investigar a produção de brasileiros em língua inglesa por meio da própria língua produzida por esses aprendizes. Não foi o caso, portanto, de investigar outros fatores, como, por exemplo, aspectos sociais, que poderiam influenciar o tema em questão.

Não poderíamos deixar de mencionar que a escolha do objeto de pesquisa foi se formando, e desse modo sendo influenciado, pelo convívio com o grupo de pesquisa *Corpus* Multilíngue para Ensino e Tradução (CoMET) durante a graduação. Foi nesse momento que começamos a ter um conhecimento formal sobre colocações, e percebemos que se tratava de algo que nossa experiência didática identificava como um problema recorrente. Foi fundamental ter tido a oportunidade de ter contato com o projeto de um dicionário bilíngue (português-inglês) de colocações verbais (TAGNIN, 2005) para definir que também examinaríamos o mesmo tipo de colocação em nossa pesquisa. Estudos sobre o tema confirmam que a produção de colocações é um problema para os aprendizes de língua estrangeira (NESSELHAUF, 2005, p. 3).

Atualmente, enfatiza-se a importância de dominar outras línguas a fim de alcançar sucesso acadêmico e profissional (HYLAND, 2009), principalmente quando essa língua é o inglês, uma vez que é a mais usada na difusão de conhecimento de diversas áreas do saber assim como no mundo dos negócios. Novos universitários chegam ao ensino superior, e percebem a importância de conhecer o inglês para terem melhor desempenho acadêmico. Diversos profissionais buscam aperfeiçoar o domínio da língua inglesa, dado que sabem que tal conhecimento, por exemplo, será avaliado em entrevistas quando buscam um novo emprego.

Aprender uma outra língua, no entanto, não é uma tarefa fácil, ainda mais quando se trata de uma língua pouco semelhante à língua materna do aprendiz, como é o caso do inglês em relação ao português. Um indivíduo brasileiro pode estudar durante anos a fim de alcançar fluência em língua inglesa, mas mesmo com anos de

estudo, ainda poderá detectar empecilhos que dificultam a comunicação. Tais dificuldades podem abranger múltiplos níveis linguísticos, desde a pronúncia – níveis fonético e fonológico – a questões que envolvem o conhecimento da situação de fala – nível pragmático. Ao observarmos as dificuldades de um aprendiz brasileiro de inglês, o que muitas vezes notamos é a influência do português na tentativa de se expressar em língua inglesa.

Um aprendiz pode falhar, por exemplo, na tarefa de perceber que em *ask a question* (em português, 'fazer uma pergunta'), há uma relação de convenção que une o verbo *ask* e o substantivo *question*. Consequentemente, um aprendiz brasileiro pode vir a produzir *do/make a question* na tentativa de se referir a 'fazer uma pergunta'. Esse é um tipo de desvio provocado pela falta de conhecimento do fenômeno das colocações verbais

Como não seria possível investigar todos os níveis de convencionalidade linguística num trabalho de mestrado, analisamos somente colocações verbais, pois essas colocações não têm sido um foco frequente de atenção em análises de inglês de aprendizes (NESSELHAUF, 2005, p. 3).

Diante do cenário que descrevemos, o objetivo de nossa pesquisa é investigar colocações verbais na produção escrita de aprendizes brasileiros avançados de inglês. Com a análise, nosso propósito também é identificar os tipos de desvios de convencionalidade linguística que ocorrem envolvendo combinações de um verbo com um substantivo. Visto que Nesselhauf (2005) conduziu com sucesso um trabalho de doutorado com objetivo semelhante, no qual examina colocações em textos escritos por aprendizes avançados alemães de inglês, consideramos coerente basear nossa pesquisa no estudo da autora por meio da adaptação de seu aparato metodológico para o domínio de um trabalho de mestrado com menor quantidade de dados. Adiante, no capítulo dedicado à metodologia, evidenciamos como utilizamos as categorias de classificação de dados propostas pela pesquisadora alemã.

Dado que combinações entre verbo e substantivos são frequentes e estão entre as mais difíceis para o aprendiz de língua (HOWARTH, 1996), nosso trabalho busca responder às seguintes perguntas de pesquisa.

- Qual a incidência de colocações verbais na produção escrita de aprendizes brasileiros de inglês no nível avançado?

- Quais tipos de desvios ocorrem?
- Quais as causas que contribuem para a ocorrência de um desvio?

A fim de alcançar nosso objetivo, investigamos redações escritas por brasileiros aprendizes de inglês no nível avançado. Os textos examinados são provenientes do *Education First-Cambridge Open Language Database* (EFCAMDAT). Com o auxílio de um programa computacional de anotação de *corpora*, criamos um esquema de anotação de segmentos de textos e identificamos não apenas as colocações verbais, mas também outros tipos de combinações entre verbos e substantivos, assim como os desvios cometidos pelos aprendizes. Essa ferramenta eletrônica também permitiu estabelecer a distribuição dos dados identificados.

CAPÍTULO 1. FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, apresentamos as bases teóricas que orientam nossa pesquisa. De modo a partir de uma visão teórica geral em direção a conceitos específicos relevantes para nosso trabalho, apresentamos inicialmente a Linguística de *Corpus* (1.1), passando por *corpora* no ensino (1.2) e *corpora* de aprendizes (1.3) para então fazermos considerações sobre colocações (1.4).

1.1 Linguística de *corpus*

Se considerarmos que a Linguística é a área do saber voltada para o estudo científico da linguagem humana, podemos afirmar que a Linguística de *Corpus* (LC) produz conhecimento sobre a língua por meio da investigação de *corpora* analisados com métodos de pesquisa específicos. Em trabalhos no campo da linguagem, a utilização da palavra *corpus* é comum para se referir a um conjunto de dados examinados pelo pesquisador. No entanto, não consideramos qualquer análise de dados linguísticos como um trabalho na área de LC. Assim, é necessário definir sua abrangência.

Em primeiro lugar, diferentemente de disciplinas linguísticas como a morfologia e a sintaxe, por exemplo, que se ocupam da investigação, respectivamente, de morfemas e construções sintáticas, isto é, de objetos de estudo específicos, a LC não atua em um nível linguístico específico. Esse é um dos motivos que faz com que não haja consenso sobre o estatuto da LC como disciplina ou metodologia[1]. McEnery e Hardie (2012, p. 1), por exemplo, definem a LC como uma área que se concentra em um conjunto de procedimentos, ou métodos, para o estudo da língua, mas reconhecem que há aqueles que não consideram a área como uma simples metodologia.

Independentemente da posição adotada, a LC faz oposição a teorias centrais no campo da Linguística. Por estar inserida em uma abordagem essencialmente empírica dos estudos da linguagem, a LC vai na contramão da perspectiva racionalista

[1] Para uma discussão sobre a questão cf. BERBER SARDINHA, 2004, p. 35.

proposta pelo Gerativismo (BERBER SARDINHA, 2004, p. 30). Na perspectiva empírica, a língua é vista como um sistema probabilístico em detrimento de um sistema "possibilístico". Em outras palavras, a probabilidade de ocorrência de certo dado linguístico é mais significativa do que sua possibilidade já que nem tudo que é possível dizer-se na língua é, de fato, usual. Esse é um aspecto fundamental que opõe metodologias racionalistas e empíricas nos estudos da linguagem. Do lado racionalista, temos a introspecção para confirmar a possibilidade de um enunciado linguístico ocorrer, ou seja, o próprio falante julga a adequação de um fenômeno. Já do lado empírico, é necessário que os dados corroborem eventuais afirmações do pesquisador.

Os dois tipos de abordagem são válidos e podem produzir resultados variados, porém a tarefa é mais árdua quando um pesquisador investiga uma língua que não domina, isto é, na ocasião em que, como falante de outra língua, não tem acesso à introspecção de um falante nativo. Nesse aspecto, o recurso a um *corpus* pode ser mais benéfico do que a simples introspecção. Segundo Tagnin (2013, p. 29), um *corpus* é "uma coletânea de textos, necessariamente em formato eletrônico, compilados e organizados segundo critérios ditados pelo objetivo de pesquisa a que se destina".

As ferramentas que permitem a investigação metodológica da LC podem ser utilizadas em diferentes áreas de estudos, inclusive no ensino de línguas. A seguir, discutimos a convergência entre *corpora* e ensino.

1.2 *Corpora* no ensino

De maneira geral, *corpora* fornecem evidências sobre o uso real da língua. A história da LC nos mostra que o uso de *corpora* eletrônicos foram um marco na confecção de dicionários. Hoje em dia, na compilação de grandes dicionários, especialmente nos de língua inglesa, é impossível pensar em outro método lexicográfico senão com o auxílio de um *corpus*.

Apesar da influência que o uso de *corpora* teve nos dicionários, inclusive naqueles voltados para aprendizes de língua, segundo O'Keeffe, McCarthy e Carter (2007, p. 21), a língua presente em materiais de ensino ainda é, muitas vezes, baseada

em percepções de como usamos a língua em vez de evidências de uso de fato. Ainda de acordo com os autores, além de fornecerem uma base empírica para que averiguemos nossas intuições sobre a linguagem, *corpora* trazem à luz características da língua que podem ter enganado nossa intuição.

Como um aprendiz de língua estrangeira não tem acesso à introspeção linguística de um falante nativo, um *corpus* pode ser um meio de entrada para a utilização da língua. Em outras palavras, mesmo desprovido da competência linguística de falante nativo, um aprendiz pode pesquisar como a língua é realmente usada.

O exame de características linguísticas em um *corpus* evidencia o que é de praxe na utilização da língua. Esse "aspecto que caracteriza a forma peculiar de expressão numa dada língua" (TAGNIN, 2013, p. 153) pode ser chamado de convencionalidade linguística. Aquilo que é convencional na língua torna-se regular e pode ser verificado com o auxílio de uma das ferramentas mais utilizadas na LC, o concordanciador.

O concordanciador produz concordâncias a partir de uma ou mais palavras de busca. A concordância é uma forma de visualizar os dados linguísticos de modo que uma ou mais palavras estejam centralizadas em uma lista com todos os contextos de ocorrência no *corpus*. Na Figura 1, temos um exemplo de concordância da palavra *make* no *corpus* que analisamos em nossa pesquisa.

1	epresentative discussed about making a robot that	make	a few facial expressions. They are working
2	have taken from it. They probably will	make	a few phone calls to make sure
3	. Interesting no? If you prefer, you can	make	a list. It's a helpful and
4	going, the police collected enough data to	make	a publicity campaign, offering a $300 reward
5	appear, I close my eyes again and	make	a wish, feeling all nature power around
6	acting suspiciously. Now we are about to	make	an appointment with the detective in charge
7	is difficult to get the robots to	make	even small movements in a natural way,
8	been working on getting the robots to	make	human expressions, such as sadness and happin
9	analyse data and it is supposed to	make	hypotheses based on its analysis. The last
10	future intentions to turn the focus to	make	it able to visit, assist patients victimized
11	air. The unceasingly singing of the birds	make	me forget all noisy car honks, and
12	hand, the service robot is going to	make	part of our lifes in a very
13	, the proof is undeniable. Thousands of people	make	reports yearly of strange lights performing i
14	the end of tomorrow. They will also	make	several interviews with the detective in char
15	people each year. And what does it	make	so perfect and a memorable symbol? The
16	; we can watch tv while the robots	make	some so rude house taske. </text>
17	wage that shows my income! Just to	make	sure I'll take last year' tax
18	will make a few phone calls to	make	sure that the information I have provided
19	certainly will scrutinize my financial history to	make	sure that I do not have any
20	, and also but not least important, to	make	the best cello bows out of it.
21	be the harshest and most expensive to	make	. The first thing you must know about
22	, it is painstaking and time-consuming to	make	the smallest movement to work. Future develop
23	think men now have enough intelligence to	make	their choices, with the internet event is
24	systems. Scientists are making huge effort to	make	things "believavable" to us .I think we

Figura 1 – Concordância de *make* em nosso *corpus* de estudo

Uma concordância permite explorar de maneira objetiva os sentidos linguísticos em um conjunto de dados, pois facilita a visualização de diversos contextos de uso. Além disso, como ferramentas que geram concordâncias são capazes de ordenar os dados que ocorrem à esquerda e à direita do elemento de busca, também é mais simples descobrir padrões linguísticos repetidos. Na Figura 1, por exemplo, vemos a recorrência de *make sure* nas linhas 17, 18 e 19.

Um dos grandes problemas na área de ensino de línguas é que frequentemente são ensinadas formas de se expressar em uma língua que não correspondem ao que é usado corriqueiramente. A situação torna-se ainda mais grave quando as palavras são apresentadas fora de contexto. Além de haver casos em que a língua presente nesses materiais não é autêntica, isto é, foi criada somente para fins didáticos, também deparamos com palavras ensinadas isoladamente, por exemplo, em listas que simplesmente agrupam um conjunto de palavras.

Com o uso de *corpus* nas pesquisas linguísticas, surgiu a abordagem de ensino chamada *Data-Driven Learning* (DDL) proposta por Johns (1988; 1991). Nessa perspectiva de ensino, aprendizes atuam como se fossem pesquisadores da língua, pois têm contato direto com o exame de um *corpus*, seja por meio da exploração dos dados em atividade guiadas ou mediante materiais que resultam da investigação prévia de um *corpus*, especialmente linhas de concordância.

Na DDL, é importante que o aprendiz tenha a habilidade de examinar os dados que lhe são fornecidos a fim de formar generalizações sobre o que está estudando. Da mesma maneira que um linguista investigaria uma concordância, por exemplo, um aprendiz também deve ser capaz de perceber se há regularidades nos dados de que dispõe.

Trata-se, portanto, de uma abordagem de ensino indutiva, em que cabe ao professor guiar o aprendiz de modo que seja possível criar saberes a respeito da língua. Não é papel do professor simplesmente transmitir um conhecimento como seria feito de modo tradicional. É essencial que o responsável pelo ensino, na perspectiva de DDL, crie condições favoráveis para que o aprendiz alcance o objetivo didático almejado. Para isso, o aprendiz pode ser auxiliado na exploração de um

corpus ou o professor pode selecionar concordâncias para serem analisadas por seu aluno[2].

1.3 *Corpus* de aprendizes

Uma pesquisa que explora um *corpus* de aprendizes não pode deixar de fazer referência ao trabalho de Sylviane Granger. Depois de anos de existência de *corpora* de língua geral, surgiram, no começo da década de 1990, os primeiros estudos com *corpus* de língua de falantes não-nativos. O primeiro, e até hoje o mais conhecido, *corpus* de aprendizes foi o *International Corpus of Learner English* (GRANGER, 1993), criado para estudar as características da linguagem de aprendizes de inglês.

Esse novo tipo de *corpus* foi uma inovação justamente porque trazia agora textos de aprendizes (GRANGER, 1996). Por conter dados de falantes não-nativos, uma das dificuldades iniciais da coleta desse tipo de material foi a espontaneidade de produção da língua. Na compilação desse *corpus*, também foram registrados dados sobre cada aprendiz, tais como gênero, idade e instrução escolar; e a situação de produção do texto (uso de material de consulta, por exemplo).

As pesquisas com *corpora* de aprendizes tiveram sucesso na investigação da linguagem de falantes não-nativos e tem contribuído para áreas como a aquisição de linguagem. Hoje, com a difusão desse tipo de pesquisa, é relativamente mais fácil coletar textos de aprendizes, tendo em vista que, por exemplo, professores podem coletar textos escritos por seus alunos. No entanto, como veremos adiante, nossa pesquisa ainda esbarrou no problema da coleta de textos em nossa própria instituição de ensino.

No início das pesquisas com *corpora* de aprendizes, foram criados critérios de compilação para as redações que pareciam exigir metadados específicos sobre os aprendizes (idade, gênero, escolaridade etc) e sobre as redações (contexto de produção, materiais de consulta, tempo etc). No entanto, com o tempo, surgiram outros *corpora* com dados linguísticos relevantes que não poderiam deixar de ser

[2] Em nossas considerações finais (capítulo 4), apresentamos um modelo de exercícios que adapta os conceitos de DDL visando a conscientização dos aprendizes em relação ao uso de certas colocações verbais. Essas atividades foram desenvolvidas com base nos resultados de nossa pesquisa.

explorados simplesmente pela falta desses metadados. Nosso *corpus* de estudo, como veremos adiante, não possui todas as informações desse tipo. Hoje, uma quantidade significativa dos textos é produzida em cursos de ensino de idiomas a partir de tópicos pré-estabelecidos para fazer parte de um *corpus* (BAKER, HARDIE e MCENERY, 2006) e, muita vezes, esses textos não possuem metadados.

1.4 Colocação

Historicamente, o termo colocação foi criado em inglês (*collocation*) devido à ausência de um vocábulo que se referisse ao fenômeno de agrupamento convencional de palavras. Em português, talvez o termo cognato colocação não tenha o mesmo efeito de sentido, a não ser para aqueles que têm familiaridade com o tema.

O conceito de colocação varia em sua definição, mas geralmente é usado para descrever a maneira pela qual as palavras se agrupam em seu uso normal em um texto (BARNBROOK, MASON e KISHNAMURTHY, 2013, p. 3)

Tagnin (2013, p. 151-152) diferencia os termos colocação e coligação. As colocações são combinações lexicais consagradas de duas ou mais palavras de conteúdo, como ‘repolho roxo’, ‘mundos e fundos’, ‘aviar uma receita’, por exemplo. Já as coligações são combinações consagradas de elementos linguísticos em que uma estrutura ou categoria gramatical se combina com a palavra base (e.g. ‘obedecer a’, ‘cumpridor de’).

Na medida em que esses padrões só são ‘percebidos’ após várias ocasiões em que um leitor depara com eles, a LC pode agilizar esse processo ao mostrar, numa concordância várias ocorrências de uma mesma colocação – como vimos na Figura 1 para *make sure*. Por essa razão, Nesselhauf (2005, p. 3) acredita que os padrões linguísticos que podem ser identificados pela LC devam ser ensinados. Muitos desses padrões, inclusive colocações, ainda não são tratados adequadamente no ensino de língua inglesa hoje. Embora as colocações tenham recebido atenção crescente no ensino de línguas nos últimos anos, ainda estamos muito longe do desenvolvimento de uma metodologia coerente e, menos ainda, de um amplo e sistemático tratamento de colocações em materiais de ensino de línguas e conteúdos programáticos. As abordagens de ensino têm por base, na melhor das hipóteses, a análise da linguagem

de falantes nativos; dificilmente são baseadas em qualquer observação sistemática da linguagem de aprendizes. Se medidas didáticas efetivas forem planejadas, elas precisam levar em consideração as dificuldades que os aprendizes têm com esse tipo de linguagem.

1.4.1 Colocação verbal

Nosso objetivo de pesquisa é analisar apenas uma categoria dentre os tipos de colocações, as colocações verbais. De acordo com Tagnin, uma colocação verbal é uma "combinação convencional, consagrada, de verbo e substantivo, podendo esse ter a função de objeto ou de sujeito do verbo" (2013). No entanto, não serão analisadas neste trabalho colocações em que o substantivo tem a função de sujeito.

Não há motivação para um verbo se combinar com um substantivo quando outro verbo similar ou sinônimo poderia ser usado – gramatical ou semanticamente falando – mas não é. Por exemplo, diz-se *show interest*, mas não *exhibit interest.* Também é importante apontar que nem sempre colocações em um língua correspondem a colocações em outras línguas.

Como nossa pesquisa é baseada no trabalho de Nesselhauf (2005), adotamos a definição que a pesquisadora utilizou em seu estudo. Em uma colocação, "o substantivo pode ser usado sem restrição arbitrária no sentido em que está sendo usado, mas o verbo é, no dado sentido, até certo ponto, arbitrariamente restrito a certos substantivos", por exemplo, *run the risk*. (NESSELHAUF, 2005, p. 33).

Apesar de nosso foco de análise serem as colocações verbais, também identificamos, no *corpus* de estudo, combinações livres e expressões idiomáticas como outros tipos de associações entre verbos e substantivos. Portanto, adotamos 'combinação' como o termo hiperônimo para se referir a qualquer arranjo em que um verbo se combina a um substantivo.

Na colocação do tipo verbal, consideramos que o verbo se une a um substantivo, e não o contrário. O substantivo funciona como a base da colocação, é a palavra conhecida do aprendiz, enquanto o verbo é o seu colocado, a palavra que ele desconhece. Em outras palavras, podemos dizer que um verbo se coloca com um substantivo ou simplesmente que um verbo e um substantivo formam uma colocação.

Quando mencionamos um ‘desvio’, estamos nos referindo a um distanciamento da convencionalidade linguística. Se um aprendiz produz, por exemplo, *cause an impression* com o sentido de ‘causar uma impressão’, consideramos que há um desvio, pois essa combinação afasta-se da convenção da língua, já que em inglês o verbo *make* coloca-se convencionalmente com o substantivo *impression*. Evidentemente, a depender do contexto em que a combinação é empregada, ela pode ser compreendida, pois, do ponto de vista da gramática da língua, é possível que *cause* e *impression* se combinem, mas não é provável que isso ocorra.

Da mesma forma que adotamos o termo ‘combinação’ para nos referirmos à associação de verbos e substantivos em geral, usamos a nomenclatura de ‘desvio combinatório’ para fazer referência aos casos em que uma combinação não é adequada segundo a convencionalidade da língua.

Tendo em vista a sobreposição entre a teoria e a prática na LC, no próximo capítulo, em que descrevemos a metodologia adotada em nossa pesquisa, também fazemos algumas considerações teóricas do nosso trabalho. Além disso, apresentamos as ferramentas que possibilitaram esta pesquisa, como também discorremos sobre o conceito de anotação textual nos cabeçalhos e nos segmentos de texto. Por fim, retomamos o conceito de colocação e introduzimos os outros tipos de combinação que também foram identificadas em nosso estudo.

CAPÍTULO 2. METODOLOGIA

Neste capítulo, apresentamos brevemente o percurso que nos levou a adotar o *corpus* de estudo eleito para a pesquisa. Também fazemos uma descrição desse *corpus* e das ferramentas de análise de dados, assim como dos procedimentos metodológicos que seguimos no estudo.

2.1 Início da pesquisa

Em nossa primeira tentativa de investigar colocações verbais na escrita de brasileiros aprendizes de inglês, julgamos conveniente coletar textos escritos por alunos de graduação da mesma instituição de ensino em que a pesquisa foi conduzida.

O plano de analisar a produção escrita de alunos de graduação foi motivado pela existência de projetos similares realizados em outras universidades, em particular o *International Corpus of Learner English* (ICLE), pioneiro entre as empreitadas do tipo. O ICLE é composto por redações escritas por alunos cursando o segundo, terceiro ou quarto ano de língua inglesa em nível de graduação (GRANGER, 2008, p. 5).

Na Faculdade de Filosofia, Letras e Ciências Humanas (FFLCH) da Universidade de São Paulo (USP), os alunos do curso de Letras que optam pela habilitação em inglês escrevem seus trabalhos em língua inglesa para as disciplinas que cursam. Tendo conhecimento de tal fato, iniciamos a coleta de textos nas disciplinas obrigatórias Comunicação Escrita 1 (primeiro semestre) e 2 (segundo semestre) oferecidas então nos períodos matutino e noturno por dois professores, totalizando quatro turmas em cada semestre letivo. Nossa escolha foi motivada pelo objetivo da disciplina de, segundo o programa do curso, "desenvolver habilidades comunicativas na língua inglesa" na modalidade escrita.

Outro fator que incentivou a coleta de dados em nossa faculdade foi a disponibilização do *Corpus* Multilíngue de Aprendizes (CoMAprend) (TAGNIN, 2006; TAGNIN, 2013). Parte integrante do projeto *Corpus* Multilíngue para Ensino e Tradução (CoMET) da FFLCH-USP, a plataforma permite o armazenamento e a

análise de textos escritos por aprendizes das cinco línguas (alemão, espanhol, francês, inglês e italiano) que fazem parte do departamento em que foi desenvolvida. Após o registro de uma disciplina, por parte de um pesquisador, professor ou ministrante de curso previamente cadastrado, bem como do cadastro dos alunos por meio de um formulário de inscrição, as redações são submetidas pelos próprios alunos inscritos. Desse modo, esses textos são arquivados e organizados de acordo com a disciplina em que estão matriculados em determinado semestre de estudo.

A situação parecia favorável à obtenção de dados para nosso estudo, afinal, os textos já eram regularmente produzidos por aprendizes brasileiros, isto é, não seria necessário buscar outros participantes para atingirmos nossa meta. Além disso, o CoMAprend apresentava-se como modo ideal de reunir formalmente o material coletado.

No entanto, para coletarmos uma quantidade de redações que permitisse fazer generalizações sobre o uso de colocações, dependíamos do consentimento dos docentes que ministravam as disciplinas em que tínhamos interesse, assim como da colaboração voluntária dos alunos de tais disciplinas para que submetessem suas produções escritas. Apesar de termos conseguido certa cooperação de ambas as partes, não atingimos uma quantidade significativa de textos coletados devido à baixa adesão dos envolvidos na coleta[3].

Nesse contexto, tomamos conhecimento, por acaso, de outro *corpus* que se revelou apropriado para o estudo das colocações verbais na escrita de aprendizes brasileiros de inglês. Na seção seguinte, descrevemos esse *corpus* e a composição do *subcorpus* que adotamos em nossa pesquisa.

2.2 O corpus de estudo: EFCAMDAT

Em nosso estudo, adotamos o *corpus Education First-Cambridge Open Language Database* (EFCAMDAT), cuja primeira versão foi disponibilizada pública e gratuitamente *on-line* em 2013 (EDUCATION FIRST E UNIVERSIDADE DE

[3] Quatro docentes da instituição deram consentimento para que realizássemos a coleta, porém poucos alunos efetivamente submeteram suas redações, o que resultou em uma quantidade de dados insuficientes para a realização da pesquisa.

CAMBRIDGE). Atualmente, entretanto, o *corpus* não está disponível para acesso público, pois os responsáveis pelo projeto estão trabalhando em sua segunda versão que, segundo o site em que está hospedado, será lançada em breve. O projeto foi desenvolvido na Universidade de Cambridge (Reino Unido) em colaboração com a organização internacional de ensino de línguas *Education First*[4]. Com 32.980.407 palavras, a primeira versão do EFCAMDAT possui 551.036 textos escritos por 84.864 aprendizes. Apesar de não haver informação sobre a(s) língua(s) falada(s) pelos alunos, é possível recuperar o país de origem dos mesmos: 172 nacionalidades estão representadas no *corpus*. É importante destacar que mais de um terço (36,9%) das redações foram escritas por brasileiros (GEERTZEN, ALEXOPOULOU e KORHONEN, 2012).

2.2.1 *Englishtown*

Os textos que compõem o *corpus* são provenientes do curso *on-line* de ensino de inglês *Englishtown* (EDUCATION FIRST), o qual é oferecido em diversos países, para falantes de diferentes línguas, pela *Education First*. Recentemente, a versão central inglesa do curso teve seu nome alterado e passou a se chamar *English Live* (EDUCATION FIRST). Contudo, o site disponível em domínio brasileiro manteve o nome original *Englishtown.* A despeito da mudança de nome, a estrutura do curso não foi modificada, portanto descrevemos a composição do curso a fim de se compreender a consequente constituição do EFCAMDAT.

O *Englishtown* é um curso de inglês geral, voltado para adultos e organizado em dezesseis níveis que compreendem todas as categorias encontradas no Quadro Europeu Comum de Referência para as Línguas (QECR) (COUNCIL OF EUROPE) conforme a Tabela 1.Cada conjunto de três níveis do *Englishtown* corresponde a uma categoria do QECR, com exceção do último nível (16), que isoladamente corresponde ao nível C2.

[4] Na busca por instituições de pesquisa para desenvolvimentos de projetos cooperativos, a *Education First* estabeleceu um convênio com a USP no ano de 2013.

Tabela 1 – Comparação entre os níveis do *Englishtown* e do QECR

Englishtown	**QECR**
1-3	A1
4-6	A2
7-9	B1
10-12	B2
13-15	C1
16	C2

Antes de começar o curso, um novo aluno passa por um teste de nivelamento que o classifica como apto a iniciar em um nível correspondente ao QECR, ou seja, esse aluno pode começar nos níveis 1, 4, 7, 10, 13 ou 16. Dessa forma, um aluno que, conforme avaliação prévia, já possui as habilidades previstas, por exemplo, na categoria A1, poderá começar o curso no nível 4 (A2).

Cada um dos dezesseis níveis possui oito unidades didáticas a serem cumpridas pelos alunos. Uma unidade didática, por sua vez, compreende diversas tarefas que os alunos devem realizar. Muitas tarefas são corrigidas automaticamente pelo próprio sistema do curso, porém, uma das tarefas, no mínimo, consiste em um texto escrito, o qual é examinado por um avaliador humano que lhe atribui uma nota em uma escala de 0 a 100. Os critérios para a atribuição da nota, todavia, não estão disponíveis, portanto não serão apresentados. Na constituição do EFCAMDAT, foram compiladas as redações vindas de cada uma das oito unidades dos dezesseis níveis, portanto o *corpus* é composto por 128 propostas de temas de redações.

Em nosso estudo, entretanto, não analisamos a totalidade do EFCAMDAT, ou seja, todos os textos das 128 propostas, tampouco incluímos todas as nacionalidades representadas no *corpus*. Uma vez que nosso interesse são colocações verbais na escrita de brasileiros aprendizes de inglês, investigamos somente as redações do nível 16 (C2) do *Englishtown*, escritas por brasileiros. Nossa decisão foi motivada pelo estudo de Nesselhauf (2005), em que a pesquisadora analisa colocações verbais na escrita de aprendizes alemães somente com o nível de inglês avançado, e verifica que tais aprendizes ainda apresentam dificuldades no uso de colocações apesar de seu grau de proficiência. Nesse sentido, McCarthy (2013, p. 42-43) já menciona a importância de um aprendiz de inglês avançado, em especial, ter conhecimento colocacional na aprendizagem de vocabulário. Nessa categoria de aprendizado em

particular, mais do que aprender palavras individualmente, o aprendiz deve saber combinar as palavras de forma adequada. Entretanto, ainda segundo o autor, em comparação com os níveis mais elementares do QECR, há menos consenso sobre o que deve ser ensinado nos níveis C1 e C2, de modo que acredita que *corpora*, tanto de língua nativa como de aprendizes, podem ser uma excelente fonte de evidências sobre o que abordar nos estágios avançados. A partir daí, podemos levantar a hipótese de que as dificuldades colocacionais dos aprendizes se devem a lacunas no conteúdo programático dos níveis avançados. No entanto, por se tratar de uma pesquisa de mestrado, restringimos nosso estudo ao material do nível C2, pois a metodologia de anotação adotada é bastante trabalhosa.

A Tabela 2 apresenta os nomes das oito unidades do nível 16 no *Englishtown*, bem como os tópicos dos textos correspondentes a cada uma das unidades.

Tabela 2 – Tópicos das redações do nível 16 (C2) do *Englishtown*

Unidade	**Nome da unidade**	**Tópico do texto**
1	*Science and technology*	*Attending a robotics conference*
2	*National identity*	*Writing about a symbol of your country*
3	*Tough choices*	*Following a code of ethics*
4	*Fame and fortune*	*Criticizing a celebrity*
5	*Creative thinking*	*Using creative writing techniques*
6	*Financial planning*	*Applying for a home loan*
7	*Dealing with stress*	*Writing a visualization script*
8	*Doing research*	*Researching a legendary creature*

2.2.2 Cabeçalho

Os textos armazenados no EFCAMDAT possuem um cabeçalho em inglês com informações pertinentes à redação. Tais informações não são escritas pelos alunos durante o curso *Englishtown*, mas atribuídas ao texto pelo compilador do *corpus* para facilitar sua classificação. O Quadro 1 apresenta um modelo de cabeçalho com suas etiquetas (*tags*) empregadas entre parênteses angulares (< >). Com o intuito de elucidar a função das etiquetas do cabeçalho, entre parênteses simples e em fonte

distinta, adicionamos ao modelo abaixo o tipo de informação expressa pela etiqueta. Os metadados presentes no cabeçalho podem estar dentro da própria etiqueta, ou seja, entre parênteses angulares, como no caso de *<writing id= "(código do texto)" level="(nível: 1-16)" unit="(unidade: 1-8)">*, para citar alguns exemplos. As informações também podem estar entre duas etiquetas, cada qual com um par de parênteses angulares. Nesse caso, a segunda etiqueta, após a informação, recebe uma barra oblíqua (/), por exemplo, *</topic>*.

```
<writing id="(código do texto)" level="(nível: 1-16)" unit="(unidade: 1-8)">
        <learner id="(código do aprendiz)" nationality="(código da nacionalidade)"/>
         <topic id="(código do tópico do texto)">(tópico do texto)</topic>
         <date>(data de submissão do texto)</date>
         <grade>(nota: 0-100)</grade>
         <text>
(corpo do texto)
         </text>
      </writing>
```

Quadro 1 – Modelo de cabeçalho dos textos no EFCAMDAT

No Quadro 2, exemplificamos o modelo acima com um texto e seu cabeçalho devidamente preenchido. Os códigos do texto (*writing id*), do aprendiz (*learner id*) e do tópico do texto (*topic id*) consistem em um conjunto de caracteres, em sua maioria numéricos, e aleatórios, porém particulares a cada um desses três elementos (texto, aprendiz e tópico do texto). As informações referentes a nível (*level*), unidade (*unit*) e nota (*grade*) variam conforme escalas específicas de cada categoria. Já a nacionalidade (*nationality*) é representada em todos os cabeçalhos analisados por duas letras (br), dado que todos os textos selecionados para o estudo foram escritos por brasileiros. A única informação que efetivamente constitui uma frase é o tópico do texto (*topic*). A data (*date*), por sua vez, é informada no formato ano-mês-dia, além de ser acompanhada pelo horário de submissão do texto. Após essas informações, entre as etiquetas de texto (*text*), encontramos a redação escrita pelo aluno[5]. O fim do

[5] No exemplo do Quadro 2, o texto é apresentado com os desvios cometidos pelo aprendiz.

conjunto de informações pertencentes a um texto é indicado por uma última etiqueta, </*writing*>.

<writing id="U18587" level="16" unit="1">
<learner id="18622336" nationality="br"/>
<topic id="7524">Attending a robotics conference</topic>
<date>2011-06-07 15:31:12.227</date>
<grade>88</grade>
<text>
There is no way to scape from technology anymore. As long as technolgy develops, people get more and more dependent of the facilities it provides to them. With the development of indutrial robots, industries have decreased the number of employees such as daily unpleasant situations that always happen between people in any industry. Industrial robots are able to work fast and efficiently, eliminating wasted time by going to the toillet or even time to have lunch. Industrial robots don't have family, don't get sick and are able to work without stopping. Humanoid is another type of robot in constant development. This is a robot with characteristics very similar to humans and that is able to display some emotion and intelligence. Service robots will be very important in areas such as medicine and and health, being able to watch people with contagious diseases as well as to take care of elderly people. Trundling around the hospital an picking up the trash such as taking care of the surveillance of the hospital are other areas that a service robot can be useful. Another field that robotics has been reaching is sports field, where robots have place to participate of sports events.
</text>
</writing>

Quadro 2 – Exemplo de texto do EFCAMDAT com cabeçalho preenchido

2.2.3 Seleção de textos

A interface do sistema do EFCAMDAT permitia a seleção de aprendizes de acordo com nacionalidade, nível e lição. Selecionamos, portanto, os textos brasileiros das oito unidades do nível 16.

Uma vez selecionados, os textos podiam ser explorados por meio do mecanismo de pesquisa do próprio site. Esse recurso consistia em buscar desde palavras específicas até combinações de classes de palavras e estruturas sintáticas tendo em vista que o *corpus* é anotado para tais categorias. Optamos, no entanto, por outra possibilidade disponível na plataforma, a de baixar os textos para o computador do pesquisador, uma vez que utilizamos uma ferramenta específica para a investigação das redações.

Uma particularidade dos textos baixados é o fato de estarem todos em um único arquivo computacional em formato HTML (*HyperText Markup Language*), um tipo de linguagem comumente usada na criação de sites. A fim de facilitar nosso trabalho, separamos cada redação em um arquivo independente com extensão TXT (*Text File*), cuja formatação mais simples exibe somente o texto pleno independente de atributos formais como tipo e tamanho de fonte ou características como negrito, itálico e sublinhado, por exemplo. Para realizar a tarefa de separação em arquivos autônomos, utilizamos a ferramenta *splitter* (Figura 2) do programa computacional *WordStmith Tools 6.0*. Por meio da etiqueta de marcação de final de texto, </*writing*> a ferramenta divide um mesmo arquivo em arquivos menores que são destinados e organizados em uma pasta.

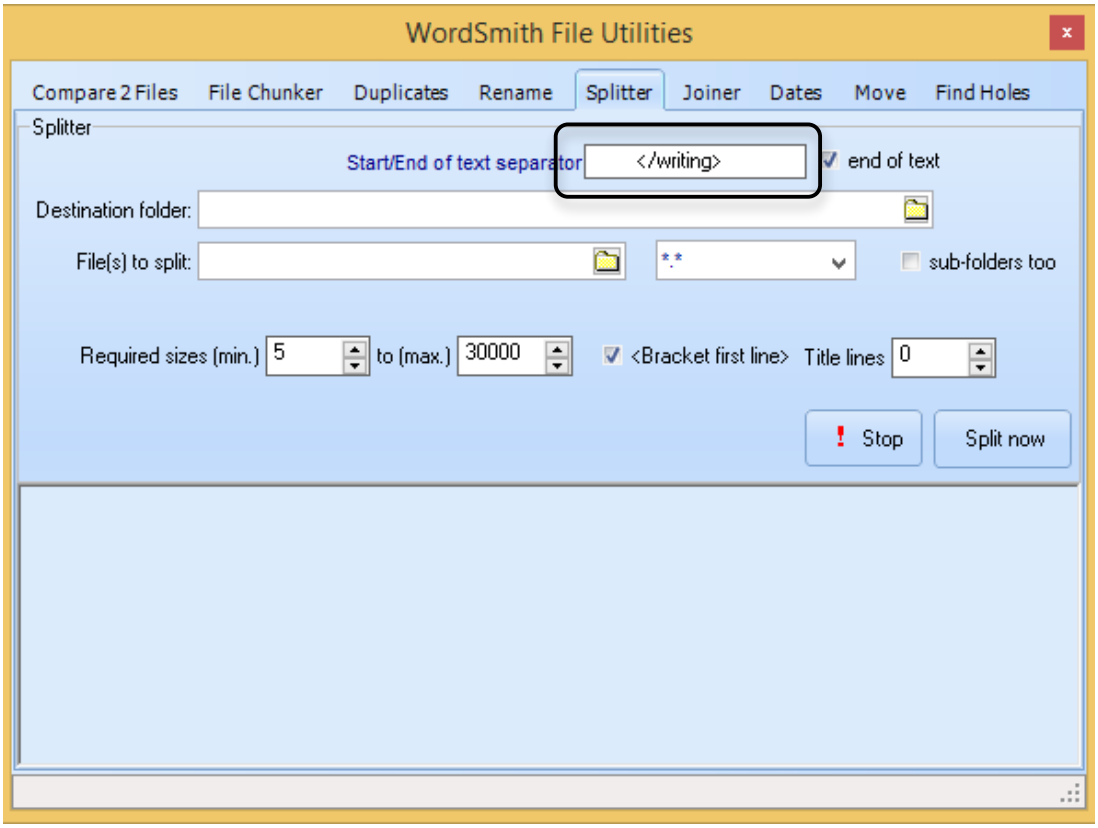

Figura 2 – Captura de tela da ferramenta *splitter* do *WordSmith Tools 6.0*

Após o procedimento de divisão de todos os textos do nível 16, escritos por brasileiros, em arquivos individuais, possuíamos 169 arquivos. No entanto, em uma primeira leitura, verificamos que algumas redações não seriam adequadas à pesquisa devido ao seu conteúdo. Apesar de figurarem com cabeçalho no *corpus*, alguns textos apresentavam problemas que nos levaram a excluí-los da nossa amostra final. Por exemplo, o sistema do *Englishtown* não permite que um indivíduo inscrito no curso avance para a unidade seguinte sem que tenha completado todas as tarefas de uma unidade, o que inclui escrever a redação. Alguns textos, no entanto, foram submetidos com apenas alguns caracteres. Uma vez que a nota mínima para ser aprovado em uma unidade não depende somente da escrita da redação, mas do desempenho no conjunto de tarefas propostas, acreditamos que alguns alunos propositalmente tenham escrito apenas alguns caracteres a fim de prosseguirem para o estágio subsequente (Quadro 3). Nesse aspecto, também temos de considerar a própria dificuldade de se lidar com o ambiente computadorizado do curso e suas possíveis falhas. Em outras palavras, textos encontrados no EFCAMDAT com apenas alguns caracteres podem ter sido submetidos acidentalmente pelos alunos, mesmo não estando completos

```
<writing id="U500909" level="16" unit="1">
        <learner id="18538337" nationality="br"/>
        <topic id="7524">Attending a robotics conference</topic>
        <date>2012-07-11 10:21:28.387</date>
        <grade>0</grade>
        <text>
        ...
        </text>
      </writing>

<writing id="U557733" level="16" unit="1">
        <learner id="18536449" nationality="br"/>
        <topic id="7524">Attending a robotics conference</topic>
        <date>2012-08-16 11:43:53.017</date>
        <grade>0</grade>
        <text>
        dfffeew qwqwdd d WDWDWQ   QADQSQSA   qawdwdS  SQS
        </text>
      </writing>
```

Quadro 3 – Dois exemplos de textos do EFCAMDAT com apenas alguns caracteres

Outro caso recorrente que encontramos entre os textos, e que nos levou a não incluir certos deles em nosso *corpus* de estudo, foi o uso de pequenas mensagens direcionadas ao avaliador da redação (Quadro 4).

```
<writing id="U528264" level="16" unit="2">
        <learner id="19441112" nationality="br"/>
        <topic id="8341">Writing about a symbol of your country</topic>
        <date>2012-07-29 20:04:55.823</date>
        <grade>45</grade>
        <text>
        I'll be sending it soon.
        </text>
      </writing>

<writing id="U460693" level="16" unit="2">
        <learner id="22673858" nationality="br"/>
        <topic id="8341">Writing about a symbol of your country</topic>
        <date>2012-06-20 01:40:15.843</date>
        <grade>38</grade>
        <text>
        Dear teacher, please do not consider this text for grading purposes. i will do it later. I am just writing this message so that i can move to the next unit. Thanks and regards,  Andre
        </text>
      </writing>
```

Quadro 4 – Dois exemplos de textos do EFCAMDAT com pequenas mensagens

Além de textos com caracteres aleatórios e com mensagens para o avaliador, também encontramos textos que eram cópias exatas, ou muito similares, do modelo de texto que o curso oferece ao aluno para cada unidade. Apesar de o sistema não permitir que o modelo de redação seja copiado facilmente pelos alunos, acreditamos que alguns devem ter encontrado uma maneira de fazê-lo. A Figura 3 mostra a parte inicial de um modelo de texto que foi integralmente copiado e submetido por um aluno (Quadro 5).

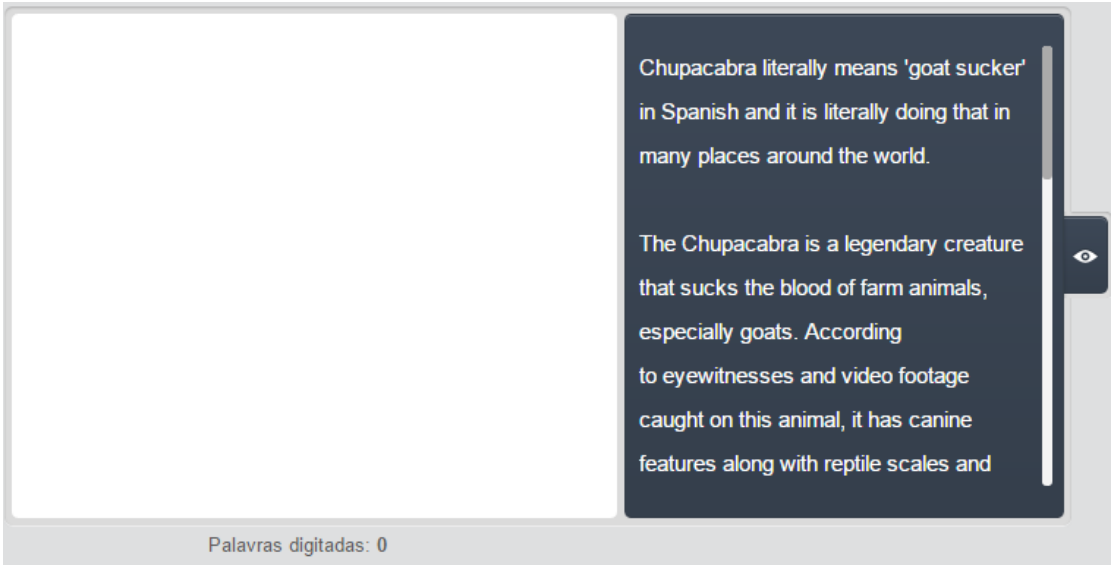

Figura 3 – Captura de tela de parte de modelo de redação no *Englishtown*

<writing id="U534957" level="16" unit="8">
<learner id="19441112" nationality="br"/>
<topic id="9186">Researching a legendary creature</topic>
<date>2012-08-02 07:29:07.950</date>
<grade>98</grade>
<text>
Chupacabra literally means 'goat sucker' in Spanish and it is literally doing that in many places around the world. The Chupacabra is a legendary creature that sucks the blood of farm animals, especially goats. According to eyewitnesses and video footage caught on this animal, it has canine features along with reptile scales and spines up and down its back. It stands 3-4 feet tall and hops like a kangaroo. Other features include a forked tongue, bulging eyes and long fangs. It is reported to hiss and screech when approached and it gives off a bad stench. Other witnesses say that its eyes glow red, which made them feel nauseous. The first known sighting was in 1995 in Puerto Rico, where eight sheep were attacked and their blood sucked dry. The only wounds on the animals were puncture wounds. Months later in a nearby area, 150 animals were reported to be killed in the same fashion. There have been numerous sightings of these creatures all over the world, From as far north as Maine in the northern US, but most reports are in the southern part of the US and in Latin America.
</text>
</writing>

Quadro 5 – Exemplo de texto do EFCAMDAT com cópia de modelo de redação

Decidimos manter textos que aparentemente podem ser cópias ou paráfrases do modelo de redação disponível[6], pois também poderiam conter colocações inadequadas. Assim, ainda que nossa decisão tenha levado a resultados em que algumas colocações empregadas adequadamente estejam repetidas, também encontramos desvios no emprego de outras combinações. Nossa decisão de manter tais redações não afetou os resultados. Pelo contrário, contribuiu para que encontrássemos mais desvios.

Na constituição de nosso *corpus* de pesquisa, a partir da parcela do EFCAMDAT, constituída apenas por textos escritos por brasileiros e que pertencem ao nível 16, excluímos 56 textos que apresentavam um dos casos descritos acima. Com isso, o *subcorpus* final consiste em 113 textos, totalizando 19.329 palavras. Os detalhes quantitativos de cada uma das unidades do nível 16, a quantidade de textos excluídos e utilizados por unidade, assim como a média de palavras por texto podem ser vistos na Tabela 3. No Apêndice A, listamos os textos utilizados no *subcorpus* por meio dos códigos (*writing id*) presentes no EFCAMDAT. Para facilitar a menção a textos específicos, empregamos, ao longo deste trabalho, o número da unidade seguido do número do texto. Dessa forma, U2-04, por exemplo, refere-se ao texto de número 4 da unidade 2.

Tabela 3 – Descrição quantitativa do *corpus* de estudo

Unidade	**Textos**			**Palavras**	**Palavras/texto**
	EFCAMDAT	**Excluídos**	**Utilizados**		
1	36	9	27	4.817	178,41
2	28	7	21	3.536	168,38
3	20	6	14	2.174	155,29
4	17	4	13	2.121	163,15
5	18	6	12	2.255	187,92
6	15	11	4	688	172,00
7	16	3	13	2.085	160,38
8	19	10	9	1.653	183,67
Total	169	56	113	19.329	171,05

[6] Não obtivemos informações sobre a autoria dos modelos.

Com a etapa de seleção dos textos cumprida e o *corpus* de estudo delimitado, o passo seguinte foi a identificação de colocações verbais nas redações de aprendizes brasileiros de inglês. Em um primeiro momento, consideramos identificar as colocações verbais recorrendo ao levantamento dos verbos mais frequentes no *corpus.* A extração das colocações pode ser feita com o auxílio de ferramentas tradicionalmente usadas para esse fim, tais como lista de palavras, que elenca as palavras mais frequentes no *corpus*, e linhas de concordância, que apresentam todos os contextos de ocorrência de uma ou mais palavras de busca (cf. ORENHA-OTTAIANO , 2012 e SHEPHERD, 2009). Alguns estudos preliminares chegaram a ser conduzidos, porém constatamos que as poucas colocações verbais identificadas por meio desse método não representavam uma quantidade significativa do total de ocorrências que poderia ser encontrado entre os dados. Os dez verbos de conteúdo mais frequentes encontrados no *subcorpus* (*have, like, use, do, make, take, arrest, see, help, question*) [7] são responsáveis por aproximadamente apenas um terço das colocações que identificamos ao final do estudo. Ainda que seja um tipo de metodologia possível, tal procedimento leva a resultados que dizem respeito somente às colocações com os verbos mais frequentes reconhecidos, isto é, deixa de considerar o fenômeno colocacional entre verbos e substantivos como um todo na amostra. Face a isso, recorremos ao uso de outra ferramenta computacional relevante para nossa pesquisa que é descrita a seguir.

2.3 UAM CorpusTool

UAM CorpusTool (O'DONNELL, 2011) é um programa computacional que permite a anotação de *corpora*. Quando falamos em anotação, estamos nos referindo ao processo de classificação de textos, ou partes de textos, em categorias específicas por meio da adição de informações atribuídas automática ou manualmente àquilo que é anotado. Os textos do curso *Englishtown* que fazem parte do EFCAMDAT, como mencionamos anteriormente, receberam um cabeçalho com informações sobre cada redação. O procedimento de adição das etiquetas (*tags*) que formam um cabeçalho

[7] Nesse levantamento, consideramos a forma lematizada do verbo, ou seja, para o verbo *have*, por exemplo, foram consideradas as formas *have, has, having* e *had.*

com dados particulares dos textos é um exemplo de anotação textual. Outro tipo de anotação, porém de palavras, é a anotação morfológica, que classifica as palavras de um *corpus* de acordo com as classes gramaticais da língua.

Após a tentativa sem sucesso de identificar as colocações verbais a partir dos verbos mais recorrentes, passamos à identificação com o auxílio das ferramentas do *UAM CorpusTool*, tendo em vista que alguns fenômenos linguísticos, os de natureza semântica e pragmática, por exemplo, que ainda não podem ser anotados automaticamente precisam ser identificados por anotadores humanos (O'DONNELL, 2008, p. 1434). No caso da nossa pesquisa, além de eventuais desvios de grafia que impedem a identificação das classes de palavra, a variedade de estruturas sintáticas em que colocações verbais podem ocorrer dificulta a anotação de modo totalmente automático. Em nossa anotação, identificamos desde uma colocação (exemplo[8] 1) em que o verbo (*gave*) e o substantivo (*chase*) estão adjacentes até o arranjo oposto (exemplo 2), em que o substantivo (*improvements*) precede o verbo (*achieving*), porém não nas imediações um do outro, pois o verbo está na oração relativa após o substantivo. Em outras palavras, a anotação automática abrangeria a colocação do exemplo 1 mais facilmente do que aquela no exemplo 2.

1. *Then, police gave chase and cornered him at 6 am, a few hours later.* (U3-03)
2. *She is very satisfied about the improvements that her robots have been achieving in areas like medicine and health.* (U01-18)

Uma vez que a anotação de modo totalmente automático não produziria o resultado desejado, aliado ao fato de que a anotação completamente manual seria exaustiva, anotamos os dados semiautomaticamente conforme descrevemos adiante.

O *UAM CorpusTool* funciona a partir de um esquema de anotação. Mais detalhes sobre o funcionamento do programa computacional serão discutidos adiante.

2.3.1 Esquema de anotação

[8] A fim de facilitar a leitura, somente as partes em discussão no texto estão sublinhadas nas sentenças dos exemplos. As frases podem ou não conter outros segmentos que foram analisados na pesquisa, porém não estão evidenciados em nosso texto para não confundir o leitor.

Nesta subseção, apresentamos o esquema de anotação utilizado em nosso estudo, abordando quatro tipos de combinação (2.3.1.1): não-combinação (2.3.1.1.1), combinação livre (2.3.1.1.2), colocação (2.3.1.1.3) e expressão idiomática (2.3.1.1.4). Em cada uma delas, explicamos os critérios adotados para sua classificação e detalhamos os níveis que subjazem às suas categorias. Ao final, são apresentadas as estruturas sintáticas em que as combinações ocorreram na pesquisa (2.3.1.2).

Um dos primeiros passos na utilização do *UAM CorpusTool* é a definição de um esquema de anotação. O esquema consiste em uma estrutura hierárquica utilizada para categorizar textos e/ou segmentos dentro de textos. Como em nossa pesquisa anotamos colocações, fizemos a anotação de segmentos dos textos. Esquemas de anotação previamente elaborados acompanham o programa computacional, contudo concebemos um novo esquema de modo a adaptar a proposta de Nesselhauf (2005) sobre combinações entre verbos e substantivos.

Apesar de nosso foco estar voltado para colocações verbais, não anotamos somente essa categoria. Todos os verbos foram anotados a fim de verificarmos a incidência das colocações verbais em relação às demais ocorrências de verbos. Isso também foi feito em consideração a possíveis pesquisas futuras e reinterpretações dos dados.

Antes de introduzirmos as categorias que fazem parte do esquema de anotação, cabem algumas observações. Em primeiro lugar, é importante salientar que toda anotação, seja ela automática ou manual, está sujeita a erros. Quando a anotação é feita automaticamente, um anotador automático pode não ser capaz de atribuir a categoria correta a um determinado dado ou até mesmo deixar de classificar algo que deveria ter considerado. Do mesmo modo, um anotador humano também pode cometer tais erros. O tipo de anotação, portanto, interfere na taxa de sucesso dos resultados, isto é, como apontamos anteriormente, um anotador humano pode estar mais habilitado do que um computador para analisar dados de acordo com certas categorias, por exemplo, semânticas. Portanto, nosso trabalho está suscetível aos mesmos riscos.

Na tentativa de superar o obstáculo descrito acima e obter resultados mais precisos, combinamos os dois tipos de anotação, automática e manual, e fizemos uma anotação semiautomática. A ferramenta *AutoCode* do *UAM CorpusTool* identifica

automaticamente verbos em inglês por meio de um dicionário que faz parte do programa[9] (Figura 4).

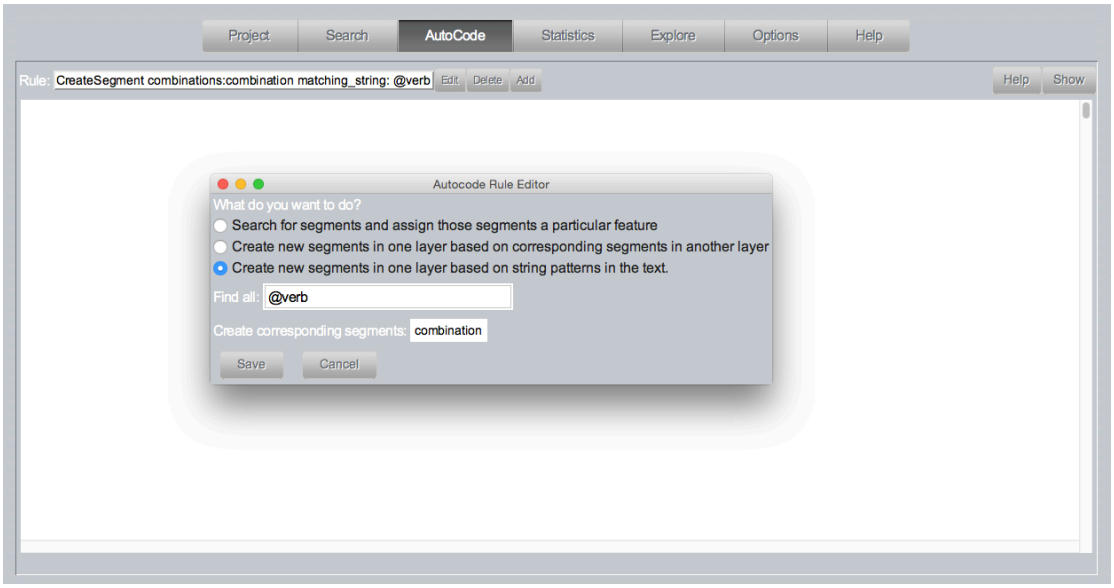

Figura 4 – Captura de tela da ferramenta *AutoCode* do *UAM CorpusTool*: anotação automática de verbos

Utilizamos inicialmente a anotação automática na tentativa de garantir que todo verbo fosse anotado. Entretanto, tendo em vista a morfologia da língua inglesa, em que uma palavra como *name*, por exemplo, pode ser usada como substantivo ou verbo, a ferramenta anota segmentos sem que necessariamente estejam sendo empregados como verbos. Outra imprecisão da ferramenta é o fato de não anotar alguns verbos como *consists, last, put*, os quais tivemos de anotar manualmente (Figura 5). Outra abordagem para anotar os dados seria a anotação automática de substantivos para que então as combinações fossem identificadas. No entanto, os substantivos são mais numerosos que os verbos e nem sempre coocorrem com verbos, o que tornaria o processo de anotação automática ineficiente em nosso caos..

[9] No campo da ferramenta *AutoCode*, o usuário deve usar o comando *@verb*.

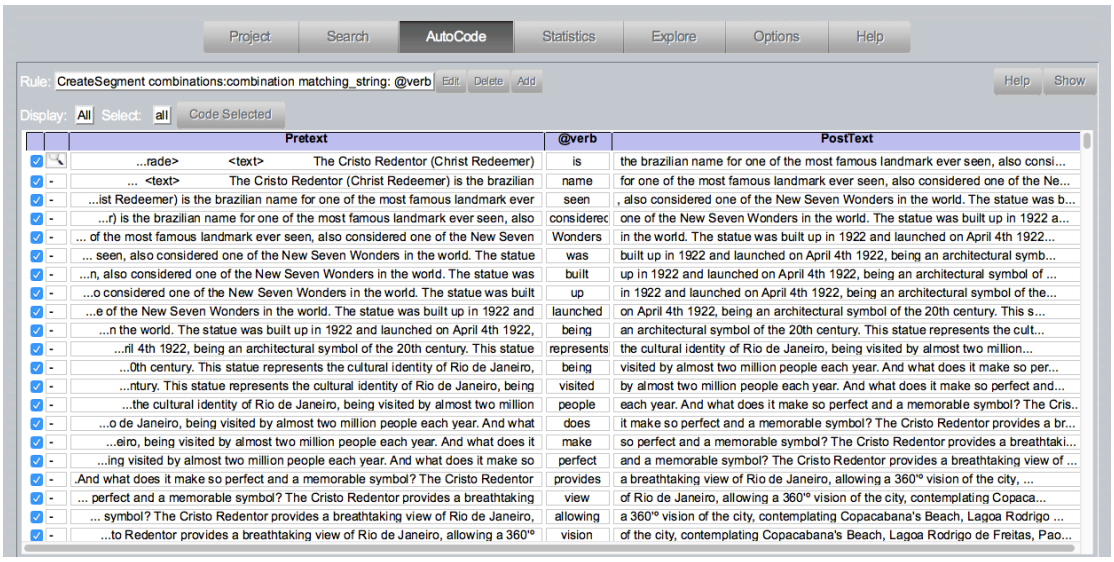

Figura 5 – Captura de tela da ferramenta *AutoCode* do *UAM CorpusTool*: verbos anotados automaticamente

Após a anotação automática, os textos anotados estão disponíveis na ferramenta *Project*. Quando abrimos um texto, os segmentos que foram identificados automaticamente como verbos estão sublinhados (Figura 6).

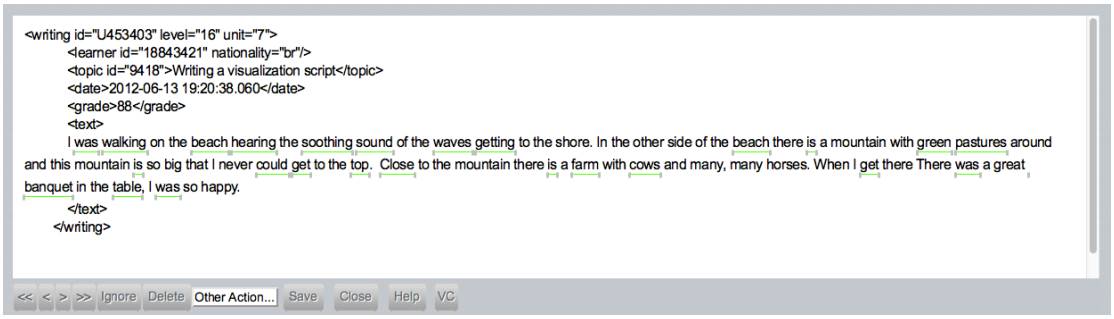

Figura 6 – Captura de tela da ferramenta *Project* do *UAM CorpusTool*: segmentos anotados automaticamente

A partir dessa primeira identificação dos verbos no *subcorpus*, o estágio manual de nossa anotação semiautomática consistiu em verificar exaustivamente cada um dos 5867 segmentos anotados automaticamente. A ferramenta permite, entre diversas aplicações, que um segmento anotado seja excluído caso o pesquisador julgue necessário. Utilizamos essa funcionalidade principalmente para descartar os segmentos anotados incorretamente como verbos. Outra utilidade do programa é a alteração de um segmento anotado. O usuário arrasta os limites da linha sob um segmento para incluir outras palavras na anotação. Usamos esse recurso para

efetivamente anotar combinações, e não somente verbos. O uso dessas duas funções da ferramenta (exclusão e alteração de segmentos) foi importante, pois anotamos verbos auxiliares (exemplo 3) e modalizações de verbos (exemplo 4) como parte do verbo principal, a despeito de a ferramenta de anotação automática ter identificado os verbos separadamente. Nos exemplos, os segmentos sublinhados correspondem a colocações.

3. *In the next 10 years, this robot will be programmed to visit quarantine patients who are suffering from contagious diseases.* (U1-14)
4. *Such as: As a soldier on wars, clean houses; we can watch tv while the robots make some so rude house taske.* (U1-10)

Todos os textos foram verificados individualmente de modo que a identificação feita automaticamente fosse corrigida[10]. No texto da Figura 6 (U7-08), por exemplo, a anotação automática identificou 25 segmentos, porém restaram somente 10 combinações após a correção manual, como mostra a Figura 7.

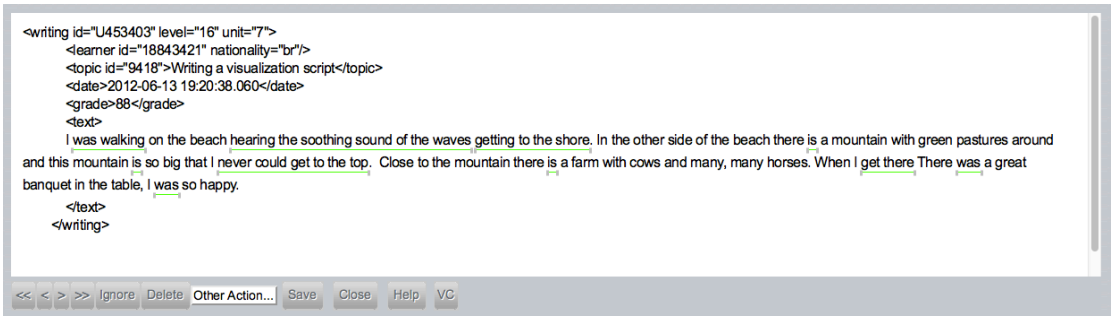

Figura 7 - Captura de tela da ferramenta *Project* do *UAM CorpusTool*: segmentos corrigidos manualmente

Também fizemos uso do recurso de anotação automática da ferramenta para definir uma classificação sintática específica. Dessa vez, inserimos a informação de que qualquer conjugação do verbo *to be* seguida de um verbo no particípio passado[11] deveria ser categorizada como voz passiva. Novamente, no estágio manual da anotação, validamos a anotação automática assim como identificamos casos que a ferramenta não anotou.

[10] Além do pesquisador, a orientadora também verificou os dados analisados.
[11] No campo da ferramenta *AutoCode*, o usuário deve usar o comando *be% @participle*.

Uma particularidade da anotação foram casos em que dois ou mais verbos ou substantivos formam colocações. No primeiro caso (exemplo 5), dois ou mais verbos (*explain* e *developing*) foram combinados com o mesmo substantivo (*ideas*), já no segundo caso (exemplo 6), dois ou mais substantivos (*thoughts* e *feelings*) foram combinados com o mesmo verbo (*expresses*). Em ambas as ocasiões, cada combinação foi anotada separadamente.

5. *In the cubing your topic will have six sides, and you will spend a feel time explain and developing you ideas of each side.* (U5-01)
6. *Creative writing is writing that expresses the writer's thoughts, feelings in an imaginative and poetic way!* (U5-07)

Por fim, é importante ressaltar que não analisamos todos os desvios encontrados no *subcorpus*. Como pode ser observado nos exemplos, há desvios de diferentes naturezas, porém concentramos nosso estudo principalmente em colocações verbais e, em segundo plano, em combinações livres e expressões idiomáticas. O nosso objetivo, portanto, é analisar apenas o caráter colocacional de uma combinação entre um verbo e um substantivo, ou seja, averiguar sua condição convencional e fraseológica. Não estão no foco do trabalho outros tipos de desvio, como os de ortografia, conjugação verbal ou concordância de número, por exemplo. Dessa forma, corrigimos e complementamos somente as anotações que envolvem colocações independentemente de serem aceitáveis ou inaceitáveis. No exemplo 7, *hability* foi substituído por *ability*, em negrito, e, no exemplo 8, a palavra *robot*, em negrito e entre parênteses, foi inserida após *one*. Nesses exemplos, as combinação entre *have* e *ability,*(exemplo 7) e entre *develop* e *robot* (exemplo 8) formam colocações aceitáveis. Apesar de haver um desvio ortográfico em um exemplo e a omissão do substantivo em outro, as colocações foram empregadas com sucesso. Já no exemplo 9, identificamos que, além do desvio em *lifes*, que foi substituído por *lives*, em negrito, o verbo que faz parte da colocação com *part of something* é *be*, também em negrito, e não *make*.

7. *As one of the improvements that have already been made, we have the mobile robot, which has the ~~hability~~ **ability** to demonstrate emotions by using facial expressions.* (U1-02)

8. *The second one **(robot)** developed by Bonnie is the service robot.* (U1-03)
9. *On the other hand, the service robot is going to ~~make~~ **be** part of our ~~lifes~~ **lives** in a very near future due to the fact that to take care of contagious diseases is not so attractive to us.*

Para a correção, recorremos a um campo disponível na ferramenta de anotação que permite que o usuário edite segmentos anotados.

Dado que anotamos todo e qualquer verbo, nosso esquema de anotação, embora baseado em uma proposta de estudo já realizado, sofreu algumas modificações. Conforme fizemos a anotação, o esquema inicialmente concebido não foi capaz de englobar os dados identificados, notadamente no tocante às estruturas sintáticas. Dessa maneira, acrescentamos categorias adicionais quando consideramos sua pertinência para o estudo. A Figura 8 retrata o esquema de anotação final que utilizamos em nossa anotação. Nas subseções que seguem, descrevemos os níveis que fazem parte da estrutura hierárquica em detalhe.

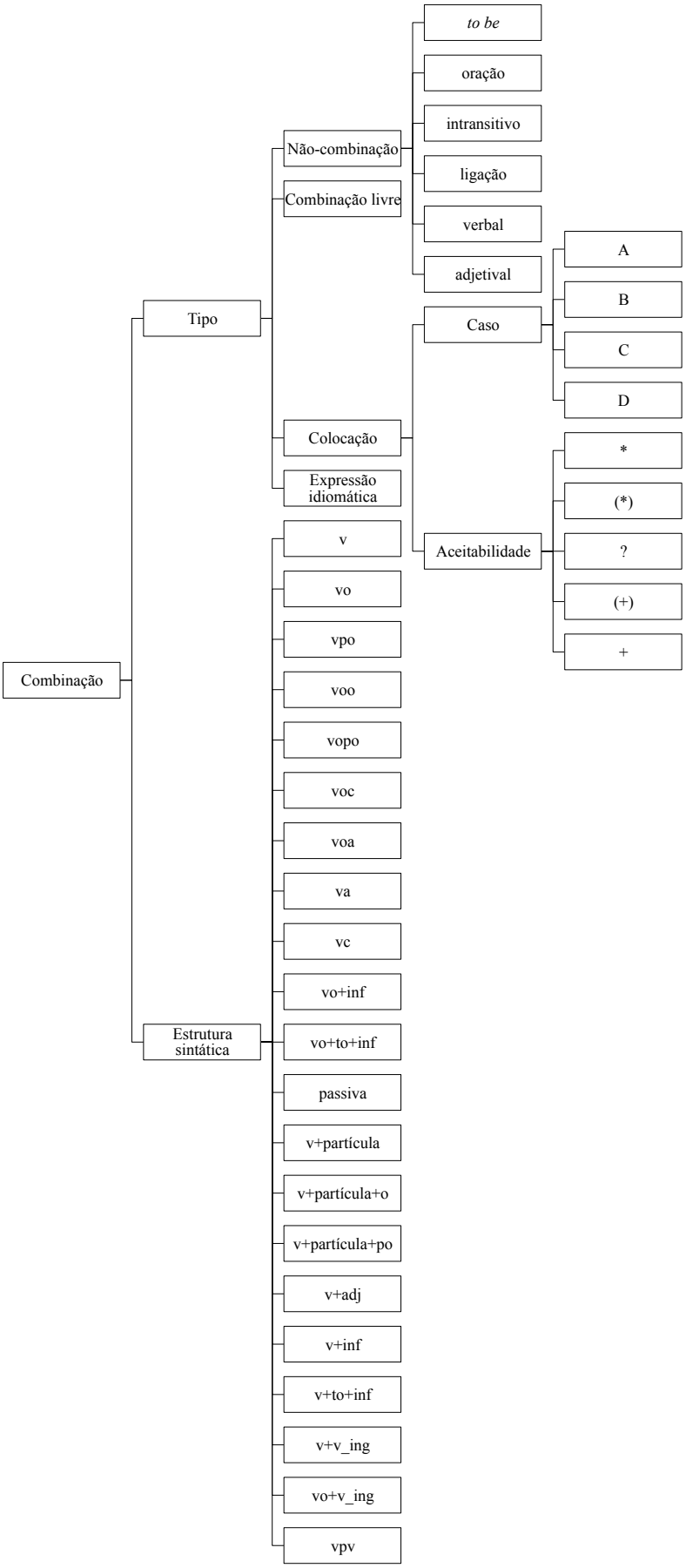

Figura 8 – Esquema de anotação adotado na pesquisa

2.3.1.1 Tipo de combinação

O primeiro nível hierárquico que constitui o esquema de anotação é o tipo de combinação (Figura 9). A decisão de anotar todos os verbos levou à adoção de uma categoria extra além das três categorias de tipo de combinação entre verbo e substantivo (combinação livre, colocação e expressão idiomática). Intitulamos essa categoria distinta não-combinação.

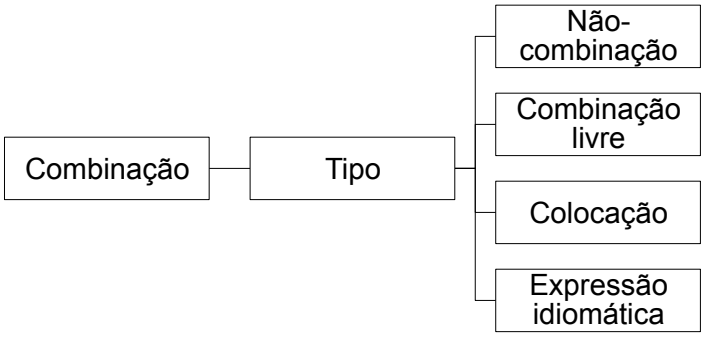

Figura 9 – Esquema de anotação: tipo de combinação

2.3.1.1.1 Não-combinação

Nesta categoria, estão classificados todos os verbos que não fazem parte dos outros três tipos de combinação entre verbo e substantivo identificados em nosso estudo. Os verbos classificados como não-combinação foram agrupados em seis subcategorias: *be*, oração, intransitivo, ligação, verbal e adjetival (Figura 10). A seguir, apresentamos a definição e exemplos dessas seis subcategorias.

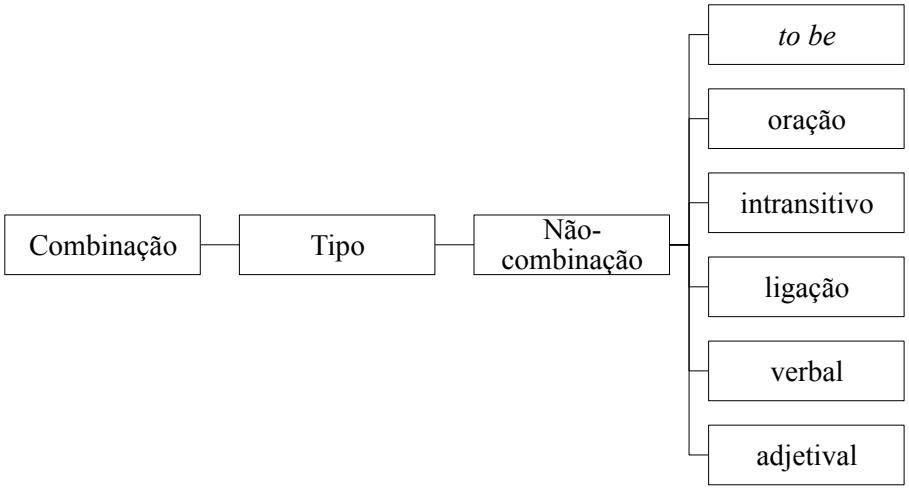

Figura 10 – Esquema de anotação: não-combinação

- *To be*: conjugações do verbo *to be* em que este não é empregado como auxiliar de outro verbo principal ou na voz passiva.

10. *Industrial robots <u>are</u> able to work fast and efficiently, eliminating wasted time by going to the toillet or even time to have lunch.* (U2-08)
11. *The dead animals, livestock mostly, were found with a pierce on their neck through which the animal blood had been sucked till there <u>was</u> none left.* (U8-02)

- Oração: verbos seguidos de uma oração (*clause*).

12. *There is no other monument able to <u>sum up</u> what Cristo Redentor means to Brazilians: it is our world-wide welcome!* (U2-18)
13. *"The local resident <u>said</u> that had seen something suspicious, so we chased him until Hamilton Road street, where he was arrested at 6am"* (U3-07)

- Intransitivo: verbos intransitivos

14. *However, we are facing huge robot developments in different areas, for instance, emotional robots, that are able to <u>smile</u>, <u>wink</u> and even <u>burp</u>!* (U1-04)
15. *Then, when I <u>am dreaming</u>, I suddenly <u>sleep</u> and recover my energy to another workday.* (U7-03)

- Ligação: verbos de ligação.

16. *A symbol of Brazilian Christianity, the statue has become an icon for Rio de Janeiro and Brazil.* (U2-04)
17. *From the first view, it seems a great and perfect project to save the young drugs dependents, but on true, she wants to explain her weakness against the drugs.* (U4-08)

- Verbal: verbos cujos complementos são outros verbos.

18. *At the moment the suspect is in police custody and they expect to have a official charge by the end of tomorrow.* (U3-07)
19. *She starts singing and he goes mad with her voice, being guided to the water, and he never returns.* (U8-03)

- Adjetival: verbos cujos complementos são adjetivos, com exceção do verbo *to be*.

20. *It is very beatiful and makes me feel comfortable and relaxed.* (U7-03)
21. *Most of the water looks very smooth, and I can see peaceful waves rolling in toward the shore.* (U7-06)

2.3.1.1.2 Combinação livre

Na categoria combinação livre, estão combinações entre verbos e substantivos em que tanto um quanto o outro, nos sentidos presentes na combinação, podem ser substituídos sem restrição arbitrária (NESSELHAUF, 2005, p. 32), ou seja, sem que haja mudança nos seus sentidos. Não há consenso, entretanto sobre o que separa combinações livres de colocações. De fato, há combinações livres que possuem caráter mais colocacional do que outras, isto é, alguns verbos estão restritos à combinação com um certo conjunto de substantivos. Em nossa pesquisa, consideramos a perspectiva do aprendiz de modo que verificamos se esse seria capaz de combinar o verbo adequado com o substantivo. Na combinação livre do exemplo

21, *read the article*, podemos substituir o verbo *read* por *see* ou *buy,* assim como podemos substituir o substantivo *article* por *magazine* ou *newspaper*, por exemplo.

22. *Hi Darth Diva, I've just read the article "Caring Celebrities or Media-Hogging opportunists?"* (U4-12)
23. *So, a man sees a mermaid brushing her hair in a river or lake.* (U8-03)

2.3.1.1.3 Colocação

Consideramos colocações aquelas combinações em que o substantivo pode ser usado sem restrição arbitrária no sentido em que está sendo empregado, porém o verbo está, até certo ponto, restrito a certos substantivos no sentido em que está sendo usado (NESSELHAUF, 2005, p. 33). Tendo em vista nosso objetivo de averiguar as colocações verbais em inglês na escrita de brasileiros, no esquema de anotação, a categoria colocação possui subcategorias que nos ajudam a compreender o fenômeno colocacional. Mais uma vez, baseamo-nos em Nesselhauf (2005) para a adoção de duas subcategorias de anotação que nomeamos uso e aceitabilidade (Figura 11).

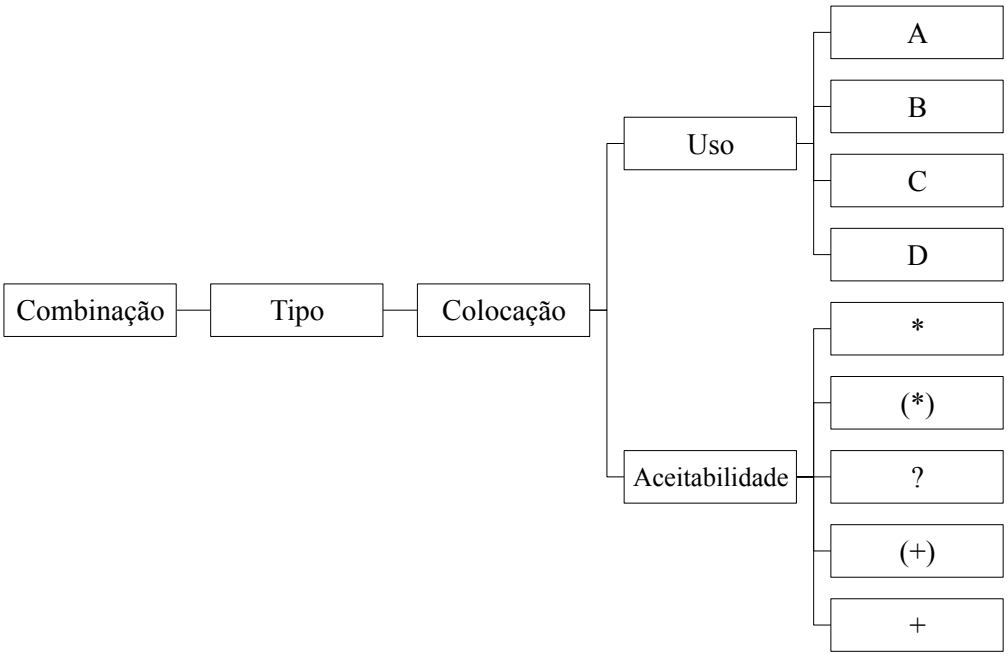

Figura 11 – Esquema de anotação: colocação

A subcategoria uso possui quatro classes que vão de A a D. Cada uma dessas classes se refere a uma situação de uso da colocação conforme descrevemos a seguir.

A. A colocação foi empregada adequadamente.

24. *Spend 30 to 50 minutes just exploring the sides.* (U5-11)
25. *At night when the stars appear, I close my eyes again and make a wish, feeling all nature power around me.* (U7-04)

B. Uma colocação diferente teria sido mais adequada em lugar da colocação empregada.

26. *As one of the improvements that have already been made, we have the mobile robot, which has the hability to demonstrate emotions by using facial expressions.* → *display/express/show emotions* (U1-02)
27. *The robots from Asia are the most advanced and they can do some interesting moviments and some tasks like washing, cooking etc.* → *make interesting movements* (U1-17)

C. Uma combinação livre (exemplo 28), uma expressão idiomática (exemplo 29) ou uma estrutura de outro tipo (exemplo 30) teria sido mais apropriada em lugar da combinação empregada.

28. *Regarding to USA, we can wait for a team of robots.* → *expect a team of robots* (U1-09)
29. *Asking is a good key to open your head.* → *have/keep an open mind* (U5-10)
30. *And whenever have a bad activities robots can do for the man.* → *there are bad activities* (U1-10)

D. A combinação empregada não pôde ser classificada por ser ininteligível.

31. *For how can we do for give our caring about these kindly people.* (U2-14)

Para verificar a condição de colocação verbal de uma combinação entre verbo e substantivo, recorremos a diversos materiais de consulta. Entre os dicionários consultados estão aqueles de língua geral *on-line Cambridge Dictionaries Online*

(CAMBRIDGE UNIVERSITY PRESS), *Collins Dictionary* (HARPERCOLLINS PUBLISHERS) e *Oxford Dictionaries* (OXFORD UNIVERSITY PRESS); e o dicionário de colocações *Oxford Collocations Dictionary for Students of English* (OXFORD UNIVERSITY PRESS, 2002). Quando os dicionários não forneceram evidências sobre certas colocações, consultamos o *Corpus of Contemporary American English* (DAVIES, 2008) e o *SkELL: Sketch Engine for Language Learning* (BAISA e SUCHOMEL, 2014).

No entanto, mesmo com a consulta a diversas fontes de pesquisa, não fomos capazes de averiguar a adequação de uma parcela das colocações encontradas no *subcorpus*. Dessa forma, a segunda subcategoria de anotação que subjaz à categoria de colocação diz respeito a um teste de aceitabilidade com falantes nativos de inglês. Um falante nativo de inglês americano e um falante nativo de inglês britânico foram consultados para julgarem as combinações não encontradas nos materiais de consulta. A tarefa dos falantes nativos foi avaliar cada combinação como aceitável, questionável ou inaceitável. Caso julgassem uma combinação como inaceitável ou questionável, deveriam, se possível, fornecer uma alternativa que lhes parecesse aceitável de acordo com sua própria intuição sobre a língua.

Em um primeiro momento, instruímos os participantes (vide Apêndice B) para que fizessem um pré-teste (vide Apêndice C), para o qual selecionamos dez frases contendo colocações que já haviam sido classificadas como adequadas com o auxílio dos materiais de consulta. Na elaboração do pré-teste, no entanto, modificamos propositalmente algumas colocações para que fossem apresentadas de modo inaceitável. Após completarem essa primeira tarefa sem dificuldades, os dois participantes avaliaram as combinações para as quais não foram encontradas evidências de aceitabilidade nos materiais de consulta.

No esquema de anotação, adotamos símbolos para identificar os resultados do teste aplicado com os falantes nativos. Além das três avaliações iniciais (aceitável, questionável e inaceitável), incorporamos avaliações intermediárias (possivelmente aceitável e possivelmente inaceitável) para contemplar os casos em que não houve consenso entre os participantes. Resumindo, uma colocação avaliada no teste de aceitabilidade pode ser classificada como:

+ aceitável;
(+) possivelmente aceitável;

? questionável;

(*) possivelmente inaceitável;

* inaceitável.

2.3.1.1.4. Expressão idiomática

Foram consideradas expressões idiomáticas aquelas combinações em que ou não são identificados sentidos independentes do verbo e do substantivo, ou então, em que ambos somente podem ser usados naquele sentido em certos contextos restritos (NESSELHAUF, 2005, p. 33), como nos exemplos 32 e 33.

32. *Please, give me a break, that's ridiculous.* (U4-02)
33. *"Trowing a brick" at someone is always a hard task to choose, because there is a possibility of commiting injustice.* (U4-09)[12]

2.3.1.2 Estruturas sintáticas

Com o objetivo de investigar a distribuição das combinações entre verbos e substantivos sintaticamente, na hierarquia do esquema de anotação, também anotamos a estrutura sintática de todos os verbos paralelamente ao tipo de combinação. Primeiramente, é importante apontar que não se trata de um trabalho minucioso de análise sintática, pois este não é o objetivo principal de nossa pesquisa. A anotação sintática superficial procura apenas enriquecer nosso estudo. Em segundo lugar, anotamos as combinações conforme a estrutura sintática em que ocorrem independentemente de apresentaram desvios. No total, identificamos 21 arranjos sintáticos distintos representados por abreviações no esquema de anotação (Figura 12). Em seguida, explicitamos as subcategorias sintáticas e as exemplificamos.

[12] Na unidade 4 do nível 16, a proposta de redação é criticar uma celebridade. Na tarefa escrita, os alunos são instruídos a usar a expressão *throw a brick* com o sentido de criticar, reprovar, repreender.

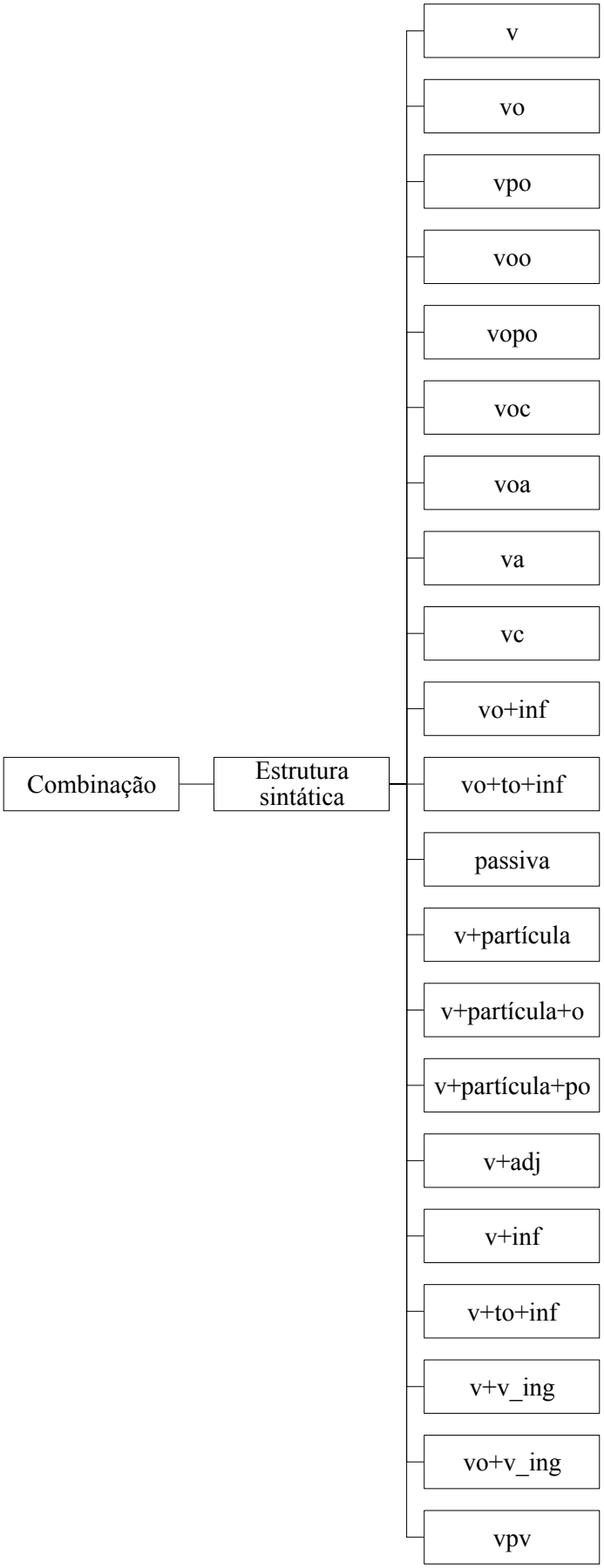

Figura 12 – Esquema de anotação: estrutura sintáticas

- v: verbo

Anotamos *pain* (exemplo 34), pois *feel the pain in my legs,* uma das correções possíveis, mantém a palavra *pain* e forma uma colocação.

34. *I walk without stop until* ***my legs*** *pain* verbo (…) (U7-10)

- vo: verbo + objeto.

Nessa subcategoria, além de casos em que, como no exemplo 35, o verbo (*make*) precede o substantivo (*list*), também incluímos a ordem inversa (exemplo 36), em que o substantivo (*cause*) precede o verbo (*defends*) devido à oração relativa.

35. *If you prefer, you can make* verbo *a list* objeto. (U5-06)
36. *Our vote is on her because of her public image and the cause* objeto *she defends* verbo. (U4-03)

- vpo: verbo + preposição + objeto

37. *In the future it will be able to visit quarantine patients that suffer* verbo *from* preposição *contagious diseases* verbo. (U1-12)

- voo: verbo + objeto + objeto

38. (...) *I have the feeling that someone is giving* verbo *me* objeto *a massage* objeto (...) (U7-05)

- vopo: verbo + objeto + preposição + objeto

39. *As a result, some artificial dogs were invented and in USA some of them take* verbo *part* objeto *in* preposição *international competitions* objeto. (U1-16)

- voc: verbo + objeto + complemento

40. (...) *service robots which promise to set* verbo *housewives and maids* objeto *free* complemento *from those boring tasks of having to take care of the house.* (U1-07)

- voa: verbo + objeto + locução adverbial

41. *However, the development has been time consuming because it is difficult to get the robots to make* verbo *even small movements* objeto *in a natural way* locução adverbial, *like humans do.* (U1-18)

- va: verbo + locução adverbial

42. *It was the 15th railroad to be built in the country and took* verbo *from 1907 to 1912* locução adverbial *to be built its extention is 366km in Amazonia it connects Porto Velho to Guajara Mirim cities founded by Madeira Mamore.* (U2-09)

- vc: verbo + complemento

43. *While the robots created in US have* verbo *as finality* complemento *the sport and in the future the scientists wants to develop a robot with AI.* (U1-17)

- vo+inf: verbo + objeto + verbo infinitivo

44. *Just leave* verbo *your ideas* objeto *flow* verbo infinitivo *and take notes of them.* (U5-09)

- vo+to+inf: verbo + objeto + *to* + verbo infinitivo

45. *If you ask* verbo *a group of people* objeto *to describe* verbo infinitivo *a robot, most of them would say they are machines that look like a human.* (U1-20)

- passiva: voz passiva

46. *The police officer in charge, Christopher Linestone affirms the arrest was made* voz passiva *after a local resident called them..* (U3-07)

- v+partícula: verbo + partícula (*phrasal verb*)

  47. *To sum* verbo *up* partícula*, I want to throw a lot of bricks at her, in order to protect the young people.* (U4-09)

- v+partícula+o: verbo + partícula (*phrasal verb*) + objeto

  48. *This robot is programmed to walk around hospital wards, picking* verbo *up* partícula *the trash* objeto *and help with monitoring and surveillance systems.* (U1-13)

- v+partícula+po: verbo + partícula (*phrasal verb*) + preposição + objeto

  49. *After some minutes, I go* verbo *ahead* partícula *with* preposição *my walk* objeto *and find the ruins of an old castle in front of me.* (U7-09)

- v+adj: verbo + adjetivo

  50. *The nightmare of the collapse became* verbo *popular* adjetivo *in people's lives.* (U2-03)

- v+inf: verbo + verbo infinitivo

  51. *From Africa, the focus is built a service robot to help* verbo *take* verbo infinitivo *care of people with contagious diseases.* (U1-09)

- v+to+inf: verbo + *to* + verbo infinitivo

  52. *It seems he's not in the good shape he likes* verbo *to* to *show* verbo infinitivo *and demonstrate, by eating in a proper manner.* (U4-04)

- v+v_ing: verbo + verbo com *ing*

53. *As the investigation kept* verbo *going* verbo com *ing*, *the police collected enough data to make a publicity campaign, offering a $300 reward for any information that could lead to the mans arrest.* (U3-09)

- vo+v_ing: verbo + objeto + verbo com *ing*

54. *By 2050, he expects to have* verbo *his robots* objeto *playing* verbo com *ing with real humans.* (U1-18)

- vpv_ing: verbo + preposição + verbo com *ing*

55. *This robot is programmed to walk around hospital wards, picking up the trash and help* verbo *with* preposição *monitoring* verbo com *ing and surveillance systems.* (U1-13)

2.3.2 Anotação

Uma vez definido o esquema de anotação, os arquivos inseridos no *UAM CorpusTool* podem ser anotados individualmente. Para anotar segmentos, basta selecionar as partes que devem ser anotadas. Como partimos da anotação automática de verbos, tivemos que, na maior parte dos casos, somente completar a seleção feita pelo programa de modo que abrangesse as combinações. Após serem selecionados, os segmentos são classificados a partir da hierarquia do esquema de anotação. Em uma lista, o usuário escolhe a categoria a que o segmento pertence e o programa segue para o próximo nível hierárquico. Na Figura 13, temos um exemplo de uma colocação (*make an appointment*) que já foi anotada para estrutura sintática (vo), e uso (A), mas que ainda precisa ser anotada para aceitabilidade.

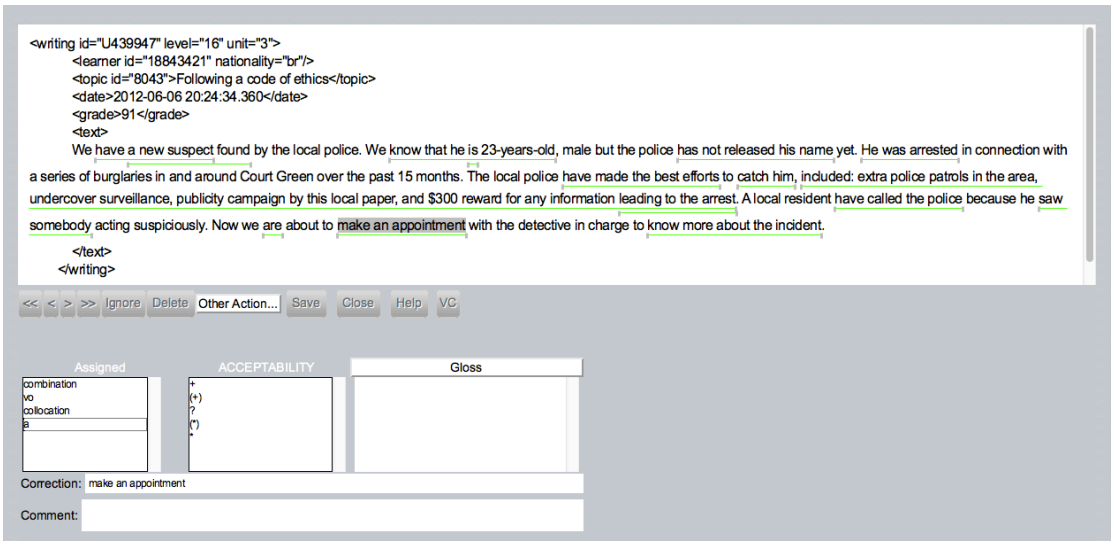

Figura 13 – Captura de tela do *UAM CorpusTool*

No presente capítulo, estabelecemos os métodos de pesquisa que guiaram nossa análise. Uma vez que analisamos dados já disponíveis, isto é, não tivemos de compilar um *corpus* para nossa pesquisa, apresentamos nosso instrumento de análise. Como pode ser notado, nosso trabalho está amplamente baseado no uso do programa computacional *UAM CorpusTool*, que permite a criação de uma esquema de anotação para então anotarmos o *corpus*. Diferentemente de outras ferramentas computacionais, o programa também fornece os resultados da anotação, os quais serão apresentados no próximo capítulo.

CAPÍTULO 3. RESULTADOS

Neste capítulo, apresentamos os resultados de nossa pesquisa. Primeiramente, são apresentados os resultados quantitativos (3.1) relativos à dispersão dos dados. Na subseção seguinte (3.2), apresentamos resultados sobre os tipos de desvio encontrados nas combinações.

3.1 Resultados quantitativos

Além da anotação, o *UAM CorpusTool* fornece levantamentos quantitativos sobre o esquema de anotação, o que permite a constatação da ocorrência dos elementos anotados. O Gráfico 1 exibe a quantidade de segmentos anotados como pertencentes a cada um dos quatro tipos de combinação estabelecidos em nosso estudo (não-combinação, combinação livre, colocação e expressão idiomática).

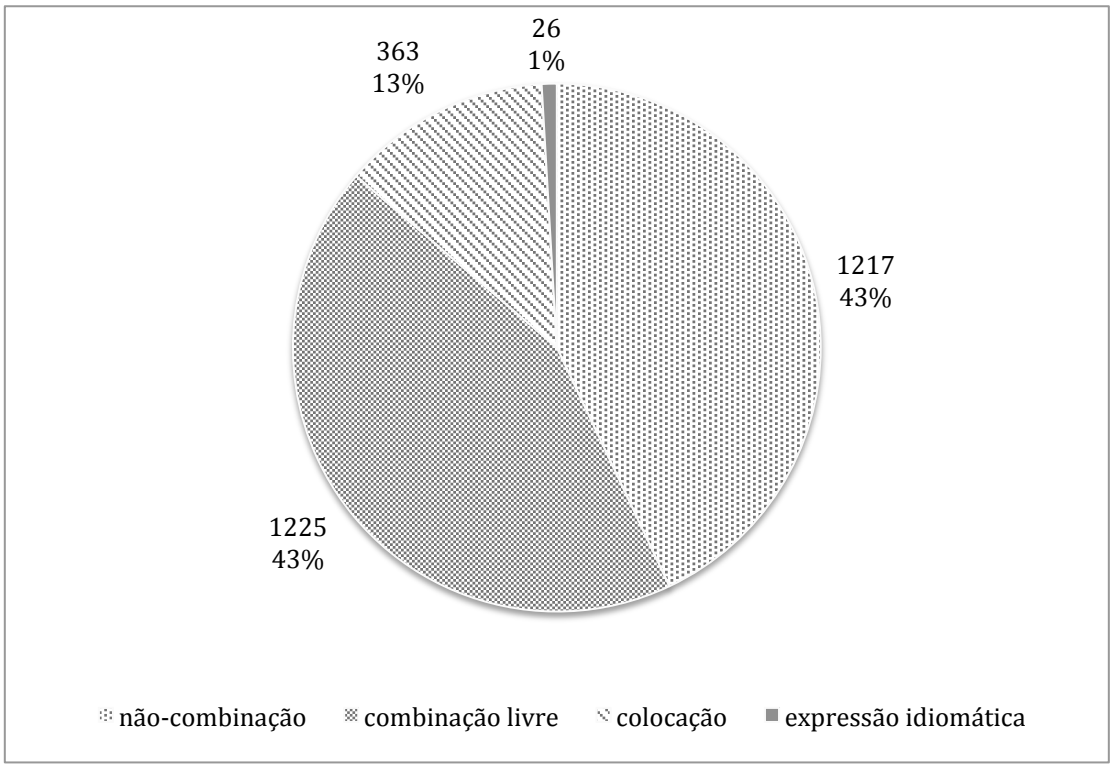

Gráfico 1 – Frequência de tipos de combinação

Podemos observar que os tipos mais frequentes, a não-combinação e a combinação livre, possuem quantidades similares em nossa amostra de dados. Contudo, devemos lembrar que a não-colocação é a categoria que engloba todos os casos que não podem ser classificados em nenhum dos outros três tipos de combinação. Portanto, se compararmos a frequência da não-combinação (43%) com a soma das demais categorias (57%), podemos afirmar que, em nosso estudo, combinações efetivas entre um verbo e um substantivo são ligeiramente mais comuns que ocorrências em que um verbo se combina com outra categoria gramatical. Em outras palavras, os aprendizes analisados em nosso *subcorpus* tendem a combinar um verbo a um substantivo em quantidade sutilmente superior a produções em que não ocorrem tal combinação. Na Tabela 4, são exibidas as subcategorias de não-combinação e seus respectivos números.

Tabela 4 – Quantidade de não-combinações

Não-combinação	**Quantidade**
to be	543
oração	133
intransitivo	348
ligação	9
verbal	132
adjetival	51
Total	1216

À parte da categoria não-combinação, verificamos a distribuição das combinações entre verbos e substantivos identificadas em relação às categorias combinação livre, colocação e expressão idiomática (Gráfico 2). Constatamos que a combinação livre (76%) é o tipo de combinação mais frequente. A colocação, em segundo lugar, representa aproximadamente um quinto das combinações estudadas, ou seja, um aprendiz tende a combinar colocacionalmente um verbo e um substantivo a cada cinco combinações que produz.

É relevante apontar que a proporção de expressões idiomáticas, por sua vez, é pouco expressiva em nossa análise. Dentre as 26 expressões idiomáticas encontradas,

12 são combinações do verbo *throw* com o substantivo *brick*. Em uma das propostas de redação do *Englishtown*, os aprendizes são instruídos a empregar a expressão *throw a brick* (*at somebody*) com o sentido de criticar alguém. Uma pesquisa no *Corpus of Contemporary American English* (DAVIES, 2008) e no *British National Corpus* (DAVIES, 2004), mesmo encontrando ocorrências da combinação, não resultou em sentidos idiomáticos como o proposto pela atividade do curso *on-line*.

Apesar do destaque às vezes dado ao ensino de expressões idiomáticas, os resultados aqui apresentados mostram que a incidência dessa categoria na produção efetiva de aprendizes brasileiros de inglês é baixa. Assim como as expressões idiomáticas, as colocações estão entre as unidades linguísticas convencionais que também tem de ser aprendidas uma a uma, pois não há regra que as rege (TAGNIN, 2013, p. 15). As colocações verbais, que, assim como as expressões idiomáticas, também tem caráter fraseológico, muitas vezes não possuem tanta relevância quanto expressões idiomáticas no ensino. Uma busca por publicações que possuem 'expressão(ões) idiomática(s)' em seu título em sites de livrarias especializadas em ensino de línguas, por exemplo, gera mais resultados do que materiais cujos nomes incluem 'colocação(ões)'.

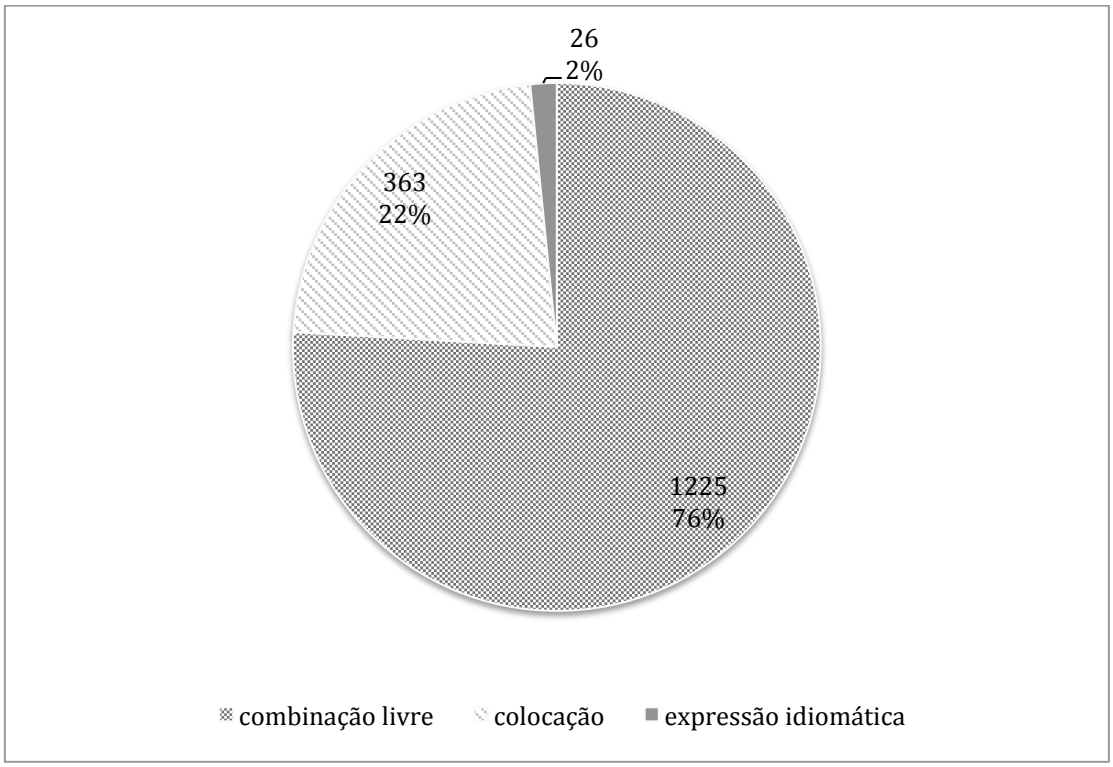

Gráfico 2 – Frequência de três combinações: combinação livre, colocação e expressão idiomática

Outro levantamento que também realizamos foi a produção de colocações de acordo com os textos e os aprendizes (Tabela 5). Em média, cerca de três colocações ocorrem em cada texto escrito pelos aprendizes. Somente um aprendiz (18843375), representado no *subcorpus* com apenas um texto, não usou nenhuma colocação. A análise de sua redação mostrou que o fato de ter escrito seu texto em formato de tópicos pode ter contribuído para que não produzisse colocações. Três aprendizes (18507049, 18727214, 19569225) tiveram todas as suas redações do nível 16 analisadas, isto é, três indivíduos completaram as oito unidades didáticas do último nível do *Englishtown* de modo a satisfazer os critérios de seleção dos textos adotados em nosso estudo. Um desses aprendizes (19569225) foi o que mais produziu colocações (30). No entanto, comparativamente à média, o aprendiz 22000339 produziu dez colocações, mesmo tendo contribuído com somente um texto para o *subcorpus*.

Tabela 5 – Quantidade de colocações e textos por aprendiz

Código do aprendiz	**Quantidade de colocações**	**Quantidade de textos**	**Média colocação/texto**
18507049	20	8	2,5
18550698	13	6	2,2
18622336	9	2	4,5
18645692	3	4	0,8
18666667	7	2	3,5
18699167	21	7	3,0
18706124	18	3	6,0
18727214	17	8	2,1
18843375	0	1	0,0
18843421	3	3	1,0
18996301	25	6	4,2
19028537	8	2	4,0
19165473	24	7	3,4
19199254	13	2	6,5
19411255	6	1	6,0
19441112	1	2	0,5
19539220	5	1	5,0
19569208	28	7	4,0

19569225	30	8	3,8
21024468	6	1	6,0
21096192	12	7	1,7
21549924	3	1	3,0
21858287	3	3	1,0
21914218	7	2	3,5
21983270	10	2	5,0
22000339	10	1	10,0
22002001	12	2	6,0
22117882	18	5	3,6
22672997	4	1	4,0
24415789	6	2	3,0
24445358	16	5	3,2
24458791	5	1	5,0
Total	363	113	3,2

No Gráfico 3, podemos contemplar a diferença na distribuição de colocações por aprendizes.

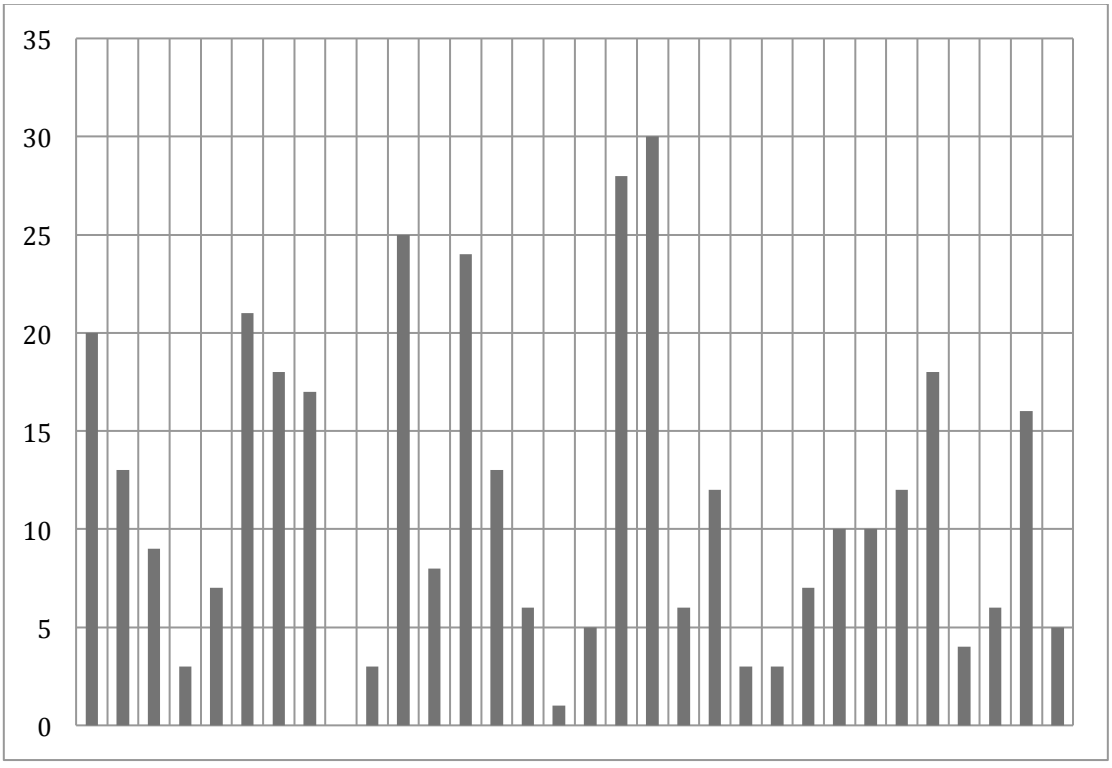

Gráfico 3 – Distribuição de colocações por aprendiz

Uma vez que consideramos, em um primeiro momento, analisar as colocações a partir dos verbos mais frequentes em sua forma lematizada, também julgamos conveniente fazer o levantamento dos verbos que mais ocorrem nas colocações identificadas. A Tabela 6 exibe os seis verbos mais frequentes nas colocações verbais de nosso *subcorpus* (*have, take, make, do, give, spend, hold*) seguidos de suas quantidades e exemplos. Esse resultado nos mostra que, se tivéssemos analisado somente as colocações com esses verbos (158), teríamos analisado menos da metade do total de colocações identificadas semiautomaticamente (363).

Tabela 6 – Quantidade de colocações por verbo lematizado

Lema	**Quant. de colocações**	**Exemplos**
HAVE	39	*HAVE a contagious disease; HAVE lunch*
TAKE	38	*TAKE years; TAKE notes*
MAKE	27	*MAKE an appointment; MAKE a list*
DO	21	*DO activities, DO a job*
GIVE	12	*GIVE an interview; GIVE an example*
SPEND	11	*SPEND the day; SPEND money*
HOLD	10	*HOLD under custody; HOLD a gun*
Total	158	

Na Tabela 7, encontramos os resultados quantitativos em relação às estruturas sintáticas em que ocorrem os quatro tipos de combinação. É importante salientar que todas as estruturas sintáticas estão agrupadas na tabela de maneira a contemplar todas as categorias definidas para tipo de colocação. Sabemos que algumas combinações não poderiam ocorrer em certas estruturas sintáticas. Uma colocação verbal, segundo nossa definição, não seria, por exemplo, classificada como 'verbo+adjetivo'. Tampouco ocorreria de uma colocação, que é justamente um tipo de combinação entre verbo e substantivo, ocorrer como apenas 'verbo'. Entretanto, em nossa anotação, anotamos os segmentos segundo a sintaxe em que ocorreram, mesmo que essa parecesse inadequada. Em um dos textos (U7-10), por exemplo, um aprendiz empregou a palavra *pain* como verbo com o sentido de *feel pain/hurt*, portanto anotamos esse segmento como uma inadequação colocacional já que poderia ser substituído por uma colocação. Dessa forma, mesmo nas categorias sintáticas da

tabela em que, em princípio, não ocorreria certa combinação, exibimos também o número 0 para mostrar que realmente não houve nenhum desvio.

Tabela 7 – Quantidade de colocações por estrutura sintática

	Não-combinação	**Combinação livre**	**Colocação**	**Expressão idiomática**
v	850	1	1	1
vo	132	773	243	12
vpo	9	131	12	1
voo	0	5	4	1
vopo	0	18	22	3
voc	0	22	5	0
voa	0	1	1	0
va	5	5	1	1
vc	0	0	1	0
vo+inf	0	7	2	1
vo+to+inf	0	14	0	0
vo+v_ing	0	2	0	0
passiva	6	207	58	0
v+partícula	33	0	0	2
v+partícula+o	4	35	13	0
v+partícula+po	0	2	0	2
v+adj	521	1	0	1
v+inf	6	0	0	0
v+to+inf	100	0	0	0
v+v_ing	16	1	0	0
vVpv_ing	4	0	0	1
Total	12176	1225	363	26

Em nossa primeira análise com o auxílio do material de consulta já mencionado, identificamos 258 colocações aceitáveis. Em contrapartida, 105 colocações não estavam na mesma situação. Dessas últimas, 53 eram claramente inaceitáveis e foram categorizadas como tal. Como não pudemos encontrar evidências nos materiais de consulta, as outras 52 colocações fizeram parte do teste de

aceitabilidade aplicado aos falantes nativos de inglês. Após o teste, 16 colocações foram, na verdade, julgadas como aceitáveis, ou seja, 36 não se encaixaram nessa categoria. A soma das 53 colocações classificadas inicialmente como inaceitáveis com as 36 problemáticas resultadas do teste nos deixou com 89 colocações que apresentam problemas contra 274 colocações aceitáveis (Gráfico 4). Podemos observar, portanto, que aproximadamente um quarto das colocações são empregadas de modo inadequado. Em outras palavras, podemos afirmar que, em média, em cada quatro colocações verbais escritas por um aprendiz brasileiro de inglês do nosso *subcorpus*, uma está empregada inadequadamente.

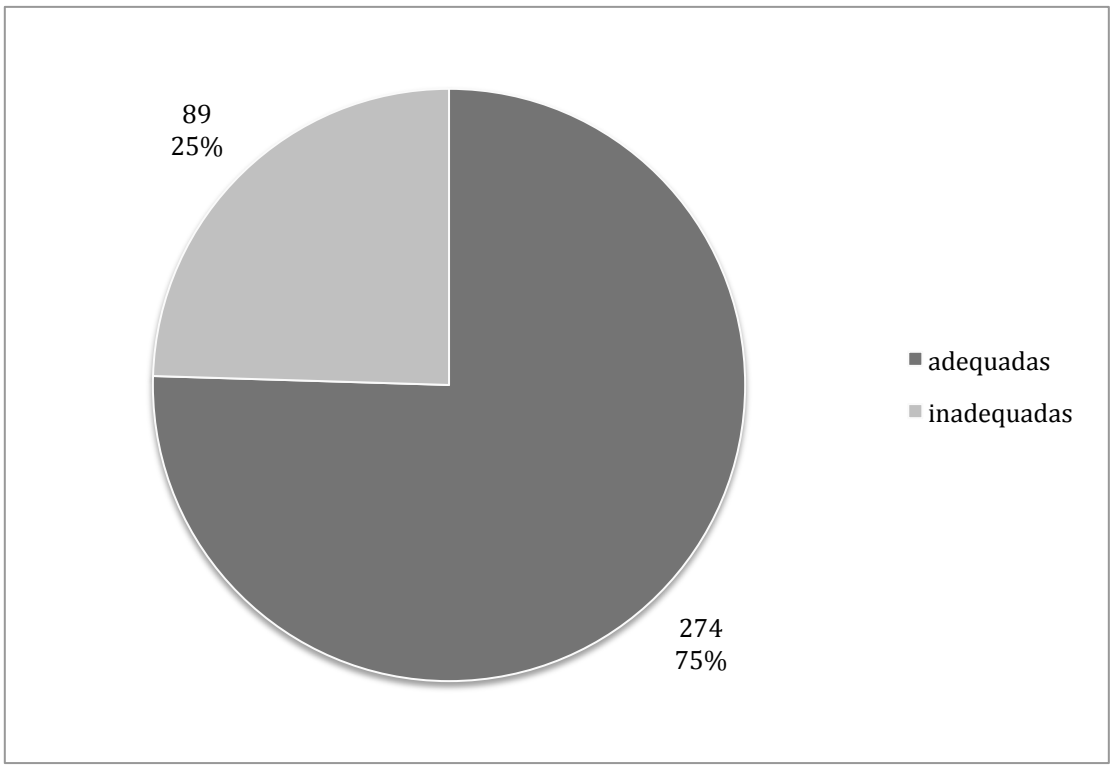

Gráfico 4 – Frequência de colocações adequadas e inadequadas

O teste de aceitabilidade com os falantes nativos de inglês também permitiu que completássemos nosso esquema de anotação com as subcategorias uso e aceitabilidade da colocação. No Gráfico 5, verificamos a incidência das colocações conforme os quatro usos estabelecidos (A, B, C e D). O uso mais frequente (A) foi aquele em que a colocação foi empregada adequadamente. Em seguida, no entanto menos expressivo, vemos o uso B, com 52 ocorrências de uma inadequação colocacional, isto é, uma outra colocação poderia ter sido empregada no lugar da combinação eleita pelo aprendiz. O uso C, por sua vez, possui 37 ocorrências de

combinações livres, expressões idiomáticas ou expressões de outro tipo que poderiam ter sido usadas no lugar dos segmentos empregados. Por fim, encontramos somente um caso que não foi classificado em nenhuma das categorias anteriores por não ter sido compreendido (*can we do for give our caring about these kindly people*), e cujo exemplo já foi mencionado anteriormente (exemplo 31).

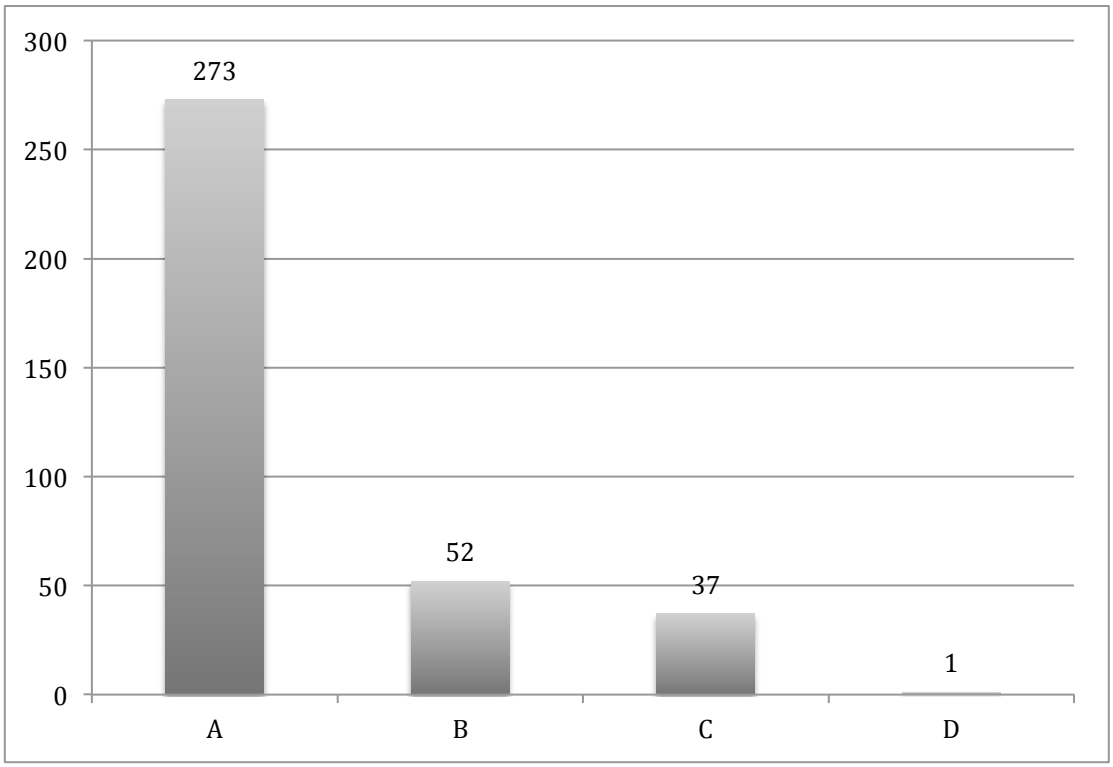

Gráfico 5 - Distribuição de colocações por uso

Para determinar as categorias intermediárias em nossa escala entre aceitável [+], questionável [?] e inaceitável [*], recorremos à seguinte classificação (Quadro 6), em que criamos as categorias 'possivelmente aceitável' [(+)] e 'possivelmente inaceitável' [(*)], representadas pelos mesmos símbolos das categorias aceitável e inaceitável, porém entre parênteses, para os casos em que não houve consenso entre os falantes nativos.

Falante Am/ Falante Br	+	?	*
+	+	(+)	?
?	(+)	?	(*)
*	?	(*)	*

Quadro 6 – Classificação de aceitabilidade

Em relação à aceitabilidade das colocações (Gráfico 6), constatamos que a maioria é tida como aceitável. Somente uma colocação foi classificada como possivelmente aceitável. Uma parcela (70) foi tida como inaceitável, enquanto pequenas quantidades foram julgadas como possivelmente inaceitável (9) e questionável (9).

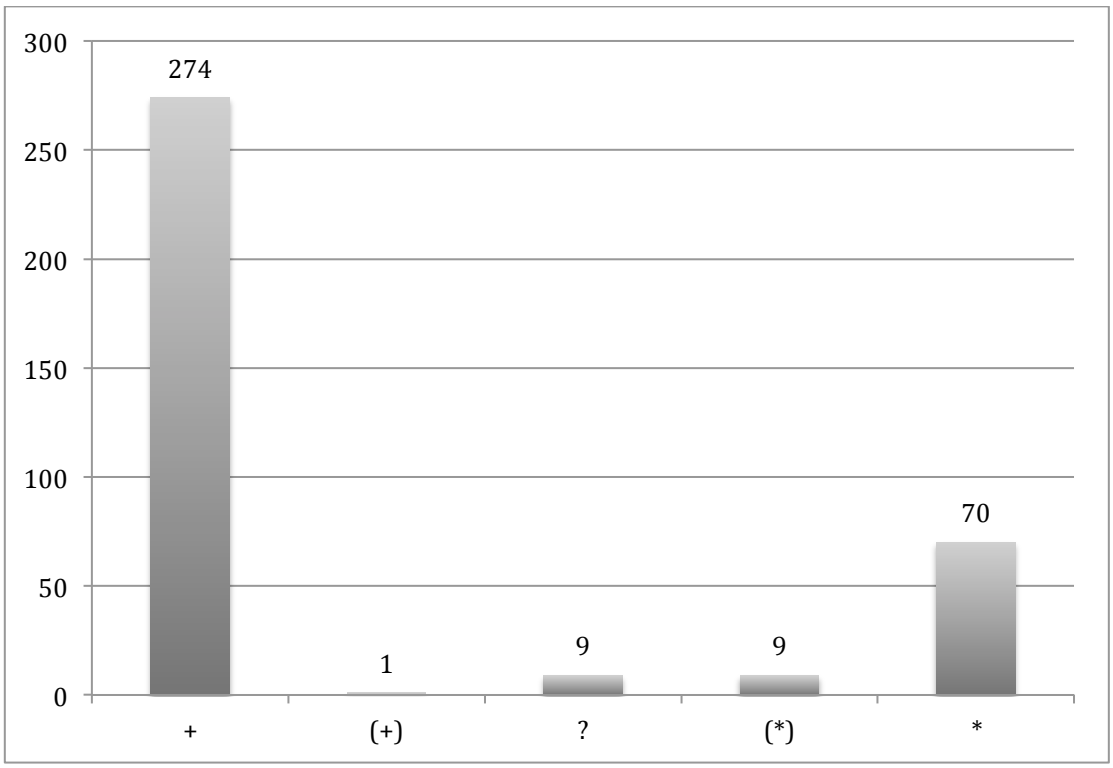

Gráfico 6 - Distribuição de colocações por aceitabilidade

No Apêndice D, encontram-se as colocações que foram classificadas como aceitáveis.

3.2 Análise das colocações

É relevante recordar que os alunos que cursam o *Englishtown* têm à sua disposição um modelo de texto para escreverem suas redações. Tendo conhecimento desse fato, em um primeiro estudo pormenorizado das colocações, pressupomos que o número de colocações julgadas como aceitáveis poderia ter sido resultado de cópia do modelo de redação. Comparamos as colocações encontradas nos modelos com as colocações produzidas pelos aprendizes e verificamos que houve pouca sobreposição dos dados. Nos raros casos em que a mesma colocação foi encontrada tanto na redação de um aprendiz quanto no modelo, o aprendiz adaptou a estrutura da colocação. Além disso, em mais de uma proposta de redação, os aprendizes são instruídos a utilizarem colocações que aparecem nos modelos de texto.

Na presente subseção, analisamos as 89 colocações não-aceitáveis de modo a criar uma tipologia de desvios combinatórios (Tabela 8). É também nesta seção que apresentamos as correções obtidas por meio do material de consulta e das sugestões dos falantes nativos para as colocações que não foram tidas como aceitáveis. Por fim, identificamos a incidência de desvios entre as estruturas sintáticas.

Tabela 8 – Distribuição de desvios combinatórios

Desvio	**Quant.**	**%**
I. Verbo inadequado	65	73,0
II. Substantivo inadequado	9	10,1
III. Verbo e substantivo adequados com estrutura inadequada	8	9,0
IV. Verbo e substantivo inadequados	3	3,4
V. Verbo como substantivo	2	2,2
VI. Substantivo como verbo	1	1,1
VII. Inclassificável	1	1,1
Total	89	

3.2.1 Desvio I: verbo inadequado

O tipo de desvio mais comumente encontrado em nossa análise é aquele em que o verbo é inapropriado enquanto o substantivo é empregado adequadamente (73%). Tanto nos usos classificados como B e C encontramos esse tipo de desvio, ou seja, nos casos em que uma colocação foi usada na correção (B) e nos casos em que combinações livre, expressões idiomáticas ou expressões de outros tipos servem como correção (C).

Esse tipo de desvio decorre justamente da natureza da colocação verbal segundo a definição adotada neste trabalho. Uma vez que o verbo é arbitrariamente restrito a certos substantivos, um aprendiz pode não conhecer suas restrições, e acaba empregando o verbo de maneira insatisfatória. Isso quer dizer que no processo de combinação entre um verbo e um substantivo, o aprendiz conhece o sentido do substantivo, mas no momento de eleger o verbo, sua escolha é inadequada, pois não domina o conhecimento colocacional.

Devido à alta incidência desse tipo de desvio, cabe fazer ainda outra subdivisão para classificar os tipos de desvios em que o verbo foi usado de forma inapropriada (Tabela 9).

Tabela 9 – Distribuição de desvio combinatório I: verbo inadequado

Desvio	**Quant.**	**%**
i. Influência do português	34	52,3
ii. Verbo similar	11	16,9
iii. *Do* e *make*	11	16,9
iv. Verbo diverso	6	9,2
v. *Take*	3	4,6
Total	65	

Na tabelas que seguem, podemos encontrar a combinação originalmente usada na redação do aprendiz e sua correção conforme o texto em que ocorre, além das classificações de uso e aceitabilidade. A correção apresenta a substituição para o desvio sublinhado na combinação. Quando as correções foram sugeridas pelos falantes nativos, indicamos entre parênteses, se o falante é britânico (Br) e/ou americano (Am). Após cada tabela, apresentamos alguns exemplos representativos do subtipo de desvio de verbo inadequado.

Na maioria dos casos de desvios provocados por verbos inadequados, os aprendizes parecem ser influenciados pela sua primeira língua, o português, no momento da escolha do verbo (Tabela 10). Em português, por exemplo, existe a colocação verbal 'denunciar o suspeito', mas em inglês *denounce*, apesar de ser um verbo cognato de 'denunciar', não forma uma colocação com *suspect*. No exemplo 56, portanto, *denounced the suspect* foi corrigido por um falante nativo para *informed on the suspect*. Temos outro exemplo do mesmo tipo em 57. Dessa vez, *use* foi empregado com *corset*, já que em português podemos 'usar' peças de roupa, mas, em inglês, combinamos o verbo *wear* com peças de vestuário, acessórios e alguns outros elementos no corpo de um indivíduo. Nesse exemplo, corrigimos *he's been allegedly using corsets* para *he's been allegedly wearing corsets*.

56. *Currently, the man is being held in custody for police questioning, at the local police department, where the undercover agent and the witness who denounced the suspect, will participate in the interrogatory*. (U3-09) → *informed on the suspect*
57. *Besides that, pictures don't lie, rumors point that he's been allegedly using corsets to keep in his ever-growing stomach!* (U4-05) → *he's been allegedly wearing corsets*

Tabela 10 – Desvio I: verbo inadequado (i. influência do português)

Texto	Combinação	Correção	Uso	Aceit.
U2-16	*purpose a gorgeous view*	*has*	B	*
U4-04	*uses spanx*	*wears*	B	*
U4-05	*he's been allegedly using corsets*	*wearing*	B	*
U5-09	*leave your ideas flow*	*let*	B	*
U2-03	*completed 71 years of age*	*turned*	B	*
U1-15	*patients victimized*	*infected*	B	?
U6-01	*have conditions*	*meet*	B	(+)
U1-16	*demonstrates a really similar appearence*	*has*	B	*
U2-14	*to put in mind that we are a*	*having in mind* (Br)*;*	B	*

	result of this process	*thinking* (Am)		
U3-09	*denounced the suspect*	*informed on* (Br)	C	?
U7-13	*could pass something similar*	*experience*	C	*
U8-08	*No one fatal victim was related*	*reported*	C	*
U8-08	*relate about this creatures*	*report*	C	*
U1-09	*can wait for a team of robots*	*expect*	C	*
U1-10	*have a bad activities*	*there are*	C	*
U1-15	*can also (...) know Major Pettigrew*	*also know* (Br); *see* (Am)	C	(*)
U2-14	*put this* (point)	*say* (Br)*; state* (Am)	C	*
U2-18	*contemplating* (...) *Central Rio*	*overlooking* (Br)*; including* (Am)	C	*
U2-18	*contemplating* (...) *Tijuca Mountains*	*overlooking* (Br)*; including* (Am)	C	*
U2-18	*contemplating* (...) *Maracana Stadium*	*overlooking* (Br)*; including* (Am)	C	*
U2-18	*contemplating* (...) *Bay of Guanabara*	*overlooking* (Br)*; including* (Am)	C	*
U2-18	*contemplating* (...) *Pao de Acucar (The Sugar Loaf)*	*overlooking* (Br)*; including* (Am)	C	*
U2-18	*contemplating* (...) *Lagoa Rodrigo de Freitas*	*overlooking* (Br)*; including* (Am)	C	*
U2-18	*contemplating Copacabana's Beach*	*overlooking; including Copacabana Beach*	C	*
U5-03	*deliver my book*	*hand in* (Br)	C	?
U7-04	*not exist anything else*	*there is not*	C	*
U7-12	*does not exist anything like*	*there is not*	C	*
U2-07	*has 38 meters*	*is 38 meters tall* (Am/Br)	C	*
U2-17	*had 820 meters as length*	*was 820 meters in length* (Br)*; long* (Am)	C	(*)
U2-17	*having 75 meters in height*	*are 75 meters in height* (Br)*; each 75 meters tall* (Am)	C	(*)
U1-09	*Give us your thoughts*	*share*	C	*

U8-08	*Big foot was retrated*	*portrayed; pictured*	C	*
U8-08	*relate the exactly same creature*	*report*	C	*
U1-11	*the planet will be carry for artificial hands*	*cared for by*	C	*

É importante apontar que, muitas vezes, o emprego de um verbo inapropriado não é completamente aleatório (Tabela 11). Nos exemplos abaixo, os verbos empregados pelos aprendizes são sinônimos dos verbos que efetivamente formam colocações com os substantivos empregados. No exemplo 58, *display, express e show* são similares a *demonstrate*, assim como, no exemplo 59, *turning* e *shifting* poderiam ser intercambiáveis em outros contextos, porém do ponto de vista colocacional, somente alguns são aceitáveis.

58. *As one of the improvements that have already been made, we have the mobile robot, which has the hability to demonstrate emotions by using facial expressions*. (U1-02) → *display; express; show emotions*
59. *We can also, on the other hand, know Major Pettigrew who used to work on military and industrial robots field now turning his focus to build a sports robot able to play with a humane team*. (U1-15) → *shifting his focus*

Tabela 11 – Desvio I: verbo inadequado (ii. verbo similar)

Texto	**Combinação**	**Correção**	**Uso**	**Aceit.**
U1-02	*demonstrate emotions*	*display; express; show*	B	*
U1-15	*turn the focus*	*shift* (Br)*; the focus can be turned* (Am)	B	?
U1-15	*turning his focus*	*shifting*	B	?
U2-20	*the damages were completely restored*	*fixed* (Br)*; the damage was restored* (Am)	B	(*)
U3-07	*releasing the charge*	*dismissing; throwing out*	B	*
U3-11	*questions that are very difficult to explain*	*answer*	B	?
U4-06	*had been troubles with alcohol and drugs*	*had trouble*	B	*

U1-04	*are facing huge robot developments*	*seeing* (Br)*; experiencing; witnessing* (Am)	C	(*)
U1-24	*are going to depict them for you*	*present* (Br)*; highlight* (Am)	C	(*)
U2-11	*have in their minds Rio de Janeiro beaches, the statue of Corcovado, the beautiful hills of Sugar Loaf, the crowded Copacabana beach or even the energy from Maracana football stadium*	*think about* (Am)	C	?
U2-18	*The statue was built up*	*erected; built*	C	(*)

A língua portuguesa também influencia a escolha do verbo nas colocações com *do* e *make* (Tabela 12), pois esses dois verbos podem ser traduções de 'fazer'. Em português, o verbo 'fazer' faz parte de diversas colocações que, em inglês, muitas vezes são empregadas com *do* e *make*. No exemplo 60, *research made* foi empregado em vez de *research done*, já que em português temos a colocação 'fazer uma pesquisa'. Contrariamente, no exemplo 61, um aprendiz usa *do some interesting movements*, quando a colocação adequada é *make a movement*. Já no exemplo 62, temos um caso em que, baseado no português 'fazer força', o aluno empregou *do some force*, porém em inglês os verbos que se colocam com *force* são *use* e *resort to*.

60. *In according with a research made by US Statistics in 50 countries in five different continents the Asia has the 32% of the robots, the North America has 18% and Africa has just 8%*. (U1-17) → *research done*
61. *The robots from Asia are the most advanced and they can do some interesting moviments and some tasks like washing, cooking etc., but the scientists want much more and say that they are ready to create a robot able to feel emotions like sadness and happiness*. (U1-17) → *make some interesting movements*
62. *I can hear the oceans noise and the water warm crush against my legs, I have to do some force to walk, but is still good*. (U7-10) → *use; resort to force*

Tabela 12 - Desvio I: verbo inadequado (iii. *do* e *make*)

Texto	Combinação	Correção	Uso	Aceit.
U1-10	*activities which is bad for integrity of human being is good to be made*	*done*	B	*
U1-16	*is going to make part of our lifes*	*be*	B	*
U1-17	*a research made*	*done*	B	*
U1-17	*can do some interesting moviments*	*make*	B	*
U3-07	*will also make several interviews*	*do; conduct; hold*	B	*
U3-09	*make a publicity campaign*	*start; run; create; develop*	B	(*)
U5-07	*have done any mistakes*	*made*	B	*
U6-03	*do a fill out an application*	*file; submit*	B	*
U7-10	*do some force*	*use; resort to*	B	*
U8-07	*This action could not be made*	*performed; carried out*	B	*
U8-08	*do he appear more scare*	*makes*	C	*

Categorizamos o quarto tipo de desvio provocado pelo emprego inadequado do verbo como 'verbo diverso', pois não reconhecemos (i) uma influência do português ou (ii) uma relação de similaridade entre o verbo utilizado pelo aprendiz e o verbo da colocação adequada. Tampouco foram casos que (iii) envolvem *do* ou *make* (iii). Os motivos que levam ao uso impróprio do verbo variam conforme as ocorrências (Tabela 13).

No exemplo 63, *displace* parece ter sido empregado pelo aprendiz provavelmente porque teve a intenção de usar *display* já que ambos os verbos possuem ortografia semelhante. Enquanto *displace* pode significar 'deslocar', *display* tem o sentido mais semelhante a *demonstrate*, que forma de fato uma colocação com o substantivo utilizado *intelligence*. Nossa hipótese é corroborada quando constatamos que *displace* foi combinado com o substantivo *expressions* na sequência. Novamente, é improvável que o aprendiz tenha tido a intenção de expressar

deslocamento. Dessa vez, *have* forma a colocação apropriada com o segundo substantivo.

63. *The first presentation was of Mr. Hiro Takashi, from Asia, who has got a service robot, he is specialized on antropomorfisism, this robot features are that it can walk and displace some emotional intelligence, and some facial expressions and even burps.* (U1-12) → *demonstrate some emotional intelligence and have some facial expressions*

Um equívoco ortográfico também parece provocar o desvio do exemplo 64. A combinação *understand surveillance* foi usada ao invés de *undercover surveillance*, que está presente nas instruções da redação no curso *on-line*. Diferentemente do exemplo anterior, a correção não forma uma colocação verbal visto que *undercover* não é um verbo, mas um adjetivo.

64. *What I can say about it, it is efforts to catch him have included: extra police patrols in the area, because the city is very dangerous, understand surveillance, publicity campaign by this local paper and more things about it.* (U3-11) → *undercover surveillance*

Na combinação *awareness resources* (exemplo 65), o aprendiz empregou dois substantivos em vez de um verbo e um substantivo. Para manter o substantivo *resources* na correção, o verbo *provide* forma uma colocação com o substantivo ao mesmo tempo em que se encaixa no contexto da frase. Contudo, também é possível empregar a colocação *raise awareness* com o primeiro substantivo usado pelo aprendiz na combinação. Dado que as duas colocações são satisfatórias, o propósito inicial do aprendiz possivelmente foi usar somente um dos substantivos com algum verbo, mas seu texto manteve a inadequação. Também contribui para essa possibilidade o fato de o aprendiz ter empregado corretamente a colocação *raise awareness* posteriormente no mesmo texto (*I really think she needs a piece of my mind to change the way to raise awareness, so, because this I throw a brick in her.*).

65. *Look, if she had been troubles with alcohol and drugs for ten years, why she don't awareness resources for this kind of young people, whom really need her and our help too?* (U4-06) → *provide resources*

Tabela 13 – Desvio I: verbo inadequado (iv. verbo diverso)

Texto	Combinação	Correção	Uso	Aceit.
U1-12	*displace some emotional intelligence*	*demonstrate*	B	*
U1-12	*displace (...) some facial expressions*	*have*	B	*
U1-20	*the scientific robots are those performed*	*made; created; built*	B	*
U4-06	*don't awareness resources*	*provide*	B	*
U1-04	*will cohabit artificial beings*	*live with*	C	*
U3-11	*understand surveillance*	*undercover* (adjetivo)	C	*

Por fim, no último agrupamento de desvio do tipo I, reunimos os exemplos de combinações em que o verbo *take* foi empregado quando, na realidade, outros verbos formam colocações adequadas (Tabela 14).

A combinação entre *take* e o substantivo *part*, empregada no exemplo 66, forma uma colocação apropriada, porém foi corrigida para *are part* por causa do sentido pretendido na frase.

66. *We''ve been talking about three type of robots which already take part of our life considering the Industrial, medical and house work.* (U1-22) → *are part of our life*

No exemplo 67, o verbo *take* é usado com dois substantivos. O desvio ocorre somente na combinação com o segundo substantivo, pois *take* não forma uma colocação com *video*, apesar de ser aceitável com *picture.*

67. *When questioned by sceptics why they did not take any picture, video or things like that, the eyewitnesses said: Everything happened so fast...* (U8-06) → *record; shoot; make a video*

Tabela 14 – Desvio I: verbo inadequado (v. *take*)

Texto	Combinação	Correção	Uso	Aceit.
U1-22	*take part of our life*	*are*	B	*
U3-10	*other burglaries over the past 15 months, taken*	*commited*	B	*
U8-06	*did not take any (...) video*	*record; shoot; make*	B	*

3.2.2 Desvio II: substantivo inadequado

O segundo tipo de desvio mais comum é aquele em que o emprego do substantivo é inadequado (Tabela 15). Com uma porcentagem de ocorrência de 10,1%, esse tipo de desvio é muito menos frequente do que o uso inapropriado do verbo (desvio I), que é responsável por aproximadamente três quartos dos desvios combinatórios. Como já destacamos anteriormente, em um colocação verbal, o sentido em que o substantivo é empregado não é restrito arbitrariamente. Do ponto de vista do verbo, é este que adquire sentido ao se combinar com o substantivo em uma colocação. Assim, as inadequações encontradas na nossa segunda categoria de desvios são causadas pelo uso equivocado do sentido do substantivo na combinação.

No exemplo 68, ainda que *have* seja um colocado de *place,* o sentido em que a colocação foi utilizada não se adequa à frase. É plausível que a escolha do substantivo tenha sido motivada pela existência de 'ter lugar' em português, pois 'ter' e 'lugar' podem ser traduções de *have* e *place*. Para alcançar o sentido pretendido, a colocação *have a chance* pode ser empregada.

68. *Another field that robotics has been reaching is sports field, where robots have place to participate of sports events.* (U1-01) → *have a chance*

O exemplo seguinte (exemplo 69), diferentemente do exemplo anterior, não apresenta uma colocação adequada cujo sentido é usado impropriamente na frase. Dessa vez, temos a ocorrência de *have as finality the sport*, que não é uma colocação uma vez que *have* convencionalmente não se coloca com *as finality*. Na correção, além de substituir *finality* por *goal,* é necessário modificar a ordem dos substantivos para *have sports as a goal.* Acreditamos que esse desvio é influenciado pela língua portuguesa, em que é aceitável algo como 'ter como finalidade.

69. *While the robots created in US have as finality the sport and in the future the scientists wants to develop a robot with AI.* (U1-17) → *have sports as a goal*

Tabela 15 – Desvio II: substantivo inadequado

Texto	**Combinação**	**Correção**	**Uso**	**Aceit.**
U1-01	*have place*	*a chance* (Am/Br)	B	*
U1-08	*have a big capability*	*the capacity*	B	*
U2-17	*raise (...) fishing*	*fish*	B	*
U5-07	*drawing the line story*	*story line* (Am)	B	*
U5-10	*open your head*	*mind*	B	*
U5-12	*answers you might have to respond*	*questions*	B	*
U7-10	*fall in sleep*	*asleep* (adjetivo)	B	*
U1-17	*have as finality the sport*	*sports as a goal*	C	*
U5-10	*don't have (...) key*	*a clue*	C	*

3.2.3 Desvio III: verbo e substantivo adequados com estrutura inadequada

Na categoria de desvio III, tanto o verbo quanto o substantivo em si foram empregados apropriadamente, porém algo na estrutura da combinação é inadequado (Tabela 16). Em nossa análise, o índice de ocorrência desse tipo de desvio (9%) é

semelhante à categoria de desvio II. Mesmo não se tratando propriamente de um desvio na seleção do verbo ou do substantivo, ainda consideramos que há um desvio combinatório, pois a estrutura da combinação é afetada e, consequentemente, seu sentido também é prejudicado.

No exemplo 70, o aprendiz empregou o verbo *put* e o substantivo *effort*, os quais formam uma colocação. No entanto, na tentativa de usar a voz passiva, também usou o pronome *it* (*It was put a lot of effort*). Tanto o falante americano quanto o falante britânico de inglês sugeriram *A lot of effort was put*. O falante americano também sugeriu a combinação com o verbo *go* (*A lot of effort went*).

70. *It was put a lot of effort to catch him, including extra police patrol in the area, publicity campaigns and $300 reward for any applicable information about him.* (U3-03) → *A lot of effort was put/went*

Neste outro exemplo (exemplo 71), a palavra usada indevidamente é o pronome *your*. Na combinação entre o verbo *come* e o substantivo *mind*, os falantes nativos julgaram que não é apropriado empregar *your*. Para formar a colocação (*come to/into mind*), ambos sugeriram o uso de *to*, enquanto somente o falante americano sugeriu *into*.

71. *The third one is the stream of conciousness, literally you have to write whatever ideas come to your mind.* (U5-09) → *come to/into mind*

Tabela 16 – Desvio III: verbo e substantivo adequados com estrutura inadequada

Texto	Combinação	Correção	Uso	Aceit.
U3-03	*It was put a lot of effort*	*A lot of effort was put* (Am/Br)*; A lot of effort went* (Am)	B	*
U5-09	*come to your mind*	*to* (Am/Br)*; into* (Am)	B	(*)
U6-03	*get to a loan*	*a*	B	*
U1-22	*take places of human*	*the place of a human being*	C	*
U3-06	*inform a man acting suspiciously*	*them of* (Am)*; them about* (Br)	C	*

U4-03	*the image people have <u>about</u> her*	*of*	C	?
U5-10	*don't have <u>way</u>*	*a way with it*	C	*
U5-10	*don't have (...) <u>others points</u>*	*a point; what to say*	C	*

3.2.4 Desvio IV: verbo e substantivo inadequados

No quarto tipo de desvio de nossa tipologia de inadequações colocacionais, tanto o verbo quanto o substantivo são inadequados (Tabela 17). Com uma baixa porcentagem (3,4%), classificamos três casos nessa categoria de desvio.

No exemplo 72, verificamos a combinação entre *make* e *house taske** (*tasks*). Um dos verbos que se coloca com *task* é *do*, que, como já discutido previamente, gera desvios combinatórios quando confundido com *make*. Entretanto, o contexto do enunciado nesse exemplo permite que identifiquemos o sentido pretendido pelo aprendiz. Na realidade, o substantivo apropriado é *housework*, que forma uma colocação com *do.* Portanto, a correção nesse caso é *do the housework.*

72. *And whenever have a bad activities robots can do for the man. Such as: As a soldier on wars, clean houses; we can watch tv while the robots <u>make some so rude house taske</u>.* (U1-10) → *do the housework*

Já no exemplo 73, o aprendiz tenta adicionar outro argumento à sua discussão quando combina *adding to* e *this point*, como se estivesse utilizando uma locução conjuntiva. Contudo, essa combinação não é convencional em inglês. A correção que adotamos (*In addition*) mantém a relação com o verbo *add* empregado pelo aprendiz.

73. *<u>Adding to this point</u>, the specialist believe that we are bound to see robots with artificial intelligence by 2050.* (U1-16) → *In addition*

Tabela 17 – Desvio IV: verbo e substantivo inadequados

Texto	**Combinação**	**Correção**	**Uso**	**Aceit.**

U1-10	*make some so rude house taske*	*do the housework*	B	*
U1-16	*Adding to this point*	*In addition*	C	*
U8-03	*have this conscious*	*are aware*	C	*

3.2.5 Desvio V: verbo como substantivo

Na categoria de desvio V, foram classificados dois casos que representam praticamente 2% dos desvios combinatórios (Tabela 18). Em ambos os exemplos, os verbos usados pelos aprendizes são desnecessários na medida em que os substantivos são verbalizados na correção. No exemplo 74, em que o aprendiz utilizou *gave him a chase*, a correção é *chased him*. No outro exemplo (exemplo 75), *repudiate* (*all the stars*) é a forma verbalizada do substantivo *repudiation*, usado pelo aprendiz em *express my repudiation about oll the stars*.

74. *With a couple of cars, the police immediately gave him a chase and eventually cornered the man in Hamilton Road.* (U3-14) → *chased him*
75. *First of all, I would like to express my repudiation about all the stars that colaborate to ONGs just to feel better.* (U4-08) → *repudiate all the stars*

Tabela 18 – Desvio V: verbo como substantivo

Texto	**Combinação**	**Correção**	**Uso**	**Aceit.**
U3-14	*gave him a chase*	*chased him*	C	*
U4-08	*express my repudiation about all the stars*	*repudiate*	C	?

3.2.6 Desvio VI: substantivo como verbo

O penúltimo tipo de desvio possui apenas um exemplo (exemplo 76) que já foi discutido anteriormente na apresentação das estruturas sintáticas usadas no esquema

de anotação. Trata-se do caso em que o aprendiz empregou *pain* como verbo, por isso adotamos o nome 'substantivo como verbo' para essa categoria. Anotamos *pain* como um desvio, pois o sentido pretendido pelo aprendiz pode ser corrigido com uma colocação para tornar-se convencional. Para manter *pain* na colocação, em nossa correção adotamos *I feel the pain in my legs*, porém também é possível que *pain* seja simplesmente substituído por *hurt* (*my legs hurt*).

76. *I walk without stop until my legs pain* (...) (U7-10) → *I feel the pain in my legs; my legs hurt*

3.2.7 Desvio VII: inclassificável

Por fim, um desvio foi categorizado como inclassificável, assim como já havia sido feito na anotação de uso (D). No exemplo 77, não sabemos ao certo o sentido da combinação. O aprendiz empregou os verbos *do*, acompanhado de *can*, e *give*. Na combinação também estão *caring* como verbo nominalizado e o substantivo *people.* Supomos que a intenção do aprendiz foi empregar algo como *what can we do to provide those kind people with care*, no entanto, como não podemos ter certeza do sentido desejado, tal dado foi categorizado como inclassificável.

77. *So, we are today this powerfull country thanks to the all nations which gave their contribution for us. For how can we do for give our caring about these kindly people.* (U2-14)

3.3 Colocações e estruturas sintáticas

Em sua análise, Nesselhauf (2005, p. 68) identificou colocações em nove estruturas sintáticas: vo, vpo, voo, vopo, voc, voa, va, vc e vcpo. Em nosso estudo, somente a estrutura vcpo (verbo + complemento + preposição + objeto) não é encontrada. Ela ocorre, por exemplo, em *fall in love with somebody*. Como aponta a própria autora, essa estrutura parece ser rara (NESSELHAUF, 2005, p. 68). Nossa

anotação sintática também difere em outro pontos. Como já havíamos explicitado na apresentação das estruturas sintáticas usadas na anotação, utilizamos algumas outras categorias para anotar colocações: v, passiva, v+partícula+o e vo+inf.

A adoção da categoria v (verbo), como já foi discutido, ocorreu devido a um dado em que *pain* foi indevidamente empregado como verbo. Adotamos as demais categorias (passiva: voz passiva; v+partícula+o: verbo + partícula + objeto; vo+inf: verbo + objetivo + verbo no infinitivo) a fim de obter resultados mais detalhados. A Tabela 19 mostra a distribuição de estruturas sintáticas em relação à aceitabilidade.

Tabela 19 – Distribuição de estruturas sintáticas e aceitabilidade

Estrutura sintática	**+**	**(+)**	**?**	**(*)**	*****	**Total**
v	0	0	0	0	1	1
vo	185	1	7	3	43	239
vpo	5	0	0	1	6	12
voo	2	0	0	0	2	4
vopo	18	0	1	1	3	23
voc	1	0	0	2	2	5
voa	3	0	0	0	1	4
va	1	0	0	0	0	1
vc	0	0	0	0	1	1
passiva	46	0	1	2	9	58
v+partícula+o	13	0	0	0	0	13
vo+inf	0	0	0	0	2	2
Total	274	1	9	9	70	363

A maioria das colocações ocorre na estrutura sintática vo (239) e na voz passiva (58). As estruturas vopo (23), v+partícula+o (13) e vpo (12) também estão entre as mais frequentes.

A fim de estabelecer uma relação entre as estruturas sintáticas e os desvios combinatórios, apresentamos, na Tabela 20, a quantidade de colocações que foram empregadas (363) e as colocações que apresentam desvios (89) por estrutura sintática. Nessa tabela, agrupamos na coluna de desvios todas as colocações que não foram classificadas como aceitáveis, ou seja, as colocações tidas como possivelmente aceitáveis [(+)], questionáveis [?], possivelmente inaceitáveis [(*)] e inaceitáveis [*].

Dessa forma, podemos estabelecer a porcentagem de desvios em cada tipo de arranjo sintático.

Tabela 20 – Colocações empregadas e desvios em relação a estruturas sintáticas

Estrutura sintática	**Col. empregadas**	**Desvios**	**% desvio**
v	1	1	100
vo	239	54	22,6
vpo	12	7	58,3
voo	4	2	50
vopo	23	5	21,7
voc	5	4	80
voa	4	1	25
va	1	0	0,0
vc	1	1	100
passiva	58	12	20,7
v+partícula+o	13	0	0
vo+inf	2	2	100
Total	363	89	

Três estruturas sintáticas (v, vc e vo+inf) chamam a atenção por apresentarem porcentagem máxima de desvio. A quantidade de colocações empregadas no *subcorpus* nesses casos, no entanto, é baixa. Tanto para a estrutura sintática v (exemplo 76) quanto para a estrutura vc (exemplo 69), há apenas um dado na amostra. Já no caso da estrutura sintática vo+inf, apesar de haver somente dois dados, é necessário fazer algumas considerações.

A estrutura vo+inf está entre as estruturas que adotamos para além do esquema de anotação inicial. Contudo, podemos agrupar essa estrutura, juntamente com a estrutura v+partícula+o (verbo + partícula + objeto) e a passiva (voz passiva), sob o domínio de vo uma vez que não deixam de ser tipos de arranjos compostos por verbos com objetos. Se considerarmos esse agrupamento, podemos afirmar que vo+inf é um tipo de vo que leva a desvios em todos os casos (exemplos 77 e 78) em nossa amostra.

78. *Just leave your ideas flow and take notes of them.* (U5-09) → *let your ideas flow*
79. *Big foot was retrated in many movies in the last decades, in relates he is always a enormous strange animal, and on the contrary of the others animals like bears, he is a biped, walk with two legs, what do he appear more scare.* (U8-08) → *makes him look scarier*

Contrariamente, não encontramos desvios em nenhuma das treze ocorrências da estrutura v+partícula+o. A análise dessas colocações mostra que nove delas são com o substantivo *trash*, sendo que oito levam o verbo frasal *pick up* (exemplo 80). As demais não possuem substantivo tampouco verbo em comum (exemplo 81).

80. *Trundling around the hospital an picking up the trash such as taking care of the surveillance of the hospital are other areas that a service robot can be useful.* (U1-01)
81. *I jotted my notes down, so I had to spend some time reorganizing it.* (U5-12)

A estrutura passiva, por sua vez, possui incidência de desvios (20,7%) semelhante aos casos de vo (22,6%), nos quais a voz é ativa. Assim, a adoção de categorias sintáticas mais detalhadas para vo parece contribuir para o compreensão da incidência de desvios.

A segunda estrutura sintática que mais apresenta desvios (80%) é voc (verbo + objeto + complemento). Com exceção da colocação *set housewives e maids free* (U1-07), as combinações nessa estrutura apresentam desvios. No exemplo 82, estão dois dos desvios com a estrutura voc. Outro exemplo ocorre em 83.

82. *It had 820 meters as length, with two towers having 75 meters in height.* (U2-17). → *was 820 long/in length* → *are 75 meters in height*; *75 meters each*
83. *Look, if she had been troubles with alcohol and drugs for ten years, why she don't awareness resources for this kind of young people, whom really need her and our help too?* (U4-06) → *has had troubles with alcohol and drugs*

Este capítulo exibiu os resultados da aplicação de nosso método de pesquisa. Apresentamos tanto resultados quantitativos quanto uma análise dos desvios

combinatórios. Por meio da abordagem quantitativa, tivemos acesso a diferentes distribuições dos dados, tais como os tipos de combinações analisadas em nossa estudo, as colocações nos textos produzidos pelos aprendizes, e as estruturas sintáticas em que elas ocorrem. Também pudemos averiguar características relevantes das colocações encontradas, que permitiram sua classificação nas categorias de uso e aceitabilidade. Dessa forma, investigamos a adequação das colocações. Mediante a análise aprofundada das colocações classificadas como inadequadas, estabelecemos uma tipologia de desvios que os categorizou em setes grupos. Por fim, também investigamos a relação entre a sintaxe das colocações e sua adequação. No capítulo seguinte, articulamos as partes de nosso trabalho de maneira a discutir as limitações e desdobramentos de nossa pesquisa.

CAPÍTULO 4. CONSIDERAÇÕES FINAIS

Em nossa pesquisa, analisamos as colocações verbais encontradas nos textos escritos por brasileiros do *corpus* EFCAMDAT. Nossa análise incidiu somente sobre o nível considerado avançado de língua inglesa. A estrutura do *corpus* reflete os níveis de ensino do curso *Englishtown*, portanto foi possível selecionar apenas o nível 16 (C2).

Recorremos à anotação semiautomática de segmentos de texto a fim de identificar todos os verbos que ocorrem nas redações. Além disso, elaboramos um esquema de anotação baseado em Nesselhauf (2005) que permitiu a categorização dos verbos e principalmente das combinações entre verbos e substantivos.

Uma vez que o esquema também possui informações sobre o uso e a adequação das combinações encontradas, fomos capazes de estabelecer uma tipologia de desvios quando esses ocorreram. Para complementar nossa análise, dados relativos à sintaxe das colocações também foram examinados.

Nesse momento, é relevante retomar as perguntas de pesquisa que nortearam nossa pesquisa. Nos parágrafos abaixo, apresentamos as três perguntas de pesquisa separadamente seguidas de um resumo dos resultados de nosso estudo que responde a elas.

- Qual a incidência de colocações verbais na produção escrita de aprendizes brasileiros de inglês no nível avançado?

Verificamos que combinações entre um verbo e um substantivo são mais frequentes do que os casos de não-combinação, ou seja, aqueles em que um verbo não se une a nenhum outro elemento ou se combina com outra classe gramatical. Somadas, combinações livres, colocações e expressões idiomáticas são responsáveis por mais da metade (57%) de usos verbais. Quando são consideradas somente essas três categorias de combinação, a incidência de colocações é de aproximadamente um quinto dos casos (22%). Entre as colocações usadas pelos aprendizes, 25% foram julgadas inadequadas, ou seja, apresentaram um desvio combinatório.

- Quais tipos de desvios ocorrem?

Em nossa tipologia de desvios combinatórios, estabelecemos sete categorias para classificar as colocações tidas como inadequadas: (i) verbo inadequado, (ii) substantivo inadequado, (iii) verbo e substantivo adequados com estrutura inadequada, (iv) verbo e substantivo inadequados, (v) verbo como substantivo, (vi) substantivo como verbo, e (vii) inclassificável. O primeiro tipo de desvio, em que o verbo é empregado inadequadamente, é o mais comum com 73% de ocorrências. Apesar de minoria apenas 25% das colocações terem sido classificadas como inadequadas, os tipos de desvios mostram que os aprendizes ainda não dominam algumas colocações básicas, como, por exemplo, *wear (use*) corsets*.

- Quais as causas que contribuem para a ocorrência de um desvio?

Visto que o uso inadequado do verbo é muito frequente dentre os desvios, criamos subcategorias para esse caso. De certa forma, a subcategorização ajuda a compreender os fatores que levam a um desvio. Na maioria dos casos, como suspeitávamos, há uma clara influência do português. O par de verbos *do/make* e o verbo *take* são exemplos dessa influência. Em alguns casos, o uso de verbos com sentidos semelhantes ao que se coloca com o substantivo também ocorre. Além disso, em nossa análise, ocorreram relativamente mais desvios em algumas estruturas sintáticas (verbo + preposição + objeto, verbo + objeto + complemento, verbo + complemento, verbo + objeto + verbo infinitivo) do que em outras, porém essas observações não podem ser tomadas como generalizações devido à baixa quantidade de dados.

Como qualquer outra pesquisa, nosso trabalho também possui limitações. Em primeiro lugar, analisamos somente um tipo de padrão convencional da língua, as colocações. Mais do que isso, investigamos apenas colocações verbais. Além disso, o nível avançado de aprendizado de inglês foi o único investigado. Assim, uma análise mais abrangente poderia levar a resultados mais precisos. É também importante lembrar que desvios de outras natureza que não dizem respeito à convencionalidade linguística ainda ocorrem apesar do nível avançado, portanto é essencial que sejam investigados.

Em segundo lugar, a metodologia de análise também é falha em alguns momentos. O recurso a avaliações por falantes nativos parece ir contra o caráter

empírico da área de pesquisa, mas, ao mesmo tempo, mostra que um *corpus*, assim como produtos resultados de pesquisa com *corpus*, como dicionários e ferramentas eletrônicas, não são capazes de fornecer todos os dados linguísticos que um pesquisador busca. Outro aspecto da avaliação de falantes nativos de inglês é o fato de não termos certeza de que eles usaram realmente sua intuição, pois, muitas vezes, em situações de teste, os avaliadores tendem a não ter o mesmo desempenho comparados a situações espontâneas de uso da língua

Em terceiro lugar, temos de salientar que analisamos um fenômeno que não considera o texto como um todo, isto é, não foi nosso objetivo fazer observações sobre os gêneros textuais ou as implicações que as propostas de redações têm no aprendizado.

Do lado positivo, podemos apontar que o fato de falarmos a mesma língua dos aprendizes, o português, ajudou a compreender e examinar melhor os dados. Os resultados obtidos em nosso estudo podem ser aproveitados para aprimorar o ensino da convencionalidade linguística em inglês.

A fim de ilustrar como os resultados de nossa pesquisa podem contribuir para o ensino de colocações verbais, propomos, ao final deste capítulo, atividades didáticas baseadas nos achados de nosso estudo. Uma vez que analisamos redações provenientes do *Englishtown*, nossa proposta busca se adaptar à estrutura desse curso.

Como já observamos, muitas atividades são corrigidas automaticamente pelo sistema do curso. Somente algumas redações são corrigidas por corretores humanos. No entanto, após receber sua nota e os comentários de seu professor, o aprendiz não faz nenhum tipo de atividade baseada em seu desempenho. Caso alcance uma nota satisfatória, poderá passar para a unidade seguinte. Se não obtém resultado satisfatório, terá de realizar a atividade escrita novamente. Tanto alunos que tenham um desempenho satisfatório quanto aqueles na situação contrária podem se beneficiar das atividades, pois os dois grupos podem ter apresentado desvios.

As atividades apresentadas aqui são meramente uma proposta de aplicação dos resultados gerados pelo nosso estudo, portanto não foram e não serão aplicadas com os aprendizes que contribuíram com suas redação para o *subcorpus*. Não temos acesso aos participantes que escreveram as redações, assim não seria possível avaliar seu desempenho em nossa proposta didática. Tampouco aplicamos as atividades a outros participantes, pois julgamos que seria difícil estabelecer o mesmo perfil dos alunos do *Englishtown*.

Tendo em vista que analisamos uma quantidade relativamente grande e variada de desvios combinatórios, nossos exemplos de atividades didáticas abordarão somente os desvios da unidade 1 do nível 16 no *Englishtown*, ou seja, após o término da redação, os alunos poderiam fazer as atividades que seguem.

Nossa proposta didática adapta a abordagem de aprendizado DDL (JOHNS, 1988), na qual é essencial que o aprendiz tenha contato com dados autênticos provenientes de *corpora*. Diferentemente da abordagem tradicional, não utilizamos linhas de concordância habituais em que o mesmo número de caracteres estão dispostos à esquerda e à direta da palavra de busca. Como já foi indicado, as colocações verbais podem ocorrer em estruturas sintáticas diversas em que verbo e substantivo podem ou não estar próximos. Além disso, questionamos a fragmentação dos enunciados nas linhas de concordância, dado que, em nossa avaliação, a suposta atenção que é dada para a procura de padrões a partir da palavra de busca não se mostra vantajosa diante do prejuízo causado pela possível ruptura das frases.

Para selecionar os exemplos de colocações verbais convencionais, utilizamos a plataforma SkELL (BAISA e SUCHOMEL, 2014), pois sua ferramenta *Examples* seleciona exemplos que tendem a ser curtos e ilustrativos devido a um instrumento chamado *Good Dictionary Example* (GDEX) que faz parte do sistema (BAISA e SUCHOMEL, 2014, p. 68). Além disso, também usamos a ferramenta *Word Sketch* que exibe os colocados de uma palavra de acordo com algumas funções indicadas nos enunciados (BAISA e SUCHOMEL, 2014, p. 65). Na busca de colocações verbais, a ferramenta mostra os verbos que formam colocações com um substantivo quando esse funciona, por exemplo, como objeto na frase. Na Figura 14, vemos um exemplo dos resultados da *Word Sketch* em que buscamos os colocados da palavra *thought*.

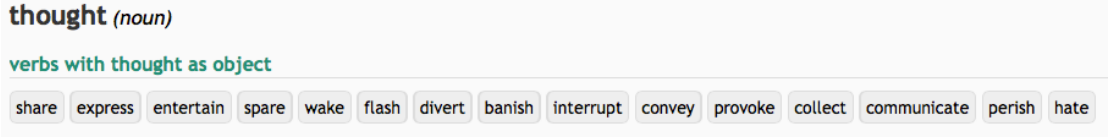

Figura 14 – Captura de tela dos verbos com *thought* como objeto na ferramenta *Word Sketch* (SkELL)

Com essas ferramentas, podemos elaborar exercícios que chamem a atenção dos aprendizes para o uso de colocações verbais aceitáveis. Para nossa proposta, selecionamos os desvios *give us your thoughts* (U1-09), *patients victimized* (U1-15) e *is going to make part of our lifes* (U1-16). Seguem os exercícios propostos:

1. Read the sentences below and underline the verb that goes with the word in bold.

Example: It's hard to express my **thoughts** in words.

a) Click the link below to share your **thoughts**.
b) Doctors concluded that these **patients** were infected with CMV before their HIV treatment.
c) Add a comment and share your **thoughts**.
d) Traffic flow is another big **part** of circulation.
e) All the **patients** were infected with a strain of bacteria called Salmonella typhimurium.
f) Water is **part** of every healthy nutrition plan.
g) Share your **thoughts** in the comment area below.
h) Slightly more than half of **patients** infected with B. burgdorferi develop Lyme arthritis.
i) Open discussion is **part** of every meeting.

2. Complete the sentences with the verb that goes with the word in bold.

a) Risk _____ **part** of every human endeavor.
b) Only a minority of **patients** _____ with H pylori will eventually develop a duodenal ulcer.
c) Please _____ your **thoughts** on this product!
d) _____ your **thoughts** in the comments section.
e) Bowen theory _____ **part** of attachment theory.
f) The search for new ways to improve treatment of **patients** _____ with HIV is constant.
g) _____ your **thoughts** and opinions in social media.
h) There must also be definitive evidence that the **patient** was originally _____ by the B. burgdorferi spirochete.
i) Rich indigenous traditions _____ **part** of daily life.

3. Rewrite these sentences to make them more natural.

a) Give us your thoughts about this theme at our website.

_____

b) This robot can assist patients victimized by contagious diseases.

_____

c) The service robot is going to make part of our lives in the very near future.

_____

Apesar de termos identificado algumas causas de desvios combinatórios, mais pesquisas são necessárias para que tenhamos uma melhor compreensão dos fatores que influenciam os desvios. Esperamos que esta pesquisa possa oferecer uma pequena contribuição para aprimorar o ensino de colocações verbais a aprendizes de inglês.

REFERÊNCIAS

BAISA, V.; SUCHOMEL, V. **SkELL:** Web Interface for English Language Learning. Eighth Workshop on Recent Advances in Slavonic Natural Language Processing. Brno: Tribun EU. 2014. p. 63-70.

BAKER, P.; HARDIE, A.; MCENERY. **A glossary of Corpus Linguistics**. Edimburgo: Edinburgh University Press, 2006.

BARNBROOK, G.; MASON, O.; KISHNAMURTHY, R. **Collocation:** applications and implications. Hampshire: Palgrave Macmillan, 2013.

BERBER SARDINHA, T. **Linguística de Corpus**. Barueri: Manole, 2004.

CAMBRIDGE UNIVERSITY PRESS. **Cambridge Dictionaries Online**. Disponivel em: <http://dictionary.cambridge.org>. Acesso em: ago. 2015.

COUNCIL OF EUROPE. Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Disponivel em: <http://www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf>. Acesso em: out. 2015.

DAVIES, M. **BYU-BNC**, 2004. Disponivel em: <http://corpus.byu.edu/bnc/>. Acesso em: out. 2015.

DAVIES, M. **The Corpus of Contemporary American English**, 2008. Disponivel em: <http://corpus.byu.edu/coca>. Acesso em: out. 2015.

EDUCATION FIRST. **English Live**. Disponivel em: <http://englishlive.ef.com>. Acesso em: out. 2015.

EDUCATION FIRST. **Englishtown**. Disponivel em: <http://www.englishtown.com.br>. Acesso em: out. 2015.

EDUCATION FIRST E UNIVERSIDADE DE CAMBRIDGE. **Education First-Cambridge Open Language Database**. Disponivel em: <https://corpus.mml.cam.ac.uk/efcamdat/>. Acesso em: 12 mar. 2014.

GEERTZEN, J.; ALEXOPOULOU, T.; KORHONEN, A. **Automatic linguistic annotation of large scale L2 databases:** The EF-Cambridge Open Language Database (EFCAMDAT). 31st Second Language Research Forum (SLRF). [S.l.]: Cascadilla Press. 2012. p. 240-254.

GRANGER, S. The International Corpus of Learner English. In: AARTS, J.; DE HAAN, P.; OOSTDIJK, N. **English Language :** Design, Analysis and Exploitation. Amsterdã: Rodopi, 1993. p. 57-69.

GRANGER, S. Learner English around the world. In: GREENBAUM, S. **Comparing English world-wide**. Oxford: Claredon Press, 1996. p. 13-24.

GRANGER, S. Learner . In: LÜDELING, A. &. K. M. **Corpus Linguistics. An International Handbook.** Berlim & New York: Walter de Gruyter, v. 1, 2008. p. 259-275.

HARPERCOLLINS PUBLISHERS. **Collins Dictionary**. Disponivel em: <http://www.collinsdictionary.com>. Acesso em: ago. 2015.

HOWARTH, P. **Phraseology in English academic writing. Some implication for language learning and dictionary making**. Tübingen: Niemeyer, 1996.

HYLAND, K. **Academic discourse**. Londres: Continuum, 2009.

JOHNS, T. Whence and wither classroom concordancing? In: BONGAERTS, T., et al. **Computer Applications in Language Teaching**. Dordrecht: Foris, 1988. p. 9-27.

JOHNS, T. Should you be persuaded: two samples of data-driven learning materials. **English Language Research Journal**, v. 4, p. 1-16, 1991.

MCCARTHY, M. and the advanced level: Problems and prospects. **English Australia Journal**, v. 29, n. 1, p. 39-50, 2013.

MCENERY, T.; HARDIE, A. **Corpus Linguistics:** method, theory and practice. Cambridge: Cambridge University Press, 2012.

NESSELHAUF, N. **Collocations in a Learner Corpus**. Amsterdam & Philadelphia: Johns Benjamins, 2005.

O'DONNELL, M. The UAM CorpusTool: Software for corpus annotation and exploration. In: CALLEJAS, C. M. B. **Applied Linguistics Now:** Understanding Language and Mind / La Lingüística Aplicada Hoy: Comprendiendo el Lenguaje y la Mente. Almería: Universidad de Almería, 2008. p. 1433-1447.

O'DONNELL, M. **UAM Corpus Tool 2.8**. Madrid: Lexical analysis software, 2011.

O'KEEFFE, A.; MCCARTHY, M.; CARTER, R. **From corpus to classroom:** language use and language teaching. [S.l.]: Cambridge University Press, 2007.

ORENHA-OTTAIANO , A. English collocations extracted from a corpus of university learners and its contribution to a language teaching pedagogy. **Acta**

Scientiarum. Language and Culture, Maringá, v. 34, n. 2, p. 241-251, Julho-Dezembro 2012.

OXFORD UNIVERSITY PRESS. **Oxford Dictionaries**. Disponivel em: <http://www.oxforddictionaries.com>.

OXFORD UNIVERSITY PRESS. **Oxford Collocations Dictionary for students of English**. Oxford: Oxford University Press, 2002.

SCOTT, M. **WordSmith Tools versão 6**. Liverpool: Lexical analysis software, 2012.

SHEPHERD, T. M. G. de aprendiz de língua estrangeira: um estudo contrastivo de n-gramas. **Veredas**, Juiz de Fora, v. 13, n. 2, p. 100-116, 2009.

TAGNIN, S. E. O. Um dicionário de colocações verbais? Para quê? In: BERBER SARDINHA, T. **A Língua Portuguesa no Computador**. Campinas: Mercaco de Letras, 2005. p. 197-214.

TAGNIN, S. E. O. A multilingual learner corpus in Brazil. In: WILSON, A.; ARCHER, D.; RAYSON, P. **Corpus linguistics around the world.** Amsterdam/New York: Rodopi, 2006. p. 195-202.

TAGNIN, S. E. O. **O jeito que a gente diz:** combinações consagradas em inglês e português. Barueri: Disal, 2013.

APÊNDICE A - LISTA DE TEXTOS E CÓDIGOS DO *SUBCORPUS*

Texto	Código
U1-01	U18587
U1-02	U42206
U1-03	U75082
U1-04	U115401
U1-05	U118793
U1-06	U161751
U1-07	U162024
U1-08	U217083
U1-09	U228043
U1-10	U327664
U1-11	U347166
U1-12	U361524
U1-13	U379470
U1-14	U385307
U1-15	U388180
U1-16	U399391
U1-17	U403249
U1-18	U429246
U1-19	U448112
U1-20	U455871
U1-21	U461478
U1-22	U473661
U1-23	U563926
U1-24	U630416
U1-25	U676609
U1-26	U701789
U1-27	U704913
U2-01	U23664
U2-02	U82732
U2-03	U144090
U2-04	U161318
U2-05	U170080
U2-06	U189249
U2-07	U243700
U2-08	U253221
U2-09	U290256
U2-10	U386292
U2-11	U387701
U2-12	U410521
U2-13	U449365
U2-14	U471745
U2-15	U521024
U2-16	U529527
U2-17	U602055
U2-18	U622617
U2-19	U643358
U2-20	U666336
U2-21	U692206
U3-01	U88397
U3-02	U166049
U3-03	U254374
U3-04	U384616
U3-05	U388844
U3-06	U409156
U3-07	U422556
U3-08	U439947
U3-09	U484870
U3-10	U492569
U3-11	U499619
U3-12	U528615
U3-13	U541285
U3-14	U616630
U4-01	U174705
U4-02	U176158
U4-03	U230199
U4-04	U316306
U4-05	U384911
U4-06	U411944
U4-07	U421608
U4-08	U428429
U4-09	U454731
U4-10	U513782
U4-11	U529857
U4-12	U572083
U4-13	U680322
U5-01	U5655
U5-02	U24707
U5-03	U104025
U5-04	U177720
U5-05	U229358
U5-06	U265011
U5-07	U360614
U5-08	U389564
U5-09	U429034
U5-10	U522929
U5-11	U677884
U5-12	U722572
U6-01	U181054
U6-02	U212797
U6-03	U540731
U6-04	U549542
U7-01	U130911
U7-02	U183296
U7-03	U230778
U7-04	U238651
U7-05	U359368
U7-06	U392565
U7-07	U450697
U7-08	U453403
U7-09	U453470
U7-10	U458400
U7-11	U531934
U7-12	U547484
U7-13	U550642
U8-01	U174123
U8-02	U183865
U8-03	U247685
U8-04	U356126
U8-05	U392938
U8-06	U454180
U8-07	U459162
U8-08	U464308
U8-09	U715523

APÊNDICE B – INSTRUÇÕES PARA OS FALANTES NATIVOS REALIZAREM O TESTE DE ACEITABILIDADE

Dear native speaker,

You are going to evaluate sentences that contain *collocations*. Here is a definition of collocation from the Oxford Dictionary of Collocation.

"Collocation is the way words combine in a language to produce natural-sounding speech and writing. For example, in English you say *strong wind* but *heavy rain*. It would not be normal to say **heavy wind* or **strong rain*. And whilst all four of these words would be recognized by a learner at pre-intermediate or even elementary level, it takes a greater degree of competence with the language to combine them correctly in productive use. To a native-speaker these combinations are highly predictable; to a learner they are anything but."

By using your native-speaker 'prediction', we would like you to evaluate combinations between a verb and a noun, for example, *make a decision* (NOT **take a decision*). Your task is to judge them as acceptable, questionable or unacceptable. If you feel a combination is acceptable, move on to the next sentence. When you feel a combination is either questionable or unacceptable, suggest a more natural-sounding combination if you can think of one. You may change the verb, the noun or even rewrite the whole combination. In case you judge a combination as questionable/unacceptable but cannot think of a way to correct it, mark the cell with an 'x'.

Here are some guidelines.

- To make the task easier, the verb and the noun are underlined. No other part of the sentence should be corrected unless it contributes to the correction of the collocation.
- Sentences are to be evaluated independently of each other.
- Do not resort to any kind of dictionary or similar material, only to your own intuition.
- Some sentences may present a collocation in the passive voice: *made a decision* → *a decision was made.*
- Do not hesitate to contact us if you have any doubt.

Danilo Murakami

APÊNDICE C – PRÉ-TESTE DE ACEITABILIDADE COM FALANTES NATIVOS DE INGLÊS

Sentence	Acceptable	Questionable	Unacceptable
The owner is currently involved in robot competitions which are getting place in the US.			
The last kind of robot already exists and it is a really productive manner of taking fun.			
The Christ the Redeemer statue scatters its arms wide open as if trying to embrace those who have been living in Rio de Janeiro.			
Nowadays, many catholics have the opportunity to celebrate baptisms in front of the statue.			
The policeman said that he didn't offer any kind of resistance; he wasn't even grasping guns.			
To my mind, the biggest problem here is that he champions his cause at every oportunity.			
I have watched a lecture about creative writing and I found useful techniques.			
They will certainly scrutinize my financial history to make sure that I do not have any debts which have not been bought.			
I can see a lot of birds flying, making loops in front of the sun.			
I think men now have enough intelligence to do their choices.			

APÊNDICE D – COLOCAÇÕES CLASSIFICADAS COMO ACEITÁVEIS

Texto	Colocação
01_001	display some emotion
01_001	display some emotion and intelligence
01_001	have lunch
01_001	picking up the trash
01_001	take care of elderly people
01_001	taking care of the surveillance of the hospital
01_002	colleting trash
01_002	has the hability
01_002	machines disigned
01_002	machines disigned, builded
01_002	machines disigned, builded and programme
01_002	one of the improvements that have already been made
01_002	support pacients
01_003	picking up the trash
01_003	show very realistic human expressions
01_003	takes 5 years
01_003	takes 5 years, the second 10 years
01_003	takes 5 years, the second 10 years and the third almost 40 years
01_003	The second one (robot) developed
01_004	all the work that has being done
01_004	doing important jobs
01_004	lead teams
01_004	picking up trash
01_004	playing matches
01_005	display human emotions
01_005	take care of patients
01_005	take out the trash
01_006	are suffering from contagious diseases
01_006	are taken place
01_006	matches played
01_006	picking up the trash

01_006	supporting security systems
01_007	have their value
01_007	increase companies productivity
01_007	increase companies productivity and profit
01_007	is not going to take the place of real human being
01_007	set housewives and maids free
01_007	take care of the house
01_007	to take care of our elders
01_008	do human chores
01_008	do mundane chores
01_009	make a few facial expressions
01_010	bad activities robots can do
01_010	can watch tv
01_010	have your benefit s
01_010	to do activities
01_011	display diary human emotions
01_012	can score goals
01_012	improve living conditions
01_012	suffer from contagious diseases
01_012	will take around 5 years
01_013	are suffering from contagious diseases
01_013	display some emotional intelligence
01_013	picking up the trash
01_013	scoring goals
01_014	are suffering from contagious diseases
01_014	can play a vital role
01_014	pick up the trash
01_014	show emotions
01_015	Is this going to take place
01_015	lose my faith
01_015	taking care of trash
01_016	having fun
01_016	is going to express emotional reactions
01_016	take part in international competitions

01_016	take part in our day by day life
01_016	the kind of work it is able to do
01_016	to take care of contagious diseases
01_016	will have an impressive success
01_018	get his research done
01_018	improvements that her robots have been achieving
01_018	make even small movements
01_018	make human expressions
01_020	analyse data
01_020	make hypotheses
01_020	perform more complex tasks
01_022	do the work
01_022	doing house chores
01_022	saving peoples lives
01_022	spend more money
01_022	spend more time
01_023	can pick the track up
01_023	can play sports
01_023	has a contagious disease
01_023	will show feelings
01_023	will show feelings, emotions
01_024	collecting infected garbages
01_024	follow direct commands
01_024	improve the efficiency
01_024	show human emotions
01_024	will take at least 10 years
01_025	does orderly duties
01_026	are having an international competitions
01_026	has a different utility for people
01_026	took 30 minutes
01_026	took the entire room
01_027	caught contagious diseases
01_027	could treat patients
01_027	display human emotions

01_027	have caught contagious diseases
01_027	not catch their diseases
01_027	picking up the trash
02_001	collect signatures
02_001	cost the equivalent of US$250,000
02_001	took nine years
02_002	gave way for the idea
02_003	This fear, however, was eliminated
02_004	weighs 635 tonnes
02_005	committing a crime
02_005	has great importantance
02_005	spreads its arms
02_006	gold gathered
02_009	took from 1907 to 1912
02_011	Laws should be ratified
02_012	doing our daily activities
02_012	having fun
02_015	gave way for the idea
02_016	speaking many languages
02_017	raise cattle
02_017	raise cattle, vegetables
02_017	took four years
02_017	weighting five thousand tons
02_020	celebrate baptisms
02_020	celebrate baptisms and marriages
02_020	has suffered the damage of weather conditions
02_020	have the opportunity
02_020	This monument was made
02_020	took 9 years
02_021	give your life for the Soccer
02_021	played soccer
03_001	made the arrest
03_001	posted a $300 reward
03_001	took many efforts

03_002	$300 reward put up
03_002	gave chase
03_002	He is being held in custody
03_003	gave chase
03_003	The man is currently held in custody
03_004	a reward put up
03_004	have joined forces
03_004	The man, whose name has not been released yet, has been held in custody
03_005	offered a reward
03_005	The suspect is now being held under police custody
03_005	The suspect will be questioned
03_006	a $ 300 reward was offered
03_006	crimes committed
03_006	efforts made
03_006	man is held in custody
03_007	arrests suspect
03_007	has a connection with a series of burglaries
03_007	the arrest was made
03_008	have made the best efforts
03_008	make an appointment
03_009	collecting information
03_009	offering a $300 reward
03_009	received a vital information
03_009	The case was firstly investigated
03_009	the man is being held in custody
03_009	the suspect was arrested
03_009	took place
03_010	campaign being held
03_010	fight crime rates
03_010	He's been held in custody
03_011	will get the information
03_012	$300 reward put up
03_012	an interview given
03_012	have a connection

03_012	The alleged man has been held in custody
03_014	A 23 years-old male, who has been wanted
03_014	did a good job
03_014	didn't offer any kind of resistance
03_014	has being held under custody
03_014	wasn't holding guns
04_001	diet people decide to have
04_001	do not have the proper means
04_001	take into consideration the unfortunate events
04_003	donate money
04_003	donate money
04_003	receive donations
04_003	the cause she defends
04_004	found out some precious informations
04_004	video made
04_005	making a good amount of money
04_006	raise awareness
04_007	call attention to herself
04_007	raise awareness
04_008	raise awareness
04_009	commiting injustice
04_009	use drugs
04_010	have had some problem
04_010	have no problem
04_011	champions his cause
04_011	wearing those corsets
04_012	she haven't given a break
04_012	telling the truth
04_013	is giving a very bad example
05_001	can do some draw
05_001	developing you ideas
05_001	explain and developing you ideas
05_001	solve the problems
05_001	will spend a feel time

05_002	spend 30-50 minutes
05_003	have a deadline
05_003	having dificulty
05_003	took me a long
05_004	asking questions
05_005	Spend 3-5 minutes
05_006	can make a list
05_006	can make your list
05_006	other good method to use
05_007	expresses the writer’s thoughts, feelings in an imaginative and poetic way
05_007	expresses the writer’s thoughtsin an imaginative and poetic way
05_008	spend around 3 to 5 minutes
05_009	answer the questions
05_009	catch people's attention
05_009	didn't take notes
05_009	have attended a lecture
05_009	some of them can be used
05_009	take notes
05_011	finish your assignment
05_011	my notes I made
05_011	Spend 30 to 50 minutes
05_012	asking me such a favor
05_012	jotted my notes down
05_012	Spend 30 to 50 minutes
05_012	spend some time
06_001	any debts which have not been paid off
06_001	been granted other loans
06_001	do not have any debts
06_001	have ever applied for and been granted other loans
06_001	to be granted a bank loan
06_001	will make a few phone calls
06_002	paying it off
06_002	repaying the loan
06_003	get a bank loan

06_003	get a bank loan
06_003	must get a Driver's License
06_003	spend almost all my time
06_004	could set a higher tax
07_001	hear the sound of the waterfall
07_001	open up my eyes
07_002	can smell the scent of the flowers
07_002	spare a minute
07_002	transmits a sensation of life and hope
07_004	covers all the area
07_004	doing loops
07_004	hear the natural sounds
07_004	make a wish
07_005	giving me a massage
07_005	have the feeling
07_007	doing relaxing activities
07_010	can hear the oceans noise
07_012	spend the day
07_013	Don't have fear
08_001	do your own research
08_001	form your own opinion
08_001	make reports
08_001	make their choices
08_001	performing impossible maneuvers
08_002	had its heyday
08_002	putting an end to the lives of many other animals
08_002	took place
08_004	Eyewitness sightings have been claimed
08_004	these sightings were first reported
08_005	Eyewitness sightings have been claimed
08_005	sightings were first reported
08_006	attacks made
08_006	did not take any picture,
08_006	This term was created

08_007	do not consume the prey
08_008	don't have resources
08_008	picture taken
08_008	take a picture
08_009	pay attention