

INVESTIGANDO COLOCAÇÕES EM CORPUS DE APRENDIZ

Valdênia C. ALMEIDA
Universidade Federal de Minas Gerais

RESUMO: Este artigo discute resultados de uma pesquisa que visou investigar como aprendizes brasileiros de inglês utilizam colocações em suas produções escritas. Mais especificamente, o estudo investigou as colocações verbo nominais com os verbos: *get*, *give*, *make* e *take* e os tipos de erros relacionados ao uso dessas colocações na escrita dos aprendizes. Os dados apresentados provêm do corpus de aprendiz Br-ICLE (*Brazilian International Corpus of Learner English*) e foram analisados, utilizando o instrumental da Linguística de Corpus. Os resultados revelam que os aprendizes brasileiros têm dificuldade em produzir e utilizar as colocações de forma adequada, quando comparados ao uso que o falante nativo faz dessas combinações de palavras. Como implicação pedagógica, tem-se a importância de promover a conscientização linguística do aprendiz acerca dos padrões da língua e dos problemas em potencial, envolvendo as colocações já referenciadas e as demais combinações de palavras.

PALAVRAS-CHAVE: corpus de aprendiz; colocação; conscientização linguística

ABSTRACT: *This article discusses the results of a research that aimed to investigate how Brazilian learners of English use collocations in their written production. More specifically, the study investigated verb-noun collocations with the verbs get, give, make and take, and identified types of collocational errors presented in their compositions. The data presented come from Br-ICLE (Brazilian International Corpus of Learner English) and were analyzed using some of the tools from Corpus Linguistics. The results indicate that Brazilian learners have difficulty in producing and using collocations when compared to native speakers. Thinking about the implications to language pedagogy it is important to raise learner's awareness of the language patterns and the potential problems involving the previously mentioned collocations and other word combinations.*

KEYWORDS: *learner corpus; collocation; linguistic consciousness*

Introdução

O termo colocação foi cunhado pelo Linguista Britânico J.R. Firth (1957), cujas palavras, para descrever o termo, ficaram bastante conhecidas entre os estudiosos da área: “*you shall know a word by the company it keeps*”. Após Firth, outros linguistas estudaram o fenômeno e novas definições de colocação foram sendo propostas. Apesar de essas novas definições apresentarem alguma diferença entre elas, todas estão relacionadas à idéia Firthiana de que as palavras andam juntas e, muitas vezes, o significado de uma palavra depende das demais palavras com as quais ela se combina. Duas definições bastante conhecidas são as de Sinclair (1991) e Cowie (1994). Segundo Sinclair (1991), a colocação é a ocorrência de uma ou mais palavras, dentro de um texto, com um espaço curto entre elas. Considera-se espaço curto como sendo a distância entre os itens lexicais relevantes (os colocados) e a palavra nódulo. Na definição de Cowie (1994) a colocação é considerada como sendo um tipo de combinação de palavras a qual pode ser distinguida de

outros tipos de agrupamentos lexicais, como as combinações livres e os idiomas. O presente estudo assume a definição de Sinclair (1991).

Recentemente, inúmeros estudos têm mostrado que as colocações são um importante aspecto no conhecimento de uma língua por seus aprendizes. O conhecimento das colocações reduz o esforço de processamento e é de grande importância para que o aprendiz produza a língua de forma mais fluente e apropriada (COKLIN; SCHMITT 2008, GRANGER; MEUNIER 2008, HOWARTH 1998, SCHMITT 2004, SINCLAIR 1991). Devido ao seu importante papel no desenvolvimento da fluência e proficiência e, conseqüentemente, na aquisição de uma língua, as dificuldades e necessidades dos aprendizes com as colocações têm sido cada vez mais reconhecidas e estudadas (HONG *et al.*, 2011, LAUFER; WALDMAN 2011, NESSELHAULF, 2005, WANG; SHAW 2008). Esses estudos revelam que, mesmo aprendizes de nível avançado, frequentemente, têm problema com as colocações.

Um tipo de colocação estudada por alguns desses pesquisadores são as colocações verbo nominais. De acordo com Nesselhauf (2005), as colocações verbo nominais estão entre as mais difíceis para o aprendiz e são de extrema importância por formarem o núcleo comunicativo das produções, uma vez que concentram a informação principal que se deseja passar e, assim sendo, elas devem ser priorizadas no ensino de línguas. Todavia, para sabermos que colocações devem ser ensinadas, é necessário identificar possíveis problemas que aprendizes de uma língua têm ao produzir e utilizar as mesmas.

Desse modo, o estudo aqui descrito, objetivou investigar o uso de colocações verbo nominais por aprendizes brasileiros de inglês de nível avançado, cursando a graduação em Língua Inglesa. Mais especificamente, o estudo investigou as colocações com os verbos *get*, *give*, *make* e *take*.

Nas seções que se seguem, apresento a metodologia de pesquisa, descrevendo os procedimentos de coleta e análise dos dados para, posteriormente, mostrar os resultados do estudo. Por fim, concluo o artigo, discorrendo acerca das implicações pedagógicas dos estudos na área.

Metodologia

O estudo utilizou um corpus de estudo e dois corpora gerais. O corpus de estudo, chamado Br-ICLE (*Brazilian International Corpus of Learner English*) é formado por composições argumentativas escritas por estudantes brasileiros universitários, aprendizes de língua inglesa em nível avançado, cursando a graduação em inglês. A versão do Br-ICLE, utilizada nessa pesquisa, conta com os textos produzidos até 2009, totalizando 159.000 palavras. A pesquisa também faz uso de dois corpora gerais, o BNC (*British National Corpus*) com 100 milhões de palavras e o COCA (*Corpus of Contemporary American English*) com 425 milhões de palavras, utilizados para avaliar a aceitabilidade

das colocações e na elaboração das atividades. Cabe ressaltar que apenas a seção acadêmica de ambos os corpora gerais foi utilizada na pesquisa.

O presente estudo fez uso da ferramenta *Concordance*, para coletar todas as colocações com os verbos *get*, *give*, *make* e *take* no corpus Br-ICLE. O concordanciador *Concordance* mostra qualquer palavra em seu cotexto de uso. Na lingüística de Corpus, utiliza-se o termo cotexto para se referir ao ambiente lingüístico, ou seja, aos itens que são usados à direita e à esquerda da palavra de busca. Assim, dentro do horizonte estabelecido pelo pesquisador para seu estudo, é possível ver o tipo de companhia que as palavras pesquisadas têm. A palavra de busca (*search word*) aparece no centro da linha de concordância e com uma cor diferente do seu cotexto, para que, assim, ela possa ser facilmente identificada pelo analista. Esse formato de apresentação é denominado de *key word in context* (KWIC). A palavra de busca também é denominada de nóduo e, os itens ao seu redor, são os seus colocados.

Na presente pesquisa, o *Concordancer* foi utilizado para gerar as linhas de concordância com cada um dos verbos estudados. Cada verbo foi inserido no campo *search word* com um asterisco, de forma tal que a busca levasse em consideração a frequência de lemas, ou seja, todas as flexões do verbo pesquisado (e.g. *take*, *takes*, *took*, *taken* e *taking*). A Figura 1 abaixo mostra a tela do *Concordancer* e algumas linhas de concordância com o verbo *take* como nóduo.

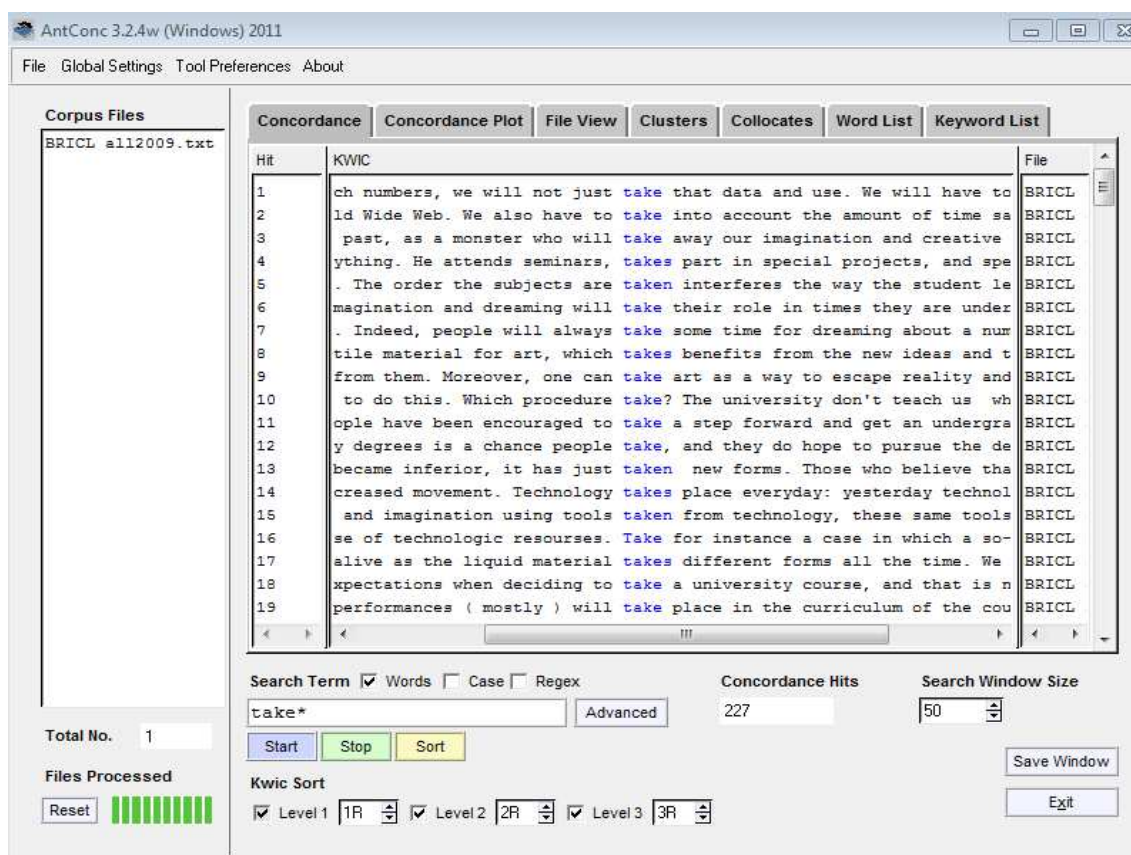


Figura 1: Tela do *Concordancer* e linhas de concordância com o verbo *take*

Primeiramente, através da ferramenta *Concordance* do AntConc, geraram-se as linhas de concordância com os verbos *get*, *give*, *make* e *take* para posterior identificação dos erros nas colocações. Como sugerido por Sinclair (1991) um horizonte de ± 5 foi utilizado, para que, assim, fosse possível considerar os casos de ocorrência de padrões que pré-modificam as colocações, como por exemplo, artigos, advérbios, adjetivos, preposições, etc, que podem ocorrer antes ou depois do verbo núcleo.

Em seguida, as linhas de concordância foram analisadas, uma a uma, para que os erros com as colocações pudessem ser identificados. Importante lembrar que, o termo erro, no presente estudo, é utilizado para se referir às formas de uso diferentes da norma padrão. Ou seja, cada erro foi julgado de acordo com sua conformidade com a norma ou aceitabilidade. Nesta etapa da pesquisa, o primeiro procedimento foi avaliar a aceitabilidade da combinação. Para tanto, os corpora gerais foram utilizados. Assim, como proposto por Nesselhauf (2005), para determinar a aceitabilidade de uma colocação, ela deveria ocorrer, no mínimo, cinco vezes em cinco textos diferentes do BNC e do COCA. Por exemplo, a colocação *get a work* não teve nenhuma ocorrência no BNC ou no COCA e, assim, ela foi tida como inadequada, sendo considerada um erro colocacional.

A próxima etapa consistiu na classificação dos erros existentes nas combinações tidas como inadequadas. Com base na classificação de Nesselhauf (2003) consideramos sete tipos de erros, que são apresentados na Tabela 1 que se segue:

Tipos de Erros		
1	Verbo	Erro na escolha do verbo
2	Substantivo	Erro na escolha do substantivo
3	Uso 1	Combinação existe, mas não é usada corretamente
4	Uso 2	Combinação não existe e não pode ser corrigida pela substituição de um dos elementos
5	Preposição	Erro na escolha da preposição, falta preposição ou preposição desnecessária
6	Determinante	Erro na escolha do artigo, falta artigo ou artigo desnecessário
7	Número	Substantivo usado no singular ao invés do plural ou vice-versa

Tabela 1: Classificação dos tipos de erros identificados nas colocações estudadas (adaptado de Nesselhauf, 2003)

Por exemplo, na combinação *get a work* consideramos que ocorreu um erro tipo 2, ou seja, houve erro na escolha do substantivo. A colocação estaria correta se o aprendiz utilizasse *job* ao invés de *work*, produzindo então *get a job*.

Resultados

Primeiramente, antes de identificarmos as colocações, verificamos a frequência com que os quatro verbos estudados eram utilizados no Br-ICLE e comparamos essa frequência de uso aos dos nativos BNC e do COCA. Para essa comparação, apenas as seções acadêmicas dos dois corpora gerais foram consideradas. O Gráfico 1, a seguir, mostra a frequência normalizada de uso dos quatro verbos nos três corpora.

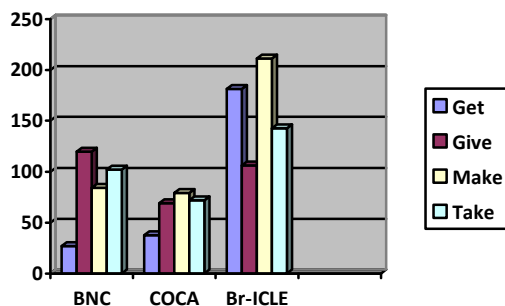


Gráfico 1: Frequência de uso de *get*, *give*, *make* e *take* no BNC, COCA e Br-ICLE.

De maneira geral, nota-se que os aprendizes brasileiros utilizam os verbos *get*, *give*, *make* e *take*, bem mais frequentemente que os nativos. Outros estudos também revelam que aprendizes geralmente sobreusam os verbos de alta frequência (ALTENBERG; GRANGER 2001, HOWARTH, 1998). Pelo gráfico, também percebemos que os aprendizes brasileiros fazem grande uso de do verbo *get*, um verbo que, de acordo com Carter e McCarthy (2006), não é considerado apropriado para ser usado no discurso acadêmico. O verbo *get* é o menos utilizado pelos nativos nos corpora BNC e COCA. Vale ressaltar que esses resultados devem ser interpretados com extremo cuidado, uma vez que eles revelam evidência quantitativa inicial, mas não há nenhuma informação qualitativa acerca do uso desses verbos pelos aprendizes. Apresentamos apenas uma idéia geral acerca da frequência de uso dos verbos.

O próximo passo foi iniciar a investigação proposta pelos objetivos do estudo, ou seja, identificar as colocações utilizadas com os verbos *get*, *give*, *make* e *take* e os erros relacionados a elas. No total, foram identificadas 1307 colocações com os verbos estudados. Sendo 478 colocações com *make*, 322 com *get*, 300 com *take* e, 207 com *give*. Desse total, 312 colocações foram identificadas como sendo inadequadas. Sendo 113 erros com *make*, 78 com *get*, 71 com *take* e, 50 com *give*. Vemos que o número de erros é proporcional ao número de colocações. Esse dado é melhor visualizado no Gráfico 2 abaixo:

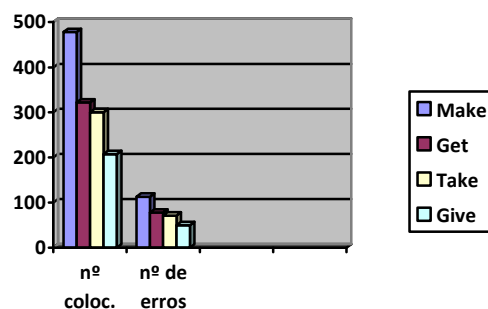


Gráfico 2: N° de colocações verbo nominais e de erros identificados no Br-ICLE

No geral, um quarto das colocações foi considerada inadequada ou errada. Os estudos de Laufer e Waldman (2011), Nesselhauf (2005) e Siyanova e Schmitt (2008) também identificaram que de um terço a um quarto das colocações produzidas pelos aprendizes estava errada.

Como apresentado na seção anterior, consideramos sete tipos de erros de acordo com Nesselhauf (2003): 1. Verbo; 2. Substantivo; 3. Uso 1; 4. Uso 2; 5. Preposição; 6. Determinante e; 7. Número.

Em nosso corpus de estudo, o caso mais comum de erro foi quanto à escolha do verbo. Muitas vezes o aprendiz usou um verbo simples onde outro verbo simples era mais apropriado, como por exemplo, *make success* (*have success*). Outras vezes, o aprendiz elaborou uma construção utilizando os verbos de alta frequência, os quais são mais gerais, onde um verbo mais específico seria mais apropriado: *Get benefit from* (*benefit from*), *take place of* (*replace*).

O segundo tipo de erro mais comum foi quanto ao uso das preposições. Os aprendizes, na maioria das vezes, têm consciência de que uma preposição deve ser utilizada, mas se confundem acerca de qual preposição usar. Por exemplo, *get in the job market* (*get into the job market*), *get married with* (*get married to*), *take for prison* (*take to prison*). Vale mencionar que, um erro bastante comum está relacionado à dificuldade do aprendiz em usar *in* ou *into*.

Apresento, a seguir, a frequência de tipos de erros com cada um dos verbos estudados. As tabelas 2, 3, 4 e 5 trazem também um exemplo de cada tipo de erro ou inadequação identificada. O exemplo de cada erro é apresentado e, em parênteses, uma correção é sugerida. A título de organização, todos os exemplos das tabelas trazem o verbo em sua forma básica, ou no infinitivo sem o *to*. Após a apresentação de cada tabela, comento alguns dos exemplos mostrados.

Tipo de erro	Exemplo	Frequência
Verbo	<i>Get courses (Take courses)</i>	39
Preposição	<i>Get to a situation (Get into a situation)</i>	16
Uso 2 (não existe)	<i>Get out the present (Get away from reality)</i>	10
Uso 1 (existe, mas usou incorretamente)	<i>Get higher (Rise)</i>	06
Substantivo	<i>Get a work (Get a job)</i>	04
Determinante	<i>Get the degree (Get a degree)</i>	03
		Total = 78 erros

Tabela 2: Frequência dos tipos de erros colocacionais com *get*

No caso de *get higher*, observamos pelo contexto, que o aprendiz estava se referindo a ‘rates’ e, assim sendo, este não era o verbo adequado a ser utilizado. Gramaticalmente, pode-se falar que a construção “*rates are getting higher*” é possível na língua inglesa. Mas ao considerarmos dados reais de uso da língua, veremos que a probabilidade dessa combinação ocorrer é baixíssima. No corpus BNC, encontramos apenas duas ocorrências de *get* com *higher* se referindo ao substantivo *rate* e, no COCA, há três ocorrências. Verificamos, nos dois corpora, que, o substantivo *rate* se coloca frequentemente com *rise*. Essa combinação ocorre centenas de vezes tanto no BNB quanto no COCA.

Tipo de erro	Exemplo	Frequência
Verbo	<i>Give a class (Teach a class)</i>	21
Preposição	<i>Give importance of (Give importance to)</i>	10
Determinante	<i>Give power to (Give the power to)</i>	10
Substantivo	<i>Give man (Give people)</i>	06
Uso 2 (não existe)	<i>Give up place for (Replace)</i>	02
Uso 1 (existe, mas usou incorretamente)	<i>Give birth to necessities (Generate)</i>	01
		Total = 50 erros

Tabela 3: Frequência dos tipos de erros colocacionais com *give*

O erro *give a class* parece vir da transferência direta do português da combinação ‘dar aula’. Todavia, em inglês, o substantivo *class* não se coloca com *give* para expressar a ação de dar uma aula. A combinação adequada seria *teach a class*.

No caso de *give man*, o objetivo do aprendiz era generalizar, ou seja, ele utilizou o substantivo ‘homem’ para se referir as pessoas em geral ou a todas as pessoas. Porém, em inglês, utiliza-se o substantivo *people* com esse propósito.

Give birth existe, mas não é usado com o substantivo *necessities*. O aprendiz queria dizer ‘gerar necessidades’. De qualquer forma, a construção é estranha, mas seria menos inadequado dizer *generate necessities*.

Tipo de erro	Exemplo	Frequência
Verbo	<i>Make actions (Take actions)</i>	67
Substantivo	<i>Make usage of (Make use of)</i>	15
Uso 2 (não existe)	<i>Make their ends meet (Make ends meet)</i>	13
Preposição	<i>Make the world go around (Make the world go round)</i>	08
Determinante	<i>Make the people (Make people)</i>	06
Uso 1 (existe, mas usou incorretamente)	<i>Make use of (Use)</i>	04
		Total = 113 erros

Tabela 4: Frequência dos tipos de erros colocacionais com *make*

Make ends meet é uma expressão idiomática e, assim sendo, trata-se de uma combinação de palavras bastante fixa. Assim, provavelmente, o aprendiz não tinha consciência de que essa expressão não permite variação.

Make use of é uma combinação bastante frequente em inglês, mas foi utilizada por um aprendiz com o substantivo *example*. Esse uso não aparece nos corpora gerais e, o mais apropriado, seria utilizar o verbo simples *use*. A sugestão do verbo *use*, nesse caso, é devido à preferência do aprendiz por este verbo. No caso de estudarmos, no BNC e no COCA, os verbos que são mais utilizados com o substantivo *example*, veríamos que há outras opções, tais como *have*, *take* e *give*.

Tipo de erro	Exemplo	Frequência
Verbo	<i>Take a sentence (Receive a sentence)</i>	32
Preposição	<i>Take part on (Take part in)</i>	14
Uso 2 (não existe)	<i>Take a sunny day (Enjoy a sunny day)</i>	08
Uso 1 (existe, mas usou incorretamente)	<i>Take it with me (Incorporate it)</i>	06
Determinante	<i>Take the control of (Take control of)</i>	04
Número	<i>Take advantages (Take advantage)</i>	04
Substantivo	<i>Take activities (Take action)</i>	03
		Total = 71 erros

Tabela 5: Frequência dos tipos de erros colocacionais com *take*

Em *take advantages* o aprendiz não tem consciência da restrição numérica do substantivo *advantage* quando parte da colocação com *take*. Ou seja, *advantage* é um

substantivo contável e seria lógico poder variá-lo quanto ao número, mas a combinação *take advantage* é bastante fixa.

Conclusão

Os resultados corroboram estudos anteriores (HONG *et al.*, 2011, LAUFER; WALDMAN 2011, NESSELHAULF 2005, WANG; SHAW 2008), os quais indicam que aprendizes de inglês de nível avançado têm dificuldades na produção e uso de colocações. Uma possível razão é que as colocações são transparentes semanticamente (ex. *make a decision, give a chance*), uma vez que elas são formadas por palavras individualmente frequentes. Sendo assim, de acordo com Laufer e Waldman (2011), quando as colocações aparecem no insumo, não são tidas como problemáticas por professores e aprendizes. Todavia, produzir colocações corretamente é difícil, pois os aprendizes tendem a se apoiar na transferência em L1 e não consideram, ou não têm consciência das restrições na combinação de palavras e formam as mensagens a partir de palavras individuais ao invés de padrões pré-fabricados (WRAY, 2002).

Talvez, por isso, encontramos, em nosso corpus de estudos, combinações como *give a class* e *do a course* (tradução direta de dar aula e fazer um curso), e de combinações como *take advantages*, na qual, provavelmente, o aprendiz não tem consciência da restrição numérica do substantivo *advantage*, quando parte da colocação com *take*.

Outro possível motivo para a dificuldade dos aprendizes em produzir as colocações com os verbos *get, give, make* e *take* é revelado nos estudos de Altenberg e Granger (2001) e Nesselhauf (2005). Segundo esses autores, os verbos de alta frequência são apresentados logo no início de qualquer currículo de ensino de inglês. O que ocorre, na grande maioria das vezes, é que, após terem sido ensinados eles tendem a ser negligenciados, o que faz com que o aprendiz corra o risco de ter um conhecimento apenas superficial do padrão lexical e gramatical dos mesmos.

Para Granger e Meunier (2008), professores têm que ajudar os aprendizes a perceberem a fraseologia da língua e seu potencial para aquisição da fluência. Howarth (1998) defende que, bem mais eficiente seria ensinar o aprendiz acerca da natureza do fenômeno colocacional e assim desenvolver a consciência dos mesmos acerca dos problemas em potencial, envolvendo as colocações e demais combinações de palavras. Concordo com as autoras e acredito que o ensino de uma língua deve promover a conscientização do aprendiz quanto às colocações e aos padrões da língua. O aprendiz deve perceber que aprender vocabulário não é somente aprender palavras novas, mas se familiarizar com as combinações das mesmas.

Referências

ALTENBERG, B.; GRANGER, S. The grammatical and lexical patterning of MAKE in native and non-native student writing. **Applied Linguistics**, v. 22, p. 173–195, 2001.

CARTER, R; McCARTHY, M. **Cambridge grammar of English**. Cambridge: Cambridge University Press, 2006.

CONKLIN, K.; SCHMITT, N. Formulaic sequences: are they processed more quickly than nonformulaic language by native and nonnative speakers? **Applied Linguistics**, v. 29, n. 1, p. 72-89, 2008.

COWIE, A. P. Phraseology. In: ASHER R.; SIMPSON, J. (Orgs.) **The encyclopedia of language and linguistics**. Oxford: Pergamon Press, 1994, p.168-71.

GRANGER, S.; MEUNIER, F. **Phraseology in foreign language learning and teaching**. Amsterdam: Benjamins, 2008.

HONG, A.; RAHIM, H.; HUA, T.; SALEHUDDIN; K. Collocations in Malaysian English learners' writing: a corpus based error analysis. **The Southeast Asian Journal of English Language Studies**, v. 17, p. 31-44, 2010.

HOWARTH, P. Phraseology and second language proficiency. **Applied Linguistics**, v. 19, n. 1, p. 24-44, 1998.

LAUFER, B.; WALDMAN, T. Verb-noun collocations in L2 writing. **Language Learning**, v. 61, n. 2, p. 647-672, 2011.

NESSSELHAUF, N. The use of collocations by advanced learners of English and some implications for teaching. **Applied Linguistics**, v.24, p. 223–242, 2003.

NESSSELHAUF, N. **Collocations in a learner corpus**. Amsterdam: Benjamins, 2005.

SCHMITT, N. (Ed.). **Formulaic sequences**. Amsterdam: Benjamins, 2004.

SINCLAIR, J. M. **Corpus, concordance, collocation**. Oxford: Oxford University Press, 1991.

SIYANOVA, A.; SCHMITT, N. L2 learner production and processing of collocation: a multi-study perspective. **Canadian Modern Language Review**, v. 64, p. 429–458, 2008.

WANG, Y.; SHAW, P. Transfer and universality: collocation use in advanced Chinese and Swedish learner English. **ICAME Journal**, v. 32, p. 201-232, 2008.

WRAY, A. **Formulaic language and the lexicon**. Cambridge: Cambridge University Press, 2002.