# ANÁLISE DE *COLLOCATIONS* NA ESCRITA ACADÊMICA DA QUÍMICA: contribuições para o ensino de inglês acadêmico nas ciências exatas

**Danilo Duarte Costa[1], Mariélen de Oliveira Amaral [1], Clever Luiz Rosa Junior [1]**

*[1]Universidade Federal dos Vales do Jequitinhonha e Mucuri (UFVJM) - Instituto de Ciência e Tecnologia (ICT), Diamantina, Brasil (danilo.costa@ufvjm.edu.br)*

*Resumo:* Este trabalho se propõe a gerar e analisar uma lista de *collocations*/colocações (frequentes combinações de palavras) presentes no inglês acadêmico da Química. A metodologia ampara-se nos pressupostos da linguística de corpus e os dados foram obtidos a partir do *Corpus of Textbooks of Chemistry*. Os resultados revelam alta frequência de colocações lexicais e gramaticais no corpus. Demonstra-se também a aplicabilidade dos resultados no ensino de Inglês para Fins Acadêmicos nas ciências exatas.

*Palavras-chave*: linguística de corpus; inglês acadêmico; química; colocações; ensino.

## INTRODUÇÃO

Ao cunhar a frase "*you shall know a word by the company it keeps*", Firth (1957) sugere que o significado de uma palavra pode ser apreendido a partir de sua relação com as demais palavras que a acompanham ou a circundam. É neste contexto que se inserem as *collocations* (doravante colocações), definidas por Hoey (2005) como "recorrentes combinações de palavras". Ou seja, uma combinação de duas ou mais palavras que, com frequência, ocorrem juntas em um texto.

Embora autores como Biber et al. (2002) entendam as colocações como uma coocorrência de palavras lexicais (e.g. a combinação de um adjetivo com um substantivo), outros pesquisadores (Granger e Paquot, 2008; Menon e Mukundan, 2012) consideram também as combinações de vocábulos onde há a presença de palavras gramaticais (e.g. preposições, pronomes). Essas sequências de palavras podem ser, portanto, classificadas em dois tipos, quais sejam: colocações lexicais (e.g. uma sequência de dois - ou mais - substantivos) e colocações gramaticais (e.g. um substantivo ou verbo junto a uma preposição).

No contexto do ensino de línguas estrangeiras, as colocações desempenham um papel fundamental no que diz respeito à aprendizagem do idioma. Durrant (2009) defende que elas formam parte essencial do léxico que precisa ser apreendido pelos alunos, ou ainda, como afirma Nation (2001), conhecer os padrões colocacionais de uma língua nos possibilita usá-la de forma fluente e apropriada.

Além disso, na esfera da escrita acadêmica (textos de, por exemplo, livros acadêmicos e artigos científicos), listas de colocações podem ser úteis para nortear e subsidiar a tomada de decisões referente ao conteúdo a ser incluído nas aulas e materiais didáticos. (Chon e Shin, 2013)

Com efeito, o inglês acadêmico, tratado de forma geral, apresenta características linguísticas próprias que o distingue de outros gêneros/registros textuais (Biber, 2006; Biber e Gray, 2010); no entanto, especificidades são também observadas ao se investigar textos das diferentes áreas do conhecimento separadamente (Charles, 2006; Hyland e Bondi, 2006).

No âmbito do ensino de inglês acadêmico nas ciências exatas, estas especificidades representam um grande desafio para os professores responsáveis pelo ensino de línguas estrangeiras. Como aponta Ward (2007), esses professores, em geral, detêm pouco conhecimento referente a, por exemplo, o vocabulário técnico de áreas como a engenharia.

Soma-se a isso o fato de que, no Brasil, há uma carência de livros e materiais didáticos que abordam as particularidades do inglês acadêmico, sobretudo em disciplinas voltadas ao ensino da habilidade leitora, como o Inglês Instrumental (Costa, 2020). Consequentemente, na tentativa contornar essa questão, professores se veem impelidos a criar seus próprios materiais (Gilmore e Millar, 2018).

Dado o exposto, utilizando ferramentas computacionais e métodos embasados na linguística de corpus, o presente estudo tem por objetivo gerar e analisar uma lista de colocações encontradas em livros didáticos de química escritos em língua inglesa, a fim de se identificar padrões lexicais (de vocabulário) e sintáticos (gramaticais) presentes nesses textos. Objetiva-se, ainda, demostrar que a identificação e análise colocações, a partir de dados

empíricos, pode contribuir para o aprimoramento do ensino de Inglês para Fins Acadêmicos nas ciências exatas, sobretudo nas disciplinas de Inglês Instrumental (*English for Reading*) e Escrita Acadêmica (*Academic Writing*).

## MATERIAL E MÉTODOS

Para a extração e identificação das colocações, utilizou-se dados empíricos obtidos a partir do TB-Chem, *Corpus of Textbooks of Chemistry* (em português, Corpus de Livros Didáticos de Química), compilado por Costa (2020) e composto por um total de aproximadamente 500 mil *tokens* (palavras). Este corpus é constituído por textos escritos em língua inglesa, retirados de livros didáticos de química adotados em cursos de graduação de universidades americanas e britânicas. A compilação do corpus, descrita em Costa (2020), seguiu rigorosos procedimentos metodológicos de forma a atender aos critérios de representatividade (Biber, 1993) e balanceamento (McEnery et al., 2006).

A noção de representatividade estabelece que uma amostra precisa incluir toda a gama de variabilidade de uma dada população, o que, no caso do TB-Chem, implica em garantir que todas as subáreas da química estejam representadas no corpus. Para tanto, foram estabelecidas as disciplinas que compõem este campo de estudo, a saber: Química Geral, Química Analítica, Química Orgânica, Química Inorgânica e Físico-Química. O balanceamento do corpus, por sua vez, diz respeito à distribuição proporcional do número de palavras entre as subáreas que o compõem. Na compilação do TB-Chem, o balanceamento foi atendido a partir da obtenção de, aproximadamente, 100 mil palavras por subárea.

Para a identificação das colocações, foi, primeiramente, gerada uma lista de palavras-chave do TB-Chem, tomadas como palavras centrais (*node words*) das colocações; ou seja, as colocações foram geradas tendo por base as palavras à direita e à esquerda das palavras-chave. Vale esclarecer que, neste estudo, considera-se palavra-chave como sendo uma palavra que, do ponto de vista estatístico e com base em sua frequência de ocorrência, se destaca por ser própria/característica de um determinado tipo de texto (Baker et al., 2006).

A lista de palavras-chave foi gerada através da comparação do TB-Chem com um corpus de referência, o OANC (*Open American National Corpus*), que possui aproximadamente 15 milhões de palavras e é composto por textos de registros orais e escritos do inglês americano. A escolha pelo OANC se deu a partir da compreensão de que, segundo Berber-Sardinha (2005), para se obter uma lista de palavras-chave, o corpus de estudo precisa ser comparado a um corpus de referência que seja, ao menos, cinco vezes maior.

A lista foi, então, obtida através do software AntConc 3.5.9 (Anthony, 2020), que calcula a distribuição e frequência das palavras do corpus de estudo (a partir do teste log-likelihood, com valor de $p < 0.05$) em comparação ao corpus de referência. As palavras-chave foram organizadas em ordem decrescente a partir de seu índice de chavicidade.

Uma vez obtida a lista com as 50 palavras-chave mais frequentes no TB-Chem, iniciou-se a identificação de seus colocados (i.e as palavras que se ligavam a sua direita e esquerda), utilizando-se o recurso *cluster*/*n-grams* desse mesmo software.

Com base em Biber et al. (2004) e em Durrant (2009), estabeleceu-se que as colocações a serem analisadas deveriam obedecer aos seguintes critérios: i) colocações com o mínimo de 40 ocorrências por milhão de palavras no corpus de estudo; ii) colocações com extensão mínima de duas e de no máximo 4 palavras; iii) colocações com ocorrência em, no mínimo, 20% dos textos do corpus de estudo, e iv) combinação de palavras que sejam parte de um mesmo sintagma.

Este último critério (iv) se faz necessário uma vez que o AntConc gerou algumas sequências de palavras que não formam uma unidade, como ilustrado abaixo no exemplo A, onde *molecules* e *the* pertencem a diferentes orações e, portanto, não formam uma colocação.

(A) (...) *or covalent diatomic* ***molecules***. ***The*** *ability to conduct electrons requires a reasonable* (...) [TB-Chem 07]

Obedecidos esses critérios, as colocações foram identificadas e analisadas, como apresentado na próxima seção.

## RESULTADOS E DISCUSSÃO

Conforme descrito na seção anterior, partiu-se de uma lista de palavras-chave para, então, se determinar quais palavras frequentemente as acompanham. Assim, considerou-se a palavra-chave como a palavra-nódulo e as palavras que vêm junto a ela (tanto à sua esquerda quanto à sua direita) como seus colocados. A combinação de uma palavra-chave com seu(s) colocado(s), portanto, formam uma colocação.

Em primeira análise, nota-se que os substantivos figuram como a vasta maioria das palavras-chave geradas a partir do TB-Chem. Porém, alguns deles se mostraram mais produtivos que outros quanto à formação de colocações. Por exemplo, substantivos como *atoms* e *energy* formaram 28 e 38 colocações, respectivamente, ao passo que *tritration* e *electrode* deram origem a somente 4 e 2 colocações cada. Além disso, todas as palavras-chave formaram colocações com palavras situadas imediatamente a sua direita e

esquerda, com a exceção de *figure*, que formou colocações com palavras a sua esquerda somente.

Como ilustração, o Quadro 1 apresenta as 5 palavras-chave (*node words*) mais frequentes no TB-Chem, juntamente com as colocações formadas. Para simplificar a visualização, a palavra-chave foi abreviada e seus colocados italicizados. Por exemplo, *of the atoms* e *carbon atoms*, são apresentadas nas formas [*of the* a.] e [*carbon* a.], respectivamente.

Quadro 1. Exemplos de colocações encontradas no TB-Chem.

| | Colocados à direita | *Node word* | Colocados à esquerda |
|---|---|---|---|
| 1 | a. *are*, a. *of (the)*, a. *in the*, a. *or*, a. *to*, a. *that*, a. *in a*, a. *is*, a. *have*, a. *with* | atoms | *(the) hydrogen* a., *(of) carbon* a., *(of) (the)* a., *oxygen* a., *two* a., *two hydrogen* a., *number of* a., *individual* a., *of hydrogen* a., *metal* a., *the carbon* a. |
| 2 | e. *of (the)*, e. *is*, e. *levels*, e. *to*, e. *than (the)*, e. *level*, e. *in*, e. *of a*, e. *released*, e. *that*, e. *as*, e. *required* | energy | *(of) the* e., *of* e., *(the) potential* e., *kinetic* e., *free* e., *thermal* e., *amount of* e., *(the) activation* e., *lower* e., *an* e., *ionization* e., *higher* e., *total* e. *the total* e., *this* e., *heat* e., *high* e., *same* energy |
| 3 | e. *in (the)*, e. *are*, e. *to*, e. *of*, e. *from*, e. *that* e. *in a* | electrons | *of (the)* e., *the* e., *two* e., *valence* e., *pair(s) of* e., *more* e., *number of* e., *energy* e., *unpaired* e., |
| 4 | m. *are*, m. *in (the)*, m. *that*, m. *of*, m. *have*, m. *with*, m. *can*, m. *to*, m. *at* | molecules | *(the) water* m., *(of) (the)* m., *in* m., *diatomic* m., *between* m., *organic* m., *or* m., *atoms and* m., *these* m., *solvent* m., *nonpolar* m., |
| 5 | c. *atom*, c. *atoms (of)*, c. *bonds*, c. *dioxide*, c. *atom is*, c. *atoms in*, c. *is*, c. *atoms of* | carbon | *(of) (the)* c., *each* c., *a* c. |

A partir do Quadro 1, observar-se que no TB-Chem há a presença não só de colocações lexicais (e.g. *termal energy*, *carbon atoms*), mas também de várias colocações gramaticais (e.g. *electrons that*, *molecules with*). Com efeito, quando analisado sob a perspectiva pedagógica, é positivo o fato de os dados revelarem um grande número de colocações gramaticais.

Ao tratar desta questão, Durrant (2009) aponta que, no âmbito do ensino da língua inglesa, um dos benefícios de se usar uma lista de colocações gramaticais diz respeito à possibilidade de podermos usá-las como forma de chamar a atenção dos alunos para palavras e padrões essenciais à aprendizagem. O autor salienta que, por vezes, colocações gramaticais são negligenciadas por pesquisadores, dado que elas não possuem sentido/significado tão explícito, como se observa nas colocações lexicais, mas defende que colocações deste tipo devem ser tratadas como parte dos objetivos de aprendizagem.

No nosso caso, por exemplo, a colocação lexical *kinetic energy (energia cinética)* claramente forma uma unidade de sentido mais explícita que a colocação gramatical *energy than (energia que)*. No entanto, a presença de *than*, com função de preposição, indica que uma comparação está sendo estabelecida no discurso. Como ilustrado no exemplo 01

(01) (…) *less stable separate atoms have higher potential* ***energy than*** *the more stable atoms* [Chem-04]

De fato, como mostrado no Quadro 1, outros colocados frequentes da palavra *energy* são os adjetivos *low (baixo)* e *high (alto)* em sua forma comparativa (acrescidos do sufixo *-er)*, formando os pares *higher energy* e *lower energy*. Portanto, por serem recorrentes, sugere-se que especial atenção seja dada a estas colocações, incluindo-as em materiais de ensino/avaliação que tratam, por exemplo, de tópicos como *comparative forms of adjetives*, ou *comparative clauses*. Afinal, para o estudante das ciências exatas, uma expressão como *higher energy* é notoriamente mais relevante e contextualizada que outras retiradas de textos e livros não-acadêmicos e comumente usadas em aula (e.g. *younger brother*, *taller man*).

É importante ressaltar que, em relação à escolha e delimitação do vocabulário ou estruturas gramaticais a serem ensinados, é desaconselhável que o professor recorra a sua intuição (Schmitt, 2010; Xiao, 2015). Nesse sentido, a análise das colocações com base em sua estrutura sintática provou ser uma boa forma de se identificar padrões comuns dessas combinações de palavras. No TB-Chem, chamam atenção algumas estruturas sintáticas recorrentes, como mostrado no Quadro 2.

Quadro 2. Estruturas sintáticas de colocados do TB-Chem.

| | Estrutura | Exemplos |
|---|---|---|
| A | substantivo + substantivo | *hydrogen atoms, water molecules, vapor pressure* |
| B | adjetivo + substantivo | *kinetic energy, chemical reaction, electromagnetic radiation* |
| C | determinante + (adjetivo) + substantivo | *these compounds, the molecular orbital, a negative charge* |
| D | sintagma preposicionado | *of molecules, as an example, in water* |
| E | substantivo + preposição/ sintagma preposicionado | *atoms with, molecule of, particles in, pairs of electrons* |
| F | substantivo + verbo | *atom is, orbitals have, molecules can, reaction occurs* |

Dentre as estruturas sintáticas apresentadas no Quadro 2, um dos padrões observados - e que pode ser explorado para fins pedagógicos - diz respeito à modificação dos substantivos. Em língua inglesa, especificamente, como tratado em Biber et al. (1999), um substantivo pode ser modificado por palavras que o antecedem, como adjetivos (exemplo 02), ou mesmo por outro substantivo (exemplo 03). Nos exemplos, as colocações estão destacadas em negrito e os pré-modificadores sublinhados.

(02) (...) *and sulfur that form **covalent bonds** and have* (...). [TB-Chem 09]

(03) *In water, the number of **hydrogen bonds** is therefore maximised by* (...). [TB-Chem 07]

No contexto do ensino do Inglês para Leitura (Inglês Instrumental), que geralmente atende a falantes do português que possuem pouca ou nenhuma proficiência em língua inglesa, a pré-modificação dos substantivos pode representar um obstáculo à compreensão do texto. Tendo o português como referência, não é raro que alunos interpretem incorretamente o significado de um sintagma nominal, sobretudo quando há a combinação de substantivo+substantivo. Por exemplo, uma colocação como *hydrogen bonds* pode ser erroneamente interpretada como 'hidrogênio de ligações', ao invés do correto, 'ligações de hidrogênio.

Dentre os padrões sintáticos analisados, observou-se também a frequente pós-modificação de substantivos, como nas estruturas substantivo+*that* (evidenciando um substantivo sendo modificado por uma oração relativa/adjetiva) e substantivo+*with* (um substantivo modificado por um sintagma preposicionado). Estas são formas intercambiáveis de se modificar um substantivo, como ilustrado nos exemplos 4a e 5a (frases autênticas retiradas do TB-Chem), e nos exemplos 4b e 5b (frases semanticamente equivalentes às anteriores, mas sendo alternadas as partículas *that* e *with*).

(4a) (...) *there are atoms **that** have enough energy to emit radiation*. [TB-Chem 04]

(4b) (...) *there are atoms **with** enough energy to emit radiation*.

(5a) (...) *molecules **with** partial or temporary charges are attracted to* (...) [TB-Chem 02]

(5b) (...) *molecules **that** have partial or temporary charges are attracted to* (...).

Curiosamente, no ensino de inglês para fins acadêmicos (sobretudo no inglês instrumental – ver, por exemplo, Souza et al., 2005), ao se tratar das orações adjetivas (*relative clauses*), é comum a afirmação de que substantivos e sintagmas nominais podem ser seguidos pelos pronomes relativos *that* ou *which*, como ilustrado nos exemplos 06 e 07.

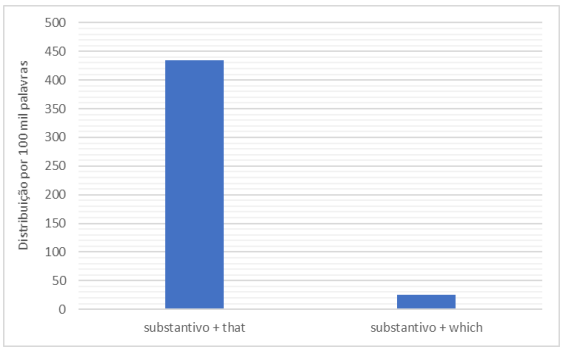
(06) *The minimum kinetic energy **that** the reactants must have in order to form* (...). [TB-Chem 07]

(07) *They possess kinetic energy **which** results in a pressure* (...). [TB-Chem 07]

Apesar de os pronomes *that* e *which* serem equivalentes nos exemplos 06 e 07, existe ali uma grande diferença no que diz respeito à frequência com que ocorrem no TB-Chem. Este é um ponto que merece atenção, pois, como defende Nation (2001), a frequência é um importante critério usado para se decidir que palavras/estruturas linguísticas devem ser apresentadas aos aprendizes de uma língua estrangeira. No nosso caso, dentre as colocações obtidas a partir do TB-Chem, a combinação substantivo+*that* (e.g. *reactions that, compounds that*) mostrou-se bastante produtiva. Por outro lado, não houve ocorrências de colocações que formassem o par substantivo+*which*.

Ao serem comparadas as ocorrências destas duas estruturas em todo o corpus de estudo – e não somente em relação às colocações – (Figura 1), observou-se que *that* é altamente mais produtivo que *which*.

Fig. 1. Distribuição de substantivos pós-modificados por *that* e *which* no TB-Chem.




Os dados apresentados na Figura 1 evidenciam a diferença das estruturas sintáticas quanto à frequência com que ocorrem nos livros de química, embora elas exerçam a mesma função na frase. A cada 100 mil palavras no TB-Chem, as combinações de substantivo+*that* e substantivo+*which* ocorrem cerca de 430 e 35 vezes, respectivamente. Por isso, no que tange o ensino, acreditamos ser importante que a estrutura mais produtiva seja privilegiada (por exemplo, sendo repetida mais vezes nos exercícios e atividades didáticas), já que, por ser mais frequente, se trata de uma construção linguística com a qual os alunos têm maior chance de se deparar.

Por fim, assim como Friginal (2018), destacamos a importância dos estudos linguísticos empíricos baseados em corpora, que nos permitem oferecer aos alunos dados relevantes sobre vocabulário e gramática, usados em contextos naturais e autênticos.

## CONCLUSÃO

Este estudo se propôs a investigar a escrita acadêmica em inglês (inglês acadêmico) da área da química, a partir da análise de uma lista de *collocations*/colocações (frequentes combinações ou sequências de palavras presentes em um texto). Os dados foram gerados a partir do TB-Chem (*Corpus of Textbooks of Chemistry*), composto por textos extraídos de livros didáticos de química escritos em língua inglesa e adotados em cursos de graduação de universidades americanas e britânicas.

Os dados revelaram que substantivos com alta ocorrência no corpus, como *atoms* e *energy*, mostraram-se altamente produtivos quanto à formação de colocações. Além disso, destaca-se que foram identificadas não só colocações lexicais (e.g. *atomic orbitals*), mas também um grande número de colocações gramaticais (e.g. *molecules with*).

Chama atenção o fato de as colocações gramaticais terem se mostrado como um bom indicador de estruturas sintáticas tipicamente encontradas no TB-Chem. Dentre estas, destaca-se a pós-modificação de substantivos, evidenciada, neste estudo, pela alta frequência de combinações do tipo substantivo+*that* (e.g. *electrons that have a negative charge*) e substantivo+*with* (e.g. *electrons with a negative charge*).

Ao se investigar padrões recorrentes de estruturas lexicais e sintáticas, demonstrou-se que as colocações podem ser utilizadas para embasar e nortear o ensino e a produção de materiais didático-pedagógicos voltados ao Inglês para Fins Acadêmicos, incluindo disciplinas como Inglês Instrumental e Escrita Acadêmica (*Academic Writing*), no contexto das ciências exatas.

Salientamos que a análise de colocações obtidas com base em dados empíricos (dados obtidos a partir de um corpus) nos permite detectar e conhecer características linguísticas específicas de textos acadêmicos; características estas que não poderiam ser identificadas simplesmente a partir de nossa intuição ou experiência.

Acreditamos, também, que os procedimentos metodológicos adotados neste estudo podem ser replicados para se investigar colocações presentes em textos das demais áreas do conhecimento.

## REFERÊNCIAS

ANTHONY, L. *AntConc* (Version 3.5.9) [Computer Software]. Tokyo, Japan: Waseda University, 2020.

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

BERBER SARDINHA, Tony. A influência do tamanho do corpus de referência da obtenção de palavras-chave usando o Programa Computacional Wordsmith Tools. *ESPecialist*, v. 26, n. 2, p. 183-204, 2005.

BIBER, D. Representativeness in corpus design. *Literary and linguistic computing*, v. 8, n.4, p. 243-257, 1993.

BIBER, Douglas. *University language*. Amsterdam: John Benjamins, 2006.

BIBER, D.; CONRAD, S.; LEECH, G. *Longman Student Grammar of Spoken and Written English*. Pearson Education. 2002.

BIBER, Douglas; CONRAD, Susan; CORTES, Viviana. If you look at…: Lexical bundles in university teaching and textbooks. *Applied linguistics*, v. 25, n. 3, p. 371-405, 2004.

BIBER, Douglas; GRAY, Bethany. Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes*, v. 9, n. 1, p. 2-20, 2010.

BIBER, D., JOHANSSON, S., LEECH, G., CONRAD, S., FINEGAN, E., QUIRK, R. *Longman grammar of spoken and written English*. London: Longman, 1999.

CHARLES, Maggie. Phraseological patterns in reporting clauses used in citation: A corpus-based study of theses in two disciplines. *English for Specific Purposes*, v. 25, n. 3, p. 310-331, 2006.

CHON, Yuah V.; SHIN, Dongkwang. A corpus-driven analysis of spoken and written academic collocations. *Multimedia-Assisted Language Learning*, v. 16, n. 3, p. 11-38, 2013.

COSTA, D. *A corpus-based investigation of the language of chemistry and physics textbooks and its implications for English for reading*. Tese de doutorado. UFMG. 2020.

DURRANT, P. Investigating the viability of a collocation list for students of English for academic purposes. *English for Specific Purposes*, v. 28, n. 3, p. 157-169, 2009.

FIRTH, John R. Papers in Linguistics 1934–1951. London: Oxford University Press. 1957

FRIGINAL, E. *Corpus linguistics for English teachers*. New York: Routledge, 2018.

GILMORE, Alexander; MILLAR, Neil. The language of civil engineering research articles: A corpus-based approach. *English for Specific Purposes*, v. 51, p. 1-17, 2018.

GRANGER, S.; PAQUOT, M. Disentangling the phraseological web. In: GRANGER, S; MEUNIER, Fanny (Ed.). *Phraseology: An interdisciplinary perspective*. John Benjamins Publishing, 2008

HOEY, Michael. *Lexical priming: A new theory of words and language*. Routledge, 2005.

HYLAND, Ken; BONDI, Marina (Ed.). *Academic discourse across disciplines*. Peter Lang, 2006.

MCENERY, A.; XIAO, Z.; TONO, Y. *Corpus-based Language Studies: An advanced resource book*. London / New York: Routledge, 2006.

MENON, S; MUKUNDAN, J. Collocations of High Frequency Noun Keywords in Prescribed Science Textbooks. *International Education Studies*, v. 5, n.6, p.149-160, 2012.

NATION, Ian SP; NATION, I. S. P. *Learning vocabulary in another language*. Cambridge: Cambridge university press, 2001.

SCHMITT, Norbert**.** *Researching vocabulary: A vocabulary research manual*. Springer, 2010.

SOUZA, A., ABSY, C., da COSTA, G., de MELLO, L. *Leitura em língua inglesa: uma abordagem instrumental*. Disal Editora. 2005.

WARD, J. Collocation and technicality in EAP engineering. *Journal of English for Academic Purposes*, v. 6, n. 1, p. 18-35, 2007.

XIAO, Richard. Collocation (chapter 6). In: BIBER, Douglas; REPPEN, Randi (Ed.). *The Cambridge handbook of English corpus linguistics*. Cambridge University Press, 2015.