Digital Receipt

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author: Barli Bram

Assignment title: Periksa similarity

Submission title: Frequent Collocation of Adjective-Noun and Noun-Adjective ...

File name: ocation_of_Adjective-Noun_and_Noun-Adjective_Corpus_Ana...

File size: 397.55K

Page count: 8

Word count: 3,397

Character count: 18,020

Submission date: 11-Jul-2022 08:36AM (UTC+0700)

Submission ID: 1868904825

FREQUENT COLLOCATION OF ADJECTIVE-NOUN AND NOUN-ADJECTIVE: CORPUS ANALYSIS

Fransiska Selvy Wulandari
Sanata Dharma University, Yogyakarta
selsaveria@gmail.com

Barli Bram
Sanata Dharma University, Yogyakarta
barli@usd.ac.id

ABSTRACT

This paper investigated frequent collocations used by English learners. Specifically, it is a corpus-based study which focuses on the frequencies of 15 pairs of adjective-noun and noun-adjective collocations. The frequency was examined in two corpora, namely the Corpus of Contemporary American English (COCA) and the British National Corpus managed by Brigham Young University (BYU-BNC). The collocations are presented in the form of pairs to give the comparisons of frequencies between frequently used collocations and less frequently used ones, as seen in the corpora although they have a similar meaning. The present study aims to help English learners to be familiar with the more appropriate word choice used in their writing and speaking.

Keywords: collocation, corpus, COCA, BYU-BNC, English learner

INTRODUCTION

Collocation merges two or more matched words to form certain meanings. Chang et al. (2008) state that collocation was introduced firstly by Firth (1957) as the word combination like 'make the decision' and 'endure the pain' (p. 258). Collocation is "a succession of two or more words that must be learned as an integral whole and not pieced together from its component parts" (Palmer, 1933; Kennedy, 2003). Nation (2001) defines collocation as "items which frequently occur together and have some degrees of semantic unpredictability' (p. 317).

Learning English collocations is one of the important measures for English learners to improve their English proficiency. English learners, especially the beginner ones, sometimes apply some uncommon collocations, such as *powerful coffee* instead of *strong coffee*. It happened simply because they rely on the dictionary and translate every word from their native language into English. Some learners are not accustomed to the frequently used collocation and preferred to use the less frequently used or untypical one. Both *powerful coffee* and *strong coffee* are grammatically correct, but *powerful* usually is more suitable to collocate with *engine* which refers to the powerful work of a machine. Collocation combines two words, but if each of the word is separated, they still stand with a similar meaning when the two words are combined, for example, *dark* and *chocolate*. The *dark chocolate's* color is dark brown/chocolate; thus, *dark chocolate* refers to a kind of food named as *chocolate* which color looks *dark* and contains more pure chocolate in it (compared to another kind of chocolate), and tastes a bit bitter.

Frequent Collocation of Adjective-Noun and Noun-Adjective Corpus Analysis

by Bram Barli

FREQUENT COLLOCATION OF ADJECTIVE-NOUN AND NOUN-ADJECTIVE: CORPUS ANALYSIS

Fransiska Selvy Wulandari
Sanata Dharma University, Yogyakarta
selxaveria@gmail.com

Barli Bram
Sanata Dharma University, Yogyakarta
barli@usd.ac.id

ABSTRACT

This paper investigated frequent collocations used by English learners. Specifically, it is a corpus-based study which focuses on the frequencies of 15 pairs of adjective-noun and noun-adjective collocations. The frequency was examined in two corpora, namely the Corpus of Contemporary American English (COCA) and the British National Corpus managed by Brigham Young University (BYU-BNC). The collocations are presented in the form of pairs to give the comparisons of frequencies between frequently used collocations and less frequently used ones, as seen in the corpora although they have a similar meaning. The present study aims to help English learners to be familiar with the more appropriate word choice used in their writing and speaking.

Keywords: collocation, corpus, COCA, BYU-BNC, English learner

INTRODUCTION

Collocation merges two or more matched words to form certain meanings. Chang et al. (2008) state that collocation was introduced firstly by Firth (1957) as the word combination like 'make the decision' and 'endure the pain' (p. 258). Collocation is "a succession of two or more words that must be learned as an integral whole and not pieced together from its component parts" (Palmer, 1933; Kennedy, 2003). Nation (2001) defines collocation as "items which frequently occur together and have some degrees of semantic unpredictability' (p. 317).

Learning English collocations is one of the important measures for English learners to improve their English proficiency. English learners, especially the beginner ones, sometimes apply some uncommon collocations, such as *powerful coffee* instead of *strong coffee*. It happened simply because they rely on the dictionary and translate every word from their native language into English. Some learners are not accustomed to the frequently used collocation and preferred to use the less frequently used or untypical one. Both *powerful coffee* and *strong coffee* are grammatically correct, but *powerful* usually is more suitable to collocate with *engine* which refers to the powerful work of a machine. Collocation combines two words, but if each of the word is separated, they still stand with a similar meaning when the two words are combined, for example, *dark* and *chocolate*. The *dark chocolate's* color is dark brown/chocolate; thus, *dark chocolate* refers to a kind of food named as *chocolate* which color looks *dark* and contains more pure chocolate in it (compared to another kind of chocolate), and tastes a bit bitter.

The recent study from Kennedy (2003) discusses the amplifier collocation in British National Corpus, exploring the use of amplifier which functions as booster (*badly, clearly, extremely),* and maximizer *(absolutely, completely, totally, fully)*, which forms the collocation such as, *completely eliminated, extremely difficult*. Laufer and Waldman (2011) in their study to investigate the use of English Verb-Noun collocation in the native speaker writing context conducted an error analysis study. Besides, they analyzed a larger corpus using a 300,000 words learner corpus and LOCNES (a young adult corpus). They found that the learners in three different levels of English proficiency master lesser collocation than the native ones.

In addition, Shin and Nation (2008) also conducted a study on collocation, focusing on the most frequent collocations in spoken English. The research was also a corpus-based research taken from BNC as its primary data source. From their study, Shin and Nation (2008) assumed that one of the effective ways to enhance the learners' language competency and native-like word choice was by studying collocations. The most popular or commonly used collocations are considered more useful because they frequently appear in various contexts (Shin & Nation, 2008: 340).

Considering the previous studies above, it is urgent to explore further collocations in English. The present study, therefore, focuses on the frequency of adjective-noun and noun-adjective collocations, specifically comparing the pairs of collocations and discovering which ones that have more occurrences and which are used more frequently in corpora. This study employs the corpus-based study which focuses on two corpora as the data sources.

Collocation constructs the learners' language production through the improvement of proficiency and native-like choice (Shin & Nation, 2008; Kennedy, 2003: 340). The word choice that learner choose to form collocation can show their language mastery and competency. The more popular one, the more it is similar to a native-like selection, and the more the learners show their English competency through their speaking or writing. There are various ways to define things, but not all of them seem natural to English native speakers (Pawler & Syder, 1983; Shin & Nation, 2008: 340). For sure, this study focuses on English collocation. For instance, '*stop here*' can also be said as '*halt the step*'. However, people do not use the phrase '*halt the step*' commonly, so '*stop here*' is considered more natural. Both phrases are grammatically correct, but the native speaker do not speak that way. This kind of case could cause a problem for the learners when they focus only on grammar, because they may produce a grammatically correct sentence but it does not seem natural or native-like (Pawler & Syder, 1983; Shin & Nation, 2008: 340). The unnatural sentence is sometimes grammatically correct, but being revised by the teacher because it is uncommonly used by native English speakers.

Collocation that contains the same words, such as "narrow in narrow escape, and slim in slim chance" is more difficult for learners to memorize because of the similarity, or the semantically related words (Tinkham, 1997; Waring, 1997; Finkbeiner & Nicol, 2003; Erten & Tekin, 2008; Webb & Kagimoto, 2011; Boers, Demecheleer, Coxhead, & Webb, 2013: 2). Memorizing the relatively different is easier than the similar one. Learners with higher English proficiency may produce the malformed collocation in the impromptu speech (Boers,

Demecheleer, Coxhead, & Webb, 2013: 3). It might happen related to the difficulties in memorizing the similar words.

Learners who possessed fine ideas sometimes get lower grades because they do not know and do not implement the important collocation and tend to use the longer sentence, which does not accurately represent their aim and contains errors in grammar (Hill, 1999; Lauder & Waltman, 2011: 468). This case commonly happens to learners who create the utterances using their own language which may be hardly understood by readers.

Collocation can be categorized as rigid collocation and elastic collocation. The rigid collocation contains words that are touching each other, for example *strong coffee*. On the other hand, the elastic collocation is separated by other word such as *endure the pain* (Chang et al., 2008: 258).

METHOD

This paper is a corpus-based study, whose data were collected from COCA (Corpus of Contemporary American English) and BYU-BNC (British National Corpus managed by Brigham Young University). The data collected from the corpus were the usage frequency of each collocation. The pairs of collocations were chosen randomly from some web sources, which provided the lists of noun-adjective or adjective-noun collocation. COCA is a corpus that contains more than 520 million words of texts which were collected from 1990-2015. Besides, BYU-BNC consists of 100 million words collected from 1980s-1993. The two big corpora were used in this research to compare which collocation in each pair that has a higher frequency (used more frequently) in the corpora.

In the data gathering process, the researchers collected 15 pairs of collocations from some web sources. There was a selection process, which required some criteria, namely the pair of collocation should have a similar meaning, and mentioned in the web sources as a frequently used collocation. The pairing was made to compare two collocations that have similar meanings, but one of them is more popularly used than another. It helps the learners to know the more commonly used or natural or native-like collocation. Then, the researchers searched the frequency of each collocation in COCA and BYU-BNC. Usually, the frequency of collocated words in COCA is larger than in BYU-BNC, considering that COCA is a larger corpus. The data presented in a corpus were limited, but they contained quantitative bias, affected by the characteristic of a corpus, and difficult to be elaborated (Cook, 1998; Widdowson, 2000; Shin & Nation, 2008: 340). The corpus data provided examples of collocations, not all possible collocations used by humans around the world. However, a corpus could be a great source of knowledge for English learners to learn collocations.

For the following steps, the researchers did the data analysis, which aimed at comparing the frequency in each collocation pairs, for example, the collocation pair *grey hair* and *white hair* in which the frequency of *grey hair* is 89 in COCA and 289 in BYU-BNC, and the frequency of *white hair* is 1172 in COCA and 127 in BYU-BNC. As seen from the comparison, found that the frequency of *white hair* is higher in COCA (1172), but lower in BNC (127), compared to the frequency of *grey hair*.

FINDINGS AND DISCUSSION

There were 15 pairs of collocations analyzed in this paper. Each pair of collocation has a similar meaning, which is used to make the comparison of which collocation in a pair that has a higher frequency (used more commonly by people). The researchers collected the pairs of collocations from many web sources. The 15 collocations in the second column in Table 1 were compared with the collocations in the third column of the table. They were compared and considered as the pair of collocation, because they have a similar meaning but different usage frequency.

Table 1. List of collocation pairs

No.	Collocations Pairs	
1	beautiful woman	pretty woman
2	express train	fast train
3	great deal	big deal
4	heavy rain	hard rain
5	heavy smoker	hard smoker
6	high temperature	tall temperature
7	new ideas	fresh ideas
8	public interest	common interest
9	quick meal	fast food
10	safe place	secure place
11	scenic view	scenic picture
12	secretary general	secretary public
13	Powerful engine	strong engine
14	strong coffee	powerful coffee
15	water resistant	waterproof

Table 2 presents the occurrence of each collocation in COCA and BYU-BNC. In each pair, the collocations written in the first column (left side) are more popular or commonly used than those written in the fourth column. It is proven by the occurrences of each collocation in the two big corpuses (COCA and BYU-BNC). The collocation pair numbers 1-11 and 13-14 are the adjective-noun collocation, except the collocation pair numbers 12 and 15, which are noun-adjective collocation. From the findings, *great deal* has the highest frequency. Its pair, *big deal,* follows it by having the second highest frequency in COCA. The phrases *great deal* and *big deal* have the same meaning. However, people tend to use the word *great* more frequently than *big* to modify *deal*. It seems that people tend to hyperbole in describing something 'really big' to *great*. People prefer to use the collocation *great deal* because of its original meaning, that *great deal* is bigger than just a *big deal*.

Table 2. Collocation pair frequencies in COCA and BYU-BNC

Collocation	Frequency		Collocation	Frequency	
	COCA	BYU-BNC		COCA	BYU-BNC
great deal	**7692**	2477	big deal	**5885**	143
new ideas	2232	494	fresh ideas	212	28
public interest	2200	883	common interest	355	147
fast food	1613	154	quick meal	64	7
beautiful woman	1310	134	pretty woman	446	46
safe place	1077	95	secure place	111	15
secretary general	893	135	secretary public	0	0
heavy rain	638	224	hard rain	111	3
high temperature	326	98	tall temperature	0	0
strong coffee	127	19	powerful coffee	1	0
express train	94	38	fast train	29	1
powerful engine	93	14	strong engine	7	0
water resistant	69	25	Waterproof	1730	400
heavy smoker	59	27	hard smoker	2	0
scenic view	39	2	scenic picture	1	0

Heavy rain (638/224) was used more commonly than *hard rain* (111/3), in line with the other pair, *heavy smoker* (59/27) was used more frequently than *hard smoker* (2/0). Both *heavy* and *hard* indicates the frequency of the rain and smoke, but people are more accustomed in saying *heavy rain* and *heavy smoker,* same as when people are accustomed in saying *heavy traffic;* compared to *hard rain,* and *hard smoker.*

On the other hand, *strong coffee* (127/19) has a higher frequency in COCA and BYU-BNC rather than *powerful coffee* (1/0). However, *powerful engine* (93/14) occurs more frequently rather than *strong engine* (7/0). Referring to the *engine* or machine, the adjective *powerful* is more suitable to modify the noun *engine*, considering that the engine works powerfully. It refers to how the *engine* works and produces the energy. When people modify *coffee*, they tend to use the adjective *strong* rather than *powerful*. *Strong* tends to describe the physical thing and people say that the coffee has the *strong* taste.

New ideas and *fresh ideas* have the similar meaning, but contextually, the collocation *fresh ideas* is an idea which is brand-new that nobody has heard about it before, but *new ideas* might be the ideas of something that somebody never gets or did before. However, the collocation *new ideas* (2232/494) used more frequently in COCA and BYU-BNC than *fresh ideas* (212/28).

People almost never use *quick meal* (64/7), because *fast food* (1613/164) has become a very popular terminology among people. However, *quick meal* has the same meaning as *fast food*. The terminology is popular considering the existence of many *fast food* restaurants around the world that use the same term *'fast food'*. Both mean a kind of food, which can be

served in a shorter time, because it has been ready to serve to overcome people's hunger. Similarly, *secretary general* (893/135) also has become a terminology used among people rather than *secretary public* (0/0). Thus, *secretary general* is used more frequently.

Besides, *express train* has become a common term among people. Both *express train* (94/38) and *fast train* (29/1) have a similar meaning. However, *express* has the higher level of speed than *fast*. Thus, *express train* is used to refer that the train rushes faster than fast or more than fast. Sometimes people also use hyperbole to call train that rushes fast to *express train*.

Water resistant is noun-adjective collocation, and its pair namely *waterproof* is an adjective. *Water resistant* have the similar meaning to *waterproof*, although *waterproof* is not the collocation. Here, the writer wants to show that the synonym of a collocation could also be non-collocation. *Waterproof* can also be *'water proof'*, and both have similar meanings. As seen in COCA and BYU-BNC, *waterproof* (1740/ 400) has a higher frequency than *water resistant* (69/25), but *waterproof* is placed in the right side of the column in Table 2, because it is not a collocation.

Besides, *high temperature* is a common collocation used by people to refer to a temperature level. *High* is used to refer to the level of temperature because people measure a temperature with a number. In referring to a number, people tend to use the word *high* when the number is high, although *high* and *tall* have nearly the same meanings. *Tall* is used more frequently to refer to a physical appearance of people or things. Thus, *high temperature* (326/98) is always used to refer to a temperature than *tall temperature* (0).

Beautiful and *pretty* have a similar meaning that refers to 'good looking'. However, to refer to a good-looking woman, it is purely people's word choice on using *beautiful* or *pretty*. In *Oxford Learner's Dictionary* online, *pretty* means 'attractive without being very beautiful', and *beautiful* means 'having beauty; pleasing to the senses of mind.' Thus, *pretty* and *beautiful* is the word choice that the speaker may choose to attribute the good-looking woman into *beautiful* or *pretty*, based on the speaker's personal preferences. However, as seen in COCA and BYU-BNC, *beautiful woman* (1310/134) is used more frequently than *pretty woman* (446/46). The adjective pairs *safe* and *secure*, *common* and *public*, and *view* and *picture* have a similar meaning. However, *safe place*, *public interest,* and *scenic picture* have a higher frequency in corpora.

CONCLUSION

The results showed that the frequencies of the 15 pairs of adjective-noun and noun-adjective collocations differed in the Corpus of Contemporary American English (COCA) and the Brigham Young University - British National Corpus (BYU-BNC). The collocations which are used more frequently by native-like people or the natural collocations are considered more appropriate word/language choice to communicate or write in English or any other languages. The language has become the culture in society which makes the commonly used one or the natural one becomes the more appropriate one, although both the popular or less popular collocations are grammatically correct. Also, the more frequently used collocation might be commonly used because of most people's personal interests. The learners need to

learn the collocations in order to increase their English mastery and proficiency. The English teachers need to assist their students to learn collocations by providing examples of collocations and giving exercises for the students to complete and discuss in class. It is also essential for the English teachers to introduce some corpora, such as COCA and BYU-BNC, to the students so that they can learn how to make use of the corpora to improve their English language in general and their collocations in particular.

ACKNOWLEDGEMENTS

REFERENCES

Boers, F., Demecheleer, M., Coxhead, A., & Webb, S. (2013). Gauging the effects of exercises on verb–noun collocations. *Language Teaching Research, 18*(1), 50-70.

British National Corpus managed by Brigham Young University (BYU-BNC) https://www.english-corpora.org/bnc/?c=bnc&q=8899086

Chang, Y., Chang, J., Chen, H., Liou, H. (2008). An automatic collocation writing assistant for Taiwanese EFL learners: A case of corpus-based NLP technology. *Computer Assisted Language Learning, 21*(3), 283-299.

Cook, G. (1998). The uses of reality: A reply to Ronald Carter. *ELT Journal, 52*(1), 57–63.

doi:10.1093/applin/amq051

Erten, İ.H., & Tekin, M. (2008). Effects on vocabulary acquisition of presenting new words in semantic sets versus semantically unrelated sets. *System*, 36, 407−22.

Finkbeiner, M., & Nicol, J. (2003) Semantic category effects in second language word learning. *Applied Psycholinguistics*, 24, 369−383.

Firth, J.R. (1957). *Papers in linguistics 1934-1951*. London: Oxford University Press.

Kennedy, G. (2003). Amplifier collocations in the British National Corpus: implications for English language teaching. *TESOL Quarterly, 37*(3), 467-487.

Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning, 61*(2), 647-672.

Nation, P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge

of the node word, and synonymy affect learning? *Applied Linguistics*, *32*(3), 259-276.

Oxford Learner's Dictionary online. https://www.oxfordlearnersdictionaries.com/

Palmer, H. E. (1933). *Second interim report on English collocation*. Tokyo: Kaitakusha.

Pawley, A., & Syder, F. (1983). 'Two puzzles for linguistic theory' in J. Richards & R. Schmidt (Eds.). *Language and Communication*. London: Longman.

Shin, D., & Nation, P. (2008). Beyond single words: The most frequent collocations in spoken English. *ELT Journal, 64*(4), 339-348.

The Corpus of Contemporary American English (COCA) https://www.english-corpora.org/coca/

Tinkham, T. (1997). The effects of semantic and thematic clustering on the learning of second language vocabulary. *Second Language Research*, *13*, 138–163. University Press.

Waring, R. (1997). The negative effects of learning words in semantic sets: A replication. *System*, 25, 261–274.

Webb, S., & Kagimoto, E. (2011). Learning collocations: Do the number of collocates, position

Widdowson, H.G. (2000). On the limitations of linguistics applied. *Applied Linguistics, 21,* 3-25.

Frequent Collocation of Adjective-Noun and Noun-Adjective Corpus Analysis

8   Eyckmans, June, Frank Boers, and Seth Lindstromberg. "The impact of imposing processing strategies on L2 learners' deliberate study of lexical phrases", System, 2016.
Publication
<1%

9   Zeltia Blanco-Suárez. " deade : Some considerations on the history of the intensifying adverb in English ", Acta Linguistica Hafniensia, 2014
Publication
<1%

10 acikbilim.yok.gov.tr
Internet Source
<1%

11 mafiadoc.com
Internet Source
<1%

12 Averil Coxhead. "Replication research in pedagogical approaches to formulaic sequences: Jones & Haywood (2004) and Alali & Schmitt (2012)", Language Teaching, 2015
Publication
<1%

Exclude quotes	On	Exclude matches	< 5 words
Exclude bibliography	On		