Semantic Preference of Verb-Noun Collocation:

Corpus-based analysis

Christiana Sidupa*, Afdol Tharik Wastono
Linguistic Department, Faculty of Humanities
University of Indonesia
Depok, Indonesia
*sidupa.christiana@gmail.com, tharikws@gmail.com

Abstract—**Todays semantic preference and semantic prosody are two notions that many scholars have been carefully studied in the field of corpus linguistics as corpora have become larger in size, and tools for extracting different lexical items for different purposes have been developed. The focus of this study was semantic preference, the relation between a word form and set of semantically-related words. This research specifically studied semantic preference of words *increase* and *improve* that seem to be problematic for language learners. Hence, the present study aimed to describe students' usage of verb-noun collocation by analyzing their argumentative essays. This, therefore, triggers the researcher to further compare it by semantically exploring how the native speakers' collocation usage of *increase* and *improve*. In the scope of corpus, the online British Academic Written English (BAWE), the research used sequential explanatory design that primary focus is to explain the phenomenon. The result of the study was that the English language learners mostly made errors in using the collocations of *increase* and *improve*. In addition, BAWE pointed out *increase* mainly collocates with quantity, such as *rate*, *number*, *level*, *cost*, *value*, *amount* and *price* whereas *improve* with *quality*. The errors are caused by the transfer of first language, Indonesian, to second language, English. To establish evidence, researcher used the Indonesian Web Corpus (IndonesianWaC).**

Keywords—argumentative essay; corpus linguistics; language learners; semantic preference; verb-noun collocation

I. INTRODUCTION

Collocations have received considerable attention in the scope of corpus linguistics. For example, most of the verb *break out* in the BNC (all inflected forms of the verb) showed semantic preference for 'situation of conflict', *'disease'*, or more broadly for 'problematic circumstances' [1]. It is based on the immediate environment of *break out* are *war, conflict, infection*, and *crisis*. Hunston and Partington explained that these semantic categories shared by the frequent collocates of a specific node item are called semantic preference [2,3]. Though semantic preference connects closely with semantic prosody, in this study the focus is semantic preference, particularly applied by Indonesian undergraduate students in their academic writing.

For learners, writing an academic text is not an easy task as problems concerning areas such as lexical selection still remain, especially for English language learners. Here, in the lexical approach conscious learning does facilitate language

learning [4]. However, teaching collocations has not yet considered as a top priority in every language course. It is reasonable, therefore, that learners continue to make such mistakes considering the teachers' focus mostly on grammar. Eventually collocations play significant role in language learning. Collocation is divided into two categories [5]. They involve lexical collocations and grammatical collocations. Lexical collocations consist of nouns, adjective, verbs and adverbs. A grammatical collocation contains a dominant word (like a verb, a noun, or an adjective) and a preposition or grammatical structure. It is classified common types of lexical collocations, which have been adopted by most researchers and are adjusted for use within the focus of this study. In this study, the focus was one type of collocation, V+N, as it is one of typical errors occurring frequently in learners' production, besides Adj + N.

There have been four main approaches to collocations. The first approach is based on native speaker's intuition to determine a collocation. The second approach is frequency-based-collocation is the occurrence of two or more words within a short space of each other in a text [6]. A short space, or span, is regarded as a distance of relevant lexical items (collocates) of the node word. A distinction is usually made as to whether the co-occurrences of the words are frequent or not. Under the umbrella of frequency-based approach, the notion of collocation is further expanded. This contributes towards a more recent definition of collocation, which states collocations as the occurrence of word combinations that is greater than by chance in its context as well as word pairs that are found together more frequently than the occurrence of their component words [7].

The third approach defines collocations as combinations of certain grammatical form regardless of whether they are 'formulaic' or otherwise. The final approach to determine a collocation is phraseological-based, in which collocation is considered as a type of word combination, which can be delimited from other types of word combinations, namely free combinations and idioms. A definition of collocations based on the phraseological approach is classified word combinations according to a collocational continuum [8], that is free collocations (combinations), restricted collocations, figurative idioms as well as pure idioms. In classifying different word combinations based on the restricted sense, it should be noted

that word combinations differ along a scale, which makes exact delimitation impossible.

The diversity of the different approaches to collocations above is in fact beneficial for researchers as it provides a fruitful variety of perspectives on the phenomenon. It is, therefore, important for researchers to consider which approach to adopt when embarking on studies on collocations. Hence, the researcher seeks to examine collocations using second approach facilitated by corpus.

II. OBJECTIVES

In this study, two verb pairs will be examined: *increase* and *improve*. Here, the students thought that *increase* and *improve* share similar denotational meaning. These items were selected because from Indonesian undergraduate students' academic writings, they used *increase* and *improve* interchangeably for different nouns as their collocations. The examples are "*We can increase our knowledge and skill by going to the schools or colleges*", "*We also can improve productivity of our employees by giving them a promotion*", "*In order to increase employees' loyalty, the employer should increase their positive emotions*".

The frequent co-occurrence of certain words takes place in a certain grammatical pattern called colligation. For example, the verb *enjoy* can be used in "V+ Doing" and "V + Noun" patterns. The selection of verbs and nouns in these two patterns are not random. We can say "*enjoy a diner*", "*enjoy a good income*" and "*enjoy one's life*", etc. but not "*enjoy death*", and "*enjoy crime*". For this study, the grammatical patterns concerning "V + Noun" patterns" will become the focus. Because of that, the present study aims to describe students' usage of verb + noun collocation in their academic essays by analyzing their written production.

Therefore, it is interesting to study collocations applied by English as Foreign Language (EFL) learners. According to Lewis and Hill, this linguistic phenomenon deserves considerable attention from Indonesian researchers as collocations are crucial to vocabulary acquisition and it is the most important process in learning a language [9,10]. In this study, the researcher seeks to find out how EFL learners' language in using Verb−Noun English collocations in their argumentative essays and the differences of semantic preferences between the use of 'increase' and 'improve' in British Academic Written English (BAWE) corpus and '*meningkatkan*' and '*menaikkan*' in Indonesian Web (IndonesianWaC) corpus.

III. METHOD

The subjects of this study are fifty college students who major in psychology, accounting, architecture, industry, communication, visual communication design, fine art and design of private university in Jakarta. They have been learning English for at least 12 years. In this private university, every student has to take the *Test of English for International Communication* (TOEIC) before enrolling the English course. The test results show that most of them are at intermediate level. Here, the students produced their around 300-word-argumentative essays in four or five paragraphs, with two different topics.

It is not easy for a non-native speaker to identify collocations. Therefore, in this study, two sets of tools were applied in the analysis. The first tool is the online British Academic Written English (BAWE) Corpus. As there is native and authentic language in that corpus, it would provide a norm for examining collocation in students' essays. The British Academic Written English (BAWE) Corpus is a good resource as a reference to see collocations used in authentic materials. It is a British academic corpus of academic works written at universities in the UK. It represents a pattern of British Academic English with fairly evenly distributed disciplinary areas in Arts and Humanities, Social Sciences, Life Sciences, and Physical Sciences and level of study (undergraduate and taught masters level). The whole corpus consists of 2761 pieces of proficient assessed student writing. The Indonesian Web Corpus (IndonesianWaC) is an Indonesian corpus made up of texts collected from the Internet.

The approach to collocation used in this study is frequency-based. The high frequency of collocations shown in the corpus is considered as one of the criteria to justify the existence of the collocation. If a phrase or word is searched for in this database and has high frequency hits, it will be the evidence of a well-formed collocation. The collocation is within a span of 4:4 words to the left and right of a node word (the search word). After that, top 15 collocates co-occurring with two node words, *increase* and *improve,* are selected and analyzed to detect their preferences.

IV. RESULTS AND DISCUSSION

Below are the results and discussions for two research questions.

A. Collocations in the Learners' Essays

The frequency of occurrence for the verbs found in the learner essays is illustrated in the following table.

TABLE I.   THE FREQUENCY OF OCCURRENCE OF THE VERB-NOUN COLLOCATION FOUND IN THE LEARNER ESSAYS

Verb Noun	knowledge	Skill	Salary	Performance	Loyalty	Productivity	Efficiency
Increase	43	37	28	14	19	3	1
Improve	40	43	50	1	3	14	1

A glance at Table 1 will show that *improve salary* is the most frequent with 50 occurrences. It indicates that the learners overuse the collocation *improve salary*. After that, it is followed by *increase knowledge* and *improve skill* with the

same number of occurrences that is 40. Some of the examples are:

(1.a) By *improving* the *salary*, the employees will work hard. (S1)

(1.b) To make employee loyal, managers should *improve* their *salaries* by observing their performance on working and help their difficulty. (S13)

(1.c) Here, we can take some courses, read books to *increase* our *knowledge* about something we are curious or start asking someone who is the expert. (S42)

(1.d) From the example if you want to *improve* your cooking *skill*, you can watch it from YouTube website or google for the recipes. (S26)

After that, *improve knowledge* is frequently used with 40 occurrences. Some of the examples are:

(1.e) However, all of us still do not really know that actually it is not hard to *improve* our *knowledge* as it seems. (S43)

(1.f) To *improve* your *knowledge*, you can try reading often. (S37)

Next, another verb that co-occurs with *skill* is *increase* with 37 occurrences. Some of the examples are: Some of the examples are:

(1.g) You can *increase* your *skills* by focusing on the things you like to do. (S39)

(1.h) If you want to *increase* your *skill*, grow your will stronger, and this action to improve your quality will be more likely to happen. (S37)

Afterwards, most learners employ *increase salary* for verb-noun collocation with 28 occurrences. Some of the examples are:

(1.i) Firstly, if the workers *increase* their *salary*, the workers will stay with the company for long term. (S3)

(1.j) When the company *increases salary*, it will make change for the company. (S46)

Thereafter, for *loyalty* the verb to be collocated with is *increase* with 19 occurrences. Some of the examples are:

(1.k) Job sharing, working time and provision for study leave can *increase* staff *loyalty* and improve an organization's performance. (S46)

(1.l) It will *increase* their *loyalty* even more if the employer makes more engaging activities to the employee who works because they like the job. (S10)

From the data, *increase performance* and *improve productivity* have the similar amount of occurrences that is 14. Some of the examples are:

(1.m) Because of the adequate salary, the employees' spirit will increase and they will do their best to *increase* their *performance* to make the company even better. (S20)

(1.n) This method is useful somehow to *increase* their *performance*, but the companies have to be cautious not to encourage employees into unhealthy competition. (S50)

(1.o) Employee loyalty is very important aspect to *improve* the *productivity* of the company itself. (S5)

(1.p) Of course, every employee is happy if he/she gets a higher salary, and maybe it will *improve* the *productivity* of the employee. (S9)

Then, *improve loyalty* and *increase productivity* have similar amount of occurrences that is 3. Some of the examples are:

(1.q) There are many ways to *improve* employee *loyalty*. (S15)

(1.r) Basically, when company can *increase productivity* it means working smarter. (S24)

Other collocational occurrences in the current study are *improve performance*, *increase productivity*, *increase efficiency*, and *improve efficiency*. Some of the examples are:

(1.s) After that, they can also *improve performance* at work that benefit the company. (S7)

(1.t) Basically, when company can *increase productivity* it means working smarter. (S24)

(1.u) Not only will it decrease turnover costs, it can also boost productivity, *increase efficiency*, and provide a much more stable work environment for everyone. (S18)

(1.v) Opportunities can be looked for to *improve efficiency* just about anywhere in the company. (S24)

B. Semantic Preferences of 'Increase' and 'Improve'

In the academic corpus BAWE, *increase rate* is the most frequent with 85 occurrences, followed by *increase risk* with 70 occurrences, and *increase number* with 66 occurrences. After that, collocates for *increase* include *efficiency* with 51 occurrences and *level* with 41 occurrences. Thereafter, other collocates that come together with *increase* are *cost, productivity*, *chance*, *sales*, *power*, *value*, *amount* in thirties occurrences. Subsequently, other collocates co-occurring with *increase* cover *production*, *profit*, and *price* in twenties occurrences.

After observing the concordance lines of *increase*, it turns out that the collocates of the verb *increase* mainly represent quantity. In this matter, *rate*, *number*, *level, cost*, *value*, *amount* and *price* connect directly with the amount or number of something. Besides that, collocates of *increase* might represent quantity indirectly. *Increase efficiency* signifies to increase the state achieving maximum productivity with minimum wasted effort or expense. Next, *increase productivity* signifies to increase the rate at which a company or country makes goods, usually judged in connection with the number of people and the amount of materials necessary to produce the goods.

Furthermore *increase sales* means to increase the quantity sold. Afterward, *increase power* indicates to increase the amount of political control a person or group has in a country. Then, *increase production* indicates to increase the total amount of something that is produced. Afterward, *increase profit* represents to increase a financial gain, especially the difference between the amount earned and the amount spent in buying, operating, or producing something. Afterwards, from the corpus most of the occurrences are dominated with *increase chances*, not *increase chance*. Here, *increase chances*

convey positive meaning that is to increase the probability of something desirable happening.

On the other hand, in academic corpus BAWE collocation *improve quality* is the most frequent with 74 occurrences, followed by *improve performance* with 51 occurrences. After that, *improve efficiency* and *improve situation* have approximately the same number of occurrences (32 occurrences). Then, other collocates are *health*, *skill*, and *life* in twenties. *Improve* also comes together with *understanding* (18 occurrences), *accuracy* (17 occurrences), *position* (17 occurrences), *condition* (17 occurrences), *service* (16 occurrences), *product* (15 occurrences), *practice* (14 occurrences), and *communication* between (12 occurrences).

As the most frequent collocates, *improve quality* means to improve general excellence of standard. Moreover, 16 occurrences are revealed for *improve quality of life* that means to improve the level of enjoyment, comfort, and health in someone's life. The collocates to the right of the verb shows that *improve* indicates several semantic preferences for 'condition or state'. It covers *situation, health, life,* and *condition. Situation* refers to a state of affairs. *Health* is condition of the body and the degree to which it is free from illness. *Life* is a person's or animal's state of health or physical fitness. Then, *condition* implies the state of something with regard to its appearance, quality, or working order.

Besides that, the collocates to the right of the verb might show that *improve* indicates several semantic preferences for 'action'. It covers *performance, efficiency, service, practice*, and *communication*. Here, *performance* refers to the action of performing a function. Next, *efficiency* conveys an action designed to achieve efficiency. Besides, *service* implies the action of helping or doing work for someone. *Practice* conveys action rather than thought or ideas. Then, *communication* expresses the act of imparting of information by speaking, writing, or using some other medium. In addition, *improve* expresses semantic preferences for 'ability', in this case *skill* and *understanding*. *Skill* refers to the ability to do something well whereas *understanding* indicates the ability to understand something.

C. Semantic Preferences of 'Meningkatkan and 'Menaikkan'

In Indonesian Web (IndonesianWaC) corpus, the most frequent occurrences with 1,055 occurrences is *meningkatkan daya*, followed by *meningkatkan kinerja* with 466 occurrences, and *meningkatkan jumlah* with 406 occurrences. Next, verb *meningkatkan* is followed by *tubuh* (390 occurrences), *kerjasama* (241 occurrences), *darah* (228 occurrences), and *kadar* (203 occurrences). Afterward, *meningkatkan* also collocates with *pemahaman* with 197 occurrences and *akses* with 142 occurrences. After that, *meningkatkan derajat* and *meningkatkan fungsi* have approximately the same number of occurrences (132). Then, *meningkatkan* also comes together with *perekonomian* (124 occurrences), *aktivitas* (93 occurrences), *jantung* (89 occurrences), and *martabat* (86 occurrences).

As the most frequent collocates, *meningkatkan* collocates with the ability to do something (*daya*). Besides, it also expresses semantic preference for achievement (*kinerja*).

Besides that, *meningkatkan* indicates semantic preferences with amount (*jumlah*) and level (*kadar, derajat, martabat*). For *meningkatkan martabat,* it signifies to increase the level of human dignity. Next, *meningkatkan* also expresses semantic preferences with human's body and its parts (*darah*, *jantung*). Furthermore, *meningkatkan* conveys semantic preferences with action (*kerjasama*, *pemahaman*, *perekonomian*, and *aktivitas*).

On the other hand, *menaikkan harga* is the most frequent occurences with 535 occurrences, followed by *menaikkan minyak* with 72 occurrences. Next, verb *menaikkan* is followed by *bunga* (54 occurrences), *suku* (51 occurrences), and *kadar* (46 occurrences). Afterward, *menaikkan upah* and *menaikkan suhu* have approximately the same number of occurrences (31). *Menaikkan* also collocates with *derajat, martabat, doa,* and *pegawai* in twenties. Then, *menaikkan* also comes together with *barangan* (17 occurrences), *permukaan* (16 occurrences), *listrik* (16 occurrences), and *harganya* (14 occurrences).

As the most frequent collocates, *menaikkan* collocates with amount (*harga*). Besides, from the corpus *harga* is followed by possessive pronoun (*-nya*). After that, *menaikkan* preferably occurs together with level (*kadar, derajat, martabat*).

Furthermore, *menaikkan* preferably occurs together with rate, amount, or degree (*bunga, suku, upah, suhu*). Here, interest relates to money paid regularly at a particular rate for the use of money lent, or for delaying the repayment of a debt. Rate connects with the amount of a charge or payment expressed as a percentage of another amount. A fixed regular payment earned for work or services, typically paid on a daily or weekly basis.

V. CONCLUSION

This study indicates that there is a significant difference in the realization of semantic preference between the online British Academic Written English (BAWE) corpus and Indonesian Web (IndonesianWaC) corpus. From the concordance lines of *increase*, it is found that its collocates mainly represent quantity, such as *rate*, *number*, *level*, *cost*, *value*, *amount* and *price*. On the other hand, *improve quality* as the most frequent collocates means to improve general excellence of standard. Moreover, 16 occurrences are revealed for *improve quality of life* that means to improve the level of enjoyment, comfort, and health in someone's life. Besides, the collocates to the right of the verb shows that *improve* indicates several semantic preferences for 'condition or state' (*situation, health, life,* and *condition*). Moreover, *improve* indicates several semantic preferences for 'action' (*performance, efficiency, service, practice*, and *communication*). Then, *improve* expresses semantic preferences for 'ability' (*skill* and *understanding*).

From Indonesian Web (IndonesianWaC) corpus, as the most frequent collocates, *meningkatkan* collocates with the ability to do something (*daya*). Then, it also expresses semantic preference for achievement (*kinerja*). After that, *meningkatkan* indicates semantic preferences with amount (*jumlah*) and level (*kadar, derajat, martabat*). Next, *meningkatkan* also expresses semantic preferences with human's body and its parts (*darah*, *jantung*). After that, *meningkatkan* conveys semantic preferences with action (*kerjasama*, *pemahaman*,

perekonomian, and *aktivitas*). As the most frequent collocates, *menaikkan* collocates with amount (*harga*). Furthermore, *menaikkan* preferably occurs together with rate, amount, or degree (*bunga, suku, upah, suhu*).

It is obvious, therefore, that the learners language or interlanguage produced by most undergraduate students in this study deals with Verb + Noun English collocations, specifically examining what verbs *increase* and *improve* collocates with what noun. The English collocations they use are influenced by their first language transfer. It, of course, is supported by using the corpus linguistic approach in order to scrutinizing how the native speakers apply the collocations.

REFERENCES

[1] D. Steward, Semantic prosody: A critical evaluation. London-New York: Routledge, 2010.

[2] S. Hunston, Corpora in applied linguistics. Cambridge: Cambridge University Press, 2002.

[3] A. Partington, "Utterly content in each other's company: Semantic prosody and semantic preference". International Journal of Corpus Linguistics, vol. 9(1), pp. 131–56, 2004.

[4] M. Lewis, Language in the lexical approach. In Lewis, Michael (ed.), Teaching Collocation: Further Development in the Lexical Approach: 126-154. Hove: Language Teaching Publications, 2000.

[5] M. Benson, E. Benson, R. Ilson, Lexicographical description of English. Amsterdam/Philadelphia: John Benjamins, 1986.

[6] J. Sinclair, Corpus, concordance, collocation: Describing English language. Oxford: Oxford University Press, 1991.

[7] M. Stubbs, "Collocations and semantic profiles: On the cause of the trouble with quantitative methods". Function of Language, vol. 2(1), pp. 1-33, 1995.

[8] P. Howarth, "Phraseology and second language proficiency", Applied Linuigistics, vol. 19(1), pp. 24-44, 1998.

[9] M. Lewis, The lexical approach: The state of ELT and a way forward. Hove: Language Teaching Publications, 1993.

[10] J. Hill, Revisiting priorities: From grammatical failure to collocational success. In M. Lewis (Ed.). Teaching collocations: Further development in the lexical approach (pp. 47-69). London: Language Teaching, 2000.