LEXICAL ANALYSIS OF THE MOST FREQUENT ACADEMIC WORD IN STUDENTS' ACADEMIC ESSAY

Christianus I Wayan Eka Budiartha and Yoga Pratama Putra Ramadhan
Sampoerna University, Jakarta Indonesia
correspondence: eka.budiartha@gmail.com

Abstract
This study aims to investigate to the most frequent academic word used in academic writings of EFL students with an essay topic "violent video games". This study employs a corpus-based analysis to profile the students' academic essay and semantic analysis to evaluate the most frequent academic word according to the New Academic Word List (NAWL) and the AntWordProfiler. From the corpus of students' academic essay, the word 'impact' was calculated as the most frequently used academic word. Further analysis was involving the Corpus of Contemporary American English (COCA) and British National Corpus (BNC) to identify the mapping of the targeted word and how the word 'impact' is being used by people who speak British English and those who speak American English. The result shows that that there is an amelioration process from the word 'impact' to its derivation 'impactful'.

Keywords: Lexical analysis, academic words, academic essay

Introduction

Among several elements of learning a language, vocabulary has been claimed to be a crucial basis. The importance of vocabulary in foreign language acquisition has beenraised for several decades. Krashen (1989), who paid attention to language acquisition, highlights that there are two reasons of why vocabulary should be introduced at the very basic level. First, vocabulary is an indicator of language ability because learners regularly make use of dictionary rather than a grammar book. Wilkins (1972) also agrees to this notion by mentioning that knowing a great amount of vocabulary can assist learners to speak more. Second, a number of words is required for being competent in a foreign language. According to Nation and Waring (1997), learners need to know a minimum of 3000 or so high frequency word families because it gives coverage of at least 95% of a running text; a word family is included base word, its inflected forms and derived forms (Bauer and Nation, 1993), for example, guide, guides, guided and, guidance. Nation (1998)adds that learning vocabulary is the most crucial process of progressing learners' knowledge. Those two reasons of the importance of vocabulary in foreign language learning has made vocabulary competence is important for foreign language users.

Learning English vocabulary in the context of EFL has a special challenge. Moghadam, Zainal and Ghaderpour (2012) states that "vocabulary competence is not an all-or-nothing relationship, but a systematic procedure in which various types of knowledge are learned until all aspects of knowledge are known for a word". Indeed, learning and then acquiring a new English words, particularly words with low frequency

involves a number of apsects. Nation (1990) suggested that there are nine aspects of the vocabulary knowledge, which are the knowledge of the spoken form of a word, the knowledge of the written form of a word, the knowledge of the parts in a word which have meaning, the knowledge of the link between a particular form and a meaning, the knowledge of the concepts a word may possess and the items it can refer to, the knowledge of the vocabulary that is associated with a word, the knowledge of a word's grammatical functions, the knowledge of a word's collocations, and the knowledge of a word's register and frequency. Nation broke down each aspect into two, the productive (active) knowledge and the receptive (passive) knowledge. The productive or active vocabulary knowledge is the language output that learners convey messages to others through speaking or writing, while the receptive or passive vocabulary knowledge is the language input that learners receive from others through listening or reading and try to understand it.

The success of communication, for both written communication and spoken communication can be determined by one's vocabulary competence (Corson, 1997; Sidek & Rahim, 2015). The vocabulary competence in written communication means the ability to access the word knowledge efficiently in comprehending text and composing text (Ali & Ayub, 2012). Past studies on vocabulary in foreign language learning have indicated that knowledge on vocabulary is one of predictors of reading ability and the capability to obtain new details from texts (Nation, 2001; Torgesen & Wagner, 2006). When a reader does not have the ability to access most words in a text, it would hinder the effectiveness and efficiency of text processing, and this condition leads to difficulties in the reader comprehending the text. (Moghadam, Zainal and Ghaderpour, 2012).

In accommodating the needs to employ lexical aspect of English learning, lexicologists have introduced what so-called "Corpus' as the media of analysis from the native speakers of English. Corpus is "…a collection of text assumed to be representative of given language put together so that it can be used for linguistic analysis…" (Bonelli, 2001). The integration of computer tools in applied linguistics researches has allowed the possibility of compilation large electronic linguistic corpora. Examples of those computer corpora are The Corpus of Contemporary American English (COCA) which consists of
520 million words of text and is equally divided among spoken, fiction, popular magazines, newspapers, and academic texts, The British National Corpus (BNC) which covers 100 million words, and The Cambridge International Corpus (CIC) which has about 100 million words.

The computer corpus databases which provides speed in data processing and analysis, has been used to examine the frequency of occurrence of language elements and patterns in authentic texts. In vocabulary research, these corpus databases are used to examine lexical features and patterns in linguistic corpora, including word collocations, formulaic language in different registers, and lexical bundles in spoken and written English. Corpus- based research has also been used to develop general and specialized word frequency lists, which have influenced vocabulary researches, second and foreign language vocabulary instructions, and assessments of lexical knowledge. Examples of word frequency lists or wordlist are The General Service List (GSL) which is developed by Michael West in 1953, which contains the 2000 most frequent English word families of a 5 million words corpus; the Academic Word List (AWL) which is developed by Averil Coxhead in 2000 which contains the 570 most frequent word families in academic written text.

Each word frequency list provides a lexical coverage for readers. Lexical/text coverage for readers refers to "the percentage of running words in the text known by the readers" (Nation, 2006). For example, the GSL provides coverage of 80% of a text, it means that learners with knowledge of all words in GSL word families would know at least 80% of the words in that text. The Academic Word List in academic written text

covers between 9.06% (Martínez, Beck, & Panza, 2009) and 11.60% (Chen & Ge, 2007; Cobb & Horst, 2004; Coxhead, 2000; Hyland & Tse, 2007; Vongpumivitch, Huang, & Chang, 2009; Ward, 2009) of a text. A person with knowledge of all words in GSL word families would know approximately 10% of the words in an academic text.

The AWL has been used in language education and research. In the foreign language learning in higher education context, Corson (1997) states the knowledge of AWL "is very important for success in higher education." Townsend, Filippini, Collins, & Biancarosa, (2012) add that the knowledge of academic vocabulary "makes an important additional contribution to academic achievement."

For students of English for Academic Purpose (EAP) in higher education, the focus of the selected words is the academic vocabulary or academic word list (AWL) which contains words that are distinctive to academic language but are found across wide range of disciplines. In English for Academic Purpose classes, students need to compose academic writing as one of compulsory lesson.

Based on drives above, this current study aims to investigate to what extent students' writings are drawn from items of Academic Word List. By evaluating the most frequently used academic words in their academic essay; the Academic Word List (AWL) can be used to define determine student proficiency levels and it give insights to developing lexical levels and thus facilitate more effective writing" (Coxhead, 2000). In particular, this study aims to answer the following questions: (1) What is the frequency of academic word list in EFL Students' writing? (2) What are the frequently used academic words in argumentative essays of English for Academic Purpose (EAP) students of Sampoerna University? (3) To what extent the frequently used academic words in argumentative essays of English for Academic Purpose (EAP) students of Sampoerna University appear in Corpus of Contemporary American English (COCA) and British National Corpus (BNC)? Pedagogically, the teaching/learning of vocabulary will be reflected upon the findings in this study.

The notion of vocabulary knowledge

Nation (1990) presents a list of nine aspect of vocabulary knowledge that native-speakers typically possess. The assumption is made that if a learner of English as Foreign Language aspires to native-like proficiency in the use of words, the vocabulary knowledge that one must acquire are:
1. The knowledge of the spoken form of a word,
2. The knowledge of the written form of a word,
3. The knowledge of the parts in a word which have meaning,
4. The knowledge of the link between a particular form and a meaning,
5. The knowledge of the concepts a word may possess and the items it can refer to,
6. The knowledge of the vocabulary that is associated with a word,
7. The knowledge of a word's grammatical functions,
8. The knowledge of a word's collocations, and
9. The knowledge of a word's register and frequency.

Furthermore, Nation also break down the aspect of vocabulary knowledge into two aspects: the receptive (passive) knowledge and the productive (2001). The receptive or passive vocabulary knowledge is the language input that learners receive from others through listening or reading and try to understand it. The receptive or passive vocabulary is composed of words that is recognized when listening or reading. Nation (2001) defines receptive vocabulary knowledge as using the vocabulary to "perceive the form of a word while listening or reading and retrieving its meaning." (p.25). In other words, receptive

knowledge is the language input that learners receive from others through listening or reading.

On the other hand, the productive or active vocabulary knowledge is the language output that learners convey messages to others through speaking or writing. According to Nation (2001), productive use of vocabulary knowledge is "wanting to express a meaning through speaking or writing and retrieving and producing the appropriate spoken or written word form." (p. 24). In other words, the productive is the language output that learners convey messages to others through speaking or writing.

Vocabulary knowledge is important for language learners because vocabulary is "central to a language and of crucial importance to the typical language learner" (Zimmerman 1997, p. 5). Research on the correlation between vocabulary knowledge and different aspects of linguistic ability, such as reading comprehension and listening comprehension (Laufer, 1992; Mecartty, 2000), and also writing (Engber, 1995) has indicated that vocabulary learning is important as vocabulary knowledge plays important component of language proficiency. (Zareva, Schwanenflugel, & Nikolova, 2005). The vocabulary knowledge can be used to assess the general level of learners' vocabulary knowledge to determine if it is appropriate to receive input for their level. (Lo & Murphy, 2010).

In research on vocabulary learning, there are two dimensions of vocabulary knowledge, depth and breadth. Depth of vocabulary knowledge refers to the quality of lexical knowledge, or how well the learner knows a word (Read, 2000). The depth of vocabulary knowledge means that if learners know a word, they should know more its meanings in particular contexts, but also know the knowledge of the word including its pronunciation, spelling, register, stylistic, and morphological features. Haastrup & Henriksen (2000) adds that the vocabulary knowledge also includes the knowledge to the extent of knowing the word's syntactic and semantic relationships with other words in the language, which includes the collocational meanings and knowledge the words' antonym, synonym, and hyponym.

On the other hand, breadth of vocabulary knowledge refers to the quantity or number of words learners know at a particular level of language proficiency (Nation, 2001). As Nation (2006) states the number of words that educated native speakers of English know is around 20,000 word families, while EFL learners need to know a minimum of 3000 high frequency words. Nation & Waring (1997) state that EFL learners need to know minimum of 3000 high frequency words because it gives coverage of at least 95% of a running text. Moreover, most research indicate that knowledge of the most frequent 5000 words should provide sufficient vocabulary to facilitate reading authentic texts. Various types of assessment are used to measure one's breadth of vocabulary knowledge, such as tests multiple-choice test, match words with definitions, and translate a word into first language.

The two dimension of vocabulary knowledge, the depth and the breadth of vocabulary are connected each other. Henriksen (1999, p. 303) states that both "…breadth and depth aspects of vocabulary knowledge should be viewed as a knowledge continuum rather than two distinct dimensions of lexical developments."

Thus, this current study will investigate the productive aspect of the learners' vocabulary knowledge by measuring the breadth of vocabulary knowledge to explain the depth of the vocabulary knowledge of the learners through comparing it to BNC and COCA.

Corpus Study

Corpus or in plural corpora can be defined as "…large bodies of machine-readable text containing thousands or millions of words." (Baker 2009, p.48). Another definition of Corpus is from Bonelli (2001, p.2) which says that corpus is "…collection of text assumed to be representative of given language put together so that it can be used for linguistic analysis". To give boundaries to the definition of corpus, therefore corpus can be defined as representative of text that contains thousands or millions of words. Based on the glossary, corpus can also defined as a large collection of authentic texts that have been gathered in electronic form according to a specific set of criteria.

The use of computer tools in applied linguistics researches have allowed linguists to make a compilation of large electronic linguistic corpora. For example, The Corpus of Contemporary American English (COCA) with 450 million words, The British National Corpus (BNC) with over 100 million words, The (COBUILD) Bank of English Corpus with over 300 million words, and the Cambridge International Corpus (CIC) with over 100million words. Online access to those large corpus databases is combined with the speed of data processing and analysis afforded by computer software.

Today's corpus databases on computer provide speed in data processing and analysis. Corpus is mainly used by linguists to examine the frequency of occurrence of language elements and patterns in authentic texts. In vocabulary research, these corpus computer databases are used to "examine lexical features and patterns in linguistic corpora, including word collocations, formulaic language in different registers, and lexical bundles in spoken and written English." (Simpson, 2011, page??). Corpus researches are also used to develop general and specialized word frequency lists, in which have influenced the vocabulary researches. For examples, word frequency lists or wordlist named The General Service List (GSL) developed by Michael West in 1953. The list contains the 2000 most frequent English word families of a 2.5 million words corpus, and the Academic Word List (AWL) which is developed by Averil Coxhead in 2000 which contains the 570 most frequent wordfamilies in academic written text.

Limiting more on the use of corpus in this study, the corpus that will be used in this study in an online corpus. The difference of this online corpus and the regular corpus that is the online corpus are available through internet and it does not need any third party programs or corpus processing programs to operate it. The corpus that will be used is COCA (Corpus of Contemporary American English) and BNC (British National Corpus) Lancaster. COCA is available on corpus.byu.edu/coca/. As the largest freely-available corpus of English, Corpus Contemporary American English contains more than 450 million words of text which includes 20 million words each year from 1990-2012. BNC is available at http://bncweb.lancs.ac.uk/. BNC is consisted of 100 million word collection of samples of written and spoken language from a wide range of sources, designed to represent a wide cross-section of British English, both spoken and written, from the late twentieth century.

Academic Vocabulary

In the context of English for Academic Purposes (EAP), it is considered more effective to identify specific words which are reasonably frequent in a wide range of academic genres but are relatively uncommon in other kinds of texts (Coxhead & Nation, 2001). The vocabularies that students of EAP are likely to encounter in reading English at university, typically fall into three main groups (Nation, 2001):

1. High frequency words. For example, The General Service List by Michael West in 1953. It is consisted of 2,000 word families in English, providing coverage of about 80% of most texts.

2. An academic vocabulary which are frequent in academic writing. An academic vocabulary list comprises of 8% to 10% of running words of academic texts. For example, The Academic Word List by Averil Coxhead in 2000.
3. A technical vocabulary which differs by subject area and covers up to 5% of texts.

Academic vocabulary is one class of vocabulary that is complex and often abstract nature (Townsend & Collins, 2008). Academic vocabulary is a component of academic English, a register of English used in academic settings and in academic texts, and it is critical for academic success (Nation & Kyongho, 1995).

The Academic Vocabulary List is a list of words containing which is drawn a corpus academic text. The well-known corpus of is the Academic Word List (AWL), created by Averil Coxhead. Coxhead's AWL is drawn from 3.5 million words of written academic texts in several disciplines (i.e., arts, commerce, law and science). Coxhead's (2000) Academic Word List (AWL) is an important contribution that identifies general, or cross- disciplinary, academic vocabulary words. The AWL contains the 570 most frequent general academic words found in academic texts.

Townsend & Collins (2008) who investigate the coverage of Academic Word List, state that Academic Word List (AWL) accounts for roughly 10% of all words in academic texts, and about 12.6 words per page. Academic vocabulary may be particularly challenging for students to both comprehend and use as they are mostly abstract and low term of frequency. Corson (1997) also demonstrates that knowledge of academic words is primarily accessed through texts, not conversation. Thus, students who have gained strong conversational skills in English but who lack extensive print exposure to academic texts will likely not have the vocabulary resources for academic reading comprehension. (Townsend & Collins, 2008)

New Academic Word List (NAWL) is the newest list of 963 words derived from an academic corpus containing about 288 million words. NAWL is developed as a core vocabulary list for second language learners. The NAWL mostly consists of the Cambridge English Corpus, which is comprised of academic journals and non-fiction, student essays, and academic discourse. NAWL also includes text from academic books and journals from the UK and US covering a wide range of disciplines and topics; Arts and Humanities, Life Sciences, Social Sciences, and Physical Sciences. This study employs NAWL as referred academic vocabulary list, as it is the newest academic word list, and it consists of more words than the Coxhead's AWL. The combination of NGSL/NAWL gives 5% more coverage than the combination of West's GSL/ Coxhead's AWL. (Browne, Culligans, & Philips, 2013).

Method
This study was corpus-based analysis, in which the published corpus becomes stepping stone to evaluate a set of corpus compiled from students' argumentative essays from a private university in Jakarta. The compiled data represents the use of vocabulary or the students' vocabulary knowledge. The corpus-based analysis utilizes a large and principled collection of natural texts as the basis for analysis, in this context Corpus of Contemporary American English (COCA) and British National Corpus (BNC). This corpus-based analysis incorporate the use of computer programs for the analysis.

The data, which consists of students' argumentative essays, is profiled using computer program Anthony's AntWordProfiler. The profiler is loaded with list of three levels of general service words –the New General Service List by Browne, Culligan and Phillips (2013) which contains the most frequent 2,800 English words-, academic words list –the New Academic Word List by Browne, Culligan and Phillips (2013) which contains 963

words derived from an academic corpus containing about 288 million words -, and supplementary words list which contains word of days, months, and numbers. The computer program shows the frequency of items from each list in the data of compiled students' argumentative essays. The most frequent word(s) is analysed through concordance web program to see if the word(s) is used in the same ways and with the same meanings with Corpus of Contemporary American English (COCA) and British National Corpus (BNC).

The subjects of this study were selected based on purposive sampling method. Purposive sampling is selecting sample from based on judgment, and/or based on prior information. (Frankael, Wallen, & Hyun, 2012). In which, in this study, the sample was decided by the recommendation of the lecturers and by the number of students in the classroom. Biber, Conrad, and Reppen (1998) state that one key aspect in defining sample for a corpus based study is to have range of linguistics variation. This means that to have a lot of samples that varied in the register related to one of the topic. Hence, in this context, having as many as of argumentative essays as sample is one way to achieve it. Biber, Conrad, and Reppen (1998) also add that corpus based study seeks to represent language or part of language. This study also seeks to represent the language used by the students in university level.

Based on the information gathered about writing argumentative essay as part of learning activity in EAP class, it was decided to take intake 2018 students in the selected University as population. This research took time consideration as one of the reason in choosing the subjects. At the time of the study, students of intake 2014 were the students who took EAP 3 class which required students to compose argumentative essays for their mid-term test. At the time of the study, the total number of intake 2018 students in the university was 157 students. The students were divided into several sections or classes, named by alphabetical order.

The lecturers of the classes, who were from The Institute of Language and Communication (ILC), recommended to choose Section A with most students, 25 students, to be investigated because it represented the population in terms of language proficiency. Another reason to choose Section A was the total students in Section A was the highest, therefore the data collected would be varied and could represent part of the language used by intake 2014 students of Sampoerna University. (Biber, Conrad, and Reppen, 1998). The procedure of this study is presented in the following flowchart:
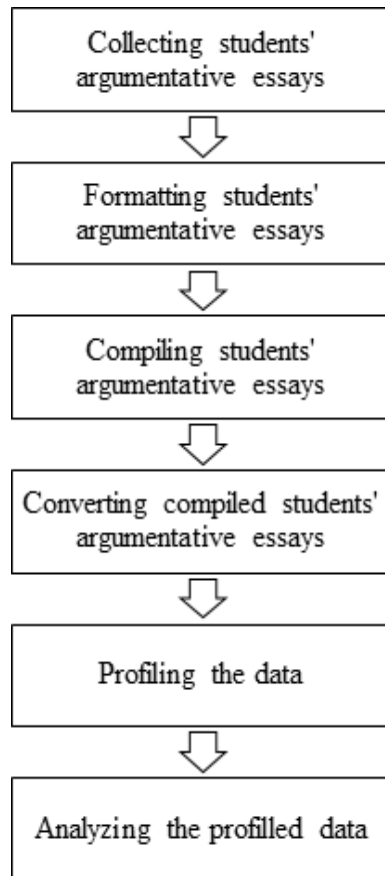
Figure 1. Flowchart of procedure

This study begins by collecting argumentative essays of the students. In collecting students' argumentative essays, they were asked to compose an argumentative essay as their mid-term test. Prior to writing the essay, the students were also required to read a text about violent video games. This was intended as receptive aspect for the students before they started writing the essay. The number of collected argumentative essays is 25 pieces of argumentative essays. This marks the completion of the first step to collect the students' argumentative essays.

This study continues by formatting the essay. In this step, the 25 argumentative essays were converted into documents of containing argumentative essays only. In which, it means that students' personal information and reference part, where students write bibliography for their essays, are removed from documents. After that, the essay documents were converted to Plain Text Format (.TXT). This process is necessary because the computer program which processes the text requires the data in Plain Text Format.

After converting data to Plain Text Format, the documents were used to be analyzed. The data is divided into three data; the main data, sub-data 1, and sub-data 2. The first one, the main data, contains a compiled 25 essays. It is consisted of 37,017 words or tokens and 3,875 types of words. From this main data, two sub-data were created. The first sub-data, sub-data 1, is data containing all argumentative essays of female students. This sub-data containing 17 essays which are consisted of 25,082 tokens and 3,167 types of words. The second sub-data, sub-data 2, is data containing all argumentative essays of female students, containing 8 essays of male students. This sub-data is consisted of 11,935 tokens and 1,982 types of words. This sub-data are created in order to compare the data between

female students and male students, which could give extra insights of this study. This ends the process of creating corpus as the data to be analyzed in this study.

The next process is profiling the data and its sub-data. The data and its sub-data are profiled using Lawrence Anthony's AntWordProfiler 1.4 available for free at http://www.laurenceanthony.net/software/antwordprofiler. This program will run a profiling test to the data and its sub-data, giving out data of numbers of lines, number of types, and number of tokens. The program also gives out data of number of types and number of tokens for words that are in NAWL (New Academic Word List). This profiling process with AntWordProfiler 1.4 gives out the most frequent words in the compiled argumentative essays.

The last step of this study is analyzing the data profile. From the data profile, the most frequent words will be analyzed further using Concordance Program. This study employs Lawrence Anthony's AntConc 3.4.4 available for free at http://www.laurenceanthony.net/software/antconc. This program gives result of concordance process which is listing every instance of the most frequent word with its immediate context of the compiled argumentative texts. Analysis is derived from the context in the data. Further analysis of the most frequently used word will be carried out using online web programs of the Corpus of Contemporary American English (COCA)and British National Corpus (BNC).

Data Collection and Data analysis

This research was conducted based on the research procedure in the previous section. Here are the profile of the corpora.

Table 1.Profile of the data

Data Name	Texts	Types	Tokens
Data of all compiled students' argumentative essays	25	3,875	37,017
Sub-data 1 (data of female students' argumentative essays)	17	3,167	25,082
Sub-data 2 (data of male students' argumentative essays)	z8	1,982	11,935

This data is consisted of all 25 essays taken from sample. There are 3,875 types of words out of 37,017 words in this data. Sub-data 1, the first sub-data which is data consisted of all female students' essays. This sub-data containing 17 essays with 25,082 tokens and 3,167 types of words. Sub-data 2, the second sub-data which is a data of all male students' essays, containing 8 essays of male students. This data is consisted of 11,935 tokens and 1,982 types of words.

The frequency distribution of academic words

Coxhead and Nation (2001, p. 254) claim that in an academic writing, the minimum average percentage of academic words in an academic texts is 10% of all the words in an academic writing corpus. In this study, the AWL (Academic Word List) which is consisted of 963 families of words, is used to analyze the data of compiled argumentative essays of students. The result is presented in the following table.

Table 2 Profile of data of compiled argumentative essays of students

No.	Word List	Token	Token%	Types	Types%	Families
1	New Academic Word List (NAWL)	703	1.9	205	5.29	181
2	New General Service List (NGSL) 1	27,369	73.94	1,384	35.72	824
3	New General Service List (NGSL) 2	3,777	10.2	631	16.28	492
4	New General Service List (NGSL) 3	1,832	4.95	304	7.85	248
5	Supplementary List	110	0.3	33	0.85	26
6	Not in the list	3,226	8.71	1,318	34.01	1,318
Total		37,017	100	3,875	100	3,089

The above table shows that there are 60.7% (total of New General Service List (NGSL)) high frequent words, 5.29% academic words, and 34.01% low frequency words available in the data of compiled argumentative essays of students.

From the 1.9% token percentage in the table, it shows that out of 37,017 token in the data of compiled argumentative essays of students, it has only 703 hits of the New Academic Word List (NAWL). The table above also shows that the academic words used by students are 205 words or only 5.29% of the total words in the data, which does not comply the minimum requirement of academic vocabulary needed in academic writing, i.e. 10% (Nation, 2001). The academic words in the data of compiled argumentative essays of students can be considered as low in frequency because it only hits up around 21.28% of the word families in the NAWL (205 types out of 963 types in the NAWL). However, the result shows that the data has 181 different academic word families out of 205 academic words types, which means the academic words in the corpus can be considered as varied. It can be said that the students are less likely to repeat the same academic word.

The result of profiling process also shows that in the data of compiled argumentative essays of students contains high number of low frequency words. This indicates that the students' argumentative essays contain high percentage of technical words. As Chung & Nation (2004) mention that a text which consists with more than 5% of technical vocabulary is hard to be comprehended by a general reader. Thus, this compiled argumentative essays of students can be evaluated as hard to be comprehended by a general reader. In other word, this of compiled argumentative essays of students is regarded as subject specific essays.

It is important to highlight in this sub-data that there are only 465 hits of academic words, meaning there are only 1.85% of the words in the corpus of argumentative essays. It is considered very low as Nation (2001) says that the percentage of academic words in an academic paper should be around 10% of the composition. The academic words found in the corpus has 165 types. It can be inferred that of all types of words in the sub-corpus, there are only 4.25% academic words (165 types of 3,875). Comparing with the main corpus, it is considered low in composition of academic words.

Another important point to be highlighted in this sub-data is that by comparing the total types of words (3,167 types) to the total data of students' argumentative essays (3,875 types), the variation of words used in female students' argumentative essay can be considered as high variation; the variation is 81.72% of the total data.

This study seeks for the most frequently academic word in the data compiled of students' argumentative essay. Lawrence Anthony's AntWordProfiler is not only giving out distribution of academic words based on NAWL list, but also giving out the list of most frequent academic words. The frequency is based on the hits to academic word family in NAWL list. Table 4.2 shows the list of top 5 most frequently used academic words in students' argumentative essay.

Table 3 Top 5 Most Frequent Academic Words in Students' Argumentative Essays

No.	Academic Word of New Academic Word List (NAWL)	Frequency
1	IMPACT	93
2	ADOLESCENT	65
3	MEDIA	51
4	CORRELATION	30
5	COGNITIVE	22

The profiling process using computer software shows that the 'impact' hits 93 times in the data. The word family of 'impact in the New Academic Word List includes, 'impact', 'impacts', 'impacted', and 'impacting'. The New Academic Word List does not include 'impactful' to be in academic list. This 'impactful' word will be discussed later in the discussion section.

This data shows that the most frequently used academic words in Students' Argumentative Essays is headword 'impact'. The headword 'impact' as the most frequently used academic word in the Students' Argumentative Essays is further analyzed in the discussion section.

This section of this study will further analyze the headword 'impact'. The analysis will divided into several parts to seek the answer of to what extent the headword 'impact' is used in the students' argumentative essay, in the Corpus of Contemporary American English (COCA), and in the British National Corpus (BNC).

'Impact', a verb, is listed as academic word in the NAWL (New Academic Word List). Merriam-Webster's Learner's Dictionary 2016, as an American English Dictionary defines, 'impact' as to have a strong and often bad effect on (something or someone) or to hit (something) with great force." While on, the other hand, Oxford Advanced Learner's Dictionary 2016 as a British English, defines 'impact' as powerful effect that something has on somebody/something. It can also be defined as the act of one object hitting another. Another definition from Oxford Advanced Learner as force with which this happens.

Both of the definition are similar in referring 'impact' to effect of something to another thing. In the context of students' argumentative essays with topic of "do violent video games contribute to youth violence in your country?", 'impact' may refer to the effect of violent video games or effect of video games on youth violence.

In its definition as verb, Merriam-Webster's Learner's Dictionary defines the word 'impact', with sense of strong or great force to something. It also relates the word 'impact' with often bad effect. On the other hand, British English (Oxford Advanced Learner's Dictionary) does not includes any sense into the word 'impact'. Unlike, the definition of the word 'impact' from Merriam-Webster's Learner's Dictionary which has negative nuance, the definition of the word 'impact' from Oxford Advanced Learner's Dictionary has neutral nuance in its relation with positive or negative sense.

The word 'impact' in the students' corpus

Concordance in corpus study of linguistics refers to process of generating the word list to see how it appears in the corpus. Further use of concordance process is to see thecontext of a word. In this study, the concordance process is carried out to see the context ofheadword 'impact' in Students' Argumentative Essays. The process is performed using a computer program named *Lawrence Anthony's AntConc 3.4.4.*

The words that co-occur with the word 'impact' is displayed. Here is the result of most frequent collocations that appears before the word 'impact' in the of students' argumentative essays.

Table 4 Words that co-occur with the word IMPACT in Students' Argumentative Essays

No.	Number of Hits	Words
1	41	THE
2	23	NEGATIVE
3	21	VIDEO
4	17	POSITIVE
5	15	GAMES

The highlighted points in this part is that among top co-occurrence words that frequently co-occur with the word 'impact', there are two adjectives that are antonymous; the word NEGATIVE and the word POSITIVE. The word NEGATIVE co-occurs 23 timesin the data, while the word POSITIVE co-occurs 17 times. This indicates that in their argumentative essays, students associate the academic word 'impact' with the word NEGATIVE more than the word POSITIVE.

In the academic section of Corpus of Contemporary American English (COCA), the top 3 words that co-occurs with the word 'impact' can be seen in the table 4.2.2.2. Words that co-occur with the word 'impact' in Corpus of Contemporary American English (COCA).

Table 5 Words that co-occur with the word IMPACT in Corpus of Contemporary American English (COCA)

No.	Word	Frequency	MI
1	NEGATIVE	656	4.15
2	POSITIVE	618	3.49

The above table is sorted by MI (Mutual Information) score. Mutual Information score refers to how strong the correlation between the two words. In other words, Mutual Information is derived from the statistical probability of the word to occur one another. According to Davies (2008), scores of Mutual Information about 3.0 or above shows a "semantic bonding" between the two words.

From table 5, the word NEGATIVE has the highest frequency in co-occur with the word 'impact', which 656. The word POSITIVE, on the other hand has lower co- occurrence frequency than the word NEGATIVE of 618. This shows that in the Corpus of Contemporary American English (COCA), the word 'impact' occurs more with the NEGATIVE than the word POSITIVE.

The word NEGATIVE also has high score of Mutual Information, 4.15 which suggests strong relation between the word NEGATIVE and the word 'impact', while the word POSITIVE has lower score of Mutual Information than the word NEGATIVE. This shows that the relation between the word 'impact' and the word NEGATIVE is stronger than the word IMPACT and the word POSITIVE.

The data from the Corpus of Contemporary American English (COCA) can be compared with the data from British National Corpus (BNC). It uses the same query methods, and the result is shown in the below table.

Table 6 the frequency of co-occurrence with the word IMPACT in British National Corpus (BNC)

No.	Word	Frequency	MI
1	NEGATIVE	31	4.27
2	POSITIVE	19	2.69

From table 6, the word NEGATIVE has the highest frequency in co-occurrence with the word 'impact', which is 31. On the other hand the word POSITIVE has lower co-occurrence frequency than the word NEGATIVE of 19. This shows that the result similar between the Corpus of Contemporary American English (COCA) and British National Corpus (BNC). It shows that the word 'impact' occurs more with the NEGATIVE than the word POSITIVE.

The word NEGATIVE also has high score of Mutual Information which is 4.27. This suggests strong relation between the word NEGATIVE and the word 'impact'. The word POSITIVE has low score of Mutual Information. The mutual information score is lower than 3. According to Davies (2008), this relationship between the word 'impact' and the word POSITIVE is not really strong in British English.

This findings show that in American English, even though Merriam-Webster Learner's Dictionary adds sense of bad or negative to the word impact, the relation of the word POSITIVE to collocate with the word 'impact'is still considered as strong in Corpus of Contemporary American English (COCA), even though the relation is weaker than collocation of the word 'impact' and the NEGATIVE.

The findings also show that in British English, the relation of the word NEGATIVE and the word 'impact' is high, while the Mutual Information score for of the wordPOSITIVE and the word 'impact' is 2.69, lower than the benchmark score of strong relation which suggest weak relation between the two words.

It can be inferred that the students relate sense of bad effect to their use of the word 'impact', shown by the higher frequency of co-occurrence of the word NEGATIVE than the word POSITIVE. Even so, students can be considered to understand that the word 'impact' does not necessarily related to bad effect as suggested in the definition of word 'impact' by Merriam-Webster Learner's Dictionary.

The word 'Impactful'

In this study, headword 'impact', as academic word in New Academic Word List, has five words in its family, which are IMPACT, IMPACTS, IMPACTED, IMPACTING, IMPACTINGS. The word IMPACTFUL, a word from headword IMPACT, is not included in the in NAWL. The word IMPACTFUL is found twice in the students' argumentative essays. Table7 shows the concordance of word IMPACTFUL in corpus of students' argumentative essays.

Table 7 The Concordance of word IMPACTFUL in Corpus of Students' Argumentative Essays

a small impact, but it could give	**Impactful** result in a research.
cases show that video games are more	**Impactful** than we think. Video games can alter

There are several reason of IMPACTFUL is not included in the New Academic Word List. The first reason is because the word IMPACTFUL is not considered as word in British English. Search on the word IMPACTFUL in British English Dictionary does not

result any word instead it refers as an American English word. Furthermore, search on the database of BNC (British National Corpus), also suggests that the word IMPACTFUL is not exist. The fact that the New Academic Word List (NAWL) is drawn based on academic texts in Cambridge English Corpus, makes the word IMPACTFUL is not included in the NAWL. Therefore, during the profiling process, the word IMPACTFUL is not listed either in New General Service List or New Academic Word List.

However, the word IMPACTFUL is existed in American English. IMPACTFUL is an adjective and it means 'having a major impact or effect'. In the Corpus of Contemporary American English, the word IMPACTFUL has total frequency of 118. It is started to be used in late 20th Century with 4 tokens of 0.04 per million words in COCA. In between 2010-2015, there are 78 hits or as the normalization score, 0.64 per million of word in COCA.

In contrast with the BNC, where the word IMPACTFUL is not found, in American English discourse community the word IMPACTFUL is acceptable. It is shown by the frequency of the word IMPACTFUL has risen as the years pass.
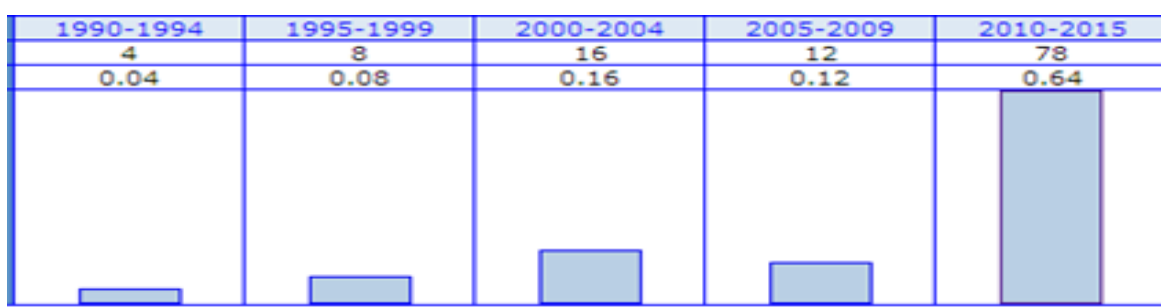
1990-1994	1995-1999	2000-2004	2005-2009	2010-2015
4	8	16	12	78
0.04	0.08	0.16	0.12	0.64

Figure 2. The mapping of the word 'impactful' according to time of usage

The word IMPACTFUL is started to be used frequently in the late 20th century. It is used in the context of news and broadcasting. There are four hits of the word IMPACTFUL found in Corpus of Contemporary American English (COCA) during 1990- 1994. This below table shows the hits of the word IMPACTFUL in the Corpus of Contemporary American English (COCA).

Table 8. Sample of the usage of word 'impactful'

just wonderful. It just shows how	**..impactful** you are.' KING: But there had to be some pain in
cases show that video games are more	**..impactful** than we think. Video games can alter
altered the cover photo document to make it more	**..impactful**. Said more harshly, they rewrote history
Huston says she is accustomed to "	**..impactful** events. " The list is long and includes the death of

The word IMPACTFUL is in the four contexts found in the Corpus of Contemporary American English (COCA), the word IMPACTFUL are used and related to negative situation. It is also used to emphasize negative or bad quality of something. As the years pass, the word IMPACTFUL is becoming more frequently used. It is between the section of year 2005-2009 to 2010-2015, the word IMPACTFUL gained more frequency in corpus.
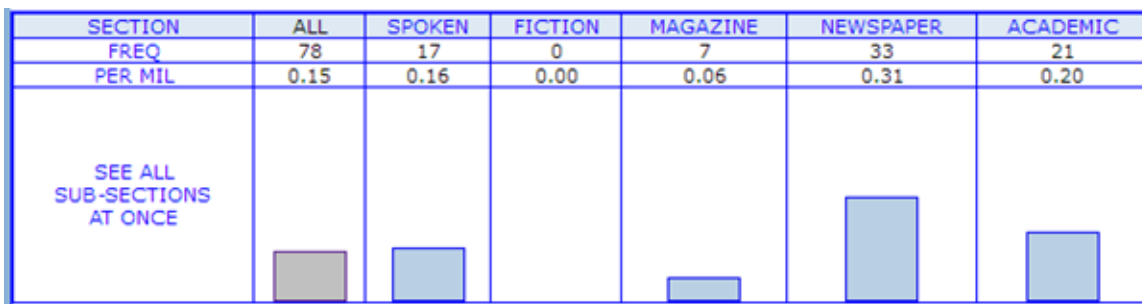
SECTION	ALL	SPOKEN	FICTION	MAGAZINE	NEWSPAPER	ACADEMIC
FREQ	78	17	0	7	33	21
PER MIL	0.15	0.16	0.00	0.06	0.31	0.20

Figure 3. The mapping of the word 'impactful' according to genre

In the 2010-2015 section, the word IMPACTFUL is mostly found in newspaper, academic, and spoken context. The spoken context here is the context of news broadcasting; radio, TV, and Internet news. This can be inferred that the word IMPACTFUL is mainly used in the context of News broadcasting media. In this case, the media are the written media, for example newspaper, and the spoken media, for example TV broadcasting.

Further investigation about the word IMPACTFUL shows that the words has undergone semantic change, which is amelioration or elevation. Amelioration in linguistics refers to the rise or improvement in the quality of the word. In this case, the early use of the word IMPACTFUL was related to negative situation. It was used to emphasize negative or bad quality of something. However, in the section 2010-2015 of Corpus of Contemporary American English (COCA), that seems not be the case. An investigation of collocation of the word IMPACTFUL with adjectives shows that the word IMPACTFUL is no longer related to negative quality, rather it has neutral quality, moreover it is more into positive quality.

From table below, it can be inferred that from the list of words that co-occur strongly with the word IMPACT that the meaning of IMPACTFUL is longer related to negative quality, rather it is related to neutral quality and positive quality. On the list, it also can be seen that the word POSITIVE co-occur strongly with the word IMPACTUL with MI 4.38 (Davies, 2008). The word NEGATIVE is not found as collocation of the word IMPACTFUL. This shows that there is no relation between the word IMPACT and the word NEGATIVE found in Corpus of Contemporary American English (COCA).

These findings about the word IMPACTFUL in the Corpus of Contemporary American English (COCA) have shown that the word IMPACTFUL, which is derived from the word IMPACT, has semantically change from a word which is related to negative quality to a word which related to neutral and positive quality.

No.	Word	Frequency	MI
1	SUCCINCT	1	11.15
2	WAS-	1	10.76
3	INCREMENTAL	1	9.19
4	EXPANSIVE	1	8.88
5	MEANINGFUL	2	7.65
6	EXPLOSIVE	1	7.61
7	DISTINCTIVE	1	7.4
8	EXTRAORDINARY	2	7.16
9	SHORTER	1	7.06
10	IMPRESSIVE	1	6.56

11	EFFICIENT	1	6.36
12	HELPFUL	1	6.34
13	SUBSTANTIAL	1	6.08
14	RELEVANT	1	6.03
15	VISIBLE	1	5.91
16	LARGEST	2	5.85
17	PRACTICAL	1	5.83
18	FAIR	1	4.77
19	POWERFUL	1	4.71
20	CRITICAL	1	4.55
21	INTERESTING	1	4.45
22	BEAUTIFUL	1	4.39
23	POSITIVE	1	4.38

Conclusion

This study has provided information that in the context of academic writing in English as Foreign Language, students have little awareness toward the composition of vocabulary in their writing. Even though there is inadequate awareness towards the balance of word list in students' academic writing, students employ varied academic words in their writing. This indicates that the students have a wide range of academic vocabulary. However, the students incorporate insufficient percentage of academic words in their writing. It is needed to be highlighted that the percentage of academic words needs to be balanced with the general service words and technical words. The reason is that the success of communication for written communication can be determined by one's vocabulary competence which is the ability to access the word knowledge efficiently in composing text, in which this includes the balance of word list in a text. (Ali & Ayub, 2012; Corson, 1997; Sidek & Rahim, 2015).

Other information which this study has provided through the analysis of the headword IMPACT is that collocations of words can be important part of vocabulary knowledge and vocabulary learning in English as Foreign Language contexts. From the comparison of the collocation, it can be concluded that in vocabulary choice, students are more influenced by the American discourse community.

These information gathered in this current study suggest that language teachers to raise awareness of academic vocabulary in academic writing. The awareness of the balance between the vocabulary lists is also need to be improved. Language teachers can make use of corpus based materials to raise this awareness. They can also introduces corpus based activities to investigate their writing in English as Foreign Language class.

This information about inadequate awareness of academic vocabulary in academic writing can also be considered by material developers and curriculum designers. Ergo, the material developers and curriculum designers can address the inadequate awareness of academic vocabulary in academic writing through planned learning series which may incorporate corpus based activities.

References

Ali, Z., & Ayub, M. A. (2012). Obstacles and successes in learning vocabulary from context. *GREduc2012*.

Anderson, R., & Freebody, P. (1981). Vocabulary Knowledge In J.T. Guthrie . *Comprehension and teaching: Research reviews*, 77-117.

Anthony, L. (2014). AntConc (Version 3.4.3) [Computer Software]. Tokyo, Japan: Waseda University. Retrieved from http://www.laurenceanthony.net/

Anthony, L. (2014). AntWordProfiler (Version 1.4.1) [Computer Software]. Tokyo, Japan: Waseda University. Retrieved from http://www.laurenceanthony.net/software/antwordprofiler/

Baker, P. (2009). *Contemporary corpus linguistics.* London: Continuum.

Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography*, 253-279.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: investigating language structure and use.* Cambridge: Cambridge University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S., & 8c Finegan, E. (1999). *Longmangrammar of spoken and written English.* London: Longman.

Browne, C. C. (2013). *The new general service list*. Retrieved from http://www.newgeneralservicelist.org.

Chapelle, C. (1998). Construct definition and validity inquiry in SLA research. In L. F. Cohen, *Interface between Second Language Acquisition and Language Testing Research* (pp. 32-70).

Chen, Q., & Ge, G.-C. (2007). A corpus-based lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles. *English for Specific Purposes*, 502-514.

Chung, T., & Nation, P. (2004). Identifying technical vocabulary. *System*, 251-263. Corson,

D. (1997). The learning and use of academic English words. *Language Learning*, 671-718.

Coxhead, A. (2000). A new academic wordlist. *TESOL Quarterly*, 213-238.

Coxhead, A., & Nation, P. (2001). The specialised vocabulary of English for academic purposes. In J. Flowerdew, & M. Peacock, *Research Perspectives on English for Academic Purposes* (pp. 252-267). Cambridge: Cambridge University Press.

Davies, M. (2008). The corpus of contemporary american english: 520 million words, 1990-present. Retrieved from http://corpus.byu.edu/coca/

Engber, C. A. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing*, 139–155.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to design and evaluate research in education* . New York: McGraw-Hill.

Gardner, D. (2007). Validating the construct of word in applied corpus-based vocabulary research: A critical survey. *Applied Linguistics*, 241–265.

Haastrup, K., & Henriksen, B. (2000). Vocabulary acquisition: Acquiring depth of knowledge through network building. *International Journal of Applied Linguistics*, 221-240.

Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition*, 303-317.

Krashen, J. D. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *The Modern Language*, 440-464.

Laufer, B. (1992). How much lexis is necessary for reading comprehension? In H. Bejoint, & P. Arnaud, *Vocabulary and applied linguistics* (pp. 126-132). Basingstoke & London: Macmillan.

Lo, Y. Y., & Murphy, V. A. (2010). Vocabulary knowledge and growth in immersion and regular language-learning programmes in Hong Kong. *Language and Education*, 215–238.

Mackey, W. F. (1965). *Language teaching analysis.* London: Longman.

Martinez, R., Beck, S. C., & Panza, C. B. (2009). Academic vocabulary in agriculture research articles: A corpus-based study. *English for Specific Purpose*, 183-198.

Mecartty, F. (2000). Lexical and grammatical knowledge in reading and listening comprehension by foreign language learners of Spanish. *Applied Language Learning 11*, 323–348.

Mirriam-Webster. (2015). *Mirriam-Webster Dictionary*. Retrieved from http://www.merriam-webster.com/

Moghadam, S. H., Zainal, Z., & Ghaderpour, M. (2005). A review on the important role of vocabulary knowledge in reading comprehension performance. *Procedia Social and Behavioural Sciences*, 555-563.

Nation, P. (1990). *Teaching and learning vocabulary.* New York: Newbury House.

Nation, P. (1998). Helping learners take control of their vocabulary learning. *GRETA*, 9- 18.

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

Nation, P. (2006). How large a vocabulary is needed for reading and listening?. *Canadian Modern Language Review*, 59-82.

Nation, P., & Kyongho, H. (1995). Where would general service vocabulary stop and special purposes vocabulary begin?. *Elsevier Science Ltd.*, 35-101.

Nation, P., & Waring, R. (1997). Vocabulary size, text coverage and word lists. In N. Schimtt, & M. M. Carty, *Vocabulary: Description, acquisition and pedagogy* (pp. 8 - 10). Cambridge: Cambridge University Press.

Oxford. (2015). *Oxford Advanced Learner's Dictionary*. Retrieved from http://www.oxfordlearnersdictionaries.com/

Read, J. (2000). *Assessing vocabulary.* Cambridge: Cambridge University Press.

Schmitt, N., & Zimmerman, C. B. (2002). Derivative word forms: What do learners know? . *TESOL Quarterly*, 145–171.

Sidek, H. M., & Rahim, H. A. (2015). The role of vocabulary knowledge in reading comprehension: A cross-linguistic study. *Procedia - Social and Behavioral Sciences*, 50-56.

Tannenbaum, K. R., Torgesen, J. K., & Wagner, R. K. (2006). Relationships between word knowledge and reading comprehension in third-grade children. *Scientific Studies of Reading*, 381-398.

Tognini-Bonelli, E. (2001). *Corpus linguistics at work.* Amsterdam: J. Benjamins.

Townsend, D., & Collins, P. (2008). English or Spanish? Assessing Latino/a children in the home and school languages for risk of reading disabilities. *Topics in Language Disorders*, 61-83.

Townsend, D., Filippini, A., Collins, P., & Biancarosa, G. (2012). Evidence for the importance of academic word knowledge for the academic achievement of diverse middle school students. *The Elementary School Journal*, 497-518.

Wilkins, D. A. (1972). *Linguistics in language teaching.* Cambridge: Cambridge University Press.

Zareva, A., Schwanenflugel, P., & Nikolova, Y. (2005). Relationship between lexical competence and language proficiency: Variable sensitivity. *Studies in Second Language Acquisition 27*, 567–595.

Zimmerman, C. (1997). Historical Trends in Second Language Vocabulary Instruction. In J. T. Coady, *Second Language Vocabulary Acquisition and A rationale for Pedagogy* (p. 11). Cambridge: Cambridge University Press.