Integrating Online Corpus and Online Dictionary to Expand Student's English Vocabulary for Tourism

Andi Rizki Fauzi
STIPARY Tourism Academy
andi_gundoel@yahoo.com

Suradi
STIPARY Tourism Academy
suradi_hadi@yahoo.com

ABSTRACT

This study was aimed to know the contribution of online corpus integrated with an online dictionary to take part in learning English vocabulary for tourism. An experimental research was conducted toward the fourth semester of Diploma III students at STIPARY Tourism Academy majoring in Hospitality. The population consists of 118 students and they were divided into 4 classes in which class A and B were randomly chosen as the samples. 100 words from the corpus created from articles in the tourism field were listed through the corpus data and the level of familiarity rated by the students. Based on the analysis result, it could be seen that Asymp. Sig (2-tailed) value was 0,000, which was less than the probability value of 0,05. It means that integrating online corpus and online dictionary could be significantly expand the students' English Vocabulary for tourism.

Keywords : Concordance, Online Corpus, and Online Dictionary

1. Introduction

Emerging technologies make a wider choice for both English learners and teachers to embed the technology into the teaching and learning process in the classroom. Computer technology which develops rapidly can assist the teachers to widely implement strategies, approaches, or techniques to meet the students' needs as well as engage the students to be more active. It can be proven by abundant research which exploited the use of the computer to boost the students' English language skills.

Nowadays, the computer is also employed to improve the students' vocabulary mastery. Vocabulary plays a vital role in language learning as it can determine whether someone is successful or not in using their language in communication. It is simply because the more vocabulary the students know, the better understanding of either written or spoken information will be. In other words, the students will get nothing from the text when they are reading or from what they are listening to if they do not know exactly the meaning of the words used. Therefore, in learning English, vocabulary is an aspect that the teachers emphasize toward their students.

In higher education, teaching and learning English vocabulary are packaged with different techniques that

enable the students to freely adopt the most appropriate one. Although learners' style preferences should be considered, the teachers should teach students to use various kinds of techniques, so that all learners, regardless of preferred style, will be able to learn vocabulary more efficiently (Oxford, 1990). For students who are studying English for specific purposes such as English for tourism, English for office, or English for businesses, they require more exposure of the words in the context of fields they are focusing on. They might be unfamiliar due to limited access toward those words although they are extremely essential to know.

The technology development also makes profound changes as learning vocabulary which relies on traditional methods such as making word lists for unfamiliar words and memorizing them in isolation without the context have been left. Online corpus now can be accessed freely and the students can use the concordance program to independently analyze the words they are learning in different contexts taken from authentic texts from thousands of articles. Based on the observation carried on in the class by showing the corpus data, there was a challenge that many words in corpus data are less-common for the students so that they have to think hardly understanding them. However, those words are extremely important to learn to make the students have a bigger vocabulary size. Therefore, the online dictionary can be applied as translating word by word in the text by looking up the printed dictionary is considered as such a time-consuming activity. This research attempts to integrate online corpus and online dictionary to expand the student's English vocabulary for tourism and examines whether this way can be a potential solution in teaching and learning English vocabulary

2. Review of Related Literature

Nowadays, English teaching and learning cannot be separated from the use of technology. The rapid development of ICT which is used in learning English can reform and explore the English teaching model and the way of learning English in the new era. To make a better effect of using ICT in education, those technologies can be integrated based on education purposes (Fauzi,2015). It means that the teachers necessarily become more creative and aware to blend the variety of ICT developments to achieve the aim of teaching such as computer and internet.

Internet usage also grows quickly as the internet can be accessed everywhere and anywhere using a computer or even a smartphone. Brändström (2011) found that teachers think that the Internet is a valuable source of information and an important additional teaching tool. The Internet can motivate the students, make teaching more fun, and allow variation in teaching. Additionally, Internet-based systems had a real benefit in teaching and learning in which the time used for studying with the Internet-based systems was less than the time used for studying with the traditional method (Krismant (2002). This indicated that the well-designed programs of Internet-based systems used could reduce instruction of teaching and learning.

In the linguistics field, the use of a computer is more necessary with the need to record the languages, such as dictionaries and text collections and it can be used as a tool in concordance. Corpus linguistics becomes a field that was growing from over the past two

decades and having contributions as the pioneer for research in many areas of communication studies and language description (Adolphs, 2012).

Johns (2013) divided teaching ESP into some phases as follows:

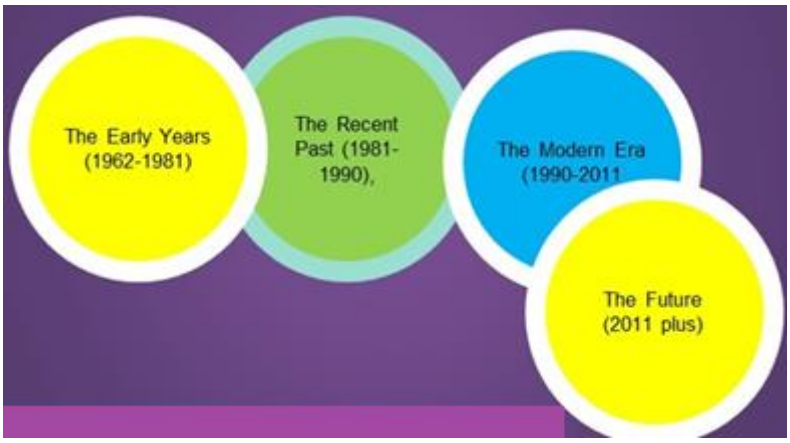


Fig 1. Phases in Teaching ESP

Johns explained that the three phases before 2011, the focus of ESP was on identifying the characteristic of the language of a specific domain and the latter task would change as the technological advance. In 2011 and the following years, the research about ESP will be facilitated with the spread of corpora and software.

The term corpus can be described as a large collection of authentic texts that have been gathered in electronic form according to a specific set of criteria (Bowker & Pearson (2002). There are some general corpora consisting of millions of words such as British National Corpus (BNC), COCA (The Corpus of Contemporary American English), Brown, and ANC (American National Corpus). The benefits of applying corpus linguistics are that corpus-based studies have provided an accurate description of the language, and its new potentials for language structure and use, and have many applications in language learning (Miangah, 2012). It contributes to rendering learning a foreign language more effective since students will be faced with real language (Cotter,1996). The method for analysis using the corpus linguistics approach is quantitative in the way of adopting some language analytical tools: Claws 4 for corpus annotation, WordSmith 3.0 and AntConc 3.2 for retrieval analysis (Kang & Yu, 2011). Those are called concordancing tools.

The use of concordancing is a way to show that L2 learners can have access to authentic language through corpus and they can discover language patterns (Bernardini. 2002). It is very essential in analyzing data because the linguists or the other users can investigate the occurrences and behavior a certain word or word forms in real-lifee contexts which have been used by native or nonnative speakers. Those data are shown in the concordance display. They are the collection of the occurrences of a word-form, each in its own textual environment (Tribble, 2012). In addition, the immediate context for the occurrences of a given word in a corpus or KWIC-Key Word in Context can be seen as well.

There were many studies that had been conducted by previous researchers about the use of Corpus. Donesch-Jezo's study aimed to investigate the ways to facilitate students' acquisition of new words. The study provided a number of examples of learning activities showing how the teaching of vocabulary items can be realized in the L2 classroom in an effective and appealing way with the use of language corpus together with concordancing software (Donesch-Jezo, 2013). Kobelinski's study explored the possibilities of the use of Corpus in the design of materials for vocabulary learning. The findings of the study suggested that participants held positive attitudes towards corpus-based approaches. However, the results also revealed that the reading tasks presented complexities for learners because of the long and challenging authentic texts (Kobelinski, 2005).

Unfortunately, the previous research only focused on the most frequent words in the corpus that can be useful as it can be used to indicate which words a learner is most likely to find in a certain field. For instance, Kaewphanngam et.al (2002) investigated 10 most frequent words at technical and sub-technical vocabulary in sub-fields of psychology.

The big question is how about the unfamiliar words that appear around the frequent words in the corpus data but they are considered as the essential words to learn in a certain field such as tourism. Additionally, to be proficient in English, the students should recognize and able to use less common words. This question leads to why this research is extremely urgent so that the students can expand their English vocabulary for tourism by using online corpus which can potentially stimulate language awareness. To help the students quicker getting the meaning of the difficult words found in the corpus context, the online dictionary is employed as a printed dictionary will be a waste of time and to analyze the words through concordances needs a big effort so that the learning vocabulary will be more efficient and effective.

Now, E-dictionary is available online and dictionary apps are available on most hand-held devices and mobile phones. E-dictionaries offer interactive ways of finding the meanings of the word in comparison to printed dictionaries (Lew, 2010). It is true as the students are provided with many features such as word family, synonyms, and collocation that can support the learning activities. However, the dictionary such as Longman dictionary has provided when the words are used in contexts but they are highly limited. The example of online Longman dictionary display can be seen as follows:

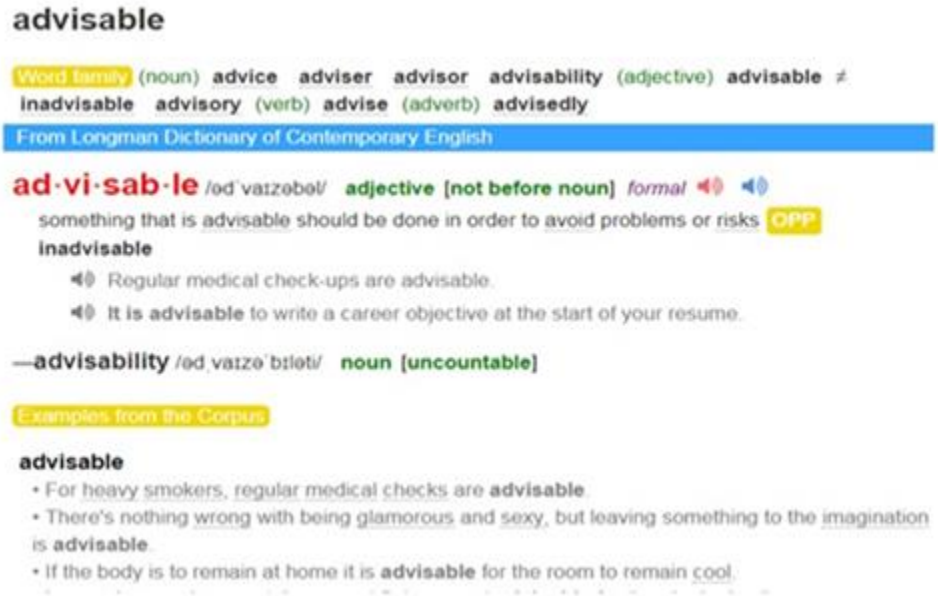


Fig 2. The Display of Online Longngman Dictionary

It is clearly seen that although the corpus examples are shown for the searched word, the numbers of examples are not sufficient for learners to get more exposure to the use of the words in different contexts. As an alternative, the students may utilize online corpus to have more examples of the words they desire to know.

3. Method of The Research

This research used an experimental research design. That is to say that this research intended to test the hypothesis and investigate the cause and effect relationship among the variables. The samples in this research consisting of two groups: the experimental group and control group in which both of the groups were given pre-test and post-test. However, the treatments using integrated online corpus and online dictionary were only given to the experimental group.

The design of this research could be illustrated below:

Table 1: The research design

Group	Pre-test	Treatment	Post-test
Experimental	O_1	X	O_2
Control	O_3	-	O_4

This research was conducted on the fourth-semester students majoring in hospitality at STIPARY Tourism Academy. Their native language was Indonesian and they are mostly between the age of 19 and 20. There are 118 students as the population in this research divided into four classes(A, B,

C,&D). A class consists of 28 students and B class consists of 26 students whereas B class and D class have 33 and 31 students respectively. Both classes have been selected as samples through random sampling. The researcher selected which one would be the experimental or control group using a lottery.

The instrument of the research was devised through the test. A proficiency test was given to the participants to find out whether two groups were homogeneous or not. Next, 100 target words were obtained by selecting unfamiliar words from the online corpus data about tourism. The selection process was based on the lowest frequency of words in the corpus and the rating result of the students for each word. The test which was in the form of matching words with its definitions was given. In the test, one hundred words chosen were classified into several tables based on their word classes and the students were asked to choose the right definitions which were randomly presented. The Procedure of the research could be seen as follows:

Fig 3. Research Procedure

The normality test used Shapiro-Wilk and Mann-Whitney Test was employed. The analysis result was used to examine the hypothesis of the research as follows.

Criteria of hypotheses:

1. Ho is accepted if the significance value (Asymp. Sig (2-tailed) is greater than 0.05 or it can be said that students who were taught with Integrating online corpus and online dictionary are not better in expanding their English vocabulary for tourism than those without.

2. Ha is accepted if the significance value (Asymp. Sig (2-tailed) is less than 0.05 or it can be said that students who are taught with Integrating online corpus and online dictionary are better in expanding their English vocabulary for tourism than those without. The process of the research can be seen in the roadmap below:

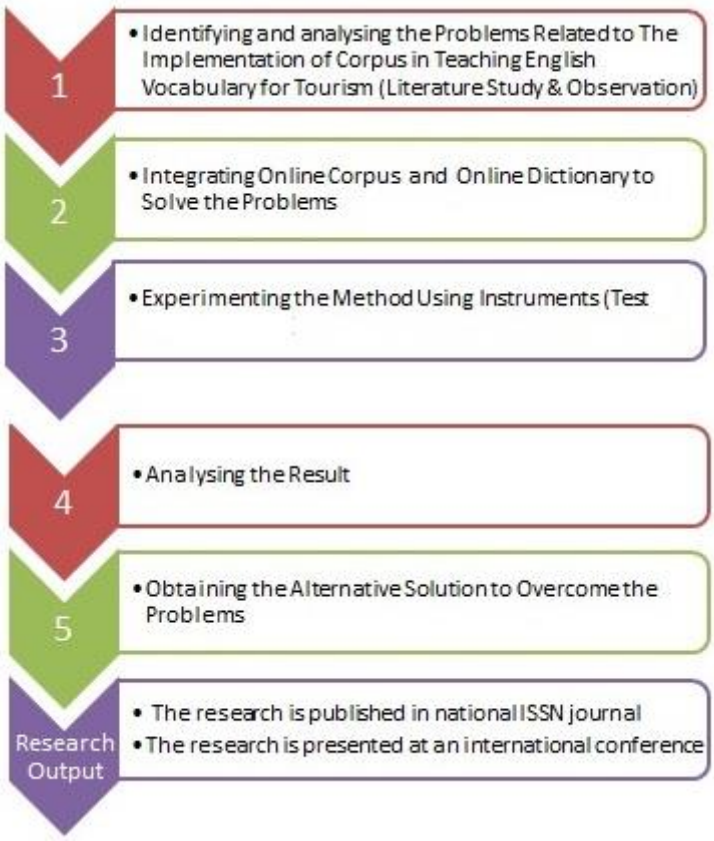


Fig 4. Roadmap of the Research

4. Result and Discussion

a. Result

1. Proficiency test of Students' English Vocabulary

Before conducting the experiment, the researcher had to know the students' proficiency level regarding their English vocabulary mastery. The test was administered online through https://www.oxfordonlineenglish.com.There were 45 students who carried out the test (24 students from the experimental group and 21 students from the control group). Based on descriptive analysis using SPSS toward the test result (see appendix 1), the students from both groups obtained an average score of 41, 5 as can be seen in the table below:

Table 2. Descriptive statistics result of English vocabulary proficiency test

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Vocabulary scores	45	22,50	55,00	41,5000	7,17872
Valid N (listwise)	45				

It means that all the students as the samples have a low level of proficiency in vocabulary mastery. In addition, the table shows that the highest score of the test result was 55 and the lowest score was 22, 5 meaning that their abilities were all on average.

2. Familiarity Test

The researcher has been collected 100 articles related to tourism topics such as adventure travel, art & culture, festivals, food & drink, and family holidays. They were taken from several websites such as www.lonelyplanet.com, www.indietravelpodcast.com, www.thetravelmagazine.net, www.travelandtourworld.com,www.traveldailynews.com, www.travelweekly.com, and traveltips.usatoday.com. The authentic texts were chosen as the hospitality students should recognize the words used in authentic texts. The small corpus was created using AntConc software.

In this research, the focus was not on the most frequent words found in the corpus but the words which have the lowest frequency. Based on the concordance output, there were 1.358 vocabularies (noun, verb, adjective, and adverb) which emerged once (see appendix 2) which are potentially unfamiliar for the students based on the consideration of the researcher. In this research, the researcher limited the number of target words. One hundred words were chosen randomly from 1.358 words (see appendix 2) and they were divided into 7 tables based on their word classes. The familiarity test was given to the students to give the rate based on their familiarity with the words. This process was essential because they would be examined whether they could expand their vocabulary or not through integrating online corpus and online dictionary.

The average rates for each word could be seen below:

Table 3. Familiarity rate of target words

tranquility	**1,96**	**trespass**	2,14	**initiate**	2,47
opulence	1,83	**tarry**	1,65	**stuffy**	2,80
outset	3,33	**undergo**	2,21	**sleek**	2,83
touristy	3,93	**trigger**	2,22	**succulent**	1,79
seclusion	2,41	**resist**	3,24	**thoughtful**	2,64
viewfinder	2,86	**traverse**	2,11	**splendour**	1,63
tourist traps	3,50	**revamp**	1,76	**sumptuous**	1,58
tendency	2,33	**swell**	2,28	**tremendous**	1,73
wanderlust	2,38	**lure**	1,72	**picky**	2,21
shoreline	2,49	**jaunt**	1,65	**tangible**	2,04
plunge	2,14	**intend**	2,88	**thrilled**	2,02
slurp	2,07	**inspire**	3,72	**sedate**	1,73

shallow	3,16	**inferior**	2,54	**loosely**	2,86
scrummy	2,75	**leery**	1,99	**subsequently**	1,94
strenuous	1,79	**intoxicating**	1,81	**sheepishly**	1,87
messy	2,63	**groovy**	2,60	**preferably**	2,20
notify	2,17	**grubby**	2,27	**swiftly**	2,21
meld	1,91	**Graceful**	3,04	**frill**	1,73
invade	1,63	**incremental**	1,87	**enclosures**	2,07
hover	2,90	**interminable**	1,98	**civilization**	2,31
menace	2,16	**hostile**	2,02	**exertion**	2,14
merit	1,71	**impregnable**	1,72	**epicures**	2,11
flaunt	3,36	**frantic**	1,94	**humidity**	2,50
glide	2,26	**inconspicuous**	1,70	**ambience**	2,70
heap	3,20	**tirelessly**	2,03	**extravaganza**	3,27
infuse	1,79	**artificially**	2,36	**hipster**	2,34
indulge	1,73	**thoroughly**	2,55	**dwellers**	1,91
leverage	1,67	**majorly**	2,90	**fraternity**	1,80
Immerse	1,51	**cautiously**	2,15	**festivity**	2,77
lug	1,66	**solely**	1,87	**exaggeration**	1,90
litter	1,91	**practically**	3,18	**hominy**	1,92
frenetic	1,98	**lightly**	3,18	**fair**	3,84
indefinite	2,01	**substantially**	2,18		
hefty	1,70	**playfully**	3,25		

From the table above, the highest rate was in the word "fair" with 3,84 meaning that the students were pretty sure that they have seen or heard the word but they were not positive and the lowest rate was in the word "immerse" with average rate of 1,51 meaning that they almost have never seen or heard the word before. Totally, the familiarity rate showed that there were no words that had more than 4 in rating value. It means that 100 words chosen were unfamiliar based on the descriptor table for each rate as follows.

Table 3. The Use of Base form Verb

Rating Scale	Descriptor
7	You recognize the word and are confident that you know the meaning of the word.
6	You think you know the meaning of the word but are not certain that the meaning you know is correct.
5	You are certain that you have seen the word but you only have a vague idea of its meaning.
4	You recognize the word as one you have seen or heard before, but you don't know the meaning of the word.
3	You are pretty sure that you have seen or heard the word but you are not

	positive.
2	You think that you might have seen or heard the word before.
1	You have never seen or heard the word before.

3. Normality and Homogeneity test

The collected data from the test results were analyzed using SPSS PASW statistics 18 and the Shapiro-Wilk test was used to know the normality of the data as the sample size less than 50. It was because of the reason that the students who followed both pre-test and post-test were only 19 students in both groups. The result of the analysis could be seen in the table below:

Table 4. The normality of the data Using Shapiro-Wilk Test

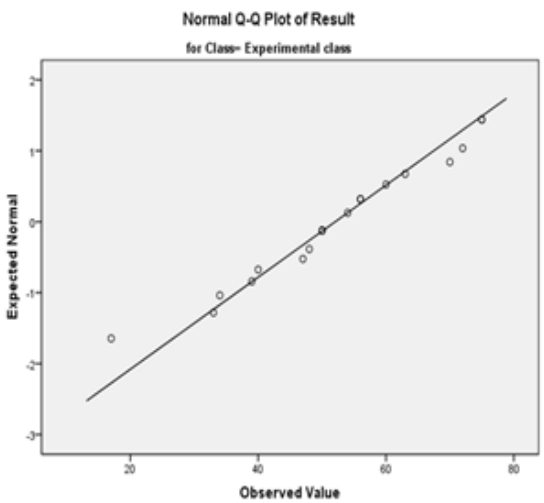
Tests of Normality

Class		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Result	Experimental class	,108	19	,200*	,964	19	,655
	Control class	,204	19	,037	,883	19	,024

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

From the table above, the significance value of the experimental group is 0,655 which is more than 0,05. It means that the data distributed normally. However, the significance value of the control group is 0,024 or less than 0,05 meaning that the distributions of the data were not normal. The distribution of both groups can also be seen by Q-Q plot graphs as follows.



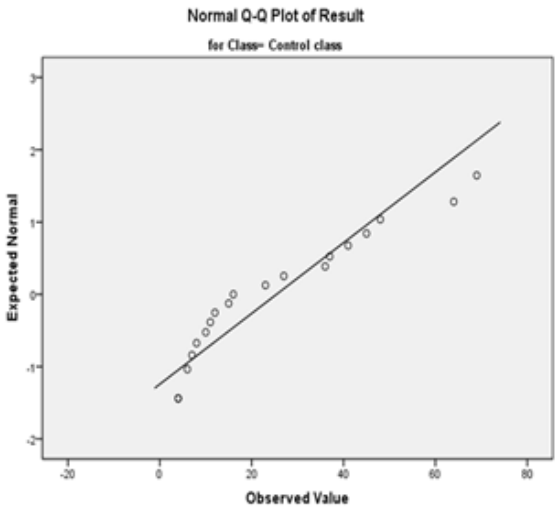


Figure 5. Normality of the data using Q-Q Plots

The figure above showed that the data taken from the post-test of the experimental group spread close to the diagonal line so that the data can be said as normal data. As opposed to the experimental group, the data from the control group mostly spread outside of the diagonal line which means they are not normally distributed.

4. The Difference Between Experimental group and Control Group Using Mann-Whitney Test

As the data from the control group did not distribute normally, the data were analyzed using non-parametric statistics with the Mann-Whitney test. The analysis result could be seen from the table below:

Table 5. The Result of MannWhitney Test

Mann-Whitney Test

Ranks

	Class	N	Mean Rank	Sum of Ranks
Result	Experimental class	19	26,13	496,50
	Control class	19	12,87	244,50
	Total	38		

Test Statistics[b]

	Result
Mann-Whitney U	54,500
Wilcoxon W	244,500
Z	-3,680
Asymp. Sig. (2-tailed)	,000
Exact Sig. [2*(1-tailed Sig.)]	,000[a]

a. Not corrected for ties.

b. Grouping Variable: Class

Based on the output shown in the table above, it can be seen that Asymp.Sig (2-tailed) value was 0,000 or less than 0,05. It ould be concluded that Ha was accepted meaning that there was a significant difference of the test result between the experimental group and control group or those who expanded their English Vocabulary for Tourism using the integration of online corpus and online dictionary and those who did not.

b. Discussion

In this research, although the vocabulary was limited at 100 words found in the article about tourism, they were all unfamiliar words. From the familiarity rate given, it can be seen that

the students had a lack of familiarity with many vocabularies used in the authentic texts in the small corpus built. It could be possible that although the students are from hospitality majors, they are more likely finding reading authentic text related to their fields is an arduous task. Unfortunately, to expand more vocabulary, the treatments must be given in more time as employing an online corpus and online dictionary need time to get significant results. Based on the observation during the treatments, creating a small corpus was very essential as the students just realized some new words and they considered that those were very important to know and must be taken as priority words to learn although it only emerged once in the corpus data.

Based on the result of the analysis, integrating online corpus and online dictionary was very successful to expand the students' vocabulary for tourism. By doing activities in the treatments such as finding the target words in the online corpus (https://corpus.byu.edu) the students get more exposure to the words they learn. It was a great tool as the online dictionary only limited a few examples of the words in sentences. This integration could make the students able to observe the use of the words in several other contexts and how the words were used in the sentences. It reinforces the research finding conducted by Bernardini (2002) that concordance can be helpful to observe language pattern.

Not only getting the benefit of recognizing how the target words used in the corpus, the guessing process which was gone through during the treatments was also powerful to help the students memorizing and understanding the words before they consulted the online dictionary. It could encourage the students to learn English from guessing activity, as they made a mistake in guessing the meaning of the target words at several concordance lines displayed, they can learn from it. Eventually, they could memorize the meaning of the words and guessed the definition of the words precisely afterward compared with the students who did not receive treatments.

By integrating online corpus and online dictionary, the corpus data which were considered as complex and challenging for language learners as stated by Kobelinski (2005) could be solved. The students could directly find the meaning of the target words or the words surrounding in the text to understand the corpus data presented. They could also develop their vocabulary size outside the target words they learned when they attempted to understand the concordance output using an online dictionary. However, providing assistance through an online dictionary to the students who had a very low level of vocabulary mastery still needed more supports. It was due to the fact that understanding the meaning presented in the monolingual dictionary (English to English) was not easy enough for them. It could be seen from the result of the test that there were still the samples who got the post-test score under 50 from the experimental group although their score was significant compared with their pre-test results.

Knowing the words that they were unfamiliar with was very important to expand the students' vocabulary mastery in tourism and it finally can steep their English proficiency up. However, the test which was devised using the matching word and its definition can only measure the receptive vocabulary acquired by the students. For that reason, further research can be conducted on whether

integrating online corpus and online dictionary can give significant improvement to the students to turn the receptive vocabulary into a productive vocabulary so that the students can expand their vocabulary and increase their ability in language productive skills such as speaking and writing.

5. Conclusion and Suggestion

Based on the result and discussion, the integration of online corpus and online dictionary was proved to expand the students' vocabulary for tourism. The statistic result showed that the significant difference of Sig (2-tailed) value between the experimental group and control group (Asymp.Sig (2-tailed) value was 0,000 < 0,05). It means that employing an online dictionary potentially assist the students in learning new vocabulary for tourism through the integration of online corpus and online dictionary.

However, the success of the use of online corpus and online dictionary could be supported by other techniques or strategies particularly those who had a low level of proficiency in vocabulary. Vocabularies in the authentic text of the corpus data were considered demanding for the students at that level. Consulting the words found in the text to the monolingual online dictionary was very helpful but to understand it, the students who had the very limited vocabulary needed more effort and time.

6. Acknowledgments

We would like to express our gratitude to the Directorate General of Higher Education as the funder of this research. We also thank all parties who have contributed to accomplishing this research. This research is part of Grant Competence of Directorate General of Higher Education that must be finished in 2019.

7. References

Adolphs, S. (2012). Corpora: Multimodal. *The Encyclopedia of Applied Linguistics*, 2006–2009. https://doi.org/10.1002/9781405198431.wbeal0233

Bernardini, S. (2002). Exploring new directions for discovery learning. In B. Kettemannann & G. Marko (Eds.), Teaching and learning by doing corpus analysis (pp. 165-182). Amsterdam: Rodopi

Bowker, L. & J. Pearson (2002). *Working with Specialized Languages*. London: Routledge.

Brändström, C. (2011). Using the Internet in Education – Strengths and Weaknesses and Instruction. *Hogskolan I Gavle*, *3*(2), 1–56.

Cotter, M. J. (n.d.). Teaching terms : a corpus-based approach to terminology in ESP classes, (1996), 499–506.

Donesch-Jezo, E. (2013). Using language corpus in teaching foreign language vocabulary. *International Multidisciplinary e-Journal*, 2 (1), 11-25.

Fauzi,A.R (2015). The Use of ICT in English Learning: How It Contributes to Student's Communicative Skill and Student-Centered Learning. *'Capacity Building for English Education*:83-95 Salatiga, 11-12 September: Satya Wacana Christian University

Johns, A. (2013). "The history of English for Specific Purposes research". In Paltridge, B. & S. Starfield (eds.), *The Handbook of English for Specific Purposes*. Chichester, UK: Wiley-Blackwell, 5–30.

Kaewaphannagam, C., Broughton, M.M. and Soranasataporn, S. (2002) 'Corpus/Based Analysis: Guidelines for Getting Practical Language Input in Materials Development'. SLLT, 11, dostupno na www.sc.mahidol.ac.th/sclg/sllt/sllt2002/ArticleSLLT(Maurice).pdf

Kang, N., & Yu, Q. (2011). Corpus-based stylistic analysis of tourism English. *Journal of Language Teaching and Research*, 2(1), 129–136. https://doi.org/10.4304/jltr.2.1.129-136

Kobelinski, S. G. (2005). a Corpus-Based Approach To Esp: Est Vocabulary in Information Technology Sara.

Krismant, W. (2002). The effects of internet-based teaching and learning systems on learners. *King Mongkut's Institute of Technology North Bangkok' 1518 Piboonsongkhram Rd., Bangsue, Bangkok 10800, Thailand*, (June), 1–12. Retrieved from http://library.iugaza.edu.ps/thesis/118478.pdf

Lew, R. (2010). Multimodal lexicography: The representation of meaning in electronic dictionaries. *Lexikos Journal*, 290–306.

Miangah, T. M. (2012). Different Aspects of Exploiting Corpora in Language Learning. *Journal of Language Teaching and Research*, *3*(5). https://doi.org/10.4304/jltr.3.5.1051-1060

Oxford, R. (1990). Vocabulary Learning : A Critical Analysis of Techniques. *TESL Canada Journal*, *7*(2), 9–30.https://doi.org/10.18806/tesl.v7i2.566

Tribble, C. (2012). Concordancing. *The Encyclopedia of Applied Linguistics*, (1890). https://doi.org/10.1002/9781405198431.wbeal0183