

Do Corpora Benefit the Level of Vocabulary of 10th Year of Basic Education Students? Case Study

Sindy J. Castillo Jaen

Pontificia Universidad Católica del Ecuador, Sede Esmeraldas

Email for correspondence: sindyane_@hotmail.com

Receipt date: March 4th, 2016

Approval date: May 20th, 2016

How to cite this article (APA Norms)

Castillo, S. (2017). Do Corpora Benefit the Level of Vocabulary of 10th Year of Basic Education Students? Case Study. International Congress on the Didactics of the English Language Journal, Vol. 2, No.1. ISSN 2550 -7036. Retrieved from <http://revistas.pucese.edu.ec/ICDEL/index>

Abstract

The low level in English classes leads to the need of looking for alternatives of new techniques and methodologies that help to solve this issue. The present case study aims at investigating the benefits that corpus (or its plural corpora) can provide to school classes in Esmeraldas, Ecuador. Through a period of one month, paper material based on information taken from a corpus was applied in a 10th level class at a fiscomisional institution in Esmeraldas. This is an experimental type of study that used mainly quantitative methods for the analysis of class documents and questionnaires, in order to explore students' response and perceptions to corpus. Also, a pre-test and post-test will be carried out in two classes, one in which the corpus will be applied (treatment group) and the other one which will continue with its regular learning process (control group), in order to know the level of students regarding vocabulary pre- and post- treatment period. Later, through a comparison between the results of both groups and the analysis of a survey, be able to state the benefits that the use of corpus could bring to the English teaching field. The findings indicate that, despite its short time of implementation, the use of corpus in classes improved the level of the student's vocabulary.

Key Words: corpus; corpora; case study; English teaching

Introduction

The low English level is an issue that does not only affect Esmeraldas. According to studies done by the EF English Proficiency Index (*EF EPI*, 2014), from 14 countries in Latin-America, 12 have a low level of English and Ecuador is not an exception. The latest EF EPI study done in 2015 was based on exams about grammar and audio, took by over 18 years old people, which results located Ecuador in the 38th spot (with an average of 51.67/100) meaning a low level of English in the country (*EF EPI*, 2015).

Many factors contribute to the low level of students, such as bad study habits, as well as the lack of good materials for learning and teaching English. Calle *et al.* (2012), state that the reason of the low level of English in students is due to the over-use of traditional teaching techniques. In her research she concludes that in order to reach an improvement in the English teaching and learning of the country, continuous training is required as well as new methodologies and materials.

Corpus (or its plural *corpora*) is a computer database that can be defined as a “linguistic methodology which is founded on the use of electronic collections of naturally occurring texts”. That is why it is considered real language. A corpus might help improve the problematic in the level of English in Esmeraldas since it is not only useful for teaching current vocabulary to students, but also to help teachers to have a better understanding of English and to design “effective materials” for their students (Conrad, 1999).

This issue leads to the hypothesis that the use of corpus in high school classes of Esmeraldas is highly beneficial for the teaching and learning process of vocabulary. This research work could have very significant pedagogical implications.

English is nowadays one of the major requirements for education in Ecuador since, as stated above, it is the language for business and technology. With its efforts to improve English in the country, Ecuador has had a notorious improvement. In the first edition of the EF EPI study back in 2011, the country scored 44.54 points over 100. In the latest edition of 2015, it reached 51.67 over 100, meaning an improvement of 7.13 points. However, the score still reflects Ecuador as a country with a low level of English (EF EPI, 2015).

This research work looks to find the benefits of using corpus, or its plural corpora, in 10th year of basic education English classes in order to present it as a potential material for teaching and learning process.

Corpus Linguistics

According to Atkins (1992), a corpus is a smaller Electronic Text Library (ETL). An ETL, as its name states, are texts compiled electronically with a common format. Consequently, Corpus is considered a “subset of an ELT” (Atkins, 1992) or “a sample of a larger collection of language” (Barnbrook cited in Kindt & Wright, 2001) but with a determine design criteria and for a specific purpose. Granger (2003, 2008; Granger, Hung, & Petch-Tyson, 2002) defines Corpus as a spoken or written collection of texts with a specific design criteria such as medium, gender, average, length, learner proficiency level, etc. In his research, O’Keeffe (2007) determined that corpora are storages of written and/or spoken texts in a computer with a detail design matrix addressed to a determine audience. That is, like Ganger, to pay attention to specific characteristics such as varieties of age, gender, location, school, level, teacher, class, nationalities, etc.

Currently, thanks to the advantages of technology, Corpora cannot longer collect only words from a single author or book, like at the beginning of its creation, but it can pull together large texts from several sources ready to be analyzed, making it feasible for linguist around the world.

O’Keeffe (2007) points out the fact that corpora give the opportunity to do a quantitative and qualitative analysis. In terms of quantitative results, corpus can show statistical data about the frequency of occurrence of a word and it can be compared with other sources. A qualitative analysis is usually for people with further knowledge about linguistics who can pay attention to more details related to the matter and reach a conclusion (Fernández, 2014). That is why corpus is considered a provider of impartial information, since it allows the researcher to do the analysis.

Conrad (1999) emphasizes the benefits of using corpus because “it facilitates the use of authentic language”. People in every day conversations make any kind of grammatical, semantic, syntactic mistakes. In fact, some constructions of a language that are grammatically perfect sometimes are not used or considered unusual in a language to the point that a native speaker can perceive such structure as from another country (Perez, Moreno & Butler, 2004). So, a corpus does not show the correct or perfect use of language but the real usage of it (Sinclair, 2004).

Types of Corpora

According to Perez *et al.* (2004), a corpus can be generally classified according to different points of reference. Generally talking, it can be classified in three ways. There are written corpora, spoken corpora, or a mixture of both written and spoken corpora. However, corpora can also be classified in terms of the varieties of language, that is the “specific domain or geographical origin, synchronic and diachronic corpora, annotated or raw, and monolingual, comparable or parallel” as sated by Perez *et al.* (2004).

Corpora vs. Course Textbooks

Carter (1998) says that, in course books the topics are more relevant than the context and reality of the dialogues. In his research he found that students cannot understand real conversations when they have to express themselves outside the classroom. On the other hand, Carter (1998) also points out that scripted dialogues are not real language but students can understand its meaning, whether unscripted ones are real language but more difficult to understand by students or, like Carters says, “Less real pedagogically” (Carter, 1998).

So which one is better? According to Carter, real language is full of “interruptions, hesitations, false starts...., vague language, and even ungrammatical forms” (Carter, 1998) so it is important for students to be familiarized with it since, like Burn says, “scripted dialogues lessen students possibility to understand interaction outside classroom” (Burn cited in O’Keeffe, 2007) referring to the difficulty of students with little

exposure to real language have when they try to understand and communicate in the target language outside the classroom.

Corpus in teaching

There are several types of data that students and teachers can focus on when they use a corpus. For example: Collocations, sense distinction, figurative usage, proper names, compounding, multiword units, morphology, culture, alteration, and cross language comparison. These types of activities can be part of a task based syllabus as well as a lexical syllabus or as a way of introducing concordancing techniques (Coffey, 2004). According to Huang (2011), even though corpora have been used in a lot English classes lately, the best way of using it is through Data Driven Learning (DDL) which refers to the use of concordance in exercises and class activities so that student develop their own conclusions about patterns and regulations in the target language (Huang, 2011). This way, students have a direct access to the data so that they can create their own meaning and use. And, as mentioned above, students will get in touch with real language instead of a scripted one.

Even though DDL is described as one of the most effective ways of using corpora in classes, Corpus-aided discovery learning (CADL) is more recommended if teachers want to develop students' autonomy. CADL is about making students discover language by encouraging them to investigate (Gavioli cited in Huang 2011). For this approach, student's ability in observation, analysis and interpretation of raw data is important. Here, corpora are used as tools rather than the basis of language teaching and language (Gavioli cited in Huang 2011).

Tasks involving corpora

As it has been already set above, a corpus seems to have a lot of benefits, but how to apply it in class? Coffey (2004) showed several ways to use it. One common example is to form groups of students and ask each group to look for a word in different corpora. Then, they have to share the information found with the teacher and classmates and come out with a conclusion. Single words are not the only thing students can look for in a corpus, but also expressions and phrases. For example if students search for the word *chestnut* in the BNC (British National Corpus) it will appear *Chestnut mare* as one of the results. This might cause students to want to know what *mare* is (Coffey, 2004) and they will ask and learn new vocabulary.

Another very common task is to use gapped sentences taken from a corpus so that students fill in the blanks. Also, a more complicated task would be to ask students to create a kind of dictionary entry for all the words or phrases they had investigated in corpus. Since corpus does not give definitions of words but real

examples of the word usage, students will have to implement their analyses and creativity. This way, student can be more conscious about the work they do with corpora and also they will have a record of the information they have researched and analyzed.

If the teacher wants to increase the difficulty level of the tasks, Coffey (2004) details an example of DDL (Driven Data Learning) regarding the word *gold*. In which he details the results taken from the data analysis of the word such as the family name (Golding, Goldman), that the word has been used in several TV commercials, that the word is used as a color and also as a metal, the main uses of the words (metal, adjectivally, in sports, value/quality, and to describe a color), etc. Coffey also makes reference to the fact that students will be able to analyze most of the information since gold is a well-known metal, of course with the sufficient time given. So students can be given a word and try to do the analysis as in this example.

As above, there are lots of other activities that can be used in the classroom. The teacher has to think about the adequate activity according to students' proficiency level and knowledge in order to obtain good results from the use of corpora.

Problems when using corpora

As stated before, Corpora gained permanent importance in the teaching field. However, in a lot of cases students and teachers do not know anything about it (Al Saeed & Waly, 2009). This may be one of the major problems teachers will find when working with corpus, since students are not familiarized with corpus software. For this, Al Saeed & Waly (2009) recommend teachers to first introduce students to free corpora such as the British National Corpus (BNC). The BNC contains around 100 million words from written and spoken language. Other available data are the Michigan Corpus of Academic Spoken English (MICASE, with around 1.8 million words of speech in various academic contexts), and the Corpus of Contemporary American English (COCA, with 410 million words) (Huang, 2011).

Some teachers think that working with corpora can be "Too challenging and time consuming" (Boulton, cited in Huang, 2011). Since they need extra preparation and training in order to use it in classes in a way student can understand, which is why for the majority of new users of corpora this tool can turn useless. However, for experienced linguists or people who deal with linguistics and computer science the use of corpora may not be complicated, which tells us that some training is needed in order to work with corpora in a feasible way.

According to Frankenberg (2012), people who start using this database without knowing how to do it usually do not obtain what they want from it. This is because new users of corpora such as students use

“dictionaries or web browsing techniques of research in corpus” (Frankenberg, 2012). However, corpora do not give a definition of word or phrases. Instead it gives the raw data needed in order to take out own conclusions of how a word, phrase, or structure is used in everyday interactions. That is why the use of Corpora is attributed to linguists only. So, using the same searching technique as in dictionaries will not work with corpora. Instead, according to Al Saeed & Waly (2009), some quite experienced students use corpora as a complement of the common searching techniques.

Even though some researchers like the ones mentioned above think that every person who wants to use corpora must be trained in order to know how to analyze and understand the data, Bouton cited in Huang, (2011) argues that too much training is not necessary. Also, Bouton recommends using paper-based materials with information from corpora as a way of working with corpora in a more practical way.

Method

This case study focuses on using materials generated from corpora in order to identify the benefits of Corpus in high schools of Esmeraldas.

This is an experimental research, because it is based on the application of worksheets containing exercises based on corpus (see Task involving corpora) at a high school and the analysis of the results. With that being said, the independent variable was the use of corpus in exercises in the classroom and the dependent variable was the benefits of corpora in the vocabulary of students. The dependent variable is determined with the comparison between a post- and pre-test. Also, the study uses mostly quantitative methods of action research such as class documents, questionnaires, and learners' thoughts in order to perceive their reactions and perceptions of the research. The study started at the end of June 2015 and finished on August of the same year.

Sample

This experiment took place in Don Bosco Educational Institution in Esmeraldas-Ecuador. This is an institution with mixed funds. It means that it is partly afforded by the government and partly afforded by the missionaries. That is why it is called a *Fiscomisional* Institution. The target student population in this research work was the tenth year of Basic Education students at Don Bosco high school. From which the samples were only 10th A (control group) and 10th B (treatment group). From these two classes, only in the treatment group the exercises worksheets with data from corpus were applied during one month. In each class there are around

30 students, females and males, all of them around 13 -14 years old. The participants are all Ecuadorian with Spanish as their mother tongue.

Process description

First, students took a pre-test in order to perceive their level of English vocabulary. The tests were based on the university of Cambridge Vocabulary list for KET (Key English Test) and the webpage examenglish.com which provides practice tests for English learners (“Exam English,” 2014). According to the government, 10th level students must have around an A2 level so the KET Vocabulary list is the most suitable since its difficulty level is A2 (University of Cambridge, 2012). There were multiple choice questions regarding the vocabulary that, according to the University of Cambridge, A2 level students must be familiar with. Students had to choose an answer by marking it with an X or \surd , or by just painting the circle next to the answer they believe was correct. There were a total of 23 questions per test.

After the pre-test, corpus was applied in classes during one month. The corpus used for this treatment was the British National Corpus (<http://bncweb.lancs.ac.uk/>), due to its easy access and simple use (IT Services, n.d.). Worksheets with several exercises were created with data taken from the BNC (see Tasks involving corpora).

With the purpose of not interrupting the lesson plans and syllabus of the teacher, previous meetings were carried out in order to set and agree the schedules and how the corpus would be applied so that the worksheets were created according to the plans of the teacher. It is to say that if the teacher planned to teach *can and can't* for that certain week, the worksheets were created about the same exact subject. That way, there was not unfairness since they both were taught the same exact subject with the difference that only the treatment class worked with authentic data taken from the corpus.

The worksheets were created with a similar format but with different content. Since the topic of that exercise was *Have to*, the same verb was searched in the BNC in order to know the verbs that are more frequently used with *have to*.

Frequency breakdown of lexical items for position "node" (4900 types and 80294 tokens)			
<	<<	>>	>
Frequency breakdown of tags only			Go!
No.	Lexical items	No. of occurrences	
1	have to be	7594	
2	had to be	5730	
3	has to be	4535	
4	have to do	2127	
5	have to go	1976	
6	had to go	1280	
7	have to pay	1169	
8	had to do	1010	
9	Have to get	962	
10	have to take	952	

Figure 1: Frequency breakdown of *have to* taken from the BNC

According to the BNC, the verbs *be*, *do*, *go*, *get*, *take* and *say* are more likely to go after *Have to*. However, it is worth to mention that none of these verbs were used in the exercise of the course book that the samples of this investigation were using. After doing the research, the worksheets were created with the data from the corpus. This system was used during one month period. First students were given the worksheets that they had to work with and submit to the teacher. Then, in the next class the worksheets were given back to the students, already graded, so that they can correct them and clear up their doubts.

When the month of treatment concluded, the post-test was taken by the students of both, treatment and control group. The test was very similar to the pre-test, but some of the questions were slightly changed. Finally, the students were also giving a survey regarding their experience and thoughts about the project.

Data collection

The data obtained from the research was collected and analyzed through excel spreadsheets and statistical graphics. The reason to choose the excel program was because it is widely available and easy to use. It is not necessary to learn new methods of manipulating data and drawing graphs. It also provides the user with a lot of control and flexibility. (Zaiontz, n.d.)

Results and Discussion

Pre-test

The following analysis is from the results taken from the pre-test and the post-test from both 10th B (treatment group) and 10th A (control group) and the comparison between them.

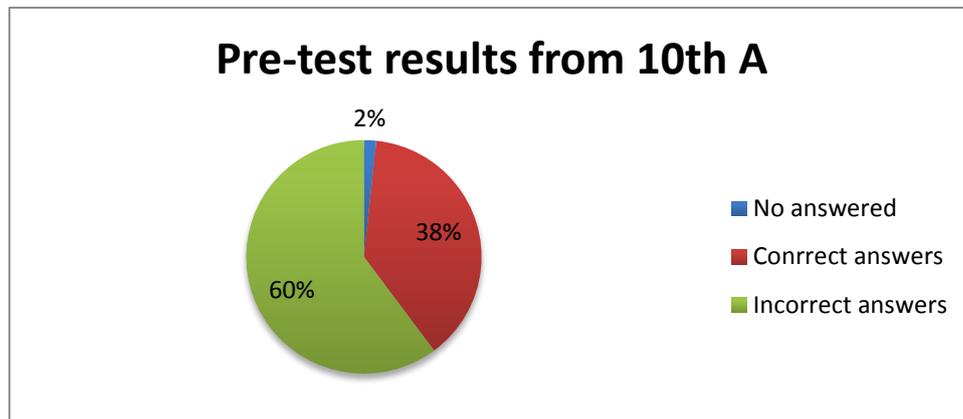


Figure 1: Performance in pre-test of 10thA students

Figure 2 represents the results in the pre-test of 10thA students, which is the class that did not work with corpora (control group). The results show that from a questionnaire of 23 questions taken to 31 students, the 38% of the questions were answered correctly. The 2% represents the questions that were not answered indicating the probability that students did not know any of the words in such questions. The 60% represents incorrect answers which is a considerable amount and indicates that students are lacking in the vocabulary area.

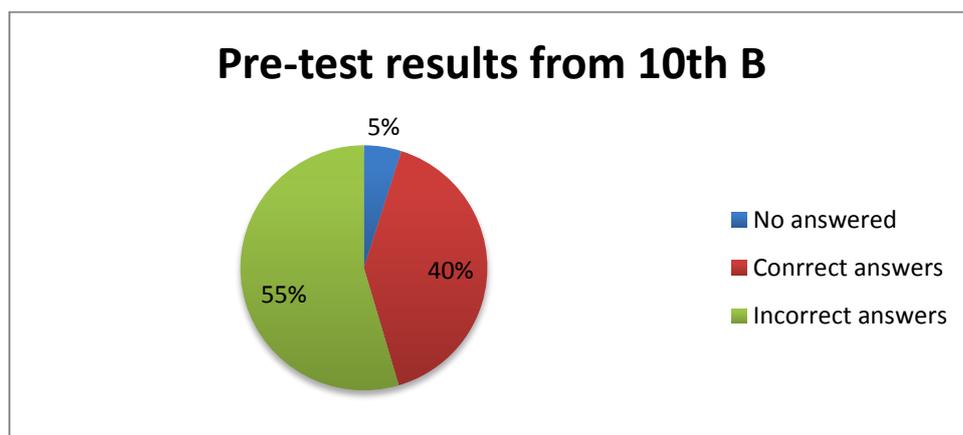


Figure 2. Performance in pre-test of 10thB students

Figure 3 represents the results from the pre-test of 10thB students (treatment group). The results are very similar to the 10thA test since their English level is also very similar. In fact, that is one of the reasons why these two classes were chosen. The graphic shows that this class had a 40% of correct answers and 55% were incorrect answers. From the 23 questions in the test, only in 5 questions the majority (more than 50%) of students answered correctly in both classes. Those questions are the ones with the easiest vocabulary which explains why the majority of students in both classes easily answered them correctly. However, there are also questions that the majority answered correctly in one group but not in the other group and others in which neither of the groups answered correctly.

Worksheets

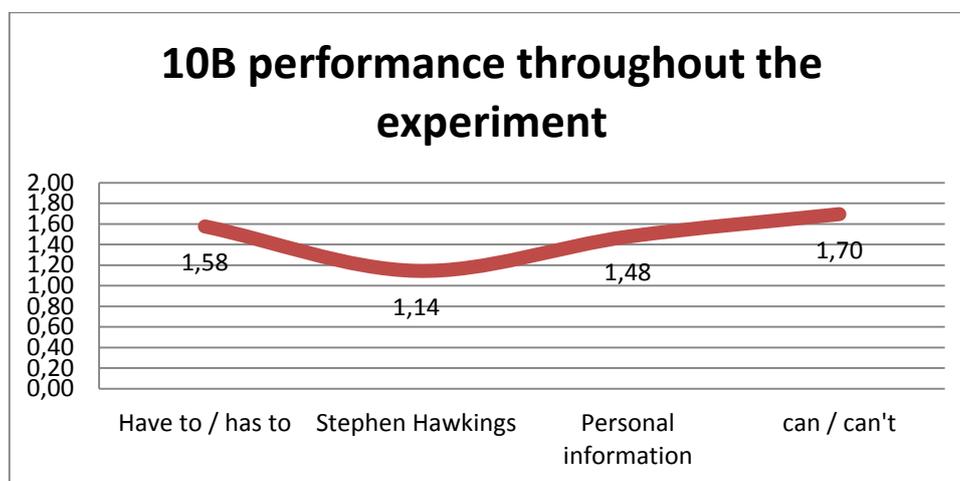


Figure 3. 10th B performance throughout the experiment

All the worksheets were scored over 2 points. As it is shown in the graphic above, the class had a good performance in all of the activities. Students had the lower grades in the *Stephen Hawking* worksheet. This is probably due to the type of activity since it was a comprehension reading and students had to answer according to what they read. Because it was a reading taken from the corpus, it had a lot of new vocabulary. As a result students had problems specifying the noun in the sentence and also they answered the questions in first person instead of in third person. It is worth emphasizing that this was previously pointed out to the students with an example, however, they made the mistake anyway. This is a difficulty that students have in all of the activities as well, sometimes due to distraction but it is something that goes beyond the scope of this research. It is worth emphasizing that students have discipline issues. It was hard for the teacher to make them listen to her. Students are hardly quiet, which made it difficult working with them. Overall, despite some mistakes, students had a good performance along all the experiment.

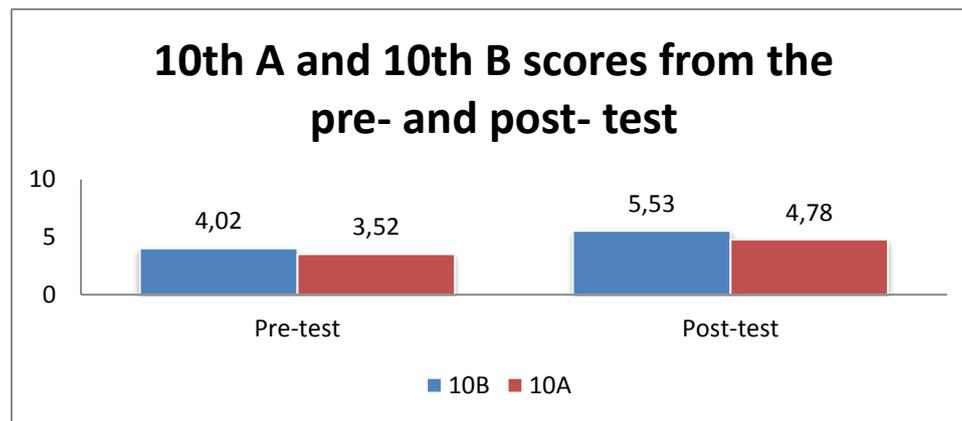


Figure 4. Post- and pre-test average score from 10th A and 10th B

Figure 5 shows the contrast between both groups in the post- and pre-test. In the pre- test 10th B scored an average of 4.02 out of 10 points and 10th A scored an average of 3.52 while in the post-test 10th B scores an average of 5.53 and 10th A scored an average of 4.78. In both tests 10th B had a better performance than 10th A. Even though 10th A did not work with corpora they managed to improve their percentages.

Survey

The survey was based on a rating scale in which 1 meant very little and 5 meant a lot. One of the reasons that encouraged this project was the lack of knowledge about corpus. Several authors state that the use of corpus is increasing more and more in the education field however that is not the case in Esmeraldas. From 28 students the 72% did not know anything about corpus before this project.

The next question was “Do you think working with corpora helped you to improve your English vocabulary?” This question had mixed results. From a total of 28 students the 15% chose number 1, meaning that they think working with corpus was of very little help or no help at all for their vocabulary improvement. The figure also shows that 2 was chosen by 22% of students, as well as number 3, meaning that they think working with corpus has helped them at some point. 19% of students chose 4 meaning that working with corpus helped them improving their vocabulary and other 22% chose 5 meaning they think working with corpora helped them a lot. It is important to emphasize that even though, as indicated by the results of the second question of the survey, students did find a lot of unknown words, more than a half of the students thought that it was not of help for their vocabulary improvement.

The final question was aimed at investigating interest of students in the process. The majority, 39% of students, chose number 5 meaning that they thought working with corpus was a lot interesting. 11% chose number 4 meaning that they thought it was less interesting. 18% chose 3, meaning it was somewhat interesting. And a 21% and 11% chose 2 and 1 respectively meaning they had very little or no interest whatsoever in the corpus. In the pedagogical field, it is important that students have interest in a material that helps improve the motivation of the class and academic results.

Conclusions

Due to problems with the English level in Esmeraldas, this research work attempted to present corpus as an alternative for teachers in the teaching and learning process. There is highly evidence that corpora can also be very useful in the education area. Vocabulary is one important element of a language. In order to understand a foreign language vocabulary needs to be learnt. Therefore, the hypothesis of this study was that corpus is highly beneficial for the process of teaching and learning vocabulary of high schools in Esmeraldas. This study does not pretend that the use of corpora is the definitive solution for improving vocabulary level. The aim was to state the benefits of using corpora in English classes in order to introduce it as an effective optional strategy that teachers could use when they create materials for their classes. After the analysis and discussion of the results, the following conclusions were reached:

The pre-test suggests that both classes have low vocabulary level. According to the LOEI, they both scored under the minimum required which is 7/10 (see figure 4). During the treatment it was notable that students not only had problems with vocabulary but also in other language elements. In the exercises, the majority of students made some mistakes especially regarding conjugation. Also the discipline is a major issue in both classes. The class was difficult to control. They do not listen to their head teacher which made even harder to work with them.

The results of the comparison between the pre- and post-test show that even though the experiment was made in a short period of time, there was an improvement in the English vocabulary of both groups. However, the increment is wider in 10thB (see Figure 4). This could be associated with the fact that 10thB used material based on information taken from the corpus, suggesting that the use of corpus in classes is beneficial for the purpose of improving their vocabulary level.

The results of the survey suggest that working with corpus can provide some benefits such as exposure to real language samples; it also challenges students, and boosts interest. Despite of the improvement, the

post-tests also showed that neither of the groups reached the minimum score required. It is undeniable that more time is needed in order to have better results.

Recommendations

It is advisable that teachers research more about corpus and acquire more knowledge about all its benefits for the teaching and learning process.

It is also advisable that teachers introduce it to students in order to help them to be more autonomous about their learning with the use of corpus.

Since corpus provides raw information, it is important to give enough time to students in order to analyze and practice the new vocabulary. A longer experiment may have improved the results of this research.

There are several types of corpora that have a lot of diverse information. It is important that teachers choose carefully and wisely the type of corpora that matches students interests and needs.

References

- Al Saeed, N., & Waly, S. (2009). *Corpus for Classrooms: Ideas for Material Design*. In Proceedings of the 10th METU ELT Convention. Ankara, Turkey.
- Atkins, S. (1992). *Corpus Design Criteria*. *Literary and Linguistic Computing*, 7(1), 1–16.
<http://doi.org/10.1093/lc/7.1.1>
- Calle, A., Calle, S., Argudo, J., Moscoso, E., Smith, A., & Cabrera, P. (2012). *Los profesores de inglés y su práctica docente: Un estudio de caso de los colegios fiscales de la ciudad de Cuenca, Ecuador*. *Biannual Journal of DIUC*, 3. Retrieved from maskana.diuc@ucuenca.edu.ec
- Carter, R. (1998). *Orders of reality: CANCODE, communication, and culture*. *ELT Journal*, 52(1), 43–56.
<http://doi.org/10.1093/elt/52.1.43>
- Coffey, S. (2004). *Investigating restricted semantic sets in a large general corpus: learning activities for students of English as a foreign language*. *Language and Computers - Studies in Practical Linguistics: Corpora in the Foreign Language Classroom: Selected Papers from the Sixth International Conference on Teaching and Language Corpora (TaLC)*, 61(1), 176.

Conrad, S. M. (1999). *The Importance of Corpus-Based Research for Language Teachers. System: An International Journal of Educational Technology and Applied Linguistics*, 27(1), 1–18.

EF English Proficiency Index. (2014). (No. 4). Retrieved from <http://www.ef.com.ec/~/media/centralescom/epi/v4/downloads/full-reports/ef-epi-2014-spanish.pdf>.

EF EPI. (2015). *EF Índice Inglés Proficiency - Inglés, Economía y Calidad de Vida*. Retrieved December 28, 2015, from <http://www.ef.com.ec/epi/insights/english-economics-and-quality-of-life/>

Exam English. (2014). Retrieved October 28, 2015, from <http://www.examenglish.com/>

Fernández, J. (2014). *The Expression of Polarity in Financial Discourse and its Application in Sentiment Analysis* (Trabajo de fin de Master). Universidad de Málaga, España.

Frankenberg, A. (2012). *Raising teachers' awareness of corpora. Language Teaching*, 45(04), 475–489. <http://doi.org/10.1017/S0261444810000480>

Granger, S. (2003). *The International Corpus of Learner English: A New Resource for Foreign Language Learning and Teaching and Second Language Acquisition Research. TESOL Quarterly*, 37(3), 538. <http://doi.org/10.2307/3588404>

Granger, S. (2008). *Learner Corpora in Foreign Language Education. In N. H. Hornberger (Ed.), Encyclopedia of Language and Education* (pp. 1427–1441). Springer US. Retrieved from http://www.researchgate.net/publication/226354169_Learner_Corpora_in_Foreign_Language_Education/file/5046351dd1ff8c57b6.pdf

Granger, S., Hung, J., & Petch-Tyson, S. (2002). *Computer Learner Corpora, Second Language Acquisition, and Foreign Language Teaching*. John Benjamins Publishing.

Huang, L.-S. (2011). *Corpus-aided language learning. ELT Journal*, 65(4), 481–484. <http://doi.org/10.1093/elt/ccr031>

IT Services, 13 Banbury Road. (n.d.). British National Corpus [Text]. Retrieved October 28, 2015, from <http://www.natcorp.ox.ac.uk/>

Kindt, D., & Wright, M. (2001). *Integrating Language Learning and Teaching with the Construction of Computer Learner Corpora*. Journal of the Nanzan Academic Society, 177–208.

O’Keeffe, A. (2007). *From corpus to classroom: language use and language teaching*. Cambridge ; New York: Cambridge University Press.

Perez, C., Moreno, A., & Butler, C. (2004). *An integrating approach to applied linguistics: academic and professional insights*. (F. Pineda Castillo, Ed.). Albolote: Editorial Comares.

Sinclair, J. M. (2004). *How to use corpora in language teaching*. Philadelphia: J. Benjamins.

University of Cambridge. (2012). KET Vocabulary List. Retrieved from <http://www.cambridgeenglish.org/images/22105-ket-vocabulary-list.pdf>

Zaiontz, C. (n.d.). *Performing Real Statistical Analysis Using Excel*. Retrieved from <http://www.real-statistics.com/>