

**PONTIFICAL CATHOLIC UNIVERSITY OF ECUADOR  
IN ESMERALDAS**



APPLIED LINGUISTICS SCHOOL

***THESIS REPORT***

***DO CORPORA BENEFIT THE LEVEL OF VOCABULARY OF 10<sup>TH</sup> YEAR OF  
BASIC EDUCATION STUDENTS? CASE STUDY IN DON BOSCO EDUCATIONAL  
INSTITUTION IN ESMERALDAS, 2015-2016.***

***TESIS DE GRADO***

***¿Corpus beneficia el nivel de vocabulario de los estudiantes de 10mo año de educación  
básica? Estudio de caso en la Unidad Educativa Don Bosco en Esmeraldas, 2015-2016***

PRIOR TO THE ACADEMIC DEGREE IN:  
APPLIED LINGUISTICS MENTION TEACHING ENGLISH

AUTHOR

Sindy Janeth Castillo Jaen

ADVISOR

Msc. Ingrid Grijalva

ESMERALDAS, 2016

## DISSERTATION COMMITTEE

“Thesis approved after complying with requirements established by the PUCESE degree regulations prior to obtaining the degree of BACHELOR OF APPLIED LINGUISTICS IN TEACHING ENGLISH”

---

President of Graduation  
Presidente tribunal de  
graduación

---

Reader 1  
Lector/a 1

---

Reader 2  
Lector/a 2

---

School Director  
Director de Escuela

---

Thesis Director  
Director de Tesis

Date/Fecha: \_\_\_\_\_

## **AUTHORSHIP**

“I, Sindy Janeth Castillo Jaen, declare that the investigation in the current thesis report research is absolutely original, authentic, and personal.

The content of this research is an exclusive legal and academic responsibility of the author and PUCESE.”

Sindy Janeth Castillo Jaen

C.I.: 0803111277

## **ACKNOWLEDGMENT**

I wish to express my deepest gratitude to the PUCESE and my teachers for helping me to accomplish this thesis. My special appreciation goes to MSc. Javier Fernández Cruz and MSc. Ingrid Grijalva. Their advices helped me to carry out my research and writing up my thesis.

## **DEDICATION**

I dedicate this thesis to God and my family for all their love and support. My special gratitude goes to my parents, Alberto and Yoconda, for all their valuable teachings which helped to become the person I am today.

## TABLE OF CONTENTS

COVER.....	i
DISSERTATION COMMITTEE .....	ii
AUTHORSHIP .....	iii
ACKNOWLEDGMENT .....	iv
DEDICATION.....	v
TABLE OF CONTENTS.....	vi
LIST OF ILLUSTRATIONS.....	viii
ABSTRACT.....	ix
RESUMEN .....	x
1. INTRODUCTION.....	1
1.1 Problem statement .....	1
1.2 Objectives.....	3
1.2.1 General .....	3
1.2.2 Specific.....	3
2. THEORETICAL BACKGROUND.....	4
2.1 Importance of English .....	4
2.2 Teaching Materials in Ecuador.....	5
2.3 Corpus Linguistics.....	6
2.4 Types of Corpora.....	8
2.5 Corpus Building .....	9
2.6 Corpora vs. Course Textbooks .....	11
2.7 Corpus in teaching.....	12
2.7.1 Tasks involving corpora .....	16

2.7.2 Problems when using corpora .....	18
3. METHODOLOGY .....	21
3.1 Type of investigation.....	21
3.2 Sample .....	21
3.3 Process description.....	22
3.4 Data collection.....	25
4. RESULTS AND DISCUSSION .....	26
4.1 Pre-test.....	26
4.2 Worksheets .....	31
4.3 Post-test .....	32
4.4 Contrast between pre- and post-test .....	36
4.5 Survey.....	38
5. CONCLUSIONS .....	42
6. RECOMMENDATIONS .....	43
7. BIBLIOGRAPHY REFERENCE .....	44
8. ANNEXES .....	49
8.1 Annex 1 .....	49
8.2 Annex 2 .....	57
8.3 Annex 3 .....	61
8.4 Annex 4 .....	62
8.5 Annex 5 .....	63

## LIST OF ILLUSTRATIONS

1 Exercise from book Engage, teacher's book 2 .....	23
2 Frequency breakdown of <i>have to</i> taken from the BNC .....	24
3. Performance in pre-test of 10 <sup>th</sup> A students .....	26
4. Performance in pre-test of 10 <sup>th</sup> B students.....	27
5. Students answers to question 3, pre-test. ....	28
6. Students answers to question 7, pre-test. ....	29
7. Students answer to question 19.....	30
8. Pre-test average score from 10 <sup>th</sup> A and 10 <sup>th</sup> B .....	30
9. 10th B performance throughout the experiment.....	31
10. Performance in post-test off 10 <sup>th</sup> A students.....	33
11. 10thB performance in post-test.....	33
12. Students answer to question 8, post-test. ....	34
13. Students answer to question 1, post-test. ....	35
14. Post-test average score from 10 <sup>th</sup> A and 10 <sup>th</sup> B.....	36
15. 10th A pre- and post-test results .....	36
16. 10 <sup>th</sup> B pre-test and post-test results .....	37
17. Post- and pre-test average score from 10th A and 10th B .....	37
18. Students answer to question 1 of the survey.....	38
19. Students answer to question 2 of the survey.....	39
20. Students answer to question 3 of the survey.....	40
21. Students answer to question 4 of the survey.....	40
22. Students answer to question 5 of the survey.....	41

## 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 will be 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 respond 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 the treatment period. So that 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 provides improvement in the vocabulary level of students.

## **RESUMEN**

El bajo nivel en las clases de inglés conduce a la necesidad de buscar nuevas técnicas y metodologías como alternativas que ayuden a resolver este problema. El presente estudio de caso investiga los beneficios que corpus pueda proveer a las unidades educativas en Esmeraldas, Ecuador. Durante el periodo de un mes, se aplicaran materiales basados en información tomada de un corpus en una clase de 10mo nivel en una institución Fiscomisional de Esmeraldas. Este estudio del tipo experimental utilizó principalmente el método cuantitativo para el análisis de documentos de clases y cuestionarios, con el fin de explorar las percepciones de los estudiantes y como estos responden al proceso. Además, se aplicará un pre-test y un post-test en dos clases, una en el cual se aplicara el corpus (tratment group) y otra la cual seguirá con su proceso de aprendizaje normal (control group) con el fin de conocer el nivel de vocabulario de los estudiantes, antes y después del proceso, y así, mediante una comparación de los resultados de ambos grupos y el análisis de una encuesta, ser capaz de exponer los beneficios que el uso de corpus pueda traer al campo de la enseñanza del Inglés. Los resultados indican que, a pesar de que tuvo un corto tiempo de implementación, el uso de corpus en las clases proporciona un mejoramiento en el nivel de vocabulario de los estudiantes.

## **1. INTRODUCTION**

Students in Esmeraldas have a low English level. According to the foreign teachers of PUCESE, in previous semesters, they had to lower the level of their lessons and their syllabi in order to work in a better way with the actual level of students. But the problem of low English level is not only in Esmeraldas. In fact, According to studies made 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 made in 2015 was based on exams about grammar and audio, took by over 18 years old people, which results located Ecuador in the 38<sup>th</sup> spot (with an average of 51.67/100) meaning a low level of English in the country (EF EPI, 2015).

The English level in Ecuador has to increase in order to become a more competitive country. The increment of the English level will have a lot of benefits for the city of Esmeraldas as well, such us a better economy, wider communication with the world, more prestige and more tourism, which is one of the most important resource of Esmeraldas. It is worth to emphasize that “nowadays English is a necessity”(Uribe, Gutiérrez, & Fernández, 2008). This research work could have very significant pedagogical implications. Teachers can take it into account when they choose activities for their lessons and syllabus designers could take it into account when they design the syllabi.

### **1.1 Problem statement**

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 found that the majority of English teachers in a high school in Cuenca City were either too old, with more than 35 years of experience but

not specialized in English, or too young without degrees, neither in English nor in Education. She believed that the lack of knowledge of the subject led the teacher to use the very traditional methods, such as the so known grammar translation method, and to base their teaching only in the given textbook. This is a big issue because although textbooks are necessary, they usually do not fulfill the needs and expectation of students (Ramirez, 2004). In addition, Calle also emphasizes the fact that some of the teachers would dictate the whole class in Spanish since they would not know any English at all. Finally, she concludes that in order to reach an improvement in English teaching, continuous training is required as well as new methodologies and materials.

The Minister of education (2014) states that high school students must have certain proficiency level set by the CEFR (Common European Framework of Reference for Languages); Level A1 at the end of 9th year Educación Básica General; Level A2 at the end of 1st year *Bachillerato*; and, Level B1 at the end of 3rd year *Bachillerato*. This means that the students graduated from high schools in Ecuador should have a B1 level (Minister of Education of Ecuador, 2014). The problem is that the material they use does not reach the aim of having B1 level students.

In the current world we live in, the internet is considered a valuable source in which we can find a lot of information that can or cannot be helpful for teaching and learning. Corpus (or its plural *corpora*) is a computer database that we can easily find on the internet. According to Granger, S., Hung, J., & Petch-Tyson, S. (2002) corpus can best be defined as a “linguistic methodology which is founded on the use of electronic collections of naturally occurring texts”. This is to say that corpus is a collection of texts, words, and transcripts of spoken language such as ordinary conversations, phone calls, radio broadcasts, and even TV shows. That is why it is considered real language.

According to the definitions detailed above, corpus might be a helpful tool to solve the problematic in the level of English in Esmeraldas. In these days, the lack of real and current vocabulary, dialogues, and situations in the English teaching system in Ecuador decreases

the motivation of students towards this subject, causing the level of English to decrease more and more. In addition, it is important to emphasize that corpora are not only useful for teaching current vocabulary to students, but also to help teacher to have a better understanding of English and to design “effective materials” for their students (Conrad, 1999).

To sum up, Ecuador has problems agreeing their aims with the reality of the country. Teachers are pushed to follow and complete an English program that does not have the current and/or real vocabulary that students need. 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.

## **1.2 Objectives**

### **1.2.1 General**

To state the benefits of using corpus in an educational institution of Esmeraldas, in order to present it as an alternative for class materials.

### **1.2.2 Specific**

- To identify students level of English vocabulary previous the implementation of the corpus based materials
- To create class materials based on information taken from Corpus
- To apply corpus based materials in a basic education class.

## **2. THEORETICAL BACKGROUND**

The current world we live in requires people to prepare more and more as the time goes by. Ecuador is currently one of the countries that has developed more in the last few years by investing in education with the purpose to go ahead with the more advanced countries. English is nowadays one of the major requirements for education in Ecuador since, as stated above, English 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, Ecuador scored 44.54 points over 100. In the latest edition of 2015, Ecuador 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) In other words, even though the improvement, Ecuador still has to keep looking for alternative such as new methodology, materials, and techniques in order to move forward regarding the learning and teaching of English.

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

### **2.1 Importance of English**

Students at schools and high schools always wonder why they have to study English. Why is it important or useful? English is not the most widely spoken language in the world, but it is spoken in more than 150 countries around the world and it has more than 480 million numbers of speakers (Comrie 1998, Weber 1997, & Summer Institute for Linguistics 1999), that is why it is considered the dominant language for business around the world or as stated by Clark in Forbes magazine, English is “the most essential language for global business success” (Clark, 2012). In fact the EF EPI declares that English is as essential for business today as it was alphabetization in previous centuries (EF EPI, 2015). This is the

reason why the government of Ecuador has as one of its goals to improve the level of English in order to “enter the international debate and the scientific-edge” (Minister of Education of Ecuador, 2013). In other words, Ecuador wants to have a better world wide communication and to be abreast with other countries regarding science.

Several journals in Ecuador make reference to the importance of English language for the development of the country. For example, Cáceres (2013) declared about the necessity of English for better job opportunities in any profession. It is also mentioned that the majority of bibliographic resources are written in this language, so if people want to obtain more information, they will find further data in English than in any other language, not only in physical papers but digitally (Internet) as well. However, students at high schools and schools don't care about better job opportunities or understanding an English book or text, they are too young for that. That is why is always important to look for new and innovative materials.

## **2.2 Teaching Materials in Ecuador**

Materials are a fundamental support in the English teaching process. In the past few years, the Ecuadorian Minister of Education decided to implement an English program called the Post Card program (Wilson, 2013). As a result, public schools had three types of English textbooks, two for general basic and the other one for baccalaureate. Ecuador demanded teachers to work with the book and to complete the whole book in one year class. The problem is that neither the syllabi nor the books were considered good enough.

According to Beltran (2013) in the journal “El Comercio”, the books were “too basic”. They use regular vocabulary from around nine years ago, forgetting that language changes constantly. In addition, the dialogues and examples used in those books do not motivate students due to their simplicity and boring situations and the vocabulary they use.

The ESOL (English for Speakers of Other Languages) in Ecuador is generally seen as something impossible to learn by students in public and private schools. As stated by the current president of Ecuador, Rafael Correa, in his weekly report to the country “English teachers in Ecuador actually do not know English” (“Deficiente la enseñanza de inglés en Ecuador,” 2012). This is due to the result of the TOEFL (Test Of English as a Foreign Language) test that English teachers in Ecuador took back in 2011. To be an English teacher it is necessary to have a minimum of B2 level, but the majority of teachers did not even reach the B1 level. This is one of the major problems in the English teaching and learning in Ecuador, because how can someone teach a language without the proper level of the field. Currently, Ecuador is working with the collection of books called “Engage”.

### **2.3 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, 1996 named 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.

Corpora were involved in the life of people since long time ago. In the past, dictionaries, and other materials were based on corpus. In the 90s, for example, dictionary makers used corpora data in order to create authentic and typical examples of a language. So even though people did not know, they were using materials based on corpus (McEnery & Xiao, 2010). Currently, many books, magazines, dictionaries, and other materials are also based

on corpus. Granger (2003) considers corpus as a valuable tool since it passed from being an enterprising idea of a few to become one of the most valuable tools for both, Linguistics and Computer science.

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. When authors refer to written and spoken texts, it means that both, text in papers and recordings, are transcribed and labeled. In the past, people have to compile the texts writing them by hand, like for example the first concordance of the Bible, which was made by around “500 monks in the thirteen century” (Tribble & Jones, 1990, named in Kindt & Wright, 2001). This indeed demanded too much effort and time. But technology advances so fast that currently It cannot only have texts in a corpus but also multimedia elements such as videos, scans and recordings in a way that they can appear in the computer screen in matter of seconds when someone queries them.

O’Keeffe (2007) also 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 lets the researcher do the analysis.

Conrad (1999) emphasizes the benefits of using corpus because “it facilitates the use of authentic language”. People who use real language are usually native speakers of a language. However, not all corpora are made from native speakers’ interaction (there are, for example, corpora of second language learners or corpora of mistakes in language, etc.). 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 consider 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).

## **2.4 Types of Corpora**

Over the years, corpora have also evolved, not only in technology but also in their scope and purpose as well, to the point that currently we can find a large variety 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, meaning only written text are compiled, spoken corpora, meaning only transcripts of recordings are compiled, or a mixture of both written and spoken corpora. However, corpora can be also classify in terms of the varieties of language, that is the “specific domain or geographical origin” (Perez, 2004) of the language used in the corpus since, for example, an Spanish corpus made in Colombia will not have the same results as a corpus made in Ecuador even though both countries have Spanish as the first language and both are south American countries.

Also, there are synchronic and diachronic corpora. Synchronic referring to the ones with the purpose of reflecting a language in a determined period of time, for example a corpus of the 80’s will based on texts used in that exact period, and diachronic with the intention of revealing a language in order that researchers can analyze the evolution of it over the years.

We can also classify corpora as annotated or raw. A raw data corpus will contain texts only with words and punctuation marks, on the other hand, annotated corpus is raw text but with extra information such as “formatting information (page breaks, paragraphs, and fonts), identifying information (author, date, genre), and linguistic information (word class,

syntactic structure, discourse markers)” (Kindt & Wright, 2001) which are helpful for a linguistic analysis.

Finally, O’Keeffe (2007) also classifies corpora as monolingual; uses only one language, comparable; two or more languages in order to compare them, and parallel corpora; one text translated in two or more languages.

## **2.5 Corpus Building**

As stated above a corpus is created by collecting texts. However, how is the whole process made? Atkins (1992) details this process in five stages.

The first two stages are concerning with preparation. The first one is the *planning stage*. Here is where the linguistic designs, costs, and administration need to be set up. For the linguistic design, specialists who know about the matter (such as sociolinguists, statistical experts, and linguistics expertise) need to be involved in order to make sure that all the “decisions regarding design, balance, and processing match with the aims of the corpus” (Atkins, 1992). However, it has to be taken into account that some design decisions such as the size of the corpus cannot be predicted since “the appropriate size of a corpus depends on its use” (Kindt & Wright, 2001). As the second stage, Atkins (1992) points out that all *copyright permissions* of the texts to be used in the corpus need to be in order. This is one of the most tedious and time-consuming stages but it is one of the most important as well.

The third stage is related to *data capture*. Sinclair (1995, quoted in Kindt & Wright, 2001) states that “the purpose of assembling a corpus is to put together data in quantity” so while larger the amount of texts, more effective the corpus is considered. As stated before, corpora can be written, spoken or both. However, Kindt & Wright (2001), think that to collect written data is easier than spoken data. Written data can be obtained already in

digital form from the internet or the population can hand it over. Then, word processing programs can be used in order to find and correct all the errors of spelling, grammar, etc. (Meunier, 1998, named in Kindt & Wright, 2001). Also, according to (Atkins, 1992) Another way of processing the collected written data is to use OCR devices, which are “a conversion of images of typewritten or printed text into machine-encoded text” (“Optical character recognition,” 2015), in order to scanned printed material and insert it in the data base of the corpus instead of transcribe them manually. On the other hand, collecting spoken data is more difficult because the recording and transcription process are time consuming and requires massive effort from human resource, and also because the people who will be recorded have to agree with it and with the fact that their speech will be in the corpus (Kindt & Wright, 2001).

The fourth stage is *text handling* which concerns with the different tools used to process the text. According to Atkins (1991), there are basic tools and advanced tools. Word frequency and concordance are basic tools. The word frequency is software that generates “lists of word type and their frequency in the corpus” (Atkins, 1991) which refers to lists of the words that are more frequently used in the corpus. The concordance is software that shows words in context. This is also calls KWIC (Key Words In Context), which is basically an index of the words in the text. In other words, not only the key word or phrase will appear on the screen but also the words that come before and after, providing the data which can help to conclude a definition of the key word or phrase (Kindt & Wright, 2001). The advanced tools, on the other hand, Atkins describe them as Lemmatization, part of speech labeling, parsing, and collocation. According to Atkins (1992) lemmatization is about presenting the lemma or base of an inflected word. This is to say that it takes out the inflected forms and reaches only the base of a word. Part of speech labeling or Tagging, as stated before, it is to add grammatical and syntactic information to a word (Kindt & Wright, 2001). Parsing is defined by Atkins as to set syntactic trees or bracketing in sentences of the corpus which comes after tagging, and identifies words functions as parts of a clause. Finally, according to Sinclair (1991, named in Kindt & Wright, 2001), collocations are words or a phrase that goes naturally together. Thereby the collocation tool will show statistical data about collocations of a language. Barnbrook (1996, named in Kindt &

Wright 2001) also details some tools such as Parsing, and Z-scores, t-scores, and mutual information which are different methods of measuring relative probability of co-occurrence of words and phrases.

The fifth and final stage Atkins calls it *user feedback* and *corpus development*. It is basically the comments, suggestions, and opinions of the people who use the corpus. This will have to be taken into account by the corpus builders in order to improve and enhance the corpus.

## **2.6 Corpora vs. Course Textbooks**

Carter (1998) contrasted two examples of conversations in a hair salon. One was taken from a real conversation and the other one from a course textbook. He noticed that the dialogue from the course textbook, apart from not reflecting the reality of hair salon conversations, it emphasized too much on the modal “can” because the subject to teach was indeed “can”. This could later lead to problems in the communication of student in the target language. For example, Neff et al. (2001, named in Al Saeed & Waly, 2009) presented some reasons of why L2 (second language) learners over use modal verbs. In their research, they examined a comparable corpus of 400.000 words between Spanish learners of English and native English speakers and they realized that Spanish learners used *can* and *must* more than any other modal. The conclusions they reached were that possibly the modal *can* was, as in Carter above, one of the first modal verb that those students learnt. Also that those students might think the modals *can* and *must* can be used in the same way and place that the Spanish modals *poder* and *deber* are used. This is probably due to the reason that course books over emphasize the use of the structure to teach in the examples and activities presented to students. The problem is that in course books the topic to teach has more relevance than the context and reality of the dialogues presented. Students could not understand real conversations when they have to express themselves outside the classroom. The interaction with a native speaker would be difficult, or, for example, when students hear or watch songs or television shows in the target language they could not be

able to understand a word of what is being said and that can cause them to perceive the subject as something useless that they will never learn. 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 Carter says, “Less real pedagogically” (Carter, 1998).

So which one is better? As stated before, McEnery & Xiao (2010) mention that in the past lexicographers used to create examples of the dictionaries because traditionally they believe learners could not understand real examples with authentic language. So, as Carter above, they presented simple examples, focusing, emphasizing and repeating only the use of a certain word or structure. The COBUILD (Collins Birmingham University International Language Database) was the first dictionary to start using authentic data taken from corpora and therefore to produce real examples from authentic language (McEnery, 2010), Which is one of the visible advantages of using a Corpus. It provides the researcher with large data from authentic language. So, if the objective of teaching English in Ecuador is to have B1 level graduated students (See Problem Formulation) it is important to teach them how real language is. According to Carter, real language is full of “interruptions, hesitations, false starts..., vague language, and even ungrammatical forms” (Carter, 1998) so it is important students to be familiarized with it since, like Burn says, “scripted dialogues lessen students possibility to understand interaction outside classroom” (Burn, 2001; named 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. That is why, like Kennedy states, second or foreign language learning is “a process of learning explicit knowledge with awareness, which requires a great deal of exposure to language data” (Kennedy, 2003; named in McEnery, 2010)

## **2.7 Corpus in teaching**

Conrad (1999) says that the use of corpus and computers has increased in English classes lately. But, before getting into corpus it is necessary to first discuss the usage of materials

in current education. According to Ramirez (2004), materials are basic and necessary in the process of teaching English. The material that has been used the most throughout the years is the textbook. Textbooks usually do not reflect the reality of students nor they fulfill their needs and expectations (Ramirez, 2004), since, as discussed above, they usually bring created and scripted dialogues and situations or, as Mindt states “one common failure of English textbooks is that they teach a kind of school English which does not seem to exist outside the foreign language classroom” (Mindt, 1996; named in McEnery, 2010). Ramirez (2004) also emphasizes the importance of material development and adaptation teaching process. In her research, Ramirez states that teachers must always adapt materials which requires a detailed process; first, identification of students' needs and problems; second, exploration of the language and skills; third, to find the right ideas and texts; fourth, appropriated instructions, exercises and activities; fifth, the reproduction of the material; finally, students' evaluation of the material. This is a helpful process in order to improve teaching and make the materials appropriated for the students.

On the Internet it is possible to find a lot of helpful information and materials. For example Boloña, (2010) who thinks that the Internet opens a large academic resource and teaching materials for teachers who want to incorporate it into their classes, investigated how students of an intermediate level in a university in Ecuador used a program on the Internet and she reached the conclusion that Internet programs have “highly beneficial effects” on the teaching/learning English process for teachers and students. Corpora can be easily found on the internet. McEnery says that the role of corpora in Language Education is “to provide more realistic examples of language usage that reflect the complexities and nuances of natural language” (McEnery, 2010) since he believes teaching should be based on real and authentic evidence rather than intuitions from lexicographers or teachers. McEnery also emphasizes that in order to make learning more effective, syllabi should be based on empirical data of real evidence with frequency of usage as its guide (Mindt 1996, cited in McEnery, 2010) meaning that educators have to create the syllabi based on information from real experiences that occur the most frequently in the target language.

Although Mindt argues that syllabi should be based on real data with “frequency of usage as a guide to priority for teaching” (Mindt 1996, named in McEnergy 2010), Meunier (2002, in McEnergy 2010) argues that frequency alone is not enough. But, there must be a balance between frequency, difficulty, and pedagogical relevance. Kaltenböck and Mehlmauer-Larcher (2005, in McEnergy 2010) also agree with this saying that frequency is not only useful in pedagogy but that teachers can use it in order to create better materials and better information to use in class and also to help syllabus designers.

McEnergy (2010) points Lexical syllabus as a good option for this goal. Sinclair (2004) detailed in his book how George (1963, named in Sinclair, 2004) analyzed a corpus back in the 60’s and he found that the common structures like present simple are not frequently used to indicate interactions. Also, more recent works such as Mindt (2000, cited in Sinclair 2004) agree with this, saying that most EFL (English as a Foreign Language) and ESL (English as a Second Language) textbooks emphasize structures such as present simple. However, people use more “the progressive present, actual present, and timeless present” (Mindt2000, cited in Sinclair 2004) to indicate interactions. Ljung (1991, named in Sinclair 2010) compared the one thousand more frequent words in an EFL textbook of secondary level with the COBUILD and he found that 20% of the words did not coincide. In conclusion, there is highly evidence that textbooks tend not to be based on structures and words that are more frequently used in the current target language.

Kindt & Wright agree with the importance of frequency implications saying that a corpus might help teachers to know “which words or phrases students are more likely to need and thus most important to teach” (Kindt & Wright, 2001). They also point out the fact that according to Aston (1995, named in Kindt & Wright, 2001) the study of authentic collocations, which are one of the advanced tools in a corpus (see Corpus Building), helps students with speed in language processing, production and fluency which will help them to approach to a native-like speech. In their research, they also argue the fact that teachers should use both native speaker corpora and learner corpora for their classes, materials, and syllabus design. The native speaker corpora is important in order to show what is common

in the target language, and the learner corpora to know the difficulties and general mistakes students have in the target language. Also, some authors like Higgins (1991, quoted in Kindt & Wright, 2001) think that the major contribution of computers in pedagogy is the concordancer. A concordancer is an index of Key Words in Context (KWIC) (Kindt & Wright, 2001). This is, whenever a word or phrase is entered in a corpus, there will emerge a list of all the sentences that contain that word or phrase. Those words or phrases are called KWIC (see Corpus Building).

As emphasized above, 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 is no more than 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 2006, named 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 2006, named in Huang 2011).

### 2.7.1 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 task a little more complicated 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.

According to Skeham (1996, named in Coffey 2004) good activities must have as goals; accuracy, complexity and fluency. As for the activities described above, they have accuracy

when students organized their data in order to give a feedback and also when students present in front of the class. Regarding fluency, concordancing technique is also seen as a form of practicing language fluency (Aston, 1995, named in Kindt & Wright, 2001). Finally, we can see complexity when students are asked to write appropriate collocations in their writings since it is something difficult but not impossible to do. So, it challenges them more.

Al Saeed & Waly (2009) presented several other examples of activities and ideas for material design for introducing corpora to students that precisely contain accuracy, complexity and fluency. The first one is called Verb Pairing Game. In this activity students are divided in two groups A and B. Group A will be giving sentences taken from corpus with the verbs *do* and *make* and they have to stand in a line facing each other. Then, group B will say their sentences and group A have to guess which verb, *do* or *make*, is more suitable for the sentence. Another activity is a fill in the gaps type activity called Computer Cloze Activity. The teacher must give students sentences with gaps taken from a corpus in a word document. Students then have to fill in the gaps. Al Saeed & Waly (2009) say it is necessary to incorporate technology in the class but this activity can also be done without technology involved. Finally, another activity for beginners would be to present data in a way that students can realize how some nouns occur more frequently with some verbs and not with others.

For intermediate students Al Saeed & Waly (2009) presented several other activities. The first one is about contextual analysis. Students are giving a number of sentences taken from a corpus and they have to try to analyze the context of a determine structure like, for example, modal verbs. Another one is a cloze activity. Students are giving sentences taken from corpus with gaps. The gaps are the topic to teach, in this case modal verbs. After students finish filling in the gaps the teacher will then show the correct answers. Then, the whole class will discuss the results. Finally, activities about self-expression are also presented. Students are given pieces of paper containing suggestion and obligation

sentences. Then each student have to say on the board if their sentence is accurate for the situation or not and why.

Saeed & Waly (2009) also presented phrasal verb activities for upper intermediate level students. For this, previous training of corpora it is necessary. The first activity is about Concordance Analysis. Teacher can ask students to look for idiomatic expression in corpora in order to analyze them and understand how native speakers use it. For the final activity students must have a previous knowledge about meanings of apology. Students are given some dialogues and they have to analyze the level of sincerity of the text.

The process of building a corpus can also be used as an activity for students. Kindt & Wright (2001) they created a learner corpus and integrated students in the very beginning of the corpus creation. They collected students typed activity about a determine subject. In this case the subject was “what do you think about ...” They highlighted the mistakes and give them back to students so that they can correct them by themselves. Later, they did the same pattern and the results showed improvement in the mistakes. In addition, students gave a positive feedback, arguing that they felt encouraged by this new activity and that it was easy for them to realize the mistakes they usually do and to correct them (Kindt & Wright, 2001).

As above, there are lots of other activities that can be used in 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.

### **2.7.2 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), which is one of the easiest corpora to work with. 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, 2010 named 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 used 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 novice users of corpora such as students use “dictionaries or web browsing techniques of research in corpus” (Frankenberg, 2012). For obtaining information from dictionaries or websites, people only need to query the word they need. This is the most common thing people do when they do not know a word. But 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.

Also, Frankenberg (2012) says that new users of corpora tend to give up on the research when they do not quickly get what they want. This is understandable since the only thing they will get from corpora is no more than raw data that they will need to analyze. Teachers and students must know that queries are e “too general or too restrictive” (Frankenberg, 2012) will not work with corpora. In this situation, what teachers must do is to make simpler the data found on corpora for those students with no experience with it (Al Saeed & Waly, 2009). Even though some researchers like Frankenberg, Al Saeed & Waly, and Carter think that every person who wants to use corpora must receive training in order to know how to analyze and understand the raw data presented in a corpus, Bouton (2010, named 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.

### **3. METHODOLOGY**

Previously, this research has discussed the necessity of applying new materials and techniques for the improvements of the teaching and learning of English in the country. This case study focuses on using materials generated from corpora in order to identify the benefits of Corpus in high schools of Esmeraldas.

#### **3.1 Type of investigation**

The lack of realness and authenticity in books used by teachers leads to the hypothesis of this study that the use of Corpus in Esmeraldas' high school classrooms is highly beneficial for the teaching and learning process of vocabulary. Therefore, the type of investigation used in this research is the experimental, 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.

#### **3.2 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. Before the start of the research, a request was submitted to the rector of the institution in

order to obtain the approval to their students being involved in the process (Annex 1). A project, detailing all the process, was also submitted. The target student population in this research work was the tenth year of Basic Education students at Don Bosco high school. The total population are four 10th levels; 10th A, 10th B, 10th C and 10th D. 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.

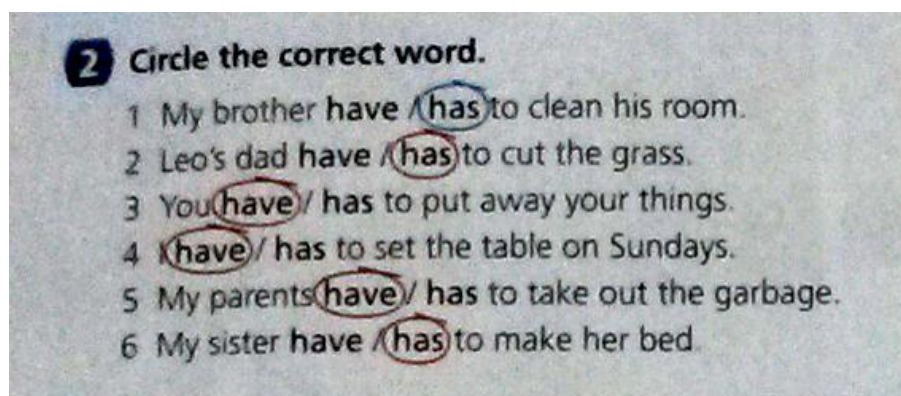
### **3.3 Process description**

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). The questions were made on a multiple choice format 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 (Annex 1).

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/bncwebSignup/>), due to its easy access and simple use (IT Services, n.d.). The way corpus was applied was by creating worksheets with several exercises 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 agreed the schedules and how the corpus will be applied so that the worksheets were created according to the plans

of the teacher. It is to say that if the teacher was planned to teach *can and can't* for that certain week, then 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 following is an example of how the exercises were created. The students are currently working with the book *Engage* so if the students worked in the book with an exercise like the following one:



1 Exercise from book Engage, teacher's book 2

Then the worksheet was created with a similar format but with different content (Annex 2). 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	<a href="#">have to be</a>	7594
2	<a href="#">had to be</a>	5730
3	<a href="#">has to be</a>	4535
4	<a href="#">have to do</a>	2127
5	<a href="#">have to go</a>	1976
6	<a href="#">had to go</a>	1280
7	<a href="#">have to pay</a>	1169
8	<a href="#">had to do</a>	1010
9	<a href="#">Have to get</a>	962
10	<a href="#">have to take</a>	952

## 2 Frequency breakdown of *have to* taken from the BNC

The results were that the verbs *be*, *do*, *go*, *get*, *take* and *say* are more likely to go after *Have to*. It is to say that, according to the BNC, those verbs are more frequently used after *Have to* for native English speaker. However, it is worth to mention that none of these verbs were used in the exercise of the book. After doing the research, the worksheets were created with the sentences and texts obtained 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 their doubts.

The topics of the worksheets were *have to*, *Stephen Hawking*, *Personal information*, and *can/can't*. All of the worksheets were only taken by the treatment group and every activity was reviewed in the following class. The first activity, *has to*, was about underlying the correct answer and filling in the gaps. The following activity was about *Stephen Hawking*. This activity was a comprehensive reading in which students had to answer questions about the reading and to choose true or false. The *Personal information* activity was another comprehensive reading in which students had to complete a table with the information they read before. The final activity was named *can/can't*. In this activity students had to answer

correctly with can or can't. A schedule of all of the activities is attached in Annex 4.

When the month of treatment was 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 (Annex 3). Copies of the worksheets as well as the pre-, post-test and survey used in this study are included in the Annexes

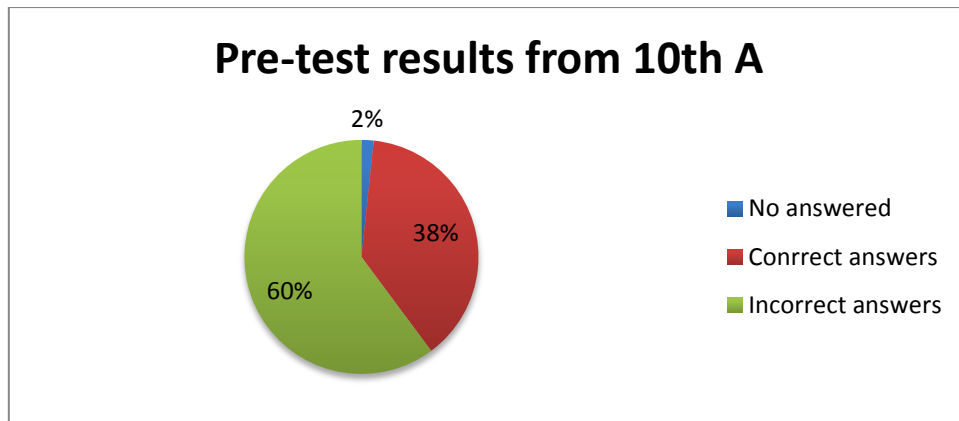
### **3.4 Data collection**

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.)

## 4. RESULTS AND DISCUSSION

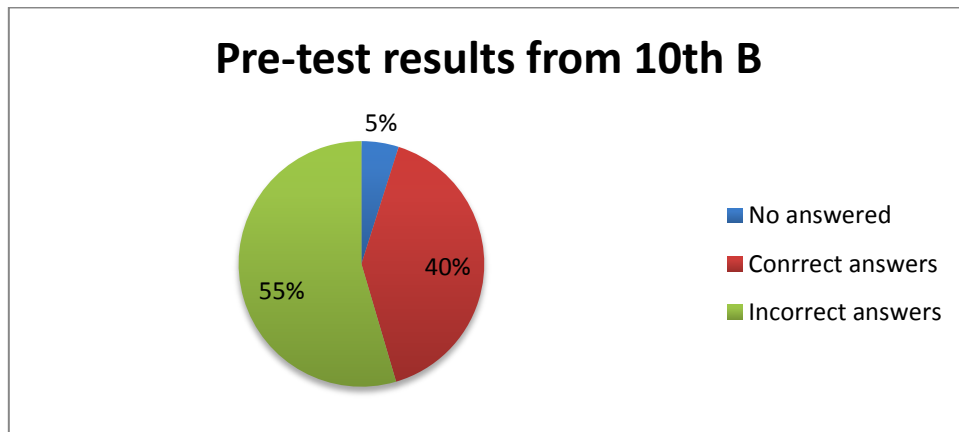
### 4.1 Pre-test

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



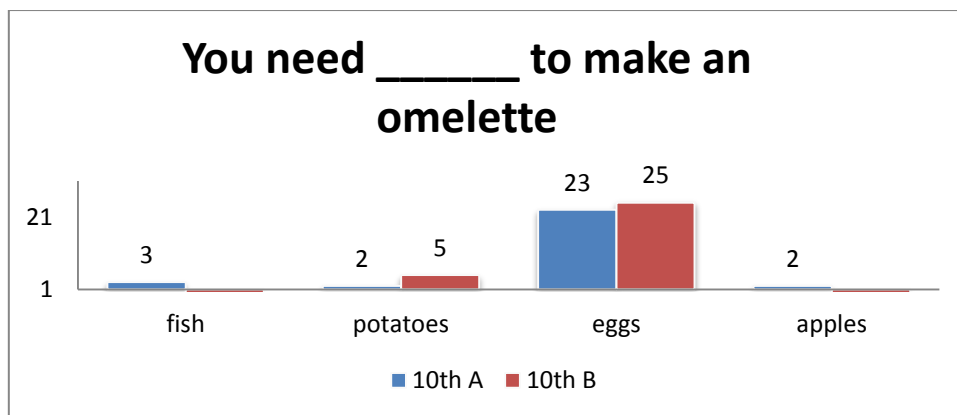
3. Performance in pre-test of 10<sup>th</sup>A students

Figure 3 represents the results in the pre-test of 10<sup>th</sup>A students, which is the class that did not work with corpora. 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.



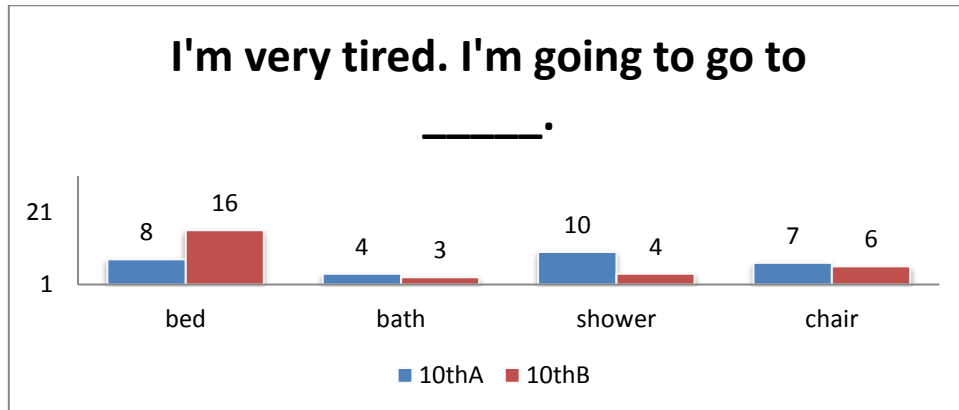
4. Performance in pre-test of 10<sup>th</sup>B students

Graphic 4 represents the results from the pre-test of 10<sup>th</sup>B students which is the class in which the experiment was implemented. The results are very similar to the 10<sup>th</sup>A 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, 2% more than 10<sup>th</sup> A. However, they had a 5% of questions that were not answered, 3% more than 10<sup>th</sup> A. Thus, the remaining 55% were incorrect answers. As stated before the test was made on a multiple choice format. The questions tested with an A2 vocabulary level. From the 23 question in the test, only in 5 questions the majority (more than 50%) of students answered correctly in both classes. However there are also questions that the majority answered correctly in one group but not in the other group. The following graphics show how students answer some of the questions in the pre-test.



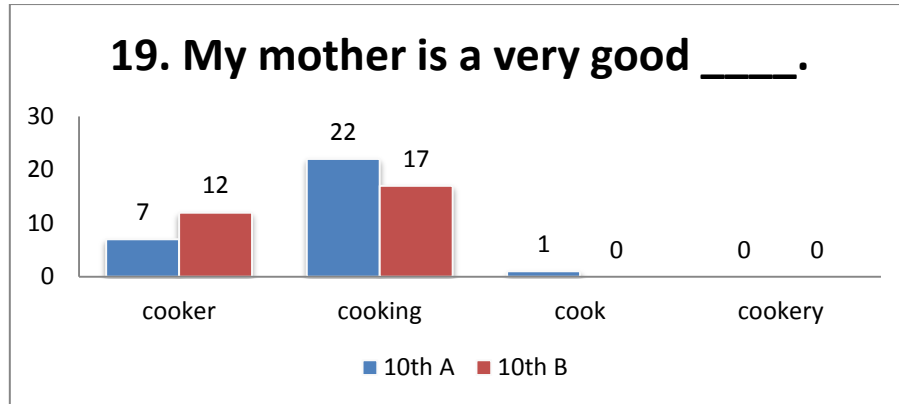
5. Students answers to question 3, pre-test.

The graphic above shows one of the questions that in their majority students answered correctly. The questions was “*you need \_\_\_\_\_ to make an omelette*” in which students had four options; *fish, potatoes, eggs* and *apples*. As it is shown, out of 31 students 23 answer correctly in 10<sup>th</sup>A, representing 74% of that class. On the other hand, 25 students answered correctly in 10<sup>th</sup> B representing an 80% of the class. There were also other questions in which the majority of students in both classes answered correctly. For example question number 5, which was “*I’m \_\_\_\_\_ I’m late. My car broke down.*” For this question students also had four options which were *wrong, sorry, bad* or *unhappy*. From which of course the correct answer was *sorry*. In 10<sup>th</sup>A a 58% answered correctly and in 10<sup>th</sup>B an 80% also answered correctly. Other question was number 9 “*Rome is the \_\_\_\_\_ of Italy*”. The options were *big, lead, capital* and *head* from which of course the correct answer was *capital*. 10<sup>th</sup>A had a 74% of students answering correctly to this questions and 10<sup>th</sup>B had an 87%. Question number 15 was “*My sister earns 200 \_\_\_\_\_ a week*” and the four options were *money, cash, dollars, notes*. From 10<sup>th</sup>A, the 74% of students picked *dollars* and from 10<sup>th</sup>B, the 77% also picked “*dollars*”. Finally, the last question, which the majority of students in each class answered correctly, was “*Bill loves repairing cars and he wants to train to be a \_\_\_\_\_*”. The options were *mechanic, doctor, painter* and *waiter* from which the correct answer was *mechanic*. 51% of students in 10<sup>th</sup>A answered correctly and 80% of 10<sup>th</sup>B also answered correctly. These questions happen to be the ones with the easiest vocabulary which explains why the majority of students in both classes easily answered them correctly. However there were other questions that happen to be answered correctly by the majority of students only in one class but not in the other one.



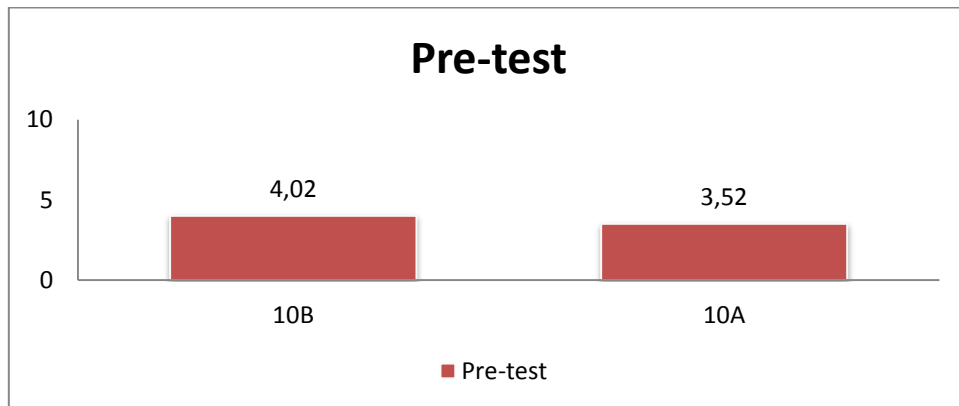
6. Students answers to question 7, pre-test.

Graphic number 6 shows how the majority (50% or more) of students in 10<sup>th</sup> B chose the correct answer however that was not the case in 10<sup>th</sup> A. This situation also happens with other questions such as number 12 and 14. Question number 12 was “*Can you \_\_\_\_\_ Spanish?*” The options given to the students were *talk*, *tell*, *say*, and *speak*. The correct answer, which is *speak*, was picked by 54% of students in 10<sup>th</sup>B and 45% of students in 10<sup>th</sup> A. Question number 14 was “*Jack is not well. I think we should call an \_\_\_\_\_*”. From the four options *illness*, *ambulance*, *accident*, and *engine*, the 61% of students in 10<sup>th</sup> B picked *ambulance* while in 10<sup>th</sup>A only the 45% picked *ambulance*. In these three questions, 10<sup>th</sup> B was the class with the majority of students choosing right the answer. However in question number 23, which was “*Jorge is Venezuelan and his wife is \_\_\_\_\_*”, 10<sup>th</sup> A was the one with the majority of student picking right their questions with a 54% while in 10<sup>th</sup> B only a 45% of students picked the right answer. The rest of the question from the test did not have good results from neither of the classes but the one with the worst result is the one displayed in the following graphic.



7. Students answer to question 19.

Graphic 7 shows the question in which students had the worst performance. The question was “*My mother is a very good \_\_\_\_*”. The four options were *cooker*, *cooking*, *cook* and *cookery*. Students may be confused because one of the rules to name professions is to add -er at the end of a verb. However, this is an exception since the correct answer is *cook*. That may be one of the reasons of the low results in this question. The results were that in 10<sup>th</sup>A only one student chose the correct answer and the majority chose *cooking* and in 10<sup>th</sup>B nobody chose the correct answer and the majority also chose *cooking*.

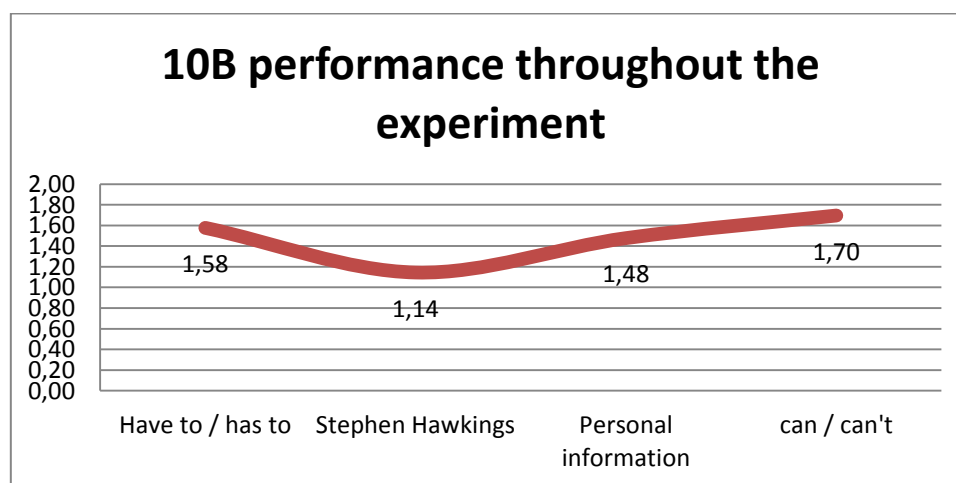


8. Pre-test average score from 10<sup>th</sup> A and 10<sup>th</sup> B

Figure 8 shows the average score of the pre-test from both classes 10<sup>th</sup> A and B. The tests were scored over 10 points. Each question was worth 0.44 points. The 10<sup>th</sup>B had an average score of 4.02 over 10 points and the 10<sup>th</sup> A had an average score of 3.52 with a difference of 0.50 which again indicates that both groups have very similar levels. However, both scores are low. According to the *Reglamento General a la Ley Orgánica de Educación Intercultural* (LOEI), 4 points or less means that the students do not reach the level of learning required since the minimum score to pass is 7 points over 10 (Minister of Education of Ecuador, 2013). The fact that these classes barely reach the 4 points shows the low level in vocabulary of students.

## 4.2 Worksheets

As stated before, 10<sup>th</sup> B, the class in which the experiment was implemented, was giving worksheets with data taken from the corpus. The following graphic displays the average score that the class had in each worksheet activity.



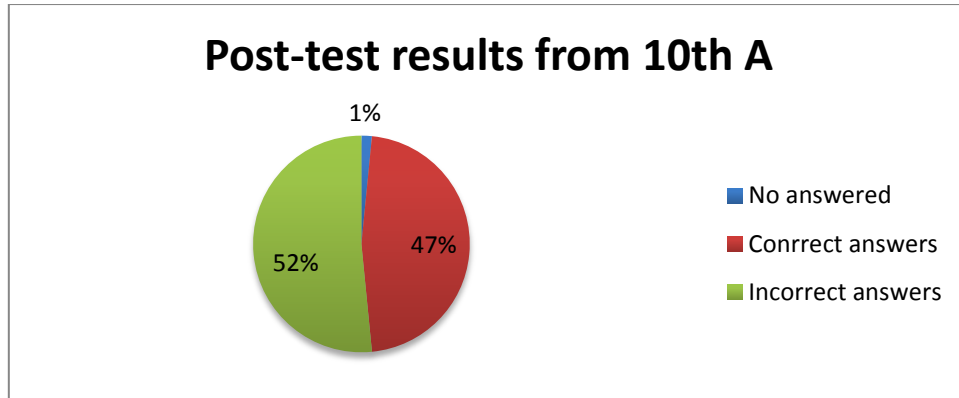
9. 10th B performance throughout the experiment

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. First off, in the *Have to /has to* worksheet,

students scored an average of 1.58 points which is overall good. The majority of mistakes in this activity were made when conjugating *have* with an auxiliary. Students had the lower grades in the *Stephen Hawking* worksheet. This is probably due to 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 answer the questions in first person which was a mistake because if the reading was in first person the answer should be in third person. It is worth emphasizing that this was previously pointed out to the students with an example however they did the mistake anyway. After all, in this activity students had an average grade of 1.14 points. The third worksheet was about personal information. This was also a comprehension reading in which students had to complete a table based on the reading. Here, it can be seen an improvement with an average score of 1.48 points. Finally, the *can/can't* worksheet had the higher score with an average of 1.70 points. In this activity some students have problem conjugating the verbs with the third person in present singular. This is a problem that students do regularly 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 hardly are quite which made it very hard working with them. Overall, despite some mistakes, students had a good performance along all the experiment.

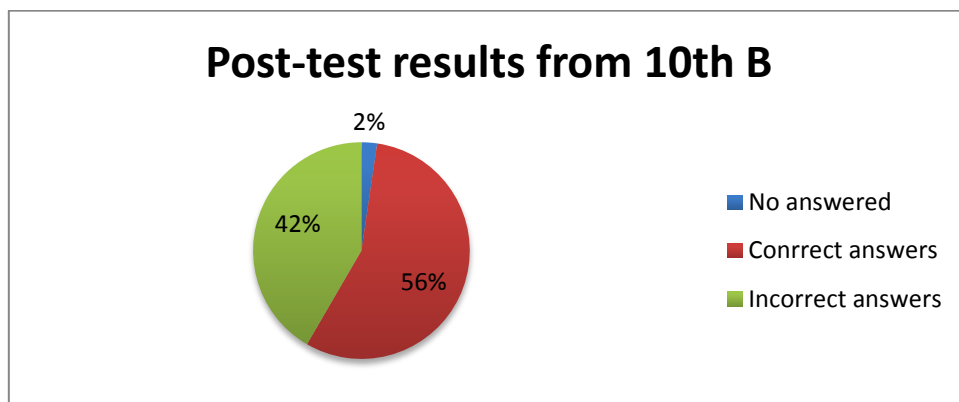
### **4.3 Post-test**

As stated before when the month of treatment was over the students took the post-test. The following graphics show the results.



10. Performance in post-test off 10<sup>th</sup>A students

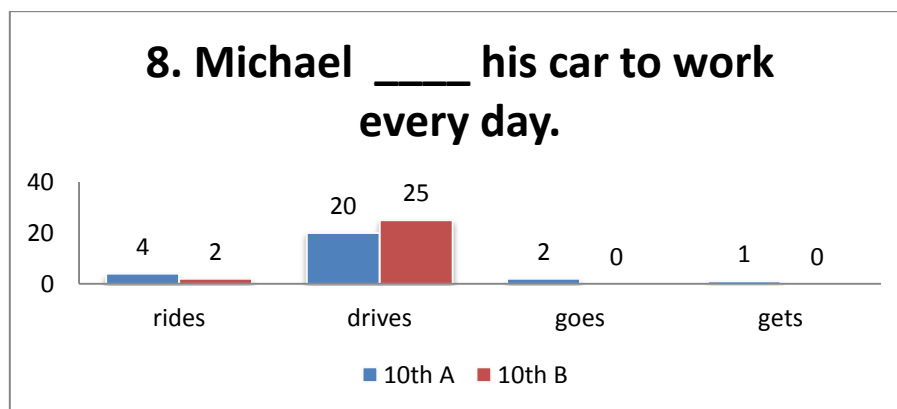
The graphic 10 indicates the results of the post-test of 10<sup>th</sup>A students. This group is the one who did not work with corpora. The results show that, from a total of 23 questions taken to 28 students, the 47% of the questions were answered correctly. The 52% of the answers were incorrect and the 1% of the questions was not answered. Now, if the results of this test are compared with the pre-test of this class a notorious improvement can be perceived since in the pre-test the average percentage of incorrect answers were 60%, the correct answers were only 38% and the not answered questions were 2%. Later, both results will be compared.



11. 10thB performance in post-test

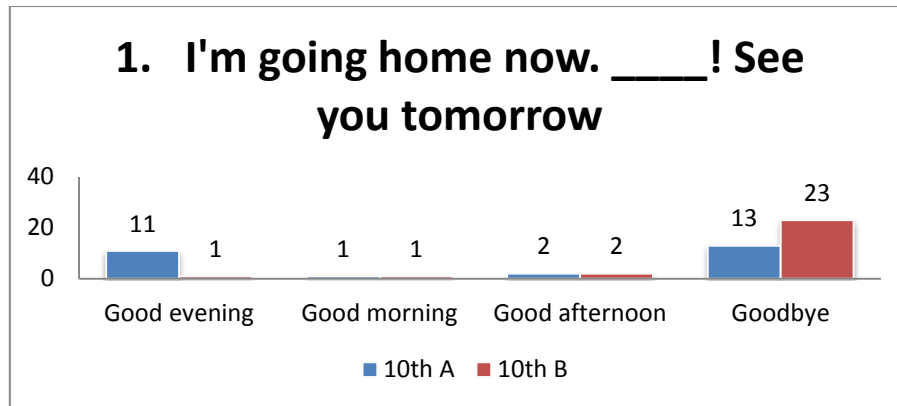
Graphic 11 presents the average percentage of the post-test of 10<sup>th</sup>B students. This is the group in which the experiment with the corpus was made. The graphic shows a notorious improvement. There is an average of 56% of correct answers which contrast with the pre-test in which the average of correct answers was only 40%. The graphic also shows that that in the post-test students had an average of 42% of incorrect answers and 2% of not answer question whereas in the pre-test they had a 55% and 5% respectively.

The questions used in this test were the same questions of the pre-test but some of them were changed. The following graphics will display the results of some of the question in order to analyze them. First of, there are the questions in which the majority of students chose the correct answer.



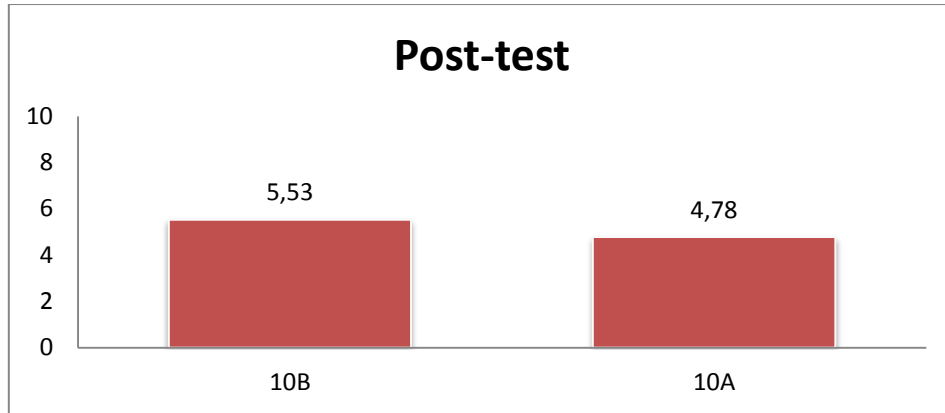
12. Students answer to question 8, post-test.

The graphic above shows the results of question 8 of the post- test. Here, it indicates that both 10<sup>th</sup> A and B chose *drives* as the answer which is correct. This is one of the question in which the majority of students (50% or more) in both groups answered correctly. There were similar results with questions such as 2, 3, 8, 9, 10, 12, 15 and 17. However there were also question in which only in one class the majority of students chose the correct answer but not in the other one.



13. Students answer to question 1, post-test.

Graphic above shows the results of question number one in the post-test. It can be seen that in 10<sup>th</sup> B the majority of students picked the right question but in 10<sup>th</sup> A there was some confusion between *good evening* and *goodnight*. It is worth to mention that in the pre-test there was a similar question to this which was “*I'm going to go to bed now. \_\_\_\_! See you in the morning*” and the correct answer was *goodnight*. However, there were some changes in the post-test therefor the question was change to the one of the graphic above. There were other questions in which only one group have better results that the other such as number four “*It's cold outside. Today you should take a \_\_\_\_ with you*”. The items were *belt, sock, scarf and glove*. This question was not changed from the pre-test. In the pre-test none of the classes had a good result with this question. However, in the post-test, 10<sup>th</sup> B had a better result with 20 students choosing *scarf*. On the other hand, 10<sup>th</sup> A only had 9 students choosing *scarf*. The same situation occurs with other questions such as 4, 5, 7, 13, 16, and 20. The only exception was with question number 23 “*Jorge is Venezuelan and his wife is \_\_\_\_*”. Only in that question 10<sup>th</sup> A had better results with 21 students choosing the correct answer while in 10<sup>th</sup> B only 12 students chose the correct answer. In the remaining questions both of the classes had week results having less than 50% of students choosing the correct answer.

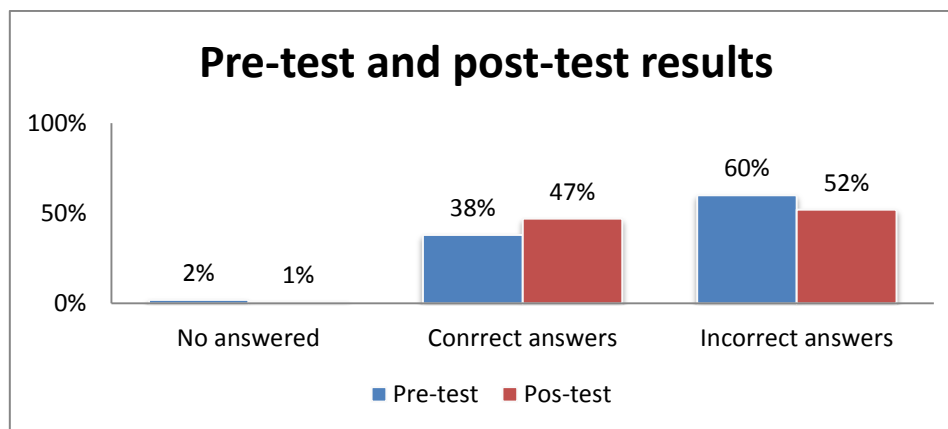


14. Post-test average score from 10<sup>th</sup> A and 10<sup>th</sup> B

Graphic 14 shows the average grades from the post-test. As stated before the tests were scored over 10 points. The graphic shows an improvement in both classes 10<sup>th</sup> A and 10<sup>th</sup> B. 10<sup>th</sup> A scored 4.78 points and 10<sup>th</sup> B scored 5.53 points. In the next section a further analysis of these grades will be made.

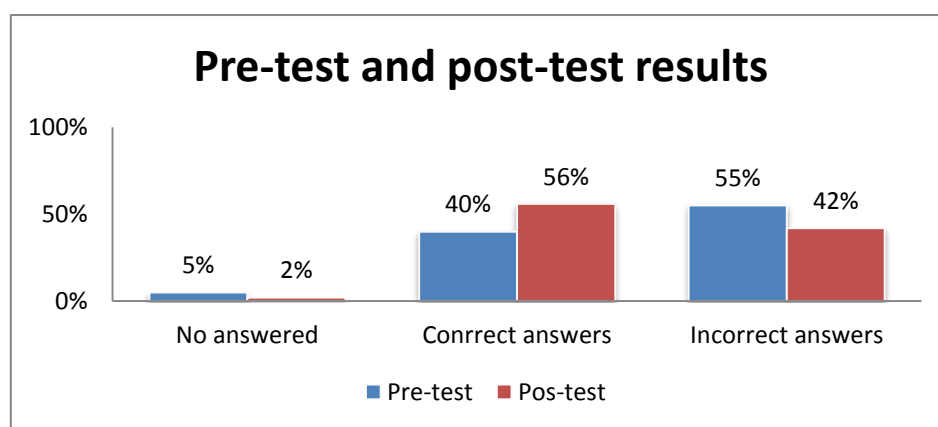
#### 4.4 Contrast between pre- and post-test

The following graphics shows the contrast between pre and post-test of 10<sup>th</sup> A and 10<sup>th</sup>B.



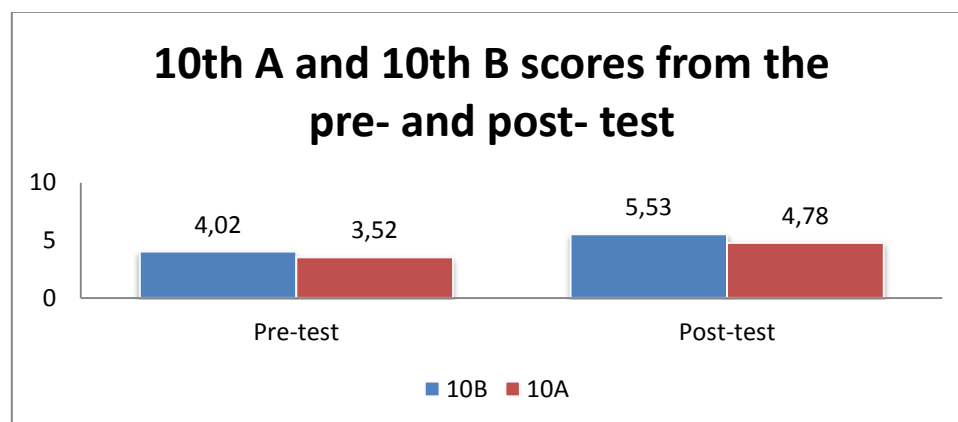
15. 10<sup>th</sup> A pre- and post-test results

The bar chart above shows very clearly the improvement. The not answered questions decreased 1%. The average correct answers increased 9%. And the average of incorrect answers decreased an 8%. Even though this group did not work with corpora they manage to improve their percentages.



16. 10<sup>th</sup>B pre-test and post-test results

As it is evident from graphic 16, 10<sup>th</sup> B had a notorious improvement in their score with an increment of 16% of correct answers. Also it shows that the average percentage of incorrect and not answered questions decrease a 13% and a 3% respectively. It is worth emphasizing that the improvement gap in 10<sup>th</sup> B is wider than 10<sup>th</sup> A as stated in the post-test section above

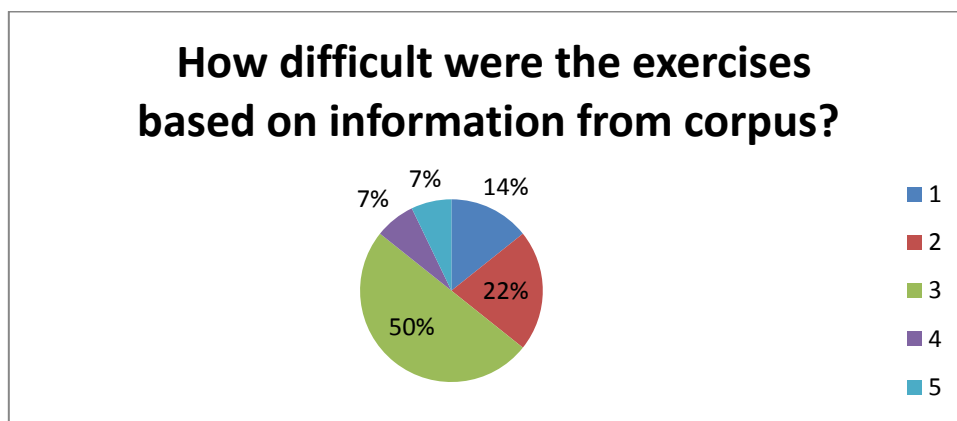


17. Post- and pre-test average score from 10th A and 10th B

Graphic 17 shows the contrast between both groups in the post- and pre-test. In the pre- test 10<sup>th</sup> B scored an average of 4.02 points and 10<sup>th</sup> A scored an average of 3.52 while in the post-test 10<sup>th</sup> B scores an average of 5.53 and 10<sup>th</sup> A scored an average of 4.78. In both tests 10<sup>th</sup> B had a better performance than 10<sup>th</sup> A.

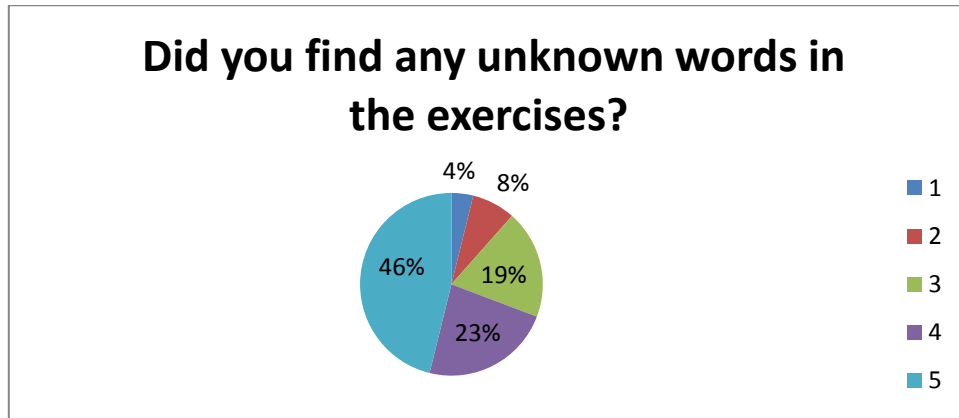
#### 4.5 Survey

This section describes and analyzes the results of a survey aimed at exploring students' thoughts on the process. The survey was based on a rating scale in which 1 meant *very little* and 5 meant *a lot*. (Annex 3)



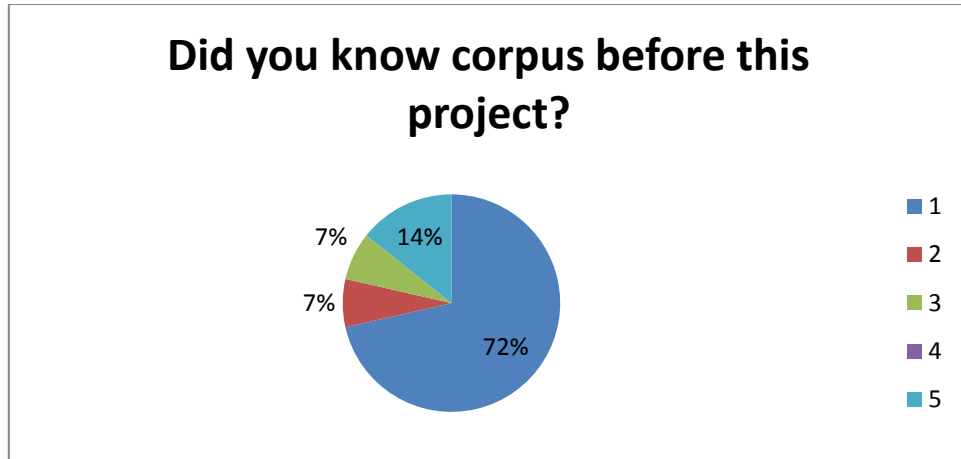
18. Students answer to question 1 of the survey

Graphic 18 shows that, from a total of 28 students, the 50% chose number 5 which means that they thought the exercises presented along the process were a lot difficult. The 22% chose number 2 meaning that they did not also found it difficult. It is worth emphasizing that all the exercises had the same format as the ones from the book that students were using and only the content was taken from the corpus. This agrees with the fact that the exercises in which students had more problems with was the exercises containing comprehensive reading which include raw text from the BNC and it had some complicated vocabulary.



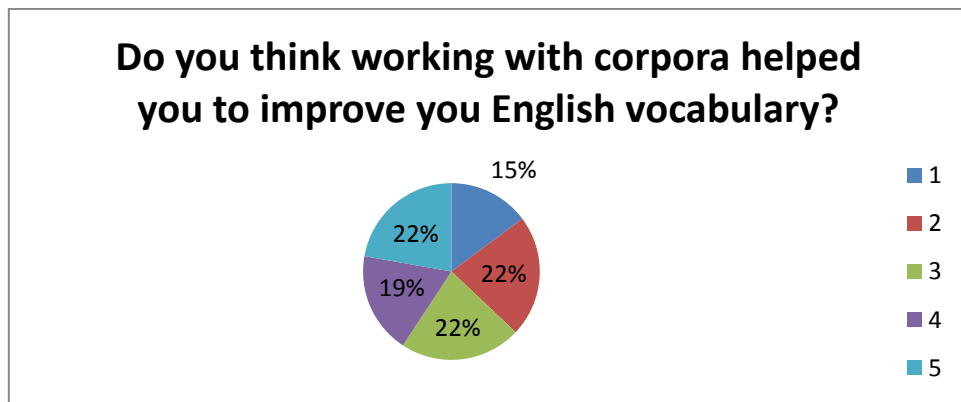
19. Students answer to question 2 of the survey

Graphic 19 indicated students answer to question 2 of the survey in which the 4% chose 1 meaning that they found very little unknown words, 8% chose 2 meaning that they found a few unknown words, the 19% chose 3 meaning that they found some unknown words, 23% chose 4 meaning that they found many unknown words, and the majority 46% chose 5 meaning that they found a lot of unknown words. This is comprehensible since corpora present raw data. As stated in the theoretical background, this could bring complications because students might find it difficult to work with too many unknown vocabulary as it is shown in graphic 19, however it also prepares students to work with more realistic vocabulary and also words apart from the ones used daily in the classroom (see corpus in teaching).



20. Students answer to question 3 of the survey

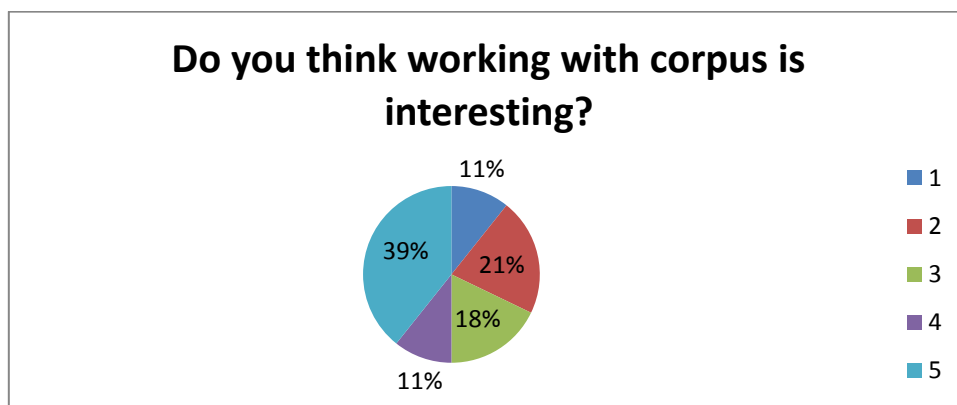
One of the reasons that encouraged this project was the lack of knowledge about corpus. Several authors state that the use of corpus is increasing more and more in the education field however that is not the case in Esmeraldas. From 28 students the 72% did not know anything about corpus before this project.



21. Students answer to question 4 of the survey

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 emphasizing that even though, as indicated by results of the second question of the survey, students did find a lot of unknown words, more than a half of students thought that it was not of help for their vocabulary improvement.



22. Students answer to question 5 of the survey

The final question was aimed at investigating interest of students in the process. The majority, 39% of students, chose number 5 meaning that they think 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 it helps to improve the motivation of the class and to obtain better results,

## 5. CONCLUSIONS

This study does not contend that the use of corpora is the definitive solution for improving vocabulary level but it is 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:

- According to the information obtained from the pre-test and the LOEI, both classes have low vocabulary level. They both score under the minimum required.
- The worksheets showed that students did not only have problems with vocabulary but also with other elements of a language.
- 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 treatment group had a wider increment of scores, suggesting that the use of corpus is beneficial for improving their vocabulary level.
- The results of the survey suggest that working with corpus can provide some benefits such as exposure to a more real language, challenges students, and increases interest. However, whether or not students are interesting in learning English will depend on other elements that are beyond the scope of this research.
- 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.

## **6. RECOMMENDATIONS**

- It is advisable that teachers research more about corpus and to 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 to analyze and practice the new vocabulary. More time is something that could have improved the outcome of this research.
- There are several types of corpora that have a lot of information. It is important that teachers choose carefully and wisely the type of corpora that matches students interests and needs.

## 7. BIBLIOGRAPHY REFERENCE

- 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>
- Beltrán, B. (2013). En los fiscales cuesta aprender inglés. *El Comercio*. Retrieved from  
[http://www.elcomercio.com/sociedad/aprender-ingles-colegios-fiscales-Ecuador\\_0\\_914908516.html](http://www.elcomercio.com/sociedad/aprender-ingles-colegios-fiscales-Ecuador_0_914908516.html)
- Boloña, M. del C. (2010). AN ACTION RESEARCH STUDY OF LEARNERS' PERCEPTION OF USING OF INTERNET ACTIVITIES THROUGH CLASS THE NICENET ICA IN THE EFL CLASSROOM IN ECUADOR, 1(2), 31–45.
- Cáceres, D. (2013). El inglés en la formación académica. *El Tiempo*. Retrieved from  
<http://www.eltiempo.com.ec/noticias-cuenca/128075-el-ingles-en-la-formacion-academica/>
- 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](mailto: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>

- Clark, D. (2012). English - The Language of Global Business? *Forbes*. Retrieved from <http://www.forbes.com/sites/dorieclark/2012/10/26/english-the-language-of-global-business/>
- 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.
- Deficiente la enseñanza de inglés en Ecuador. (2012). Retrieved from [http://www.youtube.com/watch?v=xNfiU\\_tDzTU](http://www.youtube.com/watch?v=xNfiU_tDzTU)
- 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](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.
- McEnery, T., & Xiao, R. (2010). What corpora can offer in language teaching and learning. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 364–380). London & New York: L. Erlbaum Associates.
- Minister of Educación of Ecuador. (2013). Instructivo para la aplicación de la evaluación estudiantil. Retrieved from <http://educacion.gob.ec/wp->

content/uploads/downloads/2013/10/Instructivo\_para\_evaluacion\_estudiantil\_2013.pdf

Minister of Education of Ecuador. (2013). Enseña Inglés: Formando formadores. Retrieved May 21, 2014, from <http://www.educacionsuperior.gob.ec/ensena-ingles-formando-formadores/>

Minister of Education of Ecuador. (2014, Agosto del). National Curriculum Guidance for General Basic Education and Baccalaureate: English as a Foreign Language. Retrieved June 21, 2014, from <http://educacion.gob.ec/wp-content/uploads/downloads/2014/09/01-National-Curriculum-Guidelines-EFL-Agosto-2014.pdf>

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

Optical character recognition. (2015, May 9). In *Wikipedia, the free encyclopedia*. Retrieved from [http://en.wikipedia.org/w/index.php?title=Optical\\_character\\_recognition&oldid=661517812](http://en.wikipedia.org/w/index.php?title=Optical_character_recognition&oldid=661517812)

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

Ramirez, M. (2004). English teachers as materials developers. Retrieved June 26, 2014, from <http://www.redalyc.org/resumen.oa?id=44740214>

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>
- Uribe, D., Gutiérrez, J., & Fernández, D. M. (2008). Las actitudes del alumnado hacia el aprendizaje del inglés como idioma extranjero: estudio de una muestra en el sur de España. *Porta Linguarum: Revista Internacional de Didáctica de Las Lenguas Extranjeras*, (10), 85–100.
- Wilson, J. (2013). *Postcards 1A* (2nd ed.). Ecuador: Minister of Education of Ecuador. Retrieved from [http://75.82.149.74:10003/biblioteca/Textos%20Escolares/Ecuador/Ingles/English\\_Book\\_1-Student.pdf](http://75.82.149.74:10003/biblioteca/Textos%20Escolares/Ecuador/Ingles/English_Book_1-Student.pdf)
- Zaiontz, C. (n.d.). Performing Real Statistical Analysis Using Excel. Retrieved from <http://www.real-statistics.com/>

## 8. ANNEXES

### 8.1 Annex 1

**Class:**

**Date:**

**High school:**

**Difficulty level:** A2 / Elementary

Post-test to develop a research project about the benefit of Corpora in the level of vocabulary of high school students. Case Study in Don Bosco Educational Institution in Esmeraldas. This test contains 23 questions which focus on vocabulary and language used at KET Level (A2).

*Choose the correct answer*

1. I'm going to go to bed now. \_\_\_\_\_ ! See you in the morning.
  - Good evening
  - Good morning
  - Good night
  - Goodbye
  
2. Ahmet comes from Egypt so his first language is \_\_\_\_\_
  - Arab
  - Arabian
  - Arabia
  - Arabic
3. You need \_\_\_\_\_ to make an omelette.
  - fish
  - potatoes
  - eggs
  - apples
4. It's cold outside. Today you should take a \_\_\_\_\_ with you.
  - belt
  - sock
  - scarf
  - glove
5. I'm \_\_\_\_\_ I'm late. My car broke down.
  - wrong
  - sorry

- bad
- unhappy

6. My video recorder isn't \_\_\_\_\_ could you repair it for me?

- making
- going
- doing
- working

7. I'm very tired. I'm going to go to \_\_\_\_\_ .

- bed
- bath
- shower
- chair

8. Joe \_\_\_\_\_ his bicycle to work every day.

- rides
- drives
- goes
- gets

9. Rome is the \_\_\_\_\_ city of Italy.

- big
- lead
- capital
- head

10. Could you \_\_\_\_\_ In this form for me, please?

- make
- note
- wrote
- fill

11. Come to my desk and \_\_\_\_\_ me your book, please.

- take
- get
- bring
- carry

12. Can you \_\_\_\_\_ Spanish?

talk

tell

say

speak

13. Could you \_\_\_\_\_ the situation to me again?

explain

tell

talk

excuse

14. Jack is not well. I think we should call an \_\_\_\_\_ .

illness

ambulance

accident

engine

15. My sister earns 200 \_\_\_\_\_ a week.

money

cash

dollars

notes

16. How much do these tomatoes \_\_\_\_\_?

price

cost

charge

pay

17. Bill loves repairing cars and he wants to train to be a \_\_\_\_\_

mechanic

doctor

painter

waiter

18. I'd like to work \_\_\_\_\_ and learn about a different country.

abroad

strange

- foreign
- another

19. My mother is a very good \_\_\_\_\_ .

- cooker
- cooking
- cook
- cookery

20. Are you ready to \_\_\_\_\_ your meal now, sir?

- ask for
- order
- demand
- explain

21. We're going back home the day after \_\_\_\_\_ .

- tomorrow
- yesterday
- tonight
- today

22. Did you enjoy the \_\_\_\_\_ from London to Tokyo?

- fly
- flying
- flight
- flew

23. Jorge is Venezuelan and his wife is \_\_\_\_\_ .

- Sweden
- Swedish
- Swede
- Swedes

**Name:**

**Class:**

**Date:**

**High school:**

**Difficulty level:** A2 / Elementary

Post-test to develop a research project about the benefit of Corpora in the level of vocabulary of high school students. Case Study in Don Bosco Educational Institution in Esmeraldas.

This test contains 23 questions which focus on vocabulary and language used at KET Level (A2).

*Choose the correct answer*

1. I'm going home now. \_\_\_\_\_ ! See you tomorrow.

- Good evening
- Good morning
- Good afternoon
- Goodbye

2. Lisa comes from Argentina so his first language is \_\_\_\_\_

- Argentinian
- Argentina
- Spain
- Spanish

3. You must have a \_\_\_\_\_ in a birthday party.

- cake
- potatoes
- biscuit
- apple

4. It's cold outside. Today you should take a \_\_\_\_\_ with you.

- belt
- sock
- scarf
- glove

5. I'm \_\_\_\_\_ I'm late. My car broke down.

- wrong
- sorry
- bad
- unhappy

6. My video recorder isn't \_\_\_\_\_ could you repair it for me?

- making
- going
- doing
- working

7. I'm very busy. I have to do the \_\_\_\_\_ .

- bed
- homework
- shower
- classroom

8. Michael \_\_\_\_\_ his car to work every day.

- rides
- drives
- goes
- gets

9. London is the \_\_\_\_\_ city of England.

- main
- lead
- capital
- head

10. Could you \_\_\_\_\_ in this paper, please?

- make
- note
- wrote
- write

11. Come to my desk and \_\_\_\_\_ me the newspaper, please.

- take
- get
- bring
- carry

12. Can you \_\_\_\_\_ Spanish?

- talk
- tell
- say
- speak

13. Could you \_\_\_\_\_ the homework to me again?

- explain
- tell
- talk
- excuse

14. Jack is not well. I think we should call an \_\_\_\_\_ .

- illness
- ambulance
- accident
- engine

15. My sister earns 200 \_\_\_\_\_ a week.

- money
- cash
- dollars
- notes

16. How much do these tomatoes \_\_\_\_\_?

- price
- cost
- charge
- pay

17. Bill loves repairing cars and he wants to become a \_\_\_\_\_

- mechanic
- doctor
- painter
- waiter

18. I'd like to live \_\_\_\_\_ and learn about a different country.

- abroad
  - strange
  - foreign
  - another
19. My mother is a very good \_\_\_\_\_ .
- cooker
  - cooking
  - cook
  - cookery
20. Are you ready to \_\_\_\_\_ your meal now, sir?
- ask for
  - order
  - demand
  - explain
21. We're going back home the day after \_\_\_\_\_ .
- tomorrow
  - yesterday
  - tonight
  - today
22. Did you enjoy the \_\_\_\_\_ from London to Tokyo?
- fly
  - ride
  - flight
  - flew
23. Jorge is Venezuelan and his wife is \_\_\_\_\_ .
- Sweden
  - Swedish
  - Swede
  - Swedes

---

Student's signature

## 8.2 Annex 2

NAME:  
CLASS:  
DATE:  
SCHOOL:

### 1. CIRCLE THE CORRECT ANSWER. *Have or has.*

You don't **have /has** to pay for parking or petrol

There **have /has** to be a separation between Church and State

A decision **have /has** to be made

I'll **have /has** to get out and look for it, he said.

She **have /has** to take off my shoes

They **have /has** to get in touch with you

He would **have /has** to say it very clearly

### 2. FILL IN THE GAPS WITH “HAVE TO” OR “HAS TO”

The company ..... get rid of some of the staff

I ..... get up at 5.30 tomorrow

Unfortunately they ..... to get home

Do I ..... take her home?’

You ..... say yes

He ..... have a home in London because I live and work here.

I ..... have something brand new.

Life ..... go on.

‘I .....go home,’ said Lucy.

If we break anything, we ..... pay for it



**Name:**  
**Class:**  
**Date:**

**Read the following text. Complete the table.**

Rose is 23 years old. She lives in England and her friends call her Mona.

She usually eats Ben's food (a dry, complete diet, twice a day) because it suits her digestion and she likes it. Price is a minor concern.

She first saw Dogs Today on the newsstand reads it from cover to cover and then keeps it. Her partner reads the magazine, too, and sometimes she lends a copy to a friend.

Her favourite article is Peter Neville's True Case Histories. She likes crosswords' games.

When she has a moment she probably reads Bella, Best or Take a Break, and regularly buys the Daily Mail. She sometimes goes to agricultural shows and exemption dog shows.

<b>Personal Information</b>	
Name:	Age:
Nickname:	Country:
<b>Usually activities</b>	
<ul style="list-style-type: none"><li>▪ She eats...</li><li>▪ She reads...</li><li>▪ Sometimes she also reads...</li><li>▪ She buys</li></ul>	
<b>Likes and Dislikes</b>	
She likes...	She doesn't like...

**Name:**  
**Class:**  
**Date:**

➤ **Answer the questions with *can* or *can't***

1. Can you truthfully become a child again?  
Yes, you can / No, you can't
2. Can she remember anything of that audition?  
\_\_\_\_\_ (Negative)
3. Can he remember what you did?  
\_\_\_\_\_
4. Can you hang on here for a few moments?  
\_\_\_\_\_ (Negative)
5. Can they give me one more day to come up with something?  
\_\_\_\_\_
6. Can we give him another couple of hours?  
\_\_\_\_\_
7. Please can you advise me?  
\_\_\_\_\_ (Negative)
8. Can I see it?  
\_\_\_\_\_
9. Can we discuss this tomorrow?  
\_\_\_\_\_ (Negative)
10. Can I do that?  
\_\_\_\_\_
11. Can we trust the Germans?  
\_\_\_\_\_ (Negative)

### 8.3 Annex 3

**Questionnaire concerning students' thoughts on the research project "Do Corpora Benefit the Level of Vocabulary of High School Students? Case Study in Don Bosco Educational Institution in Esmeraldas"**

**Name:**

**Class:**

**Date:**

Answer the following questions by choosing the numbers from 1 (very little!) to 5 (a lot!)

<b>Indicators</b>	<b>Scale</b>
How difficult were the exercises based on information from corpus?	1—2—3—4—5
Did you find any unknown words in the exercises?	1—2—3—4—5
Do you think working with corpora helped you to improve your English vocabulary?	1—2—3—4—5
Did you know corpus before this project?	1—2—3—4—5
Do you think working with corpus is interesting?	1—2—3—4—5

\_\_\_\_\_  
Student's signature

## 8.4 Annex 4

### Process schedule

Date	Activity	Topic	Participants	Exercises
June 30th, 2015	Pre-test		10th B, 10th A	Multiple choice questions
July 9th	Activity 1	Have to/has to	10th B	Choose the correct answer, fill in the gabs
July 14th	Activity 2	Review: have to/has to	10th B	Correction of the exercise
July 16th	Activity 3	Stephen Hawkings	10th B	Comprehensive Reading, Answer the questions, true or false.
July 20th	Activity 4	Review: Stephen Hawkings	10th B	Correction of the exercise
July 21st	Activity 5	Personal information	10th B	Comprehensive Reading, complete the table.
July 28th	Activity 6	Review: Personal information	10th B	Correction of the exercise
July 30th	Activity 7	Can/Can't	10th B	Answer the question
August 4 <sup>th</sup>	Activity 8	Review: Can/Can't	10th B	Correction of the exercise
August 4 <sup>th</sup>	Post-test		10th B, 10th A	Multiple choice questions
August 4 <sup>th</sup>	Survey		10th A	

8.5 Annex 5

