



Universidad Internacional de La Rioja  
Facultad de Educación

Master's Degree in Teaching English as a Foreign Language  
**Implementation Of The Corpus Analysis Of  
The Lexis Contained In An English Reading  
Book With The Aim Of Pre-Teaching  
Vocabulary**

MD's presented by:	Gabriela Margarita Villamarin Avila
Type of job:	Intervention proposal
Director:	Olga Fernandez Vicente
Date:	02/10/2021

## Abstract

This research aims to deal with a well-known problem in English language teaching: reading comprehension. The low comprehension level in EFL learners can be provoked for different reasons, but the lack of vocabulary knowledge is one of the most common. For this reason, this study proposes a corpus analysis of a reading book in order to identify the most frequent relevant words that appear in the text using the *Lextutor* program. Once the keywords list is obtained, seven sessions are planned to pre-teach this vocabulary before reading the book. To this purpose, various strategies are applied, such as guessing from context, word formation, antonyms and synonyms and mnemonics, among others. After that, the students have two months to read the storybook by themselves. Then, in order to assess this proposal, a checklist is applied after the first month to verify the students' level of comprehension during the reading process. Finally, at the end of the reading, three task-based assessments are posed to test if pre-teaching the vocabulary had an impact on the learners' reading comprehension.

**Keywords:** reading comprehension, corpus analysis, vocabulary pre-teaching, Lextutor program.

## Table of contents

1. Introduction.....	7
1.1. Justification .....	8
1.2. Brief analysis of the state of the art .....	9
1.3. Objectives of the study .....	15
2. Literature review .....	16
2.1. Corpus Linguistics.....	16
2.1.1. Compleat Lexical tutor.....	18
2.2. Reading skills.....	23
2.2.1. Reading comprehension .....	24
2.2.2. Project Gutenberg .....	25
2.3. Vocabulary .....	26
2.4. Pre- teaching Vocabulary .....	28
2.4.1. Vocabulary strategies .....	29
3. Intervention proposal.....	32
3.1. Aims of the proposal.....	33
3.1.1. General objective.....	33
3.2. Educational context and Target group .....	33
3.3. Timing.....	34
3.4. Methodology of the proposal .....	36
3.5. Sessions and activities.....	37
3.5.1. “Session 1” Survey .....	37
3.5.2. “Session 2” Corpus based analysis .....	38
3.5.3. “Session 3” Preparation of the vocabulary classes .....	41

Implementation Of The Corpus Analysis Of The Lexis Contained In An English Reading Book With The Aim Of Pre-Teaching Vocabulary

3.5.4.	“Session 4” Vocabulary class N. 1.....	42
3.5.5.	“Session 5” Vocabulary class N. 2.....	45
3.5.6.	“Session 6” Vocabulary class N. 3.....	47
3.5.7.	“Session 7” Vocabulary class N. 4.....	48
3.5.8.	“Session 8” Vocabulary class N. 5.....	51
3.5.9.	“Session 9” Vocabulary class N. 6.....	52
3.5.10.	“Session 10” Vocabulary class N. 7 .....	55
3.6.	Assessment .....	56
3.6.1.	Learning assessment.....	56
3.6.2.	Assessment of the proposal .....	59
4.	Discussion.....	63
5.	Conclusions.....	64
6.	Limitations and further research .....	65
	References .....	67
	Annexes .....	71

## List of figures

<i>Figure 1.</i> Comparison of the number of words between included in storybooks and textbooks .....	10
<i>Figure 2.</i> Total number of lexical frequency.....	11
<i>Figure 3.</i> Meaning performance by the two groups .....	12
<i>Figure 4.</i> Collocation performance by the two groups .....	12
<i>Figure 5.</i> Words required to cover 95% of the proficiency exams .....	13
<i>Figure 6.</i> Descriptive statistics for week by grade level.....	14
<i>Figure 7.</i> Results of the books' survey.....	37
<i>Figure 8.</i> Result of the keywords tool's analysis .....	39
<i>Figure 9.</i> Comparison to verify the frequent words.....	40
<i>Figure 10.</i> Example of first session glossary .....	42
<i>Figure 11.</i> Multi concordance exercise.....	44
<i>Figure 12.</i> Crossword vocabulary class 4.....	50
<i>Figure 13.</i> ID word exercise class N. 5 .....	52
<i>Figure 14.</i> Spinner wheel class N. 6 .....	54
<i>Figure 15.</i> Flashcard example class N.6.....	54

## List of tables

Table 1. Proposal timetable .....	35
Table 2. Vocabulary class N. 1 .....	43
Table 3. Vocabulary class N. 2 .....	45
Table 4. Vocabulary class N. 3 .....	47
Table 5. Vocabulary class N. 4 .....	49
Table 6. Vocabulary class N. 5 .....	51
Table 7. Vocabulary class N. 6 .....	53
Table 8. Vocabulary class N. 7 .....	55
Table 9. Learning diary analytical rubric.....	58
Table 10. Letter analytical rubric.....	60
Table 11. Illustration analytical rubric. ....	61
Table 12. Selecting fragments of songs analytical rubric. ....	62

## 1. Introduction

This research aims to encourage students to read of their own volition in English. As this activity could be complicated for students because of their lack of vocabulary, the teacher can carry out a corpus-based analysis of the lexis found in the chosen book to identify recurrent relevant words and pre-teach this vocabulary. This method intends to help students get a better general comprehension of what they are reading and increase their grasp of vocabulary.

The target for this research is a group of young adult learners from 18 to 25 years old with a B1 current level that study at *Universidad de las Fuerzas Armadas ESPE* located in Latacunga city, Ecuador. The coursebook used at this university includes a reading book; however, this is too short, and it doesn't contain the relevant vocabulary required at their level. For this reason, a range of five books chosen previously from the Gutenberg Project web page will be presented to the students; they will choose one book for the whole class using a survey provided by the teacher.

On the other hand, in our study we will employ an online program called *Lextutor*, designed by Tom Cobb. *Lextutor* is a web page that contains tools for linguistic analysis and the learning of a foreign language (English and French). Among other things, it has a function called *Keywords* that identifies the most frequent and relevant words in a text. Cobb (1998) claims that frequency-based word banks lead to an increase in the learners' vocabulary by highlighting which the most relevant words to learn are. Indeed, it is a well-known fact that any language tends to, so to say, recycle words that could appear more than once, depending on the context. For this reason, pre-teaching this vocabulary could be useful to get a better comprehension of the text.

After this process, the most relevant words are selected. With this information, a lesson plan for teaching this vocabulary in the first stage (pre-reading) will be created using a set of teaching vocabulary strategies such as word formation, synonyms and antonyms, suffixes and prefixes, spelling, mnemonics, guessing from context, and word parts. These strategies intend to help students to retain this information for a long time. Furthermore, the use of ICT tools, such as the aforementioned *Lextutor* program, *Kahoot*, and *Educaplay* apps, will also be required. This stage will allow students to have a better comprehension of the whole story

during the reading process. The students will develop the second stage (while- reading) because they will read the book on their own outside the class; as they were allowed to select the book, we expect their interest in reading to be slightly higher. Once half of the reading has been carried out, the students will answer a checklist of self-report lists of reading difficulties to check if the system is working.

In the last stage (post-reading), the assessment will be established to identify if they understood the whole story portrayed in the book. For this aim, the teacher will make three assessment proposals: 1. To prepare a letter to the writer explaining why they liked or disliked the book, asking questions about the plot, and including the things they would have changed. 2. To draw an illustration for each chapter or scene in the story with a short description of each of them. 3. To choose pieces of music that might portray the main events or the feelings of the characters in the book explaining their choices.

### 1.1. Justification

It is well known that the first taught skills when teaching a second or foreign language are receptive: speaking and reading. The aim is that students get familiarized with the new language in its oral and written way. It is imperative for the input provided to be as natural and understandable as possible and slightly higher than the learners' level. For this reason, it is necessary for the students to be exposed to original materials such as English books, magazines, or articles to obtain valuable knowledge and real information (Mousavian & Siahpoosh, 2018). Therefore, one effective way to provide original and valuable written input of language is through the use of storybooks that are appealing to the learners because of the interesting issues they deal with.

On the other hand, encouraging students to read books written in English is vital if we want them to learn the language. Through reading, they can absorb the structure of the language naturally. As a matter of fact, some research has proved that EFL learners who read are more likely to acquire the English skills better because, besides the reading skills, they also expand their English knowledge and abilities (Floris & Divina, 2015). If students are more exposed to reading materials, they are considered to be more likely to make considerable progress.

Furthermore, comprehending what the writer wants to transmit through the text is one of the most important goals at the moment of reading. According to Tighe & Schattschneider' 2016



work, *Examining the relationships of component reading skills to reading comprehension in struggling adult readers*, “reading comprehension is a complex skill that draws on a multitude of higher and lower order cognitive processes” (p. 2). Unfortunately, this is not always easy to achieve due to different factors; one of them is the lack of vocabulary. In order to solve this problem, pre-teaching vocabulary is an appropriate strategy for developing reading skills. This technique allows for the creation of a connection between vocabulary and comprehension instruction. Instead of decoding the meaning of unknown words, the students' brain focuses on comprehending the text and giving sense to the plot. When students reach a good grasp of vocabulary, they obtain the capacity to define the meaning and contextualize the situation while reading automatically (Mousavian & Siahpoosh, 2018).

However, identifying the specific vocabulary to be taught so as to help them in the reading process could be tough if there is not a scheme to follow. For this reason, carrying out a corpus-based analysis of the lexis contained in the book from the very beginning could be useful. Indeed, corpus-based approaches have been used as research tools in different linguistics fields as well as to develop methods to teach other languages (Biber & Reppen, 2002). With this analysis, the teacher can identify the most relevant and repeated words that could be difficult to understand at once so as to be able to tend to their students' vocabulary needs.

## 1.2. Brief analysis of the state of the art

The implementation of a corpus-based analysis, vocabulary pre-teaching, and the use of storybooks are not new when it comes to improving learners' reading comprehension skills. Different researchers have dealt with these variables and as a result, most of them have shown positive consequences and genuine benefits for teachers and students with the theories they have proposed. Thus, the viability of this proposal is highly supported by the results obtained by similar investigations.

In 2011, Hsieh, Wang, & Lee presented the study *A corpus-based analysis comparing vocabulary input from storybooks and textbooks*, where a comparison was made between the vocabulary input of storybooks and textbooks. They analyzed the corpora of 65 storybooks, and those results were compared with the vocabulary of three well-known textbooks for

elementary in the Taipei area. The results showed that the children obtained more vocabulary input with more headwords and content words like nouns, verb, adverbs and adjectives from the storybooks than from the textbooks.

Furthermore, they found that the children who listen to stories turn into independent readers at younger ages. As Wang & Lee (cited by Hsieh et al., 2011) explained in their previous work, *From listeners to readers with storytelling to bridge the gap*, “EFL children who hear stories become independent readers after only four years, while most students cannot reach this level after six years of instruction” (p. 25). This research supports the intention of using a storybook instead of a textbook in this proposal because the source of relevant vocabulary is precious, especially for a B1 level group that is interested in increasing their grasp of vocabulary. The numerical results of the comparison between the number of words included in storybooks and textbooks are the following.

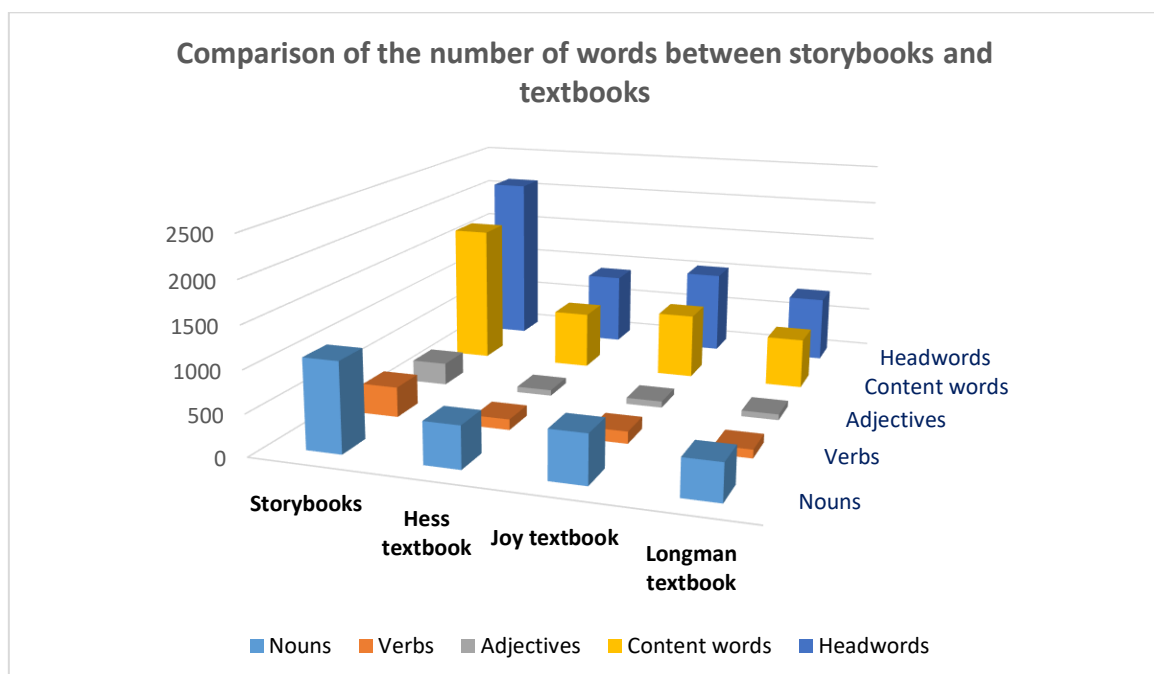


Figure 1. Comparison of the number of words between included in storybooks and textbooks (adapted from Hsieh, Wang, & Lee, 2011, p. 30)

In turn, Abduh & Rosmaladewi (2017) developed another piece of research related to corpora analysis called *Taking the Lextutor online tool to examine students' vocabulary levels in business English students* in order to determine the frequency of words in a text. To this

purpose, thirteen essays of college students were examined to identify the academic level of the vocabulary employed. On this occasion, the online program *Lextutor* was used as an analysis tool for classifying the words into categories. After a deep analysis of the data collected, they found that in academic writing, the students employed basic words instead of academic lexis and terminologies. The final results of the total number of lexical frequency are shown below.

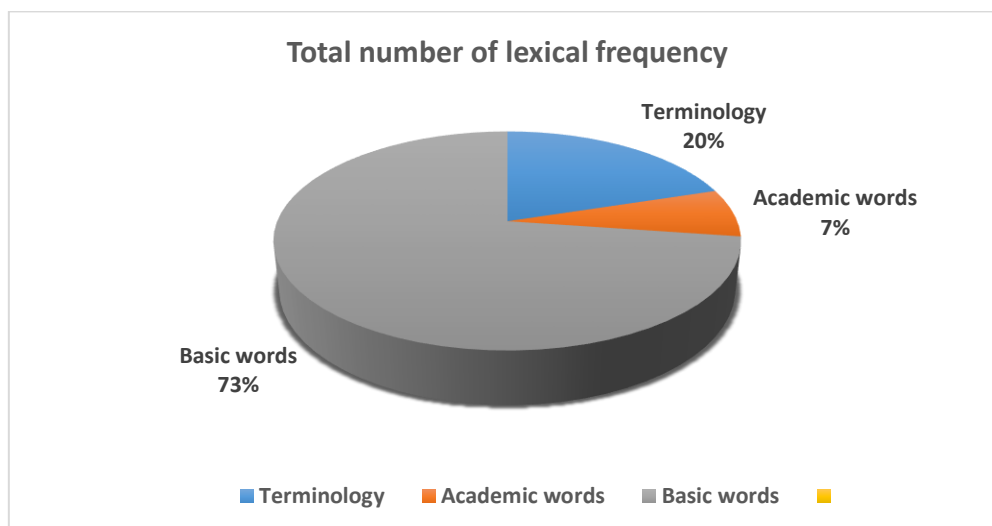


Figure 2. Total number of lexical frequency (adapted from Abduh & Rosmaladewi, 2017, p. 3)

One of the conclusions showed that “participants in this research will find it difficult to read and understand academic texts because they had mostly mastered a low level of vocabulary” (Abduh & Rosmaladewi, 2017, p. 3). Furthermore, the authors recommended the learners to use online sources like the *Lextutor* web program in order to increase their vocabulary grasp by using exercises and word classification. Taking this into account, this program was chosen to work on this proposal to classify the words to be taught at the beginning of the reading process.

Regarding the vocabulary learning process, in 2015, Al-Mahbashi, Noor, & Amir presented a study called *The Effect of Data-Driven Learning on Receptive Vocabulary Knowledge of Yemeni University Learners*; it was based on how Data Driving Learning (the approach that allows learners to manipulate, explore and learn from the language data) influenced the receptive

vocabulary and collocations knowledge of college students. This research was conducted with 60 students divided into two groups. The first group of 30 people followed the Data Driving Learning method, with activities that included logical, visual-spatial, and interpersonal intelligence and the researcher's guidance. The second group of students learned the same vocabulary working with a dictionary and the designed collocation materials. After the pre-test, post-test and delayed post-test had been applied, Al-Mahbashi et al. (2015) concluded that "the scores of the two groups increased from the pretest to the posttest. Nevertheless, learning vocabulary meaning using DDL method improved significantly more than using the dictionary" (p. 22). The results displayed the fact that the group that had worked with the Data Driving Learning method had obtained better results learning vocabulary and collocations with better information retention because of the rich contexts and constant exposition to the selected vocabulary. The results of the meaning and collocation performance by the two groups are shown below.

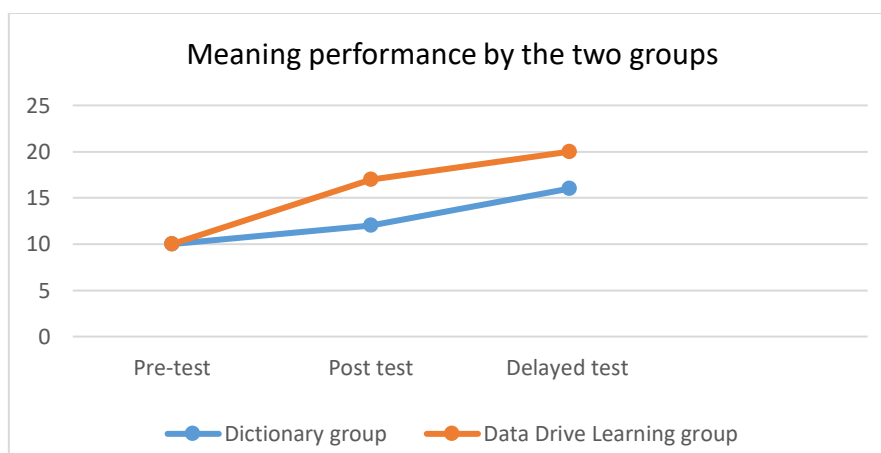


Figure 3. Meaning performance by the two groups. (adapted from Al-Mahbashi, Noor, & Amir, 2015, p. 20)

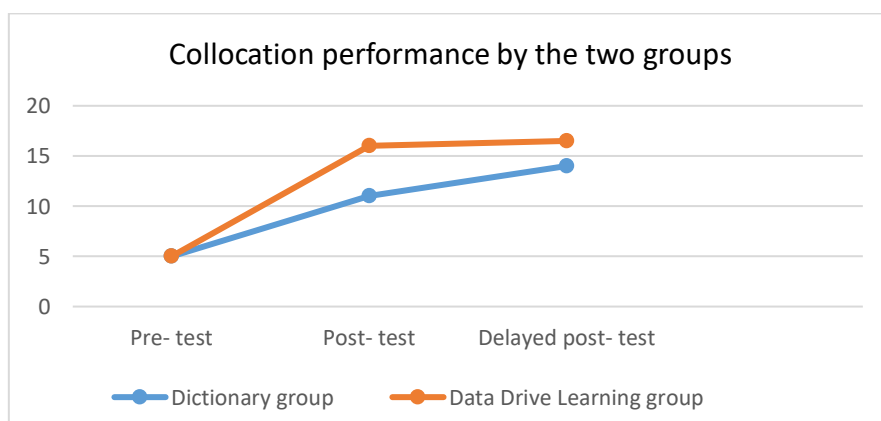


Figure 4. Collocation performance by the two groups (adapted from Al-Mahbashi, et al., 2015, p. 21)

Furthermore, Chujo & Oghigian in 2009 carried out research into corpus analysis related to vocabulary called *How many words do you need to know to understand TOEIC, TOEFL & EIKEN? An examination of text coverage and high-frequency vocabulary*. They investigated the range of vocabulary required for a 95% comprehension of proficiency tests such as TOEIC, TOEFL, and EIKEN. The data were collected from each test, and a computer program skimmed them; with this information, they created the word lists for each exam. Additionally, the lists were contrasted with the Standard Vocabulary List, the British National Corpus, and Nation's 14K word-family list. The results showed that for the TOEIC, students should know an average of 4,000 words and 3,000 word-families. For the TOEFL, 4,500 words 3,500-word families are required. Finally, for the EIKEN, it is necessary to master between 5,500 words and 4,500 word-families. Undoubtedly, this kind of research carried out into corpus-based analysis is really useful because it allows for the identification of specific vocabulary that is useful for many purposes depending on the learners' interests.

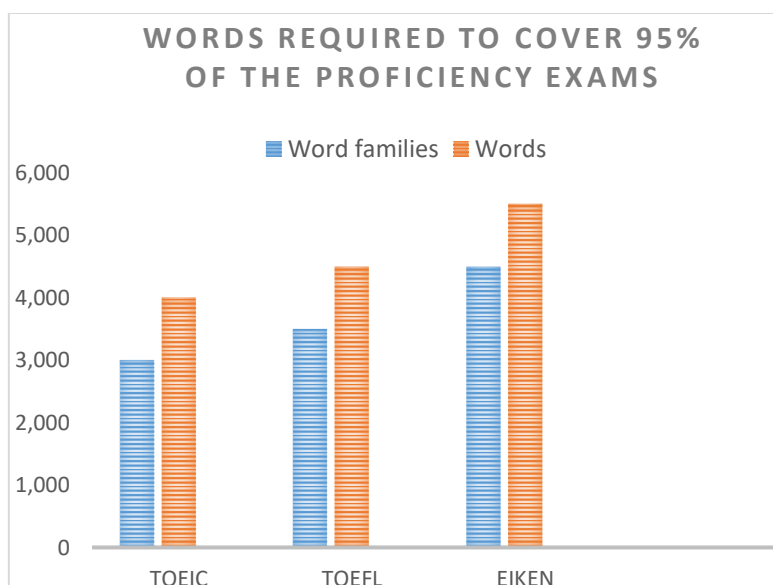


Figure 5. Words required to cover 95% of the proficiency exams. (adapted from Chujo & Oghigian, 2009, p.

142)

Another interesting study developed on the influence of pre-teaching vocabulary to improve comprehension of a narrative text was proposed by Cowell in 2012 called *Pre-teaching vocabulary to improve comprehension of a narrative text*. Three classes of an elementary school located in southeast Alabama participated; they received four days of rich vocabulary

instruction, and, on the fifth day, a multiple-choice post-test was provided. At the end of the month, a post-test was administered to the groups to identify the effect of the pre-teaching strategy and to grade the level and the follow-up test.

In the end, the groups' scores were relevant for the three weeks, but the results were different for each group depending on their year. The second-grade group achieved higher scores in reading comprehension during week one. The fourth-grade group got better scores during the third and follow-up weeks. The sixth grade didn't obtain any significant scores at any moment. During week two, none of them had relevant scores. Finally, Cowell (2012) concluded that “after the implementation of the pre-instruction, students in the treatment groups had greater comprehension on the weekly post-test during the first three weeks compared to the control group” (p. 59). The results are the following:

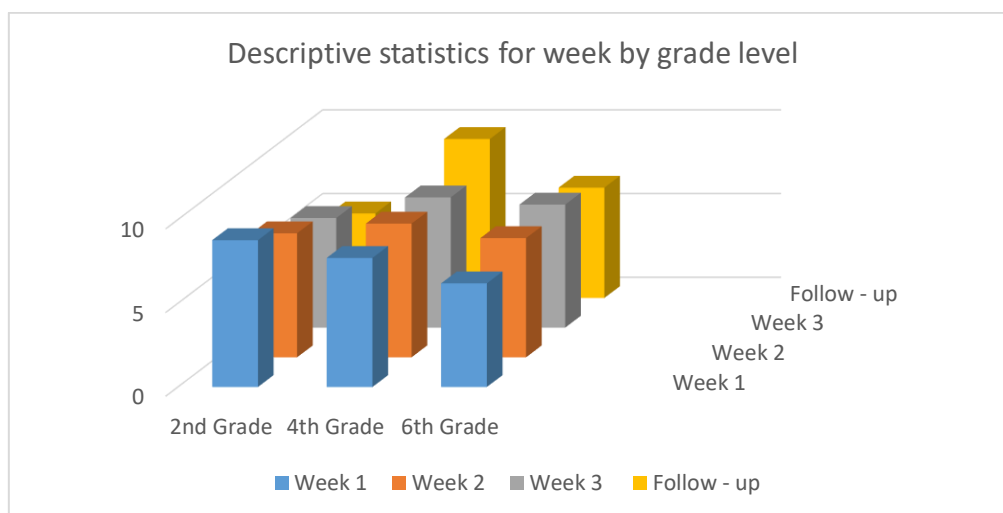


Figure 6. Descriptive statistics for week by grade level (adapted from Cowell, 2012, p. 52)

After checking the research related to the variables presented in this proposal, it is possible to get the following conclusions. Firstly, the use of storybooks is advantageous to obtain useful vocabulary, which contains more headwords and content words such as nouns, adjectives and verbs in contrast to EFL textbooks (Hsieh et al., 2011). For this reason, the use of a storybook in the proposal will be beneficial to provide the participants with the correct input of vocabulary, whenever this vocabulary is presented at the beginning of the reading process and taught beforehand. In this case, their level of comprehension will increase because they

will not need to stop their reading pace to check any unknown words, and they will understand the text in context (Cowell, 2012).

Secondly, the process of collecting data and carrying out a corpus-based analysis has been proved to be useful at the moment of determining the number of words and classifying them (Chujo & Oghigian, 2009). Moreover, working with linguistic data, manipulating and exploring it has shown positive results in the process of acquiring new vocabulary because of the constant exposition that the learners have to the language (Al-Mahbashi et al., 2015). Additionally, the *Lextutor* web program has been recommended as an analysis tool to carry out the corpus-based analysis due to the good results obtained when using it (Abduh & Rosmaladewi, 2017).

In conclusion, this proposal is highly supported by other studies with similar variables that have obtained favorable results. This proposal combines some of those variables, such as storybooks, corpus-based analysis, the *Lextutor* program, and vocabulary pre-teaching. The correct application of all of them in a precise sequence will have positive effects on the learners' reading comprehension that is the problem to be solved inside the classroom. In the end, our proposal will be verified as well as its contribution to the English teaching-learning process.

### 1.3. Objectives of the study

The main objective of this study is to improve the general reading comprehension of a storybook and to solve the lack of vocabulary knowledge in young adult learners. To this purpose, a corpus analysis will be applied in order to identify the most repeated and relevant words and pre-teach them.

The specific objectives of this proposal are the following:

- To investigate the fundamental concepts that support pre-teaching high-frequency vocabulary and the way it influences the improvement of reading comprehension and vocabulary acquisition.
- To apply the *Lextutor* program to analyze the corpus of a storybook and identify the most recurrent words that appear in a storybook. From this list, the most relevant and useful vocabulary will be selected for it to be taught.

- To look for effective teaching vocabulary strategies that ensure the appropriate enduring vocabulary acquisition. It will allow students to have appropriate vocabulary knowledge and, consequently, better reading comprehension.

## 2. Literature review

To carry out this research, it was necessary to review the main concepts that cover the variables described in this intervention proposal. For this reason, in this section, the concepts described are corpus-based analysis, reading skills, vocabulary, and pre-teaching vocabulary. These definitions help to have a deep understanding of the terminology used in this work, and it will be a guide for an easy comprehension of the proposal.

### 2.1. Corpus Linguistics

To understand this methodology, firstly, it is necessary to define what a corpus is. According to Al-Gamal & Ali (2019) "A corpus is a large collection of written or spoken texts that can be developed manually or electronically and be representative of the authentic language data" (p1). This compilation of texts is mostly employed in language research. In applied linguistics, the corpus has been part of the development of many educational resources, such as dictionaries, glossaries, grammar books, translation apps, etc. Thus, the corpus must have a defined size and content, and it must go straight to the research hypothesis. According to Nordquist (2019), some remarkable English Language Corpora are the following:

- [The British National Corpus \(BNC\)](#)

It is a collection of 100 million British English words collected from a wide range of resources. The corpus was collected from 1991 to 1994; however, it has been reviewed and the last edition dates back to 2007. The central part is written data, but some minor spoken data can be found as well. The writing data contains newspapers, journals, academic books, letters, essays, etc. On the other hand, the spoken part includes formal and informal conversation, transcriptions and spoken language from government meetings, business, and radio shows. All this information is available on its [webpage](#) (The British National Corpus, 2020).



- [The Open American National Corpus \(OANC\).](#)

It is an extensive collection of Million current American English words carried out since 1990, comparable to the British National Corpus (BNC). This project arose due to the differences found between British and American English. It contains corpora of all genres and transcriptions of oral language data. These genres include new types of language data that have arisen recently, for example, emails, tweets, web pages, blogs, and lyrics. This information is free and open to the public from its [website](#) (Open American National Corpus, 2020).

- [The Corpus of Contemporary American English \(COCA\)](#)

It is the largest collection of corpora with more than one billion words taken from texts dating from 1990 to 2019. It contains eight different genres with a fair balance of academic texts, newspapers, well-known magazines, spoken productions, fiction, blogs, TV and movie subtitles, as well as web pages. It could be considered the most widely used English corpus because 130,000 people use this [website](#) each month for different purposes. These corpora have been used for studies and language and linguistics courses. Finally, in order to obtain free access, you need to create an account and select the category that will be used (English Corpora, 2020).

Considering these concepts, it is time to study the corpus linguistics term deeply. Firstly, it is the field that investigates the linguistic phenomena through the use of the corpora. In 2019, Nordquist defined the corpus linguistics as “the study of language based on large collections of "real life" language use stored in corpora computerized databases created for linguistic research” (p 1). It focuses on the methods or procedures that are developed for the study and exploration of the language. The corpus linguistics deals with machine-readable texts or corpus in order to search and manipulate the data, to be more specific, two common forms of analysis are concordances and frequency data. Concordances (qualitative) refer to the way the words are interpreted in context, for example, they can be classified as a part of the speech or classified by their level of complexity. On the other hand, the frequency data (quantitative) deals with the number of words that appear in a corpus and defines how many times each word is repeated, creating word frequency lists. Applying these procedures, it is possible to carry out a corpus-based approach to different fields of linguistics (McEnery & Hardie, 2011).

Nowadays, the corpus analysis contributes more and more to a second or foreign language learning process. These corpora could be used in the development of study materials for all language skills such as reading, writing, listening, speaking, and vocabulary. (Al-Gamal & Ali, 2019). Furthermore, it improves the learners' linguistics skills in synonyms, antonyms, homonyms, and polysemy. The corpus linguistics methodology applies a bottom-up study of the language, which means a process of decoding from the most basic units of the text like letters, morphemes, and graphemes to sentences and texts, in an attempt to understand the language. As a result, it does not require significant expertise to work with; for this reason, it could be applied to students of all levels (Tognini-Bonelli, 2001).

As we have stated, the corpus linguistics is used to identify patterns related to grammatical and lexical features. However, it can also provide relevant evidence on register, frequency, and the way in which a language is used (Bennett, 2010). According to McCarthy (2004), by searching and analyzing corpora, it is possible to find answers to the following questions:

What are the most frequent words and phrases in English?

What are the differences between spoken and written English?

What do people use most frequently?

How many words must a learner know to participate in everyday conversation? (pp 1-2)

### **2.1.1. [Compleat Lexical tutor](#)**

If we are to apply the corpus linguistics in the classroom, we must bear in mind that a suitable corpus is vital but so is it to have a concordancing program. Bennett (2010) claims that "Concordancing programs are computer software used to access and sort data from the corpus" (p.16). When choosing one of these programs, it is necessary to bear in mind which data the user wants to get from the corpus as many of them are available depending on the user's requirements.

As we have already mentioned, Compleat Lexical Tutor (*lxtutor.ca*) is an online program for data-driven language learning created by Bafia Lufer and Paul Nation in 1995, with an off-line version called "Range" designed by Thomas Michael Cobb in 2002. The primary purpose of the program is to provide linguistic tools to help teachers and students in the language

learning process. Even though its interface is not really user- friendly, it is possible to use it to expand the vocabulary skills in the English and French languages, and the user does not need outstanding abilities to work with it.

This website is devoted to offering different interactive learning tools to the users that can be carried out by themselves when researching, teaching, and/or learning. Some of these activities can be carried out by the learners without the teacher's guidance because they include instructions and guidelines. The most useful tools for EFL learning are described below.

(To see the *Lextutor* guide with illustrations, please visit the [e-portfolio](#))

- **Group Lex (Group Lexicon).** It is a collaborative glossary to which the learners can add entries by writing the word, a sentence example, the word class, the definition, and their names. Additionally, it is possible to edit entries and create multiple-choice quizzes with this terminology.
- **The Hypertext builder.** In this tool, it is possible to upload a text and click on any word to get examples of concordances and definitions.
- **Concordances.** This section provides a list of terms or phrases related to or compared directly with the corpus or text vocabulary. It is divided into subcategories (Sevier, 2004).
  - **The Multi-Concordance + Exercise Builder.** This section permits the user to write a set of words, and the program generates concordances (fragments of texts that contain the word) in different modes. The first is plain for grouping the concordances per each word; the second is plain and gaps, which includes a blank for the word, and the third is a quiz where the learners select the correct term to fill the gaps (Sevier, 2004).
  - **The Text Concordance Input.** When the user uploads the text, the program produces a concordance index of the words uploaded and their context in alphabetical order (Sevier, 2004).
- **Ra- Read (Resource- Assisted Reading).** In this section, it is possible to read and listen to a book at the same time. There are ungraded (real English) books such as *Call of the*

*Wind* (in English) and *Boule de Suif* (in French), and 11 graded (by levels) books in English.

- ***The Corpus Correct (Corpus Grammar Correct)***. This section presents statements with grammatical errors. In the correction space section, the learners have to write the statement corrected and click on the check button to verify the answer.
- ***The Research Base***. It shows links to previous researches, papers, and websites based on the program and the "frequency-based vocabulary learning" hypothesis (Sevier, 2004).
- ***Familizer/ Lemmatizer***. In this section, the learners paste or upload the text, and the program produces a complete word-family list or a list of lemmas (headwords with their inflected and reduced (n't) forms (Nation, 1997)) for each term of the corpus.
- ***Vocabprofile***. This section classifies the words (tokens) of a corpus and matches each one to four frequency levels. The first level is the most frequent one hundred word families (K1 words, light blue color). The second one is the following most frequent one hundred words (K2 words, green color). The third one is a list of academic terms (AWL words, yellow color). Finally, the last one refers to the words that are not on the other levels (Off list words, red color).
- ***The Frequency tool***. It allows for the extraction of a high-frequent word-list from a text or corpora. The user can select the type of list between the most frequent, the least frequent, or alphabetical order (Sevier, 2004).
- ***The Text Lex Compare***. In this tool, the user uploads two texts, and the program compares them and determines: the words (tokens) and the family words that each corpus has, and those tokens they share in common.
- ***The Clickers Builder***. In this section, it is possible to produce multiple-choice quizzes with ten items. The questions are not shown on the learners' screen; thus, they have to be displayed on a presentation, paper, or told orally. Additionally, the program offers the average of the answers obtained on a bar chart.
- ***The Flashcards Builder***. It is aimed at the flashcard's creation with the term and its definition. It is possible to create from 3 to 100 cards. The learners have access to the flashcards through a link, for example, <https://lextutor.ca/flashcards/keywords->

[1.html](#). Apart from having access to the meaning of the word, the users can listen to the pronunciation of each word.

- **The Cloze builders.** This section allows for the creation of two types of cloze passages; the first one makes gaps to fill based on the words' number, which means that there is a blank after the 5th or 6th term, or the teacher can select it manually. The second type creates the gaps according to frequency levels such as academic words, Brown concordances, The British National Corpus (BNC), The Corpus Contemporary American English (COCA), etc. Furthermore, it has access to additional resources like dictionaries, audio, or video files, among others (Sevier, 2004).
- **Vocabulary Test.** This section presents a set of vocabulary tests based on the word lists of well-known researches developed by Nation (1990), Meara & Jones (1988), Read (1998), among others. Each of them has different receptive or productive activities, where the learners can check their level and practice the new vocabulary.

There is a special section called **Word games** where the learners can review new vocabulary from well-known corpora such as The British National Corpus (BNC) or The Corpus Contemporary American English (COCA), among others with practical activities described below.

- **Multi-Concordance.** In this tool, the user introduces the text; then the program displays statements of concordances (fragments of texts that contain the word) from the corpus frequency list the user chooses (Brown 1 million, Corpus-AWL, BNC written, and so on).
- **The Frequency Trainer.** This tool is aimed at measuring the learners' assumptions about frequent words. The source is the frequency lists BNC OK or COCA US. The learners have to guess the band of the frequency scheme, in other words, how many times the word appears in written or spoken language.
- **I-D Word Identification Quiz.** In this section, the words can be proposed by the teacher or the program; the tool generates concordances (statements from different sources that contain the word), and the learners have to read the statements offered and select the word that better fills the gap from a group of the letters.

- ***The Dictator***. The user pastes or uploads the text; then, the program produces a dictation spelling or quiz based on this input. Thus, the learners have to listen to the dictation and write down the statements (Sevier, 2004).
- ***Memory Measure for Spoken Language (Speech\_memory)***. The learners practice their memory skills. First, they select the mode (numbers, letters, words, and non-words), and they listen to a random series of spoken language. The task is to click on the items in the same order they were mentioned.
- ***Memory Measure for Written Words (Text\_memory)***. It is another memory game with written language. The learners read some words from a frequency list, later this list disappears, and they have to indicate the terms presented before from a set of words.
- ***Keywords extractor***. The term "keywords" is defined as "the words that are far more frequent in your text, as a proportion of its size, than they are in a reference corpus" (Lextutor.ca, 2020). In other words, the program identifies the most recurrent words of the text, and it compares this list with the [bnc coca fams speechwrite US UK per10mill](#) corpus, which is a collection of 10 million US-UK terms mixed of written and spoken language developed by Paul Nation (1995). Thus, if a word appears in this corpus list and is repeated several times in the text, the program concludes that it plays an important role in our text (Nation, Lufer, & Cobb, 1995).

As we have already mentioned, the Compleat Lexical Tutor program has been chosen for this study because it has been used in other research about corpus-based analysis giving satisfactory results (Berg & Kit, 2019). In fact, it was highly recommended by Abduh & Rosmaladewi in 2017 in their study about the students' vocabulary level in business English students. One of the main reasons is that it was created specifically for language learning. As a result, the different activities for corpus linguistics analysis can be carried out not only by teachers but also by learners. Additionally, it is available on the internet, and it is not necessary to create an account; anyone can have access to it and work with all of its functions.

On the other hand, the *Keywords Extractor* tool is employed in this research because it extracts the most frequent or repeated words of the text and selects the most relevant terms called keywords. As we have previously mentioned, this keyword selection is carried out by

comparing the frequency word list obtained beforehand from the input text with the corpus of 1 million written and spoken English words developed by Nation. The tool generates a new list of keywords with the terms that appeared in the Nation's corpus according to their relevance. In this way, the teacher ensures that the chosen vocabulary will help the learners to understand the text.

## 2.2. Reading skills

According to Santiago, Castillo & Ruiz (cited by Herazo, 2015), "Reading is understanding, and this activity involves superior mental skills such as: predict, infer, analyze, synthesize, among others" (p.1). Furthermore, it is a receptive skill in the written mode, often developed in conjunction with listening and speaking skills, but it could also be worked on in isolation. Apart from that, it has been proved that reading helps expand vocabulary, which could help listening comprehension in later stages. Furthermore, reading skills also contribute to building previous knowledge, establishing hypotheses and verifying them, and drawing inferences to understand what is suggested, to construct meanings (Herazo, 2015).

Another definition of reading was stated by Pineda & Lemus (cited by Iñiguez, 2015), and they claim that "reading is a complex, multidimensional process. It involves various levels of proactive behavior that interact with each other, as the reader tries to discern the meaning of what the author has written" (p.5). What is more, in the meta-cognitive process of reading, some phases or stages do not necessarily occur sequentially. On the contrary, they can do so in a comprehensive, interactive, and interdependent manner. García, Riggs, & Cañizares (cited by Salvador, 2005) present activities to be carried out by the reader before reading:

- Set goals.
- Activate prior knowledge about the topic.
- Determine the structure of the text.
- Make predictions about what they will learn.

During the reading (while reading) the readers will:

- Apply the available strategies.
- Summarize what the learners have read.
- Examine the meaning of words and phrases that they do not fully understand.

- Recognize key vocabulary.
- Infer meanings and concepts.
- Raise questions to be answered by reading the text.
- Notice comprehension problems.
- Evaluate the strategies to be used.

After reading, the readers will:

- Examine the reading activity to understand the content.
- Summarize the key ideas from the text.
- Evaluate and state what they learned from reading.
- Deliberate about what else the learners need to learn about the topic.
- Relate the topic with other information learned.
- Reflect on the strategies that best served the learners.

### **2.2.1. Reading comprehension**

According to McNamara (2012), reading comprehension refers to the ability to go beyond words, understand ideas, and the relationship between them, as transmitted in a text. In other words, understanding a text means efficiently extracting the required information. On the other hand, according to Salvador (2005), text comprehension is considered a constructivist activity because the learners create a mental representation of the text's content, and it is added to their knowledge base. Thus, reading comprehension would then be the competence employed when we need to perceive and understand the meaning contained in texts.

#### **Levels of understanding**

According to Ignacio and Alacbay (cited by Villamarin, 2016), there are three levels within reading comprehension which are: level one or literal, level two or interpretive, and level three or applied. We will describe them below:

- **Level one or literal.** This level involves only the superficial comprehension of the text, including facts and details, memorization, and rote learning. Therefore, the literal level of understanding is the most basic of levels in readers. In fact, they are creating mental connections, but they don't have control over them yet. The most common



comprehension questions used to induce this type of thinking are who, what, when, and where; the answers to these questions are explicitly stated.

- **Level two or interpretive.** At this level, the readers make inferences or read between the lines. Additionally, readers attach new information to previous knowledge. For this reason, they are prepared to make conjectures and understand what the author means by interpreting what he wrote in the paragraph, story, or book. As they are progressing, they will have likely memorized some facts at level one, and now they understand them and discover the writer's intention. The questions asked are open-ended and invite reflection: why? And how? The answers are not expressed directly. Therefore, the readers have to analyze the content they have to read and apply their previous knowledge and conjectures to answer the questions.
- **Level three or applied.** It consists on taking the literal part (what was stated) and then interpreting it (the meaning of what was stated), to finally applying the concepts to other situations (extending the ideas). At this level, the readers reflect on how they can apply the author's message to other real situations. Consequently, the readers develop their critical thinking. The question for this level may include "How ...?" and assumptions or conditionals such as "If these are..." "If I were..." This questioning puts interpretive reading into practice by applying the information gathered after reading.

Ignacio and Alacbay (cited by Villamarin, 2016) claim that a student has understood if he knows how to perform any of these operations:

- Explain the idea in their own words.
- Look for different examples that confirm it.
- Recognize it in different circumstances or when it is expressed in other words.
- See relationships between these ideas and other known ideas or facts.

### 2.2.2. [Project Gutenberg](#)

Project Gutenberg is a website that provides free eBooks in an online library. Michael Hart was the founder of it in 1971, and its mission is "to encourage the creation and distribution of eBooks" (Hart, 2004, p1). They want to offer eBooks to the world in all formats and languages possible. The project works renovating the physical copyrights of expired books into e-books.

Consequently, the texts published before 1923 are available on this site; by the end of 2002, they had already published more than 6,000 complete digital books (Stroube, 2003).

Apart from that, it is not necessary to have a membership or pay a fee to get the e-Books. Furthermore, the project is not supported by any political or financial power. They work with volunteers who contribute with translations to other languages, conversion to other formats, and ideas for the creation and distribution of the e-Books. Another project recently implemented by the same Gutenberg organization is "Distributed Proofreaders," which is a combination of volunteer proofreaders and optical character recognition software. It enables them to increase the volume of e-books stored in their library of this webpage.

There are three main categories of literature in Project Gutenberg: light, heavy, and references. Firstly, the light literature collection is aimed at a general target of any age, from kids to grandparents. Here it is possible to find short stories, well-known novels, and fables, among others. Some examples are *Peter Pan*, *Aesop's Fables*, *Alice in Wonderland*, and so on. On the other hand, heavy literature refers to more sophisticated texts for people looking for more complex arguments, like *Hamlet*, *Paradise Lost*, *Moby Dick*, *the Bible*, etc. Finally, the references section is devoted to educational materials such as dictionaries, thesauruses, almanacs, encyclopedias, and so on (Hart, 2004).

### 2.3. Vocabulary

According to Nation (2009), having vocabulary knowledge is a significant ability to master a language. Therefore, one of the goals of English learners is to acquire as much vocabulary as they can. In 2007, McCrostie claimed that vocabulary learning is an essential tool for the acquisition of a second or foreign language because it demonstrates communicative competence, and it guides the way to communication. Apart from that, Ghazal (2007) contributes with the following reflection: "words are the building blocks of a language since they label objects, actions, ideas without which people cannot convey the intended meaning" (p. 84). Bearing this in mind, teaching vocabulary has a significant impact on the learners' language performance due to the benefits that it implies, not only in reading but also in any skill.

However, in order to learn vocabulary, it is not enough to memorize the meaning of words in a mechanical way; the understanding and appropriate use of the words learnt in the written

and spoken context is required. If the learners know a word, they must also know its spelling, pronunciation, collocation, and appropriateness (Nation, 1990). According to Richards (1976), the process of knowing a word involves:

- Features related to the general frequency of use and its limitations in terms of syntax and situations.
- Its raw form and the word family that it can derive.
- The system of its semantic characteristics and,
- The variations that it could have in its meaning.

We must bear in mind the fact that vocabulary knowledge is created over sequences of various encounters with that word so as to be able to learn any word. One part of the vocabulary knowledge could be provided by focus teaching, and the other ones are related to meaning-focused input and output, conscious study, and fluency development tasks. Apart from that, vocabulary teaching occurs in message-focused activities that involve the four macro skills of the language (reading, listening, writing, speaking). If the learners consider that that vocabulary is relevant for the task, it will support the students when they feel it is most required (Nation & Newton, 1997).

The English language, like other languages around the world, has a large number of words; however, not all of them are used in everyday life. Thus, to determine how “useful” a word is, we can refer to word frequency. It identifies how often a word appears in common language usage. Following this idea, an example of a useful word in English is the article “the” because it represents 7% of the words in the written or spoken language (Nation & Waring, 1997). Fortunately for the learners, there is a certain amount of words (most of them are content words) that are frequently repeated, and if the speakers know these words, they will have a fair degree of written and oral comprehension. Nation & Waring in 1997 support the statement exposed before with the following reflection “Clearly, the learner needs to know the 3,000 or so high-frequency words of the language” (p. 7).

O’Keeffe (2012) claims two perspectives to organize words: to connect words by meaning and to connect words syntactically. Concerning organization by meaning, the most relevant connections between words are synonyms, antonyms, and hyponyms. Synonyms allow

learners to handle words with similar meanings. One example is the variation of language in British and American English that presents a significant number of synonyms in their vocabulary. It is highly essential to avoid overgeneralizing meaning and to underline the need to check the context in advance to avoid errors.

Antonyms provide a sense of relationship that is favorable when teaching meaning. The learner can create mental connections that facilitate the storage in the long term memory. It brings challenges because of the existing number of antonyms for a single word and their usage in different contexts. Alternatively, hyponymy is another connection that allows us to classify the terminology into hierarchical categories. It could be said that this concept is a kind of asymmetrical synonymy because its purpose is to relate the word with a category or family word. (Carter, 1987, cited by O’Keeffe, 2012). Hyponyms is an effective manner to expand vocabulary of previous knowledge. An exemplification of it could be saying that a snake is a type of reptile or a cabin is a type of house, among others.

On the other hand, another organization relevant to vocabulary instruction is the set of words connected syntactically in patterns. These can be divided into collocations, idioms, formulaic language, and lexical chunks (O’Keeffe, 2012). Molavi, Koosha, & Hosseini (2014) define collocations as “words that occur together with high frequency...” (p. 67). It means that there are words that are predisposed to go together to convey a meaning, for example, have dinner, make money, or do homework. Secondly, idioms are collections of words that transmit an implicit idea without any literal meaning; for example, “raining cats and dogs” or “be over the moon” (O’Keeffe, 2012). Most times, in order to understand them, you need to know the context. Thirdly, formulaic language refers to fixed expressions that are pre-elaborated to transmit a message or feeling such as “nice to meet you”, “see you later” or “you’re welcome”, among others. Finally, lexical chunks or multiword units are brief fixed phrases than do not contain more than six words: “you see”, “I mean”, “let me see”, etc. (O’Keeffe, 2012).

#### 2.4. Pre-teaching Vocabulary

Pre-teaching vocabulary is an instructional method that prevents the brain from decoding the meaning of unknown words. Instead of that, it focuses on the comprehension of the passage being read; it allows learners to create a vocabulary fluency and automatically identify the words’ meaning while reading (Mousavian & Siahpoosh, 2018). Therefore, this method helps

that vocabulary and comprehension to go hand in hand. If readers know the meaning of the words present in the text, they will accurately understand the main idea of what they are reading.

According to Beck, Perfetti & McKeown (1982), gaining reading comprehension could be obtained by vocabulary training. The learners understand the meaning of single words, and the mental construction of the meaning of the whole passage is carried out efficiently. Nation (2001) corroborates this idea with the following reflection: "The more words students know and recognize, the better they understand a text" (p. 64). Taking this information into account, it can be concluded that pre-teaching vocabulary could activate previous knowledge because the learners meet the terms and content found in the reading; with this information, they establish their background knowledge and start their interaction with the text.

As important as pre-teaching vocabulary is, due to its multiple benefits, determining the vocabulary to be taught. Beck et al. in 1982 claim that, "pre-teaching vocabulary may be more successful if the words to be taught are keywords from the target passage" (p. 101). When they refer to "keywords," we need to understand that they are referring to relevant words with recurrent frequency inside the text. These high-frequency words appear many times in the passages and in different contexts. In this way, readers will find that the text has familiar terminology, and they will focus on understanding what the writer wants to transmit instead of on looking up the meaning of unknown words.

As a matter of fact, it is estimated that a learner can recognize 80% of the terms in a text containing 2,000 words. However, it is required at least 95% of vocabulary comprehension for a reasonable understanding of a text and effective guessing of meaning from context (Nation & Waring, 1997). Thus, when there are many unknown words, they can provoke gaps in the meaning of the text. If they are too many, they will interfere with the text's processing, and the reader will have difficulties constructing meaning (Beck et al., 1982).

#### **2.4.1. Vocabulary strategies**

Undoubtedly, at the moment of teaching vocabulary, a considerable amount of time is consumed. For this reason, deciding whether the words to be taught are relevant or not is the most important. Therefore, it is preferable to spend time teaching high-frequency words or those that learners need to fulfill their linguistic needs, instead of low frequency or technical

words that are not useful for them (Nation & Newton, 1997). Apart from that, if there is a huge amount of vocabulary, teaching might be done by administering indirect or incidental learning through extensive reading or listening, problem-solving group activities, and formal classroom activities focused on vocabulary (Nation & Waring, 1997).

As a matter of fact, different strategies depend on the approach to teach vocabulary. For example, according to Schmitt (1997), there are two types of strategies. The first one refers to the identification of the meaning of those words that are met for the first time. It encloses social strategies and determination. The second one aims at building up that knowledge for future encounters, containing metacognitive, cognitive, memory, and social strategies, which will be detailed below.

Firstly, we must state that according to Schmitt's Taxonomy, cognitive strategies have similarities with memory ones and include recurrence and activities to study words, such as flashcards, vocabulary diary, and word lists. Then, metacognitive strategies are applied by the learners to be aware of their learning process and evaluate it through a general overview. Last but not least, social strategies refer to that work in groups that goes hand in hand with memory strategies. One example is the well-known Mnemonics, which consists on applying imaginary or grouping activities to associate the word with previous knowledge. Consequently, the general objective is to combine a variety of strategies to absorb the new vocabulary and use it in other contexts (Schmitt, 1997).

On the other hand, other second language vocabulary learning strategies were listed by Gu and Johnson in 1996; the list contains four strategies which can be enumerated as cognitive, metacognitive, memory, and activation. First, cognitive strategies contain activities such as guessing as a means for activating background knowledge and linguistic clues, taking notes of words' meaning and using it with orientation, and the smart use of the dictionary. Then metacognitive strategies work with selective attention that recognizes crucial words for passage comprehension and self-initiation that apply varied ways to make the meaning of the words transparent. On the flip side, memory strategies are subdivided into two groups: rehearsal and encoding strategies. Within the rehearsal strategies, word list repetition, oral repetition, and visual repetition are included. When it comes to encoding, we find strategies such as imagery, auditory encoding, visual encoding, contextual, semantic, and word

structure. Finally, activation strategies promote the application of the new vocabulary acquired in other contexts apart from it.

Apart from that, Nation & Meara (2010) stated other four main strategies that help students to find the meaning of unknown words and place them in their long term memory. The first one is guessing from context; the learner should pay special attention to the linguistic clues surrounding the unknown word. Considering that learning words is a cumulative process, this strategy is beneficial, but there must be enough supporting comprehensible context. The second strategy is learning from word cards; this strategy consists on associating the new word with a similar L1 and a picture of the meaning of that word in L2. Thus, the combination of both creates a reference for the learner. The third one uses Word Parts; the main objective of this strategy is to break the words into parts and analyze them one by one to get meanings and then join them as a full content. In order to carry out this strategy, the learners must know about prefixes, suffixes, and word families.

The third one is the use of the dictionary; no matter what type of dictionary it is, it requires a variety of subskills for their appropriate use, like interpreting grammatical information, phonemic transcription, and guessing from context to choose the correct alternative. The main objective is that the user learns strategies to take advantage of the dictionary as a learning tool. These strategies can include comparing the found definitions to check if there is a relation between them, to verify if the new word is related to a known word, and to visualize some sentences in order to exemplify the new terminology. Additionally, the learners need a guide to monitor what sub-skills they have and which they need to develop progressively.

According to O'Keeffe (2012), one way to accelerate vocabulary learning and retention is to increase contextual encounters (finding the same word in different real contexts). The most common encounters are generally watching television, movies, series or reading books, magazines or newspapers. Exposing learners to several contextual encounters provides them the opportunity to manipulate the vocabulary by seeing, reading, saying, or writing words that later they will retain in their long term memory. Therefore, "reading, especially, offers the opportunity for the learner to build advance vocabulary" (O'Keeffe, 2012, p 6). The reason is that readers come across more specialized and lower frequency vocabulary while they read about a particular topic.

The concepts presented before are a compilation of relevant information related to Teaching English as a Foreign Language in terms of vocabulary and reading comprehension. All this information supports our intention when creating this proposal because there are clear connections between the main variables. For example, the viability of carrying out a corpus-based analysis to get the relevant vocabulary from a text. As a matter of fact, the use and manipulation of corpora to learn a second language has been exploited in recent years (Cobb, 1998). Therefore, with analysis tools such as the *Lextutor* program, it is possible to deal with the lexis and prepare our students for it when they are learning a new language.

To sum up this section, we can say that the benefits of pre-teaching vocabulary have been shown as well as some helpful strategies that might be employed when teaching vocabulary in order to improve our learners' performance in the L2.

### 3. Intervention proposal

Reading skills have always been of great interest in teaching a second or foreign language (Nation, 2009). It is because they bring many advantages when learning a second or foreign language, as it has been previously exposed. However, finding an effective way to raise our students' interest in reading to acquire these advantages has been demanding. Learners frequently find it challenging to comprehend books because they have not been trained appropriately in all the reading stages.

Therefore, this intervention proposal suggests the need to pre-teach vocabulary before the students start to read a book. Research has shown that pre-teaching vocabulary positively affects the reading processes (Mousavian & Siahpoosh, 2018). Nonetheless, this vocabulary must be specific to the text that the students will be reading. For this aim, we propose to carry out a corpus-based analysis of the book's lexis to improve the learners' reading comprehension; the process to be followed is detailed below:

Firstly, the students select a reading book through a survey where the teacher suggests five books from the Gutenberg project in accordance with their level. Then, in order to find the high- frequency keywords that appear in the selected book, the online program *Lextutor* is employed as a tool analysis of the corpus. After that, the teacher prepares seven one-hour sessions to pre-teach this terminology, applying the most recommended strategies for this



aim. Next, the learners will have two months to read the book on their own, outside the classroom. During this process, a checklist will be applied to verify if they are getting a good comprehension of the text. Finally, the effectiveness of vocabulary pre-teaching will be assessed through a task-based activity. In the end, the learners will be offered the opportunity to choose which task to carry out according to their abilities.

### 3.1. Aims of the proposal

#### 3.1.1. General objective

The main objective is to help students increase their grasp of vocabulary to have a better general comprehension of the story told in the reading book. Bearing this in mind, we propose to carry out a corpus-based analysis and get the frequency of word recurrence of a reading book; with the analysis results, the teacher pre-teaches the most relevant and repeated words to facilitate vocabulary assimilation and improve reading comprehension.

##### 3.1.1.1. Specific objectives

- To provide the learners with a range of five reading books from the Gutenberg project for them to choose one for the whole group through a survey. The books are proposed by the teacher, taking into consideration their current level.
- To pre-teach a list containing the high- frequent and relevant words which appear in the reading book, using various proved strategies to successfully ensure learning.
- To apply a checklist of self-assessment while the students are reading the book in order to verify that their comprehension is satisfactory.
- To administer task-based activities to assess the reading comprehension of the reading book and the effectiveness of the pre-teaching vocabulary strategies applied initially. The learners will be able to choose the task according to their abilities.

### 3.2. Educational context and Target group

This proposal is designed for young adult learners from 18 to 25 years old. They study engineering at the *Universidad de las Fuerzas Armadas ESPE*, located in Latacunga city in Ecuador. As a matter of fact, they are studying English because according to their curriculum, they need to pass a B2 level according to the CEFR to obtain their bachelor's degree. In this

case, they are in the V-VI course that corresponds to the B1 level of the Language Department that belongs to the same University.

Additionally, the Language Department works with a set of textbooks called "Big Picture" by Richmond Publishing, and the content is presented in British English. Conforming to their level, they use "The Big Picture B1 Intermediate". This book includes a small reading book that is not highly useful because its plot doesn't cover the students' expectations, and they are not motivated to read it. Moreover, it is not an authentic material that presents a real contextualization of English cosmopolitanism.

Apart from that, there is a range of 20 to 35 students in each classroom. Most of the time, the learners are just Ecuadorian coming from different parts of the country. The traditional teacher-centered approach with a methodology focused on teaching grammar using L1 is the most recurrent one, and that is why the learners do not have enough vocabulary knowledge for their level and their reading comprehension is not satisfactory. There aren't any special needs.

### 3.3. Timing

This proposal will be carried in two parts. The first part included session N. 1 (survey), session N.2 (Corpus-based analysis), session N.3 (Preparation of the vocabulary classes), and they were carried out online in the second semester of 2020. The second part corresponds to sessions 4-10 (vocabulary classes), the reading process, and the assessment. They will be developed when the physical class attendance is resumed because of the limitations of online teaching. In total, the proposal will last four months. First of all, we will start with a survey which will help us choose the book to be read during the first week. As soon as the results are gathered, the corpus-based analysis of the book's lexis will be developed. Once the keywords list is completed, the teacher will prepare the lessons and the materials needed to pre-teach vocabulary during the second week. After that, in the third week, the learners will receive seven 60-minute-sessions to get familiarized with the high- frequency keywords. For the next two months, the learners will read the book on their own outside the classroom; they will complete a checklist to self-evaluate their reading performance during the process. Once they have finished reading, during the following two weeks, they will carry out the task-based

activity designed by the professor. In the end, the professor will assess the results of the proposal which will help reach some conclusions regarding its effectiveness.

**Table 1. Proposal timetable**

ACTIVITIES	First month				Second month				Third month				Fourth month			
Weeks	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
<b>A survey to choose the reading book.</b>	X															
<b>Corpus- based analysis of the book's lexis</b>	X															
<b>Preparing the classes to pre-teach vocabulary</b>		X														
<b>First class session for pre-teaching vocabulary.</b>			X													
<b>Second class session for pre-teaching vocabulary.</b>			X													
<b>Third class session for pre-teaching vocabulary.</b>			X													
<b>Fourth class session for pre-teaching vocabulary.</b>			X													
<b>Fifth class session for pre-teaching vocabulary.</b>				X												

*Table 1.* Proposal timetable. Own elaboration.

This proposal is based on a learner-centered approach because the learners are the main actors of the class, working on their own or with their classmates. The teacher has a facilitator role, who guides the students during the class and helps them to get organized. Furthermore, the vocabulary is presented in an explicit way most of the time, but with activities that lead the students to apply the knowledge in real contexts.

Finally, a continuous formative assessment is applied in each session, with activities such as quizzes, games, crosswords and pair work among others. Additionally, the teacher intends to

revise the content seen in the previous lesson and provide immediate feedback to help students in the process of knowledge consolidation. All of this process will enable both teachers and students to handle the great amount of vocabulary proposed.

### 3.5. Sessions and activities

#### 3.5.1. "Session 1" Survey

**Specific objective of this activity.** To provide the learners with a range of five reading books from the Gutenberg project. They are in charge of choosing one of them for the whole group through a survey. The books are proposed by the teacher, taking into consideration their current level.

**Development of the activity.** Firstly, the teacher has checked the books stored in Project Gutenberg; the section selected was light literature that corresponds to novels and stories to read for pleasure. In order to select the books, some steps were to consider the internet articles as well as other teachers' recommendations and the personal experience of the teacher. Moreover, the parameters to choose the book were: the text's length, the level of the vocabulary and language in general, and the plot.

Once the books were selected, a brief description of each book was provided by the teacher. This information was then included in a Google survey to which the learners had access in order to read and select which book they would like to read. Subsequently, 25 students from the B1 level course from the *Universidad de las Fuerzas Armadas ESPE* took the survey, and the results were the following.

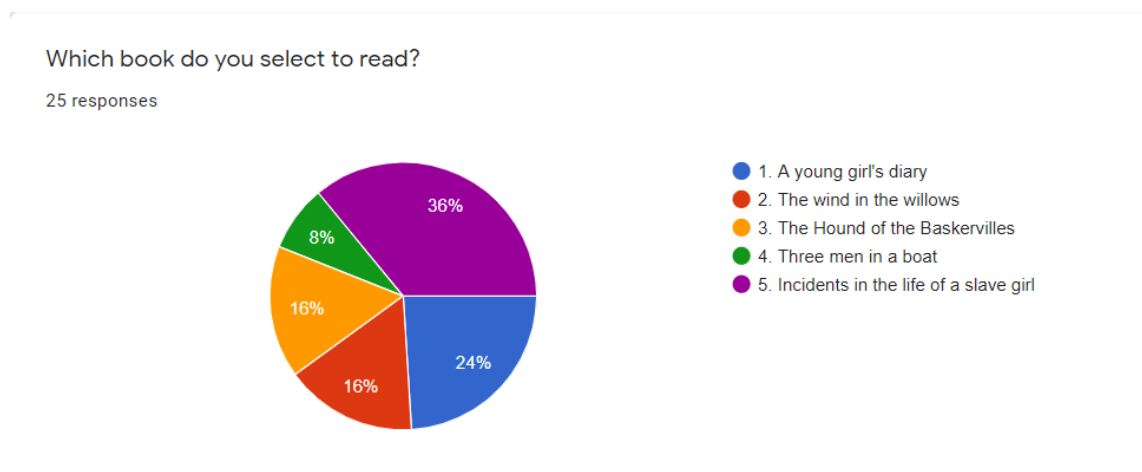


Figure 7. Results of the books' survey (taken from Google Forms, 2020)

Consequently, the results showed that 9 people were willing to read the book *The incidents in the life of a slave girl*, which corresponds to 36% and is the majority. On the other hand, 6 learners selected *A young diary girl* to read, and it was 24%, the second with the most votes. Apart from that, 4 of those surveyed selected *The wind in the widow*, representing 16%. The same thing happened with the book *The Hound of the Baskervilles*, which obtained 4 votes and 16% as well. The last one was *Three Men in a boat*, with 2 votes representing the 8%. In conclusion, the book selected by the majority was *The incidents in the life of a slave girl*, and it could be due to the fact that it portrays the real story of a slave woman and her fight for freedom, told by herself. Undoubtedly it is an appealing story for young adult learners.

**Specific resources.** Since the first part of the proposal was developed online, the contact with the students was by e-mail, and 25 students from the Language Department of the *Universidad de las Fuerzas Armadas ESPE* participated. Therefore, the survey was virtual, and the platform used was Google forms because its interface is well known and easy to follow. Once the participants answered, the results were shown and scrutinized automatically. The link to the survey is presented below, and the printed version is available in Annex 1.

**Link to the survey :**

[https://docs.google.com/forms/d/e/1FAIpQLSdl1Yrd42Gvgu\\_x8GuFGRXnGkBJldTxz2up61Xrfe7hsqel3w/viewform?usp=sf\\_link](https://docs.google.com/forms/d/e/1FAIpQLSdl1Yrd42Gvgu_x8GuFGRXnGkBJldTxz2up61Xrfe7hsqel3w/viewform?usp=sf_link)

**Duration of this activity.** As we have previously mentioned, the survey was carried out by e-mail; thus, the potential participants of the proposal took one week more or less to answer it due to the internet limitations. However, the results were processed automatically by Google Forms when all the answers were received; for this reason, the final result was ready in one day.

### 3.5.2. "Session 2" Corpus based analysis

**Specific objective of this activity.** To carry out a corpus-based analysis of the textbook *The incidents in the life of a slave girl* in order to identify the most- frequent and relevant words. To this purpose, the online *Lextutor* program was employed with the sections *Keywords* and *Text Lex Compare*. Applying these analysis tools of the *Lextutor* program, it was possible to obtain accurate information about the lexis of the text.

**Development of the activity.** During this process, the importance of not only looking for high-frequency vocabulary because it mainly contains function words became visible. Then the teacher looked for other tools to look up content words that are commonly used in the Real English corpus. For this reason, the analysis was divided into two processes to obtain real and verified results. The first process was developed with the *Lextutor* section called *Keywords* to get the most relevant vocabulary included in the text. As a matter of fact, the program selects the most frequent words and compares them with a corpus composed by 10 million words of written and spoken language created by Paul Nation [bnc coca fams speechwrite US UK per10mill](#). The program concludes that the term is a keyword in the text when it appears in the corpus list and in the text more than a few times (Nation, Lufer, & Cobb, 1995).

Taking this into account, the book was uploaded as a .txt file, and a total of 84,968 words were detected. Then the keywords list was 148, and the keyword ratio was 1,74 per 1,000 words. Bearing in mind that ratio of more than one word per 1,000 means a high keyword text and, less than that means that the text contains mainly general terms (Nation, Lufer, & Cobb, 1995); it is concluded that this text is a rich source of distinct vocabulary that could be useful for learners. The following figure shows the results of the *keyWords* tool's analysis.

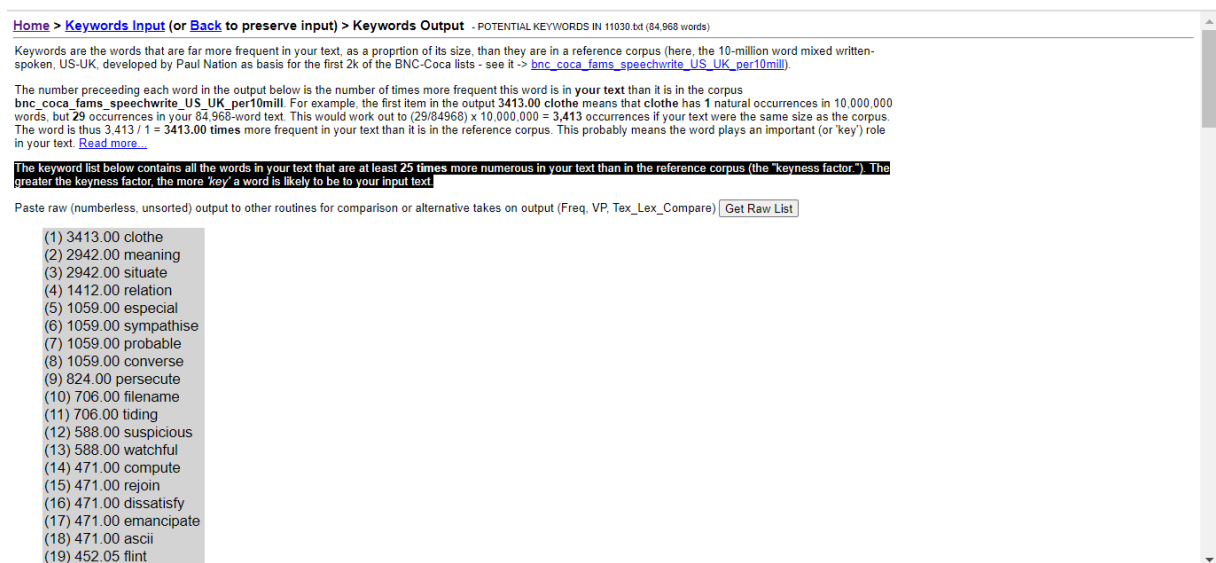


Figure 8. Result of the keywords tool's analysis (taken from *Lextutor.ca*, 2020)

On the other hand, the second process was carried out in order to verify that the keywords were highly recurrent in the text. To this purpose, the *Lextutor* section called *Text Lex Compare* was utilized. This tool allows to compare two corpora and identify the words that are shared in both. Therefore, the teacher uploaded the book's corpus and the keyword list to the program; after the analysis, it was proved that the keyword list contained the high-frequency words of the text because none of them was repeated less than four times. As a result of this analysis, ambiguity was avoided when deciding which words were relevant in the text to pre-teach them. The results of this analysis are shown in the figure below. (To see the full keyword list with the frequency number of each one, go to Annex 2).

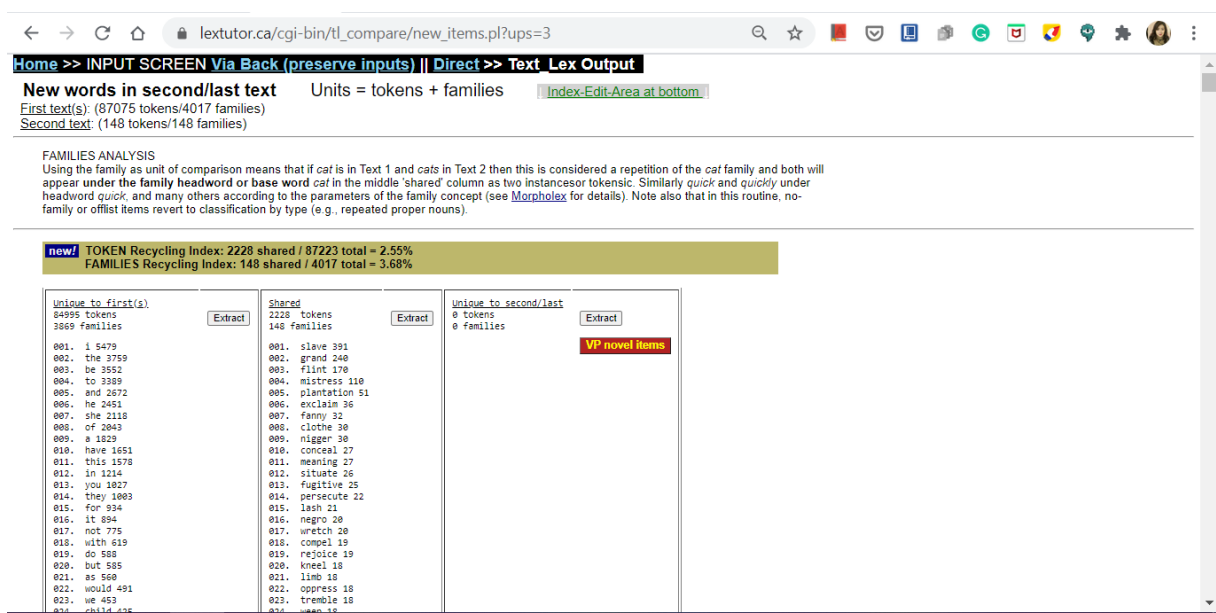


Figure 9. Comparison to verify the frequent words (taken from Lextutor.ca, 2020)

**Specific resources.** Firstly, all the resources applied in this analysis were digital. To begin with the activity, the teacher downloaded the .txt file extension of the book from the Gutenberg project webpage; then the second resource was the Lextutor program that is available on their webpage [Lextutor.ca](http://Lextutor.ca). As previously stated, one of the advantages of this program is that the user does not need to install it or create an account; it has free access for everyone. Furthermore, each section has instructions for the activities that can be done there.

**Duration of this activity.** The corpus analysis was done relatively fast because the *Lextutor* program processes the data in a short period and its tools are efficient. For this reason, this activity was developed in one day. It is essential to bear in mind that the long texts can only



be analyzed if they have a .txt extension file. As the Gutenberg Project web page provides the books in different formats, including this one, it was time-saving to process the data.

### 3.5.3. “Session 3” Preparation of the vocabulary classes

**Specific objective of this activity.** To prepare the strategies and material required to teach the vocabulary list, in order to enhance the learners’ acquiring of this vocabulary in the long term. If they internalize most of the words taught at the beginning, they will read the book faster and more easily because the time-consuming task of checking the meaning of unknown words will disappear. Consequently, the general comprehension of the reading will be much more accurate (Mousavian & Siahpoosh, 2018).

**Development of the activity.** Firstly, the teacher reviewed the vocabulary list generated by the program *Lextutor*. 15 words that are cognates, names or have well-known meanings were omitted and the final list to be presented contained 133 words (see it in Annex 2). Thus, the vocabulary will be taught in seven sessions with 20 words in each one because of the available time. The activity will last two weeks, and the classes will be sixty-minute sessions.

Furthermore, the teacher created a glossary with the keyword list in the ‘Flashcards’ section of the *Lextutor* program. The learners will have a database with the meaning of the words retrieved from the Cambridge Dictionary (2020). Additionally, the program provides the standard pronunciation of each term. Apart from the definition, they will also have the option of practicing taking a multiple choice quiz, selecting the option ‘quiz’ (Qz).

In order to avoid overwhelming the learners with the number of words, the teacher divided the glossary into 20 words for each session. In this way, they will focus on the specific terms of that day. Moreover, in the session, they will rehearse the words in order to internalize the meaning in an abiding way. The access to the glossaries is through a link provided by the teacher at the beginning or in the middle of the class, and they will have access to it at any time, for example <https://lxtutor.ca/flashcards/keywords-1.html>. Additionally, the teacher will provide the [e-portfolio](#) of the Lextutor program to the students beforehand because some activities will be carried out applying this tool and they need to know how to use it appropriately.

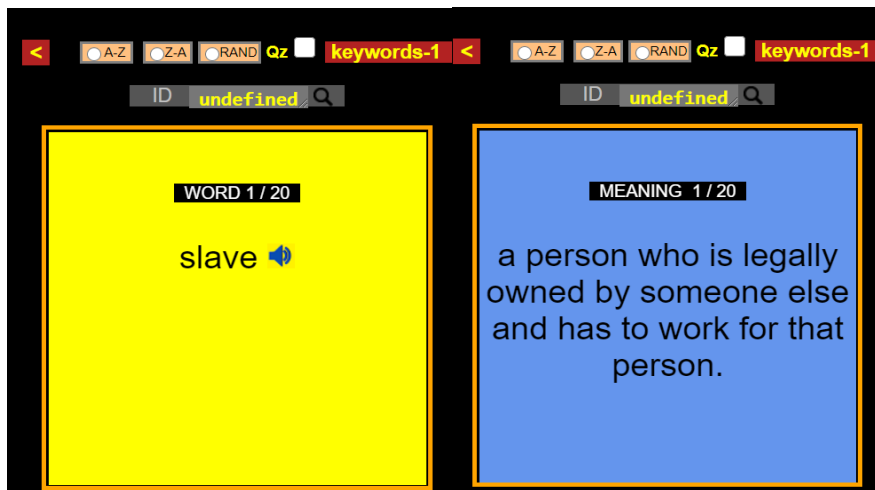


Figure 10. Example of first session glossary (taken from Lextutor.ca, 2020)

**Specific resources.** As it was previously mentioned, the *Lextutor* program was used to create the glossaries with the section called *Flashcards*. Additionally, sections like *Multi-concordance* and *Identify input* of the same program will be used by the learners in order to practice the vocabulary presented. Another resource applied was the [Online Cambridge Dictionary](#) (2020) to look up the definition of the words in English. Finally, the strategies to be used when *pre-teaching* vocabulary are the following:

- Guessing to activate background knowledge and linguistic clues (Gu & Johnson, 1996).
- Word formation, synonyms and antonyms, suffixes and prefixes and spelling (Nation, 2009).
- Mnemonics (Schmitt, 1997).
- Guessing from context, keyword technique, word parts (Nation, 2010).

**Duration of this activity:** The teacher took one week planning the pre-teaching sessions because a meticulous selection of the strategies to be applied and the resources to be used was necessary. Undoubtedly, the success of this proposal highly depends on the correct vocabulary instruction. Furthermore, the creation of the glossary was also time-consuming because of the number of words and the proper selection of their meaning.

#### 3.5.4. “Session 4” Vocabulary class N. 1

**Specific objective of this activity.** To pre-teach the high-frequency keywords obtained from the reading book. To do so, the teacher will use some proved teaching vocabulary strategies such as guessing from context and mnemonics. It will influence positively in the reading

comprehension of the text later because it will reduce the time devoted to looking up the meanings of the unknown words (Cowel, 2012).

### Development of the activity.

**Table 2. Vocabulary class N. 1**

Activities	Length	Goals	Outcomes	Resources
<b>1-General guidelines</b>	5'	To inform the students of the book selected and the outcomes of the classes.	Students are aware of the purpose of the activities proposed.	Blackboard
<b>2-Brainstorming</b>	10'	To activate previous knowledge.	Students can define what slavery is and its consequences.	Blackboard
<b>3- Guessing from context</b>	15'	To apply the guessing strategy to activate linguistic clues and increase contextual encounters.	Students are able to guess the vocabulary' meaning from context using the <i>Lextutor</i> program.	Internet Computers cellphones
<b>4- Vocabulary flashcards presentation</b>	15'	To present the vocabulary's meaning using the <i>Lextutor</i> program and make students reflect on their own concepts.	Students can compare and contrast their guessing with the real vocabulary's meaning and draw conclusions.	Internet Computers cellphones
<b>5-Mnemonics</b>	15'	To make them use the new vocabulary in another context to enhance its internalization.	Students can invent a poem on "slavery", applying the vocabulary learned.	Notebooks Pens Markers

Table 2. Vocabulary class N. 1. Own elaboration.

**Activity 1.-** The class will start informing the students that the storybook selected in the survey was *The incidents in the life of a slave girl* because it got the 36% of the votes. Then the teacher

will explain that the book's most relevant vocabulary will be reviewed to help them in the reading process, and finally, the teacher will present the outcomes of this class.

**Activity 2.-** For this activity, the teacher will write on the board the word "slavery" and ask students what comes to their mind reading this word, how they feel, and the implications of this practice in our current society. Next, the teacher will give them time to think; then, the teacher will ask them to write their ideas on the board. In the end, the whole class will analyze their ideas and share opinions about the topic orally.

**Activity 3.-** The teacher will ask the students to use their computers or cellphones and open the multi-concordance section of the *Lextutor* program: <https://www.lexutor.ca/conc/multi/>. Then the students will get in pairs. Next, the teacher will write on the board twenty words and ask students to type them in the program's searcher. As a result of this action, some sentences that contain the vocabulary will be displayed. The pairs will read the sentences and infer the meaning of the word in bold, paying attention to the surrounding context clues.

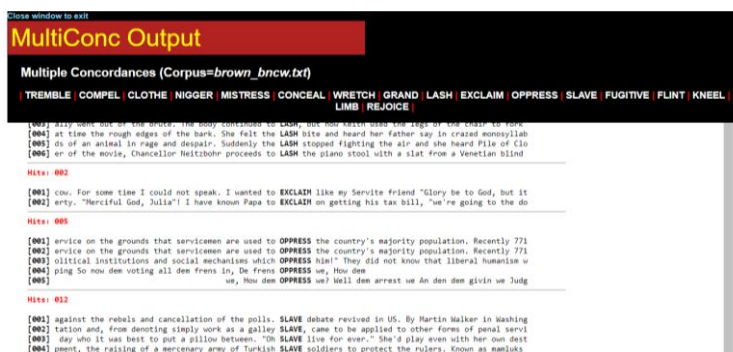


Figure 11. Multi concordance exercise (taken from Lextutor.ca, 2020)

**Activity 4.-** The teacher will provide the link to the vocabulary flashcards: <https://lexutor.ca/flashcards/keywords-1.html> and ask the pairs to review the meaning of each word presented before. Then the students will compare their guessing with the formal definition of the vocabulary; they will verify the correct and incorrect assumptions and the possible reasons for the mistakes. In the end, each pair will share with the whole class how many words they guessed correctly from context.

**Activity 5.-** Firstly, the teacher will ask students how they feel thinking about slavery, then they will imagine, if they were living in those days, what they would say to the slaves. After

this reflection, the teacher will ask them to write a poem about slavery, expressing their feelings and using as many words from the vocabulary as they can. In the end, they will post the poems on the wall in order to share their ideas with the class.

### 3.5.5. "Session 5" Vocabulary class N. 2

**Specific objective of this activity.** To pre-teach the high-frequency keywords obtained from the reading book. To do so, the teacher will use some proved teaching vocabulary strategies, for instance, working with synonyms and word formation. Hence, the keyword list of this session was rearranged to work with the activities proposed.

#### Development of the activity.

**Table 3. Vocabulary class N. 2**

Activities	Length	Goals	Outcomes	Resources
<b>1- Feedback Sentence construction</b>	10'	To review the previous class' vocabulary and provide feedback in order to help students to internalize it.	Students are able to tell a sentence using a word from the previous class' vocabulary.	Piece of papers A bag
<b>2- Vocabulary flashcards presentation</b>	10'	To present the vocabulary using the <i>Lextutor</i> program and ask them to classify it into parts of the speech.	Students meet the new vocabulary through the use of flashcards in the <i>Lextutor</i> program and they classify it into nouns, verbs, adjectives or adverbs.	Internet Computers cellphones
<b>3- Work with synonyms</b>	15'	To make students create mental connections between their previous knowledge with the new one.	Students are able to find synonyms for the new words and create pictures or symbols for them.	Notebook Pens markers
<b>4- Word formation</b>	20'	To let students explore the formation of the words with compound nouns, prefixes and suffixes.	Students are able to disassemble and reassemble words to analyze their meaning.	Notebook Pens markers

<b>5. Multiple choice quiz</b>	5'	To present the students the multiple choice quiz option in the <i>Lextutor</i> flashcards and let them to measure their own progress.	Students can answer the multiple choice quiz and reflect on their learning.	Internet Computers cellphones
--------------------------------	----	---	---	-------------------------------------

Table 3. Vocabulary class N. 2. Own elaboration.

**Activity 1.-** At the beginning, the teacher will ask students to tell aloud the words they remember from the last class. The teacher will show pieces of paper with the vocabulary seen in the previous lesson and puts them into a bag. The students will have to take one piece of paper and produce a sentence orally with the word written. If one student does not know the definition, their classmates can help them, or the teacher may provide clues to help them remember.

**Activity 2.-** The teacher will present the new vocabulary for this class using the following link <http://lextutor.ca/flashcards/keyword-2.html> . The students will revise the new vocabulary and read the definitions. After that, they will classify the words into parts of the speech, such as nouns, verbs, adjectives, or adverbs.

**Activity 3.-** First, the students will get into three-people groups. Then the teacher will ask them to look for a synonym for the ten first words on the list provided before (bestow, degrade, procure, summon, swamp, acquaint, blight, converse, sympathize, and clasp). These synonyms should be words that students already know and remember in order to create mental connections. When they have found the synonyms, they will draw a picture or symbol representing the new words. In the end, they will share their lists with the other groups to compare ideas, for example:

- Bestow= award 🏆.

**Activity 4.-** The students will continue working in groups. In this case, they will disassemble into morphemes the last ten words of the list presented before (housemaid, storeroom, watchful, disclaim, trademark, footstep, proclaim, filename, dissatisfy, runaway). To this purpose, they will analyze if they are compound nouns, prefixes, or suffixes. Using this information, they will explain the reason for their meaning, for instance:

- Housemaid= compound noun.

- House= a building where people live.
- Maid= a woman who works as a servant.

**Activity 5-** At the end of the class, the teacher will ask students to open the *Lextutor* flashcards again and show them how to activate the multiple-choice quiz. Then the students will answer this vocabulary quiz, and they will measure how many words' meanings they remember and which ones they have to work on.

### 3.5.6. "Session 6" Vocabulary class N. 3

**Specific objective of this activity.** To pre-teach the high-frequency keywords obtained from the reading book. To do so, the teacher will use some proved teaching vocabulary strategies such as working with antonyms and mnemonics. In this way, the learners will be familiar with the new terms and keep them in an abiding way.

**Development of the activity.**

**Table 4. Vocabulary class N. 3**

Activities	Length	Goals	Outcomes	Resources
<b>1- Feedback Guessing the word</b>	10'	To revise the vocabulary seen in the previous class and provide feedback in order to help students to internalize it.	Students are able to tell the definitions of the previous class' vocabulary and work in groups.	Blackboard Markers Chairs
<b>2- Vocabulary flashcards presentation</b>	10'	To present the vocabulary meaning using the <i>Lextutor</i> program.	Students get to know the new vocabulary through the use of flashcards in the <i>Lextutor</i> program.	Internet Computers Cellphones
<b>3- Working with antonyms</b>	20'	To make students create mental connections between their previous knowledge with the new one.	Students are able to find antonyms for new words using familiar terms.	Notebook Pens Markers
<b>4- Mnemonics</b>	20'	To promote the use of Mnemonic tools to	Students can create a short story applying the	Blackboard

---

enhance	the new vocabulary and	Markers
internalization of the	evaluate its correct	
new terms.	usage.	


---

*Table 4. Vocabulary class N. 3. Own elaboration.*

**Activity 1.-** To start the class, the teacher will divide the course into two groups to play a guessing game. Each group will choose a participant and place one chair backwards at the front. For this, the teacher will write on the board a word from the previous class list; the participants will not see the term, but their classmates will. Each group will tell the definition to their respective participant, and he or she will figure out what term it is. The participants have to rotate one by one. The winner will be the group that guesses more words.

**Activity 2.-** The teacher will present the new vocabulary for this class with the following link <http://lextutor.ca/flashcards/keywords-3.html>. The students will review the new vocabulary and read the definitions.

**Activity 3.-** For this activity, students will get in pairs. Then the teacher will ask them to copy the first ten words of the vocabulary list presented before. After that, students will look for the antonym of each term. In order to make mental connections, the antonym selected must be a familiar word. When they finish writing, they will draw a picture of the antonym and cross it out, for example:

emancipate ≠ slavery 

**Activity 4.-** To start, the teacher will divide the class into two groups; each group will create a short story with the last ten words of the list provided before. It does not need to have much logic, and it can have something hilarious. To this purpose, each member of the group will write a sentence on the board, and the next one must continue the story until everybody has participated. In the end, each group will read its story, and the other one will evaluate if the terms have been applied accurately.

### 3.5.7. "Session 7" Vocabulary class N. 4

**Specific objective of this activity.** To pre-teach the high-frequency keywords obtained from the reading book. To do so, the teacher will use some proved teaching vocabulary strategies,



for instance, increasing contextual encounters, completing crosswords, and self-evaluation. In this way, the learners will be familiar with the new terms and internalize them for the future.

### Development of the activity.

**Table 5. Vocabulary class N. 4**

Activities	Length	Goals	Outcomes	Resources
<b>1- Feedback Kahoot Game</b>	10'	To revise the vocabulary seen in the previous session and provide feedback in order to help students to internalize it.	Students are able to select the correct word to complete a sentence in the Kahoot game.	Internet Computers Cellphones
<b>2- Vocabulary flashcards presentation</b>	10'	To present the meaning of the selected vocabulary using the <i>Lextutor</i> program.	Students get to know the new vocabulary through the use of flashcards in the <i>Lextutor</i> program.	Internet Computers Cellphones
<b>3- Increase contextual encounters</b>	20'	To promote that the students see the words in real contexts and choose the best example for the definitions provided.	Students are able to select the most accurate sentence that matches with the new vocabulary definitions presented before.	Internet Computers cellphones
<b>4- Crossword</b>	15'	To provide the students with a crossword to practice the vocabulary presented by listening to it.	Students listen to the definition and recognize the word to complete the crossword.	Internet Computers cellphones
<b>5 Multiple choice quiz</b>	5'	To present the students the multiple-choice quiz option in the <i>Lextutor</i> flashcards and let them measure their own progress.	Students answer the multiple-choice quiz and reflect on their learning.	Internet Computers cellphones

Table 5. Vocabulary class N. 4. Own elaboration.

**Activity 1.-** At the beginning of the class, the teacher will ask students to open the Kahoot app to play a game with the vocabulary seen in the previous class (they have used the app before).

Thus, the teacher will show the screen with the incomplete sentence (taken from the book), and the students will have to select one term to fill the gap. At the end of each attempt, the teacher will provide feedback on the reason why the selected term is the correct one.

**Activity 2.-** The teacher will present the new vocabulary for this class with the following link. <http://lextutor.ca/flashcards/keywords-4.html> The students will review the new vocabulary and read the definitions. After that, they will get in pairs and test their memory by asking questions about the meaning of the words to each other.

**Activity 3.-** For the following activity, the learners will continue working in pairs. Firstly, the students will go to the *Lextutor* website and select the word games section- *Multi-concordance*: <https://www.lex tutor.ca/conc/multi/>. Then they will type in the program's searcher the vocabulary list presented before. After that, they will read the sentences generated for each word in context. Finally, the learners will discuss in pairs, and they will select the example that best matches the definition provided.

**Activity 4.-** The teacher will provide the link to a crossword to practice the last 10 words of the vocabulary list provided before. [https://es.educaplay.com/recursos-educativos/7911754-vocabulaulary\\_session\\_4.html](https://es.educaplay.com/recursos-educativos/7911754-vocabulaulary_session_4.html) The learners will listen to the definition and write the correct word.

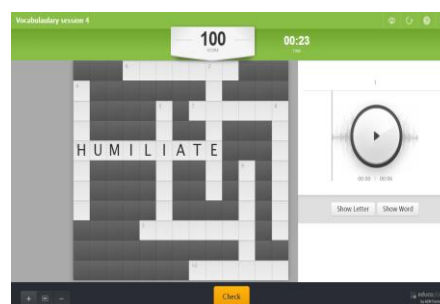


Figure 12. Crossword vocabulary class 4 (taken from Educaplay, 2020)

**Activity 5.-** At the end of the class, the teacher will ask students to open the *Lextutor* flashcards again and show them how to activate the multiple-choice quiz. Then the students will answer this quiz about the vocabulary, and they will evaluate how many definitions they remember and which ones they have to work on.

**3.5.8. "Session 8" Vocabulary class N. 5**

**Specific objective of this activity.** To pre-teach the high-frequency keywords obtained from the reading book. To do so, the teacher will use some proved teaching vocabulary strategies, such as mnemonics and contextual encounters. In this way, the learners will be familiar with the new terms, and they will read the book more easily.

**Development of the activity.****Table 6. Vocabulary class N. 5**

Activities	Length	Goals	Outcomes	Resources
<b>1- Feedback Spelling contest</b>	10'	To revise the vocabulary seen in the previous session and provide feedback in order to help students to internalize it.	Students are able to say the word when listening to the definition and spell it.	Blackboard Markers
<b>2- Vocabulary flashcards presentation</b>	15'	To present the vocabulary using the <i>Lextutor</i> program and ask them to classify it into parts of the speech.	Students get to know the new vocabulary through flashcards in the <i>Lextutor</i> program and they classify it into nouns, verbs, adjectives or adverbs.	Internet Computers Cellphones Notebook
<b>3- Mnemonics</b>	20'	To promote the use of mnemonic tools to enhance the internalization of the new terms through mental connections.	Students are able to match the word's definition with something or someone familiar to them.	Notebook Markers Pens
<b>4- Increase contextual encounters</b>	15'	To present a real context to apply the new vocabulary.	Students can select the correct words to complete statements generated by the <i>Lextutor</i> program.	Internet Computers cellphones

*Table 6. Vocabulary class N.5. Own elaboration.*

**Activity 1.-** To start the class, the teacher will explain the spelling contest. To this purpose, the course will be divided into two groups. The teacher will be the moderator and tell the

definition of the word. Then each group will select a participant, and they will say the word that corresponds to the meaning and spell it. If the term and its spelling are correct, the group will get 1 point. The participants will rotate until everybody has participated.

**Activity 2.-** The teacher will present the new vocabulary for this class with the following link.

<http://lextutor.ca/flashcards/keywords-5.html> The students will revise the new vocabulary and read the definitions. After that, they will classify the words into parts of the speech, such as nouns, verbs, adjectives or adverbs.

**Activity 3.-** For this activity, the students will reread the vocabulary and think about someone or something they remember when hearing that word. This activity should be personal and individual for each learner, for instance:

- Toil= when I worked as a builder.
- Gratify= how my neighbor feels when I help her.
- Mourn= when I cried over my dead dog.

**Activity 4.-** The students will work on the *Lextutor* program. Then the teacher will ask them to click on the section of word games- id word: <https://www.lex tutor.ca/id/> . Next, they have to write on the program's searcher the new vocabulary. Thus, a list of statements will appear and they will select the correct word to fill the gap.

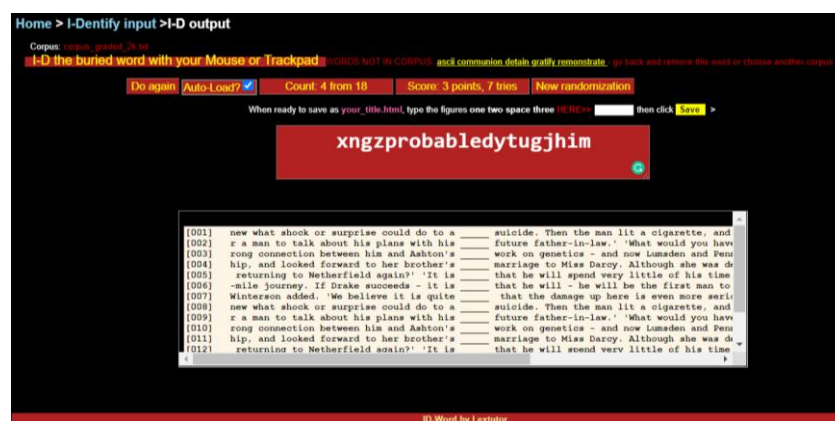


Figure 13. ID word exercise class N. 5 (taken from Lextutor.ca, 2020)

### 3.5.9. "Session 9" Vocabulary class N. 6

**Specific objective of this activity.** To pre-teach the high-frequency keywords obtained from the reading book. To do so, the teacher will use some proved teaching vocabulary strategies,

such as working with synonyms and flashcards. In this way, the learners will be familiar with the new terms, and they will read the book more easily.

### Development of the activity.

**Table 7. Vocabulary class N. 6**

Activities	Length	Goals	Outcomes	Resources
<b>1- Feedback Sentence construction</b>	10'	To review the previous class' vocabulary and provide feedback in order to help students to internalize it.	Students are able to tell a sentence using a word from the previous class' vocabulary.	Blackboard Cellphone Markers
<b>2- Vocabulary flashcards presentation</b>	10'	To present the vocabulary using the <i>Lextutor</i> program.	Students meet the new vocabulary through the use of flashcards in the <i>Lextutor</i> program.	Internet Computers Cellphones
<b>3- Work with synonyms</b>	20'	To make students create mental connections between their previous knowledge and the new one.	Students are able to find synonyms for new words and create pictures or symbols for them.	cardboards Pens Markers
<b>4- Group work</b>	20'	To provide students the opportunity of interacting with their classmates in order to enhance their learning vocabulary.	Students can recognize a word by its synonym and picture.	Flashcards

Table 7. Vocabulary class N. 6. Own elaboration.

**Activity 1.-** At the beginning of the class, the teacher will ask students to write on the board one word they remember from the previous lesson. When there are 20 words, the teacher will use the app named *Who's next* ([https://play.google.com/store/apps/details?id=de.mphil.whosnext&hl=es\\_419&gl=US](https://play.google.com/store/apps/details?id=de.mphil.whosnext&hl=es_419&gl=US)) to call the students one by one. Then they will write one sentence with this word in the first person, for example

- Astonish: What astonished me was her new car.

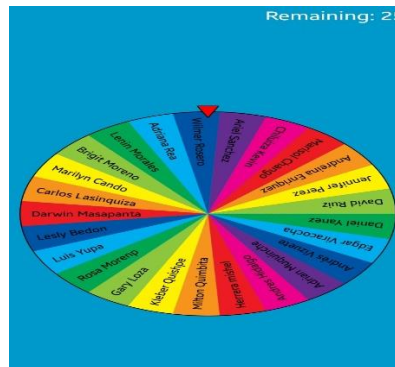


Figure 14. Spinner wheel class N. 6 (taken from Who's next, 2020)

**Activity 2.-** The teacher will present the new vocabulary with the following link. <http://lextutor.ca/flashcards/keywords-6.html> .The students will revise the new vocabulary and read the definitions. After that, they will get in pairs and test their memory, asking questions about the meaning of the words to each other.

**Activity 3.-** For this activity, the students will continue in pairs, and they will create their own flashcards. The teacher will number the teams 1 and 2. Team N. 1 will work with the first ten words, while Team N. 2 will work with the last ten words of the vocabulary provided before. To this purpose, on the first side of the flashcard, the students will write the term; then on the second side, they write a familiar synonym and draw a symbol or picture below, for instance:



Figure 15. Flashcard example class N.6. Own elaboration.

**Activity 4.-** The students will get into four-people groups, joining with another pair, and they will play with the flashcards. Thus, each pair will show the side with the synonym and the picture to the other pair. Then the other pair will say the word that it means. In the end, the winner will be the pair who guesses more words.

**3.5.10. "Session 10" Vocabulary class N. 7**

**Specific objective of this activity.** To pre-teach the high-frequency keywords obtained from the reading book. To do so, the teacher will use a proven teaching vocabulary strategy that is working with mnemonics. In this way, the learners will be familiar with the new terms and later, they will read the book easily.

**Development of the activity.****Table 8. Vocabulary class N. 7**

Activities	Length	Goals	Outcomes	Resources
<b>1- Feedback</b>	10'	To review the previous class' vocabulary and provide feedback in order to help students to internalize it.	Students are able to select the correct definition of the previous class terminology in the Lextutor quiz.	Internet Computers Cellphones
<b>2- Vocabulary flashcards presentation</b>	10'	To present the vocabulary using the <i>Lextutor</i> program and ask them to classify it into parts of the speech.	Students get to know the new vocabulary through flashcards in the <i>Lextutor</i> program and they classify it into nouns, verbs, adjectives or adverbs.	Internet Computers Cellphones Notebook
<b>3- Mnemonics</b>	20'	To encourage students to make mental connections between the words and the graphics.	Students are able to create a graphic, picture or symbol to represent a word.	Notebook Pen Markers
<b>4- Mnemonics</b>	20'	To promote the use of mnemonic tools to enhance the internalization of the new terms.	Students can create a short story applying the new vocabulary and evaluate its correct usage.	Notebook Pen Markers

*Table 8. Vocabulary class N.7 Own elaboration.*

**Activity 1.-** The teacher will ask the students to go to the following link <http://lextutor.ca/flashcards/keywords-6.html> in order to do the multiple-choice quiz about the previous class vocabulary. Consequently, the students will self-evaluate their performance and realize the terms that they have to practice more.

**Activity 2.-** The teacher will present the new vocabulary for this class with the following link <http://lextutor.ca/flashcards/keywords-7.html> The students will review the new vocabulary and read the definitions. After that, they will classify the words into parts of the speech, such as nouns, verbs, adjectives or adverbs.

**Activity 3.-** For this activity, the students will create a mental image of each word from the list above. When they have it ready, they will draw that image next to the word written. If students are not good at drawing, they can use symbols to represent the meaning of the word. In the end, the students will link the image with the term making mental connections that help them to remember it easier.

**Activity 4.-** The students will prepare a short story using the thirteen words of the list. In this case, the plot has to be original but not realistic or logical. The teacher will guide the students by providing an example of a story to give them clues. At the end of the class, they will get into five-people groups, and they will share their stories with their classmates.

### 3.6. Assessment

This section is aimed to evaluate the proposal's process through the activities presented, and at the same time, its effectiveness in solving the reading comprehension problem. For this purpose, formative assessment and summative assessment were considered, and they are described below.

#### 3.6.1. Learning assessment

It is important to mention that the proposal's process is focused on continuous formative assessment. The reason for this choice is that our main aim is to guide the learning process, observing and modifying it day to day to help students grasp the vocabulary to later improve their reading comprehension. To this purpose, assessment for learning and assessment as learning has been applied, and they are explained below.



Firstly, assessment for learning is aimed to provide feedback of the learning process both the teacher and the students in order to improve their performance in the future. Therefore, it will be applied in the seven classes proposed with activities such as multiple-choice quizzes, sentence formation, group works, contents, and games. In these cases, the students will measure how many words they remember easily to fulfill the requirements and realize which words they need to practice more. Additionally, the students will get feedback immediately from the program, from the teacher, or their classmates when they are interacting. Consequently, both the teacher and the students recognize which aspects need to be improved in future classes.

On the other hand, assessment as learning is intended to make students aware of their learning progress by reflecting on their own work. Taking this into account, it will be implemented in the learning diary that the students will write after each vocabulary class. In this diary, they will write how comfortable they feel with the class' dynamic, what activities were useful for them in order to learn vocabulary, and which ones did not work well for them. This diary will be a personal reflection of the learners, making them aware and responsible for their own learning process. The analytical rubric for the learning diary is presented below:

### LEARNING DIARY

**Objective:** To use critical reflection to analyse and gain insight into your learning process throughout the vocabulary classes.

**Instructions:** After each vocabulary class, you will write a reflection about how you feel about the class' dynamic, what activities were useful for you to learn the vocabulary and which ones did not work well. Additionally, you can write constructive comments and questions that you have about the class.

**Table 9. Learning diary analytical rubric.**

LEARNING DIARY	DESCRIPTION	MAXIMUM SCORE (POINTS)	%
<b>Depth of reflection</b>	The response demonstrates an in-depth reflection of their own learning. Their arguments are strong and well analyzed.	5	50 %
<b>Content</b>	The response includes all components. Meets or exceeds all requirements indicated in the instructions.	3	30 %
<b>Use of language</b>	Writing is straightforward, and the sentences are complete and well organized. Thoughts are expressed in a coherent manner. Minor mistakes do not interfere with the reader's comprehension.	1	10 %
<b>Organization</b>	The diary is well organized by days and easy to follow.	0.5	5 %
<b>Presentation</b>	The diary is supported by visual elements like, for example, diagrams, charts, pictures, etc. when appropriate	0.5	5 %
<b>TOTAL</b>		<b>10</b>	<b>100 %</b>

*Table 9. Learning diary rubric. Own elaboration.*

### 3.6.2. Assessment of the proposal

After the seven classes of pre-teaching vocabulary, the learners will read the storybook outside the class on their own. During this state, the first assessment will be implemented in order to evaluate the impact that the proposal will be having on the students' reading comprehension. To do so, assessment as learning will be implemented through the use of a checklist. This will help students to evaluate the strengths and weaknesses of their reading process. Apart from that, the teacher will have a real x-ray of the results that the pre-teaching vocabulary is provoking.

Link to the checklist:

[https://docs.google.com/forms/d/e/1FAIpQLSdk\\_8SjuKmzSGqc7C77RtISOVSEgPpoaMuWtLF0wNdRgfkPPg/viewform?usp=sf\\_link](https://docs.google.com/forms/d/e/1FAIpQLSdk_8SjuKmzSGqc7C77RtISOVSEgPpoaMuWtLF0wNdRgfkPPg/viewform?usp=sf_link) (Adapted from Alvazoglou & Griva, 2014).

On the other hand, the learners will have a summative assessment when they finish reading the book in order to verify how helpful the pre-teaching vocabulary strategy is and the real impact on the improvement of the students' reading comprehension. Bearing this in mind, the teacher will propose three types of assessments, and the learners will be allowed to select which one is most suitable for their abilities and learning style.

The first one is to write a letter addressed to Harriet Jacobs, the author of *Incidents in the life of a slave girl book* to express their like or dislike of the text, ask doubts about the plot, and explain the aspects that they would change in the story. The second one is to illustrate important events of the story with pictures and write a short and concise description of each of them. The last option is to select fragments of songs that better describe the book's main events and write a brief explanation of their choices. To explain it better, all of them are detailed below:

**OPTION 1.- Write a letter addressed to Harriet Jacobs, the author of *Incidents in the life of a slave girl book*.**

**Objective:** To write a letter addressed to Harriet Jacobs, the author of *Incidents in the life of a slave girl book* expressing your like or dislike of the text, ask about the story, and explain the aspects that you would change if you were the author.

**Instructions:** When you finish reading the book, reflect on the story, and answer the following questions. Then prepare the letter, including these requirements:

- Do you like or dislike the book? Explain your choice to the author with strong arguments to support it.
- Do you have questions for the author about her story? Imagine that the author is in front of you, ready to answer all your queries. Then prepare the questions carefully to receive precisely the answer that you expect.
- If you were the author, would you change any aspect or event of the story? Analyze which specific parts of the story you would like to change and why you want to do it (to make the plot more interesting, make it more dramatic, give a happy ending, etc.).
- The letter can be presented in written or digital format. However, it must be an original creation of the author.

**Table 10. Letter analytical rubric.**

WRITING A LETTER	DESCRIPTION	POINTS	%
<b>Depth of reflection</b>	The letter's arguments show a deep understanding of the story. It denotes a strong personal reflection of the plot. The questions are specific and well-directed. The changes are carefully analyzed and have specific purposes.	5	50 %
<b>Content</b>	The response includes all components detailed before. Meets or exceeds all requirements indicated in the instructions.	3	30 %
<b>Use of language</b>	Writing is straightforward and easy to follow. The vocabulary is appropriated to their level. Sentences are complete and well organized. Thoughts are expressed in a coherent manner.	1	10 %
<b>Organization</b>	The letter is well organized, natural, and easy to follow.	1	10 %
<b>TOTAL</b>		<b>10</b>	<b>100 %</b>

Table 10. Letter analytical rubric. Own elaboration.

**OPTION 2.- Illustrate the main events of the book with pictures.**

**Objective:** To illustrate the main events of the story with referential pictures and write a brief and concise description of each one.

**Instructions:** When you finish reading the book, reflect on the relevant events of the story and visualize images that describe each of them, creating a clear overview of the story. To this purpose:

The pictures have to be clear enough and tell the story to a person who has not read the book yet.

- Each picture has to have a brief written description that describes the situation or part of the story.
- The pictures must have a precise sequence without distractors.
- The final product can be presented physically or in a digital format. However, the pictures presented must be an original creation of the author.

**Table 11. Illustration analytical rubric.**

WRITING A LETTER	DESCRIPTION	POINTS	%
<b>Depth of reflection</b>	The illustration shows a deep understanding of the story. The plot is appropriately telling in an illustrative way. The pictures describe the main events clearly and in a coherent sequence.	5	50 %
<b>Content</b>	The illustration includes all components detailed before. Meets or exceeds all requirements indicated in the instructions. No important event has been omitted.	3	30 %
<b>Use of language</b>	The writing description is straightforward and easy to follow. The vocabulary is appropriated to their level. Sentences are complete and well organized. Thoughts are expressed in a coherent manner.	1	10 %
<b>Organization</b>	The illustrated story is well organized, natural, and easy to follow.	1	10 %
<b>TOTAL</b>		<b>10</b>	<b>100 %</b>

Table 11. Illustration analytical rubric. Own elaboration.

**OPTION 3.- Select fragments of songs that best describe the book's main events**

**Objective:** To select fragments of songs that better describe the book's main events and write a brief explanation of their choices.

Instructions: When you finish reading the book, think about fragments of songs that best match the story's main events. To this purpose:

- The songs' pieces must have lyrics that clearly describe a situation or the characters' feelings in the story.
- The songs' fragments have to be presented logically, following the plot of the story.
- The written explanation of the fragment choice should be supported by strong arguments based on the story.
- The final product should be presented in a video where the songs and the written part appear.

**Table 12. Selecting fragments of songs analytical rubric.**

WRITING A LETTER	DESCRIPTION	POINTS	%
<b>Depth of reflection</b>	The selection of the songs' fragments shows a deep understanding of the story. The pieces clearly describe a situation or characters' feeling that occurs in the plot. The written explanation of the fragment choice is supported by strong arguments based on the story.	5	50 %
<b>Content</b>	The songs' fragments have valuable lyrics that describe each event clearly. They meet or exceed all requirements indicated in the instructions. No significant event has been omitted.	3	30 %
<b>Use of language</b>	The writing description is straightforward and easy to follow. The vocabulary is appropriated to their level. Sentences are complete and well organized. Thoughts are expressed in a coherent manner.	1	10 %
<b>Organization</b>	The songs' fragments are presented logically, following the plot of the story.	1	10 %
<b>TOTAL</b>		<b>10</b>	<b>100 %</b>

*Table 12. Selecting fragments of songs analytical rubric. Own elaboration.*

## 4. Discussion

This study is intended to solve two well-known problems: the lack of vocabulary and the low reading comprehension level in EFL learners. To this purpose, we propose to carry out a corpus-based analysis to obtain the high-frequency keywords contained in a storybook. Thus, the teacher will teach this keywords list before the learners start reading the book. Bearing this in mind, we will discuss the contributions that this proposal could bring.

First of all, the benefits of using storybooks instead of graded or textbooks are undeniable. Hsieh et al. (2011) affirm that storybooks allow students to encounter natural language rich in and headwords and content words such as nouns, adjectives, adverbs, and verbs, which other books cannot offer. As a result, the students can access a rich and authentic lexicon that will increase their language knowledge. For this purpose, a set of story books were chosen by the teacher, and a survey to choose one of them was applied to the potential proposal participants. It showed that the storybook preferred to read is: *The incidents in the life of a slave girl*. In addition to the linguistic benefits, by allowing students to choose a reading book, it is also more probable that they are willing to read because they will be interested in the story presented. For this reason, it is essential that the teacher provides books that fulfill students' expectations, taking into account their interests (Wang & Lee, 2007, cited by Hsieh et al., 2011).

On the other hand, pre-teaching vocabulary is not a new issue of study. Cowell, in 2012, in her research *Pre-teaching vocabulary to improve comprehension of a narrative text*, proved that this strategy could contribute positively to improve the students' reading comprehension because the learners pay closer attention to the content while they are reading. The reason is they do not get overwhelmed due to the number of unknown words in the text; hence, the time employed looking their meanings up is reduced as well. Grabe and Jiang in 2013 enlisted some factors that impact reading abilities; the most notable are efficient word recognition and vocabulary knowledge. The pre-teaching vocabulary strategy will enhance these abilities' acquisition and provide learners the tools to understand a text more easily (Cowell, 2012).

Additionally, the correct selection of strategies to teach vocabulary plays a vital role in obtaining the expected results (Schmitt, 1997). What is more, applying a variety of teaching vocabulary strategies will allow students to meet the words in different ways. As a result, in

any of these encounters, the term will become familiar to the learners (Nation & Waring, 1997). Therefore, in order to prepare the classes, the teacher carried out the selection of teaching vocabulary strategies carefully. The literature review showed different strategies presented by well-known authors. However, not all of them were used because they were not suitable for the students' level, age, and interests. The strategies selected were: word formation, synonyms and antonyms, suffixes and prefixes, spelling, mnemonics, guessing from context, and word parts.

Furthermore, another essential factor to bear in mind is the vocabulary type to teach and the parameters to choose it. According to Nation (1997), it is better to teach high-frequency vocabulary because it will continuously appear along with the whole text. If the students recognize these words beforehand, their reading rate will increase, and they will have better overall comprehension. For this reason, the corpus-analysis of the book was carried out appropriately, thanks to the *Lextutor* program. The use of the *KeyWords* and the *Text Lex Compare* tools allowed us to obtain the high-frequency word list and terms with significant appearances in a well-known 1-million English Corpus developed by Paul Nation (1995).

Finally, the assessment proposed will make students aware of their own learning process. In other words, the learning diary and the checklist will make learners analyze if the proposal had an impact on their vocabulary grasp and reading process. Meanwhile, the summative assessment will evaluate the students' reading comprehension more deeply because they cannot just copy a summary; they have to create a product that shows their understanding.

## 5. Conclusions

This study aimed to help students increase their grasp of vocabulary to improve their reading comprehension. The proposal was to carry out a corpus-based analysis of the lexis in a reading book and pre-teach the relevant high-frequent vocabulary obtained. At the end of the study, we can conclude that:

Firstly, it can be said that the proper investigation of the main variables provided the teacher with a broader understanding of the concepts required for this proposal. Some aspects that at the beginning were not contemplated are considered now. For instance, the importance of not only looking for high-frequent vocabulary because it mainly contains function words, and



to look for content words commonly used in the Real English corpus. Additionally, the literal framework permitted establishing connections between the pre-teaching vocabulary strategy and its positive influence in the reading comprehension process through similar studies with positive results.

Furthermore, it was proved that *Lextutor* is an excellent program for teachers and learners because it contains various activities to learn the English language working interactively with corpora. This program was useful in order to identify the high-frequency word list from the book's corpus and to provide activities to practice the vocabulary in context. However, it is the teacher's job to verify that the final wordlist to teach is accurate and to select the most useful exercises that contribute effectively to the learning process.

On the other hand, the teaching vocabulary strategies were carefully chosen from studies that were carried out by authors with broader experience in teaching vocabulary. Some of them are innovative, but others are generally known by English teachers. Therefore, it was concluded that a great part of the vocabulary strategies' effectiveness depends on how the teacher presents and applies them in the class. Thus, it must be considered the students' needs and age, and the vocabulary selected.

To sum up, it was concluded that to carry out a corpus analysis is beneficial to discrete the vocabulary that plays an essential role in the text. If this list is taught with various vocabulary strategies that cover students' necessities, it will increase their vocabulary knowledge. Moreover, introducing these keywords before the reading process can booster a higher reading comprehension at the end.

## 6. Limitations and further research

One of the most significant limitations to develop reading comprehension is the lack of reading habit in Ecuador. It is a cultural problem that needs to be treated from an early age. Besides, working with the intrinsic and extinct motivation is another factor that needs to be included to improve the reading skill acquisition.

Another limitation to teach vocabulary is the time. As previously exposed, a large amount of time is required to implement the pre-teaching proposal. Unfortunately, most of the English classes have a maximum of six hours a week. Additionally, the curriculum is mostly based on

grammar content, and it does not pay enough attention to developing other skills. Therefore, time is not enough to cover the curriculum content and extra vocabulary classes.

On the other hand, it's undeniable that pre-teaching vocabulary is an essential resource to improve reading comprehension, but not the only one. As comprehending a text is a cognitive process, it needs to have further instruction that helps learners develop skills to decode the text's information.

In conclusion, it could be suitable to apply this proposal as experimental research in order to measure its real effectiveness in vocabulary grasp and reading comprehension. Further investigations should then be focused on implementing more strategies to improve EFL learners' reading comprehension that complement this study.

## References

- Abduh, A., & Rosmaladewi, R. (2017). Taking the Lextutor online tool to examine students' vocabulary levels in business English students. *World Transactions on Engineering and Technology Education*, 15(03), 283-286.
- Aivazoglou, E., & Griva, E. (2014). Reading skills and strategies: Assessing primary school students' awareness in L1 and EFL strategy use. *International Journal of Applied Linguistics & English Literature*, 3(5), 239-250.
- Al-Gamal, A. A. M., & Ali, E. A. M. (2019) Corpus-based Method in Language Learning and Teaching. *IJRAR- International Journal of Research and Analytical Reviews*, 6(2), 473i-476i. Retrieved from: [www.researchgate.net/publication/332876550\\_Corpus-Based\\_Method\\_in\\_Language\\_Learning\\_and\\_Teaching](http://www.researchgate.net/publication/332876550_Corpus-Based_Method_in_Language_Learning_and_Teaching)
- Al-Mahbashi, A., Noor, N. M., & Amir, Z. (2015). The Effect of Data-Driven Learning on Receptive Vocabulary Knowledge of Yemeni University Learners. *3L: The Southeast Asian Journal of English Language Studies*, 21(3), 13 -24.
- Beck, I. L., Perfetti, C. A., & McKeown, M. G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology*, 74(4), 506–521.
- Bennett, G. R. (2010). *Using corpora in the language learning classroom: Corpus linguistics for teachers*. University of Michigan Press.
- Berg, E., & Kit, M. (2019). Reference and Educational System LexSite-LexTutor as Information Technology for Foreign Language Learning. *CEUR Workshop Proceedings*, 25(2), 1-14.
- Biber, D., & Reppen, R. (2002). What does frequency have to do with grammar teaching? *Studies in Second Language Acquisition*, 24(2), 199–208.
- Chujo, K., & Oghigian, K., 2009. How many words do you need to know to understand TOEIC, TOEFL & EIKEN? An examination of text coverage and high-frequency vocabulary. *The Journal of Asian TEFL* 6 (2), 121-148.
- Cobb, T., 1998. Why & How To Use Frequency Lists To Learn Words. [online] *Lextutor.ca*. Retrieved from: <https://www.lextutor.ca/research/>

- Cowell, L. (2012). *Pre-teaching vocabulary to improve comprehension of a narrative text* (Doctoral dissertation). Auburn University, Auburn, Alabama.
- English Corpora: most widely used online corpora. Billions of words of data: free online access. (2020). Retrieved November 2020, from <https://www.english-corpora.org/coca/>
- Floris, F D; Divina, M. (2015). A study on the reading skills of EFL university students. *Teflin Journal*, 20(1), 37-47.
- Ghazal, L. (2007). Learning vocabulary in EFL contexts through vocabulary learning strategies. *Novitas-ROYAL*, 1(2), 84-91.
- Grabe, W. & Jiang, X. (2013). Assessing reading. *The companion to language assessment*, 1(1), 185-200. Retrieved from <https://onlinelibrary.wiley.com/doi/full/10.1002/9781118411360.wbcla060>
- Gu, Y., & Johnson, R. K. (1996). Vocabulary learning strategies and language learning outcomes. *Language Learning* 46 (4), 64 – 79.
- Hart, M. (2004). Project Gutenberg mission statement. *Project Gutenberg*. Retrieved from [http://www.gutenberg.org/about/background/mission\\_statement.html](http://www.gutenberg.org/about/background/mission_statement.html)
- Herazo, Z. D. (2015). La lectura y sus tipos. *Portal de las Palabras*, 1(1), 9-13.
- Hsieh, M. Y., Wang, F. Y., & Lee, S. Y. (2011). A corpus-based analysis comparing vocabulary input from storybooks and textbooks. *The International Journal of Foreign Language Teaching*, 6(1), 25-33.
- Iñiguez, J. K. (2015) La motivación por la lectura en el Ecuador fomentando nuevas ideologías. *Atlante: Cuadernos de Educación y Desarrollo*, 1(2),1-13. Retrieved from <http://atlante.eumed.net/wp-content/uploads/lectura1.pdf>
- McCarthy, D., Koeling, R., Weeds, J., & Carroll, J. A. (2004). Finding predominant word senses in untagged text. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics*, 1(4), 279-286.
- McCrostie, J. (2007). Examining learner vocabulary notebooks. *ELT Journal: English Language Teachers Journal*, 61(3), 246-255.

- McEnergy, T., & Hardie, A. (2011). *What is corpus linguistics? In Corpus Linguistics: Method, Theory and Practice (Cambridge Textbooks in Linguistics, pp. 1-24)*. Cambridge University Press. DOI:10.1017/CBO9780511981395.002
- Molavi, A., Koosha, M., & Hosseini, H. (2014). A comparative corpus-based analysis of lexical collocations used in EFL textbooks. *Latin American Journal of Content & Language Integrated Learning*, 7(1), 66-81.
- Mousavian, S., & Siahpoosh, H. (2018). The effects of vocabulary pre-teaching and pre-questioning on intermediate Iranian EFL learners' reading comprehension ability. *Global Journal of Foreign Language Teaching*, 8(3), 94-103.
- Nation, P. (1990). *Teaching and learning vocabulary*. New York: Newbury House.
- Nation, P., Lufer, B., & Cobb, T. (1995). *Compleat Lexical Tutor*. Retrieved November 2020, from <https://www.lextutor.ca/>
- Nation, P., & Newton, J. (1997). Teaching vocabulary. *J. Coady and T. Huckin (eds.) In Second Language Vocabulary Acquisition (pp. 238-254)* Cambridge University Press, Cambridge.
- Nation, P., & Waring, R. (1997). Vocabulary size, text coverage, and word lists. N. Schmitt and M. McCarthy (eds.). In *Vocabulary: Description, Acquisition and Pedagogy* (pp. 6-19) Cambridge University Press, Cambridge.
- Nation, I. S. P. (2009). *Teaching ESL/EFL Reading and Writing*. London & New York: Routledge.
- Nation, I.S.P. & Meara, P. (2010) Vocabulary. In N. Schmitt (ed.) *An Introduction to Applied Linguistics* (pp. 34-52). Edward Arnold. Second edition.
- Nordquist, Richard. (2019). Definition and Examples of Corpus Linguistics. Retrieved from <https://www.thoughtco.com/what-is-corpus-linguistics-1689936>
- O'Keeffe, A. (2012) "*Vocabulary Instruction*". In A. Burns and J. Richards (Eds) *The Cambridge Guide to Pedagogy and Practice in Language Teaching* (pp. 236-245). Cambridge: Cambridge University Press,
- Open American National Corpus | Open Data for Language Research and Education. (2020). Retrieved November 2020, from <http://www.anc.org/>

- Richards, J. C. (1976). *The role of vocabulary teaching*. TESOL Quarterly, 10 (1), 77.
- Salvador, S. P. (2005). La comprensión de textos académicos en un contexto cooperativo guiado a través de la enseñanza recíproca. *Electronic Journal of Research in Educational Psychology*, 3(5), 77-96. Retrieved from: [http://www.investigacion-psicopedagogica.org/revista/articulos/5/espanol/Art\\_5\\_60.pdf](http://www.investigacion-psicopedagogica.org/revista/articulos/5/espanol/Art_5_60.pdf)
- Schmitt, N. (1997). Vocabulary learning strategies. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: description, acquisition and pedagogy* (pp.199-228). Cambridge: Cambridge University Press.
- Sevier, M. (2004). The compleat lexical tutor, v. 4. *TESL-EJ*, 8(3).
- Stroube, B. (2003). Literary freedom: Project Gutenberg. *XRDS: Crossroads, The ACM Magazine for Students*, 10(1), 3-3.
- The British National Corpus, version 3 (BNC XML Edition). 2007. Distributed by Bodleian Libraries, University of Oxford, on behalf of the BNC Consortium. Retrieved from: <http://www.natcorp.ox.ac.uk/>
- Tighe, E. L., & Schattschneider, C. (2016). Examining the relationships of component reading skills to reading comprehension in struggling adult readers: A meta-analysis. *Journal of learning disabilities*, 49(4), 395-409.
- Tognini-Bonelli, E. (2001). *Corpus linguistics at work* (Vol. 6). John Benjamins Publishing.
- Villamarín, G. M. (2016). *La enseñanza recíproca (reciprocal teaching) en el desarrollo de la comprensión lectora del idioma inglés en los estudiantes del octavo año de educación básica de la Unidad educativa "Victoria Vásquez Cuví-Simón Bolívar-Elvira Ortega" de la ciudad de Latacunga, provincia de Cotopaxi* (Bachelor's thesis) Universidad Técnica de Ambato, Ecuador.

## Annexes

### Annex 1. Survey applied to the potential participants of the proposal.



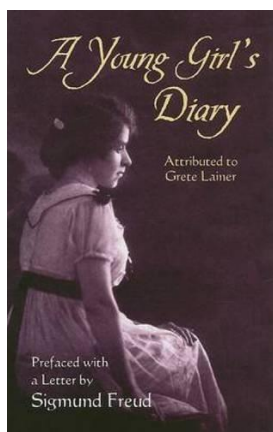
## READING BOOK SURVEY

### ENGLISH CLASS V-VI LEVEL

**OBJECTIVE:** To provide the learners with a range of five reading books from the Gutenberg project. They are in charge of choosing one of them for the whole group through a survey. The books are proposed by the teacher, taking into consideration their current level.

**INSTRUCTIONS:** Dear students, in this survey, I propose you five reading books from the Gutenberg Project, you will read each one's description, and you need to choose the one that has caught your attention. In the end, the book which has more votes will be the one that we will read as a whole class.

#### A young girl's diary



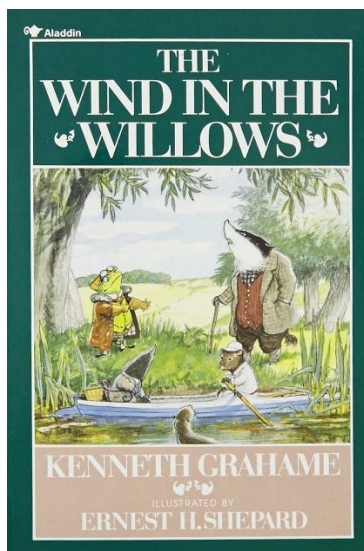
#### Author: Sigmund Freud

This is the diary of a Viennese teenager from the upper-middle-class called Grete Lanier. She and her best friend agree to write journals and exchange them as letters. The story started in 1910, and it covers since Grete is 11 until she turns out 14 more or less. Grete shares her

school experiences, her relationship with their family and friends, and her typical concerns of a girl who is becoming a teen. The author Sigmund Freud claims that this diary is authentic, and he did not edit it. That's why it starts with a letter of him explaining the content. " Never before, I believe, has anything been written, enabling us to see so clearly into the soul of a young girl... We are shown how the sentiments pass from the simple egoism of childhood to attain maturity..." (Freud, 1919, p1)

Link to the book: <https://www.gutenberg.org/files/752/752-h/752-h.htm>

### **The wind in the willows**



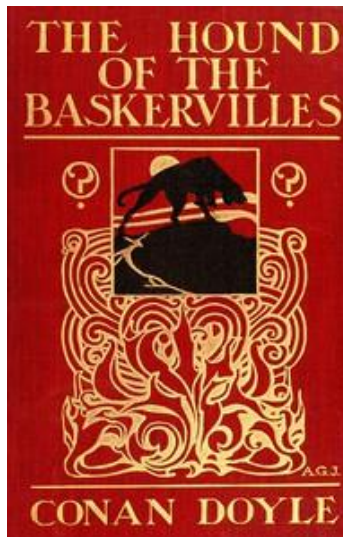
**Author: Kenneth Grahame**

This is the tale of four animal friends in the countryside: Toad, Rat, Mole, and Badger. They think and act like humans, but they keep their instinct animal. The story starts when a humble and homely Mole leaves to walk to the river park; he meets an easygoing and friendly Rat. Then they go to Toad hall and meet its owner, Toad, and spoil and impulsive millionaire. All together decide to visit a clever and isolated Badger. The four friends start an adventure with a series of events that are provoked and solved by their temperament. In the way, they learn about friendship, solidarity, and maturity.

Link to the book: <https://www.gutenberg.org/files/289/289-h/289-h.htm>



## The Hound of the Baskervilles

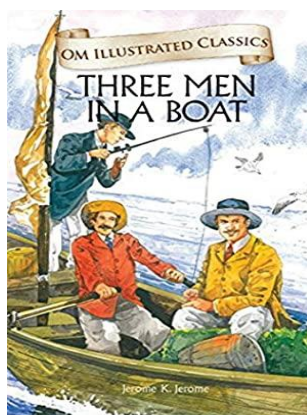


**Author: Conan Doyle**

The story is set in Devonshire, England. In 1889, Sir Charles Baskerville was found murdered under strange circumstances. The legend claims that the Baskerville family has a curse started by an ancestor who sold his soul and was killed by a hellish hound as punishment. People from the town claim that a frightening hound appears in the moor, and it was howling at Sir Charles' grove at night. Sherlock Holmes and Dr. Watson are called to investigate the mystery and protect Sir Henry Baskerville, Sir Charles's nephew, and heir. Sir Henry receives a letter warning him to keep away from the moor. Holmes and Watson pass through a series of events until they find the real assassin.

Link to the book: <https://www.gutenberg.org/files/2852/2852-h/2852-h.htm>

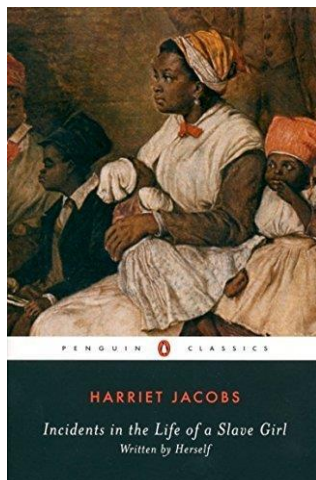
## Three men in a boat



**Author: Jerome K. Jerome**

This is the humored story of three good friends Jerome, Harrys, and George. One day, they get together and talk about the multiple illnesses they think they have; in a discussion, they agree that a great way of dealing with it is having a holiday trip. After that, they came up with the idea of preparing a boat trip over the Thames river. The three comrades and a dog called Montmorency start the trip carrying the essential things to survive and sleep in the boat. They sail from Kingston upon Thames river to Oxford, passing landmarks and villages where they camp, visit, and get new experiences while hearing Jerome's anecdotes.

Link to the book: <https://www.gutenberg.org/files/308/308-h/308-h.htm>

**Incidents in the life of a slave girl****Author: Harriet Jacobs**

The story is related by Linda Brent (Harriet Jacobs), a slave girl in the 80s. Linda describes her early happy childhood with her parents and her brother and how her grandmother got her freedom. At the age of six, her mother died, and she went to live with her mother's mistress, with whom she learns to write and read. When the mistress died, and she is willed to Dr. Flint to start the real calvary. In the rest of the story, she describes the abuses that she suffered, the forbidden love that she had, how she tried to protect her children, and how finally she got their freedom after several years.

Link to the book: <https://dev.gutenberg.org/files/11030/11030-h/11030-h.htm>

**Which book do you select to read?**

1. A young girl's diary
2. The wind in the willows
3. The Hound of the Baskervilles
4. Three men in a boat
5. Incidents in the life of a slave girl

**Annex 2 Final word list obtained from the *KeyWords* and the *Text- Lex Compare* tools of the Lextutor program.**

- The number next to the word means how many times it appears in the text.
- The highlighted words were eliminated from the list to teach because they were cognates, names, or well-known terms.

001. slave 391	020. kneel 18	039. chile 10
002. grand 240	021. limb 18	040. converse 10
003. flint 170	022. oppress 18	041. especial 10
004. mistress 110	023. tremble 18	042. piazza 10
005. plantation 51	024. weep 18	043. probable 10
006. exclaim 36	025. archive 14	044. runaway 10
007. fanny 32	026. bestow 14	045. sympathise 10
008. clothe 30	027. storeroom 14	046. clasp 9
009. nigger 30	028. toil 14	047. emancipate 9
010. conceal 27	029. accustom 13	048. entreat 9
011. meaning 27	030. degrade 13	049. peril 9
012. situate 26	031. procure 13	050. staid 9
013. fugitive 25	032. relation 13	051. strive 9
014. persecute 22	033. trademark 12	052. tyrant 9
015. lash 21	034. amen 11	053. vicinity 9
016. negro 20	035. summon 11	054. bondage 8
017. wretch 20	036. swamp 11	055. deem 8
018. compel 19	037. acquaint 10	056. flog 8
019. rejoice 19	038. blight 10	057. footstep 8

## Implementation Of The Corpus Analysis Of The Lexis Contained In An English Reading Book With The Aim Of Pre-Teaching Vocabulary

058. impart 8	083. kindle 6	108. scorch 5
059. inflict 8	084. mortify 6	109. scourge 5
060. ingratitude 8	085. nourish 6	110. solicit 5
061. loophole 8	086. orphan 6	111. terrify 5
062. startle 8	087. revolve 6	112. abate 4
063. chattel 7	088. shingle 6	113. allude 4
064. contrive 7	089. suspicious 6	114. aperture 4
065. episcopal 7	090. trample 6	115. bereave 4
066. filename 7	091. watchful 6	116. bewilder 4
067. forsake 7	092. ascii 5	117. christen 4
068. insolent 7	093. astonish 5	118. clinch 4
069. insurrection 7	094. baptism 5	119. coax 4
070. pike 7	095. communion 5	120. complaints 4
071. pious 7	096. compute 5	121. compress 4
072. proclaim 7	097. defraud 5	122. condescend 4
073. sabbath 7	098. detain 5	123. console 4
074. tiding 7	099. detest 5	124. contaminate 4
075. vigil 7	100. dissatisfy 5	125. deportment 4
076. compile 6	101. elude 5	126. dissipate 4
077. confide 6	102. gratify 5	127. equip 4
078. disclaim 6	103. meditate 5	128. exasperate 4
079. expend 6	104. mourn 5	129. exult 4
080. forbear 6	105. profess 5	130. grievous 4
081. housemaid 6	106. rejoin 5	131. grope 4
082. humiliate 6	107. remonstrate 5	132. heathen 4

133. inculcate 4	139. recount 4	145. stormy 4
134. inhabitant 4	140. refine 4	146. stow 4
135. mulatto 4	141. seafarer 4	147. twine 4
136. mutter 4	142. shutter 4	148. yearn 4
137. odious 4	143. sinful 4	
138. politic 4	144. spirited 4	