Universidad de Valladolid

FACULTAD de FILOSOFÍA Y LETRAS

DEPARTAMENTO de FILOLOGÍA INGLESA

Grado en Estudios Ingleses

TRABAJO DE FIN DE GRADO

Text Inspector corpus linguistics tool on trial: Checking accuracy for students' writings assessment

Alicia Martín Rodríguez

Tutor: Beatriz Méndez Cendón

Curso: 2022-2023

ABSTRACT

Digital tools are increasingly present in education not only to enhance teaching but also to assist educators with lesson planning and students' assessment. This undergraduate dissertation defends the use of corpus linguistics tools by language teachers to carry out their work more efficiently. In fact, the dissertation's main objective is to test one of these applications called *Text Inspector* to find out if English teachers could use it to evaluate the accuracy of students' writings. To this end, corpora compiled from undergraduate dissertation abstracts of students in Engineering, Business Administration and Early Childhood Teaching at University of Valladolid (Uva) have been introduced in the software, which automatically determines the Common European Framework of Reference for languages (CEFR) level of each group. Then, some metrics have been applied to the data to scientifically validate the reliability of the tool, finding some limitations.

Keywords: Corpus linguistics tools, *Text Inspector*, abstracts, metrics, Common European Framework of Reference for languages (CEFR), students' writings assessment

RESUMEN

Las herramientas digitales se incluyen cada vez más en Educación, no sólo para mejorar la enseñanza, sino también para planificar las clases y puntuar a los alumnos. Este trabajo final de grado defiende el uso de herramientas de lingüística de corpus por parte de los profesores de idiomas para trabajar de forma más eficiente. De hecho, el objetivo principal del mismo es probar una de estas aplicaciones, llamada *Text Inspector*, para averiguar si los profesores de inglés podrían utilizarla para evaluar los escritos de sus alumnos. Para ello, se han introducido en el software corpus compilados a partir de *abstracts* de trabajos finales de grado de estudiantes de Ingeniería, Administración de Empresas y Educación Infantil de la Universidad de Valladolid (Uva), determinando automáticamente para cada grupo su nivel del Marco Común Europeo de Referencia para las lenguas (MCER). A continuación, se han aplicado algunas métricas a los datos para validar científicamente la fiabilidad de la herramienta, descubriendo algunas limitaciones.

Palabras clave: Herramientas de lingüística de corpus, *Text Inspector*, *abstracts*, métricas, Marco Común Europeo de Referencia para las lenguas (MCER), evaluación de los textos escritos de los estudiantes

TABLE OF CONTENTS

LIST OF ACRONYMS USED

AWL: Academic Word List

BNC: British National Corpus

CEFR: Common European Framework of Reference for languages

CL: Corpus Linguistics

COCA: Corpus of Contemporary American English

EFL: English as a Foreign Language

ELT: English Language Teaching

EVP: English Vocabulary Profile

FKG: Flesh-Kincaid Grade

FRE: Flesh Reading Ease

GFI: Gunning Fox Index

ICTs: Information and Communication Technologies

KVL: Knowledge-based Vocabulary Lists

MTLD: Measure of Textual Lexical Diversity

POS: Part-of-Speech

TTR: Type-token ratio

UVa: University of Valladolid

VOCD: Vocabulary Diversity

FIGURES

TABLES

INTRODUCTION

Technology is driving major improvements in all sectors. But when it comes to education and language teaching, has its full potential been exploited? There are many educators who want to use all the innovative digital tools available, or to try any new associated methodology, but it is not a question of employing more widgets or distinct related methods but a matter of thinking critically which mechanisms can be implemented to work more efficiently as teachers.

Specifically, what has inspired this work is the fact that many language teachers spend much of their free time doing classwork because they often need to find extra texts in addition to those already found in the chosen textbooks, according to their individual students' needs. In addition, it is increasingly common that educators want to obtain authentic texts as they can be very useful for today's learners to practice with real-life contexts. Personally, I think it is difficult to determine whether these texts match the desired language level of the learner or of the groups of learners for whom they are intended.

A few months ago, I came across *Text Inspector*, a corpus linguistics tool which was developed by the professor of Applied Linguistics, Stephen Bax, in 2011. It consists of a "text checker" for English, which according to the web page, is trusted by relevant international institutions such as the University of Cambridge. By simply entering an English text into the program, it performs an automatic analysis based on more than 200 metrics such as "readability", "lexical diversity" or "lexical sophistication" and shows the general text level in terms of the A1-C2 scale of the Common European Framework of Reference for languages (CEFR).

Their developers assert that *Text Inspector* can be used not only to search accurate readings or listening texts by selecting "reading" or "listening" from the analysis options, but also to evaluate students' work by choosing the "writing" mode. I found this last function very interesting as assessment is another common task that language teachers need to do quickly but fairly, what is difficult due to the high number of elements implied in the diagnosis of learners' writings.

Therefore, the general aim of this undergraduate dissertation is to investigate if *Text Inspector* is truly able to figure out the exact CEFR level of any writing and, if so, to propose the use of this corpus linguistics tool as a way of easily solving the teacher problem previously mentioned.

So, I will analyze some students' writings with the tool to extract its CEFR level, and looking to the results, I will establish and apply some statistical measures to be able to scientifically validate *Text Inspector*.

As I have no students' writings of my own to use, I have chosen to assess the CEFR level of several undergraduate dissertation abstracts in English elaborated by students at the University of Valladolid (Uva) during the academic year 2021-2022, dissertations which I can easily find under a Creative Commons license in the Uva Repository.

The undergraduate dissertation abstracts in English will be taken from three different groups of Uva students from the degrees of Engineering, Business Management and Early Childhood Teaching to introduce them, group by group, within *Text Inspector* and to find out which CEFR level the dissertation abstracts of students of each discipline have. To do that, I will compile three corpora consisting of these dissertations abstracts of those three different degrees.

The reason why this analysis with *Text Inspector* will be carried out with the writings of up to three groups of students, and not just one, will be for the study to be scientifically valid by proving that the tool is able to evaluate different types of texts. Likewise, for each type of text, very specific characteristics will be expected, since I have selected very different groups representing the scientific-technological field, the field of economic sciences, and the field of education.

Theoretically, it is expected that all three groups of students will be classified *by Text Inspector* as proficient users (C) of the language, since they are university students about to finish their undergraduate degree who, although they are not asked to provide proof of this, are expected to have a high level of English in order to be able to handle the scientific and academic knowledge of their respective disciplines, and to write a good dissertation abstract. Furthermore, the engineers' group would score higher than the business administration group and the business administration group would score higher than the early childhood education group, since the engineers would have a high level of terminology specific to their field and would be more precise in their use of lexis than the others with a lower level of technical vocabulary in their fields.

These theoretical hypotheses that would be related to an expected theoretical efficiency for *Text Inspector* will be checked on a practical level by setting certain statistical measures that will confirm the theoretical validity of the tool previously defined. As the groups of texts are so different, the same control variables for the three types of texts will be first set so that these will not be a source of error when carrying out the study.

Hence, this is an empirical study that would consist of a theoretical part with the argued hypotheses and a practical part carried out with the establishment and application of various statistical measures to the quantitative results of the analysis made by *Text Inspector*.

The research questions that I will be answering at the end of this undergraduate dissertation are the following:

1. Is *Text Inspector* able to appropriately assess the CEFR level of the students' dissertation abstracts of different disciplines?
2. If so, is *Text Inspector* so precise and user-friendly for teachers to be suggested to evaluate their students' writings?

By doing this project, I believe that I can put into practice some of the knowledge acquired during the degree not only in the subject of "Information and Communication Technologies applied to English Studies", subject on which my undergraduate dissertation is based, but also in those of "Corpus Linguistics" and "Academic English". In addition, the basics of data analysis learned while elaborating it could be very useful to me in my future career within the education field.

Even if this paper is challenging for me, I also consider it interesting because it is innovative and rewarding. As *Text Inspector* is not a corpus linguistics tool with many years of experience, there is a lack of academic information about it, so my undergraduate dissertation is intended to fill that literary gap. Furthermore, I would like to encourage teachers to study the advantages and disadvantages of this and other corpus linguistics tools, and to use them as they can be of great benefit to themselves and their classes.

The work is structured in six chapters. Chapter one is devoted to the theoretical framework in which four sections are included: "Technological trends in English Language Teaching", "Corpus Linguistics for English language teachers", "Some corpus linguistics tools for English Language Teaching", and "The Common European Framework of Reference for Languages". Chapter two introduces the state of the art with some reference works which underpin my exposition. Chapter three explains the methodology used in this dissertation to carry out the analysis. Chapter four and five deal with the results and the discussion, respectively. And finally, chapter six offers some conclusions and considers further study.

1. Theoretical background

1.1. Technological trends in language teaching

Technological transformation has been constant over the last few decades. A key role in the development of these emerging technologies has been played by the significant changes on the Internet. Web l.0 was described as "one-way communication, a lecture or a monologue" (McLeod & Vasinda, 2008, p. 4) whereas web 2.0 is more like "a dialogue, an engaging class discussion or two-way communication" (McLeod & Vasinda, 2008, p. 4). This means that web 2.0 allows any type of user to share information and to collaborate with others in order to enhance the available knowledge.

These improvements have a major impact on the world of education, and, in particular, on Language Teaching. I will see these effects, first, from the point of view of language students, and second, from the perspective of language teachers.

The repercussion for language learners is huge. Firstly, web 2.0 propitiates “a more engaging, interactive and motivating learning environment” (Başal & Aytan, 2014, p. 372). Students are more eager to learn in an ambience they like and to which they are widely accustomed nowadays. Hence, the classroom is open 24/7. The activities available are much more varied to practice. Blogs, wikis, podcasts, social networks, or other web 2.0 tools permit a great deal of interaction with natural linguistic data, which enhances the knowledge of a particular language (Krashen’s (1982) Second Language Acquisition theory).

Secondly, and in line with new trends in learning theories such as constructivism (Piaget, 1964) and in educational methodologies, such as active learning (Dewey, 1916), technology provides more students-oriented classrooms by giving them the responsibility of their own learning (Russel & Sorge, 1998). The role of the learners as simply passive receptors of the information given by the teacher in the traditional way of vocabulary lists and grammar points is slowly but steadily receding. Students who are involved in practicing on their own the topics they are interested in have better results.

Now looking to the language teacher's perspective, his role turns more into a guide and a monitor of the learning process. Nonetheless, that does not mean it is passive. Technology is not the panacea, and the teacher needs to plan, design, and implement the correct drills to achieve the syllabus objectives: "Integrating technology into teaching requires the combination of adequate technical skills and sound pedagogical foundations" (Simon, 2008, p. 1). In any case, it would be impossible to integrate in the classroom all the hundreds of tools we have on hand today.

In addition, in almost all cases, in-depth research is still needed to prove the real effectiveness of each tool, and some tools replace others very rapidly, so research is time-consuming and not very profitable. Most of the times, teachers are the ones who will have to explore effectively and creatively ways of implementing technology in their language classes without forgetting the most renowned learning and teaching theories and methodologies to tailor the learning needs of their students. As a result, today's teachers spend a lot of their free time dealing with classroom work.

At present, there are many digital tools that can be employed to optimize that time. These tools might be used by teachers of any language, and some are becoming increasingly popular. Some examples of them are exposed below, from the most general to the most specific:

(i) *Google Drive*: Teachers can store all the information they required for their lessons, elaborate all sorts of documents including multimedia elements and hyperlinks to the resources, organize students' assignments, create quizzes with Google Form, initiate collaborative projects, etc. (Başal & Aytan, 2014).

(ii) *Blendspace*: This is a content curation tool for integrating different multimedia components in a lesson (Başal & Aytan, 2014). It takes less than a few minutes to select the elements and transform them into a presentation. This can work with "flipped classrooms", which consists of giving instruction materials for home and taking advantage of the classroom time for experiential learning (Strayer, 2007).

(iii) *Livebinder*: It is used to collect different Internet sources, to combine them and give them different uses, from uploading course material to reviewing student progress, as a repository, or for sending individual or group projects to students (Başal & Aytan, 2014).

Although all these utilities enable to work more fluently, there is a fundamental issue in language teaching to which normally teachers devote time, and it is text selection. Whether it is classroom, reinforcement, or extension material for home, choosing the correct text is of major importance. A wide number of textbooks graded by level are accessible these days, but technology allows to search the adequate passage for a determinate student, group of students, or specific needs of a given moment. This technology also permits to work with authentic texts of the real world of the level needed, which is beneficial for today's students practice of natural language.

Instead of spending infinite hours looking for texts of a determinate level already posted on online language courses with the associated dangers implied due to copyright issues, teachers can take advantage of Corpus Linguistics (CL) and corpus linguistics tools, which are often little known by teachers, to select language teaching material. This kind of applications can also facilitate the students' assessment, as it will be shown later.

1.2. Corpus Linguistics for language teachers

I will devote this section to explain what Corpus Linguistics (CL) is for those who are not very familiar with this discipline yet. My purpose is to expand CL and corpus linguistics tools knowledge, as these can make teaching labor easier to a great extent and are not so complex to use, although they are based on some corpus linguistics foundations that are recommended to acquire or to develop.

Specifically, this section deals with the particular terminology of CL. Some of the basic corpus linguistics concepts exposed here will be useful to understand better the corpus linguistics tools for teachers later commented, mostly, *Text Inspector*, the application which all this work revolves around. Therefore, it is intended as a brief introduction to CL for language teachers, above all, English teachers.

Firstly, the term "corpus" has been defined as follows:

(i) "A collection of texts assumed to be representative of a given language, dialect, or other subset of a language, to be used for linguistic analysis" (Francis, 1992, p. 7).

(ii) "A collection of naturally-occurring language text, chosen to characterize a state or variety of a language" (Sinclair, 1991, p. 171).

(iii) "A finite-sized body of machine-readable texts sampled in order to be maximally representative of the language variety under consideration" (McEnery & Wilson, 2001, p. 32).

If we see literature, there is a large number of descriptions. However, its main features are already outlined. It consists of a set of electronic written texts or a range of transcriptions from speech, which are selected with a determinate purpose of research. Besides, this must be a representative sample of the language or variety of language in use to explore. I would only add that a corpus might be "general" which means that is composed by thousands of texts of different types and millions of words and uploaded to the Internet as it is the case of the Brown Corpus, the British National Corpus (BNC), the Corpus of Contemporary

American English (COCA) etc., or "specialized" which implies less texts and words compiled with specific purposes.

Secondly, it would seem obvious to say that CL uses corpora of real texts for language investigation. However, it is not so easy to properly define CL. The idea behind the study of texts is that Sinclair (1991) detects that a word by itself does not carry meaning, but a sequence of words does. Many other scholars have made their contribution to its evolution: Leech, Hunston, Biber, Francis, Conrad, Finegan, Johansson, etc. (Bennett, 2010). Nevertheless, there is a big discussion between the supporters of CL as a theory and the ones who think of that as a methodology.

As Tognini-Bonelli (2001) claims, CL obtained theoretical status since it possesses a clearly established set of rules and principles, which cannot be confused with other linguistic branches.

The Corpus Approach (Biber, Conrad & Reppen, 2007) has four major characteristics:

1. It is empirical, analyzing the actual patterns of language use in natural texts.
2. It utilizes a large and principled collection of natural texts as the basis for analysis.
3. It makes extensive use of computers for analysis.
4. It depends on both quantitative and qualitative analytical techniques.

Others such as Swales (2006) describes it as a methodology since it looks to linguistic data for many purposes: historical, critic, pedagogic, etc. In this sense, the research used with that methodology does not have to be entirely linguistic, that is, founded on a language theory or created to test a language theory.

All this discussion leads us, therefore, to accept two types of corpus studies: "corpus-driven studies" or "corpus-based studies". The former means "a theory-generating branch in the field of linguistics" (Friginal, 2018, p.13) while the latter refers to "a research approach or method" (Friginal, 2018, p.13). In other words, the investigations within the first group test existing hypothesis in the linguistic field. Nonetheless, inquiries within the second group study research questions relevant to distinct disciplines, for example, to English Language Teaching (ELT).

In any case, corpus linguistics data scrutiny provides, above all, with a deeper and reliable understanding of a determinate language. Before the corpora appeared, linguistics depended very much on native intuitions. The analysis of a large amount of data was possible on many occasions through online freely software such as *AntConc* (Anthony, 2019). When these

kinds of programs appeared, they offered a complete view on language structure as never before, which could empirically validate or refute all the language theories gathered until the moment. Therefore, all this modern research could begin to be transferred into teaching.

According to Vannestål and Lindquist (2007), linguists have validated corpora for pedagogical issues since the 80s. The functions which these applications most often do comprise the extraction of keywords from frequency lists of words or clusters of words of a determinate number (N-Grams). This is very useful for teachers to get to know the most important words and phrases in a language and to generate vocabulary lists for learners in which the latter can work first to enhance their communication skills. Besides, the concordance tool can search the context in which a keyword is found. Hence, it supplies with examples of use of a word or phrase in real contexts applicable to actual communicative situations.

All the above promotes a pedagogical improvement in language teaching. On the one hand, research can serve for upgrading syllabuses, course books, and other teaching materials. On the other hand, students might act as investigators and work directly with corpora getting to know the most frequently used word, checking errors, or practising alternatives, among other activities. This might be also motivating for student learning. In Proctor's (2012) words about ELT, "exposing to authentic English and producing native-like English through corpus are of significance for many EFL (English as a Foreign Language) students as beginners or intermediate ones" (p. 5).

This type of language pedagogy is in connection with the new teaching approaches and methodologies already commented above in this discussion. "Learners are encouraged to become more autonomous in their studies (...) discovery learning activities are designed to favour learner-centred, open-ended and tailored-learner" (Sinclair, 2004, p. 27).

Despite all these advantages, CL is still often overlooked in language learning and teaching. That was, in principle, understandable since our society has massively had texts in electronic format for several decades, but it has not had the appropriate resources to examine them thoroughly. Technological improvements, including web 2.0, enabled powerful search engines and digital tools to scan information in a matter of seconds. However, nowadays, CL is not sufficiently exploited. Language teachers need to be aware of all these methodological changes which can boost their teaching labor.

In addition, the introduction of CL in the classroom is also beneficial for teachers, especially if they are non-native speakers of the language they teach. Corpus-based discovery activities can help them to check their perceptions about how language is used and to guess

generalizations. In this way, they may reflect on the knowledge which they have always been taken for granted (Sinclair, 2004).

1.3. Some corpus linguistics tools for English Language Teaching

This chapter will shed light on the way in which corpus linguistics tools might be employed in ELT. More specifically, it focuses on the ones which can ease the process of assessing writings. These tools are *LexTutor* by Tom Cobb, *VersaText* by James Thomas and Vít Baisa, and Text Inspector by Stephen Bax.

1.3.1. *LexTutor*

First, I will explain LexTutor (see Figure 1). This free toolkit is described by Friginal (2018), as "a veritable armory of corpus- and frequency-based tools for English and French language learners and teachers, and researchers in linguistics" (p. 235). The tools that comprise are related to frequency, concordances, etc. It even includes flashcards for lexical learning. However, for the matter in hand, I want to highlight the tool *Vocabprofile*.

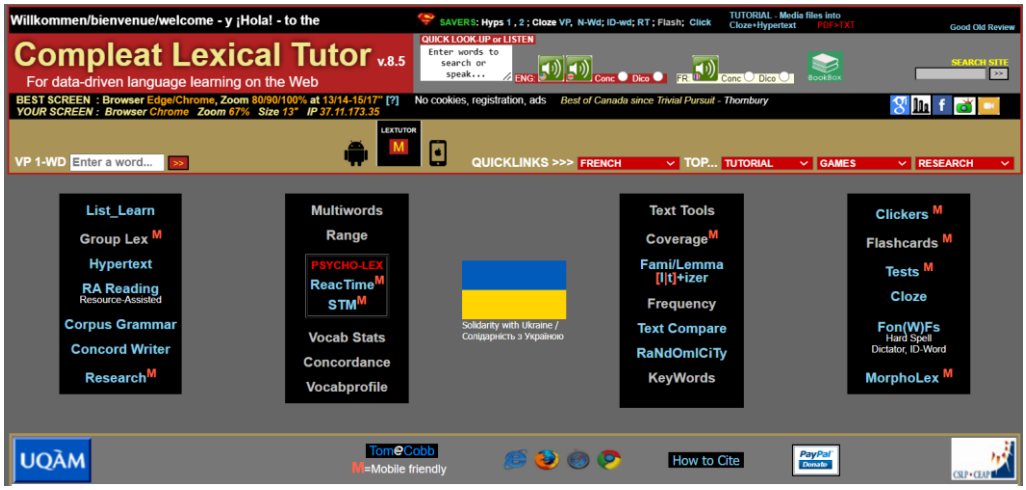
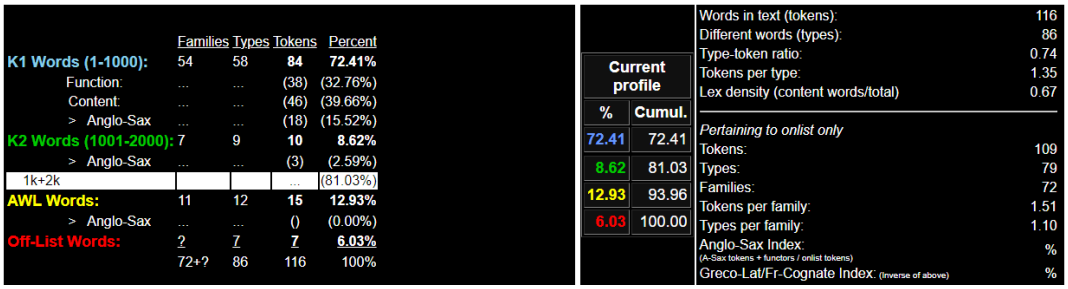


Figure 1. *Appearance of LexTutor homepage.*

Vocabprofile is based on frequencies analysis. It arranges the vocabulary of a text into K1, K2, AWL or Off-List. First, K1 is a list of the first thousand more frequent words in English. Second, K2 is a list of the second thousand most frequent words, which are related to terminology. Thirdly, AWL represents an academic wordlist. Finally, Off-List incorporates words that are not included in the other lists.

The procedure to use this toolkit is as follows. In the Classic VocabProfile section (VP-Classic), the text which I want to examine is inserted in the search box and submitted. Then,

another window appears. In the first place, it contains two tables with statistics and the percentage which represents the four lists mentioned above. Each of them is differentiated with distinct colors (see examples in Figure 2). In the second place, the text introduced is shown with the same coloring in the words (see examples in Figure 3). At the bottom, all words in each list are given (see examples in Figure 4).



	Families	Types	Tokens	Percent
K1 Words (1-1000):	54	58	84	72.41%
Function:	...	...	(38)	(32.76%)
Content:	...	...	(46)	(39.66%)
> Anglo-Sax	...	...	(18)	(15.52%)
K2 Words (1001-2000):	7	9	10	8.62%
> Anglo-Sax	...	...	(3)	(2.59%)
1k+2k			...	(81.03%)
AWL Words:	11	12	15	12.93%
> Anglo-Sax	...	...	()	(0.00%)
Off-List Words:	2	7	7	6.03%
	72+?	86	116	100%

Current profile	
%	Cumul.
72.41	72.41
8.62	81.03
12.93	93.96
6.03	100.00

Words in text (tokens):	116
Different words (types):	86
Type-token ratio:	0.74
Tokens per type:	1.35
Lex density (content words/total)	0.67
Pertaining to onlist only	
Tokens:	109
Types:	79
Families:	72
Tokens per family:	1.51
Types per family:	1.10
Anglo-Sax Index: (A-Sax tokens + functors / onlist tokens)	%
Greco-Lat/Fr-Cognate Index: (inverse of above)	%

Figure 2.

Example of output tables within Vocabprofile in LexTutor.

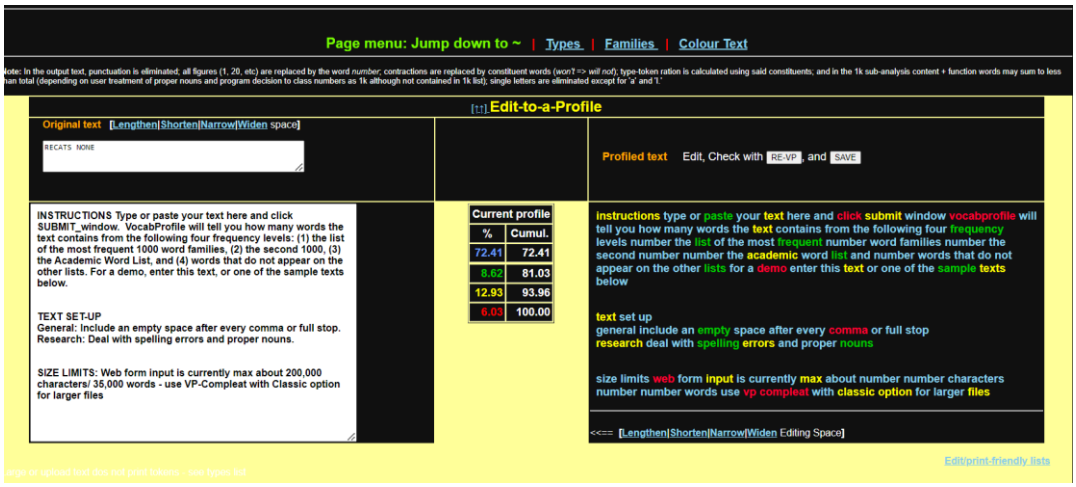


Figure 3.

Example of an output text within VocabProfile in LexTutor.

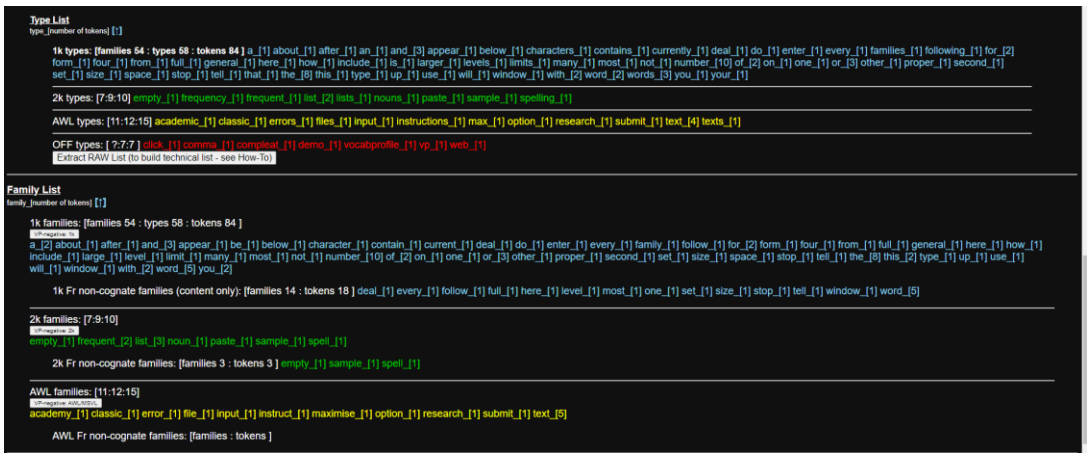


Figure 4.

Example of word-type lists within VocabProfile in LexTutor.

Its usefulness is obvious since it allows to choose passages with more or less frequent words (K1 words), a larger or smaller amount of terminology (K2 words), and many or few academic words (AWL words). In Friginal's (2018) own words, teachers can "use this information to measure the lexical sophistication of a text" (p. 240).

At the beginning, this tool might seem overwhelming, but if the teacher focuses just on the elements explained above, it does not to have to be confusing (Friginal, 2018). This tool was one of the first to be able to find out the kind of refinement of the words that conforms a text, but nowadays there are many more with other helpful features or which can specify exactly which CEFR level the texts have.

1.3.2. VersaText

It is also a free toolkit (see Figure 5), which explores the language of a single text with the help of the following tools: Wordcloud, Concordance and Profiler. It was inspired by the VocabProfile in *LexTutor*, and another application, which is Concordle. This latter produces a word cloud from the text given (Thomas, 2020). The main advantage of this toolkit respect to *LexTutor* is that is visually much more attractive and clearer. It even establishes the CEFR of the whole text.

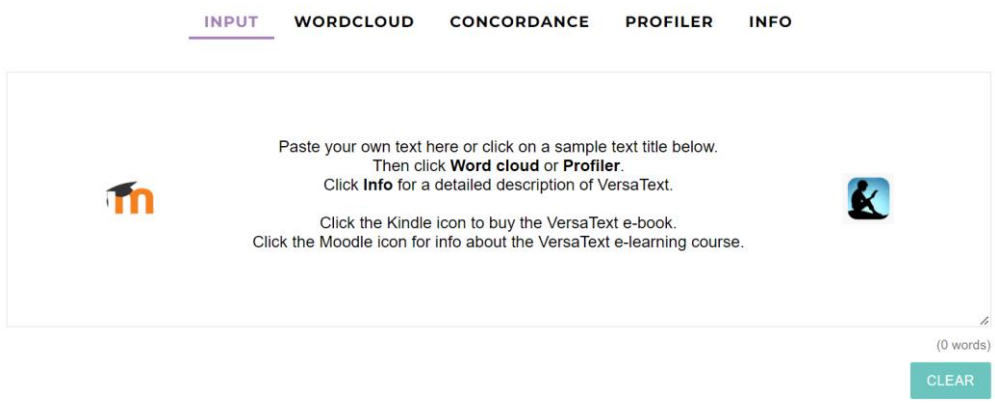


Figure 5.

Appearance of VersaText homepage.

The use of Profiler is the following. First, to paste a text into the *VersaText* box is required. The next step would be to click on the Profiler button to find information about the words in the text. The lexical statistics shown are very similar to the ones in *LexTutor*. The distinctive features are that the words are classified by Part-of-Speech (POS): noun, verb, adjective, etc. (See examples in Figure 6 and Figure 7).

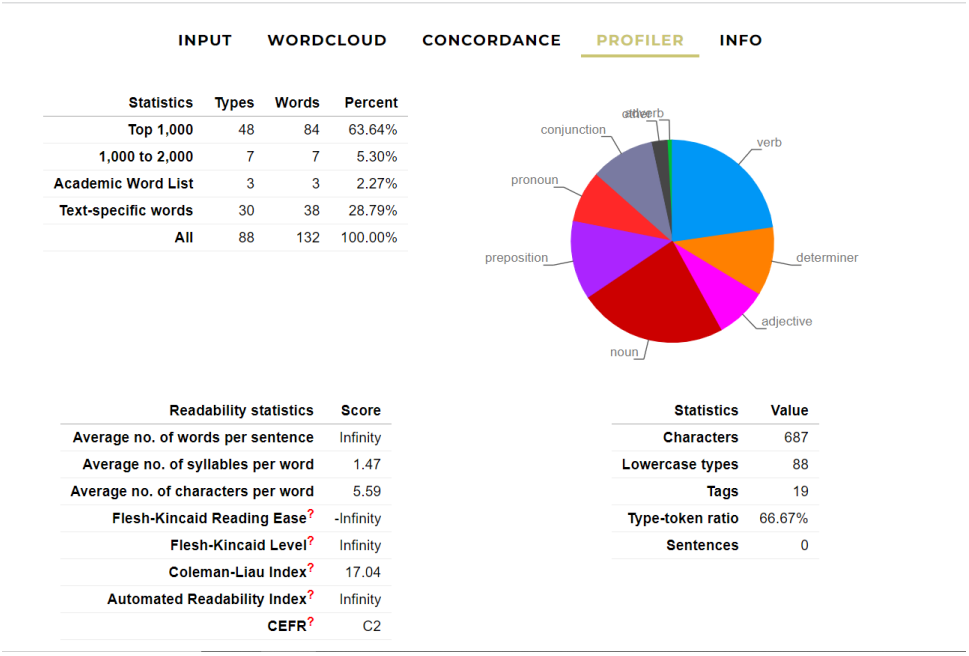
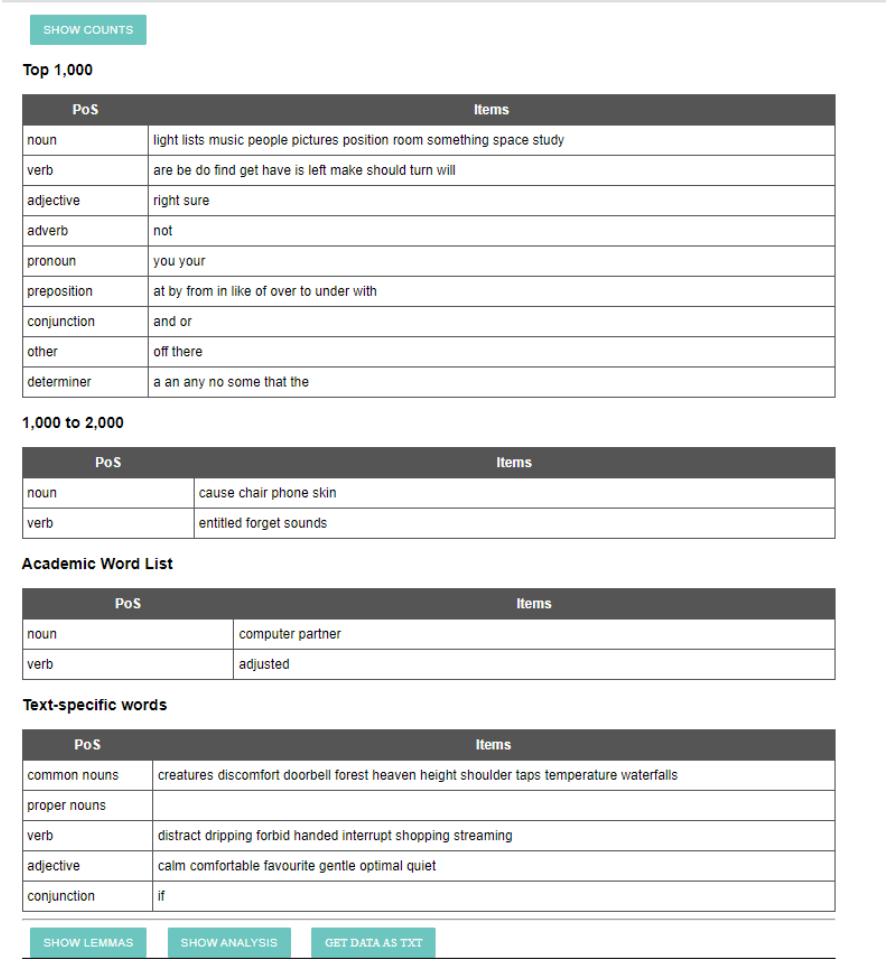


Figure 6.

Example of output statistics within Profiler in VersaText


SHOW COUNTS

Top 1,000

PoS	Items
noun	light lists music people pictures position room something space study
verb	are be do find get have is left make should turn will
adjective	right sure
adverb	not
pronoun	you your
preposition	at by from in like of over to under with
conjunction	and or
other	off there
determiner	a an any no some that the

1,000 to 2,000

PoS	Items
noun	cause chair phone skin
verb	entitled forget sounds

Academic Word List

PoS	Items
noun	computer partner
verb	adjusted

Text-specific words

PoS	Items
common nouns	creatures discomfort doorbell forest heaven height shoulder taps temperature waterfalls
proper nouns	
verb	distract dripping forbid handed interrupt shopping streaming
adjective	calm comfortable favourite gentle optimal quiet
conjunction	if

SHOW LEMMAS SHOW ANALYSIS GET DATA AS TXT

Figure 7.

Example of word-type lists within Profiler in VersaText.

1.3.3. *Text Inspector*

I cannot leave without commenting on *Text Inspector* (see Figure 8). Contrary to other complete corpus linguistics tools, *Text Inspector* can be used free of charge but only to analyze a maximum of 400 words at a time. The individual standard paid subscription is worthy to examine 10,000 text words, and it costs 59.76£ per year. The English Vocabulary Profile (EVP) created by Cambridge University Press is one of the most important utilities which is offered by *Text Inspector* to determine the CEFR level of a whole text. It contains reliable information about which words, phrases, idioms, and collocations are used at each level of the CEFR on a scale of A1-C2. However, it is available to subscribers only. There is the possibility to try the subscription option for a week at the price of 1.99£.

Text Inspector is based on the analysis of more than 200 metrics, which gives as a result an overview of the CEFR level of the text. This fact is supposed to make *Text Inspector* more appropriate to measure the students' performance in writing.

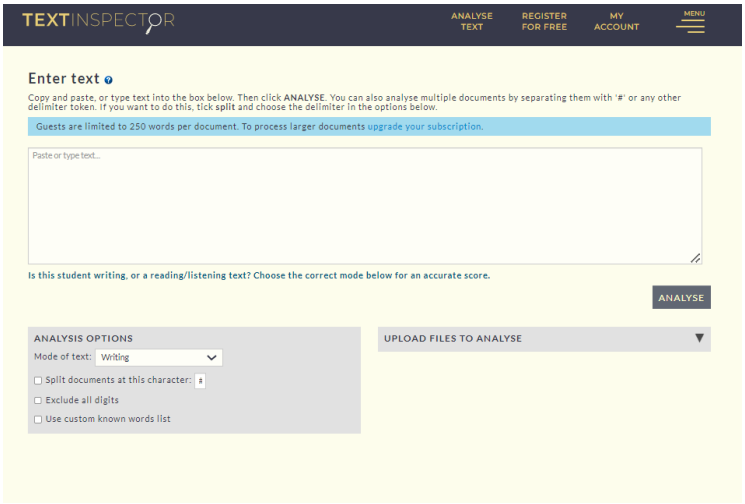


Figure 8.

Appearance of Text Inspector tool.

1.4. The Common European Framework of Reference for languages

The Common European Framework of Reference for languages (CEFR) represents a standard meter of language level in Europe, which describes the proficiency in a language achieved by an individual according to the main four language skills: 'reading' 'listening', 'writing', and 'speaking' (Council of Europe, 2001). This standard was created by thinking of a way of dividing language learning and teaching into levels that would be more achievable by the learner and assessable by the teacher. CEFR levels are shown in Figure 9. The criterion to distinguish among the levels is stated in Table 1.

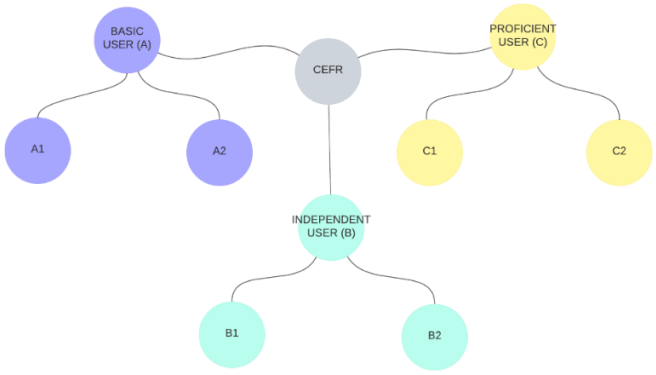


Figure 9.

CEFR levels.

Note: Adapted from *Figure 1* (Council of Europe, 2001, p. 23)

Proficient User	C2	Can understand with ease virtually everything heard or read. Can summarise information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.
	C1	Can understand a wide range of demanding, longer texts, and recognise implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices.
Independent User	B2	Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.
	B1	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes and ambitions and briefly give reasons and explanations for opinions and plans.
Basic User	A2	Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.
	A1	Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.

Table 1.

CEFR global scale.

Note: Extracted from *Table 1: Common Reference Levels: global scale* (Council of Europe, 2001, p. 24)

This framework provides the basis for the elaboration of curriculums, syllabuses, materials, and examinations. Likewise, the CEFR is introduced, for instance, in *Text Inspector* because, first, it is a valuation that is recognizable by everyone, and second, because it is less complex and faster than if the appraisal of the text were given in terms of vocabulary choice, discursive markers, or spelling, among other possible evaluative resources. In short, the CEFR attempts to encompass the performance achieved in different competences by assessing them through different descriptors (Council of Europe, 2020), and *Text Inspector* is inspired by this to give an overall assessment.

2. State of the art

This chapter is devoted to the literature that underpins my work. However, there is not much of it yet. *Text Inspector* is still a very new tool in the field of CL, and which competes somehow for the researcher's attention with the dozens of new linguistic tools that are continually coming onto the market today. The works that mention the *Text Inspector* tool and the English Vocabulary Profile (EVP), respectively, are those listed below. They are directly related to text selection and students' writing assessment regarding the CEFR.

2.1. Vocabulary profiles of authentic texts used by upper secondary English teachers: A lexical analysis of authentic texts used in EFL classrooms (Stahlberg, 2021)

This essay deals with the topic of English teachers' choice of texts in the classroom and uses technology (in this case, *LexTutor* and *Text Inspector*) to approach to that issue in a reliable way, by comparing some tentative texts against common reference descriptors. This shows that the English teacher's own selection of texts considering a reference level is nowadays more frequent as teachers no longer rely only on textbooks but guide their lessons with materials which they prepare by themselves.

The specific aim of the research in the paper is to analyse whether the vocabulary of the authentic texts worked in the Sweden course of *English 7*, which corresponds to B2.2, meets the standards of the CEFR of the Council of Europe and the Swedish National Agency for Education (Stahlberg, 2021).

The analysis is carried out by examining 26 texts from 5 teachers using the vocabulary profile tools in *LexTutor* and *Text Inspector* (Stahlberg, 2021). The frequency analysis leads to the conclusion that the texts of the 5 teachers have a similar level of vocabulary while the survey of the CEFR levels indicates that the texts are more suitable for students with C1 to C2 rather than B2.2, which is the level of *English 7*, as the 95% threshold is often achieved at C1 and the 98% threshold at C2. This means that are more advanced than required (Stahlberg, 2021). Interestingly, the essay includes a table (see Table 2) with prompts to get to know the vocabulary which would be appropriate for the different levels according to a vocabulary range descriptor (Council of Europe, 2020). This can be followed as a model for English teachers.

VOCABULARY RANGE	PROSIGN
C2	Has a good command of a very broad lexical repertoire including idiomatic expressions and colloquialisms; shows awareness of connotative levels of meaning.
C1	Has a good command of a broad lexical repertoire allowing gaps to be readily overcome with circumlocutions; little obvious searching for expressions or avoidance strategies. Can select from several vocabulary options in almost all situations by exploiting synonyms of even less common words. Has a good command of common idiomatic expressions and colloquialisms; can play with words fairly well. Can understand and use appropriately the range of technical vocabulary and idiomatic expressions common to his/ her area of specialisation.
B2	Can understand and use the main technical terminology of his/her field, when discussing his/her area of specialisation with other specialists. Has a good range of vocabulary for matters connected to his/her field and most general topics. Can vary formulation to avoid frequent repetition, but lexical gaps can still cause hesitation and circumlocution. Can produce the appropriate collocations of many words in most contexts fairly systematically. Can understand and use much of the specialist vocabulary of his/her field but has problems with specialist terminology outside of it.
B1	Has a good range of vocabulary related to familiar topics and everyday situations. Has a sufficient vocabulary to express him/herself with some circumlocutions on most topics pertinent to his/her everyday life such as family, hobbies and interests, work, travel, and current events.
A2	Has sufficient vocabulary to conduct routine, everyday transactions involving familiar situations and topics. Has a sufficient vocabulary for the expression of basic communicative needs. Has a sufficient vocabulary for coping with simple survival needs.
A1	Has a basic vocabulary repertoire of words and phrases related to particular concrete situations.
Pre-A1	*No descriptors available*

Table 2.

Vocabulary range by CEFR levels.

Note: Extracted from *Table 2. Common Reference Levels to Vocabulary Range* (Stahlberg, 2021, p. 6)

Finally, in the work, there is a reflection on the empirical study carried out. Despite the resources used in the research are recommended, with some albeit unspecified limitations on the part of *Text Inspector*, the importance of teacher training to examine the data from these tools correctly is stressed (Stahlberg, 2021). As I mentioned in the previous section of this study, the corpus linguistics tools are not very difficult to use, but they are not as user-friendly as other digital tools devoted to teachers. At the beginning, some prior knowledge of CL is highly suggested as they can be a bit overwhelming because of all the linguistic statistics that they collect. However, this should not be an impediment to giving them a chance as they are time-saving and very effective.

2.2. The English Vocabulary Profile as a benchmark for assigning levels to learner corpus data (Leńko-Szymańska, 2015)

This essay explores the English Vocabulary Profile (EVP), which is one of the instruments that *Text Inspector* nowadays uses to assign a proficiency level of the CEFR to learner corpora. Since this study was written some years ago, when there were not tools for students' writing assessment, it might seem outdated and useless, but nothing could be further from the truth. Contrary to that, it is important for the present dissertation in the sense that this investigation sets the first empirical steps to the rationality, reliability, and functionality of the analyses of corpus texts based on EVP. This idea means that the study would demonstrate EVP, which is also present in the research of my dissertation, to be a valid instrument for assigning a CEFR level to students' writings.

In any case, the purpose of the paper was to discover if the descriptors of lexical items of learner production in terms of the EVP could situate learners into a correct CEFR level. The data was taken from the International Corpus of Crosslinguistic Interlanguage, composed by English essays written in similar conditions in class by primary and secondary students of different countries, with A1–B2 levels. The 90 essays chosen were from Austria, Poland, and Spain. (Leńko-Szymańska, 2015).

Every word was manually coded through a level tag (A1–C2) according to the EVP. In addition, each text was rated as a whole by two raters. They considered nine descriptors inspired from the ones in the CEFR (Council of Europe, 2001): "overall written production", "creative writing", "reports and essays", "general linguistic range", "vocabulary range", "vocabulary control", "grammatical accuracy", "orthographic control", and "coherence". This implied much work and time. (Leńko-Szymańska, 2015). Hence, it can be argued that new tools such as *Text Inspector*, which includes the EVP, are convenient.

The hypothesis of the experiment was that the level assigned by the EVP system and the one conferred by the raters was similar, and the results demonstrated the truth of that statement (Leńko-Szymańska, 2015). The elements inspired by the CEFR served to sense that a language level, in this case, written, cannot be quickly evaluated just by lexical content, but it needs digital resources to be assess rapidly but with exactitude. Further research is needed regarding to this point, and this undergraduate dissertation goes in that direction.

3. Methodology

This chapter exposes the method of study of this undergraduate dissertation. In addition, it describes the procedures for collecting the data, and for designing and compiling the different corpora used with the criteria followed to do it. It also outlines the materials which are required such as the corpora, the digital resources, and some dictionaries. Finally, there is a separated important section where it is argued that the samples are representative and that, thus, the derived outputs from the analysis of the tool are also representative, legitimising the scientific validity of the whole study and the results obtained.

3.1. Method, procedures, and materials

An empirical study has been followed to scientifically proved the real capacity that *Text Inspector* has to provide reliable results for teachers when assessing students' writings according to the CEFR. This study includes a theoretical part with some justified hypotheses and a practical part carried out by establishing and applying various statistical measures to the quantitative results extracted from the analysis of the texts carried out by *Text Inspector*.

As I had no students' writings of my own to use, I decided to assess the CEFR level of several undergraduate dissertation abstracts in English elaborated by students at the University of Valladolid (Uva) during the academic year 2021-2022 and available in the Uva Repository under a Creative Commons license, only with the condition of giving credit to their authors in the work (for corpora references, see Appendix).

The undergraduate dissertation abstracts in English, as they have approximately the same number of obligatory words, were randomly collected. Specifically, I selected them from three different groups of Uva students from the degrees of Engineering, Business Management and Early Childhood Teaching. The intention was to introduce them, group by group, within *Text Inspector* and to find out which CEFR level the dissertation abstracts of students of each discipline had. In this way, regardless the texts field and the specific characteristics expected for such texts, the effectiveness of the tool would be tested, and bias would be avoided.

To do that, I decided to compile three corpora consisting of these dissertations abstracts of those three different degrees. Hence, this empirical study is also closely related to the discipline of CL. Specifically, it keeps a corpus-based approach since it uses corpora, as it has been just said, to test the corpus linguistics tool, *Text Inspector*.

The three corpora are described as “specialized learner corpora of L2 English speakers”, that is, of non-native speakers of English, monolingual, written, synchronic, mono-modal and closed. They are described as follows:

- Monolingual: the texts are only available in one language.
- Written: the texts are digitalised in a writing form. They are not in audio format.
- Synchronic: the texts come from a same unique moment of time.
- Mono-modal: the texts are in one format which is written. They are not audios or videos.
- Closed: the corpus is compiled for a determinate purpose, and there is no possibility to add anything else.

The corpora design is the following: The first corpus comprises fifty of these abstracts in the field of Engineering and has 6,753 words in total. The second corpus is made of forty-five abstracts from the Business Management area and consists of 6,833 words. The third corpus contains forty-six abstracts of the Early Childhood Teaching degree and includes 6,709 words (See Appendix for more information about corpora metadata).

The protocol compilation which was carried out had the next steps:

1. Finding enough learner data in UVa repository.
2. Downloading it in different files, storing and identifying them with a tag. For example, 01EABALONSOEN2022, which means in this order: text file number, field (E-Engineering, B-Business management, T-Early childhood teaching, text type (AB-abstract), author, language, and year.
3. Changing its format to plain text (.txt) in the codification “Unicode UTF-8” to be later introduced in *ReCor* to find out the sample size representativity of each corpus in terms of the number of documents and the number of words introduced.
4. Manually compiling the three corpora to be inserted in *Text Inspector*.

After that, the process to perform the qualitative and quantitative study was the following.

To start with, I checked the representativity of the corpora since if its size were not sufficiently scientifically valid, the study would have to start all over again. I will be later discussing this process in the following subsection called “Assessment of the representativity of the samples”.

Likewise, I established some theoretical hypotheses for the data of the three corpora.

- First hypothesis:

It is expected that the three groups of students will be classified as proficient users of English at level C of the CEFR, as they are university students who are expected to have a good command of the English language in their respective professional careers, which should be reflected in their dissertation abstracts.

- Second hypothesis:

Considering the most relevant metrics for the classification of students' writings at one level or another of the CEFR, such as “lexical sophistication”, “lexical diversity” of the texts, engineers would score higher than business administration and business administration would score higher than early childhood education in all of them, and in general.

The engineers' abstracts would contain a greater amount of terminology, as their authors, who are pure science students, are assumed to have a higher degree of technical precision than the other selected groups of students. This would mean that in "lexical sophistication" they would be seen to use mostly advanced level vocabulary that allows them to communicate effectively, but they would also use many proficiency or terminological words. It would also mean that in "lexical diversity" they would achieve better marks than the other groups by handling a greater number of synonyms from all CEFR levels of vocabulary.

On the other hand, abstracts on Business Administration would achieve scores in "lexical sophistication" and "lexical diversity", and also, in general, lower than those of engineers and higher than those of early childhood teaching students, since their authors have university studies in economics. They are, thus, assumed to be less technically precise than engineers, but more precise than early childhood educators, as they handle terminology specific to this area, which is normally in English, and to have greater communicative ability than the latter, as they have to be fluent in the world of business. Therefore, the greater number of words used would also belong to the advanced level, but they would produce many proficiency-level words as well.

Finally, the authors of the abstracts on Early Childhood Teaching would also have the C-level English expected as students about to graduate, but as they do not belong to the group of pure science or economics students, it is assumed that they would score lower in "lexical sophistication" and "lexical variety", as they would use on average an advanced English vocabulary like the other groups but it is assumed that they would not use as much C-level

English as they would not have as much terminology in their field. The same would be true for "lexical variety" as they would be more accustomed to dealing with less technically precise and explanatory texts with repetitive vocabulary, as those in the educational world tend to be like that.

After establishing the hypotheses, I introduced the corpora, one by one, into *Text Inspector* software. This was done within the "analyse" section of the program. I have to highlight here that for students' writing assessment, the "writing" mode must be selected at the end of the section mentioned above. When inserting each of the corpora, the tool automatically calculated the metrics, determining a score and an overall CEFR level according to such score for the abstracts included in each corpus.

Then, I carefully looked at the outputs of the analysis, which were distributed in several tables according to the different metrics provided by the tool. These outputs will be later provided in the "Results" section of this undergraduate dissertation.

At that point, I studied again the representativity of the samples by establishing some of the metrics as control variables to limit the error uncertainty which is inherent in any empirical study. There was arguably so much consistency among the quantitative outputs of the metrics set as control variables that this assessment of the representativity legitimised the validity of the samples and the study's future conclusions about the reliability of *Text Inspector*. This representativity assessment will be shown in the following subsection called "Assessment of the representative samples".

Finally, I established and applied some series of statistical measures or new metrics to some of the quantitative outputs from the analysis provided by *Text Inspector* to compare the practical efficiency of the tool with the theoretical efficiency expected when I previously proposed the theoretical hypothesis, and to eventually determine the scientifical reliability of the tool.

To finish this subsection, and regarding the materials, apart from the corpora, and the *ReCor* and *Text Inspector* softwares mentioned, other free resources were employed during the research such as *Lucidspark* and *Paint* for the preparation of figures.

3.2. Assessment of the representativity of the samples

This is one of the most important parts of this undergraduate dissertation. If the samples were not representative and valid, the results of the study would also not be reliable or useful for providing any scientific knowledge to the field of ELT.

Therefore, the representativity of the samples were proved in two ways. First, by employing the free software, *ReCor* (Corpas & Sehgiri, 2007), to check the size of each corpus, that is, to see if the samples in each corpus were accurate in both the number of documents and the number of words compiled. Second, by establishing the same control variables looking to the metrics and outputs according to them provided by *Text Inspector*. The aim was to ensure that the texts, even if they were very different, could be equally evaluated with the tool, and that they were not a source of error for the results of the empirical study. For this purpose, the general statistical assumption that these errors or deviations would be below or around 5% was followed.

3.2.1. *ReCor*:

ReCor calculates the minimum size of a corpus by analysing the lexical density in relation to the corpus increase, which is represented by C, document by document, which are denoted by d1, d2, d3, etc., as shown in the following equation: Cn= d1+ d2+d3+...+dn. This is the algorithm 'N-Cor'. The main idea is that the number of types does not increase in proportion to the number of words the corpus contains, once a certain number of texts has been achieved. Therefore, at that point, it would be enough representative, and there would no need to enter more words or documents in the corpus (Seghiri, 2014).

The tool works as follows. Once, the corpus is inserted in *ReCor*, three output files with statistics and two graphs resulted from them are created in each case. The software illustrates the level of representativeness of the three corpora in the two graphs (see Figures 10, 11, and 12). They show that lines in both graphs grow exponentially at first and then stabilise as they approach to zero. At those points, sample representativity is achieved, what happens in every corpus.

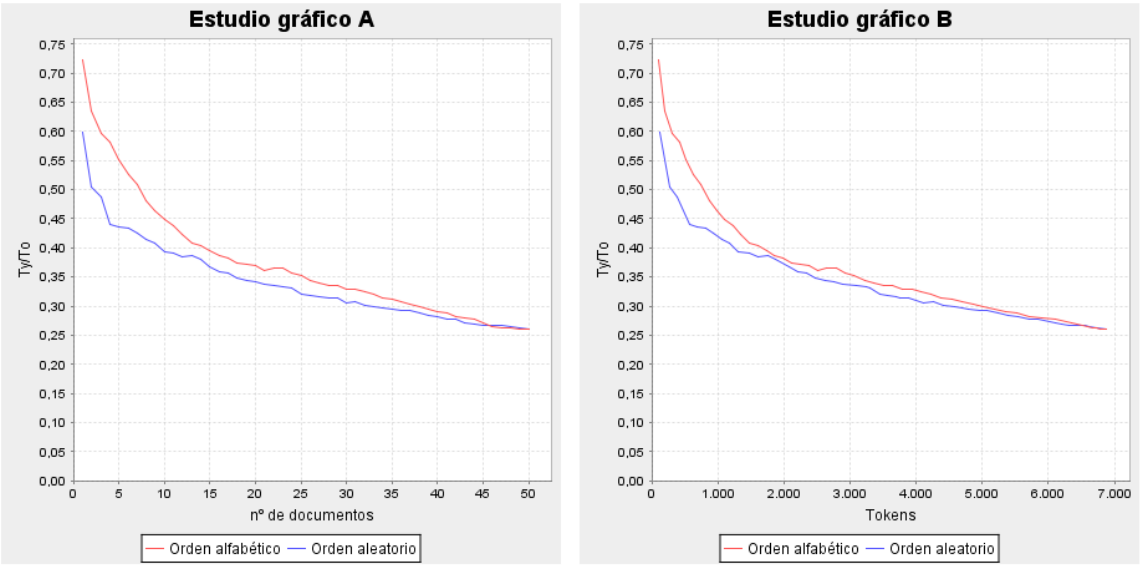


Figure 10.

ReCor representativity of Engineering corpus.

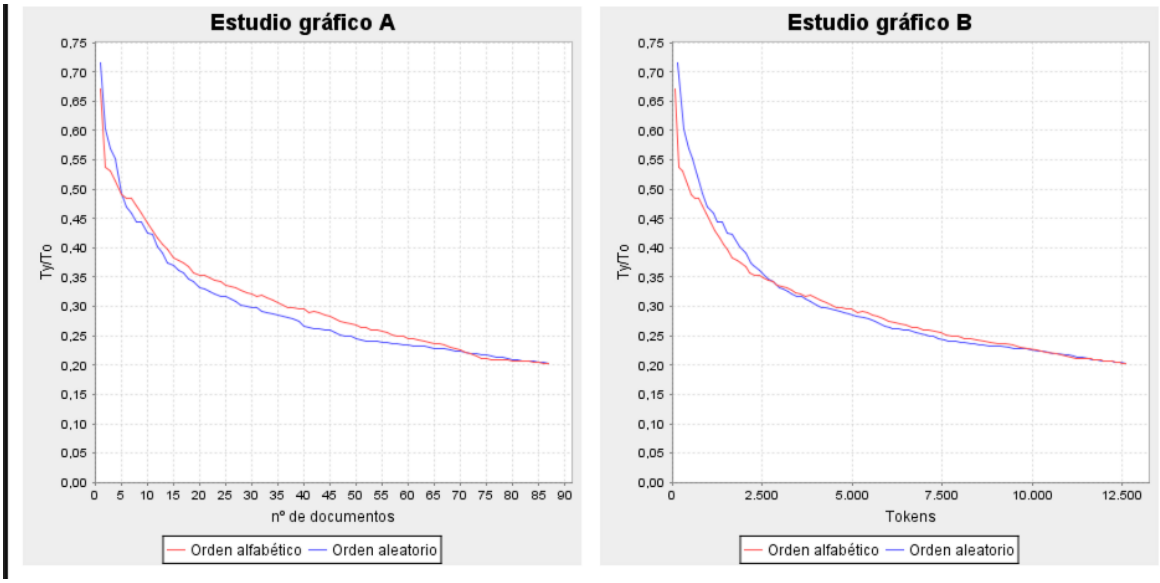


Figure 11.

ReCor representativity of Business Management corpus.

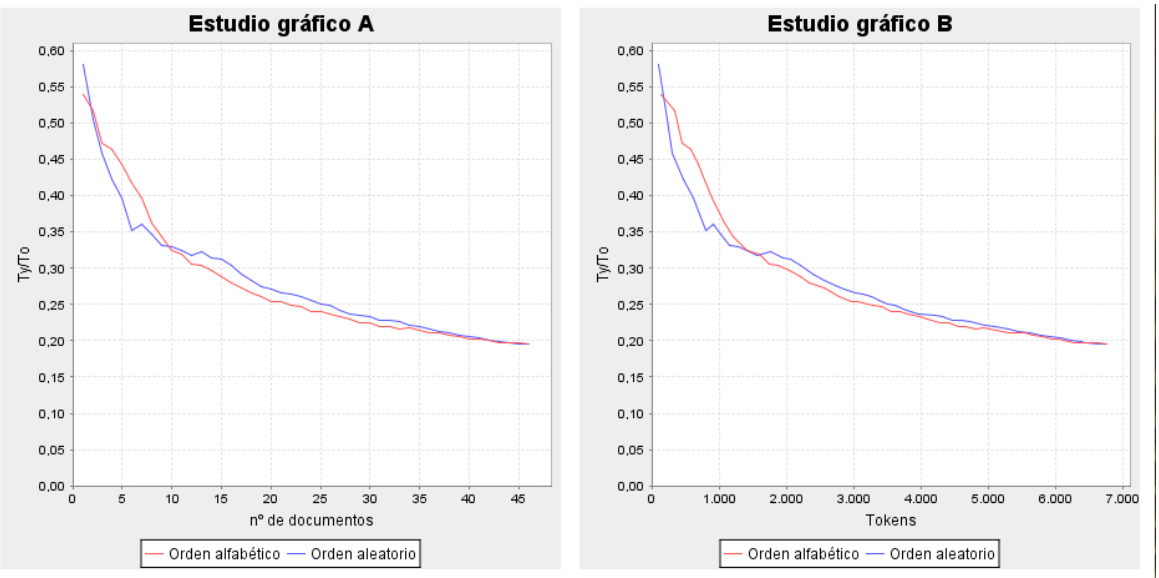


Figure 12.

ReCor representativity of Early Childhood Teaching corpus.

3.2.2. Control Variables:

The following subsection gives more quality to this study as assessing and verifying the representativity level of the samples by control variables legitimises its results, and thus the scientific validity of *Text Inspector*.

To this end, from the tables and graphics provided by *Text Inspector* after the analysis of the texts based on their metrics, I will decide which indicators on them can be considered as control variables because they offer more or less homogeneous quantitative outputs among the three groups.

The tables and graphics with its indicators and outputs are displayed within *Text Inspector* in the following sections: "statistics", "lexical diversity", "errors", "lexis:EVP", "lexis KVL", "lexis:BNC", "lexis:COCA", "lexis:AWL", "metadiscourse" and "scorecard". First, I will describe them.

In the first place, "statistics" section has some fundamental metrics, which give basic information of the learner corpora introduced. For instance, the number of tokens (excluding numbers) and the number of types (excluding numbers) of each corpus. "Tokens" means the whole number of words in the text. Instead, "types" refers to the whole number of words in the text too, but without counting repetition. The type-token ratio (TTR) is a basic indicator of the variety of the lexicon within the texts. A high TTR indicates a high lexical variety

while a low TTR means the opposite (a low lexical variety). Yet, TTR is sensitive to text length so other lexical diversity indexes are proposed in the following section "lexical diversity" to avoid any interpreting problem. Likewise, as this is learner corpora, average sentence length is also useful because it reveals the capacity of the learner to make complex sentences with coordinates and subordinates' clauses.

The "statistics" section also displays the readability score. It presents three distinct instruments to measure it to avoid false conclusions: the Flesch Reading Ease (Flesch, 1948), the Flesh-Kincaid Grade (Flesch, 1994), and the Gunning Fog Index (Gunning, 1968). First, the Flesch Reading Ease is calculated with the sentence length and the number of syllables per word. That concedes a score between 1 and 100, and each of the scores corresponds to a British education grade level. Second, the Flesch-Kincaid Grade is approximately the same as the Flesch Reading Ease but comparing with the United States grades. Finally, the Gunning Fox Index is the average of the number of words per sentence and the number of long words per word. It figures out the time counted in years of formal education a person requires to grasp a text on the first reading.

In the second place, "lexical diversity" section measures how many different words appear in a text, this time through the Vocabulary Diversity (VOCD) and the Measure of Textual Lexical Diversity (MLTD) indexes, indexes which are used together to avoid drawing false conclusions.

Thirdly, "errors" section displays the possible spelling mistakes found in the corpora.

Then, there are a number of sections called "lexis:EVP", "lexis KVL", "lexis:BNC", "lexis:COCA", and "lexis:AWL" which deal with "lexical sophistication". The sophistication or knowledge of all kinds of words, even the most advanced or specialised ones, is specifically checked by the Cambridge-designed EVP or English Vocabulary Profile tool which shows each word with its correspondent CEFR level, but also by the British Council's KVL or Knowledge-based Vocabulary Lists. Another way of checking this lexical sophistication is through corpora by looking at the frequency of use within a large corpus, for which corpora such as BCN or British National Corpus and COCA or Corpus of Contemporary American English can be used. If we want to check academic lexis, there is a ready-made wordlist called AWL or Academic Word List, which is already built into *Text Inspector,* which is very convenient for quick comparisons between results without having to go outside the software.

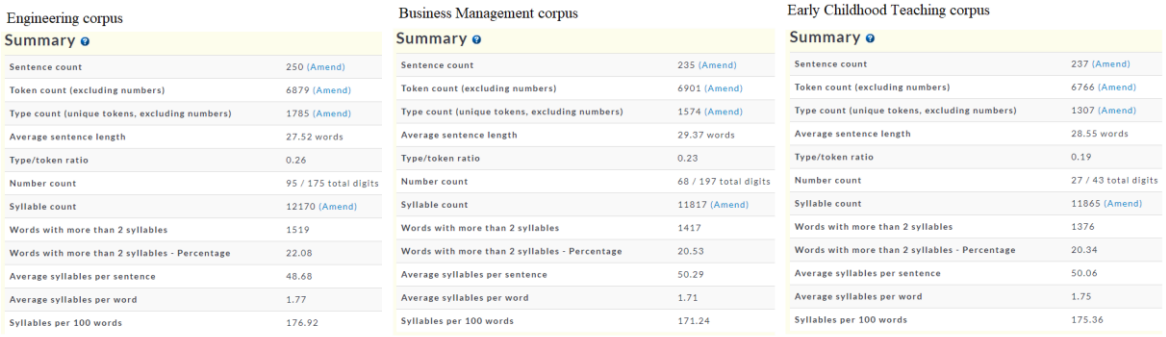
After that, there is the “metadiscourse” section with the most used kinds of discourse markers, showing how many of them appear in the texts entered, and which ones they are.

Finally, the “scorecard” section is the most important tool for the teachers. It shows the score achieved according to the texts entered, which the closer to 100%, the more similar to academic texts in the same field a native speaker would produce, and the level of CEFR associated with it. It is on this score that teachers using *Text Inspector* should look. In addition, all the other metrics can also be used to provide data on students' performance and, according to their strengths and weaknesses, teachers can help them to write consistent texts.

After having examined the tables carefully, I have chosen to fix the control variables according to all the indicators in the “statistics” section (see Figure 13 and Figure 14) since the quantitative outputs these indicators show are very similar in the three corpora, with a very small difference between them, and because, when its deviation or possibility of error is calculated, it is below or around the statistically 5% permitted, in all cases. Likewise, even the percentage error of the average of all the errors of all the indicators in these tables would be less than or around 5% and, thus, compensated.

For example, the deviation or possibility of error for the control variable of the average sentence length (in Figure 13) is calculated as (29.37-27.52):29.37=0.063. Thus, it is a 6%, very near to the maximum error of the 5%. So, that is a manageable amount of error for the analysis. Or another example, the deviation or possibility error for the control score of readability is measured in terms of the Flesch Reading Ease (in Figure 14) is calculated as (32.16-29.24):32.16=0.090. Therefore, this means that it is a 9% of manageable error which allows to carry out the empirical study.

In addition to the above, the readability scores could be good control variables also because their data in Figure 14 show the same information for the three groups of students, that is, that all the students’ abstracts are in the same intervals or very near to them. They were about to graduate in the moment of writing the abstract, so they belong to the same type of university students. (See Table 3, Table 4, and Table 5 which provide the same kind of information).



Engineering corpus

Summary

Sentence count	250 (Amend)
Token count (excluding numbers)	6879 (Amend)
Type count (unique tokens, excluding numbers)	1785 (Amend)
Average sentence length	27.52 words
Type/token ratio	0.26
Number count	95 / 175 total digits
Syllable count	12170 (Amend)
Words with more than 2 syllables	1519
Words with more than 2 syllables - Percentage	22.08
Average syllables per sentence	48.68
Average syllables per word	1.77
Syllables per 100 words	176.92

Business Management corpus

Summary

Sentence count	235 (Amend)
Token count (excluding numbers)	6901 (Amend)
Type count (unique tokens, excluding numbers)	1574 (Amend)
Average sentence length	29.37 words
Type/token ratio	0.23
Number count	68 / 197 total digits
Syllable count	11817 (Amend)
Words with more than 2 syllables	1417
Words with more than 2 syllables - Percentage	20.53
Average syllables per sentence	50.29
Average syllables per word	1.71
Syllables per 100 words	171.24

Early Childhood Teaching corpus

Summary

Sentence count	237 (Amend)
Token count (excluding numbers)	6766 (Amend)
Type count (unique tokens, excluding numbers)	1307 (Amend)
Average sentence length	28.55 words
Type/token ratio	0.19
Number count	27 / 43 total digits
Syllable count	11865 (Amend)
Words with more than 2 syllables	1376
Words with more than 2 syllables - Percentage	20.34
Average syllables per sentence	50.06
Average syllables per word	1.75
Syllables per 100 words	175.36

Figure 13.

Text Inspector Control Variables

Engineering corpus

Readability Scores

Flesch Reading Ease	29.24
Flesch-Kincaid Grade	16.02
Gunning Fog index	19.84

Business Management corpus

Readability Scores

Flesch Reading Ease	32.16
Flesch-Kincaid Grade	16.07
Gunning Fog index	19.96

Early Childhood Teaching corpus

Readability Scores

Flesch Reading Ease	29.50
Flesch-Kincaid Grade	16.24
Gunning Fog index	19.55

Figure 14.

Text Inspector Control Scores

Reading Ease Score	Descriptive Categories	Estimated Reading Grade
90 – 100	Very Easy	5th Grade
80 – 90	Easy	6th Grade
70 – 80	Fairly Easy	7th Grade
60 – 70	Standard / Plain English	8th and 9th Grade
50 – 60	Fairly Difficult	10th to 12th Grade (High School Sophomore to Senior)
30 – 50	Difficult	In College
0 - 30	Very Difficult	College Graduate

Table 3.

Interpretation of the Flesh Reading Ease score.

Note: Extracted from *Table 1: Descriptive Categories used in the Flesch Reading Ease Formula* (Hussin, 2015, p.126).

FRE score	Reading difficulty	Estimated FK reading grade level
91-100	Very easy	US Grade 5 or 11-year-old
81-90	Easy	US Grade 6
71-80	Fairly easy	US Grade 7
61-70	Standard	US Grade 8-9 or 13-15-year-old
51-60	Fairly difficult	US Grade 10-12
31-50	Difficult	US Grade 13-16
0-30	Very difficult	College graduate

FRE=Flesch reading ease, FK=Flesch-Kincaid

Table 4.

Interpretation of the Flesch Kincaid-Grade score.

Note: Extracted from *Table 1: Interpretation of the Flesch Reading Ease score* (Jindal & MacDermid, 2017, p. 85)

Fog Index	Reading Level by Grade
20+	Post-graduate plus
17–20	Post-graduate
16	College senior
15, 14, 13	College junior, sophomore, freshman
11–12	High school senior, junior
10	High school sophomore
9	High school freshman
8	8th grade
7	7th grade
6	6th grade

Table 5.

Interpretation of the Gunning Fog Index score.

Note: Extracted from *Table 3: Gunning's fog-index level* (Eleyan, Othman & Eleyan, 2020, p.7)

4. Results

This chapter deals with the quantitative outputs obtained with *Text Inspector* from the automatic analysis of the different corpora introduced on the basis of the metrics which the tool contains, and which principally assigns a score and a CEFR level to each of them. The chapter also includes the most relevant part of this undergraduate dissertation which is the display of its practical results provided after carefully examining the quantitative outputs.

4.1. *Text Inspector* outputs

For our empirical study, of the sections of the *Text Inspector* described in the previous part of the dissertation, we have chosen to focus on the "scorecards" section (see Figure 15), as it contains the scores obtained by each group of students and the level achieved.

We can see that all of them have obtained a score of more than 70% which allows them to acquire a C2 level of written English (they even put a plus sign in the engineers' C2), being 100% the level of academic English expected for a native speaker as we have already said. On the other hand, the overall scores obtained seem to leave those on Early Childhood Teaching in second place and those on Business Administration in third.

For the sake of practicality not all metrics considered by the tool are displayed here since there are over 200, but as we saw before, the most important metrics are distributed in other sections of the tool to see the results one by one in detail.

Among all the metrics that the tool contains, I would highlight the importance for the empirical study of the “lexical variety” measured with the Vocabulary Diversity (VOCD) and the Measure of Textual Lexical Diversity (MLTD), which is also in the “scorecards” section; the “lexical sophistication” in terms of the EVP which is in the “lexis:EVP” section of the tool, and the “readability score” which is inside the “statistics” section of the tool. The interpretation of the results of lexical diversity can be made through the information in Figure 16.

If we examine also lexical variety with the TTR in the “statistics” part, we can check that the lexical diversity is similar to the measured in terms of the VOCD and the MTLD, and the three groups are quite different based on this former: 0.26 in engineering abstracts, 0.23 in business management abstracts and 0.19 in early childhood teaching abstracts, being these latter the ones with a lower ratio.

With the help of all those metrics, we will be able to check the theoretical hypotheses we had presumed.

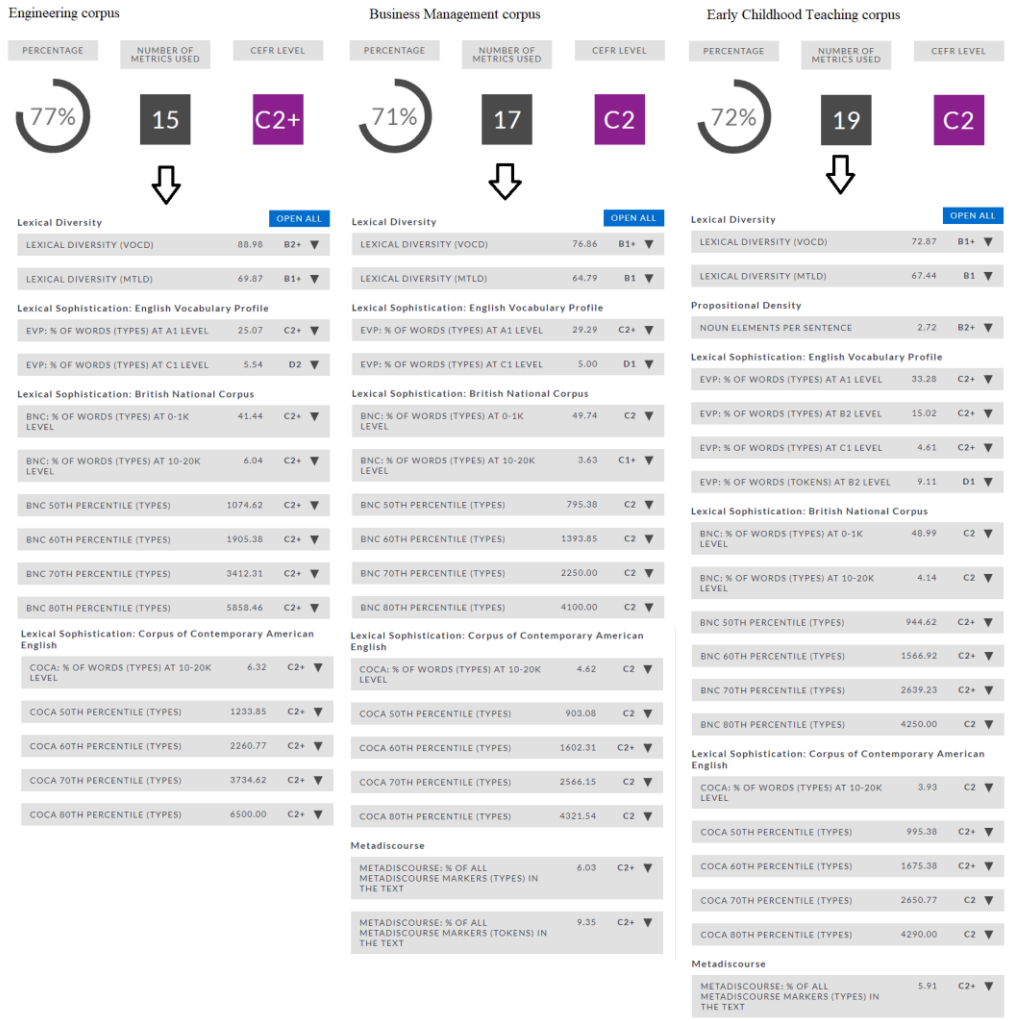


Figure 15.

Text Inspector Scores and CEFR levels

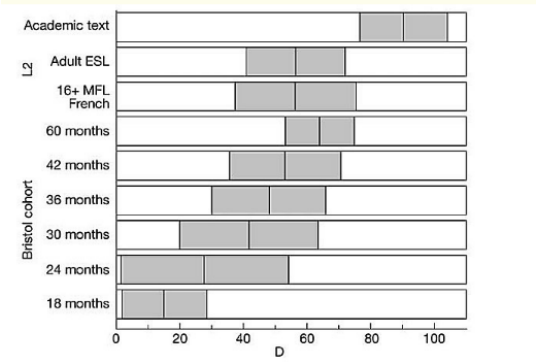
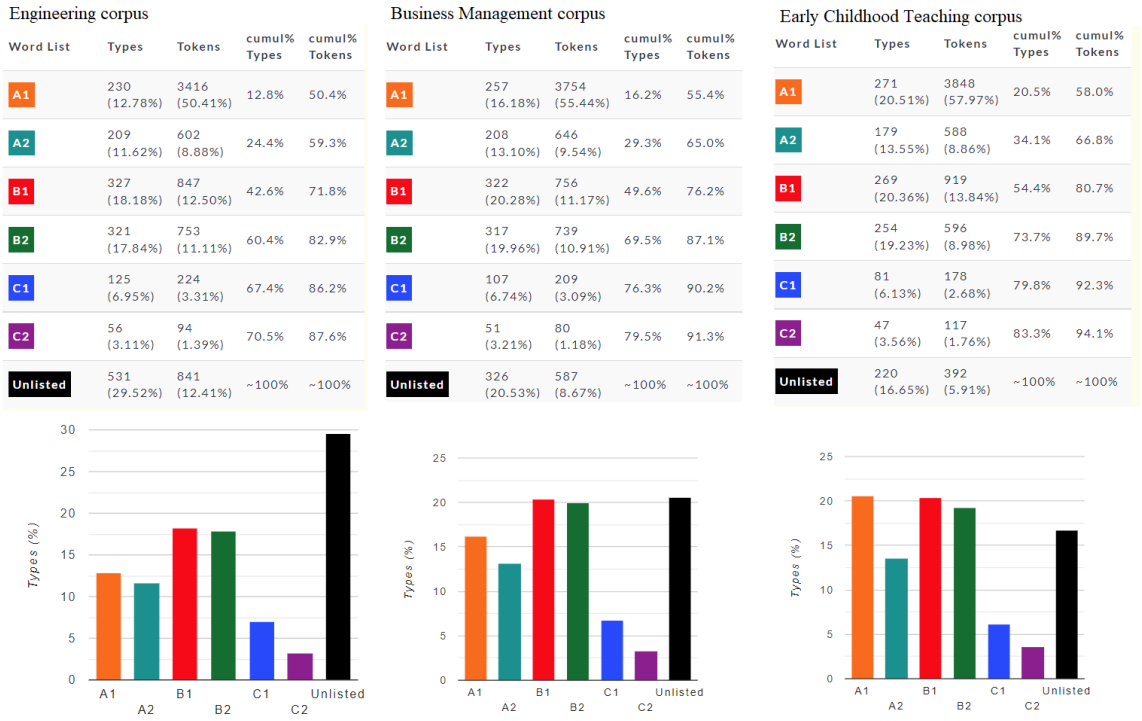


Figure 16.

Interpretation of lexical diversity.

Note: Means and sub-ranges (10th-9th percentiles) of D for various cohorts (Duran, Malvern, Richards, Chipere, 2004, p. 238)



Engineering corpus

Word List	Types	Tokens	cumul% Types	cumul% Tokens
A1	230 (12.78%)	3416 (50.41%)	12.8%	50.4%
A2	209 (11.62%)	602 (8.88%)	24.4%	59.3%
B1	327 (18.18%)	847 (12.50%)	42.6%	71.8%
B2	321 (17.84%)	753 (11.11%)	60.4%	82.9%
C1	125 (6.95%)	224 (3.31%)	67.4%	86.2%
C2	56 (3.11%)	94 (1.39%)	70.5%	87.6%
Unlisted	531 (29.52%)	841 (12.41%)	~100%	~100%

Business Management corpus

Word List	Types	Tokens	cumul% Types	cumul% Tokens
A1	257 (16.18%)	3754 (55.44%)	16.2%	55.4%
A2	208 (13.10%)	646 (9.54%)	29.3%	65.0%
B1	322 (20.28%)	756 (11.17%)	49.6%	76.2%
B2	317 (19.96%)	739 (10.91%)	69.5%	87.1%
C1	107 (6.74%)	209 (3.09%)	76.3%	90.2%
C2	51 (3.21%)	80 (1.18%)	79.5%	91.3%
Unlisted	326 (20.53%)	587 (8.67%)	~100%	~100%

Early Childhood Teaching corpus

Word List	Types	Tokens	cumul% Types	cumul% Tokens
A1	271 (20.51%)	3848 (57.97%)	20.5%	58.0%
A2	179 (13.55%)	588 (8.86%)	34.1%	66.8%
B1	269 (20.36%)	919 (13.84%)	54.4%	80.7%
B2	254 (19.23%)	596 (8.98%)	73.7%	89.7%
C1	81 (6.13%)	178 (2.68%)	79.8%	92.3%
C2	47 (3.56%)	117 (1.76%)	83.3%	94.1%
Unlisted	220 (16.65%)	392 (5.91%)	~100%	~100%

Figure 17.

Text Inspector lexical sophistication according to EVP.

4.2. Practical results of the empirical study

From the quantitative outputs provided by *Text Inspector* and based on the different metrics we have seen that we need to consider checking the theoretical hypotheses, I have elaborated some statistical measures or new metrics that will allow scientifically valid conclusions.

To do this, I have calculated some empirical means for each indicator selected which match the content of the theoretical hypotheses established in the "Methodology" section of this dissertation (see Table 6).

	Engineering corpus	**Business Management corpus**	**Early Childhood Teaching Corpus**
1st metric related to the scores in the "Scorecards"	1,748.218	1,069.504	1,014.75
2nd metric related to the lexical diversity in terms of the VOCD and the MTLD	79.425	70.825	70.155
3rd metric related to the lexical sophistication in terms of the EVP	257	226.85	188.71

Table 6.

Metrics elaborated for the empirical study

5. Discussion

In this part of the undergraduate dissertation, I will establish a comparison between the theoretical efficiency and the practical efficiency of *Text Inspector*. This is also a very important part of the study as it clearly verifies the theoretical hypotheses previously discussed.

- First hypothesis:

Confirmed. All groups have been within C-level. In fact, C2-level, which is what is expected as it is an academic level. Even, the engineers appear to have C2+, which could be a C2.2 level, which is close to that of the natives in the academic field (77% of percentage). This can be achieved in this kind of discipline since the technical language required, not applied neither to Business Management nor to Early Childhood Teaching.

- Second hypothesis:

Confirmed. Although the scores show a higher score of the educators over the economists with the calculation of the empirical means we have checked that in all indicators the order of the groups is checked as the one established in the theoretical hypotheses: first, the engineers, then the economist and, then, the educators.

According to the lexical sophistication considering the EVP, there are mostly B-words in all groups, but there are less C-words in early childhood teaching groups than in the other two groups. In addition, we see that there is a large number of words that has not a CEFR level, since they surely are not categorized by the EVP. This situation happens even more in the case of engineering (until 531 unlisted words) because it consists of specialized terminology with fewer frequencies, which *Text Inspector* does not recognize.

In the light of the above, it is reported that *Text Inspector* has very interesting aspects that seem to accurately determine the level of the texts included in the different corpora since many reasons: The texts compiled in the corpora belong to an academic level, written by adults who have ESL, and finishing their university degree. It also seems that there is accuracy in the valuation of the texts since it shows probable differences in the texts regarding their genre.

Therefore, the strengths of *Text Inspector* for the English teachers might be the clear display of the CEFR levels and scores; the lexical diversity and lexical sophistication indexes, above all, in terms of the useful tool, EVP of Cambridge to get to know which words the students are daily employing; its user-friendliness if teachers only focus on the specific sections to

look for information about their students‘ performance, etc. On the other hand, the format of the “scorecards” is very attractive, and it can be consulted by even learners without any difficulty.

Nonetheless, I consider that *Text Inspector* also has some problems to solve yet. As Stahlberg (2021) already exposed, it has some limitations. Firstly, the EVP or other frequency lists inserted do not include all the lexicon in a language, so much of the vocabulary cannot be valued in terms of the CEFR. This also happens with errors, which are not recognized by the tool. Secondly, *Text Inspector* does not include information about grammar, which I consider fundamental to also figure out the writing level of the students as they could write very elaborated vocabulary but then make grammatically incorrect sentences. The English Grammar Profile tool could be inserted in the software to give more accurately results.

5. Conclusions

In this undergraduate dissertation, I have attempted to answer two research questions:

1. Is *Text Inspector* able to appropriately assess the CEFR level of the students' dissertation abstracts of different disciplines?
2. If so, is Text Inspector so precise and user-friendly for teachers to be suggested to evaluate their students' writings?

The answer of the first one is positive. The theoretical and the practical efficiency of *Text Inspector* have been scientifically proved with empirical and representative data. The software distinguishes, and correctly grades different types of texts and it was not difficult to use.

One of the problems found is regarding to the amount of vocabulary that is not tagged by the tool, what would have to be done by hand, using dictionaries and corpora, and consisting in a hard task as we already see in (Leńko-Szymańska, 2015).

Other problem deals with the lack of a grammatical perspective since *Text Inspector* is a "text-checker" corpus linguistics tool based on lexical content overall. This means that, although *Text Inspector* has a more appealing format, and more measuring instruments, it relies on the basics of other existing tools, such as *LexTutor* or *VersaTex*t, and it does not provide other elements that teachers often use to assess their pupils, such as grammar. The inclusion of them could have been a major advantage for the company in the market.

The CEFR is constantly updating with new ways of assessing the students regarding different parameters. At each level of each new evaluable competence some features of a performance are expected. Thus, we all know that a well-written text is not only featured by lexicon and spelling. Attention-getters, communicative achievement according to the target reader, context appropriateness, organization of the text which moves smoothly from paragraph to paragraph in climatic order, convenient cohesive devices (anaphoric, cataphoric or exophoric references), tense agreement, substitution or linking words are also important to be included (Council of Europe, 2020). Leńko-Szymańska (2015) already talked about this in her essay. If we want to know the exact level of the students, more aspects require attention.

Due to that latter issue, the answer to the second research question is no. I would not recommend *Text Inspector* for official students' assessment for the moment. Despite that, it

can be of a great utility for English teachers to get to know the approximate level of their students to improve lesson plans, creating materials, etc. It can also be used with adult learners to teach them how to be aware of the elements which compose a good writing.

Anyway, further research is encouraged because due to space restrictions, this study left out important issues such as checking with samples if the words in the texts are well-tagged and categorized or comparing the learner texts with corpora written by native speakers of the same educational level. For instance, a study in meta-discursive markers could be done contrasting with those in native texts. Learners use the same markers and, in a quite similar amount, but maybe this differs from the natives. The possibility to compare with a reference corpus would be an idea for *Text Inspector* to offer more information.

All in all, this undergraduate dissertation can serve only as a preliminary of a deeper survey on the topic and can help teachers who have an interest in optimising and sharping their classes to know more tools and evaluative instruments. People usually not academically engaged in linguistics might acknowledge the relevance of grappling with English language deeply, new lines of enquiry, and technological upgrades and tendencies, which can support a better development of its teaching, and students' learning.

REFERENCES

Anthony, L. (2019). *AntConc* [computer software]. Waseda University.

Başal, A., & Aytan, T. (2014). Using web 2.0 tools in English Language Teaching. In *Conference proceedings* (p. 372–375). Libreriauniversitaria. it Edizione.

Başal, A. (2016). *The use of Web 2.0 tools in ELT.* [PDF] https://www.researchgate.net/publication/309731877_The_use_of_Web_20_tools_in_ELT

Bax, S. *Text Inspector* [computer program]. Accessed 18 November 2022 at https://textinspector.com/

Bennett, G. (2010). *Using corpora in the language learning classroom: Corpus Linguistics for teachers.* University of Michigan Press.

Biber, D., Reppen, R., & Conrad, S. (2007). *Corpus linguistics: Investigating language structure and use.* Cambridge University Press.

Cambridge University Press. (n.d.). Cambridge Dictionary. Accessed 30 November 2022 at https://dictionary.cambridge.org/

Cobb, T. VocabProfilers v.8.5 [computer program]. Accessed 18 November 2022 at https://www.lextutor.ca/vp/

Collins. (n.d.). Collins Dictionary. Accessed 30 November 2022 at https://collinsdictionary.com/

Corpas, G. & Sehgiri, M. (2007). ReCor v.2.0 [computer software]. Accessed 16 November at University of Málaga.

Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment.* [PDF]. https://rm.coe.int/1680459f97

Council of Europe (2020). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Companion Volume*. [PDF] https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4

Dewey, J. (1916). *Democracy and Education.* CreateSpace Independent Publishing Platform.

Durán, P., Malvern, D., Richards, B., & Chipere, N. (2004). *Lexical diversity and language development: Quantification and assessment.* Palgrave Macmillan Ltd.

Eleyan, D., & Othman, A., & Eleyan, A. (2020). *Enhancing Software Comments Readability Using Flesch Reading Ease Score. Information.* [PDF] https://www.researchgate.net/publication/344162323_Enhancing_Software_Comments_Readability_Using_Flesch_Reading_Ease_Score/download

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32(3), 221–233. https://doi.org/10.1037/h0057532

Flesch, R. (1994). *The art of readable writing*. Wiley.

Francis, W. N. (1992). Language corpora B.C. In *Directions in corpus linguistics: Proceedings of Nobel symposium 82, Stockholm, 4-8 August 1991* (pp. 17–32). Mouton de Gruyter.

Friginal, E. (2018). *Corpus linguistics for English teachers: new tools, online resources, and classroom activities*. Routledge.

Gunning, R. (1968). *The technique of clear writing*. McGraw- Hill Book Co.

Hussin, A. (2015). *Refining the Flesch Reading Ease Formula for Intermediate and High-Intermediate ESL Learners*. [PDF] https://www.researchgate.net/publication/329921610_Refining_the_Flesch_Reading_Ease_Formula_for_Intermediate_and_High-Intermediate_ESL_Learners

Jindal, P. & MacDermid J, C (2017). *Assessing reading levels of health information: uses and limitations of Flesch formula* [PDF] https://www.educationforhealth.net/temp/EducHealth30184-5388625_145806.pdf

Krashen, S. (1982). *Principles and practice in Second Language Acquisition*. Elsevier.

Leńko-Szymańska, A. (2015). *The English Vocabulary Profile as a benchmark for assigning levels to learner corpus data*. [PDF]

https://www.researchgate.net/publication/275021429_The_English_Vocabulary_Profile_as_a_benchmark_for_assigning_levels_to_learner_corpus_data

McEnery, T., & Wilson, A. (2001). *Corpus Linguistics: An Introduction.* 2nd Edition, Edinburgh University Press.

McLeod, J., & Vasinda, S. (2008). Critical literacy and web 2.0: Exercising and negotiating power. *Computers in the Schools*, *25*(3-4), 259–274. https://doi.org/10.1080/07380560802394815.

Olohan, M. (2004). *Introducing corpora in Translation Studies.* Routledge.

Pearson, J. (1998). *Terms in context.* John Benjamins.

Piaget, J. (1964). Part I: Cognitive development in children: Piaget development and learning. *Journal of Research in Science Teaching*, *2*(3), 176–186. https://doi.org/10.1002/tea.3660020306

Proctor F. (2012). *Functional and corpus approaches of English grammar.*

Nyx Academics LLC.

Russell, J., & Sorge, D. (1998). *Training facilitators to enhance technology integration.* Journal of Instruction Delivery Systems. Retrieved 20 November 2022 from https://eric.ed.gov/?id=EJ608467

Seghiri, M. (2014). *Too big or not too big: Establishing the minimum size for a legal ad hoc corpus* [PDF] http://www.researchgate.net/publication/269808823_Too_big_or_not_too_big_Establishing_the_minimum_size_for_a_legal_ad_hoc_corpus

Simon, E. (2008). *Foreign language faculty in the age of web 2.0.* [PDF] https://er.educause.edu/articles/2008/8/foreign-language-faculty-in-the-age-of-web-20

Sinclair, J. (1991). *Corpus, concordance, collocation.* Oxford University Press.

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

Stahlberg, J. (2021). *Vocabulary profiles of authentic texts used by upper secondary English teachers. A lexical analysis of authentic texts used in EFL classrooms* [PDF] https://www.diva-portal.org/smash/get/diva2:1527767/FULLTEXT02

Strayer, J. (2007). *The effects of the classroom flip on the learning environment: a comparison of learning activity in a traditional classroom and a flip classroom that used an intelligent tutoring system* [Doctoral dissertation, Ohio State University]. OhioLINK Electronic Theses and Dissertations Center. http://rave.ohiolink.edu/etdc/view?acc_num=osu1189523914

Swales, J. M. (2006). Corpus linguistics and English for academic purposes. In *Information technology in languages for specific purposes.* 1(2), 19-33. MA.

Thomas, J. & Baisa, V. *VersaText* [computer software]. Accessed 18 November 2022 at https://versatext.versatile.pub/

Thomas, J. (2020). *Two web-based tools for learning language from language.* [PDF] https://www.versatile.pub/uploads/8/1/6/3/81634112/two_web-based_tools_for_learning_language_from_language._thomas_2020_v2.pdf

Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. John Benjamins.

Vannestål, M. E., & Lindquist, H. (2007). Learning English grammar with a corpus: Experimenting with concordancing in a university grammar course. *ReCALL*, 19(3), 329–350.https://doi.org/10.1017/s09583440070006