International Journal on Advanced Science Engineering Information Technology

Vol.10 (2020) No. 5
ISSN: 2088-5334

Barriers in Reading Comprehension of University Students: Analysis of the Complicated Words Annotated in the VYTEDU-CW Corpus

Jenny Alexandra Ortiz-Zambrano[a], Arturo Montejo-Raéz[b]

[a] *Software career, University of Guayaquil, Cdla. "Salvador Allende" - Av. Delta and Av. Kennedy, Guayaquil, Ecuador*
E-mail: [1]jenny.ortizz@ug.edu.ec

[b]*CEATIC -Centre for Advanced Studies in Information and Communication Technology, The University of Jaén, "Las Lagunillas" campus, s/n, 23071, Jaén, Spain*
E-mail: amontejo@ujaen.es

***Abstract*— Students often require a greater understanding of the lexicon that teachers use when dictating an assignment in class or in written texts as supporting material. Identifying and labelling difficult words has allowed us to examine the problem. A sample of students from the University of Guayaquil (Ecuador) was taken to experiment in a corpus of video transcripts that correspond to the different careers. After performing the analysis of the tagged words, the conclusions reached by other research papers in lexical simplification are confirmed and corroborates the recommendations of the Easy Reading guide prepared by Inclusion Europe in 2009. The investigation determined that the words labeled as difficult were specialized words, common lexical words, slang, English words, acronyms, among others. It was difficult for students to understand its meaning; in some cases, they either ignored its definition or just had the wrong idea of the lexicon. This work aims to be a contribution to future research in the area of lexical simplification applied to the development of solutions for detecting difficult words in the university academic field. Also, the type of complex expressions identified in the VYTEDU-CW corpus were characterized by the software, which enriches this resource while opening the possibility to organize a workshop where to promote research in the detection of difficult words to the Spanish. The support to validate these solutions is available to the scientific community.**

***Keywords*— complex words; university scope; lexical simplification.**

I. INTRODUCTION

A. Barriers in Reading Comprehension

Comprehension consists of understanding the content presented in a text. One of the factors that hinder linguistic understanding is ignoring the lexicon and its meaning. The success or failure of the reader's understanding of a book will depend on the knowledge or ignorance of the terms that compose it [1]. For many people, the content of the documents represents a barrier in their understanding as it contains difficult information due to the use of sophisticated and specialized vocabulary. The presence of compound words, long sentences, or passive sentences, as well as uncommon words in the content of texts, implies a difficult understanding, especially for those with cognitive disabilities [2].

The Inclusion Europe organization, with its project "Creating paths to adult education for people with intellectual disabilities", has established European rules that help people make information easy to read and understand. This project involved not only people with intellectual disabilities but also other people with learning difficulties for different reasons, such as second language learners with low literacy [3].

Higher education institutions expect to receive students whose competences in reading comprehension have reached an appropriate level to face successful university studies. However, practice this does not happen since it is permanently evident that students have difficulty in understanding what they read [4]. This work constitutes a fourth phase in the development of the research of a doctoral project in the area of Text Simplification proposed by the University of Jaén (Spain) for the University of Guayaquil (Ecuador), as follows:

The first phase consisted of the construction of the VYTEDU corpus available at http://jour-nal.sepln.org/sepln/ojs/ojs/index.php/pln/ article/view/5512 as a resource essential for research in Spanish at the university scope [5]. The application of complexity metrics in the texts of the VYTEDU corpus was carried out in the second phase [6]. The words identified and labelled as compounds contained in the documents of the VYTEDU

corpus were registered in VYTEDU-CW [7] using a software application; this work was part of the third phase of the project. The VYTEDU-CW corpus is the result of manual labeling of the problematic words identified and tagged by university students (annotators) in the texts of transcripts of educational videos of the VYTEDU corpus. The main objective of this work is to analyze the barriers that academic texts contain that cause difficulty in linguistic understanding in university students.

B. Terminology

The concepts related to readability are diverse: understanding, complexity, simplification. Some of the definitions given by the literature support the work done:

1) Readability: It is the ease with which you can read and understand a text [8]. To achieve the legibility of a docuement is necessary to consider several factors, including the presentation of the writing, the style, the clarity of the content, the way of writing, the language used, among others. Different typefaces and punctuation marks are some of the elements that help the reader understand what he is reading. Linguistic readability involves using the most basic sentence structures, that is, subject-verb-complement, to make the text more readable. Making useful information means making information that is easy to read and understand [3].

2) Compound sentence: It is a sentence considered difficult for the reader, so its simplification would be advisable [9].

3) Complex word: word found difficult to read or understand, possibly a long or unusual term, which needs to be replaced by a simpler one [9].

4) Text Simplification: It is the process of transforming a text into a new one to reduce its lexical or syntactic complexity while retaining its original meaning [2]. The simplification process employs some strategy for adapting the text that can be the division of sentences or lexical substitution. Simplification of texts can be significant for many people, making information accessible to all [10].

5) Lexical Simplification: It consists of replacing a word in each context with a synonym that is easier to understand. Its objective is to succeed not only terms, but also confusing sentences with simple ones; that is, a word substitution task, where the purpose is to find a synonym that is, in a sense, simpler than the original word [11]. However, carrying out this task constitutes a challenge since the substitution must preserve both the meaning and the grammaticality of the original sentence [12].

II. MATERIALS AND METHOD

One of the first notable systems is LexSiS (Lexical Simplification for Spanish) [13] which executes a lexical text simplification algorithm for Spanish and represents the first computational implementation. LexSiS applied three techniques to select the most straightforward synonym, to replace the word with its simplest substitute, used a word vector model to identify the correct meaning, the criteria of simplicity established in the frequency, and the length of the word. All this served to demonstrate that research work in the area of lexical simplification can have resulted in making use of available resources, such as the body of documents in Spanish taken from the web.

After evaluating the results, LexSiS proposed simpler synonyms than the frequency baseline, also offering more lexical substitutions as well as a more significant percentage of meaning preservation. The process did not consider replacing sentences but only changing individual words. The objective of identifying complex words was to find out which words to simplify (substitute) in a sentence [12]. The research focused on the needs of non-native second language, so he sought to provide studies to produce more reliable data and create innovative solutions.

This proposal implemented a strategy for CWI (identification of complex words) based on lexicons, that is, on the vocabulary of people (their known language).This study took 40 sentences in English and involved 51 volunteers of different nationalities, asking them to select the words they found complex, that is, the meaning of which they did not understand. The CW, LexMTurk, and Wikipedia corpus were the resources that contributed to the investigation. Some sentences contained between 20 and 30 words. Very excellent results obtained in terms of lexical simplification based on new types of substitution classifiers and new resources for the limitations of modern second language approaches such as spoken text language models.

Previous study proposed a text simplifier for English: the YATS project (Yet Another Text Simplifier) to improve readability and comprehension of the text to help people with intellectual disabilities [10]. Four phases comprised the development of the proposed project: 1) document analysis, 2) detection of compound words, 3) disambiguation of the meaning of words, 4) more frequent synonyms (ranking), and understanding of language. A vector space model approach was used to obtain the most appropriate meaning of the difficult word in a defined context and implemented measures of word frequency simplicity. The experimental results showed superior performance in the lexical simplification component, as well as excellent precision in syntactic simplification.

Various techniques were implemented for the simplification of Czech texts by applying some resources, including a small corpus called "Complex" [9]. Also, other implementations using various lexical readability measures contributed to the Czech books, which meant an advance in research in lexical simplification. Strategies implemented for lexical simplification, executed, and evaluated several methods for the identification of compound words and substitution selection.

In the process, he began marking the words he had identified as complex and then asked three annotators to verify the annotations and possibly add or delete some words identified as complex. The annotators received instructions to simplify each of the sentences where it was necessary to recommend that the challenge would be that while retaining the meaning, the resulting text should be easier to understand. From the results, he concluded that the strategy for the identification of complex words based on frequency turned out to be very good since it manages to choose a

significantly more straightforward substitution in half the time compared to the strategy based on precision.

A. The Corpus: VYTEDU y VYTEDU-CW

1) The VYTEDU corpus: (Videos and Transcripts in the Educational field) was born in 2017 from a doctoral project proposal from the University of Jaén (Spain) for the University of Guayaquil (Ecuador), taking as a premise the shortage of resources for the language Spanish and particularly in the educational field at a higher level. VYTEDU constitutes a fundamental data resource to continue with the advances in research in the Spanish text simplification systems that support education. (see Fig. 1).

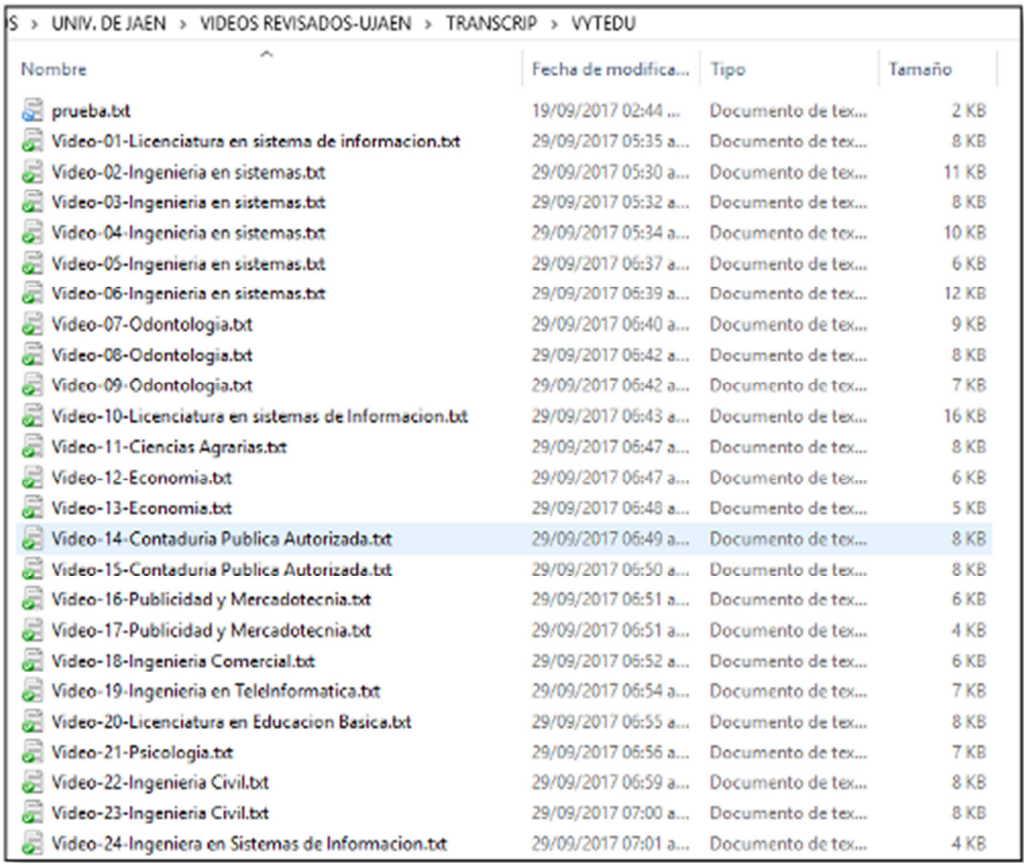


Fig. 1 VYTEDU corpus texts

The SEPLN (Spanish Society for Natural Language Processing) congress held in the city of Murcia (Spain) in 2017 [5], made its presence through a demonstration. VYTEDU contains the transcripts of 55 educational videos made to a group of teachers in the classrooms of the University of Guayaquil about the various topics that correspond to various subjects of the academic programs offered in this educational institution. This resource currently has several experiments carried out with students in the area of higher education, the results make a significant contribution to the considerations to overcome barriers in reading comprehension. Next, in the table I are the VYTEDU corpus statistics.

TABLE I

STATISTICS OF THE VYTEDU CORPUS

	Min	Max	Average	Total
Number of videos				55
Duration	0:05:01	0:21:08	0:10:18	9:26:32
Size en Mb	4,9	2.645	804,5	44.248,1
Number of words	465	2.646	1.244	68.414
Number of paragraphs	6	29	12,24	673

2) The VYTEDU-CW (Videos and transcripts in the educational field - Complex words) corpus was born in the third stage of the research project. Its objective is to create a data set with the problematic (complex) terms contained in the VYTEDU corpus documents, identified, and labeled by the students. The strategy of annotated carried out, software with free tools were built, which was called EIL (Language Research Environment), where the texts of the VYTEDU corpus were loaded. The annotated process takes into consideration the research work proposed by [13,9,16,17] on the lexical simplification project for Spanish and the lexical simplification for Czech, respectively.

A minimum number of three annotators (students) from the different careers of the University of Guayaquil were taken to label the difficult Words of the topics proposed in the corpus texts corresponding to their specialty and semester of study. The sample of students was according to their level of preparation regarding the content of the reading (see Fig. 2), which was essential to achieve the labeling objectives.

The data set that makes up VYTEDU-CW are seven fields:

- the problematic word,
- the type of word (difficult or empty),

- the name of the documents from which it comes,
- the initial location of the word in the text,
- the length of the word in characters,
- the position of the line break,
- and the number of times the problem word appears in the document.

An example is presented (see Fig. 3).

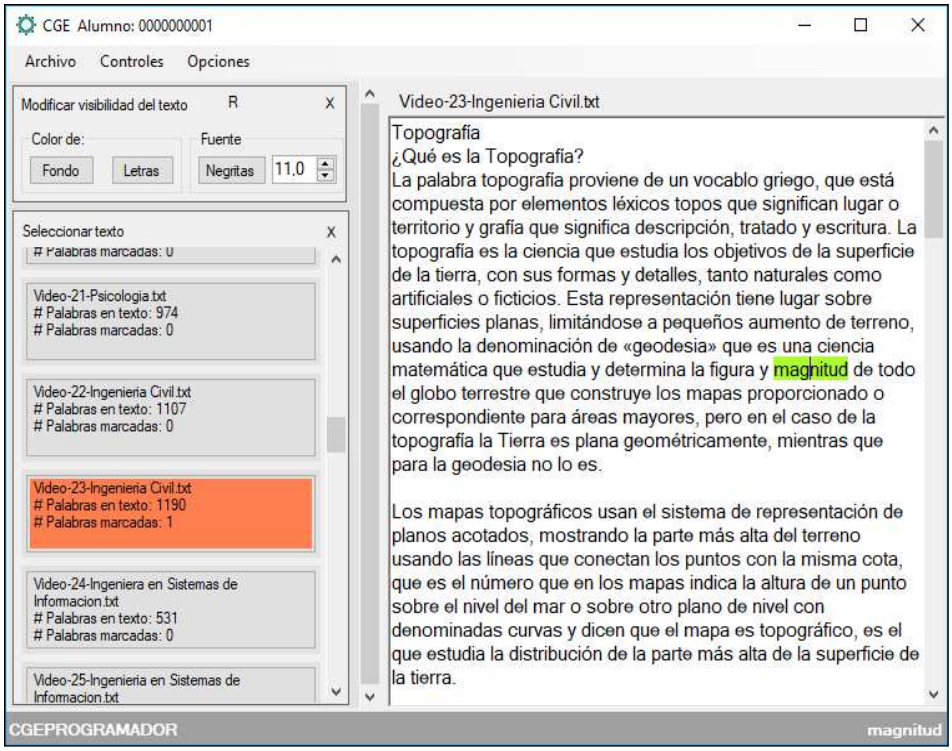
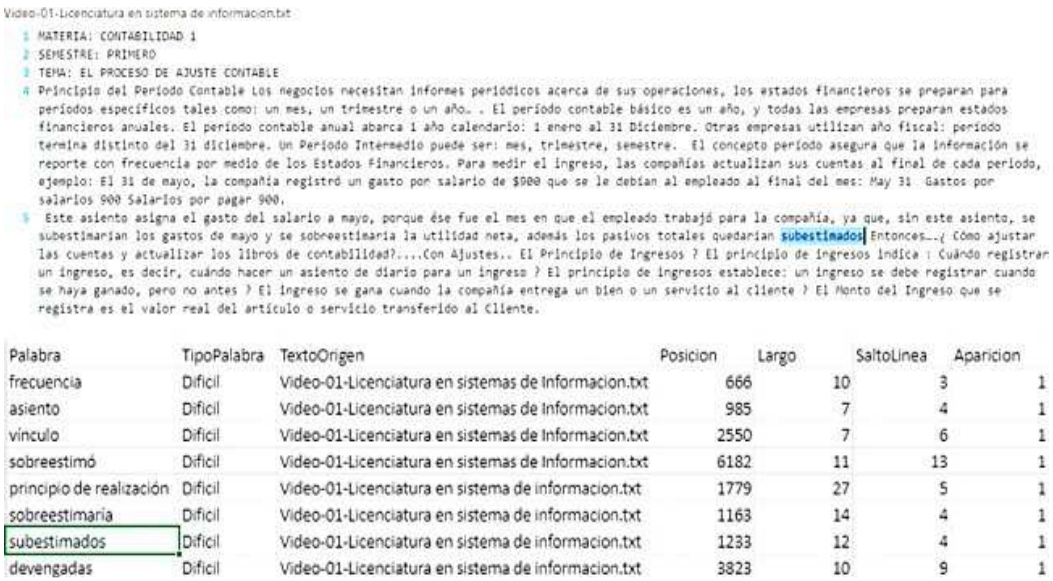


Fig. 2 Labelling of compound words in the VYTEDY corpus


Video-01-Licenciatura en sistema de informacion.txt

1 MATERIA: CONTABILIDAD 1
2 SEMESTRE: PRIMERO
3 TEMA: EL PROCESO DE AJUSTE CONTABLE
4 Principio del Periodo Contable Los negocios necesitan informes periódicos acerca de sus operaciones, los estados financieros se preparan para períodos específicos tales como: un mes, un trimestre o un año. . El período contable básico es un año, y todas las empresas preparan estados financieros anuales. El período contable anual abarca 1 año calendario: 1 enero al 31 Diciembre. Otras empresas utilizan año fiscal: período termina distinto del 31 diciembre. Un Periodo Intermedio puede ser: mes, trimestre, semestre. El concepto período asegura que la información se reporte con frecuencia por medio de los Estados Financieros. Para medir el ingreso, las compañías actualizan sus cuentas al final de cada periodo, ejemplo: El 31 de mayo, la compañía registró un gasto por salario de $900 que se le debían al empleado al final del mes: May 31 Gastos por salarios 900 Salarios por pagar 900.
5 Este asiento asigna el gasto del salario a mayo, porque ése fue el mes en que el empleado trabajó para la compañía, ya que, sin este asiento, se subestimarían los gastos de mayo y se sobreestimaría la utilidad neta, además los pasivos totales quedarían subestimados Entonces…¿ Cómo ajustar las cuentas y actualizar los libros de contabilidad?....Con Ajustes.. El Principio de Ingresos ? El principio de ingresos indica : Cuándo registrar un ingreso, es decir, cuándo hacer un asiento de diario para un ingreso ? El principio de ingresos establece: un ingreso se debe registrar cuando se haya ganado, pero no antes ? El ingreso se gana cuando la compañía entrega un bien o un servicio al cliente ? El Monto del Ingreso que se registra es el valor real del artículo o servicio transferido al Cliente.

Palabra	TipoPalabra	TextoOrigen	Posicion	Largo	SaltoLinea	Aparicion
frecuencia	Dificil	Video-01-Licenciatura en sistemas de Informacion.txt	666	10	3	1
asiento	Dificil	Video-01-Licenciatura en sistemas de Informacion.txt	985	7	4	1
vinculo	Dificil	Video-01-Licenciatura en sistemas de Informacion.txt	2550	7	6	1
sobreestimó	Dificil	Video-01-Licenciatura en sistemas de Informacion.txt	6182	11	13	1
principio de realización	Dificil	Video-01-Licenciatura en sistema de informacion.txt	1779	27	5	1
sobreestimaria	Dificil	Video-01-Licenciatura en sistema de informacion.txt	1163	14	4	1
subestimados	Dificil	Video-01-Licenciatura en sistema de informacion.txt	1233	12	4	1
devengadas	Dificil	Video-01-Licenciatura en sistema de informacion.txt	3823	10	9	1

Fig. 3 Complex words tagged by career annotators Degree in Information Systems

In the example shown in Fig. 3 the word "*underestimated*" has been identified and labeled by the annotator, thus registering its absolute position in the text (starting a character at position 1233), its length (12 characters), and the number of times it appears in the document. According to the analysis of the terms, there are phrases that the scorer has classified as complicated. An example found in the data obtained from the phrase "principle of realization", it is in the absolute position 1779, the length is 27 characters and appears only once in the document.

The global results obtained from the VYTEDU-CW corpus showed fundamental data that will contribute to research on lexical complexity and barriers to reading comprehension.

Here are the details:

- 1709-word annotations the annotators considered

challenging to understand.

- A total of 430 students accessed the software only once and interacted based on the labeling of difficult words.
- 573 admissions made through the app; some students decided to carry out experiments with other texts whose content they learned in subjects reviewed in the course.
- A total of 723 words identified as different difficult.
- 250 users tagged words that other users had not selected, that is, that did not match those written down by other students.
- There is a total of 9175 different words in the VYTEDU corpus.
- 2.5% is the percentage of words labeled as complicated concerning the total number of terms contained in the VYTEDU corpus texts.

The overall results of the VYTEDU-CW corpus are presented in Table II.

TABLE II
GLOBAL RESULTS OF THE VYTEDU-CW CORPUS

All records of the VYTEDU-CW corpus	**Total**
# of annotated words	1709
# of annotators who entered only once	430
# of entries to the application	573
Records without repetition in the VYTEDU-CW corpus	
# of annotated words	723
# of annotators	250
Corpus VYTEDU	
# total words	68248
vocabulary size (# of terms without repetition)	9175
VYTEDU Vs. VYTEDU-CW	
% of complicated words identified	2.5%

III. RESULTS AND DISCUSSION

A. Corpus VYTEDU-CW statistics by career

The VYTEDU-CW dataset also contains other fields corresponding to each text read from the VYTEDU corpus, such as:

- the word identified and labelled as complex,
- the student's identification,
- the name of the document read,
- the initial position in the text of the word complicated,
- word length,
- date and time of the creation of the annotation.

It was necessary to filter the data of the VITEDU-CW corpus by the field of interest for its analysis. Then it was required to delete the records of the students who did not label any word as difficult.

Some words labeled as difficult that are elementary, such as: *only, here, the order, the terms, the title*, among others, it was necessary to eliminate them from the list, such as the research carried out in lexical simplification [9]. The process was the same for the data for each career. Finally, the last data were obtained for their analysis of the summary of corpus VYTEDU-CW statistics as follows (See Table III).

TABLE III
SUMMARY OF CORPUS VYTEDU-CW STATISTICS

Careers	**Total number of confusing words labeled**	**Total number of severe sentences**	**VYTEDU-CW Vs. VYTEDU**	**Word length**				**Average placement of the word in the text**
			% of complicated words in the text	**min**	**max**	**average**	**Mode**	
Degree in Information Systems	130	2	5.67	3	29	9.27	11	4213.31
Networking	72	1	3.26	2	16	6.40	3	5733.74
Odontology	11	0	0.96	5	11	7.36	2	4783.91
Research Unit	784	7	65.22	1	82	10.51	47	3187.50
Law	56	0	4.78	5	17	9.82	3	4387.36
Advertising and Marketing	24	0	5.16	5	16	10.83	3	1399.04
Architecture	18	0	1.11	3	14	9	3	4854.72
Authorized Public Accounting	64	0	6.64	3	20	9.31	3	3307.33
Psychology	5	0	0.46	8	15	11	1	3315.2
Economy	24	1	2.63	3	6	8	1	2876.63
Civil Engineering	100	5	10.20	3	7	7	5	828
Tele-Computing Engineering	35	5	3.19	6	12	21	3	2635
Agricultural Sciences	53	0	5.05	4	13	13	3	4411.23
Biology	18	1	0.98	2	20	6.83	4	4249.22
Environmental Engineering	24	0	2.41	3	17	8.88	5	1943.58
Medicine	11	0	0.82	3	10	7.36	2	3829.36
Information Systems	30	0	4.26	6	12	21	3	2730.02
Degree in Basic Education	32	4	2.71	7	42	12.72	4	2481.31

B. Analysis of Corpus VYTEDU-CW by Carrera

The analysis of the words labeled as complex confirms the conclusions reached by the research work of [14] and corroborates the recommendations of the Easy Reading guide prepared by [3], as well as the guidelines proposed in the "EASY TO READ" guide in the automatic identification of complex words [15] (CWI).

These research papers have attracted attention in recent years, with the advent of deep learning approaches [16] and multilingual challenges [17], which contributes to the evaluation of words labeled as severe, such as specialized words, common lexicon words, slang, English words, acronyms, among others. Given these terms, students had a hard time understanding; in some cases, they ignored its meaning or had some idea or notion of it. Here are some examples:

1) The use of technicalities by teachers: This example happens when teaching is in their classes (technical voices used in the language of science) causes barriers in linguistic understanding, as is the case of the career in "Degree in Information Systems", some of the tagged words were *break, for, while, variable.* In the "Networking" career, the words are *cluster, Hosting, kernel*. Other examples found are such as in the career of "Odontology", the words are *deflection, homeostasis, muñón*. In the "Architecture" career, the technical words used by teachers with students who are not of the corresponding level, also cause difficulty in their understanding, for example, words tagged as difficult*: poetry, hepatization, hydrolyzing*.

2) The use of sophisticated vocabulary: Some teachers use an improved, cultured terminology, which makes it challenging to understand the content for many students in the classroom. For example, students of the "Odontology" career labeled words such as *contributive, alienation, tariff, optional, praxis, or revocation*. In the "Authorized Public Accounting" career, words such as *inference, subsidiary, or contingencies*. In the case of the "Research Unit", it was determined according to the analysis performed that the teacher uses a lexicon so sophisticated and unique that very few understand words such as *ultrasonic, taxonomic, maceration, genome, carto-print, bioethics, or synthetization*.

3) The use of teacher abbreviations: It is also a cause of difficulty in reading comprehension. It possible to determine that in the case of the students of the "Networking" career, words corresponding to their abbreviations were labeled, such as *Mbps, IPAN, GEAR, LAN*. In the case of the "Research Unit" degree, the teacher used the acronym *ISBN*, which caused difficulty (barrier) for the students in their understanding, this occurred because the teacher did not explain its meaning. Another problem that exists is that teachers propose examples in class using words that do not belong to the level of education or specialization. While they teach their classes, they use words that even the students of a certain level of studies do not know, as is the case of the students of the "Degree in Information Systems" career, words are such as *routers, tree structure, script*. In the case of the career in "Engineering in Agricultural Sciences", the term: Python, is a term of another science; its understanding is difficult in students of another specialization. In the case of the "Research Unit" degree, the teacher used computer terms with the group of students who are of different specializations and are the ones who take the course to train and be able to prepare their end-of-course research work. Analysis of the data found that many of the students had difficulties in linguistic understanding as they identified and labeled various terms. The results showed that they were students who did not belong to the specialty. Some of the words tagged severe were: virtual, pixels, logical operators, jpg.

4) The use of metaphors and figurative language. It causes difficulties when it comes to an understanding what the teacher wants to convey, and such is the case of the students of the "Research Unit" who labeled the phrase "*if you want to describe the truth, leave the elegance for the tailors.*"

*5) The use of long, hard-to-pronounce and unusual words***:** It makes it difficult to recognize them immediately. To cite several examples. In the "Research Unit", the students labeled words that represent the use of terms of difficult pronunciation, such as *qualitatively, quantitatively, chronologically, consent, particularly*. In the "Law" career, students tagged words with long terms, such as *homogeneity, consideration, methodological, interculturality appear*. In the "Advertising and Marketing" career, long words were identified, such as *imprescriptible, inalienable, decentralized.* In the *"*Economy*"* career, the words*: decentralized, totalitarianism*, they represent the use of long words. In the degree of "Civil Engineering", some of the terms that the students considered difficult to understand because they are "infrequently used words" were: *correlate them, deontology, sclerometer, workability*. In the "Agricultural Engineering" degree, some words represent the use of difficult pronunciation terms, such as *unanimously, carrageenan, paleobotany, carrageenan, oleoresins*. In the "Authorized Public Accounting" degree, a term that most students labeled a complicated word was a *concession*. In the "Degree in Information Systems" career, the words: *oversized, overestimated*, were labeled as severe because they are considered long words. In the "Agricultural Science Engineering" career, the term *unanimously* represented the use of word of hard pronunciation.

6) The use of verb nominalization: it was also a cause for word tagging. In the "Networking" race, in the word: *emphasized*, it presents the nominalization of the verb emphasize. In the "Law" career, the words labeled complex were: scrutinized, emanated, revalued. These words were nominalized verbs; that is, verbs wrote using suffixes. In the "Research Unit" career, the words: *delimited, conditioned, synthesized, standardized*, are also examples of verbs that make use of nominalization using the use of suffixes. In the "Degree in Information Systems" career, the use of suffixes in the verb caused difficulty in understanding the word: *digitized*. In the career "Engineering in Agricultural Sciences," suffixes were also used, such as the word: *enchanted*. In the "Advertising and marketing" career, the presence of suffixes made it challenging to understand the word: *decentralized*.

7) The use of compound words: it caused students difficulty in understanding. In the “Degree in Information Systems” career, the expressions: *intermediary nucleus, tree structure*, they are examples of some compound words found in the texts. In the "Research Unit" career, the words: *ultrasonic, taxonomic view*. In the "Economy" career, in the term: *stock market circuit*, the union of two or more simple words was a cause of the difficulty in reading comprehension. Another labeled example that causes a problem in linguistic understanding is in the "Civil Engineering" career document, in the compound word: *flexy traction*. The career of "Telecommunications Engineering" was no exception when presenting the *pre-established* word as a word identified as complex contained in the evaluated text.

The documents do not contain terms with the use of colloquial language, which could explain that it uses an academic vocabulary because the contents of the texts correspond to transcripts of university videos.

IV. Conclusions

In recent decades, the automatic simplification of texts has made very substantial contributions to the information society. The Lexical Simplification is a fundamental part of the Automatic Simplification of Texts; it contributes significantly in the application of techniques that contribute to the linguistic understanding of people, for one reason or another, they have difficulties in accessing information, especially those who have abilities specials. The University of Guayaquil, being a public university and the largest in the country, has diversity in students. The contribution of the scorers is highly valued.

For the labeling of the words, the students carried out the activity through several sessions in different careers. The annotators were students of different ages, low motivation, high capacities, some in situations of social risk; others come from various geographical points of the country associated with disadvantaged social, economic, or cultural conditions. It is necessary to emphasize that there are research works carried out applying Lexical Simplification in Spanish in other areas of study.

This article presents the analysis of the confusing words contained in the VYTEDU-CW corpus, becoming a pioneering work proposed to continue advancing research in Lexical Simplification for the field of higher education in Spanish. The data analysis was performed using free software tools and based on the statistical values obtained from the words labeled by the annotators. This article constitutes a valuable contribution to lexical simplification, since its development represents a way to corroborate other previous studies. Among other things, this analysis helps characterize the type of complex words identified in the VYTEDU-CW corpus, which makes it located in a privileged position to continue as future work using this corpus for the development of complicated words detection solutions in the university field. The resource is now available to validate these solutions.

The results obtained from the analysis of the words labeled as complex ratify the conclusions of other research papers, as well as the recommendations presented by the Easy Reading guide prepared by Inclusion Europe. It was evident that some of the words labeled as difficult were within the category of specialized words, that is, terms specific to a specific subject of the subject. Other words belonged to the common lexicon; additional terms were words of use diary; others were in another language; some names were acronyms, among others.

Following the recommendations of the guidelines for easy reading, we suggest working on the inclusion of content enrichment resources, such as external visual explanatory notes and, or audio records, that contribute to the reader's understanding. The analysis of this article offers the possibility of planning workshops for teachers of the University of Guayaquil, to receive training on the recommendations presented in the *Easy Reading* guide so that making use of these guidelines contribute to student learning, especially those who have difficulties in reading comprehension. In the same way, this article opens the possibility to organize a workshop to promote research in the detection of confusing words for Spanish in the educational field. The proposal will be to achieve the lexical substitution of the problematic terms contained in VYTEDU-CW as the next research objective.

References

[1] González, R. Reader in Initial University Students. *Person*, (001), 43-65 (1998).

[2] Torunoglu-Selamet, D., Pamay, T., Eryigit, G. Simplification of Turkish Sentences. The First International Conference on Turkic Computational Linguistics, 55–59 (2016).

[3] Europe, I. Information for everyone. European rules for making information easy to read and understand (online). Retrieved from https://plenainclusion.org/sites/default/files/informacion_todos.pdf (2010).

[4] Neira Martínez, A. C., Reyes Reyes, F. T., & Riffo Ocares, B. E. Academic experience and reading comprehension strategies in first-year university students. Literature and linguistics, (31), pp. 221-244 (2015).

[5] Ortiz, J., Montejo-Ráez, A. VYTEDU: A corpus of videos and transcriptions for research in the education domain. In: SEPLN-Spanish Society for Natural Language Processing, vol. 59, pp. 167–170 (2017).

[6] Ortiz, J., Varela, E.: Reading Comprehension in University Texts: The Metrics of Lexical Complexity in Corpus Analysis in Spanish. In: First International Conference 2018, vol. 959, pp. 111–123. Springer, Nature Switzerland AG (2019).

[7] Zambrano, J. O., MontejoRáez, A., Castillo, K. N. L., Mendoza, O. & Perdomo, B. VYTEDU-CW: Difficult Words as a Barrier in the Reading Comprehension of University Students. The International Conference on Advances in Emerging Trends and Technologies (pp. 167-176). Springer, Cham (2019, March)

[8] Belart, V. F. The Legibility: A fundamental factor in understanding a text. Primary care, 34(3), 143-146 (2004).

[9] Burešová, K. Text Simplification in Czech. Ph.D. Disertation (2017).

[10] Ferrés, D., Marimon, M., Saggion, H. YATS: Yet Another Text Simplifier. International Conference on Applications of Natural Language to Information Systems, pp. 335-342. Springer, Cham (2016, June).

[11] Saggion, H., Štajner, S., Bott, S., Mille, S., Rello, L., Drndarevic, B.: Making It Simplext: Implementation and Evaluation of a Text Simplification System for Spanish. ACM Trans. Accessible Comput. (TACCESS) 6(4), 14 (2015).

[12] Paetzold, G. H. Reliable Lexical Simplification for Non-Native Speakers, pp. 9–16 (2015).

[13] Bott, S., Rello, L., Drndarevic, B., Saggion, H. Can Spanish Be Simpler? LexSiS: Lexical Simplification for Spanish. Proceedings of COLING 2012, (December), pp. 357–374. Retrieved from http://www.aclweb.org/anthology/C12-1023 (2012).

[14] Garcia, Ó. (2012). Easy Reading: Writing and evaluation methods. Royal Board on Disability, 140 (2012).

[15] Alarcon, R., Moreno López, L., Segura Bedmar, I., & Martínez Fernández, P. Lexical simplification approach using easy-to-read resources (2019).
[16] Gooding, S., & Kochmar, E. Complex Word Identification as a Sequence Labelling Task. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (pp. 1148-1153), (2019, July).
[17] Finnimore, P., Fritzsch, E., King, D., Sneyd, A., Rehman, A. U., Alva-Manchego, F., & Vlachos, A. Strong Baselines for Complex Word Identification across Multiple Languages. arXiv preprint arXiv:1904.05953 (2019).