UCLV
Universidad Central
"Marta Abreu" de Las Villas

FH
Facultad de
Humanidades

Departamento de Lengua Inglesa

TRABAJO DE DIPLOMA

Mejoramiento de la competencia traductora de los estudiantes a partir de *ParaCorp2020*, un corpus paralelo (español-inglés) de textos científico-técnicos

Autora: Lisset Rodríguez Aguila

Tutores: MSc. Profesor Auxiliar Humberto Miñoso Machado

Dra. P.T. Juana Idania Pérez Morales

Santa Clara

UCLV
Universidad Central
"Marta Abreu" de Las Villas

FH
Facultad de
Humanidades

English Language Department

DIPLOMA THESIS

Enhancing Students' Translation Competence through *ParaCorp2020*, an English-Spanish Sentence-Aligned Parallel Corpus of Science and Technology Texts

Author: Lisset Rodríguez Aguila

Supervisors: Associate Profesor Humberto Miñoso Machado, BE.d

Full Professor Juana Idania Pérez Morales, PhD

Santa Clara

ACTA DE CONFORMIDAD PARA ESTUDIANTES DE PREGRADO

Universidad Central "Marta Abreu" de Las Villas

Por una parte: Lisset Rodríguez Aguila, estudiante de la carrera de: Lengua Inglesa con Segunda Lengua Extranjera (Francés) en la facultad de: Humanidades, en lo adelante **EL ESTUDIANTE**, con número de identidad permanente: 97022310612 o pasaporte: _____. Y, por otra parte: Juana Idania Pérez Morales, Jefe del Departamento Docente de: Lengua Inglesa en la ya mencionada facultad, en lo adelante **EL JEFE DE DEPARTAMENTO**, Humberto Miñoso Machado, profesor(es) encargado(s) de tutorar el Trabajo de Diploma de **EL ESTUDIANTE**, en lo adelante **EL TUTOR,** reconocen que:

I. A **EL ESTUDIANTE** se le ha aprobado como tema de investigación para su Trabajo de Diploma el titulado: Mejoramiento de la competencia traductora de los estudiantes a través de ParaCorp2020, un corpus paralelo español-inglés de textos científico-técnicos.

II. **EL ESTUDIANTE** no divulgará información concerniente a la investigación, tanto durante el desarrollo como tras la culminación de esta sin la debida autorización de **EL TUTOR** o **EL JEFE DE DEPARTAMENTO**.

III. Que el Trabajo de Diploma, fruto de la labor investigativa de **EL ESTUDIANTE** y la asesoría de **EL TUTOR**, resulta de **TITULARIDAD EXCLUSIVA** de la Universidad Central “Marta Abreu” de las Villas.

IV. **EL ESTUDIANTE,** una vez aprobada su tesis para la defensa, depositará una copia electrónica de la misma en el Repositorio Digital Institucional de la Universidad Central “Marta Abreu” de Las Villas.

V. A partir de la defensa y aprobación del Trabajo de Diploma, la publicación total, parcial o la elaboración de cualquier obra que se derive de esta investigación por parte de **EL ESTUDIANTE**, contará con la coautoría de **EL TUTOR** y viceversa, resultando de referencia obligada esta obra en cualquier otra que se elabore. El incumplimiento de esta cláusula, puede llevar consigo el inicio de procesos de plagio. Todo lo anterior de acuerdo a la normativa de Derecho de Autor vigente en Cuba.

Y para que así conste se firma la presente en la Universidad Central “Marta Abreu” de Las Villas, a los 27 días del mes de noviembre del año 2021.

_____	_____
EL ESTUDIANTE	**EL JEFE DE DEPARTAMENTO**
_____	_____
TUTOR	**TUTOR**

DEDICATION

To my best friend Adonai, for always being with me. To my beloved parents: nothing can repay all these years of love and dedication.

ACKNOWLEDGMENTS

To Adonai, for He has prepared my ways and has never left my side. To my parents and brother for all their love and unconditional support, for being my place of refuge and comfort.

To my supervisors, professors Humberto Miñoso Machado and Juana Idania Pérez Morales, because without their practical wisdom, infinite patience and invaluable guidance, this research could have never been done. You are the kind of professors one can never forget. Thank you for always encouraging me to go for more.

To my grandparents, for their permanent love, care and support.

To my friends, for their fondness and for always encouraging me in good and bad times.

To my dear classmates who have been with me sharing happiness and struggles during these six years. Maibe, thank you for being my umbrella in a rainy day.

To all those people who in one way or another contributed to the realization of this research. To the university and all the experiences that have made me fall and stand up again. To all,

THANK YOU!

ABSTRACT

Abstract:

This paper describes the use of an English-Spanish Sentence-Aligned Parallel Corpus of Science and Technology Texts for enhancing the translation competence of fourth-year students of the degree course English Language with Second Foreign Language (French), at Universidad Central “Marta Abreu” de las Villas. For this purpose, a corpus of 50 English science and technology texts and their translations into Spanish was compiled, including the domains of medicine, physics, chemistry and artificial intelligence. The dataset was collected and processed using corpus tools, such as *OmegaT* aligner, *Xbench*, and *AntConc*. The research adopted a mixed-method approach designed for collecting, analyzing, and interpreting both quantitative and qualitative data. The results from the survey applied to students evidenced some difficulties in choosing the right equivalent for some words and phrases in the process of translating science and technology texts. The design of the proposal called *ParaCorp2020* followed a process-oriented approach to translation. It provides a wide range of lexical, syntactical and textual examples found in the English texts, with their equivalents in Spanish, in order to help students, integrate the necessary knowledge and skills to successfully translate science and technology texts. Also, a word bank was derived from the corpus, as a supporting material for the acquisition and systematization of specialized terminology and a group of corpus-based activities was elaborated to aid the translation process in the classroom.

Keywords:

Alignment, parallel corpus, process-oriented approach to translation, science and technology texts, translation competence, translation process

TABLE OF CONTENTS

Table of Contents

INTRODUCTION

Introduction

Today's multicultural and multilingual society demands effective, efficient, and empathetic communication between languages and cultures. In this sense, translation has always been an important factor in life from unmemorable times until today. That is why it has increased, becoming a part in every major, and transferring the knowledge from one culture to another (Kaur, 2005). Translation is a courier for the transmission of knowledge, a protector of cultural heritage and essential to the development of global economy (Burrow-Goldhahn, 2021).

Therefore, the demands for competent translators is rapidly growing. Professionals with linguistic expertise, appreciation for other cultures, background knowledge on specialized areas, and who master translation strategies, are of great importance in society. Currently, learners of English as a Foreign Language in higher education institutions face multiple challenges related to the development of specific competences that are necessary for their professional performance as translators and interpreters. However, the acquisition of translation competence starts from the very process of language learning and professional training, mainly at university level. (Eser, 2015)

Translation competence development requires taking all the characteristic features of the competence into account. A distinction should be made between competence (the underlying system of knowledge) and performance (translating) (PACTE, 2003). Complex and difficult though it may seem, the aim of translation trainers is to provide their trainees with as rich and thorough translation education as possible (Bogucki, 2015).

Statement of the problem

In the case of the English Language with Second Foreign Language degree course at *Universidad Central "Marta Abreu" de Las Villas*, the development of the translation competence has become a priority for both, professors and learners, though students still have difficulties translating specialized texts, specifically science and technology texts.

There is a large number of scientific journals published worldwide. Several estimates point to around 30,000 with close to two million articles published each year (G.

Altbach, Philip and de Wit, Hans, 2018). With more and more works and papers on science and technology published in English, and their level of lexical complexity, the need to enhance the competence of the translators is imperious, since they are the bridge between the scientists and the readers.

The **problem** raised in this research lies not only on the difficulties students face when translating science and technology texts but also on the insufficient amount of offline bilingual consulting sources to enhance the translation competence in the classroom. Also, the students present difficulties at the time of selecting the correct equivalent in Spanish for a scientific or technical word in English, since they barely know the specialized vocabulary, and do not have enough background knowledge on science and technology related topics. Besides, little or none systematic analysis of already-made translations is carried out in class, so that students can see examples of the strategies and solutions provided by other translators, and thus enhance their translation competence. The study was carried out in the subject Translation of Science and Technology Texts taught in fourth year.

The problem evidenced leads the author to formulate the following **research question**: How could the translation competence of the fourth-year English Language students be enhanced in order to prepare them for accurately translating science and technology texts?

Consequently, the **overall aim** of this research work is:

To propose an English-Spanish Sentence-Aligned Parallel Corpus of Science and Technology Texts for the enhancement of the translation competence.

Derived from the overall aim, the **specific objectives** are:

i. To establish the theoretical foundations related to translation competence, translation of science and technology texts and its inherent difficulties, corpus linguistics and its application in translation studies, and parallel corpora as an aiding tool in the translation process.

ii. To diagnose the current situation concerning the process of translation of science and technology texts in the fourth year of the English Language with Second Foreign Language (French) degree course.

iii. To build an English-Spanish Sentence-Aligned Parallel Corpus of Science and Technology Texts with the purpose of enhancing the translation competence of fourth-year students of the English Language with Second Foreign Language (French) degree course.

Contribution

The proposal of an English-Spanish Sentence-Aligned Parallel Corpus of Science and Technology Texts is proposed to help the students of the degree course English Language with Second Foreign Language (French) acquire and develop their translation competence as undergraduates, and hopefully, in their working environment as translators. According to experts, using parallel corpora as an aiding tool in the translation process provides a broad range of vocabulary and translation solutions, which is beneficial to systematize lexical knowledge and translation strategies. Considering the importance of topic, the proposal could be viewed as an innovative and far-reaching alternative.

Structure of the paper

The paper is structured in two chapters. Chapter I deals with some essential aspects on the translation competence and its acquisition, corpus linguistics, its applications to the translation process, and its possible contributions to the didactics of translation in class. Chapter II details the proposal of a bilingual sentence-aligned parallel corpus of science and technology texts to enhance the translation competence of students, the selection of sources, and the processing and preliminary results of such parallel corpus.

CHAPTER I

Chapter I. Theoretical Framework

Science and technology texts are very difficult to translate because of the terminology, the characteristic speech, the syntactic differences between the languages, and the background knowledge that is often needed in order to fully understand the texts (Esra'a M. Muhiesen, 2019). Therefore, the constant process of acquiring and strengthening translation competence, especially when translating those texts, has never been more critical. The result of a good translation is the visible part of the process; however, what does it take to arrive in that point?

Translators use a great variety of tools and sources to carry out their work; monolingual and bilingual dictionaries, online and portable translators are among the most widely used. However, the development of linguistic studies in addition to the technical development, has introduced new sources and tools, which are a valuable aid for the translator.

Corpus linguistics, on the other hand, has also provided new paths in the study of translation. Corpora are not only used for the study of language phenomena, but as a bank of authentic examples that can work as a guide for language students, and for translators specifically. This chapter will deal with some theoretical aspects regarding the acquisition of the translation competence, the translation of science and technology texts, the relationship between translation and corpus linguistics, and how a parallel corpus can be useful in translation.

1.1 Translation

1.1.1 Translation Competence. Definition.

In the decades of the 80s and 90s of the 20th century, many linguists and translation theorists attempted to define translation competence and its underlying sub-competences. For instance, Abouzar Oraki and Gholam-Reza Tajvidi (2020), identified the translation competence as a set of different sub-competences required for translating. Both of them identified six main translation sub-competences: 1)

Linguistic competence of the languages concerned, 2) Cultural competence, 3) Textual competence, 4) Domain/subject specific competence, 5) Research competence and 6) Transfer competence (Oraki, Abouzar and Tajvidi, Gholam-Reza, 2020).

More recently, the group PACTE (*Proceso de Adquisición de la Competencia Traductora y Evaluación*, by their acronym in Spanish) of the *Departament de Traducción i d'Interpretació de la Universitat Autonóma de Barcelona in Spain*, a pioneer group in the study of translation competence with empiric and experimental studies, also define translation competence as:

"The underlying knowledge system needed to translate". It has four distinctive characteristics: it is expert knowledge and not possessed by all bilinguals, it is basically procedural knowledge, it is made up of various interrelated sub-competences, and the strategic component is very important, as it is in all procedural knowledge". (PACTE, 2003)

The various interrelated sub-competences they proposed are:

1. The bilingual sub-competence: Predominantly procedural knowledge needed to communicate in two languages. It includes the specific feature of interference control when alternating between the two languages. It is made up of pragmatic, socio-linguistic, textual, grammatical and lexical knowledge in the two languages.

2. Extra-linguistic sub-competence: Predominantly declarative knowledge, both explicit and implicit, about the world in general and specific areas. It includes (1) bicultural knowledge (about the source and target culture); (2) encyclopedic knowledge (about the world in general), and (3) subject knowledge (in special areas).

3. Knowledge about translation sub-competence: Predominantly declarative knowledge (implicit and explicit), about translation and the aspects of the profession. It includes (1) knowledge about how translation functions: type of translation units, processes required, methods and procedures used (strategies and techniques), and types of problems; (2) knowledge related to professional translation practice: knowledge of the work market (different types of briefs, clients and audiences, etc.).

4. Instrumental sub-competence: Predominantly procedural knowledge related to the use of documentation sources and information and communication technologies applied to translation: dictionaries of all kinds, encyclopedias, grammars, stylebooks, parallel texts, electronic corpora, searchers, etc.

5. Strategic sub-competence: Procedural knowledge to guarantee the efficiency of the translation process and solve the problems encountered. This sub-competence affects all the others and causes inter-relations amongst them because it controls the translation process. Its functions are (1) cognitive components such as memory, perception, attention and emotion; (2) attitudinal aspects such as intellectual curiosity, perseverance, rigor, critical spirit, knowledge of and confidence in one's own abilities, the ability to measure one's own abilities, motivations, etc.; (3) abilities such as creativity, logical reasoning, analysis and synthesis, etc.

The author of this work ascribes to the classification of the group PACTE for the translation sub-competences, and established a direct connection between them and the usefulness criteria for the compilation of the English-Spanish Sentence-Aligned Parallel Corpus of Science and Technology Texts.

According to the American Translators Association (ATA), translation competence is, in short, expert knowledge (declarative, procedural, and attitudinal) made up of various interrelated sub-competences, which requires the specialization of the individuals. It is very important to develop and enhance this competence from the very training process, because the professional exercise of translation will certainly be demanding. Hence, the need of integrating disciplines in class, and proposing innovative ways of acquiring and enhancing translation competence. (ATA, 2012)

1.1.2 Translation Competence: Its acquisition through a process-oriented approach and method

According to Kussmaul (1995), there are two approaches in translation teaching: product-oriented and process-oriented ones. The product-oriented approach includes error analysis and translation quality assessment. The process-oriented

approach provides an insight into the translation process itself and the cognitive efforts made by people who perform translation tasks. (Kussmaul, 1995)

Traditional translation lessons are based on translation assignments, which are corrected in class, with teachers criticizing or approving the students' choices and presenting their own solutions. Although the method is operational, it is less than optimal on two points: It focuses on the product rather than on the processes, which means that inferences for the correct processes are made by the students themselves, with little possibility of guidance by the teachers.

Very often, students reject the teacher's criticism and solutions because of diverging linguistic norms and because they feel attacked. This slows down the learning process. If teachers focus on the process, they can be less critical of the product and, to a large extent, avoid such problems. It is necessary that the interaction between the teacher and the students in the translation process is a dialogue, rather than an imposition of the teacher´s proposals. (Gile, 2018)

The focus of this research is based on the process-oriented approach. It is necessary to strengthen the process of translation in the English Language degree course, in order to prepare proficient translators with strategies and knowledge that allow them to face any sort of texts and render high-quality translations. The difficulties in the translation process will be many, particularly in the most specialized texts. Therefore, it is necessary to complement and strengthen the teaching of translation in every way possible, focusing more on the process.

1.1.3 Translation of Science and Technology Texts: Characteristics

According to Vieira (2020), the available science and technology publications normally have, as its audience, graduate and undergraduate students, in addition to teachers and researchers. As a result, this literature is often written in a high-level linguistic pattern, which implies complexity in vocabulary as well as in sentence structures. Therefore, the translator must focus not only in finding an accurate equivalence between words, but also in translating ideas, always considering the

specificities of the science and the cultural aspects involved in the process of translation.

Technical-scientific translation is very different from literary or socio-cultural translation. However, when translating texts dealing with these fields, the translator must pursue the same objective, which is to convey accurate and intelligible ideas to the reader. Nevertheless, technical-scientific literature has its peculiarities which need to be considered in order to understand the texts and being able to efficiently translate them. (Esra'a M. Muhiesen, 2019)

For instance, while in literary texts it is very common that a word in the source text may be translated into two synonyms in the target text in order to avoid repetition and maintain a high-level text style, in technical-scientific texts, this may be dangerous, since one of the most important aspects of this type of literature is precisely to maintain the consistency of terms, regardless of repetition. Using two different equivalents in the target text for the same word in the source text, could cause a great confusion in the reader, due to the very complexity of technical-scientific literature. Terminological consistency is a lamp for the reader that helps him to go through the text without getting lost. (Kaur, 2005)

Moreover, sometimes the translator attempts to translate a texts without first reading it carefully and test his/her understanding of such text, inverting thus the logical development of the translation process. This represents a hazard in every translation; however, the danger is doubled in technologic-scientific translation due to the already-mentioned complexity it entails.

In an annual meeting held by the American Medical Association in 2002, titled Translating Scientific Text: Practicalities and Pitfalls, Brenda Gregoline stated the main goal in technical-scientific translation as: "The work of scientific translators is to achieve one primary goal: to write information in a clear, concise, and accurate manner".

In order to fulfill this main goal, the translator must sort out certain difficulties summarized by Gregoline, as follows:

1. *difficulty of the text*, ranging from level I, in which isolated words can be identified, to level V, in which the translator brings technical and specialized cultural knowledge to the text;
2. *linguistic difficulty*, which is defined by the differences between the writing, grammatical, and semantic systems of one language and those of the other; and
3. *language difficulty*, which is calculated by the number of weeks it takes to learn a language. (Gregoline B. , 2002)

All in all, the translator of scientific and technological texts must be proficient in the target language, and possess general knowledge and awareness in terms of the specific technical-scientific words pertaining to the area to be translated. (Vieira, 2020)

Finally, it is also important that the translator respects the choices and style of the author of the source text. It is very common among novice translators the tendency of taking their preconceived ideas and style into their work, which may affect the legitimacy and fidelity of translation.

Knowing the main characteristics of technical-scientific translation and respecting the rules proposed by experienced translators, may help novice translation to foster and develop their translation competence, which is the main objective of this research.

1.2 Corpus Linguistics

1.2.1 Corpus Linguistics. Definition.

The compilation and use of corpora are not exclusive of our technological and digital society. For centuries, language researchers have compiled texts to use them for a specific purpose. Generally speaking, a corpus is a collection or body of texts, just like the complete works of an author. For instance, John Sinclair, late linguist and language scholar defined a corpus as: “a collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of the language.” (Rodríguez-Inés, 2008)

Moreover, Mona Baker, translation scholar and expert in corpus-based translation studies provides a more specific definition adapted to the current

digitalization of corpora, stating that a corpus is “any collection of running texts, held in electronic form and analyzable automatically or semi-automatically (rather than manually).” (Baker, 1995) cited in (Rodríguez-Inés, 2008) Professors Anthony McEnery and Andrew Wilson of the University of Lancaster complete this definition provided by Mona Baker by adding that a corpus is indeed a collection of texts, but finite and machine-readable, sampled to represent a language or variety. (McEnery, 2001) cited in (Rodríguez-Inés, 2008).

All in all, we can conclude that a corpus is a collection of texts that is finite, machine readable, automatically or semi-automatically analized, selected according to explicit linguistic criteria in order to be used as a sample of the language. The size of the corpus, i.e, the fact that it is finite, is determined by the purpose and therefore by the scope of such corpus.

Having defined the concept of “corpus”, then what is corpus linguistics? According to Michael Stubbs (2001), Corpus linguistics is a research methodology used for the study of linguistic phenomena through large collections of texts, which are currently machine/ readable, oral or written called corpora. In other words, it is a research methodology based on empirical evidence in how language is actually used.

A similar definition was presented by professor Sara Cushing (2017), stating that Corpus linguistics is a research methodology used for the study of linguistic phenomena through large collections of machine-readable texts, oral or written, called corpora. In other words, it is a research methodology based on empirical evidence about how language is actually used.

Now, corpora can be classified into various categories, considering the source of the content, metadata, number of languages, objective, etc. According to the EAGLES (Expert Advisory Group on Language Engineering Standards) text typology elaborated by John Sinclair (1996), the most general and first distinction made between corpora occurs between Monolingual and Multilingual (including Bilingual) corpora. Multilingual (bilingual) corpora can be sub classified into Comparable corpora (corpora compiled using similar design criteria but which are not translations) and **Parallel, or Translation Corpora**. (Sinclair,1996)

This last type of corpus is of special importance for this research, since it is the corpus that best suits as an aiding tool to strengthen translation competence. (Claudio Fantinuoli and Federico Zanettin, 2015)

1.2.2 Parallel corpus: its usefulness in the translation classroom for the enhancement of translation competence

For the purpose of this research work, it is important to establish from the very beginning the close link existing between translation and corpus linguistics. The initial thrust to descriptive corpus-based studies in translation came in the 1990s, when researchers and scholars saw in large corpora of monolingual texts an opportunity to further a target-oriented approach to the study of translation, based on the systemic comparison and contrast between translated and non-translated texts. (Claudio Fantinuoli and Federico Zanettin, 2015)

Within the whole typology of corpora with specific uses in linguistic studies, the interest of this research focuses on a specific type of corpus called parallel corpus. A parallel corpus comprises texts in one language aligned with their translations in another. Both languages need to be aligned, i.e. identifying corresponding segments, usually sentences or paragraphs, which need to be matched through a corpus tool. (Zanettin, 2002) In this research work we specifically use a sentence-aligned parallel corpus to enhance translation competence.

Parallel corpora are considered one of the most helpful tools in the enhancement of the translation competence. For instance, the user can then search for all examples of a word or phrase in one language and the results will be displayed together with the corresponding sentences in the other language. (Zanettin, 2002)

Lavion believed that “with the help of corpora, the translator can easily find out the most suitable words and expressions” (2002). Besides, they can search some fixed phrases or collocations of the source language and see how these are translated in the target language. Moreover, corpora provide a better complementary resource that allows translators to see terms in a variety of contexts simultaneously, which enables them to choose the best equivalents for each context (Guo-rong, 2010).

Likewise, Monserrat Bermúdez Bausela (2016) strongly affirms that corpora, particularly specialized corpora, are an inestimable source not only for terminology and phraseology extraction, but also for studying the textual conventions that characterize and define specific genres in the translation languages. In her paper *The importance of corpora in translation studies: a practical case*, she highlights the contribution of corpora to the study of a specialized language from the translator's point of view, which specifically enables the enhancement of the bilingual translation sub-competence.

> *"From the observation of specialized corpora, it is possible to identify specific patterns, phraseology, terminological variants, and the frequency of conceptually relevant words, cohesive features and so forth. The access to this information will allow the translator to produce quality texts".* (Bausela, 2016)

One of the main advantages of translation teaching with corpora is the fact that the presence of corpora reduces the role of the teacher's intuition in the translation classroom, and at the same time assigns more importance to the students and their documentation skills. This is very important in contexts of online learning and independent study (Rodríguez Inés, 2009, p. 131) cited in (Krüger, 2020). By providing alternative sources of authority as well as a set of authentic data, corpora can also shift the role of the teacher from that of the principal information provider to that of an information facilitator (Rodríguez Inés, 2009, p. 130, p. 133) cited in (Krüger, 2020)

Therefore, the use of a parallel corpus in the translations lessons brings about that students can learn specific vocabulary, analyze the solutions given by other translators, and identify which is the best equivalent for the context they are translating. This way, students can integrate the knowledge and strategies acquired in the classroom through the use of the parallel corpus. This knowledge will be very useful to enhance the students´ translation competence.

1.2.3 Representativity, sampling and size of a corpus

For the compilation of a corpus, there are certain criteria that need to be considered in order to attain its adequacy for the purposes for which it is compiled. For instance, there is a specific criterion called “representativeness” that is central when it comes to the compilation of a corpus.

Biber (1993) states that representativeness refers to the extent to which a sample includes the full range of variability in a population. That is, when a sample of texts is representative of the language, dialect, type of texts, thematic domain that is subject to study. This category is generally associated with other criteria such as wise sampling and size of the corpus, being a sample any selection of texts for a specific purpose.

The relationship between the size of a corpus and its representativeness is still very controversial, since there appears to be no consensus among experts due to the countless number of specific corpus cases and the variety of purposes for which they are compiled. (Biber, 1993)

It is surprising that, while for some experts there seems not to be a specific number of texts and words that assure the representativeness of a corpus, others take a radical position when it comes to the relationship size-representativeness (i.e. “the bigger, the more representative”). For instance, there are some authors who declare that ‘there is no text like more text’, ‘more data is better data’ or ‘the bigger the corpus the better’ (Church and Mercer 1993: 18-19). Similarly, Sinclair (2004) considers that ideally a corpus should be ‘big’, although the interpretation of this adjective remains open to debate because no approximate figure is given. (Corpas Pastor, Gloria and Seghiri, Miriam, 2010)

However, other equally prestigious researchers in the area of corpus linguistics have affirmed that smaller corpora provide with outstanding results when used for the study or acquisition of language in specialized areas. In fact, an increasing number of researchers, such as Bowker and Pearson (2002: 48), stress that smaller corpora with ‘a few thousand and a few hundred thousand words’ are very useful in the study of languages for specific purposes. (Corpas Pastor, Gloria and Seghiri, Miriam 2010)

Also, Mohsen Ghadessy, Alex Henry and Robert Roseberry emphasized that smaller corpora are extremely useful for studying the language of a specific field or for language teaching. (2001)

Last but not least, professor Nosley Pérez Castellano summarized the factors that enter the equation to obtain a suitable corpus in our specific context than when he stated that: "… a corpus should be as large as possible, depending on factors such as the availability of texts in electronic format, the time available or the lack of resources and, in addition, representativeness, i.e., it should match the purposes of the research". (2017)

In the case of the corpus proposed in this work, the author ascribes to the criteria of the experts who state that smaller corpora are very useful when compiled to acquire knowledge on language for specific purposes, which is the case of science and technical terminology. Therefore, we consider that size and representativeness of the parallel corpus proposed in this work are appropriate, taking into account that it was created to address the topics dealt with in the Subject Translation of Science and Technology Texts, and to help the students to acquire the terminology and translation strategies of this particular type of texts.

1.2.4 Corpus alignment criteria

In the introduction, and throughout this research work, the author has been referring to a specific type of parallel corpus: a sentence-aligned one. However, why to use a sentence-aligned parallel corpus and not another type?

First, it is important to explain that, an aligned corpus is the outcome of "finding correspondences, in bilingual parallel corpora, between textual segments that are translation equivalents". (Kraif, 2018) The alignment of a corpus can be done at a paragraph level or at a sentence level, although what is usually understood by alignment is related to the sentence level. (Claudio Fantinuoli and Federico Zanettin, 2015)

In the case of the corpus in question, the alignment was made at a sentence level because it is more user-friendly and more comfortable for the automatic work. Sentence alignment can help users better visualize the searched items in the text and

better compare specific equivalents or analyze a particular translation solution in a more precise context. (Marasek, 2015)

Jorge Leiva Rojo from the University of Málaga considers that the use of aligned parallel corpora is more advantageous than a paragraph-aligned parallel corpus for visualizing the terminology and translation strategies. (Rojo, 2018)

As a matter of fact, it is very interesting and surprising the fact that the availability of sentence-aligned parallel corpus is limited, to a certain extent, mainly due to the arduous work that needs to be done to compile the original texts and their accurate translations, as well as in the further process of alignment. (Zanettin, 2012)

Therefore, the alignment of the parallel corpus proposed in this work is an addition to the scarce number of sentence-aligned parallel corpus compiled for specific purposes; in this case, to enhance the translation of science and technology texts.

Conclusions of Chapter I

To sum up, this chapter englobes the most pertinent concepts and theoretical foundations that serve as basis to this research work. Critical notions such as translation competence, general and specialized translation, corpus linguistics and parallel corpus have been defined. In addition, a journey has been traced on the close relationship between translation and corpus linguistics, and more specifically, experts´ criteria about the valuable help provided by a parallel corpus in the exercise of translation has shed some light and given impetus to the implementation of the English-Spanish Parallel Corpus of Science and Technology Texts.

CHAPTER II

Chapter 2. Methodology for building an English-Spanish Sentence-Aligned Parallel Corpus of Science and Technology Texts

Having established the theoretical foundations for the use of parallel corpora as an aiding tool to enhance the translation competence, it is pertinent to analyze the methodology and practical steps involved in the construction of the parallel corpus proposed in the present research work.

2.1 Methodology

The design of the present research adopted a mixed-method approach. This approach analyzes the collected data quantitatively and qualitatively interprets the research results.

The research process consisted of three main stages:

The first stage was a bibliographic review in order to construct a theoretical framework to systematize the main concepts associated with the topic of the research and their relationships. Therefore, notions such as translation, approaches to translation, main difficulties in the translation process, the acquisition of translation competence, corpus linguistics, different types of corpora and corpus linguistics in translation studies were analyzed and systematized.

The second stage was devoted to the analysis of the methodology used for diagnosing the students' needs concerning the translation process and also the establishment of the steps to take in order to compile, align and manually process the corpus of science and technology texts. Therefore, students were surveyed to determine their difficulties concerning the translation process of science and technology texts and the importance they confer to the incorporation of new input sources (a parallel corpus) to enhance their translation competence.

In the third stage, a Sentence-Aligned English-Spanish Parallel Corpus of Science and Technology Texts was compiled in the light of the theoretical considerations and methodological steps previously established. In this stage, the author of this research explains in detail the practical steps involved in the

compilation of the English-Spanish Sentence-Aligned Parallel Corpus of Science and Technology Texts, the process of texts selection, alignment and manual processing of the corpus. Moreover, as a practical example of the usefulness of the corpus to enhance the translation competence of the students, a word bank of science and technology texts extracted from the corpus is proposed, as well as a sample of activities that can be developed in class.

2.1.1 Description of the sample

The sample for this research involved:

- Sixteen fourth-year students of the English Language with Second Foreign Language (French) degree course at Universidad Central “Marta Abreu” de Las Villas.
- 50 English texts on science and technology with their respective translations into Spanish for a total of 100 texts, taken from the World Health Organization (WHO) official website, the Nobel Prizes official website and the Encarta Encyclopedia Student Premium 2009. It is worth noticing that the prestige of the sites from which the texts were retrieved, supports the authenticity of the texts of the corpus, including their translations into Spanish.

2.1.2 Research Methods

In order to successfully organize and develop this research, the author used diverse theoretical and empiric methods, such as: historical and logical, analysis and synthesis, induction and deduction, structural systemic, document analysis, non-participant observation and a survey, in order to gather all the theoretical basis and practical arguments to sustain the purpose of this research and its usefulness in the specific context of the English Language with a Second Foreign Language (French) degree course at Universidad Central “Marta Abreu” de Las Villas.

The analysis of documents and the observation were crucial for attaining a deeper understanding the methodology implemented to teach the translation discipline, its objectives, and the expected profile of a translator graduated from our studies

program; as well as to evidence the difficulties and challenges faced by the students every day, particularly when they have to translate on specialized topics (science and technology).

Along with the observation and other empirical methods, carrying out a survey led the author to the conclusion that it was necessary to update the practice of translation teaching by combining the different disciplines and providing students with new and helpful tools in order to enhance their translation competence. The results of the survey introduced the possibility of using a bilingual parallel corpus as an aiding tool for the practice of translation, and to help students enhance their translation competence in the subject Translation of Science and Technology Texts, which, as stated before, is very complex.

Theoretical Methods

Historical and logical: This method was used for analyzing the principles and concepts related to translation competence, approaches to translation teaching, corpus linguistics, corpus-based approach to translation, parallel corpus, the construction of a parallel corpus and the use of it as an aiding tool in the translation process.

Analysis and synthesis: When consulting bibliographic references, this method allowed to analyze the tendencies found in the collected data to formulate the theoretical framework to develop this research.

Induction and deduction: This method was applied to analyze the bibliography to come to general conclusions regarding the most common difficulties in the translation of science and technology texts, as well as the use of corpus to aid translation.

Structural and systemic: Provided a logical and systematized guidance to the authors throughout the whole research process and in the proposal of solutions to the stated problem.

Empirical methods

Document analysis: This method was extremely useful in order to analyze and gain a deep insight on the principal documents that regulate the degree course and therefore, the discipline of translation at the English Language Department of Universidad Central “Marta Abreu” de Las Villas. Analyzing these documents was crucial for the purpose of understanding the methodology implemented to teach the translation discipline, its objectives, and the expected profile of a translator graduated from our studies program.

It is important to say that in response to the recent political and socio-economic changes being implemented in Cuba, as well as the international demand for more efficient professionals, there has been a renewal and improvement of the curriculum. The previous curriculum for the English language with a Second Foreign language (French) was Curriculum D, and the present one is Curriculum E.

The document analysis focused on the plans for both curricula, and the presence of the subject Translation of Science and Technology Texts in them, in order to support the relevance of this research for the present curriculum.

Translation in Curriculum D

Curriculum D has a more holistic approach to the preparation of the English Language students. Under its implementation, the degree year consisted of six years of study with an introductory year of intensive development of skills in the English Language. An important feature of this curriculum D is that students had to take all the linguistic as well as translation/interpreting subjects, while the Curriculum E advocates more flexibility with respect to the subjects that students need to take up.

In Curriculum D, there are seven subjects devoted to the field of translation:

1- Introduction to Translation and Socio-Cultural Translation with 64 class hours each in third year;
2- Translation of Science and Technology Texts, Official Documents Translation with 48 hours each and Computer-Assisted Translation with 32 hours in fourth year;

3- Socio-Economic Translation and Journalistic Translation with 48 class hour each in fifth year.

According to the above-stated organization of Curriculum D, the subject that the author aims to support through the use of the parallel corpus (Translation of Science and Technology Texts), is taught during the first semester of the fourth year of the English Language with a Second Foreign Language (French) degree course. Each translation subject in Curriculum D is named after the topics covered in the texts dealt with.

Translation in Curriculum E

In the Curriculum E there are some changes regarding the Translation subjects. While in Curriculum D there were seven translation subjects, in Curriculum E there are only three translation subjects. Nevertheless, the restructuring does not hinder the preparation received by the English Language students, since almost all the topics previously addressed in the translation subjects of Curriculum D, make now part of the translation subjects of Curriculum E. Only the Official Documents Translation was excluded.

Summing up, the syllabus of the Translation Syllabus of Curriculum E is established as follows:

1. Translation I, which will include Translation of Socio-Cultural Text with 64 class hours in the sixth semester (second year);
2. Translation II, which includes Translation of Science and Technology Texts with 48 class hours, Translation of Social-Political Texts with 48 class hours and Oral Translation by Simple Sight (TOASV) with 32 class hours in the seventh semester (third year);
3. Translation III, which includes Translation of Socio- Economical Texts with 48 class hours, Translation of Journalistic Texts with 48 class hours and Oral Translation by simple Sight (TOASV) with 32 class ours in the eighth semester (third year).

Translation of Science and Technology Texts in the curriculum

In Curriculum E (the present curriculum), the prior subject Translation of Science and Technology Texts is included in the subject Translation II, taught in the first semester of the fourth year of the English Language with a Second Foreign Language (French) Degree Course, which will occupy 48 class hours.

The topics addressed in the texts to be translated during the semester are:

1. Biology and biomedical sciences
2. Computer Science (Artificial intelligence and bioinformatics)
3. Chemical Sciences
4. Physical Sciences

The parallel corpus proposed in this research work has a fair representation of each topic, covered by an average of 10 texts with their translations for each topic dealt with in the Translation of Science and Technology Texts. It does not only harmoniously match the topics, but also provides a direct support to the objectives of the Translation of Science and Technology Texts, which are the following:

1. To accurately carry out the translation process from English to Spanish of the texts retrieved from textbooks, brief scientific articles, and journalistic publications on Biology/ Biomedicine, Computer Sciences, Chemical Sciences and Physical Sciences.
2. To apply the historical, social, cultural and scientific background knowledge in the understanding of the texts.
3. To use the necessary tools to translate.
4. To master the fundamental English-Spanish terminological equivalents of the scientific field dealt with in the text.

Corpus linguistics: To use the computerized tools designed for building the parallel corpus of science and technology texts with the long-term objective of applying this tool in the Translation lessons. To create a word bank of English-Spanish equivalents retrieved from the corpus.

As a methodology, corpus linguistics guides researchers in the processes of analyzing specific aspects of language or compiling corpora for specific purposes, i.e. that acquisition and developing of language skills of translation competence. (Lindquist, 2009)

Therefore, the author of this research integrated all the theoretical basis previously established on the definition and usefulness of corpus linguistics, in order to compile the proposed parallel corpus of science and technology texts. And not only to compile, but to work with and analyze the texts compiled and explore the way of inserting the parallel corpus of science and technology texts in the translation class of the fourth-year English Language with a Second Foreign Language students at Universidad Central "Marta Abreu" de Las Villas. For the compilation and further alignment and processing of the corpus, the software used were *OmegaT*, *Xbench* and *AntConc*.

Survey: The objective of the survey was to find out about the current situation concerning the difficulties hindering the translation process, the situation in respect to the availability of offline resources for translating science and technology texts, and the students' opinion concerning the use of a sentence-aligned parallel corpus in the translation lessons.

Non-participant observation: To observe the development of the translation process in the classroom and the main difficulties experienced by the students. To observe the role of the teacher as well as the role of the students.

2.1.3 Results of the survey

The survey (Appendix 1) was applied to 20 fourth-year students of the degree course English Language with Second Foreign Language (French) in the academic year 2019-2020.

The majority of students (90 percent) perceived some difficulties hindering the translation process:

1. Insufficient offline bilingual reference sources for translating in the lessons.
2. Insufficient time for translating in class.

3. Some students claimed for more tutoring hours with the professors.
4. Most of the translation is relegated to independent study.
5. Insufficient background information on the topics of science and technology before facing the translation of a text dealing with these topics.

As evidenced in the results of the survey, the translation process is affected by the fact that most of the translation is done outside the classroom, independently, and with little offline resources. It is known that the process of acquisition of translation competence should be accompanied with the help and guidance of a tutor, so that students can clear out doubts, the use of different equivalents and strategies. The insufficient background information about specific topics of science technology may hinder the comprehension process of the texts, and therefore, the accuracy of the translations.

More than a half declared insufficient prior knowledge or background information, vocabulary on the topics, and many rely on online translators for choosing lexical equivalence and precise information. Overall, they refer not to have the necessary knowledge and the skills for the translation process.

Table 1 reflects the quantitative data derived from the survey on the difficulties stated above.

Table 1. Results of the survey. Quantitative analysis.

Students' perceptions concerning the subject Translation of Science and Technology Texts			
Availability of bilingual reference sources	Enough	Little	Insufficient
	7.69%	46.15%	38.46%
Prior knowledge or background information, vocabulary and	Enough	Little	Insufficient
	23.07%	23.07%	53.84%

possible solutions to translation difficulties			
Dependence on online translators	Very little dependent	More or less dependent	Very dependent
	23.07%	30.76%	46.15%
Using only online translators is enough to improve translation competence	Yes	No	
	7.69%	92.3%	-
Students who have used a bilingual corpus as an aiding tool to translate	-	100%	-

The second question of the survey was aimed at finding out the students' opinion about the use of a sentence-aligned parallel corpus in the translation lessons as an offline resource to aid the translation of science and technology texts. The majority of students considered the proposal a very useful resource for translating this kind of texts, and the following statements support this assertion:

Student 3: *"It would be very useful as a tool, not only for counting on a reliable source for term translation, but also to see them in context and with a variety of examples, collocations and alternatives of specific professionalisms".*

Student 7: *"It would be very helpful. We could see the use of terms in context and finish our translation quicker and with more quality".*

Student 8: *"Not only it would be helpful in terms of contrasting and finding new solutions to the translation problems we face in the translation of scientific and*

technological texts, but it would help us to practice, to enhance not only the result but the process of translation, to learn and discover new translation strategies".

Moreover, the non-participant observation led us to confirm that, in effect, as Xiaoping (Isadora) Jiang and Josta van Rij-Heyligers indicated in their article "Parallel Corpus in Translation Studies: An Intercultural Approach", translators can spend almost 50% of their time consulting reference materials. Therefore, it is considered that the use of parallel corpora can enhance the speed and quality of translation, since they provide more native-like interpretations and strategies in both, source and target texts. They can also help novice translators to become aware of and learn the general patterns and preferred ways of expressing an idea. (Jiang, Xiaoping (Isadora) and van Rij-Heyligers, Josta, 2008)

Also, through non-participant observation the author could observe that students most of the time propose different equivalents after reading and analyzing the text in groups, sentence by sentence. Then, the difficult chunks are analyzed in order to find out the best solutions to it collectively. Finally, the professor and the students agree upon the most accurate equivalents, taking into account the students'opinion. This demonstrates the use of a process-oriented approach to the translation process, in which the teachers and the students collectivelly arrive to an agreement upon the equivalents and translation strategies, and the teacher guides the process without imposing his or her criterion.

2.2 Steps for building the English-Spanish Sentence-Aligned Parallel Corpus of Science and Technology Texts. Corpus design criteria.

In the design of the corpus proposed in this work, we followed the criteria of the Professor Patricia Rodríguez-Inés (2008), and the group PACTE (2003), and the author established the connection of those criteria with the sub-competences embedded in the concept of translation competence. These experts state that a parallel corpus:

- Can be built in less time than a conventional printed corpus. This criterion specifically contributes to the enhancement of the instrumental sub-competence.
- Can be reutilized for diverse purposes. This criterion specifically contributes to the enhancement of the instrumental and knowledge about translation sub-competences).
- Allows to carry out quick and systematic search of information in a wider and more varied range of texts. This criterion contributes to the development of instrumental, strategic and extra linguistic sub-competences.
- Offers a great number of lexical examples in the source and target languages in context. This criterion contributes specifically contributes to the enhancement of the bilingual sub-competence.
- Allows translators to go directly to those sections in the text where there are key words with the assistance of a corpus analysis tool. This criterion contributes specifically to the enhancement of the instrumental sub-competence)
- Facilitates the detection of information as collocational patterns. This criterion contributes specifically to the development of the bilingual-sub-competence)
- Facilitates the detection of frequency, which allows to determine the relevance of the patterns found. This criterion contributes specifically to the knowledge-about translation sub-competence.
- Facilitates and accelerates the retrieval and analysis of concrete information This criterion contributes specifically to the textual sub-competence.
- Offers certitude at the time of making strategic decisions such as lexical choices. This criterion specifically contributes to the enhancement of the strategic sub-competence)
- Provides a wide range of examples on how other translators dealt with specific translation problems, and this knowledge can be incorporated by students in order to solve similar translation problems. This criterion specifically contributes to the enhancement of the strategic and knowledge about translation sub-competences.

- Provides greater opportunities for casual learning. This criterion specifically contributes to the enhancement of the extra-linguistic sub-competence) (Rodríguez-Inés, 2008)

Having taken those criteria into consideration, the next step was to establish a planning process and to use it as a guide in the construction of the corpus. Such planning included the following stages:

i. Selection of sources
ii. Data capture and encoding
iii. Corpus processing (Sue Atkins, Corpus Design Criteria, 1992)

For building the corpus, it was necessary to use tools such as the *OmegaT* aligner (Godfrey, 2019) and the *Xbench* tool (ApSIC, 2017). All the files were converted to .txt format and subsequently encoded in UTF8 format. Once all the files were encoded in the required formats, they were uploaded them to the OmegaT tool to be aligned at a sentence level. Afterwards, they were uploaded to *Xbench* and converted into a translation memory, in order to be stored and allowing the users to carry out word search. Finally, *AntConc* tool (Lawrence, 2014) was used for analyzing the most common translation difficulties and specialized terminology, in order to organize them into a word bank.

2.2.1 Selection of sources

The texts included in the Sentence-Aligned English-Spanish Parallel Corpus of Science and Technology Texts are a very specific sample chosen for a specific purpose, in accordance with the particularities of the translation discipline in the degree course English Language with Second Foreign Language (French). The texts were retrieved from reliable sources advised by the translation professors.

In the first semester of 4th year, students work with science and technology texts in the domains of: (1) medicine and (2) scientific discoveries on physics,

chemistry and artificial intelligence. These domains are divided into the specific topics of: communicable and non-communicable diseases, artificial intelligence models and Nobel Prizes. Due to the recent impact of the Covid-19 pandemic, it has been considered relevant to include a number of texts dealing with this topic in the corpus, in the domain of medicine.

The corpus comprises 50 texts in English and their respective translations into Spanish for a total of 100 texts retrieved from the WHO (World Health Organization) official website, the Nobel Prizes official website, and the Encarta Encyclopedia Student Premium 2009. Some of these texts have been used by fourth-year students and professors of the English Language with Second Foreign Language (French) degree course in their translation lessons, and have been already corrected in group with the help of the professor.

The texts used in the corpus can be classified according to their format and style as specialized, expository or journalistic texts. They have been all published between 2009 and 2020.

The characteristics of the selected corpus follow the compilation criteria exposed by Vargas (2006). These characteristics are shown in Table 2 below:

Table 2. Characteristics of the selected corpus

Size of the corpus	125 860 word tokens
Transmission mode	Written transformed into .txt format
Level of specialization	Specialized texts of specific domain of medicine, physics, chemistry and artificial intelligence
Type of texts	Published research articles
Authorship	Varies from science specialist explaining a discovery to journalist

	reporting a specific scientific milestone
Language of the corpus	Texts originally written in English
Date of publication	Texts published from 2009 to 2020

Vargas Sierra, Ch. (2006). Diseño de un corpus especializado con fines terminográficos: el corpus de la piedra natural. *Debate Terminológico*, 2 (7/2006). París: RITERM (Red Iberoamericana de Terminología).

2.2.2 Data encoding and alignment

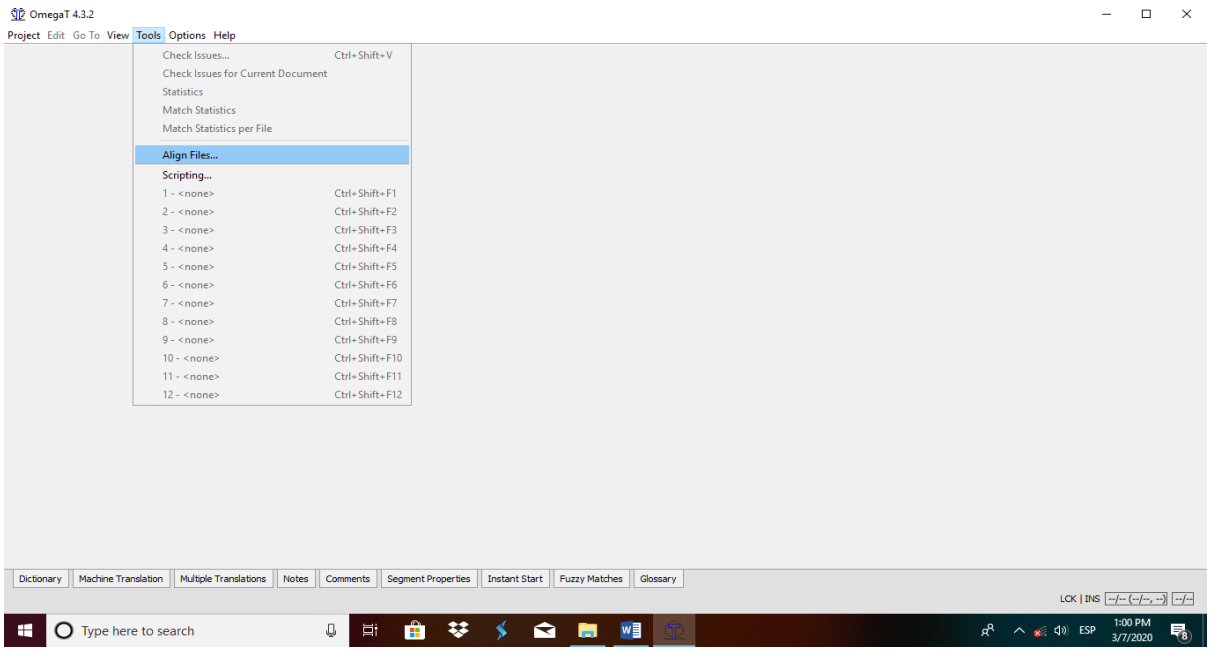
For the alignment of the corpus, the author used free, user-friendly software since, as Federico Zanettin wisely states, not everybody has the computational expertise that some traditional compilation and alignment methods require. Besides, the use of free and user-friendly software is what best fits the current state of technical advancement and internet access in our context. (Zanettin, 2012)

Therefore, the tool used to align the proposed corpus was *OmegaT*, a free translation memory application that works on Windows, macOS and Linux. It is a software intended for professional translators to carry our simultaneous processing of multiple-file projects, simultaneous use of multiple translation memories. It includes glossaries, works with more than 30 formats, has a Unicode (UTF8) support which makes it ideal to align the texts of the proposed corpus. This tool has a very reliable aligner and has a significant level of acceptance among users, who state that “it is the perfect companion for those who do not need complexity but efficiency”. (Godfrey, 2019)

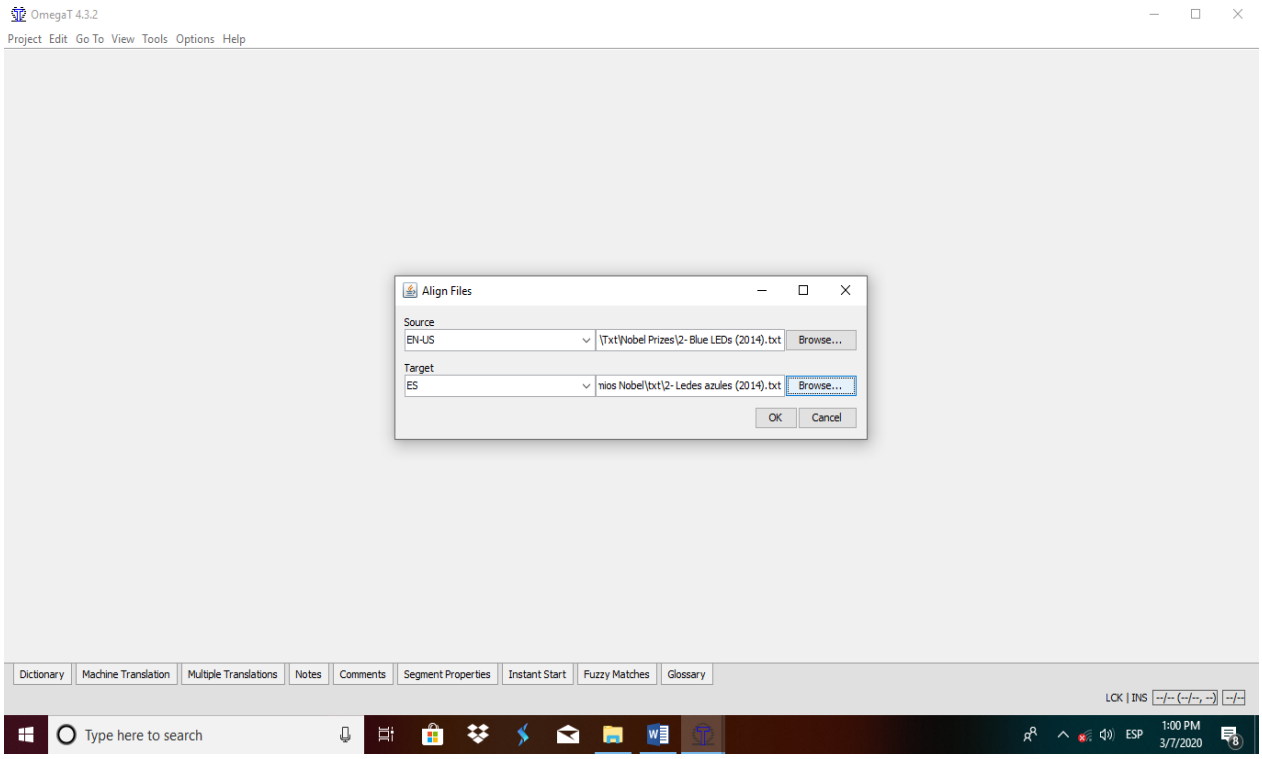
As has been referred above, one of the fundamental criterium for a set of texts to be a corpus as such, is that the texts must exist in a machine-readable format. In this case, texts in both languages were converted to txt. format or Text Encoded, and subsequently to UTF8, for them to fulfill the requirement of being machine-

readable texts. These formats made it possible to upload the texts to the *OmegaT* aligner.

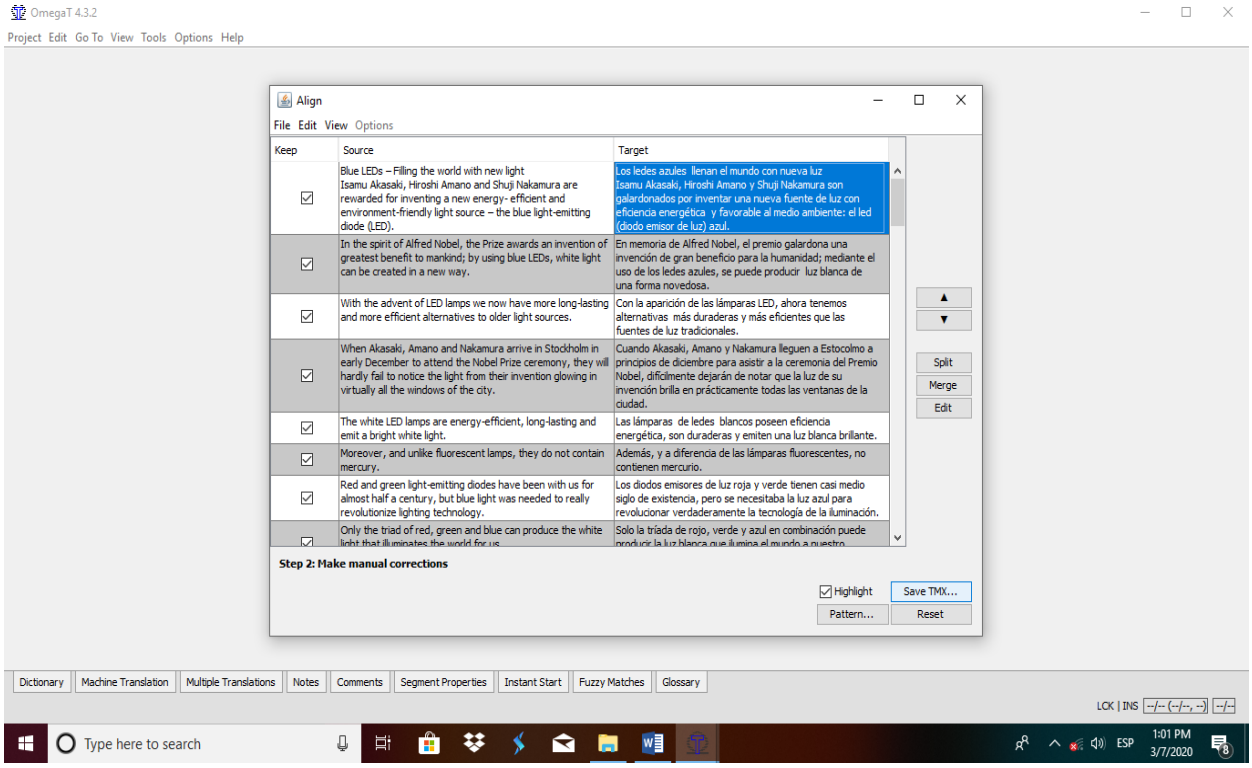
However, as accurate and effective the corpus aligned may be, it is always necessary to carry out a manual revision of the aligned texts, as shown in **Screenshots 1, 2, 3, and 4**, since the automatic alignment is not exempt of producing errors. Even with the great help of an automatic alignment tool, the process of alignment is quite cumbersome and needs to be developed very carefully in order to obtain a high-quality final product. *OmegaT* has also the function of exporting the aligned translations into TMX format, necessary to upload the texts as a translation memory in the tool used to process the corpus.



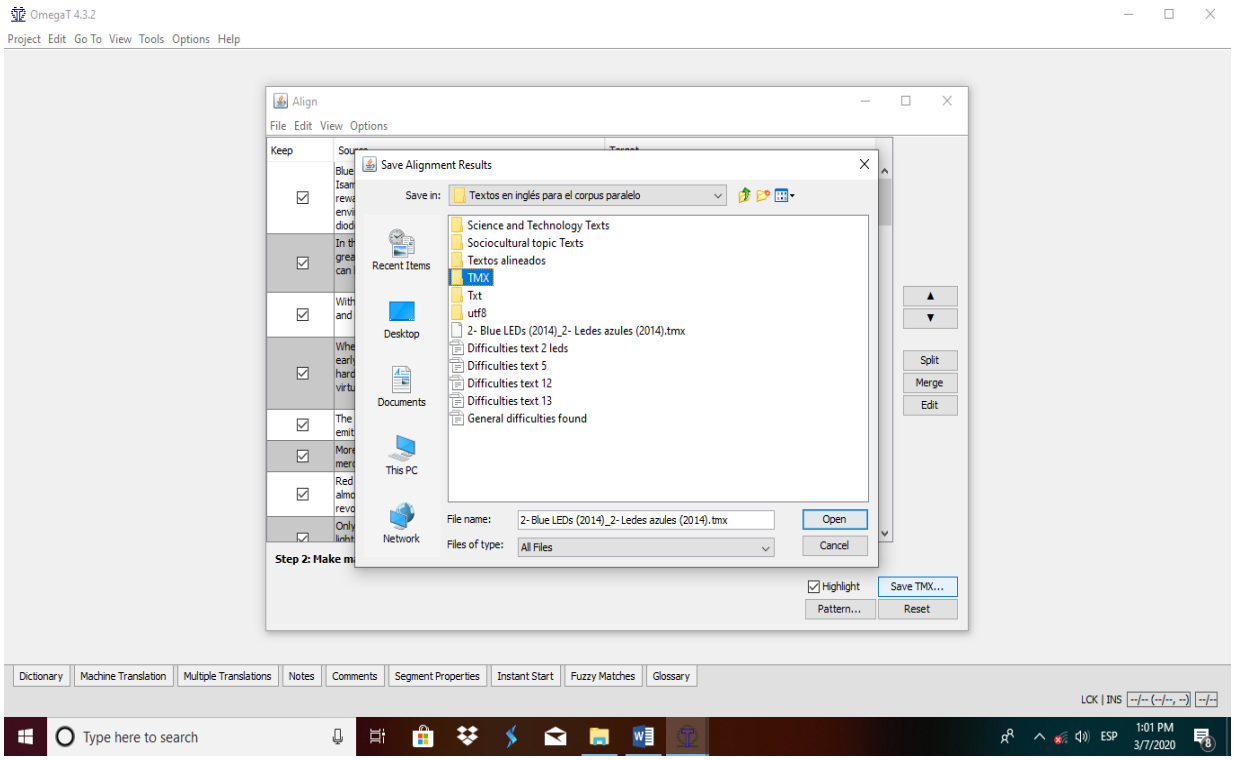
Screenshot 1. Using the “Align Files” tool of the *OmegaT* software



Screenshot 2. Selecting the texts to be aligned



Screenshots 3. Automatic alignment and manual edition of the texts



Screenshot 4. Creating a TMX file to save the aligned texts and store the corpus

After the texts were converted and saved in UTF8 format, the next step was the aligning process with *OmegaT* software. After opening it, the user must click on the option “Tools” and select the button “Align Files…”. Afterwards, a box will appear on the screen asking the user to browse for the files in English and Spanish to align them. The texts are aligned one by one.

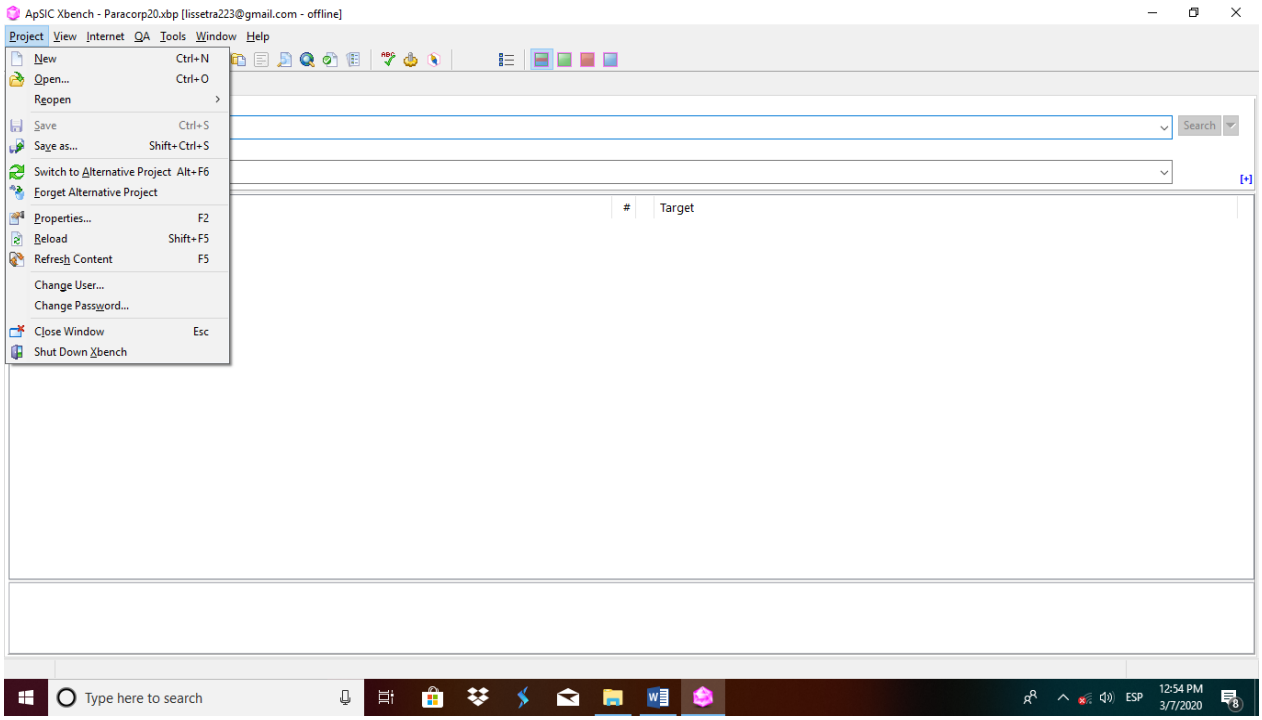
Having selected the UTF8 files and clicked on the “OK” button, the automatic alignment process will take place. As previously explained, it is necessary to check the automatic alignment since the tool is not infallible. Many of the sentence-alignments needed to be corrected, and although it was a time-consuming process, it was generally easy to carry out and the results were worth the process.

2.2.3 Corpus processing

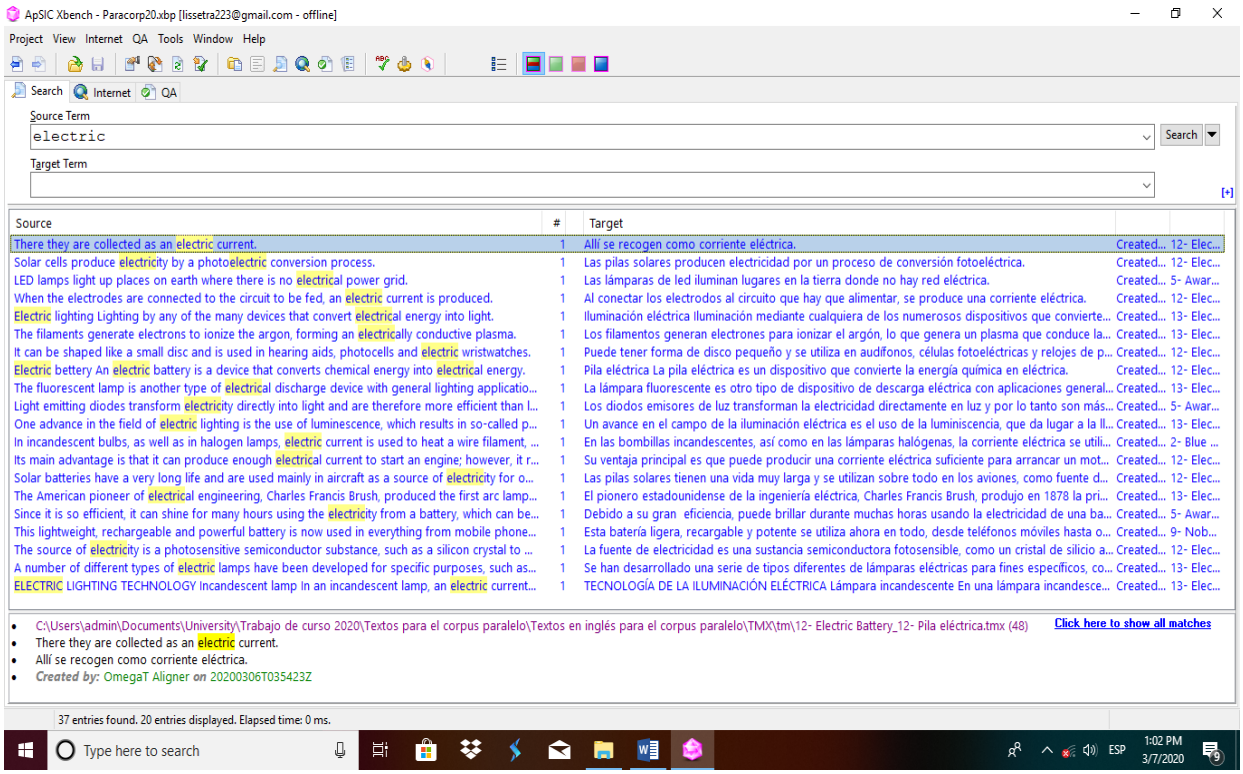
The corpus processor tool used to store the English-Spanish Sentence Aligned Parallel Corpus is called *Xbench,* which really is a user-friendly and accessible tool.

It is available in free versions on the Internet. Once the parallel corpus has been uploaded as a new project in *Xbench*, this tool allows users to carry out word and phrase search in both, the source and the target language, as shown in **Screenshots 5 and 6**.

First, one must open a “New Project” in *Xbench* and search the TMX file where the aligned texts were stored. Once the aligned texts have been open in *Xbench*, this software will save the corpus as a unity and will allow users to search words and phrases in the source and target languages. The results of the search will appear in the texts box of *Xbench* in the form of aligned sentences, which allows the user to see the words and expressions in contexts along with their equivalent. The English-Spanish Sentence-Aligned Parallel Corpus was saved in an *Xbench* file, as a translation memory, and from now on will be called *ParaCorp2020.*



Screenshot 5. Opening the TXM memory of *ParaCorp2020* with *Xbench* software

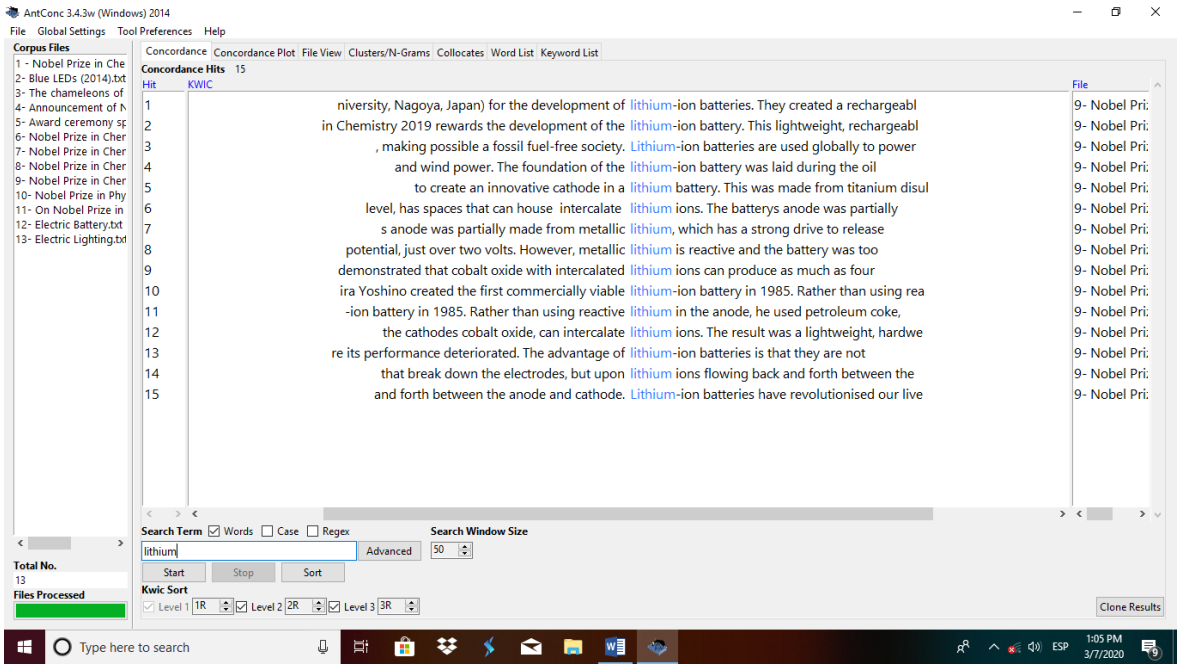


Screenshot 6. Carrying out word search in *ParaCorp2020* (using *Xbench* software)

2.3 Analysis of specialized terminology

AntConc tool was of great help when analyzing the texts selected for the English-Spanish Sentence-Aligned Parallel Corpus of Science and Technology Texts. The objective of this analysis was to identify the specialized scientific and technical terminology present in the texts, which could pose difficulties in the translation process.

The most common specialized terminology was identified by using *AntConc´*s option “Word List” and through simple sight observation, as shown in **Screenshot 7**. *AntConc* presents this word list in alphabetical order, and provides information about the frequency of their appearance in the texts analyzed.



Screenshot 7. Analysis of specialized terminology

The analysis of the technical-scientific terminology extracted from the texts included in *ParaCorp2020* helped us to have a clearer insight into the vocabulary used in texts on physics, chemistry, biology, astronomy, medicine and artificial intelligence. In the same way, it provided a clearer notion of the most common terminological and translational difficulties that our students must cope with when translating science and technology texts. The results derived from the word bank show many similarities with the linguistic phenomena and difficulties described in the previous section about the most common terminological difficulties in the translation of science and technology texts.

The presence of scientific and technical words in *ParaCorp2020* is notorious, and the following nouns exemplify this assertion: anode, cathode, light-emitting diodes, tungsten, nucleotide, rotaxane, alkaline, antigen, arthralgia, microcephaly, angioplasty, carcinogen, neoplasm, cryo-electron microscopy, etc. Also, there are many semi-scientific and technical words such as: conductor, black hole, battery, fuel, interstellar, carbon rods, antibacterial, chemotherapy, outbreak, parasite, prevalence, etc.

Moreover, there are several cases of academic abstract nouns, which rise the complexity level of the texts; a common feature of scientific writing. The word bank extracted from *ParaCorp2020* exhibits some of these abstract nouns, such as: disruption, discomfort, condition, onset, screening, radiation, staining, seizure, etc. These nouns are mostly used to indicate means, existence, tools, states, and results of actions, or to name particular phenomena or diseases, such as cancer, dementia, diabetes, encephalitis, neuropathy, stroke, glioblastoma, COVID-19, hypertension, etc.

Abstract adjectives also abound among the new vocabulary extracted from *ParaCorp2020*; words such as: halogen, electromotive, zoonotic, carcinogenic, ionizing, malignant, unorthodox, palliative, etc. There are also some adjectives that are not translated in the texts as to their plain meaning. For instance, green is translated as “ecológico”, rather than “verde”. Also, some other adjectives are rather translated into their formal and less common entry, than to the more colloquial one; for instance, cheap is translated as “menos costoso” instead of “barato”.

Another important feature of the technical-scientific vocabulary included in the word bank, it the presence of common verbs that are translated into their formal equivalent in Spanish, as well as the presence of new and specific verbs that are not of common knowledge to the translator, since they belong to the scientific jargon. For instance, fight-combatir, accelerate-agilizar, bolster-reforzar, drive-impulsar, compound-agravar, halve-reducir a la mitad o a un 50%, lessen-atenuar.

2.3.1 Word bank of scientific and technical terminology

The use of a self-compiled corpus offers many highways towards the purpose pursued. In this case, the parallel corpus compiled has the aim of helping students enhance their translation competence, which includes their preparation in lexicological terms, in order to face the intrinsic difficulties that this level of the language poses in specialized topics such as science and technology.

From the analysis of corpus-related research works, particularly those related to translation and the acquisition of competences, it was concluded that the extraction of a word bank is a very common practice derived from the practical work

carried out with a corpus, since it provides language researchers with the key vocabulary on a topic. It can be embedded as academic materials for language teaching, and for individual development of language skills, especially fluency, spelling and writing. (Rodríguez-Inés, 2008)

Therefore, derived from an analysis of the vocabulary present in the parallel corpus, carried out with *AntConc* corpus tool, it was possible to compile a word bank of the most relevant scientific and technical (ST) terminology extracted from the texts of the corpus. This word bank was organized as shown in **Annex 2.**

The terms of the word bank were organized in accordance with the topics studied in the subject Translation of Science and Technology Texts, with headings indicating so. It comprises 478 entries in English and their corresponding equivalents in Spanish, for a total of 1901 words in both languages.

This word bank will serve as an aiding material for the acquisition and systematization of specialized terminology, necessary for translating the texts in class and for the future professional development of the students as translators. Through its use, accompanied by the parallel corpus (*ParaCorp2020*) in the translation lessons, it was possible to derive a group of corpus-based activities that help integrating *ParaCorp2020* in the translation lessons, with the aim of enhancing the translation competence of the fourth-year students of English Language with a Second Foreign Language at Universidad Central “Marta Abreu” de Las Villas.

2.4 Proposal of exercises to use ParaCorp2020 in the translation lessons

In order to provide a greater scope to the use of the parallel corpus proposed in this research (*ParaCorp2020*), this section’s objective is to propose a pedagogic application for it, in order to contribute in a practical way to the enhancement of the translation competence of our students, specifically in the translation of science and technology texts.

This proposal consists of a sample of exercises, which have the purpose of motivating the students and introducing the use of *ParaCorp2020* in the translation

class as a didactic tool, in order to help students to enhance their translation competence.

The objectives of this proposal of exercises are:

i. To facilitate the acquisition of knowledge and skills through the use of *ParaCorp2020* for translation

ii. To enable the acquisition of knowledge and skills necessary to translate about ST texts.

iii. To provide a variety of contextualized examples of equivalents in both languages, so that students can incorporate the ST vocabulary that appears in the texts and that sometimes is not found in the dictionary.

iv. Provide a range of solutions to different translation problems due to differences in grammatical and syntactic structures of both languages.

v. To provide didactic activities for practicing translation, which does not only consist of "reading and translating", but also intend that students can make generalizations, discover principles and necessary skills when translating.

Each exercise sheet will work with a different text. They will serve as a lesson´s guidelines and will have an **introduction**, a **body** and **conclusions** as an ordinary lesson. The introduction comprises a warm-up section with general familiarization questions in order that the students can have a preliminary idea of the topic they will be dealing with during the lesson, as well as to check the background knowledge they have about it. After the warm-up, it is important to state which field and specific topic will be treated in the lesson, and to establish its objectives.

The **body** of the exercise sheet will comprise the presentation of the English-Spanish Sentence-Aligned Parallel Corpus of Science and Technology Texts (at least in a first lesson), and the corpus-based translation exercises to be completed by the students. Many of these exercises will be interactive, so that the teacher and

the students can search together for possible alternatives, and strengthen the translation process. The techniques used in the lesson should not be based on the prevalence of the teacher´s solutions over the students´, but on a mutual collaboration and dialogue as to find the best solutions and most importantly, that the students can systematize new terms, translation principles and solutions. They need to focus on the process and the acquisition of techniques, without neglecting the quality of the results. The acquisition and development of translation competence through the continuous practice, will contribute to the improvement of the translations. Finally, in the **conclusions**, the teacher and students will review all the translation difficulties, new **techniques, and new terminology** learned during the lesson in order to systematize and fix the knowledge acquired.

Table 3. Layout of the exercise sheets

Exercise Sheet (Layout)	
Introduction	Warm-up. State field, topic and objectives of the lesson
Body	Presentation of *ParaCorp2020* and the text to translate in class. Complete 1-10 corpus-based translation exercises. Carry out interactive revision of the translation, as well as an analysis of the difficulties and its possible solutions through the search of alternatives in the corpus and interactive debate in class.
Conclusions	Summarize the new terminology and techniques learned in class. Clear out possible doubts. Ask students for feedback about the translation exercises and the use of *ParaCorp2020*.

Examples of the exercise sheets based on the use of *ParaCorp2020*

Exercise Sheet 1

Domain: Biology

Topic: Cryo-EM

Objectives: Students should be able to:

1- Become familiar with the topic of Cryo-EM.

2- Become familiar with the terminology of the topic through the extraction of the vocabulary in the source text (English), the equivalents appearing in *ParaCorp2020* and in the word bank extracted from this corpus.

3- Find solutions to translation difficulties through the use of *ParaCorp2020.*

Exercises:

Warm-up:

1- Are you interested in current scientific advances in biology?
2- How do you think science can help in the development of the world?
3- Do you know what a molecule is?
4- Have you ever read or heard something about Cryo-EM technique?
5- Any guesses of what it can be?

- The teacher explains that in today´s lesson our task will consist in translating a text about Cryo-EM.

Body of the lesson:

1- Let us search in Wikipedia or internet what Cryo-EM is. (Have a debate in it, trying to guess what its importance and main features are)

Exercises with the corpus

1- Read carefully the following text titled: *The development of cryo-EM into a mainstream structural biology technique.*
2- Do you find it difficult to read? Why?

3- Read the text again and this time extract the terms, verbs, expressions, etc. you do not know (It is advisable to underline or highlight in the text for better visibility). E.g.: transmission electron microscopy (TEM).

4- Try to find their equivalent in the dictionaries you have. Could you find an equivalent in the dictionary for all of them? (The answer will be “no”, since the vocabulary is very specific and complex)

-The teacher introduces *ParaCorp2020* to the students and explains what it is.

5- Now, search the text *El desarrollo de la criomicroscopía electrónica en una técnica de biología estructural de uso generalizado* in the corpus. Find the equivalents in Spanish for the terms you identified as unknown in exercise 3.

6- Write a list with the terms in English and their equivalents in Spanish.

7- Now, try to translate the text on your own. If you find any difficulty in the text, do not look for a solution in the corpus until the final revision of the translation, which will be carried out in 40 minutes with the teacher and the rest of the students.

<u>Revision of the translation</u>

- If the students disagree about the translation of a phrase, or if it is difficult to find an equivalent, you can go to the corpus and try to find a solution together.
- Try to make generalizations about the difficulties found in the translation process so that students leave the classroom with an idea of how to solve them.

Exercise Sheet 2

Domain: Medicine

Topic: COVID-19

Objectives: Students should be able to:

1- Continue becoming familiar with the topic of COVID-19 pandemics.

2- Become familiar with its particular terminology through the extraction of the vocabulary in the source text (English) and the equivalents through the use of the English-Spanish Sentence-Aligned Parallel Corpus of Science and Technology Texts *ParaCorp2020* and the word bank extracted from this corpus.

3- Find solutions to translation difficulties through the use of the English-Spanish Sentence-Aligned Parallel Corpus of Science and Technology Texts *ParaCorp2020.*

Exercises

Warm-up:

1- It is no secret for every inhabitant of the planet that COVID-19 is a pandemic that has affected the world in an unprecedented manner. Share briefly with your teacher and classmates what has impacted you the most about this experience.

2- Have you incorporated new vocabulary after this world-wide phenomenon, either in your mother tongue or in English?

3- COVID-19 is a very sad event that has affected our lives in an unprecedented manner. Media worldwide has covered this issue in every way possible. As translators and interpreters, it is important for us to get acquainted with this issue and all the new vocabulary and linguistic phenomena it brings about. Let's analyze and translate the following text related to the topic of COVID-19, titled "**Global partnership to make available 120 million affordable, quality COVID-19 rapid tests for low- and middle-income countries."**

Body of the lesson:

8- Read carefully the two first pages of the text titled: "**Global partnership to make available 120 million affordable, quality COVID-19 rapid tests for low- and middle-income countries."**

9- Read the text again and this time extract the terms, verbs, expressions, etc. you do not know (It is advisable to underline or highlight in the text for better visibility). E.g.: antigen-based rapid diagnostic tests

Work with terminology

10- Search the unknown words in the dictionary as long as possible. There will be some specific terms and that will not appear in the dictionary, for those cases we have provided the word bank extracted prom the corpus *ParaCorp2020* where most of the difficult or specialized terminology related to the topic is summarized in the section **Communicable diseases: COVID-19.**

11- How would you translate the phrasal verb "make available" into Spanish? Please, search in the corpus what was the equivalent used by the translator. Do you find it interesting? Do you agree with his/her selection?

12- Find in the text the verb *"to execute".* What is the main entry for this verb in the bilingual dictionary? Do you think this is the equivalent that best fits in this case? Search in *ParaCorp2020* the equivalent used by the translator for this verb in the source text.

13- Search in the sixth paragraph of the text two verbs you think can be translated into the Spanish expression *"poner en marcha".*

Work with difficult expressions and sentence structures

14- Now, try to translate the text. You will find some difficult expressions and sentences, for instance:

The tests developed by Abbott and SD Biosensor are highly portable, reliable, and easy to administer, making testing possible in near-person, decentralized healthcare settings.

* English verbs ending in **-ing** are always difficult to translate into Spanish due to the complexity of the correct use of gerunds. Search in the corpus what was the solution given by the translator to this sentence and try to apply it to similar cases in the text.

-The teacher will carry out a similar analysis with other expressions and sentence structures signaled by the students, in order to find the best solutions for translating and making the Spanish version pleasant to the reader in style and accuracy.

-The most important goal to achieve is that through the corpus-based analysis of the texts, and the use of the word bank, the students can acquire knowledge that contributes to their translation competence and are able to apply it in future translations.

Conclusions of the lesson

For the conclusions of the lesson, it is important to summarize and try to recall all the new vocabulary and translation strategies learned through the translated text.

All these materials have been suggested to the translation professors of the English Language Department of Universidad Central “Marta Abreu” de Las Villas, for their further use in the translation lessons.

Conclusions of Chapter II

This chapter described the methodology and different stages implemented in order to compile the English-Spanish Sentence-Aligned Parallel Corpus of Science and Technology Texts *(ParaCorp2020)* proposed in this research work. It dealt with the diagnosis of the perceptions and needs of the fourth-year students of the English Language with Second Foreign Language (French) degree course in relation with the translation of Science and Technology texts. It also covered the criteria and steps followed for compiling and processing the texts of the corpus, and the preliminary results obtained from the analysis of *ParaCorp2020* through corpus-processing tools.

CONCLUSIONS

Conclusions

Being proficient translators is one of the most important goals for the students of the English Language with Second Foreign Language (French) degree course at Universidad Central “Marta Abreu” de Las Villas. The present research aims at contributing to the endless effort made by language institutions around the world to train proficient translators.

1. This research has been theoretically framed upon the acquisition of translation competence through a process-oriented approach, corpus linguistics and its bonds with translation, and the benefits brought about by the use of a parallel corpus in translation.

2. The analysis of the context evidenced that the fourth-year students of the degree course English Language with a Second Foreign Language (French) present difficulties when translating science and technology texts, which hinder their translation competence. These difficulties are associated with the selection of accurate equivalents, the insufficient knowledge on the terminology and topics related to science and technology, and the need for offline bilingual consulting sources to support the translation process.

3. The English-Spanish Sentence-Aligned Parallel Corpus of Science and Technology Texts *ParaCorp2020* was an authentic and innovative idea in response the deficiencies in the enhancement of the translation competence. It includes a representative group of English texts and their translations into Spanish, the most important terminology, several examples of translation solutions in different fields of science and technology. A word bank was derived from the corpus, as a supporting material for the acquisition and systematization of specialized terminology and a group of corpus-based activities was elaborated to aid the translation process in the classroom.

RECOMMENDATIONS

Recommendations

For the further development of this research, the author recommends:

I. To use of *ParaCorp2020* in the translation lessons of the fourth-year English Language students next semester.

II. To continue adding texts of *ParaCorp2020* in order to enlarge this tool and provide a wider range of examples of translations for the students to work in the translation lessons. Consequently, to enlarge the word bank of scientific and technical terminology.

III. To publish in Moodle platform.

BIBLIOGRAPHICAL REFERENCES

Bibliographical References

Abouzar, Oraki & Gholam-Reza, Tajvidi. (2020). Training Translators and Interpreters: The Need for a Competence-Based Approach in Designing University Curricula. *Iranian Journal of English for Academic Purposes*, 1.

Anexiam Language Services. (2016). Retrieved December 3, 2020, from https://anexiam.com/en/services/general-translation

ApSIC. (2017). Retrieved November 9, 2020, from http://www.xbench.net/

Baker, M. (1995). Corpora in Translation Studies: An Overview and Some Suggestions for Future Research. *Target*. doi:https://doi.org/10.1075/target.7.2.03bak

Bausela, M. B. (2016). The importance of corpora in translation studies: a practical case. *Research-publishing.net*, 1.

Bernardini, S. (2015). Exploratory learning in the translation/language classroom: Corpora as learning aids. *CULT Conference.* Alicante.

Biber, Douglas, Conrad, Susan and Reppen, Randi. 1998. *Corpus Linguistics: Investigating Language Structure and Use.* Cambridge: Cambridge University Press.

Bogucki, Ł. G.-R. (2015). *Ways to translation, Translation competence.* Łódź-Kraków: Łódź University Press and Jagiellonian University Press.

Burrow-Goldhahn, A. (2021). *The Importance of Translation Studies*. Obtenido de Translation studies at Exeter: https://blogs.exeter.ac.uk/translation/blog/2018/06/19/the-importance-of-translation-studies/

Bowker, Lynne and Pearson, Jennifer. (2002). *Working with Specialised Language: A Practical Guide to Using Corpora.* London: Routledge.

Church, Kenneth Ward and Mercer, Robert L. (1993). Introduction to the special issue on computational linguistics using large corpora. *Computational Linguistics* 19 (1): 1-24.

Claudio Fantinuoli and Federico Zanettin. (2015). *Creating and using multilingual corpora in translation studies. Berlin: Language Science Press.* Berlin: Language Science Press.

Comply. (2017). Retrieved on November 20th, 2020, from blog.v-comply.com/user-friendly-software/

Corpas Pastor, Gloria and Seghiri, Miriam. (2010). *Size matters: A quantitative approach to corpus representativeness.* Málaga: Universidad de Málaga.

Cushing, S. T. (2017). Corpus linguistics in language testing research. *Language Testing*, 441.

Daza, I. C. (2014). *La traducción: problemas y métodos. La traducción inglés-español.* Cádiz, España.

Eser, O. (2015). Model of Translator´s Competence from an Educational Perspective. *International Journal of Comparative Literature and Translation Studies, 3*(1), 2-4. Retrieved from https://www.journals.aiac.org.au/index.php/IJLTS/article/view/1279

Facchinetti, R. (2007). *Corpus Linguistics 25 Years on.* Amsterdam: Editions Rodopi B.V.

Fantinuoli, Claudio and Zanettin, Federico. (2015). *Creating and using multilingual corpora in translation studies. Berlin: Language Science Press.* Berlin: Language Science Press.

Foster, M. (1958). *Translation from/into Farsi and English.* Retrieved April 2007, from Parsa: http://www.parsa-ts.com/index.htm.

Frérot, C. (2016). *Corpora and Corpus Technology for Translation Purposes in Professional and academic Environments. Major Achievements and New Perspectives.* universidad de Grenoble.

G. Altbach, Philip and de Wit, Hans. (2018). Too much academic research is being published. *University World News*.

Garcia, I. (1992). A tradução do texto técnico-científico. *In Ilha do Desterro,N°.28*, 75-85.

Gerding-Salas, C. (2018). Teaching Translation: Problems and Solutions. *Translation Journal 2000.*

Ghadessy, Mohsen; Henry, Alex and Roseberry, Robert L. (2001). *Small Corpus Studies and ELT: Theory and Practice.* Amsterdam/Philadelphia: John Benjamins.

Gile, D. (2018). The Process-Oriented Approach in Translation Teaching. In C. D. Lindegaard, *Teaching Translation and Interpreting 2: Insights, aims, visions.* Amsterdam/Philadelphia: John Benjamins Publishing Company.

Godfrey, K. (2019). *OmegaT*. Retrieved March 21, 2020, from https://omegat.org

Gouadec, D. (2007). *Translation as a Profession.* Amsterdam: John Benjamins.

Gregoline, B. (2002). *Translating Scientific Texts: Practicalities and Pitfalls .* Chicago: American Medical Association.

Guo-rong, S. (2010). Corpus-based Approaches to Translation Studies. *Cross-Cultural Communications, Vol. 6, No. 4*, 181-187.

Hurtado, A. (1999). *Enseñar a traducir. Metodología en la formación de traductores e intérpretes.* Madrid: Edelsa.

Kaur, K. (2005). A Competent Translator and Effective Knowledge Transfer. *Translation Journal, 9*(4), 1-2. Retrieved 4 5, 2021, from https://translationjournal.net/journal/34edu.htm

Kraif, O. (2018). *From Translational Data to Contrastive Knowledge: Using bi-texts for Bilingual Lexicon Extractions. International Journal of Corpus Linguistics.* John Benjamins Publishing.

Krzysztof, Kolk and Krzysztof, Marasek. (2015). *A Sentence Meaning Based Alignment Method for Parallel Text Corpora Preparation.* ResearchGate. doi:10.1007/978-3-319-05951-8_22

Kussmaul, P. (1995). *Training the Translator.* Amsterdam/Philadelphia: John Benjamins Publishing Company.

Lavion. (2002). *Corpus-based Translation Study Theory, Findings, Application.* Amsterdam: Rodolpi.

Laviosa, S. (2010). *Corpora. Handbook of Translation Studies.* Benjamis Publishing.

Lawrence, A. (2014). AntConc (Version 3.4.3w). *[Computer Software]*. Tokyo, Waseda University, Japan. Retrieved November 9, 2020, from http://www.antlab.sci.waseda.ac.jp/

Lindquist, H. (2009). *Corpus Linguistics and the Description of English.* Edimburg University Press.

Livbjerg, Inge and Mees, I. M. (2002). Problem-solving at different points in the translation process: quantitative and qualitative data. *In Empirical Translation Studies: Process and Product [Copenhagen Studies in Language Series 27]* (pp. 147-190). Copenhagen: G. Hansen.

McEnery, Anthony and Wilson, Andrew. (2001). *Corpus Linguistics.* Edimburgh: Edinburgh University Press.

Muhiesen, Esra'a M. and Al-Ajrami, Mona S. (2019). Challenges in Translating Technical Texts. *Dirasat, Human and Social Sciences, 46*(1), 1-3.

Chávez Hernández, Marisleidy and García Fundora, Dianelys. (2011). The Use of an English- Spanish Parallel Corpus for the Study of Translation.

PACTE. (2003). Building a Translation Competence Model. *Triangulating Translation: Perspectives in process oriented research* (pp. 43-66). Amsterdam: F. Alves.

Pérez Castellano, N. (2017). Bilingual glossary (English-Spanish) of diplomatic terms. Master's Degree Dissertation. Universidad Central "Marta Abreu" de Las Villas.

Ratzhavova, D. (2020). *Difficulties of the translation of scientific and technical texts //.* Retrieved from https://moluch.ru/archive/147/41270/

Rodríguez-Inés, P. (2008). *Uso de Corpus Electrónicos en la Formación de Traductores (inglés-español-inglés).* Master´s Degree Dissertation. Barcelona.

Rojo, J. L. (2018). Diseño y compilación de corpus paralelos alineados: dificultades y (algunas) soluciones en el ejemplo de un corpus de textos museísticos traducidos (inglés-español). *Revista de Lingüística y Lenguas Aplicadas https://doi.org/10.4995/rlyla.2018.7912, 13*, 59-73. doi:https://doi.org/10.4995/rlyla.2018.7912

Schäffner, C. (2000). Running before Walking? Designing a Translation Programme at Undergraduate Level. *Developing Translation Competence*, 143-150.

Šeböková, S. (2010, February 10). Comparing Translation Competence. Theses.cz. Retrieved on February 22nd, 2021, from https://is.muni.cz/th/146168/ff_m/Stasectc.pdf.

Sinclair, J. (1996) *EAGLES Preliminary recommendations on Corpus Typology, EAG--TCWG-CTYP/P.* Online:http://www.ilc.pi.cnr.it/EAGLES96/corpustyp/corpusty p.html.

Stubbs, M. (2001). *Texts, corpora, and problems of interpretation: A response to Widdowson.*

Sue Atkins, J. C. (1992). Corpus Design Criteria. *Literary and Linguistic Computing, 7*(1), 1-16. doi:https://doi.org/10.1093/llc/7.1.1

Vargas Sierra, Ch. (2006). Diseño de un corpus especializado con fines terminográficos: el corpus de la piedra natural. *Debate Terminológico*, 2 (7/2006). París: RITERM (Red Iberoamericana de Terminología).

Vieira, R. J. (2020). *The Translation of Technical-Scientific Texts. A Brief Analysis.* UFSC.

Xiaoping, (Isadora) Jiang and van Rij-Heyligers, Josta. (2008). Parallel Corpus in Translation Studies: An Intercultural Approach. *The international symposium on Using Corpora in Contrastive and Translation Studies*, (pág. 2). Hangzhou.

Xiuhua, Li and Li, Li. (2015). Characteristics of English for Science and Technology. *International Conference on Humanities and Social Science Research.* Beijing: Atlantis Press.

Zanettin, F. (2002). Corpora in translation practice. *Language resources for translation work and research*, 10-14.

Zanettin, F. (2012). *Translation-driven corpora. .* Mánchester: St. Jerome.

ANNEXES

Annex 1: Survey

Survey applied to fourth-year students

Academic Year: 2019-2020

The present survey has the purpose to know your perceptions of the translation process and the difficulties you face concerning the subject Translation of Science and Technology Texts. We also want to know your opinion about the use of a bilingual parallel corpus in the translation lessons as an aiding tool to enhance your translation competence.

We thank you for your collaboration.

Note: A <u>Parallel Corpus</u> is a collection of machine-readable texts, comprising documents in one language (L1) aligned with their translations (L2).

1- How do you rate the availability of bilingual reference sources for translating in class? (Specifically good offline bilingual dictionaries, translators, glossaries, etc.)

a) __ Enough b) __ A few c) __Insufficient

2- Which of the following tools do you have at hand when translating in class or at home?

__ monolingual dictionaries __bilingual dictionaries __ bilingual corpora

__ glossaries on socio cultural texts __ google translator __ Linguee

__ encyclopedias __others (specify)

3- How do you feel when translating specific terms and collocations in Science and Technology texts?

a) __ certain of doing a good translation b) __ uncertain due to lack of knowledge

4- Do you think the time given to translate in class is enough?

a) __ Yes b) __No

5- The professor orients some independent study for searching information about the topics dealt with in science and technology texts prior to their translation in class?

__Always

__Never

__Almost always

4- How dependent are you on online translators?

a) __ Very dependent b) __ Not that much dependent c) __Very little dependent

4.1- Do you think that using ONLY online translators is enough to improve your translation competence and guarantee you an accurate translation? Support your answer in each case.

__Yes

__No

5- Are you given different solutions to translation problems when they appear in the translation process of science and technology texts?

a) __ Always b) __ Sometimes c) __Rarely d) __ Never

6- Do you use any bilingual corpus as an aiding tool to translate in the subject Translation of Science and Technology Texts?

a) __ Yes b) __No

7- To what extent a bilingual corpus could help you solving a translation difficulty in science and technology texts? (Provide as many arguments as you can)

Annex 2: Sample of the word bank

Topic	Classification (word category)	English	Spanish
Electric lighting and LEDs	**Noun/noun group**	150-watt incandescent bulb	Tubo incandescente de 150 vatios
		40-watt fluorescent tube	Tubo fluorescente de 40 vatios
		Anode	Ánodo
		Arc lamps	Lámparas de arco
		Argon	Argón
		Art of engineering	Trabajo de ingeniería
		Bulb	Bombilla
		Cathode	Cátodo
		Compact fused quartz glass tubes	Tubos compactos de vidrio de cuarzo
		Conductive plates	Placas conductoras
		Conductor	Conductor
		Devices	Dispositivos
		Efficient blue LED	Led azul eficiente
		Electric discharge lamps	Lámparas de descarga eléctrica
		Electric discharge vapor lamps	Lámparas de vapor por descarga eléctrica
		Electric lighting	Iluminación eléctrica
		Electrical circuit	Circuito eléctrico
		Electrical current	Corriente eléctrica
		Electrical discharge	Descarga eléctrica
		Electrical discharge device	Dispositivo de descarga eléctrica
		Electrical energy	Energía eléctrica
		Electrical engineering	Ingeniería eléctrica

		Electrical imput power	Potencia eléctrica
		Electrical power	Corriente eléctrica/electricidad
		Electrical power grid	Red eléctrica
		Electrical signals	Señales eléctricas
		Electrolyte	Electrólito
		Electrons	Electrones
		Energy-efficient	Con eficiencia energética/energéticamente eficiente
		Energysaving	Que ahorran energía
		Environmentally-friendly	Favorable al medio ambiente
		Filament	Filamento
		Film	Película
		Fluorescent lamps	Lámparas fluorescentes
		Glass bulb	Ampolla de vidrio
		Halogen gases	Gases halógenos
		Halogen lamps	Lámparas halógenas
		High-pressure sodium vapor lamps	Lámparas de vapor de sodio de alta presión
		High-speed gas discharge strobe lamp	Lámpara estrobscópica de descarga de gas a alta velocidad
		Hundreths of a second	Centésimas de segundo
		Incandescent lamps	Lámparas incandescentes
		Inductance coil	Bobina de inductancia
		Inert gas	Gas inerte
		Insulating material	Material aislante
		Ionization	Ionización
		Krypton	Criptón

		Layered semiconductor materials	Materiales semiconductores en capas
		Lead	Plomo
		LED Lamps	Lámparas de ledes
		LEDs	Ledes
		Light beams	Haces de luz
		Light energy	Energía luminosa
		Light source	Fuente de iluminación/ Fuente luminosa
		Light-emitting diodes	Diodos emisores de luz
		Long-lasting	Duradero
		Low-pressure mercury vapor lamps	Lámparas de vapor de mercurio de baja presión
		Lumen	Lúmenes
		Luminescence	Luminescencia
		Melting point material	Material de punto de fusión
		Neon lamps	Lámparas de neón
		Output	Potencia
		Panel lighting	Iluminación de paneles
		Phosphor coated tube	Tubo revestido con fósforo
		Photocells	Células fotoeléctricas
		Photons	Fotones
		Prerequisites	Requisitos previos
		Rare gas	Gas poco común
		Semiconductors	Semiconductores
		Shop windows	Mostradores
		Temperatures above...	Temperaturas superiores a...
		Thermal energy	Energía térmica
		Tin oxide	Óxido de estaño
		Transluscent ceramic tubes	Tubos de cerámica translúcida
		Tungsten	Tungsteno/Volframio
		Vacuum	Vacío

		Watt	Vatios
		Wavelength	Longitud de onda
		Wire filament	Filamento de alambre
Molecular biology		Antibodies	Anticuerpos
		Billions	Mil (es) de millones
		Biofuels	Biocombustibles
		Catalyst	Catalizador
		Chemistry of life	Química de la vida
		Enzyme	Enzima
		Green	Ecológico
		Harness	Aprovechar
		Pharmaceuticals	Fármacos
Molecular engineering		Axle	Eje
		Cathenane	Catenano
		Groundbreaking	Innovador
		Nanocar	Nanocoche
		Rotaxane	Rotaxano
		51 Pegasi b	Dimidio
Astronomy		Exoplanet	Exoplaneta
		Solar-type star	Estrella de tipo solar
Batteries		Alkaline or nickel-iron battery	Pila alcalina o batería de níquel y hierro
		Ammonium choride	Cloruro de amonio
		Carbon rod	Varilla de carbono
		Chemical energy	Energía química
		Chemical reaction	Reacción química
		Dry cell	Pila seca
		Electric battery	Pila eléctrica

		Electric wristwatches	Relojes de pulsera eléctricos
		Electromotive force	Fuerza electromotriz
		Energy-rich (material)	Material rico en energía
		Fossil-fuel free	Libre de combustibles fósiles
		Free electrones	Electrones libres
		Fuel cells	Pilas de combustible
		Gallium nitride	Nitruro de galio
		Hearing aids	Audífonos
		Ion conductor	Conductor iónico
		Lead dioxide	Dióxido de plomo
		Lead sulphate	Sulfato de plomo
		Lead-acid batteries	Baterías de plomo y ácido