Universidad de Valladolid

FACULTAD de FILOSOFÍA Y LETRAS

DEPARTAMENTO de FILOLOGÍA INGLESA

Grado en Estudios Ingleses

TRABAJO DE FIN DE GRADO

An English/Spanish Glossary in Proteomics: An Approach to Science

Tamara Modrego Merino

Tutor: Beatriz Méndez Cendón

2013/2014

ABSTRACT

This undergraduate dissertation deals with the compilation of a bilingual (English/Spanish) glossary in the subject field of Proteomics, the study of proteins. To compile this glossary we have followed a systematic ellaboration of terminology. Therefore, our methodology analyses what 'genre', 'text type' and 'register' are, the importance of the field of 'terminology' and the relevance of the studies in the domain of 'corpus linguistics'. We have followed Cabré's (1993) stages for the compilation of terms out of our corpus, as well as the steps to design a glossary by Frankervert García et al. (2001). The aim of this dissertation is to give the Spanish experts a bilingual terminographic tool that would help them to understand and write their papers in English in the field of Proteomics since English is the international language of scientific texts. We have found that there is not such a tool in this specialized domain. We pretend to simplify their task, so the spread of their findings in an international environment will be a little bit easier.

Key words: proteomics, terminological glossary, corpus linguistics, terminology, terminological entry.

RESUMEN

En este Trabajo de Fin de Grado (TFG) se describe y realiza la compilación de un glosario bilingüe (inglés / español) en el campo de la Proteómica, el estudio de las proteínas. Para compilar este glosario hemos seguido una elaboracion sistemática de la terminología. Por lo tanto, nuestra metodología analiza los conceptos de "género", "tipo textual" y "registro", así como la importancia de la terminología y del análisis del corpus para llevar a buen puerto la elaboración de esta herramienta terminográfica. Además, hemos seguido las indicaciones de Cabré (1993) sobre la identificación y extracción de términos del corpus, así como los pasos para compilar un glosario de Frankervert García et al. (2001). El inglés internacional es la lengua de los textos científicos; sin embargo, los expertos españoles en este campo no tienen un diccionario especializado que les facilite el trabajo de leer y escribir en inglés. Por lo tanto, este TFG pretende facilitar su labor, para que la difusión de sus resultados en un entorno internacional sea un poco más fácil.

Palabras clave: proteómica, glosario terminológico, lingüística del corpus, terminología, entrada terminológica .

INDEX

LIST OF TABLES

LIST OF IMAGES

LIST OF EXAMPLES

1. GENERAL INTRODUCTION

Proteomics is the study and characterization of proteins which are expressed in a genome in order to identify the cell proteome. Likewise, the cell proteome is the set of proteins in a cell, tissue or organ in a state of differentiation and development. This branch of Biology has become a priority research line, though being a relatively new and undeveloped field. This project was written to assist in the elaboration of publications by Spanish Proteomics researchers or experts in the field. These publications are written in English with the aim of raising awareness of the research done in our country. Therefore, this glossary will be a starting point for the development of such articles.

Need of the project

Regarding Terminology, there is a lack of Spanish glossaries in the field of Proteomics, because:

English language is taken as a reference in publications and research articles, in order to secure international prestige and dissemination of the results. Therefore, anyone who wants to write an article should do it in English, which involves a double handicap: a linguistic difficulty with the language and a complexity of the medical terminology.

There is no precedent of an English/Spanish bilingual glossary. Such technical dictionaries have been produced in English, since, as we have explained above, it is the *lingua franca* of scientific language.

Therefore, our work aims to make a glossary that allows the scientific community in the field of Proteomics to easily write research papers or publications in English.

Description and aims of the project

In order to compile a good glossary, we should have background knowledge in four different fields: ‘register’, ‘genre and textual types’, ‘corpus linguistics’ and ‘terminology’. The analysis of register, textual type and genre will permit us to know the structure and composition of scientific research articles that we will deal with. Corpus linguistics will give us an understanding of methodology for the compilation of

our bilingual corpus in English and Spanish. Furthermore, the field of terminology will help us to extract candidate terms and to choose the most adequate ones for our glossary. So, we will analyse, first, the terms in the field of Proteomics and then, the necessary steps for the development of the glossary.

Our work will consist of a bilingual corpus in English and Spanish of 100,000 words each, taken from research papers extracted from five journals and three PhDs: four research journals in English (*Biochimica et Biophysica Acta (BBA), Genomics, Proteomics & Bioinformatics (GPB), Journal of Proteomics* and *Proteomics)* and a Spanish journal (*Revista de la Sociedad Proteómica*) and three PhDs (*Proteómica Bottom-up y Top Down de Organismo poco Secuenciados, Proteómica de Expresión Diferencial en Acinetobacter baumannii Resistente a Colistina,* and *Estudios de la Alteración de Proteínas en el Tejido Adiposos Omental y Subcutáneo en Obesidad mediante Técnicas de Proteómica*). These articles have a major impact in the area of Proteomics since journals are read by many experts of this field and the PhDs provide new information and discoveries.

We will use AntConc, a computer program that allows us to extract KWIC concordance lines in a corpus for the analysis of these texts.

Content parts

This work consists of two main parts: a theoretical part and methodological part.

The first part or the theoretical section constitutes the background information for the methodology. In the first three subsections the concepts of ‘gender’, ‘register’ and ‘textual type’ will be analysed. A description of what ‘corpus linguistic analysis’ is and also of our bilingual corpus will be done in the fourth subsection. Finally, the fifth subsection deals with the concept of ‘terminology’. We will explain the relationship between terminology and translation, what terms are and their types, and how they are formed and used in language.

The second part is divided into two sections: (1) the methodology for the compilation of a corpus and (2) the methodology for the development of a glossary.

In the section on the methodology of compiling our corpus, we will explain the software program 'Antconc', as well as how we have extracted keywords. Besides, we will follow Cabré (1993) and Sager's (1990) methodology.

The next section will explain the steps according to Frankerverg-García, A. et al, eds. (2011) for the development of a glossary, as well as for the formulation of definitions and the importance of the supervision of an expert.

This dissertation will end with the results, which is the English/Spanish glossary, and the conclusion in terms of the theoretical and methodological aspects discussed.

Finally, the project ends with a list of the references used in this work.

2. THEORETICAL FRAMEWORK

There are different types of texts: political, medical, journalistic, etc. The development of the field of linguistics, rhetoric, language, etc. has influenced our point of view about text types. In this section we will study the analysis by Trosborg (1997) on register, gender and text type.

2.1 The concept of register

Malinowski's work (1923) and subsequently Firth's work (1935, 1951) have been essential in defining what register is in a text. Malinowski's theory stated that cultural context, which is from the ritual to the everyday aspects of daily life, was basic for interpreting the message. Later, Firth argued that cultural factors influenced and determined the linguistic choices, that is, Firth gave importance to the situation and culture.

Later, Reid (Reid, 1956) found that language diverges depending on the 'varieties' of language called *records*. Halliday et al. (1964) took this statement and made a classification according to the variety of the speaker: (1) dialects and (2) registers. The dialects were geographical, social or temporal variations, while registers were an open set of varieties or styles from occupational fields, such as the language of religion, legal language, medical, etc. were termed dialects.

Finally, Gregory and Carroll (1978) define register as: "a contextual category correlating groupings of linguistics features with recurrent situational features".

2.2 The concept of text type

There is a misunderstanding between genre and text type, since texts of the same gender may represent several text types, while texts of different genres and linguistically similar can have the same text type. While gender refers to full texts, text type cuts across genres. In addition, gender refers to an open set of texts, whereas text type refers to a closed number of categories.

Hatim and Mason (1990) define text type as: "a conceptual framework which enables us to classify texts in terms of communicative intentions serving an overall rhetorical purpose".

In terms of classification of text types, there are several types of distribution. Kinneavy (1971, 1980) makes a classification of text types depending on the mode and distinguishes between two types: (1) static (look for something at a specific time) and (2) dynamic (look how it changes over time). Through these two main types, he concluded four subtypes: narration, classification, description and evaluation. Description and classification are included in the static view, whereas narration and evaluation are enclosed in the dynamic view.

Werlich's classification (1976) includes four types of text types: description, narration, exposition and argumentation. Later on, Hatim and Mason (1990) added another type: instruction. Referring to translation purposes there were two subcategories: instruction with option (e.g. advertisements, manuals, etc.) and instruction without option (e.g. legislative texts, contracts, etc.).

Related to the levels of text types, Trosborg (Trosborg, 1997) states that there are two levels, macrolevel and microlevel:

> "We need a two-level typology for text types rather than a single level of types only. At the macrolevel of discourse, text type may be assumed to precede the level of text-strategic choices, thus affecting the whole strategy of the text. The choice of microlevel text types on the other hand, has to do with the textualization process, which is determined by the text producer's text strategy."

To sum up, text types cut across genres and are divided into five types: description, narration, exposition, argumentation and instruction.

2.3 Genre and style

According to Webster's Third Dictionary, genre is "a distinctive type or category of literary composition". Registers are divided into genres because they reflect the social purpose of registers and how they are used. According to Instituto Secundario Lomas de Zamora (2014), there are seven types of genres: 'scientific', 'advertising', 'journalistic', 'legal', 'commercial', 'political' and 'literary'. In our work we will deal with the scientific genre, and more specifically, with research papers.

2.3.1 Research paper

According to Sager (Sager et al. 1980) a scientific article is a "written work that reflects the personal and individual and even idiosyncratic interpretation of a topic or issue towards the introspection or the experimental analysis".

Swales's scheme (1990) explains the main parts of a research paper. These parts are: Introduction, Methods, Research and Discussion (also called the IMRD structure).

Later on, Fortanet et al. (1998) added four more parts: Title, Abstract, Graphic data and References. Furthermore, we must take into account the subject field in a research paper, for example, pharmacy, medicine, chemistry, etc.

2.4 Analysis of corpus linguistics

With the aim of preparing our terminological entries of our glossary in Proteomics, we have compiled an electronic corpus in English and Spanish. This compilation has been made according to our previous reading on papers dealing with corpus linguistics and the types of corpora. Therefore, we could establish a correct criterion related to the length of our corpus, the languages to be included and its purpose. In our case, the aim is clear: extract the terms to compile a glossary on Proteomics.

EAGLES (Expert Advisory Group on Language Engineering Standards) (EAGLES, 1996a) defines corpus linguistics 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". John Sinclair (EAGLES, 1996b) established four minimum criteria for the development of a corpus: quantity, quality, simplicity of codification and documentation. Quantity and quality are closely related, since the quality of a corpus is measured by its inclusion criteria (Ezquerra, A. and Corpas Pastor, 1994), such as temporal or geographical limits or the levels of text. Besides, quality is related to quantity, as a balanced corpus in length is more representative than one that is not. Moreover, simplicity of codification and documentation allow a precise reference corpus of the documented sources used (Corpas 2001a). The variability of these four minimum will lead to different types of corpus.

2.4.1 Types of corpus linguistics

We take the typology by Llisterri et Torruella (1999) for the classification of linguistic corpus together with the contributions by Baker (1995), Johansson (1998) and Bravo Gozalo et al. (1998). According to them, there are five different criteria: (1) percentage and distribution of the corpus, (2) specificity, (3) amount of collected text (4) coding and (5) documentation.

According to the percentage and distribution of texts, six types of corpus are distinguished:

a) *Large or extensive corpora*. These corpora do not take into account balance or representation.

b) B*alanced corpora*. They contain language varieties in similar percentages.

c) *Pyramidal corpora*. They contain texts by levels, that is, with different degrees of complexity.

d) *Monitor corpora.* They have a constant length, but it is updated with new equivalent material.

e) *Parallel corpora*. Corpora that consist of texts in the source language and their translations in one (or more) language(s) (Corpas: 1996b). If there are two languages involved, it is a parallel bilingual corpus; if there are more than two, multilingual. Its field of application is the automatic translation and contrastive linguistics (Bravo Gozalo, 1998). They have the advantage that any expression in the source text should be in the target text; on the contrary, a parallel corpus is very difficult to compile due to two aspects: (1) it only exists in very specialized areas; (2) the target corpus of such texts are difficult to obtain due to privacy policies of professional translators (Langlois 1997). An example of this type of specialized corpus is the *Canadian Hansards*, a compilation of 100 million words in English and French about the daily sessions of the Federal Parliament of Canada.

f) *Comparable corpora*. They contain similar types of original texts, that is, they have the same design criteria (size, genre, area of specialty, origin, etc.) Like the parallel corpus, this one can be monolingual or bilingual/multilingual. Its main advantage is that researchers can have similar texts in different languages with a similar communicative level (Bravo Gozalo, 1998). They are also very useful for a terminologist and a professional translator for the huge amount of information that can

be extracted. However, they have two drawbacks: (1) they are not as much related as a parallel corpus from a semantic point of view, so they may not always be useful for translators and terminologists; (2) they have problems with automatically utilisation, that is, as they are not alienated texts, they cannot have the same terminology, phraseology, etc. Its main application is linguistic comparison (Corpas Pastor, 2001), which allows to analyse the behaviour and textual forms of two or more languages. Moreover, through this type of corpus syntactic, semantic and collocational patterns can be identified. An example of comparable corpus is the *Multilingual Lexicography Project,* coordinated by John Sinclair (Sinclair, 1991b). This project consists of a comparable corpus in various European languages (English, German, Dutch, Swedish, Italian, Spanish, Hungarian and Serbo-Croatian) in order to make a multilingual dictionary.

If we take into account the specificity of texts there are six types of corpora:

a) *General corpora*. Corpora that represents the common language with everyday communicative situations, such as a transcription of a conversation.

b) *Specialized corpora.* They consist of texts from a linguistic variety.

c) *Corpora from genre*. Corpora of texts of a particular genre, such as a corpus in the legal field.

d) *Canonical corpora*. They are formed by the compilation of texts by the same author.

e) *Chronological corpora*. Corpora that include texts over a period of time.

f) *Diachronic corpora.* Texts of successive periods that analyse the linguistic evolution of a language.

Furthermore, the amount of text to be found in each corpus provides three types:

a) *Textual corpora*. They are a corpus from the common language, as well as its varieties. Its field of application is the development of dictionaries, grammars and reference works.

b) *Referenced corpora*. Fragments of texts used to analyse the state of a language. Another field of application is the development of reference works.

c) *Lexical corpora*. These types of corpus have a lexical interest and it is made of texts with very small fragments of equal length.

The following criterion refers to the coding and classification of texts in a corpus:

a) *Annotated corpora*. Corpora with text format or ASCII format, which presents a high degree of simplicity.

b) *Non-annotated corpora*. They have been tagged and rated in a linguistic or metaphysical form, either manually or automatically.

The last classification criterion refers to the documentation of texts:

a) *Documented corpora*. In this type of corpus, each text file has a DTD (Document Type Definition) or header about its origin.

b) *Undocumented corpora*. They are the opposite of the above type, that is, this corpus does not have a file to classify their origin.

2.4.2 The compilation of our corpus linguistics in the field of Proteomics

Once we have done the analysis of linguistic corpora and their types, we will discuss which one we have chosen for our project in Proteomics, as well as the extent and purpose of it.

Our corpus will be bilingual in English and Spanish, which means that we have to choose the type of corpus that best fits with our needs. Every corpus has the aim of:

> "representing all the relevant varieties of the language, and the characteristic vocabulary, so that it can be used as a basis for reliable grammars, dictionaries, thesauri & other language reference materials." (Sinclair, 1996)

Bravo Gozalo (1998) states that corpora have been used for five different purposes: (1) linguistic investigation, which makes it possible for the empiric study of languages; (2) elaboration of consulted works such as dictionaries and descriptive grammar; (3) processing of natural languages; (4) teaching languages, but it has less impact; and (5) translation, which from the 1980s is named as Translation Studies.

As we are going to compile an English/Spanish glossary, our main purpose is the elaboration of consulted works, defined in the field of Terminology and Translation.

Once we have a clear objective in this work, we will choose the most suitable type of corpus. First, we have to decide whether our corpus would be comparable or parallel. As we have mentioned above, a comparable corpus is used in linguistic

comparison, so lexical and syntactic information can be extracted. On the other hand, a parallel corpus reveals great information due to alienation of texts, but they are very difficult to compile. Furthermore, it is applied to automatic translation and applied linguistics.

In our case, we have chosen a specialized, bilingual comparable corpus, English-Spanish. It is specialized, as it focuses on a particular field, Proteomics. It is comparable because it will be focused on the elaboration of a glossary of terms. Finally it is bilingual as we can extract terms in both languages.

Our bilingual comparable corpus consists of two different languages: English and Spanish. The English corpus has been collected by using research articles from four different journals: *Biochimica et Biophysica Acta (BBA), Genomics, Proteomics & Bioinformatics (GPB), Journal of Proteomics* and *Proteomics*. These articles are addressed to a specific type of reader: an expert in the field. Due to this reason, we also had to choose for our Spanish corpus four different types of texts which were addressed to the same audience. We chose *Revista de la Sociedad Proteómica, a PhD in Proteómica Botton-up y Top Down de Organismo poco Secuenciados, a PhD in Proteómica de Expresión Diferencial en Acinetobacter baumannii Resistente a Colistina, and a PhD in Estudios de la Alteración de Proteínas en el Tejido Adiposos Omental y Subcutáneo en Obesidad Mediante Técnicas de Proteómica.* All the texts have been compiled from the Internet.

All the texts compiled in the corpus belong to the expert-to-expert communicative setting (Pearson, 1998). The journals and dissertations are written by an expert in the field and are exclusively addressed to experts in the field of Proteomics.

In the case of authorship, publishers in the English corpus have English as their mother tongue, whereas in the Spanish corpus the language is Spanish and they are native speakers of this language. Therefore, each corpus is written by speakers of their mother tongue, so this will help us in a better understanding of the syntax and lexicography, as a native speaker expresses his/herself better than one who is not.

Related to language, international English is used by the authors in the English corpus. International English is used by investigators or experts in the field with the aim of spreading those discoveries from other non-English speaking country. Méndez Cendón (2002) in her PhD Dissertation explains the usefulness of international English:

> "el inglés se considera en el campo médico no sólo como lengua internacional sino como lengua intranacional. Es decir, el inglés es una segunda lengua común a médicos e investigadores y, en general, a aquellas personas involucradas, de alguna manera, en disciplinas relacionadas con la medicina."

In Table 1, we can see the source and the date of the texts compiled, as well as the impact factor and the number of tokens in each magazine. Our corpus consists of 100,000 words in English and 100,000 in Spanish. Considering that 50% of the words in a corpus are 'noise'[1] (Bravo Gozalo, 1998), the amount of useful words varies depending on the extension. Therefore and taking into account this fact, we have decided that 100,000 words is the appropriate extension for our corpus.

ENGLISH CORPUS			
Texts	**Year**	**Impact factor**	**Nº Tokens**
Biochimica et Biophysica Acta	2007-2011	4.204	28,904
Genomics, Proteomics & Bioinformatics	2012-2013	9,74	10,514
Proteomics	2012	4.132	30,795
Journal of Proteomics	2010-2011	4.088	31,527
TOTAL	-	-	**101,008**

SPANISH CORPUS			
Texts	**Year**	**Impact factor**	**Nº Tokens**
Revista Proteómica	2009-2012	3.851	58,965
PhD "Proteómica Bottom-up y Top-Down"	2008	-	15,340
		-	14,651
PhD "Alteración de Proteínas"	2011	-	14,840
PhD "Proteómica de Expresión"	2010		
TOTAL		-	**103,796**

Table 1.List of compiled texts in English and Spanish

[1] 'Noise': words which are neither terms or candidate terms, for example: prepositions, adverbs, wh-questions, etc. (Bowker et al. 2002: 123)

To sum up, a bilingual comparable corpus is the most suitable type of corpus for our ultimate goal (a bilingual glossary in Proteomics). This type of corpus consists of different texts in English and Spanish on the same area: Proteomics. Furthermore, considering the amount of information we should use to extract terms, each corpus contains 100,000 words in order to have enough useful words. Once we have the linguistic corpus, we will discuss the importance of terminology in our work.

2.5 Terminology

According to Cabré (1993), "terminology is the discipline whose aim is the study and compilation of specialized terms". Nowadays terminology is very important in the field of communication studies. Furthermore, it is an interrelated discipline with other fields of study, since according to Sager (1990) "it is vital to the functioning of all sciences, it is concerned with designations in all other subject fields, and it is closely related to a number of specific disciplines".

The first use given to the word 'terminology' was as technical vocabulary, such as a list of terms. Today, the term is polysemic since it refers to the theory of terminology, to the work of the terminologist and also to the set of terms of a particular subject field as a result of a previous terminology work.

The International Association of Terminology summarises very clear what terminology is and its interdisciplinary character:

> "Terminology is concerned with the study and use of the systems of symbols and linguistic signs employed for human communication in specialised areas of knowledge and activities. It is primarily a linguistic discipline – linguistics being interpreted here in its widest possible sense – with emphasis on semantics (systems of meanings and concepts) and pragmatics. It is interdisciplinary in the sense that it also borrows concepts and methods from semiotics, epistemology, classification, etc." Pérez Hernández (2002) http://elies.rediris.es/elies18/31.html

2.5.1 The concept

The concepts are the mental representation of the terms whereas the terms are the linguistic representations of the concepts. A concept is "an abstract unit composed of characteristics of objects in the real world" (Pizarro, 2013-2014).

Concept formation is a process of grouping and ordering material and non-material objects which we perceive or imagine in abstract categories. Sager (1990) provides steps for forming concepts: (1) identifying the common characteristics of individual objects, (2) gathering abstract objects in larger classes, (3) define the concepts.

Therefore, concepts, terms and objects are related. Figure 1 shows this relationship.

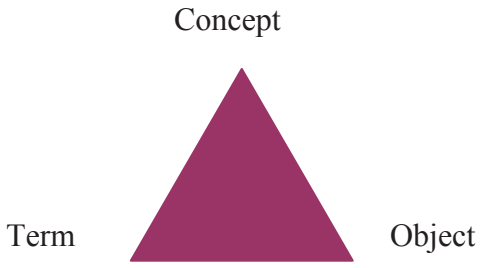


Image 1. Relation concept-term-object (Pizarro, 2013-2014)

2.5.2 Terminology and translation

Translation and terminology are related, as translators and interpreters facilitate communication between specialists (Cabré 1993). Terminology is useful for a translator when transferring content from one language to another. Thus, a translator must have prior knowledge of the subject, as well as knowledge of the target language.

Translators make use of terminology using bilingual or multilingual technical glossaries, thesauri and term banks. However, translators have to play the role of terminologists a lot of times. Sometimes, there are terms that do not appear in glossaries and they have to solve these doubts in so little time.

Cabré (Cabré 1993) explains the function of terminology and translation:

> "La terminología que los traductores necesitan, además de términos equivalentes en otras lenguas, debe contener contextos que den información sobre cómo utilizar lingüísticamente cada unidad y, en una situación óptima, datos sobre el concepto que la denominación expresa, a fin de garantizar que utilizan la forma precisa que corresponde a un determinado contenido."

Cabré states that translators need to know terminology in order to know equivalent terms, as well as information about concepts and contexts.

Now that we know about the relationship between terminology and translation, we are going to analyse what terms are.

2.5.3 Terms

Bowker and Pearson, J. (2002) state that a term is a lexical unit used and understood in a specialized field. Terms are extracted from specialized texts so they cannot be analysed without looking at the specialized discourse. However, what differentiates a term from a word? Pavel (2002) explains the difference between a terminological unit and a word:

> "A term or terminological unit in a specialized language is distinguished from a word in the general language for its univocal relation with the specialized concept that describes (phenomenon called monosemy) and the stability of that relation between form and content in texts which deals with that concept (phenomenon denominated lexicalization). Later, it is its frequency of use and the contextual environment, as well as the typographical indicators (italics, bold, inverted commas, etc.) which states the situation of the term".

That is, words occur in general language and specialized language while terms occur only in specialized language. However, they show a connection between form and content in relation to a specific field.

A term is a specialized lexical unit which is used in a specialized field and designates one concept. Terms constitute the entries of our glossary in Proteomics.

2.5.3.1 Term formation

For the elaboration of new terms, there are linguistic mechanisms established by the grammar of each language (Cabré, 1993). Thus, when a terminologist forms a new terminological unit, he makes it with the basis of a series of regulated processes.

The process of formation of terms is a series of systematic rules depending on the type of resource. From this we can distinguish the following norms:

(1) Derivation prefix rules: [prefix + [base] x] (x) (y)

(2) Derivation suffix rules: [[base] x + suffix] (x) (y)

(3) Composition rules: [[base] x # [base] x]

(4) Conversion rules: [base] x → [base]y

2.5.3.2 Types of terms

Cabré (1993) makes a classification of terms according to four criteria: form, function, meaning and origin.

From the point of view of form, there are four groups distinguished:

a) *Depending on the number of morphemes*, the terms can be simple or complex.

b) *Depending on the type of morpheme* in their formation, they are derived[2] or compound[3].

c) *Depending on the combination of words*, they can have or not terminological phrases.

d) *Depending on the complex formation of simple terms*: we can distinguish:

I. Acronyms: units formed by combining the initials of several words that are longer term. i.e. USA →United States of America

II. Initials: words formed by combining segments of a developed phrase. i.e. bit → binary digit

III. Abbreviations: forms that reproduce the initial segment of a word. i.e. N. → noun

IV. Abbreviated forms: units to reduce the speech, using the first part of a longer word or the first word of a phrase. i.e. bike → bicycle

According to the function of terms, we can distinguish:

a) *Nouns*

b) *Adjectives*

c) *Verbs*

d) *Adverb.*

In terminology, there is a disproportion of different grammatical categories. Thus, there are a large number of nouns compared to verbs or adverbs.

According to the meanings of the words there are:

a) *Objects or entities*: names

b) *Processes, operations and actions*: verbs, nominalization of verbs

c) *Property, states, qualities*: adjectives

---

[2] Derived: formed by the aggregation of lexical affixes to bases.

[3] Compound: formed by the aggregation of lexical affixes to bases. It is a formation of words formed by combinations of lexical, current or historical basis, with the possibility of adding affixes to them later.

d) *Relationships:* adjectives, verbs.

According to the linguistic origin, terms may be:

a) *Built by applying rules of its own linguistic code*

b) *Loans*: There are three types of loans:

I. Cultism: loans from the Greco-Roman history.

II. Loans: those coming from other current historical language

III. Loans from other geographical and social dialects of the same language, but they are not considered loans.

To sum up, in subsection 2.5 we have explained that terminology is a discipline whose aim is the compilation of specialized lexical units (terms). This discipline is interrelated with translation, due mainly to the need of compiling terminographic works. Concepts are abstract units that are designated by means of the terms and they are both interrelated with objects. Also, there are many types of terms and different ways of term formation.

3. METHODOLOGY OF THE EXTRACTION OF KEY TERMS

According to Cabré (1993), before producing a list of terms, it is important to define the work area and have a prepared plan in accordance with what you want to accomplish. We have accomplished this task in the previous section and now we will elaborate the terminology work[4].

The steps in the compilation of terminology are: the extraction of terminology, the elaboration of a card file and of a terminological record.

3.1 Extraction of terminology

The extraction of terminology consists of obtaining candidate terms[5] in a specialized field in which we work. Terminologists must recognize those segments or words from this area and delimit them. The theoretical background is a useful guide for the terminologist or translator, as it is not an expert in the field. For our terminology work we have extracted the terminology in two phases: first, looking for single candidates terms and then looking for complex candidate terms.

We have used Antconc, a software program that allows us to obtain concordances and their frequencies for the elaboration of a list of key terms.

3.1.1 AntConc

AntConc is a corpus analysis tool which works with many operative systems such as Windows, Macintosh, or Linux, and it is frequently used by terminologists, corpus linguists, translation students, etc. It was designed by Laurence Anthony and it can analyse mostly all the languages including Asian languages). This program has seven different parts for analysing our corpus: *Concordance, Concordance Plot, File View, Clusters, Collocates, Word List, and Keyword list*. For our extraction of key terms, we will only use *Concordance, Cluster* and *Word List.*

---

[4] Terminology work: "work concerned with the systematic collection, description, processing and presentation of concepts and their designations." (Pizarro, 2013-14)

[5] A candidate term is a term extracted from a corpus that can be or not an entry of a specialized dictionary.

AntConc works with .txt format files, so that all texts must be in this form. For our project, we have developed eight txt files, one for each journal and / or PhD Dissertation.:

a) English corpus

I. *Biochimica et Biophysica Acta (BBA)*

II. *Genomics, Proteomics & Bioinformatics (GPB)*

III. *Journal of Proteomics*

IV. *Proteomics*

b) Spanish corpus

I. *Revista Proteómica*

II. *PhD Dissertation in Proteómica Bottom-up y Top Down de Organismo poco Secuenciados*

III. PhD Dissertation in *Proteómica de Expresión Diferencial en Acinetobacter baumannii Resistente a Colistin*

IV. PhD Dissertation in *Estudios de la Alteración de Proteínas en el Tejido Adiposos Omental y Subcutáneo en Obesidad Mediante Técnicas de Proteómica.*

Once we have the .txt documents, AntConc needs a document to remove the 'noise'[6] of the corpus. Thus, we have used a stop list, that is, a group of useless words such as prepositions, articles, pronouns, etc. that make 'noise'. Image 2 shows our stop list in English.

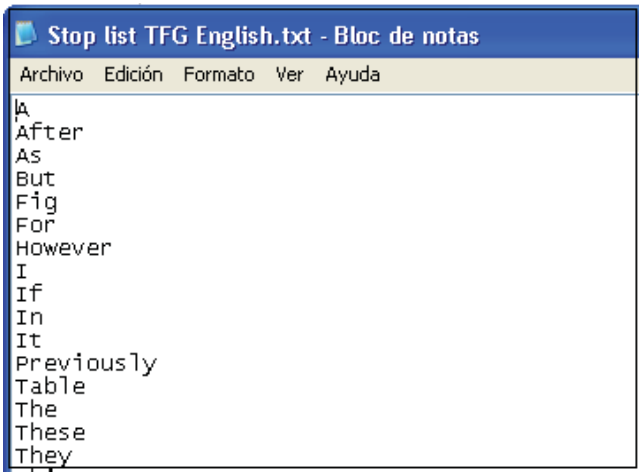


Image 2. English stop list

[6] The noise deals with the unnecessary words for the extraction of key terms.

After the stop list was generated, a lemma list was required. A lemma list is a set of words which are part of the same lemma. As an example, the term *protein* in sense is the same as *proteins* or *Protein*, but both belong to the same lemma, which is *protein*. In the same way, *eat, eats, eating, ate*, etc. are word-forms belonging to the lemma *eat*. Image 3 shows the English lemma list.

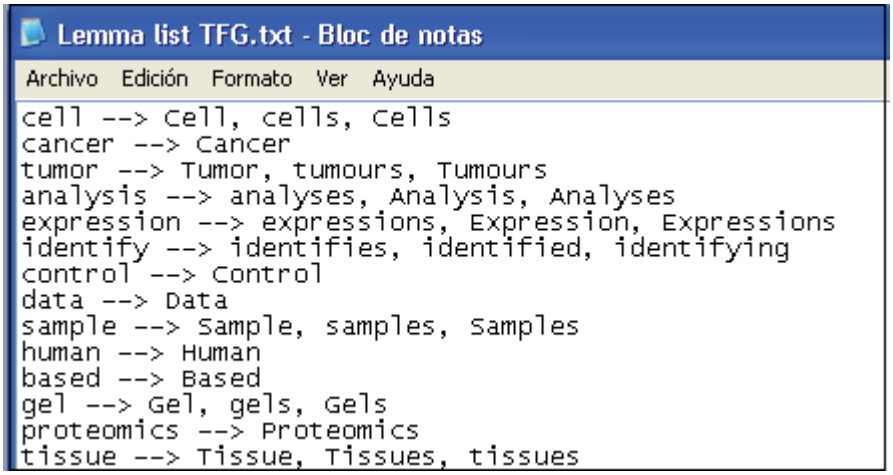
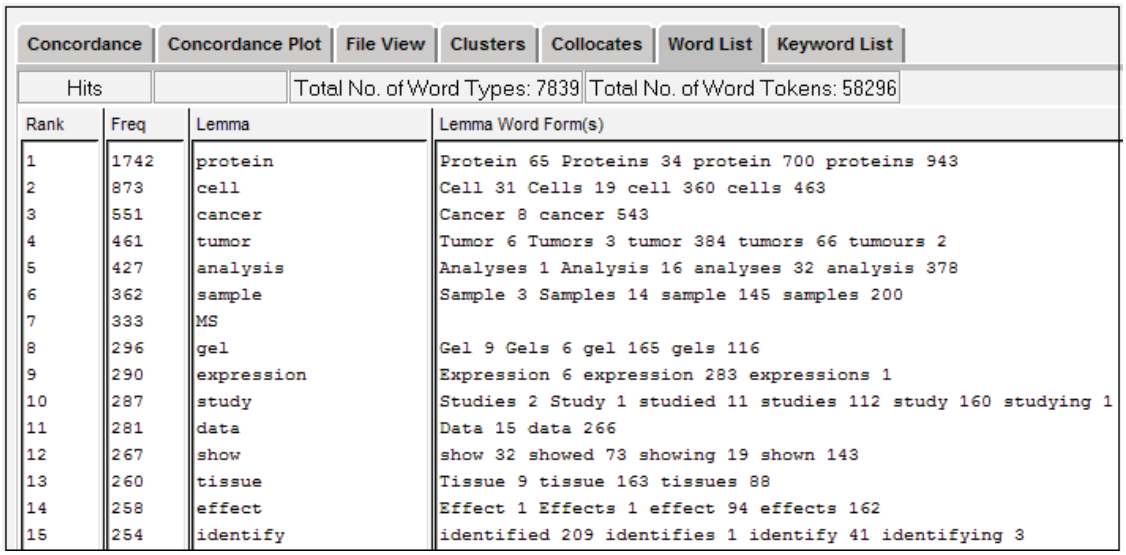


Image 3. English lemma list

Once we have the lemma list and the stop list, we can start by looking at the results that occur in the frequency list. The frequency list is a section of Antconc which comprises the number of repetitions of a word. These words are called tokens, which are the number of sequences of letters separated by spaces or punctuation. The criterion to choose our candidate terms is to see the tokens which are the most frequent, and then, the context of each token. As we have read about the topic before, we know which words can be the most important. Image 4 reveals the results that we have obtained from the English corpus about Proteomics.



Concordance | Concordance Plot | File View | Clusters | Collocates | Word List | Keyword List

Hits | Total No. of Word Types: 7839 | Total No. of Word Tokens: 58296

Rank	Freq	Lemma	Lemma Word Form(s)
1	1742	protein	Protein 65 Proteins 34 protein 700 proteins 943
2	873	cell	Cell 31 Cells 19 cell 360 cells 463
3	551	cancer	Cancer 8 cancer 543
4	461	tumor	Tumor 6 Tumors 3 tumor 384 tumors 66 tumours 2
5	427	analysis	Analyses 1 Analysis 16 analyses 32 analysis 378
6	362	sample	Sample 3 Samples 14 sample 145 samples 200
7	333	MS	
8	296	gel	Gel 9 Gels 6 gel 165 gels 116
9	290	expression	Expression 6 expression 283 expressions 1
10	287	study	Studies 2 Study 1 studied 11 studies 112 study 160 studying 1
11	281	data	Data 15 data 266
12	267	show	show 32 showed 73 showing 19 shown 143
13	260	tissue	Tissue 9 tissue 163 tissues 88
14	258	effect	Effect 1 Effects 1 effect 94 effects 162
15	254	identify	identified 209 identifies 1 identify 41 identifying 3

Image 4. English corpus results from frequency list

For the Spanish corpus, we have followed the same steps, that is, the elaboration of a stop list and of a lemma list.

3.1.1.1 Simple candidate terms

After extracting the frequency list, we have designed a table with all the results. Table 2 shows the data obtained in the English and Spanish corpus of the simple candidate terms, that is, possible terminological units made up of a word.

	ENGLISH CORPUS		SPANISH CORPUS	
ORDER	**TERMS**	**FREQUENCY**	**TERMS**	**FREQUENCY**
1	protein	1742	proteína	1223
2	MS	333	MS	286
3	proteomics	176	proteómica	299
4	biomarkers	147	proteoma	127
5	proteome	142	2-DE	101
6	HSP	141	biomarcadores	86
7	MS/MS	89	electrophoresis	83
8	antibody	79	aminoácido	75
9	mRNA	51	MS/MS	61

10	MASCOT	41	cromatografía	56
11	ubiquitin	37	MALDI-TOF	39
12	phosphorylation	34	anticuerpos	33
13	PAGE	32	LC-MS	33
14	IPG	27	SDS-PAGE	29
15	ubiquitination	24	fosforilación	28
16	ABHD11	23	Orbitrap	20
17	electrophoresis	22	HSP	24
18	aminoacid	21	IPG	17
19	IEF	18	MASCOT	15
20	PTM	18	IEF	14
21	ELISA	13	ELISA	7
22	Orbitrap	12	PMF	7
23	chromatography	9	ubiquitinación	6
24	PMF	8	ubiquitina	5
25			ARNm	3

Table 2. Simple candidate terms and their frequency.

There are a total of twenty-four English simple candidate terms and twenty-five in Spanish. The difference in frequency between the same candidate terms (the English term and its equivalent in Spanish) is due to the difficulty to find the same number of words for the same terminological unit in a comparable corpus, as I explained in the second section of this dissertation.

3.1.1.2 Complex candidate terms

From these twenty-four words in English and twenty-five in Spanish, we will elaborate a list of compound candidate terms. AntConc has an option called *Cluster*, which we use to extract compound candidate terms. Thus, a simple terminological unit is introduced in the program and, then, you specify the number of words which you want to occur in the results (i.e. min.: two → max.: five).

Table 3 shows the results of complex candidate terms and their frequency in English and Spanish respectively.

ENGLISH CORPUS			SPANISH CORPUS	
Nº	**Candidate terms**	**Fr.**	**Candidate terms**	**Fr.**
1	LC-MS	49	2-DE	46
2	protein abundance	43	MALDI-TOF	39
3	protein expression	42	proteínas identificadas	39
4	identified protein	35	LC-MS	33
5	protein spots	33	identificación de proteínas	32
6	gene expression	33	2D-DIGE	28
7	protein spots	33	expresión proteica	25
8	MALDI-TOF	32	proteómica cuantitativa	20
9	Western Blot	31	expresión génica	19
10	protein identification	27	proteínas implicadas	19
11	differentially expressed proteins	24	Western Blot	19
12	binding protein	23	proteínas secretadas	12
13	2D DIGE	22	modificaciones postraduccionales	8
14	Heat Shock Protein (HSP)	18	proteínas nitradas	7
15	protein interaction	18	trampa iónica lineal	5
16	activity-based proteomics	17	síntesis de proteínas	4
17	comparative proteomics	17	proteómica clínica	3
18	protein database	17	proteínas carboniladas	2
19	regulated protein	17	spots de proteína	2
20	associated protein	15	proteómica comparativa	1
21	protein interaction	14	kit BCA protein assay	1
22	quantitative proteomics	13		
23	protein synthesis	12		
24	BCA protein assay kit	5		
25	linear ion trap	5		

Table 3. Complex candidate terms and their frequency

There are twenty-five English complex candidate terms and twenty-one in Spanish. The variation in the results of the two languages is because journals usually have more variety of topics than PhDs, since the latter focus on a very specific aspect of the field of Proteomics.

3.1.2 Key terms

The second part of the extraction of terminology consists of a table with the terminological units extracted from the corpus. Not all the terms that occur in the corpus must be necessarily terminological units from a specific subject (Cabré, 1993). As this is a specialized area, it will not be easy for the terminologist to figure out what candidate terms are useful terminological units for the glossary. In order to do this, one must make a distinction between collocation and term. A term or terminological unit is a lexical unit, simple or compound that forms part of a specialized language (De Santiago, 2013-2014), while a collocation is a combination of lexical items that regularly co-occur in language (Plum, 2006).

Cabré (1993) establishes various criteria to analyse which words are terms and which are not:

- A term is a set logically organized around a single base.
- A term cannot insert other linguistic elements within the terminological phrase.
- A term cannot separate any parts of its whole unit.
- A term can have an antonym in the same specialty.
- A term has similar frequency of occurrence texts belonging to the same register in a particular subject field.

These characteristics distinguish a terminological unit from a collocation. Image 5 (De Santiago, 2013-2014), shows a diagram of the relationship between terminological units and collocations.

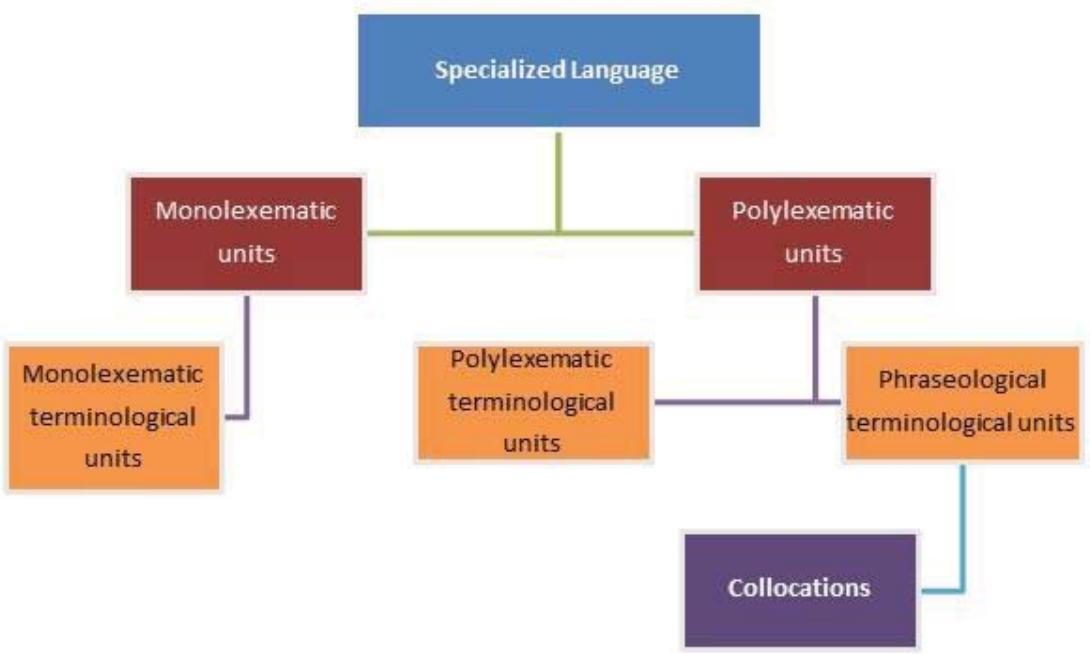


Image 5. Types of specialized vocabulary (De Santiago, 2013-2014)

Collocations are phraseological units, whereas compound terms are polylexematic terminological units made up of one base and one or more pre-modifiers (in the case of English) and post-modifiers (in the case of Spanish).

Sometimes, there is a discrepancy between the frequencies of the candidate terms. Thus, KWIC concordance lines are another criterion to identify key terms. A concordance line is fragment of a text in a corpus that helps you to specify the context of the word that occurs a certain number of times in a search pattern (Bowker et al. 2002). Antconc has an option in which you can see the concordance lines of a term. In image 4 we can see an example of a KWIC concordance line. As a result, key terms will be easier to recognize in the corpus.

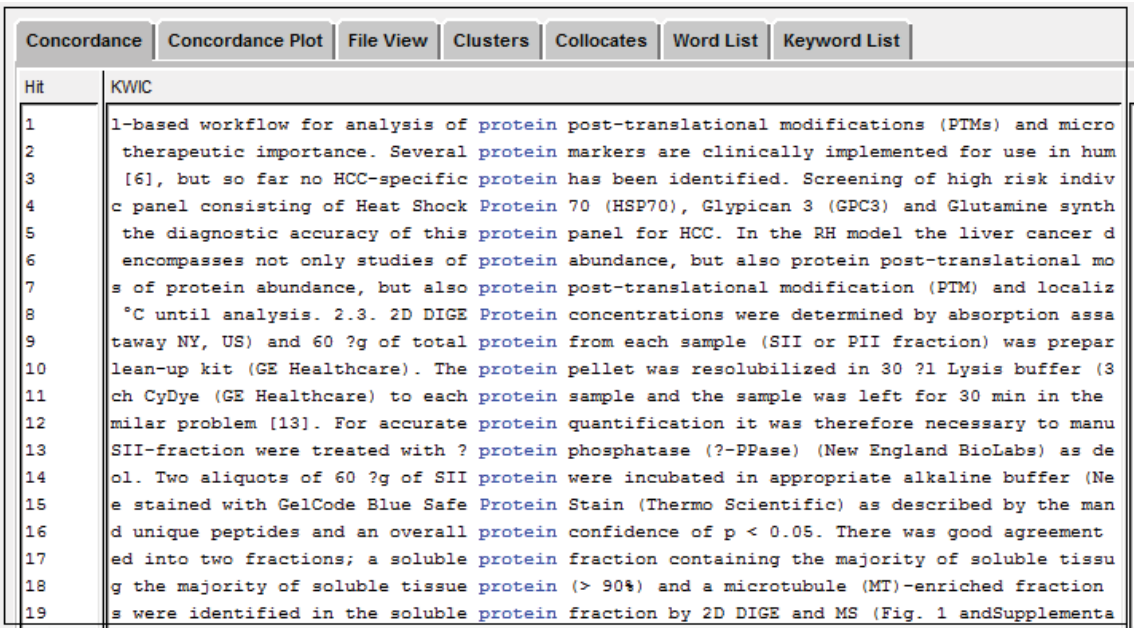


Image 6. Concordance line in the English corpus

With all of these criteria taken into account, we identified the terms that would be included in our glossary. Table 4 shows this final list of terms, simple and complex, in English and Spanish and in alphabetical order.

Nº	ENGLISH	SPANISH
1	2D DIGE	2D-DIGE
2	2-DE	2-DE
3	amino acid	aminoácido
4	antibody	anticuerpos
5	BCA protein assay kit	kit BCA protein assay
6	biomarkers	biomarcadores
7	chromatography	cromatografía
8	comparative proteomics	proteómica comparative
9	electrophoresis	electroforesis
10	ELISA	ELISA
11	gene expression	expresión génica
12	HSP (Heat Shock Protein)	HSP (Proteína del Choque Térmico)
13	IEF (IsoElectric Focusing)	IEF (Enfoque IsoEléctrico)

14	IPG (Immobilized pH Gradient)	IPG (Gradientes de pH Inmovilizados)
15	LC-MS	LC-MS
16	linear ion trap	trampa iónica lineal
17	MALDI-TOF	MALDI-TOF
18	MASCOT	MASCOT
19	mRNA	ARNm
20	MS (Mass Spectrometry)	MS (Espectrometría de masas)
21	MS/MS	MS/MS (espectrometría de masas en tándem)
22	Orbitrap	Orbitrap
23	phosphorylation	fosforilación
24	PMF (Peptide Mass Fingerprinting)	PMF (huellas de masas peptídicas)
25	protein	proteína
26	protein expression	expresión proteica
27	protein identification	identificación de proteínas
28	protein spots	spots de proteína
29	protein synthesis	síntesis de proteínas
30	proteome	proteoma
31	proteomics	Proteómica
32	PTM (Post-translational Modifications)	modificaciones postraduccionales (PTM)
33	quantitative proteomics	proteómica cuantitativa
34	SDS-PAGE	SDS-PAGE
35	ubiquitination	ubiquitinación
36	ubiquitin	ubiquitina
37	Western Blot	Western Blot

Table 4. Key terms and their translation

This table shows a total of thirty-seven terms extracted from our comparable English/Spanish corpus that will be the entries of the bilingual glossary in Proteomics.

3.2 Card file

Following the steps of Cabré (1993), we have developed a card file with the extracted terms. This file consists of the English and Spanish key terms ordered alphabetically.

Cabré states that a card file should have (1) the terminological unit, (2) grammatical category, (3) field, (4) context, (5) reference context, (6) author of the card file, and (7) date of the file. In our case, contexts are important because they give additional information (pragmatic information) to our future definitions and this type of information is very useful for terminologists and specialized translators. Therefore, it is very helpful for the translator to have examples of the contexts of each terminological unit contained in the glossary.

Table 7 shows a model of a card file that will be filled up with the terminological information extracted from the corpus.

LANGUAGE	
1. TERMINOLOGICAL ENTRY	
Gram. Category	Field
Context	
Reference	Author and date

Table 5. Example of card file. (Santamaría, 2013)

Image 7 shows an example of our card file in English and Spanish. We have followed Santamaría's (Santamaría, 2013) example of card files and have included different cards together with our corpus stored in a cd attached to this undergraduate dissertation.

ENGLISH		SPANISH	
11. gene expression		**expression génica**	
noun	Proteomics	nombre	Proteómica
The microarray results indicated that, at safe levels, arbutin can induce a change in gene expression.		*Una de estas estrategias consiste en una amplia regulación en cada uno de los pasos de la expresión génica.*	
Biochimica et Biophysica Acta	T. Modrego, 5th July, 2014	Revista Proteómica	T. Modrego, 5th July, 2014

Image 7. English-Spanish card file

3.3 Terminological record

According to Cabré (1993), a terminological record is "una pauta estructurada que permite consignar de forma ordenada las informaciones sobre cada término que serán necesarias para llevar a cabo el trabajo". These files are made from the information extracted in the card file. The main purpose of these files is to collect additional information about the terminology units.

The terminology files can be manual or automatic. Automatic files are usually represented in databases. A terminological record consists of:

a) Term entry
b) Grammatical category
c) Subject field
d) Synonym(s)
e) Definition
f) Context
    a. Example
g) Equivalent term in other languages
h) References of all the information compiled in the term record
i) Author name and date of elaboration of the term record

These terminological records vary depending on the needs of the terminologist. In our case, we have not made a terminology record, as the card file comprises enough information to complete the definitions compiled in the glossary in Proteomics.

So this subsection has dealt with the methodology for the extraction of key terms. Through this extraction of terminology, we have obtained English and Spanish complex and simple candidate terms in the subject field of Proteomics. Once we have done this, we have completed a card file with each terminological unit, whose information will help us to elaborate our glossary.

4. METHODOLOGY OF THE ELABORATION OF THE GLOSSARY

Following the methodology by Cabré (1993), once we have compiled the terminology, we put together the work from language A and language B, as well as its supervision by an expert. Image 8 shows Cabre's (1993) scheme about the bilingual systematic work.

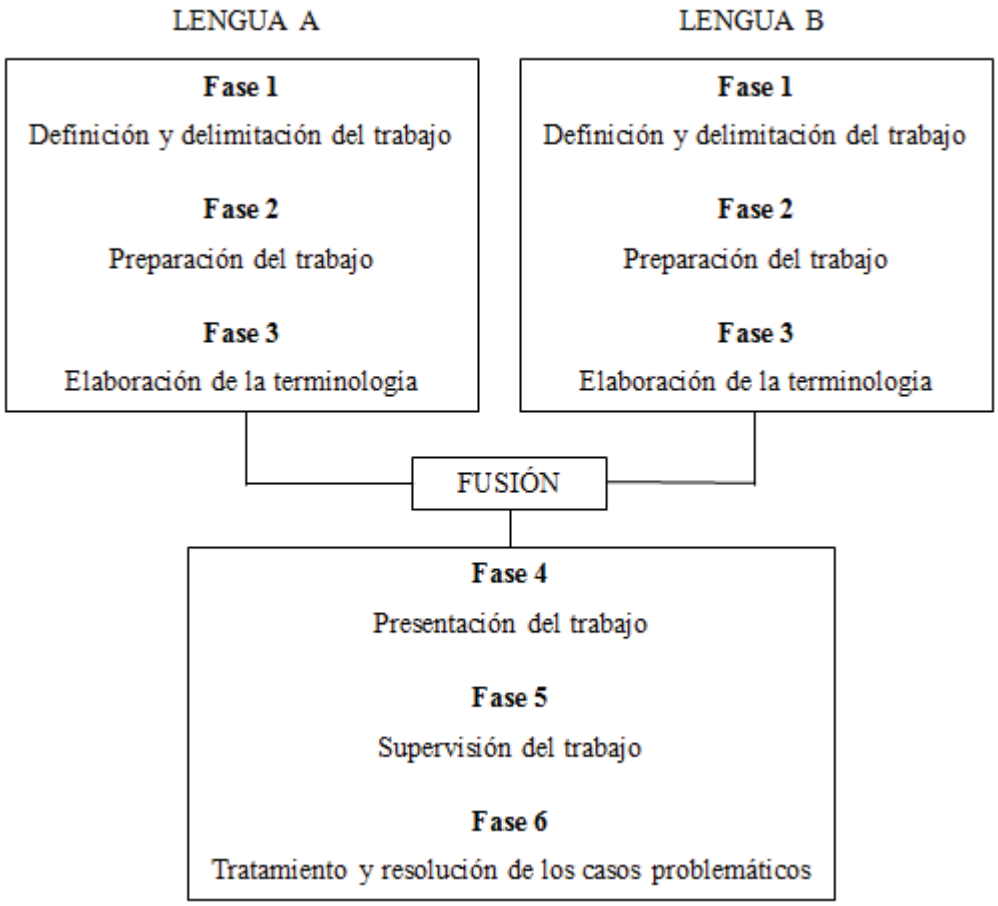


Image 8. Bilingual process of a project (Cabré, 1993)

Therefore, our aim is to compile a bilingual glossary in English and Spanish with the terms extracted from the comparable corpus. To do this, we will follow Frankerverg-García et al.'s (2001) methodology.

4.1 Glossary

According to Frankerverg-García, A. et al, eds. (2011) a glossary is a list of terms together with their definitions in one or more languages. Traditionally, they were written by hand, but with the development of technology, terminologists use corpus handling tools to speed up the glossary production process. The main advantages of producing a glossary are four: identifying equivalents in another language, acquiring terminology, knowing how to use terminology and save time looking for terms.

Frankerverg-García et al, eds. (2011) states that there are ten steps to compile a glossary:

1. Decide the domain and language
2. Locate a subject expert
3. Identify a subfield
4. Start to collect documentation
5. Extract candidate terms
6. Decide if the candidate terms belong to the subfield or not
7. Expand the list of candidate terms
8. Formulate definitions
9. Compile a new corpus in another language
10. Store the data

Example 2 shows how we have followed these steps in our glossary.

1. Domain and language: Pharmacy and English/Spanish
2. Subject expert: we have consulted for advice to Enrique Garcia Gómez
3. Subfield: Proteomics
4. Collect documentation: English and Spanish corpora
5. Candidate terms: we have done a list of simple candidate terms.
6. Decide if the candidate terms belong to the subfield: list of simple key terms
7. Expand the list: list of compound key terms
8. Formulate definitions: searched in specialized websites.
9. Compile a new corpus in another language: we have compiled a Spanish corpus
10. Store the data: each term has a definition and the grammatical category

Example 1. Example of the methodology of a glossary

Once we have analysed the process to compile a glossary, we will study how the definitions are elaborated and their characteristics.

4.1.1 Definitions

According to the Recommendations of the Swiss Chancellery (2002), a definition is a description of a concept according to its intension or extension which makes possible to differentiate this concept from other concepts in a conceptual tree/scheme of the subject field. A good definition is essential for a reliable terminological work, in our case, the glossary.

The terminological definition always consists of a single sentence and it should start with the generic term of the term to be defined (Cabré, 1993). Arntz et Pitch (1995) provide two types of definitions: (1) extensional definition and (2) intensional definition.

a) *Extensional definition*: it indicates the extension of a concept. The steps to elaborate this type of definition are three: (1) enumeration of all subordinate concepts in the same classification, (2) enumeration of all individual objects (e.g. The planets of the Solar System are: Mercury, Venus, Earth, Mars, Jupiter, Saturn, Neptune and Uranus) and (3) a statement of the rule that leads to the enumeration (e.g. "Prime numbers are characterized by being divisible only by themselves and by the unit".).

b) *Intensional definition*: it starts with the generic (superordinate) term of the term to be defined and then it should include its main (intrinsic) characteristics. That is, an intensional definition structure is: Concept = superordinate concept + distinguishing characteristics (e.g. Incandescent lamp: object (superordinate concept) that emits light in which an electric current heats a solid material to a temperature (intrinsic distinguishing characteristics).

We follow the intensional definitions in the glossary of Proteomics, since they are the most appropriate terminological definitions for a terminographic work.

The Swiss Chancellery (Recommendations of the Swiss Chancellery: 2002:27) establishes a set of criteria for a good development of a definition:

(1) conciseness,

(2) reference to a syntactic system,

(3) reference to the subject field,

(4) use of terms which have already been defined,

(5) indication of the field application of a definition,

(6) avoidance of circular definitions

(7) avoidance of negative definitions

Example 2 shows the application of these methods to one of our entries in the glossary.

***2D DIGE (two-dimensional Differential Gel Electrophoresis)*:** A modified form of 2D electrophoresis (2DE) that allows one to compare two or three protein samples simultaneously on the same gel.

Reference: http://www.ncbi.nlm.nih.gov/pubmed/17172676

Date of access: 25 May 2014

1. Conciseness: this definition is brief and still contains all the distinguishing characteristics.
2. Reference to a system: it delimits the concept itself (*"method based on two-dimensional electrophoresis"*)
3. Reference to the subject field: all the definitions follow the same objective criterion and the same subject field: Proteomics.
4. Use of terms which have already been defined: in this example the term *"two-dimensional electrophoresis"* is also defined in the glossary.
5. Indication of the field of application of a definition: it is very important to know the specific meaning of a term. In our case *"quantitative protein"*.
6. Avoidance of circular definitions: this definition is not circular as it is not defined by itself.
7. Avoidance of negative definitions: the definition does not start with negative characteristics.

Example 2. Requirements of a good definition

In addition to the methods described by the Swiss Chancellery, Cabré (1993, 210-213 and 312-313) establishes basic criteria to define properly in the field of terminology. A definition must:

a) Distinguish the concept being defined from others in the same field.

b) Use initial descriptors that have the same category as the term being described, but in a relation of semantic inclusion with the latter. For example.

Circulation (specific term): movement (generic term) of a fluid in an organism that is generally carried out through ducts."

c) Avoid circularity: For example. 'Shock': shock effect.

d) Avoid the use of a concept defined by negation: For example: 'Unequal': not equal.

e) Avoid the use of paraphrases that do not provide any information that could become detached from the term itself: For example: 'Jefferson Fracture': fracture discovered by the English doctor G. Jefferson.

f) Avoidance of metalinguistic formulas: For example: 'To devitalize': verb that describes the act of depriving life activity.

g) Being a complete linguistic expression and not an unfinished piece of content. Therefore, it should consist of a single sentence.

h) Being expressed taking into account the addressee.

Our definitions are taken from different web pages: from specialized articles about Proteomics to specialized websites on this field. Besides, they follow the criteria mentioned above by Cabré (1993) and The Swiss Chancellery.

All the definitions are in English and we have translated them into Spanish. In order to translate the English definitions, we have chosen a software program called Wordfast Pro that will be explained in the next section of this dissertation.

4.1.1.1 Wordfast Pro

Wordfast Pro[7] is: a standalone, multi-platform TM tool designed to address the needs of translators, language service providers and multinational corporations. Wordfast Pro offers a complete translation environment that stores your translated content for reuse on future projects, thereby maximizing efficiency, increasing overall consistency and reducing translation costs.

Wordfast Pro stores all the words you translate and, then, it retrieves them automatically. The main advantage of Wordfast Pro is that you do not have to translate

[7] To know more about Wordfast Pro see: http://www.wordfast.com/docs/WF_ProProductbrief.pdf and https://www.youtube.com/watch?v=RQ-PDYwE0ZE (accessed date: 17/07/2014)

the same segment twice and it guarantees consistency in translations. Furthermore, it provides the option to alternate document translating views, so that you can obtain a faster translation.

In order to translate the terminological entries of our glossary, we have created a project in Wordfast Pro, which will store the target text, that is, the Spanish entries of the specialized dictionary. Image 9 shows the creation of a new project in Wordfast Pro.

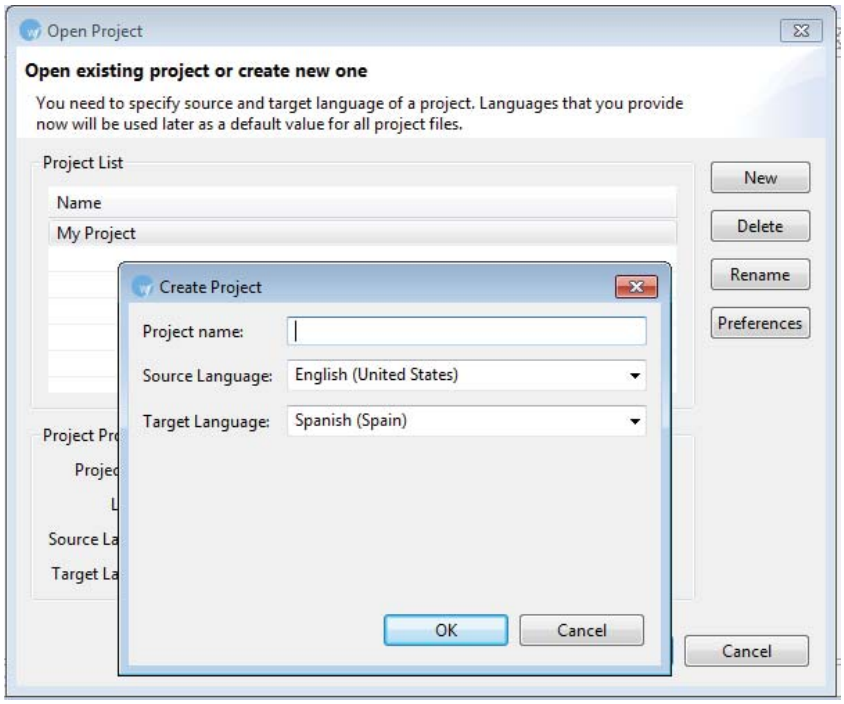


Image10. Wordfast Pro. Fragmented document

Once we have created our new project, we can insert the text to be translated, in our case, the glossary in English. This text appears in segments, which helps with the translation. Also, punctuation (e.g. *{ut1}*) is arranged so that when you click on "*copy tags in current fragment*" the format is copied in the target text. This is advantageous because the translator does not have to worry about changing the format of the translated text after translating the document. Image 10 shows how the document is divided in fragments.

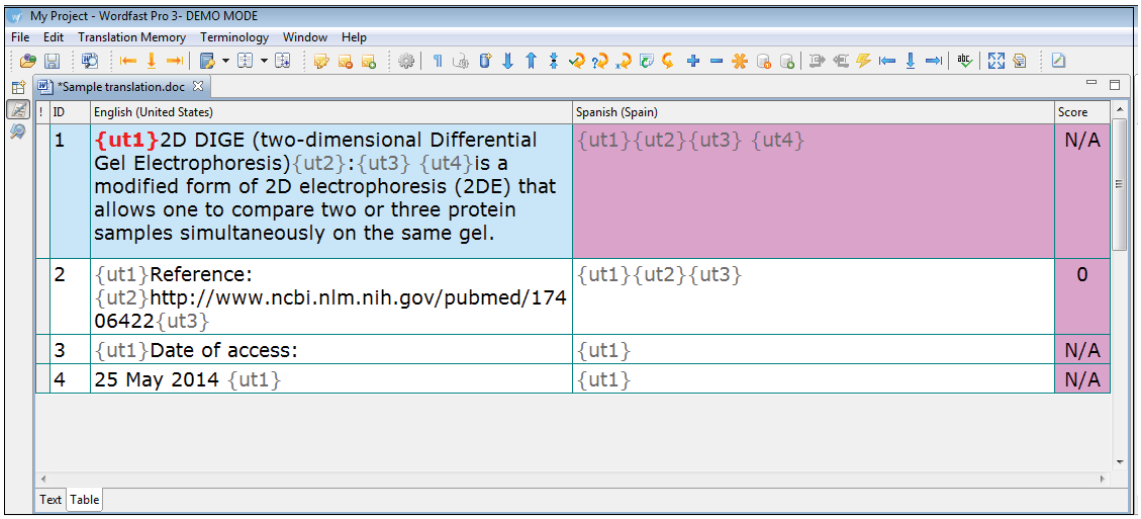


Image 10. Wordfast. Fragmented document

Once we have done the translation, Wordfast Pro generates a Word document (.doc) with the target text. Image 11 shows how both the source definition and the target definition are shown with Wordfast Pro.

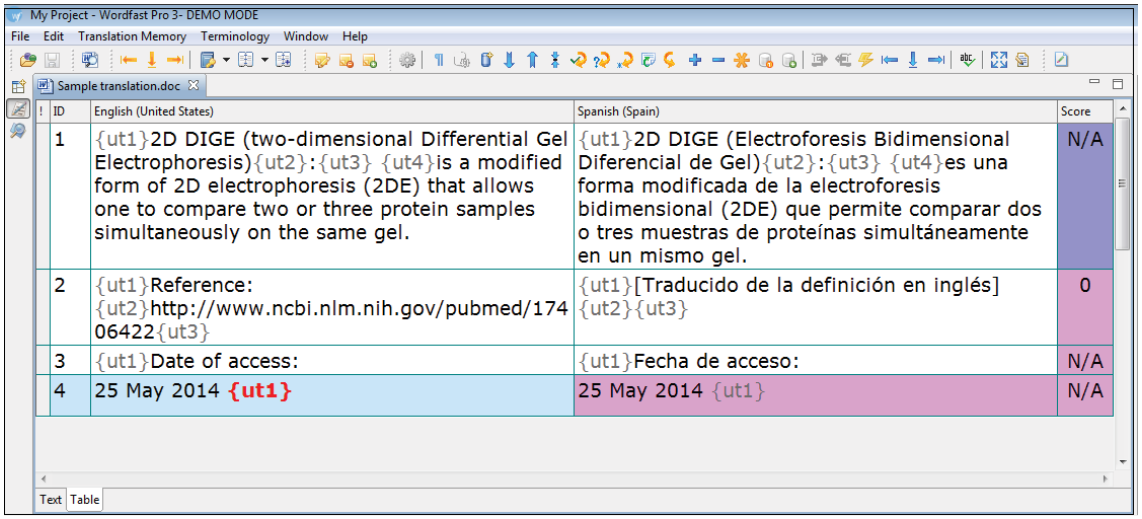


Image 11. Wordfast Pro. Translation of the document

The Wordfast Pro software program has helped us to translate faster, dividing the text into fragments and facilitating the target format.

4.2 Supervision of an expert and review

The steps of the terminology work conclude with the supervision of the project by an expert (Cabré, 1993). This review is important because both the content and the form of the work are verified. Pavel (Pavel, 2002) explains the importance of revising in Terminology:

> "La revisión […] no es una mera verificación técnica ni una corrección de pruebas, sino que consiste en examinar tanto la forma como el contenido […] y comprobar la exactitud de la correspondencia, la presencia de una equivalencia textual en las pruebas textuales y la precisión de las marcas de uso, los valores asignados, los campos temáticos y las fuentes."

The expert in the field is very useful since he/she discusses with the terminologist the relevance of the terms chosen for the terminology project, as well as the form and content of scientific definitions. In our glossary of Proteomics, our expert Enrique García Gómez, pharmacist and chemist, has supervised the translated definitions, as well as the definitions we have chosen from specialized websites.

In conclusion, we have followed the methodology by Cabré (1993) for the terminology work in which we have put together the work done before in language A and B. We have followed the indications of Frankerverg-García et al, eds. (2011) for the compilation of the glossary. In addition, we have followed Cabré (1993) and the Swiss Chancellery's guidelines to elaborate good terminological definitions and how to reject incorrect terminological definitions.

5. RESULTS

The result of this project is the "English/Spanish glossary in Proteomics". It contains thirty-seven entries together with their definitions. Regarding the format of the glossary, we have followed Navarro's (2005) format. He is the author of *Diccionario Crítico de Dudas Inglés-Español de Medicina*.

The entries in this specialized dictionary are in bold, followed by their grammatical category and their definition. Moreover, our glossary shows the reference of the definitions and the date of access to these sources since a lot of websites become obsolete in a short period of time.

Image 12 shows one of the entries of the glossary. We have included the whole glossary together with our corpus and with the card files in a CD attached to this undergraduate dissertation.

N	ENGLISH	SPANISH
3.	**amino acid**: [N.] A building block of proteins in which is coded by a codon and linked together through peptide bonds. Reference: http://www.biology-online.org/dictionary/Amino_acid Date of access: 25 May 2014	**aminoácido**: [N.] Bloque de proteínas codificadas por un codón y unidas entre sí mediante enlaces peptídicos. [Translated from the English definition]

Image 12. Terminological entry

6. CONCLUSIONS

The aim of this undergraduate dissertation was to compile a bilingual terminological glossary addressed to experts in Proteomics (the study of proteins). In the theoretical framework of this dissertation we have discussed key concepts for the elaboration of the glossary, such as the concepts of register, text type and genre. We have also described the field of corpus linguistics and have analysed what is meant by a concept and a term. In the methodology sections, we have followed the steps by Cabré (1993) for the extraction of key terms and the steps by Frankerverg-García et al. (2001) for the compilation of a glossary. We have always born in mind that the glossary is an expert-to-expert terminographic work.

With this project we can conclude that:

(1) The elaboration of a bilingual glossary requires background knowledge of corpus linguistics, terminology and translation. Before compiling a text we should differentiate that register is a contextual category with recurrent situational features and that a text type is a framework to classify texts in terms of communicative situations. Furthermore, we should know that there are several types of corpus, such as comparable, parallel corpora, general, etc. (in our case we have compiled a comparable corpus) in order to select the one that best fits with the purpose of our glossary. In addition, terminology, which is the discipline that studies and compiles specialized terms, is closely related to translation, as the final products of terminology are term banks, thesauri, technical glossaries, nomenclatures, etc. which are very useful for translators. Furthermore, specialized translators also follow the punctual investigation in terminology for their translators.

(2) Terminology work is essential to carry out this glossary, as it is necessary for the extraction of key terms: Cabré (1993) does a perfect analysis of terminology work, establishing the most important steps (extracting terminology, elaboration of card files and terminological records) for a good terminological systematic work. All of them have been essential for the development of the work, as they have given us useful information for the compilation of the glossary. Besides, the extraction of terminology, such as the card files, gives us pragmatic information that can be beneficial for translators in the future.

(3) In order to create any glossary, it is a major aspect to accomplish some terminology criteria for compiling the entries and definitions of the specialized

dictionary: a good glossary should have good definitions, as well as terminological entries adjusted to the needs of the addressee.

As a result, we have developed an English/Spanish glossary with thirty-seven terminological entries and their definitions. This small glossary in Proteomics will help advantageous for future terminological publications in this field. Thus, we can conclude that this bilingual glossary will open the door to subsequent bilingual glossaries in Proteomics, in order to facilitate the spread of research papers and publications in this field.

7. REFERENCE LIST

Alvar Ezquerra, M., Corpas Pastor, G. "Criterios de Diseño para la Creación de Corpus." *J.AM. Alvar Ezquerra y J.A. Villena Ponsoda (coords.)*1994. *Estudios para un Corpus del Español* Universidad de Málaga, Málaga. Pp. 31-40. Print

Anthony, L. *AntConc (Version 3.4.1)* [Computer Software]. Tokyo, Japan: Waseda University. 2011. http://www.antlab.sci.waseda.ac.jp/index.html (20 May 2014)

Arntz, R. et Pitcher, H. *Introducción a la Terminología*. Salamanca: Fundación Germán Sánchez Ruipérez. 1995. Print

Baker, M. "Corpora in Translation Studies: An Overview and Suggestions for Future Research". *Target* 7(2). 1993 Pp. 223-243. Print

BBA Proteins and Proteomics. Eds. Lee, I. and Helmut, H. 2002-2014. http://www.sciencedirect.com/science/journal/15709639 (20 May 2014)

Bowker, L. and Pearson, J. *Working with Specialized Language: a Practical Guide to Using Corpora*. Routledge. 2002. Print.

Bravo Gozalo, J.M et al. *La Traducción: Orientaciones Lingüísticas y Culturales*. Valladolid: Universidad de Valladolid. 1998. Print

Cabré, M. T. *La Terminología: Teoría, Metodología, Aplicaciones*. Barcelona: Antártida [etc.], 1993. Print.

Carrera Mouriño, M. *Proteómica Bottom-Up y Top-down de Organismo Poco Secuenciados. Secuenciación de Novo de Péptidos Biomarcadores para la Identifiación de Especies de la Familia* Merlucciidae. Universidad de Vigo: Departamento de Bioquímica, Genética e Inmunología. Vigo. 2008. Print.

Corpas Pastor, G. *Manual de Fraseología Española*. (Biblioteca Románica Hispánica. Manuales, 76). Madrid: Gredos. 1996. Print.

Corpas Pastor, G. "Compilación de un Corpus *ad hoc* para la Enseñanza de la Traducción Inversa Especializada". *Trans, nº5,* pp. *155-184.* 2001. Print

De Santiago, P. Class notes of the subject "Introducción a la Traducción Automática". 2013.2014. Valladolid: Universidad de Valladolid. Print

Diccionario de la Real Academia Española. Madrid: Grupo Planeta. 22nd edition. 2001. Print

EAGLES. "Preliminary Recommendations on Corpus Typology" Document Eagles (*Expert Advisory Group on Language Engineering)*. 1996. Print

Federal Swiss Chancellery. Terminology Section. *Recomendaciones de la Cancillería Suiza.* Berne: MediaCenter of the Conferation. 2nd edition. 2002. Print

Firth, J.R. *Papers in Linguistics: 1934-1951*. Oxford: Oxford University Press. 1951. Print.

Fortanet, I. et al. *Genre Studies for Academic Purposes*. *Filología/9*. Castelló de la Plana: Publicacions Universitat. 1998. Print

Jaume I. Frankerberg-García, A. et al, eds. *New Trends in Corpora and Language Learning.* Bloomsbury: Bloomsbury Academic. 2011. Print

Johansson, S. "On the Role of Corpora in Cross-linguistic Research". Johansson, S. /Oksefjell, S. (eds) *Corpora and Cross-linguistic Research: Theory, Method, and Case Studies*. Amsterdam/Atlanta: Rodopi. 3-24. 1998. Print

Genomics, Proteomics & Bioinformatics. Ed. Yu, J. 2006-2014. http://www.journals.elsevier.com/genomics-proteomics-and-bioinformatics/ (20 May 2014)

Gregory, M. and Carroll, S. *Language and Situation: Language Varieties in Their Social Contexts*. London: Routledge & Kegan Paul. 1978. Print

Halliday, M.A.K. et al. *The Linguistic Sciences and Language Teaching.* London: Longman. 1964. Print

Hatim, B. and Mason, I. *Discourse and the Translator.* London/New York: Longman. 1990. Print

Tipos de Texto, Género y Escrito. Instituto Superior de Estudios Lomas de Zamora. 2014. http://www.isel.edu.ar/resultados_investigacion/ (21 Jun 2014)

Journal of Proteomics. Ed. Juan Calvete. 2008-2014. http://www.journals.elsevier.com/journal-of-proteomics/ (20 May 2014)

Kinneavy, J. L. *A Theory of Discourse: The Aims of Discourse*. Englewood Cliffs, NJ: Prentice-Hall International. 1971. Print

Langlois, L. *Bi-text, Bi-concordance et Collocation*. MA Thesis of the University of Ottawa. 1997. Print

Lee YW, D. "Genres, Registers, Text Types, Domains, and Styles: Clarifying the Concepts and Navigating a Path the BNC Jungle" *Language Learning & Technology*. Sep 2001, vol. 5, nº 3. Pp. 37-72. Print

Llisterri, J. et Torruella, J. "Diseño de corpus textuales y orales". *Filología e Informática. Nuevas Tecnologías en los Estudios Filológicos.* Barcelona:

Seminario de Filología e Informática, Departamento de Filología Española, Universidad Autónoma de Barcelona - Editorial Milenio. 1999, pp. 45-77. Print

Malinowsky, B. "The problem of meaning in primitive languages". Supplement I to C.K. Ogden and I.A. Richards, *The Meaning of Meaning*. London: Kegan Paul. 1923. Print

Méndez Cendón, B. *Estrategias Fraseológicas en el Género Discursivo de los Artículos Científicos Médicos en Lengua Inglesa*. PhD Dissertation. Universidad de Valladolid, Valladolid. 2002. Print.

Navarro, F. *Diccionario Crítico de Dudas Inglés-Español de Medicina.* 2nd edition. Madrid: McGraw Hill. 2005. Print

Pavel, S. and Nolet, D. *Manual de Terminología*. Québec: Ministro de Obras Públicas y Servicios Gubernamentales de Canadá. 2002. Print

Pearson, J. *Terms in Context*. Amsterdam/Philadelhia: John Benjamins. 1998. Print.

Pérez Hernández, M. C. "Explotación de los córpora textuales informatizados para la creación de bases de datos terminológicas basadas en el conocimiento". *Estudios de Lingüística del Español (ELiEs)*. Volumen 18, 2002. http://elies.rediris.es/elies18/31.html (30 Jun 2014)

Pérez Pérez, R. (2011) *Estudio de la Alteración en la Expresión de Proteínas en el Tejido Adiposo Omental y Subcutáneo en Obesidad mediante Técnicas de Proteómica.* Universidad Autónoma de Madrid. Departamento de Bioquímica. Madrid. Print.

Pizarro, I. Class notes of the subject "Linguística Aplicada III". 2013-2014. Valladolid: Universidad de Valladolid. Print

Plum, G. "Text and Contextual Conditioning in Spoken English: A Genre Approach". Layout *B. Collocation and Textual Conhesion; A Comparative Corpus Study between a Genre of Written Sports Reports and a Large Reference Corpus*. University of Birmingham, Birmingham. 2009. Print

Proteómica. Sociedad Española de Proteómica. Eds. Jorrín, J. and Vázquez, J. 2005-2012. http://www2.cbm.uam.es/seprot/revista/revista.htm (20 May 2014)

Proteomics. Ed. Michael J. Dunn. 2001-2014. http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)1615-9861 (20 May 2014)

Reid, T.B.S. "Linguistics, structuralism and philology". *Archivum Linguisticum 8*. 1956. Print

Robertson, H. *The Research Essay. A Guide to Essay's Papers*. Toronto: Mc Gaw-Hill. 1995. Print

Rodríguez Falcón, M. *Proteómica de Expresión Diferencial en Acinetobacter baumannii Resistente a Colistina*. Universitat Pompeu Fabra. Departament de Ciències Experimentals i de la Salut. Barcelona. 2010. Print.

Sager, J. A. *English Special Languages: Principles and Practice in Science and Technology*. Wiesbaden: Oscar Brandstetter 1980. Print

Sager, J. *A Practical Course in Terminology Processing*. Amsterdam [etc.]: John Benjamins. 1990. Print.

Santamaría, I. Class notes: "De la teoría a la práctica terminológica: metodología para la elaboración de un trabajo terminológico". Power Point Presentation. 2008. Print

Sinclair, J. and Ball, C. "Preliminary Recommendations on Text Typology". *EAGLES Documents EAG-TCWG-TTYP/P*, June 1996. Print.

Sollaci, L.B. et al. "The Introduction, Methods, Results and Discussion (IMRAD) Structure: a Fifty-Year Survey". *Journal of the Medical Library Association*. Jul 2004, 92(3). Pp. 364-371. Print

Swales, J.M. *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press. 1990. Print

Trosborg, A. *Text Typology and Translation*. Amsterdam/New York: To John Benjamins. 1997. Print

Webster Third International Dictionary. Merriam Webster Inc. 2002. Print.

Wordfast LLC. *Wordfast Pro* [Computer Software]. 2012. New York. http://www.wordfast.com/support_relicensewfp.html (20 May 2014)

Wordfast Pro Step by Step Translation Session. Wordfast TM. YouTube. http://www.youtube.com/watch?v=RQ-PDYwE0ZE (30 Jun 2014)

This document lists the thirty seven entries of our English/Spanish glossary. Here, there are two types of grammatical categories: names (N./N.) and acronyms (Ac./Acrón.). The reason why acronyms appear instead of the full form of the term is because it is an expert-to-expert comparable corpus. Thus, this specialized dictionary is addressed to experts in the field of Proteomics.

N.	ENGLISH	SPANISH
1.	**2D DIGE (two-dimensional Gel Electrophoresis):** [Ac.] Modified form of 2D electrophoresis (2DE) that allows one to compare two or three protein samples simultaneously on the same gel. Reference: http://www.ncbi.nlm.nih.gov/pubmed/17406422 Date of access: 25 May 2014	**2-D DIGE (electroforesis bidimensional diferencial):** [Acrón.] Forma modificada de la electroforesis bidimensional (2DE) que permite comparar simultáneamente dos o tres muestras de proteínas en el mismo gel. [Translated from the English definition]
2.	**2-DE (two-dimensional electrophoresis):** [Ac.] Widely used method for the analysis of complex protein mixtures extracted from cells, tissues, or other biological samples. Reference: http://www.gelifesciences.com/webapp/wcs/stores/servlet/catalog/es/GELifeSciences-US/products/2-d-electrophoresis/ Date of access: 25 May 2014	**electroforesis 2D (electroforesis bidimensional):** [N.] Método ampliamente utilizado para el análisis de mezclas complejas de proteínas extraídas a partir de células, tejidos u otras muestras biológicas. [Translated from the English definition]
3.	**amino acid**: [N.] A building block of proteins in which is coded by a codon	**aminoácido**: [N.] Bloque de proteínas codificadas por un codón y unidas

	and linked together through peptide bonds. Reference: http://www.biology-online.org/dictionary/Amino_acid Date of access: 25 May 2014	entre sí mediante enlaces peptídicos. [Translated from the English definition]
4.	**antibody**: [N.] A protein produced by the body's immune system when it detects harmful substances, called antigens. Reference: http://www.nlm.nih.gov/medlineplus/ency/article/002223.htm Date of access: 25 May 2014	**anticuerpos:** [N.] Proteína producida por el sistema inmunológico del cuerpo cuando detecta sustancias dañinas, llamadas antígenos. [Translated from the English definition]
5.	**BCA Protein Assay Kit:** [N.] Equipment used to determine protein concentration in the range 20-200µg/ml in either a standard assay or microassay configuration. Reference: http://www.merckmillipore.com/ES/es/product/BCA-Protein-Assay-Kit,EMD_BIO-71285 Date of access: 25 May 2014	**Kit BCA Protein Assay:** [N.] Equipo usado para determinar la concentración de proteínas en el rango 20-200µg/ml, ya sea en un ensayo estándar o en un microensayo. [Translated from the English definition]
6.	**biomarkers:** [N.] Subcategory of medical signs, that is, objective indications of medical state observed from outside the patient, which can be measured accurately and reproducibly. Reference:	**biomarcador:** [N.] Subcategoría de muestras médicas, es decir, indicaciones objetivas de estado médico observar desde el exterior del paciente, que puede ser medida con precisión y de manera reproducible. [Translated from the English

	http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3078627/ Date of access: 25 May 2014	definition]
7.	**chromatography:** [N.] Separation […] of closely related compounds, by allowing a solution or mixture to seep through an adsorbent (such as clay, gel, or paper) so each compound becomes adsorbed into a separate, often colored, layer. Reference: http://www.chemicool.com/definition/chromatography.html Date of access: 25 May 2014	**cromatografía**: [N.] Separación, de compuestos estrechamente relacionados mediante una solución o mezcla que se filtra a través de un absorbente (arcilla, gel, o papel) y forman diferentes capas por cada compuesto. [Translated from the English definition]
8.	**comparative proteomics:** [N.] Analysis of the proteome changes in response to development, disease, or environment. Reference: http://www.biotechniques.com/multimedia/archive/00001/BTN_A_000112653_O_1509a.pdf Date of access: 25 May 2014	**proteómica comparativa**: [N.] Análisis de los cambios en el proteoma en respuesta al desarrollo, enfermedad, o al medio ambiente. [Translated from the English definition]
9.	**electrophoresis:** [N.] Separation of ionic molecules, (principally proteins) by the differential migration through a gel according to the size and ionic charge of the molecules in an electrical field. Reference: http://www.biology-	**electroforesi**s: [N.] Separación de moléculas iónicas, (principalmente proteínas) mediante la migración diferencial de un gel de acuerdo con el tamaño y la carga iónica de las moléculas en un campo eléctrico. [Translated from the English

	online.org/dictionary/Electrophoresis Date of access: 25 May 2014	definition]
10	**ELISA:** [Ac.] A sensitive immunoassay that uses an enzyme linked to an antibody or antigen as a marker for the detection of a specific protein, especially an antigen or antibody. Reference: http://medical-dictionary.thefreedictionary.com/ELISA Date of access: 25 May 2014	**ELISA**: [Acrón.] Inmunoensayo que utiliza una enzima unida a un anticuerpo o antígeno como un marcador para la detección de una proteína específica, especialmente un antígeno o anticuerpo. [Translated from the English definition]
11	**gene expression:** [N.] The conversion of the information from the gene into mRNA via transcription and then to protein via translation resulting in the phenotypic manifestation of the gene. Reference: http://www.biology-online.org/dictionary/Gene_Expression Date of access: 25 May 2014	**expresión génica**: [N.] Conversión de la información a partir del gen en el ARNm a través de la transcripción y traducción que resultando la manifestación fenotípica del gen. [Translated from the English definition]
12	**HSP (Heat Shock Protein):** [Ac.] Any of a group of proteins that are produced especially in cells subjected to stressful conditions (as high temperature), that serve to ensure proper protein folding, and that are held to comprise a class of molecular chaperones. Reference: http://www.merriam-	**HSP (Proteína de Choque Térmico)**: [Acrón.] Grupo de proteínas que se producen sobre todo en las células sometidas a condiciones de estrés (como alta temperatura), que sirven para garantizar el plegamiento de proteínas, y que se lleva a cabo para comprender una clase de chaperonas moleculares.

	webster.com/dictionary/heat%20shock%20protein Date of access: 25 May 2014	[Translated from the English definition]
13	**IEF (Isoelectric Focusing):** [Ac.] Electrophoresis in which the protein mixture is subjected to an electric field in a gel medium in which a pH gradient has been established; each protein then migrates until it reaches the site at which the pH is equal to its isoelectric point. Reference: http://medical-dictionary.thefreedictionary.com/isoelectric+focusing Date of access: 25 May 2014	**IEF (Enfoque Isoeléctrico):** [Acrón.] Electroforesis en la cual la mezcla de proteínas se somete a un campo eléctrico en un medio de gel en un gradiente de pH; cada proteína luego migra hasta que el pH alcanza es igual a su punto isoeléctrico. [Translated from the English definition]
14	**IPG (Immobilized pH Gradient):** [Ac.] Difference in concentration of H+ which is used for the procedure of 2D-PAGE with the aim of eliminate the problems of instability of the gradient of pH and the low capacity of charge associated to them. Reference: http://www.ibt.unam.mx/computo/pdfs/met/plataformas_de_proteomica.pdf Date of access: 25 May 2014	**IPG (Gradientes de pH inmovilizados):** [Acrón.] Diferencia de concentración del ión H+ que se utiliza en el procedimiento del 2D-PAGE con el fin de eliminar los problemas de inestabilidad del gradiente de pH y la baja capacidad de carga asociada a estos. [Translated from the English definition]
15	**LC-MS:** [Ac.] Analytical technique for identification, quantification and mass analysis of a wide variety of non-volatile or semi-volatile organic	**LC-MS:** [Acrón.] Técnica analítica para la identificación, cuantificación y análisis de masas de una gran variedad de compuestos no volátiles o

	or inorganic compounds in mixture. Reference: http://www.chemir.com/liquid-chromatography-mass-spectrometry.html Date of access: 25 May 2014	semivolátiles orgánicos e inorgánicos en una mezcla. [Translated from the English definition]
16	**linear ion trap:** [N.] Ions [that are] confined radially by a two-dimensional (2D) radio frequency (RF) field, and axially by stopping potentials applied to end electrodes. Reference: http://onlinelibrary.wiley.com/doi/10.1002/mas.20004/abstract Date of access: 25 May 2014	**trampa iónica lineal**: [N.] Iones que se limitan radialmente por una radiofrecuencia bidimensional de campo (RF), y axialmente por la detención de potenciales aplicados para poner fin a los electrodos. [Translated from the English definition]
17	**MALDI-TOF:** [Ac.] Technique of ionization which determines the level of morphological correlation between molecular and histological image. Reference: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2918004/ Date of access: 25 May 2014	**MALDI-TOF:** [Acrón.] Técnica de ionización que determina el nivel de correlación morfológica entre la imagen molecular y la histológica. [Translated from the English definition]
18	**MASCOT:** [Ac.] Test made by the standard for protein identification using mass spectrometry. Reference: http://www.matrixscience.com/server.html Date of access: 25 May 2014	**MASCOT:** [Acrón.] Prueba estándar usada para la identificación de proteínas usando la espectrometría de masas. [Translated from the English definition]

19	**mRNA (Messenger RNA):** [Ac.] Molecule in cells that carries codes from the DNA in the nucleus to the sites of protein synthesis in the cytoplasm (the ribosomes). Reference: http://global.britannica.com/EBchecked/topic/377106/messenger-RNA-mRNA Date of access: 25 May 2014	**ARNm (ARN mensajero):** [Acrón.] Molécula en las células que contienen códigos de ADN en el en la síntesis de proteínas en el citoplasma (ribosomas). [Translated from the English definition]
20	**MS (Mass Spectrometry):** [Ac.] Method for identifying the chemical constitution of a substance by means of the separation of gaseous ions according to their differing mass and charge. Reference: http://www.merriam-webster.com/dictionary/mass%20spectrometry Date of access: 25 May 2014	**MS (espectrometría de masas):** [Acrón.] Método para identificar la constitución química de una sustancia mediante la separación de iones gaseosos según sus masas y sus cargas. [Translated from the English definition]
21	**MS/MS:** [Ac.] Method for structure determination and analysis of molecules. Reference: http://scs.illinois.edu/massSpec/ion/msms.php Date of access: 25 May 2014	**MS/MS**: [Acrón.] Método para determiner la estructura y análisis de las moléculas. [Translated from the English definition]
22	**Orbitrap:** [N.] Mass spectrometer that operates by radially trapping ions	**Orbitrap:** [N.] Espectrómetro de masas que funciona atrapando

	about a central spindle electrode. Reference: http://chimica.campusnet.unito.it/didattica/att/5658.7195.file.pdf Date of access: 2014	radialmente iones sobre un electrodo de husillo central. [Translated from the English definition]
23	**phosphorylation:** [N.] [Process of] transferring a phosphoryl group from a donor to the recipient molecule. Reference: http://www.biology-online.org/dictionary/Phosphorylation Date of access: 25 May 2014	**fosforilación**: [N.] Proceso por el cual se transfiere el grupo fosforilo de un donante a un recipiente molecular. [Translated from the English definition]
24	**PMF (Peptide Mass Fingerprinting):** [Ac.] Technique used to identify proteins by matching their constituent fragment masses (peptide masses) to the theoretical peptide masses generated from a protein or DNA database. Reference: http://www.ionsource.com/tutorial/protID/fingerprint.htm Date of access: 25 May 2014	**huellas de masas péptidas (PMF)**: [N.] Técnica analítica usada para identificar proteínas mediante la correspondencia entre las masas de los fragmentos que lo forman (masas peptídicas) y las masas teóricas de los péptidos, generadas en una base de datos de proteínas o ADN. [Translated from the English definition]
25	**protein:** [N.] Molecule made up of amino acids that are needed for the body to function properly. Reference: http://www.ghr.nlm.nih.gov/glossary=protein Date of access: 25 May 2014	**proteína**: [N.] Molécula hecha de aminoácidos que son necesarios en el cuerpo para funcionar correctamente. [Translated from the English definition]
26	**protein expression:** [N.]	**expresión proteica**: [N.]

	Identification of the proteins expressed in a particular tissue, under a specified set of conditions and at a particular time, usually compared to expression in reference samples. Reference: http://www.ncbi.nlm.nih.gov/pubmed/19381602 Date of access: 25 May 2014	Identificación de las proteínas expresadas en un tejido particular, bajo un conjunto específico de condiciones y en un momento determinado, por lo general comparadas con la expresión en las muestras de referencia. [Translated from the English definition]
27	**protein identification:** [N.] Analysis in-gel proteolytic cleavage of the protein(s) sample followed by extraction and mass spectrometric analysis […] obtaining peptide sequence information. Reference: http://www.genosphere-biotech.com/Protein-Identification-Services.html Date of access: 25 May 2014	**identificación proteica**: [N.] Análisis de escisión proteolítica en gel de las proteínas seguida por la extracción y análisis por espectrometría de masas, obteniendo información de la secuencia del péptido. [Translated from the English definition]
28	**protein spots:** [N.] Small samples of proteins used in 2D electrophoresis which are excised from the gel and the protein is subjected to proteolytic treatment prior to identification by mass spectrometry. Reference: http://www.bio-rad.com/es-es/applications-technologies/protein-spot-excision-protein-identification#1 Date of access: 25 May 2014	**spots de proteína**: [N.] Pequeñas muestras de proteínas utilizadas en la electroforesis bidimensional, las cuales se extraen del gel y de la proteína y se someten a un tratamiento proteolítico antes de la identificación por espectrometría de masas. [Translated from the English definition]

29	**protein synthesis:** [N.] Creation of proteins by cells that uses DNA, RNA and various enzymes. Reference: http://www.biology-online.org/dictionary/Protein_synthesis Date of access: 25 May 2014	**síntesis de proteínas**: [N.] Creación de proteínas mediante células que usan ADN, ARN y varias enzimas. [Translated from the English definition]
30	**proteome:** [N.] All the proteins produced by an organism. Reference: http://www.ama-assn.org//ama/pub/physician-resources/medical-science/genetics-molecular-medicine/current-topics/proteomics.page Date of access: 25 May 2014	**proteoma**: [N.] Conjunto de proteínas producidas por un organismo. [Translated from the English definition]
31	**proteomics:** [N.] Large-scale study of proteins, particularly their structures and functions. Reference: http://www.news-medical.net/health/Proteomics-What-is-Proteomics.aspx Date of access: 25 May 2014	**proteómica**: [N.] Estudio a gran escala de las proteínas, en concreto de su estructura y funciones. [Translated from the English definition]
32	**PTM (Post-translational Modifications:** [Ac.] Chemical modification of a protein after its translation; one of the later steps in protein biosynthesis, and thus gene expression, for many proteins. Reference: https://www.boundless.com/biology/d	**PTM (Modificaciones Postraduccionales):** [Acrón.] Modificación química de una proteína tras su traducción; uno de los pasos posteriores de la biosíntesis de proteínas, y por lo tanto la expresión génica de muchas proteínas. [Translated from the English

	efinition/post-translational-modification/ Date of access: 25 May 2014	definition]
33	**quantitative proteomics:** [N.] Branch of proteomics used for both discovery and targeted proteomic analyses to understand global proteomic dynamics in a cell, tissue or organism. Reference: http://www.piercenet.com/method/quantitative-proteomics Date of access: 25 May 2014	**proteómica cuantitativa**: [N.] Rama de la proteómica usada para detectar y dirigir análisis proteómicos para entender la dinámica global de una célula, tejido u organismo. [Translated from the English definition]
34	**SDS-PAGE:** [Ac.] Method by which proteins are separated largely on the basis of polypeptide length, and so their molecular weight can also be estimated. Reference: http://www.science.smith.edu/departments/Biochem/Biochem_353/sdspage.html Date of access: 25 May 2014	**SDS-PAGE:** [Acrón.] Método por el cual las proteínas se separan en gran medida en una base de un polipéptido, y en consecuencia, se conoce su peso molecular. [Translated from the English definition]
35	**ubiquitin:** [N.] A small polypeptide that is involved in histone modification and is a marker for intracellular protein transport and degradation. Reference: http://medical-dictionary.thefreedictionary.com/ubiq	**ubiquitina**: [N.] Pequeño polipéptido implicado en la modificación de histonas y es un marcador para el transporte intracelular de proteínas y su degradación. [Translated from the English definition]

	uitin Date of access: 25 May 2014	
36	**ubiquinitation:** [N.] An enzymatic, protein post-translational modification (PTM) process in which the carboxylic acid of the terminal glycine from the di-glycine motif in the activated ubiquitin forms an amide bond to the epsilon amine of the lysine in the modified protein. Reference: http://www.news-medical.net/health/Ubiquitination-(Ubiquitylation).aspx Date of access: 25 May 2014	**ubiquitinación**: [N.] Modificación post-traduccional (PTM) de las proteínas enzimáticas el cual el ácido carboxílico de la glicina del extremo de la di-glicina en la ubiquitina activada forma un enlace amida con la amina épsilon de la lisina en la proteína modificada. [Translated from the English definition]
37	**Western Blot:** [N.] (also called immunoblotting) Technique used for analysis of individual proteins in a protein mixture (e.g. a cell lysate). Reference: http://www.antibodies-online.com/resources/17/1224/Western+blotting+immunoblot+Gel+electrophoresis+for+proteins/ Date of access: 25 May 2014	**Western Blot**: [N.] (también llamada inmunotransferencia) Técnica usada por el análisis de las proteínas individuales en una mezcla proteica. [Translated from the English definition]