Pamela Faber, Rafael Navajas-Pérez
University of Granada – Spain

SECONDARY TERM FORMATION AND TERM STABILITY IN GENETICS: AN ENGLISH-SPANISH CORPUS STUDY

Abstract: This research analyzed a set of representative terms in the field of Genetics, all of which have multiple and competing correspondences in Spanish. Based on the results obtained, seven secondary term formation strategies were identified, and possible reasons were given for implementing one instead of the other in different contexts. The results obtained indicate that factors that lead to the use of one strategy over the others include expediency, the desire for semantic transparency, and the morphology of the target language. In this regard, the knowledge level of the target language users was also found to motivate the choice of target language correspondence.

Keywords: Term formation; Genetics; Domain loss, Term stability.

1. Introduction

In science and technology, research findings must be rapidly disseminated. The most effective way to communicate new discoveries to the international community is to employ a global language that everyone can understand. In today's world, the language used for this purpose is English. In fact, over the years English has become the language of science and the "de facto" language of communication for those who wish their oral or written texts to travel beyond national boundaries.

This is certainly the case in international conferences where scientific and technical research findings are presented and discussed. These events are breeding grounds for temporary, self-contained cultures, specific of a certain professional group. The use of the same type of discourse and terminology affirms the speakers' membership in a certain profession or specialized field (industrial engineering, medicine, chemistry, genetics, etc.). Since the terminology represents the knowledge structures characteristic of the domain, the relation between specialized knowledge and the language used to express it is

very close. In fact, a mastery of scientific and technical language in a certain area is often perceived as being synonymous with a mastery of the knowledge designated by the terms.

This is not only true for individuals but also for language-cultures. Speaker communities in which cutting-edge research is performed and new scientific discoveries are made usually have the terms at their disposal to designate new knowledge. Generally speaking, languages such as English and Japanese are where primary term formation takes place. Precisely for this reason, English has an advantage over minority languages (e.g. Norwegian, Basque, Bulgarian, etc.), which must adopt some type of translation strategy and policy in order to assimilate and/or systematically create correspondences for newly coined scientific terms in English. All too often, secondary term formation occurs very quickly. This is particularly the case when the research was performed in a non-Anglophone country and the findings are disseminated in parallel in English and another language.

However, as observed by Tardy (Tardy 2004: 248), using English for research dissemination purposes in international research contexts has advantages as well as drawbacks. Although the use of English as a *lingua franca* enhances and facilitates information storage and retrieval, scientists from other countries often feel at a disadvantage when it comes to publishing their work. The prevalence of English, especially in relatively new scientific domains, such as Genetics, has had a significant impact on scientific discourse published in other languages. This is reflected at the syntactic, semantic, and pragmatic dimensions of specialized knowledge units and texts.

One of the most evident effects of the dominance of English means that it is often difficult to find high-level scientific research published in other languages. Scientists and engineers, even those who speak major languages (e.g. Spanish, German, Russian, Chinese), are aware that they must publish in English if they wish their work to be read by specialists in other countries. Not surprisingly, this has created a thriving market for the translation and revision of scientific and technical texts into English since few researchers wish to divulge the results of their studies to a limited readership. However, the dominant role of English in the transmission of research to the international scientific community is not exempt of negative consequences.

Swales (Swales 1997: 374) goes so far as to characterize English as a *Tyrannosaurus rex* that "gobbles up the other denizens of the academic linguistic grazing grounds". Similarly, Phillipson (Phillipson 2008: 251) suggests that there are grounds for referring to English as a *lingua frankensteinia.* Although English is generally marketed as being a language of international understanding, which foments the advancement of scientific knowledge, its exclusive use in certain disciplines can promote the marginalization of other languages (Phillipson 2008). This is reflected in the lack of terminology (or stable terminology) in other languages to designate specialized concepts originally created and named in English.

In fact, when English terminology is used to exclusively name scientific concepts in other languages and present research results, it can even lead to *domain loss*. This term refers to the reduced use of a certain language in a domain (area or context) usually because another language has taken its place. The concept of *domain* initially became popular because of Fishman (Fishman 1972), who used it to explain the codeswitching of speakers in multilingual settings, who choose to use one language instead of another in a certain area. However, according to Haberland (Haberland 2005), the term, *domain loss,* lacks precision because there are different perceptions of what constitutes a domain, which seems to refer to a wide range of contexts and activities.

Not surprisingly, domain loss is a hot topic in the Scandinavian countries, where English is increasingly used in domains such as higher education (Hultgren 2013), and is thus regarded as a threat to the national languages. For example, Björkman (Björkman 2015) points out that there is the perception that Swedish is 'losing ground' to English, and that Swedish terminology is not available for certain domains. This is also applicable to other language cultures.

Within the context of scientific communication, domain loss is generally defined as the "loss of ability to communicate in a language on all levels of an LSP field because of deficient further development of the necessary LSP resources" (Lauren at al. 2002: 25). Although the most direct effect of this kind of dispossession is the lack of terminology in a language, as shall be seen, domain loss is more subtly reflected in the systematic use of calques and multiple terminological correspondences that do not conform to the rules of term formation in the target language. This tends to occur when terminological correspondences in other languages are created with excessive speed.

This type of 'on-the-fly' term creation is especially prevalent in relatively new domains such as Genetics, a science that has developed so rapidly that the establishment of stable correspondences for the English terms still lags far behind. In this field, new terms often appear very quickly in other languages, mostly to meet academic needs (e.g. publication of university textbooks). Depending on a number of factors, a newly created correspondence may survive and eventually become the established term in the target language, or it may eventually become deprecated and be replaced by another term that better conforms to the rules of the target language. There also might be a longer or shorter period in which two or more terms co-exist in the same language, after which one might be restricted to scientific texts of a certain knowledge level or with a specific function. Alternatively, the new term/s might finally disappear altogether.

Using examples from the field of Genetics, this paper focuses on the analysis of English and Spanish terminological designations in texts for users with different knowledge levels (lay, semi-expert, and expert). The remainder of the paper is structured as follows. Section 2 gives an overview of primary and secondary term formation processes and contexts. Section 3 describes the dimensions and characteristics of the corpus as well as the computer applications used in the analysis. Section 4 presents the results obtained and discusses their significance within the context of secondary term formation. Finally, Section 5 lists the conclusions regarding term formation strategies and patterns that can be derived from this research.

2. Term formation

As previously mentioned, the rapid evolution of science and technology generates new concepts and new terms to designate them. Depending on whether a term for the new concept already exists in another language, term formation can be either primary or secondary (Sager 1990: 80). These types of term formation differ in their contextual circumstances. However, as stated by Temmerman (Temmerman 2013), both primary and secondary term creation result in motivated terms and play a role in the advancement of understanding. Both processes are also embedded in situated communication, a context influenced by intralinguistic, interlinguistic, and extralinguistic circumstances.

2.1. Primary term formation

Primary term formation occurs when a new concept requires a name. Even when there is no linguistic precedent, there is a set of term formation rules in the source language. Although the process of concept designation appears to be more or less spontaneous, it can be implicitly influenced by already existing terms within the same domain as well as by language structure. It may also be based on terminologization and metaphorical extension since both are well-known processes of lexical creation. In science, terms can be borrowed from general language (terminologization) or even from another specialized domain, and then used in a metaphor to fill a lexical gap where no term exists. This corresponds to Halliday's ideational function (Goatly 1997).

For example, in Genetics, primary term creation makes use of both processes. Genetic terms such as *messenger RNA, genetic code, transcription, translation, proofreading, gene library*, etc., which have been recycled from general language, reflect the conceptualization of DNA as a language in which language-related activities take place (Volanschi and Kübler 2011: 212-213).

Despite the fact that in science, primary term formation generally happens in English, it can also be multilingual when there is the need to divulge or name the entity in more than one language. In this sense, *parallellingualism* (divulging terms in parallel) is a key concept in Nordic language policy, and has been proposed as a way to ensure an equitable balance between English and the Nordic language(s) without the former encroaching on the latter (Hultgren 2013: 61; Hultgren 2014).

2.2. Secondary term formation

As pointed out by Sager (Sager 1997: 27), secondary term creation can occur either in a monolingual or an interlinguistic context. For example, a monolingual revision of terminology can lead to a change in the designation of a concept. Secondary term creation also occurs when scientific and technological knowledge is transferred from one language to another. In this type of transfer, which is often based on partial or total lexical borrowing, the English source language term often conditions the target language correspondence in some significant way.

This may involve the transposition of metaphors in different language communities. When the languages are more distant from each other, the transposition of metaphors may be more complex than previously assumed (Grevy 1999). However according to Humbley (Humbley 2006: 207), metaphors taken from shared cultural sources facilitate secondary term formation since then it is easier to identify the original conceptualization as well as the fields concerned.

In this sense, awareness of the processes of primary term formation in the source language can help to guide secondary term creation. Valeontis (Valeontis 2004) and Valeontis and Mantzari (Valeontis, Mantzari 2006) even go so far as to affirm that in secondary term formation, it is absolutely necessary to take into account the corresponding processes in the source language just as they are reflected in primary term formation. Apart from the viability (or possibility) of this methodology, this is rarely what occurs in secondary term formation, where English loanwords, semantic calques, and adaptations are the order of the day.

In the translation of specialized texts, newly coined terms can be problematic because they are often too recent to be included in specialized dictionaries and term banks. They thus lack established equivalents in the target language or may not have any equivalents at all. Not surprisingly, the translation of neologisms has been described as the biggest problem for non-literary and professional translators (Newmark 1988). In translation contexts, situations often arise in which there is no target language correspondence for the new term in the source language. As a result, translators must decide whether to create a neologism or to make an explanatory or descriptive translation.

Descriptive translations are made by using generic terms from a definition derived from the term's context or simply by borrowing, calquing, or adapting the source language term. If translators decide to create a neologism or propose a new term, the new term should conform to the word-formation rules in the target language system. Relevant examples are provided by Fernández-Domínguez (Fernández-Domínguez 2016), who makes a contrastive analysis of the morphological and semantic characteristics of English and Spanish terms from the olive oil industry.

The translation strategy employed also depends on the capacity of the target language to accept and assimilate foreign words. Although almost half of English

consists of words borrowed from other languages, languages, such as Spanish or French, are more reticent about absorbing words from other languages, possibly because of the existence of national language academies in each country (i.e. the *Academie Française* and the *Real Academia Española*). This tendency, however, is at odds with the rapid influx of technical and scientific terms in the modern world (Montero-Martínez at al. 2001). Lexical borrowing is so frequent because from a practical perspective, it is often the easiest solution to implement.

According to Kachru (Kachru 1994: 139), the primary motivation for lexical borrowing is to remedy a linguistic deficit in the lexical resources of a language (i.e. The Deficit Hypothesis). However, a more plausible explanation is provided by the Expediency Hypothesis (Widyalankara 2015: 2), which states that for reasons of pragmatism and convenience, a linguistic community tends to favor loanwords simply because adopting a word (or part of a word) is less work and much quicker than creating a new one.

For this reason, the structure of the English source language term may condition the creation of new lexical units in the target language. The target language terms thus created may show various degrees of *terminological dependency* (Humbley and García-Palacios 2012), which reflect the influence that English has on the formation of terminological units in other languages. Nevertheless, as shall be seen, the structure of the target language also has an impact on the creation and subsequent entrenchment of target language correspondences. This paper focuses on secondary term creation in interlinguistic contexts in rapidly evolving scientific domains such as Genetics.

3. Materials and Methods

3.1. Corpus

The corpus used in this study was composed of parallel texts in English and Spanish, and had a total of 853,479 words. The general corpus was divided into three subcorpora with English and Spanish parallel texts typical of three knowledge levels for the following user types (1) expert; (2) semi-expert; (3) lay readers.

The subcorpus of texts for lay readers included articles from recent issues of the journal, *Scientific American,* and its Spanish counterpart, *Investigación y Ciencia.* The texts on Genetics were manually extracted from the issues of these journals. Another source of texts for non-experts was *Open Mind* (https://www.bbvaopenmind.com), which publishes overviews of scientific topics in Spanish and English for the general public. This subcorpus had 389,701 words of which 201,029 were in Spanish and 188,672 in English.

The subcorpus of texts for semi-experts was composed of Genetics university textbooks in English and Spanish, such as *Genetics: A Conceptual Approach* (2nd edition) and its Spanish translation, *Genética: un enfoque conceptual.* The 20 parallel texts in this subcorpus totaled 245,729 words: 133,374 in Spanish and 122,355 in English.

Because of the difficulty of finding cutting-edge research in Genetics written in Spanish, the subcorpus of texts for experts was extracted from the repository of PhD dissertations in the DIGIBUG database of the University of Granada (Spain). The seed words used were *genética* [genetics], *genómica* [genomics], and *gen* [gene]. The documents selected were dissertations in which part of the research had been carried out in an English-speaking country, and which thus contained sections with an English translation. Also included in this subcorpus were papers on Genetics from the *Revista Española de Cardiología,* whose articles are published in English and Spanish. The 79 documents in this subcorpus totaled 218,049 words: 108,032 in Spanish and 110,017 in English.

All of the texts were converted to txt format with the application Notepad++ and manually inspected for their readability. The corpus was then uploaded to SketchEngine for analysis. When necessary, the size of the Spanish corpus was increased by means of the WebBootCat function of Sketch Engine. This tool allows the user to rapidly compile a corpus from the web, based on the seed terms entered. Furthermore, the user can choose the web sites to be included. This automatic compilation complemented the manual selection of corpus texts.

3.2. Sketch Engine

Sketch Engine (www.sketchengine.co.uk) is a corpus query system that allows the users to view concordance lines, word sketches, meaning-related words, frequencies, as well as a wide range of contextually related information.

One of its most useful functionalities is the Word Sketch, which is an automatic corpus-derived summary of a word's grammatical and collocational behavior. Word sketches have the advantage of being fully integrated with the concordancing. By clicking on a collocate of interest in the word sketch, users can access concordances from the corpus evidence, which exemplify that collocate in that grammatical relation (Kilgarriff et al. 2014).

For the purposes of this study, based on frequency data, a set of English terms in the subject field of Genetics was generated along with their Spanish translations. Each English term had more than one Spanish correspondence in the corpus texts. The analysis discusses possible reasons for this process of multiple secondary term creation.

4. Results and discussion

4.1. Results

Table 1 shows the English terms analyzed and their Spanish correspondences. These representative examples reflect that secondary term formation in Spanish is often motivated by competing strategies, which can produce various correspondences for each source language term. As reflected in Table 1, all of the source language terms have various correspondences in the target language.

The terms in boldface are the correspondences given in the Spanish version of Wikipedia. In the case of *marcador de secuencia expresada,* which appears in parenthesis, this Wikipedia term did not appear in any of the texts in our corpus.

Table 1. English terms and their Spanish correspondences

Concept	English term	Spanish correspondences
Object	DNA	***ADN**, DNA*
	RNA	***ARN**, RNA*
	genetic code	***código genético**, DNA*
	contig	***cóntigo**, contig*
	expressed sequence tag (EST)	*EST, secuencia ESTs, secuencia EST, etiqueta de secuencia expresada,* ***(marcador de secuencia expresada)***
Process	crossing over	*sobrecruzamiento,* ***entrecruzamiento**, recombinación*
	genomic imprinting	***impronta genética**, impronta genómica imprinting, marcaje genómico*
	deletion	***deleción**, delección*
	splicing	***splicing, corte y empalme, ayuste***
	DNA fingerprinting	*huella dactilar genética,* ***huella genética**, perfil genético*

The following sections analyze the secondary term formation strategies used and propose possible reasons for the proliferation of target language equivalents.

4.1.1. DNA and RNA

Not surprisingly, the most frequent terms in the corpus were *DNA* (dexoxyribonucleic acid) and *RNA* (ribonucleic acid). As is well known, DNA is a molecule that carries the genetic instructions used in the growth, development, functioning and reproduction of all known living organisms and many viruses. Although the *Diccionario de la Real Academia Espanola* accepts both *DNA* (the English term) and *ADN* (its Spanish translation) as designations for the concept, the preference for *DNA* or *ADN* in the Spanish texts of our corpus clearly depended on the level of specialization. Texts for experts tended to use *ADN* whereas those for semi-experts used *DNA*. Texts for lay readers used both though with a preference for *ADN*. Exactly the same thing occurred with *RNA* (ribonucleic acid), whose Spanish equivalents are RNA and ARN.

This seems to indicate that although *DNA* and *RNA* are popularly accepted in the target language at the non-expert and semi-expert level, experts prefer the Spanish translations, *ADN* and *ARN*. Both the English loanword and the target language translations co-exist in Spanish though in texts of different knowledge levels.

4.1.2. Contig

The English term *contig* is a clipping of *contiguous*. In Genetics, it refers to a set of overlapping DNA segments that together represent a consensus region of DNA. In Spanish, there are two correspondences that co-exist in specialized texts, though much less comfortably than *DNA* and *ADN*. The Spanish correspondences for *contig* are: (i) *contig,* the direct borrowing from English; (ii) *cóntigo,* an adaptation of the English term into Spanish. Both terms clash with Spanish grammar and morphology though for different reasons.

For example, *contig,* the loanword, is not in consonance with word-formation patterns in the target language because there are no Spanish words ending in 'g'. In the second correspondence, *cóntigo* [contig], an "o" was added to the source language term in an effort to adapt the term to Spanish. However, the result is not successful because the additional "o" makes it an exact match of *contigo* [with you], a resemblance that even the accent added to the first syllable cannot attenuate.

In our corpus, only *contig* appeared in our corpus of parallel texts. As for *cóntigo*, it was only found in a relatively low number of Internet texts, though always in combination with *contig.* This type of clarification is an indicator of term instability and lack of entrenchment in the target language.

(1) Ello conduce al concepto *contig*: un contig (o *cóntigo*, como algún autor español ha traducido...) [This leads to the concept of *contig*: a contig (or *cóntigo*, as translated by some Spanish authors…]

Despite the fact that *cóntigo* is rarely found without the English term alongside of it, it is the correspondence for *contig* given in the Spanish version of Wikipedia as well as in various online bilingual glossaries (*Glosario hablado de términos genéticos, Glosario de biotechnología para la agricultura y la alimentación, Vocabulario inglés-español de bioquímica y biología molecular,* etc.). The continued co-existence of *contig* and *cóntigo* in Spanish texts shows that these terms are still in competition with each other. The persistence of the borrowed term (*contig*) could stem from the lack of a felicitous Spanish adaptation of the term, or at least one that does not activate a similar though irrelevant word pattern in the speaker's mind.

4.1.3. Expressed sequence tag (EST)

An *expressed sequence tag* is a unique stretch of DNA within a coding region of a gene that is useful for identifying full-length genes and serves as a landmark for mapping. ESTs are collected in libraries or databases. There are five possible Spanish correspondences for this term. The first option is a direct borrowing of the abbreviation, *EST*. The abbreviation often appears alongside *etiqueta de secuencia expresada,* the literal Spanish translation. This correspondence is frequently found in textbooks, where the most expedient solution is a semantic calque of the source language term. Curiously, this

Spanish correspondence never appears as *ESE*, the Spanish abbreviation, but always as *EST*, the English abbreviation. This tendency is also evident in other equivalences such as *secuencia EST* and its variant, *secuencia ESTs,* in which *secuencia* is reiterated in Spanish, despite the fact that *sequence* is one of the words in the English abbreviation. The frequency of these correspondences highlights the entrenchment of *EST* in the specialized domain in both the source and target language.

Curiously, the Wikipedia proposal of *marcador de secuencia expresada*, [expressed sequence marker], which did not appear in our corpus, seems to be gaining ground, especially in the Latin American texts retrieved from the Internet. In fact, it has almost the same number of hits on Google as *etiqueta de secuencia expresada,* the more literal translation of the English term.

The evolution of term preferences can also be seen in various collocations of *EST*. For example, *EST* tends to combine with *library, collection,* or *database* because ESTs are invariably stored in some type of repository. In Spanish, *EST library* was initially calqued as *librería de ESTs,* based on the surface similarity between *library* and its false friend, *librería.* However, this translation was erroneous because the meaning of *librería* is not library, but bookshop. Although *librería de ESTs* can still be found in older texts, the term has now been corrected, and in more recent books and articles, the Spanish correspondence for *EST library* is *biblioteca de ESTs* [EST library] or *genoteca de ESTs* [gene bank].

4.1.4. Crossing over/crossover/recombination

In the field of Genetics, the exchange of genetic material between homologous chromosomes that results in recombinant chromosomes during sexual reproduction has three possible designations: *crossing over, crossover,* and *recombination.* As can be observed, these terms highlight different facets of the genetic process designated. More specifically, *recombination* refers to the final result of the process whereas *crossing over* and *crossover* refer to the process that produces the result. The expert texts in English tend to use *recombination,* which is more precise because *crossing over*, though the main cause, is far from being the only cause of recombination. The lay user texts also use *recombination*, possibly because it is more semantically transparent than *crossing over*. In the texts for semi-experts, all three terms appear with more or less the same frequency.

In Spanish, *crossing over* is translated primarily as *entrecruzamiento* [(literally) crossing between] and *recombinación* [recombination]. However, despite being the exact calque of the English term, *sobrecruzamiento* [crossing over] has a significantly lower frequency with only a few hits on Internet and in the CREA corpus because it is a false derivation. More specifically, **sobrecruzar* [to cross over] is not a verb in the Spanish language, whereas *entrecruzar* does exist as a verb, as reflected in the *Diccionario de la Lengua Española* of the Spanish Royal Academy.

Another sign that *sobrecruzamiento* lacks stability is the fact it often appears alongside either the English source term or one of the other Spanish correspondences:

(2) Este fenómeno se llama **sobrecruzamiento** o crossing-over. [This phenomenon is called *sobrecruzamiento* or *crossing-over*.]

(3) El **sobrecruzamiento** o crossing-over es el fenómeno responsable del intercambio y ya hemos visto que ocurre durante la primera mitad de la fase [The *sobrecruzamiento* or *crossing-over* is the phenomenon responsible for the interchange and we have already seen that it occurs during the first half of the phase]

(4) En este período se produce el **sobrecruzamiento**, entrecruzamiento o crossing-over entre cromátidas homólogas [In this period, the *sobrecruzamiento, entrecruzamiento* or *crossing-over* is produced between homologous chromatids.]

This inclusion of various target language correspondences for the same concept is an indicator of the instability or lack of entrenchment of the target language correspondence.

4.1.5. Genomic imprinting

Genomic imprinting is the phenomenon of parent-of-origin gene expression, where the expression of a gene depends upon the parent who passed on the gene. In Spanish, the correspondences are *imprinting, impronta genética, impronta genómica, marcaje genómico*. Secondary term creation processes were the following: (1) adoption of the source language term (*imprinting*) as a loanword; (2) translation of *imprinting* as *impronta* [imprint] or *marcaje* [marking] (*impronta genómica, marcaje genómico*); and (3) translation of *genomic* as *genómico* or *genética.*

In the last two cases, different translations for one part of the term led to a veritable proliferation of correspondences in the target language. This can be problematic because the resulting term pairs are rarely, if ever, synonymous. For example, in Spanish, from both a linguistic and scientific perspective, *genómico* [genomic] and *genético* [genetic] are clearly not synonyms since the first is related to a genome or an organism's entire genetic makeup, and the second is related to genetics or the study of individual genes and their role in in heritance. Nevertheless, in a wide variety of texts, *genómico* and *genética* are often used interchangeably.

As pointed out by Navarro (Navarro 2008), this can lead to inaccuracy. He gives the example of *mutación genética* [genetic mutation], which is produced by a change in the base sequence of a gene or the transformation of an allele into another. In contrast, *mutación genómica* [genomic mutation] occurs when the number of chromosomes in the genome of an organism is modified. *Mutación genómica* [genomic mutation] is thus much more specific than *mutación genética* [genetic mutation]. However, at the more practical level, many scientific writers are not aware of this difference and use them as synonyms.

In our corpus, there was a clear preference for *impronta genética* followed by *impronta genómica. Marcaje genómico* [genomic marking] lagged far behind and was clearly not favored, probably because it is not a semantic calque but rather a semantic adaptation of the English source language term. When it does appear in texts, it is almost always accompanied by one of the other correspondences:

(5) La impronta o **marcaje genómico** no supone cambios en la secuencia de nucleótidos del DNA [The *impronta* or *marcaje genómico* does not produce changes in the nucleotide sequence of the DNA.]

(6) a este fenómeno se le denomina impronta genómica o **marcaje genómico** ("genomic imprinting"). [This phenomeonon is known as *impronta genómica* or *marcaje genómico* (*genomic imprinting*).]

As can be observed in (5) and (6), a stable correspondence for *genomic imprint* does not as yet exist in Spanish. Secondly, the order of the terms in examples (5) and (6) indicates that *marcaje genómico* is the least preferred option because it is always mentioned last on each list. This insistence on semantic clarification in the Spanish texts is an implicit acknowledge of the lack of term stability in the target language.

4.1.6. Deletion

Within the context of Genetics, *deletion* is a mutation (a genetic aberration) in which a part of a chromosome or a sequence of DNA is lost during DNA replication. There are two possible Spanish correspondences for *deletion*, which are *deleción* and *delección*. Despite the fact that *delección* is not correct, both terms co-exist in Spanish scientific books and articles. Although *deleción* was the only correspondence found in our corpus, *delección* (in contexts related to *cromosoma* [chromosome]) had a relatively high frequency in the texts retrieved from Internet.

One of the reasons that *delección,* despite being erroneous, has arisen as a correspondence is probably because of its surface similarity with the source language term (*deletion).* However, the main reason for its persistence in scientific texts seems to be its structural resemblance to other familiar Spanish words such as *selección* [selection], *detección* [detection], *defección* [defection], *predilección* [predilection], etc. Here, the morphological structure of the target language seems to have influenced term formation. This logically generates a certain level of confusion and competition between the correct equivalent, *deleción,* and *delección,* which, despite being incorrect, is in greater harmony with word formation patterns in the target language.

The confusion stems from the etymology of *deletion,* which is not derived from the Latin *delectio,* but rather from *deletion, -onis*. Nevertheless, members of the speaker community are generally unaware of the etymology of a term and tend to opt for a linguistic form resembling that of other more familiar words

in the language. Still another possible reason for the persistence of *delección* is that it is an entry in lexicographic resources, such as the *Diccionario Manual de la Lengua Española, Diccionario Enciclopédico Vox,* and *Diccionario de Medicina Vox.*

(7) **delección**
s. f. Pérdida de un segmento cromosómico que lleva consigo la desaparición de l a información genética contenida en él. (*Diccionario Manual de la Lengua Española Vox*. © 2007 Larousse Editorial, S.L.).

(8) **delección**
f. biol. Aberración cromosómica que consiste en la falta de una porción del mate rial genético de un cromosoma. (*Diccionario Enciclopédico Vox*. © 2009 Larousse Editorial, S.L.).

(9) **delección**
Sustantivo femenino
Pérdida de una porción de un cromosoma o de una o varias bases de un fragmento de DNA. Las delecciones provocan graves alteraciones del código genético y en ocasiones se manifiestan en forma de retrasos mentales, microcefalia, epicantus, labio leporino e hipertelorismo, entre otras. (*Diccionario de Medicina Vox*).

In this case, term instability is produced by the structure of the target language, which is more receptive to the erroneous correspondence (as reflected in the Internet corpus). However, the exclusive use of *deleción* in our corpus of texts seems to confirm the preference for the correct term in academic contexts.

4.1.7. Splicing

In Genetics, *splicing* can be defined as the process by which the DNA of an organism is cut and a gene, perhaps from another organism is inserted. As observed by Temmerman (Temmerman 2013), *splicing* is a polysemous term that can be applied to rope and film as well as to DNA. As such, it is an example of terminologization and metaphorical extension in which a general language word is recycled and used in a specialized context, where it becomes a term.

The Spanish correspondences for *splicing* are *splicing, corte y empalme,* and *ayuste.* The first correspondence, *splicing,* is the source language term used as a loanword. In the second, *corte y empalme* [cut and join] is a correspondence stemming from the explicitation of the splicing process. The third, *ayuste,* is a term generally used in nautical contexts, and is the Spanish translation of *splice,* though in reference to two ends of a cable or rope. Borrowed from another specialized domain, this correspondence stems from the same process of metaphorical extension as the source language term.

Because of its specialized meaning, the topic of splicing was not addressed in any of the texts in the subcorpus for lay users. In the subcorpus for semi-experts, *corte y empalme* was the preferred term, perhaps because of its transparency. In contrast, in the subcorpus for experts, the preference was for *ayuste,* with

splicing, a close second. In the Spanish version of Wikipedia, all of these terms were provided as possible correspondences for *splicing.*

4.1.8. DNA fingerprinting

DNA fingerprinting is a forensic technique used to identify individuals by characteristics of their DNA. This profile is a small set of DNA variations that is very likely to be different in all unrelated individuals, thereby being as unique to individuals as are fingerprints. In this case, primary term creation is based on a metaphor taken from forensic science.

Spanish correspondences for the source language term are *huella dactilar genética, huella genética,* and *perfil genético.* The first two equivalents were the most frequently used, probably because both are close semantic calques that respect the term formation process in the source language.

The only difference between *huella dactilar genética* and *huella genética* is that the second is a clipped version of the first. More specifically, *huella dactilar* [fingerprint] is shortened to *huella* [print (i.e. of a finger/paw)]. The second contrast is between the translation of *fingerprinting* as *huella* [print (i.e. of a finger/paw)] or *perfil* [profile]. In all of the Spanish terms, *DNA* is generalized and translated as *genética* [genetic].

The preference in our corpus as well as in the Internet texts was for *huella genética* probably because it is a shortened version of the semantic calque of the source language term. It is also the correspondence in the Spanish version of Wikipedia. *Perfil genético* [genetic profile] had the lowest frequency. As is well known, *profile* is a representation of something in outline, especially a human head or face represented or seen in a side view. A possible reason for its lack of popularity is thus its semantic distance from the source language term.

4.2. Discussion

As can be observed in the results of this study, the Spanish correspondences of the source language terms had different motivations. Parallel and competing equivalents of the same source language term reflect tendencies in secondary term formation that are worth exploring in greater depth. The main types of secondary term formation were the following:

- Adoption of the source language term as a loanword in the target language (*contig, splicing, EST, imprinting*).
- Combination of a source language term with a target language term (*secuencia EST, secuencia ESTs*).
- Semantic calque of the source language term (*ADN, ARN, etiqueta de secuencia expresada, sobrecruzamiento, huella (dactilar) genética*).
- Morphological adaptation of the source language term to word formation patterns in the target language (*cóntigo, delección, entrecruzamiento*).
- Semantic adaptation of the source language term (*perfil genético, marcaje genómico, marcador de secuencia expresado*).

• Explicitation of a different meaning facet (*corte y empalme, recombinación*).
• Metaphoric extension (*ayuste*).

The first correspondence that initially arises in the target language seems to be the source language term, used as a loanword. This tendency is probably based on expediency because for practical reasons, it is quicker and less work to transform the original term into a loanword. In specialized texts, whose function is to disseminate research results in the target language, the direct use of the English term is not regarded as important because the target reader group understands English. There is thus no need for greater semantic transparency.

However, competing target language correspondences seem to arise when texts are written whose function is to explain these concepts to target language readers with less shared knowledge. As shown in our corpus examples, in such texts, when the source language term appears, it is invariably accompanied by one or more correspondences in the target language. This 'explicitation strategy', implemented for the sake of semantic clarity, can even extend to the structure of the term itself (e.g. *secuencia EST*) in which a target language word (*secuencia*) appears alongside the source language term (*EST*). This strategy in texts as well as in term structure, not only reflects the lower knowledge level of the readers, but may also be an indication of term instability, particularly when more than one correspondence is given.

Competing correspondences in the target language are also motivated by other secondary term formation strategies, such as semantic calques. Generally speaking, correspondences that were more semantically distant from the source language term (e.g. *marcaje genómico* [genomic imprint] *perfil genético* [genetic fingerprint]) were found to be less popular than those that were semantically closer (e.g. *impronta genómica* [genomic imprint], *huella genética* [genetic fingerprint]).

However, the popularity of a target language correspondence is not only influenced by semantics but also by the structure of the target language. Certain potential equivalents remain unstable because they do not easily conform to the morphological patterns of the target language (e.g. *cóntigo* [contig], *sobrecruzamiento* [crossing over]).

In contrast, other equivalents are favored because they are more in consonance with target language morphology. This is the case of *entrecruzamiento* [crossing over]. *Entrecruzamiento* is the nominalization of the verb *entrecruzar* [to cross (between)], which competes with *sobrecruzamiento* [crossing over], However, *sobrecruzamiento,* though a closer semantic calque, is less frequent, probably because it is false derivation, given that the verb **sobrecruzar* does not exist in Spanish. Still another example of the impact of the target language structure on secondary term formation is *delección* [deletion], where an erroneous correspondence is regarded as acceptable because of its similarity to word formation patterns in Spanish.

Finally, competing target language correspondences can also arise for other reasons. Especially in didactic texts, where semantic transparency is a priority, it is also possible to find target language correspondences, whose structure reflects an explicitation of different facets of term meaning (e.g. *corte y empalme* ['cut and join' for splicing]). One of the other target language correspondences for *splicing* is *ayuste*, which like the source language term, is also based on metaphorical extension. However, given its semantic opacity, it was only found in the specialized texts for experts. This seems to indicate that the most elegant correspondences that respect primary term formation processes are not always the most popular and do not necessarily find favor with text receivers with a lower level of shared knowledge.

5. Conclusions

English is the *lingua franca* of scientific communication. Many terms coined in English are not easy to adapt and assimilate into Romance languages, such as Spanish. Reasons for this include differences in the structure of both the source and target languages as well as the tendency of experts to disseminate their research exclusively in English. This can result in domain loss, as reflected in in the systematic use of calques and multiple terminological correspondences that do not conform to the rules of term formation in the target language.

In relatively new fields, such as Genetics, in which cutting-edge research is being performed, experimental findings lead to a wide variety of new concepts, which need to be named in other languages. This can often lead to a proliferation of target language terms. This research analyzed a set of representative terms in the field of Genetics, all of which have multiple and competing correspondences in Spanish. Based on the results obtained, seven secondary term formation strategies were identified, and possible reasons were given for implementing one instead of the other in different contexts.

Although a more in-depth study with a larger inventory of terms is necessary to confirm the results, our findings indicate that factors that lead to the use of one strategy over the others include expediency, the desire for semantic transparency, and the morphology of the target language. In this regard, the knowledge level of the target language users also motivated the choice of target language correspondence.

Acknowledgements

This research was carried out within the framework of project FF2014-52740-P, Cognitive and Neurological Bases for Terminology-enhanced Translation (CONTENT) funded by the Spanish Ministry of Economy and Competitiveness.

REFERENCES

Björkman 2015: Björkman, B. Attitudes towards English in University Language Policy Documents in Sweden. – In: English in Europe: Opportunity or Threat? Vol. 1: Attitudes towards English in Europe. Ed. by A. Linn, N. Bermel, G. Ferguson, and C. Hadjidemetriou. Berlin – New York: Mouton de Gruyter, 2015, pp. 115–138.

Fernández-Domínguez 2016: Fernández-Domínguez, J. A Morphosemantic Investigation of Term Formation Processes in English and Spanish. – Languages in Contrast, 2016, № 16 (1), pp. 54–83.

Fishman 1972: Fishman, J. A. The Relationship between Micro-and Macro-sociolinguistics in the Study of Who Speaks What Language to Whom and When. – In: Sociolinguistics. Ed. by J. B. Pride, and J. Holmes. Harmondsworth: Penguin, 1972, pp. 583–603.

Goatly 1997: Goatly, A. The Language of Metaphors. London: Routledge, 1997.

Grevy 1999: Grevy, C. Informationsmotorvejen og andre metaforer I computerfagsprog. – Hermes, № 23, pp. 173–201.

Haberland 2005: Haberland, H. Domains and Domain Loss. – In: The Consequences of Mobility. Ed. by B. Preisler, A. Fabricius, H. Haberland, S. Kjærbeck, and K. Risager. Roskilde: Roskilde University, 2005, pp. 227–237.

Hultgren 2013: Hultgren, A. K. Lexical Borrowing from English into Danish in the Sciences: An Empirical Investigation of 'Domain Loss'. – International Journal of Applied Linguistics, 2013, № 23 (2), pp. 166–182.

Hultgren 2014: Hultgren, A. K. Whose Parallellingualism? Overt and Covert Ideologies in Danish University Language Policies. – Multilingua, 2014, № 33 (1–2), pp. 61–87.

Humbley 2006: Humbley, J. Metaphor and Secondary Term Formation. – In: Cahiers du CIEL 2000–2003: La métaphore: du discours général aux discours spécialisés. Paris, 2006, pp. 199–212.

Humbley and García-Palacios 2012: Humbley, J., J. García-Palacios. Neology and Terminological Dependency. – Terminology, 2012, № 18 (1), pp. 59–85.

Kachru 1994: Kachru, B. B. The Speaking Tree: A Medium of Plural Canons. – In: J. A. Alatis (ed.) Educational Linguistics, Crosscultural Communication, and Global Interdependence (Georgetown Round Table on Language and Linguistics, 1994). Washington DC: Georgetown University Press, 1994, pp. 6–22.

Kilgarriff et al. 2014: Kilgarriff, A., V. Baisa, J. Busta, M. Jakubicek, V. Kovar, J. Michelfeit, P. Rychly, V. Suchomel. Sketch Engine: Ten Years On. – Lexicography ASIALEX, 2014, № 1 (1), pp. 7–36.

Lauren et al. 2002: Lauren, C., J. Myking, H. Picht. Language and Domains: a Proposal for a Domain Dynamics Taxonomy. – LSP & Professional Communication, 2002, № 2 (2), pp. 23–29.

Montero-Martínez et al. 2001: Montero-Martínez S., P. Fuertes-Olivera, M. García de Quesada. The Translator as 'Language Planner': Syntactic Calquing in an English-Spanish Technical Translation of Chemical Engineering. – Meta: Journal des traducteurs/Meta:Translators' Journal, 2001, № 46 (4), pp. 687–698.

Navarro 2008: Navarro, F. Genético, génico o genómico. Available at: http://medicablogs.diariomedico.com/laboratorio/2008/07/10/genetico-genico-y-genomico/

Newmark 1988: Newmark, P. A Textbook of Translation. New York: Prentice Hall, 1988.

Phillipson 2008: Phillipson, R. Lingua Franca or Lingua Frankensteinia? English in European Integration and Globalization. – World Englishes, 2008, № 27 (2), pp. 250–267.

Sager 1990: Sager, J. A Practical Course in Terminology Processing. Amsterdam – Philadelphia: John Benjamins, 1990.

Sager 1997: Sager, J. Term Formation. – In: Handbook of Terminology Management. Ed. by S. E. Wright, and G. Budin. Amsterdam – Philadelphia: John Benjamins, 1997, pp. 25–41.

Swales 1997: Swales, J. English as Tyrannosaurus Rex. – World Englishes, 1997, № 16, pp. 373–382.

Tardy 2004: Tardy, C. The Role of English in Scientific Communication: Lingua Franca or Tyrannosaurus Rex? – Journal of Englishfor Academic Purposes, 2004, № 3, pp. 247–269.

Temmerman 2013: Temmerman, R. Primary and Secondary Term Creation and the Process of Understanding. Presentation given at Application of Cognitive Terminology Theories in Terminology Management. Zagreb, 27–28 September 2013. Available at: http://struna.ihjj.hr/uploads/actt/Temmerman_Zagreb_2013.pdf

Valeontis 2004: Valeontis. K. The Analogue Rule – a Useful Terminological Tool in Interlingual Transfer of Knowledge. Paper presented at the 2nd Terminology Summit. Barcelona, 26-27 November 2004.

Valeontis and Mantzari 2006: Valeontis, K., E. Mantzari. The linguistic dimention of terminology: principles and methods of term formation. Paper presented at the 1st Athens International Conference on Translation and Interpretation Translation: Between Art and Social Science. Athens, 13-14 October 2006.

Volanschi and Kübler 2011: Volanschi, A., N. Kübler. The Impact of Metaphorical Framing on Term Creation in Biology. – Terminology, 2011, № 17 (2), pp. 198–223.

Widyalankara 2015: Widyalankara, R. C. The Expediency Hypothesis and Modes of Secondary Term Formation in ICT Terminology in Sinhala: A Bilingual Corpus Analysis. – International Journal of Scientific and Research Publications, 2015, № 5 (10), pp. 1–8.