Lexical variation, register and explicitation in medical translation

A comparable corpus study of medical terminology in US websites translated into Spanish

Miguel Ángel Jiménez-Crespo and Maribel Tercedor Sánchez
Rutgers University / Universidad de Granada

Differences in register, lexical use, syntactic shifts or determinologization strategies between source and target medical texts can produce usability or comprehensibility issues (Askehave and Zethsen 2000a; Tercedor and López 2012; Nisbeth Zethsen and Jensen 2012; Alarcón, López-Rodríguez, and Tercedor 2016). This study analyzes differences in lexical variation between translated and non-translated online medical texts resulting in potential register shifts, also known as “register mismatches” (Pilegaard 1997). The study uses a corpus methodology to compare (1) the frequency of Latin-Greek (LG) terms in translated medical websites in the USA and in similar non-translated texts in Spain and Latin America, and (2) the frequency of determinologization and explicitation of LG terms in both textual populations. The results show that US medical websites translated into Spanish show lower frequencies of LG terms and higher frequencies of reformulation strategies than similar non-translated ones; they are partly explained through the process of interference from source texts.

Keywords: medical translation, lexical variation, medical terminology, corpus-based translation studies, explicitation

1. Introduction

Medical Translation represents one of the most important areas of practice in professional translation (Hurtado Albir 2000; Montalt Resurrecció and González Davies 2007). It is also one of the main areas of research within TS due to its high volume of translation worldwide (Montalt Resurrecció 2011). In this context, this study follows theoretical and methodological principles of corpus-based translation studies to study lexical variation, determinologization and explicitation resulting
in register shifts between original and translated Spanish medical texts. This study is inspired by previous work on translated medical genres between Danish and (i.e., Askehave and Zethsen 2003a; Zethsen 2004; Raynor 2007; Nisbeth Jensen and Zethsen 2012; Nisbeth Jensen 2013). The basic premise behind these studies is that some of the most common medical genres addressed at lay people need to match the expectations of the audience to guarantee efficient and effective communication. This is often presented as the main goal of all textual genres in which the communicative situation involves a transfer of knowledge from expert to layman. The adjustment of the level of register expected by target audiences represents a main issue in medical translation, since register considerations between languages is so important given that “difference in usage register can trigger variations in terminology and style” (Wright 2011: 246). One of the main features in studying the differences in register between original and translated texts is the distribution and patterns of use of Latin-Greek (LG) terms, as well as their explicitations and “determinologization” or reformulation strategies, such as lowering the register (Meyer and Mackintosh 2000). Differences in these two features result in what is known as “register mismatches” between original and translated texts, that is, when register differences between languages in certain communicative situations or textual genres are incorrectly addressed in translation (Pilegaard 1997: 171). For example, medical LG terms that might be totally acceptable to a general audience in French and Spanish might be associated with higher register in general informative texts in English (Montalt Resurrecció and González Davies 2007: 242). Similarly, some terms that are totally acceptable in both English and Romance languages, such as “coagulation” or “appendicitis,” might be of too high a register if used in Nordic languages, i.e., Danish or Swedish, which prefer Nordic stems for the formation of compound medical terms (Nisbeth Jensen and Zethsen 2012).

2. Register, lay-friendliness, and determinologization in medical translation

The general public has a growing interest in health-related issues. The shift toward a patient-oriented model has resulted in a growing number of materials specifically designed for patients. In English to Spanish translation, medical documentation, medical devices and general medical information are key areas of translation practice. Thus, attention beyond expert-to-expert communication scenarios is much needed in the field of specialized translation in general, and in medical translation in particular. The growing interest in health and disease has resulted in an increase in the volume of medical translation within new communicative contexts.
and genres, as well as a desire to increase health literacy through lay-friendliness and initiatives, such as the plain language movement. In this context, it is widely acknowledged that a similar genre, such as a patient guide for a medical condition, presents variations of structure, register, terminology, tenor, and style between languages (Montalt Resurrecció and González Davies 2007: 157). In the case of texts addressed to lay audiences, certain observed differences can be related to the patient-centeredness paradigm (Putnam and Lipkin 1995; Tercedor 2017) and the subsequent preference for lower register terminological variants.

Even though LG terms are commonplace in medicine in many Western languages (Gutiérrez Rodilla 2014), the extent to which they coexist with vernacular terminological variants varies from language to language. Whereas in many scenarios English prefers a non-LG lay form, Spanish may make use of an LG term in the same contexts. For example, this is the case of the term “ear-nose-throat doctor” or “ENT,” which can only be rendered in Spanish using the LG term “otorrinoslaringólogo,” or by the shortened conversational and informal “otorrino,” which, however, still has LG roots. The reason behind the differences in register is that Latin was not incorporated to the same extent in all European languages (Zethsen 2004: 132). Thus, while Spanish, Italian and French medical terminology is eminently Latin and Greek in origin, Northern European languages possess a layered medical terminology in which many scientific LG words have lay or lower register counterparts; see, for example, the use of doublets in English such as “clotting” and “coagulation,” or “scar” and “cicatrization,” as opposed to the use of Latin-only word formations in Spanish, such as “coagulación” and “cicatriz,” respectively. Both these words are used in both highly specialized and general texts without distinction. Therefore, what “in Latin-based languages might sound too low a register is perfectly acceptable as scientific terminology in English” (Montalt Resurrecció and González Davies 2007: 242). Thus, in English one of the most common ways in which synonymy occurs in scientific and medical domains is through the co-existence of a technical term with its less specialized equivalent, as with “cephalalgia” and “headache.” It is often understood that these cases of synonymy are “a source of translation problems because languages are not symmetrical in their use: for example, what in Spanish is considered to be low register may be perfectly acceptable in English in the same text genre” (Montalt Resurrecció 2011: 80). One of the strategies to deal with this issue involves what is known as determinologization or reformulation in intralinguistic communication, and explicitation in translation processes, the topic of the next section.
2.1 Determinologization in medical texts

It is generally acknowledged that the process of reformulation, also known as determinologization, is one of the most frequently employed strategies at the lexical level in order to adapt textual genres to a non-expert readership. It involves using general language to communicate the meaning of a specialized term (Meyer and Mackintosh 2000), helping to close the gap between specialized knowledge and lay audiences. Montalt Resurrecció and Shuttleworth (2012: 16) refer to this notion as:

a process of recontextualisation and reformulation of specialized terms aiming at making the concepts they designate relevant to and understandable by a lay audience. This process is motivated by specific cognitive, social and communicative needs, and takes place as part of a broader process of recontextualisation and reformulation of discourse [...].

This process appears in medical texts both in intralinguistic translation, such as the case of research articles summarized for laymen in the Annals of Internal Medicine (Muñoz-Miquel 2012: 200–202) or also in translated texts for general audiences (Tercedor and López-Rodríguez 2012). It is also commonplace in intergeneric translation (Askehave and Zethsen 2000a, 2003), the translation of an expert textual genre for a general audience, such as the conversion of summaries of product characteristics for patient information leaflets in the pharmaceutical domain (see also Ezpeleta 2012 for Spanish studies). This latter process involves the use of lower register terminological variants and general language expressions in order to render and convert highly specialized texts into texts addressed to laymen.

Determinologization involves a large number of potential strategies that are covered under this hypernym, such as explanation, definition, reformulation, exemplification, illustration, analogy, comparison and substitution by a more popular term (Campos Andrés 2013; Montalt Resurrecció and González Davies 2007: 252–253). According to Montalt Resurrecció and González Davies (ibid), this process can involve a number of strategies described below:

1. Retaining the LG term but following it with an explanation, such as “poliuria, aumento de la cantidad de orina” [polyuria, increase in the volume of urine]
2. Retaining the LG term in parenthesis after explanations, such as “aumento en la cantidad de orina (poliuria)” [increase in the volume of urine (polyuria)]
3. Retaining the LG term immediately after a popular term, such as “orinar mucho y con frecuencia o poliuria” [urinating heavily and frequently or polyuria]
4. Avoiding LG terms altogether or replacing them with explanations or popular terms “patients can experience an increase in the volume of urine.”
All these mechanisms can help increase the readability and efficiency of translated medical texts for laymen, but they also relate to one of the main general tendencies of translation, “explicitation.” If reformulation or determinologization represent a natural mechanism in intralingual and intergeneric translation in medical genres, the translation process can potentially increase or decrease the frequency and nature of the explicitation strategies present in these texts. The study of reformulation or determinologization strategies in translated medical texts resulting in terminological variation and potential register shifts requires the corpus-based analysis of the compounding effect of reformulation as an intralinguistic translation and/or intergeneric translation (Ezpeleta 2012) and the potential tendency of translated texts to exhibit higher levels of explicitation than non-translated texts. In order to analyze the relationship between intralinguistic determinologization processes and the notion of “explicitation” in TS, a closer analysis is required.

2.2 Corpus-based studies and general tendencies of translation: Explicitation and interference

Among the main research trends of interest in corpus-based translation studies and one of the theoretical issues of interest in this paper is the relationship between (1) reformulation or determinologization in semi-specialized and non-specialized texts addressed to laypeople in intralingual and intergeneric translation processes (also referred to as general “explicitation”) (Askehave and Zethsen 2000a, 2003; Ezpeleta 2012), and (2) “explicitation” as a “general feature of translated language” (Chesterman 2004). This latter notion is understood as the byproduct of a distinctive cognitive process of language mediation rather than the direct production of content in any given language (Krüger 2013). These two different levels go hand in hand both in the intralingual and interlingual translation of the genres under study, namely the medical information websites addressed to laymen. “Explicitation” is thus understood as a process that emerges in all instances of language mediation (i.e., Blum-Kulka 1986: 19), and therefore, explicitation and determinologization resulting in patterns of lexical variation are a byproduct of intralingual translation when experts produce medical genres that are easier to understand for laymen in any language (Muñoz-Miquel 2012; Hill-Madsen 2015). Nevertheless, it is argued here that there will be differences in the patterns of explicitation in intralingual translation (and its subtypes such as intergeneric translation) across languages, in tune with the differences in register, lexical patterns, genre structures, etc. that can appear when translating medical genres (Montalt Resurrecció and González Davies 2007).
The translation process in expert-to-layman genres in medical settings may produce an additional level of explicitation: the explicitation or determinologization of medical terms that already exist in source texts can be multiplied through what is known as “translation-inherent explicitation” (Klaudy 1998). This can be attributed to the nature of the translation process itself and can add another layer of explicitation in all types of texts and genres that are the product of the intralingual translation of an expert-to-expert genre in order to produce an expert-to-layman text, such as with popularized science articles, patient information leaflets or medical information guides (i.e., Zethsen 2004; Nisbeth Jensen and Zethsen 2012).

Thus, since this study employs a comparable corpus methodology and the English source texts were not compiled, it is impossible to identify to what extent any potential differences in explicitation will be the result of the interplay of the following levels:

1. The determinologization and explicitation present in the initial intralinguistic process of writing the source medical text addressed to laymen, and
2. The potential explicitation (or implicitation) resulting from the translation process.

It should be mentioned here that the opposite of explicitation, that is, implicitation, has been recommended in the translation of medical texts from English into Spanish in the genres under study so as to match the register of medical texts originally produced in Spanish (Campos Andrés 2013: 53). In the case of medical patient guides, Campos Andrés (2013: 53) indicates that when translating this genre into Spanish, it is not always necessary to use doublets that might appear in source English texts.

This means that in this language direction it is recommended that some doublets of LG terms be eliminated or reformulated in the translation into Spanish, since many of these terms are considered general language terms, used in everyday communication and perfectly understood by lay audiences. An example would be the translation of the English segment “an ophthalmologist, the eye doctor,” which in Spanish can be adequately rendered by using “el oftalmólogo” alone. Also, in many instances and as previously mentioned in the case of “otorhinolaryngologist” and “ENT,” Spanish provides only the LG term when English has the doublet, as with the use of “coagulation or clotting,” which in Spanish is rendered as “coagulación.” TS literature also recommends studying explicitation in conjunction with implicitation (Becher 2011). While some cases might be compulsory or related to a specific language pair, others might be optional for the translator, and therefore represent instances of what could be called “translation-inherent implicitation.”
Therefore, despite the fact that the hypotheses that will be described in the next section involve an expectation of different levels of explicitation in translations, the specifics of the Spanish-English language combination should not leave aside the potential for implicitation.

3. Empirical study

Following the objectives and literature review previously described, this empirical study uses a corpus-based methodology to identify patterns of variation in lexical use between translated and non-translated texts, and in doing so it investigates issues of register, determinologization, explicitation and terminological variation in translated texts. The general hypotheses for the study are the following:

1. Translated medical texts from English into Spanish for a general audience will show different frequencies of LG terms when compared to similar Spanish non-translated texts due to interlinguistic differences at the lexical level between English and Spanish.

2. LG terms in translated medical texts from English into Spanish for a general audience will show differences in their rates of reformulation or explicitation since it is assumed that translated products will show a compounding effect of general reformulation strategies in intergeneric translation and the presence of “translation-inherent” explicitation.

3.1 Methodology

This study uses a comparable corpus methodology, that is, “a structured electronic collection of texts originally written in a particular language, alongside texts translated into that same language” (Baker 1995: 234). Such corpora have been used to test what is known as T-Translation tendencies, such as simplification (Laviosa 2002) or conventionalization (Baker 1993). They have also been used to study “explicitation” (Jiménez-Crespo 2011, 2016), a so-called S-Translation feature that is normally studied by means of contrasting the target texts to the source texts using parallel corpora, such as in Blum-Kulka (1986) and Krüger (2013). The corpus used in the present study is the Translational Web Corpus of Medical Spanish, TWCoMS (Jiménez-Crespo 2014), a corpus project conceived as a tool to study medical translation in the United States. In addition, the VariMed database (varimed.ugr.es), a repository of more than 1200 concepts with more than 5000 lexical variants in medicine, is used to randomly retrieve lexical units for our study (see 4.2).
3.2 The TWCoMS corpus

This corpus contains approximately 40 million words from medical information websites addressed to general audiences in the US and a comparable selection of websites from Mexico and Spain. The translational section was compiled in 2013/2014 and the non-translational one in 2015. In order to provide a representative snapshot of the textual population under study, namely “medical information websites,” the main criterion for the textual selection and corpus compilation was the notion of “textual genre” (Ezpeleta 2012). The corpus is thus composed exclusively of a web genre, i.e., a type of genre produced exclusively for the web (Jiménez-Crespo 2013: 66–100). Web genres display distinctive features from similar genres that were produced for print (Kennedy and Shepherd 2005; Jiménez-Crespo 2009, 2013). For example, the medical information genre exemplar of MedlinePlus represents a distinct genre when compared to the medical guides for patients that can be found in print in hospitals or clinics.

Many medical genres, such as medical histories, medical forms, informed consents, patient information leaflets, and clinical guides, online genres, such as medical information websites, online medical forums, medical company websites or medical practice websites have recently emerged as the main source of information for the general public, and thus an important sector of translation. In line with Volansky (2012) and Mauranen (2008), this paper argues that the study of general tendencies in translation needs to be reviewed in the context of specific genres and registers, since results might show differences depending on the language combination and the genre and register in question. The genre with the highest volume of translation around the world is the medical information website, a modern update to the classic patient health information leaflet. It should be mentioned that the multiplicity of functions in websites, as suggested by Jiménez-Crespo (2013: 86), is due in part to the fact that websites are normally “complex genres,” that is, genres that include other genres in their structure. For example, medical websites can include forums, medical information texts, journalistic articles, legal texts, forms or advertisement. This phenomenon is known as “genre embedding” (Martin 1995: 24). Thus, in order to exclude from the corpus textual components initially developed for print distribution, all pdf and word files were excluded in the compilation process.

The TWCoMS comprises two interrelated subcorpora in Spanish. One of them includes medical websites or portals in the United States translated into Spanish and another subcorpus of similar websites produced originally in Spain, Mexico or for a general Latin American audience. The following is a description of the two subcorpora:
1. The translational subcorpus contains 32,330,052 tokens. It comprises four distinct subsections of texts that could be considered cases of “intrasocial translations,” that is translations addressed to members of the same society speaking another language, such as the case of Spanish speakers in the US. The four sections include (1) US government websites (i.e., Center for Disease Control, Women’s Health); (2) websites with medical information from the different Departments of Health at the state level; (3) general medical websites from national organizations (i.e., MedlinePlus); and (4) medical history forms available on the Internet, a category that can be used for contrastive purposes (González-Darriba 2014).

2. The non-translational subcorpus contains medical websites originally produced in Spanish, such as Mapfresalud (Spain), Universomedico (Mexico) and Geosalud (Latin America). The comparable section contains 8,701,867 tokens. The compilation process was carried out using the open source website downloader HTTrack (https://www.httrack.com/). The corpus was tagged by including information in the header automatically in each page with HTTrack in the downloading process. In a second level of processing the entire corpus was converted to plain text format using the software tool Detagger (http://www.jafsoft.com/detagger/). This means that the entire corpus is kept in two formats, the downloaded html format and the txt one. It is important to highlight that the corpus contains full websites, that is, the entire structure of the website is downloaded and analyzed. Generally, web corpora compiled using other tools, such as WebBootCat or Sketchengine (Kilgarriff et al. 2014), lack significant textual sections such as navigation menus or other recurring textual components that appear in all websites. The difference between both methods of compilation entails serious implications for later analyses since recurring segments in navigation menus, for example, represent textually specific concepts of interest that are highlighted by producers in the macrostructure of the genre. This repetition can, depending on the analyses carried out, bias the end results by distorting statistics; on the other hand, it can contribute to our understanding of the overall lexical content of the website.

3.3 Materials

For the purposes of the analysis and to operationalize the hypotheses, thirteen terms were randomly selected from the VariMed terminological variation website (varimed.ugr.es) from an initial list of 100 concepts with LG forms in the database:
1. Dysmenorhea (en) – dismenorrea (es): painful periods
2. Dyspnoea (en) – disnea (en): difficulty breathing
3. Halitosis (en) – halitosis (es): bad breath
4. Hematuria (en) – hematuria (es): presence of blood in the urine
5. Hypoglycemia (en) – hipoglucemia (es): low blood sugar
6. Hysterectomy (en) – histerectomia: surgical removal of the uterus
7. Myalgia (en) – mialgia (es): muscle pain
8. Polydipsia (en) – polidipsia (es): excessive thirst
10. Rhinorrhea (en) – rinorrea (es): nasal dripping
11. Tachypnea (en) – taquipnea (es): fast breathing
12. Xerosis (en) – xerosis (es): dry skin
13. Xerostomy (en) – xerostomía (es): dry mouth

All these LG terms are often paired with a range of lexical variants in order to express the same concept using reformulations with different degrees of lexicalization or fixation. Even though many of these forms are recorded in VariMed, this study revealed a wider range of reformulation possibilities, including a set of reformulations that only appear in translated texts and therefore, could be the result of interference or the literal rendering of lexicalized reformulations in the source texts in English (see Section 4.3).

Figure 1. Snapshot of concordance lines in WordSmith Tools for the term “disnea.” Example taken from the Non-Translational corpus

3.4 Lexical analyses

In order to test the working hypotheses, two different analyses were conducted using the lexical analysis tool WordSmith Tools 6.0 (Scott 2012). The first analysis used the wordlist function and it involved statistically establishing the relative frequency of each LG term in each section of the corpus. This relative frequency is obtained automatically by WordSmith Tools by dividing the instances of the chosen lexical items by the total token count in each subsection of the corpus. These general
relative frequencies in the wordlist of corpora are a common measure to observe the distribution of lexical items in corpora.

The second analysis involves a mechanism to identify the potential lay-friendliness of these medical information websites: the potential use of doublets formed by the scientific LG term followed by reformulation or determinologization. This mechanism corresponds to the first three determinologization strategies indicated above (Montalt Resurrecció and González Davies 2007: 252–252), such as retaining the specialized term followed by an explanation, retaining the specialized term in parenthesis after an explanation or reformulation or retaining the specialized term after a popular term. This analysis of reformulation was conducted manually using the concordance feature of WordSmith Tools 6 by checking the concordance lines for each selected LG term and noting whether it was accompanied by determinologization. Each concordance line with a reformulation strategy was recorded in a spreadsheet and the percentage of use reformulation strategies were extracted. The normalized frequency per 100,000 words for each term was calculated dividing the occurrences of each term by the total number of tokens in each corpus and multiplying by 100,000. Statistical analysis was performed using the one-tail Man-Whitney U-test for non-parametric independent samples in GraphPad Prism 5.0.

4. Results

4.1 Contrastive analysis of frequency of use of Latin-Greek terms

This analysis involves the contrastive study of frequency of overall use of LG terms between both subsections of the TWCoMS corpus. Figure 2 shows a contrastive analysis of the relative frequencies of use of the selected terms in both subcorpora. The red lines represent the normalized frequencies per 100,000 words in the translational corpus, while the blue bars indicate the normalized frequencies per 100,000 in the non-translational corpus. It can be observed that LG terms show a higher frequency of use in the texts originally produced in Spanish. In the case of “hysterectomy” (rel. freq in translational corpus = 2.524; rel. freq in non-translational corpus = 3.850), while the frequency of use is higher in the non-translational corpus, the difference is not significant. This could be a bias introduced by the fact that the translational corpus includes a federal US website addressed exclusively at women, Womenshealth.org, while all the websites in the non-translational subcorpus are generalist in nature.
Since both subsections of the corpus are representative of the textual population targeted and both include similar genres produced under different conditions, it is assumed that the relative frequency of the concepts should in principle be similar between both textual populations. The differences identified in the percentage of frequency of these LG terms when both subcorpora are contrasted are therefore indicative of a lower presence of these terms in translated texts, most likely linked to the overall lower presence of LG terms in source English medical texts addressed to lay audiences. This finding would suggest, as others have indicated, that English and Spanish are not symmetrical in the use of these terms.

Figure 3 includes a contrastive analysis of the combined frequency of use for all terms. The table includes both the overall frequency of these terms in the corpus as well as the average frequency for all terms. The addition of all frequencies for all terms yields a combined presence in the wordlist of 0.356 in the translational corpus, while this number is 2.906 in the non-translational corpus. Statistical analysis shows that the occurrence of LG terms in the translational corpus is significantly lower than in the non-translational corpus.

Figure 3 shows that the average frequency for LG terms in each corpus is 0.075 per 100,000 tokens for the translational corpus, whereas it is 0.455 for the original non-translated one. Thus, translated medical texts from English into Spanish and addressed to lay audiences statistically display a much higher frequency of use of LG terms. This lower presence in translations of LG terms would in principle represent the opposite result of the studies in the English-Danish combination.
where the direct transfer of LG terms hindered comprehension and made some medical texts “incomprehensible to many Danish laymen” (Askehave and Zethsen 2000b: 57). The results from these analyses confirm the first hypothesis relating to translated texts showing lower frequency of LG terms when compared to similar non-translated texts.

Does this mean then that these texts might be easier to understand for the laymen than non-translated Spanish texts would be? As previously described, research in English into Spanish translation has indicated that LG terms are on average understandable to laypeople and therefore, a suggested strategy is to delete the use of the frequent doublets in English, such as “myalgia, pain in the muscles,” because laymen in Spanish can clearly understand the LG terms and therefore their presence would not raise the register of the text as it would in English (Campos Andrés 2013: 53). This would in principle mean that not only could texts translated into Spanish include the LG term when the source language text in English uses the general language one, but also that the English source texts would display a higher percentage of reformulation or determinologization strategies when compared

Figure 3. Mean (SD) of the Latin-Greek terms normalized frequencies per 100,000 words, in non-translated and translated corpora with and without term reformulation/explicitation. *** Significant difference ($p < .05$) after Mann-Whitney U-test; $p = .0001$ in non-reformulated and $p = .0002$ in reformulated samples analysis.
Miguel Ángel Jiménez-Crespo and Maribel Tercedor Sánchez

with their Spanish counterparts. The following analysis therefore investigates the frequency of reformulations of LG terms in order to identify whether the lexical norms and expectations of the source genre in English is also directly transferred into the Spanish target texts.

4.2 Reformulation and explicitation in translated medical texts

Figure 4 shows the contrastive analysis of the percentage of term occurrences accompanied by reformulation or explicitation in each subcorpus.

![Figure 4. Percentage of use per term of reformulation-explicitation with Latin-Greek terms in both sections of the corpus](image)

It can be observed that for eleven out of the thirteen selected terms, the translational subcorpus shows a higher percentage of reformulation/explicitation strategies than the non-translational one. The difference in the frequency of use varies from terms such as “myalgia” that is reformulated/ explicitated in 77.7% of cases in the translational corpus and 15.1% in the non-translational, to cases such as “xerostomy” in which the difference is minimal (translational subcorpus = 61.11%, non-translational = 60%). The only two cases in which the frequency is higher in the non-translational corpus are “hysterectomy” (translational subcorpus = 9.31%, non-translational = 14.02%) and “polydipsia” (translational subcorpus = 25%, non-translational = 39.02%). As previously mentioned, in the case of “hysterectomy” this difference could be due to the composition of the corpus, since the translational corpus includes a website exclusively addressing women’s health, but...
the other condition, “polydipsia,” affects men and women equally and so the higher frequency can not be assigned to corpus content or genre-related effects.

If the averages are computed, the overall percentage of reformulation or explicitation in the translational corpus as seen in Figure 5 is 40.6 while for the non-translational corpus it is 21.23. The presence of reformulated LG terms in the translated corpus is significantly higher ($p = .0056$) than in the non-translated corpus, as shown in Figure 5.

![Figure 5.](image)

These two analyses show a considerable difference in the percentage of reformulation or explicitation between both textual populations under analysis. This difference could therefore be explained through two different and potentially complimentary processes taking place at the same time. On the one hand, the previous contrastive analysis of LG terms, in line with the existing literature, has shown that these terms are higher in register in English than in Spanish, and therefore, the possibility is less that these terms will be included in texts addressed to lay readers in English, and consequently through a process of interference in their translations into Spanish. On the other hand, it is unclear the degree to which explicitation, as a general feature of translation, might be occurring. In this sense, there is a double process embedded in translational texts giving rise to a higher percentage of reformulation techniques in source text writing to adapt source texts to lay audiences in English, and (2) the translation process involving the use of explicitation in the interlingual
translation process in cases where the translator might attempt to adjust the register to target audiences and include a reformulation of LG terms. Using a parallel corpus methodology, comparing source and target texts, would allow researchers to identify whether the results are due to intralinguistic or interlinguistic explicitation processes. This will be one of the future objectives in this ongoing research project on the study of variation in medical terminology.

4.3 Analysis of lexical variation related to reformulations

Reformulations or explicitations represent one of the main indications of terminological variation and are identified as pointers to basic relations such as synonymy (also known as, termed) and metalinguistic elements (parenthesis, quotes, commas) (Montero and Faber 2008; Tercedor and López 2012). Therefore, any differences in the overall number and forms of reformulations in the corpus can bear implications for the study of terminological variation. An analysis was conducted in order to reveal whether the higher reformulation levels were also related to higher terminological variation. For each term, the number of terminological variants was recorded and contrasted. Figure 6 shows this contrastive analysis of the different types of reformulations/explicitations in each subcorpus. On average the non-translational corpus shows a higher mean of LG reformulated terms, 9.69 per term, as contrasted with the translational corpus, 6.54 per term. Therefore, it is clear that despite having higher levels of reformulation/explicitation in the translational corpus, the actual reformulations show a wider range of potential reformulations in the original non-translated texts.

As an example, Table 1 shows the range of reformulations for the concept with the highest volume of different reformulations in the corpus, the term “hypoglycemia.” For this term, 17 different reformulations were identified in the translational corpus.
while 24 were identified in the non-translational one. It is of interest that out of 17 different reformulations in the corpus, only three of them are actually used in the non-translational corpus. This means that 80.52% of reformulations-explicitations in the translational corpus, even when they describe the same concept, do not appear in the original texts. Very similar reformulations exist but often the more general term “azúcar” [sugar] is substituted in the non-translated texts for “glucose” [glucose], a higher register variant, a substitution that seems to agree with the overall pattern identified in this study of higher register lexical usage in the non-translated texts as compared to the translated ones.

Table 1. Contrastive analysis of reformulations of the term “hypoglycemia.” The asterisk indicates a reformulation that does not appear in the non-translated texts

<table>
<thead>
<tr>
<th>Reformulations for term: Hypoglycemia</th>
<th>Translational corpus</th>
<th>Non-translational corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td>nivel bajo de azúcar en la sangre</td>
<td>azúcar en la sangre demasiado baja</td>
<td></td>
</tr>
<tr>
<td>azúcar bajo en la sangre*</td>
<td>disminución de los niveles de glucosa en la sangre</td>
<td></td>
</tr>
<tr>
<td>bajo azúcar en la sangre*</td>
<td>baja en el azúcar</td>
<td></td>
</tr>
<tr>
<td>azúcar baja en la sangre</td>
<td>bajada de glucosa</td>
<td></td>
</tr>
<tr>
<td>concentaciones bajas de azúcar en la sangre*</td>
<td>bajo nivel de azúcar en la sangre</td>
<td></td>
</tr>
<tr>
<td>glucemia baja *</td>
<td>un nivel bajo de azúcar en la sangre</td>
<td></td>
</tr>
<tr>
<td>concentración anormalmente baja de azúcar en la sangre*</td>
<td>cuando la concentración de glucosa sanguínea es inferior a 50 mg/dL</td>
<td></td>
</tr>
<tr>
<td>niveles bajos de azúcar*</td>
<td>cuando baja la glucosa en la sangre</td>
<td></td>
</tr>
<tr>
<td>los bajos niveles de azúcar en la sangre*</td>
<td>cuando los niveles de azúcar en la sangre están demasiado bajos</td>
<td></td>
</tr>
<tr>
<td>(baja del azúcar sanguíneo)*</td>
<td>descenso de los niveles sanguíneos de azúcar</td>
<td></td>
</tr>
<tr>
<td>bajo nivel de azúcar en la sangre (baja del azúcar en la sangre)*</td>
<td>descenso del nivel de azúcar en sangre</td>
<td></td>
</tr>
<tr>
<td>baja del nivel de azúcar en la sangre*</td>
<td>disminución excesiva del nivel de glucosa en sangre</td>
<td></td>
</tr>
<tr>
<td>la disminución de los niveles de azúcar en la sangre*</td>
<td>valores de azúcar muy bajos</td>
<td></td>
</tr>
<tr>
<td>(disminución del azúcar en la sangre)*</td>
<td>valores de glucosa en la sangre muy bajos</td>
<td></td>
</tr>
<tr>
<td>niveles bajos de azúcar*</td>
<td>valores muy bajos de azúcar en la sangre</td>
<td></td>
</tr>
<tr>
<td>disminución abrupta del nivel de azúcar en la sangre* nivel bajo de glucosa en sangre</td>
<td>bajada de azúcar en sangre</td>
<td></td>
</tr>
<tr>
<td></td>
<td>bajada de los niveles de glucosa</td>
<td></td>
</tr>
<tr>
<td></td>
<td>descenso excesivo de glucosa en sangre</td>
<td></td>
</tr>
<tr>
<td></td>
<td>disminución de los niveles de glucemia en sangre</td>
<td></td>
</tr>
<tr>
<td></td>
<td>los niveles de azúcar en sangre bajos</td>
<td></td>
</tr>
<tr>
<td></td>
<td>azúcar baja en la sangre</td>
<td></td>
</tr>
<tr>
<td></td>
<td>glucosa baja en sangre</td>
<td></td>
</tr>
<tr>
<td></td>
<td>disminución de glucosa en sangre</td>
<td></td>
</tr>
</tbody>
</table>
The results in this study show that the reformulations tend to follow patterns similar to the source text with a high degree of interference resulting in renderings that do not appear in non-translated texts. For example, the English segment “low blood sugar” is rendered in the translations in order of frequency as “azúcar bajo en la sangre,” “bajo azúcar en la sangre,” and “glucemia baja,” which do not appear among the non-translated texts. A similar rendering appears in non-translated texts, “azúcar baja en sangre,” which do not use the Spanish article “la” in front of the noun “sangre” [blood]. These cases could be considered literal translations that do not appear in non-translated texts. Similarly, cases of interference appear in the following example, “concentración anormalmente baja de azúcar en la sangre,” which represents a direct translation of “unusually low concentration of sugar in the blood.”

In this case, the results could be explained in terms of a double process that translated texts show through interference and literal translation explicitations that might not be necessary in the target language, given that LG terms are more transparent and understandable to the Spanish speaking audience and doublets might not be needed; they may also represent cases of “translation inherent” explicitation, when the translator adds the explicitation in the communicative process of the translation itself.

5. Conclusions

This study was initiated with the interrelated objective of studying register, terminological variation, and lay-friendliness in translated medical texts and investigating general tendencies in translation, such as explicitation, through a comparable corpus analysis of lexical features in original and translated medical websites from English into Spanish. The results of the study confirmed the first hypothesis related to differences in the use of LG terms, with a statistically significant relationship between the type of corpus and the frequency of LG terms. Contrary to what was observed in previous studies in the English to Danish combination, LG terms are more frequent and are more widely present in expert to laymen genres in Spanish that in English source texts. Consequently, translated texts from English into Spanish in expert-to-layman medical genres do in fact display lower frequencies of LG terms than non-translated texts. Similarly, it was put forward that since reformulations and determinologization strategies are mainstream in intralinguistic and intergeneric translations in monolingual settings, the above-mentioned hypothesis would also lead to a second hypothesis, that LG terms in translated texts would display different frequencies of reformulations or explicitation due to interference
or literal translations of source texts. The testing of the second hypothesis showed that translated texts displayed a 40.59% rate of reformulation-explicitation while original texts showed only 21.23%. It was of interest that original texts showed a wider range of potential formulations in the explicitations. Also, reformulations and explicitations found in translated texts were often completely different and not found in original texts.

The study thus raises many interesting questions about whether this lower register level might contribute to lay-friendliness and usability, or whether it might undercut the authority of the text. Additional experimental studies, such as the one that has been designed as a follow up to this study (Jiménez-Crespo 2017) could be designed in order to assess contrastively the usability, understandability or readability of translated and original texts to find out whether the supposedly more lay-friendly patterns of lexical use in translations do in fact result in more suitable texts for the intended communicative purposes.

Acknowledgements

This research is part of the CombiMed Project (FFI2014-51899-R), funded by the Spanish Ministry of Economy and Competitiveness. It is also funded by a grant from the Rutgers University Council Research Grants program.

References


Authors’ addresses

Miguel Ángel Jiménez-Crespo
Department of Spanish and Portuguese
Rutgers University
15 Seminary Place, 5th Floor
New Brunswick, NJ 08901
USA
miguelji@rci.rutgers.edu

Maribel Tercedor Sánchez
Departamento de Traducción e Interpretación, Universidad de Granada
Buensuceso 11
18071 Granada
Spain
itercedo@ugr.es

Biographical notes

Miguel A. Jiménez-Crespo is an Associate Professor in the Department of Spanish and Portuguese at Rutgers University, where he directs the M.A. program in Spanish Translation and Interpreting. He holds a Ph.D. in Translation and Interpreting Studies from the University of Granada, Spain. He is the author of Crowdsourcing and Online Collaborative Translations: Expanding the Limits of Translation Studies published in 2017 by John Benjamins and Translation and Web Localization published by Routledge in 2013. He is the Assistant Editor of the Journal of Internationalization and Localization (JIAL). His research focuses on the intersection of translation theory, translation technology, digital technologies, corpus-based translation studies and translation training.

Maribel Tercedor is full professor in Translation at the Faculty of Translation and Interpreting of the University of Granada, where she teaches multimedia and scientific and technical translation. Her main research interests are in the fields of lexical and cognitive aspects of scientific and technical translation, terminology (variation and phraseology) and accessibility in translation. She is the head of the CombiMed (FFI2014-51899R) research action on lexical combinations in medical texts.