ONOMÁZEIN
Journal of linguistics, philology and translation

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
FACULTAD DE LETRAS

Effects of translation memory software on English-to-Spanish translation: evidence of normalisation in translations by undergraduate trainees

Eduardo Véliz-Ojeda
Universidad de Tarapacá
Chile

ONOMÁZEIN 66 (December 2024): 182-200
DOI: 10.7764/onomazein.66.10
ISSN: 0718-5758

Eduardo Véliz-Ojeda: Departamento de Idiomas Extranjeros, Facultad de Educación y Humanidades, Universidad de Tarapacá, Chile. | E-mail: evelizo@academicos.uta.cl

Received: December, 2020
Accepted: April, 2021

Abstract

This article reports on a translation quasi-experiment aimed at looking at how the textual and grammatical features of non-translated English and Spanish scientific texts are represented in translations from English into Spanish when they are mediated by translation memory software. Senior undergraduate translation trainees from three different Chilean universities translated abstracts for research articles in conservation biology from English into Spanish in a translation memory environment. The translations were then contrasted against a corpus of non-translated texts of the same genre written in English and Spanish, in terms of syntactic dependency relations. The results suggest that student translators, regardless of the use of TM systems, would tend to 'normalise' translation, i.e. they would be overly concerned about conforming to target language norms to the extent of overusing grammatical features.

Keywords: translation memory; corpus-based translation studies; syntactic dependency relations; normalisation; translator training.

1. Introduction

Translation memories (TM) are databases which store chunks of source texts (ST)—usually delimited by punctuation—and their translations in order to speed up the translation process of related-subject texts by "remember[ing] the content of past translations" (Melby and Wright, 2015: 662). The program searches for ST chunks that fully or partially match the text chunk being translated. If a match is found, the translator is provided with a translation suggestion they can either accept or reject; that is, the new translation is being informed by translations carried out in the past by the same or other translators (Carl and Planas, 2020). The reutilisation of previous translations is believed to increase productivity and terminological consistency (Leblanc, 2013); however, the benefits of the use of TM software will be relative to the degree of repetition or matches between the segments stored in the TM and the new text being translated. Indeed, Pym (2020: 439) refers to the assumption of increased productivity of TM as "promotional discourse".

Studies on the effects of the use of TM software (Martín-Mor, 2011; Dragsted, 2006; Martín-Mor, 2019) have concluded that dissimilar translation environments (with and without TM software) and uneven levels of experience (professional and novice translators) yield different translation results. Based on theories of cognitive psychology and working memory, Dragsted (2006) argues that the unnatural focus on the sentence imposed on translators by the use of TM affects translation both as a process and a product. For instance, translators spend more time on revising each segment while translating. Consequently, the time spent on the final revision is reduced, lessening the importance of the post-translation stage. Dragsted (2006) also stresses that post-translation observations of informants on the translation process lead to the conclusion that translators' awareness of TM advantages and disadvantages is a decisive factor in the number of sentence structure alterations in target texts.

Martín-Mor (2011) investigates the effects of the use of TM systems on both the translation process and translated texts, and observes that TM-translated texts show a higher degree of linguistic interference. He agrees with Dragsted (2006) regarding the effects TM may cause on the translation process and observes that the frequency of linguistic interference varies depending on the translation environment in which a TT is produced (with or without TM software). Martín-Mor (2011) contrasts linguistic interference caused by the use of TM in the translation environment against users' expectations, concluding that TMs may affect the readability, cohesion and textual coherence of translations. Nevertheless, in a more recent empirical study with professional translators only, Bundgaard and others (2016: 106) found that TM systems have a "restraining influence" in that "the translator resists the influence of the tool by interrupting the usual segment-by-segment method encouraged by translation technology". At the same time, they argue that technology has an "aiding influence" which contributes to meeting specific translation requirements. Bundgaard and others's findings suggest that translator trainers should strive to develop

in their trainees a level of awareness and cognitive behaviour that is as close as possible to those achieved by more experienced translators.

Earlier in 2005, Bowker warned about the uncritical attitude of translators towards the proposals of TM systems which undermine the quality of the translation. She suggests that more training for novice translators is needed in the appropriate use of translation technology to "question the suitability of the [decontextualized] TM's proposals" (Bowker, 2005: 19). In the same vein, Dragsted (2006) suggests the need for adjusting the segmentation system to agree more with the cognitive translation unit, or the "segment that is limited by the capacity in working memory and identifiable through pauses in the translation process" (Dragsted, 2006: 445), and recommends that student translators should use TM with caution and be made aware of the potential negative effects of translation technology. Additionally, Martín-Mor (2011) proposes further research to determine whether appropriate training for novice translators in the use of translation technology would enable them to avoid the unwanted effects of the use of TM. It can thus be inferred that there is a pressing need to investigate how to make student translators aware of the interference phenomenon and train them to use translation technology wisely, favouring a more natural translation process and reaching better translation standards.

Given the recent developments in translation technology, machine translation and other tools and resources are being integrated into TM systems, obscuring the boundaries between the roles of the translator and the machine, a phenomenon scholars call "the blurring of technologies" (Doherty, 2016; Kenny, 2020; O'Hagan, 2020). Therefore, the discussion appears to have moved on to analysing machine translation and focusing on issues such as pre- and post-editing and, most recently, the development of neural machine translation systems. As a result, scholars are now calling for the inclusion of the development of "technology literacy" (Drechsel, 2019) and "machine translation literacy" (Bowker and Buitrago Ciro, 2019) in the training of translators and interpreters. Despite the development of other translation technologies, TMs are "at the core of CAT [computer-aided translation] tools" (Doherty, 2016: 950) and are, therefore, worth studying in the context of translator training.

In the context described above, the aim of this study is to investigate how the textual and grammatical features of non-translated scientific texts written in English and Spanish are represented in English-to-Spanish translations produced by senior undergraduate translation students using TM software. In order to determine this, the study was framed around the questions: What are the textual features of abstracts for research articles in conservation biology written in English and Spanish? Do student translators replicate the source language (SL) genre features in their translations using TM software? Or do students' translations stray from target language (TL) genre features? The study was developed as part of the author's doctoral research project, and was approved by Monash University's Human Ethics Low Risk Committee.

The structure of the article is as follows: section 2 focuses on the concept of cross-linguistic influence and its relation with translation and translation technology; section 3 describes the methodology and provides a brief summary of a previous study aimed at determining the syntactic dependency features of abstracts for research articles in conservation biology written in English and Spanish, and how the results were compared to the results of the present study; section 4 discusses the findings, and section 5 provides a set of concluding remarks and avenues for future research.

2. Theoretical framework

2.1. Cross-linguistic influence

Linguistic interference has traditionally been associated with the acquisition of a second or foreign language and the learning problems caused by the influence of the learner's mother tongue. Moreover, it has been widely studied in the context of languages in contact, with the focus placed on the linguistic changes that interference might cause in one of the languages in contact (Matras, 2013).

Numerous works up to the present cite Uriel Weinreich's study (1953) on language contact as the classic work on interference and thus attribute the coinage of the term *linguistic interference* to this author. According to Weinreich, two or more languages are in contact when they are used alternately by the same persons, and the primary locus of contact are thus the bilinguals or the individuals using those languages. It is, therefore, in situations of language contact where interference phenomena occur. However, when employed by monolingual speakers of a language, the use of elements previously introduced to the linguistic system by influence of another system would not be considered as interference, since "the consumer of imported goods only rarely has the same awareness of their origin as the importer or investigator" (Weinreich, 1953: 11-12). In this context, Weinreich (1953: 1) defines linguistic interference as "[t]he instances of deviation from the norms of either language which occur in the speech of bilinguals as a result of their familiarity with more than one language; i.e. as a result of language contact".

Although interference can be evident in linguistic production, it derives from the interplay of structural or internal (linguistic) factors and non-structural or external (extralinguistic) factors. Among the structural factors of interference, Weinreich (1953) distinguishes interference at three levels: phonic, grammatical, and lexical. *Phonic* interference would occur when a speaker identifies a phoneme in the secondary language (usually L2) as one pertaining to the primary language (usually L1) and, thus, reproduces this phoneme following the phonetic rules of the primary system. In terms of *grammatical* influence, the author makes a distinction between interference affecting *morphemes* and those modifying *grammatical relations*. *Lexical* interference refers to the various ways in which the vocabulary of

a language can be transferred to another language, or the replication of lexical or semantic functions of an element of one language in another.

Since the present study investigates whether TM software imposes replication of source language syntactic elements in the target language, it is interference at the level of grammatical relations that demands more attention here. Weinreich (1953) asserts that interference of relations can be classified into that which results in the conveyance of unintended meaning, and that in which there is violation of existing patterns, producing nonsensical utterances or speech that might only be understood by implication. Another type of interference would be the use of obligatory relations which do not exist or are not equally compulsory in the receiving system. This third type is rather theoretical in that it cannot be accurately tested, essentially considering that the linguistic elements transferred to the receiving system do exist in this language but can be recognised as foreign only because they sound strange, exotic or monotonous to the ear of a native speaker. By the same token, Kabatek (1997) uses the term *negative interference* to refer to the cases in which elements of the receiving language are missing by influence of another language. The imposition of grammatical relations in a language where these relations do not necessarily occur with the same frequency is of great interest to the purposes of this research project. An example of such an imposition that is embedded in the present study is the replication in Spanish translation of English cohesive devices, such as word repetition, which would prompt the translator to overuse independent (paratactic) sentences and avoid more frequent cohesive devices available in the Spanish language such as subordination.

Hickey (2013) discusses linguistic influence of one system over another in terms of transfer, where, in the context of language switching, bilingual speakers transfer features of their original language to another. Hickey also distinguishes between *supportive transfer*, which occurs when a specific element is present in and pertains to the two languages in contact, and *innovative transfer*, where the presence of an element that is characteristic of one language is present in but not typical of another, constituting an innovation. Hickey's definition of innovative transfer agrees with Weinreich's concept of linguistic interference but interference is perceived by Hickey as bearing an evaluative meaning, especially as applied in second language acquisition. This is reaffirmed by Kellerman (1995), who argues that interference implies an error, whereas transfer, or *crosslinguistic influence*, assumes a non-prescriptive approach to the phenomenon, and by Matras (2013: 67; emphasis in original), who asserts that contact-related change has to do with functional purposes in communication, and therefore "contact phenomena are in this respect seen as *enabling* rather than *interfering* with communicative activity". In addition, Kabatek (1997) considers interference not necessarily as an error but as a technique used by speakers to identify themselves with a community and define their social position through the selection of linguistic variations. Concerning the evaluative load of the concept of linguistic interference, it is convenient here to take a stance on whether the term, as proposed by Weinreich (1953) and revised by

Kellerman (1995) and Matras (2013), among others, is appropriate for the purposes of the present research. The approach taken in the present study considers interference, transfer or cross-linguistic influence, not intrinsically as an error, but as a phenomenon that might eventually hamper communication in translation or have undesired effects on translation reception. The terms preferred in this work are *transfer* or *influence*, unless discussed in association with other authors' propositions.

2.2. Cross-linguistic influence in translation

Toury (2012: 310), in his attempt to establish laws of translational behaviour, refers to linguistic influence in translation as *discourse transfer*, and states that "in translation, phenomena pertaining to the make-up of the source text tend to force themselves on the translators and be transferred to the target text", this constituting the *law of interference*. This assertion therefore implies that interference would be inherent to any translational act. A related concept is that of *universals of translation*, or features characteristic of translations, not found in original texts, that are, however, independent of the influence of the languages involved in translation (Baker, 1993: 243). Pym (2014) identifies an interrelation between these two concepts and regards universals as identified linguistic tendencies which are explained by laws in terms of social, cultural or psychological factors. Although Baker's definition seems to set interference apart from the concept of universals of translation, if the use of technology in the translation environment caused some sort of influence, it would be sensible to closely analyse the implications of such an influence in the translator's cognition and the quality of translation.

Toury (2012: 311) suggests that interference is rooted at mental processes and is "an external manifestation of a general cognitive law" that makes discourse transfer universal. However, the choices made in translation are not the same for all translators nor is the output of translation the same in all cultures, because tolerance of interference is influenced by the socio-cultural conditions under which translation is carried out. The amount of interference observable in a given translation would then be dependent on "whether the source text was approached and processed as *one entity*, a holistic message in an act of communication, or as an *organization of lower-level linguistic entities*" (Toury, 2012: 31; emphasis in original). Therefore, interference is expected to be more pervasive when lower-level features of the ST are taken as the main units of meaning for the production of a translation but less foreseeable if translation is guided by a functional communicative purpose. Yet interference could be so natural to translation that it may even be present in the work of renowned, experienced translators.

Being dependent on socio-cultural factors, resistance to discourse transfer may result in translations showing less manifestations of this phenomenon, which is often resolved in the revision or editing stages of the translation process (Toury, 2012). Nevertheless, not all languages hold the same prestige in a given culture. Hence, interference may be more or less

tolerated based upon the "relative prestige" of the SL in the target culture (Toury, 2012: 314). Another factor that may determine tolerance of, or resistance to, interference is text-type, meaning that certain types of texts, because of their context of production or reception and their users' expectations, may tolerate more or less amounts of linguistic transfer. Finally, Toury (2012: 315) recognises the possibility of interference occurring in varying degrees at different text levels, stating that "even for one and the same text, neither interference nor tolerance of it are necessarily the same with respect to all linguistic and textual levels".

Cross-linguistic influence has also been examined in the context of scientific translation. Franco Aixelá (2009) discusses interference specifically in technical translation, and expands Toury's classification of the phenomenon to the level of genre conventions (structural and pragmatic interference). He stresses the fact that interference has been traditionally regarded as something to avoid in translation for the sake of fluency and correction in the TL. This would apply mostly to technical translation (with the exemption of sworn translation), since, he adds, there have been advocates of "controlled interference" (Franco Aixelá, 2009: 76) in the translation of sacred and canonical texts, and some theorists, such as Venuti (1998), call for the use of foreignising strategies in literary translation to resist the cultural domination exercised by mainstream groups over minority cultures.

Kranich (2014) offers a review of the study of *language contact through translation* (LCTT) and argues that linguistics has failed to investigate the effects of language contact in written interaction, much as translation studies has failed to provide explanations of language change through translation. By providing an account of the results yielded in different studies involving lexical, morphological, syntactic, and pragmatic/stylistic innovations in translated texts, she concludes that the most significant factors of interference in present-day LCTT are the orientation of the translator (towards overt or covert translation), socio-political dominance relations, prestige of the SL, and attitudes towards the SL. Kranich (2014) further proposes that most interference in present-day LCTT occurs at the pragmatic and stylistic level, for example in terms of author-readership interaction, and that the evidence available so far suggests that innovation does not spread to monolingual text production but remains limited to translations.

One of the causes of interference is the tension between the force exerted by the proposals of the ST and the expectations of linguistic correction and proper writing in the target context (Franco Aixelá, 2009), allegedly ingrained in translator training and reaffirmed by the professional narratives of experienced translators. In this regard, Franco Aixelá points out that the translator is required to strike a balance between these two forces, considering that favouring TL linguistic correction would take away the technicality of the TT, whereas guiding the translation task by the "frequency" of occurrence of certain linguistic elements would yield a translation that may not be well-received by highly specialised readers. This suggests that there is a need to investigate further readability and acceptability of technical translation,

with a view to assisting the translator's decision-making process. Other motives for interference that affect technical and scientific translation in particular would be "the creation and preservation of a specific terminology or jargon, the nonexistence of a given term or structure in the TL, and the prestige of the source culture" (Franco Aixelá, 2009: 79). Given the generally accepted status of English as the language of science and technology, its prestige may have consequences in the way interference is perceived in translation from and into English.

In reviewing the literature on linguistic interference, it is possible to infer a number of methodological suggestions to guide the analysis of interference in translation in terms of both the object (texts) and subjects (translators) of study. Weinreich (1953: 29) notes that, for the analysis of grammatical interference, "[t]he main requirement is that in a given contact situation, both languages be defined in the same terms". This implies that the main linguistic features of the genres and the languages to be analysed should be described thoroughly, with a view to picturing a degree of a priori structural differences and similarities that this study aims to test.

Since "[i]t is thus in a broad psychological and socio-cultural setting that language contact can be understood" (Weinreich, 1953: 4), linguistic analysis of cross-linguistic influence could be thoroughly understood only if correlated with extra-linguistic factors associated with it. This is a very important point to make, since the context of language contact in translation differs significantly from the loci of contact described in the first conceptualisations of interference and the empirical work carried out in the field of linguistics. To name a few differences, translation not only occurs from L1 to L2 but frequently from L2 to L1, involves written communication rather than oral speech production, and is usually mediated by technology. Other contextual factors that might determine the degree of language influence in translation include the environment in which the translation is performed, training and professional experience of translators, globalisation, cultural resistance to—or tolerance of—interference, genre standardisation, and language standardisation, such as in the case of Spanish and its regulating body, the *Real Academia Española*.

Martín-Mor (2011) investigates the effects of the use of TM software on both the translation process and translated texts in terms of linguistic interference. He concludes that the distribution of linguistic interference in translated texts varies depending on the translation environment in which a TT is produced (with or without the use of TMs). Moreover, Martín-Mor (2019) observes that linguistic interference is more pervasive in translations produced by students than in those performed by experienced professionals. The author analyses linguistic interference at the level of formal correspondence ("close translations" versus "distant translations") (Martín-Mor, 2011: 387) and its impact, suggesting that acceptability of interference would depend on the expectations of the discourse community receiving the translated text. Overall, Martín-Mor's work (2011) suggests that TMs may affect readability, cohesion and textual coherence of translations.

3. Methodology

Chesterman (2017 [1993]: 175) distinguishes two types of norms governing translation: *professional norms* and *expectancy norms*, and argues that expectancy norms should be in part defined in relation to "*good* native texts" (emphasis in original). As a result, he points out, translations should be at least partly evaluated in terms of how close or far they are from the conventional features of non-translated writing in the target language. In this line, the present study intends to contrast the syntactic features of translations against the features of non-translated texts.

3.1. A previous corpus-based study of non-translated English and Spanish abstracts

In a previous study (Véliz-Ojeda, 2020) carried out in the context of the same doctoral research project the present study is a part of, a bilingual comparable corpus was built, including abstracts for research articles published in highly cited journals in English and Spanish. The aim of this corpus was to help determine the main differences between the text genre in English and Spanish as found in authentic non-translated texts. The results of this corpus-based study were then compared to the results obtained in the quasi-experiment the present paper reports on.

The corpus was parsed using UDPipe, "a trainable pipeline which performs [automatic] sentence segmentation, tokenization, POS [parts of speech] tagging, lemmatization and dependency parsing" (Straka and Straková, 2017: 88), based on the Universal Dependencies (UD) framework. This is an unusual procedure in translation studies, where the universal syntactic annotation scheme acts as a *tertium comparationis*, a sort of external point of reference that allows one to compare syntactic structures in two or more languages. This procedure makes it possible to predict the areas in which the two languages differ and, consequently, may serve to anticipate aspects in which translated texts might exhibit traces of linguistic transfer, such as those proved to be more salient in TM-translated texts by Dragsted (2006) and Martín-Mor (2011).

The abstract was chosen over full research articles, as it can be considered as an independent genre. The abstract provides a summary of the entire research article without depending on its other sections to fulfil its purpose (Martín-Martín, 2005). Therefore, abstracts are more homogenous in terms of rhetorical and textual structure. This maximises the comparability between languages, and between translated and non-translated texts. Having identified the distinctive features of scientific abstracts, it was necessary to make a clear distinction between rhetorical and grammatical (syntactic) features. Although rhetorical conventions may be realised by particular textual and syntactic elements, this study analyses the influence of TM systems on syntactic features (as measured in terms of syntactic dependency relations), leaving aside the study of rhetorical aspects of the genre.

The study of features such as text structure, rhetorical moves, hedging and personal attribution, among others, within the abstract genre is beyond the scope of the present study. These rhetorical features of research article abstracts have been widely studied, as with the work of Martín-Martín (2003, 2005) that extensively elaborates on the rhetorical differences between research article abstracts written in English and Spanish. Moreover, since translators are reluctant to make changes to the global organisation of texts (Hoey and Houghton, 2001), the exploration of translation-induced rhetorical changes seemed difficult to evaluate within the timeframe and aims of this study.

Once parsed, the English and Spanish sub-corpora were compared in order to identify syntactic differences. A corpus-driven approach was used here to study potential transfer in translation. This means that, although the researcher did have an a priori idea of what could be found in the study, the features to be analysed were not chosen intuitively but based on the results of this first analysis of the bilingual comparable corpus. For the comparison, the statistical method *Random Forest* (Breiman, 2001) was used. By means of trial and error analysis of the available data, random forests determine if a specific variable is a useful predictor (Tagliamonte and Baayen, 2012). In other words, once a subset of data is selected for analysis, the method randomly selects a set of variables and compares them to establish the best predictors for comparison, and this is repeated a large number of times (Ivaska and Siitonen, 2017).

The results of the comparison showed that the three main differences between the two sub-corpora lie in the highest frequency of three syntactic features in the Spanish text samples: the relation between a noun and its determiner (DET); the relation between a noun and a preposition or English genitive (CASE), and the relation for nominal modifiers of nouns (NMOD)—in English, prepositional complements and genitive's complements.

3.2. Translation experiment

Once the features of non-translated SL and TL texts were determined, a translation quasi-experiment was carried out with three groups of advanced translation trainees from three Chilean training institutions. Each group was divided into a control group who used Microsoft Word and no TM software, and an experimental group who used SDL Trados Studio. The source texts used were independent research article abstracts published in English journals to be translated by the informants into Spanish, their native tongue. In an effort to identify differences in terms of syntactic features, the resulting translations were then compared to non-translated SL and TT texts from the corpus previously built and analysed, as described in section 3.1.

The quasi-experiment took place in August 2018 with a total number of 42 participants. The informants were undergraduate English-to-Spanish translation students from three Chilean universities: Universidad Católica de Temuco (UCT) (18), Universidad de Tarapacá

(UTA) (11), and Universidad Arturo Prat (UNAP) (13). All the participants were native speakers of Spanish who entered translation programs with little or very basic skills in the English language, immediately after graduating from secondary school. Given the students' profile, a significant part of the curriculum is oriented to developing English language proficiency and to enhancing students' skills in Spanish as a native language, leaving little room for translation-related subjects in the first two years, then gradually intensifying translation-specific training until the fifth (last) year. Although the informants were students, from the description of the academic programs participating in the study, it follows that they can be considered junior translators expected to have developed a certain degree of competence in scientific translation and the use of TM software.

In a 60-90 minute session, students were asked to translate the abstract of a scientific article of around 200-300 words from English into Spanish using TM software. The TM software used was SDL Trados Studio, since it was available to the informants at their institutions and is the most widely used TM software by both the industry and training institutions. No TMs were provided to the students in order to avoid previous translations from influencing students' decisions in terms of TT segmentation.

Students were provided with information about the potential readers of their translations, scientists, and the researcher collected the translations from the participants' computers once they had finished the task. At the same time, another group of students were asked to perform the same task using Microsoft Word instead of TM software. Students at each university were divided into two sub-groups in order to have one experimental group and one control group in each institution. To control variables such as foreign and native language proficiency and experience in translation and software use, only students of the final (fifth) year of the programs were invited to participate.

In order to ensure independent observation, the texts that served as STs for the translation experiment were different from those included in the English sub-corpus, but were extracted from the same journals that supplied the samples for the English sub-corpus. So as to prevent idiosyncratic textual features from influencing the results of the experiment, texts from different authors were used, and no text was translated by more than one informant in the experimental groups. The same texts were translated by the informants in the control groups. This allowed for comparison of translations with and without the use of a TM system in specific cases to illustrate the potential influence of the software on the translational product, and strengthen the results, interpretations, and their generalisability.

4. Findings

The translations obtained in the quasi-experiment were parsed and analysed for DET, CASE, and NMOD. Subsequently, three comparisons were made: translations produced using MS

Word were compared to those carried out with the use of TM software; all translations were compared to their STs, and all translations were contrasted against the Spanish-subcorpus of non-translated abstracts. The results of this analysis do not seem to support the initial expectation that translations would be highly influenced by their STs and fall somewhere between the English and Spanish originals with regard to the three linguistic features under examination (see table 11). The average occurrences of the three syntactic features are relatively similar for both TM and MS Word translations, with the TTs produced without TM software showing a slight underrepresentation of DET, CASE, and NMOD. Therefore, the results would suggest that the hypothesis that the TM segmentation system encourages underrepresentation of syntactic features (CASE, DET and NMOD) in Spanish translations should be rejected.

TABLE 1

General information about the selected films

SYNTACTIC DEPENDENCY FEATURE	TM TRANSLATIONS	MS WORD TTS	ENGLISH STS	SPANISH NTS** IN COMPARABLE CORPUS
DET	49.95	43.38	15.09	38.08
CASE	56.38	53.76	28.05	42.99
NMOD	39.90	39.047	18.14	30.29
Number of texts	21*	21*	22	100

*20/22 STs were translated both with and without a TM system; of the remaining 2 STs, 1 was translated with a TM system and 1 was translated without a TM system.

** NTs = non-translated texts.

One possible interpretation of the results is that they could reflect that the student translators who participated in the experiment are highly competent in recognising grammatical differences between SL and TT and in dealing with the translation of the syntactic issues at hand. In fact, by exploring the instances of the three syntactic relations, it was verified by the researcher that, in most cases, their use was obligatory in Spanish. Still, when both TM and MS Word translations are compared to the Spanish non-translated texts in the comparable corpus discussed in section 3.1, all the translations show a frequency of occurrence of DET, CASE and NMOD higher than non-translated texts. An explanation for this could be that there is a tendency on the part of the student translators to normalise, in Baker's (1993) terms, the language of the translations. In other words, the informants could have striven to adhere to the norms of the TL system to the point of exaggeration or overreaction, as put forward by Bernardini and Ferraresi (2011).

In their corpus-based study aimed at unveiling features of translated texts, Bernardini and Ferraresi (2011) analyse the presence of Anglicisms in Italian technical writing and translation. They found that the translators participating in their study tend to be more "conservative" than native Italian writers in the use of English words and morphological features when there is an alternative that is considered "normal" or more "typical" of the Italian language. The text samples making up the corpus dealt with a very restricted topic in the field of computing, and it is thus surprising that, instead of finding traces of interference, what the findings of Bernardini and Ferraresi's study confirm is that translators' choices tend to normalise the language more than native Italian writers. What is even more striking and conclusive is that the translations analysed in the study were produced by "amateur" translators who were more experienced in the computing field than in linguistics or translation. These results led the authors to the conclusion that "the very act of translation may induce one to take a more conservative, normalizing attitude" (Bernardini and Ferraresi, 2011: 242) and add to the evidence for normalisation as a "translation universal" (Baker, 1993: 243-244) or for Toury's (1995: 267) "law of growing standardization". Although Bernardini and Ferraresi (2011) focus on the frequency of lexical and morphological calques in translation and do not analyse syntactic features such as those under scrutiny in the present study, their findings and conclusions can provide a new line of interpretation for the results of the translation experiment. Moreover, the study under discussion has a methodological implication for the present work in that the former proves the value of using comparable corpora, that is to say, translated text material for comparison against non-translated texts produced in the translations' TL, to escape from the circularity of parallel corpora (the contrast of TTs against their STs).

If different outcomes had stemmed from each institution, a thorough analysis would have had to be carried out in order to determine if variations in the results can be explained by curricular differences or any other factors. In this regard, as shown in table 2 below, no significant differences were detected among institutions.

TABLE 2

Average occurrences of syntactic dependency features per institution and experimental/control group

INSTITUTION	UCT		UTA		UNAP		
FEATURE	TM	MSW	TM	MSW	TM	MSW	ALL TRANSLATIONS
DET	54.22	54.55	49.71	50.17	42.6	44.33	49.16
CASE	58.00	58.89	56.29	54.33	53.6	52.67	55.07
NMOD	39.89	43.88	40.71	38.17	38.8	38.5	40.11

MSW = Microsoft Word.

Assuming that there are differences between Spanish originals and Spanish translations (due to either transfer or normalisation), I considered it convenient for the purposes of this study to build another corpus, one of professional Spanish translations that could serve as the benchmark against which students' translations could be compared. In this gold standard corpus, research article abstract authors were not native speakers of Spanish, thus the need for intervention of a professional translator. For sample selection, similar criteria as for the selection of the previous corpora were considered, that all or most of the authors be native speakers of English, as judged by their name and institutional affiliation. This would ensure that the authors themselves were not the ones who carried out the translation, and that professional translation practitioners had been hired for the translation of the abstracts composing the gold standard corpus. The source for this new corpus was one of the journals that fed the English sub-corpus, *Conservation Biology*. Although this journal publishes articles in English only, it requires that authors submit a Spanish translation of the English abstract. Since *Conservation Biology* is highly cited and enjoys prestige in the field, it can be assumed that the translations are carried out by professional translators and are therefore of a high standard. I therefore built a corpus of 22 abstract translations extracted from *Conservation Biology*, the same number of STs used in the translation experiment and conducted the same syntactic dependency feature analysis carried out on the previous translated and non-translated corpora. Table 3 below shows the average occurrences of DET, CASE, and NMOD in the gold standard corpus.

TABLE 3

Average occurrences of DET, CASE, and NMOD in Gold Standard corpus

FEATURE	GOLD STANDARD AVERAGE	TRANSLATION EXPERIMENT TTS	SPANISH SUB-CORPUS OF NON-TRANSLATED TEXTS
DET	46.18	49.16	38.08
CASE	64.73	55.07	42.99
NMOD	53.95	40.11	30.29

As can be observed, averages in the gold standard corpus for CASE and NMOD are significantly higher than those found in the TTs obtained in the translation experiment; however, the average occurrence of DET is relatively lower in the gold standard than in the translations from the experiment. The results of this analysis would then partially support the hypothesis of the present study, at least for CASE and NMOD. As compared to the average in the Spanish sub-corpus of non-translated texts, both the gold standard and the translation experiment TTs show an even higher degree of occurrence of the studied features. This is an interesting finding in that it shows that both professional translations and TM-mediated

students' translations differ from Spanish originals, this being perhaps a confirmation of the presence of translationese in scientific translation.

By taking a closer look at the samples of published translations considered for the gold standard, from the point of view of a native Spanish speaker and professional translator such as this author, it can be established that the translations are generally of a good standard; however, evident traces of syntactic transfer can still be recognised in the translated texts, for example the transfer of ellipsis when it is not the preferred case in Spanish, and the unnecessary use of the passive voice. This could suggest that (a) English as the "lingua franca" of science has a marked influence on the way abstracts are translated into Spanish in terms of syntax, or (b) the translation samples were most probably not translated by experienced, professional translators. If (b) were the case, I would need to consider the possibility of building a new corpus to be the gold standard for comparison, one in which I could ensure that the samples correspond to authentic pieces of professional translation work. To do this, professional English to Spanish translators would be required to participate in the present study. However, as discussed above, the differences between translations produced by professional and trainee translators have already been studied and are not within the scope of the present study.

5. Conclusions

The results of the quasi-experiment show that the use of CASE, DET and NMOD in the translations conforms to the frequency of these features described for the TL, thus, rejecting the expectation that they would be under-represented as per the influence of the SL and TM software. This may be interpreted as the translators being either competent in the use of the features characteristic of the TL or as over-application of the syntactic rules of the TL (normalisation). Therefore, the initial expectation that the segmentation system of TM software has an influence on the way translations are written, particularly when translators are trainees, could not be proven true in the conditions of the present study. A valuable and interesting avenue for research would be the exploration of the way the translations produced in this quasi-experiments are received by real specialist readers, that is, Spanish-speaking scientists in conservation biology. Translation reception could be a valid measure of quality (Véliz-Ojeda, 2020), and, if described in terms of learning objectives, expectancy norms could inform training and help graduates to produce translations that live up to the readers' expectations (Chesterman, 2017 [1993]).

6. Acknowledgements

This study was funded through a Becas Chile scholarship granted by the Chilean National Agency for Research and Development and was carried out during research leave awarded

to the author by Universidad de Tarapacá. Special appreciation goes to professor Silvia Bernardini and Dr. Ilmari Ivaska, who generously hosted the author at the University of Bologna and assisted him with the design and analysis of the corpus-based study.

7. References

Baker, Mona, 1993: "Corpus Linguistics and Translation Studies: Implications and Applications" in Mona Baker, Gill Francis and Elena Tognini-Bonelli (eds): *Text and Technology: In Honor of John Sinclair*, Amsterdam: John Benjamins Publishing Company, 233-250.

Bernardini, Silvia, and Adriano Ferraresi, 2011: "Practice, Description and Theory Come Together: Normalization or Interference in Italian Technical Translation?", *Meta: Translators' Journal* 56 (2), 226-246.

Bowker, Lynne, 2005: "Productivity Vs Quality? A Pilot Study on the Impact of Translation Memory Systems", *Localisation Focus: The International Journal of Localisation* 4 (1), 13-20.

Bowker, Lynne, and Jairo Buitrago Ciro, 2019: *Machine Translation and Global Research: Towards Improved Machine Translation Literacy in the Scholarly Community*, Bingley: Emerald Publishing Limited.

Breiman, Leo, 2001: "Random Forests", *Machine Learning* 45 (1), 5-32.

Bundgaard, Kristine, Tina Paulsen Christensen and Anne Schjoldager, 2016: "Translator-Computer Interaction in Action: An Observation Process Study of Computer-Aided Translation", *The Journal of Specialised Translation* 25, 106-130.

Carl, Michael, and Emmanuel Planas, 2020: "Advances in Interactive Translation Technology" in Erik Angelone, Maureen Ehrensberger-Dow and Gary Massey (eds): *The Bloomsbury Companion to Language Industry Studies*, London and Newy Yor: Bloomsbury Publishing, 361-386.

Chesterman, Andrew, 2017 [1993]: "From 'Is' to 'Ought': Laws, Norms and Strategies in Translation Studies" in Andrew Chesterman (ed): *Reflections on Translation Theory: Selected Papers 1993-2014*, Amsterdam: John Benjamins Publishing Company, 167-183.

Doherty, Stephen, 2016: "The Impact of Translation Technologies on the Process and Product of Translation", *International Journal of Communication* 10, 947-969.

Dragsted, Barbara, 2006: "Computer-Aided Translation as a Distributed Cognitive Task", *Pragmatics and Cognition* 14 (2), 443-464.

Drechsel, Alexander, 2019: "Technology Literacy for the Interpreter" in David B. Sawyer, Frank Austermühl and Vanessa Enríquez Raído (eds): *The Evolving Curriculum in Interpreter and Translator Training: Stakeholder Perspectives and Voices*, Amsterdam: John Benjamins Publishing Company, 259-268.

Franco Aixelá, Javier, 2009: "An Overview of Interference in Scientific and Technical Translation", *The Journal of Specialised Translation* 11, 75-88.

Hickey, Raymond, 2013: "Language Contact: Reconsideration and Reassessment" in Raymond Hickey (ed): *The Handbook of Language Contact*, Hoboken: Wiley-Blackwell, 1-28.

Hoey, Michael, and Diane Houghton, 2001: "Contrastive Analysis and Translation" in Mona Baker (ed): *Routlege Encyclopedia of Translation Studies*, London and New York: Routledge, 45-49.

Ivaska, Ilmari, and Kirsti Siitonen, 2017: "Learner Language Morphology as a Window to Crosslinguistic Influences: A Key Structure Analysis", *Nordic Journal of Linguistics* 40 (2), 225-253.

Kabatek, Johannes, 1997: "Dime Cómo Hablas Y Te Diré Quién Eres: Mezcla De Lenguas y Posicionamiento Social", *Revista de Antropología Social* 6, 215-236.

Kellerman, Eric, 1995: "Crosslinguistic Influence: Transfer to Nowhere?", *Annual Review of Applied Linguistics* 15, 125-150.

Kenny, Dorothy, 2020: "Technology and Translator Training" in Minako O'Hagan (ed): *The Routledge Handbook of Translation and Technology*, London and New York: Routledge, 498-515.

Kranich, Svenja, 2014: "Translations as a Locus of Contact" in Juliane House (ed): *Translation: A Multidisciplinary Approach*, Hampshire and New York: Palgrave Macmillan, 96-115.

Leblanc, Matthieu, 2013: "Translators on Translation Memory (Tm): Results of an Ethnographic Study in Three Translation Services and Agencies", *Translation & Interpreting* 5 (2), 1-13.

Martín-Martín, Pedro, 2003: "A Genre Analysis of English and Spanish Research Paper Abstracts in Experimental Social Sciences", *English for Specific Purposes* 22 (1), 25-43.

Martín-Martín, Pedro, 2005: *The Rhetoric of the Abstract in English and Spanish scientific Discourse: A Cross-Cultural Genre-Analytic Approach*, Bern: Peter Lang.

Martín-Mor, Adrià, 2011: *La Interferència Lingüística En Entorns De Traducció Assistida Per Ordinador: Reserca Empíricoexperimental*. Doctoral thesis, Universitat Autònoma de Barcelona.

Martín-Mor, Adrià, 2019: "Do Translation Memories Affect Translations? Final Results of the Trace Project", *Perspectives* 27 (3), 455-476.

MATRAS, Yaron, 2013: "Contact, Convergence, and Typology" in Raymond HICKEY (ed): *The Handbook of Language Contact*, Hoboken: Wiley-Blackwell, 66-85.

MELBY, Allan K., and Sue Ellen WRIGHT, 2015: "Translation Memory" in Chan SIN-WAI (ed): *The Routledge Encyclopedia of Translation Technology*, London and New York: Routledge, 662-677.

O'HAGAN, Minako, 2020: *The Routledge Handbook of Translation and Technology*, London and New York: Routledge.

PYM, Anthony, 2014: *Exploring Translation Theories*, London and New York: Routledge.

PYM, Anthony, 2020: "Quality" in Minako O'HAGAN (ed): *The Routledge Handbook of Translation and Technology*, London and New York: Routledge, 437-452.

STRAKA, Milan, and Jana STRAKOVÁ, 2017: "Tokenizing, Pos Tagging, Lemmatizing and Parsing Ud 2.0 with Udpipe", paper presented at the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies, 88-99.

TAGLIAMONTE, Sali A., and R. Harald BAAYEN, 2012: "Models, Forests, and Trees of York English: Was/Were Variation as a Case Study for Statistical Practice", *Language Variation and Change* 24 (2), 135-178.

TOURY, Gideon, 1995: *Descriptive Translation Studies and Beyond*, Amsterdam: John Benjamins Publishing Company.

TOURY, Gideon, 2012: *Descriptive Translation Studies and Beyond. Revised Edition*, Amsterdam: John Benjamins Publishing Company.

UNIVERSAL DEPENDENCIES: "Universal Dependencies" [http://universaldependencies.org/, date of retrieval: 18 November 2018].

VÉLIZ-OJEDA, Eduardo, 2020: *Translator Education in Chile: Effects of Translation Memory Systems on the Reception of Scientific Translations*. Doctoral thesis, Monash University.

VENUTI, Lawrence, 1998: *The Scandals of Translation: Towards an Ethics of Difference*, London and New York: Routledge.

WEINREICH, Uriel, 1953: *Languages in Contact: Findings and Problems*, New York: Linguistic Circle of New York.