

Using bilingual parallel corpora in translation methodology: an analysis of students' translation competences in the UCMA-MUST corpus (English-Spanish)

**Los corpus paralelos bilingües como herramienta metodológica
en traducción: análisis de la competencia traductora de los estudiantes
en el corpus UCMA-MUST (inglés-español)**

Jorge Braga Riera^a & Nava Maroto^b

^a Universidad Complutense de Madrid/ITEM. jbragariera@filol.ucm.es

^b Universidad Politécnica de Madrid/Tradaval. mariadelanava.maroto@upm.es

Received: 30.04.2019. Accepted: 02.08.2019

Resumen: El proyecto MUST (MUltilingual Student Translation) (Granger y Lefer, 2018: 72) tiene como objetivo compilar un corpus paralelo multilingüe de traducciones realizadas por estudiantes sin experiencia en el campo de la traducción. En el marco de este proyecto, el equipo UCMA (de la Universidad Complutense de Madrid) aporta un subcorpus bilingüe de textos especializados traducidos por estudiantes en los ámbitos audiovisual y de las humanidades, entre otros. Teniendo en cuenta las cinco subcompetencias propias de la adquisición de la competencia traductora (lingüística, extralingüística, de transferencia, profesional, psicofisiológica y estratégica [PACTE, 2001 y 2003]), este estudio persigue detectar las debilidades y fortalezas presentes en las tareas de traducción inglés-castellano de los estudiantes. Para ilustrarlas, se ha utilizado el sistema de anotación TAS propuesto por Granger y Lefer (2018) y Granger, Lefer y Penha Marion (2018), junto con una muestra de textos humanísticos a cargo de traductores en formación universitaria (postgrado) que se encuentran almacenados en el corpus UCMA.

Palabras clave: UCMA-MUST; traducción especializada; corpus paralelos; competencia traductora; etiquetado.

Abstract: The Multilingual Student Translation (MUST) Project (Granger & Lefer, 2018: 72) aims to compile a multilingual parallel corpus of translations carried out by inexperienced learners. Within this project, the MUST partner UCMA (Complutense

» Braga Riera, Jorge & Maroto, Nava. 2019. "Using bilingual parallel corpora in translation methodology: an analysis of students' translation competences in the UCMA-MUST corpus (English-Spanish)". *Quaderns de Filologia: Estudis Lingüístics* XXIV: 39-58. doi: 10.7203/QF.24.16298

University of Madrid) contributes to the global MUST project with a bilingual sub-corpus containing student-translated specialised texts in the fields of media and humanities, among others. Bearing in mind the five sub-competences that contribute to the acquisition of a translation competence, namely linguistic, extra-linguistic, transfer, professional, psychophysiological and strategic (PACTE 2001, 2003), this study aims to detect the weaknesses and strengths present in the students' tasks when translating from English into Spanish. To do so, the TAS annotating system put forward by Granger & Lefer (2018) and Granger, Lefer & Penha Marion (2018), together with a sample of humanistic texts translated by trainee translators at a university (postgraduate) level and retrieved from the UCMA corpus has been used as means of exemplification.

Keywords: UCMA-MUST; specialised translation; parallel corpora; translation competences; tagging.

1. Introduction: the MUST project

In recent years, we have witnessed the proliferation of studies that resort to corpus linguistics as a useful methodological tool in translation (Johansson, 2007; Kruger *et alii*, 2011; Granger & Lefer, 2017), broadening its potential applications for translative and contrastive studies. However, and as stated by Rica (in press), not many of these studies focus on the pedagogical applications corpora may have in translation teaching and learning when the target texts under investigation are not produced by professional translators but inexperienced learners or trainees with a poor knowledge in the translation field.

Within this context, the bilingual parallel corpus being constructed under the MULTilingual Student Translation (MUST) Project, launched in 2016 by Sylviane Granger and Marie-Aude Lefer at the Catholic University of Louvain, aims to fill this gap. By collecting a sizeable multilingual student translation corpus (plus standardised metadata), this international project attempts to “share translations produced by students and to process them using a standardised set of tools and guidelines intending to optimising translation teaching and advancing empirical research” (Granger & Lefer, 2018: 72). The project is singular in the sense that it is truly multilingual (15 languages involved), it represents a wide range of text types, genres and topics, and it includes both L2>L1 and L1>L2 translations (Granger & Lefer, 2018).

All the data collected by the 32 partners from 14 countries are gathered and searchable via Hypal4MUST (Obrusnik, 2014), an adapted version of the Hypal software tool for the processing of parallel texts. Hypal4MUST allows for the integration of standardized structured metadata, as it incorporates POS tagging, sentence alignment (both automatic and manual), and annotation and corpus search features within the same web interface. By using this interface, the process of uploading, aligning and annotating the original texts and the translations provided by the students is enabled in a fairly intuitive way.

An innovative computer-aided Translation-Oriented Annotation System (TAS) is currently being developed for the characterization of the texts gathered using Hypal4MUST. The main distinctive features of this new system are two: on the one hand, it allows for the annotation of translation procedures, therefore “catering for theoretically oriented

research” (Granger & Lefer, 2018). On the other hand, not only errors can be tagged, but also successful translation strategies.

Within this far-reaching project, the MUST partner UCMA, based in Complutense University of Madrid (Spain)¹, contributes to the global MUST project with a bilingual subcorpus (English-Spanish) containing student-translated general and specialised texts in the fields of media, literature, humanities, science, law and economy. Students’ profiles comprise translation trainees, foreign language students with a major in English, engineers studying EFL and MA students, all of them with different English levels ranging from B1 to C1.

This paper will make use of the UCMA subcorpus in order to attempt to illustrate the weaknesses and strengths present in the students’ performance, while testing TAS as a potentially useful annotation system both for assessment and research.

2. Theoretical framework

2.1 *Corpus Linguistics and Translation*

Corpus linguistics has been long established as a significant paradigm in translation theory and practice since Baker (1993) put forward her belief that corpora could be a suitable methodology for translation studies (Corpus-based translation, or CBT) (Fantinuoli & Zanettin, 2015: 1). This notion was later pursued by Guy Aston and the subsequent CULT (Corpus Use and Learning to Translate) conferences (Beeby *et alii*, 1999: 1). A clear symptom of the overall relevance of corpus studies for translation theory and practice is the emergence in recent years of proposals for the construction of appropriate corpora and their effective exploitation for research and educational purposes. In fact, large corpora are even being used to develop new models of machine translation systems (Kruger *et alii*, 2011: 1).

However, and as, again, Fantinuoli and Zanettin (2015) rightly point out, corpus-based research finally depends on the availability of tools and resources which help both in the corpus design and compilation as

¹ The Spanish UCMA team is made up by the following members: Arsenio Andrades Moreno, Jorge Braga Riera, Nava Maroto García, Sara Martínez Portillo, Juan Pedro Rica Peromingo (lead investigator) and Ángela Sáenz Herrero.

well as in data analysis. Reusability of the resources created is also considered as a vital issue in CBT studies, as corpus compilation is a hard and time-consuming task and the resulting data are many times used only once. Corpus annotation and alignment are key concepts too, as a fully annotated and sentence-aligned version must also be a goal before the comparative process is carried out (Volk, 2019).

More specifically, the use of corpora for translator training purposes has also been a significant concern in CBT, with studies by Olohan (2004) and Zanettin *et alii* (2003). Indeed, Olohan (2004) devotes a whole chapter of her book *Introducing Corpora in Translation Studies* to the use of corpora in translator training. Lynne Bowker and Peter Bennisson (2003), for their part, suggested the creation of a learner corpus, the Student Translation Archive, where students could submit their translations electronically. This archive could be used for the extraction of ad hoc corpora. According to Olohan (2004), the mere availability and accessibility of these data are beneficial, since this could allow for the analysis of certain aspects of students' performance. Also relevant are the studies carried out by Fictumova *et alii* (2014, 2017), focusing on error statistics for individual student translations including the use of corpora and TM tools.

Nevertheless, and despite the benefits that this method of analysis conveys, the compilation of large multilingual learners' corpora has not traditionally been a major concern for CBTS². In this sense, a corpus such as the one being compiled within the MUST project not only will offer full availability but will also be an invaluable, reusable resource for researchers and translation trainers, who will be able to withdraw linguistic and extra-linguistic information that can be used for mere consultation or educational and academic purposes. This goes in line with some relatively recent proposals in this sense, such as Sánchez-Gijón's (2009) use of corpora as a documentation resource or the study by Rodríguez Inés (2009), which revolves around how the process (and not just the product) can also become the aim for analysis.

² Still, some further examples can be mentioned such as PELCRA, the Student Translation Archive (STA) and other European projects as MeLLANGE Learner Translation Corpus or RusLTC. In Spain, ENTRAD (University of Zaragoza) is a computerised corpus precisely designed for the enhancement of the teaching-learning process in translation training.

2.2 *On translation competence*

In the last two decades, the development of translation competence has been envisaged as a keystone in translator training –hence the significant number of authors who have tried to describe its nature (Carrasco, 2019). This emerging interest is not surprising, given the implications of this concept (understood as an expert, underlying system of knowledge that is required to translate adequately) as pivotal in pedagogical approaches to translation. Still, there is no doubt that one of the current dominant models in this respect is the holistic approach put forward by the PACTE (Process in the Acquisition of Translation Competence and Evaluation) group, which studies written translation. More specifically, this model contemplates translation competence acquisition as a process of restructuring and developing five sub-competences of translation competence. These sub-competences, which were initially labelled as linguistic, extra-linguistic, transfer, professional, psychophysiological and strategic (PACTE, 2001), were later redefined (PACTE, 2003) as *bilingual*, *extra-linguistic*, *strategic*, *instrumental* and *knowledge about translation*, to which psycho-physiological components were added. Even though such a detailed list has raised some controversy among scholars (see Malmkjaer, 2009: 13), this rich taxonomy unquestionably gives a complete account of what one is expected to master to become a translator.

The bilingual sub-competence refers to “the underlying system of knowledge and abilities necessary for linguistic communication in both languages” (PACTE, 2003: 48). This competence, which includes the pragmatic, socio-linguistic, textual, grammatical and lexical knowledge in the two languages, can be shared with other bilinguals and professionals (Jiménez Crespo, 2017: 244). The extra-linguistic sub-competence, however, is defined as the “implicit or explicit knowledge about the world in general and specific areas of knowledge” (PACTE, 2003: 48): bicultural knowledge, encyclopaedic knowledge and subject knowledge. The strategic sub-competence includes “procedural knowledge to guarantee the efficiency of the translation process and solve the problems encountered”. This is an essential sub-competence, since “it affects all the others and causes inter-relations amongst them because it controls the translation process” (PACTE, 2003: 48). Functions included here are the choice of the adequate method, evaluation of processes

and results, the compensation for the deficiencies in the other sub-competences and the identification of translation problems (and procedures applied to solve them). For its part, the instrumental sub-competence refers to “the use of documentation sources and information and communication technologies applied to translation: dictionaries of all kinds, encyclopaedias, grammars, style books, parallel texts, electronic corpora, searchers, etc.” (PACTE, 2003: 59). This sub-competence cannot be understood without the use of technology, which has exponentially increased in the last two decades (Jiménez Crespo, 2017: 245). Knowledge about translation includes knowledge about how translation functions: “Types of translation units, processes required, methods and procedures used (strategies and techniques), and types of problems” (PACTE, 2003: 59) but also knowledge related to professional aspects and the work market. Finally, the psychophysiological components comprise cognitive components (such as memory, perception, etc.), attitudinal aspects (intellectual curiosity, perseverance, critical spirit, etc.) and abilities such as creativity or logical reasoning.

The psychophysiological components and the instrumental sub-competences are out of the scope of this paper, since additional metadata are necessary to prove the knowledge of documentation resources and ICTs (Information and Communication Technologies) as applied to translation, as well as all the cognitive aspects linked to the task. Thus, they will be covered in future studies.

3. Aims and methodology

3.1 Objectives

As said above, this paper attempts to illustrate the weaknesses and strengths present in the selected students’ tasks (see below). To this aim, attention will be paid to three specific sub-competences: on the one hand, the *bilingual* sub-competence, and to what extent the selected subjects keep the linguistic, sociolinguistic, pragmatic, and discursive competences acquired in the past: the pragmatic conventions needed to carry out language acts and speech functions; the needed sociolinguistic conventions (including language register and dialectal varieties); coherence and cohesion mechanisms according to genre conventions; and aspects related to vocabulary, morphology and syntax. On the other hand,

some focus will be given to the *extra-linguistic* sub-competence, trying to prove to what extent the bicultural knowledge, the encyclopaedic knowledge and the subject knowledge (in the area of Humanism) are reflected (or missing) in the resulting translations. On top of this, the *knowledge about translation* techniques and strategies will also be taken into consideration (although knowledge related to the work market has been excluded in this analysis).

For this purpose, a sample of a humanistic text translated by two trainee translators at a postgraduate level and retrieved from the UCMA corpus has been used as means of exemplification. Methodologically, the TAS annotation method MUST has been used, not only as a means to identify and implement those strengths and weaknesses detected in the translated versions but also as a way to critically assess its applicability as a tagging system.

3.2 Description of the TAS annotation method at MUST

TAS (Translation-oriented Annotation System) consists of 60 tags sub-divided in three categories, and each of them can be further segmented in sub-categories (Granger & Lefer, 2018; Granger, Lefer and Aguiar de Souza, 2018). The first one, *ST-TT transfer*, allows for the annotation of discrepancies between the source text (ST) and the target text (TT) and/or between the TT and the translation brief. The second, *Language*, identifies features of the TT that are erroneous and/or inappropriate independently of the ST. There is a third module related to *Translation procedures*, envisaged for the annotation of the strategies used to solve translation problems, which can be observed when comparing the translation with its source text. While the first two modules have already been implemented, the third is still under development.

The following tags are included under the ST-TT transfer part of the TAS system: content transfer (omission, distortion, addition, indecision); lexis: translating untranslatable, untranslated translatable, term translated by non-term, non-term translated by term; discourse/pragmatics: connectors, theme-rheme; register/culture: register mismatch, cultural mismatch; translation brief: inconsistent with glossary, formatting.

As for the language part of the TAS annotating system, it comprises the following tags: grammar: inflectional morphology, tense and aspect, voice, word order determiner, preposition, pronoun, concord, comple-

mentation (adjective, noun, verb, adverb); lexis and terminology: single-word non-term, single-word term, multiword non-term, multiword term; cohesion: linkword, pronoun reference; mechanics: spelling; punctuation; units, dates, numbers; style and situational context: heavy, redundant, contextual variant, degree of (in)formality.

Two meta-tags have been conceived to indicate whether a given solution is positive or whether there is a suspected source-language intrusion. TAS also features comment boxes both for each annotation and a final box for general comments. In addition, a correction box is also available. The following figure (Fig. 1) illustrates the TAS annotation system as featured in the editor Hypal4MUST.

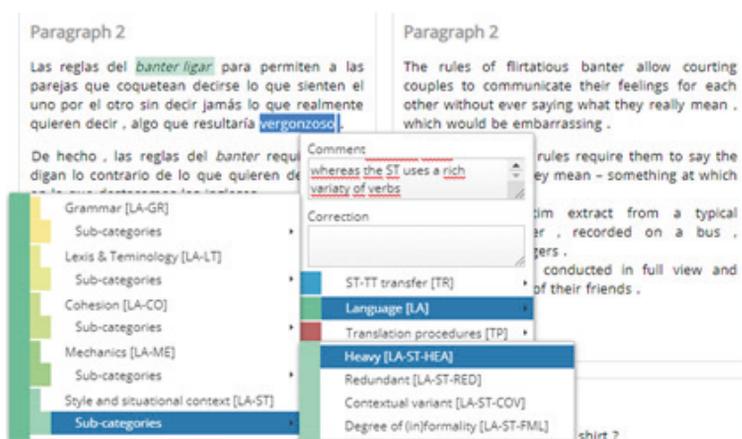


Figure 1. Example of the TAS annotation system in the editor Hypal4MUST

3.3 Corpus sample description and participants

As stated above, a sample humanistic text (within the sociological field) retrieved from the UCMA corpus has been resorted to for illustrative purposes. The text chosen is a 389 word-long excerpt from Kate Fox's *Watching the English: The Hidden Rules of English Behaviour*, edited in 2004 by Hodder Headline. The sample has been taken from the chapter devoted to behaviour codes ruling sex (pp. 324-352), to be precise the part focusing on the hidden rules of banter (p. 336). The whole extract can be seen in Appendix A.

Two sample translations into Spanish have been selected as object of analysis. These two translations of this text were requested as an assignment for the 30-hour subject “Translation of Humanistic Texts”, which is part of the Specialist Course in Translation for Specific and Professional Purposes (English-Spanish) offered by the Institute of Modern Languages and Translators at Complutense University of Madrid. The space limitations of this paper account for the fact that the analysis will be limited to only two texts: the first one (T1) was done by a student with a degree in Translation and Interpreting (S1), with Spanish as her L1 and English as her L2; the second subject (S2) is a student with a Spanish degree in International Relations, with Russian as her L1 and English and Spanish as her working languages (L2).

4. Analysis

The ST-TT transfer part of the annotation system designed for MUST (TAS) gives proof of some discrepancies between the source text and the students’ solutions in the target texts. In our sample, these divergences do not apply to the Discourse/Pragmatics and Register/Culture sections since both Spanish translations show no inconsistencies in the use of connectors and information structure, nor any cultural or register mismatches. Additionally, both participants have complied with the specifications of the task (translation brief). As far as content transfer is concerned, only one instance of omission has been registered (S2, see Table 1 below), while additions and indecisions (i. e., several options given for one translation) are not present in the resulting texts.

ST	TT (Student 1)	TT (Student 2)
The key ingredients of flirtatious banter are all very English:		Los ingredientes clave de las bromas coquetas:

Table 1. Omission

There are, however, six instances of distortion (3 cases in S1, another 3 in S2). In some of them, the resulting text is inappropriate (see Table 2), whereas in others the meaning provided is inexact (Table 3):

ST	TT (Student 1)	TT (Student 2)
Without ever saying what they really mean .		Requieren que digan lo contrario de lo que quieren decir .
But a normal, unremarkable , everyday English courtship sequence.		Una manera de banter, sin complicaciones , bastante típica en la vida diaria de Inglaterra.

Table 2. Distortion (incorrectness)

As seen in the first example in Table 2, S2 wrongly opts for lo *contrario* ('the opposite') to render "without ever saying...", and chooses *sin complicaciones* ('hassle-free, uncomplicated') to express the content of "unremarkable".

ST	TT (Student 1)	TT (Student 2)
A typical flirtatious encounter, recorded on a bus.	Situación típica de coqueteo sacada de un autobús.	
The exchange was conducted in full view and hearing of a group of their friends.	El intercambio se desarrolló a plena vista y oído de un grupo de amigos suyos.	La conversación se realizó enfrente un grupo de amigos.
I was collecting examples of real-life chat-up routines .		Estaba coleccionando ejemplos de conversaciones en la vida real.

Table 3. Distortion (inexactness)

Table 3 contains other three examples of ST-ST transfer distortion. First, *sacada de* ('taken from') does not identically convey the action expressed by "recorded". Secondly, "exchange" means *diálogo* or *conversación* in this context rather than *intercambio* ("exchange", S1), while the solutions provided for "in full view and hearing" are unidiomatic in both cases. Finally, "chat-up routines" cannot be rendered simply as *conversaciones* ('conversations') as S2 proposes.

But it is the Language Part that comprises the highest number of instances. Participants seem to use cohesive devices correctly, and the same can be said for punctuation and spelling (mechanics), with only one exception: the double use of hyphens and inverted commas for dialogue construction by S1, when Spanish typographic conventions typically require dashes to indicate separate speakers' interventions.

Grammar, lexis and terminology, as well as Style and Situational Context, are the wealthiest categories in number of examples: S1 shows three grammar inaccuracies, with S2 doubling this figure; lexically, S1 registers 5 inappropriate solutions, with 9 cases in S2; lastly, S1 shows 12 Spanish phrases that need stylistic reformulation, as opposed to just the three examples registered by S2.

Grammatically, a controversial issue is the use of determiners and prepositions, as seen in the following examples:

ST	TT (Student 1)	TT (Student 2)
And banter specifically excludes...		<i>Banter excluye específicamente...</i>
From the conversation among their friends afterwards, it was clear...		Mientras la conversación con sus amigos quedó claro que...
This pair had been attracted to each other for some time...	Este par de dos se gustaban durante un tiempo.	

Table 4. Grammar and sentence errors

The first translated phrase includes an error in the use of the article, as *el* is required in Spanish as a modifier of the noun “banter”. The second example portrays a wrong rendering of the preposition “from” as *mientras* (which means “meanwhile” in Spanish), resulting in an ungrammatical sentence in the recipient language. The third example not only shows a dubious tense form (*gustaban*) but also an incorrect choice of *durante* (“during”), which is deemed unnecessary in the target version. This choice might also be the consequence of negative transfer in an attempt to maintain the time particle “for” of the original. This case should require double annotation, which is currently not allowed in the

system, since morphosyntactic calques are not contemplated as an individual error category (something possible, however, in the lexis and terminology category for semantic calques).

Lexically, the most common errors are due to the influence of a formally similar word in the ST, wrong selection of collocations and the choice of single or multiword terms. Table 5 contains five illustrative cases in this respect:

ST	TT (Student 1)	TT (Student 2)
The key ingredients of flirtatious banter are all very English.		Los ingredientes principales de las bromas coquetas son todos muy ingleses.
Sad geek.		Friki penoso.
Here is a verbatim extract from a typical flirtatious encounter.	A continuación se encuentra un extracto con palabras textuales ...	
I was collecting examples of real-life chat-up routines .		Estaba coleccionando ejemplos de conversaciones en la vida real.
Recorded on a bus.	Grabado en un autobús.	

Table 5. Lexical and terminological errors

The source language interference reveals itself in solutions such as *coleccionando* for “collecting” (*recopilando* would be a more appropriate option) or *grabado* for “recorded”, since in this context the recording action suggests preservation or registration in writing rather than in sound or digital means. *Friki penoso* is not a happy collocation in Spanish to describe “sad geek”. Other lexical errors are *bromas coquetas* for “flirtatious banter”, *se encuentra* for “Here is” (*se incluye/se ofrece* are better alternatives) or *palabras textuales* for “verbatim” (‘literal’).

But as anticipated above, it is the style and situational context part that reveals the largest number of examples, since the target texts offer clumsy chunks that need some reformulation, either because the text is stylistically dense or redundant, or because of the degree of (in)formality:

ST	TT (Student 1)	TT (Student 2)
The key ingredients of flirtatious banter .	Los elementos claves del banter para ligar .	
... communicate their feelings for each other without ever saying what they really mean decirse lo que sienten el uno por el otro sin decir jamás lo que realmente quieren decir para que se comuniquen mutuamente sin decir lo que realmente quieren decir .
An item .	Este par de dos .	
Subsequent discussion.	Conversación subsecuente .	
Real-life chat-up routines.	de hechos reales sobre hábitos de seducción .	

Table 6. Style and situational context errors

Banter para ligar can be regarded as a redundant phrase since *para ligar* is unnecessary for a full understanding of the sentence. The second example in Table 6 reflects two cases of a stylistically heavy text, mainly due to the repetition of the verb *decir*, which contrasts with the lexical variety in the ST (“communicate”, “say”, “mean”). The three remaining cases clearly display instances of words that are too informal for the context (*par de dos*) or too formal for the context provided, as *subsecuente* for “subsequent” or *hábitos de seducción* (‘seduction habits’) for “chat-up routines”.

6. Discussion and conclusions

The use of multilingual corpora in translation has proved to be highly beneficial as a valuable source for investigation. In this sense, the MUST Project offers a helpful compilation of bilingual parallel translations done by learners with multiple possibilities not only for research purposes but also for educational applications. By resorting to a sample of two humanistic translations (English-Spanish) carried out by trainee translators, and using the TAS annotating system proposed by MUST, this paper has attempted to look into the weaknesses and strengths of the participants’ tasks, and determine to what extent the three translation competences under study (following the taxonomy by PACTE)

were present in them. Despite the brevity of the sample, some conclusions can be drawn from the contrastive analysis:

The ST-TT transfer seems to be less problematic in the texts selected, with just a few instances of distortion. By contrast, the Language Part gathers the largest amount of inappropriate solutions: while, roughly speaking, cohesion, punctuation and spelling are not a primary concern, complications arise when grammar, lexis and terminology, as well as style come into play. In this respect, S2 is the one with most difficulties grammatically but, contrary to expected, stylistically she shows a better performance than S1 (whose mother tongue is precisely Spanish). Both translations contain lexical inaccuracies, in most cases possibly due to SL interference, although the inappropriate use of documentation sources might well be behind some wrong decisions (i. e. *chacoteo*). Hence, and following PACTE, particular attention must be paid to the participants' bilingual sub-competence (mainly language register aspects related to vocabulary and syntax) and their knowledge about translation techniques (above all in the cases where negative transfer is present). The resulting translations, however, do not reflect failure to comply with the extra-linguistic sub-competence.

Regarding the application of the TAS system provided by MUST, this has proved to be an advantageous tool in the course of assessing translations, as it allows for exhaustive tagging. Nevertheless, some pitfalls have emerged during the annotation process: to start with, the possibility of tagging right choices is not contemplated at this stage without resorting to meta-tags, which makes it difficult to assess, in a simpler manner, the strengths of the students' performance. Also, double annotation is not easy to implement, which brings about problems when various errors gather in the same word/phrase, as the annotator can subjectively contemplate these errors from different perspectives: for instance, the term "banter" can be rendered using an equivalent Spanish word or sentence, but it can also be kept as such in the TT provided its meaning is clearly understood from the context (for example by means of compensatory strategies). In the same way, mistakes can have an origin different from all the options provided by the TAS system; for example, the wrong use of typographical conventions for dialogue construction by S1 can be contemplated as a Language problem but also as a result of (negative) source language interference in the transfer process, a possibility that is not envisaged here (again, meta-tags appear as

the only solution to this aspect). This leads to a further concern, which is the possible variety of interpretations (and consequently of tags) and the amount of time that must be devoted to carrying out a comprehensive annotation analysis of a single translated text.

Studies based on many more texts will undoubtedly offer diverse possibilities of analysis, as they will be able to detect different types of errors depending on the factors contemplated (for example the participants' mother tongue or the language they have as L2). More than that, the didactic prospect of the corpus once the annotation process is completed is endless, given its potential for (i) the design for translation class materials, (ii) class use during the process of translation instruction and text correction and (iii) use by students as an online tool for untutored practice, apart from its implications about its potential validity for diverse text types and the subjects' level of expertise, in an attempt to help students acquire their own competence (and that could be subsequently extrapolated to the professional field).

From a research perspective, the application of this tool is unquestionably beneficial to pinpoint the intervening factors in the process of translation, as well as the reasons lying behind the resulting taxonomy of errors.

Bibliography

- Beeby, Allison; Rodríguez-Inés, Patricia & Sánchez-Gijón, Pilar. 2009. Introduction. In Beeby, Allison; Rodríguez-Inés, Patricia & Sánchez-Gijón, Pilar (eds.) *Corpus Use and Translating*. Amsterdam & Philadelphia: John Benjamins, 1-8.
- Bowker, Lynne & Bennison, Peter. 2003. Student translation archive: Design, development and application. In Zanettin, Federico; Bernardini, Silvia & Stewart, Dominic (eds.) *Corpora in Translator Education*. Manchester: St. Jerome, 103-117.
- Carrasco, José Andrés. 2019. Teaching *English for Translation and Interpreting*: A framework of reference for developing the translator's bilingual sub-competence. *Complutense Journal of English Studies* 27. In press.
- Fantinuoli, Claudio & Zanettin, Federico. 2015. *New directions in corpus-based translation studies*. Berlin: Language Science Press. https://play.google.com/books/reader?id=t_arCQAAQBAJ&hl=es&pg=GBS.PA1 [Accessed April 2019].

- Fictumova, Jarmila *et alii*. 2014. Translation error typology for quality feedback: Czeching MeLLANGE through HYPAL. *Proceedings of TIFO 2014: Translation and Interpreting Forum Olomouc*.
- Fictumova, Jarmila *et alii*. 2017. Teaching Specialized Translation. Error-tagged Translation Learner Corpora. *Sendebär* 28: 209-241.
- Fox, Kate. 2004. *Watching the English: The Hidden Rules of English Behaviour*. London: Hodder Headline.
- Granger, Sylviane & Lefer, Marie-Aude. 2017. MUST: A collaborative corpus collection initiative for translation teaching and research. In Granger, Sylviane; Lefer, Marie-Aude & Aguiar de Souza Penha Marion, Laura (eds.) *Book of Abstracts. Using Corpora in Contrastive and Translation Studies Conference* (5th edition). Louvain-la-Neuve: Centre for English Corpus Linguistics / Université Catholique de Louvain, 72-73.
- Granger, Sylviane & Lefer, Marie-Aude. 2018. The Translation-oriented Annotation System: A tripartite annotation system for translation research. *International Symposium, PaCor*. Madrid, November 5-7.
- Granger, Sylviane; Lefer, Marie-Aude & Aguiar de Souza Penha Marion, Laura. 2018. *MUST Translation-oriented Annotation System (TAS) Version 1.0*. Internal document.
- Jiménez Crespo, Miguel Ángel. 2017. *Crowdsourcing and Online Collaborative Translations*. Amsterdam: John Benjamins.
- Johansson, Stig. 2007. *Seeing through Multilingual Corpora. On the Use of Corpora in Contrastive Studies*. Amsterdam: John Benjamins.
- Kruger, Alet; Wallmach, Kim & Munday, Jeremy. 2011. *Corpus-Based Translation Studies. Research and Applications*. London: Bloomsbury.
- Malmkjaer, Kirsten. 2009. What is translation competence? *Revue française de linguistique appliquée* 1(XIV): 121-134.
- Obrusnik, Adam. 2014. Hypal: A user-friendly tool for automatic parallel text alignment and error tagging. *Eleventh International Conference Teaching and Language Corpora*, 67-69.
- Olohan, Maeve. 2004. *Introducing Corpora in Translation Studies*. London: Routledge.
- PACTE. 2001. La competencia traductora y su adquisición. *Quaderns. Revista de Traducció* 6: 39-45.
- PACTE. 2003. Building a translation competence model. In Alves, Fabio (ed.) *Triangulating Translation: Perspectives in process-oriented research*. Amsterdam: John Benjamins, 43-66.
- Rica, Juan Pedro. (In press). *Corpus-based Studies and Audiovisual Translation: Subtitling*. Series: New Trends in Translation Studies (edited by J. Díaz Cintas). Frankfurt: Peter Lang.

- Rodríguez-Inés, Patricia. 2009. Evaluating the process and not just the product when using corpora in translator education. In Beeby, Allison; Rodríguez-Inés, Patricia & Sánchez-Gijón, Pilar (eds.) *Corpus Use and Translating*. Amsterdam/Philadelphia: John Benjamins, 129-149.
- Sánchez-Gijón, Pilar. 2009. Developing Documentation Skills to build do-it-yourself corpora in the specialised translation course. In Beeby, Allison; Rodríguez-Inés, Patricia & Sánchez-Gijón, Pilar (eds.) *Corpus Use and Translating*. Amsterdam/Philadelphia: John Benjamins, 109-128.
- Volk, Martin. 2019. Innovations in parallel corpus alignment and retrieval. In Doval, Irene & Sánchez Nieto, Teresa (eds.) *Parallel Corpora for Contrastive and Translation Studies*. Amsterdam/Philadelphia: John Benjamins, 79-50.
- Zanettin, Federico. 2014. Corpora in Translation. In House, Juliane (ed.) *Translation: A Multidisciplinary Approach*. Basingstoke: Palgrave Macmillan, 178-199.
- Zanettin, Federico. 2016. *Translation-Driven Corpora: Corpus Resources for Descriptive and Applied Translation Studies* (Translation Practices Explained). London: Routledge.

Appendix A: source text

The Rules of Banter

In most other cultures, flirtation and courtship involve exchanges of compliments: among the English, you are more likely to hear exchanges of insults. Well, mock-insults, to be precise. ‘Banter’, we call it, and it is one of our most popular forms of verbal interaction generally (on a par with moaning), as well as our main flirting method. The key ingredients of flirtatious banter are all very English: humour, particularly irony; wordplay; argument; cynicism; mock-aggression; teasing; indirectness – all our favourite things. And banter specifically excludes all the things we don’t like and that make us uncomfortable: emotion, sopiness, earnestness and clarity.

The rules of flirtatious banter allow courting couples to communicate their feelings for each other without ever saying what they really mean, which would be embarrassing. In fact, the banter rules require them to say the opposite of what they mean – something at which the English excel. Here is a verbatim extract from a typical flirtatious encounter, recorded on a bus, between two teenagers. The exchange was conducted in full view and hearing of a group of their friends.

‘You gotta licence for that shirt? Or are you wearing it for a bet?’

‘Huh! Look who’s talking – I can see your knickers, you slag!’

‘It’s a thong, you nerd – not that you’d know the difference. And that’s the closest you’ll get to it.’

‘Who says I’d want to? What makes you think I fancy you? You’re such a slag!’

‘Better than being a sad geek!’

‘Bitch!’

‘Geek!’

‘Sla – Oh, that’s my stop – you coming out later?’

‘Yeah – come round about eight.’

‘Right.’

‘Bye.’

From the conversation among their friends afterwards, it was clear that this pair had been attracted to each other for some time, had just

started ‘sort of going out’ together (in that rather vague, non-dating way the English do these things), and were expected to become ‘an item’ in the near future. Even if I had not heard this subsequent discussion, I would have recognized the exchange of insults as a typical flirtation – perhaps not the wittiest or most articulate flirtatious banter I’ve come across, but a normal, unremarkable, everyday English courtship sequence. I only recorded it in my notebook because I happened to be doing a study on flirting at the time, and was collecting examples of real-life chat-up routines.