Learner Translations in Contrast: An English-Basque-Spanish Case Study

Traducciones de aprendices en contraste: un estudio de caso inglés-euskera-castellano

Marlén Izquierdo
marlen.izquierdo@ehu.eus
Universidad del País Vasco (UPV/EHU)

Naroa Zubillaga
naroa.zubillaga@ehu.eus
Universidad del País Vasco (UPV/EHU)

Abstract: This article reports on a contrastive descriptive study of learner translations into Basque and Spanish of the same source text originally written in English. It is a preliminary case study with a twofold aim: first, to bring to contrast translations into different target languages (TLs) that have been produced under the same circumstances; and two, to identify and test the usefulness of corpus metadata in the interpretation of the translation product. The study was conducted within the framework of the MUST (MUltilingual Student Translation) project. Accordingly, all learner translations were first aligned, and then annotated with regard to three different error categories: transfer, language, and meta-text. The juxtaposition and contrast of the translational phenomena identified in each target language showed that the Basque data set reveals more language-related problems, while the Spanish data set features slightly more transfer-related errors. In interpreting the results, we benefited from the learners' metadata our MUST corpora are enriched with. In particular, we observed that those learners who considered Basque to be their one and only mother tongue performed the best in terms of language.

Resumen: El presente artículo da cuenta de los resultados derivados de un estudio descriptivo-contrastivo de traducciones de estudiantes al euskera y al castellano de un mismo texto originariamente escrito en inglés. El objetivo de dicho estudio preliminar era doble: primero, contrastar traducciones a

distintas Lenguas Meta (LMs) producidas en circunstancias iguales; y segundo, identificar y testar la utilidad de un corpus de metadatos en la interpretación de la traducción en cuanto producto. El estudio se realizó en el marco del proyecto MUST (MUltilingual Student Translation). Para ello, en primer lugar, se alinearon todas las traducciones de los estudiantes y, a continuación, se realizó la anotación de tres categorías de error diferentes: transferencia, lengua y metatexto. La yuxtaposición y el contraste de los fenómenos traductológicos identificados en cada lengua meta mostraron que el grupo con lengua meta euskera revela más problemas relacionados con la lengua, mientras que el grupo con lengua meta español presenta ligeramente más errores asociados con la transferencia de contenido. En la interpretación de los datos, nos beneficiamos de los metadatos de los estudiantes con los que están enriquecidos nuestros corpus MUST. En concreto, observamos que los alumnos que consideraban el euskera como su única lengua materna eran los que obtenían mejores resultados lingüísticos.

Palabras clave: Estudio descriptivo-contrastivo de traducción, Corpus de aprendices, Inglés-español-euskera, Diglosia

Introduction

Empirical approaches to translation research have traditionally focused on product-oriented studies that observe and describe real world translation phenomena from a corpus approach. TRACE, TRALIMA-ITZULIK or ACTRES are some examples of research groups working on corpus-based translation studies in Spain. Echoing Olohan's (2003) call for "contextualising translation by combining corpus-based investigations with other kinds of methodologies and analyses" (p. 419), this study tries to broaden the scope of empirical research by combining methodologies proper of two neighbouring disciplines, namely, corpus-based contrastive linguistics (CBCL) and corpus-based translation studies (CBTS). The former explores differences and similarities between at least two languages on the basis of a *tertium comparationis* that enables the contrast. The latter examines strategies, resources, and norms in the rendering of content from a source language (SL) into, at least, a target one (TL). The notion of equivalence is key to the quality evaluation of translated texts.

Many scholars have underlined the need to integrate in such empirical observations more social, contextual, and cognitive data (De Sutter and Lefer, 2019). In this regard, for a few years now the convergence between corpus-based and process-oriented translation studies is shaping current empirical translation studies (TS) (Kotze, 2019), thus connecting the three branches of TS, namely, product-, process-, and function-oriented research (Holmes, 1972). This effort requires new-generation corpora that are "more carefully

designed to take consideration of translators' backgrounds and the circumstances of text production" (Kotze, 2019, p. 356). A prime example of such an endeavour is the MUltilingual Student Translation (MUST) project, a learner translation corpus enriched with standardised metadata related to the source text, the translation, and the students (Granger and Lefer, 2020). It has been within the framework of this project that members of the TRALIMA-ITZULIK research group have started to investigate translations made by students from various languages, such as English, German, or Spanish, into Basque (Sanz Villar, 2024; Izquierdo Fernández and Zubillaga Gómez, 2022). Accordingly, this study aims to contribute to empirical translation research into Basque with a contrastive descriptive translation case study from a learner corpus approach, a novel type of study in the field. In particular, we will compare the translations of the same source text into Basque and Spanish as rendered by translator trainees to answer the following questions:

> i) Does a given source text (ST) pose the same problems to learners translating it into two contact languages?

> ii) Do the same chunks trigger the same or different translational phenomena in each target language (TL)?

Considering the diglossia situation in which the languages under contrast are used for translation and/or other communicative purposes, and benefiting from the students' metadata, the current study will interpret the findings in the light of the learners' linguistic background to answer our third research question:

> iii) To what extent do translation products differ depending on whether the learner's mother tongue is Basque (EU) or Spanish (ES)?

1.   CORPUS-BASED CONTRASTIVE LINGUISTICS AND TRANSLATION STUDIES

The disciplines of contrastive linguistics (CL) and translation studies (TS) have long been related as two branches of "interlingual linguistics" that enabled empirical approaches to language through contrast and/or contact (James, 1980). Ever since the scientific development −and settlement− of corpus linguistics in the 1990s, CL and TS have evolved over the course of time in a parallel and complementary way, shaping the field of cross-linguistic research (Izquierdo Fernández, 2007). The envisioned potential of corpus linguistics "in refining understanding of how languages relate to one another" (Granger *et al*., 2003) has taken shape in various ways: in a large number of tailored, periodic conferences (*e.g.* UCCTS); in widely-cited, top-edited publications (Baker, 1995; Granger *et al*., 2003; Doval Reixaand Sánchez-Nieto, 2019; Lavid-López *et al*., 2021), and journals (*Languages in Contrast*, *Target*, *IJCL*, *Corpora*, or *Learner Corpus Research*, to name few); in a wealth

of open-ended research (Granger and Lefer, 2020); and, most importantly, through a steady increase of applications developed within both fields (see Doval Reixa and Sánchez-Nieto, 2019). As a result, a great deal of present-day cross-linguistic research is by definition corpus-based, relying on various types of corpora to fine-tune theories with descriptive adequacy.

Both comparable and parallel corpora have played a key role in the cross-fertilization of CL and TS, in various ways: i) unveiling language-related aspects that might escape the human eye but are observable through comparable/parallel concordance data; ii) conferring robustness and maturity to empirical research thanks to statistics, corpus-driven techniques, and systematised annotation schemes at various levels; iii) opening up research avenues by combining different types of data; and iv) enabling data triangulation (Marco Borillo, 2019), among other possibilities.

A productive and extended procedure in joint corpus-based contrastive linguistics (CBCL) and corpus-based translation studies (CBTS), to which we will refer for short as CBCL-and-CBTS, entails the use of a comparable corpus to analyze a given language pair, and the use of a parallel corpus to examine the translation from one of the languages into the other (Rabadán Álvarez, 2007). Insights from the comparable contrastive analysis are then considered control data in the interpretation of parallel-based findings. Less common, however, has been the use of translated data as contrastive data. CBCL-and-CBTS research, on the other hand, is no longer restricted to a pair of languages; multilingual corpora are more and more frequently used in cross-linguistic research (Hunston, 2002; Johansson, 2007). Moreover, various directions of analysis are considered (*e.g.* PAGES project), as well as various modes of communication (*e.g.* EPTIC project). Most importantly, the nature of the language data under analysis has also diversified, with learner language being a central object of study from the point of view of CBCL-and-CBTS. The use of learner corpora has contributed positively to a reapproachment between the abovementioned disciplines, not only between one another, but also with the field of language pedagogy. As Granger and Lefer (2021) state, learner corpora are believed to strengthen synergies between contrastive linguistics and translation studies, as the present study seeks to prove.

To add one more trend that draws from all the interdisciplinarity above, this case study reports on a contrastive descriptive translation study, where the same source text in English (EN) is translated into two target languages, namely, Basque (EU) and Spanish (ES) by translator trainees. In other words, we will contrast students' translations arising from the same source text in two contact languages, a type of analysis never attempted, to the best of our knowledge.

2. LEARNER CORPORA IN EMPIRICAL TRANSLATION RESEARCH

As repositories of authentic, acted-out cross-linguistic correspondences, parallel corpora, as well as comparable corpora, are of direct benefit not only to translators but also to learners of a foreign language, be it for general purposes or for specific purposes, such as translation (Bowker, 1999). Notwithstanding this advantage, to truly progress towards a meaningful teaching-learning experience it is necessary to observe first-hand what learner language use is like. In short, the more learner data we examine, the more insights we gain into their linguistic/translator competence and, therefore, the easier it might be to raise student awareness of their learning/training needs.

Accordingly, building corpora of students' translations is motivated by the need to extend the fruitful combination of learner corpus research (LCR) and CBTS. Among the first initiatives to combine translator training research and CBTS, Bowker's corpora created by translators (CCBT) stands out as "a type of learner corpora that can be used to investigate difficulties encountered by trainee translators" (Bowker, 2003, p. 169). Similarly, the Translation Teaching and Learning Corpus (TTLC) project was carried out to "give priority to actual learner needs and integrate both language-based and process-based translating skills" (Tiayon, 2004, p. 119). Other ensuing successful projects worth highlighting have been the UPF learner translation corpus, featuring the language pair English-Catalan (Espunya Prat, 2014); special mention deserves the undergraduate learner translator corpus (ULTC) that features translations from English or French into Arabic (Alfuraih, 2020). This kind of proposals unquestionably meant a breakthrough in empirical translation research, even though there was still a lot of room for improvement. The language combinations lacked variety, translations were analysed in only one direction, and inconsistencies or lack of systematicity in error-annotation prevented the comparability of findings across research projects, let alone a generalization of results with real pedagogical application (Granger and Lefer, 2020). In addition, learner corpus research focused most of its attention on L2 teaching-learning, to the detriment of translation training. This apparent neglect, however, would nevertheless benefit student translation corpora in the years to come, making available for researchers long-tested and widely attested protocols in corpus design, methodology, analysis, and application.

In order to overcome the abovementioned shortcomings, the Multilingual Student Translation project (MUST) was launched in 2016 as an ambitious, yet well-thought, initiative that has grown over the past years and at the time of writing brings together researchers in the fields of translation, contrastive linguistics, and language/translation pedagogy from 38 universities (Granger and Lefer, 2020). The MUST project provides

researchers with relevant metadata about students that might help understand the process and outcome of translation, as well as to help understand translation as a social process. Students' metadata are related to their mother tongues, the foreign languages they know, or the languages used in their education, among others. We are particularly interested in the students' linguistic background, given that the two target languages of our contrastive descriptive translation study, namely, Basque (EU) and Spanish (ES) are languages in contact in a diglossic environment. The relevance of this fact is twofold. First, the study of "languages in contact" has always been framed within cross-linguistic research (James, 1980; Izquierdo Fernández, 2007). Second, we would hypothesise that the social reality that shapes our students' communicative competence, as characterised by diglossia, might have an effect on the translation process, arguably projected or reflected on the translation outcome that we will juxtapose and contrast (see section 6).

3.   DIGLOSSIA IN THE BASQUE COUNTRY

The territory where Basque is spoken is divided in two nations (Spain and France) and three regions or communities. The sociolinguistic situation of the Basque language differs from one territory to the other, since the official status of the language is different, and the measures oriented to its promotion vary as well. In the Basque Autonomous Community (a.k.a. Euskadi), Basque is the official language together with Spanish and all children learn Basque at school. Nevertheless, since the approval of the 138/1983 Basque Law ruling the use of official languages at non-university education, there are three different educational models, namely A, B and D, according to which the role of Basque as the language medium of instruction differs. Out of these models, it is only Model D that provides integral teaching in Basque except for the subjects of Spanish and a foreign language, most usually English. This model has become today the most popular one (Urdalleta Lete, 2023). Nevertheless, while this formal instruction enables language acquisition, it does not guarantee a settled habitual use of the language. In fact, according to the latest sociolinguistic survey, 53% of the population over 16 can speak or understand Basque (Basque Government, 2016, p. 4) but just 20.5% use it on a daily basis and as their main language (Basque Government, 2016, p. 23).

The case of the Autonomous Region of Navarre is particularly complex. The region is divided into three linguistic zones, where the status of Basque varies from fully to partially or no official. Consequently, the presence of Basque at schools is enhanced to different degrees. Nowadays, 23.2% of people over 16 can speak or understand Basque in Navarre (Basque Government, 2016,p. 4), but just 6.6% use it as their main language (Basque Government, 2016, p. 23).

This diglossic situation has its impact on the actual use of, even attitude towards, the Basque language by learners. For example, Barnes Mason and García Fernández (2011) found out that most of the children at school age use Spanish regardless of the educational model. In this regard, extensive research on the issue of minority language acquisition has been done on Basque (Austin, 2009; Almgren and Manterola Garate, 2016), also in a trilingual context with English (Leonet *et al*., 2019). Nevertheless, to the best of our knowledge, not much attention has been paid to the role of Basque (in contrast with Spanish) in the development of translators' competence or students of translation training. The few studies analysing Basque as a language of translation that have been done so far focus on literary translations published by professional translators (Zubillaga Gómez et al., 2015; Zubillaga Gómez, 2016; Sanz Villar, 2018).

4. METHODOLOGY

Framed within the MUST project (see section 2), we carried out a description of translational options benefiting from source text (ST) to target text (TT) alignments on two parallel corpora. The study is mainly qualitative in that translations are annotated for errors, as well as for particularly good translational options. In addition, some quantitative data is provided for the sake of contrast. The last stage of the study considers the learners' linguistic background, which might explain the contrastive results of the translation annotation process.

In the following sub-sections, we will detail which data was used for the study, and the actual procedure followed to analyse it.

4.1. The Data

Data was retrieved from two MUST sub-corpora, *i.e*., English-Basque (EN-EU) and English-Spanish (EN-ES). Each of these corpora is a multiple translation corpus (Espunya Prat 2014), as there is more than one translation for the same ST. In turn, each corpus contains several translations as output of different learning activities.

For this particular study, we selected data from one specific *ad hoc* translation task, designed in exactly the same way for students translating into EU and those doing it into ES. The ST, common to both language pairs, was an opinion article written by Eli Davies and published by *The Guardian*, on its online version, in January 2021 (Davies, 2021). The topic dealt with in the text was the loneliness pandemic brought about by COVID19. The article was originally 994 words long, but it was shortened to a bit less than 600 words to make the translation task suitable for a continuous assessment activity. Students were given 105 minutes to accomplish the task during class time, in

a computer-equipped classroom, and were allowed to use tools and resources to assist their translation, such as online dictionaries, machine translation engines, or corpora, among others. This task would be marked, *i.e.*, it would add on to the students' final mark in the subject, a fact all the translator trainees were aware of. The task represented 15% of the total grade and was corrected by the teachers, who later provided the learners with feedback. Therefore, the task was realised in a parallel but independent way by learners enrolled in translation courses for each language pair. In this way, we managed to select parallel data with the greatest comparability degree possible, thus resembling a comparable parallel corpus (Hareide, 2019). Table 1 shows the corpora composition:

	EN-EU	EN-ES
No. translations	15	17
No. words/translated sample	7016	10547
No. words/translation	467.7	620.4
Degree year	3rd	3rd
Degree course	*Translation Practice IV: English-Basque I*	*Translation Practice IV: English-Spanish IV*
Task timing (minutes)	105	105
CAT Tools allowed?	Yes	Yes
For marking?	Yes	Yes

Table 1. Comparable parallel data sets
Source. Elaborated by the authors

To compile two parallel corpora that were in turn comparable, a *tertium comparationis* was established to enable the contrast of the two translational data sets. This was done on the basis of i) the students' profile described in demographic (gender, age, educational level, and training), as well as in linguistic terms (*e.g.*, years of L2 learning and bilingual mindset); ii) task instructions and conditions for completion (*e.g.*, time, setting, software used and grading); iii) size of data sets, notwithstanding the minor differences in the number of translations collected per language pair, *i.e.*, number of translations, and consequently number of types;[1] and a final aspect that warrants comparability of our data would be iv) data annotation, as described in the following section.

The EN-ES group included 17 students (3 males, 18%, and 14 females, 82%) enrolled in the third grade of Translation and Interpreting at the University of the Basque Country. 13 of them (76%) considered ES their L1,

---

[1]The number of translations does not reflect the number of students taking the course, but the number of the students who consented that their translations be included in MUST.

while the remaining 4 (24%) acknowledged both ES and EU as their mother tongue. The ES translated sample is over 10 thousand words large. The subject for which they conducted the translation task was *Prácticas de traducción IV: inglés-español IV* (Translation practice IV: English-Spanish IV). All 17 students in this group chose ES as their A language and EN as their B language.[2] Most of them came from the Basque Autonomous Community, which means that the majority knew some EU, as both ES and EU are compulsory in the Basque educational system. In fact, even the students who reported having only ES as their mother tongue (76%) have mostly been schooled in Basque, as indicated in the metadata.

Regarding the EN-EU group, it consisted of 15 students (3 males, 20%, and 12 females, 80%), also in the third grade of Translation and Interpreting at the University of the Basque Country. The EU translated sample amounts to slightly more than seven thousand words. The subject for which they conducted the translation task was *Itzulpen Praktikak IV: ingelesa-euskara I* (Translation practice IV: English-Basque I). While the number of translation practical courses into their, respective, A language is the same for both student sets, the opportunity to work on EN as a source language is thrice as big for trainees whose A language is Spanish when compared to those with Basque as their A language (Ministerio de Educación, Cultura y Deporte, 2011). On the basis of such a difference, we could hypothesise that the learners translating into Basque may commit more transfer-related errors, on the assumption that they have a poorer command of the source language (see section 6).

4.2. The procedure

We approached our analysis with the classical procedure followed in contrastive linguistics, namely, selection, description, juxtaposition, and contrast (Kreszwoski, 1990). The selection of our data has been detailed in 4.1. We will focus on the description, juxtaposition, and contrast of translational options in the following lines.

4.2.1. Description

At the second stage of our analyses, we annotated our data using the translation-oriented annotation system (TAS) tailored to the MUST project, and accessible through the Hypal4Must interface (Fictumova *et al*., 2017; Granger and Lefer, 2020). Although there is currently an upgraded second

---

[2]As is usual in Translation Studies at universities, all students have to choose their A, B and C languages when they enroll. "A" language is their 'first language', "B" the second one and "C" their third one. All subjects are organised in such a way that students translate from "B" and "C" into their "A" language, except for some inverse translation subjects, where students also practice translating from their "A" language into their "B" language.

version of the annotation system, namely TAS 2.0, this was released after we had already annotated our data using the very first version. TAS 1.0., as well as TAS 2.0., does not focus only on translational errors; it gives users the possibility of tagging translational options that are considered particularly good.

By employing TAS 1.0., we could classify translational options into four categories, namely, ST-TT transfer (TR), language (LA), translation procedures (TP), and meta-text (MT). At the time of annotation, the third category was still under construction, so we described the translational phenomena observed in the TTs using the categories of TR, LA, and MT. These categories represent a first level of annotation of our data.

Going further down to greater detail, every category breaks down into a second level of analysis where a number of sub-categories label a given type of TR, LA or MT error or phenomenon. As such, the TR category is further divided into 5 sub-categories as illustrated in Figure 1.
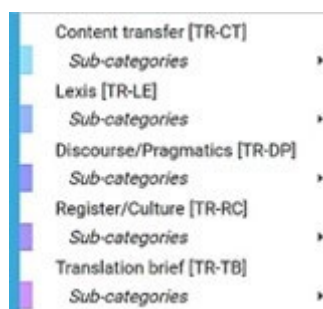

Figure 1. ST-TT transfer sub-categories
Source. Hypal4Must (Granger and Lefer, 2020)

The first sub-category specifies whether the ST-TT Transfer error is one affecting "content" [TR-CT]. The second one (Lexis TR-LE) would be useful to annotate that a wrong lexical translational option has been chosen. In case that the transfer triggered deviations at the level of discourse and/or pragmatics, the third sub-category (TR-DP) would mark it so. When deviations affect register and/or cultural aspects from ST to TT, there is also a possibility of annotation (TR-RC). Finally, the fifth sub-category within level-one transfer phenomena would relate to changes in the translation brief (TR-TB). However, our learners did not have to use one for their translation task.

Likewise, there are five LA sub-categories as shown in Figure 2. Every sub-category tags errors that take place at different levels of linguistic

analysis: grammar (LA-GR), lexis and terminology (LA-LT), cohesion (LA-CO), spelling and punctuation (LA-ME), and finally, style (LA-ST).

Grammar [LA-GR]
   Sub-categories   ›
Lexis & Teminology [LA-LT]
   Sub-categories   ›
Cohesion [LA-CO]
   Sub-categories   ›
Mechanics [LA-ME]
   Sub-categories   ›
Style and situational context [LA-ST]
   Sub-categories   ›

Figure 2. LA sub-categories
Source. Hypal4Must (Granger and Lefer, 2020)

The third category, MT, breaks down into two types of annotations (Figure 3):

Positive [MT-PLUS]
Suspected SL intrusion [MT-SLI]

Figure 3. MT sub -categories
Source. Hypal4Must (Granger and Lefer, 2020)

The first annotation addresses positive translational solutions. The other, by contrast, addresses translational solutions that are considered negative as suspected to stem from SL intrusion.

Each of the TR and LA sub-categories may be further detailed at a third level of delicacy; our data was annotated at all three levels, which was not free from difficulty, given the high granularity of TAS 1.0.

In fact, the description stage was rather challenging to carry out for two main reasons. First, as a three-tier translation-oriented annotation system, TAS 1.0. was rather complex to implement, reason why TAS 2.0. has been refined to diminish granularity and it features fewer sub-categories (Granger and Lefer, 2021). Second, the annotation of each language data set was to be done independently from each other, which could double the opportunity for discrepancy between language annotators. Accordingly, to conduct the annotation in an individual but parallel and comparable way, a basis for some degree of inter-rater consistency was laid. Aiming for the highest degree of inter-rater reliability, two coders first considered a few translations into each language together, to have the possibility of interpreting TAS 1.0. jointly, discussing their understanding of each tag, and deciding what TAS annotation

seemed to be appropriate for certain errors that were either common to both language pairs, or recurrent, or ambiguity-ridden. In short, we aimed to ensure coherence and consistency in our TAS management.

Once the description stage was completed, the number of annotations per language data set was 351 annotations in the EU translations and 399 annotations in the ES data set. On average, there are 21 annotations per text in the EN-EU translations and 22.1 in the EN-ES translations. There are, therefore, fewer annotations in the first combination than in the second. What this might be due to will be discussed further in the article (see section 5).

To ease readability, in this study, we will juxtapose annotations at the first and second level (see section 4.2.2.). In the proper contrastive stage (see section 5), we will be commenting on third-level annotations to discuss similarities and, most importantly, differences between Basque and Spanish translations of the same original English text.

4.2.2. Juxtaposition and contrast

In the bosom of the MUST project, a powerful interface and browsing software was developed to serve both teaching-learning and research purposes, namely, Hypal4Must. This is an extension of the *Hybrid Parallel Text Aligner* (Hypal), a web-based interface originally developed to align parallel texts (Obrusník, 2014). Although initially tested on the Czech-English pair, Hypal was conceived of as a language-independent tool, which has enabled its customisation for MUST purposes, now homing multiple corpora and a teaching tool section (Granger and Lefer, 2021). Hypal4Must is programmed to enable a number of corpus functionalities such as data upload, lemmatisation, and part-of-speech tagging; text alignment, at the paragraph and sentence level; text annotation; and corpus browsing, yielding both statistical as well as qualitative information.

Figure 4 juxtaposes information related to the number of annotations at the first level in each target language:
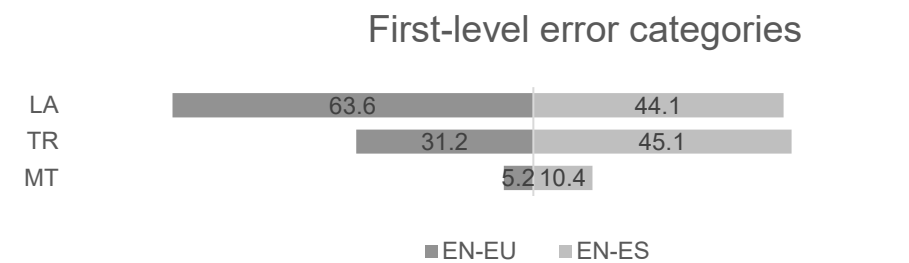
First-level error categories

	EN-EU	EN-ES
LA	63.6	44.1
TR	31.2	45.1
MT	5.2	10.4

■EN-EU   ■EN-ES

Figure 4. Juxtaposition of first-level annotations in EU and ES translations
Source. Elaborated by the authors

As illustrated, the LA-related annotations outnumber TR-related ones in EN-EU translations (left column). In fact, students having EU as their target language show 32.4% more language annotations (LA, 63.6%) than transfer annotations (TR, 31.2%). LA-related annotations rank second. On the other hand, far from there being a clear distinction between LA and TR errors in the EN-ES data set, these categories share a similar percentage; TR-related annotations represent 45.51%, while 44.08% of the annotations involve LA. In other words, there are very few more TR annotations (181) than LA ones (176). Finally, MT is clearly the least frequent category in both languages. Yet, a striking cross-linguistic difference is observed, for there are double the number of annotations in EN-ES than in EN-EU. A further difference lies in the phenomena tagged under such a category; while all MT instances in the EN-EU data set feature source-language intrusion cases into the TL, only in the EN-ES translations were remarkably good translational options also annotated.

Similar to Figure 4, Figure 5 juxtaposes the second-level annotations in each language-pair translation set:
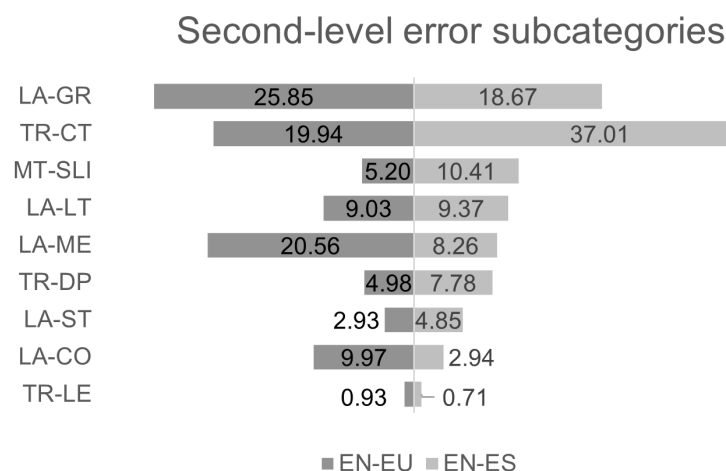
Second-level error subcategories

	EN-EU	EN-ES
LA-GR	25.85	18.67
TR-CT	19.94	37.01
MT-SLI	5.20	10.41
LA-LT	9.03	9.37
LA-ME	20.56	8.26
TR-DP	4.98	7.78
LA-ST	2.93	4.85
LA-CO	9.97	2.94
TR-LE	0.93	0.71

■ EN-EU ■ EN-ES

Figure 5. Juxtaposition of second-level annotations in EU and ES translations
Source. Elaborated by the authors

We observe that within the LA category the most recurrent annotation hints at grammar issues (LA-GR). Even though this is common to both data sets, grammatical errors stand out among EU student translations. Other grammatical errors were also identified in terms of the cohesion and coherence of the TT (LA-CO); lexis and terminology choices (LA-LT); mechanics of the text, *i.e.*, spelling and punctuation, among others (LA-ME); and issues having to do with the style of the translation/translator (LA-ST), or whether the TT feels redundant or heavy, among other possibilities. The occurrence of such errors differs across the two language pairs, with EU student translations displaying first LA-GR, followed by LA-ME, LA-CO, LA-LT, and LA-ST. In the EN-ES data set, however, LA-CO annotations represent the smallest group. LA-LT ranks second, followed by LA-ME and LA-ST categories. Table 2 illustrates such a diverging order of frequency of the same LA-bound error subcategories, together with examples:

	EN-EU	EN-ES	Examples[3]
LA-GR	1st	1st	
			(1) ST. I've been thinking a lot about loneliness over the last few years as I've drifted in and out of various forms of it myself.

---

[3] For the sake of contrast, we have chosen ST chunks that trigger errors in each TL whenever available.

			TT (EU). *Asko pentsatu dut bakardadeaz azken urteotan,* **ni neu** *ere bakardadearen hainbat aldaera pairatu baititut* [Much thought I have loneliness (about) last years (in the), me myself also loneliness (of) various forms suffered because - the subject "me myself" has to be in ergative form, *nik neuk*, as it is a subject of transitive verb]. TT (ES). *He estado reflexionando acerca de la soledad estos últimos años y he experimentado con ella y sus distintas formas* **hasta que experimentarla** *en su mayor esplendor, como es lógico, en 2020.* [...and its forms different **until that to experience** it...]
LA-ME	2nd	3rd	(2) ST. The historian Fay Bound Alberti, who has written a "biography" of the condition, argues that… TT (EU). *Fay Bound Alberti historialariak baldintzaren inguruko «biografia» idatzi du Ø eta pentsatzeko modu horrek...* [Fay Bound Alberti the historian the condition about "biography" written has Ø and to think the way this… - in Basque you would put a comma before 'and'] TT (ES). *El historiador Fay Bound Alberti Ø que ha escrito una "biografía" de la situación, sostiene...* [missing comma]
LA-CO	3rd	5th	(3) ST. …and perhaps that's why we find it so difficult to talk about or admit to. TT (EU). *…eta agian horrexegatik egiten zaigu hain zaila horren inguruan hitz egitea edota Ø onartzea* [...and perhaps that is why make it us so difficult about that to speak and to accept - the second infinitive verb *onartzea* or "accept" needs a pronoun in absolute form: *hura*]. TT (ES). *…y quizás es por eso nos resulta tan difícil hablar de ello o admitirlo* [linking "that" [QUE-conjunction] is missing between *eso* and *nos*].
LA-LT	4th	2nd	(4) ST. …by the global health crisis and its accompanying <u>lockdown</u>. TT (EU). ----- TT (ES). *…la crisis sanitaria mundial y el* **bloqueo** *que la acompañaba.* [...the crisis health worldwide and the **block** that escorted it]
LA-ST	5th	4th	(5) ST. …to <u>work on</u> my PhD

	TT (EU). ------ TT (ES). *…para **sacarme** un doctorado.* […to **take out** a PhD]	

Table 2. LA-error types in EU and ES: order of frequency and examples[4]
Source. Elaborated by the authors

On the other hand, the various types of annotations tagging errors related to the actual transfer from the SL to the TL are equally distributed in EN-EU and EN-ES student translations. Among the various TR errors identified, content (TR-CT) annotations stand first, followed by annotations referring to problems in the building of the TT at the level of discourse and pragmatics (TR-DP). Finally, the smallest percentage of transfer-related annotations relates to the lexical choice made to render a given meaning from the ST into the TT (TR-LE). Even though TAS 1.0. broke down this category into more error types referring to register and cultural issues (TR-RC), or the translation brief (TR-TB), none of them were observed in our translations.

	EN-EU	EN-ES	Examples
TR-CT	1st	1st	(6) ST. I've been thinking a lot about loneliness over the last few years as I've drifted in and out of various forms of it myself. TT (EU). *Azken urteotan **bakardadean asko pentsatu dut**, ni neu bakardadeak dituen hainbat formetatik sartu eta atera bainaiz* [Last years **alone much thought I have**,.... – "thinking alone" is something different as "thinking about"] TT (ES). *He estado reflexionando acerca de la soledad estos últimos años **y** he experimentado con ella y sus distintas formas hasta que experimentarla en su mayor esplendor, como es lógico, en 2020.* […about loneliness these last years **and** have experienced with it…]
TR-DP	2nd	2nd	(7) ST. Like so many living alone during lockdown, I've felt incredibly isolated. TT (EU). ------ TT (ES). *Yo también me he sentido apartado durante el confinamiento, al igual que muchas personas.* [I too have felt casted during the lockdown, same as many people - inverse clause order!]

---

[4] For the sake of contrast, we have chosen ST chunks that trigger errors in each TL whenever available.

TR-LE	3rd	3rd	(8) ST. Pages appeared on the NHS and Red Cross websites…. TT (EU). **NHS eta Gurutze Gorriko webguneetan...** [NHS and Red Cross(of) websites… - NHS has not been translated] TT (ES). *Aparecieron secciones páginas web del NHS y en la Cruz Roja que….* [Appear sections pages web of the NHS and…]
TR-RC	-	-	
TR-TB	-	-	

Table 3. Order of frequency of TR-error types in EU and ES
Source. Elaborated by the authors

Example (6) shows a change of content in the transfer from original English into target Basque and Spanish. In the TT (EU) example, the translation means that it is "while being alone" (*bakardadean*) that a lot of thinking has been done (*asko pentsatu dut*). Although affecting another chunk of the original sentence, a distortion of content is also observed in TT (ES) whereby a subordinate causal clause, "…**as** I've drifted…" is rendered into Spanish as a complex clause introduced by coordinating conjunct *y* ("and"). Not only is the content changed, but also the actual discourse, for the underlying causal relation between main and subordinate clauses of the ST disappears altogether in the TT (ES), with a coordination of two clauses at the same propositional level. This phenomenon is also observed in (7) TT (ES), where the order of the two clauses is inverted. Finally, Example (8) illustrates the least frequent translational option in either language, which consists in leaving an original item untranslated. The instances found among our data affect mainly the abbreviation "NHS", *i.e.*, National Health System, which is left as such in both translation data sets.

5. DISCUSSION

As an answer to our first research question, our data reveals that the same ST does not necessarily pose the same problems for students translating it into two −contact − target languages. Cross-linguistically, the distribution of annotations per error categories would suggest that trainees working into Basque have had more difficulties dealing with the target language, for Basque translations seem to be linguistically worse done than the Spanish translations. Such results would not refute a hypothesis posited earlier in the article (see section 4.1.) as to whether the EN-EU learners could have greater difficulties in decoding the original text, given that they have not worked on English as a source language as much as their EN-ES counterparts.

By contrast, Spanish translations appear to deviate from the ST in terms of source content to a greater extent, which would hint at the learners' difficulties in the craft of "transferring" meanings. Let us discuss the overall difference in the categorization of the annotations per language pair.

First, the translator trainees' competence in the target language may account for the nature of the errors tagged, mainly those with a high recurrence rate and observed to be common to all translations in the data set. This is believed to be so in the case of grammar problems among the Basque translations, where they stand out. In particular, the most frequently annotated language error in the EN-EU group concerned grammar inflection (15.5 % of all the annotations). Basque is a highly inflected language, which turns out to be a challenge when it comes to acquiring the language, whether as an L1 or L2 (Ezeizabarrena Segurola, 2012). The noun-phrase is inflected in 17 different ways for case and it is, furthermore, an ergative-absolutive language, where the case used for the agent of a transitive verb is ergative, formally marked by *-k*. This is exactly what many students have forgotten in their translations, as illustrated in Example (1) in Table 2, where the subject should have been *ni**k** neuk*. The result is, therefore, an ungrammatical construction.

EN-ES grammar problems, which rank second in frequency after lexical and terminological issues (LA-LT), affect mainly the choice of verb tense, rendering not an ungrammatical construction but mostly an inaccurate translation as (9) illustrates.

(9) ST. I've felt incredibly isolated.

   TT (ES). ***Me sentí*** *increíblemente aislada.*

On the other hand, most LA-related errors in this data set have to do with the translation of multiword and idiomatic expressions. Where the ST features an idiomatic expression, trainees are expected to use an equivalent expression in the TL whenever possible. This has not always been the case, as exemplified in the translation of "struck a chord" in (10).

(10) ST. When I read these words, they struck a chord.

   TT (ES) a. *Cuando escuché estas palabras, me tocaron la fibra sensible.*

   TT (ES) b. *Estas palabras me conmovieron*.

While some students produced an equivalent term, in meaning and form, such as *tocar la fibra sensible* (TT [ES] a), other options were single-word verbs conveying only the meaning, as in (TT [ES] b).

Flawed interpretation of the source text may also give shape to the actual annotations. Distortion of the original content is clearly a pervasive problem for the learners translating into Basque as well as for those working into Spanish. Most often, this distortion yields an inaccurate transfer as observed in Example (6) in Table 3. The chunk "I've been thinking over loneliness" has been mistranslated into Basque by more than half the students. We suspect that the students have actually understood the meaning of the ST but not managed to render it in Basque. In fact, we would argue that the actual translational option stems from cognitive interference of Spanish; the English phrase "think about loneliness" is equivalent to Spanish *pensar en la soledad*, literally, 'think in the loneliness'. Arguably, the Spanish preposition is negatively transferred into the Basque translations. In a bilingual scenario, this error could be interpreted in the light of "L2 intrusion," a tag that would be parallel to TAS "source language intrusion" (MT-SLI). Considering the diglossic status of Basque and Spanish, we would claim that the dominant language intrudes, affecting both translation process and product (Sanz Villar, 2018). As a matter of fact, 17 instances of MT-SLI were found in the EN-EU data.

Arguably, the fact that the "well translated" perspective (MT-PLUS) was taken into account only when tagging EN-ES translations, as opposed to the EN-EU translations where only the MT-SLI cases were acknowledged, would be due to a rater's "idiosyncratic vision of good practice" (Graham *et al*., 2012, p. 4). In other words, the researchers did not show equal awareness of the need or relevance of marking good translational options, which hints at a slightly different interpretation of the corpus data.

Aware that many variables may shape the translational output of two different language data sets, in our second research question we aimed to observe whether one given chunk would trigger the same or different translational phenomena in each TL. To this end, we narrowed our search to the fifth paragraph of the text, as this seemed to be rather problematic in both translation tasks. It triggered 84 annotations in the EN-ES group and 78 in the EN-EU group. Our corpus is aligned at the sentence level, so Table 4 shows the sentences composing the fifth paragraph with the overall number of corresponding annotations:

5th paragraph sentences	EN-EU	EN-ES
A few months into the Covid-19 outbreak, people started to talk about a **corresponding loneliness pandemic**.	LA-ME (4) LA-GR (3) MT-SLI (2) LA-CO (1) TR-CT (1)	LA-LT (4) LA-GR (3) LA-ME (1) LA-ST (1) TR-CT (4)

	LA-LT (1) **12**	TR-DP (1) **14**
Pages appeared on the **NHS** and Red Cross websites advising how to cope with the isolation thrust upon us by the global health crisis and its accompanying lockdown.	MT-SLI (8) TR-LE (7) LA-ME (5) LA-GR (3) LA-LT(1) **24**	LA-GR (6) LA-ME (2) LA-ST (1) LA-LT (1) TR-CT (5) TR-LE (4) TR-DP (1) MT-SLI (5) **25**
But before this year, various reports claimed that **loneliness had reached dangerous and even life-threatening epidemic levels**, and in 2018 Theresa May launched a UK government **loneliness strategy.**	LA-GR (9) LA-ME (9) TR-CT (3) **21**	LA-GR (14) LA-LT (2) LA-ST (1) LA-ME (3) TR-CT (7) TR-DP (2) **29**
Such concerns have always been particularly heightened during winter and around Christmas, a time when charities and politicians frequently urge festive revellers to think about and reach out to the lonely and vulnerable.	LA-GR (7) LA-ME (6) LA-LT (4) LA-CO (2) MT-SLI (2) **21**	LA-GR (5) LA-LT (3) LA-ME (2) LA-ST (1) MT-PLUS (3) TR-CT (1) TR-DP (1) **16**
TOTAL ANNOTATIONS	**78**	**84**

Table 4. Annotations of most problematic paragraph and common problem-triggers (in bold) into EU and ES
Source. Elaborated by the authors

The order of the first-level annotation categories differs between the groups; there are 55 LA (70.5%), 12 MT (15.4%), and 11 TR (14.1%) annotations among the Basque translations. In the Spanish data set, LA stands first with 50 (59.5%) annotations, followed by 26 TR (31%) cases and finally 8 MT (9.5%) annotations. Likewise, the most recurrent LA error committed by Basque learners is one either of spelling or punctuation (LA-ME), very closely followed by grammatical mistakes (LA-GR). By contrast, this paragraph triggered mainly LA-GR problems into Spanish, followed by errors at the lexical and terminological level (LA-LT). Greater differences were observed within the TR category, with Basque translations featuring mainly lexical errors (TR-LE), as opposed to a majority of content distortions (TR-CT) in the Spanish translation of the paragraph. Finally, for once the MT annotations in the Basque translations outnumber those in the Spanish data, with the qualitative difference that the former report only source language

intrusion while the latter required tagging occurrences of "good translations" (MT-PLUS).

On the other hand, we observe discrepancies but also similarities if we focus exclusively on the chunks that posed problems into both TLs, namely, "corresponding loneliness pandemic" (sentence 1), "pages appeared" and "NHS" (sentence 2), and "loneliness had reached dangerous and even life-threatening epidemic levels" and "loneliness strategy" (sentence 3).

Cross-linguistically, "corresponding loneliness pandemic" has been dealtwith differently. While "corresponding" seems to be the problem for Basque trainees, many Spanish translations reveal problems at the compound level.

TT (EU). *COVID19aren agerralditik hilabete batzuetara, birusarekin zetorren "bakardadearen pandemiaz" hitz egiten hasi zen* [COVID19 (of) outbreak months some (after), virus (with) coming 'loneliness (of) pandemic' (about) to speak started – the phrase "the loneliness pandemic coming with the virus" seems not natural in Basque].

TT (ES). …*hablar de una correspondiente pandemia solitaria.*

"Pages appeared" and "NHS", either together or independently, have triggered the same kind of problems into both EU and ES. Many translations of the first chunk are mostly tagged as MT-SLI, in view that they sound "literal" irrespective of the TL. By contrast, the second item has remained untranslated several times (Example 12):

TT (EU). **NHS** *eta Gurutze Gorriko webguneetan orrialdeak* **agertu ziren** [NHS and Red Cross (of) websites (on) pages appeared – "NHS" has not been translated and "pages appeared" does not sound natural].

TT (ES). **Aparecieron páginas** *en los sitios web del* **NHS** *y…*

The translation of the two items in the third sentence has brought to light the learners' difficulties with phraseological complexity, as already seen in Example (11). As a result, ill-built phrases abound among both translation data sets, mostly featuring non-grammatical solutions that would call into question the learners' competence in the TL.

TT (EU). *hainbat txostenek zioten bakardadea epidemia mailara iritsi zela eta* **arriskutsuak zein hilgarria** *zela* [various reports said loneliness (the) epidemic level (to) arrived and dangerous (plural) as well as lethal was – both adjectives "dangerous" and "lethal" do not agree in the same way: plural and singular].

TT (ES). *la soledad había alcanzado niveles peligrosos **e incluso, amenazantes epidemiológicos…***

In Example (14), the equivalent compound in either language would require a preposition that is either an ill choice or missing in the target text.

TT (EU). *Theresa May momentu horretan lehen ministroa zenak «**bakardadearen estrategi**» izeneko kanpaina jarri zuen abian* [Theresa May moment that (in) prime minister was (who) "loneliness (of) strategy" called campaign put forward – when reading "strategy of loneliness", one could think that the aim is reaching loneliness].

TT (ES). *...en 2018 Theresa May lanzó una **"estrategia de soledad"** del gobierno británico.*

In the light of the data so far discussed, it would not seem safe to state that one given chunk gives way to the same kind of errors by learners working into two contact languages. Notwithstanding this difference, it has been observed that certain chunks operating at the lexical level are more likely to lead to similar errors.

The high rate of LA errors among the EU translations has been an interesting finding. Related to this, our third and last research question goes beyond the translation product and aims to delve into the translation process, on the assumption that the linguistic background of the trainees might explain their decisions and the translational options so far identified. Indeed, the issue of what a language user's mother tongue is in a diglossic context is fraught with controversy as much as it is filled with provoking thoughts. As briefly explained in 3 above, the sociolinguistics of the Basque Country is rather complex. For our research purposes, it pertains to acknowledge that the diglossic language situation in which Basque is spoken is one where Spanish is felt as dominant or "high" language, whereas Basque would be the dominated or "low" language (Ferguson, 1959). This diglossic situation has its impact on the actual use of, even attitude towards, the Basque language by learners. In fact, it may trigger different feelings. On the one hand, Basque being the original language of the territory, it may be felt as the only mother tongue, even though all Basques are unquestionably competent in Spanish. On the other hand, there are also people who feel comfortable in both languages, and others who barely use Basque although they have most likely been formally instructed in the language, especially among the age range our learners would belong to. Taking this complex sociolinguistic situation into account, we decided to look closer at those metadata about the mother tongue in the EN-EU group and relate them to the annotations for language. Figure 6 illustrates the type and percentage of first-level TAS categories per EN-EU trainee. Out of the 15 students, four of them, namely students 4654, 4655,

4656, 4659, reported that Basque was their one and only mother tongue. Such students "consume" Basque daily, in various facets of life; Basque is their formal medium of instruction, their home, peer, and leisure language.
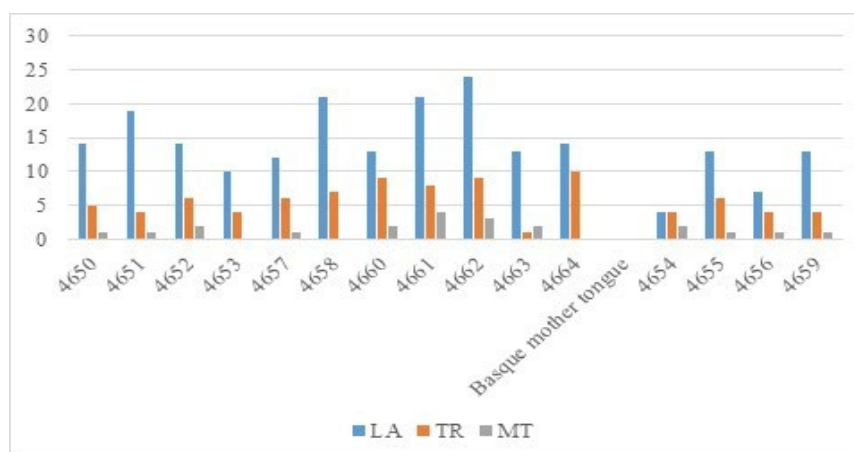

Figure 6. TAS first-level annotations per EN-EU trainee
Source. Elaborated by the authors

As can be seen, it is two of the students who claimed Basque to be their only mother tongue that have received the smallest number of language annotations (students 4654 and 4656). Accordingly, we would argue that, in a diglossic environment, knowing the "low" language does not guarantee a proficient command by the speakers. Rather a native-like command stemming from an all-round use of the language appears to be essential. Previous research on language acquisition corroborates this fact (Ezeizabarrena Segurola *et al.*, 2009; Ezeizabarrena Segurola, 2012; Zawiszewski *et al*, 2011). It seems, therefore, that exposure to the language, its varieties and different registers is a key factor to master the language. Given the exploratory nature of this case study, these results cannot be generalised. However, they hint at a possibility worthy of further exploration; our contrastive descriptive study might contribute to the field of empirical translation studies by identifying a niche of research where the actual amount and type of use of contact languages in bilingual settings and the language quality of translation students may be correlated.

CONCLUSION

The aim of this article was to report the preliminary results of a contrastive descriptive translation study from a learner corpus approach. Within the framework of the MUST project, two sets of students were asked

to translate the same source text from English into Basque and/or Spanish, the learners' assumed mother tongue or the A language they chose in their university degree. Generally speaking, translation trainees working into Spanish seem to have more problems with content transfer, while Basque students show difficulties at using the target language. Nevertheless, it is also observed that translation at the lexical level is prone to less divergence, across target languages, than at the discourse level.

Another conclusion we reach is that TAS 1.0 cannot escape the subjectivity of the human eye that necessarily examines each and every translation in search of errors and/or options worth annotating. As a matter of fact, TAS 2.0. has deleted some levels of annotation as well as ambiguous tags, displaying less granularity. In the same way that human skills and limitations give shape to a translation, it is a human mind with an idiosyncratically evaluative attitude that judges the degree of correctness, acceptability and/or appropriateness of the translation under correction.

In addition to describing the actual translational options suggested by the learners, we also aimed to understand the outstanding difficulties in producing grammatically correct Basque translations. At this stage, the students' metadata our corpora are enriched with were essential to figure out that a poorer exposure to the Basque language in the diglossic context our learners live and learn might account for the many LA-related problems. Being the dominated language, Basque is not experienced as much or as diversely as is Spanish. On the other hand, being the dominant language, Spanish may intrude in the trainee's Basque output as a consequence of some cognitive interference in bilingual settings.

Our study has been an exploratory one, yet it contributes to existing research in various ways. First, it adds a new trend in contrastive studies whereby the translations into two target languages of the same source text are contrasted. In addition, it integrates learner data into contrastive and translation studies. Most importantly, it has taken a step further by trying to understand translation products from a wider perspective, using the trainees' metadata as indicative of the process of translation. Where the trainees claim Basque to be their one and only L1, they perform better. Finally, as expected, this preliminary study has shed light on some limitations that need addressing before embarking on larger contrastive descriptions of learner translations. For example, our metadata being common to the MUST project, it has been insufficient to fully reflect the sociolinguistic or socio-political status of minority languages like Basque. Even a clear explanation regarding what L1, L2 or foreign language should be understood as to trainees seems appropriate prior to collecting their metadata. On the light of the results, we also believe that understanding TAS tags is one part of the annotation process only; foreseeing

the co-text where certain tags may be relevant is also a factor for the coders to agree on.

Finally, further research may be taken on the results obtained. For example, a similar analysis of larger phraseological data would attest whether it is really the case that cross-linguistic differences at the lexical level are fewer. Likewise, a comparison between TAS-annotated human and machine translations could help us understand how determinant the human variable may be. On a pedagogical note, the same Hypal4Must software we have used to process and annotate the translations remains to be used in the translation class. The benefit of this would be for the trainees to learn from their own translations as well as previously done learner translations. Browsing a students' corpus, no matter the language combinations, trainees would become aware of potential equivalents, common errors, and a range of translational strategies or solutions.

REFERENCES

Alfuraih, R. F. (2020). The undergraduate learner translator corpus: a new resource for translation studies and computational linguistics. Lang Resources & Evaluation, 54, 801-830. https://doi.org/10.1007/s10579-019-09472-6

Almgren, M., & Manterola Garate, I. (2016). The development of narrative skills in learners of Basque as a second language. Education Inquiry, 7(1), 27-46. https://doi.org/10.3402/edui.v7.27627

Austin, J. (2009). Delay, interference and bilingual development: The acquisition of verbal morphology in children learning Basque and Spanish. International Journal of Bilingualism, 13(4), 447-479. 10.1177/1367006909353234

Baker, M. (1995). Corpora in translation studies: An overview and some suggestions for future research. Target, 7(2), 223-243. https://doi.org/10.1075/target.7.2.03bak

Barnes Mason, J., & García Fernández, I. (2011). Vocabulary growth and composition in monolingual and bilingual Basque infants and toddlers. International Journal of Bilingualism, 17(3), 357-374. 10.1177/1367006912438992

Basque Government. (1983). Decreto 138/1983, por el que se regula el uso de las lenguas oficiales en la enseñanza no universitaria en el País Vasco. https://www.legegunea.euskadi.eus/eli/es-pv/d/1983/07/11/138/dof/spa/html/webleg00-contfich/es/

Basque Government. (2016). VI. inkesta soziolinguistikoa. https://www.irekia.euskadi.eus/uploads/attachments/9954/VI_INK_SOZLG-EH_eus.pdf?1499236557

Bernal, J. M. (2001). Diglosia y funciones sociales de las lenguas en Grecia (1830-1941). Minerva: Revista de Filología Clásica, 15, 115-116.

Bowker, L. (1999). Exploring the potential of corpora for raising language awareness in student translators. Language Awareness, 8(3-4), 160-173. https://doi.org/10.1080/09658419908667126

Bowker, L. (2003). Corpus-based applications for translator training: exploring the possibilities. In S. Granger, J. Lerot, & S. Petch-Tyson (Eds.) Corpus-based approaches to contrastive linguistics and translation studies (pp. 169-183). Rodopi. https://doi.org/10.1163/9789004486638_014

Davies, E. (4th of January, 2021). To solve the problem of loneliness, society needs to look beyond the nuclear family. The Guardian. https://www.theguardian.com/commentisfree/2021/jan/04/loneliness-society-nuclear-family-lockdown-communities

De Sutter, G., & Lefer, M.A. (2019). On the need for a new research agenda for corpus-based translation studies: A multi-methodological, multifactorial and interdisciplinary approach. Perspectives, 28(1), 1-23. https://doi.org/10.1080/0907676X.2019.1611891

Doval Reixa, I., & Sánchez-Nieto, M. (Eds.). (2019). Parallel corpora for contrastive and translation studies. New resources and applications. John Benjamins.

Espunya Prat, A. (2014). The UPF learner translation corpus as a resource for translator training. Lang Resources & Evaluation, 48, 33-43. https://doi.org/10.1007/s10579-013-9260-1

Ezeizabarrena Segurola, M. J. (2012). The acquisition of the (in)consistent ergative marking in Basque: L1 and early L2. Lingua, 122(3), 303-317. https://doi.org/10.1016/j.lingua.2011.11.009

Ezeizabarrena Segurola, M. J., Manterola Garate, I., & Beloki Lizarralde, L. (2009). Euskara H2 goiztiarraren ezaugarrien bila: adizkiak eta gramatika-kasuak haurren ipuin-kontaketetan. Euskera, 54(2-1), 639-681.

Ferguson, C. (1959). Diglossia. Word, 15, 325-340. https://doi.org/10.1080/00437956.1959.11659702

Fictumova, J., Obrusnik, A., & Stepankova, K. (2017). Teaching specialised translation. Error-tagged translation learner corpora. Sendebar, 28, 209-241.

Graham, M. J., Milanowski, A. T., & Miller, J. B. (2012). Measuring and promoting inter-rater agreement of teacher and principal performance ratings. Center for Educator Compensation Reform. https://files.eric.ed.gov/fulltext/ED532068.pdf

Granger, S., & Lefer, M. A. (Eds.). (2020). The complementary contribution of comparable and parallel corpora to crosslinguistic studies. Languages in Contrast, 20(2).

Granger, S., & Lefer, M. A. (October, 2021). Translation-oriented annotation system manual version 2.0. Centre for English Corpus Linguistics – UCLouvain.

Granger, S., Lerto, J., & Petch-Tyson, S. (2003). Corpus-based approaches to contrastive linguistics and translation studies. Rodopi.

Hareide, L. (2019). Comparable parallel corpora: A critical review of current practices in corpus-based translation studies. In I. Doval Reixa, & M. Sánchez-Nieto (Eds.), Parallel Corpora for contrastive and translation studies (pp. 19-38). John Benjamins.

Holmes, J. (1972). The name and nature of translation studies. In J. Holmes (Ed.), Translated! Papers on literary translation and translation studies (pp. 67-80). Rodopi.

Hunston, S. (2002). Corpora in applied linguistics. CUP.

Izquierdo Fernández, M. (2007). Corpus-based cross-linguistic research: directions and applications. Interlingüística 17, 520-527.

Izquierdo Fernández, M, & Zubillaga Gómez (2022). Empirical translation studies: contrasting learner translations in a diglossic environment. Paper presented at the *6th Learner Corpus Research Conference, 22-24 September 2022, Padova (Italy)*.

James, C. (1980). Contrastive analysis. Longman.

Johansson, S. (2007). Seeing through multilingual corpora: on the use of corpora in contrastive studies. John Benjamins.

Kotze, H. (2019). Converging what and how to find out why: An outlook on empirical translation studies. In H. Kotze (Ed.), New empirical perspectives on translation and interpreting (pp. 333-371). Routledge.

Kreszwoski, T. P. (1990). Contrasting languages. The scope of contrastive linguistics. Mouton de Gruyter.

Lavid-López, J., Maíz-Arévalo, C., & Zamorano-Mansilla, J. R. (Eds.). (2021). Corpora in translation and contrastive research in the digital age. John Benjamins.

Leonet, O., Cenoz Iragui, J., & Gorter, D. (2017). Challenging minority language isolation: translanguaging in a trilingual school in the Basque Country. Journal of Language, Identity and Education, 16(4), 216-227. https://doi.org/10.1080/15348458.2017.1328281

Marco Borillo, J. (2019). Living with parallel corpora. The potentials and limitations of their use in translation research. In I. Doval, & M. Sánchez-Nieto (Eds.), Parallel corpora for contrastive and translation studies. New resources and applications (pp. 39-56). John Benjamins.

Ministerio de Educación, Cultura y Deporte. (2011). Memoria de grado de Traducción e Interpretación en la UPV/EHU. https://gestion-alumnos.ehu.es/tmp/Memoria%20Verificada%2006-02-2014.pdf

Obrusník, A. (2014). Hypal: A user friendly tool for automatic parallel text alignment and error-tagging. Paper for the 11th Teaching and Language Corpora Conference (TALC 2014).

Olohan, M. (2003). How frequent are the contractions?: A study of contracted forms in the translational English corpus. Target, 15(1), 59-89. 10.1075/target.15.1.04olo

Rabadán Álvarez, R. (2007). Divisions, description and applications: The interface between DTS, corpus-based research and contrastive

analysis. In Y. Gambier, M. Shlesinger, & R. Stolze (Eds.), Doubts and directions in translation studies: Selected contributions from the EST Congress, Lisbon 2004 (pp. 237-252). John Benjamins

SanzVillar, Z. (2018). Interference and the translation of phraseological units in a parallel and multilingual corpus. META, 63(1), 72-93. https://doi.org/10.7202/1050515ar

Sanz Villar, Z. (2024). German-to-Basque translation analysis of multiword expressions in a learner translation corpus. Íkala, Revista De Lenguaje y Cultura, 29(1), 1-21. https://doi.org/10.17533/udea.ikala.354417

Tiayon, C. (2004). Corpora in translation teaching and learning. Language Matters, 35(1), 119-132. https://doi.org/10.1080/10228190408566207

Urdalleta Lete, I. (8 September, 2023). Hegoaldeko hamar ikasletik seik ikasiko dute D ereduan. Berria. https://www.berria.eus/euskal-herria/hegoaldeko-hamar-ikasletik-seik-ikasiko-dute-d-ereduan_1341536_102.html

Zawiszewski, A., Gutiérrez Sigut, E., Fernández Fernández, B., & Laka Mugarza, I. (2011). Language distance and non-native syntactic processing: Evidence from event-related potentials. Bilingualism: Language and Cognition, 14(3), 400-411. 10.1017/S1366728910000350

Zubillaga Gómez, N. (2016). (In)direct offense. A comparison of direct and indirect translations of German offensive language into Basque. Perspectives: Studies in Translatology, 24(3), 486-497. 10.1080/0907676X.2015.1069858

Zubillaga Gómez, N, Sanz Villar, Z. & Uribarri Zenekorta, I. (2015). Building a trilingual parallel corpus to analyse literary translations from German into Basque. In C. Fantinuoli & F. Zanettin (Eds.) New directions in corpus-based translation studies (pp. 71–93). Language Science Press.

APPENDIX 1

Source text (ST) provided to the learners for their translation into Basque and Spanish:

To solve the problem of loneliness, society needs to look beyond the nuclear family

Eli Davies

Like so many living alone during lockdown, I've felt incredibly isolated. It's time to rethink the way communities work.

"I have been trying, for some time now, to find dignity in my loneliness," wrote the poet and critic Maggie Nelson in her 2009 book Bluets. When I read these words, they struck a chord.

I've been thinking a lot about loneliness over the last few years as I've drifted in and out of various forms of it myself, the most extreme form coming, unsurprisingly, in 2020. Is there dignity to be had in it? Perhaps not, and perhaps that's why we find it so difficult to talk about or admit to.

A few months into the Covid-19 outbreak, people started to talk about a corresponding "loneliness pandemic". Pages appeared on the NHS and Red Cross websites advising how to cope with the isolation thrust upon us by the global health crisis and its accompanying lockdown. But before this year, various reports claimed that loneliness had reached dangerous and even life-threatening epidemic levels, and in 2018 Theresa May launched a UK government "loneliness strategy". Such concerns have always been particularly heightened during winter and around Christmas, a time when charities and politicians frequently urge festive revellers to think about and reach out to the lonely and vulnerable.

There isn't much talk in all this, though, of what loneliness actually is, what it feels like or where it comes from. In these scenarios it is an affliction: distant, othered and slightly frightening, coming to us in the form of elderly people at Christmas, the recently widowed, those who are unloved or forgotten about. But thinking about it as some kind of disease is wrongheaded. The historian Fay Bound Alberti, who has written a "biography" of the condition, argues that this way of thinking suggests "that it's coming from the outside, rather than being something that is a social problem."

Loneliness, then, is partly produced by the way we organise the world – and to address it we need to seriously rethink how we approach our public spaces, housing arrangements and relationships. This includes questioning our dependency on certain forms of relationship – the couple and the nuclear family – as units of social organisation.

I first started properly thinking about all this around the summer of 2017. I had recently come out of a 12-year relationship and, not unrelatedly, had moved to Ireland to work on my PhD, that most solitary of endeavours. After years of coupled domesticity, I was living alone. Solitude is not the same as loneliness, of course – as Nelson puts it, "loneliness is solitude with a problem" – and some of this was OK: I read, I walked, I wrote, I went out and made new

friends. But my isolation, coupled with the rawness of a recent heartbreak, frequently was a problem.

At such times I would message friends and family back home, filling my phone screen with three, four, five WhatsApp chats, and in these exchanges they described their own struggles: too much to do, not enough time or space for themselves, an excess of people and stuff to look after. The contrast in our predicaments seemed, at times, completely absurd and, above all, wasteful. I often wondered whether it wouldn't make sense to bolt my household on to one of theirs, to redistribute some of my caring resources and they, in turn, to share some of the human company that I often craved.

As indicated, this is a shortened version of the original text, which can be found at To solve the problem of loneliness, society needs to look beyond the nuclear family | Eli Davies | The Guardian.