

GERMAN-TO-BASQUE TRANSLATION ANALYSIS OF MULTIWORD EXPRESSIONS IN A LEARNER TRANSLATION CORPUS

ANÁLISIS TRADUCTOLÓGICO (ALEMÁN-VASCO) DE UNIDADES FRASEOLÓGICAS
EN UN CORPUS DE APRENDICES DE TRADUCCIÓN

UNE ANALYSE DE LA TRADUCTION DES UNITÉS PHRASÉOLOGIQUES ENTRE L'ALLEMAND
ET LE BASQUE DANS UN CORPUS DES TRADUCTIONS DES ÉTUDIANTS

ANÁLISE DA TRADUÇÃO DE UNIDADES FRASEOLÓGICAS DO ALEMÃO PARA O BASCO
NUM CORPUS DE TRADUÇÕES DE ALUNOS

Zuriñe Sanz-Villar

Associate Professor, University
of the Basque Country/ Euskal
Herriko Unibertsitatea —UPV/EHU,
Autonomous Community of the
Basque Country, Spain.
zurine.sanz@ehu.eus
[https://orcid.
org/0000-0002-2281-2574](https://orcid.org/0000-0002-2281-2574)

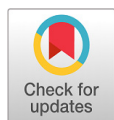
ABSTRACT

This paper presents the results of the annotation in a learner translation corpus consisting of German source texts and student translations to Basque. The analysis was carried out with the purpose of identifying trainee translators' strengths and weaknesses when translating multiword expressions, such as compounds, collocations, and idioms. The data comprised eight German source texts and sixty-eight Basque translations from undergraduate students enrolled at the University of the Basque Country. From the total number of annotations (1214), which include not only errors but also cases of interference from the source language and positive outcomes, around 27 % are related to multiword expressions. The results of the translation analysis show that there are variables — such as the use of machine translation systems, the level of specialisation of the source text, the type of multiword expression to be translated or the absence of a literal counterpart in the target language — that may affect the translation of such units and lead to erroneous solutions and/or interference in the outputs produced by the trainee translators. From a pedagogical point of view, these findings will have a direct impact on the translation classes and will be very valuable for designing corpus-based in-class activities.

Keywords: trainee translators, German-to-Basque translation, translation analysis, multiword expressions, phraseological units, learner translation corpus

RESUMEN

Este artículo presenta los resultados de las anotaciones en un corpus de aprendices de traducción compuesto por textos originales en alemán y traducciones de estudiantes al vasco. El objetivo del análisis fue identificar las fortalezas y debilidades de los traductores en formación al traducir unidades fraseológicas, como palabras compuestas, colocaciones y locuciones. El conjunto de datos estuvo formado por



Received: 2023-07-26 / Accepted: 2023-11-03 / Published: 2024-01-31

<https://doi.org/10.17533/udea.ikala.354417>

Editor: Luanda Sito, Universidad de Antioquia, Medellín, Colombia.

Copyright, Universidad de Antioquia, 2024. This is an open access article, distributed in compliance with the terms of the Creative Commons license BY-NC-SA 4.0 International.



Íkala, Revista de Lenguaje y Cultura

MEDELLÍN, COLOMBIA, VOL. 29 ISSUE 1 (JANUARY-APRIL, 2024), PP. 1-21, ISSN 0123-3432

www.udea.edu.co/ikala

ocho textos originales en alemán y 68 traducciones al vasco realizadas por estudiantes de grado de la Universidad del País Vasco (UPV/EHU). Del número total de anotaciones (1 214), entre las que se encuentran no solo errores, sino también casos de interferencia de la lengua de origen y resultados positivos, cerca de un 27 % tiene que ver con unidades fraseológicas. Los resultados del análisis traductológico muestran que hay variables —como el uso de sistemas de traducción automática, el grado de especialidad del texto origen, el tipo de unidad fraseológica o la ausencia de un equivalente literal en el idioma de llegada— que pueden influir en la traducción de dichas unidades y llevar a soluciones erróneas o a casos de interferencia en las traducciones realizadas por los aprendices de traducción. Desde un punto de vista pedagógico, estos hallazgos tendrán un impacto directo en las clases de traducción y serán muy valiosos para el diseño de actividades basadas en corpus.

Palabras clave: aprendices de traducción, traducción del alemán al vasco, análisis de traducciones, unidades fraseológicas, corpus de aprendices de traducción

RÉSUMÉ

Cet article présente les résultats des annotations sur un corpus de traductions d'étudiants composé de textes sources allemands et de traductions d'étudiants en basque. L'analyse visait à identifier les forces et les faiblesses des traducteurs en formation lorsqu'ils traduisent des unités phraséologiques, tels que des composés, des collocations et des idiomes. L'ensemble de données comprenait huit textes sources en allemand et 68 traductions en basque réalisées par des étudiants de l'Université du Pays basque (UPV/EHU). Sur le nombre total d'annotations (1 214), qui comprennent non seulement des erreurs mais aussi des cas d'interférence avec la langue source et des résultats positifs, environ 27 % concernent des unités phraséologiques. Les résultats de l'analyse de la traduction font apparaître des variables —telles que l'utilisation de systèmes de traduction automatique, le degré de spécialisation du texte source, le type d'unité phraséologique à traduire ou l'absence d'équivalent littéral dans la langue cible— qui peuvent influencer la traduction de ces unités et conduire à des solutions erronées ou à des interférences avec les produits des traducteurs en formation. D'un point de vue pédagogique, ces résultats auront un impact direct sur les cours de traduction et seront très utiles pour la conception d'activités en classe basées sur des corpus.

Mots clef : traducteurs stagiaires, traduction de l'allemand vers le basque, analyse des traductions, unités phraséologiques, corpus des traductions des étudiants

RESUMO

Este artigo apresenta os resultados de apontamentos em um corpus de traduções de alunos que consiste em textos originais em alemão e traduções de alunos para o basco. A análise buscou identificar os pontos fortes e fracos dos tradutores estagiários ao traduzir unidades fraseológicas, como compostos, locuções e colocações. O conjunto de dados consistia em oito textos de origem em alemão e 68 traduções para o basco feitas por alunos de graduação da Universidade do País Basco (UPV/EHU). Do número total de anotações (1 214), que incluem não apenas erros, mas também casos de interferência do idioma de origem e resultados positivos, cerca de 27 % referem-se a unidades fraseológicas. Os resultados da análise da tradução mostram variáveis - como o uso de sistemas de tradução automática, o grau de especialização do texto de origem, o tipo de unidades fraseológicas a ser traduzidas ou a ausência de um equivalente literal no idioma de destino - que podem

influenciar a tradução dessas unidades e levar a soluções errôneas ou interferir nos produtos dos tradutores estagiários. Do ponto de vista pedagógico, essas descobertas terão um impacto direto nos cursos de tradução e serão muito valiosas para a elaboração de atividades em sala de aula baseadas em corpus.

Palavras chave: tradutores estagiários, tradução do alemão para o basco, análise de traduções, unidades fraseológicas, corpus de traduções de estudantes

Introduction

The translation of phraseological units¹ by trainee translators has been studied by different authors but is still an underexplored area of research (Sanz-Villar, 2022, p. 268; Serrano Lucas, 2010, pp. 197–198). In a study by Serrano Lucas (2010), a methodological proposal based on the task-based approach is made for teaching phraseology in the context of translation didactics. Leiva Rojo (2013) focuses on the importance of considering phraseological units when assessing and reviewing the quality of texts translated by translation students. In the context of general translation classes, Valero Cuadra (2015) and Albaladejo Martínez (2015) examine the translation of collocations by trainee translators. It is worth mentioning that both papers were published in a volume dedicated to phraseology, didactics and translations (Mogorrón Huerta & Navarro Domínguez, 2015). In the same year, Marcelo Wirnitzer and Amigo Extremeña (2015) presented a pilot study analysing the process of translating phraseological units —PUS— (more specifically, phrasal verbs, conversational routines, collocations and idioms) by trainee translators, as well as the product itself.

Castagnoli (2023) analyses variation in translations in a corpus of an English source text (ST) and 35 translations to Italian produced by translation trainees; additional professional translators' outputs are included in the study. To this end, the translation paradigms of pre-selected items — idiomatic and non-idiomatic multiword units (MWUs), among others — were examined. One aim was “to observe the forms variation can take and how it may be related to the linguistic items involved as well as to individual translator experience” (2023, p. 120). In this respect, the author concludes that variation is greater in the translation of idiomatic units, especially when a literal

rendering of the ST item does not exist in the target text (TT; 2023, p. 120).

The first attempt to analyse the translation of multiword expressions (MWEs) by trainee translators in the language combination German-Basque was made by Sanz-Villar (2022). This study shows the results of a translation analysis conducted in a small corpus containing 24 different texts (6 838 tokens) and created with a tool called TALigner (TRALIMA-ITZULIK, 2019). The texts were collected over four academic years, from 2014 to 2018. Multiword verbs were pre-selected and manually extracted from the German STs, and their counterparts in students' translations were analysed. The results of that study showed that the insertion of words into the structure of the verbal patterns leads students to misunderstand the meaning of the ST and that both source language (SL) interference and interference from a third language is observed in students' translations (Sanz-Villar, 2022, pp. 283–284); not only because of the presence of this third language, Spanish, during the translation process, but also due to the use of resources for translators, such as machine translation (MT) systems.

Interference has also been observed in other studies analysing translations to Basque. The corpus created by Sanz-Villar (2018) contained German STs and professional translations to Basque, and it was concluded that “[w]hen translating from a prestigious language A to a minority language B, if language B coexists unequally with a dominant language C, then, according to different variables, different types of interference from language C into language B can occur to different degrees” (2018, p. 90). Aierbe Mendizabal (2008) analysed the translation of specialised phraseological units in administrative texts and noted that Basque is very dependent on Spanish administrative language and that the influence of Spanish on Basque TTs is far-reaching.

The present article analyses translations from German (A) to a minority language, Basque (B),

1 In this paper, the terms “phraseological unit” (PU) and “multiword expression/unit” (MWE and MWU, respectively) are used interchangeably.

which is in a diglossic situation with a dominant language, Spanish (C). Therefore, interference may be expected not only from the SL, but also from Spanish, further intensified in the students' translations due to the increasing use of MT systems. The lack of direct resources — understood as “sets of previously gathered linguistic data which are made available in some electronic format so that they can be used or looked up by translators” (Alcina, 2008, p. 98) — in German-to-Basque translations leads trainee translators in this language combination to depend on indirect resources (in the language combination German-Spanish or German-English).

This paper sets out to analyse the translation of MWES translated from German to Basque by trainee translators based on a learner translation corpus (LTC), which was compiled within the framework of the MUST (Multilingual Student Translation) project (Granger & Lefer, 2020). In 2019, several members of the TRALIMA-ITZULIK research group from the University of the Basque Country (UPV/EHU) joined this project, and have been contributing data for the following language combinations: Basque to Spanish, English to Basque/Spanish and German to Basque. The goal will be to present the general results of the annotations made in students' translations in terms of errors, interference, and positive outcomes, and then to focus on the translation analysis of MWES. First, these units are noteworthy due to the number of annotations related to MWES. Through the annotation process, course trainers can identify students' difficulties and design tasks to help them overcome their shortcomings (Espunya, 2013, p. 130). Secondly, these units are good candidates for analysing interference (Sanz-Villar, 2018).

Section 2 presents the MUST initiative, the Translation-oriented Annotation System (abbreviated as TAS) created within this project and the notion of MWES. Section 3 describes the compilation and annotation process of the LTC and the characteristics of the corpus regarding metadata. Section 4 outlines the results of the study, and it

will be concluded with a discussion on pedagogical implications based on the outcomes and with some final remarks (Section 5).

Theoretical Framework

This section presents the MUST initiative and mentions the TRALIMA-ITZULIK research group's contribution to this project. Following that, the annotation system (TAS 1.0) will be described as a significant component of MUST, with a focus on MWES.

Learner Translation Corpus

LTCs containing original texts and translations made either by foreign language learners or trainee translators can be regarded as the synergetic product of two previously separate fields — learner corpus research (LCR) and corpus-based translation studies (CBTS; Granger & Lefer, 2020, p. 1184). These two research strands both emerged in the late 1980s and early 1990s, but the first LTC was only created in the early 2000s. The idea of compiling corpora with students' translations is not new (*inter alia*, Castagnoli et al., 2011; Sánchez Nieto, 2012; Espunya, 2013; Wurm, 2013), but as argued by Granger and Lefer (2020, p. 1184), most of the corpora created have been local initiatives with the exception of MELLANGE (Multilingual eLearning in Language Engineering) and MUST (Multilingual Student Translation). In line with the MELLANGE corpus, the aim of the MUST project has been “to collect a large Multilingual Student Translation corpus, to design a rich set of standardised metadata, to create a standardised annotation system for translated language and to make all the data available to the contributing partners via a web-based interface for research and teaching” (Granger & Lefer, 2020, p. 1186).

Castagnoli *et al.* (2011, p. 237) identified the ten most frequent errors in the MELLANGE LTC, which includes 232 annotated translations to Catalan, German, English, Spanish, French, and Italian. As will be explained in the results section, it is noteworthy that the two most frequent

errors in the MELLANGE LTC concern incorrect lexis and terminology (first place) and distortion (second place). Also, Wurm (2013) concludes that lexical errors correspond to the most common category in her LTC, which is consistent with the results obtained within two other similar projects (2013, pp. 399-400).

As mentioned above, in 2019 members of the TRALIMA-ITZULIK research group joined the MUST initiative. Since this paper presents the results of the annotation of the German-to-Basque subcorpus, it is necessary to describe the annotation system.

TAS 1.0

The MUST project team uses the term “Translated-oriented Annotation System” rather than error annotation system because the taxonomy is not limited to errors but includes additional metatags — as explained in this section — and may in the future incorporate the option of tagging translation procedures (Granger & Lefer, 2018). In a large-scale project such as this, it was important to develop a flexible, language-independent system. Depending on the language pair and scope, it can be decided how detailed the annotation should be. The taxonomy should also be valid for both research and teaching purposes.

TAS 1.0 is based on two traditions: error taxonomies for ST-TT, such as the CELTRAC taxonomy (based on the abovementioned MELLANGE translation error typology) and annotation systems “developed for the analysis of learners’ free writing” (Granger & Lefer, 2020, p. 1194). It contains 60 tags and is divided into four main blocks, as can be observed in Figure 1: *ST-TT transfer*, *language*, *metatags*, and *translation procedures*. However, the latter is not used in this corpus since it was under preparation at the time of conducting the research. A second taxonomy, Translation-oriented Annotation System (TAS2.0), has existed since autumn 2021. However, the results presented here are based on the TAS1.0 taxonomy, the one available when annotating the mentioned subcorpus.

ST-TT transfer refers to “discrepancies between the source text (ST) and the target text (TT) and/or between the TT and the translation brief” (Granger & Lefer, 2018). Tagging language errors means identifying segments that are erroneous from the perspective of the target language (TL), independently of the ST. With the *Metatag* label it is possible to tag positive outcomes and/or suspected SL intrusion.

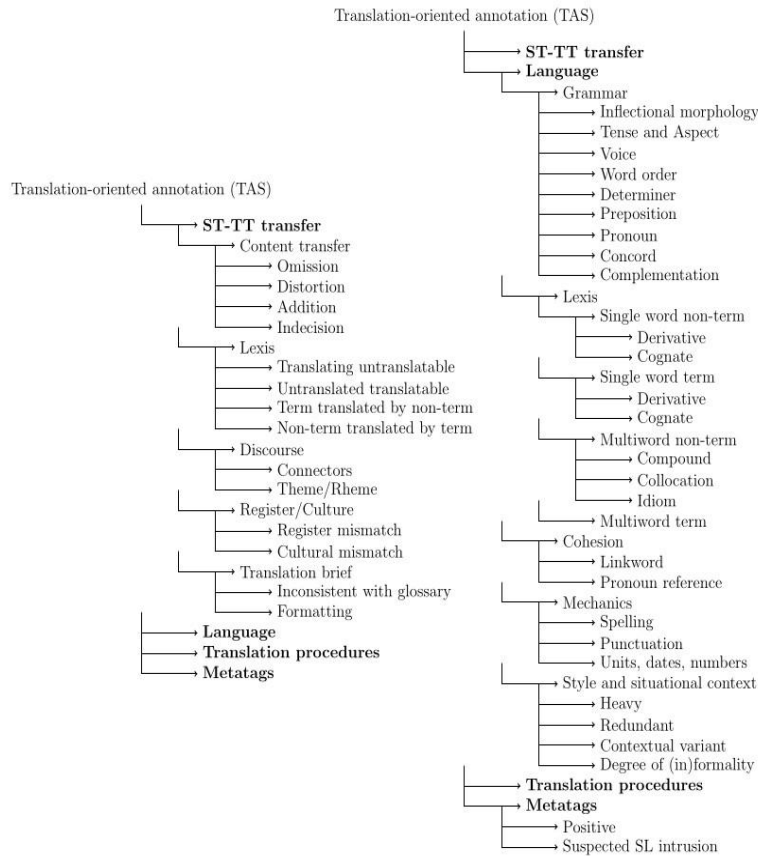
Subcategories are included for each of the main categories; that is, each category is multi-layer (Granger & Lefer, 2018), as shown in Figure 1. There are five subcategories for each of the two parts (*ST-TT transfer* and *Language*) of the TAS that serve to annotate errors: *Content transfer*, *Lexis*, *Discourse*, *Register/Culture*, and *Translation brief* in the former and *Grammar*, *Lexis*, *Cohesion*, *Mechanics*, and *Style and situational context* in the latter. Two subcategories are included in the *Metatags* block to indicate whether an output represents a good translation solution (*positive*) or whether the student’s translation solution has been negatively influenced by the SL (*suspected SL intrusion*). As mentioned in the TAS manual (Granger & Lefer, 2021, p. 28), the *suspected SL intrusion* metatag differs from the previously mentioned error tags in that it does not describe the nature of errors, but the possible source of an infelicitous translation solution. The annotation may be even more detailed, and the depth of the TAS is at the discretion of the annotator.

MWES in the TAS

There is still a lack of consensus on the use of the metalanguage in the field of phraseology, as mentioned, for example, in the METRAFAS project, whose aim is to establish a common terminology in the phraseological field and thus contribute to a more homogeneous use of the metalanguage.² As mentioned in Sanz-Villar (2022), “[t]he terminology used by researchers when naming and

2 Refer to point 4 of the project website for more details: <https://www.cirp.gal/proxectos/proxecto-fraseologia-galega.html>

Figure 1 Categories of TAS 1.0



defining their object of study varies depending on the approach or the way multi-word units (MWU) are analysed”.

Corpas Pastor’s well-known classification proposes three main phraseological spheres, namely collocations, idioms and phraseological utterances (Corpas Pastor, 1996). The first two are included in the TAS as subcategories of “multiword non-term”, as can be observed in Figure 1. Collocations are defined as “usage-determined or preferred syntagmatic relations between two lexemes in a specific syntactic pattern. Both lexemes make an isolable semantic contribution to the word combination but they do not have the same status. Semantically autonomous, the ‘base’ of a collocation is selected first by a language user for its independent meaning. The second element, i.e. the ‘collocate’ or ‘collocator’, is selected by and

semantically dependent on the base” (Granger & Paquot, 2008, p. 43). As for idioms, semantic non-compositionality is their most salient feature, but “[l]ack of flexibility and marked syntax are further indications of their idiomatic status” (Granger & Paquot, 2008, p. 43).

The TAS considers compounds as MWEs, although their inclusion within the field of phraseology is not uncontroversial and “pose[s] problems because of their [compounds’] uncertain status as single or multi-word units” (Granger & Paquot, 2008, p. 32). These are understood as “multiword units that constitute one semantic unit although they may be written in one, two or more words, as well as other fully fixed sequences such as complex prepositions, complex adverbs, complex conjunctions and complex verbs, including phrasal verbs” (Granger & Lefer, 2021, p. 21).

During the error tagging of the present study, it was found that the translation of some solid compounds — such as *Familienstand* (‘marital status’) — caused translation errors (as in *familia-egoera*, ‘family situation’, where the meaning of the original compound is not properly rendered in the translation). The inclusion or exclusion of these units in the framework of a translation analysis of MWES, while beyond the scope of this paper, remains an interesting area of research.

Method

As part of the MUST project, a web-based interface called Hypal4MUST, based on the Hypal software and developed by Obrusnik (2014), allowed

the creation and annotation of parallel corpora. Once corpora were compiled, the uploaded texts were queried within the tool’s interface. This section first describes the process of compiling the corpus. It then explains how the MWES were queried in the tool and exported for further analysis. Finally, the metadata concerning both translation tasks and participants will be described.

Compilation of Corpus

The first step consisted of creating the task (i.e. the STs were uploaded and accepted by the MUST coordinators). As can be seen in Figure 2, in the German-to-Basque subcorpus, eight tasks were created.

Figure 2 Uploaded and Accepted Task in the German-Basque Subcorpus

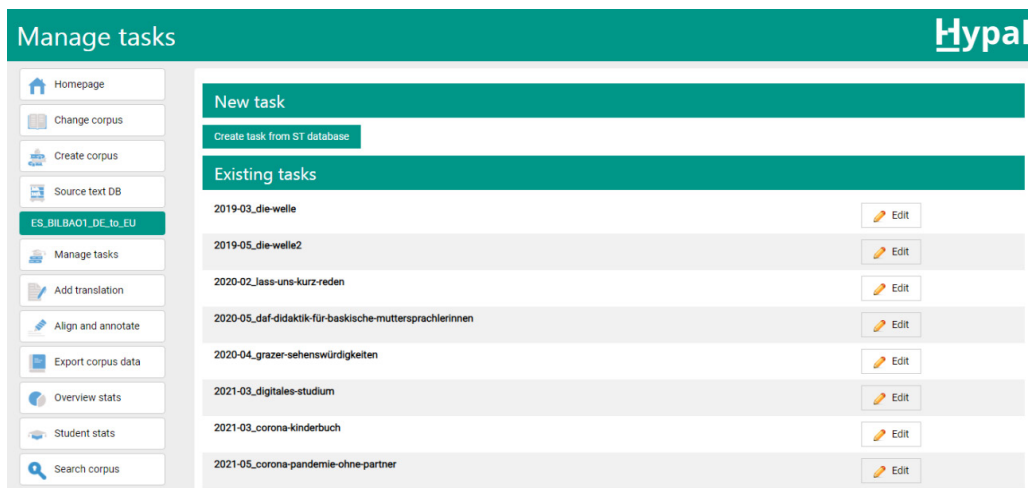


Figure 3 Bitext and Student Metadata in Hypal4MUST

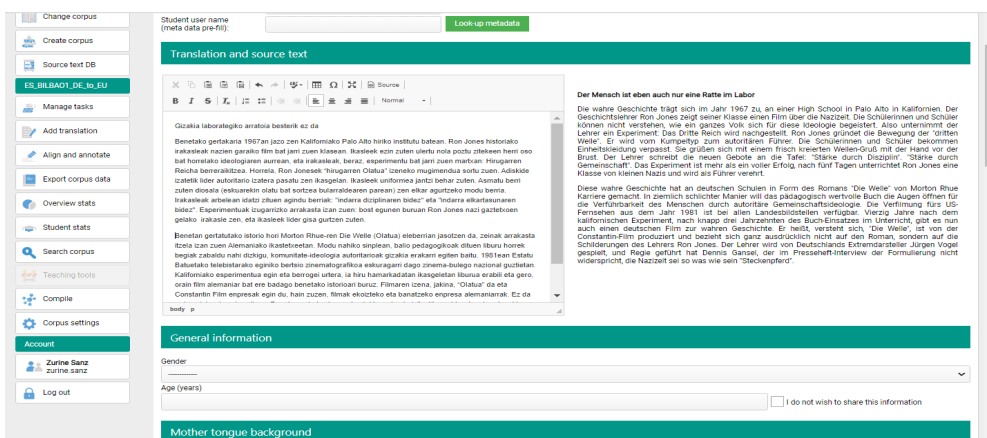


Figure 4 Manual Alignment in Hypal4MUST

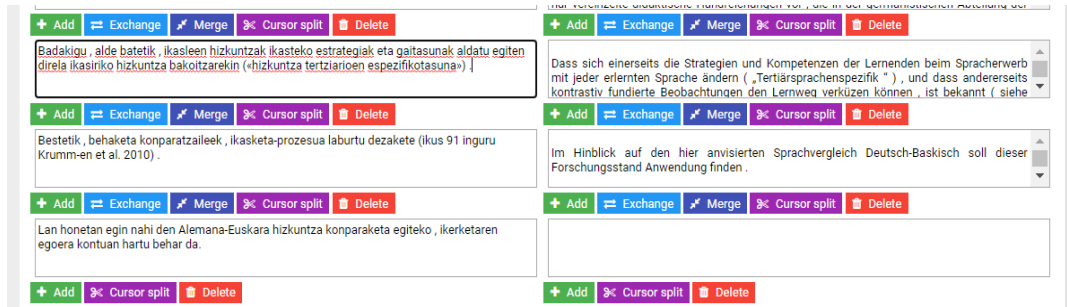
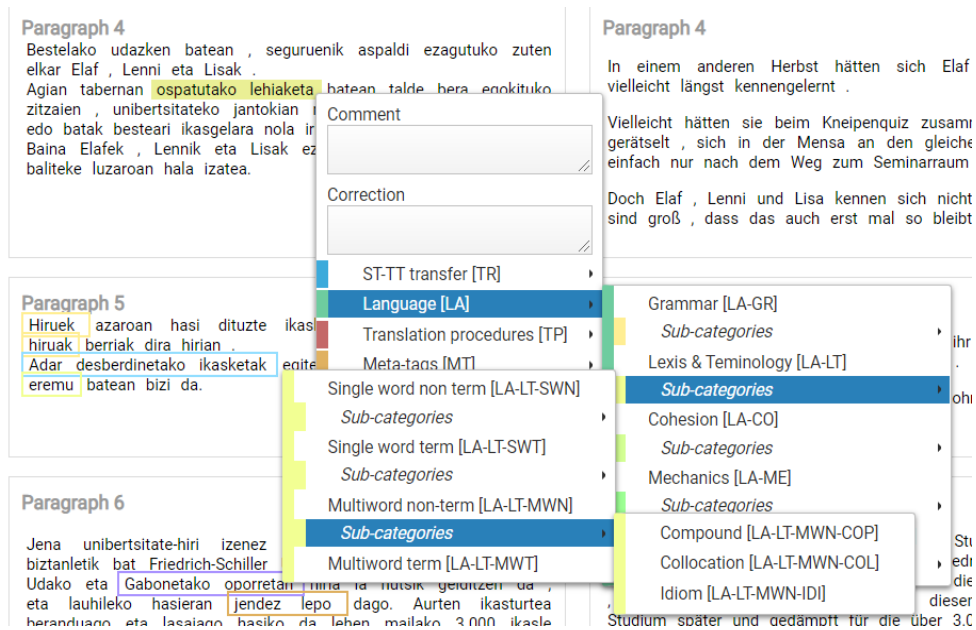


Figure 5 Translation Annotation in Hypal4MUST



Either the teacher or the student can later upload the translations for each task to Hypal4MUST. Together with the translated text, student metadata was included; the ST and TT were aligned at paragraph and sentence level and TTs were enriched with annotations based on the TAS. Figure 3 shows an uploaded translation together with student metadata below that need to be filled in.

Figure 4 shows the options (add, cursor split, merge, etc.) for making manual adjustments to the automatic alignment.

The annotation categories of the TAS can be seen in Figure 5. As can be observed, in addition to

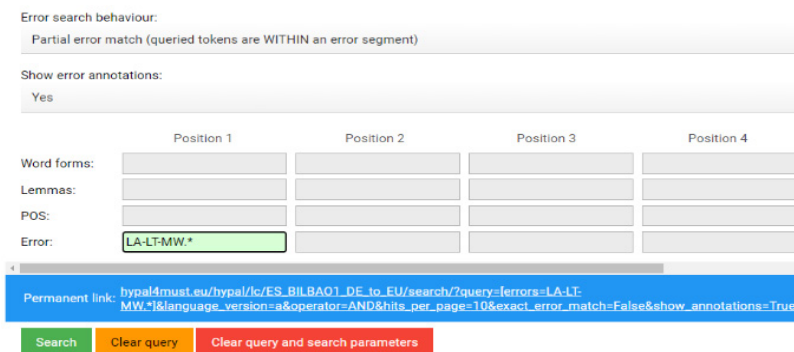
annotating segments on the basis of the TAS, comments and corrections can also be added.

Corpus Query

Once the translations had been aligned and annotated, the corpus was compiled and general results regarding the annotation were obtained by way of the in-built bilingual concordancer. This enabled the corpus data to be extracted and subsequently downloaded in Excel format (with error annotations and metadata).

During the process of annotating the TTs, not only the mentioned “multiword” subcategory but

Figure 6 Querying by Error Annotation in Hypal4MUST



also other were used to annotate MWE-related passages of the TTs:

- *ST-TT transfer -> distortion, omission*: when the meaning of the ST MWE was not properly rendered in the TT or when it was omitted.
- *ST-TT transfer -> cultural mismatch*: when a culturally bound MWE in the ST was not properly translated in the TT.
- *Language -> Lexis -> Multiword term and non-term*: when the use of the terminological or non-terminological MWE in the TL was incorrect.
- *Metatags -> Positive*: when the use of MWEs in the TTs was regarded as especially good.
- *Metatags -> Suspected SL intrusion*: there was interference from a MWE of the ST (or a third language).

Thus, the corresponding annotations were queried in Hypal4MUST using the abbreviated forms of each tag, as can be seen in Figure 6. The figure shows the search for language errors tagged as multiword terms and non-terms (LA-LT-MW.*). The wildcard ‘.’ matches any number of characters.

The same process was followed with the other levels liable to contain MWE annotations. However, not all distortion errors, for instance, contained MWEs. Therefore, relevant results for MWE analysis were selected manually. Quantitative and

qualitative information regarding the translation analysis of MWEs is provided below.

Metadata: Tasks and Participants

As Table 1 shows, eight tasks were created in all (i.e. eight source texts were uploaded and enriched with the corresponding metadata) and sixty-eight translations (22,184 tokens) were annotated using 1214 tags. Most of the STs (6) were from the general language category (journalistic texts), while two were specialised texts. However, ST5 was a promotional text that did not present a high level of specialisation, and ST4, which contained academic prose, although specialised, was on a topic with which students were familiar. There were three translation tasks that were undertaken by students under exam conditions and later marked by the teacher. The rest were either unmarked home assignments or in-class activities. The shortest texts were examinations and the longest text contained 493 words.

The number of translations uploaded for each text depended on the number of students enrolled in each academic year, the students’ willingness to take part in the project and whether or not they translated the texts uploaded to Hypal4MUST. At the beginning of each academic year, the MUST project was presented to students by the teacher. Students who agreed to contribute their translations and metadata to the project signed the MUST informed consent form. This has been approved by the Ethics Committee of the Institute for

Table 1 Tasks Created in Hypal4MUST

	Title	Language type, Supergenre, Genre	Type of task	Marked	Number of words	Academic year	Number of translations
ST1	Der Mensch ist eben auch nur eine Ratte im Labor [The human being is just a rat in the laboratory]	General language, journalistic, review	In-class examination	Yes	278	2018-19	5
ST2	Auf Wiedersehen, Kinder [Goodbye, children]	General language, journalistic, review	In-class examination	Yes	281	2018-19	5
ST3	Lass uns kurz reden [Let's talk briefly]	General language, journalistic, news/reportage	Home assignment	No	415	2019-20	10
ST4	DaF-Didaktik für baskische MuttersprachlerInnen [DaF (German as a foreign language) didactics for Basque native speakers]	Specialised language, academic prose, abstract	In-class activity	No	330	2019-20	11
ST5	Grazer Sehenswürdigkeiten [Sights of Graz]	Specialised language, promotional text, tourist guide/brochure	In-class activity	No	386	2019-20	13
ST6	Digitales Studium: Die beste Zeit ihres Lebens. Eigentlich [Digital studies: The best time of their life. Actually]	General language, journalistic, news/reportage	Home assignment	No	445	2020-21	8
ST7	Corona-Kinderbuch: China ist sauer über Hamburger Verlag [A children's book on the coronavirus: China is angry with Hamburg-based publisher]	General language, journalistic, news/reportage	In-class examination	Yes	327	2020-21	8
ST8	Corona-Pandemie ohne Partner: Gemeinsam einsam sein [The Covid-19 pandemic without a partner: together, but alone]	General language, journalistic, news/reportage	Home assignment	No	493	2020-21	8

Language and Communication of UCLouvain (approval number: CE-ILC/2022/09).

During the translation assignments, students were allowed to use any tool or resource they deemed necessary. The aforementioned metadata were stored with others (student-related metadata, for instance) in the tool.

Translations were compiled from 26 undergraduate students of Translation and Interpreting, distributed over three academic years: 2018–19, 2019–20 and 2020–21. In general, students in this German-to-Basque translation course were in their sixth semester, meaning that they still had three semesters to go (including the sixth semester) before graduation. They had attended three

translation-practice courses, which involved their two first languages (Basque and Spanish), and in the sixth semester they were translating from two foreign languages: German to Basque and either English or French to Basque. This was the case for students who did not spend a semester or academic year abroad. Of the 26 students mentioned, only 23% had completed a translation internship.

According to students' metadata, in 80.8% of the cases, their self-rated proficiency in the SL was intermediate. The remainder (5 students) rated their level of German as advanced. Half (13) of the total number had visited German-speaking countries, with the duration of stays ranging from 1 to 11 months. As for the TL, Basque, most (24 students) said they had a native command of the language. For the remaining two students, Spanish was their only first language, self-rating their command of the language as advanced. The majority of the students (69.2%) said they had two first languages (Basque and Spanish) with the remainder stating that they had one first language (either Spanish, French or Basque).

In summary, the students' translation experience was still limited when attending the German-Basque translation course, and for the majority, their command of the SL, German, was intermediate. However, the fact that half of them had spent time in German-speaking countries and/or had studied it in school before entering university shows that the groups do not usually have a homogeneous command of the SL. According to the metadata obtained from students, most stated that they had two first languages, Basque and Spanish, and all but two said they had a native-level command of the TL.

Results

In this section the annotation results of the German-to-Basque LTC are presented. The 'General results' subsection will feature results of all categories of the TAS. The second subsection ("The translation of MWES") will focus on the tagged units containing MWES.

General results

Table 2 shows the general results for first-level annotations. Language errors accounted for 50% of all annotations. The number of transfer errors, 436, was also considerable. In addition to the errors, 169 passages were marked as either positive or "too literal" — that is, traces of the SL (or another language) were observed.

Table 2 Quantitative Results of First-Level Annotation

Language	609 (50%)
ST-TT transfer	436 (36%)
Metatags	169 (14%)
Total number of annotations	1214 (100%)

The distribution of tags in different subcategories can be seen in Figure 7. In the *Language* category, many errors of grammar (156) and mechanics (179) were found (the latter concerned punctuation marks and spelling, for instance). Within this category, however, the most prominent subcategory was *Lexis* (with 239 errors). As mentioned above, LTC were analysed in both Castagnoli *et al.* (2011) and Wurm (2013). Different tagsets were used during the annotation and therefore the results are not comparable, but lexical errors are among the most frequent in both studies.

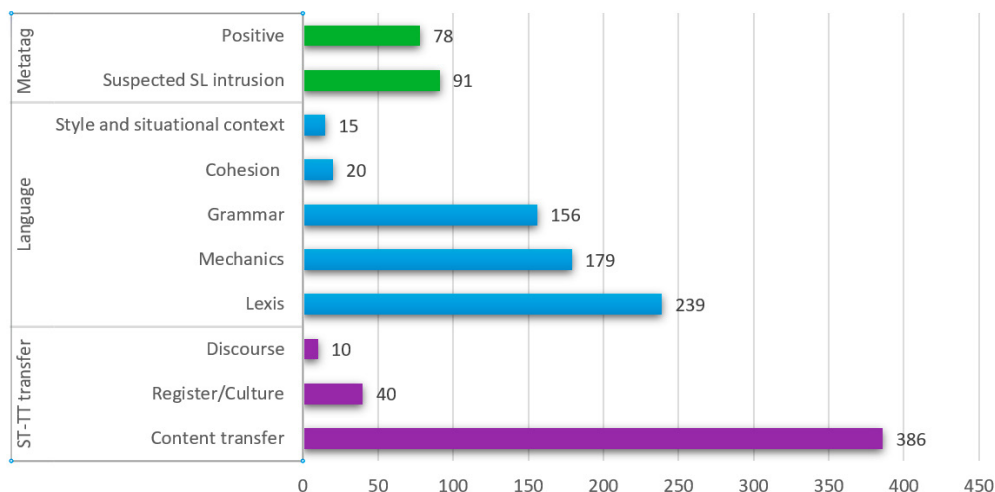
The large number of content transfer errors corresponded to student metadata and their command of the SL. According to the questionnaires that the students filled in at the beginning of each semester, what they usually fear most in this language combination is not properly understanding the meaning of the ST and not being able to render its meaning in the TL.

Noteworthy too were the positive outcomes (78) and suspected SL intrusions (91), both with a similar level of representation in the corpus.

The Translation of MWES

As explained in the section on methodology, the tagged units containing MWES were selected and

Figure 7 Quantitative Results of Second-Level Annotations



analysed manually. Table 3 shows the categories of the TAS containing MWE-related annotations, ordered by frequency.

Examples extracted from the corpus of each of the categories from Table 3 are presented in Table 4. In the first example, the Basque counterpart of the German MWE *dünnhäutig reagieren* represents an unusual word combination: *gaitzikor erantzun* ('to respond suspiciously'). There are no results for this word combination in the reference corpus of contemporary Basque (Sarasola et al., n.d.), and the word *gaitzikor* ('suspicious') has a very low frequency (25). In the second example, part of the meaning of the German collocation *kerzengerade sitzen* ('to sit bolt upright') was omitted in the Basque TT. The German MWE *lassen Sie sich überraschen* (literally 'let yourself be surprised') was translated, in one case, with a somatic MWE in Basque: *aho bete hortz utzi* (literally, 'to leave someone with the mouth full of

teeth', meaning 'to leave someone speechless'). This was tagged as positive, because the meaning was conveyed accurately, and another MWE was used in the TT, which fits well in the promotional text from which the example was extracted (ST5). The following example shows the same sentence translated to Basque literally. *Utzi zaitex harritzen* ('let yourself be surprised') is not only a literal translation of the German MWE but also of the Spanish expression *déjate sorprender*. Finally, *Semesterferien* ('semester break') was adapted to the target culture as summer and Christmas break. However, it is noteworthy that the winter break for German students typically falls around February and March and not during Christmas.

Table 5 shows the distribution of annotations by task. It is necessary to take into account the unequal number of participants submitting a translation for each task (see Table 1). However, ST4 was notable for the high number of error annotations (66 in total) and SL intrusion cases (16) compared to the other tasks of the same academic year, as well as the low number of positive features (1). Out of the tree tasks of the academic year 2020-21, the only marked task made under exam conditions, ST7, contained the least number of errors (21, as opposed to 30 in ST6 and 27 in ST8). A summary of the main features of each category will be given in the next section.

Table 3 MWE Annotations

Multiword terms and non-terms	128
Content transfer (distortion and omission)	119
Positive	40
Suspected SL intrusion	26
Cultural mismatch	11
Total	324

Table 4 Examples of Each Relevant Category of the TAS

Multiword non-term	Wenn es um die Ursachen der Corona-Pandemie geht, reagiert China sehr dünnhäutig . [When it comes to the origins of the Covid-19 outbreak, China reacts very thin-skinned.]	Koronavirusaren pandemiaren jatorriari dagokionez, Txinak gaitzikor erantzuten du. [When it comes to the origins of the Covid-19 outbreak, China responds suspiciously.]
Content transfer	Elaf, Jeans und schwarzes Top, die Laptop-Kamera abgeklebt, sitzt kerzengerade auf ihrem Stuhl am Schreibtisch. [Elaf, jeans and a black top, the laptop camera taped off, sits bolt upright in her chair at the desk.]	Elafek galtza bakeroak eta top beltza ditu jantzita. Idazmahiko [sic] aulkian dago zain, ordenagailuko kamera itzalita. [Elaf wears jeans and a black top. She waits in the desk chair with the computer camera off.]
Positive	Spazieren Sie einfach los und lassen Sie sich überraschen . [Just take a walk and be surprised.]	(...) ibili eta alde zaharrak aho bete hortz utziko zaitu. [(...) walk and the old town will leave you speechless.]
Suspected sl intrusion	Spazieren Sie einfach los und lassen Sie sich überraschen .	(...) joan paseatzera eta utzi zaitetz harritzen . [(...) take a walk and let yourself be surprised]
Cultural mismatch	In den Semesterferien ist die Stadt ausgestorben, (...). [During the semester break, the city is deserted.]	Udako eta Gabonetako oporretan hiria ia hutsik gelditzen da, (...). [During the summer and Christmas holidays, the city is almost empty.]

Table 5 Task-Based Distribution of Annotations

14

		Multiword Non-Terms	Multiword Terms	Content Transfer	Cultural Mismatch	Positive	Suspected sl Intrusion
2018-19	ST1	3	0	4	4	6	0
	ST2	4	0	12	2	5	0
2019-20	ST3	8	11	21	0	0	3
	ST4	7	32	27	0	1	16
	ST5	23	0	20	2	11	5
2020-21	ST6	4	6	18	2	5	0
	ST7	4	9	8	0	6	2
	ST8	15	2	9	1	6	0
		68	60	119	11	40	26

Multiword Terms and Non-Terms

Lexical MWE errors were marked as such because they represented inappropriate expressions in Basque. The number of errors tagged as non-term multiword (68) was slightly higher than term multiword errors (60), probably due to the non-specialised character of most STs.

With regard to multiword terms, around 50% of the errors (32) were found in student translations

of the abovementioned specialised text (ST4), which was an abstract on the teaching of German as a foreign language for native speakers of Basque at the UPV/EHU. The text describes the context in which German is taught at the Faculty of Arts of the UPV/EHU. During the translation of this text, students encountered terms that belong to their own culture but had been adapted in the ST for a German-speaking readership. Hence, they were familiar with the reality described in the ST; however, errors arose because, instead of using

Table 6 Translation of a Multiword Verb

	Mittelfristig zielen wir auf ein Forschungsprojekt, dass (...). [For the medium term, we are aiming at a research project that (...).]	Epe ertainera, ikerketa-proiektu bat espero dugu, (...). [For the medium term, we hope a research project (...).]
ST4		

the usual terminology in Basque, they tended to translate them literally. Frequently the meaning of the terms (such as *Hauptfach*, *Nebenfach*, *Nebenfachstudierende*, *Hilfssprache...*) was not properly transferred, even though they were familiar with this context. On other occasions, the problem may have been the degree of specialisation. For instance, the term *kontrastive Analyse* was frequently translated in Basque as *analisi kontrastatu* (7 out of 11), whereas in fact *analisi kontrastibo* is the term generally used in this field. A search for both terms in a reference corpus of contemporary Basque (Sarasola *et al.*, n.d.) yields 15 occurrences for *analisi kontrastibo* and none for *analisi kontrastatu*. Despite the low frequency, the books on linguistics from which they are extracted are very reliable.

In the case of multiword non-terms, they were usually tagged as errors because students produced unusual and inappropriate word combinations in the translations. For instance, in the example in Table 6, the multiword verb *zielen auf* ('aim at'), referring to *ein Forschungsprojekt* ('a research project') in this case, was translated with a single verb, *espero* ('hope', 'expect'), and a noun that functions as the object, *ikerketa-proiektu* ('research project'). As a result, the meaning was not properly rendered in the translation and from the perspective of the TL, the sentence was incomplete. The same error was found in five translations (out of a total of eleven).

We also find collocations that were translated too literally or erroneously, as in Table 7. In the case of *Kontakte herstellen* (literally 'to make contact', but meaning to socialise with others), most students (6 out of 8) tended to translate the noun *Kontakte*

Table 7 Translation of a Collocation

	Aber im 21. Jahrhundert gibt es ja noch Möglichkeiten, Kontakte herzustellen (...). [But in the 21 century, there are still ways to socialise (...).]	Baina, XXI. mendean badaude jada harremanetan hasteko erak (...). [But in the 21 century there are ways to begin socialising (...).]
ST8		

with the Basque word *kontaktu* ('contact') or *harreman* ('relation') and then combine it with verbs such as *izan* ('to be'), *hasi* ('to begin'), *jarraitu* ('to keep'). Only two students deviated more from the original and conveyed the meaning of the sentence using *jendea ezagutu* ('to get to know people').

Distortion

Informal MWEs may have caused problems for trainee translators in this language combination. The case of the German MWE (*gar*) *kein Bock haben* ('not being in the mood for'), where two students misinterpreted the meaning of the original, is noteworthy. In the first translation of Table 8, the meaning of the MWE was exaggerated and differed from the ST: no desire to live. In the second, another meaning of the word *Bock* was used and the student inadequately translated the MWE as 'not a single person'. According to the German monolingual dictionary DWDS (Berlin-Brandenburgischen Akademie der Wissenschaften, n.d.), *Bock* can also mean "störrischer Mensch, Dickkopf" ('stubborn person, obstinate').

Table 8 Translation of a Colloquial MWE

	„12 Uhr aufgestanden, gar kein Bock “, schreibt einer. [“Got up at 12, not in the mood”, writes one.]	“12tan esnatuta, bizitzeko gogoirik [sic] gabe”, idatzi du norbaitek. [“Waking up at 12, no desire to live”, writes someone.]
ST6		«12:00ak puntuan, eta tiporik ez » idatzen du batek. [“12 o'clock and not a single person”, writes one.]

Compounds may also have caused difficulties to students from the perspective of the transfer from one language into the other. In ST3 we find the compound *Nullerjahre* ('Noughties'). In this case, the output of the machine translation (MT) system may have influenced the translation. Six out of ten trainee translators translated it as "in the 90s" (*90eko hamarkadan, 90. hamarkadan*) and so did DeepL in its German-to-Spanish translation output (*en los años noventa*).

These (and other) examples show that MT systems struggle not only with collocation- or idiom-like multiword expressions but also with compounds, as may be the case of *Nebenfachstudierende* from our corpus. This refers to students with German as their second foreign language. However, DeepL translated this expression as *alumnos menores*, and some students (4 out of 11), probably taking this intermediary version as the new ST, translated it to Basque as *ikasle gazte(ago)* ('young(er) students') or *ikasle adingabe* ('underage students'), without considering the context. We find this MWE in the abstract (ST4). Only one student chose to explain what the MWE means: *bigarren hizkuntza alemana daukatenentzat* ('for those who have German as a second language').

Idiomatocity may be another factor influencing the translation of MWEs. In the example in Table 9, it is possible that additional linguistic and instrumental factors influenced the translation: there was no literal MWE in the TL, the context did not help to elucidate its meaning, and the outputs of MT systems made no sense. The MWE in Table 9, *seine Runden drehen* ('to do a few lengths (in the lake)'), was translated in a variety of ways: *paseoa egin* ('to go for a walk') or *korrika egin* ('to run'), to mention but two. In the previous sentences, we do find references to water — *Seestück* and *am Wasser* — but they appear not to have been helpful for most of the students. The text is about the German film *Die Welle* (*The Wave*), and this part describes the daily routine of the chief character, Wenger.

In some cases, we found MWEs that may have different definitions, whose meaning is dependent

Table 9 Translation of an Idiomatic MWE

	Die Geschichte beginnt als Seestück. Wenger (Jürgen Vogel) lebt mit seiner Frau Anke (Christiane Paul) in einem Blockhaus am Wasser. Jeden Morgen dreht er in vollkommener Einsamkeit seine Runden, danach wird gefrühstückt, und seit ein paar Tagen ist Anke	Istorioa itsas pintura bat bezala hasten da. Wenger (Jürgen Vogel) eta bere emazte Anke (Christiane Paul) ur ertzetan bizi dira egurrezko etxe batean. Wengerrek goizero gosaldu baino lehenago bakardade osoan egiten du korrika, eta duela egun pare batetik
ST2	schwanger. [The story begins as a lakeside story. Wenger (Jürgen Vogel) lives with his wife Anke (Christiane Paul) in a log cabin by the water. Every morning he does a few lengths in complete solitude, then they have breakfast, and for a few days now Anke has been pregnant.]	Anke haurdun dago. [The story begins like a sea painting. Wenger (Jürgen Vogel) and his wife Anke (Christiane Paul) live by the water in a wooden house. Every morning Wenger goes for a run in total solitude before having breakfast, and for a few days now Anke has been pregnant.]

on the context and which caused transfer problems for all students without exception. This is the case of the binomial *hin und her (gehen)* in the example in Table 10. The ST refers to the situation described in the previous paragraph, where it is explained how first-year students are waiting in Zoom and writing in the chat before the online class begins. We found six very different translation options, but they all contained transfer errors. The first option included in the table refers to one of the students who is nosing around. The second makes reference to an explicit movement: she is

Table 10 Translation of a Binomial

	So geht das ein bisschen hin	Kuxkuxeatzen ibili da pixka bat. [She/he has been snooping for a little while.]
ST6	und her. [It goes back and forth for a little while.]	Alde batetik bestera mugitzen da. [She/he moves from one side to the other.]
		Gutzigorabehera badao. [It works somehow.]

moving from one side to the other. The third says “it works”, maintaining to some extent the ambiguity of the original, but leaving the reader with the uncertainty of knowing what it is that works.

Positive

Translations marked as positive were regarded as such because the outputs were particularly good or because they utilised resources typical of the Basque language. In the example of Table 11,

Table 11 Example of Positive Outcomes

ST2	Jeden Morgen dreht er in vollkommener Einsamkeit seine Runden, danach wird gefrühstückt, und seit ein paar Tagen ist Anke schwanger. [Every morning he does a few lengths in complete solitude, then they have breakfast, and for a few days now Anke has been pregnant.]	Wenger goizero joaten da igeri egitera bakar-bakarrik , eta, ondoren, gosaltzen du. Bada egun pare bat Anke hardun dela. [Every morning Wenger goes swimming in complete solitude and then he has breakfast. For a few days now Anke has been pregnant.]
-----	--	---

the student, instead of sticking to the original, made use of reduplication — an often-used resource in Basque (Ibarretxe-Antuñano, 2012, p. 138) — which in this case served to express an emphatic use (i.e. *bakar-bakarrik*, ‘in complete solitude/loneliness’).

There are translations that are more idiomatic than the German STs. This is the case of the first example in Table 12, where the student used a somatic MWE in Basque: (*jende*)z *lepo* (where *lepo* means ‘neck’ and the overall figurative meaning is ‘crowded’).

This translation was tagged as positive, but it should be noted that phraseological gain does not always result in a better translation. In the last example in the same table, the German somatic MWE (*die Augen öffnen*, literally ‘open the eyes’) was not translated with an equivalent Basque somatic MWE, which exists, but with a verb

Table 12 Examples of Positive Outcomes

ST6	In den Semesterferien ist die Stadt ausgestorben, am Semesteranfang überdreht . [During the break, the city is deserted, at the beginning of the semester, packed.]	Udako eta Gabonetako oporretan hiria ia hutsik gelditzen da, eta lauhileko hasieran jendez lepo dago. [During the summer and Christmas holidays, the city is almost empty and at the beginning of the semester, it is crowded.]
ST1	In ziemlich schlichter Manier will das pädagogisch wertvolle Buch die Augen öffnen für die Verführbarkeit des Menschen durch autoritäre Gemeinschaftsideologie. [In a rather simple way, the pedagogically valuable book aims at opening people’s eyes to the seductiveness of authoritarian community ideology.]	Oso liburu esanguratsua da pedagogikoki, eta haren asmoa modu sinplean honakoaz ohartaraztea da: gizakia erraz erakartzen du komunitate-ideologia autoritarioak. [Pedagogically, it is a very significant book, and its goal is to draw attention to the following in a simple way: people are easily attracted to authoritarian community ideology.]

(*ohartarazi*, ‘to draw attention to’) that properly captured the meaning of the original.

Suspected SL Intrusion

As previously discussed, when translating from German to Basque, students are used to working with an intermediate language, usually Spanish. Furthermore, as the students themselves recognised when completing the metadata forms, the majority use MT systems. In this language combination, they first consult the German-to-Spanish MT, and may then look for the output of the Spanish-to-Basque MT system. The possible interference from the pivot language may be regarded as textual interference since traces of the MT text may be found in the Basque translations. Traces of such textual interference may be observed in the two examples in Table 13. DeepL, currently the most widely used MT system among our trainee translators, translated the first sentence from the

Table 13 Interference from Spanish

	Das Kurzdeutsch habe sich verselbstständigt , werde längst auch von deutschen Muttersprachlern aller Bildungsschichten genutzt. [“Short German” has taken on a life on its own and has long been used by native German speaker of all educational levels.]	Alemaniera laburrak bizitza propioa hartu du, eta Alemaneko hiztun natiboek erabili dute bizitzako arlo guztietan. [“Short German” has taken his own life and it has been used by native German speakers in all areas of life.]
ST3		
	Zunächst gilt für die große Mehrheit, dass sie Kenntnisse in Spanisch und Baskisch mitbringen ; [First of all, the vast majority have knowledge of Spanish and Basque.]	Lehenik eta behin, gehienek gaztelaniaren eta euskararen ezagutza dute , (...) [First of all, most of them have knowledge of Spanish and Basque.]
ST4		

18

table as follows: “El alemán breve ha cobrado vida propia”. *Cobrar vida propia* is an MWE meaning ‘to take on a life on its own’. In the Basque translations of this sentence, two students (out of 10) chose to retain and literally translate the Spanish MWE to Basque: *bizitza propioa hartu* (literally, ‘to take the own life’).

The translation of the second example (ST4) also shows traces of another Spanish MWE, *tener conocimientos de algo* (literally, ‘to have knowledge of something’), which was found in the DeepL translation: “En primer lugar, la gran mayoría de ellos tiene conocimientos de español y de euskera” (‘first of all, the vast majority of them have knowledge of Spanish and Basque’). Another reason for choosing this direct translation of the Spanish MWE may have been the register of the ST. Because ST4 was a formal, specialised text, the trainee translator, in an attempt to maintain the same level of formality, may have thought that they were maintaining the register in the translation by using this noun and verb collocation (instead of a single verb — in this case *jakin* — as others have appropriately done).

Cultural Mismatch

The group of MWEs belonging to this section comprised predominantly solid compounds, which were often mistranslated when translated literally due to their attachment to the source culture, as shown in Table 4.

Discussion and Conclusions

The goal of the present paper, carried out in the framework of the MUST project, was to present the results of the annotation in an LTC consisting of German STs and trainee translators’ outputs to Basque. As described in the results section, the category with the highest number of error annotations is *Language*, and within this category, lexical errors are quantitatively the most important. This has been observed in other studies (Castagnoli *et al.*, 2011; Wurm, 2013). In the *Language* category, the second subcategory with the most errors is *Mechanics*, which includes both spelling and punctuation errors. In the paper by Fictumova, Obrusnik and Stepankova (2017), in which a Czech-to-English LTC was compiled and examined, punctuation was regarded as a problem that needed to be “addressed in the curriculum” (2017, p. 225). On the basis of our results, the same applies to the language combination examined in this article.

More specifically, the present study has contributed to the analysis of MWEs in translation didactics. The total number of annotations related to these units is 324, spread over 8 different translation tasks from three different academic years. The uneven number of translations per task makes comparison between tasks difficult, but it can be highlighted that the specialised text (ST4) representing an abstract was the most demanding. As has already been explained, there were many mistranslated terms (32, see Table 5) in this text. As for non-terms, examples of erroneously translated idioms (Table 9), collocations (Table 7), and multiword verbs (Table 6) were given.

Activities to increase awareness of different types of MWEs and of the abovementioned problematic cases can be addressed in class. More ambitiously, a German-Basque bilingual dictionary of MWE could be designed, whose main recipients would be translation trainees: “Learner translation data could be used to design similar notes and incorporate them into bilingual dictionaries, especially learners’ bilingual dictionaries” (Granger & Lefer, 2016 as cited in Granger & Lefer, 2020, p. 1195).

Castagnoli (2023) argued that variation is greater in the translation of idiomatic expressions and especially when a literal translation is not possible in the TL. The analysis of variation was not the goal of the present study; but based on the distortion errors described in this paper, it could be suggested that in some cases not having a literal translation in the TL may increase variation and errors in the translations (see Table 10). This is not a conclusion, but a matter that may be worthy of further analysis in the future.

Certain specific MWEs were generally problematic. This was the case, for instance, with the binomial MWE *hin und her gehen* (see Table 10) and the multiword verb *zielen auf* (see Table 6). The latter example corroborates the results of a previous study (Sanz-Villar, 2022), where the translation of multiword verbs was analysed, as indicated in the introduction of this paper. Marcelo Wirnitzer and Amigo Extremera (2015, p. 381) mention in their paper that students are generally familiarised with English phrasal verbs, but less frequent phrasal verbs (or verbs with particles, as they call them) were translated erroneously more frequently.

The examples in the section on distortion errors may indicate that relying on MT systems’ outputs when translating MWEs can sometimes lead to mistranslations or interference. A critical use of resources for translators such as MT systems is necessary, as stated by Rabadán and Gutiérrez Lanza (2020, pp. 379-380): “Contemporary translation training relies on technology, from translation memories and machine translation to the more

modest grammatical and spell-checkers, to reduce the time and effort invested in the task. However, as with any use of language and translation technology, successful performance requires that the user can evaluate the outcome. A variety of (post)-editing strategies can be applied to both human and machine translation outputs, which require critical human assistance”.

As far as interference is concerned, it is not limited to the influence that the German ST can exert on the Basque TT but can also be affected by the influence of intermediary MT versions in a third language, mostly Spanish, as shown in Table 13. With regard to this type of interference, referred to as instrumental interference in Sanz-Villar (2018), hands-on activities to become familiar with and use monolingual and bilingual resources (albeit indirect) related to MWEs can be helpful.

The importance of interference awareness among translation trainees is highlighted by Rabadán and Gutiérrez Lanza (2020, p. 380): “Whether translating or (post)-editing, awareness of language-pair-dependent problems underlies successful performance. Human translation, partially informed by machine-mediated translation, is a given in student workflows, but errors can easily go unnoticed if cross-linguistic competence is not properly developed”. Thus, developing awareness of source- and third-language interference when translating MWEs may benefit students’ cross-linguistic competence.

It may also be useful to present in class the different translation options we may have when translating MWEs. Students may have the erroneous impression that phraseological maintenance or gain ensures a better translation (as explained in the example in Table 12). Exercises including quality evaluation of different translation options of the same ST may be employed to design corpus-based teaching material.

In a previous study (Sanz-Villar, in press), it has been found that for some students literally

translating a Spanish MWE (when translating a literary text from German to Basque) was a conscious decision. The mentioned study analysed not only students' translations but also the results of questionnaires students filled out after completing the translation tasks. In the future, if the process of translation is to be taken into account, using more than one method to collect data will be crucial.

The use of MT systems, the type of MWE to be translated, the absence of a literal counterpart in the TL or the type of ST can be important variables that need to be taken into account when analysing the translation of MWEs. The metadata collected during the creation of this corpus also takes into account other variables, such as the SL and TL command of the translation trainees. It may be interesting to consider the effect of these variables on the output, for which it would be necessary to have more specific information on the level of proficiency of both the foreign language and native language, beyond the self-assessment.

References

- Aierbe Mendizabal, A. (2008). La traducción a la lengua vasca de las unidades fraseológicas especializadas del lenguaje administrativo. In M. I. González Rey (Ed.), *A multilingual focus on contrastive phraseology and techniques for translation* (pp. 27–44). Dr. Kovaç.
- Albaladejo Martínez, J. A. (2015). Fraseología especializada en traducción general. In P. Mogorrón Huerta & F. Navarro Domínguez (Eds.), *Fraseología, didáctica y traducción* (pp. 273–290). Peter Lang.
- Alcina, A. (2008). Translation technologies. Scope, tools and resources. *Target*, 20(1), 79–102. <https://doi.org/10.1075/target.20.1.05alc>
- Berlin-Brandenburgischen Akademie der Wissenschaften. (n.d.). Bock. In *Digitales Wörterbuch der deutschen Sprache (DWDS)*. <https://www.dwds.de/wb/Bock#1>
- Castagnoli, S. (2023). Exploring variation in student translation. *International Journal of Learner Corpus Research*, 9(1), 97–125. <https://doi.org/10.1075/ijlcr.22010.cas>
- Castagnoli, S., Ciobanu, D., Kunz, K., Volanschi, A., & Kübler, N. (2011). Designing a learner translator corpus for training purposes. In N. Kübler (Ed.), *Corpora, language, teaching, and resources: From theory to practice* (pp. 221–248). Peter Lang.
- Corpas Pastor, G. (1996). *Manual de fraseología española*. Gredos.
- Espunya, A. 2013. Investigating lexical difficulties of learners in the error-annotated UPF learner translation corpus. In S. Granger, & G. Gaëtanelle (Eds.), *Twenty years of learner corpus research. Looking back, moving ahead* (pp. 129–137). Presses Universitaires de Louvain.
- Fictumova, J., Obrušnik, A., & Stepankova, K. (2017). Teaching specialized translation. Error-tagged translation learner corpora. *Sendebár*, 28, 209–241.
- Granger, S., & Lefer, M-A. (2021). Translation-oriented annotation system manual (Version 2.0). CECL Papers 3. Centre for English Corpus Linguistics/Université catholique de Louvain. https://cdn.uclouvain.be/groups/cms-editors-cecl/cecl-papers/TAS-2.0_annotation_manual_2021-10-26.pdf
- Granger, S., & Lefer, M-A. (2020). The multilingual student translation corpus: a resource for translation teaching and research. *Lang Resources & Evaluation*, 54, 1183–1199. <https://doi.org/10.1007/s10579-020-09485-6>
- Granger, S., & Lefer, M-A. (2018, November 5–7). The translation-oriented annotation system: A tripartite annotation system for translation research [Conference presentation abstract]. *Parallel Corpora: Creation and Applications (PaCor)*, Madrid. https://eventos.ucm.es/_files/_event/_18651/_editorFiles/file/Book%20of%20abstracts_ECET-T-PaCor2018b.pdf
- Granger, S., & Paquot, M. 2008. Disentangling the phraseological web. In S. Granger, & F. Meunier (Eds.), *Phraseology: An interdisciplinary perspective* (pp. 27–49). John Benjamins. <https://doi.org/10.1075/z.139.07gra>
- Ibarretxe-Antuñano, I. (2012). Análisis lingüístico de las onomatopeyas vascas. *Oihenart: Cuadernos de Lengua y Literatura*, 27, 129–177.
- Leiva Rojo, J. (2013). La traducción de unidades fraseológicas (alemán-español / español-alemán) como parámetro para la evaluación y revisión de traducciones. In C. Mellado Blanco (Coord.), P. Buján, N. M. Iglesias, M. C. Losada, & A. Mansilla (Eds.), *La fraseología del alemán y el español: lexicografía y traducción*, (pp. 31–42). Peniope.
- Marcelo Wirnitzer, G., & Amigo Extremeña, J. J. (2015). La traducción de fraseologismos en el aula de Traducción General. In G. Corpas Pastor, M. Seguiri Domínguez, R. Gutiérrez Florido, & M. Urbano Mendaña (Eds.), *Nuevos horizontes en los estudios de traducción e interpretación* (pp. 373–387). Tradulex.

- Mogorrón Huerta, P., & Navarro Domínguez, F. (2015). *Fraseología, didáctica y traducción*. Peter Lang. <https://doi.org/10.3726/978-3-653-05309-8>
- Obrusnik, A. (2014, July 20–23). Hypal: A User-Friendly Tool for Automatic Parallel Text Alignment and Error Tagging [Conference presentation abstract]. *Eleventh International Conference Teaching and Language Corpora (TaLC 11)*, Lancaster. <https://ucrel.lancs.ac.uk/talc2014/doc/TALC2014-abstract-book.pdf>
- Rabadán, R., & Gutiérrez-Lanza, C. (2020). Developing awareness of interference errors in translation. An English-Spanish pilot study in popular science and audiovisual transcripts. *Lingue e Linguaggi*, 40, 379–404.
- Sánchez-Nieto, M. (2012). La doble interpretación aspectual de predicados en la traducción alemán-español de secuencias narrativas: análisis de un corpus de traducciones estudiantiles. *TRANS. Revista de Traductología*, 16, 79–99. <https://doi.org/10.24310/TRANS.2012.v0i16.3213>
- Sanz-Villar, Z. (in press). Los fraseologismos en el aula de prácticas de traducción: estudio de caso sobre el impacto de los traductores automáticos.
- Sanz-Villar, Z. (2022). German-into-Basque translation of verbal patterns. In C. Mellado Blanco (Ed.), *Productive patterns in phraseology and construction grammar* (pp. 265–286). Walter de Gruyter. <https://doi.org/10.1515/9783110520569-011>
- Sanz-Villar, Z. (2018). Interference and the translation of phraseological units in a parallel and multilingual corpus. *Meta: Journal des traducteurs / Meta: Translators' Journal*, 63(1), 72–93. <https://doi.org/10.7202/1050515ar>
- Sarasola, I., Salaburu, P., & Landa, J. (n.d.). *Eguno Testuen Corpua (ETC)*. <https://www.ehu.eus/etc/>
- Serrano Lucas, L. C. (2010). Metodología para la enseñanza de la fraseología en traducción: la ficha fraseológica como tarea final. *Paremia*, 19, 197–206.
- TRALIMA-ITZULIK. (2019). *TAligner* (Version 3.0) [Computer software]. <https://addi.ehu.es/handle/10810/42445>
- Valero Cuadra, P. (2015). Fraseologismos en el aula de traducción general. In P. Mogorrón Huerta, & F. Navarro Domínguez (Eds.), *Fraseología, didáctica y traducción* (309–320). Peter Lang.
- Wurm, A. (2013). Eigennamen und Realia in einem Korpus studentischer Übersetzungen (КОРТЕ). *Trans-kom*, 6(2), 381–419. https://www.trans-kom.eu/bd06nr02/trans-kom_06_02_06_Wurm_Eigennamen.20131212.pdf

How to cite this article: 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>