

## Análisis de las traducciones de aprendices en la dirección inglés-euskera. Primeros resultados

### Students' translation analysis in the English-Basque language combination. First results

---

NAROA ZUBILLAGA GÓMEZ

Universidad de País Vasco / Euskal Herriko Unibertsitatea (UPV/EHU), Facultad de Letras / Letren Fakultatea. Departamento de Filología Inglesa, Alemana y de Traducción e Interpretación/Filologia Ingelesa, Alemana, Itzulpengintza eta Interpretazioko Departamentua, Paseo de la Universidad / Unibertsitateen Pasealekua, 5, 01006, Vitoria-Gasteiz. España.

Dirección de correo electrónico: [naroa.zubillaga@eus.eus](mailto:naroa.zubillaga@eus.eus)

ORCID: <https://orcid.org/0000-0001-8181-2622>

Recibido/Received: 14/12/2022. Aceptado/Accepted: 13/2/2024.

Cómo citar/How to cite: Zubillaga Gómez, Naroa, «Análisis de las traducciones de aprendices en la dirección inglés-euskera. Primeros resultados», *Hermēneus. Revista de Traducción e Interpretación*, 27 (2025): pp. 553-575.

DOI: <https://doi.org/10.24197/4kt0yz54>

Artículo de acceso abierto distribuido bajo una [Licencia Creative Commons Atribución 4.0 Internacional \(CC-BY 4.0\)](#). / Open access article under a [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](#).

**Resumen:** El objetivo de este artículo es informar sobre los primeros resultados de un estudio de traducción basado en corpus con traducciones de aprendices. Se han incluido traducciones de los cursos 2019/2020 y 2021/2022 de la asignatura de traducción inglés-euskera y de los tipos periodístico, académico y turístico. Para la compilación, anotación y análisis estadístico del corpus, se ha utilizado la metodología del proyecto MUST. Esta metodología incluye metadatos relativos al perfil sociolingüístico de los estudiantes, lo cual resulta especialmente interesante en contextos como el de la lengua vasca. Los resultados en este primer estudio muestran como los estudiantes en esta dirección lingüística incurrir sobre todo en errores de la propia lengua meta.

**Palabras clave:** Estudio basado en corpus; traducción inglés-euskera; traducción periodística; traducción académica; traducción turística.

**Abstract:** The aim of this article is to report the first results of a corpus-based students' translation study. Translations included come from students of 2019/2020 and 2021/2022 years studying English-Basque translation and correspond to three text types: journalistic, academic and tourist texts. For the compilation, annotation and statistical analysis of the corpus, the MUST project methodology has been used. This methodology includes metadata related to the sociolinguistic profile of the students, which is especially interesting in contexts such as the Basque language. The results in this first study show how students in this linguistic direction incur, above all, errors in the target language itself.

---

**Keywords:** Corpus-based study; English-Basque translation; journalistic translation; academic translation; touristic translation.

**Sumario:** Introducción; 1. Traducción de textos periodísticos, académicos y turísticos en lengua vasca; 2. Metodología; 3. Resultados, 3.1. Análisis de una muestra más específica; Discusión y conclusiones; Referencias bibliográficas.

**Summary:** Introduction; 1. Translation of journalistic, academic and touristic text types in Basque; 2. Methodology; 3. Results, 3.1. Analysis of a more specific example; Discussion and conclusions; References.

---

## INTRODUCCIÓN

Tanto la lingüística contrastiva como los estudios de traducción han evolucionado al basar su investigación en corpus. Así, ambos campos contribuyen al desarrollo de la investigación interlingüística a través de una metodología común que es la lingüística de corpus (Izquierdo, 2006). Y es que la lingüística de corpus se ha convertido en la metodología empírica base de los estudios de traducción, y publicaciones de gran calidad de los últimos años son prueba de ello (Doval y Sánchez-Nieto, 2019; Lavid-López *et al.* 2021).

La rama educativa de los estudios de traducción también ha aplicado la metodología basada en corpus y, así, contamos con corpus de aprendices, aunque no es habitual encontrar constancia de las aplicaciones de estos corpus y los resultados derivados de estos, salvando alguna que otra excepción (Castagnoli, 2009).

Dado que nuestro estudio se contextualiza en la docencia de los estudios de traducción en lengua vasca, cabe mencionar que el euskera cuenta con un gran número de corpus de referencia desarrollados tanto por la Universidad del País Vasco / Euskal Herriko Unibertsitatea (UPV/EHU) como por otras instituciones y que sirven como herramienta a traductores así como a investigadores (Bilbao y Makazaga, 2015, p. 327), pero hasta la fecha nunca se ha compilado un corpus de aprendices de traducción, a pesar de que los estudios de traducción se pusieran en marcha en el año 2000 dentro de la UPV/EHU.

A fin de salvar esas carencias, cuatro miembros del grupo de investigación Tralima/Itzulik<sup>1</sup> se sumaron al proyecto MUST, MULTilingual Student Translation. Se trata de un proyecto cuyo objetivo es almacenar las traducciones realizadas por el alumnado, alinearlas y

---

<sup>1</sup> <https://www.ehu.eus/eu/web/tralimaitzulik/home>

etiquetarlas para su posterior análisis. Es un proyecto de envergadura internacional, en el que, hasta la fecha, participan 38 universidades de todo el mundo. Son muchas las combinaciones lingüísticas que tienen cabida en este proyecto, pero todos utilizan el mismo programa para la compilación de corpus (Hypal4Must, que explicaremos más detenidamente en el apartado sobre la metodología), así como el mismo sistema de anotación de errores.

Además, otra ventaja que tiene el programa de compilación y explotación de corpus Hypal4Must es que utiliza metadatos. Tradicionalmente, los estudios empíricos de traducción se han centrado sobre todo en el producto. En 2003, Olohan abogó por la contextualización, proponiendo combinar metodologías basadas en corpus con otro tipo de metodologías y análisis (2003). Desde entonces, son varios los investigadores que han resaltado la necesidad de combinar los estudios empíricos con datos de carácter más social, cognitivo y de contexto (Sutter y Lefer, 2019). El proyecto MUST es ejemplo de la convergencia entre estudios basados en corpus y estudios más sociológicos, ya que incluye metadatos sobre, por ejemplo, el perfil lingüístico de los estudiantes.

Las participantes en el proyecto MUST por parte de Tralima/Itzulik contamos con experiencia previa en estudios basados en corpus (Izquierdo, 2008; Manterola, 2012; Zubillaga, 2015; Sanz, 2018), y gracias al hecho de que nos hemos embarcado en este proyecto, esperamos obtener resultados basados en las traducciones de nuestros estudiantes.

## **1. TRADUCCIÓN DE TEXTOS PERIODÍSTICOS, ACADÉMICOS Y TURÍSTICOS EN LENGUA VASCA**

Aunque en los últimos años han abundado los estudios académicos de traducción en el ámbito de la lengua vasca (Barambones, 2009; Jaka, 2011; Manterola 2012; Zubillaga, 2013; Sanz, 2015; Ibarluzea, 2017), todos ellos se centran en la traducción audiovisual o en la traducción literaria. Hasta la fecha no se han realizado estudios de traducción con textos de aprendices en lengua vasca, tampoco con otro tipo de textos como los que hemos analizado para este artículo: el periodístico, el académico y el turístico.

Según reconoce Hernández (2008), la escasez de datos sobre la traducción periodística dificulta la labor de describir la realidad de este ámbito. Añade que este género «no ha recibido gran atención por parte de los estudios traductológicos» (Hernández, 2008, p. 359). La misma autora

cuenta con otros estudios sobre la traducción periodística (Hernández, 2019; Hernández, 2021). Sin embargo, en lo que se refiere a la traducción periodística en lengua vasca, no encontramos estudio alguno. La única excepción es el artículo que nos ofrece Irene Arrarats sobre la labor de corrección que realizan en el periódico en lengua vasca Berria (Arrarats, 2004). Arrarats no se refiere directamente a la traducción periodística, pero sí que menciona un aspecto que resulta interesante para nuestro estudio de traducciones realizadas por aprendices: comenta que, en su labor como correctora, se ha dado cuenta de que los periodistas-traductores recién egresados cometen siempre los mismos errores en lengua vasca, errores que muy a menudo están relacionados con la interferencia del castellano o del francés (Arrarats, 2004, pp. 107-108).

En lo que se refiere a textos académicos, Bereziartua y Boillos proponen en su publicación una propuesta docente para impartir el euskera académico en las universidades (Bereziartua y Boillos, 2020). Las autoras recuerdan que es importante tener en cuenta la presencia que ha tenido la lengua en el ámbito académico en las últimas cuatro décadas, puesto que tanto la estandarización de la lengua como el ámbito académico a la que se ha visto expuesta cuentan con un breve recorrido. Cabe señalar que la primera universidad del País Vasco (Euskadi) fue la universidad privada de Deusto (1886), luego se fundó la universidad pública del País Vasco (UPV/EHU) en 1980 y, más tarde, la universidad privada de Mondragón (1997), por lo que se empezó a utilizar la lengua vasca en el ámbito académico en la década de 1980. Hay que tener en cuenta, además, la presencia de lenguas dominantes como el castellano, el francés o el inglés, no solo por el estatus del que gozan, sino también por la cantidad de publicaciones académicas con las que cuentan. Aún y todo, la cantidad de alumnos que escoge cursar sus estudios en lengua vasca ha crecido exponencialmente en los últimos años (Mees, 2010), por lo que resulta de vital importancia trabajar y desarrollar el aspecto más académico de esta lengua. Aunque en los últimos años han aumentado los recursos y las herramientas para poder producir textos académicos en lengua vasca y el trabajo de Bereziartua y Boillos demuestra que se dan pasos para la docencia de un registro académico de la lengua, no hay ningún trabajo de referencia en lo que se refiere a la traducción de textos académicos en lengua vasca.

Ocurre lo mismo con la traducción de textos turísticos. En la combinación lingüística inglés-español, contamos con la tesis doctoral de Santamaría (2014), quien estudia en concreto la traducción de las guías de

viaje, justamente el tipo de texto que los alumnos objeto de estudio de este artículo han tenido que traducir. Según Santamaría, existen diferencias entre el inglés y el español, por ejemplo, el estilo más informal del inglés en comparación con el más formal del español (Santamaría, 2014, p. 341). Sin embargo, en la combinación lingüística con la lengua vasca no encontramos ningún estudio de este tipo.

Teniendo en cuenta que en los tres ámbitos de estudio mencionados (el periodístico, el académico y el turístico) no existen estudios previos ligados a la traducción en lengua vasca, se podría concluir que queda mucho por hacer, máxime en un contexto sociolingüístico como el de la lengua vasca, en el que siempre se ha trabajado como mínimo con tres lenguas (euskera, castellano y francés) y cada vez más con el inglés (Auzmendi *et al.*, 2003, pp. 5-6). En el presente estudio presentamos los primeros resultados de un estudio basado en corpus con traducciones de aprendices para textos periodísticos, académicos y turísticos, por lo que se trata de una primera aportación, que podría funcionar como punto de partida para futuros estudios.

Sin embargo, cualquier aportación en el ámbito de la traducción en lengua vasca no debería obviar la situación diglósica que vive la lengua, ya que influye directamente en la competencia que tienen los usuarios de esa lengua. Así, sin olvidar la necesidad de combinar puntos de vista de contexto en estudios de corpus y, además, teniendo en cuenta que el proyecto MUST nos permite analizar los resultados desde un punto de vista más sociolingüístico, los objetivos de trabajo marcados en el presente estudio han sido los siguientes:

- a) Observar los errores que comete el alumnado en la combinación lingüística EN-EU en los textos periodísticos, académicos y turísticos, a fin de contrastar las dificultades que ocasiona cada uno de ellos entre los aprendices.
- b) Contrastar el rendimiento del alumnado con lengua materna euskera y lengua materna castellano. En un contexto sociolingüístico en el que existen dos lenguas oficiales, parece lógico pensar que habrá diferencias según los sujetos hayan tenido una u otra lengua materna.

## 2. METODOLOGÍA

Tal y como se ha adelantado en la introducción de este artículo, la metodología utilizada para este estudio se basa en el programa

Hypal4Must del proyecto internacional MUST, *MUltilingual Student Translation*.

Hypal4Must es un programa que sirve para almacenar, alinear y etiquetar traducciones realizadas por estudiantes (Fictumova *et al.*, 2017). Primero de todo, el o la docente que participe en el proyecto MUST debe haber creado un corpus en su combinación lingüística (en nuestro caso, EN-EU). Una vez creado el corpus, se escoge el texto que el alumnado deberá traducir. Los textos que se escojan para crear tareas de traducción deben ser validados por las directoras del proyecto MUST. Una vez que se haya recibido el visto bueno, el o la docente puede crear la tarea.

El alumnado puede traducir y guardar su traducción directamente en el programa, después de haber creado una cuenta en él. Además, tiene que introducir una serie de metadatos relativos a su perfil lingüístico y a las herramientas utilizadas para la traducción. Si por alguna razón el alumnado no puede realizar la tarea directamente en el programa, el o la docente también puede subir por su cuenta las traducciones y completar los metadatos uno por uno. En nuestro caso, el año que los docentes de Tralima/Itzulik empezamos a participar en el proyecto MUST, el alumnado del curso 2019/2020 ya había realizado unas tareas de traducción, por lo que las traducciones fueron introducidas por la docente. Sin embargo, las traducciones relativas al curso 2021/2022 fueron introducidas directamente por el alumnado.<sup>2</sup>

Los textos que el alumnado ha traducido en ambos cursos son exactamente los mismos (un texto periodístico, un texto académico y un texto turístico), pero el número de traducciones recabadas para cada uno de los cursos es muy distinto:

	Curso 2019/2020		Curso 2021/2022	
	Número de traducciones	Número de palabras	Número de traducciones	Número de palabras
Texto periodístico	15	6.472 (eu) / 7.740 (en)	5	2.130 (eu) / 2.580 (en)
Texto académico	13	2.217 (eu) / 2.587 (en)	5	834 (eu) / 995 (en)

<sup>2</sup> En el curso 2020/2021 también se recabaron datos para el proyecto MUST, pero se utilizaron para realizar un estudio contrastivo junto con la doctora Izquierdo: Izquierdo y Zubillaga (2025).

Texto turístico	15	4.954 (eu) / 6.150 (en)	4	1.274 (eu) / 1.640 (en)
--------------------	----	----------------------------	---	----------------------------

*Tabla 1.* Número de traducciones y palabras para cada texto en los cursos seleccionados para el estudio, 2019/2020 y 2021/2022.

Durante el curso 2021/2022 asistieron a clase muy pocos alumnos. Aunque los matriculados fueron catorce, la mayoría de ellos estaban en el extranjero con la beca Erasmus, ya que los años anteriores no les resultó tan fácil hacerlo, debido a la pandemia causada por el COVID19. Por otra parte, en ambos cursos hay una tarea en la que se han recabado menos traducciones, debido a la falta de asistencia de algunos alumnos: en el curso 19/20, hay dos traducciones menos en la tarea de traducción del texto académico y en el curso 21/22 una traducción menos en la tarea de traducción del texto turístico. El alumnado de ambos cursos ha traducido exactamente los mismos textos.

El texto periodístico es un texto publicado en el periódico británico *The Guardian* en su versión digital;<sup>3</sup> al tratarse de un texto bastante largo, el alumnado no tradujo todo el texto entero, sino parte de él. El texto académico es el *abstract* extraído de un artículo (Pontrandolfo, 2015, pp. 137-160) y el texto turístico se ha extraído de una página web que da ideas sobre qué hacer en la ciudad de Dublín.<sup>4</sup> Se han escogido diferentes géneros de texto para poder contrastar si alguno de ellos provoca más anotaciones.

Una vez subidas las traducciones al programa, se alinean a nivel de párrafo y frase. En este caso se trata de alinear todas las traducciones derivadas de cada uno de los tres textos; es decir, cada uno de los textos originales tiene más de un texto meta, por lo que podríamos hablar de un corpus de traducción múltiple (Espunya, 2014, p. 35). Después, el o la docente puede pasar a anotar cada una de las traducciones. El sistema de anotación es el que Hypal4Must utiliza para todos los participantes en el proyecto. Aunque ya está en marcha la versión actualizada y mejorada del sistema de anotación (TAS 2.0), la utilizada en las traducciones del presente artículo es la versión 1.0 (Granger y Lefer, 2020).

Este sistema de anotación común diferencia tres grandes bloques: anotaciones relativas a la propia lengua de llegada, *language* (LA),

<sup>3</sup> Véase <https://www.theguardian.com/global-development/2020/feb/08/more-than-half-of-women-in-zimbabwe-have-faced-sex-tortion-finds-survey>

<sup>4</sup> Véase: <https://www.visitdublin.com/>

anotaciones relativas a la transferencia de contenido, *transfer* (TR) y existe además una etiqueta meta, *meta-tag* (MT), con el que poder anotar interferencias de la lengua origen o los casos positivos. Estos tres bloques constituyen, en realidad, el primer nivel de anotación, ya que cada una de las categorías se divide en más subcategorías, para así poder especificar el tipo de anotación de lengua (LA), transferencia (TR) o anotación meta (MT).

Así, la categoría de lengua de llegada (LA) se divide en cinco subcategorías: gramática (LA-GR), léxico y terminología (LA-LT), cohesión (LA-CO), ortografía y puntuación (LA-ME) y estilo (LA-ST).

La categoría relativa a la transferencia (TR) también se divide en otras cinco categorías de segundo nivel: errores de contenido (TR-CT), de léxico (TR-LE), de discurso o pragmáticos (TR-DP), de registro o culturales (TR-RC) o errores que se asocian al incumplimiento de la tarea de traducción (TR-TB).

La categoría meta (MT) se divide en dos categorías de segundo nivel: la anotación dirigida a anotar las soluciones de traducción positivas<sup>5</sup> (MT-POS) y la que anota la posible interferencia de la lengua origen (MT-SLI).

Las categorías primeras relativas a la lengua de llegada (LA) y a la transferencia (TR) pueden dividirse en una tercera subcategoría. Sin embargo, a fin de no dilatar las explicaciones demasiado, no se van a detallar cada una de las subcategorías. En el siguiente apartado dedicado a los resultados se adjunta una tabla (tabla 3) donde se aprecian los acrónimos relativos a las anotaciones de tercer nivel, pero se explicará el significado de aquellos que más se han anotado en el mismo texto del artículo.

Una de las ventajas del programa Hypal4Must es que, una vez compilado el corpus, el o la docente tiene acceso a una serie de estadísticas relativas a las anotaciones realizadas. En el caso de las traducciones realizadas y subidas directamente por el alumnado al programa, también ofrece datos relativos a la evolución de cada alumno.

En nuestro caso, dado que solo cuatro y cinco alumnos han realizado y subido directamente sus traducciones al programa, utilizaremos esos datos como casos representativos para así poder mostrar la evolución de dichos alumnos.

---

<sup>5</sup> Aunque se trate de un sistema de anotación de «errores», existe una anotación para opciones positivas. Así, la anotación de errores informa, en realidad, sobre los fenómenos de traducción identificados en las traducciones del alumnado.



### 3. RESULTADOS

Con objeto de contrastar los errores cometidos por el alumnado en cada uno de los textos, a continuación, se incluyen gráficos y tablas para observar y comentar los resultados con más precisión.

El gráfico 1 muestra los resultados de las anotaciones de primer nivel, es decir, aquellas que se distinguen en tres grupos: anotaciones de lengua (LA), de transferencia (TR) y las metaanotaciones (MT).

El número total de anotaciones varía, y es mucho mayor en el caso del artículo: las anotaciones recogidas para el encargo del artículo componen el 20,13 % de todas las anotaciones del corpus EN-EU, las anotaciones del *abstract*, el 9,24 % de las anotaciones del corpus, mientras que las anotaciones para el texto turístico componen el 8,74 % de las anotaciones del corpus EN-EU.<sup>6</sup> De todos modos, cabe recordar que el artículo era el texto más largo, por lo que es lógico que surjan más anotaciones en este encargo.

Cabe destacar, por otra parte, que las anotaciones más comunes son las relativas a la lengua y a la transferencia, y que casi no se ha utilizado la metaanotación; solamente se han recogido 0,57 % de metaanotaciones en el artículo periodístico y 0,07 % en el texto turístico; no se han recogido metaanotaciones para el *abstract*. De todos modos, las anotaciones más habituales casi siempre serán las relativas a la lengua y a la transferencia entre lenguas, pues son los niveles que más subapartados recogen. En el caso de la metaanotación, tan solo abarca dos anotaciones posibles: la anotación relativa a la interferencia del texto origen y la relativa a la anotación positiva, para los casos en los que los docentes quieran premiar la traducción dada por el/la estudiante.

En cualquier caso, lo que llama la atención es que el número de anotaciones de lengua es el doble o más que el doble en los tres casos (50 % más en el caso del artículo, 75 % más en el caso del *abstract* y 59 % más en el caso del texto turístico). Resulta especialmente llamativa la diferencia tan grande entre anotaciones de lengua y de transferencia en el caso del *abstract*.

---

<sup>6</sup> El corpus EN-EU se compone de todos los encargos realizados y recopilados hasta ahora en esta dirección lingüística. No todos los encargos se han incluido en este estudio; es por eso que la suma de los porcentajes no llega al 100 %.

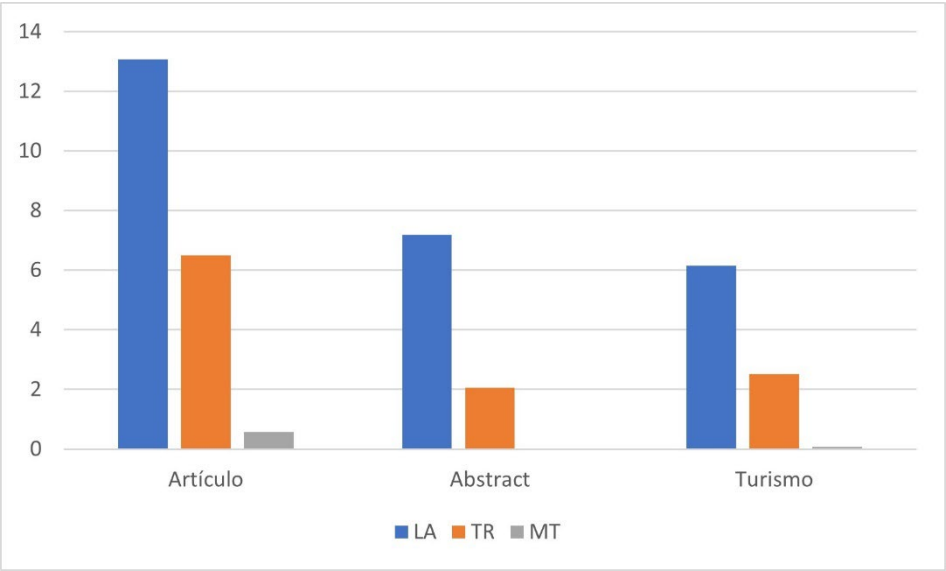


Gráfico 1. Anotaciones de primer nivel recogidas en los tres encargos.

Es evidente que el alumnado incurre sobre todo en errores relativos a la lengua meta o que, por otra parte, la docente se ha centrado más en ese tipo de errores. A fin de observar más detenidamente qué tipo de errores se han anotado, la tabla 2 muestra en porcentajes todos los errores de segundo nivel. Se han marcado en verde, azul y naranja (de más a menos) los tres errores más habituales.

	Artículo	Abstract	Turismo
LA-GR	8,8 %	3,88 %	4,12 %
TR-CT	5,91 %	1,94 %	1,31 %
LA-ME	3,18 %	0,37 %	0,87 %
TR-RC	0,37 %	0,02 %	0,79 %
LA-LT	0,38 %	2,53 %	0,48 %
LA-ST	0,24 %	0,41 %	0,66 %
TR-LE	0,15 %	0,09 %	0,41 %
TR-DP	0,06 %	0	0
MT-SLI	0,57 %	0	0,07 %
LA-CO	0,46 %	0	0,02 %

Tabla 2. Anotación de errores de segundo nivel.

Tanto en el caso del artículo periodístico como en el del *abstract* o del texto turístico, el primer error más habitual ha resultado ser de lengua, pero han sido, más específicamente, errores relativos a la gramática. Podía haberse dado el caso de que alguno de los géneros provocase errores de lengua pero de distinta índole. Sin embargo, los tres tipos de texto coinciden en el tipo de anotación más habitual. Si nos fijamos en la segunda anotación más habitual, el texto periodístico y el turístico coinciden en la anotación de transferencia-contenido, lo cual significa que ha habido errores en los que el alumnado ha alterado, de alguna manera, el significado del contenido textual. En el caso del *abstract*, esa misma anotación ha resultado ser la tercera anotación más habitual, ya que la segunda anotación más habitual también ha sido la de lengua, más en concreto, la relativa al léxico. Se podría suponer que este tipo de anotación es habitual en el caso del *abstract*, donde la terminología tiene una mayor relevancia. En artículo periodístico y el texto turístico también coinciden en la tercera anotación más habitual, que es la anotación de lengua, pero la relativa a la mecánica de la lengua: puntuación y ortotipografía. Cabe destacar que el texto periodístico y el turístico coinciden totalmente en el tipo de anotaciones, mientras que el *abstract* se distingue en la segunda y la tercera anotación más habitual.

La tabla 3 nos muestra el tercer nivel de anotación de errores, es decir, la anotación específica de todo tipo de subgrupos en cada una de las anotaciones principales.

	Artículo	<i>Abstract</i>	Turismo
TR-CT-OMI	2,27 %	0,27 %	0,02 %
LA-GR-WOR	2,84 %	1,42 %	0,76 %
LA-GR-INF	3,13 %	1,27 %	0,6 %
TR-CT-DIS	2,5 %	1,39 %	0,76 %
LA-GR	1 %	0,72 %	1,46 %
LA-ME-PUN	1,33 %	0 %	0,59 %
TR-CT-IND	1,1 %	0,28 %	0,54 %
LA-ME-SPE	1,74 %	0,37 %	0,24 %
TR-RC-CUL	0,37 %	0,02 %	0,79 %
LA-LT	0,38 %	2,53 %	0,41 %
LA-GR-PRO	0,49 %	0,07 %	0,65 %

TR-LE-TNT	0,15 %	0 %	0,35 %
LA-ST	0,11 %	0,41 %	0,13 %
LA-GR-VOI	0,09 %	0 %	0 %
LA-ME	0,07 %	0 %	0 %
LA-GR-TNS	0,92 %	0,39 %	0,39 %
LA-GR-COM	0,11 %	0 %	0,24 %
TR-DP-CON	0,06 %	0 %	0 %
LA-ST-FML	0,08 %	0 %	0 %
LA-GR-CCD	0,06 %	0 %	0,04 %
LA-ME-UND	0,04 %	0 %	0,04 %
LA-GR-DET	0,15 %	0 %	0 %
MT-SLI	0,57 %	0 %	0,07 %
LA-CO-LIN	0,46 %	0 %	0,02 %
LA-ST-RED	0,06 %	0 %	0 %
TR-CT-ADD	0,04 %	0 %	0 %
TR-LE-UNT	0 %	0,09 %	0 %
LA-ST-COV	0 %	0 %	0,06 %
TR-LE	0 %	0 %	0,04 %
LA-ST-HEA	0 %	0 %	0,48 %
LA-LT-MWN	0 %	0 %	0,07 %
TR-LE-UN	0 %	0 %	0,02 %

Tabla 3. Anotaciones de errores de tercer nivel.

Observamos que las tres anotaciones más repetidas en el artículo periodístico y en el *abstract* son muy similares. En ambos casos, las dos primeras etiquetas más frecuentes han sido de lengua, aunque en el caso del *abstract* el problema principal de lengua está asociado al léxico, fenómeno que ya habíamos observado y comentado en las anotaciones de segundo nivel. La segunda anotación más repetida para ambos tipos de texto es la relativa a la etiqueta LA-GR-WOR (gramática-sintaxis) y la tercera ha sido la relativa a la etiqueta TR-CT-DIS (contenido-distorsión del significado).

En el caso del texto turístico, la primera anotación más común también es la propia lengua y la gramática, aunque no se haya especificado el tipo de error gramatical. Cabe mencionar que, aunque el sistema de anotación permite detallar hasta en tres subapartados el tipo de anotación, el o la docente puede decidir no especificar tanto la anotación, por ejemplo, si no

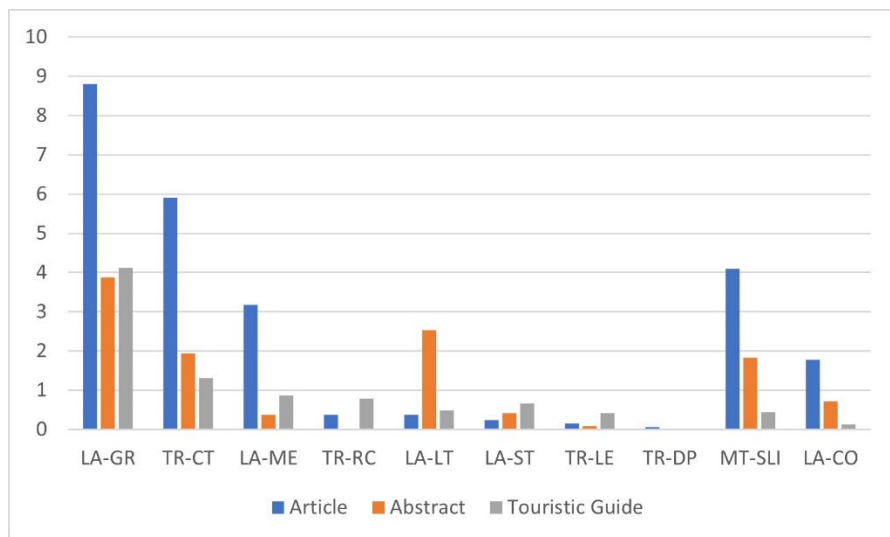
está seguro o segura a qué tipo de subapartado pertenece la anotación en cuestión. En cualquier caso, lo que sí llama la atención es que la lengua meta y, más específicamente, la gramática, resulta ser un problema también en los textos turísticos. Si observamos las segundas y las terceras anotaciones más numerosas, las diferencias no son tan grandes: la segunda anotación más habitual lo es tan solo por un 0,03 % más. Cabe resaltar, sin embargo, que, en el caso de este tipo de textos, nos encontramos con la etiqueta de TR-RC-CUL para la segunda anotación más habitual. Esta anotación es la relativa a la traducción de referencias culturales, anotación que no se encuentra entre las más habituales en el texto periodístico (solo comprende el 0,37 % de las anotaciones) ni en el *abstract* (el número es aún menor: 0,02 % de las anotaciones).

A raíz de todas estas observaciones, se podría decir que en el caso de los tres tipos textuales la lengua meta provoca más anotaciones que los fenómenos de transferencia y que, más específicamente, la gramática suele ser una de las anotaciones más habituales. Asimismo, existe una correlación entre el tipo de texto y la anotación en el caso del *abstract* (ya que la anotación LA-LT, léxico y terminología, ha resultado ser la más habitual) y en el caso del texto turístico (la anotación TR-RC-CUL relativa a las referencias culturales ha resultado ser la segunda anotación más habitual).

### 3. 1. Análisis de una muestra más específica

Tal y como se ha explicado en el apartado dedicado a la metodología, tanto el alumnado del curso 2019/2020 como el del curso 2021/2022 tradujo los mismos textos, pero los estudiantes del curso 2021/2022 subieron directamente sus traducciones a la plataforma, de modo que primero debieron darse de alta en el sistema Hypal4Must. Esto permite al docente extraer estadísticas relativas a cada estudiante. A fin de sacar el máximo rendimiento al presente estudio, analizaremos en más detalle los resultados de los cinco estudiantes que participaron en el curso 2021/2022, para así observar si los resultados globales comentados hasta ahora en torno a las anotaciones más comunes para cada tipo de texto coinciden o no. Por otra parte, dado que los estudiantes introducen datos sociolingüísticos relativos a su lengua materna, observaremos más en detalle la evolución de un estudiante con lengua materna euskera y la evolución de otro con lengua materna castellano.

Observemos, primero, en el gráfico 2, las anotaciones más comunes (anotaciones de segundo nivel) para cada tipo de texto.



*Gráfico 2.* Anotaciones más habituales para cada tipo de texto entre los estudiantes del curso 2021/2022.

Según nos muestra el gráfico 2, las anotaciones de segundo nivel más habituales entre los estudiantes del curso académico 2021/2022 para el *abstract* y el texto turístico no son muy diferentes: en el caso del *abstract*, las anotaciones más habituales siguen siendo las relativas a la lengua-gramática, el léxico y el contenido, aunque cabe mencionar que también se pueden observar casi la misma cantidad de anotaciones de interferencia que de contenido; en el caso del texto turístico, se repiten las anotaciones de lengua-gramática, transferencia-contenido y lengua-mecánica.

En el caso del artículo periodístico, los resultados sí son algo diferentes: la primera anotación más habitual sigue siendo la de lengua-gramática, la segunda anotación también es la relativa al contenido, pero observamos que la tercera anotación más habitual es la metaanotación, más en concreto, los casos de interferencia de la lengua origen. Esta anotación casi no estaba representada en los resultados globales y esto puede deberse a una atención más detallada por parte de la docente a este tipo de fenómenos. Veamos un ejemplo:

(TO: EN) *Widespread corruption and deteriorating economy have contributed to rise in sexual bribery, say researchers.*

(TM: EU) *Ikertzaileen esanetan, ustelkeria hedatuak eta ekonomiaren hondatzeak sobornatze sexuala areagotzea suposatu du* [Según dicen los investigadores, la corrupción extendida y el deterioro de la economía ha supuesto un aumento del soborno sexual]

En este ejemplo, la frase original en inglés utiliza la expresión verbal *X contributes to Y*, para expresar la consecuencia de unos problemas. El o la estudiante ha utilizado la expresión *suposatu du* (ha supuesto), que dicho en castellano suena perfectamente normal, pero no ocurre lo mismo en euskera. De hecho, esta expresión está recogida en el registro *kalkoen behatokia* (el observatorio de los calcos),<sup>7</sup> que administra y mantiene la Universidad del País Vasco (UPV/EHU) con el fin de que los docentes u otros usuarios estén informados sobre los calcos que han de evitar en sus escritos.

Efectivamente, la anotación SLI *source language interference* o interferencia de la lengua meta (subapartado de la metaanotación), no se refiere en este caso a la lengua original del texto, al inglés, sino al castellano, lengua materna primaria o secundaria para nuestros estudiantes.

Tanto el castellano como la lengua vasca son oficiales en la Comunidad Autónoma Vasca y en parte de Navarra, pero conviven en una situación diglósica para el euskera. Los estudiantes en el curso de prácticas de traducción inglés-euskera tienen dominio de la lengua vasca, pero, lógicamente, no están libres de la influencia y, por tanto, interferencia del castellano como lengua dominante. Estudios previos basados en corpus han encontrado indicios de interferencia cognitiva del castellano en traducciones literarias realizadas del alemán al euskera (Zubillaga, 2015; Sanz, 2015). Utilizamos el término «interferencia cognitiva», porque no se trata de la interferencia creada por la lengua original del texto de partida (inglés, en el caso del presente estudio), sino de la lengua que interfiere como original al tratarse de una lengua activada en la cognición del o de la estudiante (el castellano, en el caso de los estudiantes participantes en este estudio).

<sup>7</sup> Ficha número 668, donde el observatorio llama la atención del uso inadecuado de esta expresión: <https://www.ehu.eus/chg/kalkoak/?id=668> (recuperado el 18/11/2022)

El sistema de anotación de Hypal4Must no da opción de detallar la procedencia de la interferencia (de la lengua origen de los textos o de una lengua puente), aunque sería apropiado hacerlo, a fin de poder abarcar este tipo de fenómenos que podemos observar en situaciones sociolingüísticas más diferentes a los de la norma.

Sin embargo, Hypal4Must sí que recoge datos de carácter sociolingüístico en el perfil de cada estudiante (*cf.* metodología) y, a continuación, con el fin de ofrecer una visión más de contexto a este primer estudio basado en corpus, presentamos la comparación de las anotaciones del estudiante 732, quien introdujo en sus metadatos tener como lengua materna 1 el euskera y lengua materna 2 el castellano y los datos del estudiante 792, quien dijo tener como única lengua materna el castellano.

Un estudio previo realizado de forma colaborativa y que tuvo como objeto comparar las traducciones realizadas de un mismo texto origen en inglés tanto al castellano como a la lengua vasca ha constatado mayor número de anotaciones de lengua-gramática entre los estudiantes traduciendo a la lengua vasca (Izquierdo y Zubillaga, 2025), pero, aún más interesante, al observar más detenidamente solamente el alumnado que tradujo a la lengua vasca, aquellos estudiantes que aseguraban tener solo el euskera como lengua materna realizaron como media menos errores de lengua-gramática que los estudiantes que aseguraban ser bilingües.

En los gráficos 3 y 4, podemos contrastar los errores anotados como LA-GR-INF, es decir, errores gramaticales de flexión.

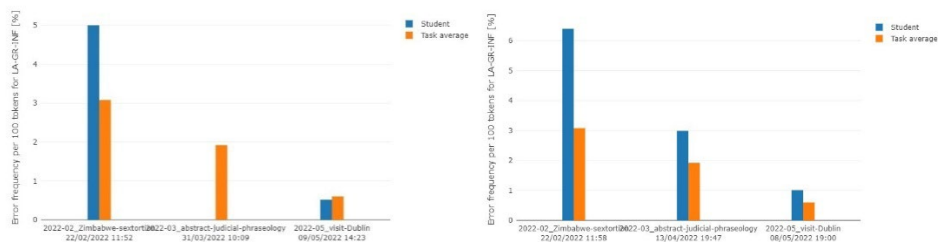
Cabe recordar que la lengua vasca es una lengua aglutinante, por lo que las palabras se forman uniendo lexemas y afijos, cada uno de ellos con su significado referencial o gramatical. Es, además una lengua ergativo-absolutiva, ya que marca con una -k (y sus variantes) el sujeto de un verbo transitivo. Así, diferenciaríamos *Gizonak etxea erosi du* [el hombre ha comprado la casa], donde el sujeto es definido y singular; *Gizonek etxea erosi dute* [los hombres han comprado la casa], donde el sujeto es definido y plural; *Edonork eros dezake etxea* [cualquiera puede comprar una casa], donde el sujeto es indefinido.

El dominio de las diferentes flexiones y sobre todo el caso del ergativo produce dificultades en algunos estudiantes, sobre todo en aquellos que hayan tenido o tengan menos exposición a la lengua. Algunas investigaciones realizadas en el ámbito de la adquisición de la lengua corroboran este hecho: dos estudios (Ezeizabarrena *et al.*, 2009 y Ezeizabarrena, 2012) analizaron la producción oral del caso ergativo en euskera de los niños escolarizados en un programa de inmersión en



euskera, con el fin de comprobar si existían diferencias entre los niños con euskera como L1 y euskera como L2. El primer estudio observó a niños de entre dos y cinco años, y encontró diferencias dependiendo de si el euskera era L1 o L2, concretamente, los niños con euskera como L2 omitían el ergativo en su producción (Ezeizabarrena *et al.*, 2009). El segundo estudio observó a niños a la edad de ocho años y concluyó que estas diferencias persistían: «children with less exposure may be expected to undergo longer OECS [etapa de caso ergativo opcional]» (Ezeizabarrena, 2012, p. 316). Estos estudios se refieren a la producción oral de niños bilingües y, por tanto, se refieren a un desarrollo lingüístico que se produce en etapas más tempranas, pero se ha detectado algo similar en etapas más tardías, entre jóvenes de 20 años (Zawiszewski *et al.*, 2011).

La exposición parece ser un factor clave, por lo que en este estudio hemos querido contrastar el número de anotaciones LA-GR-INF en el caso de un/a estudiante con euskera como primera lengua materna y el caso de un/a estudiante con castellano como única lengua materna.



Gráficos 3 y 4. Anotaciones LA-GR-INF del estudiante con lengua materna euskera (izquierda) y lengua materna castellano (derecha).

El gráfico 3 (izquierda) corresponde al/ a la estudiante con euskera como lengua materna primera, y observamos que sus anotaciones de LA-GR-INF son inferiores que la media en las tres tareas, menos en la del texto periodístico, donde sí acumula más errores de flexión que la media (38,4 % más de anotaciones). En la tarea de traducción del *abstract*, sin embargo, no acumula ni una anotación de este tipo. El o la estudiante con castellano como única lengua materna acumula más anotaciones de flexión morfológica en las tres tareas: acumula 52 % más en la traducción del artículo periodístico, 35,8 % más en la traducción del *abstract* y 40,6 % más en la traducción del texto turístico.

## DISCUSIÓN Y CONCLUSIONES

Miembros docentes e investigadores de Tralima/Itzulik nos unimos al proyecto MUST MUltilingual Student Translation con el objeto de sistematizar la corrección y colecta de las traducciones de nuestros aprendices y así tener material para realizar estudios basados en corpus de traducciones de aprendices. Una de las ventajas del proyecto es que los estudiantes incluyen metadatos de índole sociolingüístico, lo cual enriquece los datos y, por ende, el estudio traductológico. Uno de los objetivos de este artículo ha sido demostrar la utilidad de estudios basados en corpus con traducciones de aprendices, más aún si el estudio incluye metadatos sociolingüísticos.

En el presente estudio se han incluido traducciones de veinte estudiantes de los cursos 2019/2020 y 2021/2022 del grupo de prácticas de traducción inglés-euskera. Se han incluido traducciones de tres tipos de texto en concreto: un texto periodístico, un *abstract* y un texto turístico.

Los resultados demuestran un alto porcentaje de anotaciones sobre la lengua meta en los tres tipos de texto. Sin embargo, sí que se puede inferir una especie de relación entre el tipo de texto y tipo de anotación en el caso del *abstract* y en el del texto turístico. En el caso del *abstract*, se observa una mayor cantidad de anotaciones del tipo de lengua, pero, más en concreto, relativa al léxico escogido. En el caso del texto turístico, abundan anotaciones relativas a la traducción de referencias culturales, lo cual no ocurre en la misma medida en los otros tipos de texto.

La acumulación de un tipo de error y anotación en concreto en algunos de los textos podría demostrar la idoneidad de estructurar la adquisición de la competencia traductora según el encargo y el tipo de problema que se trabaja en cada uno de ellos (Hurtado, 2015; PACTE, 2018). Los textos académico y turístico del presente estudio han provocado problemas de tipo específico, por lo que serían adecuados para trabajar dichos problemas en clase.

Por otra parte, y en consonancia con otro estudio previo realizado (Izquierdo y Zubillaga, 2025), se vuelve a observar una gran cantidad de anotaciones de lengua en las prácticas de traducción inglés-euskera, pero, además, se ha podido observar que parecen existir diferencias entre estudiantes con lengua materna primera euskera y estudiantes con lengua materna primera castellano.

No es intención de este artículo realizar un análisis empírico contrastivo, ni tampoco sacar conclusiones de carácter sociolingüístico, ya

que el número de datos utilizado no es representativo. Lo que sí quedaría demostrada es la idoneidad de utilizar métodos basados en corpus como el del proyecto MUST, donde se pueden recolectar las traducciones de aprendices con las anotaciones del o de la docente y los respectivos metadatos incluidos por el alumnado. Se trata de material valioso que se puede utilizar tanto con fines docentes (utilizando ejemplos en clase) como con fines de investigación. Este estudio ha sido una primera contribución dentro del estudio de traducciones realizadas por aprendices en la combinación inglés-euskera.

### REFERENCIAS BIBLIOGRÁFICAS

- Arrarats, Irene (2004). Zuzenketa lana eguneroko prentsan: azken 13 urteotako esperientzia. *Senez*, 27, 105-112. <https://eizie.eus/eu/argitalpenak/senez/20041031/zuzenarrarats> ; <https://doi.org/10.1023/B:JEAL.00000007392.55963.3d>
- Auzmendi, Lurdes, Biguri, Koldo, Del Olmo, Karlos, Joseba Erkizia, Joseba, Mendiguren, Xabier y Zubizarreta, Iñaki (2003), *Itzulpengintza, Interpretazioa, Zuzenketa. Kulturaren Euskal Plana*. Recuperado el 4/12/2022 de [https://www.euskadi.eus/contenidos/informacion/pv\\_traducccion/eu\\_10490/adjuntos/Ponencia\\_traduc\\_e.pdf](https://www.euskadi.eus/contenidos/informacion/pv_traducccion/eu_10490/adjuntos/Ponencia_traduc_e.pdf)
- Barambones, Josu (2009). *La traducción audiovisual en ETB-1: un estudio descriptivo de la programación infantil y juvenil*. UPV/EHU.
- Bereziartua, Garbiñe y Boillos, Mari Mar (2020). Euskarazko alfabetatze akademikoa helburu: irakaskuntza sistematizatzeko urratsak. *El Guiniguada*, 29, 148-163. Disponible en <https://ojsspd.c.ulpgc.es/ojs/index.php/ElGuiniguada/article/view/1098> ; <https://doi.org/10.20420/ElGuiniguada.2020.345>
- Bilbao, Gidor y Makazaga, Jesus Mari (2015). EHUskaratuak: un corpus de traducciones académicas en una lengua minoritaria. En María Teresa Sánchez Nieto *et al.* (Ed.), *Metodologías y aplicaciones en la investigación en traducción e interpretación con corpus* =

*Methodologies and applications in corpus-based and corpus-driven Translation and Interpreting research* (pp. 323-338). Universidad de Valladolid.

Castagnioli, Sara (2009). *Regularities and Variations in Learner Translations: A Corpus-Based Study of Conjunctive Explicitation*. Universidad de Pisa.

De Sutter, Gert y Lefer, Marie Aude (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, Irene y Sánchez-Nieto, María Teresa (2019). Parallel corpora for contrastive and translation studies: New resources and applications. *Research in Corpus Linguistics*, 8 (1), 195-200. <https://doi.org/10.1075/scl.90>

Espunya, Anna (2014). The UPF learner translation corpus as a resource for translator training. *Language Resources and Evaluation*, 48, 33-43. <https://doi.org/10.1007/s10579-013-9260-1>

Ezeizabarrena, Mari Jose (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, Mari Jose, Manterola, Ibon y Beloki, Leire (2009). Euskara H2 goiztiarraren ezaugarrien bila: adizkiak eta gramatika-kasuak haurren ipuin-kontaktetan. *Euskera*, 54 2-1, 639-681. <https://doi.org/10.59866/eia.v2i54.244>

Fictumova, Jarmila, Obrušnik, Adam y Stepankova, Kristyna (2017). La enseñanza de la traducción especializada. Corpus textuales de traductores en formación con etiquetado de errores. *Sendebarr*, 28, 209-241. <https://doi.org/10.30827/sendebarr.v28i0.5419>

Granger, Silviane y Lefer, Marie Aude (2020). The multilingual student translation corpus: a resource for translation teaching and research.

- Language Resources and Evaluation*, 54, 1183-1199. <https://doi.org/10.1007/s10579-020-09485-6>.
- Hernández Guerrero, María José (2008). La traducción periodística en los diarios españoles de información general. En Luis Pegenaute *et al.* (Eds.), *Actas del III Congreso Internacional de la Asociación Ibérica de Estudios de Traducción e Interpretación. La Traducción del Futuro: Mediación Lingüística y Cultural en el siglo XXI*, (pp. 359-368). PPU. [https://www.aieti.eu/wp-content/uploads/AIETI\\_3\\_MJHG\\_Traduccion.pdf](https://www.aieti.eu/wp-content/uploads/AIETI_3_MJHG_Traduccion.pdf)
- Hernández Guerrero, María José (2019). La traducción en las nuevas formas de periodismo. *MonTI. Monografías de Traducción e Interpretación*, 5, 72-93. <https://doi.org/10.6035/MonTI.2019.ne5.3>
- Hernández Guerrero, María José (2021). News translation strategies. En Esperança Bielsa (Ed.), *The Routledge Handbook of Translation and Media* (pp. 232-249). Taylor & Francis. <https://doi.org/10.4324/9781003221678-18>
- Hurtado, Amparo (2015). The acquisition of translation competence. Competences, tasks, and assessment in translator training. *Meta*, 60 (2), 256-280. <https://doi.org/10.7202/1032857ar>
- Ibarluzea, Miren (2017). *Itzulpengintzaren errepresentazioak euskal literatura garaikidean: eremuaren autonomizazioa, literatur historiografiak eta itzultzaileak fikzioan* [Tesis doctoral]. Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU). <https://addi.ehu.es/handle/10810/26629>
- Izquierdo, Marlén (2006). Corpus-based cross-linguistic research: directions and applications. *Interlingüística*, (17), 520-527.
- Izquierdo, Marlén (2008). The ACTRES Parallel Corpus: an English-Spanish translation corpus. *Corpora*, 3 (1), 31-41. <https://doi.org/10.3366/E1749503208000051>

- Izquierdo, Marlén y Zubillaga, Naroa (2025), Learner translations in contrast: An English-Basque-Spanish case study. *Hikma*, 24 (1), 1-31. <https://doi.org/10.21071/hikma.v24i1.17070>
- Jaka, Aiora (2011), *Itzulpenari buruzko gogoeta eta itzulpen-praktika Joseba Sarrionandiaren lanetan*. UPV/EHU.
- Lavid-López, Julia, Maíz-Arévalo, Carmen y Zamorano Mansilla, Juan Rafael (2021) (Eds.). *Corpora in Translation and Contrastive Research in the Digital Age*. John Benjamins. <https://doi.org/10.1075/btl.158>
- Manterola, Elizabete (2012). *Euskal literatura beste hizkuntza batzuetara itzulia. Bernardo Atxagaren lanen itzulpenen arteko alderaketa*. UPV/EHU.
- Mees, Jürgen (2010). Eleaniztuna eta euskalduna, baita mundu akademikoan ere posible ote? *Bat: Soziolinguistika Aldizkaria*, 75, 83-94.
- Olohan, Maeve (2003). How frequent are the contractions?: A study of contracted forms in the translational English corpus. *Target*, 15(1), 59-89. <https://doi.org/10.1075/target.15.1.04olo>
- PACTE Group, Hurtado Albir, Amparo (investigadora principal), AGalán-Mañas, Anabel, Kuznik, Anna, Christian Olalla-Soler, Christian, Rodríguez-Inés, Patricia y Romero, Lupe (grupo investigador, en orden alfabético) (2018). Competence levels in translation: working towards a European framework. *The Interpreter and Translator Trainer*, 12 : 2, 111-131. <https://doi.org/10.1080/1750399X.2018.1466093> ; [https://ddd.uab.cat/pub/artpub/2018/194868/intratra\\_a\\_2018\\_v12n2p111.pdf](https://ddd.uab.cat/pub/artpub/2018/194868/intratra_a_2018_v12n2p111.pdf)
- Pontrandolfo, Gianluca (2015). Investigating judicial phraseology with COSPE: a contrastive corpus-based study. En Claudio Fantinuoli y Federico Zanettin (Eds.), *New Directions in Corpus-Based Translation Studies* (pp. 137-160). Language Science Press

- Santamaría, Alexandra (2014). *Traducción de las guías de viaje (inglés-español): análisis contrastivo de la información práctica* [Tesis doctoral]. Universidad de las Palmas de Gran Canaria (ULPGC). [https://accedacris.ulpgc.es/bitstream/10553/11913/2/0699306\\_00000\\_0000.pdf](https://accedacris.ulpgc.es/bitstream/10553/11913/2/0699306_00000_0000.pdf)
- Sanz, Zuriñe (2015). *Unitate fraseologikoen itzulpena: alemana-euskara. Literatur testuen corpusean oinarritutako analisia*. UPV/EHU.
- Sanz, Zuriñe (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>
- Zawiszewski, Adam, Gutiérrez, Eva, Fernández, Beatriz e Laka, Itziar (2011). Language distance and non-native syntactic processing: Evidence from event-related potentials. *Bilingualism: Language and Cognition*, 14 (3), 400-411. <https://doi.org/10.1017/S1366728910000350>
- Zubillaga, Naroa (2013). *Alemanetik euskaratutako haur- eta gazte-literatura: zuzeneko nahiz zeharkako itzulpenen azterketa corpus baten bidez*. UPV/EHU.
- Zubillaga, Naroa (2015). Building a trilingual parallel corpus to analyse literary translations from German into Basque. En Claudio Fantinuoli y Federico Zanettin (Eds.), *New directions in Corpus-Based Translation Studies* (pp. 61-81). Language Science Press.