CADERNOS de TRADUÇÃO

Metodología para el diseño de un asistente semiautomático de redacción y de traducción de fichas descriptivas de embutidos del español al inglés

Methodology to develop a writing and translating aid tool to transfer dried-meat product cards from Spanish into English

María Teresa Ortego Antón
Universidad de Valladolid
Soria, Castilla y León, España
mariateresa.ortego@uva.es
https://orcid.org/0000-0003-3379-4700

Resumen: La traducción y la redacción del español al inglés en el sector agroalimentario y el desarrollo de aplicaciones y herramientas que asistan durante este proceso se ha convertido en un campo prioritario para los Estudios de Traducción. A partir de la compilación de un corpus virtual comparable español-inglés compuesto por fichas descriptivas de producto describimos la metodología empleada para desarrollar GEFEM (GEnerador de Fichas de EMbutidos). Determinamos el prototipo de estructura retórica en cada lengua con la ayuda del Etiquetador Retórico© desarrollado por el grupo ACTRES. A continuación, con el Visor de Corpus Comparables© identificamos las líneas modelo, los principales patrones léxico-gramaticales, la terminología y la fraseología. Con la información lingüística se desarrolla el software, que da como resultado una herramienta que asiste a la redacción de fichas descriptivas de embutidos del español al inglés, que contiene un menú en el que se selecciona la macroestructura del texto y se corresponde con la estructura retórica y, a continuación, varias alternativas para desarrollar el contenido de cada apartado.
Palabras clave: traducción; corpus comparable; herramienta de redacción; español-inglés; embutidos.

Abstract: Translating and writing agri-food texts from Spanish into English and the development of tools to aid users during these tasks has become a key issue for Translation Studies. Based on the compilation of a virtual Spanish-English comparable corpus composed of dried-meat product cards,

the methodology applied to develop GEFEM is described. First, the prototypical rhetorical structure was identified with the Rhetorical Move Tagger© developed by ACTRES research group. Next, model lines were established with the Comparable Corpus Browser© and the most recurrent lexical grammatical patterns, terminology, and phraseology were also determined. With linguistic data the software was developed. The result is a semiautomatic corpus-based writing aid tool which assist users through writing and translating Spanish dried-meat product cards into English. The software includes a menu to select the macrostructure of the text and it corresponds with the rhetorical structure. Next, several options are offered to develop each section.
Keywords: translation; comparable corpus; writing-aid tool; Spanish-English; dried-meat products.

1. La traducción en el sector agroalimentario

Uno de los pilares de la economía española se corresponde con la industria agroalimentaria (MAPA, 2022, p. 4–5) y, dentro de esta, la industria cárnica es la primera en lo que respecta a facturación y a empleos directos. A esto hay que añadir que las empresas chacineras se caracterizan por ser pymes que se localizan en zonas rurales y tienen tradición familiar. Para poder comercializar sus productos en el exterior requieren de servicios de redacción y de traducción del español al inglés.

Dentro de los Estudios de Traducción, la traducción en el sector agroalimentario constituye un campo del saber que todavía necesita ser explorado con mayor profundidad. De hecho, fruto del interés surgido en la última década, somos testigos de publicaciones centradas en este campo, por ejemplo, números especiales en revistas como *Terminology* (Temmerman & Dubois, 2017), *Terminàlia* (SCATERM, 2017) o *Perspectives* (Li *et al*., en prensa), las ediciones de volúmenes sobre lenguas de especialidad y traducción en el sector agroalimentario (Rivas Carmona & Veroz González, 2018) sobre la terminología de los embutidos (Ortego Antón, 2019) o sobre la traducción y la interpretación para los profesionales de la agroalimentación (Peñuelas Gil & Ortego Antón, 2024a), entre otras. También se han publicado varias tesis doctorales (Ramírez Almansa, 2020; Ruiz Romero, 2020) en la Universidad de Córdoba, pionera en la Península Ibérica a la hora de incluir este ámbito en los programas formativos sobre traducción e interpretación en el nivel de la enseñanza superior.

Por otro lado, ciertos productos han atraído el foco de la investigación desde una perspectiva contrastiva en español e inglés, por ejemplo, el aceite de oliva (Montoro del Arco & Roldán Vendrell, 2013; Roldán Vendrell, 2010; Sanz Valdivieso & López Arroyo, 2022), el turrón (Santamaría Pérez, 2015, 2016, 2017) o el vino, con el proyecto WeinApp de la Universidad de Córdoba, coordinado por la Dra. Balbuena Torezano (Castellano Martínez, 2024). A estos estudios hay que sumar la investigación llevada a cabo por el grupo interuniversitario ACTRES (Análisis Contrastivo y Traducción Inglés-Español)[1], coordinado por la Dra. Rabadán Álvarez, que ha desarrollado estudios contrastivos en el par de lenguas inglés-español centrados en el sector agroalimentario en general (Rabadán *et al*., 2021), en las notas de cata del vino (López Arroyo & Roberts, 2016, 2017a, 2017b, entre otros), en las descripciones de queso en línea (Labrador & Ramón, 2015, 2020; Ramón & Labrador, 2018), en las infusiones (Pérez Blanco & Izquierdo, 2020, 2021, 2022) o en los embutidos

[1] Disponible en: https://actres.unileon.es/wp/es/home-espanol/ (Fecha de acceso: 23/02/2024).

y adobados (Ortego Antón, 2019, 2020, 2021a, 2021b, 2024; Peñuelas Gil & Ortego Antón, 2024b). Los resultados derivados de estos estudios están permitiendo desarrollar e implementar aplicaciones lingüísticas basadas en el procesamiento del lenguaje natural que asistirán en la traducción y redacción del español al inglés a las empresas del sector agroalimentario durante la exportación y les darán una mayor visibilidad en la web del conocimiento, por ejemplo, generadores de escritura bilingües, como BiTexCook, GDQ, FITEVI, GDEGA, PROMOCIONA-Té o el caso que abordamos en este trabajo, GEFEM[2].

No obstante, ante los avances de la inteligencia artificial, se podría llegar a pensar que la traducción automática neuronal podría ser la solución para dar respuesta a las necesidades de las pymes cárnicas. Sin embargo, los géneros textuales se caracterizan por utilizar patrones diferenciados que dependen de la cultura meta. Por tanto, la promoción de un determinado producto tiene que tener en cuenta las diferencias interculturales para garantizar que los textos meta satisfacen la normativa y las expectativas de la comunidad meta, "no solo en lo relativo al significado, sino también al registro, el estilo, las variantes geográficas, etc." (Durán Muñoz & Corpas Pastor, 2020, p. 164), de manera que la traducción automática no satisface las necesidades de las pymes chacineras (Ortego Antón, 2024, p. 70).

> El éxito de la comunicación no dependerá únicamente de la transmisión precisa de la información especializada relevante del campo del saber, sino también del cumplimiento de las convenciones culturales en todos los niveles. Para ello, es clave emplear la terminología adecuada y las convenciones típicas del género en cuestión (Pérez Blanco & Izquierdo, 2021, p. 148, nuestra traducción[3]).

Conscientes de la gran demanda de los productos procedentes del sector chacinero fuera de nuestras fronteras (González Fernández, 2021; Osuna Macías, 2020) y del fracaso de las aplicaciones de traducción automática en este género textual, en este trabajo pretendemos presentar la metodología que hemos empleado para desarrollar GEFEM, un asistente de traducción y de redacción del español al inglés de fichas descriptivas de embutidos basado en la compilación, análisis y explotación del corpus comparable C-GEFEM. El fin último de esta herramienta es asistir a los redactores técnicos y a los traductores a producir textos que se adecuan a las convenciones retóricas y léxico gramaticales de la lengua inglesa con la terminología, el registro y la macroestructura que satisfaga las necesidades de los hablantes de la lengua inglesa.

2. Los corpus como herramientas de traducción

El empleo de una metodología basada en el diseño, la compilación y la explotación de corpus virtuales, definidos estos como "un corpus que no contiene muchos textos pero dichos textos se adecuan al campo del saber, al género y a la variedad textual" (Corpas Pastor, 2008, p. 91, nuestra

[2] Disponible en: https://actres.unileon.es/demos/generadores/applications.html#generators2Section (Fecha de acceso: 23/02/2024).

[3] "Successful communication will depend not only on the accurate transmission of relevant subject-specific information within the professional domain, but also on compliance with cultural conventions, both at the big and small cultural levels. To this end, acceptable usage language, plus an awareness of genre conventions, are paramount".

traducción[4]) está ampliamente extendida en los Estudios de Traducción (Beeby *et al*., 2009; Berber Sardinha, 2002; Bowker, 2002; Corpas Pastor & Seghiri, 2017; Fantinuoli, 2016; Ortego Antón, 2024; Sánchez Carnicer, 2022; Sánchez Ramos, 2019, 2020; entre otros) por sus múltiples ventajas, entre las que destacan la objetividad, la reutilización, la posibilidad de conferirles múltiples usos, la facilidad para explotarlos y el acceso y la gestión de ingentes cantidades de información en muy poco tiempo (Corpas Pastor & Seghiri, 2009, p. 77).

Aunque han surgido infinidad de taxonomías para clasificar los corpus atendiendo a sus características y a su aplicabilidad según el contexto traductológico, en este trabajo nos limitaremos a emplear corpus comparables, definidos por EAGLES (1996, nuestra traduccioón[5]) como "está constituido por textos similares en más de una lengua o variedad lingüística". Entre sus características, Rabadán y Fernández Nistal (2002, p. 53) subrayan que los textos incluidos deben funcionar de forma similar en el plano de la situación comunicativa; es decir, recogen contenidos similares, están redactados en fechas cercanas y desempeñan una función semejante desde el punto de vista discursivo, de manera que la lengua meta no está influenciada por la lengua origen, lo que suele ponerse de manifiesto en las traducciones.

Por tanto, los corpus comparables son muy útiles para el estudio de la terminología y de la fraseología especializada y, además, constituyen una fuente muy valiosa para los estudios contrastivos en general y del léxico especializado en particular.

No obstante, el éxito de la comunicación interlingüística depende del uso del inglés como *lingua franca,* así como de los siguientes factores:

> Producir textos completos en una lengua extranjera respetando las particularidades retóricas, las normas y las convenciones de un determinado género; guían al usuario en el formato del género en cuestión, sugieren unidades semánticas y frases completas, en lugar de términos o elementos individuales. Las unidades que se ofrecen al usuario son resultado de un análisis cualitativo y cuantitativo de un corpus, así que el texto resultante no solo cumplirá con las convenciones normativas de la gramática, la estructura y el formato, sino que también respetará las particularidades del género de la lengua de llegada (Moreno-Pérez & López-Arroyo, 2021, p. 259–260, nuestra traducción[6]).

Entre las herramientas típicas, destacan los asistentes de redacción, que proporcionan recomendaciones, generalmente sobre terminología o estilo para mejorar los textos redactados en lengua extranjera para que parezcan nativos; las plantillas, que son modelos para un determinado tipo o género textual, son útiles para redactar y organizar el material, así como para elaborar las oraciones, los párrafos o la estructura y, por último, los generadores de escritura, que se caracterizan por ser aplicaciones que producen textos completos en la lengua meta siguiendo las particularidades retóricas, normas y convenciones de un determinado género (Moreno-Pérez & López Arroyo, 2021, p. 259–260). El producto reflejará las convenciones típicas en la lengua meta.

[4] "A corpus in which there are not many texts but that the few texts included are suited to the field of knowledge, genre and textual variety".
[5] "One which selects similar texts in more than one language or variety".
[6] "To produce full texts in a foreign language following the rhetorical particularities, norms and conventions of a given genre; they guide the user through the format of the genre in question, suggesting full semantic units and phrases, rather than terms or individual elements. The units offered to the user are based on quantitative and qualitative corpus analysis of that specific genre, so the resulting text will not only be correct in grammar, structure and format, but also reflect the particularities of the genre in the language being used".

Así pues, una vez expuestas las ventajas del uso de corpus virtuales comparables, así como el espectro de herramientas existentes, procedemos a describir la metodología de diseño de GEFEM, un GEnerador de Fichas descriptivas de EMbutidos basado en la compilación y explotación de C-GEFEM, cuyo análisis nos permitirá identificar la estructura retórica, las líneas modelo, los parámetros léxico gramaticales y la terminología prototípica que caracterizan a las fichas descriptivas de embutidos en las lenguas española e inglesa.

3. Metodología para el diseño de GEFEM

3.1 La compilación y explotación de C-GEFEM: el corpus virtual comparable español-inglés sobre fichas descriptivas de embutidos

En primer lugar, siguiendo un protocolo similar al previamente empleado por Seghiri (2017) y Ortego Antón (2019, 2020, 2024), diseñamos y compilamos C-GEFEM, que es un corpus virtual comparable español-inglés compuesto por 100 fichas descriptivas de embutidos redactadas en español y otras 100 en inglés.

Puesto que el sector de los embutidos es muy amplio, nos hemos centrado en seleccionar fichas descriptivas de tres productos —chorizo, salchichón y lomo— publicadas en Internet de 2016 a 2018, lo que nos asegura la autenticidad y, a la vez, hacen posible que procedan de una amplia variedad de autores para cumplir con el criterio de la representatividad. El subcorpus de C-GEFEM en español incluye fichas descriptivas de embutidos publicadas por empresas cárnicas españolas, en tanto que el subcorpus en inglés está compuesto por textos de Reino Unido, EE. UU., Canadá, Irlanda y Australia.

Una vez establecidos los parámetros, la compilación de C-GEFEM se compone de cuatro fases:

1. Búsqueda de textos en Internet para encontrar las fichas descriptivas de embutidos en páginas web de empresas chacineras de reconocido prestigio.
2. Descarga de los textos manualmente en formato HTML.
3. Conversión de los textos a TXT con codificación UTF-8 para que puedan ser procesados por el software de gestión de corpus.
4. Almacenamiento de los textos en la carpeta C-GEFEM, que se divide en dos subcarpetas, los originales en una carpeta denominada BIBLIOTECA DIGITAL (formato XML) compuesta por dos subcarpetas: inglés (EN) y español (ES), y los archivos en TXT en la carpeta CORPUS, clasificados por lenguas.

Figura 1: Estructura de C-GEFEM

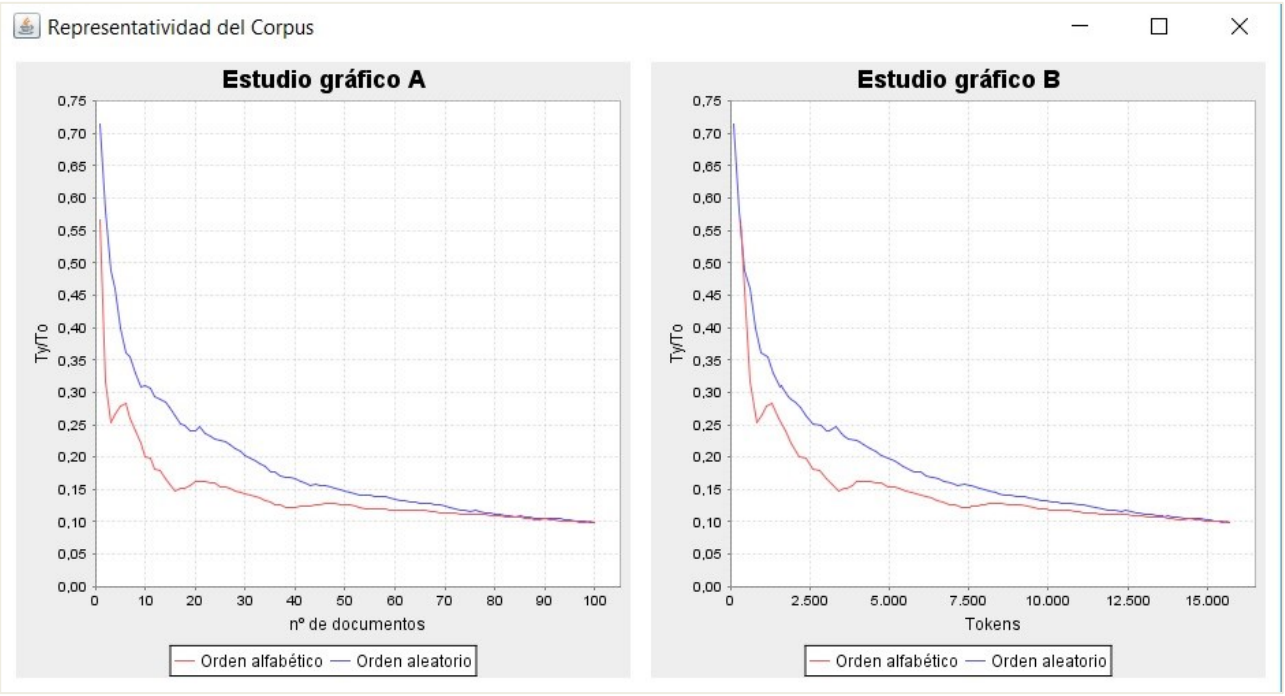
Fuente: Autora (2024)

El resultado es un corpus virtual comparable bilingüe de fichas descriptivas de embutidos, integrado por 25 425 *tokens*[7] o casos en inglés y 14 196 *tokens* en español, que es representativo a nivel cualitativo gracias a los parámetros de diseño y al protocolo de compilación seguido. La diferencia en el número de palabras en cada lengua se debe a que, en lengua inglesa, se especifican las características del producto, el embalaje y la preparación y uso, en tanto que en español la información es más sintética, probablemente por las diferencias culturales existentes.

Para concluir el proceso, hemos comprobado la representatividad cuantitativa con el programa ReCor[8] (Corpas Pastor & Seghiri, 2010; Seghiri, 2006), que calcula el número mínimo de palabras que debe incluir el corpus para ser representativo en lo relativo a la terminología básica en este género. El programa es un archivo ejecutable con una interfaz intuitiva en la que se cargan los subcorpus y el programa genera dos gráficas (Estudio gráfico A y Estudio gráfico B). Por lo que respecta al Estudio gráfico A, este presenta en el eje horizontal el número de archivos del corpus, mientras que en el eje vertical se muestra el cociente tipos/casos (Ty/To). Se recogen dos funciones, una para los archivos ordenados alfabéticamente (línea roja) y otra (línea azul) para los archivos elegidos aleatoriamente, de forma que nos aseguramos, mediante doble comprobación, que el orden de los textos no repercute en la representatividad del corpus. Cuando ambas funciones se estabilizan podemos afirmar que el corpus es representativo. Simultáneamente se genera otra gráfica (Estudio gráfico B) en la que se ofrece los casos en el eje horizontal. A partir de dicho eje se puede extraer el número de palabras mínimo que debe incluir el corpus para ser representativo en lo relativo a la terminología básica empleada en este género.

[7] Número de palabras que contiene un corpus.
[8] Pueden ampliar detalles respecto a la representatividad cuantitativa de C-GEFEM en Ortego Antón (2020, p. 178-179).

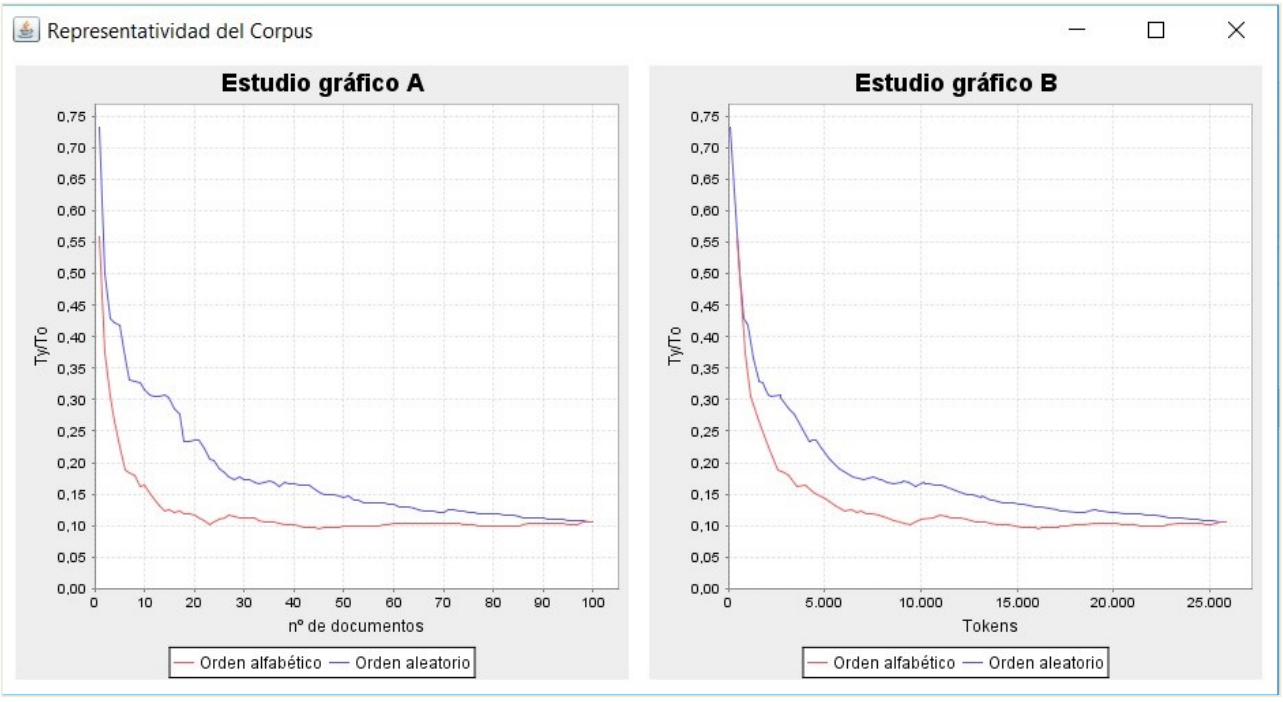
Figura 2: Representatividad cuantitativa del subcorpus en español de C-GEFEM calculada con ReCor



Fuente: Autora (2024)

Los datos de la Figura 2 muestran que el subcorpus en español de C-GEFEM es representativo a partir de los 80 documentos (Estudio gráfico A) y de las 12500 palabras (Estudio gráfico B). Por lo que respecta al subcorpus en inglés de C-GEFEM (Figura 3), este es representativo a partir de los 90 documentos (Estudio gráfico A) y las 25000 palabras (Estudio gráfico B).

Figura 3: Representatividad cuantitativa del subcorpus en inglés de C-GEFEM calculada con ReCor



Fuente: Autora (2024)

Una vez compilado C-GEFEM, procedemos a explicar cómo hemos explotado los datos del corpus siguiendo una metodología *top-down*; es decir, a partir de un análisis multiestratificado en el que accedemos al texto desde lo general hasta lo específico: la estructura retórica prototípica, las líneas modelo y el glosario con la terminología y la fraseología.

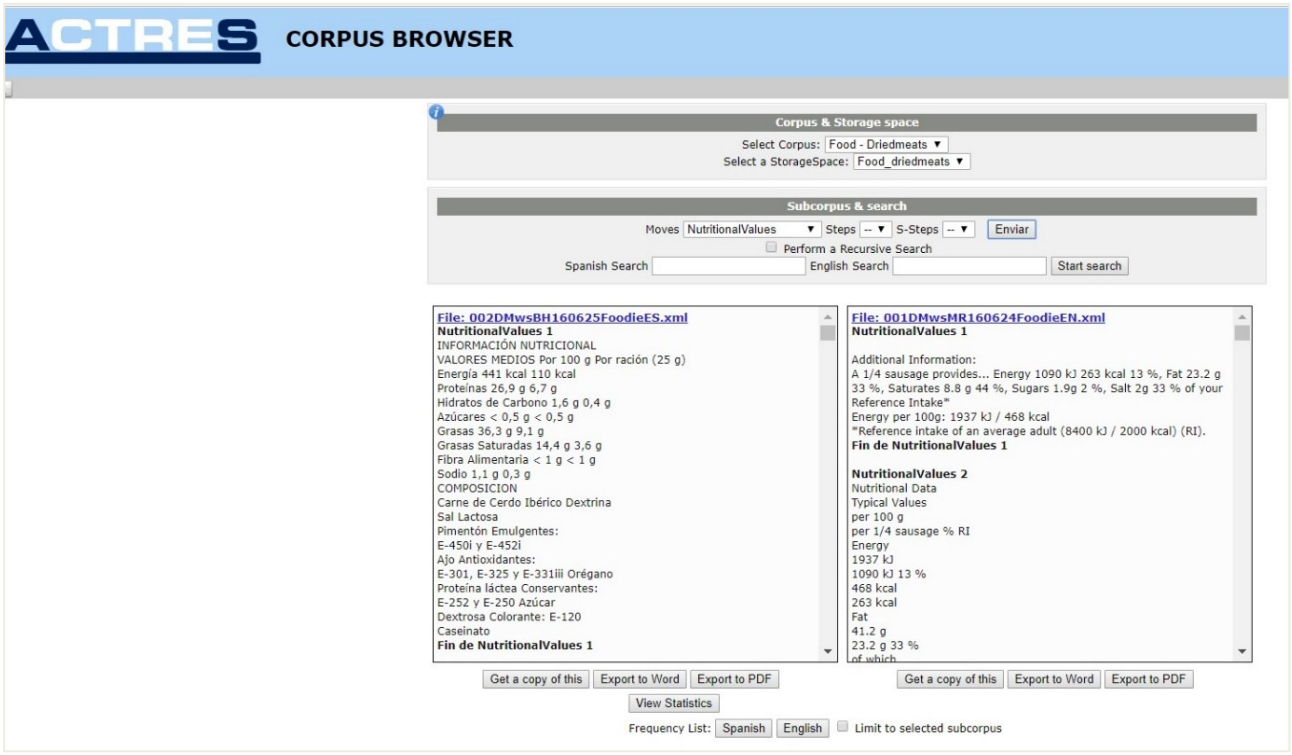
3.2 El establecimiento del prototipo de estructura retórica

Para establecer el prototipo de estructura retórica, definida esta como "la organización jerárquica de un texto. Incluye varias secciones y subsecciones del texto, movimientos y pasos" (López Arroyo & Roberts, 2015, p. 155, nuestra traducción[9]), hemos seguido la metodología propuesta por Biber *et al.* (2007). Estos autores consideran que los géneros textuales se caracterizan por estar formados por una serie de componentes retóricos denominados *moves* o movimientos, definidos como "una unidad disursiva o retórica que realiza una función comunicativa coherente" (Biber *et al.*, 2007, p. 23, nuestra traducción[10]). A su vez, estos movimientos pueden dividirse en varios *steps* o pasos, cuya función es "lograr la finalidad del movimiento al que pertenece" (Biber *et al.*, 2007, p. 24, nuestra traducción[11]). Esta metodología permite identificar las características lingüísticas de los movimientos, así como la descripción de su estructura y distribución léxica, además de ofrecer su posición en relación a otros movimientos y posibilitar el desarrollo de un determinado género textual.

Por tanto, establecemos una batería de etiquetas retóricas asociadas a posibles movimientos y pasos y las introducimos en el Etiquetador de movimientos retóricos®[12]. A continuación, etiquetamos manualmente cada uno de los textos asignando las etiquetas relativas a los movimientos y pasos, seguidamente, procedemos a comparar los distintos movimientos y pasos en español y en inglés con el Visor de corpus comparables bilingües®[13], que incluye un menú que permite analizar y contrastar la información retórica, así como un analizador de concordancias.

[9] "The hierarchic organization of a text. It involves the various sections and subsections of a text, moves and step".
[10] "A discoursal or rethorical unit that performs a coherent communicative function".
[11] "To achieve the purpose of the move to which it belongs".
[12] Disponible en: https://actres.unileon.es/wp/es/rhetorical-move-tagger-es/ (Fecha de acceso: 23/02/2024).
[13] Disponible en: https://actres.unileon.es/wp/es/comparable-corpus-browser-es/ (Fecha de acceso: 23/02/2024).

Figura 4: Visor de Corpus Comparables©



Fuente: Autora (2024)

Comprobamos el porcentaje de movimientos y pasos en cada uno de los subcorpus anotados, la frecuencia, el porcentaje de textos en los que se incluyen dichos movimientos y pasos, así como el número de palabras total de cada movimiento o paso. Los datos resultantes de dicha comprobación nos han permitido desarrollar el prototipo de estructura retórica formada por movimientos y pasos en cada una de las lenguas de trabajo como se muestra en la Tabla 1. La frecuencia de uso la hemos representado con estrellas, siendo cinco estrellas (*****) el símbolo de obligatoriedad (81% - 100%), cuatro estrellas (****) una alta aparición (61% - 80%), tres estrellas (***) una frecuencia media (41% - 60%), dos estrellas (**) poca frecuencia (21% - 40%) y una estrella (*) una escasa aparición (1% - 20%).

Tabla 1: Prototipo de estructura retórica en español y en inglés

Prototipo de estructura retórica en español	Frec.	Prototipo de estructura retórica en inglés	Frec.
1. Denominación del producto (*****)	100 %	1. Product name (*****)	98 %
2. Imagen embutido (*****)	96 %	2. Weight (***)	47 %
3. Marca (***)	41 %	3. Product image (*****)	87 %
3.1. Nombre (***)	29 %	4. Product code (**)	27 %
3.2. Logo (*)	2 %	5. Conceptual information (***)	53 %
3.3. Descripción (***)	41 %	6. Description (*****)	81 %
4. Información conceptual (**)	29 %	7. Information	
5. Código del producto (**)	21 %	7.1. Brand (**)	27 %
6. Descripción del producto (*****)	96 %	7.1.1. Brand description (**)	27 %

7. Información del producto		7.2. Storage (***)	59 %
7.1. Peso (*****)	84 %	7.3. Origin (***)	46 %
7.2. Ingredientes (*****)	90 %	7.3.1. Packed country (*)	10 %
7.2.1. Aditivos (**)	16 %	7.4. Preparation and use (****)	60 %
7.3. Alérgenos (*****)	19 %	7.5. Packaging info (***)	46 %
7.4. Información nutricional (****)	66 %	7.6. Recycling (*)	13 %
7.5. Curado (**)	35 %	7.7. Other information (**)	17 %
7.6. Conservación (*****)	85 %	8. Ingredients (****)	78 %
7.7. Utilización (**)	40 %	8.1. Additives (*)	10 %
8. Origen (*)	7 %	8.2. Allergens (***)	49 %
9. Envasado (***)	46 %	8.3. Suitable for (**)	12 %
10. Fabricante (***)	52 %	9. Nutritional values (***)	59 %
10.1. Dirección (***)	51 %	10. Manufacturer (*)	13 %
		10.1. Manufacturer address (**)	36 %
		11. Return to address (**)	24 %
		12. Review (**)	26 %
		12.1. Comments (**)	22 %
		13. Follow on (*)	9 %

Fuente: Autora (2024)

Una vez establecido el prototipo de estructura retórica, cuya descripción ha sido abordada en detalle previamente (Ortego Antón, 2019, p. 111–118, 2020, p. 188–189), nos gustaría señalar que la estructura propuesta tendrá carácter dinámico en GEFEM; es decir, los usuarios podrán modificarla en función de sus necesidades, puesto que los diferentes movimientos y pasos no son ni obligatorios ni restrictivos y es el usuario quien decide qué datos desea incorporar en función de sus necesidades.

3.3 La identificación de las líneas modelo

La siguiente fase se corresponde con la identificación de las líneas modelo, que Pérez Blanco e Izquierdo (2021, p. 157) definen como las oraciones típicas donde el contenido y el formato son estándar porque comparten patrones fraseológicos y léxico-gramaticales. Con el Visor de Corpus Comparables© manualmente revisamos cada uno de los movimientos y pasos, analizamos las ocurrencias y detectamos las líneas modelo más frecuentes para cada movimiento y paso. El resultado se corresponde con patrones obligatorios, que se representan con paréntesis, de manera que es necesario insertar una palabra o grupo de palabras de una lista de selección. También se emplean corchetes para señalar que la información en ellos recogida es opcional y puede omitirse. Por último, las llaves indican que una de las dos opciones delimitadas por la barra ha de seleccionarse. Por ejemplo, el paso "Alérgenos" —*Allergens* en inglés— tiene dos líneas modelo posibles y se enumeran los ejemplos de uso para facilitar la tarea al usuario:

- Ejemplo 1: *[Dietary information]. {Allergen / Allergy} information / Allergy advice}. For allergens, see {{highlighted / capitalised} ingredients / ingredients in bold}. {{Contains / May contain} (ALÉRGENO) / (ALÉRGENO) free}.*
    - *Allergen Information: For allergens, see highlighted ingredients.*
    - *Allergy Advice: For allergens, see highlighted ingredients.*
    - *Dietary Information. Allergy Advice. For allergens, see ingredients in bold.*
- Ejemplo 2: *[Dietary information]. {{Contains / May [also] contain} (ALÉRGENO) / (ALÉRGENO) free / This product is (ALÉRGENO) free}.*
    - *Dietary Information. May contain Milk. May Contain Soya. May contain traces of soya and milk.*
    - *Dietary Information. Contains Milk, May Contain Nuts, May Contain Soya / Soybeans.*
    - *This product is Wheat, Gluten and Dairy Free.*

Tras exponer las líneas modelo, para los casos en los que el usuario debe elegir un término, GEFEM también incorpora un diccionario con la terminología más frecuente y su fraseología, que se describe en el siguiente apartado.

3.4 La extracción de la terminología y su fraseología

Para obtener los términos más frecuentes de C-GEFEM hemos utilizado TermoStat Web 3.0. (Drouin, 2003), un extractor terminológico semiautomático que ofrece candidatos a término. A continuación, hemos validado estos candidatos aplicando los criterios propuestos por L'Homme (2020, p. 72–75) en cada lengua y manualmente hemos establecido los equivalentes al inglés teniendo en cuenta los contextos de uso. Aunque se podría pensar que este proceso podría ser más simple utilizando un corpus paralelo, estudios previos (Ortego Antón, 2019, p. 187, 2021b, p. 105) ponen en evidencia que muchos de los equivalentes propuestos en los textos traducidos no se emplean en los textos redactados originalmente en lengua inglesa.

Seguidamente procedemos a la recogida de los términos y de sus equivalentes clasificándolos en función de los distintos campos semánticos —aditivos, alérgenos, elementos nutricionales, empaquetado, ingredientes, materiales, origen y país— acompañados de la categoría gramatical y de un ejemplo de uso, como se muestra en la Figura 5.

Figura 5: Términos pertenecientes al campo semántico de ingredientes

	A	B	C	D	
1	ES	TIPO	EN	TIPO	EJEMPLO
2	aceite de maíz	N	corn oil	N	Pork, Water, Salt, Nonfat Dry Milk, Dextrose, Paprika,
3	aceite de oliva	N	olive oil	N	Pork, Salt, Paprika, Olive Oil, Garlic, Antioxidant (Sodiu
4	aceite vegetal	N	vegetable oil	N	Pork, Water, Salt, Nonfat Dry Milk, Dextrose, Paprika,
5	ácido ascórbico	N	ascorbic acid	N	Pork, Water, Salt, Nonfat Dry Milk, Dextrose, Paprika,
6	ácido cítrico	N	citric acid	N	Pork, Water, Salt, Nonfat Dry Milk, Dextrose, Paprika,
7	agua	N	water	N	Pork, Water, Salt, Nonfat Dry Milk, Dextrose, Paprika,
8	ajo	N	garlic	N	Pork, Salt, Paprika, Olive Oil, Garlic, Antioxidant (Sodiu

Fuente: Autora (2024)

A partir de esta base de datos hemos desarrollado e-DriMe (Ortego Antón, 2021a), un diccionario electrónico inglés-español basado en corpus sobre la terminología de los embutidos que sigue los principios de la Teoría Funcional de la Lexicografía (Bergenholtz & Tarp, 2002, 2003) y en la semántica léxica aplicada a la terminología (L'Homme, 2020). Una vez descrita la explotación de C-GEFEM, procedemos a explicar el desarrollo del software.

4. GEFEM: el asistente de redacción de fichas descriptivas de embutidos del español al inglés

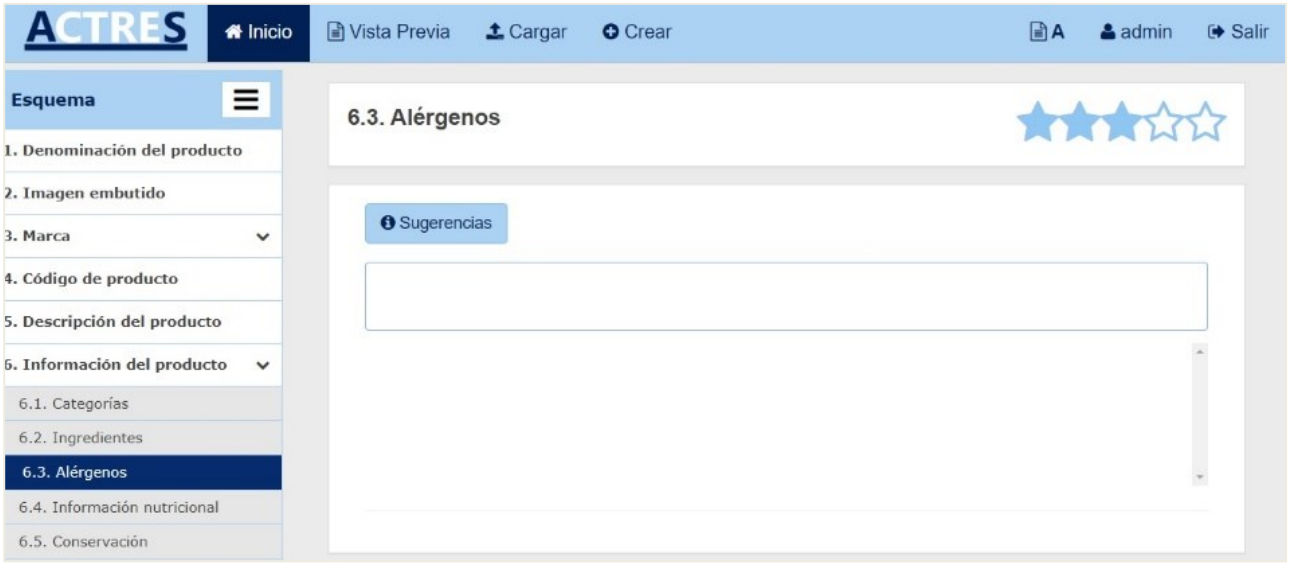
GEFEM[14] es un software dinámico basado en la web con una interfaz intuitiva que va guiando al usuario durante la redacción de una ficha descriptiva de embutido, es gratuito y está alojado en los servidores de la Universidad de León. En primer lugar, la aplicación permite crear un documento en blanco o cargar un documento en el que el usuario ya había trabajado previamente, como se muestra en la Figura 6. Si optamos por la primera opción, "Crear documento", la aplicación nos muestra un menú en el lado izquierdo de la pantalla que permite desplazarse entre los distintos movimientos y pasos, como se observa en la Figura 7.

Figura 6: Interfaz de inicio de GEFEM

Fuente: Autora (2024)

[14] Disponible en: https://actres.unileon.es/demos/generadores/applications.html#generators2Section (Fecha de acceso: 24/02/2024).

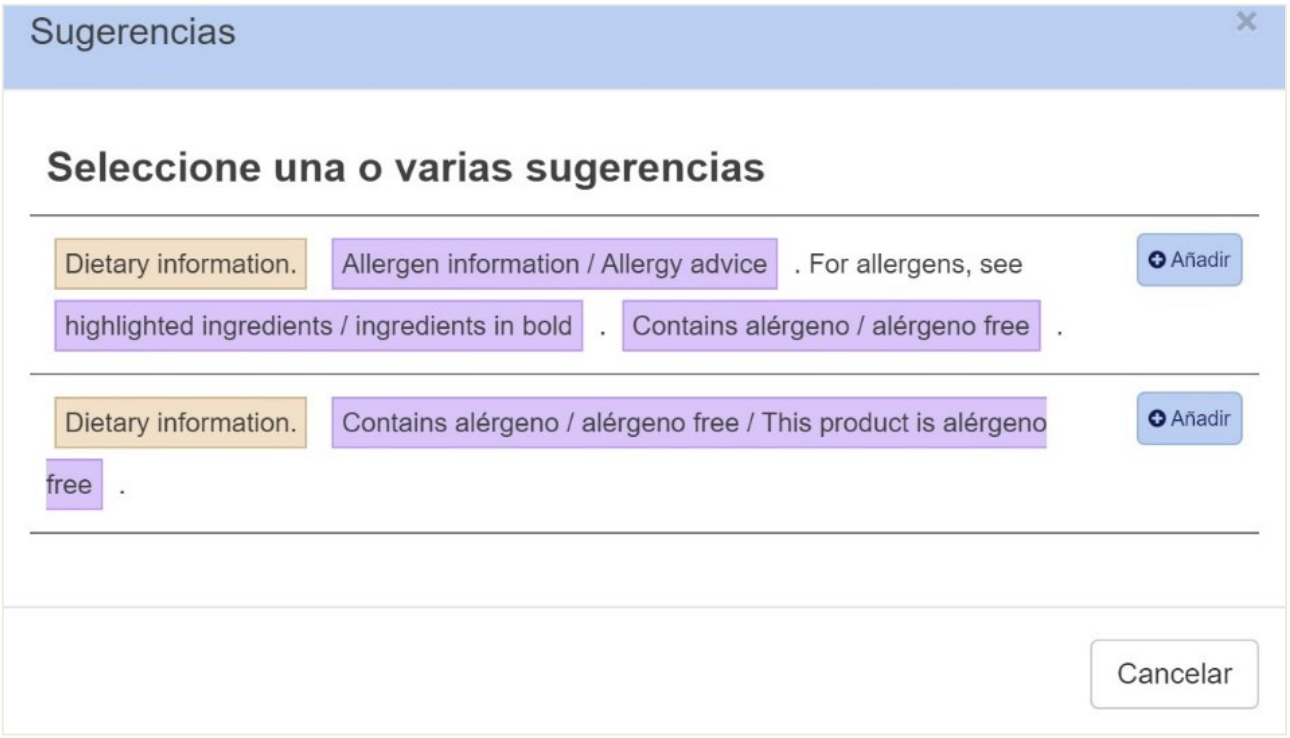
Figura 7: Menú desplegable con la estructura de la ficha de producto



Fuente: Autora (2024)

Por ejemplo, seleccionamos el paso "Alérgenos", hacemos clic en el botón "Sugerencias", se despliega una ventana donde aparecen las dos líneas modelo y el usuario debe decantarse por una de ellas pulsando "Añadir", como se muestra en la Figura 8.

Figura 8: Líneas modelo del paso "Alérgenos"



Fuente: Autora (2024)

Además, la herramienta ofrece distintos elementos que van guiando al usuario con botones de colores diferentes. Por ejemplo, el color verde indica que hay que insertar una palabra o grupos

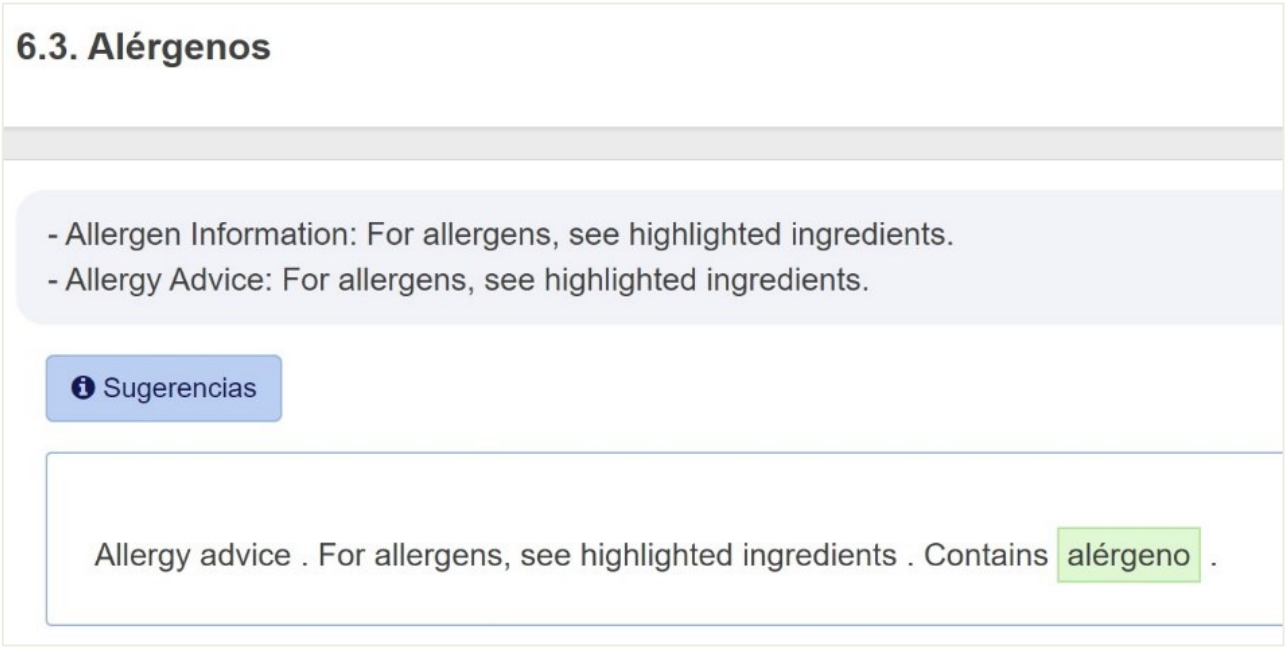
de palabras, el morado muestra dos o más opciones entre las que el usuario debe escoger una y el naranja se corresponde con un fragmento de texto opcional.

En este punto se puede apreciar que es necesario que el usuario tenga unos conocimientos mínimos de lengua inglesa, dado que estamos ante una herramienta de asistencia a la redacción y a la traducción, que no debe confundirse con los sistemas de traducción automática, que trasvasan directamente el contenido sin tener en cuenta las convenciones típicas de la lengua de llegada.

Por tanto, el usuario va completando el contenido de la ficha descriptiva de producto siguiendo las instrucciones de los distintos recuadros y, cuando tiene que escoger una unidad léxica de la base de datos terminológica, esta aparece en verde, como se puede observar en la Figura 9.

Al hacer clic en "alérgeno", se despliega una ventana en la que el usuario introduce los primeros caracteres del término en español y se ofrecen los equivalentes en inglés, como se muestra en la Figura 10.

Figura 9: Ejemplo de unidad léxica procedente de la base de datos terminológica



Fuente: Autora (2024)

Figura 10: Ejemplo de menú desplegable para escoger una unidad léxica

Fuente: Autora (2024)

Cuando el usuario ha completado todos los campos, hace clic en el botón de vista previa del menú superior y obtendrá la ficha descriptiva de producto en lengua inglesa, en formato DOCX, que se ajusta a las convenciones lingüísticas y culturales del inglés.

Para finalizar, estamos ante una herramienta de asistencia a la traducción y a la redacción del español al inglés que satisface las necesidades de los profesionales de la comunicación en este campo del saber y, entre sus potenciales usuarios, se encuentran las pymes del sector cárnico. La herramienta proporciona información fiable, basada en corpus, precisa y representativa, lo que resulta más barato que la posedición de la traducción automática.

5. Conclusiones

A lo largo de este trabajo hemos ofrecido la descripción de GEFEM, una herramienta que asiste a los usuarios durante la traducción y la redacción del español al inglés de fichas descriptivas de embutidos basada en la explotación de C-GEFEM, un corpus virtual comparable en las lenguas española e inglesa.

El desarrollo de aplicaciones lingüísticas como GEFEM hace posible que los resultados de la investigación puedan transferirse al tejido productivo, de manera que las pymes puedan tener acceso a aplicaciones y herramientas que mejoran su potencial de exportación y de internacionalización ante la imposibilidad de contratar profesionales lingüísticos y, así, puedan ofrecer información de calidad en la visualización de su actividad en la web del conocimiento.

Desde la perspectiva académica, somos conscientes de que es necesario seguir formando a los futuros traductores y redactores multilingües en las particularidades de los distintos géneros lingüísticos para que no se limiten únicamente a trasvasar el contenido a otra lengua, sino también a adaptar los textos que se traducen a las convenciones de la cultura de la lengua de llegada para que el texto funcione en la situación comunicativa de la mencionada cultura.

Para finalizar, GEFEM está en proceso de mejora, puesto que se está implementando su vinculación con sistemas de traducción automática y, al mismo tiempo, seguimos desarrollando herramientas de ayuda a la redacción multilingüe para otros campos del sector agroalimentario.

Referencias

Beeby, A., Rodríguez Inés, P., & Sánchez Gijón, P. (2009). *Corpus Use and Translating*. John Benjamins.

Berber Sardinha, T. (2002). Corpora electrônicos na pesquisa em tradução. *Cadernos de Tradução*, *9*(1), 15–59.

Bergenholtz, H., & Tarp, S. (2002). Die moderne lexikographische Funktionslehre: Diskussionsbeitrag zu neuen und alten Paradigmen, die Wörterbücher als Gebrauchsgegenstände verstehen. *Lexicographica*, *18*, 253–263.

Bergenholtz, H., & Tarp, S. (2003). Two opposing theories: On H.E. Wiegand's recent discovery of lexicographic functions. *Hermes Journal of Linguistics*, *31*, 171–196.

Biber, D., Connor, U., & Upton, T. A. (2007). Discourse on the Move. Using Corpus Analysis to Describe Discourse Structure. John Benjamins.

Bowker, L. (2002). Computer-Aided Translation Technology: A Practical Introduction. University of Ottawa Press.

Castellano Martínez, J. M. (Ed.). (2024). *Lengua, Cultura y Traducción: Andalucía como destino enoturístico.* Peter Lang. https://doi.org/10.3726/b20627

Corpas Pastor, G. (2008). Investigar con corpus en traducción: los retos de un nuevo paradigma. Peter Lang.

Corpas Pastor, G., & Seghiri, M. (2009). Virtual Corpora as Documentation Resources: Translating Travel Insurance Documents (English-Spanish). In A. Beeby, P. Rodríguez Inés & P. Sánchez-Gijón (Eds.), *Corpus Use and Translating* (pp. 75–107). John Benjamins.

Corpas Pastor, G., & Seghiri, M. (2010). Size Matters: A Quantitative Approach to Corpus Representativeness. In R. Rabadán *et al.* (Eds.), *Lengua, traducción, recepción: en honor de Julio César Santoyo* (pp. 111–145). Universidad de León, Área de Publicaciones.

Corpas Pastor, G., & Seghiri, M. (Eds.). (2017). *Corpus-based Approaches to Translation and Interpreting: From Theory to Applications*. Peter Lang. https://doi.org/10.3726/b10354

Drouin, P. (2003). Term Extraction Using Non-technical Corpora as a Point of Leverage. *Terminology*, *9*(1), 99–117.

Durán Muñoz, I., & Corpas Pastor, G. (2020). Corpus-Based Multilingual Lexicographic Resources for Translators: An Overview. In M. J. Domínguez Vázquez *et al.* (Eds.), *Studies on Multilingual Lexicography* (pp. 159–178). De Gruyter. https://doi.org/10.1515/9783110607659-009

EAGLES. (1996). *Preliminary Recommendations on Corpus Typology*. Documento técnico EAGLES EAG-TCWG-CTYP/P.

Fantinuoli, C. (2016). Revisiting corpus creation and analysis tools for translation tasks. *Cadernos de Tradução*, *36*(1), 62–87. https://doi.org/10.5007/2175-7968.2016v36nesp1p62

González Fernández, J. (2021). El mercado del jamón y la charcutería en China: Estudio de mercado. ICEX.

Labrador, B., & Ramón, N. (2015). 'Perfectly smooth creamy and full flavoured': Online cheese descriptions. *Procedia: Social and Behavioural Sciences, 198*, 226–232. https://doi.org/10.1016/j.sbspro.2015.07.440

Labrador, B., & Ramón, N. (2020). Building a Second-language Writing Aid for Specific Purposes: Promotional Cheese Descriptions. *English for Specific Purposes*, *60*, 40–52. https://doi.org/10.1016/j.esp.2020.03.003

Li, S., Fuentes-Luque, A., & Desjardins, R. (Eds.). (En prensa). *Perspectives*, *32*.

L'Homme, M. C. (2020). *Lexical Semantics for Terminology: An Introduction*. John Benjamins. https://doi.org/10.1075/tlrp.20

López-Arroyo, B., & Roberts, R. P. (2015). The Use of Comparable Corpora: How to Develop Writing Applications. In M. T. Sánchez Nieto (Ed.), *Corpus Based Translation and Interpreting Studies: From Description to Application* (pp. 147–156). Frank & Timme.

López-Arroyo, B., & Roberts, R. P. (2016). Differences in wine tasting notes in English and Spanish. *Babel*, *62*(3), 370–401. https://doi.org/10.1075/babel.62.3.02lop

López-Arroyo, B., & Roberts, R. P. (2017a). El lenguaje metafórico en las fichas de carta de vino en inglés y en español. *Hermēneus: Revista de Traducción e Interpretación*, *19*, 139–163. http://uvadoc.uva.es/handle/10324/27782

López-Arroyo, B., & Roberts, R. P. (2017b). Genre and Register in Comparable Corpora: An English/Spanish Contrastive Analysis. *Meta*, *62*(1), 114–136. https://doi.org/10.7202/1040469ar

MAPA. (2022). *Informe anual de la industria alimentaria española periodo 2021-2022.* https://www.mapa.gob.es/es/alimentacion/temas/industria-agroalimentaria/20220728informeanualindustria2021-20222t22ok_tcm30-87450.pdf

Montoro del Arco, E. T., & Roldán Vendrell, M. (2013). Terminología, normalización y comunicación: Las categorías del aceite de oliva en español, inglés y chino. *Terminology, 19*(1), 62–92. https://doi.org/10.1075/term.19.1.03mon

Moreno Pérez, L., & López Arroyo, B. (2021). A Typical Corpus-based Tools to the Rescue: How a Writing Generator Can Help Translators Adapt to the Demands of the Market. *MonTI, 13*, 251–279. https://doi.org/10.6035/MonTI.2021.13.08

Ortego Antón, M. T. (2019). La terminología del sector agroalimentario (español-inglés) en los estudios contrastivos y de traducción especializada basados en corpus: los embutidos. Peter Lang. https://doi.org/10.3726/b15808

Ortego Antón, M. T. (2020). Las fichas descriptivas de embutidos en español y en inglés: un análisis contrastivo de la estructura retórica basado en corpus. *Revista Signos Estudios de Lingüística, 53*(102), 170–194.

Ortego Antón, M. T. (2021a). e-DriMe: A Spanish-English frame-based e-dictionary about dried-meats. *Terminology*. *27*(2), 330–357. https://doi.org/10.1075/term.20013.ort

Ortego Antón, M. T. (2021b). Los corpus como herramientas para la traducción español-inglés en el sector chacinero en el siglo XXI. In E. Sartor (Ed.), *Los corpus especializados en la lingüística aplicada: traducción y enseñanza* (pp. 89–112). Universitas Studiorum Editrice.

Ortego Antón, M. T. (2024). The Design of Torrezno TRAD: the Semiautomatic Spanish-English Writing and Translation Aid Tool. In I. Peñuelas Gil & M. T. Ortego Antón (Eds.), *Interpreting and Translation for Agri-food Professionals in the Global Marketplace* (pp. 69–84). De Gruyter. https://doi.org/10.1515/9783111101729-004

Osuna Macías, J. Á. (2020). *El mercado del jamón y de los embutidos porcinos en Estados Unidos: Estudio de mercado.* ICEX. https://www.icex.es/icex/wcm/idc/groups/public/documents/documento/mdiw/odyz/~edisp/doc2020863140.pdf

Pérez Blanco, M., & Izquierdo, M. (2020). A multi-level contrastive analysis of promotional strategies in specialized discourse. *English for Specific Purposes*, *58*, 43–57. https://doi.org/10.1016/j.esp.2019.12.002

Pérez Blanco, M., & Izquierdo, M. (2021). Developing a Corpus-informed Tool for Spanish Professionals Writing Specialized Texts in English. In J. Lavid-López *et al.*, (Eds.), *Corpora in Translation and Contrastive Research in the Digital Age* (pp. 147–173). John Benjamins. https://doi.org/10.1075/btl.158.06per

Pérez Blanco, M., & Izquierdo, M. (2022). Engaging with Customer's Emotions: A Case Study in English-Spanish Online Food Advertising. *Languages in Contrast*, *22*(1), 43–76. https://doi.org/10.1075/lic.19016.per

Peñuelas Gil, I., & Ortego Antón, M. T. (Eds.). (2024a). *Interpreting and Translation for Agri-food Professionals in the Global Marketplace.* De Gruyter. https://doi.org/10.1515/9783111101729

Peñuelas Gil, I., & Ortego Antón, M. T. (2024b). El prototipo de estructura retórica en español y en inglés de las fichas descriptivas de torrezno y adobados. In J. Sánchez Carnicer & L. Arce Romeral (Eds.), *Nuevos avances tecnológicos en la teoría y práctica de la Traducción e Interpretación.* Peter Lang.

Rabadán, R., & Fernández Nistal, P. (2002). *La traducción inglés-español: fundamentos, herramientas, aplicaciones.* Universidad de León/ITBYTE.

Rabadán, R., Pizarro Sánchez, I., & Sanjurjo González, H. (2021). Authoring Support for Spanish Language Writers: A genre-restricted Case Study. *RESLA, 34*(2), 677–717. https://doi.org/10.1075/resla.19048.rab

Ramírez Almansa, I. (2020). *Terminología y traducción en contextos especializados (alemán-español): Vitivinicultura.* [Tesis doctoral]. Universidad de Córdoba. https://helvia.uco.es/xmlui/handle/10396/19636

Ramón, N., & Labrador, B. (2018). Selling cheese online. *Terminology, 24*(2), 210–235. https://doi.org/10.1075/term.00019.ram

Rivas Carmona, M. M., & Veroz González, M. Z. (Eds.). (2018). *Agroalimentación: lenguajes de especialidad y traducción.* Comares.

Roldán Vendrell, M. (2010). Bases para la terminología bilingüe del aceite de oliva. Comares.

Ruiz Romero, M. A. (2020). *Tipología textual y traducción en el ámbito agroalimentario: definición de perfil y ejercicio profesional.* [Tesis doctoral]. Universidad de Córdoba. http://hdl.handle.net/10396/20185

Sánchez Carnicer, J. (2022). *Traducción y discapacidad: un estudio comparado de la terminología inglés-español en la prensa escrita.* Peter Lang. https://doi.org/10.3726/b19567

Sánchez Ramos, M. M. (2019). Corpus paralelos y traducción especializada: ejemplificación de diseño, compilación y alineación de un corpus paralelo bilingüe (inglés-español) para la traducción jurídica. *Lebende Sprachen*, *64*(2), 269–285.

Sánchez Ramos, M. M. (2020). Documentación digital y léxico en traducción e interpretación en los servicios públicos (TISP): fundamentos teóricos y prácticos. Peter Lang. https://doi.org/10.3726/b16632

Santamaría Pérez, I. (2015). *Diccionario LID del turrón.* LID Editorial.

Santamaría Pérez, I. (2016). Diseño, implementación y elaboración de una terminología multilingüe del ámbito del turrón, mazapanes y otros dulces. *Cuadernos ASPI, 6*, 75–94.

Santamaría Pérez, I. (2017). La terminología del torró. *Terminàlia, 15*, 59–60.

Sanz Valdivieso, L., & López Arroyo, B. (2022). The phraseology of wine and olive oil tasting notes: A corpus based semantic analysis. *Terminology, 28*(1), 37–64. https://doi.org/10.1075/term.20035.lop

SCATERM. (2018). Dossier: Gastronomia i terminologia Semblança: Isidra Maranges i Prat (1919-2012). *Terminàlia*, 15.

Seghiri, M. (2006). Compilación de un corpus trilingüe de seguros turísticos (español-inglés-italiano): aspectos de evaluación, catalogación, diseño y representatividad. [Tesis doctoral]. Universidad de Málaga. http://hdl.handle.net/10630/2715

Seghiri, M. (2017). Metodología de elaboración de un glosario bilingüe y bidireccional (inglés-español/español-inglés) basado en corpus para la traducción de manuales de instrucciones de televisores. *Babel*, *63*(1), 43–64. https://doi.org/10.1075/babel.63.1.04seg

Temmerman, R., & Dubois, D. (2017). Food and terminology: Expressing sensory experience in several languages. *Terminology, 23*(1), 1–8. https://doi.org/10.1075/term.23.1

Notas

Contribución de autoría

Concepción y elaboración del manuscrito: M. T. O. Antón
Recolección de datos: M. T. O. Antón
Análisis de datos: M. T. O. Antón
Discusión y resultados: M. T. O. Antón
Revisión y aprobación: M. T. O. Antón

Datos de la investigación

No se aplica.

Financiación

El presente trabajo se ha realizado en el marco del proyecto nacional de I+D titulado "Lenguajes naturales controlados, comunicación colaborativa y producción textual bilingüe en entornos 3.0" (PID2020-114064RB-I00), coordinado por la Dra. Noelia Ramón García (Universidad de León) y parcialmente en el seno de los proyectos nacionales de I+D titulados "Multi-lingual and Multi-domain Adaptation for the Optimisation of the VIP system" (VIP II) (PID2020-112818GB-I00), coordinado por la Dra. Gloria Corpas Pastor (Universidad de Málaga) y "App para entrenar en posedición de traducción automática neuronal mediante la gamificación en entornos profesionales (GAMETRAPP) (TED2021-129789B-I00), coordinado por la Dra. Cristina Toledo Baez (Universidad de Málaga).

Derechos de uso de imagen

No se aplica.

Aprobación de comité de ética en investigación

No se aplica.

Conflicto de intereses

No se aplica.

Declaración de disponibilidad de datos de investigación

Los datos de esta investigación, que no están expresados en este trabajo, podrán ser proporcionados por el/los autor(es) bajo solicitud.

Licencia de uso

Publisher
Cadernos de Tradução es una publicación del Programa de Posgrado en Estudios de Traducción de la Universidad Federal de Santa Catarina. La revista *Cadernos de Tradução* está alojada en el Portal de Periódicos UFSC. Las ideas expresadas en este artículo son responsabilidad de sus autores y no representan necesariamente la opinión del equipo editorial o de la universidad.

Editores del número especial
Andréia Guerini – Fernando Ferreira Alves – Orlando Grossegesse

Editor de sección
Willian Moura

Corrección de normas
Alice S. Rezende – Ingrid Bignardi – João G. P. Silveira – Kamila Oliveira

Histórico
Recibido el: 29-05-2023
Aprobado el: 17-02-2024
Revisado el: 27-02-2024
Publicación: 04-2024

María Teresa Ortego Antón

Metodología para el diseño de un asistente semiautomático de redacción y de traducción de fichas descriptivas de embutidos del español al inglés
Methodology to develop a writing and translating aid tool to transfer driedmeat product cards from Spanish into English

Cadernos de Tradução
vol. 44, núm. 1, Esp. e94647, 2024
Universidade Federal de Santa Catarina,
ISSN: 1414-526X
ISSN-E: 2175-7968

DOI: https://doi.org/10.5007/2175-7968.2024.e94647