



Universidad de Valladolid

FACULTAD DE TRADUCCIÓN E INTERPRETACIÓN

Grado en Traducción e Interpretación

TRABAJO FIN DE GRADO

**Diseño y elaboración de una base de datos
terminológica bilingüe (español-inglés) sobre
embutidos a partir del análisis del corpus paralelo
P-GEFEM**

Presentado por Claudia Ioana Macrea

Tutelado por la Dra. M.^a Teresa Ortego Antón

Soria, junio de 2018

RESUMEN

En el presente Trabajo Fin de Grado pretendemos realizar un estudio de la terminología empleada en el sector agroalimentario, en el campo de los derivados cárnicos y, en concreto, de los embutidos, a partir del análisis de los datos del corpus paralelo en español e inglés P-GEFEM. Este estudio es fundamental, dado que este sector requiere cada vez más de los servicios de traducción y que aún no existe una uniformidad entre las opciones de traducción al inglés de cada término especializado. Con los datos que se desprendan del análisis de P-GEFEM elaboraremos una base de datos terminológica bilingüe (ES/EN) sobre embutidos que facilite la labor traductológica de los traductores de este campo del saber.

PALABRAS CLAVE

Embutidos, base de datos terminológica, corpus, inglés, español.

ABSTRACT

This final project aims to carry out a study of the terminology used in the agri-food sector, in the field of meat products and, particularly, dried meats, based on the analysis of data from the Spanish-English parallel corpus P-GEFEM. This study is crucial, as this sector increasingly requires translation services and there is a lack of univocity for translation equivalents of specialized terms. Based on the results of the analysis of the parallel corpus P-GEFEM, a bilingual Spanish-English terminology database on chorizo sausages will be compiled to aid translators during the translation workflow in this field of knowledge.

KEYWORDS

Dried meats, terminology database, corpus, English, Spanish.

ÍNDICE DE CONTENIDO

1. INTRODUCCIÓN	6
1.1. Justificación	6
1.2. Competencias	7
1.3. Objetivos	9
2. EL USO DE CORPUS COMO HERRAMIENTA DE TRADUCCIÓN	10
3. METODOLOGÍA DE TRABAJO	13
3.1. Diseño y compilación del corpus paralelo bilingüe (ES-EN).....	13
3.1.1. Búsqueda y localización de textos paralelos	15
3.1.2. Descarga de la información.....	15
3.1.3. Normalización o codificación	16
3.1.4. Almacenamiento.....	17
3.2. Metodología de análisis	19
3.2.1. Frecuencias léxicas	19
3.2.2. Colocaciones de las unidades léxicas	21
3.2.3. Metodología de elaboración de la base de datos	24
4. ANÁLISIS Y RESULTADOS	28
4.1. Frecuencias léxicas	28
4.2. Colocaciones	29
4.3. Base de datos terminológica	30
5. CONCLUSIONES	32
6. BIBLIOGRAFÍA	33
7. ANEXO	34

ÍNDICE DE TABLAS

Tabla 1. Competencias generales del Grado en Traducción e Interpretación	7
Tabla 2. Competencias específicas del Grado en Traducción e Interpretación.....	8

Tabla 3. Clasificación de los corpus lingüísticos.....	11
Tabla 4. Equivalentes en inglés de los 10 términos más recurrentes.....	28

ÍNDICE DE FIGURAS

Figura 1. Espacio de trabajo en Google Drive.....	13
Figura 2. Gestión de páginas web.....	15
Figura 3. Normalización de los TO.....	17
Figura 4. Normalización de los TM.....	17
Figura 5. Gestión de los textos recopilados.....	18
Figura 6. Frecuencias léxicas ES/EN.....	19
Figura 7. Búsqueda de frecuencias léxicas (ParaConc 269).	20
Figura 8. Equivalencias en inglés de las unidades léxicas más frecuentes.	21
Figura 9. Colocaciones de las unidades léxicas más frecuentes y sus equivalentes en inglés.	22
Figura 10. Búsqueda general de colocaciones del término «chorizo».....	23
Figura 11. Búsqueda avanzada de colocaciones en español del término «chorizo».....	23
Figura 12. Búsqueda de términos y equivalentes.	24
Figura 13. Definición de base de datos terminológica.....	26
Figura 14. Análisis del corpus en español con AntConc.	27
Figura 15. Ejemplo de entrada terminológica (entrada 1: chorizo).	34
Figura 16. Ejemplo de referencia cruzada (término <i>sausage</i>).....	35

ÍNDICE DE ABREVIATURAS

- AT: *advertising texts*
- DM: *dried meats*
- DP: descripción de producto
- EN: inglés
- ES: español
- ID: identificación del documento

- LM: lengua meta
- LO: lengua origen
- PD: *product description*
- TAO: traducción asistida por ordenador
- TEN: traducciones en inglés
- TM: texto meta
- TO: texto origen
- TP: texto promocional
- URL: Uniform Resource Locator

1. INTRODUCCIÓN

El presente Trabajo Fin de Grado (TFG) se encuadra dentro de la tipología investigadora y es producto de los resultados que se han obtenido en el proyecto de investigación titulado «Creación de una base terminológica bilingüe (EN/ES) especializada en embutidos» que hemos llevado a cabo durante el curso académico 2017-2018 en el Departamento de Lengua Española a través de una Beca de Colaboración de Estudiantes en Departamentos Universitarios del Ministerio de Educación, Cultura y Deporte, tutorizada por la Dra. M.^a Teresa Ortego Antón.

Asimismo, este trabajo se ha realizado en el seno de la red temática TRAJUTEC: TRAducción JURídica y TECnología de la Universidad de Málaga, coordinada por la Dra. Miriam Seghiri Domínguez.

1.1. Justificación

La industria agroalimentaria es uno de los principales sectores económicos en España y «la opción más prometedora a medio y largo plazo en amplias regiones de tradición agrícola y ganadera que, debido al éxodo poblacional durante décadas, se encuentran en el momento actual en riesgo de una alarmante despoblación» (Ortego Antón y Fernández Nistal, 2018). Además, ha experimentado un gran crecimiento a nivel internacional en los últimos años, por lo que las empresas de este sector requieren cada vez más de los servicios de traducción del español al inglés para poder dar a conocer sus productos en otros países.

No obstante, la labor traductológica en este campo no está desarrollada en su totalidad, dado que aún no existe una uniformidad entre las opciones de traducción al inglés de cada término especializado. De modo que consideramos necesario el desarrollo de herramientas lingüísticas que nos permitan acceder a los equivalentes de estos términos especializados y automatizar el proceso de traducción. Por tanto, hemos decidido realizar un estudio sobre el comportamiento morfosintáctico y fraseológico del léxico de este campo tanto en español como en inglés a partir de una metodología de análisis basada en corpus, cuyo producto resultante se corresponde con una base de datos terminológica que se pueda utilizar integrada en herramientas TAO para facilitar la labor del traductor en este campo del saber.

Por último, otra de las razones por las que hemos decidido realizar este trabajo de investigación es la necesidad de profundizar en los conocimientos relativos al estudio terminológico, a la traducción especializada y a la traducción asistida por ordenador, adquiridos durante el Grado de Traducción e Interpretación.

1.2. Competencias

En este trabajo se pretenden conseguir las siguientes competencias que se corresponden con las mencionadas en la guía docente de la asignatura Trabajo fin de Grado (TFG):

COMPETENCIAS GENERALES	
G1	Que los estudiantes hayan demostrado poseer y comprender conocimientos en el área de estudio (Traducción e Interpretación) que parte de la base de la educación secundaria general, y se suele encontrar a un nivel que, si bien se apoya en libros de texto avanzados, incluye también algunos aspectos que implican conocimientos procedentes de la vanguardia de su campo de estudio.
G2	Que los estudiantes sepan aplicar sus conocimientos a su trabajo o vocación de una forma profesional y posean las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de su área de estudio –Traducción e Interpretación.
G3	Que los estudiantes tengan la capacidad de reunir e interpretar datos relevantes (normalmente dentro de su área de estudio) para emitir juicios que incluyan una reflexión sobre temas esenciales de índole social, científica o ética.
G4	Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado.
G5	Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía.

Tabla 1. Competencias generales del Grado en Traducción e Interpretación.

COMPETENCIAS ESPECÍFICAS	
E8	Conocer y gestionar las fuentes y los recursos de información y documentación en Lengua A/B/C necesarios para el ejercicio de la traducción general/especializada B/C.
E14	Conocer la teoría de la ciencia de la documentación y su aplicación en los procesos de traducción.
E16	Manejar las últimas tecnologías documentales aplicadas a la traducción: sistemas de gestión y recuperación de información electrónica.

E17	Dominar conceptos básicos sobre el funcionamiento de herramientas informáticas que faciliten su utilización y su integración en la labor del traductor.
E18	Utilizar las herramientas informáticas básicas como instrumento específico de ayuda a la traducción en las diferentes fases del proceso traductológico.
E19	Desarrollar un método de trabajo organizado y optimizado gracias al empleo de herramientas informáticas.
E26	Conocer el valor de las aportaciones tecnológicas en la labor de traducción.
E27	Aplicar conocimientos sobre herramientas informáticas de tratamiento de textos a proyectos de traducción general/especializada.
E34	Aplicar a proyectos de traducción general/especializada las herramientas profesionales de traducción asistida por ordenador (TAO) y de gestión terminológica.
E35	Conocer el trabajo terminológico en cada una de sus fases y aplicarlo a su labor traductora.
E36	Adquirir conocimientos básicos de terminología.
E37	Relacionar los conocimientos de terminología con otras disciplinas lingüísticas y no lingüísticas.
E38	Extraer información conceptual de textos especializados y representarla gráficamente.
E39	Dominar las normas terminológicas, los tipos y organizaciones de normalización y estandarización.
E40	Diseñar y elaborar diccionarios y bases de datos terminológicas.
E41	Adoptar una postura crítica a la hora de aceptar y/o rechazar calcos y préstamos terminológicos, especialmente neológicos.
E46	Mostrar curiosidad hacia la mediación lingüística, desde un punto de vista científico y profesional.
E47	Mostrar habilidades de gestión y de evaluación de la calidad de la información recabada y que servirá de sustento empírico de un proyecto de investigación.
E49	Desarrollar la capacidad de aplicar los conocimientos y competencias adquiridos durante el grado sobre algún aspecto de la mediación lingüística a la práctica y a la investigación.
E50	Conocer las habilidades y métodos generales y específicos de investigación y aplicarlos a proyectos concretos del área de la Traducción e Interpretación y de las Humanidades en general.
E51	Conocer los fundamentos interdisciplinares que servirán de marco teórico para el trabajo de fin de grado.
E52	Asegurar la calidad del trabajo en el marco de unos plazos establecidos.
E72	Saber distinguir los principales recipientes del conocimiento científico (hipótesis, leyes, principios, teorías) y los principales métodos (observación, experimentación, contrastación...).

Tabla 2. Competencias específicas del Grado en Traducción e Interpretación.

1.3. Objetivos

El objetivo principal de este trabajo es profundizar en el estudio de los términos en español que se emplean en la industria agroalimentaria para hacer referencia a los productos cárnicos y de sus correspondientes equivalentes al inglés.

Asimismo, pretendemos alcanzar los siguientes objetivos específicos:

- Abordar la Lingüística de Corpus desde la perspectiva de los Estudios de Traducción.
- Profundizar en el campo de las herramientas de gestión terminológica.
- Analizar y extraer los resultados obtenidos a partir del corpus paralelo bilingüe (ES/EN) P-GEFEM.
- Elaborar una base de datos terminológica bilingüe (ES/EN) sobre embutidos a partir de los datos extraídos del corpus paralelo bilingüe (ES/EN) P-GEFEM.

2. EL USO DE CORPUS COMO HERRAMIENTA DE TRADUCCIÓN

La Lingüística de Corpus es la disciplina que permite realizar un estudio o análisis de la lengua no solo en el plano teórico, sino también en contexto. Podríamos definirla según McEnery y Hardie (2012: 1) como «*an area which focuses upon a set of procedures, or methods, for studying languages [...]*» o, según Corpas Pastor (2008: 49), como «una metodología de análisis que se puede aplicar a cualquier disciplina lingüística [...]». Como podemos observar, ambas definiciones abordan la Lingüística de Corpus como una metodología, dado que esta no se centra en la explicación teórica de la lengua, sino en analizar su uso en diferentes contextos a través de «*some set of machine-readable texts which is deemed an appropriate basis on which to study a specific set of research questions*» (McEnery y Hardie, 2012: 1) o, lo que es lo mismo, los corpus lingüísticos. Estos últimos son una de las herramientas más empleadas en el ámbito de la Traducción, ya que nos permiten analizar o estudiar el uso de las lenguas en contexto, la estructura de los textos y las traducciones de forma contrastiva según cuál sea la finalidad del estudio. Por tanto, es comprensible que la Lingüística de Corpus, gracias a sus características, sea una de las disciplinas que conforman la Traducción.

Los corpus lingüísticos podrían definirse como «*a collection of texts in electronic format which are processed and analyzed using software specially created for linguistic research*» (Zanettin, 2012: 7). Efectivamente, como se afirma en esta definición, se trata de una recopilación de textos electrónicos que poseen unas características determinadas y que pueden procesarse o analizarse mediante programas específicos, que nos permiten, a su vez, observar, por ejemplo, las frecuencias y colocaciones léxicas y los equivalentes o traducciones de los distintos términos en la lengua de destino según el contexto.

Asimismo, de acuerdo con Zanettin (2012: 11), los corpus lingüísticos podrían clasificarse de la siguiente manera:

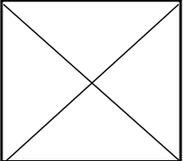
Comparable, monolingual Originals + Translations Language A + Language A	Comparable, bilingual Originals + Originals Language A + Language B	Parallel, bilingual Originals + Translations Language A + Language B
Reciprocal (bilingual, bidirectional, parallel) <div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;"> Originals Language A Translations Language A </div> <div style="text-align: center;">  </div> <div style="text-align: center;"> Translations Language B Originals Language B </div> </div>		

Tabla 3. Clasificación de los corpus lingüísticos.

Por tanto, según esta tabla, tenemos dos tipos de corpus principales: los corpus comparables y paralelos. Los corpus comparables pueden ser monolingües (incluyen textos originales y traducciones en una sola lengua), bilingües (incluyen textos originales en dos lenguas) y multilingües (incluyen textos originales en más de dos lenguas). A su vez, los corpus paralelos también pueden ser bilingües (incluyen textos originales y sus respectivas traducciones en la LM) y multilingües (incluyen textos originales y sus traducciones en las distintas LM). Además, estos corpus se pueden compilar en formato papel o electrónico. Lo más común y lógico en la actualidad es que se compilen en formato electrónico, dado que este formato permite que los textos sean analizados de manera rápida y eficaz por medio de herramientas informáticas como es el caso de los analizadores de concordancias o de los programas de alineación.

Por último, como ya hemos mencionado anteriormente, los corpus son herramientas muy importantes en Traducción, dado que poseen muchas ventajas y que ayudan al traductor a que lleve a cabo su labor traductológica de manera eficaz. Según Sánchez Ramos (2017: 140-141), algunas de las ventajas más significativas del uso de corpus en Traducción serían las siguientes:

- Facilitan la especialización del traductor.
- Permiten resolver los problemas lingüísticos que puedan surgir durante el proceso traductológico, ya que podemos observar la fraseología, la terminología, los

términos más frecuentes y la estructura del texto que emplea una lengua en un contexto determinado.

- Un acceso rápido cuando surja algún problema traductológico, en el caso de que se hayan compilado en formato electrónico.

Asimismo, permiten elaborar bases de datos terminológicas y glosarios (monolingües, bilingües, trilingües, etc.), recursos muy útiles para traducir.

3. METODOLOGÍA DE TRABAJO

Para elaborar una base de datos terminológica bilingüe (ES-EN) sobre embutidos, en concreto, sobre el chorizo, hemos tenido que establecer unas fases y el orden en el que se realizarían dichas fases para poder trabajar de manera ordenada y eficaz. Por tanto, decidimos crear un espacio de trabajo en Google Drive para que todos los componentes del proyecto pudieran acceder a la información y a los documentos del mismo (Figura 1).

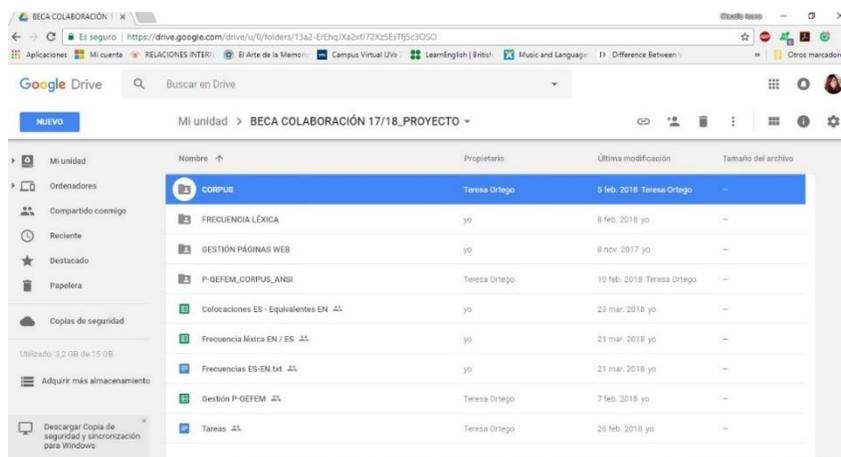


Figura 1. Espacio de trabajo en Google Drive.

Después, elaboramos una lista de tareas que se explicarán con más detalle a continuación y que se podría resumir de la siguiente manera:

- Diseño y compilación de un corpus paralelo (ES-EN).
- Análisis del corpus.
- Elaboración de la base de datos lingüística bilingüe (ES-EN).
- Elaboración del glosario bilingüe (ES-EN).

3.1. Diseño y compilación del corpus paralelo bilingüe (ES-EN)

Para el diseño y compilación de un corpus lingüístico, en primer lugar, se deben establecer unos criterios de diseño que establezcan la finalidad de este y que permitan que sea un corpus de calidad y representativo. De modo que para compilar el corpus paralelo bilingüe (ES-EN) para el proyecto de investigación, decidimos basarnos en los criterios de diseño y el protocolo de compilación de corpus ofrecidos por Seghiri (2011 y 2017). Según Seghiri (2017: 125), los criterios de diseño de un corpus son el género, la autoría, la temática

y la antigüedad de los textos que lo componen. En nuestro caso, como ya hemos mencionado anteriormente, se han empleado estos mismos criterios.

Al tratarse, como ya se ha mencionado antes, de un proyecto de investigación que tenía por objeto elaborar una base de datos terminológica bilingüe especializada sobre embutidos, en concreto el chorizo, decidimos que lo mejor sería buscar textos para el corpus en las páginas web de empresas que se encargan de producir embutidos nacionales y de la comercialización de estos, no solo a nivel nacional, sino también internacional. Este factor es muy importante, ya que aquellas empresas que exportan sus productos a otros países tienen o suelen tener sus páginas web traducidas al inglés o a otras lenguas, lo que nos ha permitido poder elaborar un corpus paralelo. Además, se trata de fuentes de información a las que es muy fácil acceder y que son fáciles de recopilar. Tan solo necesitamos un ordenador y conexión a Internet.

Pese a que en este tipo de fuentes podemos encontrar distintos tipos de texto, nos decantamos por los textos de descripción de producto (DP), que contienen información sobre el producto, y los textos promocionales (TP), que dan publicidad al producto o a la empresa. Con respecto al último criterio, la antigüedad de los textos, no tuvimos que establecer un intervalo de tiempo exacto, dado que las fuentes eran páginas web de empresas que tienden a actualizarse. Por último, establecimos el tamaño del corpus: 100 textos en español o textos origen (TO) y sus correspondientes traducciones al inglés o textos meta (TM).

Una vez establecidos los criterios de diseño, el corpus tenía las siguientes características: género (DP y TP), autoría (empresas de producción y comercialización de embutidos, en particular de chorizo) y temática (embutidos).

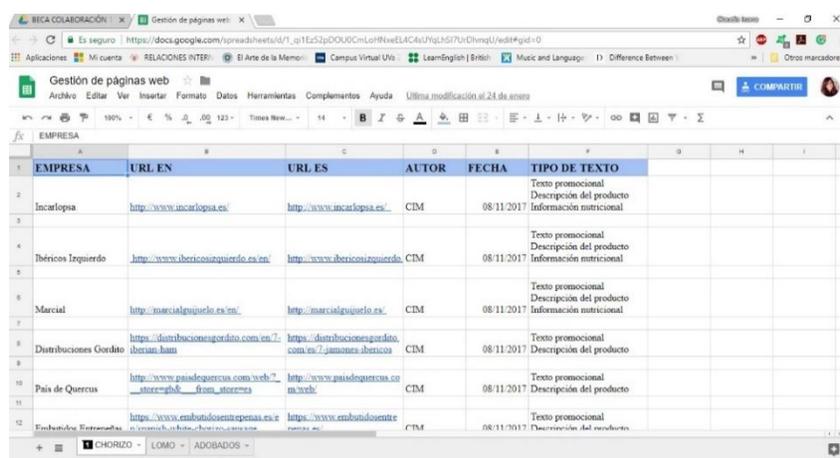
Después, procedimos a la compilación del corpus basándonos en el protocolo de compilación de Seghiri (2011 y 2017) que consta de cuatro fases generales:

- búsqueda y localización de los textos paralelos;
- descarga de la información;
- normalización o codificación;
- almacenamiento.

A continuación, vamos a explicar con más detalle en que consiste cada una de estas fases y cómo hemos llevado a cabo la labor de compilación de corpus en este caso concreto.

3.1.1. Búsqueda y localización de textos paralelos

En esta primera fase, realizamos una búsqueda exhaustiva de las páginas web de las diferentes empresas españolas relacionadas con los embutidos. A medida que realizábamos la búsqueda, teníamos que descartar aquellas páginas que no estuvieran traducidas al inglés. Ya desde el principio sabíamos que este iba a ser uno de nuestros mayores problemas a la hora de compilar el corpus paralelo bilingüe, ya que muchas de las páginas solo están disponibles en español y algunas de las que están traducidas al inglés no tienen textos paralelos en sí, sino que tienen traducido tan solo el nombre del producto y no la descripción de este. Por tanto, decidimos crear un documento de Excel, titulado *Gestión de páginas web*, para gestionar las páginas web de las empresas consultadas. En este documento incluimos el nombre de la empresa, la URL de la página en español y en inglés, el autor de la búsqueda, la fecha en la que se realizó dicha búsqueda y el tipo de texto (DP y TP) que se podría recopilar de cada una posteriormente (Figura 2).



EMPRESA	URL EN	URL ES	AUTOR	FECHA	TIPO DE TEXTO
Incarlopa	http://www.incarlopa.es/	http://www.incarlopa.es/	CIM	08/11/2017	Texto promocional Descripción del producto Información nutricional
Ibericos Izquierdo	http://www.ibericosizquierdo.es/	http://www.ibericosizquierdo.es/	CIM	08/11/2017	Texto promocional Descripción del producto Información nutricional
Marcial	http://marcialgispuelo.es/	http://marcialgispuelo.es/	CIM	08/11/2017	Texto promocional Descripción del producto Información nutricional
Distribuciones Gordito	https://distribucionesgordito.com/en?_ibercan_barr	https://distribucionesgordito.com/es?_jamones_ibericos	CIM	08/11/2017	Texto promocional Descripción del producto
País de Quercus	http://www.paisdequercus.com/web?store=gb&from_store=es	http://www.paisdequercus.com/web/	CIM	08/11/2017	Texto promocional Descripción del producto
Embutidos Entramates	https://www.embutidosentramates.es/en/panes-bollos-biscuits-caseros	https://www.embutidosentramates.es/es/	CIM	08/11/2017	Texto promocional Descripción del producto

Figura 2. Gestión de páginas web.

3.1.2. Descarga de la información

Tras buscar y gestionar las páginas web útiles para la elaboración del corpus, se procedió a la descarga de los textos paralelos. Este proceso se llevó a cabo de manera manual e individualizada, dado que, al tratarse de páginas web, era más fácil obtener los textos necesarios. Al final obtuvimos 100 TO con sus correspondientes TM en formato HTML (.html), PDF (.pdf) y de texto plano (.txt).

El siguiente paso consistió en la conversión de los textos en formato HTML (.html) y PDF (.pdf) a formato de texto plano (.txt), ya que los programas que se emplearían para analizar el corpus (ParaConc 269 y AntConc) reconocen este formato de texto. Para ello, utilizamos el conversor gratuito en línea Online-Convert.com¹. En cuanto a los textos que ya estaban en (.txt), hay que mencionar que se elaboraron de la siguiente manera: copiábamos la información que nos interesaba de la página web correspondiente y la pegábamos en el documento en formato (.txt).

Después de terminar esta tarea, se procedió a la revisión de los textos recopilados, es decir, se eliminó la información irrelevante o la publicidad en caso de que hubiera. Asimismo, estos textos planos fueron guardados en codificación UTF-8 en un primer momento, pero al dar error en la lectura con el programa ParaConc 269 (desaparecía la letra «ñ» y las tildes), tuvimos que cambiar la codificación de los textos a ANSI.

3.1.3. Normalización o codificación

En esta fase se nombraron los documentos en (.txt) para que se pudieran identificar fácilmente. En nuestro caso, hemos optado por la siguiente codificación en inglés compuesta por números y letras:

- **Número:** empezamos a enumerar los textos por el «101». El número nos indica la antigüedad de recopilación y la cantidad de textos recopilados. Esta regla se aplicó tanto a los TO como a los TM para así facilitar la identificación entre ellos.
- **DM (Dried Meats):** indica la temática del corpus, los embutidos.
- **ws (website):** indica el soporte del que se recopilan los textos.
- **Iniciales de la empresa:** así podemos saber a qué empresa pertenece la información recopilada.
- **AT (advertising text) o PD (product description):** nos indican la tipología textual del texto recopilado.
- **Fecha (aammdd):** indica la fecha de recopilación.
- **Foodie:** indica el tipo de corpus. Como se puede ver, indica que se trata de un corpus sobre alimentos.

¹ <https://www.online-convert.com/es>

- **ES o TEN:** «ES» indica que se trata de los TO, escritos en español, y «TEN» que son las traducciones en inglés de dichos textos.

En este proyecto, esta normalización o codificación se realizó de forma manual y el resultado final sería el siguiente (Figura 3 y 4):

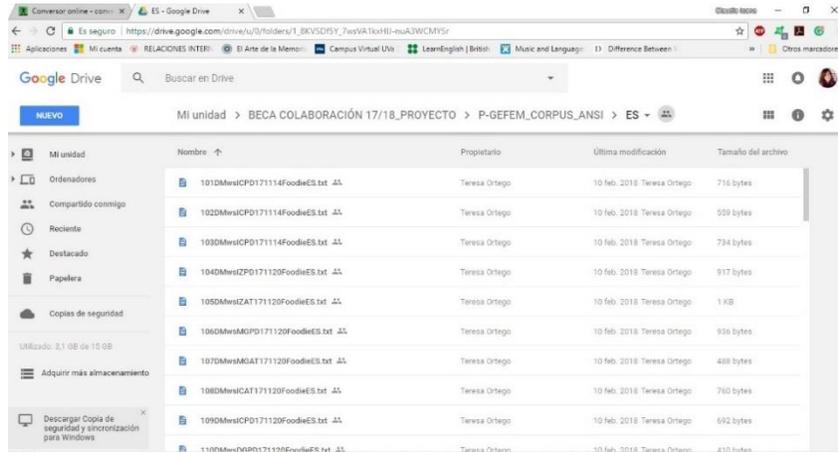


Figura 3. Normalización de los TO.

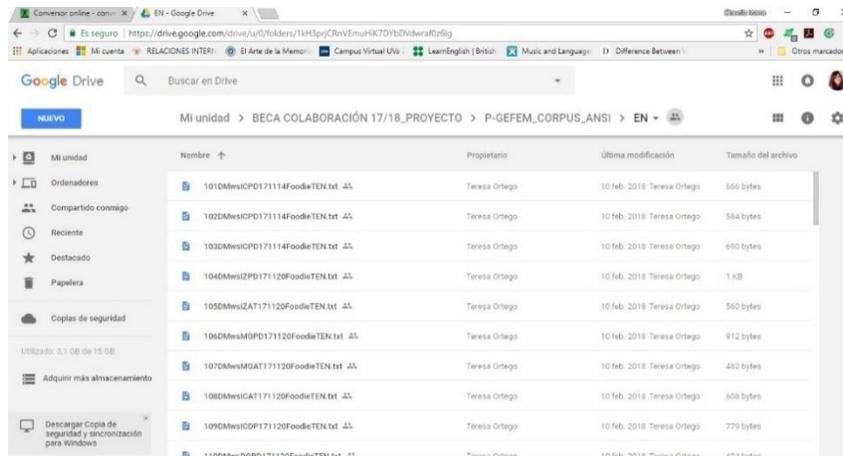


Figura 4. Normalización de los TM.

3.1.4. Almacenamiento

Una vez denominados los documentos, tuvimos que almacenarlos en la carpeta correspondiente dentro del corpus. Para ello, se siguió el criterio que también menciona Seghiri (2017). Se almacenaron los TO en una carpeta titulada «ES» y las traducciones en otra carpeta titulada «TEN». Estas dos carpetas serían las dos subcarpetas que formarían parte

de la carpeta denominada «P-GEFEM_CORPUS». Las dos subcarpetas indican las lenguas empleadas para la compilación del corpus.

Asimismo, debemos mencionar que según íbamos recopilando los textos paralelos y realizando las distintas fases del protocolo de compilación, completábamos al mismo tiempo un documento de registro de los textos recopilados en Excel denominado *Gestión P-GEFEM*.

Este documento consta de tres hojas: una en la que se mencionan los códigos de las empresas y en las otras dos el registro de los textos recopilados en cada idioma por separado. En cada una de estas dos hojas, se menciona la identificación del documento (ID), la URL para acceder al texto en línea, las siglas del nombre de la persona que realiza la búsqueda y recopilación, la empresa a la que pertenece la información, el nombre de la empresa principal en el caso de aquellos productos que son de una empresa distinta a la que aparecen publicados, la fecha de descarga o recopilación, el nombre del producto, el país y la denominación que recibe el producto por la empresa (Figura 5).

ID	URL	Researcher ID	Author	Company	Download date	Product	Country	Company denomination
101DMvsiCPD171114FoodieES	http://www.incarlopsa.es/produccion/chorizo-cular-ibero-extra	CIM		Incarlopsa	14/11/2017	chorizo	Spain	Chorizo Cular Ibérico Extra
102DMvsiCPD171114FoodieES	http://www.incarlopsa.es/produccion/chorizo-cular-ibero-extra	CIM		Incarlopsa	14/11/2017	chorizo	Spain	Chorizo Ibérico Extra
103DMvsiCPD171114FoodieES	http://www.incarlopsa.es/produccion/chorizo-cular-ibero-extra	CIM		Incarlopsa	14/11/2017	chorizo	Spain	MINI Chorizo Ibérico extra
104DMvsiZQPD171120FoodieES	http://www.ibericosizquierdo.es/comercio/iberico-cular-bellota-2/	CIM		Ibéricos Izquierdo	20/11/2017	chorizo	Spain	Chorizo cular Ibérico de bellota
105DMvsiZQAT171120FoodieES	http://www.ibericosizquierdo.es	CIM		Ibéricos Izquierdo	20/11/2017	chorizo	Spain	
106DMvsiMGPD171120FoodieES	http://marcialguijuelo.es/produccion/chorizo-iberico-bellota/	CIM		Marcial Guijuelo	20/11/2017	chorizo	Spain	Chorizo Ibérico de bellota Marcial
107DMvsiMGAT171120FoodieES	http://marcialguijuelo.es/	CIM		Marcial Guijuelo	20/11/2017	chorizo	Spain	
108DMvsiCAT171120FoodieES	http://www.incarlopsa.es/sistema-servicios/chorizo-sarta-ibero	CIM		Incarlopsa	20/11/2017	chorizo	Spain	
109DMvsiCPD171120FoodieES	http://www.incarlopsa.es/produccion/chorizo-sarta-ibero	CIM		Incarlopsa	20/11/2017	chorizo	Spain	Chorizo Sarta Ibérico

Figura 5. Gestión de los textos recopilados.

Gracias a este archivo, hemos podido evitar repetir los mismos textos o búsquedas en las páginas web de las mismas empresas.

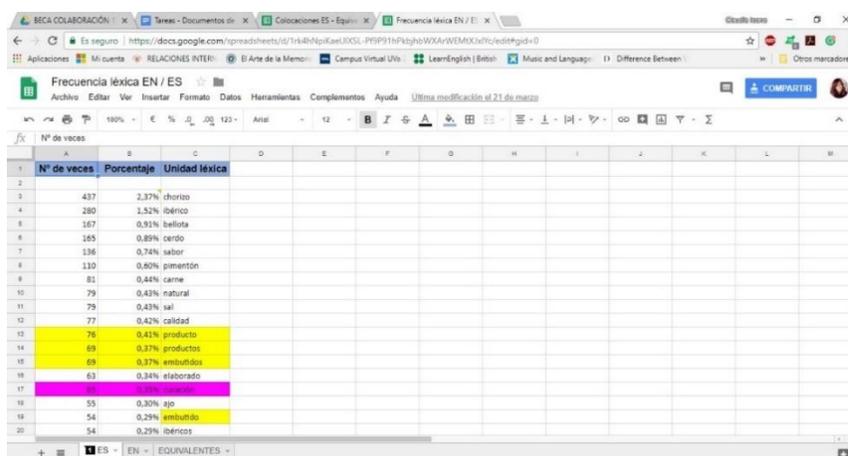
Después de terminar todo el proceso de compilación y gestión de corpus, obtuvimos un corpus paralelo bilingüe compuesto por 200 textos (100 TO y 100 TM) de los cuales 20 son textos promocionales y 80 descripciones de producto. Además, este corpus es un corpus virtual o *ad hoc*, ya que según la definición de Seghiri (2011: 16) es «aquel que se crea exclusivamente a partir de recursos electrónicos disponibles en la red Internet [...]».

3.2. Metodología de análisis

El siguiente paso de nuestra investigación era el análisis del corpus paralelo que habíamos compilado. Consistió en la búsqueda por frecuencia de los términos que más se repetían en los géneros textuales que habíamos recopilado, en la búsqueda de colocaciones léxicas, tanto en LO como en LM, y en la búsqueda de los términos equivalentes en la LM. Para llevar a cabo esta tarea empleamos el programa informático ParaConc 269 que permite analizar los corpus paralelos.

3.2.1. Frecuencias léxicas

Primero, creamos en el espacio de trabajo un documento de Excel, al que titulamos *Frecuencia léxica ES/EN*, en el que incluimos las frecuencias léxicas de los principales términos en ambas lenguas en hojas separadas. En cada hoja aparecería la unidad léxica, el número de veces que se repite y el porcentaje de repetición (Figura 6).



Nº de veces	Porcentaje	Unidad léxica
437	2,37%	chorizo
280	1,52%	ibérico
167	0,91%	bellota
165	0,89%	cerdo
136	0,74%	sabor
110	0,60%	pimentón
81	0,44%	carne
79	0,43%	natural
79	0,43%	sal
77	0,42%	calidad
76	0,41%	producto
69	0,37%	productos
69	0,37%	embutidos
63	0,34%	elaborado
55	0,30%	apio
54	0,29%	embutido
54	0,29%	ibéricos

Figura 6. Frecuencias léxicas ES/EN.

Para obtener dichas frecuencias, cargamos los textos en el programa ParaConc 269 y realizamos la búsqueda de las frecuencias léxicas mediante la opción *Frequency Order*, a la que podemos acceder haciendo clic en *Frequency* (Figura 7) y descargamos la búsqueda en formato (.txt) para poder manipular fácilmente la información.

Spanish (Modern Sort)			English (United States)		
Count	Plat	Word	Count	Plat	Word
1606	0.2022%	in	891	5.1062%	the
707	3.8396%	y	629	3.0102%	and
609	2.9472%	en	496	2.3761%	of
468	2.5418%	la	408	2.4481%	choizo
437	2.3729%	chorizo	395	2.1361%	in
384	2.0665%	el	328	1.6031%	a
289	1.5683%	con	303	1.5181%	it
280	1.5207%	ibérico	272	1.3214%	baian
252	1.3837%	que	252	1.3121%	with
250	1.3785%	a	244	1.2641%	to
231	1.2548%	los	189	1.1346%	from
230	1.2492%	un	164	0.8405%	if
207	1.1242%	se	159	0.8246%	our
167	0.9070%	bellota	140	0.8400%	post
165	0.8925%	cerdo	126	0.7560%	meal
163	0.8852%	del	123	0.7200%	usage
163	0.8852%	su	121	0.7200%	natural
148	0.8038%	es	115	0.6900%	spanish
141	0.7668%	el	112	0.6720%	coast
138	0.7580%	per	109	0.6460%	g
136	0.7386%	sabor	105	0.6300%	for
134	0.7278%	carne	95	0.5700%	this
131	0.7110%	los	80	0.5500%	made
126	0.6842%	una	50	0.3500%	paglika
110	0.5974%	pimentón	85	0.5100%	are
102	0.5540%	más	84	0.5040%	product
101	0.5486%	g	84	0.5040%	products
81	0.4390%	carne	84	0.5040%	we
79	0.4291%	natural	80	0.4800%	is
78	0.4231%	el	80	0.4800%	de
77	0.4182%	calidad	77	0.4620%	salt
76	0.4128%	producto	77	0.4620%	that
69	0.3740%	embalaje	75	0.4500%	or
69	0.3740%	productos	70	0.4200%	quality
67	0.3630%	...	67	0.4000%	...

Figura 7. Búsqueda de frecuencias léxicas (ParaConc 269).

Una vez realizada la búsqueda, nos dispusimos a analizar las frecuencias. Tuvimos que descartar aquellas unidades léxicas que se corresponden con las preposiciones, los pronombres, los determinantes, etc., ya que lo que nos interesaba eran la frecuencia de aquellos términos relacionados con la temática de nuestra investigación. Tras realizar la «limpieza de frecuencias», incluimos las necesarias en el documento de Excel. Llevamos a cabo el mismo proceso para las frecuencias de ambas lenguas.

Cuando terminamos con las frecuencias, decidimos seleccionar las 10 unidades léxicas más frecuentes (*chorizo, ibérico, bellota, cerdo, sabor, pimentón, carne, natural, sal y calidad*) para poder llevar a cabo un estudio exhaustivo de las colocaciones léxicas.

No obstante, antes de proceder con la búsqueda de colocaciones léxicas, decidimos determinar cuáles serían los términos equivalentes en la LM de estas 10 unidades léxicas (Figura 8).

ES			EN		
Nº de veces	Porcentaje	Unidad léxica	Nº de veces	Porcentaje	Unidad léxica
437	2,37%	chorizo	400	2,45%	chorizo
280	1,52%	ibérico	272	1,63%	iberian
167	0,91%	bellota	64	0,38%	bellota
185	0,89%	cerdo	140	0,84%	pork
136	0,74%	sabor	47	0,28%	pigs
110	0,60%	pimentón	62	0,37%	sauvage
81	0,44%	carne	48	0,29%	flavour
79	0,43%	natural	92	0,55%	paprika
79	0,43%	sal	26	0,16%	pepper
77	0,42%	calidad	48	0,29%	meats
			120	0,73%	natural
			77	0,46%	salt
			70	0,42%	quality

Figura 8. Equivalencias en inglés de las unidades léxicas más frecuentes.

3.2.2. Colocaciones de las unidades léxicas

Una vez obtuvimos las frecuencias léxicas y los equivalentes de las mismas en la LM, nos dispusimos a realizar la búsqueda de las colocaciones de las unidades léxicas mencionadas en el apartado anterior. Sin embargo, antes de proceder con la búsqueda, creamos en el espacio de trabajo un documento de Excel que denominamos *Colocaciones ES – Equivalentes EN*. En este documento incluimos las colocaciones de cada una de las 10 unidades léxicas más frecuentes en español y sus equivalentes. Para llevar a cabo una labor organizada, decidimos dividir el documento en diez hojas, una hoja para las colocaciones de cada una de las unidades léxicas (*chorizo*, *ibérico*, *bellota*, *cerdo*, *sabor*, *pimentón*, *carne*, *natural*, *sal* y *calidad*). En cada hoja aparecen las colocaciones en español, el número de veces que se repiten, los equivalentes de las colocaciones en inglés y la frecuencia de las mismas (Figura 9).

	B	C	D
	ES	Frecuencia	EN
1			
2	chorizo	408	chorizo
3		122	sausage
4		28	Spanish Chorizo
5		1	Spanish sausage
6		28	chorizo sausage
7			
8	chorizo ibérico	50	Iberian Chorizo
9		11	Chorizo Ibérico
10		16	Iberian Sausage
11		1	Iberian spanish Chorizo
12		2	Iberian cured sausage
13			
14	chorizo ibérico bellota	2	Bellota Iberian Chorizo
15		1	Bellota Iberian Spanish Chorizo
16		2	Iberian Sausage Bellota Chorizo
17		9	Chorizo Iberico Bellota
18		3	acom-fed Iberian Sausage
19		2	Spanish Acom-fed Iberian Chorizo
20		1	Spanish acom chorizo
21			

Figura 9. Colocaciones de las unidades léxicas más frecuentes y sus equivalentes en inglés.

Para obtener las colocaciones, empleamos el programa ParaConc 269. Primero, cargamos los textos del corpus paralelo, tanto en inglés como en español, en el programa y después realizamos la búsqueda paralela del primer término en español (chorizo) de nuestra lista de frecuencias léxicas mediante la opción *Parallel Search*. A continuación, buscamos las colocaciones en español e inglés de este término mediante la opción *Collocate Frequency Data* → *All*, a la que podemos acceder haciendo clic en el botón *Frequency* que ya hemos mencionado en el apartado anterior. Una vez obtenida la búsqueda (Figura 10), decidimos imprimir el resultado para facilitarnos la búsqueda de las colocaciones. Llevamos a cabo el mismo proceso para todas las unidades léxicas más frecuentes que habíamos seleccionado.

Asimismo, después de hacer las búsquedas generales de estas colocaciones, realizamos búsquedas avanzadas de colocaciones (Figura 11) para cada término, dado que así es más fácil obtener las colocaciones de forma automática y es mucho más visual. Para realizar este tipo de búsqueda, tuvimos que ir a la opción *Frequency* de la barra de herramientas de ParaConc 269, hacer clic en la opción *Advanced Collocation* y después en *Spanish (Modern Sort)*, tras la cual se abre un cuadro en el que se deben escoger las características que tienen que tener las colocaciones y la cantidad de elementos o colocaciones que deseamos que nos muestre el programa. Una vez obtenidas las colocaciones de cada unidad léxica, guardábamos el archivo en formato (.txt) para que nos fuera más fácil manipular la información y poder acceder siempre que quisiéramos a la misma. Realizamos el mismo proceso para los textos en inglés, pero con la opción *English (United States)* y

llevamos a cabo este tipo de búsqueda para todos los términos elegidos, al igual que hicimos en la búsqueda general.

Spanish (Modern Sort) - [chorizo]				English (United States) - [chorizo]			
2 Left	1 Left	1 Right	2 Right	2 Left	1 Left	1 Right	2 Right
18 de	41 el	64 bécico	29 bellota	15 beñen	20 spanish	20 sausage	20 chorizo
9 bellota	10 un	62 de	26 de	13 spanish	25 the	16 spanish	17 a
7 chorizo	15 nuestro	11 cular	16 bécico	13 the	15 spanish	15 from	12 de
7 en	12 del	11 vela	13 bon	10 bellota	11 cured	13 bécico	10 bellota
4 elaboración	10 de	9 esta	8 chorizo	9 acorn-fed	8 sur	13 bécico	8 from
4 la	7 esta	5 ponillo	7 chorizo	8 chorizo	8 pork	13 la	8 the
4 medación	6 origen	6 para	6 a	7 cured	7 bellota	8 cular	8 sausage
4 picante	5 bécico	7 bécico	6 esta	7 bécico	7 beef	8 made	7 a
4 price	4 código	7 sala	6 urta	7 of	7 vela	8 made	7 a
3 arino	3 característico	5 inconfundible	5 cular	7 sausage	6 guarnido	6 beñen	6 esta
3 color	3 ma	4 guarnido	5 trocheado	6 acorn	6 ponillo	6 vela	5 bécico
3 codo	3 picante	4 heradusa	5 en	6 ponillo	6 acorn	5 t	5 made
	3 piñenta	3 barbacoa	4 para	6 pepper	4 code	4 an	4 international
	3 sala	3 del	4 pueblo	5 origin	4 ruy	4 esta	4 kg
		3 may	4 time	4 an	4 of	4 heradusa	4 pueblo
			4	4 g	4 partly	4 made	4 quality
			3 salsa	4 in	4 unappreciated	4 sala	4 copy
			3 papafol	4 maturation	3 barbacoa	4 spanish	4 to
			3 guatilo	4 presentation	3 fed	4 the	3 beñen
			3 sabroso	4 product	3 bécico	3	3 it
			3 típico	4 recommend	3 piece	3 chorizo	3 bon
				4 very	3 beer	3 del	3 sogno
				3 blind	3 spread	3 spanish	3 smeg

Figura 10. Búsqueda general de colocaciones del término «chorizo».

Count	Text
17 (0.403%)	[chorizo bécico de
25 (0.720%)	[chorizo bécico bellota
19 (0.470%)	[chorizo de bon
12 (2.740%)	[chorizo de bellota
9 (2.059%)	[chorizo cular bécico
9 (2.059%)	[chorizo vela ponillo
7 (1.607%)	[chorizo bécico en
7 (1.607%)	[chorizo bécico esta
6 (1.372%)	[chorizo bécico cular
6 (1.372%)	[chorizo para sala
5 (1.144%)	[chorizo del pueblo
5 (1.144%)	[chorizo bécico bellota
5 (1.144%)	[chorizo ponillo trocheado
5 (1.144%)	[chorizo ponillo se
5 (1.144%)	[chorizo muy sabroso
5 (1.144%)	[chorizo y el
4 (0.915%)	[chorizo queso típico
4 (0.915%)	[chorizo de lazo
4 (0.915%)	[chorizo guarnido a
4 (0.915%)	[chorizo bécico español
3 (0.688%)	[chorizo de guatilo
3 (0.688%)	[chorizo de bon
3 (0.688%)	[chorizo bécico chorizo
3 (0.688%)	[chorizo bécico elaborado
3 (0.688%)	[chorizo inconfundible origen

Figura 11. Búsqueda avanzada de colocaciones en español del término «chorizo».

Tras realizar la diferentes búsquedas necesarias, nos dispusimos a analizar las colocaciones que habíamos obtenido. Una vez analizadas, las incluimos en la hoja del documento de Excel correspondiente. Cuando terminamos, buscamos los equivalentes en inglés de todas las colocaciones obtenidas para cada término. Para esta tarea, también empleamos ParaConc 269, dado que al tener cargados todos los textos del corpus paralelo, podíamos acceder fácilmente a los equivalentes del término que estuviéramos buscando tan solo con clicar encima de cualquiera de las frases originales resultantes de la búsqueda y mirar el término en inglés que aparecía en la frase equivalente del TM (Figura 12).

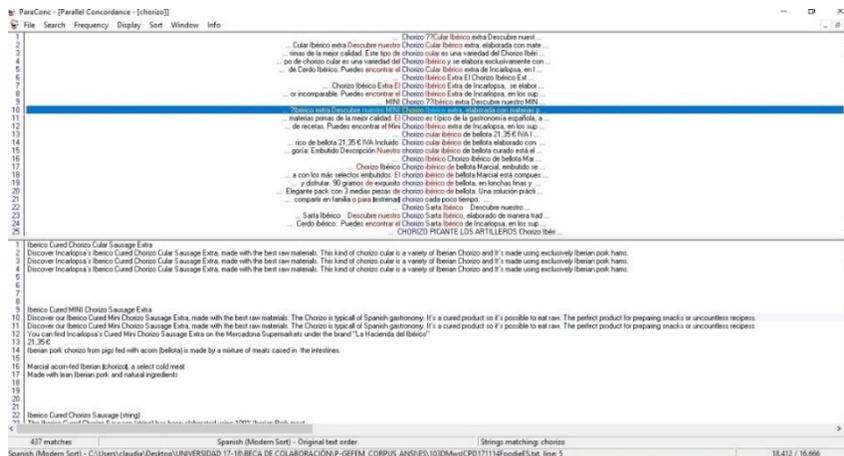


Figura 12. Búsqueda de términos y equivalentes.

Después de obtener los equivalentes en inglés, los incluimos en el documento de Excel *Colocaciones ES – Equivalentes EN*. En este documento también añadimos, como ya hemos mencionado, la frecuencia con la que aparecían los términos en nuestro corpus paralelo.

3.2.3. Metodología de elaboración de la base de datos

Una vez terminado el análisis del corpus paralelo bilingüe, comenzamos a elaborar la base de datos terminológica bilingüe basada en un archivo de definición de base de datos terminológica existente. Dicha definición de base de datos contiene los siguientes campos en inglés (Figura 13):

- Nivel de entrada:
 - **FIELD**: hace referencia al campo general del que forma parte el producto, que en este caso se corresponde con el campo *derivados cármicos*.
 - **SUBFIELD**: se trata del subcampo al que pertenecen los diferentes productos cármicos, que en este caso se corresponde con el subcampo *embutidos*.
 - **DATE**: se corresponde con la fecha de creación de la entrada terminológica.
 - **REV DATE**: se debe incluir la fecha de revisión de la entrada terminológica.

- **AUTHOR**: se corresponde con el nombre del autor de la entrada terminológica. En nuestro caso decidimos incluir las siglas del nombre del autor.
- **CLIENT**: se debe incluir el nombre del cliente al que va dirigida la base de datos. Para este campo también decidimos emplear las siglas del nombre del cliente.
- Nivel de término:
 - **TERM REF**: se debe incluir la referencia del texto en el que aparece el término de la entrada en cada uno de los idiomas. En nuestro caso, decidimos incluir el nombre del texto en el que aparece dicho término.
 - **GRAM**: hace referencia a la categoría gramatical del término (n: nombre, v: verbo, adj: adjetivo, etc.).
 - **ABB**: se tienen que incluir las abreviaturas del término, si las tiene.
 - **ABB REF**: se corresponde con la referencia o el nombre del texto en el que aparece dicha abreviatura.
 - **DEF**: se define el término.
 - **DEF REF**: se tiene que incluir la referencia o el nombre del texto en el que aparece dicha definición.
 - **CONT**: hace referencia al contexto en el que se utiliza el término. Por tanto, en este campo se tiene que incluir una oración en la que se emplee el término.
 - **CONT REF**: se corresponde con la referencia o el nombre del texto en el que aparece la frase con el término en contexto.
 - **EX**: se debe incluir un ejemplo de uso del término.
 - **EX REF**: se corresponde con la referencia o el nombre del texto en el que aparece el ejemplo de uso del término.
 - **DER**: se tienen que incluir los derivados del término, si los tiene.
 - **DER REF**: se corresponde con la referencia o el nombre del texto en el que aparecen los distintos derivados del término.
 - **VAR**: incluyen las variantes terminológicas del término, en caso de que las tenga.

- **VAR REF:** se corresponde con la referencia o el nombre del texto en el que aparecen las distintas variantes terminológicas del término.
- **OBS:** incluyen las observaciones relacionadas con el término que se consideren oportunas.

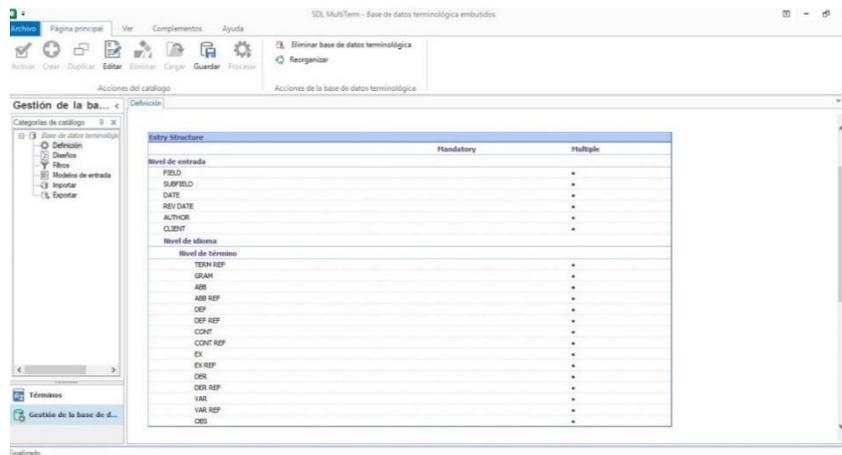


Figura 13. Definición de la base de datos terminológica.

Después de crear la base de datos terminológica basada en este archivo de definición de base de datos, procedimos a la creación de las entradas terminológicas (18 entradas en total). En cada entrada completamos los campos correspondientes, tanto a nivel de entrada como a nivel de término. En lo que respecta los campos a nivel de término, decidimos completar en todas las entradas los siguientes campos que consideramos obligatorios: TERM REF, GRAM, DEF, DEF REF, EX, EX REF, VAR (en caso de que el término tenga alguna variante) y VAR REF.

Asimismo, realizamos referencias cruzadas de aquellos términos que tienen alguna variante terminológica. Para poder llevar a cabo este proceso, tuvimos que crear una entrada para cada variante. De esta manera se puede observar claramente las diferentes formas que se utilizan para hacer referencia al mismo término en ambas lenguas y el uso de dichos términos en contexto.

Por último, para realizar la búsqueda de las definiciones y los ejemplos de uso de los términos, empleamos el programa AntConc, un analizador de concordancias gratuito que nos permite analizar de forma fácil e intuitiva el corpus (Figura 14). El único inconveniente que

tiene es que no permite cargar los textos que conforman el corpus en ambas lenguas. Por lo que tuvimos que buscar los datos necesarios para cada lengua por separado.

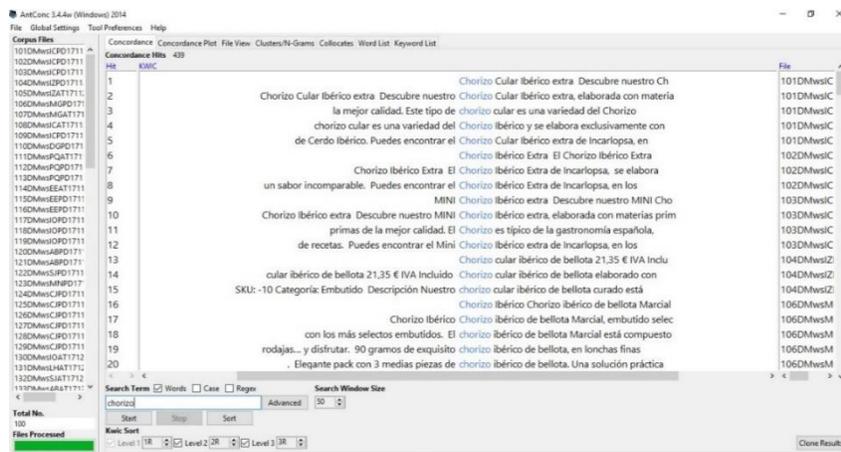


Figura 14. Análisis del corpus en español con AntConc.

4. ANÁLISIS Y RESULTADOS

Los resultados que se desprenden del análisis fueron muy llamativos, ya que en muchos de los casos había más de un equivalente en la LM para un mismo término.

4.1. Frecuencias léxicas

Después de incluir las frecuencias en el documento de Excel correspondiente, comenzamos a estudiarlas junto con sus equivalentes. Durante este proceso nos dimos cuenta de que, en algunos casos, en particular en los tres primeros términos que se pueden observar en la siguiente tabla, también se han mantenido los términos en español en la LM.

ES	EN
chorizo	<i>chorizo</i> <i>sausage</i>
ibérico	<i>iberian</i> <i>ibérico</i>
bellota	<i>bellota</i> <i>acorn</i>
cerdo	<i>pork</i> <i>pig</i>
sabor	<i>taste</i> <i>flavour</i>
pimentón	<i>paprika</i> <i>pepper</i>
carne	<i>meat</i>
natural	<i>natural</i>
sal	<i>salt</i>
calidad	<i>quality</i>

Tabla 4. Equivalentes en inglés de los 10 términos más recurrentes.

Esta característica es bastante habitual en este ámbito, sobre todo con respecto a estos términos u otros que estén relacionados con el nombre de los diferentes productos que son considerados productos *gourmet*. Esta última, es una de las características más importantes por las que en la mayoría de los casos no se traduce el nombre del producto, como es el caso del *chorizo*, por ejemplo. Trataremos este aspecto detalladamente más adelante.

4.2. Colocaciones

En lo que respecta las colocaciones de las unidades léxicas más frecuentes, hemos observado que también se da el fenómeno mencionado en el apartado anterior. En la gran mayoría de los casos, la colocación no se traduce al inglés, sino que se prefiere dejar el término en español. Un ejemplo podría ser el término *chorizo ibérico*, que, en los textos que conforman nuestro corpus paralelo en inglés, tiene los siguientes equivalentes: *Iberian Chorizo*, *Chorizo Ibérico*, *Iberian Sausage*, *Iberian sausage*, *Iberian spanish Chorizo* e *Iberian cured sausage*. Como se puede observar, cada palabra que compone la mayoría de estos términos está escrita con inicial mayúscula, respetando así el uso de las mayúsculas en inglés. No obstante en el caso de *Iberian spanish Chorizo*, podemos ver que existe un error, ya que el término *spanish* debería ir con inicial mayúscula, dado que en inglés las nacionalidades se escriben con inicial mayúscula. En este caso, podríamos decir que se trata de una errata o de un despiste del traductor encargado de traducir el texto en el que aparece este término. Por otro lado, tenemos otros términos como *Iberian sausage* o *Iberian cured sausage* que solo llevan inicial mayúscula en la primera palabra de la unidad léxica que lo componen. Por último, observamos el uso de un equivalente compuesto por una palabra en inglés y otra en español: *Iberian Chorizo*. Este fenómeno podría deberse a que el término *chorizo* no tiene traducción en inglés y, por tanto, se opta por el uso del término en español o a que, al tratarse de un producto *gourmet* español, se conoce con su nombre original, por lo que no es necesaria su traducción.

Todas estas características de traducción que hemos mencionado, tienen lugar en la traducción de muchas de las colocaciones obtenidas.

Por último, nos gustaría destacar otro aspecto interesante de las colocaciones que hemos obtenido para el término *chorizo* en español. Con este aspecto, nos referimos a cómo aparecen colocados el término *chorizo* con los términos *ibérico* y *bellota*. Según nuestro análisis, hemos obtenido las siguientes colocaciones: *chorizo ibérico bellota*, *chorizo ibérico bellota cortado (lonchas)*, *chorizo de bellota*, *chorizo ibérico bellota sarta*, *chorizo de bellota ibérico*, *chorizo ibérico de bellota*, *chorizo cular ibérico de bellota* y *chorizo cular de bellota 100 % ibérico*. Como se puede observar, en muchos casos se omite la preposición «de» que debería haber delante del término *bellota*. Este fenómeno podría deberse sobre todo a un ahorro de espacio, ya que estos términos aparecen también en los embalajes y etiquetas de

estos productos. Asimismo, podría tratarse tan solo de una decisión de la propia empresa que produce estos productos.

4.3. Base de datos terminológica

Como ya hemos mencionado, la base de datos terminológica sobre embutidos que hemos elaborado para este proyecto de investigación consta de 18 entradas. Dichas entradas hacen referencia a los siguientes términos: *chorizo*, *sausage*, *Spanish chorizo*, *Spanish sausage*, *chorizo sausage*, *iberian chorizo*, *chorizo ibérico*, *iberian sausage*, *iberian cured sausage*, *iberian acorn chorizo*, *chorizo ibérico de bellota*, *bellota iberian chorizo*, *bellota iberian Spanish chorizo*, *iberian sausage bellota chorizo*, *chorizo iberico bellota*, *acorn-fed iberian sausage*, *Spanish acorn-fed iberian chorizo* y *Spanish acorn chorizo*. La mayoría de estos términos se corresponden a los equivalentes en inglés de un mismo término, de modo que las referencias cruzadas de los términos que componen la base de datos quedaría de la siguiente manera:

- El término *chorizo* con *sausage*, *Spanish chorizo*, *Spanish sausage*, *chorizo sausage*.
- El término *iberian chorizo* con *chorizo ibérico*, *iberian sausage*, *iberian acorn chorizo* y *iberian cured sausage*.
- El término *iberian acorn chorizo* con *chorizo ibérico de bellota*, *iberian chorizo*, *bellota iberian chorizo*, *bellota iberian Spanish chorizo*, *iberian sausage bellota chorizo*, *chorizo iberico bellota*, *acorn-fed iberian sausage*, *Spanish acorn-fed iberian chorizo* y *Spanish acorn chorizo*.

A continuación, vamos a tomar de ejemplo la primera entrada (entrada 1) que hemos elaborado en esta base de datos y el resto de entradas a las que está vinculada para realizar un pequeño análisis de las entradas resultantes de este proyecto. Esta entrada (Figura 14) es un gran ejemplo de las variantes terminológicas y de las referencias cruzadas, y se corresponde con el término *chorizo*, que en inglés, además de *chorizo*, se conoce como *sausage* (entrada 2), *Spanish chorizo* (entrada 3), *Spanish sausage* (entrada 4) y *chorizo sausage* (entrada 5). Como ya hemos mencionado, estas entradas están enlazadas entre sí, de manera que se puede acceder fácilmente a la entrada correspondiente a cada variante terminológica y así poder observar cómo se utiliza cada una en contexto. Es importante

mencionar que las entradas que se refieren a estas variantes solo contienen el ejemplo de uso de las mismas en la lengua que corresponda y no la definición (Figura 15, ver Anexo), ya que la definición se puede ver en la entrada de la variante terminológica principal, que en este caso es la del término *chorizo*. No obstante, hemos incluido en cada entrada relacionada con el término general la información de todos los campos que hemos completado para el término en español. Lo hemos considerado necesario, ya que la persona que acceda a esta base de datos terminológica puede consultar cualquier término sin tener en cuenta cuál es el término general, por lo que cada entrada necesita de la definición y del ejemplo de uso en español, no solo del término equivalente.

5. CONCLUSIONES

En el presente trabajo hemos podido observar que existe una gran necesidad de estudio de los términos especializados pertenecientes al sector agroalimentario, en concreto, del ámbito de los derivados cárnicos, así como de los equivalentes existentes en inglés de los mismos. Esto se debe a que en muchos de los casos no existen unos términos equivalentes fijados en inglés, sino que coexisten varios equivalentes para un mismo término en español, por lo que los traductores de los textos de este ámbito optan muchas veces por utilizar el término en español o realizar una traducción parcial o total del mismo.

Asimismo, algunas de las traducciones de estos términos al inglés no se componen en su totalidad de términos ingleses, sino de términos en inglés y en español, como es el caso de *bellota iberian chorizo*. Como ya se ha mencionado, esto podría deberse al hecho de que hacen referencia a productos *gourmet* españoles, por lo que se pretende traspasar esta característica también a la LM, o también porque así lo decide la empresa que produce y exporta dichos productos.

De modo que consideramos que se debería profundizar más en este ámbito y estudiarlo también a partir de un corpus comparable, ya que de esta manera podríamos observar cuáles son los términos que se emplean en los TO en español, así como en los TO en inglés y el empleo de dichos términos en un contexto. De esta manera, podríamos establecer y unificar los equivalentes en inglés de los términos relacionados con los embutidos y elaborar una base de datos terminológica o un glosario bilingüe con los mismos, para así facilitar la labor del traductor de este ámbito.

6. BIBLIOGRAFÍA

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

Corpas Pastor, G. y Seghiri, M. (2010). Size matters: A quantitative approach to corpus representativeness. En Rabadán, R., Fernández López, M. y Guzmán González, T. (Coords.), *Lengua, traducción, recepción: en honor de Julio César Santoyo*. León: Universidad de León, 111-145. Recuperado de <https://bit.ly/2IUj5hT>

McEnery, T. y Hardie, A. (2012). *Corpus Linguistics*. Cambridge: University Press.

Ortego-Antón, M. T. y Fernández-Nistal, P. (2018). Estudio contrastivo de la terminología de embutidos en inglés y en español con ParaConc y tICorpus a partir del corpus paralelo P-GEFEM y del comparable C-GEFEM. En Seghiri, M., *El uso de corpus lingüísticos como herramienta pedagógica para la enseñanza-aprendizaje de lenguas, traducción e interpretación: marco teórico, metodología y aplicaciones*. Berna: Peter Lang.

Sánchez Ramos, M. (2017). Metodología de corpus y formación en la traducción especializada (inglés-español): una propuesta para la mejora de la adquisición de vocabulario especializado. *Revista de Lingüística y Lenguas Aplicadas*, 12, 137-150. <https://doi.org/10.4995/rlyla>

Seghiri, M. (2011). Metodología protocolizada de compilación de un corpus de seguros de viaje: aspectos de diseño y representatividad. *RLA: Revista de lingüística teórica y aplicada*, 49(2), 13-30. Recuperado de <https://bit.ly/2qus5nm>

Seghiri, M. (2017). Corpus e interpretación biosanitaria: extracción terminológica basada en bitextos del campo de la Neurología para la fase documental del intérprete. *Panace@: Revista de Medicina, Lenguaje y Traducción*, 18(46), 123-132. Recuperado de <https://bit.ly/2JJxy1y>

Zanettin, F. (2012). *Translation-driven corpora*. Manchester: St. Jerome Publishing.

7. ANEXO

The screenshot displays the SDL MultiTerm interface for a terminological database. The main window shows the entry for 'chorizo' with the following details:

- Entry number:** 1
- FIELD:** Derivados cárnicos
- SUBFIELD:** Embutidos
- DATE:** 16/04/2018
- AUTHOR:** CIM
- CLIENT:** TOA

The entry is presented in two language versions:

- English:**
 - TERM REF:** 148DMwsJMAT171219FoodieTEN.bt
 - GRAM:** n
 - DEF:** [...] is made with chopped pork meat and pork fat, seasoned with paprika and garlic, all stuffed into natural gut.
 - DEF REF:** 148DMwsJMAT171219FoodieTEN.bt
 - EX:** The ingredients of the Joselito Chorizo are: lean pork, sea salt, paprika and sugar.
 - EX REF:** 185DMwsJSPD180124FoodieTEN.bt
 - VAR:** [sausage](#), [Spanish chorizo](#), [Spanish sausage](#), [chorizo sausage](#)
 - VAR REF:** 130DMwsCJAT171204FoodieTEN.bt, 155DMwsVJPD180122FoodieTEN.bt, 179DMwsNVAT180124FoodieTEN.bt, 148DMwsJMAT171219FoodieTEN.bt
- Spanish:**
 - TERM REF:** 168DMwsJPPD180123FoodieES.bt
 - GRAM:** n
 - DEF:** [...] es un embutido elaborado con magro y grasa de cerdo de bellota 100 % ibérico, picado y adobado que se ha sometido a un proceso de secado y maduración naturales.
 - DEF REF:** 168DMwsJPPD180123FoodieES.bt
 - EX:** Este tipo de chorizo cular es una variedad del Chorizo Ibérico y se elabora exclusivamente con jamones de Cerdo Ibérico.
 - EX REF:** 101DMwsICPD171114FoodieES.bt

The interface includes a search bar at the top with 'chorizo' entered, and a list of terms on the left side. The status bar at the bottom indicates 'Finalizado'.

Figura 15. Ejemplo de entrada terminológica (entrada 1: chorizo).

SDL MultiTerm - Base de datos terminológica embutidos

Archivo | Página principal | Ver | Complementos | Ayuda

English | Buscar | (Ningún filtro) | (Ningún modelo de e) | Flags layout

Invertir | Modo jerárquico | Búsqueda de términos | Modo de filtro forzado | Agregar nueva | Editar | Guardar | Eliminar | Agregar copia | Fusionar | Cancelar

Spanish | Búsqueda normal | Solicitar término de destino | Todo el formulario | 100 | Navegación

Idiomas | Buscar | Filtros | Edición | Modelo de entrada | Mostrar

Términos

Examinar

- acom-fed iberian sausage
- bellota iberian chorizo
- bellota iberian Spanish chor
- chorizo
- chorizo ibérico
- chorizo iberico bellota
- chorizo ibérico de bellota
- chorizo sausage
- iberian acom chorizo
- iberian chorizo
- iberian cured sausage
- iberian sausage
- iberian sausage bellota chor
- sausage**
- Spanish acom chorizo
- Spanish acom-fed iberian c
- Spanish chorizo
- Spanish sausage

Examinar | Lista de resp...

Términos

Gestión de la base de d...

chorizo | sausage

Entry number: 2
 FIELD: Derivados cárnicos
 SUBFIELD: Embutidos
 DATE: 16/04/2018
 AUTHOR: CIM
 CLIENT: TOA

English

sausage

TERM REF: 130DMwsCJAT171204FoodieTEN.bt
 GRAM: n
 EX: The patatera is a sausage typical of northern Extremadura, in particular a sausage-although its appearance and flavor reminiscent of a sausage, which is in the form of a horseshoe and whose production is used the baked potato (50%), along with lean (10%) and fat (40%) of Iberian pig.
 EX REF: 148DMwsJMAT171219FoodieTEN.bt
 VAR: [chorizo](#), [Spanish chorizo](#), [Spanish sausage](#), [chorizo sausage](#)
 VAR REF: 148DMwsJMAT171219FoodieTEN.bt, 155DMwsVJPD180122FoodieTEN.bt, 179DMwsNVAT180124FoodieTEN.bt, 148DMwsJMAT171219FoodieTEN.bt

Spanish

chorizo

TERM REF: 168DMwsJPPD180123FoodieES.bt
 GRAM: n
 DEF: [...] es un embutido elaborado con magro y grasa de cerdo de bellota 100 % ibérico, picado y adobado que se ha sometido a un proceso de secado y maduración naturales.
 DEF REF: 168DMwsJPPD180123FoodieES.bt
 EX: Este tipo de chorizo cular es una variedad del Chorizo Ibérico y se elabora exclusivamente con jamones de Cerdo Ibérico.
 EX REF: 101DMwsICPD171114FoodieES.bt

CIM

Figura 16. Ejemplo de referencia cruzada (término sausage).