MONOGRÁFICO

Estudios de **Traducción**
ISSN: 2174-047X

EDICIONES COMPLUTENSE

https://dx.doi.org/10.5209/estr.80584

El trabajo terminográfico basado en corpus: el caso del recurso *DicoAdventure*

Isabel Durán Muñoz[1]

Recibido: 20 de febrero de 2022 / Aceptado: 24 de marzo de 2022

Resumen[2]**.** Los corpus textuales se han convertido en la actualidad en un recurso fundamental para el trabajo terminográfico, ya que permiten recabar información real, fiable y de calidad para estudiar cualquier campo de especialidad. Están presentes en todas las fases del trabajo, desde la extracción de terminología hasta la inclusión de información lingüística, semántica y pragmática en las entradas de los diccionarios. Por este motivo, el corpus especializado *Advencor* es la principal fuente de información de DicoAdventure, un proyecto de investigación que tiene como objetivo elaborar un recurso bilingüe (inglés-español) basado en la léxico-semántica sobre el turismo de aventura, el diccionario DicoAdventure. En este trabajo presentamos las dos primeras fases de la metodología, a saber: la compilación del corpus y la extracción y selección de unidades especializadas para ilustrar parte del trabajo terminográfico basado en corpus.
Palabras clave: corpus; turismo de aventura; trabajo terminográfico; léxico-semántica; DicoAdventure

[en] Corpus-Based Terminography: The Case of *DicoAdventure*

Abstract. Corpora have now become a fundamental resource for terminological work, since they facilitate a collection of real, reliable and quality information to study any specialised domain. They are present in all phases of the work, from the extraction of terminology to the inclusion of linguistic, semantic and pragmatic information in dictionary entries. For this reason, the *Advencor* specialised corpus is the main information source of DicoAdventure, a research project that aims to develop a bilingual resource (English-Spanish) based on the lexicon-semantics for the language of adventure tourism, the DicoAdventure dictionary. In this paper, we present the first two phases of the methodology, namely: corpus compilation and extraction and selection of specialised units to illustrate part of the corpus-based terminological work.
Keywords: corpus; adventure tourism; terminological work; lexico-semantics; DicoAdventure.

Cómo citar: Durán Muñoz, I. (2022). El trabajo terminográfico basado en corpus: el caso del recurso *DicoAdventure. Estudios de Traducción*, 12, 109-118.

1. Introducción

Los corpus textuales se consideran hoy herramientas imprescindibles para cualquier trabajo terminográfico, ya que han demostrado ser recursos muy productivos para extraer información lingüística que permite elaborar productos terminológicos de calidad. El corpus hace posible la consulta y el procesamiento de grandes cantidades de información en un tiempo muy reducido y, a menudo, de forma automática o semiautomática gracias a herramientas informáticas como TermoStat Web 3.0 o Sketch Engine[3]. Sin embargo, las ventajas que aporta el corpus en un trabajo terminográfico no se limitan a la consulta de información en la fase de elaboración de la terminología, sino que se trata de una herramienta que acompaña al terminógrafo en todas las fases de su trabajo, es decir, se utiliza

[1] Universidad de Córdoba
iduran@uco.es
https://orcid.org/0000-0002-6795-498X

[2] El presente trabajo se ha realizado en el seno del proyecto DicoAdventure: *Diseño y desarrollo de un recurso electrónico especializado bilingüe (Inglés, Español) sobre el turismo de aventura a partir de marcos semánticos* (Ref. UCO-1380857), cofinanciado por el Programa Operativo FEDER 2014-2020 y por la Consejería de Economía, Conocimiento, Empresas y Universidad de la Junta de Andalucía. Además, de forma parcial también por los proyectos VIP II (PID2020-112818GB-I00) e INMOCOR (Ref. P20-00109).

[3] *TermoStat Web 3.0* es un extractor de terminología de acceso gratuito desarrollado en la Universidad de Montreal (http://termostat.ling.umontreal.ca/). *Sketch Engine* es un gestor de corpus que incluye diferentes funcionalidades (extracción de terminología y listas de palabras frecuentes, análisis de estructuras gramaticales, colocaciones, etc.) de licencia comercial (https://www.sketchengine.eu/).

desde la fase inicial de preparación del proyecto hasta la fase final de validación. Así pues, la documentación y, en consecuencia, los corpus y las herramientas de análisis de corpus permiten llevar a cabo las siguientes tareas (Durán Muñoz 2011: 45):

- La adquisición de conocimiento conceptual y la familiarización del terminógrafo no experto del dominio mediante la consulta de documentos, a fin de identificar la estructura interna del dominio, las relaciones con otros campos de especialidad y las fuentes de conocimientos adecuadas.
- La identificación de unidades terminológicas dentro de un discurso de especialidad.
- El análisis y la preparación de entradas mediante la adquisición de información lingüística, pragmática y semántica extraídas del corpus, como por ejemplo definiciones, contextos o colocaciones.
- La detección y descripción de nuevos conceptos, así como la identificación de etiquetas léxicas que se están atribuyendo a dichos conceptos dentro del dominio. En algunos casos, también la propuesta de un neologismo adecuado (cuando todavía no se ha acuñado un término en una lengua).
- La estandarización de términos sinónimos o cuasisinónimos que son utilizados por diferentes expertos.
- Y la clarificación de dudas y preguntas sobre inconsistencias y otros asuntos que, de lo contrario, solo se podría llevar a cabo mediante consultas a expertos del dominio.

Para poder conseguir los beneficios que aportan los corpus textuales al trabajo terminográfico desde un punto de vista lingüístico, pragmático y semántico, es necesario que estos estén compilados de forma apropiada y según unos criterios generales y específicos coherentes y preestablecidos (Sinclair 1996, Meyer 2001; Durán Muñoz 2011), teniendo en cuenta que "the corpus needs to be as linguistically and conceptually rich as possible" (Meyer y Mackintosh 1996: 266). Por este motivo, la fase de compilación del corpus es uno de los puntos más relevantes de la metodología basada en corpus y es necesario prestar mucha atención a esta etapa del trabajo.

Teniendo en cuenta las ventajas que aportan los corpus textuales al trabajo terminográfico, el proyecto DicoAdventure (Ref. UCO-1380857), en el que se enmarca este artículo, considera el uso del corpus una herramienta esencial para llevar a cabo las diferentes tareas que conforman la metodología utilizada. En este contexto, el objetivo de este capítulo consiste en describir el corpus compilado en este proyecto, el corpus Advencor[4], así como presentar las fases de la metodología seguida en el proyecto, centrándose particularmente en la extracción y selección de las unidades especializadas, en concreto verbos de movimiento, que forman parte del recurso *DicoAdventure*.

Para ello, el trabajo se estructura de la siguiente manera: en primer lugar, se realiza una revisión de las principales contribuciones que el corpus ha aportado a la Terminografía; a continuación, se presenta el corpus Advencor en el seno del proyecto DicoAdventure y la metodología aplicada en dicho proyecto, para centrarse, en los siguientes apartados, en la compilación y en la gestión del corpus para el trabajo terminográfico. Finalmente, se aportan una serie de conclusiones sobre el trabajo presentado y el uso del corpus en Terminografía.

2. La Terminografía basada en corpus

La Lingüística de corpus ha sido una de las disciplinas lingüísticas que más ha influido en la Terminología y, por ende, en la Terminografía, es decir, en su rama aplicada, cuyo objetivo es la elaboración de recursos especializados. Tanto es así que hoy en día la idea de la contextualización de los términos a través de los corpus textuales está totalmente aceptada en la comunidad de terminógrafos, y los corpus son considerados recursos indispensables para cualquier trabajo de esta naturaleza, como mencionábamos anteriormente.

La Terminología moderna ha incorporado el empleo de los corpus textuales como fuente de información esencial en su metodología, ya que estos han permitido conseguir avances y mejoras en la calidad del producto final, sobre todo con relación a las posibilidades que ofrece este para la investigación de material terminológico real. En este contexto, en Terminografía, el uso de los corpus textuales se ve motivado por dos razones fundamentales (Durán Muñoz 2012: 79-80):

Por un lado, el trabajo terminográfico no consiste en la invención de denominaciones para unos conceptos previamente establecidos como propugnaban los estructuralistas, sino en "la identificación y recopilación de los términos que los especialistas utilizan en realidad" (Cabré Castellví 1993: 113). Por este motivo, si un terminógrafo debe estudiar los términos que los especialistas de un campo de especialidad en cuestión utilizan en su trabajo diario, deberá consultar directamente a dichos especialistas o realizar un estudio detallado de las producciones lingüísticas que estos crean para comunicarse entre ellos o con otros actores.

De estas dos opciones, la primera no es siempre posible, ya que a menudo resulta complicado disponer de los especialistas adecuados y, cuando se puede acceder a ellos, frecuentemente encuentran dificultades a la hora de ex-

[4] Este corpus se ha utilizado en investigaciones previas, aunque no se ha realizado una descripción exhaustiva del mismo. Véanse Durán Muñoz (2019, 2021) y Durán Muñoz y L'Homme (2020).

plicar el significado y el uso del lenguaje que emplean, al fin y al cabo, de forma intuitiva. En palabras de Meyer y Mackintosh (1996: 264):

> While experts obviously *know* their domains, they do not all *explain* their knowledge *clearly* (whether orally or in writing), *completely* (it is up to the knowledge acquirer to make sure that all important areas in the field are covered), or *consistently* (experts often disagree with each other or change their minds).

Por ello, la segunda opción, es decir, la consulta de documentación especializada en forma de corpus textual es más accesible, rápida y directa y, por tanto, determinante en el trabajo terminográfico.

Por otro lado, el segundo motivo que hace imprescindible el uso del corpus en Terminografía se refiere a la dimensión conceptual de los términos. Para poder identificar y recopilar los términos que los especialistas emplean en la realidad, los terminógrafos necesitan estudiar las estructuras de conocimiento (los conceptos y sus relaciones) que los términos representan. Es decir, los terminógrafos deben familiarizarse con el tema específico de su trabajo y adquirir los conocimientos básicos que le permitan estructurar y delimitar adecuadamente el alcance de la obra que pretenden elaborar, lo que Cabré Castellví (1999: 144) denomina "competencia cognitiva". De la misma forma que en el caso anterior, los terminógrafos pueden dirigirse a especialistas en el ámbito de especialidad en cuestión, con lo que se encontrarán con problemas similares a los citados anteriormente, o consultar documentación especializada. En la mayoría de las ocasiones será más fácil para el terminógrafo familiarizarse con el ámbito de especialidad, con sus conceptos y estructuras a través de la documentación especializada, así como consultar a los especialistas para realizarles preguntas acerca del uso, significado, definición, etc. específico de una unidad terminológica que pueda ver en un texto y, por tanto, en su contexto real.

La necesidad de la utilización de documentación especializada en el trabajo terminográfico sistemático queda patente con estas dos razones expuestas. Asimismo, el empleo de corpus textuales se vuelve imprescindible en esta labor si nos referimos al estudio de la variación denominativa y conceptual (Cabré Castellví 1999: 144) y al estudio de la formación de nuevos términos.

La labor terminográfica basada en la documentación especializada o, incluso, en corpus textuales no es nueva, ya que autores de reconocido prestigio en el ámbito terminológico como Rondeau (1983: 71), Dubuc (1980: 24) y Picht y Draskau (1985: 167) se refieren a la utilización desde siempre de fuentes de información lingüística y conceptual. No obstante, estos hacen referencia al análisis manual de los textos utilizados como fuentes de información, por lo que el trabajo terminográfico se vuelve tedioso y extenuante, además de complicado a la hora de realizar, por ejemplo, estudios sobre la variación denominativa o de localizar y ordenar la información lingüística, conceptual y pragmática de los textos. Por este motivo, la mayor parte del trabajo terminográfico actual se realiza de forma semiautomática[5], ya que se encuentra asistido por herramientas informáticas que permiten agilizar y facilitar la labor del terminógrafo.

Este cambio de metodología (de manual a semiautomática) se ha debido principalmente a la revolución tecnológica de las últimas décadas que ha dado lugar a grandes avances, sobre todo en el campo de la informática. Gracias a estos avances, se han incrementado sobremanera la velocidad y la capacidad de almacenamiento de los ordenadores, se han creado herramientas informáticas destinadas a procesar, analizar y recuperar un gran número de textos y, además, se ha facilitado el acceso y la recuperación de textos electrónicos de diferentes contenidos, formatos y ubicaciones a través redes de información existentes, especialmente la red Internet. Todo lo cual ha permitido (y permite) la elaboración, el tratamiento y el procesamiento de corpus textuales de gran tamaño y riqueza almacenados en ordenadores, lo que se denominan corpus textuales informatizados o electrónicos. De esta manera, se ha hecho posible que el terminógrafo pueda tener a su disposición grandes cantidades de texto en formato electrónico, de forma que sus observaciones no estén basadas solo en sus intuiciones lingüísticas, sino en el estudio detallado del uso lingüístico, lo que Leech (1992: 106) reconoce como "a new way of thinking about language".

3. El corpus en el proyecto *DicoAdventure*

El proyecto de investigación DicoAdventure (Ref. UCO-1380857-F), en el que se enmarca este trabajo, gira en torno al ámbito de la Lingüística de corpus, la Léxico-semántica basada en la semántica de marcos y la Terminografía, por lo que se trata de una investigación eminentemente multidisciplinar y en línea con las principales corrientes de trabajo terminográfico actuales.

El principal objetivo de este proyecto es el diseño y desarrollo de un recurso electrónico bilingüe (inglés, español) flexible e integrador, denominado *DicoAdventure*[6], sobre el turismo de aventura que permitirá la adquisición del conocimiento especializado de forma intuitiva y sencilla a partir de marcos semánticos. Para llevar a cabo este objetivo,

---

[5] A pesar de que las herramientas informáticas permiten un manejo automático de la terminología, como por ejemplo para la extracción de términos, de relaciones conceptuales, de cognados, etc., es necesaria la revisión y corrección posterior de los resultados extraídos automáticamente por parte de los terminólogos.

[6] Para consultar el recurso *DicoAdventure* (en construcción), visite http://olst.ling.umontreal.ca/dicoadventure/

es imprescindible alcanzar otros objetivos secundarios, como son el estudio basado en corpus y en la semántica de marcos de la terminología del discurso especializado del turismo de aventura desde un punto de vista morfológico, semántico y pragmático. Por cuestiones de espacio, en este trabajo solo nos centraremos en la parte del corpus, es decir, en su compilación y en su gestión para el trabajo terminográfico.

Con respecto a la metodología que se aplica en este proyecto, todas las fases parten de la compilación y de la gestión del corpus Advencor, un corpus comparable bilingüe (inglés, español) de textos promocionales del sector turístico de aventura (véase sección 4, a continuación) y se dividen en las siguientes:

1. Compilación de un corpus especializado
2. Extracción y selección de unidades especializadas
3. Anotación semántica y sintáctica de contextos de cada unidad terminológica seleccionada
4. Definición de la estructura argumental de cada unidad
5. Detección y distinción de significados de las diferentes unidades
6. Identificación y definición de marcos semánticos
7. Identificación de equivalentes

Como se mencionaba anteriormente, esta metodología se enmarca en la terminología basada en la léxico-semántica (Faber y L'Homme 2014; L'Homme y Robichaud 2014; L'Homme 2016, 2018 y 2020) y en la semántica de marcos (Fillmore 1976 y 1982, Fillmore y Baker 2010). Ambas corrientes parten de la concepción de que los conceptos están relacionados entre sí de forma tan directa que, al activar uno, se activan otros que aparecen frecuentemente en los mismos contextos. Por ejemplo, en el caso de una actividad de aventura como puede ser la escalada, en el momento en el que activa el concepto "escalada", se activan también otros conceptos relacionados, como son el equipo que se utiliza en esa actividad, el lugar donde se realiza, las personas involucradas, etc. Esto hace que el análisis semántico de los conceptos en este tipo de estudios sea un punto muy determinante en este tipo de trabajos. En este trabajo, nos enfocamos en la parte del corpus para el trabajo terminográfico y no tanto en la parte semántica[7], puesto que también se trata de otro aspecto esencial para alcanzar resultados adecuados en este tipo de estudios léxico-semánticos.

En todas las fases indicadas anteriormente, el corpus se vuelve una herramienta imprescindible, ya que permite detectar ejemplos reales de uso del lenguaje especializado, en este caso del lenguaje del turismo de aventura, además de captar las relaciones semánticas entre conceptos, observar los diferentes significados que tienen las unidades especializadas, etc.

De estas seis fases, este trabajo se centra en las fases 1 y 2, es decir, en la compilación del corpus especializado y en la extracción y selección de unidades especializadas, ya que son las primeras tareas que se llevan a cabo en cualquier trabajo terminográfico en la actualidad y, como tales, son esenciales para que las demás tareas se puedan desarrollar de forma adecuada. No obstante, a la hora de seleccionar las entradas para el recurso *DicoAdventure*, es necesario realizar algunas tareas de desambiguación de significados y de análisis de contextos, que entrarían en las fases 4 y 5, como explicaremos en los siguientes apartados.

3.1. La compilación del corpus Advencor

El corpus compilado para este estudio, el corpus Advencor, es un corpus especializado en formato electrónico, bilingüe (inglés / español) y comparable, es decir, contiene textos originales en ambas lenguas que presentan las mismas características. Está formado por textos promocionales sobre el turismo de aventura y se trata de una compilación *ad hoc* para el uso específico de este proyecto.

En cuanto al tamaño del corpus, el subcorpus inglés cuenta con 1 189 409 *tokens*, mientras que el subcorpus español 1 326 337 *tokens*. Ambos son representativos del discurso del turismo de aventura y están equilibrados internamente, lo que garantiza una calidad adecuada para obtener resultados satisfactorios.

En la Tabla 1, se describen las características principales de cada uno de los subcorpus.

[7] En Durán Muñoz y L'Homme (2020) y Durán Muñoz (2016) se lleva a cabo un análisis desde un punto de vista más semántico de las unidades especializadas del turismo de aventura.

Tabla 1. Resumen de características del corpus Advencor

N.º de *tokens*	1 189 409 (inglés) 1 326 337 (español)
Tipo de corpus	Especializado / electrónico / comparable
Modo	Escrito
Idioma	Inglés / español
Dominio / Subdominio	Turismo / Turismo de aventura
Género	Promocional
Longitud de textos	Textos completos
Objetivo	Análisis terminológico y fraseológico
Situación comunicativa	Divulgativo (especializado a no especializado)
Fecha de publicación	2015-2020 ~~Reciente~~
Fuente de textos	Páginas web
Autoría	Entidades públicas o privadas de habla inglesa (agencias de viaje, empresas, etc.)[8]

Tanto el subcorpus inglés como el español se compilaron de forma semiautomática mediante el uso de *Sketch Engine*[9] (Kilgarriff, Rychlý, Smrz y Tugwell 2004), aunque todo el proceso de compilación fue supervisado cuidadosamente para evitar páginas web y datos irrelevantes o inapropiados que pudieran sesgar el análisis final.

El proceso de compilación del corpus que se siguió en el compilador automático se puede dividir en las cuatro siguientes fases:

Paso 1. Selección de las palabras clave para el dominio de interés en las lenguas de trabajo. En nuestro caso, las palabras clave seleccionadas se referían a las actividades de aventura que frecuentemente se ofrecen como parte del discurso del turismo de aventura (ver Tabla 2).

Tabla 2. Palabras clave utilizadas para compilar el subcorpus inglés

adventure sport	*adventure tour*	*adventure tourism*
adventure activity	*kayak*	*parasailing*
adventure activities	*cycling*	*parachuting*
trekking	*mountain biking*	*skydiving*
hiking	*riding*	*caving*
canyoning	*potholing*	*paragliding*
mountaineering	*climbing*	*hang gliding*
kayaking	*bungee jumping*	*dogsledding*
hike	*speleology*	*canoeing*
trek	*rappel*	*adventure*
canyon	*rafting*	*zip lining*

Paso 2. Creación automática de combinaciones de palabras clave para las consultas. En esta fase, las palabras clave indicadas se combinaron aleatoriamente entre sí en diferentes conjuntos de varias palabras generados automáticamente por el programa (por ejemplo, "*adventure tourism kayak trekking*"). Estas combinaciones ofrecidas por el sistema pueden eliminarse si se considera oportuno, aunque normalmente son adecuadas. En nuestro caso, se aceptaron todas las ofrecidas.

Paso 3. Generación de una lista de URL potencialmente relevantes. Esta lista de URL incluye las combinaciones seleccionadas en el paso anterior y, al igual que antes, esta lista propuesta por el sistema puede revisarse y eliminar todas aquellas que no corresponden con los objetivos del proyecto antes de compilar el corpus. En esta fase, se llevó a cabo una revisión exhaustiva de las URL propuestas y se eliminaron todas aquellas de Wikipedia, Amazon, redes sociales (por ejemplo, Facebook y Pinterest), Youtube, Scribd, eBay, etc.; aquellas que no estaban escritas originalmente en inglés o en español, y aquellas que no estaban publicadas por instituciones públicas o privadas (incluyendo agencias o empresas de viajes), tales como artículos, blogs, etc. Una vez finalizada esta revisión manual, el 30 % de las URL propuestas por el sistema fueron descartadas, lo que demuestra la importancia de la revisión manual durante el proceso de compilación automática.

[8] Algunos de las páginas web que se han incluido en este corpus Vi*sit Scotland (ht*tps://www.visitscotland.com/), Vi*sit Peak District (ht*tps://www.visitpeakdistrict.com/), Re*i Adventure (ht*tps://www.rei.com/adventures), Lo*st Earth Adventure (ht*tps://www.lostearthadventures.co.uk/), Mo*untain Wings (htt*p://www.mtnwings.com/) o Dingle Adventure Race (http://dingleadventurerace.com/).

[9] URL: https://app.sketchengine.eu/ [último acceso: 15 de enero 2022].

Paso 4. Recopilación del corpus de las páginas web seleccionadas. La lista de sitios web seleccionados se utilizó para construir el corpus automáticamente. El programa limpió el archivo resultante y eliminó el texto no deseado (como hiperenlaces, imágenes, etc.) para obtener material de calidad. En esta fase, el sistema permite descargar el corpus en el ordenador o guardarlo en el mismo programa para poder gestionarlo con las opciones que ofrece, como son el extractor de terminología, buscador de concordancias, etc.

Como se puede observar, a diferencia de la compilación manual, que requiere mucho tiempo de búsqueda de textos, la compilación automática de corpus ofrece la posibilidad de compilar corpus muy grandes de forma rápida y sin esfuerzo. Sin embargo, es necesario llevar a cabo una revisión manual minuciosa y cuidadosa durante el proceso de compilación (concretamente en los pasos 2 y 3) para refinar las búsquedas y garantizar resultados exitosos, lo que incrementa el tiempo total de compilación del corpus.

3.2. Extracción y selección de las unidades especializadas

Una vez finalizado el proceso de compilación de los dos subcorpus; se utilizó la opción Keywords de Sketch Engine para la extracción de los candidatos a término; que es el primer paso para seleccionar las unidades especializadas del discurso del turismo de aventura.

Antes de abordar el proceso de extracción terminológica como tal, es necesario delimitar qué se entiende por "extracción terminológica". Taljard y De Schryver (2002) lo definen como el proceso mediante el cual se utiliza un programa de ordenador para detectar y extraer automáticamente términos potenciales (o candidatos a términos) de corpus electrónicos. Sin embargo, en todos los enfoques que se pueden aplicar para realizar este proceso, ya sea el lingüístico (basado en los patrones de formación de palabras), el estadístico (basado en la frecuencia de uso de los términos en el corpus de trabajo) o el híbrido (la combinación de ambos), las personas siguen teniendo un lugar preponderante a la hora de revisar y de decidir si los términos sugeridos por el programa en cuestión tienen o no el estatus de término (2002: 46). En otras palabras, una revisión manual después de la extracción automática de candidatos a términos es fundamental para garantizar la calidad de los resultados.

La recuperación y correcta identificación de estas unidades especializadas es la etapa clave para determinar qué conceptos deben incluirse como entradas en los recursos terminológicos. Sin embargo, como señala Cabré Castellví (1993), no todos los términos que aparecen en el texto especializado de una disciplina deben figurar en la terminología que queremos estudiar o tratar, puesto que estos dependerán de los objetivos y de los posibles usuarios del proyecto en cuestión.

Los parámetros que se establecieron para la extracción terminológica de ambos subcorpus en este proyecto fueron tres: 1) se realizó una extracción únicamente de unidades verbales de ambos subcorpus, 2) se estableció un mínimo de dos ocurrencias por unidad para evitar un uso casual, y 3) se utilizaron como corpus de referencia el corpus English Web 2013 ('enTenTen13') para el subcorpus inglés y el corpus Spanish Web 2018 ('enTenTen18') para el subcorpus español. A continuación, la lista resultante fue revisada manualmente con el objetivo de seleccionar las unidades adecuadas para este proyecto. Teniendo en cuenta que la primera fase del proyecto tiene como meta el estudio de los verbos de movimiento de este discurso especializado, se descartaron los siguientes candidatos ofrecidos por el programa:

1) las unidades verbales extraídas que no tenían relación con el discurso del turismo de aventura, por ejemplo, *pride, roof, shop*;
2) las unidades verbales extraídas que no hacían referencia a movimiento, por ejemplo, *enroll, cancel, reserve*;
3) los candidatos que representaban formas distintas del mismo lema, por ejemplo, *paraglided, paraglides, paragliding* del lema *paraglide*.
4) los candidatos que fueron lematizados como unidades verbales de forma incorrecta, por ejemplo, *NZD, yrs, byte, our*.

La extracción terminológica con *Sketch Engine* y su posterior revisión dio como resultado un total de 157 verbos de movimiento para el subcorpus inglés y de 117 para el subcorpus español. Como ejemplo, la Tabla 3 muestra los cinco verbos más frecuentes en el subcorpus inglés.

Tabla 3. Los cinco verbos de movimiento más frecuentes en el subcorpus español

Forma lematizada	Frecuencia	Especificación	Variantes
climb	1899	100,15	climb, climbs, climbed, climbing
hike	1141	80,69	hike, hikes, hiking
trek	915	72,64	trek, treks, trekked, trekking
raft	745	65,59	raft, rafted, rafting
jump	811	58,88	jump, jumps, jumped, jumping

Aunque la mayoría de estos candidatos no plantearon ningún problema, algunos de ellos requirieron un análisis más exhaustivo a los contextos en los que aparecían para desambiguar sus significados y determinar su naturaleza terminológica dentro del turismo de aventura, lo que se realizó de forma generalizada en las fases 4 y 5 de la metodología del proyecto.

Los contextos de las diferentes unidades verbales que sirvieron para el análisis de los diferentes significados se extrajeron con la opción *Concordance*, el buscador de concordancias que ofrece el programa *Sketch Engine* (Figura 1).

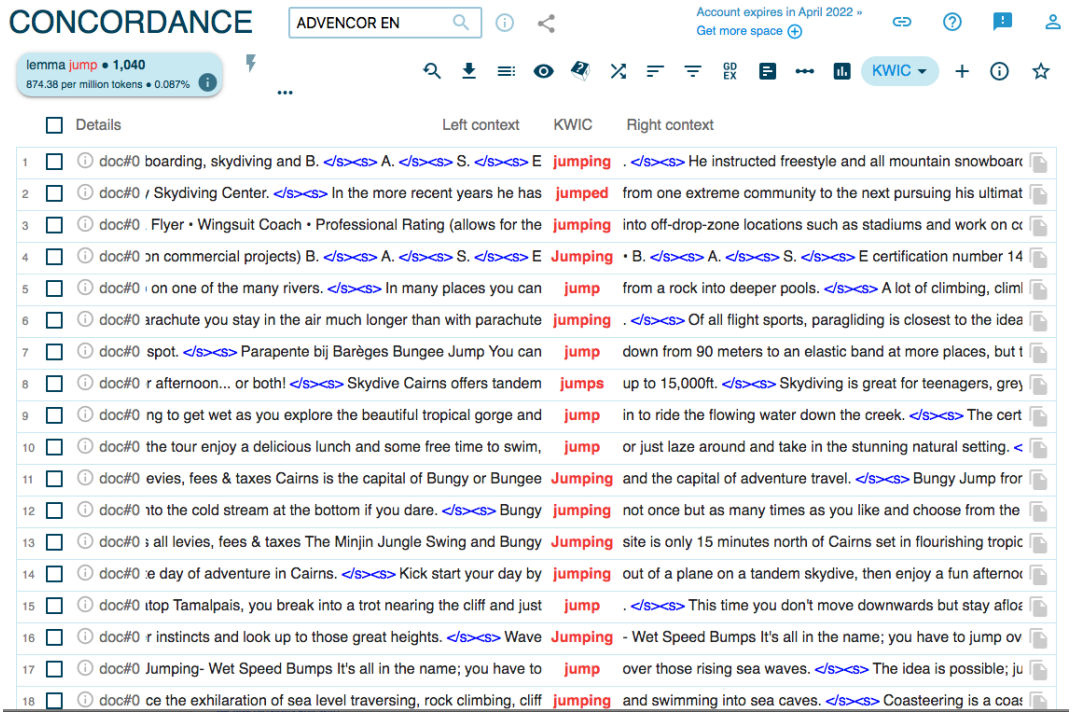

Figura 1. Líneas de concordancias del verbo *jump* con *Concordance*

Un ejemplo claro se encontró con el verbo *to jump* (*saltar*, en español) (Figura 1). Sin tener en cuenta los contextos de uso real de este verbo en su versión en inglés o en su versión en español en el discurso del turismo de aventura, podría parecer lo suficientemente familiar como para ser rechazado directamente al considerar que no forma parte de este discurso especializado. Sin embargo, los contextos de este verbo revelaron que se trata de un término relevante en este ámbito y que debe ser tomado en consideración, puesto que tiene una vinculación muy estrecha con otros conceptos directamente relacionados con el turismo de aventura, tal y como se puede observar en los ejemplos siguientes:

(1) You can *jump* down from 90 meters attached to an elastic band at more places, but the location in Luz St Sauveur (10 min) is unique[10].
(2) This venture involves visitors *jumping* from a platform positioned more than 135m above the river below.
(3) Kick start your day by *jumping* out of a plane on a tandem skydive, then enjoy a fun afternoon of white water rafting in the beautiful Barron Gorge!
(4) In many places you can *jump* from a rock into deeper pools.

Los ejemplos 1, 2, 3 y 4 se refieren a actividades de aventura específicas de este dominio: *puenting* (1 y 2), paracaidismo (3) y descenso de barrancos (4), en los que el verbo *to jump* se vincula a diferentes aspectos esenciales de cualquier actividad de aventura, como son el punto de partida, es decir, el punto desde el cual comienza la acción (en este caso, *platform, plane*); el equipo o los instrumentos que se utilizan para llevar a cabo una actividad concreta (en este caso, *elastic band*); el lugar donde ocurre la acción (en este caso, *river*); el punto de destino, es decir, el lugar hacia donde se dirige la acción (*deeper pools*), y la forma en la que se ejecuta la actividad (*tandem skydive*), así como otras actividades de aventura relacionadas como, por ejemplo, *rafting*. Este análisis de contextos nos permitió hacer una selección de los verbos de movimiento relevantes e incluir, por tanto, solo los verbos que positivamente mostraban un significado vinculado al turismo de aventura.

El análisis de contextos extraídos del corpus también fue necesario para detectar diferentes significados de los verbos de movimiento, como es el caso de *to ascend* (*ascender*) o *to fall* (*caer*). En el primer caso, se detectaron dos significados diferentes del verbo *to ascend*: uno, relativo al movimiento real que hacen las personas cuando suben

[10] La cursiva está añadida por la autora.

(*You'll* ascend *a different via ferrrata each day)*, y otro, el movimiento ficticio de las cosas (normalmente, una carretera, un sendero, etc.) cuando parece que se elevan (*A number of trails* ascend *to the top, including the most popular, the 13-mile Barr Trail*). En este caso, el diccionario *DicoAdventure* incluye dos entradas diferentes, una para cada significado del verbo.

En el segundo caso, con el verbo *to fall*, se detectaron también dos significados diferentes (y, por tanto, también cuenta con dos entradas diferentes en el diccionario), aunque en este caso ambos hacían referencia a movimientos reales: por un lado, el significado del verbo que hace referencia a una caída involuntaria (*Don't worry about your balance, you cannot* fall *down in any case.*) y, por otro, a una caída voluntaria desde un avión, un puente, etc. cuando se está practicando paracaidismo o *puenting* (*Skydivers can also undertake group jumps—these are the ones where divers* fall *in circles or various other formations.*). Al igual que ocurre con el subcorpus inglés, también se detectaron los mismos significados en el subcorpus español.

Finalmente, los contextos también permitieron detectar significados de verbos de movimiento sin relación con el turismo de aventura, como es el caso del verbo *to pull* (*tirar*) o el verbo *to reach* (*alcanzar*).

(5) *Pull* on your bathing suit and water shoes, don a helmet and prepare to get wet and wild.

(6) Feel free to *reach* out and contact us with any questions you have about gear, sponsorships, thru hiking, or becoming a patron of Tandem Trekking!

Tal y como se demuestra con los ejemplos anteriores, el corpus especializado, en este caso el corpus ADVENCOR, sirvió de gran ayuda tanto para extraer los candidatos a términos como para seleccionar los significados de los verbos de movimiento que se incluirán en el recurso *DicoAdventure*.

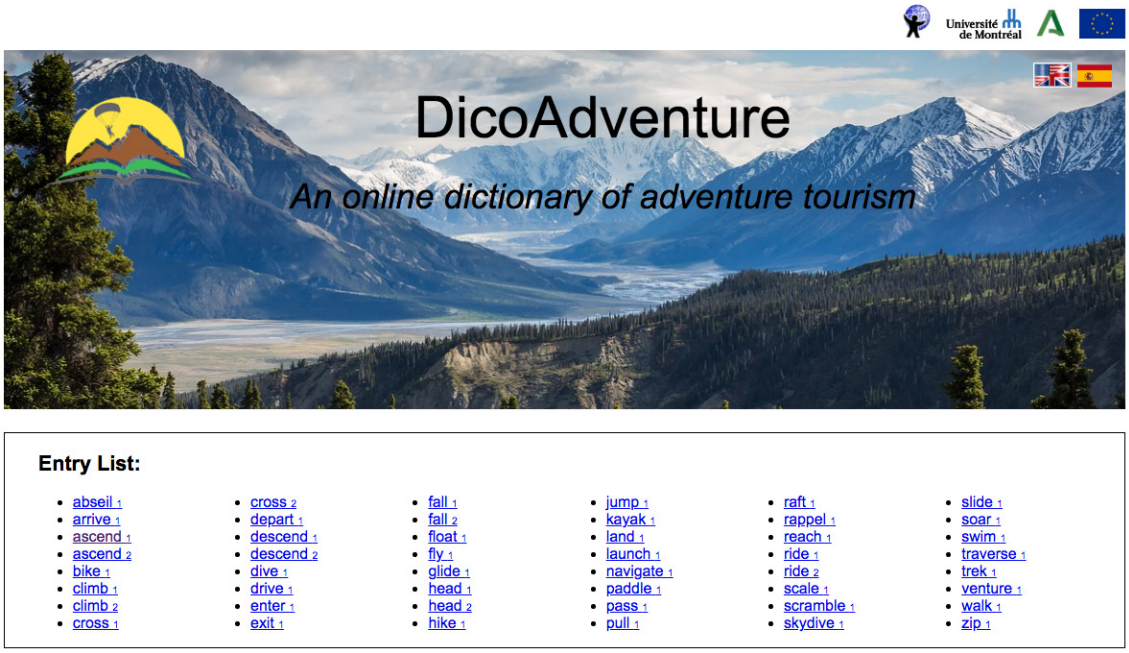

Figura 2. Entradas actuales del recurso *DicoAdventure*

En esta primera fase del proyecto, el número de entradas del diccionario es limitada y se centran únicamente en los verbos de movimiento (Figura 2), aunque seguimos trabajando para aumentar este número de entradas tanto en lo que respecta a las unidades verbales como a otras categorías gramaticales, como son los sustantivos, los adjetivos o los adverbios frecuentes de este campo de especialidad.

4. Conclusiones

A lo largo de este trabajo ha quedado patente la relevancia que tiene el corpus para cualquier trabajo terminográfico hoy en día, ya que, gracias a este recurso se puede manejar material real de forma rápida y sencilla. Esto, no obstante, no podría ser factible si no fuera por los programas de gestión de corpus que hay disponibles en la actualidad, como pueden ser *TermoStat Web 3.0* o *Sketch Engine*.

Además de garantizar unos resultados adecuados, el corpus permite llevar a cabo un trabajo sistemático y completo en cualquier discurso de especialidad, siempre y cuando, por supuesto, se trate de un corpus representativo del discurso en cuestión y cumpla unos requisitos internos y externos que garanticen su calidad (Sinclair 1996; Meyer 2001; Durán Muñoz 2011). Por este motivo, la fase de compilación de corpus de cualquier trabajo terminográfico debe llevarse a cabo de forma muy meticulosa, ya sea de forma manual o de forma (semi)automática mediante el uso de compiladores automáticos.

El proceso de extracción terminológica se concibe como la etapa en la que se reconocen, se delimitan y se recuperan todos los candidatos que pueden ser considerados términos dentro de un discurso de especialidad concreto, en este caso el turismo de aventura, y de acuerdo con los objetivos del proyecto en cuestión. En este sentido, debe realizarse de forma minuciosa, puesto que, durante todo este proceso, se decidirán las entradas que tendrá el futuro recurso terminológico.

Además, como hemos visto con los ejemplos de los verbos *to jump, to pull, to ascend, to fall o to reach*, el corpus sirve para desambiguar significados de términos, para detectar diferentes significados dentro y fuera del discurso de especialidad o para determinar si todos los significados de los términos son especializados mediante el análisis exhaustivo de los contextos reales en los que se encuentran dichas unidades.

Finalmente, debemos tener en cuenta que, independientemente del grado de automatización que tenga cualquier tarea que realicemos con el corpus, ya sea la compilación, la búsqueda de concordancias, la extracción de terminología o la detección de equivalentes, los terminógrafos debemos llevar a cabo una revisión exhaustiva de los resultados para garantizar la calidad de estas actividades y, por ende, del recurso terminológico final.

Referencias

Cabré Castellví, M. T., *La Terminología: Teoría, metodología, aplicaciones*. Ampurias: Editorial Antártida 1993.

Cabré Castellví, M. T., *La terminología, representación y comunicación: Elementos para una teoría de base comunicativa y otros artículos*. Barcelona: Universitat Pompeu Fabra 1999.

Dubuc, R., *Manuel pratique de terminologie*. Montreal: Linguatech 1980.

Durán Muñoz, I., "Criterios específicos para la elaboración y diseño de los corpus especializados para la terminografía", en Carrió Pastor, M. L. y Candel Mora, M. A. (eds.), *Las tecnologías de la información y las comunicaciones: Presente y futuro en el análisis de córpora. Actas del III Congreso Internacional de Lingüística de Corpus*. Valencia: Universitat Politècnica de València 2011, 43— 50.

Durán Muñoz, I., *La ontoterminografía aplicada a la traducción. Propuesta metodológica para la elaboración de recursos terminológicos dirigidos a traductores*. Berlín: Peter Lang 2012.

Durán Muñoz, I., "Producing Frame-Based Definitions: A Case Study", *Terminology* 22, 2 (2016), 223-249. doi: https://doi.org/10.1075/term.22.2.04mun

Durán Muñoz, I., "Adjectives and their Keyness. A Corpus-based Analysis in English Tourism". *Corpora* 14, 3 (2019), 351-378. doi: https://doi.org/10.3366/cor.2019.0178

Durán Muñoz, I., "DicoAdventure y la terminología del turismo de aventura: Propuesta de diccionario en línea", en Barceló Martínez, T.; Delgado Pugés, I. y García Luque, F. (eds.), *Tendencias actuales en traducción especializada, traducción audiovisual y accesibilidad*. Valencia: Tirant Lo Blanch 2012, 395-417.

Durán Muñoz, I. y L'Homme, M. C., "Diving into adventure tourism from a lexico-semantic approach: An analysis of English motion verbs", *Terminology* 26, 1 (2020), 33-59. doi: https://doi.org/10.1075/term.00041.dur

Faber, P. y L'Homme, M. C., "Lexical semantic approaches to terminology: An introduction", *Terminology* 20, 2 (2014), 143-150. doi: https://doi.org/10.1075/term.20.2.01int

Fillmore, C. J., "Frame Semantics and the nature of language",. *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech* 280 (1976), 20-32. doi: https://doi.org/10.1111/j.1749-6632.1976.tb25467.x

Fillmore, C.J., "Frame Semantics", *Linguistics in the Morning Calm* (1982), 111-137. doi: https://doi.org/10.1515/9783110199901.373

Fillmore, C. J. y Baker, C., "A Frames Approach to Semantic Analysis", en Heine, B. y Narrog, H. (eds.), *The Oxford Handbook of Linguistic Analysis*. Oxford: Oxford University Press 2010, 313-339.

Kilgarriff, A.; Rychly, P.; Smrž, P. y D. Tugwell. "The Sketch Engine", *Proceedings of EURALEX*. Lorient, France (2004), 105-116. doi: https://doi.org/10.1007/s40607-014-0009-9

Leech, G. "Corpora and Theories of Linguistic Performance", en Svartvik, J. (ed.) *Directions in Corpus Linguistics. Proceedings of Nobel Symposium* 82. Berlín/Nueva York: Mouton de Gruyter 1991, 105-134. doi: https://doi.org/10.1515/9783110867275.105

L'Homme, M.C. y Robichaud, B. "Frames and terminology: representing predicative units in the field of the environment", *Cognitive Aspects of the Lexicon (Cogalex 2014)*, *Coling 2014*, Dublín (Irlanda) 2014. doi: https://doi.org/10.3115/v1/W14-4723

L'Homme, M. C. "Terminologie de l'environnement et Sémantique des cadres", *Congrès Mondial de Linguistique Française — CMLF 2016.* SHS Web of Conferences 27 (2016), 1-14. doi: https://doi.org/10.1051/shsconf/20162705010

L'Homme, M. C., "Maintaining the balance between knowledge and the lexicon in terminology: a methodology based on frame semantics", *Lexicography: Journal of ASIALEX* 4, 1 (2018), 3-21. doi: https://doi.org/10.1007/s40607-018-0034-1

L'Homme, M. C., *Lexical semantics for terminology: An introduction*. Ámsterdam: John Benjamins 2020.

Meyer, I. "Extracting knowledge-rich contexts for terminography. A conceptual and methodological framework", en Bourigault, D.; Jacquemin, C. y L'Homme, M. C. (eds.), *Recent Advances in Computational Terminology.* Filadelfia: John Benjamins 2001, 279-302.

Meyer, I. y Mackintosh, K., "The Corpus from a Terminographer's Viewpoint", *International Journal of Corpus Linguistics* 1, 2 (1996), 257-285. doi: https://doi.org/10.1075/ijcl.1.2.05mey

Picht, H. y Draskau, J., *Terminology: An Introduction*. Guildford: Universidad de Surrey 1985.

Rondeau, G., *Introduction à la terminologie*. Chicoutimi (Québec): Gaëtan Morin 1983.

Sinclair, J., *Preliminary Recommendations on Corpus Typology. EAGLES Document EAG-TCWG-CTYP/P* (1996). http://www.ilc.cnr.it/EAGLES96/corpustyp/corpustyp.html [último acceso: 15 de febrero 2022].

Taljard, E. y de Schryver. G. M., "Semi-automatic Term Extraction for the African Languages, with Special Reference to Northern Sotho", *Lexikos* 12 (2002), 44-74. doi: https://doi.org/10.5788/12-0-760