

APRENDIZAJE DE INGLÉS PARA FINES ESPECÍFICOS: CREACIÓN DE UN DICCIONARIO DIGITAL MONOLINGÜE EN EL AULA DE TURISMO

BEATRIZ CHAVES-YUSTE

Universidad Complutense de Madrid

MANUEL MÁRQUEZ-CRUZ

Universidad Complutense de Madrid

1. INTRODUCCIÓN

En los últimos años, ha ido creciendo el interés por analizar lingüísticamente las opiniones de los usuarios respecto a sus experiencias en alojamientos turísticos alojados en plataformas virtuales tales como *Tripadvisor*, *Cozy Cozy* or *Airbnb*. Estas plataformas albergan las opiniones de los usuarios de una manera pública, por lo que los propios consumidores desempeñan el rol de “prosumidores” (Toffler, 1980), en tanto y en cuanto promueven y producen contenido en línea respecto a los alojamientos comentados. Se convierte así en una potente publicidad gratuita para las diferentes empresas que ofrecen sus prestaciones y pueden encontrar una difusión rápida y masiva de sus servicios por la mera generosidad de su clientela. Las reseñas relacionadas con los diferentes ámbitos del turismo (alojamientos, restauración, transporte, actividades de ocio, sitios de interés turístico y cultural, entre otros), han despertado gran interés no solamente en los potentes consumidores, sino también, en la comunidad científica. En este último caso, existen numerosos trabajos que se han centrado en el ámbito del análisis del discurso, orientados a analizar la estructura narrativa de reseñas publicadas en plataformas como *Tripadvisor* (Vásquez, 2011, 2012; Zhang y Vásquez, 2014; Feng y Ren, 2019), el grado de satisfacción en *Airbnb* (Hernández López, 2019), la mitigación en los actos de habla (Hernández Toribio y

Mariottini, 2015; Mancera, 2018; Mancera y Pano, 2021) o el análisis de las interacciones anfitrión-huésped (Hernández López, 2019).

Como señaló Bennet (2010), el inglés para fines específicos (de aquí en adelante, IFE), entendido como el uso de la lengua por una comunidad discursiva específica que comparte un campo profesional común (Asher y Simpson, 1994), la lingüística de corpus permite la compilación de corpus que proporcionan datos sobre el uso de dicha lengua muy útiles para su manejo, ya que proporciona un marco para estudiar disciplinas especializadas de la lengua, con el objetivo de descubrir y poner de relieve sus regularidades a distintos niveles, incluyendo conjuntos léxicos y terminología especializada. La compilación de corpus *ad hoc* permite incluso una mayor especialización que puede dar respuesta a las necesidades particulares de los estudiantes de IFE. Más aún, cuando estos corpus se compilan a raíz de la terminología encontrada en plataformas online, reflejan el constante cambio y evolución que la terminología puede sufrir en muchos ámbitos de especialización (Boulton, 2012). Se ha demostrado que la compilación de términos específicos de un área de conocimiento a través de corpus sistematiza y simplifica los materiales docentes, ofreciendo un vocabulario técnico de calidad al alumnado universitario (Wu, 2014).

Diferentes investigaciones se han centrado en la creación de corpus de textos especializados en medicina (Wang et al., 2008; Hsu, 2013; Lei y Liu, 2016). Sin embargo, estos corpus no han tenido en cuenta una descripción detallada de la terminología que recoge en cuanto a su morfología, por ejemplo.

2. MARCO TEÓRICO

2.1. LINGÜÍSTICA DE CORPUS: SKECTH ENGINE

Si bien es cierto que ya existía un trabajo con corpus previo al uso de los ordenadores para el análisis de los datos lingüísticos (Villayandre, 2008), la Lingüística de Corpus (en adelante, LC) entendida como “una aproximación al estudio de los hechos lingüísticos de orientación empírica y basada en el análisis detallado de gran cantidad de datos (los

corpus)” (Rojo, 2021: 48), ha supuesto toda una revolución instrumental en el ámbito de la investigación lingüística durante los últimos decenios, gracias, precisamente, al componente computacional que ha ayudado a agilizar y hacer más efectivas el almacenamiento y la consulta de los corpus. Se ha discutido sobre su naturaleza teórica, metodológica o como su concepción como subdisciplina lingüística, como demuestra la amplia bibliografía al respecto (Svartvik, 1992; Leech, 1992; McEnery y Wilson, 1996; Stubbs, 1996, 2001; Kennedy, 1998; Tognini-Bonelli, 2001; McEnery, Xiao y Tono, 2006; Gries, 2006; Guilquin y Gries, 2009; Parodi, 2010) sin que haya consenso al respecto. Ahora bien, independientemente de su consideración, es innegable que, tal y como afirma Rojo (2021), la LC se caracteriza por (i) ser empírica, (ii) utilizar como base de los análisis corpus textuales amplios que proporcionan datos obtenidos a partir de los textos producidos de forma natural, (iii) emplear ordenadores como soporte para trabajar con dichos corpus, (iv) usar técnicas analíticas cualitativas y cuantitativas y (v) tener como finalidad la realización de análisis sistemáticos y exhaustivos de los datos proporcionados por los corpus. En cuanto al concepto de corpus, nos parece que la definición de Rojo es la que mejor sintetiza su naturaleza, pues indica que un corpus es un conjunto de textos, orales o escritos, producidos en condiciones naturales, conjuntamente representativos de una lengua o una variedad lingüística, en su totalidad o en algunos de sus componentes, que se almacenan en formato electrónico y se codifican con la intención de que puedan ser analizados científicamente (Rojo, 2021)¹⁰. Dicho corpus debe cumplir los siguientes requisitos (Villayandre, 2008):

- Formato electrónico.
- Autenticidad de los datos.
- Selección de los textos que conforman el corpus de acuerdo a unos criterios condicionados por la finalidad del corpus.
- Representatividad de la variedad de la lengua objeto de estudio.
- Tamaño adecuado a la finalidad del corpus.

¹⁰ Para otras definiciones de corpus, vid. Villayandre (2008).

En cuanto a la tipología de los corpus, es sumamente amplia y responde a diferentes criterios (Rojo, 2021; Villayandre, 2008): pueden ser generales (o de referencia) o ser representaciones de las variedades de una lengua (corpus dialectales); pueden ser representativos de una época (sincrónicos) o de un largo periodo (diacrónicos); según el medio pueden ser orales o escritos; pueden tener un carácter general o especializado, como sucede con los corpus técnicos¹¹; se distinguen, también corpus abiertos (aquel que se concibe como un corpus continuamente actualizable) o cerrados, o, según la lengua o lenguas empleadas, monolingües, bilingües o multilingües; en cuanto a la anotación de los corpus, se distingue también entre corpus anotados y codificados o los que no lo están, permitiendo los primeros recuperar datos de una manera más concreta y precisa.

Así pues, la idoneidad del texto permite, mediante un análisis propio de la LC, observar el número total de palabras (*tokens*) o de formas distintas (*types*), la frecuencia total o absoluta de una forma en el corpus, la frecuencia normalizada o relativa (si se tiene en cuenta las veces que aparece la forma en relación con el tamaño del corpus), las clases de palabras, las palabras claves de un corpus (si se tiene en cuenta la frecuencia de una palabra en un corpus concreto comparando dicha frecuencia con la de otro corpus más amplio tomado como referencia), los contextos en los que aparecen las palabras, los patrones morfosintácticos más frecuentes, las combinaciones de palabras y, en caso de que el corpus haya sido marcado con dicha información, los usos sociolingüísticos, diafásicos, diastráticos, diatópicos o diacrónicos del léxico, entre otros.

El trabajo con los corpus en la LC permite un trabajo en dos direcciones. Se puede partir del análisis del corpus para obtener teorías (técnica conocida como *corpus driven*) o se puede utilizar el corpus para comprobar teorías (*corpus based*). En ambos casos, el análisis del corpus es fundamental puesto que proporciona los datos materiales, esto es, los hechos lingüísticos, para formular o comprobar teorías lingüísticas.

¹¹ Nuestra propuesta se fundamenta en la construcción de un corpus técnico, puesto que se trata de crear un corpus de aprendientes de un dominio concreto, como es el IFE en el ámbito de Turismo.

Debido al aporte que supone para el estudio de las lenguas el trabajo con corpus, desde el ámbito de la Lingüística Computacional y del Procesamiento del Lenguaje Natural se han desarrollado diversas herramientas que permiten compilar de forma automática y semiautomática corpus, así como su análisis. Resulta especialmente relevante la técnica de compilación semiautomática puesto que evita el ruido documental que produce la presencia de información irrelevante o de calidad cuestionable (Costa et al., 2016). Para proceder a tales efectos se han desarrollado herramientas como *Terminus*, *Corpografo*, *Babouk*, *iCorpora*, *BootCat* y *WebBootCat* (Toledo y Camero, 2021), entre otras. Es especialmente relevante el uso de *WebBootCat*, utilizada por *Sketch Engine*, cuyo funcionamiento se basa en la herramienta *Corpus Architect* que permite la compilación de corpus desde diferentes fuentes, entre ellas, la web, sin la necesidad de que el usuario sea un experto informático en la compilación de corpus, tal y como se explicará a continuación.

*Sketch Engine*¹² es una herramienta de análisis de corpus diseñada para trabajar con diferentes lenguas. Se trata de un software comercial, por lo que su uso requiere del pago de una licencia individual o institucional. Permite consultar y explorar corpus ya precargados en la herramienta¹³, además de importar, consultar y explorar corpus propios. La herramienta proporciona dos opciones para la importación del texto o los textos que conformarán el corpus: subiendo a *Sketch Engine* nuestros propios textos (el sistema soporta textos en los siguientes formatos: .csv, .doc, .docx, .htm, .html, .ods, .pdf, .tar.bz2, .tar.gz, .tei, .tgz, .tmx, .txt, .vert, .xlf, .xliff, .xml, .zip) o buscándolos en internet (mediante la búsqueda de términos o lemas denominados *seed words*, a través de direcciones URL desde las que la herramienta compila de forma automática el corpus o por medio de la compilación completa del contenido de una página web).

Una vez cargados los textos, la *Sketch Engine* permite consultar los corpus para inferir, por ejemplo, diferentes patrones lingüísticos y especificidades de las lenguas, además de contar con herramientas de anotación automática y análisis estadísticos parametrizables por el usuario, lo que

¹² <https://www.sketchengine.eu/>

¹³ Toledo y Camero (2021) la herramienta tiene cargados 517 corpus de 143 lenguas.

permite realizar consultas en todo el corpus o en subcorpus definidos por el usuario (Del Olmo y Arias, 2021).

El sistema presenta un panel con las siguientes herramientas¹⁴:

- Word Sketch: permite la búsqueda de patrones morfosintácticos de un lema, devolviendo como información la posición del lema en relación con la función sintáctica.
- Tesoro: permite obtener una lista de sinónimos de un término concreto en un corpus.
- N-gramas: sirve para obtener las expresiones multipalabras más frecuentes del corpus, determinando el número de unidades que constituirá la expresión.
- Diferencia Sketch: permite la comparación del comportamiento lingüístico (patrones morfosintácticos) entre dos lemas.
- Concordancia: recoge las palabras en sus respectivos contextos.
- Lista de palabras: se obtiene una lista de las palabras más frecuentes del corpus.
- Palabras claves: Proporciona una lista de las palabras que caracterizan al corpus foco de estudio, contrastando dichas palabras con un corpus de referencia mucho más amplio.

¹⁴ Junto a las herramientas que se describen a continuación, el sistema proporciona las herramientas Tendencias, Concordancia paralela, Términos bilingües y Diccionario OneClick que, dados los límites del capítulo y la nula repercusión para la confección del modelo lexicográfico que proponemos, omitimos su explicación.

FIGURA 1. Vista previa del tablero principal de Sketch Engine con las diferentes herramientas que proporciona el sistema para el análisis de corpus



Fuente: Producción propia

2.2. LEXICOGRAFÍA DIDÁCTICA

La lexicografía es una disciplina milenaria cuya vertiente práctica se ha orientado tradicionalmente a la compilación mediante diferentes formatos del vocabulario total o parcial de una o varias lenguas, con el objetivo de satisfacer la necesidades cognitivas y comunicativas de los miembros de una comunidad. El formato de los datos ha variado desde las primeras listas de palabras o glosas que facilitaban la comprensión de aquellos términos complejos, ya sea por su antigüedad, ya por su carácter técnico, religioso, etc. (Boisson et al., 1991; Lara, 1997; Boulanger, 2003; Béjoint, 2016), tal y como hicieron los escribas del siglo V a.C. en la Antigua Grecia con las obras de Homero (Tarp, 2019). Una de las ramas de la lexicografía es la denominada “lexicografía didáctica” (o *pedagogical lexicography* en inglés), que tiene como finalidad desde una perspectiva práctica la confección de diccionarios que faciliten el proceso de aprendizaje de una lengua, mientras que desde una perspectiva teórica (la

conocida como metalexigrafía) se ocupa del estudio crítico de los diccionarios desde la perspectiva de la historia, la estructura y el uso de estas obras, entre otros temas, con el objetivo de mejorar la producción de diccionarios. La lexicografía didáctica distingue dos tipos de obras: los diccionarios escolares, que son aquellos que han sido confeccionados para dar respuesta a las necesidades cognitivas y comunicativas de los hablantes nativos de una lengua, y diccionarios de aprendizaje, que son aquellos otros que han sido confeccionados para satisfacer las necesidades cognitivas y comunicativas de los hablantes no nativos de una lengua. Como es evidente, los primeros diccionarios mencionados son monolingües, mientras que los segundos comprenden un formato monolingüe, bilingüe, semibilingüe o multilingüe, según el número de lenguas codificadas en el diccionario. En ambos casos, los diccionarios suelen responder a las necesidades cualitativas y cuantitativas propias de un nivel de aprendizaje, distinguiéndose entre diccionarios para aprendientes de nivel inicial, intermedio y avanzado. Estas obras se confeccionan teniendo en cuenta aspectos que tienen que ver con el número de voces lematizadas, el vocabulario empleado en las definiciones, la tipografía del conglomerado de textos que componen el artículo lexicográfico, la iconoestructura del diccionario, etc. De hecho, Heuberger (2016) distinguió una serie de rasgos que deben estar presentes en los denominados como diccionarios de aprendizaje y que consideramos que también deben formar parte de los diccionarios escolares: el uso de vocabularios definatorios controlados, definiciones, ilustraciones, ejemplos, información gramatical, información de uso (pragmática), colocaciones, pronunciación, accesibilidad (a los datos), textos preliminares y apéndices (materiales finales).

En el caso concreto de los diccionarios de aprendizaje de inglés, su origen se remonta a los años treinta del siglo pasado, cuando se publican diferentes obras lexicográficas con la finalidad de mejorar el proceso de aprendizaje del inglés como segunda lengua. Nos referimos al *The New Method English Dictionary* (1935) de West y Endicott, al *Idiomatic and Syntactic English Dictionary* (1942) de Hornby, Gatenby y Wakefield y al *Oxford Advanced Learner's Dictionary of Current English* (1952), que es una reimpresión del *The Advanced Learner's Dictionary of Current English* (1948) que es, a su vez, una reimpresión de la obra de

Hornby, Gatenby y Wakefield. Las grandes editoriales siguieron la línea marcada por estas obras y continuaron publicando materiales lexicográficos en los años venideros. No obstante, las limitaciones del papel como soporte de publicación y el acceso gratuito y online a diccionarios publicados en internet ha sumido en un periodo de crisis a la industria editorial del diccionario (Tarp, 2019). Si bien es cierto que las editoriales no están atravesando su mejor momento, el formato digital además de poner fin a esa limitación espacial del papel, el gran lastre de la lexicografía tradicional, ha mejorado el acceso a los datos lexicográficos, con el desarrollo de herramientas que facilitan la confección de diccionarios en sus diferentes fases, como son los *Dictionary Writing Systems*, lo que ha repercutido positivamente (i) en la ampliación del abanico de usuarios de los diccionarios, dado el acceso libre que proporcionan algunos de los diccionarios publicados en internet, y (ii) en la confección de obras lexicográficas que responden a necesidades particulares de los aprendientes, como veremos en este capítulo. Todo ello, sin que sea necesario que quienes se encuentran al mando de los proyectos lexicográficos tengan una formación informática específica.

2.3. DICTIONARY WRITING SYSTEMS: LEXONOMY

Los *Dictionary Writing Systems* (en adelante, DWS) se conciben como sistemas que ayudan a mejorar el flujo de trabajo durante el proceso de confección de un diccionario: desde el diseño de los artículos lexicográficos hasta la publicación del propio diccionario. Una gran parte de estos sistemas presentan una interfaz amigable e intuitiva dotadas de un conjunto de herramientas administrativas que facilitan la gestión del proyecto lexicográfico. Entre las funcionalidades que deben proporcionar estos sistemas se encuentran (Kallas et al., 2019):

- Soporte para recopilar automáticamente los datos en diferentes formatos.
- Soporte para gestionar y almacenar datos y para su extracción automática.
- Soporte para publicar datos.
- Soporte para gestionar la participación de miembros y colaboradores, internos y externos, del proyecto lexicográfico.

En su estructura se distinguen tres componentes:

- Un editor de textos cuya interfaz permite diseñar y editar entradas.
- Una base de datos en la que se almacenan los datos lexicográficos.
- Un conjunto de herramientas administrativas que facilitan la gestión del proyecto lexicográfico y el proceso de publicación del producto final. (Atkins y Rundell, 2008).

Son tres componentes con finalidades distintas pero que conceptualmente se encuentran estrechamente unidos, puesto que interactúan en las diferentes fases que comprende la confección del diccionario.

El editor de textos es la herramienta usada para la compilación y edición de los textos del diccionario. A través de un sistema de formularios, se van introduciendo los textos de las diferentes estructuras del diccionario. Por lo general, suelen presentar una pantalla de acceso a los datos dividida en dos secciones: una que ofrece una visión global de los textos con los que se está trabajando (e. g., una lista de los artículos lexicográficos con los que va a trabajar el miembro del proyecto) y una visión individualizada de cada uno de los elementos con los que se está trabajando (e. g., un artículo lexicográfico con las diferentes categorías que lo componen para que se puedan cumplimentar o modificar).

En la base de datos se almacena el conjunto de textos que dan forma al diccionario. El motor de consultas de la base de datos proporciona un acceso total o parcial a las estructuras lexicográficas: se pueden consultar todas las entradas o parte de ellas, según la selección de criterios de consulta (e. g., la categoría gramatical) o recuperar los ejemplos, las definiciones, las formas de un lema, las marcas lexicográficas, etc.

Las herramientas administrativas permiten la distribución del trabajo del proyecto lexicográfico en sus diferentes fases, designando las distintas tareas a realizar a cada uno de los miembros del proyecto lexicográfico (determinando, e. g., el acceso a un número concreto de entradas o a determinadas partes de cada entrada lexicográfica, en atención, por ejemplo, a la especialidad del lexicógrafo). Contribuye, también a facilitar el proceso de revisión de los materiales y a determinar los permisos de acceso a los datos, tanto a nivel interno (diseño, confección y

administración del proyecto), como a nivel externo (acceso a los datos del producto final, esto es, el diccionario publicado en un formato digital) o los permisos de exportación e importación de datos o las correcciones masivas.

A pesar de los avances que se ha producido en la implementación de este tipo de herramientas, la cantidad y la calidad de las herramientas que proporciona el editor variará según el tipo de DWS, distinguiéndose entre comerciales¹⁵ (son sistemas con altas prestaciones en cuanto al número de herramientas y a su funcionalidad) y de código abierto o acceso libre (presentan un número más restringido de herramientas de administración y edición, así como una menor capacidad de almacenamiento de datos)¹⁶. En cuanto a la codificación de datos, se utilizan sistemas genéricos, como los editores de XML, que permiten codificar mediante un sistema genérico o estándar de etiquetas jerarquizado de apertura y cierre la estructura de un diccionario, de manera completamente abierta y a criterio del lexicógrafo¹⁷, o se utiliza un sistema propio, que resulta de la adaptación de editores genéricos o estándares a las necesidades de los proyectos lexicográficos (Kallas et al., 2019; Rubio et al. 2021; Abel, 2012). Como es evidente, los DWS comerciales suelen ser más completos en lo que se refiere a su funcionalidad que los de software abierto.

A pesar de lo dicho en las líneas precedentes, los DWS de libre acceso son herramientas lo suficientemente potentes para gestionar un proyecto lexicográfico de media envergadura. Uno de estos sistemas en abierto es *Lexonomy*.

Lexonomy es un ejemplo de DWS en abierto. Se concibe como una plataforma web diseñada para facilitar las labores de compilación y publicación de diccionarios de tamaño medio¹⁸. Utiliza el lenguaje XML (*Extensible Markup Language*) para el tratamiento de los datos lexicográficos, esto es, una estructura interna de datos en forma de etiquetas anidadas de apertura y cierre organizadas jerárquicamente en forma de árbol.

¹⁵ Un ejemplo de DWS comercial es TLex: <https://tshwanedje.com/tshwanelex/>

¹⁶ Un ejemplo de DWS comercial es Lexonomy: <https://www.lexonomy.eu/>

¹⁷ El uso, por ejemplo, de Oxygen XML Editor: <https://www.oxygenxml.com/>

¹⁸ Para más información vid. <https://www.lexonomy.eu/docs/elex2017.pdf>

Este sistema de marcado no presenta restricciones -más allá de su propia gramática, definida, eso sí, por el usuario, según las necesidades de marcación del documento- en cuanto al número de etiquetas utilizadas y a la nomenclatura de las mismas, lo que permite personalizar la información contenida en el diccionario en atención a diferentes y variadas necesidades. Se trata de un sistema estándar que permite compartir posteriormente los datos lexicográficos con otros sistemas como las bases de datos, otros DWS o los navegadores de internet.

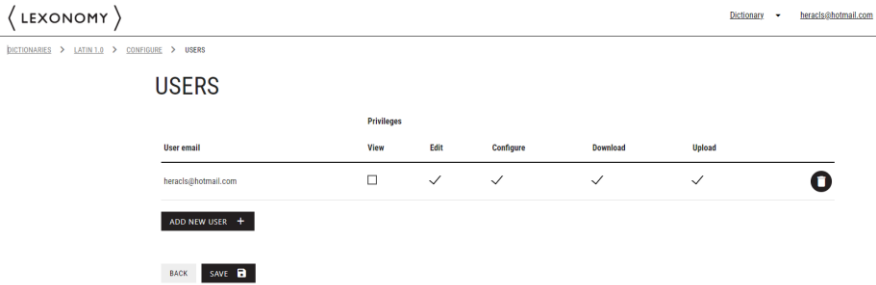
Lexonomy se caracteriza por presentar un entorno *friendly*, que facilita el uso del editor de diccionario y la administración del proyecto sin necesidad de tener altos conocimientos informáticos. Tal y como describíamos en los párrafos anteriores, consta de una base de datos donde se almacena toda la información utilizada para confeccionar el diccionario, un editor con el que se gestiona el diseño del diccionario y la forma de los diferentes textos que conforman el diccionario y un administrador de tareas que sirve para gestionar la estructura del proyecto y el acceso interno y externo a los datos (Figuras 2, 3, 4 y 5):

FIGURA 2. Vista de la estructura de los artículos lexicográficos verbales de un proyecto de diccionario bilingüe de latín-español confeccionado con *Lexonomy*

The screenshot displays the 'ENTRY STRUCTURE' configuration page in the Lexonomy application. At the top, the breadcrumb trail reads 'DICCIONARIOS > LATIN1.0 > CONFIGURE > ENTRY STRUCTURE'. The main title is 'ENTRY STRUCTURE'. On the left, a tree view shows the XML structure with elements like <entry>, <headword>, <partOfSpeech>, <valency>, <sense>, <valency_trans>, <translation1>, <example1>, <Morph_descrip1>, <morph1>, <comp_morph1>, <trad_morph1>, <th>, <examp11>, <translation2>, and <examp12>. The right panel is titled 'Element' and shows a configuration table for the 'entry' element. The table has columns for the element name, 'min', and 'max'. The 'entry' element is currently empty. Below this, there are sections for 'Attributes' (with an 'Add' button), 'Content' (with radio buttons for 'CHILD ELEMENTS', 'TEXT', 'TEXT WITH MARKUP', 'VALUE FROM LIST', 'EMPTY', and 'MEDIA', where 'CHILD ELEMENTS' is selected), and 'Child elements'. The 'Child elements' section lists: <headword> (min: 1, max: 1), <partOfSpeech> (min: 1, max: 1), <valency> (min: 1, max: 1), and <sense> (min: 1, max: 1). Each row has a green 'i' icon and a red 'x' icon. At the bottom of the child elements section is an 'Add' button.

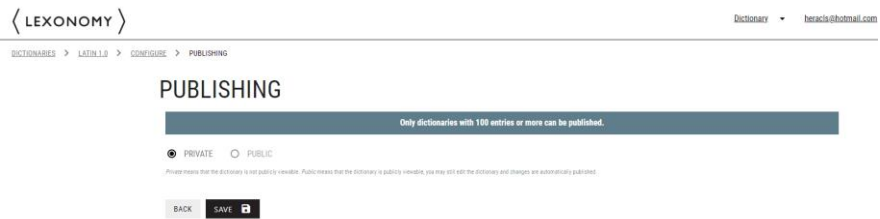
Fuente: Producción propia

FIGURA 3. Vista de la configuración del acceso interno a los datos de un proyecto de diccionario bilingüe de latín-español confeccionado con Lexonomy



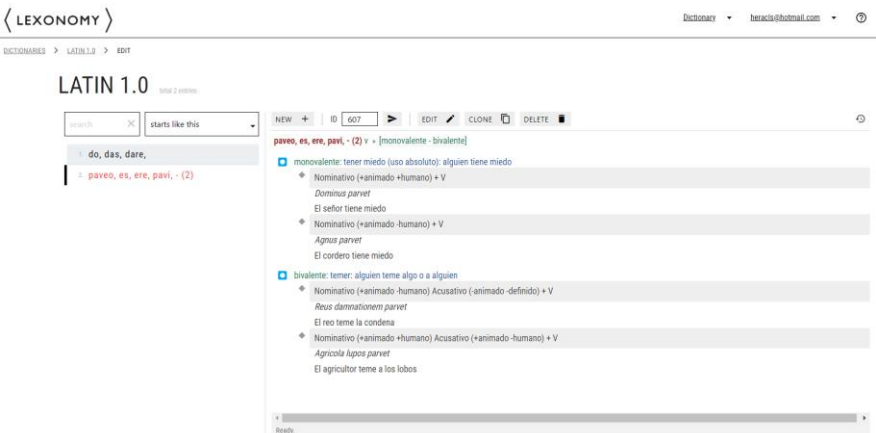
Fuente: Producción propia

FIGURA 4. Vista de la configuración del acceso externo a los datos de un proyecto de diccionario bilingüe de latín-español confeccionado con Lexonomy



Fuente: Producción propia

FIGURA 5. Vista de la publicación previa del artículo lexicográfico s.v. *pereo* de un proyecto de diccionario bilingüe de latín-español confeccionado con Lexonomy



3. OBJETIVOS

El objetivo principal de este trabajo es el de mejorar la adquisición de vocabulario específico del campo de turismo, concretamente, del léxico relacionado con los alojamientos turísticos, y por ende, mejorar la competencia lingüística de los estudiantes de Inglés I del grado de turismo.

Para ello, partiremos de la consecución de dos objetivos secundarios. En primer lugar, crear un corpus con el software *Sketch Engine* con los comentarios extraídos de la plataforma digital *Cozy Cozy* especializada en alojamientos turísticos para hacer una selección de las entradas léxicas más frecuentes y estudiar así este léxico propio del campo específico del turismo. En segundo lugar, una vez confeccionado y analizado el corpus, compilar un diccionario digital en abierto utilizando la herramienta de *Lexonomy*, para que durante su confección, el alumnado experimente un aprendizaje incidental, y durante su uso y fácil manejo, un aprendizaje intencional.

4. METODOLOGÍA

El presente estudio se basa, en un primer lugar, en la selección y compilación de lemas relacionados con los alojamientos turísticos para estudiar el léxico de este campo de interés dentro de los parámetros del inglés para fines específicos (turismo). El análisis se enfocará en el uso de lemas frecuentes relacionados con los alojamientos compartidos por angloparlantes respecto a varios destinos populares de habla anglosajona (Nueva York, Los Ángeles, Londres, Edimburgo y Dublín) durante un periodo específico (todas las reseñas recogidas durante los tres primeros trimestres del año 2023 -enero-septiembre).

El método de investigación respetará los cuatro principios básicos de la lingüística de corpus desarrollados por Biber et al. (1998):

- Análisis empírico de patrones del lenguaje en textos auténticos.
- Utilización de un volumen extenso de textos auténticos como base del análisis.
- Empleo de herramientas digitales para su análisis.
- Uso de análisis cuantitativo y cualitativo.

4.1. PARTICIPANTES

La muestra de estudio se compondrá de un grupo de cien estudiantes divididos aleatoriamente en dos grupos (grupo de control y grupo experimental). El grupo de control compilará un glosario en papel con los términos más frecuentes utilizados en las opiniones recogidas de los alojamientos disponibles en cinco ciudades turísticas de habla inglesa. Por su parte, el grupo experimental trabajará con la herramienta *Sketch Engine* para la creación del corpus que más tarde recogerán en el diccionario digital con la herramienta *Lexonomy*. Este diccionario digital será publicado en abierto para que se encuentre a disposición de todo usuario interesado. Para organizar el trabajo, tanto el grupo de control como el experimental se organizarán en grupos de trabajo cooperativo heterogéneos, teniendo en cuenta el nivel de la lengua y las características propias del alumnado.

4.2. INSTRUMENTOS

Para testar la eficacia de la creación del corpus y su explotación mediante la compilación de un diccionario digital *ad hoc*, testaremos si la competencia lingüística en lengua inglesa de los discentes, dentro de su ámbito de especialización de turismo, ha mejorado tras la experimentación. Con este fin, utilizaremos dos instrumentos de medición cuantitativos: un pre y un post test. Tras la recogida de datos, se realizará un análisis estadístico inferencial para comprobar si la experimentación arroja diferencias significativas. Tanto el pre como el post test evaluarán las cuatro destrezas lingüísticas (comprensión oral y escrita, expresión oral y escrita) y el uso de la lengua, todo ello relacionado con el campo del turismo. Se les asignará una calificación de 0 a 10 puntos, siendo de 0-4.99 puntos, suspenso, de 5 a 6.99 puntos, aprobado, de 7 a 8.99 puntos, notable y de 9 a 10 puntos, sobresaliente. Cada una de estas cinco secciones tendrá una puntuación máxima de 2 puntos y comprenderán diferentes ejercicios para testar las diferentes destrezas y manejo de la lengua.

En lo que se refiere a la actividad en sí, la primera fase consiste en compilar y analizar el corpus. Para ello se utilizará *Sketch Engine* (la Universidad Complutense de Madrid tiene una licencia anual institucional):

mediante la herramienta NUEVO CORPUS se creará un corpus compilado a partir de las URL de las páginas de *Cozy Cozy* (en inglés) con los destinos seleccionados. A continuación se utilizarán el resto de herramientas para determinar las colocaciones y combinaciones de palabras más usuales (Word Sketch), las expresiones multipalabras (N-gramas), las palabras más frecuentes (Lista de Palabras) y palabras que determinan la temática del corpus, tomando como corpus de referencia el corpus English Web 2021, que contiene más de 52 millones de palabras (Palabras clave), los contextos en los que aparecen las palabras más relevantes del corpus (Concordancia) y las colocaciones más usuales entre términos similares (Diferencia Sketch), para, finalmente, recopilar un tesoro de sinónimos de aquellos términos que por semánticas puedan presentar ciertas dificultades en su definición o equivalencia al español (Tesauros).

Una vez confeccionado el corpus, pasaremos a utilizar el DWS *Lexonomy*, una herramienta lexicográfica de acceso abierto y *user-friendly* que facilita la labor de diseñar, compilar y publicar un diccionario creado *ad hoc* por el alumnado. En primer lugar, requiere el registro gratuito por parte de los diferentes grupos en la plataforma. A continuación, con el comando *edit* se deben crear y editar las diferentes entradas lexicográficas y se ha de dar visibilidad a su creación dando los permisos necesarios. Paulativamente, los discentes configurarán el diccionario incluyendo los datos necesarios recogidos en las CODING TAGS.

4.3. PROCEDIMIENTO

Esta investigación comprende dos frases claramente diferenciadas: i) la creación y validación de un corpus de IFE de turismo, específicamente sobre alojamientos turísticos con las opiniones recogidas en la página web *Cozy Cozy* mediante la herramienta *Sketch Engine* y ii) confección y uso de un diccionario digital *ad hoc* con dichas entradas con el DWS *Lexonomy*.

Por ello, en primer lugar, se presentará la actividad a los estudiantes de Inglés I del Grado en Turismo. Tras firmar un consentimiento informado a través de *Google Forms*, se les explicará que recibirán una formación complementaria en varios seminarios que tendrán lugar en un horario diferente al de las horas lectivas de la asignatura en cuestión: recibirán

un taller sobre el uso de la herramienta *Sketch Engine* disponible en <https://www.sketchengine.eu/>. Este recurso permite el análisis de corpus a nivel semántico, morfológico y sintáctico. Se hará uso de las herramientas de concordancias, palabras clave, palabras frecuentes y word Sketch. La selección de los términos se llevará a cabo teniendo en cuenta las palabras clave y la frecuencia de aparición de los términos. Ya que es un software que puede utilizarse en línea, se evitarán así posibles errores técnicos que puedan surgir por la incompatibilidad de diferentes sistemas operativos. Además, *Sketch Engine* permite opciones personalizadas para buscar lemas concretos y mostrar concordancias. La frecuencia y la distribución de los términos son opciones útiles para investigar el uso común de dichos términos y la uniformidad con la que se halla en el corpus.

A continuación, una vez creado el corpus, se formará al alumnado en un segundo taller que versará sobre un DWS de acceso abierto que permite el almacenamiento, gestión y publicación de diccionarios online. Se trata de *Lexonomy*. Tras recibir esta formación específica, los discentes serán capaces de confeccionar y publicar sin problema un diccionario monolingüe en inglés del ámbito de turismo. Para ello, seguirán los principios de la lexicografía orientada al aprendizaje de lenguas.

5. RESULTADOS ESPERADOS

De forma análoga a investigaciones previas en las que se ha confeccionado un corpus para mejorar la adquisición de léxico especializado en diferentes áreas de conocimiento, considerado por muchos investigadores como la pieza más fundamental de los programas de IFE (Nababan, 1993), esperamos resultados satisfactorios. Hay que tener en cuenta que el significado de un término puede variar dependiendo del registro utilizado y la confección, en primer lugar, de un corpus, y, en segundo lugar, de un diccionario *ad hoc* para un campo específico de trabajo (turismo) facilita enormemente la adquisición de la polisemia de dichos términos y su uso adecuado.

Por otro lado, también se espera que la motivación y la participación activa del alumnado incremente con esta experimentación. Poco a poco los discentes serán más conscientes de que son los verdaderos

protagonistas de su aprendizaje y que deben desempeñar un rol activo y responsable en su aprendizaje. El aprendizaje incidental, aquel que tiene lugar a la vez que se confecciona el corpus y el diccionario repercutirá, sin lugar a dudas, en la competencia lingüística de los discentes, pues serán capaces de utilizar la lengua de una forma más correcta haciendo uso de un elenco mayor de vocabulario técnico propio de su campo de estudio. Más aún, la consulta del diccionario creado *ad hoc* facilitará en gran medida su aprendizaje, pues responderá directamente a las necesidades observadas. El hecho de ser un diccionario digital de acceso abierto permitirá su consulta en diferentes dispositivos (teléfono móvil, tablet, ordenador) toda vez que se tenga acceso a internet.

En último lugar, también esperamos que el alumnado trabaje adecuadamente en grupos de trabajo cooperativo. A través de esta metodología, los discentes aprenderán los unos de los otros, aprenderán a aceptar las diferencias y tomar decisiones grupales, respetar los diferentes puntos de vista y mejorar sus capacidades de aprendizaje.

6. DISCUSIÓN

Diferentes investigaciones se han centrado en la confección de un corpus para estudiar vocabulario específico de distintos ámbitos y han defendido su uso en cursos de aprendizaje de lenguas basándose en las necesidades de los estudiantes (Watson, 2001). Como señala Wu (2014), la elaboración de un corpus propio ayuda tanto a los profesores como los alumnos a manejar con más éxito las dificultades de aprendizaje que se pueden encontrar con el vocabulario de IFE. Alentar a los estudiantes a confeccionar su propio corpus resulta ser una técnica de aprendizaje de lenguas efectiva.

Yoon (2005) utilizó satisfactoriamente un corpus para enseñar escritura académica y observó que el empleo de un corpus despertó el interés de los discentes por la gramática de la lengua meta. Gilmore y Millar (2018) estudiaron el potencial pedagógico del uso de corpus de IFE en ingeniería civil. Demostraron que es un acercamiento útil en la enseñanza de IFE pues responde a las necesidades propias de su alumnado.

No obstante, debemos prestar atención a un número de variables, tales como los factores motivacionales, el nivel de la lengua de los discentes o las estrategias de aprendizaje que pueden modular significativamente el aprendizaje de IFE. Los alumnos de IFE con frecuencia muestran una motivación moderada, quizá por las experiencias negativas experimentadas en el aprendizaje de la lengua con fines generales en la anterior etapa educativa (educación secundaria y bachillerato). Por ello, la confección de corpus y de diccionarios digitales que responden a sus necesidades creados *ad hoc* pueden simplificar enormemente la adquisición del léxico especializado y mejorar su motivación, convirtiéndolos en protagonistas de su propio aprendizaje y no meros receptores de la información a aprender. A través de los múltiples ejemplos que se recogerán tanto en el corpus como en el diccionario, los alumnos aprenderán no solamente léxico, sino también estructuras sintácticas y conocimientos morfológicos. Estudiar la formación de palabras ayudará a ahondar en la comprensión del vocabulario y en su ortografía, fundamental en muchos textos especializados en las diferentes ramas de conocimiento. Como señala Ellis (2008), los cursos que incentivan el aprendizaje del léxico, adquirido en gran medida de una forma incidental, arroja mejores resultados.

Lam (2004) hizo un análisis léxico-gramatical de un corpus de la industria turística, que comprendía vocabulario especializado hallado en guías de viaje, información turística y catálogos de viajes de diferentes páginas web. Proporcionó los diferentes campos semánticos de este campo de estudio y contextualizó algunos términos mediante concordancias, así como ejemplificando patrones de colocaciones y prosodias semánticas. Describió los beneficios de utilizar el corpus en el aprendizaje del inglés con fines turísticos y exaltó la relevancia de su aprendizaje mediante los principios de la lingüística de corpus y la lexicografía.

7. CONCLUSIONES

Con esta propuesta didáctica pretendemos facilitar el aprendizaje de ILE de los alumnos del grado de Turismo. Con frecuencia hemos venido observando la dificultad para adquirir realmente vocabulario propio de su

campo de trabajo y consideramos que la elaboración de un corpus y de un diccionario digital abierto por ellos mismos ayudará significativamente a mejorar la adquisición de la lengua inglesa. Por consiguiente, pretendemos ofrecer una educación de calidad y capacitar a los alumnos para su plena y satisfactoria inserción en el mundo laboral en un futuro cercano, mejorando no solo su competencia lingüística sino también digital, competencia también clave en la educación actual, siguiendo así los preceptos establecidos por los Objetivos de Desarrollo Sostenible de la Agenda 2030 (OCDE, 2021).

A través de esta práctica esperamos también una mayor motivación e implicación del alumnado en su proceso de aprendizaje de la lengua inglesa. Los alumnos se han de sentir los plenos protagonistas de su aprendizaje y han de ser conscientes de su papel activo y único en dicho proceso. Esperamos arrojar datos satisfactorios una vez haya concluido y hayamos analizado los datos recogidos, mostrando pruebas objetivas de la adquisición del léxico específico del ámbito de los alojamientos turísticos a la par que mejorar su utilización de la lengua, empleando estructuras sintáctico-discursivas más complejas que faciliten la comunicación en su campo de trabajo.

Esta propuesta comprende ciertas limitaciones, como son, en primer lugar, el léxico del corpus, limitado al vocabulario relacionado con los alojamientos turísticos. Por ello, aunque el objetivo lexicográfico no es crear un diccionario completo para el aprendizaje de IFE de turismo, entendemos que es un primer paso para la confección futura de dicho diccionario. En futuras experimentaciones, dado que el diccionario será de acceso abierto y público, pretendemos aumentar el número de lemas introduciendo otros ámbitos del turismo (e. g., restauración, transporte, actividades de ocio, etc.) haciendo uso de otras plataformas que también proporcionen la información y los comentarios facilitados por sus usuarios/clientes.

Por otro lado, consideramos que sería conveniente confeccionar un corpus similar pero utilizando otros destinos turísticos en diferentes países anglófonos. De este modo, quizá podamos enriquecer el corpus con expresiones propias de diferentes variedades de la lengua.

8. AGRADECIMIENTOS/APOYOS

Agradecemos a LALINGAP, grupo de investigación de la Universidad Complutense de Madrid al que pertenecemos, su inestimable contribución al desarrollo de la investigación sobre el aprendizaje y enseñanza de lenguas, parte de cuyos resultados se plasman en el presente capítulo.

9. REFERENCIAS

- Abel, A. (2012). Dictionary writing systems and beyond. En S. Granger y M. Paquot (Eds.), *Electronic Lexicography*, (pp. 83-106). Oxford University Press.
- Asher, R. E. y Simpson, J. M. Y. (1994). *The Encyclopedia of Language and Linguistics*. Pergamon.
- Atkins, B. T. S. y Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. Oxford University Press.
- Biber, D., Conrad, S. y Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.
- Boulton, A. (2012). Corpus consultation for ESP: A review of empirical research. En A. Boulton, S. Carter-Thomas y E. Rowley-Jolivet (Eds.), *Corpus-Informed Research and Learning in ESP*, (pp. 261-292). John Benjamins Publishing Company.
- Costa, H., Corpas, G., Seghiri, M. y Mitkov, R. (2015). iCorpora: Compiling, Managing and Exploring Multilingual Data. En G. Corpas, M. Seghiri, R. Gutiérrez, M. Urbano (Eds.), *Nuevos horizontes en los Estudios de Traducción e Interpretación / New Horizons in Translation and Interpreting Studies*, (pp. 74-76). Tradulex.
- Del Olmo, E. y Arias, I. (2021). Un estudio empírico con Sketch Engine sobre la interfaz sintáctico-ática para la identificación de la estructura temática intraoracional en español. *Revista De Humanidades Digitales*, 6, 129–150. <https://doi.org/10.5944/rhd.vol.6.2021.30965>
- Ellis, R. (2008). *The study of second language acquisition*. Oxford University Press.
- Gilmore, A. y Millar, N. (2018). The language of civil engineering research articles: A corpus-based approach. *English for Specific Purposes*, 51, 1-17. <https://doi.org/10.1016/j.esp.2018.02.002>
- Gries, S. T. (2006). Introduction. En S. T. Gries y A. Stefanowitsch, (Eds.), *Corpus in Cognitive Linguistics: Corpus-Based Approaches to Syntax and Lexis*, (1–17). Mouton.

- Guilquin, G. y Gries, S. T. (2009). Corpora and Experimental Methods: A State of-the-Art Review. *Corpus Linguistics and Linguistic Theory*, 5 (1), 1–26. <http://dx.doi.org/10.1515/CLLT.2009.001>
- Hernández López, M. O. (2019). What makes a positive experience? Offline/online communication and rapport management in Airbnb positive reviews. *Pragmatics and Society*, 10 (2), 177-204. <http://dx.doi.org/10.1075/ps.16060.ohe>
- Hernández Toribio, M. I. y Mariottini, L. (2015). Actos de habla y atenuación 2.0: Tripadvisor. *Círculo de Lingüística Aplicada a la Comunicación*, 73, 15-32. <https://doi.org/10.5209/CLAC.59057>
- Heuberger, R. (2016). Learners' Dictionaries. En P. Durkin (Ed.), *The Oxford Handbook of Lexicography*, (pp. 25-43). Oxford University Press.
- Hsu, W. (2013). Bridging the vocabulary gap for EFL medical undergraduates: The establishment of a medical word list. *Language Teaching Research*, 17 (4), 454-484. <https://doi.org/10.1177/1362168813494121>
- Kallas, J., Koeva, S., Kosem, I., Langemets, M. y Tiberius, C. (2019). ELEXIS D1.1. Lexicographic Practices in Europe: A Survey of User Needs. https://elex.is/wp-content/uploads/2020/06/Revised-ELEXIS_D1.1_Lexicographic_Practices_in_Europe_A_Survey_of_User_Needs.pdf
- Kennedy, G. (1998). *An introduction to corpus linguistics*. Longman.
- Lam, P. Y. W. (2004). A corpus-driven lexico-grammatical analysis of English tourism industry texts and the study of its pedagogic implications in English for Specific Purposes. In E. Hidalgo, L. Quereda, y J. Santana, *Corpora in the Foreign Language Classroom*, (pp. 71-90). Brill.
- Leech, G. (1992). Corpus and theories of linguistic performance. En J. Svartvik (Ed.), *Directions in Linguistics: Proceedings of Nobel Symposium 82* (pp. 105-122). Mouton de Gruyter.
- Lei, L. y Liu, D. (2016). A new medical academic word list: A corpus-based study with enhanced methodology. *Journal of English for Academic Purposes*, 22, 42-53. <https://doi.org/10.1016/j.jeap.2016.01.008>
- Mancera, A. (2018). La atenuación lingüística en las reseñas digitales de hoteles y restaurantes en español. *Círculo de Lingüística Aplicada a la Comunicación*, 73, 53-76. <https://doi.org/10.5209/CLAC.59059>
- Mancera, A. y Pano, A. (2021). La opinión pública en la red. Análisis pragmático de la voz de los ciudadanos. *Iberoamericana*.
- McEnery, T. y Wilson, A. (1996). *Corpus Linguistics*. Edinburgh University Press.

- McEnery, Xiao, R. y Tono, Y. (2006). *Corpus-Based Language Studies*. Routledge.
- Nababan, P. W. J. (1993). *E.S.P. Materials Preparation in a Foreign Language Situation*. ERIC Digest.
- Organización para la Cooperación y el Desarrollo Económico (OCDE). *OECD Skills Outlook 2021: Learning for life*. OECD Publishing.
<https://doi.org/10.1787/0ae365b4-en>
- Parodi, G. (2010). *Lingüística de corpus: De la teoría a la empiria*. Iberoamericana Vervuert.
- Rojo, G. (2021). *Introducción a la lingüística de corpus en español*. Routledge.
- Rubio, R. J., Bonilla, J. E. y Bernal, J. A. (2021). Dictionary Writing Systems y otras herramientas informáticas para la elaboración, administración y publicación de diccionarios. *Lingüística y Literatura*, 80, 340-360.
<https://doi.org/10.17533/udea.lyl.n80a20>
- Stubbs, M. (1996). *Text and corpus analysis. Computer-assisted studies of language and culture*. Blackwell.
- Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. Blackwell.
- Svartvik, J. (1992). Corpus linguistics comes of age. En J. Svartvik (Ed.) *Directions in Corpus Linguistics. Proceedings of Nobel Symposium 82*, (pp. 7-13). Mouton de Gruyter.
- Tarp, S. (2019). La ventana al futuro: Despidiéndose de los diccionarios para abrazar a la Lexicografía, RILEX. *Revista sobre investigaciones léxicas*, 2 (II), pp. 5-36. <https://doi.org/10.17561/rilex.v2.n2>
- Toffler, A. (1980). *The third wave*. William Morrow.
- Tognini- Bonelli, E. (2001). *Corpus linguistics at work*. Benjamins.
- Toledo, M. C. y Camero, A. (2021). Compilación semiautomática con Sketch Engine de un corpus ad hoc comparable bilingüe (inglés-español) de contratos de distribución exclusiva (DISTRIBUCOR). En E. Sartor (Ed.), *Los corpus especializados en la lingüística aplicada: enseñanza y traducción*, (pp. 15-41). Universitas Studiorum.
- Vásquez, C. (2011). Complaints online: The case of TripAdvisor. *Journal of Pragmatics*, 43 (6), 1707-1717.
<https://doi.org/10.1016/j.pragma.2010.11.007>
- Vásquez, C. (2012). Narrativity and involvement in online consumer reviews. The case of TripAdvisor. *Narrative Inquiry*, 22 (1), 105-121.
<https://doi.org/10.1075/ni.22.1.07vas>

- Villayandre, M. (2008). Lingüística con corpus (I). *Estudios Humanísticos. Filología*, 30, 329–349. <https://doi.org/10.18002/ehf.v0i30.2847>
- Wang, J., Liang, S. y Ge, G. (2008). Establishment of a medical academic word list. *English for Specific Purposes*, 27 (4), 442-458. <https://doi.org/10.1016/j.esp.2008.05.003>
- Watson, T. R. (2001). Building and using your won corpus and concordance. *Thai TESOL Bulletin*, 14 (2). <http://arts.kmutt.ac.th/call/doc/Building%20and%20Using%20Your%20Own%20Corpus%20and%20Concordance.pdf>
- Wu, L. F. (2014). Motivating college students' learning English for Specific Purposes courses through corpus building. *English Language Teaching*, 7 (6), 120-127. <http://dx.doi.org/10.5539/elt.v7n6p120>
- Yoon, H. (2005). An investigation of students' experiences with corpus technology in second language academic writing. [Tesis Doctoral, The Ohio State University]. Repositorio Institucional OSU. <https://www.learntechlib.org/p/120367/>
- Zhang, Y. y Vásquez, C. (2014). Hotel responses to online reviews: Managing consumer dissatisfaction. *Discourse, Context and Media*, 6, 54-64. <http://dx.doi.org/10.1016/j.dcm.2014.08.004>