USO DEL CORPUS SEAH PARA LA ENSEÑANZA DEL ESPAÑOL DE LA ARQUITECTURA Y CONSTRUCCIÓN[1]

SARA PICCIONI
Università degli Studi "G. D'Annunzio" Chieti-Pescara
MARCELLA SOLINAS
Università degli Studi "G. D'Annunzio" Chieti-Pescara
VALENTINA CASTRIGNANÒ
Università Ca' Foscari Venezia

Introducción

En el marco del Programa Erasmus +K023 financiado por la Unión Europea, en 2020 la Universidad "Gabriele D'Annunzio" de Chieti-Pescara ha inaugurado —en colaboración con otras universidades europeas— un proyecto multilingüe (español, francés, alemán, italiano, ruso) bajo el nombre de SEAH[2] (Sharing Europen Architectural Heritage: Innovative Language Teaching Tools for Academic and Professional Mobility in Architecture and Construction) para fomentar, principalmente en contextos académicos, el aprendizaje de la lengua extranjera en el sector de la Arquitectura y la Construcción a través del desarrollo de recursos lingüísticos innovadores (D'Angelo, 2021).

Para responder a las directrices de la Unión Europea que apuntan a promover la diversidad lingüística, a impulsar la movilidad internacional de los ciudadanos para la implementación de un diálogo intercultural y la valorización del patrimonio multilingüe y multicultural de cada uno de los Estados Miembros, se hace necesario realizar una programación glotodidáctica ponderada, capaz de facilitar el éxito de los proyectos de movilidad y reducir los obstáculos que, en ocasiones, dificultan los intercambios universitarios o condicionan la adquisición de conocimientos técnicos y profesionales adecuados al nivel científico requerido. A menudo los estudiantes de movilidad no poseen un conocimiento de la lengua extranjera adecuado para llevar a cabo las tareas propias de un programa de estudio superior, programa que supone el manejo no solo de la lengua estándar, sino también de un lenguaje técnico-científico especializado. El uso exclusivo del inglés se revela insuficiente para responder a las exigencias formativas de estos estudiantes, quienes —debido a las peculiaridades de determinadas tradiciones científico-culturales, prácticas profesionales o marcos jurídicos nacionales— no siempre pueden utilizarlo como lengua franca y necesitan, por lo tanto, adquirir competencias especializadas en la lengua del país de llegada. A este propósito se hace fundamental proporcionar a los estudiantes programaciones y herramientas didácticas que hagan hincapié en las reales exigencias formativas de los participantes en los proyectos de intercambio y que, además, favorezcan e implementen el plurilingüismo en la comunicación científico-académica.

Dadas estas premisas, el proyecto SEAH ha tenido como objetivo principal la creación de una serie de módulos lingüísticos disponibles *online* y accesibles en el sitio SEAH[3] a través de una *open educational resources platform* para el autoaprendizaje del lenguaje académico en el sector de la arquitectura y construcción (AC de aquí en adelante) en los cinco idiomas del proyecto. Como paso preliminar para la

[1] El presente artículo es fruto del trabajo conjunto de las autoras. Por razones de reconocimiento académico, se precisa que los § "Introducción", 2 y 6 corren a cargo de Marcella Solinas, los § 1, 3, 4 y "Conclusiones" están redactados por Sara Piccioni, mientras que el § 5 es de Valentina Castrignanò.

[2] El proyecto SEAH es coordinado por Mariapia D'Angelo; Sara Piccioni y Maria Chiara Ferro coordinan respectivamente las fases relativas al desarrollo de corpus especializados (Intellectual Output 1 - Language Corpora Design, Creation and Distribution - French, German, Italian, Russian and Spanish for Architecture and Construction) y de la creación de los módulos de autoaprendizaje lingüístico (Intellectual Output 2 - Developing and Testing Innovative Language Learning Materials - French, German, Italian, Russian and Spanish for Architecture and Construction). Participan en el proyecto la Università "G. D'Annunzio" Chieti-Pescara (Italia, universidad coordinadora), la Université Bordeaux Montaigne (Francia), la Universidad Politécnica de Madrid (España), la Polotsk State University (Bielorrusia) y la Masarykova University (República Checa).

[3] https://www.seahproject.eu/index.php [consulta: 22/11/2022].

realización de los módulos didácticos se ha compilado un corpus multilingüe especializado que incluye textos relativos a distintos dominios (académico, didáctico, profesional y divulgativo) en el campo de la AC (Piccioni, D'Angelo y Ferro, 2021). Este recurso se ha utilizado tanto para identificar los rasgos propios del lenguaje de la AC útiles para las necesidades formativas de los estudiantes de intercambio, como para crear actividades didácticas basadas en corpus encaminadas a familiarizar a los estudiantes con el uso de los corpus para el aprendizaje autónomo de una lengua extranjera.

En este marco, el presente trabajo pretende, por un lado, ilustrar las premisas teóricas y metodológicas en que se fundamenta el uso de corpus en la enseñanza/aprendizaje de las lenguas de especialidad (§ 1 y 2) y su incorporación en el proyecto SEAH (§ 3 y 4), y, por otro, ilustrar ejemplos de uso didáctico del corpus ES_SEAH para el aprendizaje de la lengua española de la AC (§ 5) y la didáctica de la traducción especializada (§ 6).

1. Corpus lingüísticos y didáctica de LE/L2

La incorporación de metodologías de corpus en la enseñanza de LE/L2 se enmarca en el giro empírico que ha interesado a la investigación lingüística en décadas recientes y responde a un nuevo interés por el uso de datos lingüísticos auténticos y métodos inductivos en el aprendizaje lingüístico. En este ámbito, se han identificado usos indirectos y directos de los corpus lingüísticos (Leech, 1997; McEnery y Xiao, 2011). Los primeros se refieren a la explotación —en ámbito pedagógico y editorial— de datos lingüísticos extraídos de corpus para el diseño de currículos y la realización de manuales didácticos y obras de referencia: piénsese, p. ej., en el pionero diccionario *COBUILD* (HarperCollins, 1995) o en los manuales y obras de consulta de la Cambridge University Press, realizados a partir del Cambridge English Corpus[4]. En estos entornos, los corpus permiten la identificación de las formas léxico-gramaticales más frecuentes en hablantes nativos y de los errores más comunes entre los aprendices no nativos, brindando criterios empíricos para la selección y secuenciación de temas y actividades en los programas de enseñanza y en los materiales pedagógicos creados.

Los usos directos de los corpus, en cambio, se despliegan en las tres modalidades identificadas por Leech (1997) y denominadas: 1) *teaching about*, referido a la enseñanza de la lingüística de corpus como asignatura académica a la par que otras disciplinas, como la sintaxis o la pragmática; 2) *teaching to exploit*, es decir, proporcionar a los estudiantes experiencia práctica en el uso de corpus para explorar los aspectos lingüísticos de interés de forma autónoma o guiada (McEnery, Xiao y Tono, 2006); y 3) *exploiting to teach*, que correspondería al uso de métodos basados en corpus en cursos de lingüística que normalmente se darían sin el auxilio de corpus, p. ej., sociolingüística o análisis del discurso.

De estas tres modalidades, la que más se ajusta a los objetivos del presente trabajo y del proyecto SEAH es la segunda (*teaching to exploit*), que promueve el llamado "aprendizaje guiado por datos" (*data-driven learning* o DDL, Johns, 1991) o *discovery learning* (p. ej., Gavioli, 2001; Huang, 2011) y se ha tradicionalmente aplicado en entornos de enseñanza/aprendizaje de lenguas para fines académicos y profesionales. Se trata de enfoques didácticos centrados en el aprendiz y en sus exigencias, que se desarrollan en tres fases sucesivas de aprendizaje inductivo, que parten de la observación (de concordancias u otras evidencias lingüísticas) para pasar a la clasificación (de rasgos salientes de los usos lingüísticos observados) y terminan con la generalización de las reglas extrapoladas (Johns, 1991). Estas tres fases —correspondientes aproximadamente a las "tres íes" (*illustration*, *interaction*, *induction*) del modelo inductivo de Carter y McCarthy (1995)— promoverían un aprendizaje más eficaz y duradero en tanto que basado en el razonamiento del propio aprendiz, al mismo tiempo que fomentarían su autonomía tanto en la identificación de problemas y dudas como en la formulación de soluciones y respuestas.

Frente a las ventajas hasta aquí señaladas, no se pueden pasar por alto algunas limitaciones importantes de estos enfoques, que han obstaculizado su adopción en contextos escolares y universitarios. Entre estos, mencionamos las dificultades relacionadas con la creación o consulta de corpus adecuados para los objetivos de aprendizaje propios de cada entorno, la necesidad de una

[4] No es de acceso público, pero para más información acerca del corpus *vid.* https://www.cambridge.es/en/about-us/cambridge-english-corpus [consulta: 22/11/2022].

formación específica para los profesores (Abad Castelló, 2019) y el acceso desigual (según los contextos) a ordenadores, *software* dedicado y conexiones a internet de alta velocidad. En cuanto a las desventajas de tipo pedagógico y cognitivo, se señala la falta de contexto situacional de los patrones lingüísticos observados que —en los programas de concordancias— se visualizan de forma fragmentaria y descontextualizada dificultando o, incluso, impidiendo los procesos inductivos necesarios para la identificación de las reglas (p. ej., Sinclair, 1991, p. 34; McEnery y Wilson, 1996, p. 79). Además, los procesos que implican la observación, la comprensión y el análisis de una gran cantidad de datos lingüísticos pueden suponerle al aprendiz un coste desproporcionadamente alto en términos de esfuerzo y tiempo frente a los beneficios de aprendizaje (Thompson, 2006), sin garantías de llegar a conclusiones correctas (Boulton, 2016). La misma disponibilidad de textos auténticos —generalmente considerada una ventaja— puede resultar contraproducente sobre todo para el aprendiz con un nivel de lengua no suficientemente alto, que se beneficiaría más de textos graduados según su nivel o simplificados (Flowerdew, 2009).

2. Corpus y didáctica del español

En el mundo hispano el empleo de los corpus en las clases de ELE es muy reciente y resulta marginal en la práctica docente. Las primeras propuestas para aprovechar los corpus en la didáctica ELE se remontan a principios de los años 2000, cuando una serie de autores (p. ej., Izquierdo Gil, 2002; San Mateo, 2003; Alonso Pérez-Ávila, 2007; Campillos, Gozalo y Moreno, 2007; Buyse, 2011) empezaron a remarcar la eficacia de la enseñanza basada en datos.

Sin embargo, hay que constatar que hasta hoy el uso de los corpus en la didáctica del español como LE/L2 sigue siendo más bien minoritario y fragmentado. Las motivaciones de esta laguna residen a menudo, como ya se ha mencionado arriba, en la falta de formación especializada de los docentes en el uso pedagógico de los corpus (Abad Castelló, 2019), que obstaculiza la explotación didáctica de toda la información que puede ofrecer un corpus y en el largo tiempo de preparación de dichas actividades. De ahí la necesidad de implementar la investigación didáctica en este campo para enriquecer la publicación de material didáctico basado en corpus listo para el uso.

A pesar de estas dificultades, en los últimos años, gracias también a la difusión cada vez más generalizada de las TIC en las aulas, el interés por el aprovechamiento de los corpus en la didáctica del español L2/LE se ha incrementado mucho, como demuestran las crecientes publicaciones sobre el tema (*vid.* Cruz Piñol, 2017; Brown, Paz y Brown, 2021; Elvira-García, 2021; Sartor, 2021), la mayor presencia de paneles dedicados a la lingüística de corpus en los congresos internacionales —ASELE es un ejemplo emblemático de ello— y el creciente espacio que la didáctica de corpus va ganando en ámbito universitario.

Actualmente la explotación de los corpus en la didáctica ELE se centra sobre todo en actividades que favorecen la consulta espontánea en clase para solucionar dudas lingüísticas a través de la búsqueda de ocurrencias y concordancias en textos auténticos, en ejercicios de investigación léxica (frecuencias), en la observación repetida de casos reales de la lengua que permitan al estudiante extraer reglas de manera inductiva y en todas aquellas actividades que tienen como objetivo impulsar la conciencia metalingüística de los aprendices[5].

Algunos de los corpus de español más famosos y utilizados por los profesores en la enseñanza de ELE gracias a su accesibilidad en línea son: el Corpus Diacrónico del Español (CORDE)[6], el Corpus de Referencia del Español Actual (CREA)[7] y el Corpus del Español del Siglo XXI (CORPES XXI)[8] de la Real Academia Española, el Corpus del Español de Mark Davies[9] y los corpus de aprendices, como el

[5] Muestras recientes de material didáctico ELE creado a partir de los corpus se pueden encontrar en Paz (2013), Benavides (2015), Cruz Piñol (2017), Abad Castelló (2019), Brown, Paz y Brown (2021) o Pastor Cesteros (2023).
[6] https://corpus.rae.es/cordenet.html [consulta: 22/11/2022].
[7] https://corpus.rae.es/creanet.html [consulta: 22/11/2022].
[8] https://www.rae.es/banco-de-datos/corpes-xxi [consulta: 22/11/2022].
[9] https://www.corpusdelespanol.org/ [consulta: 22/11/2022].

CEDEL2[10], que recogen textos de aprendientes de la lengua y son utilizados principalmente para trabajar la interlengua y el análisis de los errores[11].

Relativamente a la enseñanza de las lenguas para fines específicos, la aplicación de corpus a la didáctica resulta una herramienta muy valiosa para observar, entre otras cosas, el léxico funcional, las colocaciones y las unidades terminológicas en un género textual concreto. En el sector de los corpus especializados señalamos —como muestra— tres ejemplos: el corpus multilingüe GENTT, que recopila géneros jurídicos, médicos y técnicos (García Izquierdo, 2005); el Corpus PUCV-2006 del Español (Parodi, Ibáñez y Venegas, 2009), de géneros académicos y profesionales, y el corpus WebLesp[12] (Piccioni y Pontrandolfo, 2021), de comunicación especializada y divulgativa en los ámbitos de la medicina, el derecho, la economía y las ciencias medioambientales. Hoy podemos añadir a estos importantes recursos también el corpus multilingüe para la AC ES_SEAH[13].

3. El corpus multilingüe SEAH

Como ya se ha indicado en los § "Introducción" y 1, el objetivo principal del proyecto SEAH ha sido la creación de cursos en línea[14] para el autoaprendizaje del lenguaje académico-profesional de la AC en alemán, español, francés, italiano y ruso. Además de aprovechar varios recursos multimedia para ofrecer actividades encaminadas a la adquisición de los principales aspectos léxico-gramaticales del lenguaje de la AC, los módulos didácticos creados se basan en datos lingüísticos extraídos de corpus electrónicos creados *ad hoc* para las finalidades de SEAH.

De hecho, en el marco del proyecto se compiló un corpus multilingüe especializado de textos representativos de algunos subdominios del campo de la AC, que se utilizó posteriormente para la creación de materiales didácticos. El desarrollo de los cursos de lenguas SEAH se basa tanto en usos indirectos como directos de los corpus (*cf.* § 1). Los usos indirectos han permitido, por un lado, la selección de los rasgos léxico-gramaticales que se iban a incluir en las actividades didácticas propuestas, a partir de consideraciones relativas a su frecuencia en los corpus; por otro lado, se han explotado para la extracción de oraciones y textos para el desarrollo de ejercicios y actividades de comprensión lectora y auditiva basados en material lingüístico auténtico.

Los usos directos de los corpus (en la ya mencionada modalidad *teaching to exploit*) han dado pie a la elaboración de actividades en las que a los alumnos se les propone utilizar los corpus de primera mano con el doble objetivo de guiarles en el aprendizaje de estructuras lingüísticas o elementos léxicos concretos, así como de dotarles de las habilidades necesarias para utilizar los corpus de forma autónoma para resolver dudas concretas y explorar fenómenos lingüísticos de interés.

La construcción de los cinco corpus SEAH (alemán, español, francés, italiano y ruso) se llevó a cabo después de una consulta inicial con las partes interesadas del proyecto (docentes de los departamentos de AC, profesionales del ámbito de la AC y profesores de idiomas que se dedican a la enseñanza del lenguaje especializado de la AC), que permitió identificar: (1) las áreas temáticas de los textos que se iban a incluir en los corpus (seleccionándolas entre aquellas a las que están expuestos principalmente los estudiantes de AC que participan en programas de movilidad); (2) tipo de habilidades lingüísticas y comunicativas necesarias (comprensión/producción escrita/oral); (3) géneros discursivos que los estudiantes y/o profesionales en movilidad deben saber producir o comprender en una lengua extranjera.

Las aportaciones de los expertos consultados definieron la arquitectura general del corpus, que se basa en una categorización multinivel de los textos seleccionados según su canal de recepción, el dominio en el que se inscriben y el género discursivo. En su configuración final, los corpus constan de alrededor de 1,2 millones de *tokens* cada uno (para un total de más de 6 millones de *tokens*), conformados por textos escritos y transcripciones de textos orales (en una proporción del 80 % y 20 % aproximadamente),

[10] http://cedel2.learnercorpora.com/ [consulta: 22/11/2022].
[11] Para una reseña exhaustiva de los diferentes recursos electrónicos relacionados con la lingüística de corpus *vid.* Cruz Piñol (2017).
[12] https://corpora.unich.it/weblesp/#open [consulta: 22/11/2022].
[13] https://www.seahproject.eu/resources.php [consulta: 22/11/2022].
[14] Los módulos son accesibles en https://www.seahproject.eu/learning.php [consulta: 22/11/2022].

pertenecientes a cuatro dominios (académico, didáctico, profesional y divulgativo) y a 22 géneros discursivos distintos (*cf.* Piccioni, D'Angelo y Ferro, 2021 para mayor detalle).

Los corpus, lematizados y anotados lingüísticamente, están disponibles de forma abierta[15] a través de la plataforma NoSketchEngine (Rychlý, 2007), que permite generar listas de palabras, concordancias, colocaciones y palabras clave.

En los tres apartados que siguen se ilustran de forma más detenida la arquitectura del corpus de lengua española (ES_SEAH, § 4) y ejemplos de sus usos directos para la realización de actividades para la didáctica de la lengua española (§ 5) y de la traducción especializada español-italiano (§ 6).

4. El corpus ES_SEAH

El corpus ES_SEAH se centra en cinco áreas temáticas correspondientes a los dominios léxicos cubiertos en los módulos de lengua desarrollados en la segunda fase del proyecto: (1) Rehabilitación, (2) Edificaciones y Obras de Ingeniería Civil, (3) Edificios No Residenciales, (4) Edificios Residenciales y (5) Materiales de Construcción.

Las áreas temáticas están representadas por textos pertenecientes a los siguientes géneros y dominios:

Tabla 1 - Composición del corpus ES_SEAH.

Canal	Dominio	Género discursivo	*Tokens*
Escrito	Académico	Monografías	65 910
	Académico	Artículos científicos	5118
	Académico	Tesis	87 573
	Didáctico	Material didáctico	10 310
	Didáctico	Enunciados de proyecto	165 049
	Profesional	Descripción de proyecto	345 928
	Profesional	Informe de proyecto	178 602
	Profesional	Descripción de materiales	23 967
Total Escrito			882 457
Oral	Académico	Defensa de tesis	13 532
	Divulgativo	Podcast	5005
	Divulgativo	Conferencias públicas	8544
	Profesional	Presentaciones de proyectos	140 513
Total Oral			167 594
Total ES_SEAH			1 050 051

Como muestran la Tabla 1 y la Imagen 1, el corpus consta de un 84 % de textos escritos y un 16 % de textos orales. En cuanto a la distribución de dominios y géneros, el corpus escrito está compuesto principalmente por textos profesionales (62 %), en particular descripciones de proyectos, mientras que el dominio académico (monografías y artículos científicos) y didáctico (enunciados de proyectos y material didáctico) ocupan el 18 % y el 20 % respectivamente. La sección oral consiste casi en su totalidad de textos profesionales (84 %, presentaciones de proyectos), con proporciones residuales de textos académicos (8 %, defensas de tesis) y divulgativos (8 %, podcasts y conferencias públicas).

[15] Los corpus se pueden consultar en la página https://corpus.unich.it/sito/seah-corpus-it.html [consulta: 22/11/2022].

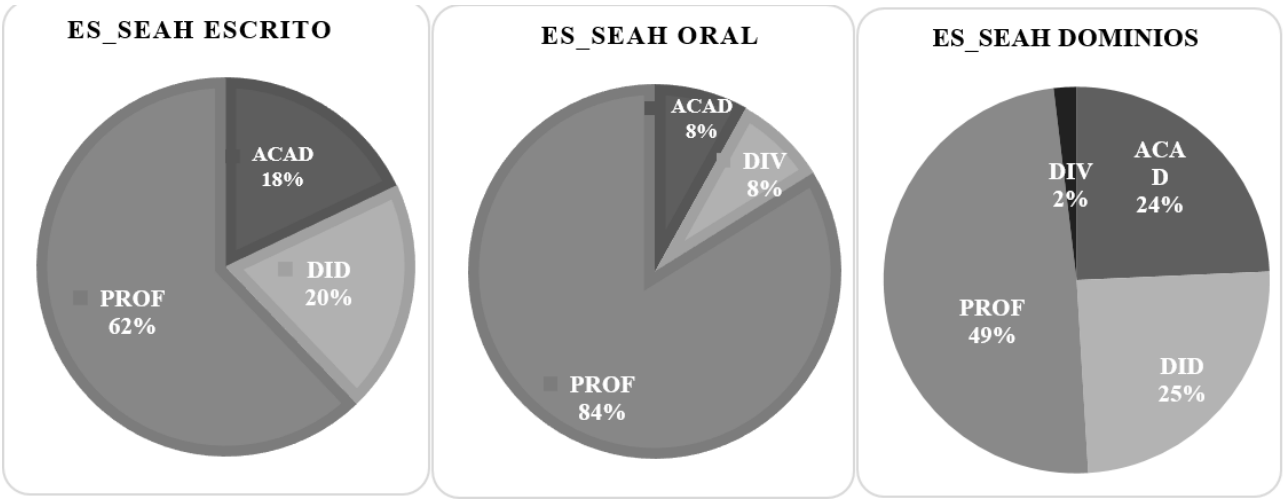


Imagen 1 - Distribución de dominios en las secciones escrita y oral y en el conjunto del corpus ES_SEAH.

En su conjunto, casi la mitad del corpus (49 %) está compuesta por textos profesionales, mientras que el resto se reparte casi perfectamente entre los dominios académico (24 %) y didáctico (25 %). La proporción de textos divulgativos es mínima (2 %), respondiendo así a las necesidades de aprendizaje de los estudiantes de movilidad que, como futuros profesionales en formación, requieren habilidades lingüísticas en la comprensión y producción de textos profesionales y académicos (para adquirir los contenidos especializados de su área de estudio), así como la capacidad de gestionar material didáctico para participar con éxito en los cursos de la universidad de destino.

5. Usos del corpus ES_SEAH para el aprendizaje del español de la AC

En el presente apartado se ilustran de forma más detallada los usos directos del corpus ES_SEAH para la realización de actividades didácticas para la lengua española de la AC.

Las actividades presentadas se encuentran en la sección "Mi glosario" de los módulos lingüísticos SEAH, que consta de un documento PDF editable en el que se proponen unos ejercicios basados en corpus, cuyo objetivo es ayudar a los aprendices a familiarizarse con esta herramienta y guiarles tanto en la exploración y el enriquecimiento del léxico especializado como en la profundización de algunos aspectos morfosintácticos particularmente productivos en el lenguaje de la AC (p. ej., la construcción de verbos de régimen preposicional, las colocaciones sustantivo + adjetivo o verbo + adverbio, los procesos morfológicos de formación de palabras).

Concretamente, los ejemplos que se ilustran a continuación corresponden a las actividades propuestas en la sección "Mi glosario" de las Unidades 2 y 3 (Las Ciudades y sus Infraestructuras[16] y Rehabilitación de Edificios Antiguos[17]). Empezando por el análisis de aspectos léxicos y morfológicos, se ha aprovechado la anotación morfosintáctica del corpus (que permite extraer automáticamente expresiones correspondientes a distintas categorías morfosintácticas, p. ej., sustantivos, adjetivos, verbos, etc.), para proponer una tipología de ejercicios que guía a los alumnos en la exploración de los mecanismos de formación de palabras propios de la lengua española. Además, al centrarse en mecanismos muy productivos en el vocabulario de la AC, las actividades fomentan la autonomía de los aprendices, animándolos a utilizar el corpus de forma independiente para ampliar su propio vocabulario especializado.

La primera actividad que se propone (*cf.* ES_SEAH, "Mi glosario", Unidad 3) se centra en el mecanismo neológico de la derivación: se pide a los alumnos que busquen los adjetivos que terminan por el sufijo *-al* en el corpus ES_SEAH. Para la realización de esta actividad van a utilizar la función "Wordlist", siguiendo las instrucciones ilustradas a continuación (imágenes 2 y 3):

[16] Accesible en https://www.seahproject.eu/see-training.php?id=20 [consulta: 22/11/2022].
[17] Accesible en https://www.seahproject.eu/see-training.php?id=27 [consulta: 22/11/2022].

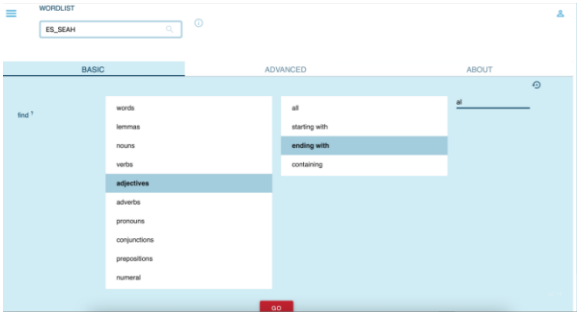


Imagen 2 - Instrucciones para buscar los adjetivos terminados en *-al* en ES_SEAH.

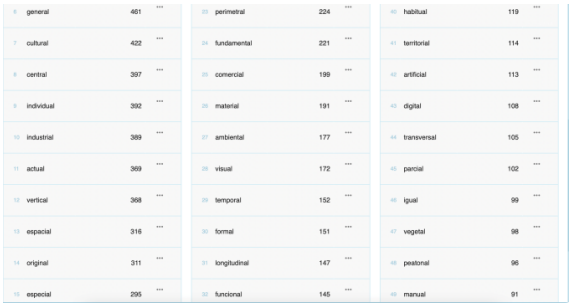


Imagen 3 - Lista de adjetivos que terminan por *-al* en el corpus ES_SEAH.

Después de visualizar y analizar la lista de adjetivos, se les pide a los alumnos que identifiquen aquellos que indican una posición en un espacio (p. ej., *central, vertical, perimetral, longitudinal*, etc.).

La resolución de este ejercicio les permitirá profundizar ciertos aspectos conceptuales (que se integrarán con los meramente lingüísticos), que no sólo son necesarios para ampliar sus conocimientos disciplinares, sino que les ayudarán en el aprendizaje de la lengua extranjera precisamente a partir del corpus. La misma actividad se ha propuesto para la exploración de otras unidades léxicas, como los adjetivos que terminan por *-ble* o los sustantivos que terminan por *-ía*. En estos casos concretos, los alumnos, siguiendo las instrucciones ilustradas anteriormente, realizarán su búsqueda de forma totalmente autónoma, para descubrir nuevas palabras que les permitan ampliar su vocabulario.

Otro ejercicio muy eficaz para explorar los mecanismos de formación de palabras es aquel en el que se les pide a los aprendices crear respectivamente el sustantivo y el adjetivo correspondientes a un verbo dado en infinitivo (*cf.* ES_SEAH, "Mi glosario", Unidad 3). En esta actividad, tienen que utilizar la función "Concordancias". Escribiendo en la sección de búsqueda la raíz de la palabra que se quiere analizar seguida por un asterisco, se pueden observar todas las diferentes palabras que empiezan o terminan por esa misma raíz, como se ilustra a continuación (Imagen 4):

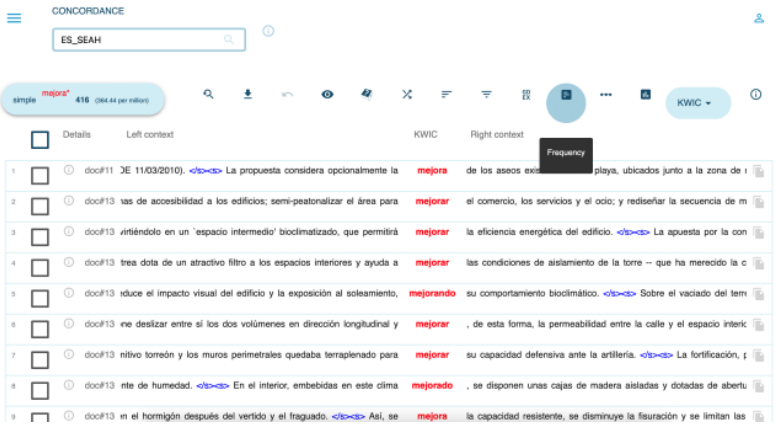


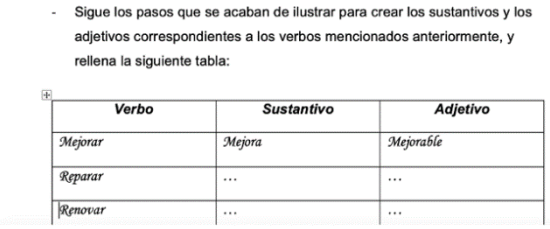
Imagen 4 - Concordancias de las palabras que empiezan con la raíz *mejor-*

Después de visualizar las concordancias, se les pide a los alumnos que seleccionen "Frequency" y después "WordForms" e identifiquen el sustantivo y el adjetivo correspondientes (Imagen 5).



Imagen 5 - Listas de palabras que empiezan con la raíz *mejor-*

Por último, para fomentar el aprendizaje autónomo, se les pide que repitan el mismo procedimiento para explorar otros verbos relacionados con la rehabilitación de edificios (Imagen 6):


- Sigue los pasos que se acaban de ilustrar para crear los sustantivos y los adjetivos correspondientes a los verbos mencionados anteriormente, y rellena la siguiente tabla:

Verbo	Sustantivo	Adjetivo
Mejorar	Mejora	Mejorable
Reparar	…	…
Renovar	…	…

Imagen 6 - Actividad para promover el aprendizaje autónomo.

Estas actividades facilitarán la adquisición de la rutina de búsqueda, que les permitirá acceder de forma autónoma al vocabulario necesario para desarrollar sus estudios o actividades profesionales.

Lo que se acaba de exponer ejemplifica cómo el corpus constituye una herramienta eficaz para practicar aspectos teóricos, fomentando un aprendizaje que —lejos de ser meramente mnemónico— permita a los discentes investigar, descubrir y consolidar o perfeccionar nociones previamente adquiridas.

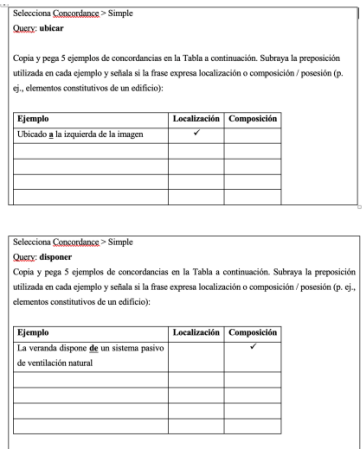
Una actividad muy interesante que se ha propuesto en el glosario de la Unidad 2 (*cf.* ES_SEAH, "Mi glosario", Unidad 2), consiste en la consolidación de una estructura léxico-gramatical que suele ser problemática para los estudiantes de una lengua extranjera: los verbos de régimen preposicional. Una vez más, el corpus ha demostrado ser una herramienta muy útil para investigar y comprender el funcionamiento de estos verbos. Seleccionando la sección "Concordance" y después "Advanced" y utilizando la sintaxis de búsqueda CQL (*corpus query language*), se han buscado las secuencias constituidas por verbos no auxiliares en infinitivo seguidos de preposiciones ([tag="VMN.*"] [tag="S.*"][18]), obteniendo así muchos ejemplos de uso de esta estructura (Imagen 7).

[18] Como se puede observar, en las búsquedas CQL cada palabra de la secuencia que se busca debe colocarse entre corchetes. Luego, a cada palabra se le asocia una etiqueta, es decir, un atributo de información específico. En el caso que nos ocupa, en [tag= "VMN.*"], la "V" hace referencia a la parte de la oración que se va a analizar (el verbo); la "M" alude al tipo de verbo en el que se va a centrar la búsqueda (un verbo no auxiliar) y la "N.*" indica el modo verbal (en este caso, el infinitivo); por su parte, en el caso de [tag= "S.*"], "S.*" designa la categoría preposicional.



Imagen 7 - Verbos preposicionales en el corpus ES_SEAH.

Para obtener una visualización más clara de los verbos preposicionales, se ha creado una lista de frecuencia de las formas extraídas del corpus ("Frequency" y después "WordForms"), a partir de la cual se les pide a los estudiantes que se fijen en algunos verbos propios del ámbito de la AC, verbos que, en español, se utilizan para expresar una localización o la composición de un edificio (p. ej., *situar en, ubicar en, contar con, disponer de*, etc.). Para cada uno de ellos, los estudiantes deberán seleccionar cinco ejemplos de concordancias, identificando las preposiciones que se suelen asociar a tales verbos e indicando si expresan localización o composición de un edificio (Imagen 8):



Selecciona Concordance > Simple

Query: **ubicar**

Copia y pega 5 ejemplos de concordancias en la Tabla a continuación. Subraya la preposición utilizada en cada ejemplo y señala si la frase expresa localización o composición / posesión (p. ej., elementos constitutivos de un edificio):

Ejemplo	Localización	Composición
Ubicado **a** la izquierda de la imagen	✓	

Selecciona Concordance > Simple

Query: **disponer**

Copia y pega 5 ejemplos de concordancias en la Tabla a continuación. Subraya la preposición utilizada en cada ejemplo y señala si la frase expresa localización o composición / posesión (p. ej., elementos constitutivos de un edificio):

Ejemplo	Localización	Composición
La veranda dispone **de** un sistema pasivo de ventilación natural		✓

Imagen 8 - Ejemplo de actividad propuesta.

A partir de la resolución de estos ejercicios, los alumnos podrán visualizar el uso de la construcción verbo + preposición y, tras realizar las observaciones necesarias, serán capaces de formular hipótesis, deducir reglas y reconocer aquellos matices de significado que surgen de la asociación de un mismo verbo con diferentes preposiciones —rasgos a veces difíciles de ilustrar basándose en las reglas de la gramática tradicional y en la observación de construcciones aisladas—.

En resumidas cuentas, las actividades que se acaban de presentar son solo algunas de las muchas que se han propuesto en los diferentes glosarios del curso de español SEAH. Como se ha visto, el objetivo es guiar a los estudiantes para que creen glosarios personales relacionados con los elementos y los conceptos propios de la AC y favorecer el uso autónomo del corpus para aprender la lengua. Sin embargo, se han tenido en cuenta las posibles dificultades que los estudiantes podrían encontrar al trabajar con esta herramienta, ya que se trata de aspirantes profesionales de la AC, a los que no se les pueden exigir competencias metalingüísticas avanzadas, ni conocimientos previos acerca del uso de corpus lingüísticos (puede que no sepan qué es una concordancia, una lista de frecuencia, etc.). Por esta razón, todas las actividades propuestas se han ilustrado paso a paso con instrucciones, con el fin de que los estudiantes aprendan a utilizar las diferentes funcionalidades del corpus (búsquedas avanzadas con la sintaxis CQL, análisis de las concordancias y listas de frecuencias, etc.) y puedan acceder de forma guiada, primero, y

autónoma, después, al vocabulario necesario para realizar con éxito sus actividades académico-profesionales.

6. Usos del corpus ES_SEAH en la didáctica de la traducción especializada

El corpus SEAH, creado básicamente para los estudiantes y especialistas de AC, resulta una herramienta valiosa también para los estudiantes de traducción especializada. Tratándose de un corpus multilingüe comparable, que por lo tanto incluye textos originales en varias lenguas comparables por temas y géneros discursivos, se presta bien para la explotación didáctica en ámbito traductológico, ya que constituye una fuente de información documental, lexicográfica y textual y favorece la comparación interlingüística (Sartor, 2021) tanto como la adquisición de la llamada competencia léxica traductora (Sánchez Ramos, 2013).

En este apartado presentaremos algunas breves tareas introductorias que se pueden realizar al principio de un curso de traducción para acercar paulatinamente al alumnado al uso del corpus y que sirven como preparación a la traducción de un texto especializado. La actividad se basa en el uso directo del corpus SEAH en su versión en español (ES_SEAH) y en italiano (IT_SEAH).

Para empezar, se les pide a los estudiantes que busquen, a través de la lista de frecuencia, los primeros cincuenta sustantivos (*nouns*) en ES_SEAH e IT_SEAH. Los resultados generan una lista de términos ordenados por frecuencia decreciente de aparición.

A partir de la observación de estas primeras ocurrencias, los alumnos empezarán a reflexionar sobre las diferencias y analogías entre la versión italiana y la versión española del corpus, ya que muchos de los términos que aparecen en los dos listados difieren. Empiezan así, de manera inductiva, a hacer consideraciones no sólo de carácter lingüístico, sino también cultural.

Luego, se les pide a los estudiantes que seleccionen aquellos términos más "técnicos" comunes a los dos listados (p. ej., *espacio/spazio*, *centro/centro*, *planta/piano*, *sistema/sistema*) y busquen sus colocaciones en los dos corpus. El resultado será una lista de colocaciones de términos "clave" muy frecuentes en las dos lenguas, a partir de la cual los estudiantes pueden extraer un primer pequeño glosario bilingüe a su disposición, basado en textos auténticos, fruto de una investigación activa y no de una mera búsqueda pasiva, y a menudo inadecuada, en un diccionario bilingüe.

Sucesivamente se les puede pedir a los estudiantes que seleccionen algunas de las colocaciones encontradas y que busquen sus concordancias en los dos corpus, lo que les permitirá explorar el comportamiento de dichas palabras en su contexto y cotejar las versiones en español y en italiano.

Se llega así a una dimensión más textual de la actividad que constituye la base mínima para cualquier trabajo de traducción. A estas alturas, se invita a los alumnos a identificar, siempre buscando en el corpus, los dominios y los géneros a los que pertenecen los textos, la frecuencia con la cual las diferentes colocaciones se encuentran en cada dominio y si hay diferencias entre el corpus español y el italiano.

Estas simples actividades inducen al estudiante a reflexionar sobre el funcionamiento afín, pero desigual, de las dos lenguas y nos muestran cómo, incluso con tipos de búsquedas básicas al alcance también de los estudiantes pocos familiarizados con el uso de corpus, se puede aprovechar dicha herramienta para interrogarla y para obtener información de tipo documental, textual, fraseológico y terminológico.

Conclusiones

Los corpus SEAH, anotados lingüísticamente y accesibles en línea de forma libre, resultan ser un recurso muy valioso para el desarrollo de actividades de autoaprendizaje por al menos dos razones. Por un lado, el acceso a grandes colecciones de textos representativos del lenguaje especializado de la AC permite recuperar unidades terminológicas difíciles de encontrar en corpus de referencia de la lengua general (p. ej., *aluminio lacado*, *ladrillo silicocalcáreo*, *vidrio semirreflectante*, etc.); además, la posibilidad de observar el uso de los diversos lemas en el contexto de la oración favorece la formulación de hipótesis y la deducción de reglas, que ayudarán a los estudiantes en la adquisición de otras unidades léxicas y de los mecanismos morfosintácticos en los que participan. Por otro lado, el aprendizaje de diferentes funcionalidades de

búsqueda de corpus (p. ej., búsquedas avanzadas con sintaxis CQL, ordenación de concordancias y procesamiento de listas de frecuencia, análisis de colocaciones, etc.) favorece el uso autónomo del corpus por parte del alumno, que podrá volver a consultar la herramienta para resolver las dudas lingüísticas que surjan.

Referencias bibliográficas

Abad Castelló, M. (2019). Uso de corpus lingüísticos por y para profesores de español como lengua extranjera. *redELE,* 31, pp. 148-167.

Alonso Pérez-Ávila, E. (2007). El corpus lingüístico en la didáctica del léxico del español como LE. *Boletín de ASELE (Asociación para la Enseñanza del Español como Lengua Extranjera)*, 37, pp. 11-25.

Benavides, C. (2015). Using a Corpus in a 300-level Spanish Grammar Course. *Foreign Language Annals*, 48 (2), pp. 218-235.

Boulton, A. (2016). Integrating corpus tools and techniques in ESP courses. *Asp*, 69, pp. 113-137.

Brown, A. V., Paz, Y. B. y Brown, E. K. (2021). *El léxico-gramática del español. Una aproximación mediante la lingüística de corpus*. Londres/Nueva York: Routledge.

Buyse, K. (2011). ¿Qué corpus en línea utilizar para qué fines en la clase de ELE? En J. de Santiago Guervós, H. Bongaerts, J. J. Sánchez Iglesias y M. Seseña Gómez (Eds.), *Del texto a la lengua: La aplicación de los textos a la enseñanza-aprendizaje del español L2-LE. Actas del XXI Congreso Internacional de ASELE. Universidad de Salamanca, 29 de septiembre-2 de octubre de 2010*. Universidad de Salamanca: Asociación para la Enseñanza del Español como Lengua Extranjera, pp. 277-288. https://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/21/21_0277.pdf [consulta: 22/11/2022].

Cambridge English Corpus. https://www.cambridge.es/en/about-us/cambridge-english-corpus [consulta: 22/11/2022].

Campillos, L., Gozalo, P. y Moreno, A. (2007). El corpus C-ORAL-ROM en la enseñanza de ELE. En E. Balmaseda Maestu, F. García Andreva, E. Pérez Jiménez y M. Morimoto (Eds.), *Las destrezas orales en la enseñanza del español L2-LE. XVII Congreso Internacional de la Asociación del Español como lengua extranjera (ASELE). Logroño, 27-30 de septiembre de 2006*. Universidad de La Rioja: Asociación para la Enseñanza del Español como Lengua Extranjera, pp. 1115-1128.

Carter, R. y McCarthy, M. (1995). Grammar and the spoken language. *Applied Linguistics*, 16 (2), pp. 141-158.

CEDEL2 = Lozano, C. (2022). *Corpus Escrito del Español L2*, versión 2.0. http://cedel2.learnercorpora.com/ [consulta: 22/11/2022].

CORDE = Real Academia Española. *Corpus Diacrónico del Español*. http://corpus.rae.es/cordenet.html [consulta: 22/11/2022].

CORPES XXI = Real Academia Española. *Corpus del Español del Siglo XXI*. https://www.rae.es/corpes/ [consulta: 22/11/2022].

CREA = Real Academia Española. *Corpus de Referencia del Español Actual*. http://corpus.rae.es/creanet.html [consulta: 22/11/2022].

Cruz Piñol, M. (2017). *Lingüística de corpus y enseñanza del español como 2/L*. Madrid: Arco Libros.

D'Angelo, M. (2021). Discorso accademico-disciplinare in prospettiva glottodidattica: una proposta multilingue *corpus-based* per il linguaggio dell'architettura e delle costruzioni. En M. D'Angelo y M. Ozbot (Eds.), *Lingue, testi e discorsi. Studi in onore di Paola Desideri*. Florencia: Cesati, pp. 347-361.

El Corpus del Español = Davis, M. (2001-2018). *El corpus del español*. https://www.corpusdelespanol.org/ [consulta: 20/11/2022].

Elvira-García, W. (2021). *Uso de corpus en clase de ELE: la lengua real como modelo*. Barcelona: Difusión.

ES_SEAH = SEAH Corpora Project. https://www.seahproject.eu/resources.php [consulta: 20/11/2022].

Flowerdew, J. (2009). Corpora in language teaching. En M. H. Long y C. J. Doughty (Eds.), *The handbook of language teaching*. Oxford: Wiley-Blackwell, pp. 327-335.

García Izquierdo, I. (Ed.) (2005). *El género textual y la traducción. Reflexiones teóricas y aplicaciones pedagógicas*. Berna: Peter Lang.

Gavioli, L. (2001). The learner as researcher: Introducing corpus concordancing in the classroom. En G. Aston (Ed.), *Learning with corpora*. Houston: Athelstan, pp. 108-37.

HarperCollins (1995). *Collins COBUILD English dictionary,* 2.ª ed. Londres: Collins COBUILD.

Huang, L.-S. (2011). Corpus-aided language learning. *ELT Journal*, 65 (4), pp. 481-484.

IT_SEAH = SEAH Corpora Project. https://www.seahproject.eu/resources.php [consulta: 20/11/2022].

Izquierdo Gil, M. C. (2002). Los córpora electrónicos en la enseñanza del vocabulario LE: la utilidad pedagógica de las concordancias. En A. Gimeno Sanz (Ed.), *Actas del XII Congreso Internacional de ASELE: tecnologías de la información y de las comunicaciones en la enseñanza de la E/LE*. Valencia: Publicaciones Universidad Politécnica de Valencia, pp. 201-212.

Johns, T. (1991). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. *English Language Research Journal*, 4, pp. 27-45.

Leech, G. (1997). Teaching and language corpora: A convergence. En A. Wichmann y S. Fligelstone (Eds.), *Teaching and language corpora*. Londres: Longman, pp. 1-23.

McEnery, T. y Wilson, H. (1996). *Corpus Linguistics*. Edimburgo: Edinburgh University Press.

McEnery, T. y Xiao, R. (2011). What Corpora Can Offer in Language Teaching and Learning. En E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning*. Londres/Nueva York: Routledge, pp. 364-380.

McEnery, T., Xiao, R. y Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. Londres: Routledge.

Parodi, G., Ibáñez, R. y Venegas, R. (2009). El Corpus PUCV-2006 del Español: identificación y definición de los géneros discursivos académicos y profesionales. *Estudios Filológicos*, 44, pp. 123-147.

Pastor Cesteros, S. (2023). *Español académico como LE/L2 destrezas, competencias y movilidad universitaria*. Londres/Nueva York: Routledge.

Paz, Y. (2013). Aplicación pedagógica de la lingüística de corpus a una clase de gramática avanzada de español como segunda lengua. *MarcoELE*, 16, pp. 1-17.

Piccioni, S., D'Angelo, M. y Ferro, M. C. (2021). I Corpora SEAH di comunicazione specializzata nel settore dell'Architettura e delle Costruzioni. Struttura, compilazione e usi. *Linguistica,* 61 (2), pp. 99-125.

Piccioni, S. y Pontrandolfo, G. (2021). WebLesp: corpus de comunicación digital especializada en español. Arquitectura, compilación y usos. *RLA — Revista de lingüística teórica y aplicada*, 59 (1), pp. 13-37.

Rychlý, P. (2007). Manatee/bonito - a modular corpus manager. *1st Workshop on Recent Advances in Slavonic Natural Language Processing*, pp. 65-70.

San Mateo, A. (2003). Los corpus lingüísticos y la enseñanza de E/LE: el uso de los pasados en español. *Frecuencia — L*, 22, pp. 52-58.

Sánchez Ramos, M. del M. (2013). El léxico en el aula de traducción: diseño de un modelo de adquisición de la competencia léxica traductora (inglés-español). *Tonos. Revista de Estudios Filológicos*, 24.

Sartor, E. (Ed.) (2021). *Los corpus especializados en la lingüística aplicada: enseñanza y traducción*. Mantua: Universitas Studiorum (Pliegos Hispánicos).

Sinclair, J. M. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Thompson, P. (2006). Assessing the contribution of corpora to EAP practice. En Z. Kantaridou, I. Papadopoulou y I. Mahili (Eds.), *Motivation in Learning Language for Specific and Academic Purposes*. Macedonia: University of Macedonia.