Proyecto SEAH: corpus lingüístico de arquitectura y construcción para la enseñanza de idiomas en el contexto de la movilidad académica y profesional[1]

Paloma Úbeda Mansilla
M.ª Luisa Escribano Ortega
Elena Isabel Romero Cañabate
Universidad Politécnica de Madrid

"El uso de corpus para el estudio de la lengua se considera una metodología empírica de trabajo, basada en el empleo de datos reales, de muestras de uso de la lengua"
(Villayandre 2008, 330)

1 Introducción

En el campo de la enseñanza del español como lengua extranjera (ELE) son muchos los investigadores que han trabajado en lenguas de especialidad, como Buyse (2021) y Ferreira, Elejalde y Blanco (2022); sin embargo, tanto en el ámbito de ELE como en el de inglés como segunda lengua (ESL), es escasa la bibliografía sobre lenguas con fines específicos enfocadas al lenguaje técnico, de la ciencia y la tecnología.

Existen estudios que introducen la metodología de trabajo con corpus; concretamente, Rojo (2021) y Parodi, Cantos-Gómez y Howe (2022) sirven como punto de referencia a los profesores de lenguas. En el ámbito de corpus escritos de especialidad contamos con el trabajo de Soneira Beloso (2013) en ESL, cuya tesis plantea que cada comunidad de conocimiento tiene un tipo de discurso distinto y una identidad. A partir de un corpus escrito de revistas de arquitectura, analiza el léxico desde el punto de vista de la formación de los neologismos para crear un glosario especializado de inglés y gallego (lengua materna de la autora) en el campo de la arquitectura. Más recientemente, en el campo de ELE Castrignanò (2021) analiza las concordancias de palabras que aparecen en el corpus SEAH (*Sharing European Architectural Heritage*) de español para, a partir de aquí, diseñar ejercicios de vocabulario y gramática para estudiantes extranjeros.

Por su parte, en lo que respecta a los corpus orales, podemos destacar la tesis doctoral de Úbeda Mansilla (2001) en el ámbito de ESL, en el que se recopila una muestra de conversaciones en inglés realizadas en estudios de arquitectura de habla inglesa. Este corpus oral es analizado desde un enfoque cognitivo prestando especial atención a la metáfora como extensión del proceso de categorización del habla. En esta línea de análisis de corpus orales de especialización, Rodríguez Cuadrado (2022) lleva a cabo una investigación en el campo de ELE sobre la comunicación arquitectónica a partir de un corpus de presentaciones orales realizadas por alumnos de la Escuela Técnica Superior de Arquitectura de Madrid. Concretamente, trata de analizar el discurso de los estudiantes en la defensa de los mejores trabajos presentados en las asignaturas de proyectos arquitectónicos. Este subgénero se caracteriza por la necesidad de efectividad en la percepción del mensaje, ya que la exposición tiene que realizarse en un tiempo muy limitado.

[1] Las autoras quieren mostrar su enorme agradecimiento a las siguientes personas que nos han facilitado ideas, documentos, soluciones y recursos de manera generosa para la recopilación del corpus. Sin ellas este trabajo habría sido mucho más difícil o imposible de gestionar al nivel de palabras y gestión de tiempo con el que contábamos para esta fase: Margarita de Luxan, Carmen Espegel, Chema Lapuerta Montoya, Rosario Segado, Ángela Juarranz, Paloma Poveda, Teresa Pastor Novella, Lara Alegre Rodríguez, Mar Martínez Rodríguez.

Así pues, tanto Castrignanò (2021) como Rodríguez Cuadrado (2022) abren una línea de trabajos basados en un corpus de español, elaborado como parte del proyecto europeo SEAH (*Sharing European Architectural Heritage: innovative language teaching tools for academic and professional mobility in Architecture and Construction*)[2]. El objetivo de este proyecto es el desarrollo de materiales didácticos con fines académicos y profesionales a partir de la creación de un corpus de textos técnicos, de arquitectura y construcción, en cinco idiomas: alemán, español, francés, italiano y ruso. Las universidades participantes son la Universitá d'Annunzio di Chieti-Pescara (Italia), promotora y coordinadora del proyecto, y la Universidad Politécnica de Madrid (España); la responsable de la UPM se unió al proyecto con un grupo de universidades europeas: Université Bordeaux Montagne (Francia), Masarykova Univerzita (República Checa) y Polotsk State University (Bielorrusia). Junto a estas cinco universidades, participa Internet Web Solutions (IWS), una empresa encargada del desarrollo tecnológico, con experiencia en este tipo de asociaciones, y afincada en Málaga (España).

Durante el primer año se creó el corpus que sirve de base para la creación de diferentes módulos de idiomas académico-profesionales para francés, alemán, italiano, ruso y español como lengua extranjera, en el campo de la arquitectura y la construcción (CA), así como los módulos que serán de acceso libre a estudiantes y profesionales. Uno de los objetivos subyacentes al proyecto es el fomento de la movilidad de estudiantes de estas áreas de conocimiento a través de una formación en lenguas extranjeras previa a la etapa de intercambio académico, necesaria para optimizar los resultados. En el ámbito de la tecnología de enseñanza, el proyecto cuenta con Internet Web Solutions (IWS), que presta el soporte tecnológico a todos los socios.

2 Organización del trabajo

El consorcio, liderado por la Universitá d'Annunzio, está formado por cinco grupos de trabajo, uno en cada una de las instituciones de educación superior, con amplia experiencia docente e investigadora en los campos de la lingüística aplicada, la enseñanza de lenguas extranjeras, la lingüística de corpus y el aprendizaje de idiomas. Cada grupo ha trabajado independientemente, pero de forma eficaz y coordinada gracias al trabajo de IWS, que ha aportado uniformidad a todo el proyecto.

Concretamente, el proyecto se plasma en tres productos intelectuales (*Intelectual Outcomes,* IOs), utilizando el lenguaje *Erasmus+*. En el primer año (IO1), SEAH creó cinco corpus lingüísticos digitales, compuestos por un millón de palabras en textos representativos de la comunicación profesional y académica en arquitectura y construcción; cada uno de estos corpus incluye textos en uno de los cinco idiomas objeto de estudio: francés, alemán, italiano, ruso y español.

En el segundo año (IO3), cada universidad asociada ha desarrollado cinco módulos *online* de enseñanza en el idioma de los corpus ya elaborados, destinados a estudiantes universitarios o profesionales, estudiantes de alguna de las lenguas extranjeras participantes en el proyecto. Se cuenta, por tanto, con un total de 25 módulos.

[2] El proyecto SEAH (*Sharing European Architectural Heritage: innovative language teaching tools for academic and professional mobility in Architecture and Construction*) fue concedido en octubre de 2020 y financiado por el programa ERASMUS + Strategic partnership (E+KA203) (número de referencia: CCG06-UPM/HUM-0559240240 2020-1-IT02-KA203-079511). El equipo está formado por de Luxán García de Diego, Escribano Ortega, Espegel Alonso, Gómez Blanco, Gómez Ortiz, Rodríguez Cuadrado, Romero Cañabate y Úbeda Mansilla. Página web: https://www.seahproject.eu/.

El tercer producto intelectural (IO2) se ha llevado a cabo a lo largo de todo el proyecto, por parte del socio tecnológico, IWS, y consiste en desarrollar, probar, actualizar y difundir la plataforma Multilingual Open Educational Resources (REA), que se ha utilizado para la distribución de los recursos creados dentro del proyecto. Ha sido también el responsable de fijar el cronograma general de implementación del proyecto, de definir aquellos contenidos que había que homogeneizar en todas las actividades, siempre teniendo en cuenta las peculiaridades de cada lengua y de cada universidad participante.

3 El corpus de español

La Universidad Politécnica de Madrid, responsable del corpus de español y de los módulos lingüísticos posteriores, fue invitada a participar como consecuencia de la experiencia adquirida en la elaboración del *Manual de Español Académico para la Arquitectura y la Construcción B1-B2* (Úbeda y Escribano 2015), resultante del grupo de innovación educativa PIE 2014/1825. Este manual resultó ser inspirador para las promotoras del proyecto SEAH y ha convertido al equipo español en motor del proyecto, proponiendo y pilotando tanto la elaboración del corpus como la elaboración de los módulos lingüísticos.

Para conseguir un millón de palabras en español relacionadas con el ámbito de la arquitectura y construcción, se consideró importante incluir tanto documentos escritos como orales (75 % escritos y 25 % orales). Para ello, se tuvo en cuenta que los textos debían ser siempre especializados y útiles para estudiantes y profesionales; evitar textos divulgativos o publicitarios, con autores nativos de español, tanto peninsular como español de América; equilibrar textos producidos por mujeres y por hombres; y buscar los géneros texturales más frecuentes en la comunicación profesional. Siguiendo estas indicaciones, en agosto de 2021 el equipo español había creado un corpus con 830 000 palabras dentro de la modalidad escrita y 170 000 dentro de la oral (Tabla 1).

CORPUS	**Documentos**	**Procedencia/Fuentes**
ESCRITO 350 000 palabras	Descripción/Memorias de proyectos. Temas: rehabilitación, edificación, ingeniería civil, edificios residenciales y no residenciales.	Revista especializada en arquitectura *El Croquis*
ESCRITO 180 000 palabras	Enunciados de proyectos elaborados por profesores de Proyectos Arquitectónicos.	Arquitectos de proyectos de la Escuela de Arquitectura de Madrid
ESCRITO 200 000 palabras	Textos técnicos: descripción de materiales, descripciones de procedimientos y estrategias técnicas.	Alumnos y profesores de la Escuela de Arquitectura de Madrid
ESCRITO 100 000 palabras	TFG y TFM de estudiantes destacados.	Alumnos de la Escuela de Arquitectura de Madrid
ORAL 20 000 palabras	Conferencias sobre proyectos arquitectónicos, estructuras de edificación e instalaciones.	Colegio Oficial de Arquitectos de Madrid y de la Escuela de Arquitectura de Madrid
ORAL 100 000 palabras	Proyectos en Construcción (*Project Pitch*)	Tesis docturales (ETSAM) Presentaciones de estudiantes de arquitectura
ORAL 50 000 palabras	*Podcast*	Programas de radio

Tabla 1. División de número de palabras por tipologías textuales (elaboración propia)

4 El corpus de español en una imagen

Como se puede observar en la Imagen 1 (véase página 392), el soporte del póster presentado en XXXI Congreso Internacional de ASELE representa un edificio con cuatro niveles (cimientos, edificio, detalles constructivos, y ático y cubierta) y un pequeño anexo en el que se presenta al equipo. El equipo de investigación, en esta primera fase, está formado por tres arquitectas, cinco filólogas y un técnico. Se trata, por tanto, de una representación metafórica del resultado del primer año del Proyecto SEAH.

En el primer nivel, se encuentran los cimientos, constituidos por los antecedentes, objetivos y metodologías, junto con el equipo responsable de la parte española. Consideramos fundamental este primer nivel de definición, que estuvo muy marcado por la universidad coordinadora desde los inicios de la elaboración del proyecto. La clara definición de objetivos resultó muy útil y permitió aprovechar el tiempo para dedicarlo a los niveles posteriores.

El segundo nivel, que se corresponde ya con el edificio en sí, incluye las especificaciones de los corpus de los distintos idiomas (alemán, español, francés, italiano y ruso) y la variedad de herramientas tecnológicas para la recogida de los resultados del trabajo de todos los equipos. Juntas constituyen una buena forma de comunicación y un soporte tecnológico práctico y útil. En la balconada se incluyen las palabras que hasta la fecha de celebración del congreso se habían recogido. En la actualidad el corpus se ha completado y contiene un total de 992 186 palabras en español.

El tercer nivel está formado por los detalles constructivos, es decir, la selección e incorporación del corpus de español, que hemos comentado brevemente en el apartado 3. Queremos valorar el enfoque de la perspectiva de género durante esta primera fase, ya que se ha intentado, con gran esfuerzo, conseguir una muestra tanto para la tipología escrita como para la oral de 50 % de arquitectos y 50 % de arquitectas.

El cuarto y último nivel está conformado por el ático y la cubierta, que serán los módulos didácticos que desarrollaremos en el segundo año del proyecto (IO3). Como cabe suponer, este nivel está definiéndose en estos momentos. Las reuniones de estudio previas se vienen desarrollando desde septiembre 2021 por asamblea de todos los socios. Todos estos trabajos preliminares, así como las pruebas y unidades piloto que se están desarrollando en los diferentes idiomas, ayudarán a definir las tareas para, así, poder cumplir con el objetivo final.

Por último, hemos querido añadir a nuestro edificio un espacio de reflexión con las fortalezas y debilidades del proyecto, que podrían quizás subsanarse o incorporarse a la segunda fase del proyecto. Como fortalezas destacamos una selección de textos para el corpus de temática variada, la incorporación de textos de arquitectos latinoamericanos al corpus y la rigurosidad al elegir textos científicos, suprimiendo los divulgativos, que se pueden encontrar fácilmente en revistas de actualidad. Como debilidades señalamos la dificultad de encontrar permisos para usar textos, por lo que las fuentes documentales no son tan variadas como nos hubiera gustado. También ha sido difícil encontrar un volumen representativo de textos escritos únicamente por mujeres; pese a la dificultad, este es un objetivo que hemos alcanzado tanto en el corpus oral como en el escrito.

Finalmente, nuestro póster-edificio (véase página siguiente) ha recogido una serie de palabras que permiten intuir el tipo de vocabulario específico de nuestro corpus por medio del etiquetado de algunos elementos gráficos como *contraventana*, *tejado a un agua*, *cornisa*, *ventana abuhardillada*, entre otras.

5 Conclusiones

El póster presentado anteriormente resume en una imagen el trabajo realizado en el primero año de los dos que tiene como duración el proyecto SEAH. El objetivo final de este proyecto es, por un lado, crear material de trabajo para alumnos internacionales y por otro, establecer una plataforma de referencia para profesores o eruditos en este campo de especialización.

Realizar este póster ha sido una experiencia enriquecedora, ya que nos ha obligado a observar lo que se ha obtenido durante la primera fase del proyecto. El póster representa, por tanto, un edificio o, mejor dicho, la estructura de un edificio constituido por un millón de palabras compiladas en español, que ya están en la web de prueba y que esperamos sea útil para los aprendientes de ELE interesados en los ámbitos de la arquitectura y la construcción. Gracias a la oportunidad que hemos tenido con la elaboración del póster, también hemos podido obtener una visión de conjunto que nos sitúa en la segunda y definitiva fase del proyecto, donde se está generando una muestra de material didáctico que será de utilidad para los estudiantes internacionales o para los hablantes de ELE interesados en algún tema de arquitectura.

En conclusión, es importante señalar el reto que supone la puesta en marcha de un producto homogéneo en el que participan cinco equipos de distintas lenguas, distintos países y distintas tradiciones culturales; sin embargo, ello también supone una gran oportunidad de realizar el trabajo ingente de crear un corpus especializado en arquitectura y construcción, que servirá para mejorar y ampliar nuestra experiencia docente.

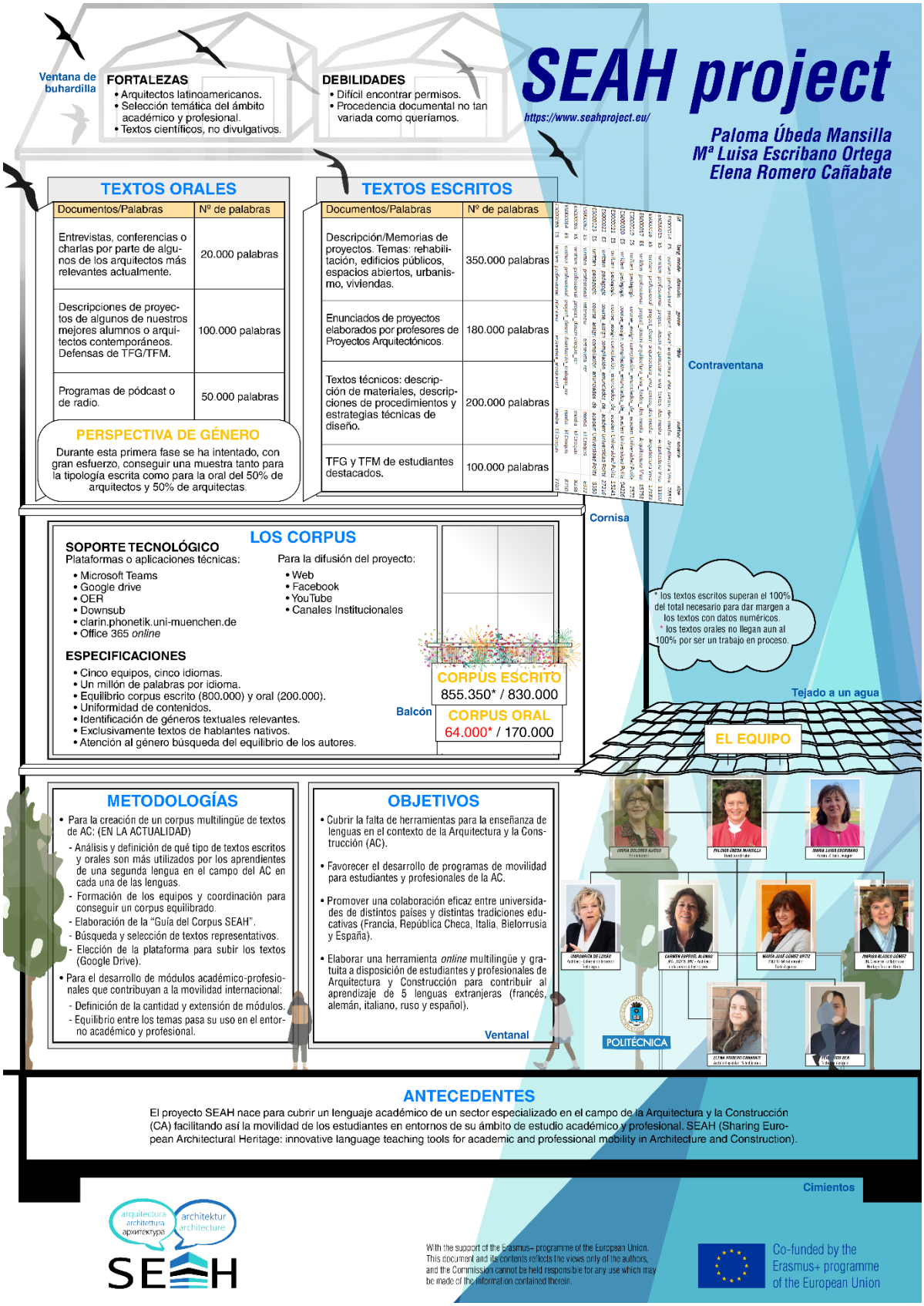

Imagen 1. Póster presentado en el congreso (elaboración propia)

REFERENCIAS BIBLIOGRÁFICAS

Buyse, K. (2021). Corpus textuales de nativos para investigar sobre la enseñanza/aprendizaje del español LE/L2. En M. Cruz Piñol (Ed.), *e-Research y español LE/L2. Investigar en la era de las tecnologías* (pp. 165-190). Londres/ Nueva York: Routledge.

Castrignanò, V. (2021). *Proposta didattica per l'insegnamento dello spagnolo per l'Architettura* [Trabajo fin de grado inédito]. Chieti/ Pescara: Università Degli Studi "G. d'Annunzio" Chieti – Pescara.

Ferreira Cabrera, A.A., Elejalde Gómez, J. y Blanco San Martín, L. (2022). Diseño e implementación del Corpus de Aprendientes de Español como Lengua Extranjera (CAELE). *Círculo de Lingüística Aplicada a la Comunicación*, 90, 137-155. https://doi.org/10.5209/clac.71174.

Parodi, G., Cantos-Gómez, P. y Howe, C. (2022). *Lingüística de corpus en español / The Routledge Handbook of Spanish Corpus Linguistics*. Londres: Routledge.

Rodríguez Cuadrado, R. (2022). *La palabra construida: Un estudio de las defensas de proyectos en la ETSAM.* [Trabajo fin de máster inédito]. Madrid: Universidad Politécnica de Madrid.

Rojo, G. (2021). *Introducción a la lingüística de corpus en español*. Nueva York: Routledge.

Soneira Beloso, B. (2013). *A Lexical description of English for Architecture: a corpus-based approach* [Tesis Doctoral]. Santiago de Compostela: Universidad de Santiago de Compostela.

Úbeda Mansilla, P. (2001). *Estudio de un corpus de textos conversacionales en inglés realizado en estudios de arquitectura: Su aplicación al diseño de un programa de inglés para arquitectos* [Tesis Doctoral]. Madrid: Universidad Complutense de Madrid.

Úbeda Mansilla, P. y Escribano Ortega, M.L. (2015). *Arquitectura y Construcción: Español Académico y Profesional. Nivel Intermedio B1-B2.* Madrid: Edinumen.

Villayandre Llamazares, M. (2008). Lingüística con corpus (I). *Estudios Humanísticos. Filología*, 30, 329-349. https://doi.org/10.18002/ehf.v0i30.2847.