



29 DE NOVIEMBRE

# JORNADA DE CORPUS Y ENSEÑANZA DE LENGUAS

**Sala de Juntas, 2º piso. Rectorado, UdL**

Reunión científica en el marco del proyecto de investigación  
FFI2016-80280-R. Financiado por



GOBIERNO  
DE ESPAÑA

MINISTERIO  
DE CIENCIA, INNOVACIÓN  
Y UNIVERSIDADES



UNIÓN EUROPEA  
FONDO  
EUROPEO DE  
DESARROLLO  
REGIONAL

*"Una manera de hacer Europa"*

**LIBRO DE RESÚMENES**

*Diseño y edición digital:* M<sup>a</sup> Ángeles Calero Fernández

*Diseño de la imagen de la portada:* Maite Miramón Escribano

*Gestión de las comunicaciones y las revisiones:* Chen Lei y Maite Miramón Escribano

*Web de la Jornada:* <http://www.lecel.udl.cat/es/congresos/pagina-1/>

Grupo LECCEL – SGR 1644-2017  
Departamento de Filología Clásica, Francesa e Hispánica  
Universidad de Lleida, mayo 2019



## Índice alfabético de autoras y autores

Aderdouch Derdouch, Samia .....	4
Calero Fernández, M <sup>a</sup> Ángeles y Maribel Serrano Zapata.....	5
García, Roberto, Albert Berga, Juan Manuel Gimeno .....	6
Gómez-Devís, M. Begoña.....	8
Iglesia Martín, Sandra e Isabel Gibert Escofet.....	9
Mira Conejero, Ángela .....	10
Miramón Escribano, Maite, Chen Lei y Liu Zhuoyu .....	11
Moreno Villanueva, José Antonio y Alicia Madrona Cao.....	12
Ortiz Cruz, Demelsa.....	14
Terrado Pablo, F. Javier, Zhang Peng y Zhang Bohua .....	15

Viernes, 29 de noviembre, **18.30 h.**

Aderdouch Derdouch, Samia

**El aprendizaje del ELE en Barcelona. Un estudio de caso centrado en las mujeres marroquíes que viven en Santa Coloma de Gramenet**

El objetivo de esta comunicación es presentar los resultados de mi investigación sobre el proceso de aprendizaje del ELE de las mujeres marroquíes inmigrantes que viven en Santa Coloma de Gramenet, centrándome en las actitudes lingüísticas que subyacen a este proceso de aprendizaje. Me ha interesado analizar cuáles son las principales dificultades y facilidades de estas mujeres al abordar el estudio del ELE, ver si recién llegadas de Marruecos aprenden la lengua o si retrasan el aprendizaje, a qué edades suelen empezar a estudiar, hasta qué niveles llegan, durante cuánto tiempo suelen estudiar español, si interrumpen el proceso de aprendizaje y, caso de que lo hagan, cuáles son las razones. Además, también he querido indagar acerca de cómo influyen las lenguas que conocen en el aprendizaje de la lengua. Para dar respuesta a estos objetivos, se han llevado a cabo 100 entrevistas a mujeres marroquíes de Santa Coloma, de las cuales 76 están estudiando ELE y 24 han abandonado sus estudios.

Viernes, 29 de noviembre, 12.00 h.

Calero Fernández, M<sup>a</sup> Ángeles y Maribel Serrano Zapata

### **Análisis de errores y etiquetado en los corpus de aprendices**

Los criterios con los que se diseña, se selecciona, se cataloga y se etiqueta un corpus lingüístico determinan las aplicaciones que dicho corpus podrá tener. En el caso de los corpus de aprendices, marcarán las posibilidades de análisis de la interlengua y la explotación didáctica de las muestras lingüísticas recogidas. Un corpus electrónico de aprendices de lengua es una base de datos lingüísticos auténticos en los que se constata el uso real de la lengua. El análisis de este tipo de corpus ofrece información sobre frecuencia de uso correcto e incorrecto, por lo que constituye un valioso recurso para la investigación en enseñanza de lenguas extranjeras.

En esta comunicación se revisarán las distintas taxonomías del error establecidas hasta el momento, así como las potencialidades y las dificultades que presentan para su adaptación a las necesidades de etiquetado de los corpus de aprendices. Al mismo tiempo, se determinará cuál es la más adecuada para el desarrollo de un sistema de etiquetado aplicable a este tipo de corpus.

Viernes, 29 de noviembre, 12.30 h.

García, Roberto, Albert Berga, Juan Manuel Gimeno

### **TextAnnot, una herramienta Web para la gestión y anotación de corpus lingüísticos**

La gestión y anotación de corpus lingüísticos se beneficia en gran medida de la utilización de herramientas informáticas. En la actualidad se trata de la forma de proceder habitual, tanto para la gestión del corpus, como para su publicación en línea. Lo que no es tan habitual [1] es que la herramienta de gestión y anotación del corpus sea también una herramienta Web que facilite el trabajo de forma colaborativa. Este es justamente el objetivo de TextAnnot, una herramienta Web para la gestión y anotación de corpus lingüísticos, que también permite la publicación en línea del corpus sobre el que se está trabajando.

TextAnnot se ha desarrollado en el marco del proyecto de investigación “Elaboración y catalogación de un corpus de textos escritos en ELE producidos por estudiantes sinohablantes” (ECCOTEELE) [2] para dar soporte a la creación del “Corpus de Interlengua Española de Aprendices Sinohablantes” (CINEAS) [3], un conjunto de textos escritos por estudiantes de español de origen chino con diferentes grados de dominio del idioma.

Los textos son tareas escritas a mano, luego digitalizados y anotados a partir de un conjunto de metadatos como sexo (o género), curso, universidad, etc. Finalmente, los textos se transcriben en documentos de texto estructurados utilizando XML, lo que permite cargarlos en TextAnnot incluyendo tanto el texto como todos los metadatos utilizados. También es posible realizar directamente la transcripción a través de un formulario Web proporcionado por TextAnnot.

A partir de este punto, los textos ya pueden ser gestionados por TextAnnot. Las funcionalidades implementadas se enfocan en tres perfiles de usuario: lingüista, usuario del corpus y administrador.

Los lingüistas pueden crear o cargar muestras, como se ha mencionado anteriormente, y también editar las muestras ya cargada junto a sus metadatos. El texto de las muestras también puede ser anotado, en el caso de CINEAS, para resaltar los errores de acuerdo con un sistema de categorías y subcategorías desarrollado a tal efecto en el proyecto.

Los textos gestionados y anotados se pueden publicar directamente desde TextAnnot de manera que están accesibles para los usuarios del corpus. Estos disponen de mecanismos para buscar concordancias en los textos de las muestras, que también se pueden filtrar según los metadatos y anotaciones asociados. Además, se proporcionan estadísticas globales y particulares para campos de metadatos y anotaciones, por ejemplo, porcentaje de textos según el curso del informante o con

[1]Tools for Corpus Linguistics, <https://corpus-analysis.com>

[2]Elaboración y catalogación de un corpus de textos escritos en ELE producidos por estudiantes sinohablantes (ECCOTEELE), Ministerio de Economía y Competitividad, Gobierno de España, FFI2016-80280-R, <http://www.lecel.udl.cat/es/investigaciones/pagina-2/>

[3]Corpus de Interlengua Española de Aprendices Sinohablantes (CINEAS), <https://cineas.udl.cat>

errores de concordancia de género. Estas estadísticas se actualizan automáticamente a medida que se añaden y gestionan las muestras del corpus.

Los usuarios administradores pueden, a su vez, definir los metadatos que se quieren asociar a cada muestra, si son públicos o si se tienen que generar estadísticas para ellos. Lo mismo sucede con las etiquetas a utilizar para la anotación. Los administradores pueden definir jerarquías de etiquetas que luego podrán ser utilizadas por los lingüistas para anotar los textos. Esto permite un gran nivel de flexibilidad de la herramienta, que se pueden adaptar a diferentes tipos de muestras o de tareas de análisis y anotación de textos, más allá de la anotación de errores en el caso de CINEAS.

Para poner en práctica esta flexibilidad, y facilitar la adopción de TextAnnot, la herramienta está disponible de manera gratuita junto con todo su código fuente. Se busca, no solo que sea utilizada en otros proyectos de construcción y gestión de corpus lingüísticos, sino también la creación de una comunidad de usuarios que contribuyan a su desarrollo y mantenimiento futuros.

Viernes, 29 de noviembre, 17.00 h.

Gómez-Devís, M. Begoña

**Disponibilidad léxica en niños de valencianos de 6 años: corpus léxico infantil**

Esta contribución presenta, en el marco de los estudios de disponibilidad léxica, una nueva perspectiva que permite investigar el léxico disponible en edades tempranas (6-7 años) a partir de encuestas orales, así como el estudio de posibles rasgos significativos en la configuración de dicho corpus: carácter oral de la prueba asociativa y del contacto de lenguas (castellano y catalán de Valencia).

Viernes, 29 de noviembre, 16.00 h.

Iglesia Martín, Sandra e Isabel Gibert Escofet

**Diccionario de español para alumnos sinohablantes: percepciones de los potenciales usuarios sobre sus necesidades**

La recopilación de datos relacionados con las necesidades lexicográficas de los estudiantes sinohablantes de español es una de las fases primordiales para establecer el tipo de categorías de información que será necesario incluir en el diseño de una base de datos con la que trabajar para la confección de un diccionario. Puesto que este estudio se inscribe en el Proyecto Diccionario de español para aprendices sinohablantes (fase primera), (Grupo de Investigación LEXELE-URV), los objetivos que se persiguen son, por una parte, comprobar la relación de los aprendices sinohablantes con la información lexicográfica y, por otra parte, enlazar las necesidades subjetivas expresadas por los estudiantes con el tipo de información necesaria en un diccionario que sea capaz de cubrir estas necesidades, sobre todo, teniendo en cuenta que hasta la fecha, no han visto la luz demasiados estudios específicos sobre necesidades de información lexicográfica por parte de estudiantes sinohablantes de ELE. Para alcanzar estos objetivos se han administrado dos encuestas a 100 estudiantes de nivel B1-B2. La primera encuesta va dirigida a estudiantes que no han tenido formación específica en el uso del diccionario y la segunda a estudiantes que han recibido un curso de formación de 30 horas. Para el diseño del estudio se ha adoptado una metodología mixta. Los resultados obtenidos nos permiten reflexionar sobre la información que debe ser incluida necesariamente en los artículos lexicográficos de un diccionario pensado los aprendientes de ELE sinohablantes de nivel B1-B2 y sobre las razones por las que los posibles usuarios de este diccionario podrían necesitar este tipo de información. Las encuestas incluyen cuestiones que van desde la información contextual de la palabra a la traducción en chino, pasando por la posibilidad de incorporar imágenes. En la era de Internet, los resultados que arrojan las encuestas son, cuanto menos, sorprendentes en muchos aspectos, puesto que se puede encontrar relación entre el uso del diccionario y otras herramientas de consulta.

Viernes, 29 de noviembre, 19.00 h.

Mira Conejero, Ángela

### **Creación de un corpus para el estudio del español como lengua de instrucción**

Cada vez son más los inmigrantes que llegan a España acompañados de sus familias, entre los que se incluyen niños y jóvenes en edad de escolarización obligatoria, pero cuya lengua materna no es el español. Esto ha supuesto nuevos retos educativos tanto para los docentes, como para las Administraciones, que han implantado diversas medidas para paliar el desconocimiento de la lengua por parte de este colectivo, como es el caso de las aulas de apoyo para alumnado no hispanohablante.

Sin embargo, se suele creer que basta con que los alumnos adquieran las habilidades básicas para comunicarse y no se otorga la suficiente importancia a la enseñanza de la lengua de instrucción con miras a una próspera integración en el aula y en el plan de estudios.

Partiendo de esta idea, se llevó a cabo un trabajo que pretendía, por una parte, identificar los aspectos lingüísticos más característicos de la lengua de instrucción empleada en los libros de texto usados en el aula y, por otra, determinar si existen elementos concretos de la lengua de instrucción asociados a cada materia. Con ese fin, se seleccionaron dos centros públicos de enseñanza secundaria de la provincia de Alicante, que nos proporcionaron los manuales que utilizaban para las asignaturas de matemáticas e historia, de los cuales se seleccionaron dos temas. A continuación, se procedió a la creación de un corpus atendiendo a la clasificación de Xirinachs (1990) y recogiendo el léxico especializado (vocabulario técnico y vocabulario académico), las nominalizaciones, la conexión (marcadores del discurso), las formas no personales del verbo (infinitivo, gerundio y participio), las oraciones en pasiva (pasiva perifrástica y pasiva refleja), los verbos en 3ª persona del singular o del plural y los acrónimos propios de la ciencia.

Del análisis individual de cada manual se obtuvieron las características de la lengua de instrucción propias de ese material. De la comparación de manuales de una misma materia se observaron coincidencias que se pueden generalizar. De la comparación de todas las unidades del corpus se puede conseguir un listado general de los elementos de la lengua de instrucción más frecuentes, que podrían utilizarse para crear unidades didácticas que aborden la lengua de instrucción como parte de los contenidos de dichas aulas de apoyo.

Viernes, 29 de noviembre, 17.30 h.

Miramón Escribano, Maite, Chen Lei y Liu Zhuoyu

**La enseñanza de la fraseología española en universidades chinas y su adquisición a partir de los datos del CINEAS (Corpus de Interlengua Española de Aprendices Sinohablantes).**

El objetivo de esta comunicación es revisar la presencia de la fraseología en el manual de enseñanza que se suele utilizar en las universidades chinas: Español Moderno. En concreto, se va a vaciar las unidades fraseológicas (colocaciones, locuciones y enunciados fraseológicos –paremias y fórmulas rutinarias–) de las actividades del nivel A2 de dicho manual para saber qué tipo de expresiones se enseñan, qué input reciben los estudiantes, y qué tipología de actividades se ofrecen.

Por otra parte, se pretende cotejar el uso real que los estudiantes hacen de estas mismas expresiones después de haberlas estudiado. Para ello, vamos a utilizar el corpus CINEAS (Corpus de Interlengua Española de Aprendices Sinohablantes) y comparar el uso, tanto acertado como erróneo, de dichas unidades fraseológicas en los niveles A1 y A2.

La metodología en la que se basa esta comunicación reposa en la lingüística de corpus, el análisis de errores y la fraseodidáctica. En lo que se refiere al marco teórico fraseológico, se sigue la clasificación expuesta por Gloria Corpas (1996), que tiene una visión ancha de la fraseología, es decir, incluye un amplio espectro de unidades fraseológicas, desde colocaciones hasta enunciados completos. Respecto al análisis de errores, se van a seguir los cuatro primeros pasos de la propuesta de Santos Gargallo (2004) para este proceso, a saber, a) compilación de datos de corpus; b) identificación de los errores; c) descripción de los errores, y d) clasificación de acuerdo a una taxonomía.

Esta taxonomía de errores entra en estrecha relación tanto con la lingüística de corpus como con el marco teórico de la fraseología, puesto que aborda el tema de la dificultad de crear etiquetas para clasificar los errores fraseológicos en un corpus de estudiantes sinohablantes. Salvando dichas dificultades, la propuesta clasificatoria, y con la que se trabajan los errores que pueden aparecer en el corpus, es la siguiente: a) uso correcto; b) adición en la forma; c) omisión en la forma; d) formación inadecuada; e) impropiedad idiomática, f) impropiedad en la adecuación.

Viernes, 29 de noviembre, 16.30 h.

Moreno Villanueva, José Antonio y Alicia Madrona Cao

**Los índices de disponibilidad léxica como herramienta complementaria para la selección del leuario de un diccionario de ELE**

Este estudio se inscribe en un proyecto de mayor amplitud («Diccionario de español para aprendices sinohablantes: fase primera»; FFI2016-76417-P) cuyo propósito último es la elaboración de un diccionario de aprendizaje de nueva planta que logre dar respuesta a las demandas de información lexicográfica de los estudiantes sinohablantes de español como lengua extranjera, más específicamente de los aprendices de nivel B1-B2. En la primera fase del proyecto, tras analizar cuáles son las necesidades de este grupo de usuarios, uno de nuestros principales objetivos es determinar el conjunto de entradas del diccionario (Bargalló, 2019). Establecer el leuario de un diccionario entraña muchas dificultades, entre otras razones porque no existe un criterio plenamente objetivo para su determinación; de hecho, los diccionarios no suelen explicitar los criterios de selección adoptados. Esta tarea cobra aún más relevancia cuando, como es el caso, se concibe como un diccionario que pretende atender de manera específica a las necesidades de producción de los estudiantes en los niveles intermedio y avanzado. A este respecto, las investigaciones sobre el dominio léxico señalan que, al término del nivel B2, un aprendiente maneja en torno a 9000 palabras, cifra que constituye una referencia — en ningún caso una limitación— para nuestro proyecto.

Entre los criterios de selección léxica que se emplean para determinar el leuario de un diccionario figura habitualmente la frecuencia —o la alusión al léxico más frecuente—, criterio que, si bien permite extraer el léxico básico de una lengua, presenta notables debilidades conforme se pretende ampliar la selección, como tendremos ocasión de mostrar en este trabajo. Por esta razón, planteamos la necesidad de emplear de manera complementaria otros corpus léxicos, entre ellos los inventarios de Nociones específicas del Plan curricular del Instituto Cervantes (2006) y los índices de disponibilidad léxica.

En relación con este último aspecto, en anteriores trabajos (Moreno y Madrona, 2017, 2019) hemos constatado que los índices de léxico disponible pueden ser de gran ayuda en esta labor de selección, sobre todo en determinadas áreas léxicas de los ámbitos público, educativo y profesional, donde resultan especialmente útiles para contrastar y triangular resultados.

Para ilustrarlo, en este estudio se analiza el léxico disponible en cuatro centros de interés que entroncan con las necesidades de comunicación de nuestros usuarios finales tanto en el contexto académico como en su vida cotidiana: educación, salud, política y justicia. Con este propósito, se ha administrado una encuesta de disponibilidad léxica a dos grupos de informantes de nivel universitario: 25 estudiantes nativos del grado de Lengua y Literatura Hispánicas y el máster de Enseñanza de ELE de la Universitat Rovira i Virgili y 25 estudiantes chinos que cursan nivel B2 de español

en el Centro de Estudios Hispánicos de esta misma universidad. A partir de esas encuestas, se identifican los términos de los centros de interés seleccionados que tienen mayor prevalencia entre los estudiantes y se profundiza en las diferencias en cuanto a la competencia léxica de los grupos analizados. Posteriormente, se establecen correlaciones con el léxico incluido en los inventarios de nociones específicas del Plan curricular del Instituto Cervantes (2006) y los índices de frecuencia de Corpes XXI-Enclave RAE y Davies & Davies (2018) a fin de determinar qué términos son susceptibles de figurar en el lecionario de nuestro diccionario.

Viernes, 29 de noviembre, **10.30 h.**

Ortiz Cruz, Demelsa

**¿Rabito o rápido? Análisis de errores en la producción escrita de sinohablantes de ELE**

La presente comunicación tiene como objetivo ofrecer los resultados del análisis de errores de producciones escritas por sinohablantes aprendientes de español como lengua extranjera (ELE) en contexto de inmersión lingüística en la ciudad de Zaragoza. Para ello, se han analizado cerca de 50 textos, para ofrecer un estudio cualitativo, cuantitativo y contrastivo, teniendo en cuenta, además del tipo de error atestiguado (entre otras cuestiones), las condiciones de creación en las que se produjeron estas muestras de expresión escrita con el fin de determinar si tales condicionan (o no) la aparición de dichos errores de manera cuantitativa y/o cualitativa.

Viernes, 29 de noviembre, 11.00 h.

Terrado Pablo, Javier, Zhang Peng y Zhang Bohua

### **Ausencia de metataxis en la producción escrita de estudiantes chinos de ELE**

La metataxis es el cambio estructural al que el traductor debe someter una frase en el proceso de versión a otra lengua.

En las fases iniciales del aprendizaje de una lengua, el aprendiz tiende a verter en estructuras sintácticas de su propia lengua los contenidos que desea expresar en la lengua meta, asumiendo, eso sí, el nuevo léxico.

De algún modo el aprendiz, si se propone construir frases de cierta complejidad, se convierte inevitablemente en un traductor improvisado, porque desconoce todavía la riqueza estructural de la lengua meta y no logra verter sus ideas en los moldes adecuados.

Tiene lugar entonces, por decirlo así, una clonación de la estructura de origen en la estructura de la lengua meta, algo cercano a lo que normalmente se llama “traducción palabra por palabra”.

La ausencia de metataxis no es un fenómeno grave entre lenguas semejantes. En cambio, en lenguas de tipología alejada, como son el chino y el español, la ausencia de este proceso de cambio estructural puede abocar a la creación de frases aberrantes y a la incompreensión.

En el caso de los estudiantes chinos de español como lengua extranjera es frecuente que las frases de sus composiciones dejen traslucir estructuras que pueden explicarse a partir de la sintaxis de la lengua china. Las frases españolas resultantes chocan entonces con el sentimiento lingüístico del hablante nativo español y se consideran poco aceptables, aun cuando puedan ser comprensibles.

En nuestra comunicación ilustraremos esto con ejemplos extraídos del corpus CINEAS, proyecto de investigación de la Universidad de Lleida dirigido por la profesora María Ángeles Calero. Veamos uno de los ejemplos analizados:

1) *Además montamos bicicleta alrededor del lago* (0895\_A2m2c6n20170601)

La frase es perfectamente comprensible pero un corrector de español probablemente le daría una forma como la siguiente:  
1.b) Además rodeamos el lago montados en bicicleta.

¿Cuál es la explicación de la estructura de esta frase? En su lengua materna, la estudiante que la produjo está acostumbrada a concebir la bicicleta (zìxíngchē) como un objeto que utiliza para la acción de montar (qí). En consecuencia le asigna en la frase española la función de complemento directo. Además, el lago lo concibe como un lugar que tiene que expresar mediante un complemento precedido y seguido de partículas (zài hú biān), como es habitual en chino. Es probable que la estudiante haya construido en su mente una frase como:

我们在湖边骑自行车

Wǒmen zài hú biān qí zìxíngchē

Frente a esa concepción, el corrector español o el profesor que deba revisar su redacción puede proponerle una frase en la que el lago se expresa mediante un complemento directo de la acción de rodear y la bicicleta como un lugar expresable mediante un complemento preposicional.

Como observa la profesora Zhang Peng: "Un determinado entorno geográfico natural, histórico y cultural da vida a diversos y específicos patrones de lenguaje y formas de pensamiento. Las grandes diferencias culturales entre China y el mundo occidental se manifiestan en el modo de pensar y también en los patrones de lenguaje". La noción de 'metataxis' puede ayudar a explicar las correspondencias entre tales patrones.