

Hacia un dispositivo informático basado en Corpus para la Enseñanza del Español Lengua Extranjera (DICEELE).

Molina Mejia, Jorge Mauricio, Grajales Ramírez, Andrés Felipe y Pemberty Tamayo, José Luis.

Cita:

Molina Mejia, Jorge Mauricio, Grajales Ramírez, Andrés Felipe y Pemberty Tamayo, José Luis (2019). *Hacia un dispositivo informático basado en Corpus para la Enseñanza del Español Lengua Extranjera (DICEELE)*. *Revista Internacional de Tecnología, Conocimiento y Sociedad*, 7 (1), 1-13.

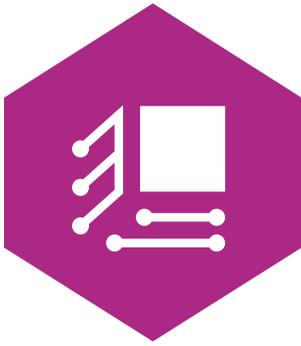
Dirección estable: <https://www.aacademica.org/jorge.mauricio.molina.mejia/50>

ARK: <https://n2t.net/ark:/13683/pqc6/ety>



Esta obra está bajo una licencia de Creative Commons.
Para ver una copia de esta licencia, visite
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>.

Acta Académica es un proyecto académico sin fines de lucro enmarcado en la iniciativa de acceso abierto. Acta Académica fue creado para facilitar a investigadores de todo el mundo el compartir su producción académica. Para crear un perfil gratuitamente o acceder a otros trabajos visite: <https://www.aacademica.org>.



VOLUMEN 7 NÚMERO 1

Downloaded by Jorge Mauricio Molina Mejía on Sat Nov 02 2019 at 19:16:46 PM CDT

Revista Internacional de

Tecnología, Conocimiento y Sociedad

Hacia un dispositivo informático basado en
Corpus para la Enseñanza del Español
Lengua Extranjera (DICEELE)

JORGE MAURICIO MOLINA MEJÍA, ANDRÉS FELIPE GRAJALES RAMÍREZ Y JOSÉ LUIS PEMBERTY TAMAYO

**REVISTA INTERNACIONAL DE TECNOLOGÍA,
CONOCIMIENTO Y SOCIEDAD**

Primera Edición Common Ground Research Networks 2019
University of Illinois Research Park
2001 South First Street, Suite 202
Champaign, IL 61820 USA
Tel.: +1-217-328-0405
www.cgespanol.org

ISSN: 2474-588X (versión impresa)
ISSN: 2174-8985 (versión electrónica)

© 2019 (artículos individuales), autor(es)
© 2019 (selección y contenido editorial), Common Ground Research Networks

Todos los derechos reservados. Excepto propósitos de estudio, investigación, crítica o revisión permitidos bajo la legislación de derechos de autor, ninguna parte de este trabajo puede ser reproducida, en ningún formato, sin el consentimiento explícito por escrito del editor. Para otros tipos de permisos y dudas, por favor, escriba a: suporte@cgespanol.org

La *Revista Internacional de Tecnología, Conocimiento y Sociedad* es una publicación académica arbitrada bajo el proceso de revisión por pares.

Hacia un dispositivo informático basado en Corpus para la Enseñanza del Español Lengua Extranjera (DICEELE)

(Towards a Corpus-based Computing Device for Teaching Spanish as a Foreign Language)

Jorge Mauricio Molina Mejía,¹ Universidad de Antioquia, Colombia
Andrés Felipe Grajales Ramírez, Universidad de Antioquia, Colombia
José Luis Pemberty Tamayo, Universidad de Antioquia, Colombia

Resumen: Presentamos en este artículo, un dispositivo informático para la enseñanza del español a aprendientes extranjeros. Un sistema de aprendizaje de lenguas que se basa en corpus textuales auténticos que han sido previamente seleccionados según los niveles del Marco Común Europeo de Referencia para las lenguas (MCER). Nuestro interés es mostrar la posibilidad que ofrece la lingüística de corpus en el campo de la enseñanza y aprendizaje de lenguas extranjeras y de cómo esta puede trabajarse a partir de enfoques teóricos como el de la lingüística textual.

Palabras clave: aprendizaje de lenguas asistido por ordenador (ALAO), lingüística de corpus, lingüística textual, español lengua extranjera (ELE), lingüística computacional

Abstract: In this article, we introduce a plan for a computer device to facilitate the teaching and learning of Spanish as a Foreign Language. This system is based on a linguistic corpus of texts properly selected according to the levels of Common European Framework of Reference for Languages (CEFR). We expect to show the possibility that corpus linguistics provides to the teaching and learning of foreign languages and how these fields can work together with a theoretical approach as text linguistics.

Keywords: Computer Assisted Language Learning (CALL), Corpus Linguistics, Text Linguistics, Spanish as a Foreign Language (SFL), Natural Language Processing

Introducción

La enseñanza y aprendizaje de lenguas extranjeras, se han ido beneficiando en estas últimas décadas del aporte de las Tecnologías de la Información y de la Comunicación para la Enseñanza (TICE) y, más concretamente, en el campo del Aprendizaje de Lenguas Asistido por Ordenador (ALAO). Podemos decir que existe, además, un interés por parte de los docentes como de los aprendientes² de lenguas en utilizar estas herramientas dentro del proceso de enseñanza-aprendizaje. Por otra parte, el uso de corpus lingüísticos para la enseñanza de lenguas extranjeras también ha sido foco de gran importancia durante varios años, incluso con la realización de congresos internacionales (Pitkowski & Vásquez Gamarra, 2009). Sin embargo, la conjunción de ambas herramientas, la informática y la lingüística de corpus, no es tan frecuente para la enseñanza del Español como Lengua Extranjera (ELE), como sí se da en otras lenguas. Podemos ver algunos ejemplos en lo que respecta al Francés Lengua Extranjera (FLE) (Álvarez Álvarez, 2004; Ardila Restrepo et al., 2012; Molina Mejía, 2015; Molina Mejía & Antoniadis, 2015; Salamanca Lamouroux, 2014). No obstante, somos conscientes de la importancia que ha

¹ Corresponding Author: Jorge Mauricio Molina Mejía, Calle 67 #53-108, Departamento de Letras: Filología Hispánica, Facultad de Comunicaciones, Universidad de Antioquia, Medellín, Antioquia, Colombia. email: jorge.molina@udea.edu.co

² Hablaremos a lo largo de este artículo del término *aprendientes*, término que se utiliza en el campo de la enseñanza y aprendizaje de lenguas extranjeras, tal y como lo propone el autor J.-P. Cuq (2003). Dejando de lado las nociones de *estudiantes* (demasiado general), o de *aprendices* (más relacionada con el campo de la formación técnica o tecnológica).

adquirido la enseñanza de ELE en las últimas décadas, como lo constata el Instituto Cervantes en su más reciente informe³, así este tipo de herramientas no sean todavía usuales en el contexto colombiano⁴.

Este artículo pretende mostrar un proyecto en proceso que se realiza en la Universidad de Antioquia (Medellín, Colombia), que involucra la investigación en la enseñanza y el aprendizaje del ELE y la manera de trabajarlos a través de un dispositivo informático basado en lingüística de corpus. Este dispositivo aplicará actividades didácticas con los fenómenos más recurrentes de la lingüística textual y el análisis discursivo. Para ello se usará un corpus seleccionado, etiquetado (morfosintáctica y discursivamente) y conformado por documentos auténticos que representen el uso real de la lengua española. Gracias a esto se podrán planificar las actividades y la programación del dispositivo para su posterior montaje en una plataforma accesible.

La idea de tal proyecto surge de las investigaciones del profesor J. M. Molina Mejía (2015) en la Universidad Grenoble Alpes (Francia) en colaboración con el profesor G. Antoniadis (Molina Mejía & Antoniadis, 2015), para la enseñanza y el aprendizaje del FLE; bases que se aplicarán en este proyecto, pero enfocadas en el aprendizaje del ELE. A partir de estas bases, consideramos la posibilidad de desarrollar un sistema que ayude a los aprendientes y profesores de ELE con actividades didácticas de manera inductiva. Esto es, hacer que los aprendientes piensen de manera metalingüística y metacognitiva acerca de la lengua, tal como lo proponen Cicurel (1985) y Molina Mejía (2015) para el FLE. Como se dijo anteriormente, este tipo de proyectos para la enseñanza del ELE no son muy frecuentes. Algunos de estos proyectos, enfocados en el español, se han realizado en España y Canadá, principalmente, trabajos hechos en la Universidad Autónoma de Madrid (Campillos Llanos *et al.*, 2006), la Universidad Antonio de Nebrija (Pérez-Ávila, 2006) y de la Universidad de Montreal (Pitkowski & Vásquez Gamarra, 2009), donde se recalca la importancia del uso de corpus para la enseñanza de lenguas.

Reconocemos la importancia que la lengua española ha adquirido en los últimos años, la cual ha desarrollado un incremento en el número de sus hablantes y continúa haciéndolo. Este incremento se refiere, naturalmente, a hablantes de español como lengua materna, pero, sobre todo, a estudiantes extranjeros que pretenden aprenderla. Generalmente, estos extranjeros viajan a países hispanohablantes para aprender la lengua directamente en institutos especializados, rodeados de un ambiente que los sumerge en la lengua. Uno de estos países a los que los extranjeros van a aprender el español es Colombia. A esta necesidad han respondido diferentes universidades e institutos y, por ejemplo, en la ciudad de Medellín se abren cada vez más centros de idiomas en los que se ofrece el ELE, y varias universidades ofrecen estos cursos a estudiantes de intercambio que vienen a Colombia para aprender la lengua. Otro factor a considerar es la existencia de numerosas lenguas indígenas en Colombia; en donde, por ejemplo, si un hablante indígena desea ingresar al sistema de educación superior, debe manejar apropiadamente la lengua española.

Marco teórico y propuesta de investigación

³De acuerdo al último reporte del Instituto Cervantes (2017) acerca del estado del español en el mundo, el grupo de usuarios potenciales del español excede los 571 millones de hablantes. De estos, 477 millones poseen al español como lengua materna y más de 21 millones la aprenden de manera institucional como lengua extranjera (Instituto Cervantes, 2017).

⁴Ya desde comienzos de la década del 2010, se veía esta falta de utilización de instrumentos informáticos en el proceso de enseñanza del ELE en Colombia (García *et al.*, 2011).

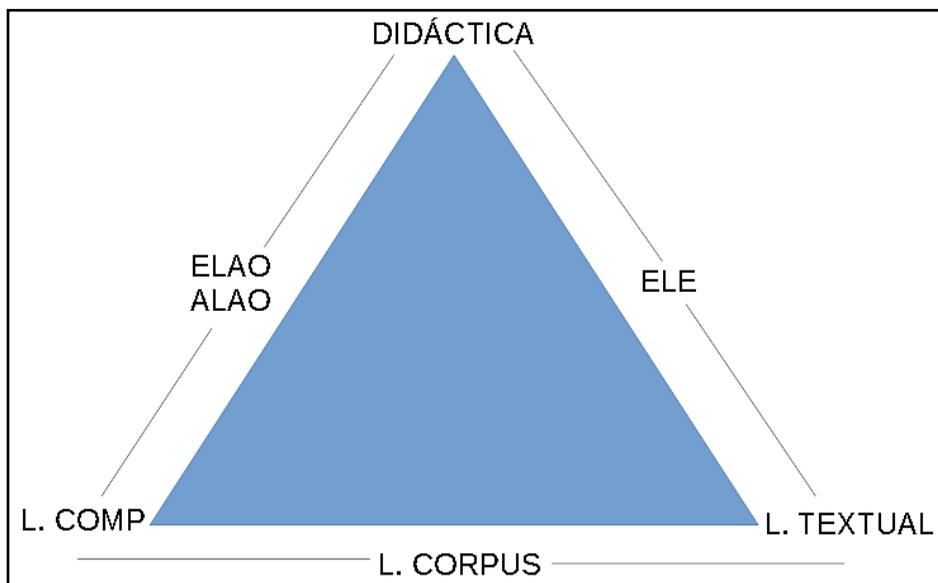


Figura 1: Modelo teórico del proyecto DICEELE
Fuente: Molina Mejía et al. (2017).

El semillero de investigación *Corpus ex Machina*, como parte del GES⁵, se ofrece para promover desde la Universidad de Antioquia el estudio del ELE; al trabajar la lingüística computacional y la lingüística de corpus, el semillero busca maneras de aplicar la lingüística para facilitar el estudio de la lengua.

El principal punto de partida de la investigación es la situación actual de la Universidad de Antioquia, ya que cada año se recibe en ella a muchos estudiantes que llegan al país con la intención de mejorar su competencia en lengua española. Es de esta manera como se propone este dispositivo informático para que pueda facilitar la enseñanza y el aprendizaje del ELE en los cursos de lengua extranjera y segunda lengua que deberían recibir estos estudiantes. Además, no solamente se habla de la población extranjera, ya que el país cuenta con una extensa variedad de lenguas indígenas y dos lenguas criollas.

Los conceptos tratados a continuación son fundamentales para la concepción de este proyecto, cuyas bases se encuentran en la didáctica, la lingüística computacional y la lingüística textual, que, a través del ALAO usan los corpus dirigidos a la enseñanza y aprendizaje del ELE (ver Figura 1).

Dispositivos informáticos para la enseñanza y aprendizaje de ELE basados en corpus

En primer lugar, la idea de trabajar la enseñanza de lenguas a través de sistemas basados en corpus lingüísticos no es nueva. Esta parte de los primeros trabajos de Johns (1991) donde nombra el campo *Data-Driven Learning*⁶ y de los trabajos de Aston (1995), que proponían la utilización de corpus dentro de la enseñanza de lenguas. Estos autores sostienen que los aprendientes de lenguas van a adquirir las estructuras lingüísticas de la lengua-meta gracias al empleo de documentos auténticos. De esta manera, los aprendientes entenderán mejor la lengua y, además, aprenderán a expresarse de una forma adecuada, teniendo para ello el modelo

⁵Grupo de Estudios Sociolingüísticos (GES), Facultad de Comunicaciones, Universidad de Antioquia.

⁶El concepto cambia de acuerdo a Boulton y Tyne (2014): "Aprendizaje basado en corpus".

lingüístico de primer orden: los textos producidos por hablantes nativos y cuyo objetivo no es la enseñanza, sino, informar o ser leídos por otros hablantes nativos. E. Pitowsky (2009), haciendo referencia al *Data-Driven Learning*, afirma que este campo ayuda a “estudiar la lengua integrada en el contexto discursivo, a través de ejemplos reales y precisos de uso; en contraposición al empleo de la introspección y los métodos intuitivos tradicionales” (p.31). Por otro lado, aunque no sea una idea nueva, este campo de trabajo, se ha desarrollado poco en el área del español.

En la actualidad, existe una fuerte tendencia hacia la concepción y desarrollo de sistemas para la enseñanza y aprendizaje de lenguas abiertos y parametrables (Antoniadis, 2008; 2010; Molina Mejía, 2015). Esto es, dispositivos informáticos para la enseñanza de lenguas que permitan al docente preparar y diseñar sus actividades y a los aprendientes realizar dichas actividades con la guía del docente. Sin embargo, se propone también una enseñanza y aprendizaje de tipo híbrido, donde se combinen cursos presenciales con aprendizaje autónomo guiado a distancia (Charlier *et al.*, 2006). Los sistemas de ALAO permiten, así, una utilización que va más allá de la herramienta que se usa en clase como soporte didáctico, transformándose en un instrumento de ayuda al aprendizaje, con la mediación del docente (Rézeau, 2001). Además, desde el punto de vista del docente, puede entenderse el sistema como una ayuda para la planificación de las actividades, así como su posterior evaluación, brindando además una forma muy efectiva de recoger datos sobre los temas que tienen una mayor o menor comprensión por parte de los estudiantes. Según esto, los dispositivos informáticos para la enseñanza de lenguas extranjeras deben responder a estas finalidades didácticas:

- a) Ayudar a los docentes en su proceso de planificación de actividades.
- b) Hacer el aprendizaje de la lengua extranjera más interesante y participativo para los aprendientes.
- c) Permitir un aprendizaje a partir de datos auténticos.

Lingüística computacional y corpus lingüísticos

Como ya se mencionó anteriormente, la búsqueda de este sistema está relacionada con el manejo de un corpus de textos, el cual necesita ser recolectado, almacenado y procesado. Es precisamente en este punto donde juega un papel importante la lingüística computacional, que aporta herramientas para el manejo informático de estas fases de realización, al mismo tiempo que permite la automatización de varias tareas, lo que ofrece, a su vez, un manejo de mayores cantidades de datos sin que esto se traduzca en tiempos demasiado extensos. Una de las principales tareas que cumplen estas herramientas es el etiquetado morfosintáctico; este procedimiento, en particular, es de suma importancia al momento de diseñar sistemas para la enseñanza y aprendizaje de lenguas que funcionen a partir de datos lingüísticos, como lo señala Antoniadis (2008; 2010).

En lo que respecta a las características de los textos es importante diferenciar el tipo de corpus que se utilizará: el corpus especializado. A diferencia del corpus representativo (corpus con gran cantidad de datos provenientes de diferentes medios), los corpus especializados se caracterizan por su variedad y calidad de la muestra (textos variados y etiquetados apropiadamente según su aplicación) y no por su cantidad. Un corpus que se encuentra bajo estas condiciones puede ser utilizado para la enseñanza de lenguas (Teubert, 2009). En este sentido, tenemos en cuenta que un corpus es una colección de textos que se constituyen en una muestra real de la lengua, como proponen desde los años noventa Sinclair (1996) y Leech (1997), y más recientemente McEnery y Hardie (2012) y Souque (2014). Lo importante se constituye en la variedad tipológica de los textos elegidos (argumentativos, explicativos, informativos, etc.) y la variedad de las fuentes de las que se obtengan. Si se tienen en cuenta estos argumentos nos hallaremos frente a un corpus especializado para la enseñanza del ELE.

Una vez que este tipo de corpus se ha conformado se pasa al etiquetado, según las necesidades del sistema. Lo que llamamos etiquetado es la actividad de introducir en el texto unas marcas que ayuden a clasificar sus elementos con el fin de hacer que las características

específicas de cada caso consten en un lenguaje informático, el cual puede ser posteriormente explotado por medio de *software*. Normalmente, se pasa primero por un etiquetado morfosintáctico, para lo cual se utilizan analizadores que realizan este trabajo de forma automática. El etiquetado morfosintáctico consiste en darle a cada palabra del texto una caracterización morfológica en cuanto a lo que representa en términos de género y número o tiempo, modo y aspecto; lo que se denomina etiquetador o analizador morfosintáctico es, pues, un *software* capaz de realizar este proceso de manera automática. Para este etiquetado nosotros utilizamos *FreeLing*, puesto que, gracias a estudios comparativos de diferentes etiquetadores automáticos realizados por *Corpus ex Machina*, *FreeLing* es el más acorde a nuestras necesidades y con un pequeño margen de error para el español.

Posteriormente es necesario introducir etiquetas que ayuden a tratar cuestiones más específicas para el proyecto, que serían, en este caso, los fenómenos textolingüísticos y discursivos. Una herramienta muy valiosa en la elaboración de este segundo etiquetado es el lenguaje XML (*eXtensible Markup Language*), que permite anotar esas nociones (anáforas, catáforas, conectores, marcadores discursivos, etc.) de forma manual, puesto que no contamos con un etiquetador automático que se encargue de estas cuestiones más específicas, en parte por ser estas más complejas de reconocer de manera sistemática y unívoca.

Para terminar el tratamiento informático de los textos es necesario que, después de etiquetar enteramente el corpus, se realice un documento estructural que contenga la sintaxis de todo el corpus; es decir, un conjunto de reglas que le comuniquen a la computadora la manera en la que serán explotados todos los datos introducidos con anterioridad; esto permite el acceso posterior a estos datos por medio de otro *software* y a su vez la verificación de que las etiquetas introducidas se ajusten a las reglas definidas. Este documento se denomina DTD⁷ y se realiza una única vez para aplicarse sobre todos los textos del corpus.

La lingüística textual en la didáctica del ELE

En lo que respecta al enfoque teórico a partir del cual funcionará nuestro sistema, partiremos de la lingüística textual y del análisis del discurso. Puesto que consideramos, al igual que otros autores (Adam, 2011a; 2011b; Chiss & David, 2012; Lundquist, 2013), que estos campos teóricos pueden ayudar en gran medida a los aprendientes de lenguas. Esto porque de esta manera el sistema no se enfoca únicamente en el análisis oracional, sino que va más allá, al análisis textual y por ende discursivo. Acogemos en estos campos los temas propuestos por autores tales como Adam (2011a; 2011b) y Lundquist (2013) y que explicaremos a continuación:

- a) *Coherencia y cohesión textuales y discursivas*: Esta temática comprende todo el trabajo que puede realizarse a partir del estudio de las nociones de referencia y correferencia en los textos, así como en el discurso. Para ello los autores antes citados proponen realizar un trabajo enfocándose en las anáforas y las catáforas, evitando con ello las repeticiones y mejorando el trabajo de los aprendientes de lenguas. Otro de los temas importantes en este ítem es el de los conectores lógico-temporales y argumentativos, los cuales permitirán a los aprendientes de lenguas mejorar la calidad de sus escritos y de su discurso hablado al proveerles los elementos adecuados para expresarse en español.
- b) *Estructuración textual y discursiva*: Adam (2011) y Lundquist (2013) proponen realizar un trabajo en el orden de las partes constitutivas del texto según el tipo de texto o de discurso a realizar. Así, por ejemplo, no se realizará un texto o un

⁷DTD (*Document Type Definition*) o “Definición del Tipo de Documento”, es el documento que provee la información gramatical que contiene el corpus, y que nos sirve, además, para la programación de las actividades del dispositivo informático.

discurso de tipo argumentativo de la misma forma que un texto o discurso explicativo, etc.

- c) *Progresión temática textual y discursiva*: Aquí proponemos que ellos aprendan a construir buenas oraciones y textos en general, así como hablar con corrección a partir de un trabajo realizado según las nociones de tema y rema.

A partir de estos campos textuales y discursivos pretendemos que los aprendientes adquieran nociones gramaticales y a la vez elementos de orden fraseológico (conectores, elementos anafóricos, etc.) que les permitan mejorar su nivel de dominio de la lengua-meta, tanto en lo que respecta a la comprensión de realizaciones complejas encontradas en la vida cotidiana y académica, como a lo que se refiere a sus propias producciones textuales y discursivas.

Hacia la concepción de DICEELE: propuesta metodológica

La sección anterior nos permitió observar, a nivel teórico, los campos disciplinares que sustentan nuestro trabajo, nuestro objetivo en esta nueva sección es el de mostrar la manera de llevar a cabo este proyecto a nivel del desarrollo de los aspectos metodológicos. Para conseguir tal fin, nuestro trabajo se divide en tres grandes fases: a) Conformación del corpus; b) modelización del dispositivo; y c) programación del dispositivo. Estas constituyen, *grosso modo*, el enfoque metodológico que planteamos con el fin de desarrollar y poner en marcha el sistema DICEELE.

Fase 1: Conformación y etiquetado de un corpus lingüístico de textos

Consideramos que tanto la conformación como el etiquetado del corpus, que será la columna vertebral de nuestro dispositivo, son de vital importancia para la correcta realización de las actividades que se propondrán a los aprendientes de ELE. Consideramos que para que exista un corpus rico y variado, y que además sea representativo de la lengua española, debemos hacer un buen trabajo de recopilación de textos y luego de esto un etiquetado que nos permita la correcta utilización de los textos.

A continuación, explicaremos en qué consisten las sub-fases recopilación y etiquetado del corpus para la enseñanza de ELE.

Recopilación del corpus

Esta primera sub-fase consiste en la conformación de lo que será la base para el dispositivo informático: el corpus de textos. Es la fase que tiene por objetivo principal la búsqueda de textos auténticos que sirvan para la enseñanza y el aprendizaje de ELE. Este corpus deberá poseer algunas características específicas y ser tratado debidamente para ser explotado por el dispositivo informático. Como dijimos antes, se trata de un corpus especializado para ELE y por lo tanto necesitamos que posea las siguientes características:

- *Textos de diferentes géneros y extraídos de distintas fuentes*: esto con el fin de privilegiar la variedad y poder diseñar las actividades con los fenómenos lingüísticos que proponemos trabajar con los aprendientes.
- *Textos clasificados según el Marco Común Europeo de Referencia para las lenguas (MCER) dentro de los niveles B1, B2 y C1*: el nivel de dificultad de los textos es esencial para trabajar las nociones lingüísticas textuales y discursivas, puesto que, en estos niveles ya existe un nivel de comprensión no muy básico que permite trabajar con aprendientes de nivel intermedio y avanzado⁸. La dificultad de los textos no la

⁸La selección de los niveles está sujeta a criterios de pertinencia pues, es a partir del nivel umbral (B1) que los textos dejan de estar conectados directamente con situaciones inmediatas y campos semánticos desconectados (familia, ocupaciones) y comienzan a hacer referencia a cuestiones abstractas y a discursos relativamente elaborados.

determinamos nosotros, sino que buscamos textos que ya estén clasificados según los niveles propuestos por el MCER.

- *Textos libres de derechos de autor*: es indispensable que el corpus se encuentre constituido por textos de libre uso, debido a cuestiones presupuestales. Esto ha dificultado en gran medida el proceso de recopilación de textos pues muchos de sitios de Internet que ofrecen textos catalogados, según los niveles antes mencionados, exigen una retribución económica para poder utilizarlos.

Consideramos, entonces, que una vez que se pueda tener un número considerable de textos que den cuenta de diferentes géneros textuales y de distintos usos de la lengua española se puede pasar a la fase de etiquetado que presentamos a continuación.

Etiquetado del corpus

Después de tener conformado el corpus (aproximadamente ciento cincuenta textos de los diferentes niveles propuestos) se pasa a la labor de etiquetarlos en lenguaje XML. Antes de ser etiquetados, los textos son revisados para eliminar todo tipo de elementos que perjudiquen el tratamiento del texto, como signos o etiquetas previas y posibles errores tipográficos, ortográficos o de puntuación, etc., esto con el fin de facilitar el proceso de etiquetado automático. Luego de realizar el etiquetado con el analizador morfosintáctico *FreeLing* es necesario revisar y corregir los posibles errores del etiquetado tanto morfológico como sintáctico, ya que el procesamiento del lenguaje natural no es perfecto y estos analizadores tienen un pequeño margen de error que el humano debe encontrar y corregir con sus conocimientos lingüísticos; lo cual puede convertirse en una labor bastante compleja. Nosotros decidimos no trabajar sobre el error para la enseñanza del ELE en este proyecto, es así como en nuestro corpus no deberían aparecer errores.

Como se dijo con anterioridad, el etiquetado de los fenómenos textuales y discursivos se realiza de forma manual, esto debido a que en la actualidad no existe una herramienta de análisis textual que etiquete de manera automática fenómenos como la correferencia textual o el uso de marcadores lógico-temporales y discursivos, por ejemplo.

De esta manera nos encontramos con un corpus conformado para la enseñanza del ELE, seleccionado cuidadosamente y etiquetado de una forma que permite el diseño de actividades didácticas para los aprendientes.

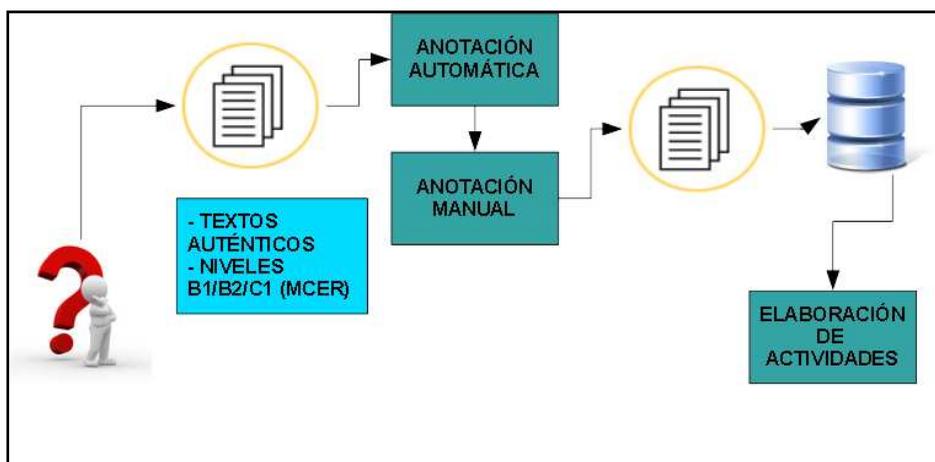


Figura 2: Etiquetado del corpus.
Fuente: Molina Mejía (2015).

Fase 2: Modelización del dispositivo informático

Una vez que tenemos un corpus que se encuentra perfectamente constituido y con todos los fenómenos lingüísticos que deseamos que los aprendientes adquieran, podemos pasar a la fase de modelización del dispositivo informático. En esta fase se planean las actividades que permiten explotar el corpus recogido y previamente etiquetado. Es el momento en el que se diseñarán las actividades que le serán útiles tanto al docente como a los aprendientes.

Entendemos por modelización, el proceso que conduce a la construcción de las diferentes partes que componen un dispositivo informático, utilizando para ello un “modelo lingüístico” y un “enfoque teórico” (en nuestro caso la lingüística textual). Un modelo lingüístico puede ser explicado como la manera en la cual se distribuyen los elementos inherentes a la lengua para conformar una serie de actividades que han sido dispuestas en un escenario pedagógico.

La modelización, en otras palabras, es una forma de abstracción y esquematización de las actividades didácticas. Se abstrae el fenómeno lingüístico para comprender cómo este puede ser aplicado en la enseñanza. Así, entonces, se realiza el diseño de las interfaces de los usuarios para que los docentes puedan realizar las actividades y los aprendientes puedan acceder a ellas. Al final del proceso se evalúan las actividades, como una simulación de lo que harían los usuarios reales, esto con el fin de observar si estas sí cumplen su objetivo de enseñar la lengua o no. Las actividades de formación son planificadas de acuerdo a lo propuesto por Molina Mejía (2015) y se dividen en cuatro partes (Figura 3):

- a) *Actividades de descubrimiento de la noción lingüística a estudiar:* Este tipo de actividades se realiza de una forma inductiva, lo que ayuda a los aprendientes de ELE a descubrir una serie de fenómenos textuales y/o discursivos inherentes a la lengua, y que quizás ellos conocían en su propia lengua.
- b) *Actividades de conceptualización teórica:* Estas actividades hacen referencia a las explicaciones de orden teórico que los docentes van a emplear para explicitar las nociones trabajadas. La idea es proveer unas herramientas que permitan estas explicaciones de una forma sencilla y clara.
- c) *Actividades de sistematización de la noción trabajada:* Una vez la noción explicada a los aprendientes, estos van a ejercitarse para tener una mayor claridad de dicha noción. Para ello, el dispositivo permite realizar una serie de actividades que les permite adquirir de una mejor manera nociones complejas como la correferencia textual (empleo de anáforas y catáforas de orden textual y discursivo), los marcadores lógico-temporales y discursivos, etc.
- d) *Actividades de aplicación de los conocimientos adquiridos:* En la última etapa, los docentes de ELE tendrán la posibilidad de pedirles a sus aprendientes el poner en práctica los conocimientos adquiridos. La idea es que estos últimos puedan enviar a través de la plataforma su trabajo y que los docentes puedan corregirlo y reenviárselo utilizando el mismo dispositivo.

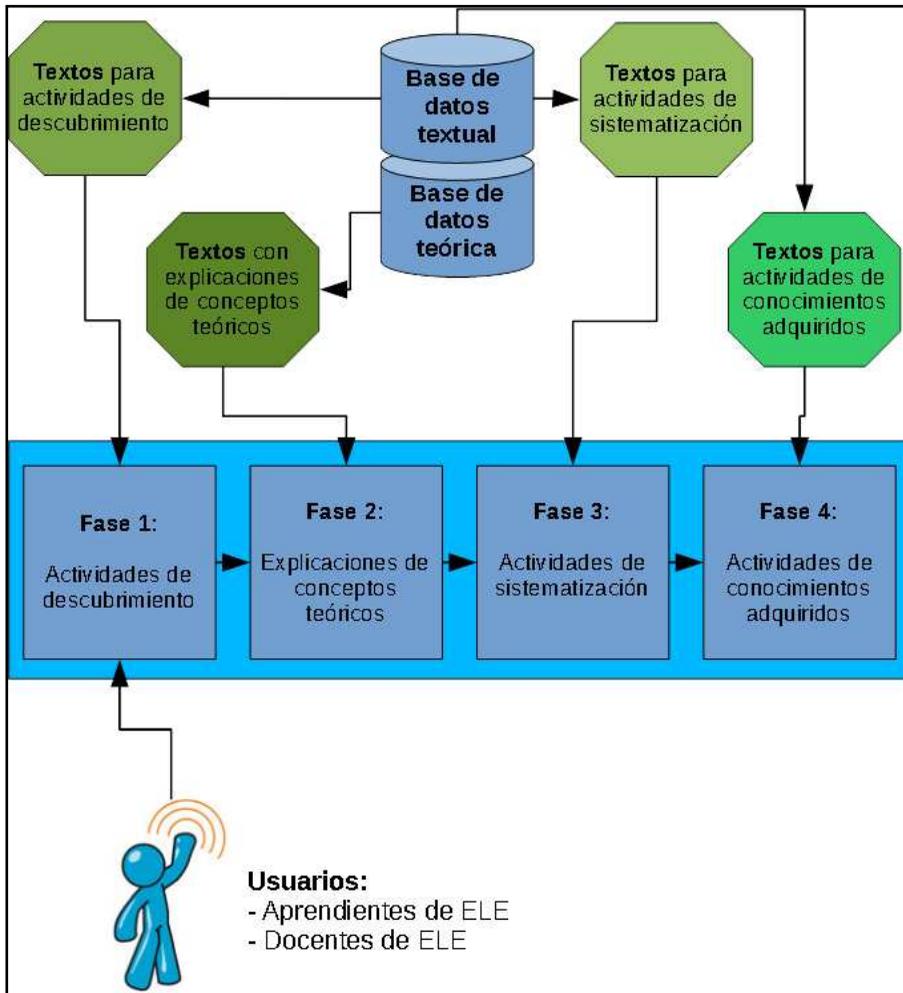


Figura 3: Modelización de las actividades.

Fuente: Molina Mejía (2015).

Un ejemplo de modelización de actividades para el FLE (Molina Mejía, 2015), adaptado al ELE puede ser visto en la imagen anterior (Figura 3).

Como podemos apreciar en la imagen anterior, el sistema se encuentra diseñado para que tanto los docentes de lengua, así como sus aprendientes, tengan acceso a las actividades. De su parte, el docente puede crear las actividades, seleccionando para ello el texto que más le convenga. Él puede, luego, modificar o borrar las secuencias didácticas. A través de una plataforma diseñada para los aprendientes, estos últimos podrán desarrollar de la siguiente manera las actividades propuestas: primero, el dispositivo le muestra al estudiante un texto que posee la noción lingüística, donde se utiliza la noción, por ejemplo, de conectores, pero no le habla directamente de los conectores, es decir, no se utiliza específicamente el metalenguaje de la noción. Entonces, después de que el estudiante puede analizar intuitivamente el texto, se le explica la noción de conectores a través del acercamiento teórico. Lo que permite al aprendiente la sistematización y aplicación de la noción en las actividades que serán evaluadas por el profesor a través del sistema.

Fase 3: Programación y puesta en marcha del dispositivo informático

Una vez que se haya realizado el trabajo de corpus y de modelización del sistema es posible pasar a la etapa de ejecución informática del dispositivo, gracias a su programación y montaje en una plataforma informática. Para esta fase es posible contar con la ayuda de ingenieros o programadores a quienes se les explicará la fase de modelizado para el montaje del sistema, de lo que surgirá:

- *Programación que permita explotar el corpus ya etiquetado:* En tal sentido se hace necesario que la modelización se halle en completa concordancia con el corpus previamente etiquetado. Esto significa el empleo de un documento que contendrá todas las especificaciones de la gramática que manejará en sistema⁹.
- *Interfaces que permitan el acceso de los usuarios al sistema:* Una interfaz que permita a los docentes crear sus secuencias didácticas por medio de las diferentes herramientas que propone el sistema. Una segunda interfaz que permita a los aprendientes la realización de las actividades y el poder descargar los soportes teóricos, además del envío de las actividades al docente.
- *Diferentes fases que lleven a los usuarios de una actividad a otra:* Serie de fases que conduzcan desde actividades con un menor grado de dificultad hacia actividades más complejas. Lo importante es que no haya saltos demasiado enormes en este sentido, para que los aprendientes no se encuentren desorientados al momento de la realización de todas las fases y secuencias didácticas.
- *Retroalimentación para cada actividad:* Consideramos que una buena retroalimentación en cada una de las diferentes etapas de la actividad es de suma importancia. Los aprendientes siempre necesitan una constante evaluación que los vaya guiando a través del recorrido de cada secuencia didáctica. Es por ello importante proporcionarles una evaluación de tipo formativo y no simplemente sumativo.
- *Elementos que permitan el envío de información para ser evaluada y la recepción de correcciones por parte del docente:* El sistema propuesto permite el envío de información y el almacenamiento de dicha información para que tanto docentes como aprendientes puedan acceder a ella.

Finalizadas estas tres etapas, contaremos con un dispositivo computarizado que pueda servir como elemento de apoyo para una clase de ELE. Para comprobarlo, lo ideal sería realizar un test del dispositivo con aprendientes reales, que bien pueden ser los estudiantes de intercambio de la Universidad de Antioquia que vienen a aprender el español o como resultado de convenios con otros institutos o universidades que podrían estar interesados en ser parte del proyecto.

Conclusiones y perspectivas

A pesar de que la investigación sobre uso de corpus para la enseñanza del ELE, a través de dispositivos informáticos, no se ha desarrollado en gran medida, reconocemos la relevancia que el uso de estas nuevas tecnologías puede tener para la enseñanza de lenguas y comprendemos la importancia de la lengua española en el mundo de hoy y el interés que esta lengua despierta en los extranjeros.

También admitimos que este tipo de sistemas para la enseñanza de lenguas son perfectamente aplicables al área del ELE, y que pueden funcionar tan bien como han funcionado para otras lenguas extranjeras, tomando las mismas bases, tal y como lo hemos hecho en el caso del FLE. Esto se debe a que la lingüística de corpus y la lingüística computacional son tomadas aquí como herramientas metodológicas de trabajo y los mismos principios son aplicables a diferentes lenguas. Algo similar ocurre con la base teórica que seleccionamos para tratar la lengua española con estos métodos: la lingüística textual. Así como en el proyecto

⁹Hablamos aquí de la DTD mencionada con anterioridad.

mencionado atrás sobre FLE, también nos basamos en la lingüística de texto y del discurso para trabajar la lengua y esto no afecta en esencia la calidad o los resultados de nuestro trabajo, puesto que los fenómenos y características de la lengua que seleccionamos se dan en estas y en otras muchas más lenguas, ya que la lingüística textual trabaja el lenguaje en su nivel pragmático, y no es tan específico como otros niveles de la lengua (fonético, morfológico, etc.) en los que habría que enfocarse en otras características más propias de la lengua-meta.

DICEELE sería un dispositivo capaz de facilitar y hacer más dinámicas las clases de ELE. Está diseñado para hacer que los aprendientes piensen sobre el idioma que están aprendiendo y fortalezcan sus conocimientos técnicos de la lengua española, pero al mismo tiempo, para que tengan resultados en la realización práctica del lenguaje. Este dispositivo exige, por lo menos, un nivel básico en lengua española; no lo construimos para aprender el lenguaje desde cero, porque son actividades diseñadas para estudiantes de nivel intermedio.

Finalmente, una vez concluido este proyecto, no pretendemos quedarnos con este resultado, ya que la idea de un dispositivo informático basado en corpus para la enseñanza del ELE puede expandirse más allá de las actividades textuales y discursivas. Por mencionar solamente una propuesta de trabajo para continuar con el proyecto DICEELE: el grupo de investigación GES ha realizado múltiples trabajos en el estudio sociolingüístico del español de Medellín a través del corpus Preseca-Medellín, el cual estamos etiquetando para facilitar su consulta. Este corpus podría ser utilizado para el estudio de una variedad del español, estudiando español oral, e incluso la posibilidad de trabajar con audios. Esta sería una futura etapa DICEELE para los estudiantes de ELE de un nivel más avanzado.

REFERENCIAS

- Adam, Jean-Michel. 2011a. *La linguistique textuelle*. París: Armand Colin.
- . 2011b. *Les textes types et prototypes*. París: Armand Colin.
- Álvarez Álvarez, Alfredo. 2004. «Las TIC en el aprendizaje del Francés Lengua Extranjera (FLE)». *Revista de Educación*, 335:497–512.
- Antoniadis, Georges. 2008. *Du TAL et son apport aux systèmes d'apprentissage des langues: Contributions*. Habilitación para dirigir investigaciones. Grenoble: Université Stendhal-Grenoble 3.
- . 2010. De l'apport pertinent du TAL pour les systèmes d'ALAO. L'exemple du projet MIRTO. *2e Congrès Mondial de Linguistique Française (CMLF-2010)*. 12–15 julio. USA: La Nouvelle-Orléans.
- Ardila Restrepo, María Elena., Luz María, Becerra Naranjo y Natalia Isabel Cañas Gil. 2012. «Maestro, traductor y tecnologías de la comunicación en la Escuela de Idiomas de la Universidad de Antioquia: Realidades y retos». *Ikala, revista de lenguaje y cultura* 17 (1): 61–79.
- Aston, Guy. 1995. Corpora in Language Pedagogy: Matching Theory and Practice. In *Principle and Practice in Applied Linguistics: Studies in honour of H. G. Widdowson*, Capítulo 17, edited by Cook, G. & Seidlhofer, B., 257–270. Oxford: Oxford University Press.
- Boulton, Alex, and Tyne Henry. 2014. *Des documents authentiques aux corpus: démarches pour l'apprentissage des langues*. París: Didier.
- Campillos Llanos, Leonardo., Gozalo Gómez, Paula, y Antonio Moreno Sandoval. 2006. «El corpus C-ORAL-ROM en la enseñanza de ELE». *Actas del XVII Congreso Internacional de la ASELE: Las destrezas orales en la enseñanza del español L2-LE*, 1115-28, Logroño.
- Charlier, Bernadette., Deschryver, Nathalie, y Daniel Peraya. 2006. « Apprendre en présence et à distance: une définition des dispositifs hybrides. » *Distances et savoirs* 4: 469-496.

- Chiss, Jean-Louis, y David Jacques. 2012. « Didactique du français et étude de la langue: Linguistique textuelle et didactique du français. » *Le Français aujourd'hui* (Hors-série) : 155–173. Paris: Armand Colin.
- Cicurel, Francine. 1985. *Parole sur parole: le métalangage en classe de langue*. Paris: CLE International.
- Cuq, Jean-Pierre. 2003. *Dictionnaire de didactique du français: Langue étrangère et seconde*. Paris: CLE International.
- García, David., Javier García, y Yudi Buitrago. 2011. “Estado del arte de ELE en Colombia, una mirada holística: metodologías y enfoques, material didáctico, variedades lingüísticas y dificultades en el área”. *Marco ELE: Revista de didáctica ELE*, 13.
- Instituto Cervantes. 2017. El español: lengua viva. En *Informe 2017*: Madrid. URL: https://cvc.cervantes.es/lengua/espanol_lengua_viva/pdf/espanol_lengua_viva_2017.pdf
- Johns, Tim. 1991. From Printout to Handout: Grammar and Vocabulary Teaching in the Context of Data-Driven Learning. In *Classroom Concordancing. English Language Research Journal*, 4 edited by Johns, T. & P. King, 27–45.
- Leech, Geoffrey. 1997. Introducing Corpus Annotation. In *Corpus Annotation: Linguistic Information from Computer Text Corpora*. Capítulo 1, edited by Garside, R.; Leech, G. and McEnery, A, 1–18. New York: Longman.
- Lundquist, Lita. 2013. *Lire un texte académique en français*. Paris: Éditions Ophrys.
- McEnery, Tony y Andrew Hardie. 2012. *Corpus Linguistics: Method, Theory and Practice*. New York: Cambridge University Press.
- Molina Mejía, Jorge Mauricio. 2015. *ELiTe-[FLE]?: Un environnement d'ALAO fondé sur la linguistique textuelle, pour la formation linguistique des futurs enseignants de FLE en Colombie*. Tesis doctoral. Grenoble: Université Grenoble-Alpes.
- Molina Mejía, Jorge Mauricio, y Georges Antoniadis. 2015. Un environnement informatique pour la formation des formateurs en FLE Colombiens, fondé sur la linguistique textuelle. In *Alsic [En ligne]*, “Des machines et des langues”, 18 (2) edited by J.-R. Lapaire & F. Demaizière (eds): URL: <http://alsic.revues.org/2843>.
- Molina Mejía, Jorge Mauricio., Pemberty Tamayo, José Luis, Grajales Ramírez, Andrés Felipe, y Alejandra Bermúdez Cardona. 2017. DICEELE: Dispositivo Informático Basado en Corpus para la enseñanza de ELE. *IX Coloquio Internacional sobre Investigación en Lenguas Extranjeras – CIILE 2017*. 23 de junio de 2017, Granada: Universidad de Granada (España).
- Pérez-Ávila, Elena. 2006. *El corpus lingüístico en la didáctica del léxico en el aula de E/LE*. Memoria de maestría. Madrid: Universidad Antonio de Nebrija.
- Pitkowski, Elena, y Javier Vásquez Gamarra. 2009. El uso de los corpus lingüísticos como herramienta pedagógica para la enseñanza y aprendizaje de ELE. *TINKUY*, 11 (Mayo): 31–51.
- Rézeau, Joseph. 2001. *Médiatisation et médiation pédagogique dans un environnement multimédia: Le cas de l'apprentissage de l'anglais en Histoire de l'art à l'université*. Tesis doctoral. Bordeaux: Université Victor Segalen-Bordeaux 2.
- Sinclair, John. 1996. Preliminary Recommendations on Corpus Typology. In *Document Technique EAGLES (Expert Advisory on Language Engineering Standards)*.
- Souque, Agnès. 2014. *Modèle de vérification grammaticale automatique gauche-droite*. Tesis doctoral. Grenoble: Université Grenoble-Alpes.
- Teubert, Wolfgang. 2009. La linguistique de corpus: une alternative [version abrégée]. En *Semen [En ligne]*, 27. URL: <http://semen.revues.org/8914>.

SOBRE LOS AUTORES

Jorge Mauricio Molina Mejía: Doctor en Informática y Ciencias del Lenguaje por la Universidad Grenoble-Alpes (Francia), docente de cátedra en el Departamento de Letras: Filología Hispánica y coordinador del semillero de investigación *Corpus ex Machina*, Facultad de Comunicaciones, Universidad de Antioquia, Medellín, Antioquia, Colombia.

Andrés Felipe Grajales Ramírez: Estudiante del pregrado en Letras: Filología Hispánica y miembro del semillero de investigación *Corpus ex Machina*, Facultad de Comunicaciones, Universidad de Antioquia, Medellín, Antioquia, Colombia.

José Luis Pemberty Tamayo: Estudiante del pregrado en Letras: Filología Hispánica y miembro del semillero de investigación *Corpus ex Machina*, Facultad de Comunicaciones, Universidad de Antioquia, Medellín, Antioquia, Colombia.

La *Revista Internacional de Tecnología, Conocimiento y Sociedad* ofrece un espacio para el diálogo y la publicación de teorías y prácticas innovadoras que relacionan la tecnología, el conocimiento y la sociedad. Su ámbito de aplicación es interdisciplinar y proporciona un punto de encuentro entre tecnólogos preocupados por los asuntos sociales y sociólogos interesados en la tecnología.

Dirigida a las personas interesadas en la dinámica de las tecnologías sociales y su impacto social, la revista se guía por los ideales de una sociedad abierta en la que la tecnología se orienta a satisfacer las necesidades humanas y servir los intereses comunitarios. Examina la naturaleza de las nuevas tecnologías, sus conexiones con la comunidad, su uso como herramientas para el aprendizaje y su lugar en una "sociedad del conocimiento".

La perspectiva de los trabajos presentados comprende desde los grandes análisis que abordan preocupaciones globales y universales hasta los casos de estudio pormenorizados que se ocupan de las aplicaciones sociales de la tecnología a nivel local.

Los artículos abarcan un terreno amplio, algunas veces de orientación técnica y otras de orientación social; unas veces adoptan una perspectiva teórica y otras una aproximación práctica; a veces reflejan un análisis objetivo y desapasionado, y en otras ocasiones sugieren estrategias para la acción.

La revista resulta de interés para académicos pertenecientes a los campos de la informática, la historia y filosofía de la ciencia, la sociología del conocimiento, la sociología de la tecnología, la innovación, la educación y las humanidades. La participación está abierta a estudiantes, investigadores, desarrolladores de tecnologías, formadores, consultores tecnológicos, etc.

La *Revista Internacional de Tecnología, Conocimiento y Sociedad* es una revista académica sujeta a un proceso riguroso de revisión por pares.