



10. Informática y lingüística: una unión para la enseñanza del español como lengua extranjera¹

María Adelaida Zapata Granados
adelaida.zapata@udea.edu.co

Daniel Taborda Obando
daniel.tabordao@udea.edu.co

Jorge Mauricio Molina Mejía
jorge.molina@udea.edu.co

Resumen

El presente proyecto de investigación, que se inscribe en la línea temática: apropiación social de las TIC para la formación en contextos diversos, aborda los diferentes procesos lingüísticos y didácticos que intervienen en la creación y puesta en marcha de un dispositivo informático para la enseñanza y el aprendizaje del Español como Lengua Extranjera (ELE). Tal dispositivo funciona a partir de un trabajo realizado sobre un corpus lingüístico especializado (nivel B1, B2 y C1 del Marco Común Europeo de Referencia - MCER), conformado por textos propios del lenguaje natural, los cuales debidamente seleccionados y procesados de manera morfológica y sintáctica; además de tener en cuenta los fenómenos más recurrentes tanto de la lingüística textual como del análisis del discurso. Por otra parte, se tratarán las diferentes etapas y actividades que constituirán el enfoque didáctico del proyecto, lo que permitirá mostrar cómo opera Diceele desde sus aspectos didácticos, de enseñanza y aprendizaje del español como segunda lengua dirigido a profesores y estudiantes, tanto extranjeros como los pertenecientes a las comunidades indígenas colombianas que pretendan aprenderlo y enseñarlo. En este sentido, este corpus puede transformarse en un importante insumo en el momento de ocuparse de una de las necesidades culturales más demandadas en la actualidad; la enseñanza de ELE.

Palabras clave: Diceele, lingüística de corpus, lingüística computacional, lingüística textual, didáctica de las lenguas extranjeras.

¹ Dispositivo Informático basado en Corpus para la Enseñanza del Español Lengua Extranjera (Diceele). Proyecto del Semillero de Investigación *Corpus Ex-Machina* adscrito al Grupo de Estudios Sociolingüísticos de la Facultad de Comunicaciones, Universidad de Antioquia y a la Red de Semilleros de la misma universidad (RedSin), Medellín, Colombia. Proyecto en curso con inicio en octubre 2017 y finalización en octubre 2019; financiado por el Comité para el Desarrollo de la Investigación (CODI), Universidad de Antioquia.



Abstract

This research project, which forms the thematic line: social appropriation of ICTs for training in diverse contexts, addresses the different linguistic and didactic processes involved in the creation and implementation of a computer device for teaching and learning of Spanish as a Foreign Language (SFL). Such a device operates on the basis of work carried out on a specialized language corpus (level B1, B2 and C1 of the Common European Framework of Reference - CEFR), consisting of texts of natural language, which are duly selected and processed in a morphological and syntactic manner; in addition to taking into account the most recurrent phenomena of both textual linguistics and discourse analysis. On the other hand, the different stages and activities that will constitute the didactic approach of the project will be discussed, which will show how DICEELE operates from its didactics, teaching and learning aspects of Spanish as a second language aimed at teachers and students both from abroad and those belonging to Colombian indigenous communities, who intend to learn and teach it. In this sense, this corpus can become an important input when dealing with one of the most demanded cultural needs today; SFL's teaching.

Keywords: Diceele, Corpus Linguistics, Computational Linguistics, Textual Linguistics, Didactics of Foreign Languages.

.....

● Introducción

En el campo de la enseñanza y aprendizaje de lenguas extranjeras, el empleo de sistemas y dispositivos informáticos no es para nada nuevo. Es así como cada vez son más los docentes que emplean este tipo de sistemas con el fin de enriquecer sus cursos, ya sea a partir de la utilización de las llamadas TICE –Tecnologías de la Información y de la Comunicación para la Enseñanza (Molina Mejía, 2009), de escenarios pedagógicos informatizados (Mangenot y Louveau, 2006) o de sistemas de ALAO– Aprendizaje de Lenguas Asistido por Ordenador (Rézeau, 2001). Por otra parte, se puede apreciar que los aprendientes de lenguas realizan, cada vez más, una profunda utilización de herramientas que se encuentran en la red tanto en términos de aprendizaje como para la profundización de los temas estudiados de forma presencial en el aula de clase.

No obstante, la mayor parte de los sistemas propuestos para la enseñanza o el aprendizaje de lenguas extranjeras han sido desarrollados, generalmente, utilizando oraciones o frases “inventadas” o “creadas” por los diseñadores de estos sistemas; como es el caso de los ejercicios creados gracias al programa *Hot Potatoes*. El problema principal de este tipo de ejercicios reside en el escaso contacto que tiene el aprendiente con la lengua real (Rézeau, 2001; Molina Mejía, 2015). Es por ello que la idea de crear un sistema fundamentado en un trabajo que tenga como objeto de utilización la lengua real, es decir, aquella que proviene de corpus especializados con



casos reales del lenguaje natural, ha sido poco desarrollada hasta ahora en lo que respecta al campo del Español como Lengua Extranjera (ELE).

Es por lo anterior, que el trabajo que aquí presentamos trata de Diceele: Dispositivo Informativo basado en Corpus para la Enseñanza del Español como Lengua Extranjera, que tiene por objeto la enseñanza y el aprendizaje del español por parte de aprendientes extranjeros y de las comunidades indígenas colombianas en situaciones reales de comunicación. Este proyecto favorece tanto al profesor como al estudiante, en la medida en que les permite integrarse a las nuevas tecnologías informáticas para aprender una lengua. De esta manera, Diceele responde a un momento en el que hay una exorbitante demanda de las Tecnologías de la Información y la Comunicación (TIC), con un programa elaborado con las condiciones necesarias: calidad y autenticidad.

El presente texto se organizará de la siguiente manera: primero se presentará el proyecto Diceele; ¿qué es Diceele?, ¿qué hace?, ¿para qué sirve?, ¿cómo surgió? y ¿en qué se basa? Segundo, el lector dispondrá de la metodología conformada por las herramientas y métodos lingüísticos e informáticos usados para el desarrollo del proyecto, así como un breve apartado sobre didáctica que develará a qué se refiere el enfoque didáctico del proyecto. Y tercero, una explicación de las etapas uno (1) a la cuatro (4) con las respectivas actividades. Este último apartado tiene como finalidad representar los aspectos prácticos de Diceele en cuanto se refiere a las actividades de orden didáctico con las que los aprendientes van a acercarse a la lengua española.

● ¿En qué consiste el Proyecto Diceele?

Diceele es un proyecto que pretende la creación de un dispositivo informático basado en el uso de un *corpus lingüístico especializado* (documentos auténticos) para la enseñanza y el aprendizaje de ELE. Este corpus se encuentra conformado por textos propios del lenguaje natural, que han sido debidamente seleccionados y procesados en lo morfológico y sintáctico, y que abordan, además, los fenómenos más recurrentes de la lingüística textual y del análisis del discurso.

El proyecto encuentra su origen en los trabajos realizados en el área del Francés Lengua Extranjera (FLE) por parte del profesor J. M. Molina Mejía (2014; 2015) y de su trabajo en colaboración con el profesor G. Antoniadis (Molina Mejía y Antoniadis, 2014; 2015) de la Universidad Grenoble Alpes. Desde estas bases, se consideró la posibilidad de trabajar a partir de la anotación de corpus lingüísticos y de desarrollar un sistema que sea de ayuda durante el proceso de formación de aprendientes de lenguas extranjeras partiendo, para ello, de un dispositivo informático que permita poner en marcha actividades de carácter didáctico y formativo, a partir de un enfoque inductivo y deductivo. La idea es que, al aprender con este enfoque didáctico, los aprendientes se vean abocados a pensar de forma metalingüística y metacognitiva la lengua estudiada.



● Metodología

Ahora bien, para el desarrollo global del proyecto se hace necesario optar por un trabajo focalizado en tres fases:

La primera fase es la recolección y el etiquetado del *corpus lingüístico*, para lo cual se requiere:

- Recopilar los textos auténticos que correspondan a los niveles del Marco Común Europeo de Referencia para las lenguas (MCER). El corpus empezaría por los niveles B1, B2 y C1. Lo que equivale a los niveles intermedio y avanzado, en los cuales el aprendiente tiene necesidad de emplear un mayor número de elementos, tanto textuales como discursivos, necesarios para la comunicación en la lengua estudiada.
- Transformarlos en el tipo de archivo que aceptan los etiquetadores morfosintácticos que se emplean en el etiquetado automático. Esto es, el texto plano o documentos en formato **.txt*.
- Etiquetar, es decir, asignar un código en lenguaje informático a cada uno de los elementos morfológicos y sintácticos de todos los textos.
- Revisar y corregir los errores de etiquetado que se hubiesen podido producir para no inducir a los aprendientes a errores de orden lingüístico.
- Anotar de forma manual los fenómenos de tipo textual y discursivo para los cuales aún no existe un etiquetado automático. Aquí entran algunas nociones como la coherencia y la cohesión textual y discursiva, donde actúan unidades como las anáforas, las catáforas, los conectores, los marcadores lógico-temporales y discursivos, etc.

La segunda fase es la modelización del dispositivo informático.

Esto permitirá:

- Diseñar las interfaces de los usuarios, docentes y aprendientes, así los primeros pueden crear las actividades y los segundos pueden acceder a ellas y realizarlas.
- Preparar las actividades de formación, que se dividirán en las cuatro etapas propuestas por Molina Mejía (2015): 1) etapa de descubrimiento de la noción lingüística a estudiarse; 2) etapa de trabajo teórico; 3) etapa de sistematización de la noción; y 4) etapa de aplicación de los conocimientos adquiridos.
- La evaluación de las actividades.

La tercera fase es la programación y puesta en obra del dispositivo informático. Una vez que se ha realizado el trabajo de corpus y de modelización del sistema, se procede a su programación y montaje en una plataforma informática. Para ello se requiere:

- Un proceso de programación que permita explotar el corpus lingüístico ya etiquetado.
- Un diseño de las interfaces que permitan el acceso de los usuarios del sistema.
- Diferentes fases que lleven a los usuarios de una actividad a otra.



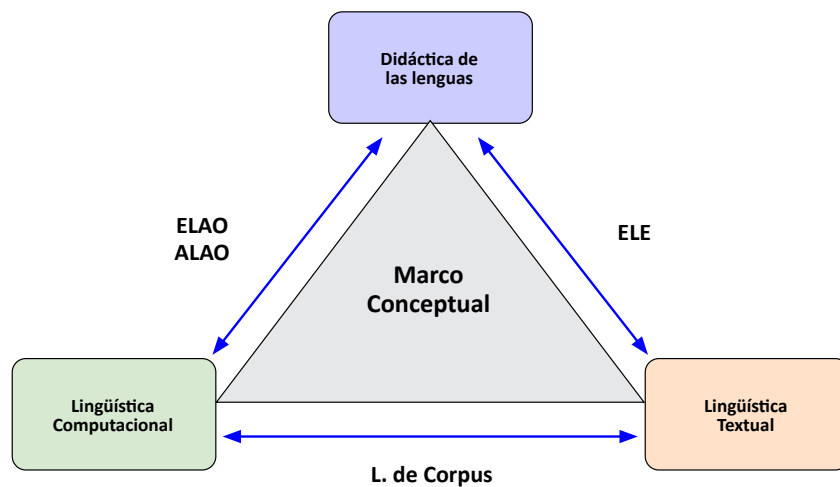
- Retroalimentación para cada actividad.
- Elementos que permitan el envío de información para ser evaluada y la recepción de correcciones por parte del docente.

La finalidad, una vez terminadas las tres fases, es tener un sistema funcional que se pueda evaluar con aprendientes reales.

● ¿Por qué es importante Diceele?

Para entender la importancia de Diceele como un medio innovador para la enseñanza de ELE es necesario abordarlo desde seis ítems: 1) la enseñanza del Español como Lengua Extranjera; 2) la didáctica; 3) la lingüística textual y análisis del discurso; 4) la lingüística de corpus; 5) la lingüística computacional; 6) el Aprendizaje y Enseñanza de Lenguas Asistido por Ordenador (ALAO-ELAO). A continuación, un gráfico que muestra estos elementos:

Figura 1. Marco conceptual del proyecto Diceele



● Enseñanza del Español Lengua Extranjera (ELE)

En primer lugar, se puede constatar el desarrollo que ha ido teniendo en los últimos decenios el campo de la enseñanza de ELE. En este sentido, y como lo señala el Instituto Cervantes de España en su informe del 2017, el español es una lengua que se encuentra en pleno auge. Ello implica que son cada vez más los aprendientes extranjeros que llegan a países de habla hispánica para aprender el idioma. Colombia, al igual que otros países de América Latina, no ha sido



la excepción, nótese entonces la manera en que este país se ha ido posicionando como uno de los destinos predilectos para el aprendizaje del español, luego del final del conflicto armado con algunos grupos insurgentes. Es así como en varias ciudades del país tanto las universidades como los centros de idiomas ofrecen cursos de ELE. Es importante destacar que los aprendientes de español no solo buscan adquirir la lengua, sino que, además, están interesados en aprender más acerca del acervo cultural de los pueblos hispanohablantes, y en este sentido el proyecto Diceele puede ofrecerles ambos elementos. Puesto que cuenta con un corpus especializado en diversidad de temas y variantes dialectales, no es lo mismo hablar del español de Colombia que del español de Chile, España, México, Perú, etc.

En segundo lugar, debemos aclarar que la idea de establecer sistemas de enseñanza de lenguas basados en corpus lingüísticos no es nueva, esto inicia con los primeros trabajos de T. Johns (1991), y del campo por él llamado *Data-Driven Learning* (o aprendizaje sobre corpus, según Boulton y Tyne, 2014), y de los trabajos de G. Aston de mediados de los años noventa (1995) que proponían la utilización de corpus lingüísticos dentro de la enseñanza de lenguas. Estos autores sostienen que los aprendientes de lenguas van a adquirir las estructuras lingüísticas de la lengua objeto, gracias al empleo de documentos auténticos. Esto les ayudará no solamente a entender mejor la lengua, sino a expresarse de una forma correcta y adecuada a partir de esta; teniendo para ello un modelo 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.

Es importante destacar que este campo de trabajo de la enseñanza a partir de la lingüística computacional y de corpus, y que en lenguas como el inglés y francés tiene un gran desarrollo en países como Estados Unidos, Inglaterra y Francia, ha sido más bien poco desarrollada en el área del español. En Colombia apenas se está explorando desde hace unos cuantos años la noción de corpus realizados a partir de sistemas informáticos. No obstante, en este momento el interés está puesto en las posibilidades de incluir estos sistemas en el campo de la enseñanza y aprendizaje de lenguas. Diceele toma en cuenta los trabajos citados anteriormente (Molina Mejía, 2014; 2015; Molina Mejía y Antoniadis, 2014; 2015) en el área de la formación de docentes de francés puesto que, de allí se busca su aplicabilidad para el ELE.

Además, se resalta que la lingüística computacional proporciona las herramientas para realizar un tratamiento informático de las lenguas naturales de una manera más rápida y con un cierto grado de precisión. Este tratamiento posibilita obtener corpus anotados y etiquetados de manera morfológica y sintáctica, lo que es imprescindible en el diseño de sistemas para la enseñanza y aprendizaje de lenguas que funcionen a partir de corpus lingüísticos, como lo señala Antoniadis (2008; 2010). Por otra parte, los dispositivos informáticos que resultan de estos procedimientos aplicados sobre corpus se deben realizar con tres finalidades:

- Ayudar a los docentes en su proceso de planificación de actividades de una forma fácil.
- Hacer el aprendizaje de la lengua extranjera más interesante y participativo por parte de los aprendientes.



- Permitir un aprendizaje a partir de datos auténticos, es decir, de textos que den cuenta del uso real de la lengua en la que se expresan los hablantes.

Sin embargo, es importante que un corpus especializado, como este, se distinga de los grandes corpus (llamados representativos) en lo que respecta a su variedad y a la calidad de la muestra (textos variados y un etiquetado que tenga en cuenta todos los fenómenos de la lengua a estudiarse). 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 J. Sinclair (1996) y G. Leech (1997), y más recientemente T. McEnery y A. Hardie (2012), A. Souque (2014) y Molina Mejía (2015). Un corpus que se encuentra bajo estas condiciones puede ser utilizado para la enseñanza de lenguas. Lo importante está constituido por la variedad de los textos elegidos (textos argumentativos, explicativos, informativos, etc.), la variedad de las fuentes (periódicos, revistas de divulgación científica, extractos de novelas, ensayos breves, etc.). Si se tienen en cuenta estos argumentos, consideramos que tenemos al frente un verdadero corpus especializado para la enseñanza del ELE.

Una vez que un corpus con dichas características ha sido recolectado, el paso a seguir es el etiquetado según las necesidades del sistema. Normalmente, se pasa primero por un etiquetado morfológico y sintáctico que se realiza gracias a los analizadores o etiquetadores automáticos (*TreeTagger*, *FreeLing*, etc.). Luego, un etiquetado manual a partir de lenguajes de balizado como XML (*eXtensible Markup Language*) que permite anotar las nociones que aún no se pueden anotar de forma automática (anáforas, catáforas, conectores, marcadores discursivos, etc.). Una vez que se ha etiquetado todo el corpus se crea un documento estructural que contiene la sintaxis de todo el corpus, este documento llamado DTD (*Document Type Definition*) permite validar todos y cada uno de los textos que conforman el corpus.

En tercer lugar, existe en la actualidad una fuerte tendencia hacia la concepción y el desarrollo de sistemas para la enseñanza y el aprendizaje de lenguas que son abiertos y parametrizables (Antoniadis, 2008; 2010; Molina Mejía, 2014; 2015). Esto es, dispositivos informáticos para la enseñanza de lenguas que deberían permitirle al docente preparar sus actividades y a los aprendientes realizarlas con la orientación del docente. Para ello, se proponen sistemas adecuados para una enseñanza de tipo híbrido, en la que se combinan cursos presenciales (en el aula de clase, o en la sala de informática), y en autonomía guiada (el docente guía el proceso a distancia) (Charlier *et al.*, 2006). Los sistemas de ALAO permiten, así, una utilización que va más allá de la “simple” herramienta que se utiliza en clase como soporte didáctico, transformándose en un instrumento de ayuda al aprendizaje, con la mediación del docente (Rézeau, 2001).

Queda claro que, el proyecto Diceele se enmarca dentro de unas dinámicas que se hacen cada vez más complejas de ignorar, el aprendizaje y la enseñanza de ELE a partir de herramientas que utilizan textos auténticos realizados por hablantes nativos como fundamento de su corpus. Diceele constituye, por lo tanto, una herramienta que puede hacer frente a las necesidades, cada vez más imperiosas, del proceso del aprendizaje del español, con la utilización de sistemas de enseñanza híbrida (presencial o a distancia).



● ¿Qué se entiende por didáctica de las lenguas?

La didáctica, en un principio, consolidada por Comenio en el siglo XVII, se entendió como “el arte de enseñar”. No obstante, un entendimiento más contemporáneo la define como una subdisciplina de la pedagogía que reflexiona los procesos de aprendizaje. Nosotros, sin embargo, nos quedaremos con la primera definición, ya que esta muestra la intención didáctica de manera más precisa en la consolidación de este proyecto. Comenio dice: “No requiere otra cosa el arte de enseñar que una ingeniosa disposición del tiempo, los objetos y el método” (1998. p. 36). Es decir, para enseñar han de usarse no sólo los conocimientos, sino todas las herramientas (objetos), las condiciones (tiempo) y una manera precisa e innovadora que optimice dicha enseñanza (el método).

La didáctica contemporánea se ha convertido en una suerte de reflexión teórica de difícil resolución donde es complejo establecer puntos comunes o acuerdos. Hoy la didáctica podría definirse así:

[...] una subdisciplina de la pedagogía que se ocupa de investigar y reflexionar de un modo teórico, práctico y aplicado sobre –las situaciones de– enseñanza que involucran, obviamente, los procesos de docencia y aprendizaje y que tienen como propósito más elevado la formación (Bildung). (Runge, 2013, p. 204)

El abordaje de esta didáctica como reflexión puede ayudar al maestro a desempeñar y articular acciones encaminadas hacia la mejora de su actividad de enseñanza. Este enfoque se pregunta acerca del qué enseñar, cómo, por qué enseñar tal cosa y no otra, dónde enseñar, cuándo, con qué métodos, entre otras (Runge, 2013). Se puede concluir en este punto que la didáctica desde esta perspectiva alcanza todo su potencial en contextos educativos escolares donde el maestro tiene que pensar y repensar su ejercicio en aras de la educación.

No obstante, como ya se ha dicho, no es ese el concepto de didáctica que se trabajará aquí. Si bien la herramienta Diceele también es pensada para maestros inmersos en educación escolarizada, también es funcional para autodidactas y comunidades indígenas que quieran aprender español como segunda lengua. Por tanto, sería un error pensar que se agota en el simple hecho de enseñar en instituciones educativas a cargo de los profesores de español. Para hacer la diferencia más clara, la idea es que los aprendientes de español se vean abocados a reflexionar de manera metalingüística la lengua objeto, no a cavilar sobre el proceso de aprendizaje y enseñanza de esa lengua. Tal tarea le compete a la pedagogía y a la filosofía, no al aprendiente de ELE.

Diceele trabaja más bien con la didáctica aplicada a las tareas de aprender español, en cómo los ejercicios de aprendizaje se tornan en secuencias prácticas con las que el aprendiente puede interactuar, acercarse al objeto, acercarse a la teoría y acercarse a la práctica del idioma. No ahondaremos en estas secuencias ya que son un asunto que se tratará en detalle más adelante. Lo importante en este punto es aclarar que la naturaleza didáctica del proyecto es práctica y reflexiva hacia el objeto lengua, en este caso el español, pero no reflexiva hacia el proceso. Si así



fuera, Diceele sería un proyecto cuyo aporte se decantaría al área de la pedagogía con implicaciones lingüísticas acerca de la enseñanza de una lengua, y no como es realmente: un trabajo de la lingüística computacional y textual que usa cierta disposición de herramientas brindadas por la informática para facilitar el aprendizaje de español como segunda lengua, cuya organización de objetos y actividades se les ha presentado de una “manera didáctica”.

En resumen, la didáctica de las lenguas como subdisciplina de la pedagogía es, en palabras breves, una reflexión acerca del enseñar y el aprender. Un enfoque que dista un poco de este proyecto: una disposición de objetos, tiempo y método para lograr acercar el saber al aprendiente. Queda pues claro que el proyecto Diceele trabaja la didáctica desde una perspectiva práctica y reflexiva de la lengua que se va a aprender y no teórico-reflexiva del proceso mediante el cual la lengua fue aprendida.

● Lingüística de corpus

Se parte desde la siguiente cita para explicar el contexto histórico que exigió la emergencia de una metodología que permitiera el estudio de la lengua teniendo en cuenta el contexto, por ende, la variación de esta. Sinclair (1991) ilustra la reducción que realizaba el enfoque generativista del lenguaje, como una visión idealizada:

Sedienta por falta de información adecuada, la lingüística languideció –de hecho– se volvió totalmente introvertida. Se hizo una moda mirar hacia adentro de la mente más que hacia la sociedad. La intuición se volvió la clave y se enfatizó la similitud de la estructura del lenguaje y varios modelos formales. El rol comunicativo del lenguaje fue escasamente mencionado. (p.1)

El generativismo negó el estudio de la lengua en su uso real y lo redujo al análisis de su producción y comprensión como facultad innata. La idea del generativismo surgió en Estados Unidos en 1960 como un movimiento lingüístico. Noam Chomsky creó la gramática generativa: un sistema de análisis del lenguaje que situó la sintaxis en el centro de la investigación lingüística y con la que cambió por completo la perspectiva, los programas y métodos de investigación en el estudio del lenguaje, actividad que lo elevó definitivamente a la categoría de ciencia moderna. Una gramática generativa se distingue de las gramáticas tradicionales que se caracterizan por su carácter prescriptivo. Es así como en este marco histórico surgió una fuerte preocupación por la ausencia del estudio del uso de la lengua, un estudio descriptivo que fuese más allá de la suposición y con esto, el interés en los estudios basados en corpus.

Así, la lingüística de corpus es un método de estudio de la lingüística para analizar la lengua a partir de su uso real. Esto se realiza utilizando un conjunto de datos organizados, denominado corpus lingüístico y que, en conjunto con la lingüística computacional, posibilitan el manejo de considerables cantidades de información. Se resalta aquí que las anteriores metodologías no son lo mismo, pero sí trabaja de manera colectiva. La lingüística de corpus es utilizada para la inves-



tigación en diferentes niveles de la lengua: sintáctico, morfológico, fonético, entre otros, por lo que ofrece múltiples aplicaciones, en este caso (Diceele), la enseñanza de ELE. En este sentido, una de sus características es la operación con un enfoque metodológico diverso, no restrictivo.

● Corpus lingüístico

La evolución de la ciencia y la tecnología ha permitido el avance de la informática, esta surgió como el centro de atención sobre el cual giraron cada uno de los desarrollos tecnológicos del siglo pasado. Se puede definir, entonces, como la ciencia que estudia la información, y los medios de automatización y de transmisión para poder tratarla y procesarla. Se podría decir que la materia prima de la informática es la información, mientras que su objetivo formal es el tratamiento de la misma. Por esto, la informática asociada con la lingüística posibilita la creación de corpus lingüísticos. Los corpus lingüísticos son unos insumos que permiten tener cantidades extensas de información; a partir de textos que posibilitan a los investigadores del área de la lingüística tener volúmenes significativos para el estudio del discurso, del uso real de la lengua, lo que posibilita analizar los fenómenos que se presentan en la lengua teniendo en cuenta el contexto y sus particularidades. De esta manera se evita caer en la formulación de casos supuestos, o inventados por el docente. Si bien el uso de los corpus lingüísticos ha contribuido al estudio de diferentes áreas que hacen parte de la lingüística, con Diceele el corpus se encuentra orientado hacia la enseñanza de ELE, esto es, con un fin pedagógico.

El significado de la palabra *corpus* proviene del latín y es definida por varios diccionarios, entre ellos: *Gran Diccionario de Uso del Español Actual*, el corpus es “conjunto de textos, procedentes del lenguaje oral o escrito o de ambos, recopilados de fuentes variadas y ordenados y clasificados según determinados criterios, de tal manera que, sobre ese conjunto, es posible realizar estudios e investigaciones lingüísticas o literarias.” También, el *Diccionario de la Lengua Española de la Real Academia*: “conjunto lo más extenso y ordenado posible de datos o textos científicos, literarios, etc., que pueden servir de base a una investigación”, teniendo en cuenta esta última definición, es importante aclarar que no necesariamente un corpus debe ser realizado de la manera más extensa posible, eso depende del propósito que se tiene con dicho corpus.

Teniendo en cuenta lo anterior, las características que posibilitan las particularidades de los corpus para el estudio del estado de la lengua; de su uso real (descriptivo), son: variación geográfica, variación generacional, variación de tipo de discurso y variación de género. Esto con el objetivo de hacer más detallado el corpus para el análisis de los casos o fenómenos que tiene la lengua, pues debido a su naturaleza es difícil separarla de un contexto y de su uso en los diferentes discursos y formas de este. En particular, los corpus reflejan el contexto en el que se utiliza la lengua e intentan ser un modelo de la realidad lingüística, muestran el uso que sus hablantes hacen de ella. Es decir, procuran ser representativos de una lengua, o de una variedad de ella. Torruella y Llisterra (1999) señalan que para que un corpus sea representativo del funcionamiento de una lengua natural, es necesario que el corpus sea neutro. Aunque reconocen que dicha



neutralidad no corresponde siempre con la realidad en la medida en que “siempre dirigimos la mirada o el pensamiento hacia aquello que consciente o inconscientemente, queremos ver o demostrar” (p. 2).

De manera general, los corpus sostenidos con programas informáticos son una herramienta eficaz y rápida en la búsqueda tanto de una palabra, un conjunto o una serie de palabras en un contexto determinado. Los programas permiten explorar, en particular, el uso de frases y vocabulario utilizado en un área específica, el nivel de uso, la frecuencia o variación de un lema en contextos reales provenientes de diversos tipos de textos, analizar la colocación de las palabras, obtener muestras de cuestiones gramaticales, así como el uso real de una palabra o expresión en un país determinado, en la obra de un autor o en un cierto período de la historia del español, entre otros. Es por todo lo anterior, que, por medio de los corpus, se puede enseñar la lengua a través de ejemplos reales de uso, porque esta herramienta informática ofrece el contexto cultural y situacional de los términos.

● **Lingüística textual y el análisis del discurso**

Si bien el término de “lingüística textual” se le debe al lingüista de origen rumano Eugenio Coseriu, a mediados de los años 1950, no será hasta algunas décadas más tarde cuando otros estudiosos del lenguaje verán en este enfoque una posibilidad para la enseñanza y el aprendizaje de lenguas (Combettes, 1983, Adam, 1990; Lundquist, 1990). Y es que la lingüística textual es un enfoque teórico que toma el texto y no la oración como su unidad de trabajo (Lundquist, 1980; 1990; Adam, 1990; 2011). Es así como las oraciones adquieren un papel preponderante, ya no al ser analizadas desde una perspectiva estructuralista (de sus componentes), sino, al ser estudiadas como parte de un todo, es decir, como unidades que conforman el texto y que establecen en él relaciones de contextualidad y cotextualidad.

En este sentido, tanto lingüística textual como análisis del discurso se encuentran intrínsecamente ligados entre sí. Ya que es imposible desligar el discurso hablado del discurso escrito. Puesto que ambos son creaciones del ser humano y en estos se está haciendo parte de la concepción del mundo, que tienen los individuos cuando se comunican entre ellos.

Dentro de las nociones importantes que tienen que ver con la lingüística textual tenemos:

- La coherencia y la cohesión textuales.
- La correferencia textual.
- La progresión temática.
- Las tipologías y los géneros textuales.

Estas son las nociones que nutren las actividades propuestas en el marco del proyecto Di-ceede. En primer lugar, la coherencia y la cohesión textuales son importantes para la correcta



comprensión del texto o del discurso que se está produciendo. En este sentido, el sistema permite a los aprendientes trabajar las partes del texto, el orden de las oraciones en los párrafos, la secuencialidad de las ideas, gracias a los conectores, etc. En segundo lugar, el sistema permite que los aprendientes descubran fenómenos como la correferencia textual que tiene que ver con el empleo de referentes y correferentes (anafóricos y catafóricos) de orden textual y discursivo. En tercer lugar, tenemos el tema dedicado a la progresión temática, que se encuentra ligada a dos fenómenos: el tema y el rema; y a tres tipos de progresión, según la manera en la que se aborden estos dos fenómenos: lineal, constante o con temas derivados. Finalmente, la cuarta parte se encuentra ligada a los tipos de texto y a los géneros textuales; pues consideramos importante el hecho de ofrecer a los aprendientes diversidad de géneros textuales (narración, argumentación, exposición, etc.), así como variedad en los tipos de texto (periodístico, novela, cuento, científico, etc.).

● **Lingüística computacional**

Para dar una definición más precisa de la lingüística computacional es necesario abordarla a partir de sus elementos constituyentes: “lingüística” y “computacional”. La lingüística es la ciencia que estudia, analiza y examina un gran sistema llamado lengua, tanto el uso real de esta como sus fenómenos históricos, sus variaciones contextuales y su evolución a través del tiempo.

La lingüística tiene cuatro niveles formales de estudio. El nivel fonético y fonológico, el morfológico, el sintáctico y el semántico. El primero estudia todo lo relacionado con los sonidos tanto producidos por los órganos fonoarticulatorios (fonética) como los sonidos ideales o funcionales dentro de cada lengua (fonología). El morfológico estudia la forma de las palabras, es decir, estudia sus elementos constituyentes (morfemas) para esclarecer el devenir histórico de las palabras. El tercer nivel es el sintáctico, estudia la relación entre los elementos constitutivos de la oración. Y, por último, el nivel semántico que se encarga del significado de las palabras y la oración.

Aclarado esto, se puede hablar de “computacional”. Según el diccionario de la *Real Academia Española*, computacional se define como: perteneciente o relativo a la computación, o bien, dicho de un estudio o de un proceso: Que se adapta a ser tratado mediante computadoras. De esa manera y teniendo en cuenta todo lo anterior, se puede afirmar que la lingüística computacional es la aplicación e incorporación de la informática a los procesos y manejos del lenguaje humano. En otras palabras, tratar la lengua desde el lenguaje de un ordenador.

Ahora bien, la lingüística computacional, a partir de su relación con la lingüística, puede ser tratada como una subdisciplina de la *lingüística teórica*, en la medida en que elabora modelos formales del lenguaje humano que se implementan de manera informática. También, la lingüística computacional se vincula con *lingüística aplicada*: una rama de la lingüística que lleva los resultados y métodos de la investigación a campos como la enseñanza, la interpretación y la traducción, entre otras (Guinovart, 1998). Así las cosas, la lingüística computacional es parte de la



sociedad, parte de la industria en cuanto se accede a una base de datos sobre conceptos, cuando se traduce un texto desde el ordenador, en el corrector ortográfico del teléfono móvil, etc.

● **Aprendizaje y Enseñanza de Lenguas Asistida por Ordenador (ALAO-ELAO)**

Se trata de una unión de la lingüística y de la informática pensada desde la didáctica y que hace uso de las Tecnologías de la Información y la Comunicación, que posibilita la creación de herramientas con las condiciones para el servicio de los aprendientes. El ALAO-ELAO está constituido por actividades que se dividen en etapas y que hacen parte de un proceso de aprendizaje que implica un orden lógico e intencionado, diseñado para que el alumno en cada etapa adquiriera elementos que le permitan desarrollar habilidades y continuar con el proceso. Las etapas contienen diferentes herramientas que apoyan el propósito de cada una, inicialmente vídeos o juegos que activen nuevos conocimientos o activen los saberes previos, si los tienen, en este caso, Dicele es un dispositivo que exige saberes previos del español, puesto que parte del nivel B1 del Marco Común Europeo de Referencia.

Es importante aclarar que el ALAO-ELAO, en este dispositivo, no tiene el propósito de despojar al docente de su campo pedagógico, es una herramienta que apoya su enseñanza para un óptimo aprendizaje. Si bien el ordenador, al igual que el profesor, tiene la información, el docente debe estar presente para el apoyo del proceso, esto es muy importante para la llamada época de la información, puesto que esta se encuentra de manera exorbitante en la red. Con este nuevo espacio educativo se da paso a la autonomía del estudiante para que disponga de su tiempo para la creación de nuevos contenidos.

● **Dicele hacia un enfoque didáctico**

Dicele es un proyecto compuesto de diferentes actividades que posibilitan la interacción del aprendiente con la herramienta informática. A continuación, se presentan en un esquema las actividades que propone el dispositivo informático (Figura 2):

Las actividades están diseñadas a partir de dos bases de datos, una textual y otra teórica. La primera consiste en el almacenamiento del corpus especializado. De esta base se extraen los textos que son utilizados para la formulación de las actividades de la etapa uno (1), tres (3) y cuatro (4). En donde la etapa uno (1) consiste en las actividades creadas para que el aprendiente realice ejercicios a partir del descubrimiento, es decir, el usuario infiere la función de una palabra dentro del texto. Esto se da a gracias a la interfaz del sujeto con el programa: una relación que permite poner en práctica los conocimientos previos, e incluso desde el desconocimiento de la lengua a trabajar, gracias a sus fenómenos particulares.



Figura 2. Esquema de actividades de Diceele



Fuente: elaboración propia

Aquí, la didáctica se lleva a cabo cuando se le posibilita al estudiante darse cuenta de sus dificultades al momento de poner en práctica la lengua estudiada. A esto hace referencia el término descubrimiento, no solo a lo que sabe, sino a eso que no sabe, por lo que se espera que se motive a continuar el proceso de aprendizaje de la mano del dispositivo. La etapa tres (3) es la que corresponde a la práctica. Aquí, el estudiante aplica los conocimientos adquiridos después de una teoría ofrecida por la etapa dos (2), la que se explicará más adelante. En esta tercera etapa se da una interacción entre el aprendiente y el programa, ya que la actividad diseñada ostenta unos ejercicios para que el usuario los desarrolle y luego el sistema corrige los posibles errores, a través de un proceso de retroalimentación que puede ser seguido por los aprendientes.

Por ejemplo, del siguiente párrafo del cuento *Espuma y nada más* de Hernando Téllez, extraído del corpus, correspondiente al nivel C1 según el MCER:

Tomé la navaja, levanté en ángulo oblicuo las dos cachas, dejé libre la hoja y empecé la tarea, de una de las patillas hacia abajo. La hoja respondía a la perfección. El pelo se presentaba indócil y duro, no muy crecido, pero compacto. La piel iba apareciendo poco a poco. Sonaba la hoja con su ruido característico, y sobre ella crecían los grumos de jabón mezclados



con trocitos de pelo. Hice una pausa para limpiarla, tomé la badana, de nuevo yo me puse a asentar el acero, porque soy un barbero que hace bien sus cosas.

Se plantean ejercicios como: el pronombre en tercera persona del femenino singular, resaltado en el párrafo hace referencia a:

- a. El ruido característico
- b. Los grumos de jabón
- c. La hoja
- d. La piel

Donde la respuesta correcta es la hoja, que corresponde a la opción C. Se recuerda que son ejercicios textuales, es decir, identificar las palabras que tienen una función dentro del texto, en este caso, que hace referencia a otra palabra mencionada con anterioridad (función anafórica). Ahora bien, si el aprendiente eligiese una respuesta incorrecta, el sistema indicaría el posible error para que esta respuesta sea reevaluada. De ese modo, sistema y usuario intervienen conjuntamente en un proceso de retroalimentación. Nótese que el enfoque didáctico se evidencia en los mecanismos que posee el dispositivo para instruir al estudiante cuando yerra en los diferentes casos del uso de la lengua. Además, como se mencionó con anterioridad, es importante resaltar que el corpus posee variedad de la lengua española, por lo que también se aprenderá sobre la arbitrariedad de esta.

En la etapa cuatro (4) se realiza la evaluación cuantitativa de lo que el estudiante ha aprendido durante todo el proceso. En este punto, la naturaleza de la actividad evaluadora no posibilita un proceso de retroalimentación que permita al estudiante saber por qué se ha equivocado. Dicho de otra forma, no hay una explicación conceptual ni una sugerencia. De manera que la etapa cuatro es la más formal y estructural de todo el proyecto puesto que se agota en acreditar una respuesta correcta y resaltar una incorrecta. Ahora bien, teniendo en cuenta el ejercicio anterior, si el estudiante elige la opción D, por ejemplo, el programa se limitará a señalar el error con una convención para dichos casos, a diferencia de la etapa tres (3), en la que sí existe un proceso de retroalimentación.

Ahora se explicará la segunda base de datos, que contiene la teoría que posibilitará llevar a cabo la etapa número dos (2), puesto que de esta se selecciona el contenido conceptual que exige el tema y el nivel de aprendizaje. Por ejemplo, los deícticos en el nivel B2 del MCER. De acuerdo con esto, de esta base se desprenden las actividades de la etapa dos (2), que en este caso consisten en la explicación de los conceptos de acuerdo con el nivel. En B1, por mencionar un caso, expresar acciones habituales, describir situaciones pasadas, expresar una acción ocurrida en una unidad de tiempo terminada o hablar de una acción que sucedió una sola vez. En B2, los usos de ser y estar: ser / estar + adjetivo (bueno, horrible...) / adverbio (bien/mal). Ser de buena educación/estar mal visto + infinitivo)/ser lógico/natural. Y en C1, por ejemplo, elementos compositivos que se asocian con sustantivos: prefijos y sufijos de uso frecuente, adjetivos con ambos géneros, construcciones comparativas del tipo más/menos de, más/menos que y cuanto



más/menos..., usos de ser y estar+adjetivo calificativo. Y entre otros muchos ejemplos de cada nivel. Lo importante en esta etapa es enseñar la teoría que se aplicará en la etapa tres (3).

● Conclusión

Es bien sabido que el aprendizaje de una lengua constituye un proceso largo y difícil, en el que cada aprendiente demuestra diferentes aptitudes que aceleran, mantienen o retrasan la adquisición de la lengua que se está estudiando. Y es en ese punto que las Tecnologías de la Información y la Comunicación (TIC) ofrecen varias ventajas al momento de afrontar este proceso, como brindar una interfaz que se adapte a la individualidad y al ritmo de aprendizaje del aprendiente, mejorar sus habilidades de comprensión lectora y auditiva, adquisición de vocabulario y pericia en la pronunciación.

Es por lo anterior, que el proyecto Dicele pretende ser una herramienta pertinente y competente a la hora del aprendizaje y enseñanza de la lengua española. No solo por su corpus lingüístico especializado para la enseñanza de español, sino por la versatilidad que tiene para interactuar con el usuario, su enfoque didáctico que toma en cuenta los saberes previos o los vacíos teóricos de los nuevos aprendientes de español, su actualidad al momento de hacer uso de las TIC, no solo en las aulas de clase, sino fuera de ellas, y su capacidad para ser usado tanto por docentes de español como por aprendientes autodidactas de esta lengua.

Cabe resaltar el carácter didáctico que distingue cada una de las fases que componen Dicele y la importancia de este, tanto para la enseñanza, como para el aprendizaje, puesto que este proceso requiere de estrategias que lo faciliten. Además, se puede decir que este proyecto se preocupa por situar el aprendizaje de la lengua debido a su corpus constituido por textos con usos reales de esta, haciendo uso del español hablado en diferentes países (variantes dialectales).

También, con dicho proyecto se pretende avanzar en el tema de los instrumentos informáticos y de su ayuda a la lingüística para la enseñanza y el aprendizaje del español como lengua extranjera. Con esto no se pretende despojar al maestro de su campo de acción, por el contrario, se busca facilitar su labor de mediar la enseñanza de manera práctica y metodológica a partir de las nuevas tecnologías (TIC), donde el aprendiente tiene la oportunidad de demostrar sus conocimientos previos y reforzar sus falencias teóricas.

● Referencias

- Adam, J.-M. (1990). *Éléments de linguistique textuelle: Théorie et pratique de l'analyse textuelle*. Pierre Mardaga: Liège.
- Adam, J.-M. (2011). *La linguistique textuelle*. Armand Colin: Paris.
- Antoniadis, G. (2008). *Du TAL et son apport aux systèmes d'apprentissage des langues: Contributions*. Habilitation à diriger des recherches, Université Stendhal - Grenoble 3, Grenoble.



- Antoniadis, G. (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 de 2010, Nueva Orleans, USA.
- Aston, G. (1995). Corpora in Language Pedagogy: Matching Theory and Practice. En Cook, G. & Seidlhofer, B., éditeurs: *Principle and Practice in Applied Linguistics: Studies in honour of H. G. Widdowson*, capítulo 17, pp. 257-270. Oxford University Press: Oxford.
- Boulton, A. y Tyne, H. (2014). *Des documents authentiques aux corpus: démarches pour l'apprentissage des langues*. Langues & didactique. Didier: París.
- Charlier, B.; Deschryver, N. y Peraya, D. (2006). «Apprendre en présence et à distance: une définition des dispositifs hybrides». En *Distances et savoirs*, 2006/4 Vol. 4 pp. 469-496.
- Church, K. & Mercer, R. (1993). Introduction to the special issue on computational linguistics. Using large corpora. En *Computational Linguistics* 9(1), 1-24.
- Combettes, B. (1983). *Pour une grammaire textuelle: la progression thématique*. Collection : Formation continuée. De Boeck - Duculot, Bruxelles.
- Comenio, A. (1998). *Didáctica Magna*. Editorial Porrúa: México. Disponible en: <http://www.pensamiento-penal.com.ar/system/files/2014/12/doctrina38864.pdf>
- Fillmore, Ch. (1992). Corpus linguistics and computer-aided armchair linguistics. En J. Svartvik (Ed.), *Directions in Corpus Linguistics*. Berlín: Mouton de Gruyter, pp. 35-60.
- Gómez Guinovart, J. (1998). Fundamentos de Lingüística Computacional: bases teóricas, líneas de investigación y aplicaciones. En *Bibliodoc: anuari de biblioteconomia, documentació i informació* [en línea], 1998, pp. 135-146. <http://www.raco.cat/index.php/Bibliodoc/article/view/56629/66051>
- Instituto Cervantes (2015). El español: lengua viva. Informe 2015: Madrid. <http://www.cervantes.es/imagenes/File/prensa/El%20espaol%20una%20lengua%20viva.pdf>
- Johns, T. (1991). "From Printout to Handout: Grammar and Vocabulary Teaching in the Context of Data-Driven Learning". En: Johns, T. & King, P., (eds): *Classroom Concordancing. English Language Research Journal*, volume 4, pp. 27-45.
- Kennedy, G. (1998). *An introduction to corpus linguistics*. New York: Longman.
- Leech, G. (1997). Introducing Corpus Annotation. En: Garside, R; Leech, G. & McEnery, A. (eds) : *Corpus Annotation : Linguistic Information from Computer Text Corpora*. Capítulo 1, pp. 1-18. Longman: New York.
- Lundquist, L. (1980). *La cohérence textuelle : syntaxe, sémantique, pragmatique*. Nyt Nordisk Forlag: Copenhague.
- Lundquist, L. (1990). *L'analyse textuelle: méthode, exercices*. DJOF Publishing: Copenhague.
- Mangenot, F. & Louveau, E. (2006). *Internet et la classe de langue*. Colección: Techniques et pratiques de classe. CLE International: Paris.
- McEnery, T. & Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge University Press: New York.
- McEnery, T. & Wilson, A. (1996). *Corpus linguistics*. Edinburg: Edinburg University Press.
- Molina Mejía, J. M. (2009). *Les TICE et le développement de l'autonomie de l'apprenant de FLE : Le cas d'une université publique colombienne*. Tesis de maestría. Université Stendhal - Grenoble 3: Grenoble.
- Molina Mejía, J. M. (2014). Conception d'un environnement informatique fondé sur la linguistique textuelle et destiné à la formation des enseignants de FLE en Colombie. In Caron, P.-A. & Champagnat, R.



- (eds): *Actes des 5ièmes Rencontres des Jeunes Chercheurs en Environnements Informatiques pour l'Apprentissage Humain*, pp. 23-29, La Rochelle. Université de La Rochelle.
- Molina Mejía, J. M. (2015). *ELiTe-[FLE]² : Un environnement d'ALAO fondé sur la linguistique textuelle, pour la formation linguistique des futurs enseignants de FLE en Colombie*. Tesis de Doctorado. Université Grenoble Alpes: Grenoble.
- Molina Mejía, J. M. & Antoniadis, G. (2014). Toward the Constitution of a Hybrid Learning Environment for the FFL Teacher's Training in Colombian Universities Based on Text Linguistics. En G. Quiroz & P. Patiño (eds): *LSP in Colombia: Advances and Challenges*. Capítulo 14, Vol. 175, pp. 233-249. Colección Linguistic Insights. Éditions Peter Lang: Bern, Berlin, Bruxelles, Frankfurt am Main, New York, Oxford, Wien.
- Molina Mejía, J. M. & Antoniadis, G. (2015). "Un environnement informatique pour la formation des formateurs en FLE Colombiens, fondé sur la linguistique textuelle". En *Alsic [En ligne]*, Vol. 18, n° 2 | 2015, "Des machines et des langues" (nombreo coordinado por J.-R. Lapaire & F. Demaizière). URL: <http://alsic.revues.org/2843>; DOI : 10.4000/alsic.2843.
- Rézeau, J. (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 de Doctorado. Université Victor Segalen - Bordeaux 2: Bordeaux.
- Runge, A. (18 de octubre de 2013). Didáctica: una introducción panorámica y comparada. *ITINERARIO EDUCATIVO*. N° 62, p. 201-240. Disponible en: <http://revistas.usb.edu.co/index.php/Itinerario/article/view/1500>
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Sinclair, J. (1996). Preliminary Recommendations on Corpus Typology. En *Document Technique EAGLES (Expert Advisory on Language Engineering Standards)*.