

## RESUMEN

En las últimas décadas ha habido un creciente interés por el aprendizaje del español como lengua extranjera. El impacto en la sociedad española de acontecimientos socioeconómicos y socioculturales (internacionalización de la economía española, apertura de Institutos Cervantes, alumnos Erasmus, turismo lingüístico, y mayor interés por la cultura hispánica) ha propiciado nuevas líneas de investigación derivadas de la enseñanza del español como lengua extranjera. La investigación que presentamos trata de resolver las dificultades en el aprendizaje del español y servir como guía a los profesores para poder diagnosticar, clasificar y resolver los problemas característicos de los aprendices de español. Las particularidades fonéticas, morfológicas, léxicas, gramaticales y pragmáticas de su lengua materna interfieren y dificultan el proceso del aprendizaje del español.

Este estudio ha radicado en que nos encontramos por un lado ante un nuevo fenómeno aplicado a las segundas lenguas: el blog y su emergencia en los medios sociales. Por otro lado, la Lingüística Aplicada (en particular, la lingüística de corpus con su enfoque empirista) resulta adecuado para observar los errores de los estudiantes en un contexto informal y notablemente subjetivo. La investigación presenta diferentes teorías de la adquisición de segundas lenguas, se estudia el Análisis Contrastivo, el Análisis de Errores e Interlengua, se describe el proceso de aprendizaje, se analizan los factores individuales (edad, aptitud, actitud, inteligencia y personalidad), externos (el *input*, el *output* y la instrucción formal) e internos (relacionados con el conocimiento del mundo por parte del individuo) que pueden influir en el aprendizaje.

La investigación está orientada en sus diferentes fases a recopilar, analizar y proporcionar resultados a partir de las composiciones escritas en los blogs de los alumnos. Utilizamos el blog como género escrito por tratarse de un medio transparente y simultáneo de escritura y publicación cuyos rasgos discursivos se traducen en características lingüísticas que reflejan el carácter personal, informal y espontáneo de este género. Por estas razones, la blogosfera constituye un acertado sistema para observar los errores que cometen los aprendices de español.

En concreto, la investigación persigue el diseño de una metodología basada en los principios de la lingüística de corpus en sus diferentes etapas, desde la recopilación del corpus, su etiquetado (identificación y clasificación de los errores) y anotación hasta la extracción de los resultados.

El principal resultado de la tesis es la creación de una taxonomía y su etiquetado subdividido en cuatro grandes subapartados: errores léxicos, gramaticales, discursivos y gráficos. El total de errores recogidos en las 766 composiciones es de 17.827 repartidos entre los cuatro subapartados. Se presentan tanto los números de errores y su frecuencia como cifras absolutas y porcentajes de cada tipo de error. Se han extraído del corpus de blogs ejemplos de dichos errores. El etiquetado propuesto es extenso y reflejo de la diversidad de los numerosos posibles errores cometidos. El error es un paso hacia el aprendizaje.

**Palabras clave:** Adquisición de Segundas Lenguas, Español como Lengua Extranjera, Análisis de errores, Taxonomía y Etiquetado de errores, Lingüística de Corpus, Corpus de Blogs de Aprendices de Español.