

El análisis de errores como evidencia de la interlengua de los estudiantes de inglés como lengua extranjera*

Error analysis as evidence of interlanguage of students of english as a foreign language

María Victoria Pardo**

RESUMEN

Este artículo presenta el análisis de los tres primeros errores más comunes en la producción escrita del inglés como lengua extranjera (ILE) en un corpus de aprendices recolectado en la Universidad del Norte de Barranquilla, Colombia. El corpus fue recopilado siguiendo los parámetros de la lingüística de corpus computacional y la extracción de errores se hizo con software especializado en análisis lingüístico WordSmith Tools 5 Scott (2008) y Lancsbox (Brezina et al., 2015). El análisis se basó en la teoría de Corder (1981) y James (1998) sobre el análisis de errores (AE) y usó el etiquetador de la Universidad de Louvaina (Dagneaux et al., 2005a) para obtener resultados comparables con trabajos similares. De acuerdo con los resultados, los tres errores más relevantes en la escritura de estudiantes universitarios pertenecen a las categorías de Forma (FS), Gramática (G), y Léxico (LS).

Palabras clave:
Análisis de errores, corpus de aprendices computacional (CAC), corpus, ILE, etiquetador de errores.

SUMMARY

This article presents the analysis of the first three most common errors in the written production of English as a foreign language (ILE) in a corpus of learners collected at the Universidad del Norte in Barranquilla, Colombia. The corpus was compiled following the parameters of computational corpus linguistics and the re-

Keywords:
Error analysis, computational learner corpus (CAC), corpus, ILE, error tagger.

* Este proyecto tuvo financiación por parte de COLCIENCIAS de Colombia (Convocatoria 647 de 2014). El análisis de errores en un corpus de documentos escritos por hablantes de español aprendices de inglés como lengua extranjera: Un estudio de corpus en estudiantes universitarios. La línea de investigación corresponde al grupo TNT (Traducción y Nuevas Tecnologías) de la Universidad de Antioquia.

** Colombiana. Doctora en Lingüística de la Universidad de Antioquia, Colombia. Profesora de tiempo completo de la Universidad del Norte en Barranquilla, Barranquilla, Colombia. mvpardo@uninorte.edu.co

moval of errors was done with specialized software for linguistic analysis, WordSmith Tools 5 Scott (2008) and Lancsbox (Brezi-
na et al., 2015). The analysis was based on the theory of Corder
(1981) and James (1998) on error analysis (AE) and used the
University of Louvaina tagger (Dagneaux et al., 2005a) to obtain
comparable results with similar works. According to the results,
the three most relevant writing errors committed by university
students belong to the categories of Form (FS), Grammar (G),
and Lexicon (LS).

Introducción

En el aprendizaje de un segundo idioma el estudiante recorre diferentes etapas hasta que logra culminar en determinado nivel de adquisición. Durante el periodo en que el estudiante no es hablante independiente de la lengua extranjera se desarrolla la “interlengua” definida por Selinker (1972) como un idiolecto desarrollado por el estudiante de una segunda lengua que preserva algunas características de la primera lengua y generaliza algunas reglas de la escritura y del habla. De acuerdo con Selinker (1972), la interlengua es un sistema lingüístico único que permite identificar en qué etapa de adquisición de la lengua se encuentra el estudiante. Una forma de conocer dicho nivel en los discentes es a través del análisis de errores (AE), una metodología que permite establecer las características de la producción oral o escrita de los estudiantes (Corder, 1981). Por medio del AE se puede establecer si la etapa en la cual se encuentran los educandos es acorde con el nivel que están cursando, para así generar intervenciones de tipo pedagógico y didáctico.

Previamente al AE, en las décadas de 1950 y 1960, se aplicaba el análisis contrastivo (AC) (Lado, 1957) para el estudio de idiomas extranjeros. Este análisis describe características comparables de la lengua materna con el idioma de destino y compara las formas y los significados resultantes de los dos idiomas, para mostrar las similitudes y diferencias y predecir las posibles interferencias. A principios de la década de 1970, debido a la falta de confiabilidad, el AC se consideraba un modelo obsoleto de descripción del lenguaje. Sin embargo, es imposible negar totalmente los efectos de la lengua materna en el idioma de destino. El AE se convirtió así, en una nueva forma de describir la interlengua del estudiante (Alexopoulou, 2005; 2010; Elejalde y Ferreira, 2018; Ferreira et al., 2014). La gran diferencia entre el AC y el AE consiste en que en este último la lengua materna no desempeña el papel principal, sino que parte de la premisa de que los errores pueden describirse en función del idioma de destino, sin necesidad de buscar la explicación en la lengua materna de los estudiantes (James, 1998).

Analizar los errores de los estudiantes que cursan ILE es una forma de aislar y comprender este fenómeno para mejorar las prácticas de enseñanza y aprendizaje, con el fin de que sean más relevantes y acordes a las necesidades de los alumnos. El presente trabajo contribuye

con una visión de la descripción de la interlengua de estudiantes de ILE a nivel universitario. Es un punto de vista que debe considerarse en la planificación y el diseño de los materiales y estrategias de enseñanza de los alumnos.

La investigación hace un análisis descriptivo y clasificatorio de los tres errores más relevantes de la producción escrita de estudiantes de ILE a nivel universitario. Se presenta una vista general de los errores en contexto y se realiza el análisis de cada error. Así mismo, compara los resultados con otras investigaciones que usen un etiquetador con categorías similares.

Marco teórico

A fines de la década de 1960, el AE se convirtió en una alternativa adecuada al AC. Esta es una tendencia que aún se mantiene, ya que con el uso de las nuevas tecnologías se han logrado obtener AE en grandes cantidades de datos imposibles de lograr por medios manuales (Díez-Bedmar, 2011; Elejalde y Ferreira, 2018; Pardo, 2020). Corder (1967), hizo los principales aportes en torno al AE:

- a. Los errores muestran las características del sistema de lengua que está usando el alumno. Evidencia lo que el alumno ha asimilado —que en muchos casos es diferente de lo que los maestros le han aportado—, porque un aporte no es igual a un concepto asimilado.
- b. Existe una “competencia transicional” (Corder, 1981) que se muestra en los errores cometidos por los aprendices de un idioma en los que se desarrolla un sistema de lenguaje independiente. Esto es lo que Selinker (1972) llama interlengua.
- c. Los errores son diferentes de las equivocaciones.
- d. Los errores son significativos en tres aspectos: evidencian lo que necesita ser enseñado; indican al investigador cómo se lleva a cabo el aprendizaje; y además son un medio por el cual los alumnos verifican y desmienten sus hipótesis respecto de una segunda lengua.

De acuerdo con James (1998) existe una posible ecuación entre **error** y **equivocación** y la distinción hecha por Krashen (1982) entre **adquisición** y **aprendizaje**, la cual se presenta en la Tabla 1.

Tabla 1
Equivocación/error vs adquisición/aprendizaje

Adquisición	Equivocación	Aprendizaje	Error
La adquisición de un idioma (por ejemplo, el caso de un hablante nativo) es un proceso inconsciente. Las reglas se adquieren inconscientemente, por lo tanto, quien aprende no está en un estado de ignorancia y no cometerá errores, sino equivocaciones.		Aprender un idioma es el resultado de la instrucción formal. Requiere conciencia acerca de sus reglas (Krashen, 1982) . En ese caso, el alumno comete errores, pero podría corregirlos o evitarlos utilizando conocimiento explícito.	
Las equivocaciones son el resultado de falta de atención. No son sistemáticas y posiblemente se autocorrijen porque el hablante conoce la regla.		Los errores son sistemáticos. Son el resultado de la ignorancia de la regla. Si el alumno no ha aprendido particularidades de un idioma, el resultado será un error.	

Fuente: Adaptado de James (1998, p. 85).

Ahora bien, debido a que los errores ocurren de manera repetitiva, es necesario que su análisis sea sistemático, con el fin de identificar su verdadera incidencia en la producción de los discentes. En cuanto a la recolección de los datos para el AE, es necesario hacerla siguiendo parámetros que permitan establecer la variedad de la interlengua de los estudiantes.

Corpus generales y corpus de aprendices

Los corpus son colecciones de textos auténticos transcritos en formato digital (pueden incluir textos orales) corresponden una muestra representativa de una lengua o una variedad de lengua (McEneary et al., 2006). Dentro de la variedad de corpus existen los corpus de aprendices que surgieron a fines de la década de 1980 y que son colecciones electrónicas de datos naturales o casi naturales producidos por estudiantes de lengua extranjera o de un segundo idioma (L2), recopilados de acuerdo con criterios de diseño explícitos (Granger et al., 2015). Estos corpus permiten la realización de un análisis científico válido de la producción del alumno y tienen las mismas características atribuidas a otros corpus con una diferencia en los datos porque, en este caso, la fuente es la producción oral o escrita de los estudiantes. El aumento de los corpus de aprendices a fines de la década de 1980 se debió, en parte, a su potencial para la investigación de la producción auténtica

de los estudiantes. Los corpus de aprendices son colocados en formato electrónico previamente a ser analizados, con el fin de utilizar un software especializado para tal fin.

Anotación del corpus

La anotación del corpus es el proceso de agregar información interpretativa y lingüística a un corpus electrónico de datos de lenguaje hablado o escrito (Leech, 2005). La información añadida viene en forma de etiquetas, que se pueden definir como entidades individuales agregadas a una parte o partes del discurso. Las etiquetas son únicas y pueden identificar características del corpus analizado. El etiquetado es un procedimiento fundamental para lograr un análisis más eficiente con el fin de obtener estadísticas y tendencias predominantes de los datos. La anotación de un corpus se puede considerar como un valor añadido, ya que se convierte en una fuente de información lingüística para realizar futura investigación (Garside, Roger; Leech, Geoffrey; McEnery, 1997) (Leech, 1997).

Existen dos tipos diferentes de anotación: enmienda y categorización (Rosen et al. 2014). En el primer caso se buscan las intenciones del autor, se validan una o más hipótesis y se corrigen los errores en los textos, mientras que en la categorización de errores se hace anotación de errores con etiquetas de categorías predefinidas. En la Tabla 2 se pueden observar las ocho categorías de errores del etiquetador usados en la presente investigación, (Dagneaux et al., 2005b).

Tabla 2
Categorías de errores con sus etiquetas

Categoría	Código. (letra que representa la categoría de error)
1. Errores de forma.	F
2. Errores gramaticales (errores que rompen las reglas generales de la gramática inglesa).	G
3. Errores léxico-gramaticales (errores donde las propiedades morfosintácticas de una palabra han sido violadas).	X
4. Errores de léxico (errores que involucran las propiedades semánticas de algunas palabras y frases).	L
5. Palabra redundante o faltante y otros errores que involucran el orden de las palabras.	W
6. Errores de puntuación.	Q
7. Errores de estilo.	S
8. Infelicities; no son propiamente errores sino expresiones poco o no aceptadas pragmáticamente en un contexto.	Z

Fuente: Adaptada y traducida de Dagneaux et al. (2005b).

Software de extracción de datos

Un software de extracción permite la búsqueda de patrones en el lenguaje y facilita las estadísticas en una gran cantidad de documentos. En la presente investigación, el software de extracción fue WordSmith (Scott, 2008), ya que permite obtener estadísticas completas de los datos. Otro software utilizado en la extracción de datos fue LancsBox de la Universidad de Lancaster (Brezina et al., 2015). Este es un software de nueva generación para el análisis de datos del lenguaje. Funciona con corpus existentes dando estadísticas y la posibilidad de obtener gráficos.

Previo trabajo relacionado en AE

Aunque para fines de la década de 1960, el AE se consideraba una alternativa aceptable al AC, hasta ahora, la mayor parte del trabajo realizado a través de esta metodología ha sido planificado y producido en Europa y Estados Unidos. En América Latina solo en años recientes algunos investigadores han trabajado en el AE Elejalde y Ferreira, (2017)

utilizando la metodología de corpus de aprendices computarizado para compilar una cantidad representativa de datos. En el presente artículo, se analizan algunos trabajos relacionados con el campo del AE, pero sería imposible dar cuenta completa del trabajo realizado en todo el mundo. Se dio preferencia al trabajo de América Latina o España, ya que los alumnos de este estudio comparten la lengua materna de los alumnos de esos países.

El *International Corpus of Learner English (ICLE)* Virtanen, (1996) de la Universidad de Lovaina en Bélgica es el ejemplo más significativo de corpus de aprendices. Este incluye producción de estudiantes cuyas lenguas maternas son muy variadas: francés, holandés, alemán, sueco, chino, español, finlandés, japonés y checo, entre otros. No solo incluyen textos escritos, sino orales previamente transcritos en formato digital.

En España, MacDonald (2016) analizó errores en textos argumentativos escritos por 304 estudiantes universitarios españoles de inglés. Los textos fueron compilados a partir de contextos técnicos y de humanidades. Los resultados muestran que los errores gramaticales son los más frecuentes.

En Chile, Ferreira et al. (2014) presentan un estudio relativo al AE asistido por computador basado en un *Corpus de aprendientes de español como lengua extranjera ELE* de la Universidad de Concepción. De acuerdo con los resultados, los errores de mayor frecuencia son los de ortografía acentual, preposiciones, concordancia gramatical, el verbo y los artículos. También en Chile, Elejalde y Ferreira, (2017) analizaron los errores de transferencia en textos escritos en español como lengua extranjera. La lengua materna de los participantes era el inglés y el alemán, en tanto que el etiquetado se realizó de manera manual y el análisis con software Corpus Tool¹ y Sketch Engine (Kilgarriff y Rychly, 2003). De acuerdo con los resultados, las autoras concluyen que los errores de transferencia más frecuentes en ambas lenguas son la preposición, el artículo, el verbo y el pronombre.

1 Software desarrollado en el contexto del Ministerio de Educación y Ciencia de España bajo el número de licencia HUM2005-01728/FILO (The WOSLAC Project).

En Puerto Rico, en tanto, Morales-Reyes y Gómez (2016) analizaron los errores de estudiantes de inglés como lengua extranjera. Evaluaron la adquisición del artículo en inglés en 30 niños de habla hispana que aprendían la lengua como segundo idioma (SL). Los resultados muestran que los niños transfieren el conocimiento de su primera lengua de manera similar a como lo hacen los adultos.

En España, Diez-Bedmar (2011) hizo el análisis de los principales errores que cometen los estudiantes españoles al escribir en inglés para un examen de ingreso universitario. El hecho más interesante en este caso es el uso del mismo etiquetador de errores de la Universidad de Lovaina en la versión 1.1 (Dagneaux et al., 1996). La autora compara los hallazgos de este estudio con trabajos previos sobre AE y encuentra algunas tendencias comunes.

En Argentina, (Sánchez et al., 2014) realizaron un estudio en un grupo de 24 hablantes nativos de español acerca de la concordancia entre género y número del sustantivo principal y el adjetivo predicativo en español. Analizaron los errores de concordancia y los patrones de omisión. Según sus hallazgos, el promedio de error no fue diferente en las dos condiciones sintácticas.

En Colombia, Vásquez (2008) analizó una composición escrita utilizando cuatro categorías de error: omisión, adición, información errónea y orden erróneo. Este análisis encontró errores de interferencia o transferencia, así como errores de desarrollo, entre otros. El autor concluye que la lengua materna afecta el proceso de aprendizaje, la sintaxis y el significado de la lengua meta.

La literatura analizada muestra, entonces, un aumento en el trabajo respecto de AE, donde los investigadores intentan diseñar su propia metodología para analizar los tipos de errores que encuentran. No hay acuerdo en cuanto al uso de un etiquetador de errores colectivo para fines de investigación, y esa tendencia es problemática, porque es imposible hacer un trabajo comparativo cuando las categorías y taxonomías de errores no coinciden entre sí. En América Latina, y especialmente en Colombia, existe una gran necesidad de investigación respecto de AE para obtener datos que conduzcan a una mejora en las prácticas de enseñanza aprendizaje de las lenguas extranjeras y que conlleve al diseño de materiales más apropiados a las necesidades de los estudiantes.

Metodología

El objetivo de esta investigación es presentar un análisis descriptivo y clasificatorio de los tres errores más relevantes de la producción escrita de estudiantes de ILE a nivel universitario. Así mismo se busca, en lo posible, comparar los resultados con otras investigaciones que usen el mismo tipo de etiquetador. El estudio se circunscribe a la lingüística aplicada, utilizando la metodología de la lingüística de corpus en la recopilación, almacenamiento, recuperación y análisis de datos y la metodología de AE para localizar, describir, obtener estadísticas y clasificar los diferentes tipos de errores encontrados. Se trata, por último, de una investigación descriptiva con datos recopilados durante el segundo semestre de 2015.

El presente corpus fue marcado con etiquetas de error para obtener las estadísticas de los errores por categorías y tipo. Para el presente estudio se hizo la anotación de errores utilizando el *Manual de etiquetado de errores versión 1.2* (Dagneaux et al., 2005a). Este etiquetador tiene ocho categorías y un total de 56 tipos de errores que pueden ser observados en el Anexo 1.

Descripción de la población

Los sujetos participantes son estudiantes de la Universidad del Norte que pertenecen a diferentes programas de pregrado (música, arquitectura, ingenierías, psicología y leyes, entre otras) registrados en un curso de ILE de acuerdo con su nivel de desempeño dado por un examen clasificatorio. La lengua materna de todos los sujetos es el español y la edad promedio es de 23 años. Los cursos de inglés se ofrecen para cubrir la demanda de los estudiantes de pregrado que deben alcanzar una competencia mínima de B2 en un idioma extranjero, de acuerdo con el *Marco Común Europeo de Referencia* (Council of Europe, 2001). En esta universidad el inglés se ha configurado como el idioma con más estudiantes registrados, debido a que es la lengua enseñada en los colegios y por ser la lengua de preferencia en los negocios.

Procedimiento seguido en la elección de la muestra representativa

En total 2.138 sujetos estaban registrados en los niveles A1 a B2 de los cuales, 636 realizaron una encuesta inicial que estableció algunos as-

pectos socioculturales que podrían tener incidencia en la adquisición del ILE. De ese grupo inicial, 515 sujetos autorizaron el uso de sus datos para fines de investigación. Para obtener una representación de toda la población, se eligió un subconjunto representativo utilizando la siguiente fórmula de Survey Monkey (*Survey Monkey*, 2017):

Tamaño de la muestra:

$$\frac{z^2 \times p(1-p)}{e^2} \\ 1 + \left(\frac{z^2 \times p(1-p)}{e^2 N} \right)$$

Tamaño de la población = N

Margen de error = e (porcentaje puesto en decimal). En este caso, es del 4%.

z-score = z (desviación estándar de la media). Con un margen de error del 5%, el tamaño mínimo de la muestra se establecía en 508, pero considerando el hecho de que 515 sujetos autorizaron el uso de sus datos para fines de investigación, la investigadora decidió analizar todo este subconjunto. Por lo tanto, en este caso particular, el margen de error es aproximadamente del 4%.

Recolección de los datos

Esta sección presenta los procedimientos metodológicos seguidos en la recopilación de datos y la anotación de errores. Los sujetos participantes fueron ubicados en diferentes niveles de acuerdo con los resultados de una prueba de nivel.

1. **Recolección del corpus.** El proceso de escritura se desarrolló durante un curso de 64 horas donde los sujetos recibieron instrucciones para desarrollar un párrafo basado en una pregunta acerca de un tema general para el nivel B1 y un ensayo de comparación y contraste para el nivel B2. En diferentes sesiones recibieron comentarios por parte de los profesores e hicieron correcciones. El corpus está formado por el trabajo final desarrollado en una clase de dos horas durante la cual debían escribir el trabajo definitivo para revisión.
2. **Codificación.** A cada archivo se le asignó un código para mantener la trazabilidad. Esta nomenclatura permitió la identificación de los textos producidos sin revelar sus identidades.

3. Los textos escritos a mano fueron transcritos a Word MS y luego se convirtieron en textos TXT para hacer la anotación de errores.
4. La anotación de errores se hizo utilizando el *Manual de Etiquetado de Errores versión 1.2* (Dagneaux et al., 2005a). Se utilizó este etiquetador de la Universidad de Lovaina ya que cuenta con categorías y subcategorías que permiten obtener la granularidad fina en los datos, además porque al utilizar este tipo de etiquetadores estándar se espera en el futuro realizar análisis comparativos a gran escala. Después del etiquetado todos los archivos se colocaron en WordSmith Tools (Scott, 2008) para alinear los textos por tipo de error y obtener las estadísticas en cada caso. También se utilizó LancsBox (Brezina et al., 2015) para obtener estadísticas de errores individuales y posibles gráficas. Los registros se guardaron en hojas de Excel para comenzar el análisis. El etiquetado se hizo siguiendo los lineamientos del manual, colocando las etiquetas entre paréntesis antes del error y la corrección entre dos signos de dólar.

Ejemplo: When I look (GA) a \$an\$ advertisement with celebrities.

Se observa un error que pertenece a la categoría gramatical y el tipo de error afecta el uso del determinante, en este caso, el artículo definido/indefinido.

5. Clasificación de errores por categorías. La versión del manual de etiquetado tiene ocho categorías de errores, las que se dividen en subcategorías para un total de 56 tipos de error. Cada categoría se representa con una letra —Tabla 1—. La primera letra de la etiqueta muestra la categoría de error. La segunda letra generalmente indica una subcategoría. Los errores fueron clasificados de acuerdo con las categorías y tipos para obtener las estadísticas de estos.

Problemas encontrados durante el proceso

A pesar de que todo el personal de apoyo fue instruido de antemano acerca del proyecto de investigación, en muchos casos los estudiantes no fueron informados.

En cuanto al sistema de anotación, el etiquetador de errores cuenta con una categorización y tipificación de los errores de acuerdo con las categorías que permite realizar clasificaciones con gran acierto, pero se deben tener instrucciones más claras acerca de su uso, un

documento o una página con preguntas frecuentes en caso de que el usuario lo necesite, ya que esperar una respuesta del proveedor toma demasiado tiempo, especialmente por la diferencia en el horario estándar. La investigadora, en muchos casos, terminó aprendiendo por ensayo y error, y eso resultó en una pérdida de tiempo. Los diseñadores del software asumen que todos los investigadores saben cómo usarlo, pero ese no es siempre el caso. Por ejemplo, cuando un archivo está incluido en una carpeta contenida en otra carpeta, el etiquetador no lo abrirá. Se debe crear un acceso más directo. Por otro lado, los errores ortográficos debieran etiquetarse automáticamente. Como razonamiento final, se insiste en la necesidad de diseñar un etiquetador de errores que esté disponible para fines de investigación y no con carácter lucrativo.

Presentación de resultados

El AE se llevó a cabo siguiendo la categorización de errores de acuerdo con el *Manual de etiquetado de errores versión 1.2*. (Dagneaux et al. 2005b) y de acuerdo con las teorías de Corder (1981) y Jameshe (1998). En cada caso, cada error recibió una interpretación respecto de lo que el alumno intentó decir y la oración fue reconstruida. Los errores se recuperaron utilizando las herramientas del software WordSmith (Scott, 2008) para obtener estadísticas de los errores. La Tabla 3 muestra el total de errores en el corpus.

Tabla 3
Total de errores etiquetados en el corpus y vista de alineación por tipo de error

Line	Text	Error	Count	%	WordSmith	Count	%
18	the manner (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
19	WordSmith (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
20	Year. On the other hand, (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
21	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
22	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
23	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
24	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
25	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
26	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
27	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
28	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
29	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
30	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
31	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
32	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
33	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
34	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
35	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
36	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
37	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
38	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
39	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
40	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
41	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
42	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
43	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
44	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
45	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
46	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
47	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
48	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
49	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
50	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
51	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
52	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
53	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
54	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
55	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
56	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
57	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
58	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
59	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
60	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
61	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
62	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
63	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
64	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
65	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
66	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
67	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
68	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
69	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
70	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
71	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
72	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
73	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
74	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
75	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
76	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
77	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
78	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
79	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
80	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
81	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
82	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
83	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
84	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
85	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
86	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
87	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
88	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
89	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
90	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
91	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
92	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
93	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
94	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
95	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
96	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
97	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
98	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
99	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%
100	(200) (S.S.S.) (200) (S.S.S.) (200) (S.S.S.)	WordSmith	102	2.00%	WordSmith	102	2.00%

Fuente: Elaboración propia.

El corpus lo conforman 149.325 tokens, 12.164 tipos y 12.337 lemas. El total de errores en el corpus fue de 14.531, es decir, 9,57 errores por cada 100 palabras, o una media de 28.05. El registro de los errores por categorías se evidencia en la Tabla 4.

Tabla 4
Total de errores por categoría

Categoría	Código categoría	Errores por categoría
1. Errores de forma.	F	1.931
2. Errores gramaticales.	G	6.192
3. Errores léxico-gramaticales.	X	257
4. Errores de léxico.	L	2.662
5. Palabra redundante, palabra faltante y orden de las palabras.	W	1.988
6. Errores de puntuación.	Q	946
7. Errores de estilo.	S	519
8. Infelicidades (<i>Infelicities</i>).	Z	36

Fuente: Elaboración propia.

Este mismo total de errores se visualiza de mayor a menor incidencia en la Figura 1.



Figura 1. Vista gráfica de la incidencia de errores por categoría

Fuente: Elaboración propia.

Las tres categorías con más incidencia de errores son Gramática (G) con 42% de errores en el corpus, seguida de la categoría Lexis (L) con 18% y Palabras (W), que se refiere a errores en el uso de las palabras con 13% del total del corpus. La Tabla 5 presenta el número de errores por composición en cada categoría y la media en cada caso.

Tabla 5
Incidencia de los errores por categorías

Errores, totales y media. Errores por categoría	Totales en corpus	Media en corpus
Gramática.	6,192	$M = 11,95$
Léxico.	2,662	$M = 5,13$
Palabras.	1,988	$M = 3,83$
Forma (deletreo, morfología).	1,931	$M = 3,72$
Puntuación.	946	$M = 1,82$
Oraciones.	519	$M = 1,00$
Léxico-gramaticales.	257	$M = 0,49$
Infelicidades: uso de la lengua inusual al contexto o extraño.	36	$M = 0,069$

Fuente: Elaboración propia.

Un análisis más completo revela el número de errores de los estudiantes por tipo de error (no por categoría). Los primeros tres tipos de error con más incidencia en el corpus son Forma deletreo (FS) con 1.917 errores; Gramática, Artículos (GA) con 1.713 errores; y Léxico, Individual (LS) con 1.255. La Figura 2 muestra el conteo de errores por tipo.



Figura 2. Número de errores por tipo

Fuente: Elaboración propia.

Se observa la representación de los tres primeros errores FS, GA y LS en relación con los otros tipos de errores en el corpus. Estos tres errores tienen incidencia en los niveles B1 y B2, lo que indica que la interlengua de los estudiantes mantiene algunas características a pesar del avance en el nivel de adquisición.

Análisis de resultados

El siguiente análisis se basa en los tres tipos de error de mayor incidencia en el corpus. Debe notarse que cada tipo de error pertenece a una categoría diferente, como se muestra en la Tabla 6.

Tabla 6
Tres errores con la mayor incidencia en el corpus

Categoría	Tipo de error	Total	% en categoría	% en corpus
Forma	FS	1,917	99	13,19
Gramática	GA	1,713	27	11,78
Léxico	LS	1,255	47	8,63

Fuente: Elaboración propia.

Los tres errores con más incidencia pertenecen a tres categorías diferentes. La categoría Forma (F) comprende los errores FM (Morfología), FS (Deletreo) y FSR (deletreo regional). Esta categoría ocupa el cuarto lugar, pero el error Forma Deletreo FS ocupa el primer lugar de incidencia con el 13,19% en el corpus. La categoría Gramática (G) comprende 25 tipos de errores. Esta categoría ocupa el primer lugar de incidencia, pero el error GA (Artículo Gramatical) ocupa el segundo lugar de incidencia con el 11,78% en el corpus. Este error afecta el uso de los artículos definidos, indefinidos o cero artículos. La categoría Léxico (L) ocupa el segundo lugar en el corpus, pero el error Léxico Individual LS ocupa el tercer lugar en la incidencia de errores del corpus. Este error afecta el uso connotativo, conceptual y de colocación del léxico.

Análisis individual de cada error

Primer error: Forma Deletreo (FS)

Este tipo de error incluye los siguientes casos: errores de deletreo, omisión o uso equivocado de mayúsculas, palabras no existentes, préstamo de palabras, palabras homófonas y omisión o uso inadecuado de guiones.

Para el caso del presente corpus, FS es un error reincidente, siendo los errores de deletreo el 90% del total etiquetado en esta categoría. Veamos algunos ejemplos del corpus².

1)	12813	... An example is when you (FS) polute \$pollute\$ with gar
2)	12814)0 \$if\$ it appears (GA) 0 \$a\$ (FS) celebritie \$celebrity\$ (G
3)	12816 to make (LCC) and \$an\$ (FS) independ \$independent\$ fa
4)	12818	t \$deal with\$ the problems or (FS) difficults \$difficulties\$
5)	12819	be happy (LS) to \$or\$ have a (FS) confortable \$comfortable\$
6)	12820	home and live on their own. (FS)first \$First\$ reason, (LP)
7)	12844	Because you know all the (FS) responsability \$responsibility\$

Todos los ejemplos anteriores tienen errores de deletreo, excepto el ejemplo número 6, donde no se usa mayúscula luego de punto seguido. En los ejemplos 5 y 7 las palabras son deletreadas como palabras similares del español.

De acuerdo con los hallazgos, el error FS tiene gran incidencia durante todas las etapas de aprendizaje de la lengua inglesa desde el nivel B1 hasta B2. Una posible hipótesis en la ocurrencia sistemática de este error es porque el sistema fonético de la lengua materna y la segunda lengua no son contrastables, aunque pueden sonar de manera similar. Existe una leve disminución de este tipo de error al iniciar el nivel B2, pero sigue afectando la producción escrita por no utilizar de manera correcta los signos de puntuación. Los sujetos en niveles intermedios y avanzados presentan gran incidencia de oraciones divididas y uso incorrecto de mayúsculas. Estos errores impiden fluidez en sus escritos y en algunos casos los hace incomprensibles.

Segundo error: Artículo Gramatical (GA)

El error GA pertenece a la categoría gramatical. Se refiere a errores que rompen con las reglas gramaticales de la lengua inglesa, específicamente, al uso inadecuado del artículo definido e indefinido. Esta es una indicación de que los sujetos no han dominado el uso de los artículos. Se podría tratar de “transferencias negativas” que causan divergencias entre la primera y la segunda lengua y que como lo mencio-

2 Para un acceso directo al corpus en línea se puede visitar <http://grupotnt.udea.edu.co/django/clec/corpus/>

nan Elejalde y Ferreira, (2017) ocasionan problemas a nivel oracional y gramatical. En más del 80% de los casos en los que se etiquetó este error, los sujetos usaron el artículo antes de un sustantivo. Esta es una característica del español donde el uso de un artículo es obligatorio antes de un sustantivo. Veamos un ejemplo.

1) 11068 the government said that (GA) the \$0\$ life is a right, a

De acuerdo con las reglas de gramática inglesa, en este caso no es necesario el uso de un artículo definido, porque **la vida** es un nombre incontable y aquí se refiere a **la vida** en general, que no es la vida de alguien en particular. Por lo tanto, el uso del artículo no es necesario. Contrariamente, en la misma situación en español **vida** como sustantivo debe estar precedido de un artículo. Este es un posible error de transferencia, porque si decimos la misma oración en español, necesitamos usar un artículo. En español se diría “el gobierno dijo que **la vida** es un derecho...”

En el siguiente ejemplo, el alumno olvidó usar el artículo indefinido. También podría ser un error de transferencia porque en español, a diferencia del inglés, no es necesario usar un artículo cuando se habla de profesiones. Veamos un ejemplo.

2) 11215 ... if you want to be (GA) 0 \$a\$ journalist (LP) 0 \$

Si quiere decir la misma oración en español sería: “...si quiere **ser periodista**...” en español no hay necesidad de un artículo.

Los errores en el uso del artículo definido/indefinido afectan en gran manera la producción escrita, ya que se mantienen en el mismo segundo lugar en los niveles B1 y B2. A pesar de más horas de instrucción, este tipo de error no disminuye en la estadística. De acuerdo con los casos analizados se observa que hay una posible interferencia de la lengua materna. El uso o no del artículo definido en inglés se convierte así en un tema de atención durante todos los niveles de aprendizaje. Al parecer no es suficiente con hacer claridad respecto de las reglas de uso del artículo, sino que es necesario implementar material que garantice el aprendizaje y uso de esas normas.

Tercer error: Léxico Individual

Pertenciente a la categoría de Léxico (L) el tercer error con el puntaje más alto en el corpus fue Léxico, Individual (LS) con 1.255 ocurrencias.

LS da cuenta de errores léxicos conceptuales, de colocación o connotativos en palabras individuales. El manual incluye en esta subcategoría compuestos sólidos y con guiones. Veamos algunos ejemplos del corpus³.

10)	5057	... Even worst, if it is an (LS) acceptance \$admission\$...
11)	5058	... also the risky ways to (LS) affront \$face\$ it. Before
12)	5059	... cheat \$cheating\$ (LS) value \$calculate\$...
13)	5060	... For the boys the (LS) candies \$chewing\$ gum

Los ejemplos 10, 11 y 12 se refieren a errores de léxico en el uso conceptual de las palabras. El ejemplo 13 se refiere a un error de colocación léxica, en este caso, usando una palabra con sentido similar. Según el inglés estándar, se acepta *chewing gum* (goma de mascar), pero no *candies gum* (dulces de mascar).

Los errores de Léxico Individual (LS) se mantienen en tercer lugar durante toda la etapa de nivel B1, pero en el nivel B2 este error pasa al primer puesto de incidencia en el corpus, desplazando a FS (Forma Deletreo) que venía ocupando el primer lugar. El incremento de este error en el nivel B2 está acompañado por la aparición de errores de concordancia del nombre en singular y plural. El incremento del error LS en el nivel B2 se debe posiblemente a la búsqueda de nuevo vocabulario en la que los sujetos adaptan términos tal y como lo refieren Elejalde y Ferreira, (2018) sobre los errores interliguales en los que el aprendiente recurre a la lengua materna para suplir las falencias. En la Tabla 7 se muestra el comparativo de errores en los niveles B1 y B2.

Tabla 7

Tabla comparativa de errores en los niveles B1 y B2

Level B1.1		Level B1.2		Level B1.3-B2.1		Level B2.2-B2.3	
E. Type	%	E. Type	%	E. Type	%	E. Type	%
FS	18,39	FS	16,17	FS	20,31	LS	11
GA	14,74	GA	11,19	GA	15,66	GA	0,08
LS	7,77	QM	6,49	GNN	7,57	WRS	0,07
GNN	6,58	LS	5,97	LS	6,20	GWC	0,07
LP	6,11	LP	5,49	SU	6,02	GNN	0,06

Fuente: Elaboración propia.

3 Para un acceso directo al corpus en línea se puede visitar <https://grupoint.udea.edu.co/clec/corpus>

Comparación de los resultados obtenidos con resultados de estudios previos

Como se mencionó anteriormente, uno de los objetivos para hacer anotaciones con el *Manual de etiquetado* de la Universidad de Lovaina es ofrecer resultados comparables con investigaciones en todo el mundo. Aunque existen trabajos en el AE, la mayoría de los investigadores diseña sus propias herramientas para etiquetar errores y eso hace que sea muy difícil o imposible hacer comparaciones.

Como resultado de la búsqueda de trabajos similares en AE se hizo el hallazgo del trabajo de Díez-Bedmar, (2011) quien hizo un análisis similar utilizando un sistema de anotación de la Universidad de Lovaina en versión 1.1, que es anterior a la del presente trabajo (versión 1.2). Algunas categorías de la versión previa siguen siendo las mismas en la segunda versión y otras han cambiado.

La Tabla 8 presenta algunas de las diferencias respecto de la versión 1.1 en comparación con la versión 1.2, utilizada en la presente investigación. Para obtener más detalles sobre la versión 1.1 consulte (Díez-Bedmar 2011).

Tabla 8
Principales diferencias en etiquetadores de 1996 Versión 1.1 vs 2005 Versión 1.2

Etiquetador de Universidad de Lovaina versión 1.1 (1996)	Etiquetador Lovaina versión 1.2 (2005)
Número de categorías: 7	Número de categorías: 8
Número de etiquetas de errores: 43	Número de etiquetas de errores: 56
Categorías:	Categorías:
1. Forma (F)	1. Forma (F)
2. Gramática (G)	2 Gramática (G)
3. Léxico-gramatical (X)	3. Léxico-gramatical (X)
4. Léxico (L)	4. Léxico (L)
5. Palabras redundantes, faltantes y orden de palabras (W)	5. Palabras redundantes, faltantes y orden (W)
6. Registro (R)	6. Puntuación (Q)
7. Estilo (S)	7. Estilo (S)
	8. Infelicidades (Z)

Fuente: Elaboración propia.

Díez-Bedmar (2011) hizo un análisis de 302 composiciones en busca de los principales errores escritos en inglés cometidos por estudiantes

españoles en la prueba de ingreso a la Universidad de Jaén en junio de 2008. Por otro lado, en la presente investigación se analizan 515 composiciones recolectadas en el año 2015. Es importante notar que el trabajo escrito de los sujetos en cada estudio es diferente en cuanto a los temas, los antecedentes de los alumnos y el nivel de inglés, entre otros aspectos. En el caso del corpus de Diez-Bedmar, no se especifica el nivel de los estudiantes, pero se supone que podría ser un nivel de inglés preintermedio o avanzado, por las implicaciones que conlleva un examen universitario de ingreso. En el caso del presente trabajo, el corpus se obtuvo de los estudiantes inscritos en diferentes carreras de pregrado y que se registraron en un curso de ILE como parte de los requisitos de la institución educativa. Los niveles en este caso van desde A1 hasta un B1. Por las razones anteriores, los datos obtenidos no son completamente comparables. El único paralelo se hizo a partir de los resultados obtenidos en algunas categorías similares. La presente comparación es informativa ya que sería necesario tener un rango de variables más similar para poder replicar este estudio. Por lo tanto, esta comparación no es de ninguna manera concluyente. La Tabla 9 muestra algunos aspectos comparables entre el trabajo de Diez-Bedmar (2011) y la presente investigación.

Tabla 9

Comparación de errores en las versiones 1.1 y 1.2 del etiquetador de la Universidad de Lovaina

Tipo de error	Diez-Bedmar (2011)	Trabajo presente
LS	$M = 3,23$	$M = 3,94$
FS	$M = 3,13$	$M = 6,02$
GP	$M = 1,62$	GA No comparable

Fuente: Elaboración propia.

Como se observa en la Tabla 9, solo los errores Léxico, individual LS y Forma, Deletreo FS son comparables en este caso. Se puede ver que las medias (M) de error en LS y FS son más altas en el presente trabajo que las obtenidas por Diez-Bedmar (2011). También para notar: hay una diferencia en la tercera etiqueta con la media más alta para Diez-Bedmar en Gramática Pronombre GP, pero en el presente trabajo es Gramática Artículo GA. Aunque ambas etiquetas pertenecen a la categoría Gramática, Gramática Artículo GA se refiere a errores en el uso de artículos, mientras que Gramática Pronombre GP se refiere a errores que involucran el uso inapropiado de pronombres.

Conclusiones

El objetivo principal de este estudio fue identificar y analizar los tres errores más relevantes en la producción escrita de estudiantes de ILE a nivel universitario. También se buscaba obtener resultados comparables con trabajos similares. Por tal razón, los errores fueron etiquetados utilizando la versión 1.2 de la Universidad de Lovaina (Dagneaux et al., 2005a; 2005b) con el fin de obtener resultados comparables con trabajos similares a nivel mundial.

En referencia al primer objetivo, los errores con los puntajes más altos fueron Forma Deletreo FS, Gramática Artículo GA, Léxico Individual LS. Los resultados del puntaje total de las ocho categorías muestran que los aspectos gramaticales son los más problemáticos para los estudiantes de ILE a nivel universitario. Respecto de los tipos específicos de errores, el puntaje de Forma Individual FS, presenta la media más alta de errores por composición en esa categoría, lo que indica que los estudiantes no han dominado las reglas de ortografía y, en muchos casos, esos errores se deben a la transferencia de la lengua materna. En el caso de GA, los resultados son un síntoma de una interlengua que todavía está en proceso de cambio. Lo sorprendente es que los errores en el uso del artículo se mantienen con la misma intensidad en el nivel B2. En relación con los errores de Léxico Individual LS, un aspecto que desencadena la incidencia de este error es la selección inapropiada de vocabulario, seguida del uso equivocado de las colocaciones.

El segundo objetivo relativo a la necesidad de obtener resultados comparables se logró parcialmente con el uso del marcador de la Universidad de Lovaina. Aunque los resultados de este estudio no son completamente comparables con los de estudios anteriores, incluido Diez-Bedmar (2011), aquí se muestran algunas tendencias que podrían ser útiles para otros estudios que utilicen la misma metodología con un codificador de errores similar. El uso de este tipo de etiquetador para describir los diferentes errores escritos del trabajo de los estudiantes es una forma de motivar a otros investigadores para que realicen un trabajo comparable. Las tendencias que se muestran son informativas, por lo que se necesitarían más estudios para asegurar la comparabilidad de los resultados.

Se evidencia cómo el AE, a partir del uso de un corpus computacional, constituye una forma confiable de evaluar la interlengua de los estudiantes para mejorar las prácticas o materiales de enseñanza.

Agradecimientos:

A la universidad del Norte por permitir la recolección de los datos.

Al grupo TNT (Traducción y Nuevas Tecnologías) de la Universidad de Antioquia y su director, Doctor Gabriel Quiroz por su apoyo incondicional.

Referencias bibliográficas

- Alexopoulou, A. (2010). El papel de la transferencia en los errores léxicos. *Revista Nebrija de Lingüística Aplicada a La Enseñanza de Las Lenguas*, 5(9), 27-36.
- Alexopoulou, Angélica. (2005). Aproximación al tratamiento del error en la clase de E/LE desde la perspectiva del análisis de errores. *Estudios de Lingüística Aplicada*, 23(41).
- Brezina, V., McEnery, T., & Wattam, S. (2015). Collocations in context: A new perspective on collocation networks. *International Journal of Corpus Linguistics*, 20(2), 139-173. <https://doi.org/10.1075/ijcl.20.2.01bre>
- Corder, S. (1967). The significance of learner's errors. *IRAL - International Review of Applied Linguistics in Language Teaching*, 5(1-4), 161-170. <https://doi.org/10.1515/iral.1967.5.1-4.161>
- Corder, S. (1981). Error Analysis and Interlanguage. In *Oxford University Press*.
- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR)*. Cambridge University Press.
- Dagneaux, E., Denness, S., Granger, S., & Meunier, F. (1996). *Error tagging manual Version 1.1* (1.1). Louvain-la-Neuve: Centre for English Corpus Linguistics, Université Catholique.
- Dagneaux, E., Denness, S., Granger, S., Meunier, F., Neff, J., & Thewissen, J. (2005a). *Error Tagging Manual Version 1.2*. Centre for English Corpus Linguistics, Université Catholique de Louvain.
- Dagneaux, E., Denness, S., Granger, S., Meunier, F., Neff, J., & Thewissen, J. (2005b). *Error Tagging Manual Version 1.2* (pp. 1-46).

Centre for English Corpus Linguistics, Université Catholique de Louvain.

- Díez-Bedmar, M. B. (2011a). Spanish pre-university students' use of English: CEA results from the University Entrance Exam. *International Journal of English Studies*, 11(2), 141–158. <http://revistas.um.es/ijes/article/view/149681>
- Díez-Bedmar, M. B. (2011b). Spanish pre-university students' use of English: CEA results from the University Entrance Exam. *International Journal of English Studies*, 11(2), 141–158.
- Elejalde, J., & Ferreira, A. (2017). *Errores de transferencia en textos escritos del español como lengua extranjera producidos en ambientes de aprendizajes mediatizados por la tecnología* [Universidad de Concepción]. http://repositorio.udec.cl/bitstream/11594/2698/3/Tesis_Errores_de_transferencia_en_textos.pdf
- Elejalde, J., & Ferreira, A. (2018). Modelo para un tratamiento de errores de transferencia en la escritura en ELE producida en comunidades virtuales1. *Folios*, 48, 137–152. <file:///C:/Users/vicky/Downloads/tratamientodeerroresdetransferencia.pdf>
- Ferreira, A., Elejalde, J., & Vine, A. (2014). Análisis de errores asistido por computador basado en un Corpus de Aprendientes de Español como Lengua Extranjera. *Revista Signos*, 47(86), 385–411. <https://doi.org/10.4067/S0718-09342014000300003>
- Garside, Roger; Leech, Geoffrey ; McEnery, T. (Ed.). (1997). *Corpus Annotation: Linguistic Information from Computer Text Corpora*. Routledge. <https://doi.org/10.5715/jnlp.21.95>
- Granger, S, Gilquin, G, Meunier, F. (Ed.). (2015). *The Cambridge Handbook of Learner Corpus Research* (Sylviane G). Cambridge University Press.
- James, C. (1998). *Errors in Language Learning and Use. Exploring Error Analysis*. Routledge.
- Kilgarriff, A., & Rychly, P. (2003). *Sketch engine*. <http://www.sketchengine.co.uk/>
- Krashen, S. (1982). *Principles and Practice in Second Language Acquisition*. University of Southern California. <https://doi.org/10.2307/328293>
- Krashen, S. D. (1982). Principles and Practice in Second Language Acquisition. In *The Modern Language Journal* (Vol. 73, Issue 2). <https://doi.org/10.2307/328293>

- Lado, R. (1957). *Linguistics Across Cultures: Applied Linguistics for Language Teachers*. University of Michigan Press 1957.
- Leech, G. N. (2005). *Developing linguistic corpora a guide to good practice* (M. Wynne (Ed.)). Oxford:Oxbow Books.
- MacDonald, P. (2016). "We All Make Mistakes!". Analysing an Error-coded Corpus of Spanish University Students' Written English. *Complutense Journal of English Studies*, 24(0), 103-129. <https://doi.org/10.5209/CJES.53273>
- McEnery, A., Xiao, R., & Tono, Y. (2006). *Corpus-Based Language Studies: An advanced Resource Book*. Routledge.
- Morales-Reyes, A., & Gómez, I. (2016). Transfer and Semantic Universals in the L2 Acquisition of the English Article System by Child L2 Learners. *Language Acquisition*, 23(1), 57-74. <https://doi.org/10.1080/10489223.2015.1067318>
- Pardo, V. (2020). La construcción de un corpus computarizado de errores escritos con textos de estudiantes universitarios en Colombia. *Lingüística y Literatura*, 41(78), 36-55. <https://doi.org/https://doi.org/10.17533/udea.lyl.n78a02>
- Rosen, A., Hana, J., Tindlov, B., & Feldman, A. (2014). Evaluating and Automating the Annotation of a Learner Corpus. *Language Resources and Evaluation*, 48(1), 65-92. <https://doi.org/10.1007/s10579-013-9226-3>
- Sánchez, M. E., Sevilla, Y., & Bachrach, A. (2014). Agreement processing in control and raising structures. Evidence from sentence production in Spanish. *Lingua*, 177, 60-77. <https://doi.org/10.1016/j.lingua.2015.12.014>
- Scott, M. (2008). *WordSmith Tools Version 5 [Lexical analysis software]*.
- Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics in Language Teaching*, 10(1-4), 209-232.
- Survey Monkey. (2017).
- Vásquez, D. (2008). Error Analysis in a Written Composition. *Profile*, 10, 135-146.
- Virtanen, T. (1996). Exploiting the International Corpus of Learner English. *AFinLAN Vuosikirja*, 54, 157-166.

Anexo 1

Tipos de errores

FM	FM Form, Morphology
FS	FS Form, Spelling
FSR	FSR Form, Spelling, Regional
GA	GA Grammar, Articles
GADJCS	GADJCS Grammar, Adjectives, Comparative / Superlative
GADJN	GADJN Grammar, Adjectives, Number
GADJO	GADJO Grammar, Adjectives, Order
GADVO	GADVO Grammar, Adverbs, Order
GDD	GDD Grammar, Determiner, Demonstrative
GDI	GDI Grammar, Determiner, Indefinite
GDO	GDO Grammar, Determiner, Possessive
GDT	GDT Grammar, Determiner, Other
GNC	GNC Grammar, Nouns, Case
GNN	GNN Grammar, Nouns, Number
GPD	GPD Grammar, Pronouns, Demonstrative
GPF	GPF Grammar, Pronoun, Reflexive/Reciprocal
GPI	GPI Grammar, Pronoun, Indefinite
GPO	GPO Grammar, Pronoun, Possessive
GPP	GPP Grammar, Pronoun, Personal
GPR	GPR Grammar, Pronoun, Relative/ Interrogative
GPU	GPU Grammar, Pronoun, Unclear reference
GVAUX	GVAUX Grammar, Verbs, Auxiliaries
GVM	GVM Grammar, Verbs, Morphology
GVN	GVN Grammar, Verbs, Number
GVNF	GVNF Grammar, Verbs, Non-Finite / Finite
GVT	GVT Grammar, Verbs, Tense
GVV	GVV Grammar, Verbs, Voice
GWC	GWC Grammar, Word Class
LCC	LCC Lexis, Conjunctions, Coordinating

LCLC	LCLC Lexis, Connectors, Logical, Complex
LCLS	LCLS Lexis, Connectors, Logical, Single
LCS	LCS Lexis, Conjunctions, Subordinating
LP	LP Lexical Phrase
LPF	LPF Lexical Phrase, False friends
LS	LS Lexical Single
LSF	LSF Lexical Single, False friends
QC	QC Punctuation, Confusion
QL	QL Punctuation, Lexical
QM	QM Punctuation, Missing
QR	QR Punctuation, Redundant
SI	SI Sentence, Incomplete
SU	SU Sentence, Unclear
WM	WM Word Missing
WO	WO Word Order
WRM	WRM Word Redundant Multiple
WRS	WRS Word Redundant Single
XADJCO	XADJCO LeXico-Grammar, Adjectives, Complementation
XADJPR	XADJPR Lexico-Grammar, Adjectives, Dependent Preposition
XCONJCO	XCONJCO LeXico-Grammar, Conjunctions, Complementation
XNCO	XNCO LeXico-Grammar, Nouns, Complementation
XNPR	XNPR LeXico-Grammar, Nouns, Dependent Preposition
XNUC	XNUC LeXico-Grammar, Nouns, Uncountable / Countable
XPRCO	XPRCO LeXico-Grammar, Prepositions, Complementation
XVCO	XVCO LeXico-Grammar, Verbs, Complementation
XVPR	XVPR LeXico-Grammar, Verbs, Dependent Preposition
Z	Z Infelicities