



Revista Signos

ISSN: 0035-0451

revista.signos@ucv.cl

Pontificia Universidad Católica de Valparaíso
Chile

Ferreira Cabrera, Anita; Elejalde Gómez, Jessica; Vine Jara, Ana
Análisis de Errores Asistido por Computador basado en un Corpus de Aprendientes de Español como
Lengua Extranjera
Revista Signos, vol. 47, núm. 86, diciembre-, 2014, pp. 385-411
Pontificia Universidad Católica de Valparaíso
Valparaíso, Chile

Disponible en: <http://www.redalyc.org/articulo.oa?id=157032730003>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica
Red de Revistas Científicas de América Latina, el Caribe, España y Portugal
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

Análisis de Errores Asistido por Computador basado en un Corpus de Aprendientes de Español como Lengua Extranjera¹

*Computer Aided Error Analysis of a Computer based on Learner
Corpus for Spanish as a Foreign Language*

Anita Ferreira
Cabrera

UNIVERSIDAD DE CONCEPCIÓN
CHILE
aferreir@udec.cl

Jessica Elejalde
Gómez

UNIVERSIDAD DE CONCEPCIÓN
CHILE
jelejalde@udec.cl

Ana Vine
Jara

UNIVERSIDAD DE CONCEPCIÓN
CHILE
avine@udec.cl

Recibido: 14-X-2013 / **Aceptado:** 18-IV-2014

Resumen

En este artículo se presenta un estudio sobre Análisis de Errores asistido por computador (CEA) basado en un Corpus de Aprendientes de Español como Lengua Extranjera (ELE). El corpus se compone de 84 resúmenes: 40 textos en modalidad expositiva, 22 en narrativa y 22 en argumentativa, producidos por 22 estudiantes extranjeros de nivel B1 del Programa de ELE de la Universidad de Concepción (ele.udec.cl). Las tareas de escritura a través del computador corresponden a la producción escrita de resúmenes de 250 palabras cada uno a partir de la lectura de textos con temáticas científicas y culturales. La metodología se basa en el Análisis de Errores Asistido por el Computador y en los procedimientos de Corpus de Aprendientes de Lenguas en Formato Electrónico en lo que se refiere a la recopilación del corpus, a la anotación lingüística del corpus y al procesamiento automático de los datos a través de la herramienta computacional Nvivo. El propósito es delimitar los tipos de errores más frecuentes que cometen estudiantes de ELE de nivel B1. Los resultados del procesamiento del corpus que arroja este estudio evidencian que los errores de mayor frecuencia corresponden a la ortografía acentual, seguido por las preposiciones, la concordancia gramatical, el verbo y los artículos. Estos hallazgos tendrán implicaciones para la delimitación, identificación y tratamiento de los errores a través de un Sistema Tutorial Inteligente (STI) para ELE.

Palabras Clave: Análisis de Errores Asistido por Computador, Corpus de Aprendientes de Lenguas en Formato Electrónico, Sistemas Tutoriales Inteligentes, Español como Lengua Extranjera.

Abstract

This paper presents a Computer Aided Error Analysis (CEA) study based on Spanish as a Foreign Language Learners' Corpus. The corpus is made up of 84 summary texts: 40 of them are of expository modality, 22 narrative and 22 argumentative. These were written in Spanish by 22 international students from diverse university study programs with a B1 language proficiency level, enrolled in a b-learning Spanish as a foreign language course at Universidad de Concepcion, Chile (ele.udec.cl). The writing computerized tasks involved the production of 250-word-summaries based on scientific, historical and cultural topics that learners were asked to read. The methodology included Computer Aided Error Analysis and Computer Learner Corpora proceedings for the corpus construction, linguistic annotation and data processing using the NVIVO software tools. The aim is to determine error types, with the highest frequency of occurrence, committed by learners of Spanish as a foreign language. The corpus processing results suggest errors with the highest frequency correspond to orthographical stress-marking errors, followed by grammatical errors; such as, prepositions, grammatical agreement, verbs and articles. These findings will have implications in the delimitation, identification and treatment of errors with the use of an intelligent tutorial system for Spanish as a foreign language.

Key Words: Computer-aided-error-analysis (CEA), Computer Learner Corpora (CLC), Intelligent Tutorial Systems for Foreign Languages (ITS for FL), Spanish as a Foreign Language.

INTRODUCCIÓN

La investigación en Corpus de Aprendientes de Lenguas en Formato Electrónico (del inglés *Computer Learner Corpora*, CLC) data de la década de los 80. Esta línea de investigación fue creada como un vínculo entre la investigación en Lingüística de Corpus (LC) y la Adquisición de Segundas Lenguas (ASL) (Granger, 2002). Su objetivo principal es el estudio y análisis de la interlengua de aprendientes de lenguas segundas (L2) o extranjeras (LE). En este sentido, la investigación busca responder a planteamientos sobre los procesos y fenómenos de la interlengua, la adquisición y el aprendizaje de una L2 o LE. Los errores, entendidos como desviaciones de las normas de la lengua meta, son parte natural del aprendizaje de la lengua (Ellis, 1997). Estas desviaciones son sistemáticas y forman parte del proceso de adquisición y aprendizaje de una L2. Los errores revelan patrones de desarrollo de los sistemas de interlengua de los estudiantes que aprenden una segunda lengua, señalando donde ellos sobregeneralizan una regla o donde transfieren de manera inapropiada una regla de la lengua materna a la segunda lengua.

En este artículo, presentamos un estudio de Análisis de Errores Asistido por el Computador basado en los procedimientos de Lingüística de Corpus en lo que atañe al área de Corpus de Aprendientes de Lenguas en Formato Electrónico. El estudio se sustenta en investigaciones previas sobre errores de lengua y estrategias de *feedback* correctivo en Español como Lengua Extranjera (ELE) (Ferreira, 2006; Ferreira,

Moore & Mellish, 2007) y más recientemente en la arquitectura e implementación de un Sistema Tutorial Inteligente (STI) para el tratamiento de errores de ELE (Ferreira & Kotz, 2010; Ferreira, Salcedo, Kotz & Barrientos, 2012; Kotz & Ferreira, 2013). El objetivo es determinar las frecuencias de los errores lingüísticos y etiológicos observados en un corpus de resúmenes en formato electrónico producidos por aprendientes de ELE de nivel B1. La delimitación de dichos errores nos permitirá seleccionar los errores más frecuentes de todo el corpus electrónico con el objeto de que sean tratados a través de estrategias de *feedback* correctivo escrito en el contexto de un STI para el ELE que está en implementación en el contexto de una investigación mayor.

El artículo se organiza en las siguientes secciones: En la sección 1, nos referimos a los principales fundamentos teóricos en materia de análisis de errores y lingüística de corpus. En la sección 2, abordamos el estudio de análisis de errores. En la sección 3, presentamos los resultados de las frecuencias de errores de todo el corpus para luego delimitar la selección de los errores con mayor frecuencia que deberían ser implementados en el STI para ELE. Por último, presentamos algunos comentarios finales y proyecciones sobre los avances y logros obtenidos en esta investigación.

1. Fundamentos teóricos

1.1. El análisis de errores en el marco de la Lingüística de Corpus

Si bien la teoría del Análisis de Errores (AE) ha sido ampliamente discutida y criticada, hoy en día ha tomado una nueva posición frente a la rigurosidad metodológica de su análisis (Dagneaux, Denness & Granger, 1998; Alba Quiñones, 2009a; Vázquez, 2009; Reppen, 2010). Esto gracias a los avances tecnológicos, cuyas aplicaciones en la investigación lingüística han favorecido el desarrollo de herramientas metodológicas y computacionales para el estudio de la lengua en uso. Entre estas aplicaciones puede mencionarse la evolución de los métodos y técnicas para la recolección y tratamiento de textos en la Lingüística de Corpus (LC).

1.2. La Lingüística de Corpus (LC)

La Lingüística de Corpus se focaliza en el estudio de la lengua en contextos de uso real para observar los fenómenos ocurrentes. Su base metodológica se sustenta en la utilización de corpus electrónicos, cuyo valor se refleja en la autenticidad de los datos obtenidos. Los corpus se pueden procesar de forma automática o semiautomática para obtener resultados sobre la observación de fenómenos lingüísticos (Leech, 1992; Granger, 2002; Parodi, 2008, 2010). La autenticidad de las muestras se manifiesta en la forma en que se obtienen los datos, dando como resultado la posibilidad de observar el estado y uso real de la lengua. La rigurosidad metodológica de la LC ha apoyado la investigación en diferentes ámbitos interdisciplinarios, como por ejemplo, en Lingüística Aplicada en el área de la Adquisición de Segundas Lenguas. En este

sentido, el análisis de corpus se ha constituido en una de las principales fuentes de investigación para la adquisición y enseñanza-aprendizaje de lenguas segundas (L2) y extranjeras (LE), permitiendo observar e identificar las tendencias de uso de las palabras más frecuentes y, por ende, identificar los errores más frecuentes y recurrentes de la interlengua de aprendientes (IL) (Granger, 2002). Los resultados de dichos estudios han contribuido en el diseño de modelos de tratamiento de los errores para el mejoramiento de la competencia lingüística, así como también en el ámbito de la enseñanza de la lengua en lo que corresponde al diseño metodológico de tareas y actividades significativas para el aprendizaje de lenguas extranjeras.

1.3. Los corpus electrónicos de aprendientes de lengua

Un corpus electrónico de aprendientes de lengua es la colección de datos lingüísticos auténticos (textos orales o escritos) donde se constata el uso de la lengua objeto de estudio (L2 o LE) (Granger, 2003, 2004). El análisis de estos corpus provee resultados de frecuencias de usos adecuados o erróneos, concurrencias de las frecuencias, recurrencias y tendencias de uso (Dagneaux, Denness & Granger, 1998; Granger, 2003, 2004, 2009; Parodi, 2008, 2010; Reppen, 2010). Es así como el corpus de aprendientes se constituye en un valioso recurso para la investigación en el ámbito de la ASL y en la enseñanza de L2. Acorde con Granger (2003), los corpus de aprendientes son útiles cuando el enfoque de investigación está orientado al problema, como por ejemplo, el análisis del error.

1.3.1. Características de los CLC

Las características de los corpus de aprendientes de lenguas son innumerables según el tipo de investigación. Por esta razón, en el diseño y construcción de CLC debe considerarse aspectos tales como: los rasgos del texto, la recolección de los datos y las variables del aprendiente (ver Tabla 1).

Tabla 1. Adaptada y traducida de Tono (2003): Diseño y consideraciones para la construcción de un corpus de aprendientes.

Características para el diseño de un corpus de aprendientes					
Rasgos del texto		Recolección de los datos		Variables del aprendiente	
Modo	Escrito/ oral	Tipo y modo de recolección	Concurrente/ longitudinal Manual/digital	Internas: Cognitiva/ afectiva	Edad/estilo de aprendizaje Motivación/ actitud
Género	Diario/ensayos personales	Elicitación	Espontánea/ Por instrucción	L1 conocimiento	Lengua materna del aprendiente
Estilo	Narrativo/ argumentativo	Uso de referencias	Diccionario/ texto fuente	L2= segunda extranjera	L2/ LE
Tópico	General/ específico	Limitación de tiempo	Libre/controlado	Nivel de proficiencia Evaluación: Externa	Test de proficiencia

Como se puede observar en la Tabla 1, las características sobre los rasgos del texto se refieren al tipo de mensaje lingüístico y sus aspectos principales. Esto es, el modo (texto escrito u oral), el género como el diario, la carta u otros, el estilo o tipo de modalidad discursiva que presenta el mensaje (narrativo, argumentativo o expositivo) y finalmente, el tópico del cual trata el texto. Por otra parte, en la recolección de los datos, es importante señalar el modo de recolección de estos. Algunos métodos de elicitación son las tareas realizadas de forma física (escritos hechos a mano) o la grabación de conversaciones espontáneas. En el caso de las tareas escritas a mano, la forma de digitalización para construir el corpus escrito es la siguiente: 1) a través de la creación de documentos PDF donde se escanea el texto original escrito a mano y 2) posteriormente se transcribe el texto en un procesador de texto con formato .txt. Este proceso se realiza con el fin de conservar el escrito original lo más auténtico posible para realizar la transcripción y los análisis llevados a cabo por programas de procesamiento de corpus y datos lingüísticos.

Otro modo de recolección es por medio del diseño de tareas realizadas a través del computador, ya sea de forma oral o escrita. Para el caso del corpus escrito se recomienda utilizar el bloc de notas o programas que permitan guardar en formato txt y, que a su vez, no tenga la opción de un corrector ortográfico. De esta manera, se cautela que el texto escrito por el estudiante se realice de forma natural y sin intervención del corrector del programa.

Las variables del aprendiente son características importantes ya que deben corresponderse con los propósitos investigativos. Por ejemplo, si se pretende realizar una descripción o un contraste en relación con los niveles de proficiencia, la lengua materna del aprendiente y el contexto de aprendizaje de la lengua objeto de estudio (L2 o LE), la selección de los sujetos debe realizarse acorde con dichos criterios.

Además de estas características, existen otras referidas al ‘tamaño’, ‘la representatividad’ y ‘la extensión del corpus’. Estas tres características dependerán del propósito de la investigación y de la disponibilidad de la muestra (Granger, 2003, 2009; Parodi, 2008, 2010; Reppen, 2010). En el caso de los estudios exploratorios, se considera pertinente que estas características sean de mayor magnitud que la de estudios de tipo descriptivo y con fines específicos académicos. Parodi (2007: 105) explica que:

“el desafío de contar con un corpus representativo de una variedad determinada de lengua –incluso de un único registro específico de tal o cual lengua– es una cuestión compleja debido a la enorme diversidad y variedad inherente a cada lengua particular”.

En la misma dirección, Granger (2003) y Reppen (2010) proponen que la representatividad podrá adecuarse según la línea de investigación y el objeto de estudio.

1.4. El Análisis de Errores Asistido por Computador

Junto con el creciente avance de los procedimientos metodológicos de la LC, el Análisis de Errores implementó en sus estudios técnicas derivadas de la LC. Por consiguiente, El Análisis de Errores Asistido por Computador (del inglés *Computer-aided-error-analysis* (CEA)) es un enfoque de investigación basado en corporas electrónicos de aprendientes y en los procedimientos de la LC para la identificación, clasificación y descripción de los errores. Este tipo de estudio permite presentar los errores y su frecuencia en el contexto del texto a través de herramientas de análisis computacionales automáticos (Dagneaux et al., 1998; Granger, 2002, 2004).

Una de las etapas en CEA es diseñar un sistema de etiquetas de anotación de errores y etiquetar todos los errores en el corpus de aprendientes en diversos niveles de categorización. Este método tiene la ventaja de permitir un estudio enfocado al error donde las posibilidades de aplicación a otras disciplinas pueden derivarse del resultado de este (por ejemplo, en interdisciplinas como *Computer Assisted Language Learning* (CALL), *Intelligent Tutorial Systems for FL* (ITS), ASL, etc. (Granger, 2002).

1.4.1. Criterios y dimensiones de la taxonomía de errores

Las dimensiones se refieren a los criterios de descripción, clasificación y explicación de la modificación de la lengua objeto de estudio (LO) presentes en el corpus de aprendientes (Corder, 1967; Granger, 2004; Díaz-Negrillo & Domínguez, 2006). En este sentido, la inclusión de las dimensiones en una taxonomía de error corresponde a: (1) Criterio de Clasificación lingüística: nivel lingüístico en el que se encuentra el error (léxico, categoría gramatical) y (2) Taxonomía de modificación de la LO: se refiere a las alteraciones o desviaciones observadas en los errores tales como ‘la omisión’, ‘la adición’, ‘la sustitución’ y ‘el orden’. La combinación de estas dos dimensiones permite la construcción de una taxonomía que da cuenta tanto del error en un nivel lingüístico, como de su respectiva categorización en el tipo de error cometido por el aprendiente.

1.4.2. Características del sistema de anotación de errores

Los sistemas de anotación de errores están basados en la combinación de varias dimensiones propuestas para las taxonomías de error. Granger (2004) propone una serie de criterios para la elaboración de un sistema de anotación con la finalidad de mantener la estandarización de las etiquetas y su codificación: (1) Informativo pero práctico: debe tener información detallada para proveer datos útiles de los errores de los aprendientes. (2) Reutilizable: las categorías deben ser lo suficientemente generales para ser usadas en otras lenguas. (3) Flexible: debe permitir la adición o eliminación de etiquetas en el estado de anotación del corpus. (4) Consistente: la taxonomía y el sistema de anotación deben compararse por separado en archivos diferentes.

1.5. Análisis de Errores en Español como Lengua Extranjera

A partir de la década de los 90, se ha llevado a cabo una serie de estudios en ELE con el objeto de identificar y clasificar los errores acorde con el Análisis de Errores (Vázquez, 1991, 2009; Alba Quiñones, 2009a). Vázquez (1991), con la finalidad de determinar los errores cometidos por los estudiantes alemanes, realizó una descripción de los errores más frecuentes. El énfasis estuvo en el nivel morfosintáctico, a partir del material escrito (composiciones libres, cartas, tesis) y oral (entrevistas y conversaciones con los estudiantes). Su investigación se centró en los errores que se producen en la interlengua de los estudiantes y no sobre los errores previstos de antemano (a través del análisis contrastivo). Además, incluye la comparación de errores en distintos momentos del aprendizaje, se trata de errores observados en el nivel principiante que luego reaparecen en etapas posteriores. A partir de los resultados obtenidos en sus estudios, Vázquez (1991, 2009) propone una taxonomía de clasificación de errores, que engloba un amplio número de criterios (lingüísticos, etiológicos, comunicativos, pedagógicos, pragmáticos y culturales).

Por su parte, Alba Quiñones (2009a) investigó los errores cometidos por estudiantes alemanes de ELE con el propósito de identificar los errores en el subsistema léxico-semántico, específicamente, errores del significante y del significado. Para ello, se basó en una prueba compuesta por un cuestionario y una composición. Los resultados de su estudio señalan que los errores más frecuentes fueron los ‘cuasisinónimos’, debido en su mayoría a la interferencia tanto de la L1 como de la L2, y los elementos de campos distintos con algunos contextos comunes, como es el caso de los verbos ser-estar en su uso atributivo. Estos resultados indican que el nivel léxico-semántico es más permeable a la transferencia tanto de la L1 de los aprendientes como de una L2, anterior a la que están aprendiendo (en este caso, otra lengua aprendida antes del español). Cabe destacar que los estudiantes de la muestra presentaban dominio en otras LE, por lo que el español vendría a ser una L3 o L4.

2. El estudio

El enfoque de investigación de este estudio corresponde al Análisis de Errores Asistido por el Computador, basado en los procedimientos de la Lingüística de Corpus en lo referido a Corpus de Aprendientes en Formato Electrónico. Se propone examinar los errores de escritura cometidos por estudiantes de ELE al desarrollar una tarea de producción escrita de un resumen a través del computador. El objetivo es determinar los tipos de errores más frecuentes en un corpus de aprendientes de ELE de nivel B1, en formato electrónico, para ser implementados en un STI para ELE.

2.1. Objetivos específicos

- a) Construir una anotación para etiquetamiento de errores lingüísticos y etiológicos de un corpus de aprendientes de ELE en formato electrónico.

- b) Determinar las frecuencias de errores en un corpus de resúmenes producidos por aprendientes de ELE de nivel B1.

2.2. Muestra de aprendientes

La muestra se constituyó por 22 estudiantes universitarios extranjeros de la Universidad de Concepción, Chile. Los estudiantes pertenecían a los distintos programas de intercambio estudiantil de nivel de pre y postgrado. Estaban inscritos en el curso de ELE nivel B1 que el Programa de Español como Lengua Extranjera de la Universidad de Concepción, ELE-UdeC, ofrece semestralmente (<http://ele.udec.cl>). Las edades de los alumnos fluctuaban entre los 19 y 36 años de edad (23 años en promedio) y tenían como L1 las siguientes lenguas: el 50% (11 alumnos) tenía como L1 el alemán; el 22,7% (5 alumnos), el inglés, el 18,2 % el francés (4 alumnos), y el 9% (2 alumnos), el portugués.

En cuanto al nivel de proficiencia en español, este fue B1 determinado a través de dos mediciones: (1) se consideró el nivel declarado por los estudiantes a través de un cuestionario, luego este nivel fue corroborado mediante la aplicación de (2) la prueba de proficiencia para el nivel B1, CELE-UdeC, desarrollada por el Programa de Español ELE-UdeC (Ferreira, Vine & Elejalde, 2013).

2.3. Metodología de la Investigación: Análisis de Errores y Lingüística de Corpus

Como se ha mencionado anteriormente, la metodología se sustenta en el enfoque de investigación del Análisis de Errores Asistido por el Computador y en los procedimientos de Lingüística de Corpus, específicamente, de Corpus de Aprendientes en formato electrónico en lo que se refiere a la recopilación del corpus, a la definición de una anotación lingüística y al procesamiento de los datos a través de la herramienta computacional NVIVO. La delimitación y clasificación de los errores se basa en las propuestas de análisis de errores para ELE de Vázquez (1991, 2009) y Alba Quiñones (2009b). De acuerdo con el modelo metodológico se consideraron las siguientes etapas.

2.3.1. Recopilación del corpus

El corpus de textos se constituye de 84 resúmenes escritos: 40 textos expositivos, 22 narrativos y 22 argumentativos. Para la elicitación de este corpus textual se tuvieron en cuenta los siguientes pasos:

1. **La aplicación de un formulario de registro:** Este instrumento consistió en una serie de preguntas realizadas a los alumnos con el objeto de determinar datos, tales como: la edad, el nivel de español, los años de estudio del español, el lugar en que realizó dichos estudios, el manejo de otras lenguas extranjeras y su nivel de competencia en ellas, etc.

2. **El análisis de las necesidades lingüísticas:** Con el propósito de conocer las temáticas de interés de los estudiantes, se realizó un análisis de necesidades lingüísticas. Para ello, se solicitó a los estudiantes el envío de un texto sobre un tema de su interés (de una página aproximadamente) a la profesora.

A partir del análisis de las necesidades de los estudiantes se delimitó los tipos de textos y temas para la lectura y escritura de los resúmenes. Los textos de lectura tenían una extensión aproximada de 1.500 palabras. Antes de comenzar con el proceso de escritura de resúmenes se entregó una cápsula instruccional a los estudiantes respecto al tipo de texto resumen y su superestructura. Los alumnos tenían un tiempo de 40 minutos para leer el texto fuente y 50 minutos para escribir un resumen de 250 palabras, como mínimo. Cada resumen fue escrito en el computador en formato txt, de este modo se evitó que los estudiantes utilizarán correctores ortográficos automáticos. Cada uno de los textos se entregó en papel, de este modo, se facilitó la lectura y se cauteló que los aprendientes no copiaran y pegaran partes del texto fuente en el resumen que debían escribir. Al finalizar, los resúmenes eran enviados a una cuenta de correo electrónico creada para este fin. La tarea de escritura se realizó en el laboratorio de Aplicaciones Tecnológicas en Lingüística Aplicada (LATLA) en cuatro momentos distintos. En la Tabla 2, se esbozan los pasos realizados para lograr que los estudiantes pudieran escribir los resúmenes de los distintos tipos de textos.

Tabla 2. Descripción de las actividades.

Sesión N°	Descripción de la actividad
0	Pre-tarea (Análisis de necesidades): Cada estudiante envía un resumen a la profesora de un tema de interés de su disciplina.
1	Cápsula instruccional: Se explicaron los lineamientos sobre la superestructura textual de un resumen: título, introducción, desarrollo (ideas principales, conectores), conclusión. Elementos necesarios en todo proceso de escritura de un resumen.
2	Texto 1 (narrativo): Se entregó un conjunto de 3 textos narrativos para que el estudiante seleccionara uno, sobre el cual haría el resumen. Estos textos trataban sobre distintos músicos chilenos: (1) Vida de Violeta Parra, (2) Biografía de Víctor Jara, (2) La trayectoria del grupo Los Prisioneros.
3	Texto 2 (argumentativo): Los estudiantes leyeron un texto argumentativo sobre las distintas posiciones sobre el proyecto de ley que obliga a las emisoras a difundir música folclórica chilena.
4	Texto 3 (expositivo): Se les entregó un extracto de un artículo científico sobre la urbanización de las cuencas cercanas a Concepción para que realizaran el resumen.
5	Texto 4 (expositivo): Se les entregó un extracto de un artículo científico sobre el pasado minero de Lota, a partir del cual debían elaborar el resumen.

A continuación, se presenta un ejemplo del formato de instrucciones que se les entregó a los alumnos, específicamente, del texto narrativo correspondiente a Violeta Parra.

INSTRUCCIONES
En tu Universidad de origen, el Departamento de Relaciones Internacionales ha solicitado que envíes un resumen de la historia de un músico popular chileno que haya contribuido al acervo cultural del país.
Con este propósito ponemos a tu disposición tres textos sobre tres importantes músicos populares de Chile. Elige uno de ellos y sigue las instrucciones que se te presentan para cumplir con tu tarea.
<ul style="list-style-type: none"> • Lee el texto sobre la historia de la gran cantante popular chilena, Violeta Parra. • Luego, escribe en el computador un resumen de acuerdo con la estructura revisada en clases: título, introducción, desarrollo con ideas principales y de apoyo, conclusión.
- Este resumen debe tener como mínimo una extensión de 250 palabras.
<ul style="list-style-type: none"> • Una vez que finalices el resumen, envíalo a la siguiente dirección: espanolacademico@gmail.com

Figura 1. Plantilla de instrucciones para la escritura del resumen.

2.3.2. Identificación y clasificación de los errores

En primera instancia se llevó a cabo una revisión y análisis del corpus textual con el objeto de precisar una taxonomía para la identificación y clasificación de los errores encontrados. Para la clasificación de los errores se utilizó como base los criterios Lingüístico y Etiológico (Vázquez, 1991; Alba Quiñones, 2009b).

Tabla 3. Criterios para la clasificación.

Criterio	Tipo de error
Lingüístico	errores de adición
	errores de omisión
	errores de falsa selección
Etiológico	errores interlinguales
	errores intralinguales

Como se puede observar en la Tabla 3, dentro del criterio lingüístico se considera un “error de adición” cuando el estudiante agrega algún elemento, ya sea de tipo gramatical o léxico, que no es necesario en el texto, ejemplo: “entonces es necesario para tener sistemas para tratar el exceso del agua” (Sujeto 11, texto expositivo (a)), en este caso el estudiante agregó una preposición innecesaria. Por el contrario, un “error de omisión”, como su nombre lo indica, se refiere a la omisión de algún elemento lingüístico en el texto, por ejemplo: “El artículo discute que este proyecto de ley sea muy importante para el futuro de ____ música chilena” (Sujeto 16, texto argumentativo), en esta oración el estudiante omitió el artículo definido. Mientras que un error de falsa selección corresponde a la utilización poco pertinente de algún término o elemento gramatical, por ejemplo: “Su madre estuvo una persona optimista” (Sujeto 12, texto

narrativo). En esta oración el estudiante utilizó el verbo estar (pretérito perfecto simple, 3ª persona singular) cuando el contexto exige el verbo ser (pretérito perfecto simple, 3ª persona singular).

En el criterio etiológico se consideró como un “error de tipo interlingual” aquel que presenta influencia de la lengua materna del estudiante, ya sea de manera parcial o completa en la palabra u oración. Se consideraron cinco errores de este tipo: uso de L1, analogía semántica, calcos, traducción literal y orden de palabras. Por ejemplo, un error clasificado como uso de la L1 es el siguiente: “hay también canales de radio que tienen el focus en música chilena” (Sujeto 22, L1 inglés, texto argumentativo). El “error de tipo intralingual” corresponde a una problemática de la naturaleza propia de la lengua que se está aprendiendo, en este caso, el español. En este nivel intralingual, se consideraron tres tipos de errores: sobregeneralización, hipercorrección y simplificación. Por ejemplo, un error de sobregeneralización es “la imagen de un minero: manos robustos” (Sujeto 8, texto expositivo (b)). En este caso, el estudiante desconoce que el sustantivo ‘mano’ es femenino y aplica una regla de sobregeneralización de los sustantivos masculinos, haciendo concordar al adjetivo con la terminación del sustantivo.

2.3.3. Sistema de anotación de errores en ELE

Sobre la base de la delimitación del tipo de errores que se considerarían para la clasificación, se procedió a construir un sistema de anotación para el etiquetamiento de los errores del corpus de aprendientes de ELE en formato electrónico. Para ello, se consideraron los criterios de elaboración de anotaciones para etiquetajes de corpus, propuestos por Granger (2003, 2004). Para la clasificación de los errores fue necesario establecer los niveles y subniveles sobre los cuales se recolectaría la información necesaria para su posterior análisis. En la Tabla 4 se ilustran los niveles y subniveles de clasificación considerados para el sistema de anotación de los errores de ELE.

Tabla 4. Categorías y subcategorías de clasificación.

Ejemplos de etiquetas del sistema de anotación							
Criterio	Nivel	Categoría	Subcategorías		ADICIÓN	FALSA SELECCIÓN	OMISIÓN
Lingüístico	Palabra oración	Gramática	Categorías gramaticales		ADI-prep	FALSA-SE-LEC-prep	OMI-prep
			Coherencia textual	Concor. Sintáctica	FALSA-SELEC-CG-GEN		
				Puntuación	ADI-coma	FALSA-SE-LEC-coma	OMI-coma
				Conectores	ADI-Co-nect	FALSA-SE-LEC-Conect	OMI-Co-nect
	Palabra	Léxico	Léxico creado por derivación		FALSA-SELEC-LEX-Creado-deriv		
			Léxico -morfología		FALSA-SELEC-LEX-MORF		
			Léxico innecesario		ADI-LEX-Inne		
		Ortografía	Acentual		ADI-Orto-grave	INAPROD-Orto-grave	OMI-Orto-grave
			Literal		ADI-Orto-Lit_c	FALSA-SE-LEC-Orto-lit_c	OMI-Orto-lit_c
			Dierética		ADI-Orto-hiato	FALSA-SELEC-Orto-hiato	OMI-Orto-hiato
Etiológico	Palabra oración	Interlin-gual	Uso L1		Uso L1		
			Uso L2		Uso L2		
			Orden de palabras		Orden de palabras		
			Analogía semántica		Analogía semántica		
			Traducción literal		Traducción literal		
		Intralin-gual	Sobregeneralización		Sobregeneralización		
			Simplificación		Simplificación		
Otros	Errores de tipo				Error tipo		

2.3.4. Procesamiento del corpus

En el procesamiento del corpus de resúmenes en formato electrónico se procedió acorde con las fases de investigación ilustradas en la Tabla 5. El proceso de etiquetamiento de los 84 textos (40 textos expositivos, 22 narrativos y 22 argumentativos) consistió en identificar los errores a través de los anotadores, quienes debían marcarlos en la versión digital del texto, empleando el *software* especializado Nvivo 10. La utilización de este *software* tiene la finalidad de facilitar el procesamiento y análisis de los datos de forma automática. A continuación se detalla los pasos seguidos en la Tabla 5.

Tabla 5. Fases de la investigación.

Fase	Descripción
1	Proceso de etiquetado de errores del corpus
2	Consolidación de los archivos, etiquetados y anexados en el <i>software</i> Nvivo
3	Resultados emitidos por el <i>software</i> Nvivo

2.3.4.1. Proceso de etiquetado de errores del corpus a través del Software Nvivo

Para el etiquetado de errores se empleó el *software* de análisis de datos cualitativos Nvivo 10, con el fin de realizar el procesamiento y análisis del corpus de forma automática. Así en palabras de Dagneaux et al. (1998) la ventaja de utilizar *software* para el procesamiento automático de los datos es el de obtener rápidamente resultados de diferentes tipo de errores en contexto y la frecuencia de estos.

El programa Nvivo está diseñado por *Qualitative Software Research* (QSR) con el propósito de apoyar la investigación mixta, basada en datos cualitativos, asistida por computador (del inglés, *Computer Assisted Qualitative Data Analysis Software*, (CAQDA)). El *software* apoya al investigador en relación con la compilación de los recursos (textos escritos u orales), la organización de estos, la creación de la taxonomía de etiquetas para identificar fenómenos, el procesamiento semiautomático de cualquier tipo de recurso y el análisis automático de los datos. La obtención de estos datos con el *software* Nvivo, a partir de los recursos cualitativos, arroja resultados que pueden procesarse de forma cuantitativa y automática. Para ello, Nvivo utiliza herramientas como: 1) la ‘búsqueda de consulta automática’ para el procesamiento de resultados a partir del cruce de diferentes variables, 2) la visualización del ‘contexto de etiquetamiento’ y 3) el análisis de ocurrencias, aplicación de fórmulas de frecuencias y paquetes estadísticos, además de otras posibilidades computacionales.

El análisis automático que proporciona el *software* contribuye a la identificación del uso y las tendencias de un fenómeno observado durante el estudio de un corpus textual. Además, es importante señalar que la opción ‘contexto de etiquetamiento’, muestra el error inserto en el texto, cuya función permite un análisis contextualizado sobre de los fenómenos encontrados y su ocurrencia (Dagneaux et al., 1998).

2.3.4.2. Consolidación de los archivos

Cada uno de los evaluadores, una vez concluida la tarea de etiquetado de errores, entregó su archivo para que éste fuera consolidado en un único fichero. El proceso que se siguió para esto, consistió en un procedimiento de depuración y luego de consolidación de archivos en un solo fichero utilizando la función de importar que incorpora Nvivo 10. De este modo, los seis archivos fueron agrupados en un fichero único que contenía los 84 resúmenes.

3. Resultados del estudio

Los resultados generales de frecuencia de errores obtenidos en el corpus textual de 84 resúmenes de nivel B1 se presentan en la Tabla 6. Se puede apreciar la distribución y frecuencias de los distintos tipos de errores identificados y etiquetados por los evaluadores con criterios lingüísticos (gramática, léxico y ortografía) y etiológicos (interlinguales e intralinguales). De un total de 2.278 errores etiquetados en toda la muestra, el 87% corresponden a errores lingüísticos y 10% a errores etiológicos. En los errores lingüísticos, los errores de mayor frecuencia son los gramaticales con un 60% (errores de categorías gramaticales y de coherencia textual), seguidos por los errores de ortografía (acentual, literal y diacrítica) 18% y por los de léxico 9%. En cuanto a los errores etiológicos 157 son interlinguales y 62 intralinguales.

Según estos resultados se puede señalar que la mayoría de errores observados en los 84 textos resúmenes corresponden a errores lingüísticos y entre estos los más representativos son los errores gramaticales. En la Tabla 6 se observa que los errores gramaticales más frecuentes corresponden a los errores de categorías gramaticales (sustantivos, verbos, artículos, preposiciones, etc.): 791 errores (35%), seguidos por los errores de coherencia textual: 567 errores (26%).

En lo concerniente a los errores etiológicos interlinguales, los de mayor frecuencia fueron los errores de uso de la L1 con una frecuencia de 95 errores y el orden de palabras con una frecuencia de 32 errores. En los intralinguales, los errores de sobregeneralización fueron los que presentaron una mayor frecuencia, 52 errores.

El número menor de errores etiológicos comparados con los encontrados en lo lingüístico evidencia que las problemáticas pareciesen estar en el plano de la precisión lingüística.

Tabla 6. Resultados generales de toda la muestra.

RESULTADOS GENERALES										
Criterio	Nivel	Categoría	Subcategorías		Fr. Sub	Fr.cat	% Sub	Fr.Abs	% Cat.	% Total
Lingüístico	Palabra oración	Gramática	Categorías gramaticales		791	791	35%	1358	60%	87%
			Coherencia textual	Concor. Sintáctica	313	567	25%			
				Puntuación	198					
				Conectores	56					
	Palabra	Léxico	Léxico creado por derivación		93	217	9%	217	9%	
			Léxico -morfología		121					
			Léxico innecesario		3					
		Ortografía	Acentual		317	413	18%	413	18%	
			Literal		51					
			Dierética		45					
Etiológico	Palabra oración	Interlingual	Uso L1		95	157	7%	219	10%	10%
			Uso L2		10					
			Orden de palabras		32					
				Analogía semántica						
			Traducción literal		7					
		Intralingual	Sobregeneralización		52	62	3%			
			Simplificación		10					
Otros	Errores de tipo				71	71	3%	71	3%	3%
Totales					2278		100%	2278	100%	100%

En el Gráfico 1 se muestra las frecuencias totales por cada uno de los grupos de errores etiquetados en todo el corpus textual.

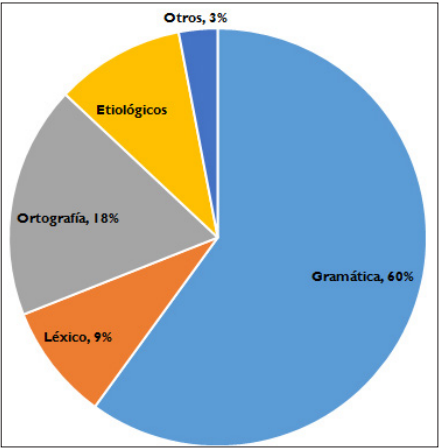


Gráfico 1. Resultados de frecuencia total en la muestra.

En consonancia con los resultados generales del procesamiento del corpus, se requiere indagar de manera más específica en torno a los errores de gramática y ortografía, dado que este tipo de errores fueron los que obtuvieron las mayores frecuencias en el análisis de errores del corpus de resúmenes. Pues bien, si nos centramos solo en los errores gramaticales en lo concerniente a las categorías gramaticales, el Gráfico 2 ilustra que los errores de mayor frecuencia corresponden a las preposiciones con un 39%, a los verbos con un 23% y a los artículos con un 17%. Esto quiere decir, que los estudiantes cometieron un mayor número de errores en sus textos cuando usaron las preposiciones en español, los verbos y los artículos.

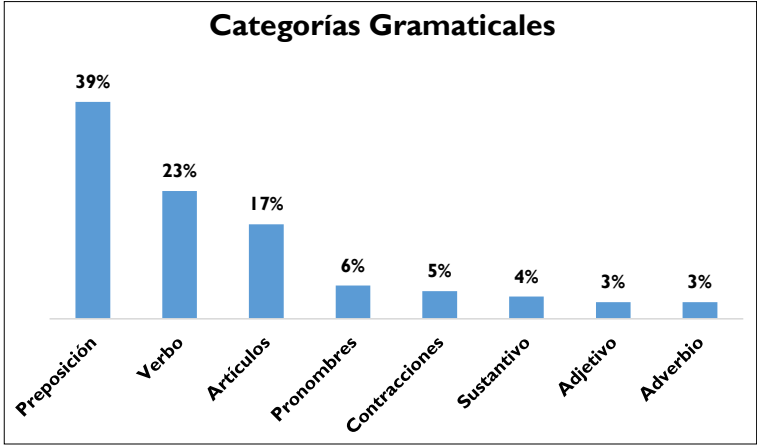


Gráfico 2. Frecuencia de errores en Categorías Gramaticales.

En la Tabla 7 se muestra que, en el caso de las preposiciones, las mayores problemáticas observadas en los textos fueron la falsa selección de la preposición (122 errores) seguida por la omisión de la preposición (120 errores). Ejemplos:

- Falsa selección de la preposición: “El artículo describe, principalmente, cuales son los peligros en contra a las cuencas en Chile” (Texto Expositivo- Sujeto16).
- Omisión de la preposición: “Ayuda a la vegetación en el área. Pero también ayudaría la ciudad y... la sociedad cuando hay inundaciones o mucha” (Texto Expositivo- Sujeto1).

Tabla 7. Frecuencia de errores en Categorías Gramaticales.

Errores de las Categorías gramaticales					
Categoría	Tipo error Gramática		Fr. Abs	Fr. Cat	%
Preposición	FALSA-SELEC-prep		122	309	39%
	OMI-prep		120		
	ADI-prep		67		
Verbo	Tiempos verbales	FALSA-SELEC-modo Subjuntivo	34	180	23%
		FALSA-SELEC-T-pret-imperf	22		
		FALSA-SELEC-T-pret-perf-simple	8		
	Verbo estar	FALSA-SELEC-verbo estar	17		
	Verbo ser	FALSA-SELEC-verbo ser	21		
		ADI-verbo ser	5		
	Morfología del verbo	FALSA-SELEC-lex-morf-deriv-irreg	5		
		FALSA-SELEC-lex-morf-deriv-reg	9		
		FALSA-SELEC-lex-morf-radic-verb-irreg	26		
	Formas no personales del verbo	FALSA-SELEC-gerundio	13		
		FALSA-SELEC-infinitivo	11		
		FALSA-SELEC-participio	9		
Artículos	Art. Def	OMI-art-def	93	131	17%
		ADI-art-def	16		
		FALSA-SELEC-art-def	11		
	Art.Indef	OMI-art-indef	6		
		ADI-art-indef	5		
Pronombres	Pron.Relat	ADI-pron-relat	15	50	6%
		FALSA-SELEC-pron-relat	13		
		OMI-pron-relat	11		
	Pron. Pers	OMI-pron-pers	6		
	Pron.átono	FALSA-SELEC--pron-átono	5		
Contracciones	Contracción al	FALSA-SELEC-contrac_al	12	36	5%
	Contracción del	FALSA-SELEC-contrac_del	13		
		OMI-contrac_del	11		
Sustantivo		FALSA-SELEC-sust	25	34	4%
		OMI-sust	9		
Adjetivo		FALSA-SELEC-adj	27	27	3%
Adverbio		FALSA-SELEC-adv	14	24	3%
		ADI-adv	10		
Total			791	791	100%

En cuanto al verbo, el número mayor de errores encontrados en el uso de los tiempos verbales corresponde a la falsa selección del modo subjuntivo (34 errores), falsa selección del tiempo pretérito imperfecto (22 errores), falsa selección de los verbos ser (21 errores), estar (17 errores). Ejemplos:

- Falsa selección del modo subjuntivo: “...Es necesario para tener espacios que pueden soportar las lluvias intesa...” (Texto Expositivo– Sujeto 11).
- Falsa selección de los verbos ser y estar: “Durante esta época, Lota estaba la primera ciudad industrial del país.” (Texto Expositivo – Sujeto 17).

En cuanto a la categoría artículo, los errores más frecuentes son los definidos (120 errores): omisión del artículo (93 errores), adición del artículo (16 errores) y falsa selección del artículo (11 errores). Ejemplos:

- Omisión del artículo: “Originalmente los jóvenes quieren atraer solo ... atención con este nombre.” (Texto Narrativo-Sujeto 22).
- Adición del artículo: “Víctor Jara empezó a estudiar la contabilidad, pero cuando su madre.. (Texto Narrativo - Sujeto 7).

Estos resultados son concordantes con otros estudios sobre frecuencias de errores en español (Vázquez, 1991; Alexopoulou, 2005, 2006) en donde las categorías gramaticales como las preposiciones, los verbos y, la concordancia gramatical de género y número presentan mayores problemáticas en el proceso de adquisición y aprendizaje del español como lengua extranjera.

En lo que respecta a la subcategoría de errores gramaticales, denominada errores de coherencia textual, la Tabla 8 evidencia que los errores de mayor frecuencia son los de concordancia sintáctica (289 errores), de estos 208 son de falsa selección de género gramatical y 81 errores de número gramatical. Ejemplos:

- Falsa selección de género gramatical: “..este sistema está perturbado por la urbanización, se necesita otras espacios para soportar adecuadamente la extirpación de.....” (Texto Expositivo- Sujeto 21).
- Falsa selección de número gramatical: “Como los mapuche dicen en mapudungún, Lota era una vez un...” (Texto Expositivo- Sujeto 11).

Tabla 8. Frecuencia de errores en la Concordancia Sintáctica.

Errores de Concordancia Sintáctica					
Categoría	Subcategoría	Tipo de error	Fr. Error	Fr. Sub	%
Concordancia sintáctica	Concordancia sintáctica	FALSA-SELEC-CG-Gen	208	289	92%
		FALSA-SELEC-CG-num	81		
	Concordancia entre sintagmas	FALSA-SELEC-CG-suj-pred	13	24	8%
		OMI-sintag-nominal	6		
		FALSA-SELEC-atributo ser o estar	5		
Totales			313	313	100%

En cuanto a los errores de ortografía, la Tabla 9 da cuenta de las frecuencias en materia de ortografía acentual (317 errores, 77% del total de errores de ortografía acentual), literal (51 errores, 12%) y dierética (45 errores, 11%). En la ortografía dierética, la mayor problemática estuvo en la omisión de tildes en hiatos (36 errores). En la ortografía literal en lo referido a la omisión y adición de mayúsculas (26 errores).

Tabla 9. Frecuencia de errores en la Ortografía.

Ortografía				
Subcategoría	Tipo de error	Fr. Error	Fr. Sub	Fr%
Ortografía acentual	OMI-Orto-aguda	104	317	77%
	INAPROP-aguda	59		
	OMI-Orto-esdrújula	98		
	INAPROP-esdrújula	6		
	INAPROP-grave	26		
	OMI-Orto-grave	24		
Ortografía dierética	OMI-Orto-hiato	36	45	11%
	ADI-Orto-hiato	4		
	FALSA-SELEC-Orto-hiato	4		
	OMI-Orto-triptongo	1		
Ortografía literal	OMI-orto-mayúscula-inicial	16	26	6%
	ADI-orto-mayúscula-inicial	10		
	FALSA-SELEC-orto-lit_c	10	20	5%
	FALSA-SELEC-orto-lit_s	10		
	OMI-orto-lit_z	5	5	1%
Totales		413	413	100%

La Tabla 10 evidencia las frecuencias de errores donde se delimitan las frecuencias de cada uno de los tipos de errores acentuales. Se observa que los errores de mayor frecuencia son los de omisión de tildación en las palabras agudas (104 errores), seguidos por los de omisión en la tildación de palabras esdrújulas (98 errores) y los de tildación inapropiada en las palabras agudas (59 errores).

Tabla 10. Frecuencia de errores en Ortografía Acentual.

Errores de ortografía acentual					
Categoría	Subcategoría	Tipo de error	Fr. Error	Fr. Sub	%
Ortografía acentual	Aguda	OMI-Orto-aguda	104	163	51%
		INAPROP-aguda	59		
	Esdrújula	OMI-Orto-esdrújula	98	104	33%
		INAPROP-esdrújula	6		
	Grave	INAPROP-grave	26	50	16%
		OMI-Orto-grave	24		
Totales			317	317	100%

Así mismo en el Gráfico 3 se puede apreciar que las palabras agudas (51%) son las que presentaron mayores complicaciones en cuanto a la tildación, seguidas por las palabras esdrújulas (33%) y las graves (16%). Ejemplos:

- Omisión de tildación en las palabras agudas: “Para países con una lengua distinto de ingles es mas importante proteger la musica nacional.” (Texto Argumentativo-Sujeto 10).
- Omisión en la tildación de palabras esdrújulas: “Refierendose a Concepción el autor informe sobre la explosión demografica como consecuencia de los procesos de industrialización ...” (Texto Expositivo - Sujeto 2).

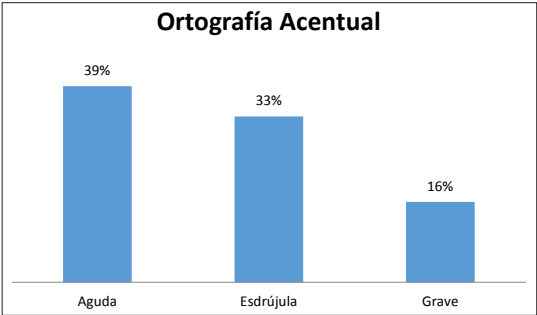


Gráfico 3. Frecuencia de errores en Ortografía Acentual.

Finalmente, sobre la base de las frecuencias de errores hasta aquí delimitadas, a través del análisis y procesamiento del corpus, podemos señalar que los tipos de errores observados con mayor frecuencia en todo el corpus de resúmenes de ELE corresponden a los errores de ortografía (26%), preposiciones (25%), concordancias sintácticas (23%), verbos (17%) y artículos (10%) (ver Gráfico 4).

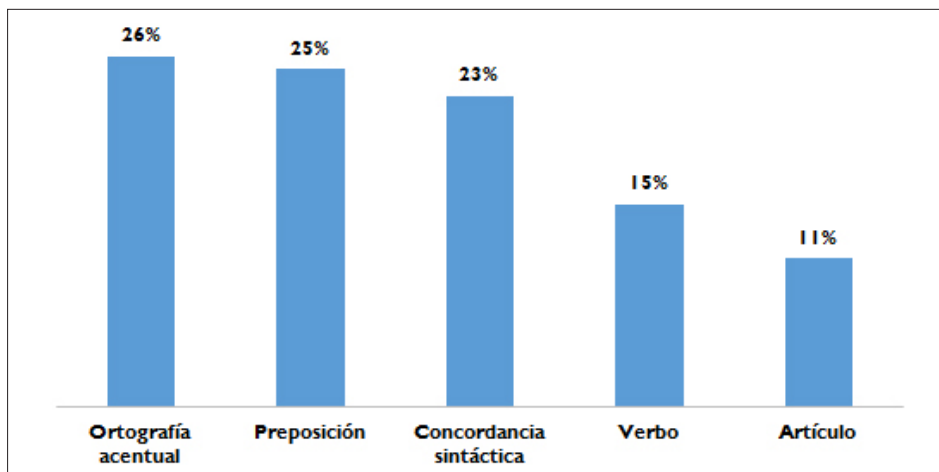


Gráfico 4. Errores más frecuentes en el Corpus Textual.

Como se puede observar en la Tabla 11, en la ortografía acentual los problemas mayores los presentan la tildación de las palabras agudas (163) y palabras esdrújulas (104). En las preposiciones (309 errores) los errores con mayor frecuencia corresponden a los de falsa selección (122 errores), omisión (120 errores) y adición (67 errores) de preposiciones. En cuanto a la concordancia gramatical, los errores más frecuentes son de concordancia de género (208 errores) y de concordancia de número (81 errores).

En lo referido al verbo, las mayores frecuencias se distribuyen en cuanto a los tiempos verbales (64) y al uso del verbo ser y estar (43). En lo que al artículo respecta, los errores de mayor frecuencia corresponden al uso del artículo definido (120), específicamente en lo que compete a la omisión del artículo (93).

Tabla 11. Errores más frecuentes en el corpus de aprendientes de ELE.

Errores más frecuentes de todo el Corpus Textual					
Categoría	Tipo de error		Fr.error	Fr.cat	%
Ortografía acentual	OMI-Orto-aguda		104	317	26%
	INAPROP-aguda		59		
	OMI-Orto-esdrújula		98		
	INAPROP-esdrújula		6		
	INAPROP-grave		26		
	OMI-Orto-grave		24		
Preposición	FALSA-SELEC-prep		122	309	25%
	OMI-prep		120		
	ADI-prep		67		
Concordancia sintáctica	FALSA-SELEC-CG-Gen		208	289	23%
	FALSA-SELEC-CG-num		81		
Verbo	Tiempos verbales	FALSA-SELEC-modo Subjuntivo	34	180	15%
		FALSA-SELEC-T-pret-imperf	22		
		FALSA-SELEC-T-pret-perf-simple	8		
	Verbo estar	FALSA-SELEC-verbo estar	17		
	Verbo ser	FALSA-SELEC-verbo ser	21		
		ADI-verbo ser	5		
	Morfología del verbo	FALSA-SELEC-lex-morf-deriv-irreg	5		
		FALSA-SELEC-lex-morf-deriv-reg	9		
		FALSA-SELEC-lex-morf-radiv-verb-irreg	26		
	Formas no personales del verbo	FALSA-SELEC-gerundio	13		
		FALSA-SELEC-infinitivo	11		
		FALSA-SELEC-participio	9		
Artículos	Art. Def	OMI-art-def	93	131	11%
		ADI-art-def	16		
		FALSA-SELEC-art-def	11		
	Art.Indef	OMI-art-indef	6		
		ADI-art-indef	5		
Totales			1226	1226	100%

CONCLUSIONES

Este artículo ha centrado su atención en el Análisis de Errores asistido por el Computador (CEA) basado en Corpus de Aprendientes de Lenguas en Formato Electrónico (CLC). Esto con el propósito general de determinar de manera más precisa y real los errores de mayor frecuencia observados en un corpus de resúmenes de aprendientes de ELE de nivel de competencia B1. En este contexto, la importancia que tiene el uso del enfoque metodológico del CEA basado en el CLC es el de obtener de forma automática y rápida resultados sobre los diferentes tipos de errores y su frecuencia. Así como lo menciona Dagneaux et al. (1998) la utilización de este tipo de programas para la identificación, clasificación y descripción de los errores permite la recuperación automática de datos de grandes cantidades de información con el objeto de realizar un análisis focalizado en las tendencias de uso de la lengua en cuestión. Una vez construido el sistema de anotación y etiquetado del corpus, la asistencia de programas computacionales para la recuperación de información, permite diferentes tipos de obtención, exploración y explotación de los datos. Para ello, se puede utilizar análisis con aplicación de fórmulas de frecuencias a los resultados del corpus y extraer de estos detalles específicos del error etiquetado en contexto.

Esta evolución de los procedimientos metodológicos de la LC, del CLC implementados en el AE ha favorecido el estudio de los errores en el uso real de la lengua del aprendiente, reestableciendo así la importancia del AE como área de estudio en el ámbito de la lingüística aplicada (Dagneaux et al., 1998). Los estudios de AE anteriores eran limitados por el tipo y cantidad de recolección de datos, los cuales correspondían, muchas veces, al uso de la lengua del aprendiente de forma aislada y descontextualizada. En consecuencia, la inclusión de corpora en formato electrónico ratifica el análisis del error como una metodología adecuada para investigar el uso real de la L2 o LE. En este aspecto, las investigaciones realizadas en CEA han demostrado que dicho enfoque provee una manera de descubrir características importantes, en particular, en áreas donde se encuentra la mayor frecuencia de los errores (Dagneaux et al., 1998; Granger, 2003).

El propósito de este trabajo ha sido el de contribuir tanto en el ámbito del Análisis de Errores de escritura en ELE como en un área temática mayor en que se circunscribe esta investigación. Esto es, en el contexto de los Sistemas Tutoriales Inteligentes (STI) para Lenguas Extranjeras (LE) en lo que compete al diseño, la implementación y la evaluación de dichos sistemas. Los resultados del procesamiento del corpus de resúmenes que arroja este estudio sugieren que un STI para ELE en el nivel B1 debería reconocer y tratar los errores observados en: (1) la ortografía acentual, (2) el uso de la preposición, (3) en la concordancia sintáctica (4) en el uso de verbos y (5) en el artículo. Estos resultados permitirán apoyar el proceso de reconocimiento y tratamiento de errores mediante técnicas de procesamiento de lenguaje natural, las cuales posibilitan el reconocimiento de los errores (*parser*) y la entrega de una

ayuda focalizada y efectiva (*feedback* correctivo escrito). De esta forma, los estudiantes pueden resolver de manera autónoma sus errores de lengua y mejorar su precisión lingüística en la escritura.

Como trabajo futuro se planifica realizar estudios de tipo cualitativo que aporten con explicaciones de los errores aquí encontrados y sus relaciones con variables individuales como el nivel de lengua y la lengua materna de los sujetos. Sería interesante, además, indagar en los errores cometidos por los estudiantes en su interlengua considerando para ello, tanto su L1 como el dominio de otras lenguas (L2, L3). Al respecto, Alba Quiñones (2009a) señala que el análisis de errores permite detectar desde dónde se posicionan los aprendientes para realizar estas hipótesis, ya que todo dependerá de cuál es la L1 del aprendiente como también de las otras lenguas que domina (L2, L3).

Otras contribuciones importantes de este estudio en el área del Análisis de Errores Asistido por Computador es: (1) La delimitación de un Sistema de Anotación para el Tratamiento de Errores de ELE. Para ello, se consideraron los criterios de elaboración de notaciones para etiquetajes de corpus, propuestos por Granger (2004). Efectivamente, sobre la base de la literatura especializada, se construyó un sistema de anotación para el etiquetamiento de los errores de un corpus de ELE que ha resultado ser eficiente para la clasificación y etiquetamiento de los errores de ELE. (2) La colección de un corpus en formato electrónico de Español como Lengua Extranjera etiquetado que posibilitará estudiar y analizar muestras reales de ELE y sus problemáticas lingüísticas, enriqueciendo así el ámbito de la adquisición y enseñanza del español como lengua extranjera. Dicho corpus se incrementa semestralmente con la recolección de nuevos textos en formato electrónico de aprendientes de ELE del Programa Español como Lengua Extranjera de la Universidad de Concepción (<http://ele.udec.cl>).

Finalmente, queremos señalar a modo de proyecciones, que los hallazgos encontrados en el estadio actual de esta investigación en materia de errores lingüísticos, nos permitirán avanzar en la delimitación de las relaciones entre tipos de errores, niveles de lengua y *feedback* correctivo escrito. Esto permitirá depurar el modelo metodológico del tratamiento de errores que hemos implementado en contextos de enseñanza de la lengua mediatizada por la tecnología.

REFERENCIAS BIBLIOGRÁFICAS

- Alba Quiñones, V. (2009a). *La enseñanza del español en centros de secundaria alemanes: Análisis de errores semánticos*. [en línea]. Disponible en: http://www.mecd.gob.es/dctm/redele/Material-RedEle/Revista/2009_16/2009_redELE_16_03DeALba.pdf?documentId=0901e72b80dd738b
- Alba Quiñones, V. (2009b). El análisis de errores en el campo del Español como Lengua Extranjera: Algunas cuestiones metodológicas. *Revista Nebrija de Lingüística Aplicada*, 5(3), 1-16.
- Alexopoulou, A. (2005). El error: Un concepto clave en los estudios de adquisición de segundas lenguas. *Revista de Lingüística Teórica y Aplicada*, 43(1), 75-92.
- Alexopoulou, A. (2006). Los criterios descriptivo y etiológico en la clasificación de los errores del hablante no nativo: Una nueva perspectiva. *Porta Linguarum*, 5, 17-35.
- Corder, S. P. (1967). The significance of learner's errors. *International Review of Applied Linguistics in Language Teaching*, 5(1-4), 161-170.
- Dagneaux, E., Denness, S. & Granger, S. (1998). Computer-aided error analysis. *System*, 26(2), 163-174.
- Díaz-Negrillo, A. & Domínguez, J. F. (2006). Error tagging systems for learner corpora. *Revista Española de Lingüística Aplicada*, 19, 83-102.
- Ellis, R. (1997). *Second language acquisition*. Oxford: Oxford University Press.
- Ferreira, A. (2006). Estrategias efectivas de feedback positivo y correctivo en el español como lengua extranjera. *Revista Signos. Estudios de Lingüística*, 39(62), 379-406.
- Ferreira, A., Moore, J. D. & Mellish, C. (2007). A study of feedback strategies in foreign language classrooms and tutorials with implications for Intelligent Computer-Assisted Language Learning Systems. *International Journal of Artificial Intelligence in Education*, 17(4), 389-422.
- Ferreira, A. & Kotz, G. (2010). ELE-Tutor inteligente: Un analizador computacional para el tratamiento de errores gramaticales en Español como Lengua Extranjera. *Revista Signos. Estudios de Lingüística*, 43(73), 211-236.
- Ferreira, A., Salcedo, P., Kotz, G. & Barrientos, F. (2012). La Arquitectura de ELE-TUTOR: Un Sistema Tutorial Inteligente para el Español como Lengua Extranjera. *Revista Signos. Estudios de Lingüística*, 45(79), 102-131.
- Ferreira, A., Vine, A. & Elejalde, J. (2013). Hacia una prueba de nivel en español como lengua extranjera. *Revista de Lingüística Teórica y Aplicada*, 51(2), 73-103.

- Granger, S. (2002). A bird's-eye view of learner corpus research. En S. Granger, J. Hungand & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 3-33). Philadelphia: John Benjamins.
- Granger, S. (2003). Error-tagged learner corpora and CALL: A promising synergy. *CALICO Journal*, 20(3), 465-480.
- Granger, S. (2004). Computer learner corpus research: Current status and future prospects. *Language and Computers*, 52(1), 123-145.
- Granger, S. (2009). The contribution of learner corpora to second language acquisition and foreign language teaching. En K. Aijmer (Ed.), *Corpora and Language Teaching* (pp. 33-13). Philadelphia: John Benjamins.
- Kotz, G. & Ferreira, A. (2013). La precisión gramatical mediada por la tecnología: El análisis y tratamiento automático de errores. *Literatura y Lingüística*, 27, 219-242.
- Leech, G. (1992). Corpora and theories of linguistic performance. En J. Svartvik (Ed.), *Trends in Linguistics Studies and Monographs Directions in Corpus Linguistics* (pp.105-122). Nueva York: Mouton de Gruyter.
- Parodi, G. (2008). Lingüística de corpus: Una introducción al ámbito. *Revista de Lingüística Teórica y Aplicada*, 46(1), 93-119.
- Parodi, G. (2010). *Lingüística de Corpus: De la teoría a la empiria*. Frankfurt: Iberoamericana/Vervuert.
- Reppen, R. (2010). Building a corpus: What are the key considerations. En A. O'Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 31-38). Nueva York: Routledge.
- Tono, Y. (2003). Learner corpora: Design, development and applications. En D. Archer, P. Rayson, A. Wilson & T. McEnery (Eds.), *Proceedings of the Corpus Linguistics 2003 Conference* (pp. 800-809). Lancaster (UK): University Centre for Computer Corpus Research on Language.
- Vázquez, G. (1991). *Análisis de errores y aprendizaje de Español/Lengua Extranjera*. Frankfurt am Main: Peter Lang.
- Vázquez, G. (2009). Análisis de errores, el concepto de corrección y el desarrollo de la autonomía. *Revista Nebrija de Lingüística Aplicada a la Enseñanza de Lenguas*, 5, 113-122.

NOTA

- 1 El estudio de errores de escritura en Español como Lengua Extranjera que se presenta en este artículo se ha desarrollado en el contexto del proyecto de investigación FONDECYT 1140651 “El Feedback Correctivo Escrito Directo e Indirecto en la adquisición y aprendizaje del Español como Lengua Extranjera”.