Análisis de errores con implicaciones para un sistema tutorial inteligente de ELE[1]

Anita Ferreira Cabrera
Universidad de Concepción

Resumen

En este artículo se presenta un estudio sobre análisis de errores con el propósito de seleccionar los errores de lengua más frecuentes que serán tratados a través de estrategias de feedback correctivo escrito generadas por un Sistema Tutorial Inteligente (STI) para ELE. El corpus textual se compone de 85 resúmenes: 41 textos académicos, 22 narrativos y 22 argumentativos, producidos por 22 estudiantes extranjeros de nivel B1. En el enfoque metodológico, se presenta los procedimientos para la identificación, clasificación y procesamiento de los datos a través del software NVIVO. Los resultados sugieren que un STI para el Nivel B1 de ELE debería reconocer y tratar los errores gramaticales como las preposiciones, los verbos ser y estar, los artículos, la concordancia gramatical y la tildación de palabras agudas, graves y esdrújulas.

1. Introducción

El Análisis de los Errores (AE) ha evolucionado a lo largo de su historia (Santos Gargallo, 1992). De predecir y explicar tipologías de errores basadas en taxonomías gramaticales, se pasó a estudiar los efectos pragmáticos en el oyente y a valorar los errores en función de su efecto comunicativo. Este estudio corrobora la necesidad del error a lo largo del proceso del aprendizaje de una lengua extranjera (LE) porque este forma parte del mismo y, por lo tanto, es una evidencia de este proceso (Ellis, 1997). 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. Por tanto, 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 (Spada y Lightbown, 1999).

La investigación que aquí presentamos se basa en estudios previos sobre errores de lengua y estrategias de feedback correctivo en Español como Lengua Extranjera (ELE) (Ferreira 2003, 2006, 2007; Ferreira, Moore y Mellish, 2007) y más recientemente en la arquitectura e implementación de un Sistema Tutorial Inteligente (STI) para el tratamiento de errores de ELE (Ferreira y Kotz, 2010; Ferreira, Salcedo, Kotz y Barrien-

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 1110812 "Un Sistema Tutorial Inteligente para la focalización en la forma en la enseñanza del español como lengua extranjera".

tos, 2012; Kotz y Ferreira, 2013). En consecuencia, el objetivo es determinar los errores lingüísticos más frecuentes observados en la escritura de resúmenes de estudiantes de ELE nivel B1. El propósito final es seleccionar dichos errores para ser tratados a través de estrategias de feedback correctivo escrito en el contexto de un STI para el ELE que estamos construyendo.

El artículo se organiza de la siguiente forma: en la sección 2 nos referiremos a los principales fundamentos teóricos en materia de análisis de errores. En la sección 3 abordamos el estudio descriptivo sobre errores de producción escrita en ELE. En la sección 4 presentamos los resultados del estudio tendientes a delimitar los errores que serán tratados por el STI que está en construcción para ELE. Por último, presentamos algunos comentarios finales y proyecciones de este estudio.

2. Fundamentos teóricos

2.1. Tipologías de Errores

Hasta ahora no existe una tipología de errores universal, disponible para cualquier sistema lingüístico que se desee tratar. Esto se debe a que todas las lenguas son diferentes y los tipos de errores identificados pueden variar considerablemente. Además, los tipos de errores varían según la lengua nativa del alumno, la lengua meta, el nivel de competencia lingüística del alumno y el fenómeno gramatical que se desea tratar. Así pues, una tipología de errores nunca puede ser elaborada independientemente del objetivo para el cual ha sido concebida. El uso de una clasificación de errores tiene las siguientes ventajas: (1) La descripción de errores y sus respuestas pueden estar basadas tanto en errores individuales como en clases de errores. (2) Los errores que pertenecen a una misma clase pueden ser tratados de manera similar. (3) La clasificación de los errores permite a los sistemas basados en parser no sólo guardar los errores sino también calcular el número errores que pertenecen a una misma clase.

La anticipación de los errores es una tarea compleja si se tiene en cuenta las infinitas posibilidades de desviaciones de las normas gramaticales. Una apropiada clasificación de errores es un requisito importante para un adecuado reconocimiento de los errores por parte de un STI y la generación de feedback correctivo.

2.2. Análisis de Errores en Español como Lengua Extranjera

A partir de la década de los 90, se ha realizado una serie de estudios en ELE con el objeto de identificar y clasificar los errores acorde con el análisis de errores (Vázquez, 1992; Fernández, 1997; Santos Gargallo, 1992). Los estudios presentan diversas situaciones respecto a la lengua materna de los alumnos que han configurado las muestras (Alba Quiñones, 2009). Algunos de ellos centran su atención en muestras de estudiantes con una sola lengua materna, mientras que otros amplían este enfoque, considerando a aprendientes de distintas L1. Vázquez (1992) desarrolló su trabajo de AE con estudiantes de la misma L1, en su caso los aprendientes de ELE eran alemanes, del mismo modo que en el estudio de Alba Quiñones (2009), donde los participantes también eran

alemanes, y de Santos Gargallo (1992) en el que participaron estudiantes serbocroatas. Fernández (1997) realizó su trabajo con estudiantes de distintas L1, específicamente de alemán, francés, árabe y japonés. Ambas perspectivas aportan y enriquecen la investigación en el ámbito del AE.

3. Estudio descriptivo cuantitativo

El presente estudio es de tipo descriptivo cuantitativo se basa en el análisis de errores y en la lingüística de corpus. Se propone examinar los errores de escritura cometidos por estudiantes de ELE al desarrollar una tarea de producción de un texto resumen. El **objetivo general** es determinar cuáles son los errores de mayor frecuencia cometidos por los estudiantes de un nivel de competencia intermedio de español como lengua extranjera (B1) según el Marco Común Europeo de Referencia para las Lenguas.

3.2. Muestra

La muestra está constituida por 22 estudiantes universitarios extranjeros de la Universidad de Concepción. Los estudiantes que llegan a la universidad vienen por programas de intercambio estudiantil tanto a nivel de pre como postgrado. Ellos se inscriben en los diferentes cursos que la Universidad ofrece, entre ellos, el curso de ELE para nivel B1. Las edades de los alumnos fluctúan entre los 19 y 36 años de edad (23 años en promedio) y tienen como L1 las siguientes lenguas: el 50% (11) tenía como L1 el alemán; el 22,7% (5), el inglés, el 18,2 % el francés (4), y el 9% (2) el portugués. En cuanto a los niveles de proficiencia en español, los estudiantes presentan un nivel B1, dicho nivel se obtuvo a través de dos mediciones: se consideró (1) 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, desarrollada por el Programa de Español como Lengua Extranjera de la Universidad de Concepción, ELE-UdeC.

3.3. Metodología de la Investigación

la metodología utilizada para la delimitación de los errores en el corpus lingüístico se sustenta en el análisis de errores propuesto por Corder (1967) y más contemporáneamente por las propuestas de análisis de errores para ELE de Vázquez (1992); Fernández (1997); Santos Gargallo (1992); y Alba Quiñones (2009). De acuerdo con estas propuestas, se consideraron las siguientes etapas:

3.3.1. Recopilación del corpus

El corpus de textos está conformado por 85 resúmenes escritos: 41 textos académicos, 22 narrativos y 22 argumentativos. Para la elicitación de este corpus textual se tuvieron en cuenta los siguientes pasos:

La aplicación de un formulario de registro: este instrumento consistió de 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.

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. Se les pidió a los alumnos que enviaran a la profesora temas de interés académico.

Para escribir cada resumen, los estudiantes contaban con instrucciones generales que se explicaron al inicio de la sesión n°2. 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 formato txt, de este modo se evitó que los estudiantes utilizaran correctores ortográficos automáticos. Al finalizar, los resúmenes eran enviados a una cuenta de correo electrónico creada para este fin: espanolacademico@gmail.com. La tarea de escritura se realizó en el laboratorio de computación del Centro de Formación y Recursos Didácticos (CFRD) en cuatro momentos distintos.

3.3.2. Identificación y catalogación 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 catalogación e identificación de los errores encontrados. Para la catalogación de los errores se utilizaron como base los criterios Lingüístico y Etiológico (ver Tabla N°. 1).

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

Tabla N.° 1: Criterios de Clasificación de Errores

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"* (S11, académico), en cuyo 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"* (S16, 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"* (S12, 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).

3.3.3. Sistema de Notación 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 notación para el etiquetamiento de los errores del corpus textual. Para la clasificación de los errores fue necesario establecer etiquetas y los niveles y subniveles sobre los cuales se recolectaría la información necesaria para su posterior análisis. En la Tabla Nº.2 se explicitan los niveles y subniveles de clasificación.

Categorías	Subcategorías		Descripción
Palabra	Categorías gramaticales		Sustantivos, preposiciones, adjetivos, tiempos verbales, verbos, adverbios, artículos, pronombres, formas personales del verbo, verbo haber, verbo ser o estar y contracciones al y del.
	Léxico		Léxico creado por derivación, morfemas radical y derivativo, palabras innecesarias.
Oración	Concordancia		Sujeto y predicado.
Frase	Nominal y verbal		-Concordancia entre los elementos de la frase: número y género. -Concordancia en el predicado.
	Subyacentes e innecesarias		-Concordancia entre las frases que usan adjetivos y adverbios. -Adición de frases innecesarias.
	Ortografía	Acentual	Agudas, graves, esdrújulas y sobreesdrújulas.
		Literal	Correcto uso de las letras.
		Dierética	Uso de los diptongos, triptongos y los hiatos.
	Puntuación	Punto	Punto aparte, seguido, dos puntos.
		Coma	Uso de la coma en diferentes casos.
Párrafo y texto	Conectores textuales		Consecutivos, ilativos, argumentativos, adversativos, entre otros.

Tabla N.º 2: Niveles y subniveles de clasificación

3.3.4. Procesamiento del Corpus Lingüístico

Los textos fueron analizados acorde con diferentes categorías correspondientes a los niveles de análisis lingüístico: palabra, frase, oración párrafo y texto. La tarea de etiquetamiento de los 85 textos (41 textos académicos, 22 narrativos y 22 argumentativos) consistió en identificar los errores a través de la intervención de un evaluador humano, quien debía marcarlos en una versión digital del texto, empleando el *software* especializado NVivo 10, con el fin de facilitar el procesamiento estadístico posterior. Una vez que se detectaba un error, seleccionaban la palabra o grupos de palabras del texto en que el error estaba presente y lo marcaban con la etiqueta correspondiente al error de entre las que consideraba la taxonomía y que estaban previamente ingresadas al archivo de NVivo 10. Con el fin de facilitarle a cada evaluador la tarea de etiquetar los errores, se preparó un archivo de base de NVivo 10 para ser utilizado en el estudio. En éste, dentro del apartado

recursos de la aplicación, se ingresaron como *elementos internos* las carpetas en donde almacenar los recursos que se utilizarían, es decir, los archivos con los resúmenes. Las carpetas creadas fueron cuatro: "Académico", "Argumentativo" y "Narrativo". En esta etapa del trabajo, se empleó un árbol que tenía 434 nodos. Para efectos del etiquetamiento por parte de los evaluadores, sólo se utilizaron los nodos finales de dicho árbol.

4. Resultados del estudio

Como se señaló en la parte introductoria de este artículo, la investigación se inserta en un proyecto de investigación mayor (FONDECYT 1110812), cuyo propósito principal es desarrollar un STI que permita apoyar y mejorar la precisión lingüística de los aprendices de ELE en sus tareas de escritura. El tratamiento de los errores de escritura será apoyado través de un andamiaje de estrategias de feedback correctivo escrito generadas por un STI para ELE. Por esa razón, se requiere indagar de manera más específica en torno a los errores de gramática y ortografía, puesto que estos tipos de errores obtuvieron las mayores frecuencias en el corpus textual. La Tabla Nº. 3 presenta los resultados generales obtenidos en el corpus textual de los 85 resúmenes de nivel B1. 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 2184 errores etiquetados en toda la muestra, el 87.6% corresponden a errores lingüísticos y 9.5% a errores etiológicos. En los errores lingüísticos, los de mayor frecuencia observados son los errores gramaticales con un 59.1% (errores de categorías gramaticales y de coherencia textual), 17.1% errores de ortografía (acentual, literal y dierética) y 11.4% errores de léxico. En cuanto a los errores etiológicos 137 son interlinguales y 71 intralinguales. En consecuencia, acorde con estos resultados se puede señalar que la mayoría de errores observados corresponden a errores lingüísticos y entre estos los más representativos son los errores gramaticales.

En cuanto a los **errores gramaticales**, se puede constatar también en la Tabla Nº. 3 que los errores más frecuentes son aquellos correspondientes a las distintas clases de palabras del español, esto es, los **errores de categorías gramaticales** (sustantivos, verbos, artículos, preposiciones, etc.): 839 errores, (38.4%), seguidos por los **errores de coherencia textual**: 452 errores, (20.7%).

En lo concerniente a los **errores etiológicos interlinguales** (ver Tabla Nº. 3), los de mayor frecuencia fueron los errores de **uso de la L1,** 65 errores, (3%) y en lo referido a los **intralinguales,** los **errores de sobregeneralización**, 60 errores, (2.7%). Este número mucho menor de errores etiológicos comparados con los encontrados en lo lingüístico está acorde con el nivel de lengua que se está analizando, las problemáticas, entonces, más bien parecen estar en el plano de la precisión lingüística que en el del proceso de adquisición o interlengua. Otro aspecto no menor que también podría explicar esta situación está dada en el hecho de que el español que presentan los estudiantes lo han aprendido, en su mayor parte, en contextos interactivos de clases presenciales, se trata, entonces, de un proceso de enseñanza-aprendizaje de la lengua más que de un proceso de adquisición o inmersión total tenido en sus países orígenes, antes de integrarse en los cursos de nuestro Programa de Español como lengua Extranjera, ELE-UdeC (http://ele.udec.cl).

FRECUENCIA DE LOS ERRORES EN TODA LA MUESTRA										
CRITERIO	NIVEL	CATEGORÍA	SUBCATEGORÍA		Fr. Abs	Fr.abs cat	Fr. Relat	Fr. Abs cat	Fr. Relat	Fr. Relat
LINGÜÍSTICO	Palabra y oración	Gramática	Categorías gramaticales		839	839	38,4%	1291	59,1%	87,6%
			Coherencia textual	Concor. Sintáctica	304	452	20,7%			
				Puntuación	91					
				Conectores	57					
	Palabra	Léxico	Léxico inadecuado morfemas		138	249	11,4%	249	11,4%	
			Léxico por derivación		105					
			Léxico innecesario		6					
	Palabra	Ortografía	Acentual		278	374	17,1%	374	17,1%	
			Literal		55					
			Dierética		41					
ETIOLÓGICO	Palabra y oración	Interlingual	Uso L1		65	137	6,3%	208	9,5%	9,5%
			Orden de palabras		38					
			Analogía semántica		9					
			Traducción literal		25					
	Palabra y oración	Intralingual	Sobregeneralización		60	71	3,3%			
			Hipercorrección		5					
			Simplificación		6					
Otros	Errores de tipeo				62	62	2,8%	62	3%	2,8%
Totales					2184	2184	100%	2184	100%	100%

Tabla No. 3: Resultados generales de toda la muestra

Si nos centramos sólo en los **errores gramaticales** en lo concerniente a las **categorías gramaticales**, en el Gráfico N°1 se ilustra que los errores de mayor frecuencia corresponden a las **preposiciones** con un 39%, a los **verbos** con un 26% y a los **artículos** con un 18%. Esto quiere decir que los estudiantes cometieron un mayor número de errores en sus textos cuando usaron las preposiciones en español, los verbos y artículos.

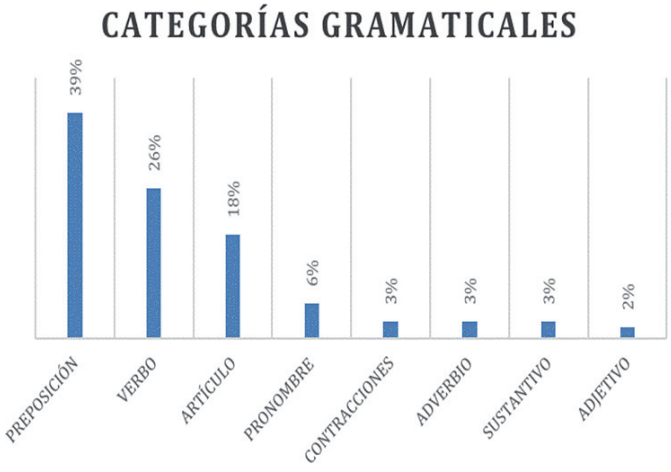

Gráfico N° 1: Frecuencia de Errores en Categorías Gramaticales

En el Gráfico N°. 1 se muestra las frecuencias totales por cada uno de los grupos de errores etiquetados en todo el corpus textual. Se aprecia con claridad que los problemas de precisión lingüística que mayormente se observan en los textos producidos por los estudiantes de Nivel B1 de la muestra están a nivel de la **gramática** (59%) y de la **ortografía (17%)**.

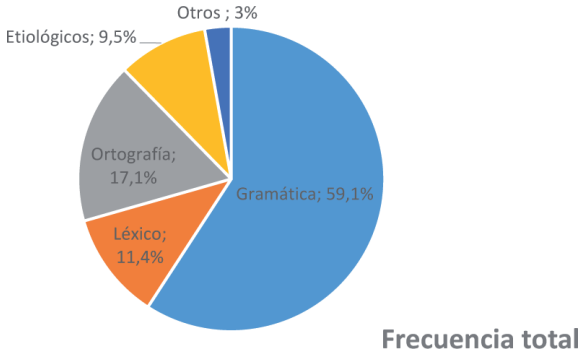

Gráfico No. 2: Resultados Generales de toda la Muestra

En lo que respecta a la subcategoría de errores gramaticales, denominados **errores de coherencia textual**, la Tabla Nº. 4 evidencia los errores de mayor frecuencia con respecto a la **concordancia sintáctica entre sujeto y predicado** (276 errores), de estos 204 son de **falsa selección de género gramatical** y 72 errores de **número gramatical**.

Subcategoría	Tipo error	Fr abs	Fr %
Concordancia sintáctica	FALSA-SELEC-CG-Gen	204	67%
	FALSA-SELEC-CG-num	72	24%
Concordancia entre sintagmas	FALSA-SELEC-CG-suj-pred	22	7%
	OMI-sintag-nominal	6	2%
Totales		304	100%

Tabla No. 4: Frecuencia de Errores en la Concordancia Sintáctica

En lo relativo a los errores de **ortografía acentual**, se puede apreciar en el Gráfico Nº. 2 que las **palabras agudas (42%)** son las que presentaron mayores complicaciones en cuanto a la tildación, seguidas por las **palabras esdrújulas (37%)** y las **graves (19%).**

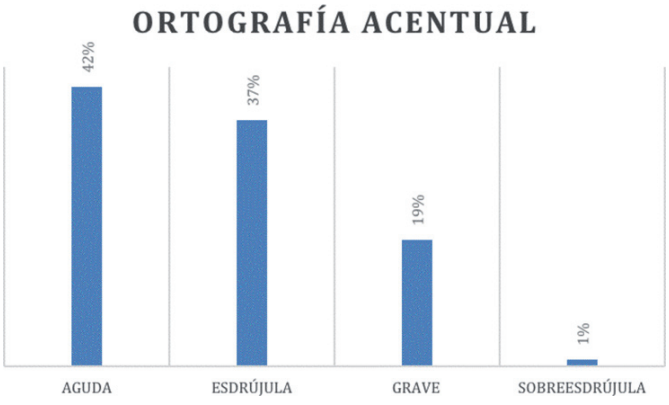

Gráfico No. 3: Frecuencia de errores en Ortografía Acentual

Sobre la base del análisis y los resultados hasta aquí ilustrados y con el propósito de delimitar un número de errores que permita avanzar en el diseño del módulo del tratamiento de errores del STI ELE-TUTORA, se seleccionó los 5 tipos de errores con mayor frecuencia de toda la muestra para constituirse en los candidatos para el diseño, implementación y procesamiento de los errores a través del STI ELE-TUTORA: preposición, ortografía, concordancia gramatical, verbo y artículo.

Comentarios finales

Este artículo ha centrado su atención en el análisis de errores para el tratamiento de errores gramaticales a través de un analizador automático de un sistema tutorial inteligente que funciona en el contexto de una plataforma de enseñanza/aprendizaje de español como lengua extranjera. La fundamentación teórica y empírica del estudio realizado han evidenciado que la tendencia en el nivel de proficiencia para corregir los errores gramaticales debería privilegiar a las preposiciones, la ortografía acentual en lo que se refiere a las palabras agudas, esdrújulas y graves, la concordancia gramatical, el verbo y los artículos. Esta propuesta debería conducir al mejoramiento de la precisión gramatical de las formas lingüísticas seleccionadas para ser procesadas por el tutor inteligente. Como trabajo futuro se planifica efectuar estudios acerca de la percepción de los alumnos ante el reconocimiento de este tipo de errores y el *feedback* correctivo otorgado por el sistema.

Finalmente, queremos señalar que el propósito de este trabajo es el de contribuir a la investigación en el área de los sistemas tutoriales inteligentes para lenguas extranjeras y, de esta manera, responder al propósito final de apoyar la enseñanza/aprendizaje del español mediante sistemas computacionales dotados de técnicas de procesamiento de lenguaje natural que sean capaces de entregar ayuda adecuada y efectiva a fin de que los estudiantes puedan resolver de manera autónoma sus errores de lengua y mejorar así su precisión lingüística.

Bibliografía

ALBA, Virginia (2009): «El análisis de errores en el campo del español como lengua extranjera: algunas cuestiones metodológicas», *Revista de Nebrija de Lingüística Aplicada*, 5 (3), 1-16.

CORDER, Pit (1967): «The significance of learner's errors», *IRAL*, 5, 161-170.

ELLIS, Rod (1997): *Second language acquisition*, vol. 2, Oxford: University Press.

FERNÁNDEZ, Sonsoles (1997): *Interlengua y Análisis de Errores en el aprendizaje del español como lengua extranjera*, Madrid: Edelsa.

FERREIRA, Anita (2003): «Feedback strategies for second language teaching with implications for Intelligent Tutorial Systems», Ph.D.Thesis, University of Edinburgh, Edinburgh, UK.

FERREIRA, Anita (2006): «Estrategias de feedback positivo y correctivo en el español como lengua extranjera», *Revista Signos*, 39 (62), 379-406.

FERREIRA, Anita (2007): «Estrategias efectivas de feedback correctivo para el aprendizaje de lenguas asistido por computadores», *Revista Signos*, 40 (65), 521-544.
FERREIRA, Anita, Johanna MOORE y Chris MELLISH (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, Anita y Gabriela KOTZ (2010): «ELE-Tutor Inteligente: Un analizador computacional para el tratamiento de errores gramaticales en Español como Lengua Extranjera», *Revista Signos*, 43 (73), 211-236.
FERREIRA, Anita, SALCEDO, Pedro, KOTZ, Gabriela y Fernanda BARRIENTOS (2012): «La arquitectura de ELE-TUTOR: Un sistema tutorial inteligente», *Revista Signos,* 45 (79), 102-131.
KOTZ, Gabriela y FERREIRA, Anita (2013): «La precisión gramatical mediada por la tecnología: El análisis y tratamiento automático de errores», *Revista Literatura y Lingüística*, 27, 217-240.
SANTOS GARGALLO, Isabel (1992): *La enseñanza de segundas lenguas. Análisis de errores en la expresión escrita de estudiantes de español cuya lengua nativa es el serbo-croata*, Madrid: Universidad Complutense.
SPADA, Nina y Patsy LIGHTBOWN (1999): «Instruction, first language influence, and developmental readiness in second language acquisition», *The Modern Language Journal*, 83 (1), 1-22.
VÁZQUEZ, Graciela (1992): *Análisis de errores y aprendizaje de español/lengua extranjera*, Frankfurt: Peter Lang.