



Literatura y Lingüística

ISSN: 0716-5811

literaturalinguistica@ucsh.cl

Universidad Católica Silva Henríquez

Chile

Kotz Grabole, Gabriela; Ferreira Cabrera, Anita

La precisión gramatical mediada por la tecnología: el análisis y tratamiento automático de errores

Literatura y Lingüística, núm. 27, 2013, pp. 219-242

Universidad Católica Silva Henríquez

Santiago, Chile

Disponible en: <http://www.redalyc.org/articulo.oa?id=35228136012>

- 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

La precisión gramatical mediada por la tecnología: el análisis y tratamiento automático de errores*

Gabriela Kotz Grabole**
Anita Ferreira Cabrera***

Resumen

En este artículo de carácter descriptivo se plantea como tema el reconocimiento y el tratamiento de errores gramaticales a través de un analizador automático o *parser* que tiene por objetivo principal el mejoramiento de la precisión lingüística de las formas gramaticales en aprendientes de español como lengua extranjera. Precisamente, se describe la manera en que se detectan y clasifican automáticamente los errores y la forma en que estos son tratados a través de la retroalimentación correctiva.

Palabras clave: español como lengua extranjera, Sistemas Tutoriales Inteligentes, error, feedback correctivo, parser.

Grammatical accuracy mediated by technology: analysis and automatic processing of errors

Abstract

This article describes grammatical error recognition and treatment through a parser, whose main objective is to improve the linguistic accuracy of Spanish grammatical forms in learners of Spanish as a foreign language. The article describes the automatic detection and classification of errors, and the treatment through corrective feedback.

Keywords: Spanish as a foreign language, Intelligent Tutorial Systems, error, corrective feedback, parser.

Recibido: 30-07-2012

Aceptado: 27-12-2012

* La investigación que se presenta en este artículo se circunscribe en el contexto del Proyecto FONDECYT 1110812, *Un sistema tutorial inteligente para la focalización en la forma en la enseñanza del español como lengua extranjera*, CONICYT-CHILE.

** Autora. Magíster en Lingüística. Doctora (c) en Lingüística, Universidad de Concepción, Concepción, Chile. gkotz@udec.cl

*** Coautora. Doctora en Lingüística, Doctora en Inteligencia Artificial, Universidad de Concepción, Concepción, Chile. aferreir@udec.cl

1. Introducción

Las teorías sobre la adquisición de lenguas han ido evolucionando con el tiempo, así como también el concepto de error. La actitud del docente hacia los errores lingüísticos de sus alumnos ha ido cambiando y depende de cómo este conciba la adquisición de una lengua. No solo la actitud del profesor ha cambiado, sino también la actitud del aprendiente respecto de su propio aprendizaje y de sus errores.

A pesar de la existencia de distintas ideas acerca de la corrección de los errores de lengua permanece indiscutible el argumento que para que su tratamiento sea efectivo y lleve a un mejoramiento sustancial en la precisión lingüística se requiere de atención experta y personalizada.

Uno de los mayores problemas en todo sistema educativo es la dificultad para atender de manera individual a un gran número de alumnos; sobre todo, si presentan problemáticas diferentes. En consecuencia, y con el objetivo de apoyar el proceso de aprendizaje, se apuesta por un tratamiento automatizado de los errores a través de un sistema que sea capaz de analizar de manera automática el enunciado del estudiante y entregar una retroalimentación efectiva, focalizada y flexible.

El presente trabajo forma parte de un proyecto mayor que consiste en la construcción de un Sistema Tutorial Inteligente (STI) para la enseñanza / aprendizaje del español como lengua extranjera (ELE) focalizado en la forma que tiene sus fundamentos en estudios y proyectos anteriores en los que, entre otros logros, se destaca la definición de un modelo para el tratamiento de los errores y un componente de estrategias de *feedback* (Proyecto Fondecyt 1080165¹, Ferreira, 2003, 2006, 2007; Ferreira, Moore & Mellish, 2007), y que contempla, a su vez, la construcción de un analizador sintáctico computacional (*parser*) para la detección de errores gramaticales (Ferreira & Kotz, 2010; Ferreira, Salcedo, Kotz & Barrientos, 2012).

Este artículo se plantea como objetivo principal la descripción del reconocimiento y tratamiento de errores gramaticales a través de un analizador sintáctico de oraciones en el contexto de un Sistema Tutorial Inteligente para la enseñanza del español como lengua extranjera.

1 Proyecto FONDECYT 1080165. Modelo Blended Learning basado en el enfoque por tareas y aprendizaje cooperativo para la enseñanza del Español como Lengua Extranjera. CONICYT-CHILE.

2. Marco teórico

2.1. El concepto del error según las teorías de adquisición de lenguas

Las teorías de adquisición de segundas lenguas (ASL) han cambiado con el tiempo; así como las primeras décadas del siglo XX estuvieron marcadas por el conductismo que entendía el aprendizaje de una lengua extranjera como la adquisición de una serie de hábitos donde los errores debían evitarse a toda costa; a partir de los años 70 nuevos estudios demuestran que muchos de los errores que cometen los alumnos provienen de problemas en la aplicación de las reglas de la lengua meta y, de esta manera, nace el concepto de *interlengua* (Selinker, 1972), término que adquiere mayor difusión con Corder (1974, 1981), quien fundó las bases del modelo de Análisis de Errores (del inglés, *Error Analysis*, EA).

Este modelo defiende 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).

Esta teoría establece, además, una distinción entre errores (*errors*) y faltas (*mistakes*): Los errores se cometen por falta de competencia lingüística y son propios de los hablantes no nativos o estudiantes de una lengua extranjera; mientras que las faltas son producto del cansancio, del estrés o de la falta de concentración y son comunes en los hablantes nativos.

Si bien la teoría del Análisis de Errores (AE) fue muy criticada, principalmente por concentrarse solo en los fenómenos negativos de los procesos de adquisición lingüística del alumno, presenta importantes méritos debido a su concentración en la superficie observable de los fenómenos, en el riguroso análisis de los textos producidos por los aprendientes y en la importancia que se le da a los errores de los alumnos no nativos para la comprensión de los procesos de aprendizaje de las lenguas extranjeras (Richards, 1984).

Aunque esta teoría tiene más de 40 años, su modelo ha ido evolucionando y adaptándose a las diferentes corrientes de ASL, y su metodología de tres fases; reconocimiento, descripción y explicación de los errores, sigue vigente y ha dado lugar a la agrupación de los errores en diversas taxonomías y clasificaciones (Richards, 1984; Santos

Gargallo, 1993; James, 1998; Vásquez, 1999, 2000; Alexopoulou, 2005; Alba Quiñones, 2009).

La existencia de una adecuada taxonomía de errores cobra especial importancia en el diseño e implementación de plataformas de enseñanza / aprendizaje de lenguas con tecnología y *ICALL*, sobre todo, en lo que se refiere a la programación de un sistema para la detección y procesamiento automático de errores (Granger, 2003; Heift & Schulze, 2007).

2.2. El enfoque en la forma

En general, hay un consenso en que la corrección de la forma es una parte importante en la adquisición de las lenguas; sin embargo, no está claro cómo se puede llevar a cabo este objetivo. Hay diferentes interpretaciones de cómo se deben presentar las formas lingüísticas en la práctica de la enseñanza en el aula y una gran disparidad de la manera en que se entiende la atención a la forma.

En este sentido, Long (1991) aclara que la *atención a la forma* no implica una *atención a las formas* lingüísticas individuales y establece la diferencia entre la atención a *las formas* característica de los enfoques anteriores de la nueva atención a *la forma*, que consiste en un cambio ocasional de la atención a las características del código lingüístico provocado por un problema puntual en la comunicación.

Siguiendo esta tendencia, las investigadoras Doughty y Varela (2009) proponen una *atención comunicativa a la forma*, es decir, un enfoque dual que se centra en la forma dentro de la interacción comunicativa, es decir, una atención a la forma relativamente implícita centrada en el contenido.

2.3. El *feedback* correctivo

La retroalimentación o *feedback* se da de manera natural durante la interacción en el contexto del aprendizaje de lenguas, sobre todo cuando el aprendiente nota que no puede expresarse correctamente en la lengua que está aprendiendo (Swain, 1998).

El término *feedback* es aplicable a cualquier tipo de retroalimentación dada al alumno, por el docente o por otro alumno, ante una respuesta, ya sea correcta o incorrecta (Good & Brophy, 2000).

La respuesta a un enunciado correcto se denomina *feedback* positivo; es decir, se afirma que la intervención del estudiante es adecuada. Este

se puede dar en forma de reconocimiento, aceptación, repetición o parafraseo de la respuesta correcta (Ferreira et al., 2007).

El *feedback* correctivo, en cambio, es definido como una indicación para el alumno de que la estructura expresada no es parte de la lengua meta (Carroll & Swain, 1993) y puede manifestarse de manera explícita o implícita. Se entiende por *feedback* explícito aquel que tiende a proveer la forma correcta e incluye correcciones y explicaciones metalingüísticas (Ellis, Loewen & Erlam, 2006). El *feedback* implícito, en cambio, apunta a estrategias de negociación y es preferido por las corrientes interaccionistas de ASL (Gass & Mackey, 2007).

Siguiendo las ideas interaccionistas y de la negociación de la forma, Ferreira (et al.) (2007) proponen una clasificación de estrategias de *feedback* correctivo en dos grupos: El Grupo 1 (*Giving-Answer Strategies, GAS*) corresponde a estrategias en las que el profesor entrega de manera directa la forma correcta o esperada. Estas incluyen la corrección explícita del error, la repetición del error o parte del enunciado erróneo y la reformulación. El Grupo 2 (*Prompting-Answer Strategies, PAS*), por el contrario, está constituido por estrategias en las que el profesor no entrega la respuesta correcta, sino que incentiva al estudiante a corregir el error por sí mismo; a este grupo corresponden las siguientes estrategias: claves metalingüísticas, elicitaciones y solicitudes de clarificación.

Si bien la investigación en ASL muestra una amplia gama de taxonomías de estrategias de *feedback* usada por los profesores para el tratamiento de los errores de los alumnos (Chaudron, 1977; Lyster & Ranta, 1997; Lyster, 1998; Ferreira, 2006), los estudios acerca de los efectos de los distintos tipos de *feedback* son relativamente recientes (Ferreira, 2003, 2006, 2007; Lyster & Mori, 2006; Hyland & Hyland, 2006; Ellis, 2011; Sheen, 2011).

En cuanto a la interrogante sobre cuáles son las técnicas de *feedback* correctivo más efectivas, estudios indican que las estrategias que tienden a la autorreparación del error, tales como la solicitud de clarificación, las claves metalingüísticas y las elicitaciones; así como las correcciones indirectas han dado mejores resultados que las técnicas que entregan la respuesta correcta. También se ha demostrado que el *feedback* focalizado (por ejemplo, en uno o en pocos tópicos gramaticales) tiene mayor efectividad que el no focalizado. (Ferreira, 2006, 2007; Sheen, 2007, 2011; Bitchener & Knoch, 2008; Ellis, Sheen, Murakami & Takashima, 2008).

Sin embargo, hay una serie de variables que hay que tener en cuenta para el análisis de efectividad del *feedback*, por ejemplo, el nivel de competencia en la lengua extranjera (Lyster & Ranta, 1997; Panova & Lyster, 2002; Ammar & Spada, 2006; Ferreira, 2007).

La efectividad del tipo de *feedback* no solo depende del nivel de proficiencia en la L2. En efecto, hay una serie de factores que afectan el *feedback* correctivo, entre los que se puede mencionar las características de las lenguas que se corrigen, las condiciones en que se entrega la corrección, las diferencias individuales de los alumnos, factores motivacionales y estilos de aprendizaje (Ferreira, 2006; Heift & Schulze, 2007; Sheen, 2011; Ferreira et. al., 2012).

Con todo, el estudio del *feedback* se ha tornado crucial con el desarrollo de las nuevas metodologías de enseñanza / aprendizaje de lenguas, con los enfoques modernos basados en tareas y centrados en el proceso, así como también en los nuevos modelos de aprendizaje *be-learning* y a distancia de los últimos 20 años (Hyland & Hyland, 2006; Morales & Ferreira, 2008).

2.4. Los parsers para el tratamiento automático de los errores en la enseñanza de lenguas extranjeras

El *parsing* es un proceso por medio del cual se convierte el texto de entrada en otras estructuras (generalmente árboles sintácticos) para capturar su jerarquía implícita y, de esta manera, permitir el análisis sintáctico de la frase u oración (Lavid, 2005; Jurafsky & Martin, 2009).

Para que la entrada pueda ser procesada sintácticamente debe pasar primero por varias etapas de pre-procesamiento. Dentro de estas, la más importante es la de etiquetado de partes del habla o *PoS-tagging*.

Las técnicas de etiquetado se pueden agrupar básicamente en tres tipos o métodos: los basados en reglas, los estadísticos o probabilísticos y los híbridos basados en transformaciones.

Los etiquetadores basados en reglas utilizan conocimiento lingüístico (*knowledge-driven taggers*) expresado en forma de reglas muy sofisticadas que se escriben manualmente.

Los etiquetadores estadísticos (*data-driven taggers*), por el contrario, se caracterizan por calcular probabilidad de una etiqueta para una

palabra o *token* en base a sus partes o *n-gramas*. Entre ellos se encuentran los modelos de *bi-gramas* o *tri-gramas*, en los que la probabilidad de una etiqueta se estima por el contexto de las etiquetas anteriores. Para ello, es necesario entrenar el sistema con un corpus etiquetado previamente.

Los sistemas híbridos combinan métodos estadísticos y basados en reglas para intentar recoger los aspectos positivos de cada uno de ellos (Siddiqui & Tiwary, 2008; Jurafsky & Martin, 2009).

Una vez preparado el texto se realiza el proceso de análisis sintáctico a través de un *parsing*.

Hay muchas técnicas de *parsing*. Sea cual fuese la estrategia utilizada, la eficacia de un parser está dada por la capacidad de almacenar resultados parciales, es decir, por ser capaz de guardar los análisis correctos para no tener que repetirlos. Este es el caso de los *parser* con tablas de almacenamiento (o *chart-parsing*). En ellas, el parser puede mantener un registro de los constituyentes reconocidos en una posición determinada del análisis y de las reglas que se han aplicado parcialmente con éxito, aunque no estén completas (Lavid, 2005).

Con el fin de ahorrar costos de procesamiento se puede realizar, además, un análisis superficial o fragmental en lugar de un análisis en profundidad, como el que efectúan los denominados agrupadores sintácticos o *chunkers*, que buscan detectar frases nominales, verbales, adjetivas, adverbiales básicas más útiles y productivas para el análisis.

El *parser* por sí solo tiene como única función analizar sintácticamente una oración correctamente formulada; sin embargo, para que el sistema funcione en un contexto de un sistema de enseñanza de lenguas debe analizar también las oraciones que contienen errores.

Existen varias técnicas para que el *parser* analice oraciones mal formadas. Una de las técnicas es la construcción, junto con la gramática de la lengua meta, de un set de reglas agramaticales (o *mal rules*) que contienen errores predictibles de una determinada intelengua. (Harbusch & Kempen, 2010; Amaral & Meurers, 2011).

Dado que es imposible predecir todos los errores que puede cometer un alumno en un momento dado de su aprendizaje, se puede construir un sistema de reglas, similar al anterior, denominado *metarreglas* o *buggy rules* (del inglés, reglas de relajación), que se enfoca en la no

consideración de ciertas restricciones gramaticales en determinados contextos definidos por el sistema (Dodigovic, 2005; Heift & Schulze, 2007; Amaral & Meurers, 2011).

Los errores especificados, es decir, las cadenas de etiquetas con errores que reconocerá el sistema, son asociados a un mensaje de error o *feedback*, que puede provenir de una base de datos o ser generado a través de un generador de *feedback*.

Durante el procesamiento, la entrada del usuario es analizada por el parser y comparada con las gramáticas del sistema. Si ésta coincide con una cadena de etiquetas de la gramática correcta, se entrega una retroalimentación positiva; por el contrario, se dará un *feedback* correctivo, si la entrada se relaciona con una secuencia de la gramática de errores.

3. Diseño e implementación del analizador sintáctico ELE-Tutor²

Si bien una de las fortalezas de los sistemas de *ICALL* es la posibilidad de trabajar los aspectos gramaticales de la lengua, se corre el riesgo de caer en un enfoque demasiado conductista de la enseñanza. Por ello, para no perder el equilibrio entre la forma y el contenido, se adoptó un enfoque comunicativo en la forma que combina la precisión y la fluidez en la comunicación.

Debido a que el *parser* que se describe en este trabajo forma parte de un sistema tutorial para la enseñanza de español como lengua extranjera y, en consecuencia, debe procesar entradas erróneas, un elemento importante es el diagnóstico y el tratamiento de los errores. Este debe tener especificaciones muy precisas, como asimismo ser capaz de entregar una respuesta flexible en el reporte de los mismos. Para cumplir con este objetivo fue necesario construir una taxonomía de errores definida por el nivel de competencia en la lengua extranjera que se desea apoyar.

3.1. Construcción de una taxonomía de errores

El primer paso del proyecto *parser* ELE-TUTOR fue la toma de decisiones acerca del concepto y definición de error que se va a manejar,

2 ELE-Tutor forma parte de un STI en construcción en el contexto de los proyectos FONDECYT 1080165 y 1110812. CONICYT-CHILE.

qué tipo de errores se van a tratar y de qué manera se va a llevar a cabo su tratamiento.

En consecuencia, se definió como error aquellas desviaciones de las reglas que se espera que el alumno cometa en determinada fase de su aprendizaje con mayor seguridad y sistematicidad, es decir, errores de interlengua según el concepto de la teoría del AE.

Debido a que este parser se focaliza en formas gramaticales complejas del ELE del nivel B2 de acuerdo al Marco Común Europeo de Referencia para las lenguas o MCER (Consejo de Europa, 2002), el tipo de errores que trata, son errores de gramática.

Una vez definido el contexto se construyó una taxonomía de errores utilizando un criterio deductivo, descriptivo y lingüístico; para ello, se consideraron errores de gramática cometidos en clases de español como lengua extranjera (nivel B2), se los clasificó y se hizo un análisis de su frecuencia. Luego, se tuvo que ajustar esa taxonomía de errores al contexto del tutor.

Para que esta taxonomía pueda ser operacionalizada en un *parser*, se consideró positivo y apropiado utilizar, además, como base la taxonomía de Heift y Schulze (2007) (Tabla 1) empleada en su tutor basado en parsing construido para la enseñanza del alemán. Este sistema consiste básicamente en la combinación de una clasificación lingüística y una clasificación en base de las características superficiales de los errores. La taxonomía propuesta consta de dos ejes: un eje que contiene los errores llamados de *estrategia de superficie* y otro con los errores en las clases gramaticales o *partes del habla*.

Partes del habla	Estrategias de superficie
determinante	selección errada, adición, omisión
preposición	selección errada, adición, omisión
verbo	selección errada, adición, omisión
verbo	error en la forma, en la conjugación, modo y tiempo
adverbio	selección errada
pronombre	selección errada, adición, omisión
-----	orden
concordancia	---
error en la categorización	selección errada

Tabla 1: Taxonomía de errores basada en Heift y Schulze (2007)

Una vez que se elaboró la taxonomía de errores se procedió al relevamiento de los mismos desde una muestra real, un curso de gramática para el español como lengua extranjera de nivel intermedio. Los errores elicitados fueron clasificados según la taxonomía, como muestra la tabla 2.

Parte del habla (PoS)	Estrategia de superficie	Nro. de errores	Porcentaje
determinante	selección errada	3	2%
determinante	adición	6	4%
determinante	omisión	12	8%
preposición	selección errada	9	6%
preposición	adición	0	0%
preposición	omisión	16	10,67%
verbo	selección errada	1	0,67%
Verbo (conjugación)	malformación	5	3,33%
Verbo (modo)	selección errada	2	1,33%
verbo (tiempo verbal)	selección errada	24	16%
verbo	adición	4	2,67%
verbo	omisión	5	3,33%
adverbio	selección errada	2	1,33%
adverbio	omisión	1	0,67%
pronombre	selección errada	1	0,67%
pronombre	adición	2	1,33%
pronombre	omisión	14	9,33%
error en orden	orden	12	8%
Errores de concordancia	---	26	17,34%
error en la categorización	selección errada	5	3,33%
TOTALES		150	100%

Tabla 2: Errores elicitados para el nivel intermedio

Como se puede observar en la tabla 2, la mayor parte de los errores cometidos correspondió a errores de concordancia, pero hay que tener en cuenta que este tipo de error se debe a concordancias diferentes, tanto a nivel del sintagma (ejemplo: determinante-artículo o nombre-adjetivo) como al de oración, como es el caso de la concordancia entre el sujeto y el verbo. Además se aprecia un alto porcentaje de errores en la selección

del tiempo verbal (16%). Si se observa el porcentaje total de errores en verbos, este asciende a 27,33%. También se muestran como relevantes los problemas en preposiciones (16,67% en total), en los determinantes (16%) y en los pronombres (11,33%).

3.2. Jerarquización de los errores

Cuando se diseña una taxonomía de errores se debe tener en consideración el propósito para el cual ha sido creada. Para ello, hay varios aspectos que se deben analizar, como los tipos de ejercicios, los temas gramaticales que se tratarán y el nivel de competencia de la lengua extranjera por parte del usuario.

Otro aspecto importante es el tratamiento de los errores múltiples. El analizador detecta todos los errores, pero para no desmotivar al alumno se debe tratar un solo error a la vez; para ello se estableció una jerarquía de errores. Se tomaron en cuenta dos criterios para la jerarquización: el de la gravedad y el de la frecuencia.

Se consideró como el error más grave, es decir, el primero que se debe tratar, cuando el alumno no responde a las expectativas del ejercicio; por ejemplo, que no responda a las instrucciones, aun cuando la frase esté bien construida.

Luego se los jerarquizó por orden de importancia según la frecuencia obtenida durante la elicitación de los errores; los errores más frecuentes deben tener prioridad en el tratamiento, ejemplo, errores en la concordancia y en los verbos.

La tabla 3 muestra la jerarquización de errores que fue tomada como base tanto para la construcción del parser como para el módulo de *feedback*.

Nivel de jerarquía	Tipo de error
1	El alumno no cumple con las expectativas del ejercicio
2	Errores de concordancia
3	Errores en los verbos
4	Omisiones y adiciones
5	Errores en el orden

Tabla 3: Jerarquía de errores para su tratamiento

3.3. Reconocimiento de errores a través del parser

3.3.1. Etiquetado

Cuando el alumno-usuario ingresa una respuesta a un ejercicio presentado por el ELE-TUTOR se efectúa el etiquetado de la entrada mediante la técnica de *n-gramas*. Para este parser en particular se utiliza un sistema de etiquetado estadístico de *tri-gramas* que consiste en la revisión del contexto de dos etiquetas precedentes. La oración etiquetada es comparada con la base de datos etiquetada manualmente y utilizada para el análisis sintáctico y morfológico.

3.3.2. Análisis Morfológico

Para el análisis morfológico se utiliza la técnica de *chunking*. Mediante esta técnica se detectan y agrupan determinadas identidades o secuencias específicas de texto, y de esta forma, se pesquisan errores morfológicos dentro de una oración mediante reglas similares a las *buggy rules* (del inglés, reglas de relajación), pero que, en este caso, se centra en la búsqueda de secuencias de etiquetas específicas dentro de una oración que representa un error morfológico, como por ejemplo: una oración con error de tiempo o número. Los errores morfológicos son informados al sistema mediante códigos de errores.

3.3.3. Análisis sintáctico

El análisis sintáctico está basado en una gramática de contexto libre o de estructura de frase ampliamente utilizada en el procesamiento del lenguaje natural y en la lingüística computacional, ya que permite reconocer o generar una gran variedad de oraciones. La desventaja de este tipo de gramáticas es que pueden reconocer o generar oraciones agramaticales y, para evitar este problema, es necesario "imponer restricciones al número de reglas y de categorías de la gramática" (Lavid, 2005, 107).

Para el análisis se utiliza la técnica de *chart-parsing* o tablas de almacenamiento por las ventajas que presenta la utilización de la programación dinámica, que es más eficiente ante problemas de ambigüedad y aumenta la velocidad de procesamiento.

Con el fin de que el *parser* reconozca frases con errores y pueda dar un informe que luego generará un *feedback* al usuario-alumno, se procedió

de la misma manera que en el análisis morfológico, con una gramática de errores.

Los códigos de errores, tanto para el análisis sintáctico como para el morfológico, fueron programados a través de *Python*, un lenguaje de programación de código abierto utilizado con frecuencia para el procesamiento de lenguaje natural.

En la Tabla 4 se muestra, a modo de ejemplo, algunos códigos de errores generados automáticamente por la unión de cadenas de texto. Por ejemplo: VN%+[‘numero esperado (S,P)’]. Si el ejercicio requiere que se conteste en singular el código será "VN%S". Los valores que varían se colocan entre corchetes.

Jerarquía	Código generado	Reporte de error
1	VT%+[‘tiempo esperado (FUT,IMP,etc)’]	Tienes un error en el verbo de tu oración. Este debe ir en [tiempo esperado (FUT,IMP,etc)].
2	VN%+[‘numero esperado (S,P)’]	El verbo debe estar en [singular / plural].
2	VP%+[‘persona esperada (1,2,3)’]	Al parecer tienes un error en la persona del verbo. El verbo debe ir en [primera / segunda / tercera persona].
1	VNE	Al parecer no conjugaste el verbo de tu oración. Revisa de nuevo.
2	COGSUJ%+[i]+%+[‘texto frase nominal’]	Recuerda que en la frase nominal debe haber concordancia en el género gramatical.
6	ADETSUJ%+[i]+%+[‘texto frase nominal’]	En tu respuesta hay una palabra de más en el sujeto de la oración. Revisa de nuevo.

6	ADETPRED%+[i]+%+['texto frase nominal']	En tu oración hay una palabra de más en el predicado. Concéntrate en los determinantes.
---	---	---

Tabla 4: Código de errores y mensajes asociados

3.4. *Feedback*

Una vez que se tienen los errores tanto sintácticos como morfológicos, se genera un reporte que da cuenta de los errores encontrados. Este se guarda en una base de datos a la cual tiene acceso el profesor, lo que le permite un mejor monitoreo de sus alumnos y de su aprendizaje.

Los códigos de errores están asociados, a su vez, a una combinación de *feedback* correctivo con claves metalingüísticas y *feedback* de elicitación, ya que estudios previos del equipo de investigación demuestran los efectos positivos, por un lado, del *feedback* de andamiaje sobre todo en los nuevos contextos de aprendizaje, donde el alumno ejerce mayor autonomía y el rol del profesor es el de mediador (Morales & Ferreira, 2008) y, por otro lado, que estos tipos de *feedback* que tienden a la autorreparación de los errores tienen mejores resultados para alumnos con un nivel de competencia lingüística intermedia o avanzada (Ferreira, 2003, 2006 y 2007).

La efectividad del tipo de *feedback* depende, además, de las diferencias individuales de los alumnos, tales como factores motivacionales y estilos de aprendizaje (Heift & Schulze, 2007; Sheen, 2007, 2011; Ferreira et al., 2012; Barrientos et al., 2012). Por ese motivo, para la retroalimentación se utilizó una combinación de *feedback*, correctivo con claves metalingüísticas y *feedback* de elicitación, para cubrir los requerimientos de varios estilos de aprendizaje.

Además, estudios han demostrado que el *feedback* correctivo metalingüístico suele ser muy efectivo cuando está combinado con corrección directa de los errores de los alumnos. Este tipo de *feedback*, al ser notado, contribuye a que los alumnos entiendan sus errores (Ferreira, 2007; Sheen, 2007). Por este motivo, se le ha dado prioridad al *feedback* metalingüístico automático, posible gracias a la inclusión de técnicas de procesamiento de lenguaje.

Por último, no se debe dejar de mencionar que los mensajes de *feedback* son claros, precisos y concisos; en su mayoría ocupan una sola línea. Las frases cortas y el léxico seleccionado se adecuan a la competencia lingüística de un alumno de nivel intermedio.

3.5. Ejemplos

A continuación se presentan algunas salidas del ELE-TUTOR con el fin de ejemplificar el funcionamiento del *parser* y sus análisis realizados.

La Figura 1 muestra la entrada del alumno en un ejercicio de futuro; la respuesta está correcta, lo que es señalado con un *feedback* positivo en color verde. Estos mensajes están asociados a la lista de respuestas correctas de la base de datos del programa. La entrada del alumno es comparada con esta base y, si coincide con alguna, se entrega un mensaje positivo. Se dispone de una variedad de mensajes de *feedback* de este tipo, que se entregan de manera aleatoria, para otorgar mayor dinamismo al sistema.



Figura 1: Salida del tutor ELE-TUTOR para respuesta correcta

En la Figura 2 la respuesta está incorrecta y esto se señala con el *feedback* correctivo metalingüístico en color rojo: "tienes un error en el verbo de tu oración. Este debe ir en condicional". Nótese que la oración está gramaticalmente correcta, pero el tipo de ejercicio es sobre el condicional, por lo que el sistema indica al estudiante que debe concentrarse en la temática gramatical que se está trabajando. En efecto, una de las

particularidades del programa es que no solo identifica errores en la forma gramatical, sino también de coherencia sobre el tema gramatical del ejercicio. Esto es posible gracias a un conjunto de reglas adicionales para cada ejercicio que exige determinados rasgos en el enunciado de la respuesta.



Figura 2: Salida del tutor ELE-TUTOR con feedback correctivo metalingüístico (con oración gramaticalmente correcta)

Cuando el estudiante vuelve a cometer el mismo error, el *feedback* que se entrega es de elicitación (en color negro), ya que se asume que la ayuda metalingüística no ha dado los resultados esperados. La Figura 3 muestra un segundo intento de responder el mismo ejercicio de condicional. El *feedback* de elicitación "¿Qué harías tú si estuvieras en el lugar de tu amiga?" indica de manera implícita que debe usar la forma condicional. Cada ejercicio contempla hasta dos posibilidades de respuesta por error y cada error puede ser corregido hasta dos veces. Al tercer intento, el sistema entrega la respuesta correcta e invita a la resolución de otro ejercicio con el fin de evitar la desmotivación del estudiante.



Figura 3: Salida del tutor ELE-TUTOR con *feedback* de elicitación

Es necesario enfatizar que, aunque el programa es capaz de reconocer todos los errores que se presentan en una respuesta del estudiante, el programa solo trata un error por vez, con el fin de que el estudiante atienda concentradamente a la forma gramatical focalizada, tal como se puede observar en la figura 4.

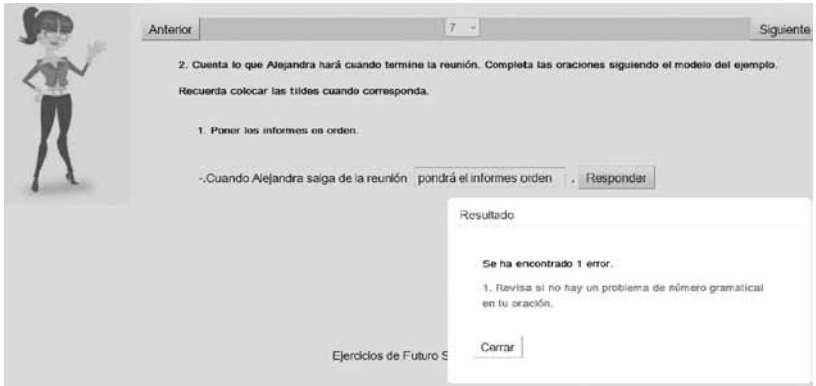


Figura 4: Salida del tutor ELE-TUTOR para respuestas con más de un error.

En efecto, la figura 4 muestra una respuesta que contiene más de un error, sin embargo, el *feedback* que se presenta es por el error de concordancia, aun cuando se cometió además un error de omisión de preposición. Esto es posible debido a que el *parser* está programado de acuerdo a una jerarquización de errores y, por lo tanto, despliega el mensaje para el error de mayor importancia según la jerarquía.

Sin embargo, todos los errores reconocidos ingresan a la base de datos del sistema a la que tiene acceso el profesor para poder monitorear mejor al alumno y poder apoyarlo de mejor manera en su proceso de aprendizaje.

4. Comentarios finales

Este artículo ha centrado su atención en el reconocimiento y tratamiento de errores gramaticales a través de un analizador automático que funciona en el contexto de una plataforma de enseñanza / aprendizaje de español como lengua extranjera.

La fundamentación teórica y empírica y los estudios realizados en contexto de clases reales han evidenciado que la propuesta conduce al mejoramiento de la precisión gramatical de las formas lingüísticas que se requieren desarrollar en el nivel de competencia lingüística intermedio (B2 según el Marco de Referencia Europeo para las lenguas).

En una taxonomía de errores, cada tipo debería estar perfectamente especificado y separado de los demás. Esto se puede lograr atribuyendo a cada caso rasgos definidos y un lugar fijo en la jerarquía previamente fijada. A medida que se van incluyendo más niveles en la taxonomía, mayor es la posibilidad de que se produzcan solapamientos; sin embargo, una tipología sin subclasificaciones no sería lo suficientemente robusta y resultaría poco funcional.

Uno de los problemas más importantes al que nos vimos enfrentados en el diseño de este sistema es la dificultad para tratar de manera automática algunos errores por las limitaciones del corpus. Dado que este ha sido etiquetado y que las etiquetas solo evidencian rasgos morfosintácticos, el *parser* no logra detectar errores de origen léxico-semántico, por lo que no se pudieron crear reglas para ese tipo de errores. En todo caso, se creó un pequeño corpus de palabras asociadas a los ejercicios y así forzar el procesamiento solucionando en parte el problema del corpus.

En estos momentos se está proyectando efectuar estudios acerca de los componentes actitudinales y la percepción de los alumnos acerca del tratamiento de sus errores, es decir, de su reacción ante el *feedback*, tanto en el contexto semi-presencial como no presencial.

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 objetivo final de apoyar a la enseñanza/aprendizaje del español mediante STI dotados de técnicas de procesamiento de lenguaje natural que son 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.

5. Bibliografía

- Alba Quiñones, V. de (2009). 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). Aproximación al tratamiento del error en la clase de E/LE desde la perspectiva del Análisis de Errores. *Estudios de Lingüística Aplicada* 23 (41): 101-125.
- Amaral, L & Meurers, D. (2011) On Using Intelligent Computer-Assisted Language Learning in Real-Life Foreign Language Teaching and Learning. *ReCALL. Cambridge University Press* 23 (1), 4-24.
- Ammar, A. & Spada, N. (2006). One size fits all? Recasts, Prompts, and L2 Learning". *SLA* 28, 543-574.
- Barrientos, F., Ferreira, A. & Salcedo, P. (2012). Modelado del estudiante para el STI ELE-TUTOR: diseño de un componente adaptativo para apoyar la competencia lingüística del español como lengua extranjera. *Boletín de Filología*.
- Bitchener, J. & Knoch, U. (2008). The value of written corrective feedback for migrant and international students. *Language Teaching Research Journal* 12, 409-431.
- Carroll, S. & Swain, M. (1993). Explicit and Implicit Negative Feedback: An Empirical Study of the Learning of Linguistic Generalizations. *Studies in Second Language Acquisition* 15 (3), 267-296.

- Chaudron, C. (1977). A descriptive model of discourse in the corrective treatment of learners' errors. *Language Learning* 27, 29-36.
- Consejo de Europa (2002). *Marco Común Europeo de Referencia para las Lenguas: Aprendizaje, Enseñanza, Evaluación*. Online. Recuperado el 7 de febrero de 2012: <http://cvc.cervantes.es/ensenanza/biblioteca_ele/marco/cvc_mer.pdf>
- Corder, S. P. (1974). Error Analysis. En Allen, J. P. & Corder, S.P. (Eds.). *The Edinburgh Course in Applied Linguistics, Volume 3 - Techniques in Applied Linguistics* London: Oxford University Press, 122-131.
- ____ (1981). *Error Analysis and Interlanguage*. Oxford: Oxford University Press.
- Dodigovic, M. (2005). *Artificial Intelligence in Second Language Learning: Raising Error Awareness*. Clevedon: Multilingual Matters Ltd.
- Doughy, C. & Varela, E. (2009). Atención comunicativa a la forma. En Doughy, C. & Williams, J. (Ed.). *Atención a la forma en la adquisición de segundas lenguas en el aula*. Madrid: Edinumen, 127-152.
- Ellis, J. (2011). Peer feedback on writing: Is on-line actually better than on-paper? *Journal of Academic Language & Learning* 5 (1), 88-99.
- Ellis, R. (1997). *The Study of Second Language Acquisition*. Oxford: Oxford University Press.
- Ellis, R., Loewen, S. & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in second language acquisition* 28, 339-368.
- Ellis, R., Sheen, Y., Murakami, M. & Takashima, H. (2008). The effects of focused and unfocused written corrective feedback in an English as a foreign language context. *System* 36, 353-371.

- Ferreira, A. (2003). *Feedback strategies for second language teaching with implications for intelligent tutorial systems*. Tesis Doctoral, Universidad de Edimburgo, Edimburgo: Escocia.
- ____ (2006). Estrategias efectivas de feedback positivo y correctivo en el español como lengua extranjera. *Revista Signos* 39 (62), 379-406.
- ____ (2007). Estrategias efectivas de feedback correctivo para el aprendizaje de lenguas asistido por computadores. *Revista Signos* 40 (65), 521-544.
- Ferreira, A., Moore, J. & 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* 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* 45/79, 102-131.
- Gass, S. & Mackey, A. (2007). *Conversational interaction in second language acquisition: A collection of empirical studies*. Oxford: Oxford University Press.
- Good, T. & Brophy, J. (2000). *Looking in Classrooms*. (5ª Ed.). Nueva York: Harper Collins.
- Granger, S. (2003). Error-tagged Learner Corpora and CALL: A Promising Synergy. *CALICO Journal* 20 (3), 465-480.
- Harbusch, K. & Kempen, G. (2010). Automatic online writing support for L2 learners of German through output monitoring by a natural-language paraphrase generator. En Levy, M., Blin, F., Bradin Siskin, C. & Takeuchi, O.

- (Eds.). *WorldCALL: International perspectives on computer-assisted language learning*. Nueva York: Routledge.
- Heift, T. & Schulze, M. (2007). *Errors and intelligence in computer-assisted language learning. Parsers and Pedagogues*. Nueva York: Routledge.
- Hyland, K. & Hyland, F. (2006). *Feedback in second language writing: Context and Issues*. Cambridge: Cambridge University Press.
- James, C. (1998). *Errors in language learning and use: exploring error analysis*. Harlow: Pearson Education Limited.
- Jurafsky, D. & Martin, J. (2009). *Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Nueva Jersey: Prentice Hall Inc.
- Lavid, J. (2005). *Lenguaje y nuevas tecnologías. Nuevas perspectivas, métodos y herramientas para el lingüista del siglo XXI*. Madrid: Ediciones Cátedra (Grupo Anaya S.A.).
- Lyster, R. (1998). Recast, repetition and ambiguity in L2 classroom discourse. *Studies in Second Language Acquisition* 20 (1), 51-81.
- Lyster, R. & Mori, H. (2006). *Interactional feedback and instructional counterbalance*. SLA 28, 269-300.
- Lyster, R. & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation ofform in communicative classrooms. *Studies in Second Language Acquisition*, 19, 37-66.
- Long, M. (1991). Focus on Form: A design feature in language teaching methodology. En de Boot, K., Ginsberg, R. & Kramsch, C. (Eds.). *Foreign language research in cross-cultural perspective*. Amsterdam: Benjamin, 39-52.
- Morales, S. & Ferreira, A. (2008). La efectividad de un modelo de aprendizaje combinado para la enseñanza del inglés como lengua extranjera: estudio empírico. RLA 6 (2), 95-11.

- Panova, I. & Lyster, R. (2002). Patterns of Corrective Feedback and Uptake in an Adult ESL Classroom. *TESOL Quarterly* 36, 573-595.
- Richards, J. (1984). [1974]. *Error Analysis, Perspectives on Second Language Acquisition*. Londres y Nueva York: Longman.
- Santos Gargallo, I. (1993). *Análisis contrastivo, análisis de errores e Interlingua en el marco de la lingüística contrastiva*. Madrid: Síntesis.
- Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics* 10 (3), 209-231.
- Sheen, Y. (2007). The effects of corrective feedback, language aptitude, and learner attitudes on the acquisition of English articles. En Mackey, A. (Eds.). *Conversational Interaction in Second Language Acquisition*, Oxford: Oxford Applied Linguistics.
- Sheen, Y. (2011). *Corrective feedback, Individual Differences and Second Language Acquisition*. Nueva York: Springer.
- Siddiqui, T. & Tiwary, U.S. (2008). *Natural language processing and information retrieval*. Nueva Dehli: Oxford University Press.
- Swain, M. (1998). Focus on form through conscious reflection. En Doughty, C. & Williams, J. (Eds.). *Focus on form in classroom second language acquisition*. Cambridge: Cambridge University Press, 64-81.
- Vásquez, G. (1999). *¿Errores? ¡Sin falta!* Madrid: Edelsa.
- _____. (2000). *La destreza oral*. Madrid: Edelsa.