

# Corrector de gramática para estudiantes autodidactas de inglés como lengua extranjera

Jim Lawley<sup>1</sup>

UNED

Ralph Martin

Cardiff University

## Resumen

Las nuevas tecnologías han aumentado notablemente las oportunidades de los estudiantes de inglés como lengua extranjera (ILE) para aprender fuera de clase. Sin embargo, hoy por hoy, la única corrección fiable de sus composiciones escritas la proporciona el profesor. Este artículo analiza hasta qué punto los programas informáticos pueden ayudar a detectar y a corregir los errores de las composiciones de los alumnos de nivel medio-bajo y medio. Se señala que los correctores ortográficos dan buenos resultados, pero que los correctores gramaticales, que se basan en un análisis sintáctico, a menudo confunden a los estudiantes. También se describe el principio y el funcionamiento de un nuevo tipo de corrector gramatical (el Corrector Gramatical de la UNED), que se fundamenta en un sistema de «búsqueda y coincidencia». Los estudios preliminares, que emplean una versión de este programa diseñada para estudiantes cuya lengua materna es el español, sugieren que dicho corrector gramatical puede ser más fiable y útil. Así, el artículo analiza cómo una versión más avanzada de este corrector gramatical, que incluyera expresiones regulares, podría aplicarse en muchos más casos, conservando un alto nivel de fiabilidad. Si esto es así, dentro de un tiempo se podrá superar el vacío que existe actualmente para la corrección de textos, que es, probablemente, el mayor problema al que se enfrentan actualmente quienes se enfrentan al estudio de una lengua de modo autodidacta.

---

<sup>1)</sup> Los autores están muy agradecidos a Patrick Campbell-Preston, por su ayuda en la implementación de la idea original. También les gustaría expresar su gratitud a José I. Rodríguez Aguado, Blanca E. Villoria Díez y Francisco J. Paredes Arranz, del Departamento de Inglés del IES «Antonio Tovar», en Valladolid, por su generosidad al permitirnos utilizar su corpus.

*Palabras clave:* corrector gramatical, ILE, español como lengua materna, estudio autodidacta, auto-corrección, aprendizaje por ordenador, expresiones regulares, secuencias incorrectas, palabras problemáticas.

**Abstract:** *A Grammar Checker for Autonomous Students of English as a Foreign Language*

New technologies have greatly increased the opportunities for students of English as a foreign language (EFL) to receive input outside the classroom. At present, however, the only reliable source of corrective feedback on their compositions is a teacher. This article examines the extent to which computer programs can help lower-intermediate-level and intermediate-level students of English as a foreign language (EFL) detect and correct mistakes in their written compositions. It notes that spelling-checkers give good results but that parser-based grammar checkers often mislead students. It describes the basis and operation of a new kind of grammar checker (the UNED Grammar Checker) which is based on a «search and match» paradigm. Preliminary studies, using a version of this program prepared for Spanish mother-tongue students, suggest that such a grammar checker may be more reliable and useful. The paper then discusses how an improved version of this grammar checker incorporating regular expressions could cover many more cases, while still giving a high level of reliability. If so, in time the absence of corrective feedback on free-form written production, currently probably the greatest problem facing language learners studying autonomously, will be overcome.

*Key words:* grammar checker, EFL, Spanish mother-tongue, autonomous study, self-correction, computer-assisted learning, regular expressions, incorrect sequences, problem words.

## Introducción

La revolución tecnológica de las dos últimas décadas ha hecho que las clases de inglés ya no sean el principal medio que tienen los estudiantes para relacionarse con esta lengua. Es fácil encontrarse con el inglés en la televisión por satélite, en los DVD, en Internet, en el correo electrónico, en los *chat*, en los periódicos y revistas más modernos, a través del teléfono, en la música pop inglesa y americana, y también en medios más tradicionales, como los libros. Conforme las tecnologías continúan desarrollándose, estos medios de acceso al inglés crecen y tienden a ser más baratos y más fácilmente accesibles.

Al mismo tiempo, las nuevas perspectivas sobre la naturaleza del proceso de aprendizaje de la lengua, sugieren que los ejercicios que se centran en la precisión y la corrección son muy importantes para ayudar a los estudiantes a ordenar y a dar sen-

tido a la nueva lengua (ver Granger y Tribble, 1998, para una síntesis y una discusión sobre esto). Los tipos de ejercicios que se fueron abandonando durante las décadas de los ochenta y de los noventa, como la traducción y el dictado, parece que pueden tener un gran valor para los alumnos, especialmente para los que trabajan de modo independiente en su casa. Por ejemplo, un estudiante puede traducir una frase del español al inglés y luego compararla con la traducción correcta que figura al final del libro. Asimismo, los estudiantes que trabajan solos pueden escribir el dictado de una grabación y corregirlo después a partir de la versión escrita.

Sin embargo, no resulta tan sencillo trabajar con la mayoría de los libros que contienen cursos de aprendizaje de ILE. Para completarlos con éxito, requieren la presencia de un profesor y de otros compañeros (para las numerosas actividades que deben realizarse en pareja o en grupo), y no suelen tener en cuenta el tipo de estudio más autodidacta que acabamos de describir. La mayor parte de las actividades del libro de ejercicios requieren la orientación y la corrección del profesor, y las secciones que se incluyen al final de los libros no contemplan la auto-evaluación. Allwright y Bailey señalan que:

(...) algunos libros de texto modernos y accesibles (...) pueden ser poco útiles para los estudiantes (...) si, por ejemplo, éstos tratan de estudiar de un modo autodidacta o de recuperar una lección que se han perdido. Los estudiantes pueden llegar a percibir que estos materiales son de poca ayuda, y que constituyen uno de los motivos por los que nunca van a tener éxito como estudiantes (Allwright y Bailey, 1991, p. 163).

Estas limitaciones en los libros de texto diseñados para las escuelas pueden incluso explicar, al menos en parte, por qué los resultados en las situaciones convencionales de aprendizaje de una lengua extranjera son pobres; después de años de estudio, muchos estudiantes dejan la escuela decepcionados por la poca utilidad del nivel alcanzado (Carrol, 1975, Stern, 1983, citado en Skehan, 1996). Skehan (1989) sugiere que, en realidad, sólo los más capacitados alcanzan niveles satisfactorios de competencia comunicativa.

Por tanto, en muchos aspectos, puede ser preferible la situación del estudiante adulto que trabaja solo en casa. Efectivamente, los estudiantes que optan por un modo autodidacta de estudio es probable que aprendan más que aquellos que sólo adquieren conocimientos cuando hay un profesor delante, ya que tendrán más ocasiones para estudiar. Como hemos señalado, los estudiantes autodidactas tienen numerosas

oportunidades de aprendizaje y de corrección de sus tareas (por ejemplo, mediante la traducción y los ejercicios de dictado). Sin embargo, para que el proceso avance, las investigaciones también enfatizan la importancia de la corrección de las composiciones escritas por los estudiantes (Tschichold, 2003); es decir, la corrección que tradicionalmente proporcionan los profesores. Los errores que comete un estudiante en las composiciones constituyen importantes oportunidades de aprendizaje (Edge, 1989), y los estudiantes de nivel medio-bajo y medio, casi por definición, cometen muchos fallos en sus composiciones escritas. La composición de 204 palabras que figura más abajo, escrita por un estudiante español, adulto y de nivel medio, en respuesta a la pregunta «¿Preferirías vivir en la ciudad o en un pueblo en el campo?», constituye un ejemplo típico:

*I think both ways of live are pleasant. At present I'm living in a village, but I've lived in a big town too.*

*When I was living in Madrid, life was very nice, because you have a lot of things near you. Theatbre, cinema, shopping-centers... all is easy to find and to have. But in the other hand, life was also difficoult because it was very noisy and dirty, and people had not ever the opportunity - and the time - to meet each other.*

*Living in a village everything is quieter. You have a bealtier life; the air here is not so insane as in the big cities. You can lead a sort of life more according to the nature laws. For instance, you can go swimming whenever you wants, you can be among the trees in the wood etc. Actually I have the possibility to choose the kind of choice you ask for. And at the present I think I prefer to live in this concrete village where I'm living now, because I feel that I dont need anymore that I have. Perphaps in the future I shall change my mind, but for the moment, at the present day, I'm very happy living bear.*

(Creo que ambos estilos de vida son agradables. Actualmente vivo en un pueblo, pero también he vivido en una gran ciudad.

Cuando vivía en Madrid, la vida era muy cómoda, porque tienes muchas cosas cerca. Teatro, cine, centros comerciales... todo es fácil de encontrar y de obtener. Pero, por otro lado, la vida también era difícil, porque había mucho ruido y suciedad, y la gente nunca tenía la oportunidad -ni el tiempo- para reunirse. Viviendo en un pueblo todo es más tranquilo. Tienes una vida más saludable; el aire no está tan viciado como en las grandes ciudades. Puedes llevar un tipo

de vida más cercana a las leyes de la naturaleza. Por ejemplo, puedes ir a nadar siempre que quieras, puedes estar en el bosque entre los árboles, etc. La verdad es que tengo la oportunidad de elegir entre las opciones por las que se pregunta. Y actualmente creo que prefiero vivir en este pueblo en particular en el que vivo ahora, porque creo que no necesito nada más de lo que tengo. Tal vez en el futuro cambie de opinión, pero por el momento, a día de hoy, soy muy feliz viviendo aquí).

Las decisiones en torno a qué constituye exactamente un error son notablemente complejas y, en algunas ocasiones, subjetivas (ver, por ejemplo, la discusión en Lennon, 1991). Pero cuatro profesores expertos en ILE estuvieron de acuerdo en diecinueve errores de la composición y en lo que el alumno debería haber escrito en su lugar. Estos diecinueve errores, con las correcciones que sugirieron los profesores, son los siguientes:

**El estudiante escribió**

*ways of live*  
*theatre*  
*shopping-centers*  
*all is easy*  
*in the other hand*  
*difficoult*  
*people had not ever*  
*bealtier*  
*not so insane*  
*more according to*  
*the nature laws*  
*you wants*  
*I have the possibility to*  
*at the present*  
*this concrete village*  
*I dont need*  
*anymore that I have*  
*perphaps*  
*I'm very happy living bear*

**El estudiante debería haber escrito**

*ways of life*  
*theatre*  
*shopping -centres*  
*everything is easy*  
*on the other hand*  
*difficult*  
*people never had*  
*healthier*  
*not so unhealthy*  
*more in accordance with*  
*the laws of nature*  
*you want*  
*I can*  
*at present*  
*this particular village*  
*I don't need*  
*any more than I have*  
*perhaps*  
*I'm very happy living here*

Dado que estos 19 errores constituyen importantes oportunidades de aprendizaje, la pregunta que debemos hacernos es: ¿puede un programa informático ayudar a los estudiantes a trabajar sin un profesor, detectando y corrigiendo sus propios errores? Si los estudiantes pudieran corregir sus propias composiciones de esta manera, serían capaces de aprender sin necesidad de profesores, simplemente escribiendo en inglés y estudiando las correcciones que hiciera el programa.

De hecho, seis de los errores *-theatrers, shopping-centers, difficult, bealtier, don't, perhaps-* son subrayados en rojo por el corrector ortográfico de Microsoft Word mientras se escribe este artículo. En todos los casos, la alternativa propuesta en primer lugar por el corrector ortográfico, coincide con la palabra que el estudiante debería haber escrito. Otros ejemplos más extensos confirman que los correctores ortográficos pueden resultar muy útiles para los estudiantes de ILE, detectando incluso algunos errores que normalmente no serían considerados como errores ortográficos: por ejemplo, *beautifuls* y *homeworks*, que, desde el punto de vista del profesor, serían errores morfológicos. Sin lugar a dudas, desde un punto de vista pedagógico, los correctores ortográficos resultarían aún más útiles si no sólo sugirieran correcciones, sino que también proporcionaran explicaciones y consejos. Así, en el caso de *shopping-centers*, para el estudiante sería de ayuda que el corrector ortográfico le explicara que *center* está correctamente escrito en inglés americano, que le mostrara otros ejemplos de diferencias ortográficas entre el inglés británico y el americano y que hiciera referencia a la importancia de la coherencia. En el caso de *dont*, el aprendizaje se vería fortalecido si en la corrección se explicara que el apóstrofo es necesario para sustituir la o de *not*, y se llamara la atención sobre otros casos similares como *isn't*. De este modo, tendríamos un esbozo para diseñar un corrector ortográfico específicamente dirigido a estudiantes de ILE. Sin embargo, aunque no esté propiamente diseñado para ellos, el corrector ortográfico de Microsoft Word habría sido claramente de gran ayuda para el autor de la composición que estamos considerando, detectando y ayudándole a corregir seis de los diecinueve errores de su escrito. Por supuesto, los correctores ortográficos llevan varios años ofreciendo una ayuda fiable.

Sin embargo, los 13 errores restantes de la composición del estudiante no pueden ser detectados por un corrector ortográfico. Son errores gramaticales y léxicos, no ortográficos. Un programa que pueda ayudar a los estudiantes en sus errores gramaticales y léxicos se llama corrector gramatical a; sin embargo, los correctores gramaticales se han elaborado tradicionalmente a partir de programas de identificación y de análisis sintáctico. Tschichold explica que:

Al igual que todas las aplicaciones de procesamiento de la lengua natural (PLN), los correctores gramaticales manejan la lengua de un modo muy diferente a como lo hacen los humanos. Para obtener un análisis que tenga sentido (para los humanos) de un texto, lo someten a una serie de módulos independientes que lo segmentan, lo identifican y lo analizan sintácticamente. Los dos procesos de identificación (asignando exactamente una parte del discurso a cada palabra) y análisis (análisis sintáctico) constituyen los mayores obstáculos en el proceso de análisis. Los identificadores más avanzados tienen una tasa de acierto del 95 al 98 por ciento y su desarrollo es cada vez más difícil y costoso. Teniendo en cuenta que esos porcentajes se han logrado en textos sin errores, es de esperar que el porcentaje de identificaciones incorrectas aumente cuando el identificador se enfrente a un texto parcialmente erróneo de alguien no nativo (Tschichold, 1999a, p. 7).

Como se sugiere más arriba, las composiciones escritas por estudiantes de nivel medio-bajo pueden contener numerosos errores y, por lo tanto, no resulta sorprendente que los correctores gramaticales basados en programas de identificación y análisis sintáctico sean de poca ayuda para los estudiantes de ILE de este nivel de conocimiento. El corrector gramatical de Microsoft Word, por ejemplo, no habría resultado de gran ayuda al autor de la composición que figura más arriba. De hecho, en 12 de los 13 errores que permanecen después de la corrección ortográfica, el corrector gramatical de Microsoft Word no encuentra ningún error. Sólo en el caso de *living bear* detecta algo erróneo, pero su sugerencia de que el estudiante cambie la expresión por *living bears* o *livings bear*, resulta igualmente inútil. Es evidente que el identificador está etiquetando *bear* como un verbo y *living* como su sujeto, y el análisis sintáctico está señalando que existe un problema de concordancia entre el sujeto y el verbo. Pero claramente, el estudiante simplemente ha confundido *bear* con su homófono *here*, y debería cambiar *living hear* por *living here*. Por lo tanto, el corrector gramatical de Microsoft Word se equivoca en su corrección de *living bear*. Se ha analizado la ayuda que otros correctores gramaticales pueden proporcionar a los estudiantes de ILE, comprobando que tienen defectos parecidos, y que señalan a los estudiantes errores que no han cometido (para una discusión más completa en torno a estas limitaciones, ver Bolt, 1992; Pennington, 1992 y 1993; Jacob y Rogers, 1999; Lawley, 1999 y Tschichold, 2003). No sorprende que los estudiantes consideren que estos programas son frustrantes y una pérdida de tiempo.

## El corrector gramatical de la UNED – Versión actual

En 2001 se comenzó a trabajar en la Universidad Nacional de Educación a Distancia (UNED), en Madrid, en el diseño de un programa de corrección gramatical que no se basara en un análisis sintáctico, sino que adoptara un nuevo enfoque para detectar específicamente los errores cometidos por los estudiantes de ILE de habla española. Se analizó un corpus de 24.000 palabras (200 composiciones, entre las que no se encontraba la que se ha tenido en cuenta en este artículo) escritas por estudiantes de nivel medio-bajo de 16 y 17 años, y los errores de las composiciones fueron identificados y analizados. Después, se eliminaron los errores que podían ser detectados y corregidos con ayuda del corrector ortográfico de Microsoft Word. En muchos de los casos restantes se observó que el estudiante había empleado una cadena de dos o tres palabras que resulta difícil encontrar en un nativo de habla inglesa. Por ejemplo, un estudiante escribió «*I like Madrid because is beautiful*». La consulta del Collins Cobuild Corpus<sup>2</sup>, de Internet, con 56 millones de palabras de habla inglesa, confirma la idea de que *because is* es siempre un error. En otros casos, los estudiantes empleaban secuencias que es muy difícil encontrar en el inglés nativo con el mismo significado que le da el estudiante. Por ejemplo, un estudiante escribió «*...and we could then go of the theatre*» (cuando *...and we could then leave the theatre* habría expresado mejor lo que quería decir). Una consulta al corpus de Internet confirma que la secuencia *go of* existe en el inglés nativo, pero sólo en frases como *let go of y make a go of*. Denominamos a las cadenas del tipo *because is* y *go of* (que tampoco aparecen en el corpus, o no aparecen con el significado que le dan los estudiantes, y cuyo uso correcto puede considerarse excepcional) como «Secuencias Incorrectas», y se elaboró una base de datos de unas 350. La corrección a cada secuencia incorrecta se escribió en español. Por ejemplo, en el caso de *because is*, la corrección señala la omisión del sujeto del verbo *is*. En el caso de *go of*, la corrección aclara que no se puede emplear como sinónimo de *leave*, pero que es correcta en secuencias como *let go of* y *make a go of*. Sin embargo, algunas cadenas que no aparecían en el Cobuild Corpus no fueron incluidas en esta base de datos de «secuencias incorrectas». Por ejemplo, un estudiante escribió «*I didn't use this things*». La cadena *this things* no aparece en el corpus y, por tanto, podía ser incorporada a la base de datos de «secuencias incorrectas». Sin embargo, resulta probable que ese tipo de estudiante escriba, por ejemplo, *this beds* y *this days* y cualquier otra combinación de *this* seguido de un nombre

<sup>2</sup> <http://titania.cobuild.collins.co.uk/form.html>

en plural. Se consideró que, antes de inflar la base de datos incluyendo varios cientos de cadenas de *this* seguido de un nombre en plural, era mejor crear otra base de datos de «palabras problemáticas». Se definieron así aquellas que se consideran asociadas al error, porque muchos estudiantes cometen errores cuando las utilizan. Otros ejemplos de «palabras problemáticas» son algunas que dan lugar a confusión (*false friends*), como *actually* (la palabra española actualmente significa *at present, no actually*), y palabras como *enough*, que causa problemas de orden (*He's not enough old to go...*). Se reunió un total de 150 palabras de este tipo en la base de datos de «palabras problemáticas».

Estas dos bases de datos, de «secuencias incorrectas» y «palabras problemáticas» forman la esencia del corrector gramatical de la UNED que, de un modo restringido, se puso a disposición de los estudiantes en diciembre de 2003. A los alumnos que usan el corrector gramatical de la UNED se les informa de que deben preparar su composición mediante el procesador de textos habitual, utilizando el corrector ortográfico, pero no el gramatical, ya que pueden encontrar correcciones erróneas con este último. Después deben copiar y pegar su composición en el corrector gramatical de la UNED. La pantalla que se muestra más abajo (Figura I) muestra la interfaz del programa.

FIGURA I. El Corrector Gramatical de UNED, después de hacer clic sobre Secuencias Incorrectas



Aquí observamos la composición que se está analizando, después de pegarla en la ventana superior del corrector gramatical de la UNED. Al hacer clic en «secuencias incorrectas», se destacan en azul las secuencias de palabras que aparecen tanto en la composición como en la base de datos de «secuencias incorrectas». Al colocar el cursor sobre la secuencia destacada, aparece en la parte inferior la explicación del error o la posible dificultad. En el caso de la secuencia destacada *in the other hand*, se explica que la frase correcta para indicar el contraste en un discurso no es *in the other hand* sino *on the other hand*. También se señala que la frase *in the other hand* puede resultar apropiada cuando *hand* se emplea de un modo literal. La única secuencia incorrecta de la composición que detecta la actual versión del corrector gramatical de la UNED es *in the other hand*; no encuentra ninguno de los otros errores. Esto puede parecer poco satisfactorio, pero se debe señalar que el corpus empleado para crear la base de datos de esta primera versión fue muy reducido -hay muchísimos errores que los estudiantes cometen habitualmente y que, sencillamente, no aparecen en una muestra de sólo 24.000 palabras-. Resulta más significativo señalar que el programa no comete errores ni realiza falsas correcciones, y que la corrección que se ofrece en el caso de *in the other hand* resulta apropiada y pertinente: parece

FIGURA II. El Corrector Gramatical de UNED, después de hacer clic sobre Palabras Problemáticas



muy probable que hubiera permitido al autor de la composición corregir *in the other band* y sustituirlo por *on the other band*.

Tras hallar las «Secuencias Incorrectas», el estudiante debería solicitar la detección de «Palabras Problemáticas», haciendo clic sobre el botón correspondiente. La pantalla que aparece más abajo (Figura II), muestra el resultado de la búsqueda de «Palabras Problemáticas»; en la parte inferior de la pantalla se muestra la información en torno a la palabra sobre la que se ha situado el cursor, *actually*, una de las palabras que puede dar lugar a confusión.

La explicación sobre cómo utilizar el corrector gramatical (que aparece en la pantalla cuando el estudiante hace clic en «Cómo utilizar el corrector gramatical») aclara que todas estas palabras se destacan simplemente porque se asocian con un error; es decir, porque muchos estudiantes cometen errores cuando las utilizan, no necesariamente porque se hayan empleado incorrectamente en esta ocasión. El programa las destaca en cualquier texto en el que aparezcan. Los estudiantes que no están seguros de haber empleado la palabra correctamente pueden consultar la información que se ofrece sobre el uso correcto e incorrecto de la palabra. Esta información explica los habituales usos indebidos de la palabra y cómo emplearla correctamente; los estudiantes pueden entonces decidir si la han utilizado apropiadamente o no. Aquí, el aspecto pedagógico fundamental es que se permite a los estudiantes evaluar lo que han escrito y asumir una responsabilidad en ello. A menudo, los estudiantes no necesitarán consultar la información, porque están convencidos de haber empleado la palabra con propiedad. En otras ocasiones, los estudiantes pueden decidir leer la información, pero sin modificar lo que han escrito. En estos casos, evidentemente, los estudiantes no habrán malgastado el tiempo, ya que el que han empleado en revisar cómo se debe usar la palabra, puede ayudarles a utilizar la palabra con mayor seguridad la próxima vez. En otras ocasiones, pueden decidir realizar modificaciones. La sección «Palabras Problemáticas» resulta especialmente útil porque puede sacar a la luz errores que ni siquiera un corrector gramatical basado en el análisis sintáctico, ni un profesor, podrían detectar. Como ejemplo, se debe señalar que en el caso de la oración (tomada de la composición analizada en este artículo) *Actually I have the possibility to choose the kind of choice you ask for (...)*, la oración está gramaticalmente bien construida, y resulta muy difícil para un profesor decidir si la palabra *actually* está correctamente empleada o no. Es decir, no está claro si el estudiante quiere decir *In fact I have the possibility* o *At present I have the possibility*. Dado que ambos significados son semánticamente admisibles, los profesores que corrigieron esta composición no lo señalaron como un error. Parece que sólo el estudiante que lo ha escrito

puede decir con seguridad qué es lo que en realidad se quería decir: este es un argumento de peso en defensa de la auto-corrección. En general, se pretende que la base de datos de «Palabras Problemáticas» facilite la auto-corrección, dado que funciona como un diccionario de términos que a menudo se emplean de un modo incorrecto. Al trabajar con palabras sueltas, en vez de con las abstractas descripciones gramaticales de los correctores de análisis sintáctico, la corrección puede ser más específica, proporcionando este tipo de información: «Has elegido utilizar esta palabra. Observa los casos en los que esta palabra se emplea correcta e incorrectamente». Parece que a los estudiantes les resulta más fácil entender y aplicar los consejos sobre cómo se emplean las palabras sueltas, que las descripciones gramaticales. Por supuesto, también es importante que la información esté escrita en la lengua materna de los estudiantes y que se evite la jerga metalingüística. De esta manera resulta más sencillo implicar al estudiante. Los ensayos preliminares sugieren que los estudiantes no ponen objeciones a las advertencias, ni siquiera cuando resultan innecesarias, siempre que puedan comprenderlas sin dificultad (un ensayo de este tipo queda reflejado en Lawley, 2005).

En el caso de la composición que se está analizando, la información que se da sobre *live*, *all* y *that* (en la frase *anymore that*) habría ayudado al estudiante a corregir tres errores más. Por lo tanto, si el estudiante hubiera empleado el corrector ortográfico de Microsoft Word y después hubiera detectado las «Secuencias Incorrectas» y las «Palabras Problemáticas», diez de los 19 errores podrían haber sido subsanados: el corrector ortográfico habría ayudado a corregir seis la utilidad «Secuencias Incorrectas» una y la utilidad «Palabras Problemáticas», tres.

A primera vista, el número de palabras (27) que se destacan al hacer clic en «Palabras Problemáticas» puede parecer excesivo, especialmente si tenemos en cuenta que sólo en tres casos las palabras están efectivamente mal empleadas. Sin embargo, nuestros ensayos preliminares sugieren que los estudiantes superan rápidamente cualquier rechazo inicial. Debe señalarse que algunas de las palabras problemáticas, como *very*, *for*, *this* y *because*, aparecen varias veces en la composición, y la información que se proporciona es, por supuesto, la misma en todos los casos. Además, muchas de estas palabras que aparecen con frecuencia, dan lugar a una información muy breve y sencilla (por ejemplo, en el caso de *very*, simplemente se recuerda al estudiante que esta palabra se emplea fundamentalmente para intensificar adjetivos, y que no tiene el mismo significado que a *lot of*). Dado que la información es sencilla y los estudiantes la verán con frecuencia (por ejemplo, *very* aparece tres veces en la composición), pronto les resultará familiar y la ignorarán. También se prevé que en

futuras versiones del corrector gramatical de la UNED, los estudiantes puedan «anular» algunas palabras de la base de datos «Palabras Problemáticas», de modo que no sean destacadas como «Palabras Problemáticas». Esto evitará que el corrector distraiga al estudiante destacando palabras problemáticas que ya se han comprendido correctamente, y permitirá que el estudiante se centre en los casos que aún no domina.

Por lo tanto, el corrector gramatical de la UNED ayuda a detectar sólo cuatro de los 13 errores no ortográficos de la composición que se analiza en este artículo. Sin embargo, conviene señalar que esta composición, aunque contiene muchos errores, está escrita por un estudiante universitario adulto, y es bastante más ambiciosa y sofisticada que las que forman el corpus de 200 composiciones escritas por estudiantes de 16 y 17 años que se empleó para crear las bases de datos del corrector gramatical de la UNED. Los estudios preliminares mostraron que los estudiantes españoles de ILE de 16 y 17 años que emplean esta primera versión del corrector gramatical de la UNED pueden detectar, junto con el corrector ortográfico, alrededor de un 60% de los errores de sus composiciones y, lo que es más importante, que el corrector no suele despistarlos señalando errores que no han cometido (Lawley, 2005). Estos resultados fueron considerados lo suficientemente alentadores como para justificar el gasto de seguir adelante e implementar el software del nuevo corrector gramatical. Se puede consultar el corrector gramatical de la UNED en <http://www.innova.uned.es>.

## El corrector gramatical de la UNED: Versión futura

La composición que se reproduce al inicio de este artículo no se eligió para mostrar todas las posibilidades de esta versión inicial del corrector gramatical de la UNED, sino para poner de manifiesto su potencial. Resulta evidente que su programa de «Búsqueda y Coincidencia», que destaca las palabras y frases de la composición que coinciden con las recogidas en alguna base de datos, podría mejorarse simplemente enriqueciendo estas bases de datos con más palabras y frases obtenidas a partir del análisis de un corpus mucho mayor de escritos de estudiantes. Por ejemplo, en el caso de «Palabras Problemáticas», la composición que se está analizando proporciona las palabras *insane* y *concrete* para ser incluidas en la base de datos. Ambas dan lugar a confusión: la palabra española *insano* significa *unhealthy*, no *insane*, y *concrete* sig-

nifica particular, no concrete. El software fue cuidadosamente implementado de forma que pueda seguir funcionando a una velocidad de interacción conforme aumenta el tamaño de estas bases de datos.

En esa composición también hay algunas frases que pueden ser incluidas en la base de datos de «Secuencias Incorrectas», bien porque no aparecen en el corpus de Internet, o bien porque sólo aparecen empleadas de una forma que es muy improbable que usen los estudiantes; si es necesario, estas formas de utilizar la frase pueden quedar reflejadas en la información que se proporciona. En tres casos (*of live*, *nature laws* y *more according to*) parece que estas cadenas se emplearían con mayor corrección en sentido literal; es decir, de la misma manera que cadenas como *in the other hand*, etc.

Sin embargo, en muchos casos los fallos son sólo ejemplos de todo un conjunto de errores similares. En vez de buscar una palabra o frase literales en la composición, se pueden emplear modelos para describir ese tipo de errores y buscarlos. Estos modelos pueden ser definidos mediante el cálculo, empleando «Expresiones Regulares» (ver, por ejemplo, Friedl, 2002). Las expresiones regulares pueden contener tanto caracteres que se representan a sí mismos, como metacaracteres que representan rangos de caracteres, alternativas, repeticiones, etc. Proponemos que se empleen expresiones regulares en una futura versión del programa, que se espera que sea mucho más eficaz. Esto evitará que el programa de la base de datos tenga que indicar explícitamente qué frases en concreto, dentro de un determinado conjunto, deberían estar en las bases de datos, y que tenga que enumerarlas. También hará posible que el programa pueda buscar numerosas variantes de las frases erróneas con mucha más rapidez que si estuvieran enumeradas individualmente en la base de datos. Otra ventaja es que la corrección también puede ser compartida por los conjuntos de errores, reduciendo el tamaño de los archivos.

En su forma más sencilla, las expresiones regulares son simples cadenas de caracteres que muestran qué es lo que se busca, por lo que, de un modo simplificado, se puede decir que las actuales bases de datos consisten en expresiones regulares; es decir, *in the other hand*, *of live*, *nature laws*, etc., son expresiones regulares. Sin embargo, como las expresiones regulares también ofrecen la posibilidad de incluir metacaracteres, proporcionan mucha más flexibilidad al proceso de búsqueda y coincidencia. Un ejemplo extraído de la composición que hemos tomado como muestra probará con claridad este funcionamiento en la práctica. En el caso de *all is easy*, como no hay casos de *all is easy* en el corpus de 56 millones de palabras de inglés contemporáneo, sería posible incluir sencillamente esta cadena en la base de datos de «Secuencias Incorrectas». Sin embargo, la experiencia sugiere que el estudiante que escribe *all is easy* es probable que también escriba *all is difficult*, *all is*

*nice*, y también *all was easy*, *all was difficult*, etc. Desgraciadamente, no podemos decir simplemente que las secuencias *all is* o *all was* (y por lo tanto *all isn't* y *all wasn't*) son, por sí mismas, necesariamente incorrectas: aparecen con bastante frecuencia cuando la palabra que precede a la secuencia es *at* o *after* o *above* o *of* (por ejemplo: *Communism above all is a mindset which... o ...the fact that he has any money at all is due to...*) y cuando está seguida por *well* o *not well* (por ejemplo: *all was not well there...*). Por lo tanto, es deseable que el programa detecte *all is*, *all isn't*, *all was* y *all wasn't*, excepto cuando están precedidas por una de estas cuatro palabras: *at*, *after*, *above*, *of* o cuando están seguidas por *well* o *not well*. Debemos emplear varias expresiones regulares para alcanzar este resultado. En primer lugar, es necesario detectar «alguna palabra *all is* otra palabra», y a continuación comprobar si «alguna palabra» y «otra palabra» cumplen los requisitos mencionados más arriba. Para detectar «alguna palabra *all is* otra palabra», se necesita la siguiente expresión regular:

```
(\w+)\s+all\s+is|isn't|was|wasn't\s+(\w+)
```

Aquí, la combinación de los metacaracteres `\s` y `+` representa uno o más espacios en blanco; `\w+` representa uno o más «caracteres de palabras» (es decir, caracteres que pueden aparecer legítimamente como parte de una palabra); y el metacarácter `|` indica alternativas. Los paréntesis de cada `\w+` hacen que se «capture» el texto que coincide con el de la expresión regular, proporcionándonos el texto que corresponde a lo que en nuestra explicación hemos denominado «alguna palabra» y «otra palabra». Así, ya se tiene cada una de estas partes del texto, y se puede aspirar a una mayor detección de coincidencias mediante expresiones regulares que utilicen las frases apropiadas (*of|after|above*) y (*well|not|swell*). Lamentablemente, una única expresión regular no es suficiente para excluir coincidencias de palabras específicas, aunque puede excluir las coincidencias de caracteres aislados. Éste es el motivo por el que se deben emplear múltiples expresiones regulares; por otro lado, debemos señalar que en otras pruebas de expresiones regulares, la coincidencia significa ignorar el caso, y la ausencia de coincidencia significa que el caso resulta de interés.

Ahora vamos a considerar los otros errores de la composición que hemos tomado como ejemplo. En el caso de *at the present*, el corpus sugiere que la cadena es admisible cuando va seguida de un nombre como *time*, *moment* o *day*. Se puede seguir un procedimiento similar al descrito para el caso de *all is*, para asegurar que *at the present* sólo coincida cuando no va seguido de tales nombres.

La cadena *had not ever* no aparece en el corpus. Parece probable que un estudiante que escribió *had not ever*, pueda escribir en otro momento *have not ever* y *has not ever*. Tampoco aparecen en el corpus ninguno de estos casos. Empleando la siguiente expresión regular, podemos detectar los tres:

$\backslash\text{sha}(\text{ve|d|s})\backslash\text{s}+\text{not}\backslash\text{s}+\text{ever}\backslash\text{W}$

Aquí,  $\backslash\text{W}$  representa cualquier carácter que no forma parte de una palabra (como un espacio, un signo de puntuación, etc.). Una expresión regular muy parecida coincidirá con *have the possibility to*, *had the possibility to* y *has the possibility to*:

$\backslash\text{sha}(\text{ve|d|s})\backslash\text{s}+\text{the}\backslash\text{s}+\text{possibility}\backslash\text{s}+\text{to}\backslash\text{W}$

En el caso de *you wants* resulta probable que en otras ocasiones el estudiante escriba *you looks*, *you thinks*, etc. Es decir, puede añadir equivocadamente una *s* al final de un verbo cuyo sujeto es *you*. Al construir la expresión regular, hay que tener presente que el infinitivo de algunos verbos termina en *ss*, por ejemplo: *kiss*, *miss* y *mess*. Por lo tanto, *you kiss*, *you miss*, etc., evidentemente son correctos. La siguiente expresión regular coincide con cadenas en las que *you* va seguido de una palabra terminada en *s*, pero no en *ss*:

$\backslash\text{s}[\text{Yy}]\text{ou}\backslash\text{s}+[\text{a-z}]+[\text{a-rt-z}]\backslash\text{s}\backslash\text{W}$

Detecta un espacio en blanco, luego *you* o *You*, después uno o más caracteres de espacio y, a continuación, una o más letras minúsculas, seguidas de cualquier letra excepto la *s*, seguida por la *s*, y ésta seguida por cualquier carácter que no pertenezca a una palabra (como un espacio, un signo de puntuación, etc.).

La cadena *living bear* no aparece en el corpus y, efectivamente, en las 40 líneas proporcionadas para *bear* y seleccionadas aleatoriamente, esta palabra nunca va precedida por otra terminada en *-ing*. Las cadenas en las que *bear* va precedido por una palabra terminada en *-ing* coinciden con la siguiente expresión regular:

$\backslash\text{w}+\text{ing}\backslash\text{s}+\text{hear}\backslash\text{W}$

De esta forma, el programa detectará *living bear*, *working bear*, *staying bear*, etc. Desgraciadamente, es evidente que no detectará *live bear*, *work bear*, *stay bear*, etc.

Si el estudio de un corpus más amplio muestra que estos usos erróneos de *bear* constituyen un problema para un cierto número de estudiantes, entonces se puede incluir la palabra en la base de datos de «Palabras Problemáticas».

En general, el empleo de expresiones regulares permitirá quitar muchas palabras de la base de datos de «Palabras Problemáticas», y permitirá también que los errores relacionados con estas palabras sean tratados como «Secuencias Incorrectas». Por ejemplo, en el caso de *very*, el error más habitual, como se ha mencionado más arriba, ocurre cuando precede a nombres en plural (por ejemplo, *there are very churches in Madrid*). Muy pocos adjetivos o adverbios terminan en s, y los que lo hacen, terminan en -ous (*famous, obvious, dangerous, etc.*) o en -ss (por ejemplo: *cross, amiss*). Los usos de *very* precediendo a palabras que terminan en s pero no en -ous ni en -ss son detectados, en primer lugar, mediante la siguiente expresión regular (que localiza *very* seguido por una palabra terminada en s):

$$\backslash s[Vv]ery\backslash s+([a-z]+s)\backslash W$$

y después mediante la comprobación de que la palabra capturada en el paréntesis no coincida con ninguna de las siguientes expresiones regulares:

$$[a-z]+ss$$

y

$$[a-z]+ous$$

El uso de *very* en frases como *these were the very churches we had intended to visit...* puede ser mencionado en la corrección.

De esta forma, los estudiantes destacarían los usos incorrectos de *very* al hacer clic en «Secuencias Incorrectas», en vez de destacar todos los usos de *very* al hacer clic en «Palabras Problemáticas». Algo parecido ocurre con *for*, que los estudiantes emplean habitualmente de un modo incorrecto; en esta composición se emplea en frases fijas, *for instance, for the moment, y ask for*. Como hemos visto, no resulta difícil utilizar expresiones regulares para evitar esas detecciones no deseadas. Más aún, las expresiones regulares también harán que sea más fácil capturar, por ejemplo, no sólo *because is*, sino también *because was, because has been, etc.*, y proporcionar la misma corrección para cada caso.

En la composición que venimos analizando, había 13 errores después de que los seis errores ortográficos fueran corregidos mediante el uso del corrector ortográfico de Microsoft Word, y aún quedaban doce después de que el corrector gramatical de la UNED hubiera detectado *in the other hand*. En una segunda versión del corrector gramatical de la UNED, con bases de datos ampliadas y con la introducción, cuando fuera necesario, de expresiones regulares relevantes, estos doce errores podrían ser detectados. Diez serían localizados en la sección «Secuencias Incorrectas» y dos en «Palabras Problemáticas».

El análisis de diez composiciones más, con un mini-corpus de casi 2.000 palabras, reforzó la idea que se desprende del estudio de la anterior composición: la combinación del corrector ortográfico y de un programa de corrección gramatical que utilice expresiones regulares podría ayudar a estos estudiantes a localizar y corregir un alto porcentaje de los errores de sus escritos. La mayor parte de los fallos a la hora de localizar errores ocurren con equivocaciones en los tiempos verbales (por ejemplo, escribir *I walk* cuando lo correcto sería *I walked*). Y lo que es más importante, con un diseño cuidadoso de las expresiones regulares, esperamos obtener un bajo número de correcciones erróneas y equivocaciones, aunque debe señalarse que, inevitablemente, habrá ocasiones en las que la corrección proporcione información sobre errores que los estudiantes tienden a cometer, más que advertencias sobre los errores que realmente han cometido, y en las que el usuario deberá decidir si lo que está escrito es correcto o no. Como se menciona más arriba, implicar de esta manera al estudiante en el estudio favorece el proceso pedagógico. En vista de las ventajas encontradas, se ha tomado la decisión de elaborar una segunda versión del corrector gramatical, incorporando expresiones regulares de carácter general.

## Conclusiones

La anterior discusión sugiere que los errores en las composiciones de los estudiantes de nivel medio-bajo y medio, se manifiestan frecuentemente mediante una cadena de caracteres (letras, espacios y, en ocasiones, signos de puntuación) que puede ser buscada y detectada por un corrector ortográfico o por un corrector gramatical basado en expresiones regulares. El resultado es el corrector gramatical de la UNED, que, utilizado junto con un corrector ortográfico, detecta muchos de los errores de esas com-

posiciones y proporciona una información adecuada sobre los mismos, posibilitando que los estudiantes realicen correcciones sin la ayuda de un profesor. También resulta evidente que el corrector gramatical de la UNED experimentaría una gran mejora mediante la incorporación de expresiones regulares que utilicen metacaracteres: se detectarían más errores y con mayor precisión. A la vez, es indudable que el estudio de un corpus mucho más extenso de escritos de los estudiantes enriquecerá enormemente las bases de datos. Queda igualmente patente que, a la hora de diseñar un corrector gramatical fiable para estudiantes de nivel medio-bajo y medio, resulta mucho más prometedor un enfoque basado en expresiones regulares que uno basado en el análisis sintáctico.

Por lo tanto, las primeras conclusiones nos muestran que el análisis de un corpus de composiciones de estudiantes y la utilización de expresiones regulares para detectar errores constituyen un enfoque prometedor para enfrentar el problema, hasta ahora irresoluble, de permitir a los estudiantes de ILE que trabajen de un modo independiente para localizar y corregir sus propios errores cuando escriben en inglés. Por ejemplo, la utilidad «Secuencias Incorrectas» del corrector gramatical de la UNED ya detecta con total fiabilidad muchos fallos comunes que los actuales correctores gramaticales no pueden localizar, y la utilidad «Palabras Problemáticas» proporciona información sobre palabras muy frecuentes, haciendo posible que los estudiantes decidan, sin la ayuda de un profesor, si lo que han escrito es correcto o no. Por lo tanto, hay buenas razones para pensar que en un futuro próximo, uno de los mayores obstáculos a los que tienen que enfrentarse los estudiantes que trabajan sin profesores será eliminado. Por supuesto, todos los estudiantes de ILE, incluyendo los que tienen profesores, son, o al menos deberían ser, estudiantes autodidactas en algunas ocasiones. El nuevo corrector gramatical no sólo pretende conseguir que los estudiantes que no tienen un profesor puedan auto-corregirse, sino también que los estudiantes que tienen profesores puedan entregar composiciones con menos errores. Tal vez esto muestre que los errores que puedan ser localizados por el programa requieran menos tiempo en clase, y que se pueda insistir más tanto con los errores que el programa no localiza, como con las cuestiones que el programa no puede abarcar, como la calidad y la sucesión de las ideas. En cualquier caso, está claro que queda mucho por investigar. No sólo será necesario llevar a cabo un estudio de un corpus mucho mayor, sino también más trabajos sobre la interacción de los estudiantes con el programa y sobre los resultados de esa interacción a largo plazo.

Traducción: Ignacio García-Romanillos García

## Referencias bibliográficas

- ALLWRIGHT, D.; BAILEY, K. M. (1991): *Focus on the Language Classroom: An Introduction to Classroom Research for Language Teachers*. Cambridge, Cambridge University Press.
- BOLT, P. (1992): «An Evaluation of Grammar Checking Programmes as Self-help Learning Aids for Learners of English as a Foreign Language». En *Computer-Assisted Language Learning*, 5, 1-2, pp. 49-91.
- FRIEDL, J. E. F. (2002): *Mastering Regular Expressions*. O'Reilly, Sebastopol, California.
- GRANGER, S.; TRIBBLE, C. (1998): «Learner Corpus Data in the Foreign Language Classroom: Form-focused Instruction and Data-driven Learning». En S. GRANGER: *Learner English on Computer*. London, Longman, pp. 199-209.
- JACOBS, G.; RODGERS, C. (1999): «Teacherous Allies: Foreign Language Grammar Checkers», en *CALICO Journal* 16, 4, pp. 509-531.
- LAWLEY, J. (1999): «A New Computer-Mediated Writing Aid for Spanish students of English as a Foreign Language». En P. FABER y otros: *English Language Teacher Education in Europe*. Peter Lang, Germany, pp 31-43.
- (2005): «A Preliminary Report on a New Grammar Checker to help Students of English as a Foreign Language». En *Arts and Humanities in Higher Education* 4, 1.
- LENNON, P. (1991): «Error: Some Problems of Definition, Identification and Distinction». En *Applied Linguistics*, 12, 2, pp. 180-196.
- PENNINGTON, M. (1992): «Beyond off-the-shelf computer remedies for student writers: alternatives to canned feedback». En *System* 20, 4, pp. 423-437.
- (1993): «Computer-assisted writing on a principled basis: the case against computer-assisted text analysis for non-proficient writers». En *Language and Education*, 7, 1, pp. 43-59.
- SKEHAN, P. (1989): *Individual Differences in Second Language Learning*. London, Edward Arnold.
- (1996): «Second language acquisition research and task-based instruction». En J. WILLIS; D. WILLIS: *Challenge and Change in Language Teaching*. Oxford, Heinemann, pp. 17-30.
- TSCHICHOLD, C. (1999): «Intelligent Grammar Checking for CALL», en *ReCALL Special issue*, 11, pp. 5-11.
- (1999): «Grammar Checking for CALL: Strategies for Improving Foreign Language

- Grammar Checkers». En K. CAMERON: *CALL: Media, Design and Applications*. Lisse, Swets and Zeigler, pp. 203-221.
- (2003): «Lexically Driven Error Detection and Correction», en *CALICO Journal*, 20, 3, pp. 549-559.