



El desarrollo de la expresión escrita mediante el uso de correctores gramaticales en español como 2L: la efectividad de *CorrectMe*

Rubén Chacón-Beltrán¹

Recibido: 8 de octubre de 2017/ Aceptado: 17 de junio de 2019.

Resumen. Este artículo presenta algunos resultados preliminares sobre la efectividad de un prototipo de corrector ortográfico y gramatical diseñado para aprendices de español como L2 con la finalidad de proporcionar al usuario información sobre sus errores mientras escribe. Esta herramienta, denominada *CorrectMe*, permite a los usuarios contrastar sus redacciones con un corpus normativo amplio que proporcionará datos sobre la frecuencia de aparición de palabras y pares de palabras, también conocidos como bigramas, en un corpus de español correcto. Desde el punto de vista metodológico se realiza un énfasis en la forma lingüística que llamará la atención del usuario sobre determinados errores, o posibles errores, proporcionándole así la oportunidad de procesar y corregir sus redacciones en el momento. Para analizar la efectividad de *CorrectMe* se han procesado dos corpus lingüísticos de aprendices de español como Segunda Lengua de 8.994 y 13.018 palabras, correspondientes a los niveles B1 y C1 del MCERL respectivamente.

Palabras clave: Desarrollo de la escritura, correctores gramaticales, español como 2L, lingüística de corpus.

[en] The development of second language writing with the help of grammar checkers in Spanish as an L2: the effectiveness of *CorrectMe*

Abstract. This article presents some preliminary results on the effectiveness of a software, a prototype of spell and grammar checker especially designed for learners of Spanish as an L2. The aim of the software is to provide students with specific information about their mistakes at the time of writing. This tool, which is called *CorrectMe*, allows the users of the software to contrast their own writing with a large normative corpus that will provide information on the frequency of words and pairs of words, also known as bigrams in a corpus of correct Spanish. From a methodological point of view there is a focus-on-form that should call students attention on specific mistakes and potential mistakes, providing them with a chance to self-correct their writing at the spot. In order to analyse the effectiveness of *CorrectMe*, two Spanish language learner corpora have been processed, containing 8.994 and 13.018 words, respectively corresponding to levels B1 and C1 according to the CEFR.

Keywords: Development of writing, grammar checkers, Spanish as an L2, corpus linguistics.

Cómo citar: Chacón-Beltrán, Rubén (2019). El desarrollo de la expresión escrita mediante el uso de correctores gramaticales en español como 2L: la efectividad de *CorrectMe*, *Círculo de Lingüística Aplicada a la Comunicación* 79, 203-216, <http://dx.doi.org/10.5209/CLAC.65656>

¹ UNED: Universidad Nacional de Educación a Distancia (España). Correo electrónico: rchacon@flog.uned.es

Índice. 1. La destreza escrita en el aprendizaje de lenguas. 2. La ortografía en el aprendizaje de lenguas. 3. El enfoque *n-gram*. 4. Metodología de investigación. 5. Análisis de los resultados. 5.1. Análisis de palabras individuales. 5.2. Análisis de bigramas. 6. Conclusiones. 7. Implicaciones para investigaciones futuras. 8. Referencias.

1. La destreza escrita en el aprendizaje de lenguas

La corrección de la producción escrita es esencial para el desarrollo de la competencia lingüística de los aprendices y requiere por una parte la identificación de los errores y, por otra, la provisión de retroalimentación que permita al aprendiz cambiar su suposición errónea sobre la lengua que está aprendiendo. En este sentido, la corrección manual por parte del profesor resulta primordial aunque presenta dos problemas principales; (a) por una parte requiere mucha dedicación de tiempo por parte del profesor que con frecuencia debe corregir las composiciones de decenas de estudiantes y, por otra parte, (b) es asincrónica por lo que hay un intervalo de tiempo inevitable entre que el estudiante escribe su redacción y recibe las correcciones o comentarios sobre la misma. Esta situación suele resultar en que el estudiante difícilmente incorpora en su interlengua los comentarios que hace el profesor pues la ventana cognitiva que estaba abierta mientras el estudiante elaboraba su redacción y mentalmente elaboraba hipótesis sobre lo que es correcto o no en la lengua se encuentra ya cerrada.

En los últimos años se ha identificado un incremento considerable en la investigación sobre el uso de correctores gramaticales, esto es, herramientas informáticas que ayudan durante el proceso de escritura proporcionando información de diverso tipo a la persona que escribe. Son numerosos los estudios que han analizado diversos aspectos la producción escrita del aprendiz de segundas lenguas y su relación con el uso de correctores gramaticales (Chacón-Beltrán, 2009, 2017; Haddon, 2009; Hernández García, 2012, 2017; Lawley, 2004, 2009; Manchanda, Athavale & kumar Sharma, 2016; Potter & Fuller, 2008). Un número considerable de estudios se ha centrado en la provisión de retroalimentación pedagógica que ayude al estudiante a avanzar en el proceso de aprendizaje del inglés (Bitchener & Knoch, 2010; Bitchener, Young & Cameron, 2005; Chandler, 2003; McMartin-Miller, 2014; Parr & Timperley, 2010; Sundeen, 2014; Stevenson & Phakiti, 2014). En menor medida también se ha investigado el aprendizaje de español (San Mateo, 2016).

Se aprecia una notable diferencia entre la cantidad de estudios llevados a cabo sobre el aprendizaje del inglés en comparación con otras lenguas que, por tratarse de lenguas con menor alcance internacional, o con menor número de aprendices de la lengua en cuestión como segunda lengua o lengua extranjera, han recibido una atención limitada. Por el momento, se han realizado algunos trabajos sobre el diseño y uso de correctores gramaticales en relación con lenguas como el francés (Burston, 2001), el griego (Gakis, Panagiotakopoulos, Sgarbas, et al, 2016), el portugués (Rino, Di Felippo, Pinheiro, et al., 2002.) o el español (San Mateo Valdehíta, 2016, 2017). En el presente trabajo nos centraremos precisamente en este último caso, es decir, el uso de este tipo de herramientas informáticas para el aprendizaje del español, una lengua que por su alcance internacional y la cantidad de estudiantes que la aprenden como segunda lengua, merece mayor atención de la recibida hasta la fecha. El grupo de investigación TISAAL de la UNED ha

desarrollado un *software* especializado, un corrector gramatical que funciona como un procesador de textos y que contrasta en tiempo real las redacciones de los estudiantes con un corpus normativo. En esencia, se trata de un prototipo de corrector gramatical que se denomina *CorrectMe*.

2. La ortografía en el aprendizaje de lenguas

De entre todas las destrezas que engloba el aprendizaje de la escritura, si hay un componente que tradicionalmente ha resultado soslayado es la enseñanza de la ortografía como se pone de manifiesto en Lawley (2015) y Godwin-Jones (2011). Dadas las limitaciones de tiempo y la complejidad de la tarea que se afronta, parece que se dedica más atención a otros componentes del aprendizaje de la escritura. Lawley (2015) describe el desarrollo de una herramienta informática para ayudar a los aprendices de inglés con la identificación de errores ortográficos tras contrastar de forma automatizada las redacciones de los estudiantes con un corpus normativo amplio. En una segunda fase, se elabora retroalimentación pedagógica personalizada para cada error que proporcione a los aprendices la información necesaria para corregir sus errores. En su estudio, Lawley (ibid.) describe una herramienta que se usa satisfactoriamente para identificar errores y posteriormente la emplea con un grupo de estudiantes que la usa con resultados favorables. Se ha usado la misma tecnología y una metodología de trabajo similar, para comprobar los resultados que pueden obtenerse en el aprendizaje de español.

3. El enfoque *n-gram*

El enfoque denominado en inglés *n-gram* en el análisis de composiciones se ha empleado ya con éxito en varias investigaciones con el inglés como objeto de estudio para analizar distintos elementos lingüísticos desde el punto de vista del aprendizaje o la traducción (Athanaselis, Bakamidis & Dologlou, 2006; Bestgen y Granger, 2014; Briscoe, Medlock & Andersen 2010; De Cock, Grange, Leech y McEnery, 1998; Lawley, 2015; Wu y Su, 2006; Yannakoudakis 2013; Yannakoudakis, Briscoe y Medlock, 2011), pero también en el análisis del español (Nazar y Renau, 2012; San Mateo Valdehíta, 2015). El término *n-gram* ha sido propuesto en referencias bibliográficas previas publicadas en inglés para describir la división de un texto escrito en segmentos de dos o más palabras consecutivas. Así se habla de *bigrams* (bigramas), *trigrams* (trigramas), etc. para referirse a las combinaciones de dos, tres o más palabras, donde *n* representa el número de palabras involucradas. Esta metodología de análisis consiste en la aplicación de un algoritmo (Sinclair, 1991) para calcular la frecuencia de las palabras así como la medida en que pares adyacentes, o pares de palabras, aparecen juntos en textos escritos. Además, se calcula de acuerdo con la frecuencia individual de cada palabra y la cantidad de veces que cada palabra debería aparecer en el corpus si las palabras pudiesen combinarse de forma aleatoria. De este modo se puede obtener un coeficiente que indica la medida en que dos palabras se atraen y por tanto es esperable que aparezcan juntas. Para la investigación descrita en el presente estudio se ha empleado un corpus de referencia de cien millones de palabras de español escrito correcto y corregido que se ha desarrollado en el seno del Grupo de

Investigación TISAAL de la Universidad Nacional de Educación a Distancia (UNED) para el desarrollo del prototipo de corrector gramatical para el aprendizaje de español denominado *CorrectMe*.

Esta herramienta informática consiste en un procesador de textos sobre el que se puede escribir directamente o se puede pegar un texto proveniente de otro procesador de textos. A continuación, el usuario pasa su reacción por varios filtros, el primer de ellos destinado a encontrar los errores de tipo ortográfico, el segundo centrado en identificar secuencias de dos o más palabras que no suelen ir juntas en español, el tercero se dedica en palabras que aunque no contienen un error en sí, son potencialmente problemáticas y por tanto es recomendable que la persona que escribe se cerciore de que el uso es correcto, por ejemplo el caso de cognados engañosos. En todos estos casos, al pulsar sobre la palabra o palabras resaltadas de forma automática se abre un desplegable en la parte inferior de la pantalla con la información de carácter metalingüístico que el usuario necesitará para corregir el error o la palabra problemática detectada por *CorrectMe*. Por último, con el filtro denominado “pares de palabras” esta herramienta contrasta la redacción del estudiante con un corpus normativo del que ya dispone *CorrectMe* y resalta de forma automática las secuencias de palabras o *bigramas* que como resultado de este análisis contrastivo aparecen en escasas ocasiones en el corpus normativo.

El corpus de referencia que se ha recopilado para su incorporación en *CorrectMe* consiste en un corpus sincrónico de textos narrativos escritos en español entre 1980 y 2012 de diferente tipología textual como noticias, artículos y novelas, de temática general y no excesivamente técnica, evitando también localismos o buscando siempre un registro formal escritos por periodistas o escritores (San Mateo Valdehíta, 2016).

A continuación se describe y ejemplifica el enfoque de análisis consistente en la frecuencia léxica y frecuencia de pares adyacentes o *bigramas*. Para explicar el concepto de probabilidad de palabras según el algoritmo desarrollado por Sinclair (1991) podríamos usar dos palabras muy frecuentes en español como son X e Y. En el corpus utilizado por *CorrectMe* la palabra X aparece X veces y la palabra Y aparece Y veces. Si las palabras ocurriesen al azar, es decir, sin la presencia de reglas gramaticales que regulen su uso, en términos estadísticos, las probabilidades de que ambas palabras apareciesen juntas sería de XX veces. Si esa combinación de palabras aparece menos de lo que sería esperable, nos estaría indicando que ese bigrama es poco frecuente y por tanto susceptible de tratarse de un error.

Como se indica en Lawley (2015: 26) es relativamente fácil, desde un punto de vista técnico, desarrollar un corrector gramatical que incorpore el análisis mediante los *bigramas* o pares de palabras, es decir, una herramienta informática que contenga la división en *bigramas* de un corpus normativo compuesto en este caso por cien millones de palabras (San Mateo Valdehíta, 2016) para contrastarlo con la producción escrita del usuario que se encuentra aprendiendo español. La metodología de trabajo consiste en (a) analizar la frecuencia de cada palabra en el citado corpus; (b) analizar la frecuencia de aparición de cada bigrama en ese corpus; y, (c) analizar la cantidad de veces que cada bigrama debería aparecer en el corpus si unas palabras se conectasen a otras de forma aleatoria en vez de seguir reglas gramaticales para ello. El procesamiento de esta ingente cantidad de información permite que con posterioridad, en tiempo real, se pueda contrastar

cualquier frase que se introduzca en *CorrecteMe* con este corpus de referencia y obtener información relevante sobre esa frase. A modo de ejemplo se analiza la siguiente oración: **Me gustaron las ciudades que visitamos para los magníficos palacios y edificios.*

Esta oración se puede dividir en pares de palabras, o *bigramas*, del siguiente modo.

Bigramas

1. Me gustaron
2. gustaron las
3. las ciudades
4. ciudades que
5. que visitamos
6. visitamos para
7. para los
8. los magníficos
9. magníficos palacios
10. palacios y
11. y edificios.

Tabla 1. La frase **Me gustaron las ciudades que visitamos para los magníficos palacios y edificios* dividida en *bigramas*.

La base de datos de *CorrectMe* analiza los *bigramas* descritos en la Tabla 1 y proporciona de forma automática la siguiente información relativa a la frecuencia de aparición de las palabras.

Frecuencia combinación	Frecuencia palabra 1	Frecuencia palabra 2	Probabilidad combinación	Umbral
Me gustaron=46	Me= 500612	gustaron= 137	0.69	66.67
gustaron las=14	gustaron= 137	las= 856448	1.17	11.97
las ciudades=2034	las= 856448	ciudades= 4523	38.74	52.5
ciudades que=171	ciudades= 4523	que= 3295376	149.05	1.15
que visitamos=20	que= 3295376	visitamos= 152	5.01	3.99
visitamos para=0	visitamos= 152	para= 593842	0.9	0
para los=12054	para= 593842	los= 1315294	7810.77	1.54
los magníficos=93	los= 1315294	magníficos= 688	9.05	10.28
magníficos palacios=8	magníficos= 688	palacios= 788	0.01	799.92
palacios y=105	palacios= 788	y= 2813605	22.17	4.74
y edificios=111	y= 2813605	edificios= 3717	104.58	1.06

Tabla 2. Análisis de frecuencias de la oración **Me gustaron las ciudades que visitamos para los magníficos palacios y edificios* por *CorrectMe*.

Tomando como muestra el primer ejemplo de la Tabla 2, se puede ver que (a) *Me gustaron* aparece 46 veces en el corpus, (b) la palabra *Me* aparece 500.617 veces, (c) la palabra *gustaron* aparece 137 veces en el corpus y si las palabras *Me* y *gustaron* apareciesen de forma aleatoria en vez del ordenadas de acuerdo con las reglas gramaticales, según el algoritmo de probabilidad descrito por Sinclair (1991) debería aparecer 0.69 veces. En cambio, en este corpus de 100 millones de palabras aparece un número de veces mucho mayor, esto es, 66'67 veces más de lo esperado por lo que las estadísticas muestran que hay un alto grado de atracción entre ambas palabras. Este alto grado de atracción entre ambas palabras nos indica también que es muy probable que la secuencia *Me gustaron* sea correcta en español. Cuanto más alto sea el umbral, más posibilidades de que la frecuencia sea correcta y al contrario, es decir, cualquier umbral situado por debajo de 1 nos indica que con toda probabilidad se trate de un *bigrama* que no es frecuente, en el uso correcto de la lengua y por tanto tiene muchas posibilidades de representar un error. Esto ocurre con el *bigrama* *visitamos para* de la Tabla 2, aunque las palabras *visitamos* (152 veces) y *para* (593.842 veces) ocurren de forma individual en el corpus utilizado, su umbral es 0 por lo que con toda probabilidad se trata de un error, de hecho es un error frecuente entre los aprendices de español ya que debería ser *Me gustaron las ciudades que visitamos por los magníficos palacios y edificios*. En cualquier caso, sería erróneo pensar que todos los *bigramas* que no aparezcan en el corpus o que tengan un umbral bajo, quizá inferior a 1 son erróneos. Véase por ejemplo que en la frase analizada anteriormente en la Tabla 2, el *bigrama* *ciudades* (4.523 veces) *que* (3.295.376 veces) tiene un umbral de tan solo 1.5 y no por ello es erróneo. Simplemente resulta poco frecuente en nuestro corpus de 100 millones de palabras.

CorrectMe proporciona información al estudiante indicando tres niveles de probabilidad de que se trate de un error, es decir, se sombreadá en rojo el *bigrama* que resulta muy sospechoso, en naranja el que se considere sospechoso y en amarillo el que se considere ligeramente sospechoso, como puede verse en la imagen 1 a continuación.

El análisis estadístico mostrado con anterioridad en la Tabla 2 puede ser útil, y de hecho lo es, para la detección automática mediante la herramienta *CorrectMe* de errores de tipo ortográfico y gramatical en textos escritos por hablantes nativos competentes como se describe en San Mateo Valdehíta (2016), pero también para detectar errores en las redacciones de aprendices de español como segunda lengua pues una vez detectados los errores es posible preparar retroalimentación pedagógica específica para cada error que se active cuando se ha detectado el error por medios informáticos.

A partir de lo expuesto en los apartados anteriores cabe preguntarse cuál sería la efectividad de una herramienta informática como *CorrectMe* para la detección de errores de tipo ortográfico y de *bigramas* para considerar la posibilidad en futuras investigaciones de proporcionar retroalimentación pedagógica al menos para los errores de tipo ortográfico o gramatical que resulten más frecuentes en corpus de aprendices de español como segunda lengua.



Imagen 1: Apariencia de la herramienta *CorrectMe*. En el centro aparece el área reservada para la escritura, a la derecha hay instrucciones de uso y en la parte superior se pueden ver los diferentes filtros o secciones en los que se está trabajando.

4. Metodología de investigación

Se seleccionó una muestra aleatoria del corpus CORANE de estudiantes con distintas lenguas maternas para los dos niveles lingüísticos que se analizaron, concretamente B1 y C1. En la siguiente tabla se indican los detalles del corpus recopilado. El corpus CORANE fue recopilado para el estudio de la enseñanza-aprendizaje de E/LE a partir de textos escritos por los alumnos y contiene 957 composiciones escritas por estudiantes de E/LE de los niveles A2, B1, B2 y C1 (Cestero Manceras & Penadés Martínez, 2009).

Lengua materna de los informantes	Número de redacciones: Nivel B1	Número de redacciones: Nivel C1
Inglés	10	10
Japonés	10	10
Coreano	5	5
Portugués	5	5
Chino	4	4
Sueco	4	4
Alemán	3	3
Número de palabras	8.994 palabras	13.018 palabras
Tamaño total del corpus		22.012 palabras

Tabla 3: Descripción de la muestra tomada del corpus CORANE, clasificada según nivel de competencia y lengua materna de los informantes.

Una vez recopiladas ambas muestras, seleccionadas aleatoriamente del corpus CORANE, según los detalles de la muestra descrita en la Tabla 3, se procedió a introducirlos para su análisis de forma separada por nivel, es decir B1 y C1, en el software *CorrectMe* siguiendo un procedimiento de corta y pega por lo que el corpus analizado por esta herramienta informática se correspondía exactamente con la muestra original del corpus CORANE.

Como se ha explicado anteriormente, el prototipo *CorrectMe* contiene dos filtros separados que permiten contrastar por una parte las palabras individuales del texto introducido con el corpus normativo de palabras correctas que permanece en su base de datos pensado especialmente para detectar problemas de ortografía, y un segundo filtro que permite, en una fase posterior, contrastar todos los *bigramas* contenidos en el texto introducido con los *bigramas* del corpus normativo incluidos en la base de datos del software, con el fin de detectar errores más allá del nivel de la palabra. De este modo, se realizan en realidad dos análisis separados de palabras individuales y de *bigramas*.

Dado que uno de los objetivos de esta investigación es contrastar la efectividad de este enfoque automatizado para la detección de errores con un enfoque manual en el que sería el profesor quien realizaría la labor de identificar y marcar estos errores, se ha realizado un segundo análisis manual de cada corpus con la finalidad de marcar igualmente los errores en el nivel de la palabra, que generalmente responde a errores de tipo ortográfico y a errores a nivel de *bigrama*, que frecuentemente se trata de errores de concordancia. Para clarificar esta cuestión, la Tabla 4 presenta una lista de errores, algunos detectados de forma manual y otros detectados automáticamente por *CorrectMe*.

	Errores ortográficos detectados por CM	Errores ortográficos <u>no</u> detectados por CM (detección manual)	Errores de <i>bigramas</i> detectados por CM	Errores de <i>bigramas</i> <u>no</u> detectados por CM (detección manual)
B1	*bailabamos *patío *incortés *después *polizia *cayéndose *Sabado *lluviendo *viejones *ocorrido	*vení *confía *rapido *llege *fantastico *mís *compañía *prometó *todavía *esistencia	*el ciudad *me hermana *todos países *misma tema *su moverse *para me *del años *la anoche *un ironía *esto mundo	*un discoteca *fue lleno *nuestras amigos *una desayuna *interesa todos *libros antiguas *en demasiado *mi vuelto *muy sorpresa *me también
C1	*cientas *aveces *enborachaba *colégio *água *habia *apreciendo *disastre *cientas *humilidad	*narcotrafico *apaque *quando *viente *suenó *avuela *Religion *holandéses *averguenzo *movimento	*es lleno *se como *muy enorme *los paredes *tiene mismo *fuera podido *de trabaja *yo fue *el problemas *este relación	*un herramienta *son vigilado *han restauradas *dudan a *a lado *existe muchas *dármelo cuenta *yo pensó *ni mismo *muchos gente

Tabla 4. Errores de ortográficos y de bigramas detectados por *CorrectMe* (CM), y errores detectados por manualmente.

4.1. Preguntas de investigación

Una vez seleccionados los dos corpus de aprendices de dos niveles educativos distinguidos, B1 y C1, según el Marco Europeo de Referencia para las Lenguas se formularon las siguientes preguntas de investigación:

- a) ¿En qué medida la herramienta informática *CorrectMe* puede identificar errores ortográficos en dos corpus lingüísticos producidos por estudiantes de niveles B1 y C1 en contraposición a la identificación por parte del profesor?
- b) ¿En qué medida la herramienta informática puede identificar combinaciones de dos palabras, o bigramas, erróneas en dos corpus lingüísticos producidos por estudiantes de niveles B1 y C1 en contraposición a la identificación por parte del profesor?

5. Análisis de los resultados

Se procedió a analizar separadamente los errores de cada nivel educativo, B1 y C1, comprobando en primer lugar los que eran detectados por *CorrectMe* y en segundo lugar los que eran detectados manualmente por el investigador. Para realizar este análisis el planteamiento fue que tanto el número de errores ortográficos como de combinaciones de palabras erróneas detectados por el ojo humano sería superior a los detectados por la herramienta informática. Se trata de un prototipo informático que está aún en fase de desarrollo con capacidad limitada y por otra el número de errores posibles es infinito, por lo que nunca se conseguirá una efectividad absoluta. No obstante, en aras de avanzar con la efectividad de herramientas informáticas que permitan ayudar a los estudiantes a mejorar su producción escrita de forma autónoma es preciso indagar y averiguar el grado de efectividad de estas herramientas en comparación con el ojo humano en diferentes momentos del desarrollo informático para poder establecer así una progresión estadística de los resultados que son posibles en función de: (a) la magnitud del corpus de aprendices analizado; (b) el tamaño del corpus normativo utilizado por la aplicación informática, y; (c) la cantidad de errores posibles que se hayan introducido en la base de datos de *CorrectMe*.

5.1. Análisis de palabras individuales

Por lo que respecta a los errores ortográficos en palabras individuales, se puede apreciar que de un total de 341 errores totales detectados en el corpus de aprendices de español de nivel B1, 194 errores, es decir, un 56'9%, fueron detectados por *CorrectMe*, mientras que 147 errores, el 42'1% no fueron identificados por la aplicación informática. Por lo que respecta al nivel C1, se puede ver un incremento considerable en el número de errores identificados a nivel de la palabra que asciende a 168 errores, es decir, un 68'8% mientras que *CorrectMe* no detectó 76 errores, esto es, un 31'2%. Aunque el resultado de errores detectados supera al número de errores no detectados, diferencia más acuciada en el nivel C1, cabe suponer que es posible que estos errores no sean identificados porque sea necesario ampliar el tamaño del corpus normativo usado como referencia. Una complejidad que a lo largo del presente estudio se ha encontrado afecta a la presencia de flexión

verbal, así como de género y número que se encuentra en español que hace que se reduzcan las posibilidades de que una palabra aparezca en un corpus normativo. En el caso de una lengua como el inglés, por ejemplo, el adjetivo sólo tendrá una forma base sin flexión para género o número (*beautiful*), mientras que en español tendríamos cuatro posibilidades (bonito, bonita, bonitos, bonitas), esto es extensible a mayor escala a la flexión verbal.

	B1		C1	
Errores detectados por <i>CorrectMe</i>	n=194	56'9%	n=168	68'8%
Errores <u>no</u> detectados por <i>CorrectMe</i>	n=147	42'1%	n=76	31'2%
	Total=341		Total=244	

Tabla 5: Número de errores detectados y no detectados por *CorrectMe* en palabras individuales, y sus porcentajes, de acuerdo con cada uno de los dos corpus de aprendices seleccionados, B1 y C1.

5.2. Análisis de bigramas

En el caso de errores contenidos en *bigramas*, cabe destacar que el porcentaje de errores detectados por *CorrectMe* es elevado, concretamente en el nivel B1 se han detectado 119 errores, lo que correspondería a un 81'5% mientras que no se han detectado 27 errores, esto es un 18'5%. En el nivel C1, los resultados son igualmente altos para la detección de errores, situándose en 63 errores, es decir, un 76'8%, mientras que no se han detectado 19 errores que ascenderían a 23'2%.

Los datos proporcionados por el presente estudio resultan reveladores en cuanto a que muestran cómo un número elevado de errores pueden ser detectados de forma automática en cuestión de milésimas de segundos, esto es el 56,9% del nivel B1 y el 68,8% del nivel C1 en el ámbito léxico. Estos errores detectados ascienden hasta el 81,5% del nivel B1 y el 76,8% del nivel C1 en lo que respecta a *bigramas*.

	B1		C1	
Errores detectados por <i>CorrectMe</i>	n=119	81'5%	n=63	76'8%
Errores <u>no</u> detectados por <i>CorrectMe</i>	n=27	18'5%	n=19	23'2%
	Total=146		Total=82	

Tabla 6: Número de errores detectados o no por *CorrectMe* en bigramas, y sus porcentajes, de acuerdo con cada uno de los dos corpus de aprendices seleccionados, B1 y C1.

6. Conclusiones

Los datos proporcionados por este estudio son valiosos para los desarrolladores de *CorrectMe*, en primer lugar, porque permiten revelar la utilidad relativa del uso de la herramienta informática, y en segundo lugar porque esta herramienta consigue encontrar con facilidad errores sobre los que posteriormente se podrá incidir

pedagógicamente desarrollando retroalimentación pedagógica específica para cada error que, además, puede ser codificado informáticamente de modo que se active inmediatamente y sea de utilidad para los aprendices de español.

Sin duda, la herramienta actual también resulta de utilidad para ayudar a los aprendices a centrar su atención en errores presentes en su propia producción escrita a partir de la cual pueden indagar e investigar de qué error se trata, todo ello como primer paso para realizar el procesamiento lingüístico necesario para cambiar las suposiciones erróneas sobre la lengua, y seguidamente dar paso al aprendizaje.

A la vista de la efectividad para la identificación de errores en el nivel de la palabra y de los *bigramas* erróneos, es posible elaborar retroalimentación que ayude a los estudiantes a subsanar y prevenir estos errores en el futuro y codificarlos informáticamente como se ha hecho en otro corrector gramatical diseñado para el aprendizaje del inglés como se indica en Chacón-Beltrán (2017), Lawley (2015; 2016), Senra Silva (2010). Este tipo de retroalimentación proporcionada específicamente para cada error ortográfico y *bigrama* erróneo supone un avance considerable en la metodología de la enseñanza de español dado que permite a los estudiantes trabajar de forma autónoma y recibir retroalimentación pedagógica justo en el momento de escribir su redacción cuando están el pleno procesamiento psicolingüístico de su mensaje que es precisamente cuando cualquier modificación en su interlengua será más efectiva.

Por lo que respecta a la figura del profesor, una herramienta de este tipo también supone una ventaja considerable dado que los estudiantes podrán trabajar de forma autónoma y presentar redacciones para su corrección en las que un porcentaje alto de errores ya habrán sido corregidos y para los que el programa informático probablemente podrá proporcionar una retroalimentación más extensa y personalizada que la del propio profesor. Sin duda, esto entraña un ahorro considerable de tiempo que el profesor podrá dedicar a otros aspectos relacionados con el aprendizaje de la lengua en el entorno del aula (Chacón-Beltrán 2009; Lawley 2009).

7. Implicaciones para investigaciones futuras

La investigación sobre el desarrollo de destrezas escritas en el aprendizaje de segundas lenguas, en este caso español, debe estar encaminada a proporcionar herramientas de aprendizaje que sin necesidad de sustituir al profesor permita a este realizar una labor significativa sobre el aprendizaje de los estudiantes. Así pues, disponer de *software* especializado para aprender una destreza compleja como es la expresión escrita de forma libre se vislumbra como un objetivo prioritario que se puede alcanzar con herramientas que: a) identifiquen errores de forma automática y b) proporcione retroalimentación correctiva de inmediato y sobre cada error.

Aunque la cantidad de errores posible de un aprendiz se vislumbra infinita, es cierto que los aprendices con una primera lengua común muestran tendencia a producir el mismo tipo de errores por transferencia de sus L1. Esto nos lleva a deducir que aunque pueda haber errores típicos ocasionados por una característica lingüística propia de la lengua objeto de estudio, lo más frecuente es que aprendices con una misma L1 tiendan a cometer los mismos errores. Errores que el

profesor en un entorno tradicional corrige una y otra vez a cada estudiante proporcionando probablemente una retroalimentación menos detallada de la que puede proporcionar de forma automática un corrector gramatical.

Una cuestión pendiente que deben analizar los desarrolladores de *CorrectMe* es cómo se incrementan o disminuyen los porcentajes de errores que detecta *CorrectMe* dependiendo del nivel de los estudiantes y si es algo que se puede optimizar con el empleo de un corpus normativo más amplio. Esta es una cuestión de la que has sido recientemente analizada para el inglés en Chacón-Beltrán (2017).

Los resultados del presente estudio muestran una vez más cómo el procesamiento del lenguaje natural y su codificación en términos informáticos no supera al procesamiento realizado por humanos aunque también muestran los resultados que esta diferencia es cada vez menor. Se pueden desarrollar instrumentos informáticos que sustituyan en algunos aspectos la presencia del profesor tradicional, lo que permite al estudiante continuar aprendiendo de forma guiada y ofreciéndole apoyo en actividades de escritura que suelen realizarse de forma individual y solitaria (Rino 2002; Potter 2008; Manchada, 2016; Lawley 2015, 2016). Estudios adicionales que contrasten la labor del profesor y la de herramientas informáticas autónomas son también necesarios.

Referencias

- Athanaselis, T., Bakamidis, S. & Dologlou, I. 2006. An automatic method for revising ill-formed sentences based on N-grams. *Speech Prosody*. ISCA Archive [en línea]. Disponible en: http://www.isca-speech.org/archive/sp2006/papers/sp06_080.pdf.
- Bestgen, Y., & Granger, S. 2014. Quantifying the development of phraseological competence in L2 English writing: An automated approach. *Journal of Second Language Writing*, 26, 28-41.
- Bitchener, J., & Knoch, U. 2010. Raising the linguistic accuracy level of advanced L2 writers with written corrective feedback. *Journal of Second Language Writing*, 19(4), 207-217. <http://dx.doi.org.ezproxy.uned.es/10.1016/j.jslw.2010.10.002>
- Bitchener, J., Young, S., & Cameron, D. 2005. The effect of different types of corrective feedback on ESL student writing. *Journal of Second Language Writing*, 14(3), 191-205. <http://dx.doi.org.ezproxy.uned.es/10.1016/j.jslw.2005.08.001>
- Briscoe, T., Medlock, B. & Andersen, Ø. 2010. *Automated assessment of ESOL free text examinations. Technical Report*. University of Cambridge Computer Laboratory, 790 [en línea]. Disponible en: <http://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-790.pdf>.
- Burston, J. 2001. Exploiting the Potential of a Computer-Based Grammar Checker in Conjunction with Self-Monitoring Strategies with Advanced Level Students of French. *CALICO Journal*, 18/3. doi: 10.1558/cj.v18i3.499-515
- Chacón-Beltrán, R. 2009. Learner autonomy and lifelong learning: Technological solutions in the European Higher Education Area. In M. L. Pérez Cañado, & F. Michavila (Eds.), *English language teaching in the European Credit Transfer System: Facing the challenge* (pp. 187-195) Peter Lang.
- Chacón-Beltrán, R. 2017. Free-form writing: computerized feedback for self-correction. *ELT Journal*, 71/2, 141-149. <https://doi.org/10.1093/elt/ccw064>

- Chandler, J. 2003. The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of Second Language Writing*, 12(3), 267-296. [http://dx.doi.org.ezproxy.uned.es/10.1016/S1060-3743\(03\)00038-9](http://dx.doi.org.ezproxy.uned.es/10.1016/S1060-3743(03)00038-9)
- De Cock, S., Granger, S., Leech, G., & McEnery, T. 1998. An automated approach to the phrasicon of EFL learners. In S. Granger (Ed.), *Learner English on computer* (pp. 67–79). London, UK: Longman.
- Gakis, P., Panagiotakopoulos, C., Sgarbas, K., & Tsalidis, C. 2015. Analysis of lexical ambiguity in modern greek using a computational lexicon. *Digital Scholarship in the Humanities*, 30(1), 20-38. doi:10.1093/lc/fqt035
- Godwin-Jones, R. 2011. Autonomous language learning. *Language Learning and Technology*, 15(3), 4-11.
- Haddon, A. 2009. *Using Word Processing Software to Reduce Common English Errors*. Kwansai Gakuin University humanities review, 13, 73-82.
- Hernández García, F. 2012. 'Palabras problemáticas y frases incorrectas: una solución autónoma para detectar lo indetectable'. *RAEL* 11/1, 41-55.
- Hernández García, F. 2017. 'La detección y corrección de errores en la deixis temporal del verbo en redacciones escritas en inglés como lengua extranjera.' *ELIA*, 17, 183-207. <http://dx.doi.org/10.12795/elia.2017.i17.08>
- Lawley, J. 2004. A preliminary report on a new grammar checker to help students of English as a foreign language. *Arts and Humanities in Higher Education: An International Journal of Theory, Research and Practice*, 3(3), 331-342.
- Lawley, J. 2009. An EFL grammar checker that really works: Making bologna come true. In M. L. Pérez Cañado, & F. Michavila (Eds.), *English Language Teaching in the European Credit Transfer System: Facing the challenge* (pp. 197-204) Peter Lang.
- Lawley, J. 2015. New software to help EFL students self-correct their writing. *Language Learning & Technology*, 19(1), 23-33.
- Lawley, J. 2016. Spelling: Computerised feedback for self-correction. *Computer Assisted Language Learning*, 29(5), 868-880.
- Cestero Mancera, A. M. & Penadés Martínez, I. 2009. *Corpus de textos escritos para el análisis de errores de aprendices de E/ELE, CORANE*. Universidad de Alcalá de Henares.
- Manchanda, B., Athavale, V. A., & kumar Sharma, S. 2016. Various Techniques Used For Grammar Checking. *International Journal of Computer Applications & Information Technology*, 9(1), 177.
- McMartin-Miller, C. 2014. How much feedback is enough?: Instructor practices and student attitudes toward error treatment in second language writing. *Assessing Writing*, 19, 24-35. <http://dx.doi.org/10.1016/j.asw.2013.11.003>
- Nazar, R. & Renau, I. 2012. Google Books N-gram corpus used as a grammar checker. Proceedings of EACL 2012: Second Workshop on Computational Linguistics and Writing. Avignon, France [en línea]. Disponible en: <http://www.aclweb.org/anthology/W12-0304>.
- Parr, J. M., & Timperley, H. S. 2010. Feedback to writing, assessment for teaching and learning and student progress. *Assessing Writing*, 15(2), 68-85. <http://dx.doi.org.ezproxy.uned.es/10.1016/j.asw.2010.05.004>
- Potter, R., & Fuller, D. 2008. My new teaching partner? Using the grammar checker in writing instruction. *English Journal*, 36-41.
- Rino, L. H. M., Di Felippo, A., Pinheiro, G. M. (et al.) 2002. Some Issues on the Development of an Automatic Grammar Checker for Brazilian Portuguese. *Estudos Lingüísticos*, 31. <http://www.gel.org.br/estudoslinguisticos/volumes/31/htm/acomunic.htm>

- San Mateo Valdehíta, A. 2016. A bigram corpus used as a grammar checker for spanish native speakers. *Revista Signos*, 49(90), 94-118.
- San Mateo Valdehíta, A. 2017. Utilidad de un corpus de bigramas empleado como corrector ortográfico y gramatical para el estudiante de español como L2, en *Foreign Language Teaching and Learning Approaches*, Cambridge Scholars Publishing.
- Senra Silva, I. 2010. Designing computer-generated pedagogical feedback for Spanish students of EFL. *RESLA*, 23: 281-296.
- Sinclair, J. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Stevenson, M., & Phakiti, A. 2014. The effects of computer-generated feedback on the quality of writing. *Assessing Writing*, 19, 51-65.
<http://dx.doi.org.ezproxy.uned.es/10.1016/j.asw.2013.11.007>
- Sundeen, T. H. 2014. Instructional rubrics: Effects of presentation options on writing quality. *Assessing Writing*, 21, 74-88.
<http://dx.doi.org.ezproxy.uned.es/10.1016/j.asw.2014.03.003>
- Wu, S. H., & Su, C. Y. 2006. An evaluation of adopting language model as the checker of preposition usage. In *Proceedings of the 18th Conference on Computational Linguistics and Speech Processing* (pp. 369-386). Disponible en: <http://aclweb.org/anthology/O/O06/O06-1023.pdf>