

# New Methodological Approaches to Foreign Language Teaching

Edited by

Laura Torres-Zúñiga  
and Thomas H. Schmidt

Cambridge  
Scholars  
Publishing



New Methodological Approaches to Foreign Language Teaching

Edited by Laura Torres-Zúñiga and Thomas H. Schmidt

This book first published 2017

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

Copyright © 2017 by Laura Torres-Zúñiga, Thomas H. Schmidt  
and contributors

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN (10): 1-4438-8615-7

ISBN (13): 978-1-4438-8615-4

# TABLE OF CONTENTS

LIST OF TABLES.....	ix
LIST OF FIGURES .....	xi
INTRODUCTION.....	1
<b>PART I: APPLIED LINGUISTICS INSIGHTS</b>	
CHAPTER ONE .....	7
<i>Language Acquisition and Teaching of Conceptual Competence in L2: The Case of Motion Events</i>	
Moiken Jessen and Ferrán Suárez	
CHAPTER TWO .....	31
<i>Proposing a Framework to Analyse and Evaluate Textbook Vocabulary Activities in the Foreign Language Classroom</i>	
Nausica Marcos Miguel and Mari Félix Cubas Mora	
CHAPTER THREE .....	53
<i>Cued vs Free Recall for Assessing L2 Vocabulary Acquisition</i>	
François Pichette	
CHAPTER FOUR .....	67
<i>New Perspectives on the Use of Articles in L2 English</i>	
Justyna Leśniewska	
CHAPTER FIVE.....	83
<i>Measuring Communicative Academic Performance</i>	
James Milton and Reem Alsager	
CHAPTER SIX.....	101
<i>Communicative Use of Language in ELT Activities: A Sample Analysis of a Spanish ELT Coursebook from the Implicit/Explicit Perspective</i>	
Gema Alcaraz-Mármol	

CHAPTER SEVEN .....	115
<i>La Situación Actual de la Enseñanza de Lenguas Extranjeras en Edades Tempranas a Nivel Europeo</i>	
Ana Andúgar, María Tornel y Beatriz Cortina-Pérez	
CHAPTER EIGHT .....	129
<i>Videojuegos para la Educación y la Formación: Tipos, Características y Propuesta para la Enseñanza de Terminología Especializada en Lengua Extranjera</i>	
José Ramón Calvo-Ferrer y Ricardo Casañ Pitarch	
<b>PART II: LINGUISTIC PERSPECTIVES</b>	
CHAPTER NINE .....	145
<i>Multilingual System for Automatic Term Extraction</i>	
Carlos Perinián-Pascual and Eva M. Mestre-Mestre	
CHAPTER TEN .....	167
<i>Implementing Technology-Supported Problem-Based Learning in a Context of English for Specific Purposes</i>	
Ana Sevilla-Pavón	
CHAPTER ELEVEN .....	187
<i>*Blush* *Facepalm* *Sigh*: Teaching Pragmatics through the Use of Stage Directions on Twitter</i>	
Carmen Aguilera Carnerero	
CHAPTER TWELVE.....	207
<i>Primeros Pasos Hacia la Profesionalización: Los Encargos Reales en el Aula de Traducción</i>	
Purificación Meseguer Cutillas	
CHAPTER THIRTEEN .....	219
<i>El Uso de Corpus de Conversaciones para la Enseñanza de Aspectos Lingüísticos, Culturales y Coloquiales: Utilidad y Desafíos</i>	
José Manuel Giménez García	
CHAPTER FOURTEEN.....	235
<i>Teaching Business English through Pedagogical Translation: Oral Mediation Activities</i>	
Jorge Soto Almela	

CHAPTER FIFTEEN .....	251
<i>A Corpus-Driven Approach to Emotion Metaphors and its Application to Teaching</i>	
Ángela Almela Sánchez-Lafuente	
CHAPTER SIXTEEN .....	269
<i>Utilidad de un Corpus de Bigramas Empleado como Corrector Ortográfico y Gramatical para el Estudiante de Español como L2</i>	
Alicia San-Mateo-Valdehita	
CHAPTER SEVENTEEN .....	283
<i>La Adquisición de las Competencias del Grado en Traducción e Interpretación y su Adecuación al Mercado Laboral</i>	
Paula Cifuentes Férrez	
 <b>PART III: CULTURAL AND LITERARY CO(N)TEXTS</b>  	
CHAPTER EIGHTEEN .....	299
<i>Students' Language Attitudes towards Peninsular Spanish Varieties: Pedagogical Implications for Spanish as a Foreign Language Materials Selection and Design</i>	
José Andrés Carrasco and Margarita Navarro	
CHAPTER NINETEEN .....	317
<i>Studying Language Contact Phenomena in a Literary Text: Lucky Lady, by Jean-Marc Dalpé</i>	
M <sup>a</sup> Isabel Escalona Fernández	
CHAPTER TWENTY .....	329
<i>Pre-Raphaelitism in the Classroom: A Proposal for Interdisciplinary Analysis and Awareness-Raising</i>	
José María Mesa Villar and Raquel Galián Sabater	
CHAPTER TWENTY-ONE .....	343
<i>Reflections on the Teachings of a Children's Literature Classic: Rudyard Kipling's Mowgli's Story Under Scrutiny by Spanish EFL Trainee Teachers</i>	
Antonio José Miralles Pérez	

CHAPTER TWENTY-TWO .....	361
<i>Enseñanza y Aprendizaje del Francés como Lengua Extranjera a Través de la Práctica Metódica de la Traducción. Ejemplos de un Ejercicio Práctico de Traducción de un Texto Literario</i>	
Antonio José de Vicente-Yagüe Jara e Isabel María Nieto Castejón	
CHAPTER TWENTY-THREE .....	377
<i>Teaching War Poetry through Creative Writing</i>	
Sara Prieto García-Cañedo	
CHAPTER TWENTY-FOUR .....	387
<i>Hacia Nuevas Orientaciones en la Enseñanza de la Literatura en Clase de Lengua Extranjera: la Aportación de la Pedagogía de Proyecto</i>	
Benoît Filhol	
CHAPTER TWENTY-FIVE .....	401
<i>Aproximación al Estudio de la Identidad de la Mujer Española a Través del Cine – una Propuesta Didáctica</i>	
Patricia Coloma Peñate	
Contributors.....	413

## INTRODUCTION

This volume brings together a number of studies by a series of authors in the common field of foreign language teaching that explore how the traditional academic disciplines of Applied Linguistics, Linguistics, Translation, Literature and Cultural Studies can contribute to or be integrated into the teaching of a foreign language by means of innovative methodologies, techniques and instruments used in different language learning and teaching contexts. The chapters deal with the teaching of the four most widely spoken foreign languages in the European Union: English, French, German, and Spanish.<sup>1</sup> It is the aim of this book to give to the foreign language teaching profession an opportunity for the sharing and comparison of strategies across languages as well as across levels, since the studies comprehend teaching contexts in primary, secondary and higher education.

The volume opens with a first section of articles by scholars in Applied Linguistics that share some significant insights and findings in the context of second/foreign language acquisition. After an opening chapter with an innovative approach to teaching German manner verbs through Cognitive Linguistics, the second chapter offers a framework for the evaluation of vocabulary activities in the foreign language classroom. This is followed by an analysis of the effectiveness of cued and free recall techniques when assessing the acquisition of L2 vocabulary. The following chapters offer new perspectives on the teaching of articles in second languages and the measuring of the communicative academic performance of children by considering their Cognitive Academic Language Proficiency. In the next chapter the true communicative use of language in the English as a foreign language classroom is put to the test through the analysis of a Spanish coursebook in terms of implicit versus explicit learning. Then, the subsequent review analyses the current situation of foreign language teaching in primary school levels across Europe in order to compare the projects in different countries. This first section finishes with a vindication of videogames as educational tools and the proposal of their use for the teaching of specialized terminology.

---

<sup>1</sup> European Commission (2012), *Special Eurobarometer 386. Europeans and Their Languages*, p. 19. [http://ec.europa.eu/public\\_opinion/archives/ebs/ebs\\_386\\_en.pdf](http://ec.europa.eu/public_opinion/archives/ebs/ebs_386_en.pdf)

The chapters in the second section of the book adopt a more linguistic perspective and offer the point of view of translation studies too. The first article introduces a new tool to extract specialized lexical units from domain-specific corpora to be used in different languages, which enables teachers and students to check word frequency, concordances and collocations. The next paper demonstrates how the requirements for an English for specific purposes university course for high-achieving students can be met by a new approach implementing problem-based learning with the explicit support of new technological means. New technologies are also the basis for the subsequent study, which from a corpus-based analysis of computer-mediated communication offers a detailed description of the concept of *stage directions* and their potential use for the teaching of pragmatics. This is followed by a study that considers the pedagogical value of commissioning real translation orders to students of translation studies, given all their different requirements and quality demands; this project is accompanied by a qualitative survey of students' opinions on motivation and learning effects. The following chapter proposes the use of a corpus of informal dialogues between students for the teaching of cultural and colloquial aspects in foreign language classrooms, in this case for learners of Spanish language.

While the next chapter advocates for the use of so-called pedagogical translation in oral mediation activities for English for specific purposes, planning a whole program of activities for its use in the classroom, another corpus-driven study recommends the introduction of concordance software into the linguistics classroom with the purpose of analysing and classifying metaphor instantiations in literary texts. Also corpus-based is the following essay, which analyses different automatic text correctors in Spanish and puts forward its own program based on bigrams as a tool for providing useful information, self-correction and self-learning opportunities to the final user. The last article in this second section reviews the competences acquired by students of translation and interpretation studies and their appropriateness for their incorporation to the job market.

In the third part of the book cultural contexts and literary texts are taken into account when teaching language. This section starts with an enquiry on the attitudes of native and non-native speakers of Spanish towards different regional varieties of this language and the possible implications of this bias when creating and selecting teaching materials. The following three chapters introduce visual art and literary works in the foreign language classroom not only to help language learning but as tools to raise students' awareness of cultural and intertextual dynamics and to improve their critical skills; they present introductory classroom-oriented



frameworks for analysis and critical reflection. Literature and translation are combined in the next chapter, which makes use of the translation of literary texts in teaching and learning French as a foreign language with exercises wherein special analysis is done on archaisms and spelling, use of capital letters, syntax and verb use. The next two pedagogical proposals defend both innovative approaches for the integration of the teaching of literature and foreign language teaching: one for French through project-based learning and its different benefits for learners, the other by means of creative writing in English. The volume closes with a last methodological approach for Spanish as a foreign language, which brings cinema into the classroom as the medium to both improve the language knowledge of students and to help them understand and handle different concepts—such as memory, trauma, disobedience, or authority—that are fundamental for the construction of the contemporary Spanish female identity.

Laura Torres-Zúñiga  
Thomas H. Schmidt

## CHAPTER SIXTEEN

# UTILIDAD DE UN CORPUS DE BIGRAMAS EMPLEADO COMO CORRECTOR ORTOGRÁFICO Y GRAMATICAL PARA EL ESTUDIANTE DE ESPAÑOL COMO L2

ALICIA SAN-MATEO-VALDEHÍTA

### **Introducción**

El desarrollo de la expresión escrita del estudiante de L2 está, en gran medida, condicionado por las posibilidades que tenga de recibir corregidos los textos que produce. Con el objetivo de dotar de autonomía en ese sentido al aprendiz de español como L2, se ha desarrollado una herramienta informática, que le permite detectar y corregir los errores ortográficos y gramaticales cometidos en sus textos escritos. Presentamos aquí las posibilidades que ofrece, para la detección y corrección de errores, el análisis estadístico de la frecuencia de las palabras y de los pares de palabras de un corpus de textos escritos en español de cien millones de palabras. Este sistema ha sido empleado en los últimos años por diferentes correctores automáticos, que han venido a complementar a los ya existentes, que precisaban del etiquetado y del análisis morfosintáctico previo para llevar a cabo su función. El funcionamiento de estos nuevos instrumentos es sencillo, como veremos a continuación, y la mayor parte se ha creado para corregir textos escritos en inglés.

El capítulo está dividido en cinco apartados: en el primero presentamos varios correctores de textos que emplean análisis estadísticos; después, expondremos el funcionamiento del corrector para textos escritos en español y, a continuación, cómo se lleva a cabo el proceso de corrección y los tipos de errores que podremos detectar con la información proporcionada. En el cuarto apartado incluimos las limitaciones de la

herramienta y, finalmente, las conclusiones y algunas sugerencias de mejora.

### Correctores automáticos

Además del corrector ortográfico y gramatical del procesador de textos (i. e. Microsoft Word) que, si bien es capaz de señalar un gran número de errores ortotipográficos y de faltas de ortografía, no detecta, entre otros, errores en palabras que llevan tilde diacrítica (*el/él, mi/mí, tu/tú*, etc.) o en palabras con doble acentuación (*gusto/gustó*), ni errores producto de la alteración del orden de grafemas (*acto/cato, al/la, paso/sapo*), o de la conmutación (*pico/piso, jefe/jefa*), adición (*año/paño, mudo/mundo*) o supresión de estos (*alma/ama, cortejar/cotejar*) (Ariza & Tapia, 1997-1998, p. 379-391), el estudiante de español como L2 tiene a su disposición correctores de ortografía, gramática y estilo en línea diseñados específicamente para él. Dos de ellos son Stilus® (Villena, González, González & Muriel, 2002) y SpanishChecker® (Nadasdi & Sinclair, 2001-2015), pero tampoco estos ayudan a localizar y corregir todos los errores.

Por otro lado, encontramos en la bibliografía estudios sobre programas que no se basan en el etiquetado de las palabras ni en el análisis morfosintáctico, como los mencionados, sino que emplean métodos estadísticos para llevar a cabo su función. La mayoría se ha creado para el estudiante de inglés como L2 con el fin de que pueda corregir sus propios textos (Gamon, Leacock, Brockett, Dolan, Gao, Belenko & Klementiev, 2009; García-Heras, 2007; Hernández García, 2012; Islam & Inkpen, 2011; Lawley, 2003, 2015; Lawley & Martin, 2006; Moré, 2006; Naber, 2003); otros programas sirven para corregir redacciones del examen del Title of English as a Foreign Language (TOEFL), como ALEK (Assessing Lexical Knowledge) (Chodorow & Leacock, 2000), o cualquier otro ensayo (Briscoe, Medlock & Andersen, 2010; Yannakoudakis, 2013). También los hay más específicos, como los creados para identificar un error concreto, p. ej., los relacionados con el uso de las preposiciones en inglés (Wu & Su, 2006), o con el orden de palabras dentro de la oración (Athanaselis, Bakamidis & Dologlou, 2006; Lin, Soe & Thein, 2011).

Para detectar errores en textos escritos en español por estudiantes de L2 por medio de análisis estadísticos, solo contamos con la propuesta de Nazar y Renau (2012), que emplea el corpus Google Books *N*-gram; y San-Mateo-Valdehita (2016) analiza una herramienta similar, pero aplicada a la corrección de textos escritos por hablantes nativos de español.

Estos informes, tanto los de los correctores de inglés como los de español, concluyen que estas herramientas que estudian el contexto que

rodea a cada palabra utilizando análisis estadísticos son capaces de detectar errores que pasan desapercibidos para otro tipo de correctores.

A continuación presentaremos el funcionamiento de un programa de esas características y comprobaremos las ventajas que ofrece al estudiante de español como L2.

### Funcionamiento del corrector *Correct Me*

La herramienta, desarrollada por la Universidad Nacional de Educación a Distancia (España), que ayuda a la detección y corrección de errores ortográficos y gramaticales en textos escritos en español, cuenta con una base de datos procedente de un corpus de cien millones de palabras que incluye extractos de novelas, cuentos y ensayos, escritos en español desde 1980 hasta 2012, y noticias y artículos de opinión publicados en periódicos nacionales y revistas de actualidad durante 2012. No contiene poemas, obras de teatro ni novelas dialogadas, con el fin de que no aparezcan expresiones propias de la lengua oral que no son apropiadas en la escrita en todos los casos. El corpus ha sido elaborado por un equipo de profesores de la UNED a partir de textos disponibles en formato digital y provenientes de los diferentes países de habla española; sobre todo se han seleccionado aquellos que no incluyen regionalismos ni expresiones alejadas del nivel formal de la lengua. Por lo tanto, para seleccionar los textos se ha tendido en cuenta tanto el criterio cronológico como el geográfico, y el tipo de texto y el nivel de formalidad, con el fin de que fuera un corpus sincrónico, que reflejara el uso actual de la lengua en los diferentes lugares en los que se habla, cuando se trata de temas no excesivamente especializados en el ámbito formal.

El programa almacena las frecuencias de aparición en el corpus de cada palabra y de cada par de palabras (“bigrama” o “digrama”; en inglés, *bigram*), y crea una base de datos con toda esa información. También calcula el número de veces que se esperaría encontrar cada bigrama en el corpus si las palabras aparecieran aleatoriamente (es decir, halla la probabilidad del par) (Sinclair, 1991), tal y como refleja la fórmula siguiente (Figura 16-1), donde  $P$  es la probabilidad;  $a$  y  $b$  representan, respectivamente, a la palabra 1 y a la palabra 2;  $T$  es el número total de palabras del corpus y  $F$ , la frecuencia.

$$P(ab) = \frac{T}{\left(\frac{T}{F(a)}\right) \times \left(\frac{T}{F(b)}\right)}$$

Figura 16-1. Fórmula para calcular la probabilidad del bigrama.

Para detectar los errores de un texto, el *software* primero comprueba la ortografía de las palabras utilizadas y, después, divide el texto en bigramas y analiza si estos aparecen en el corpus más (o menos) veces de lo que sería esperable según su probabilidad; es decir, calcula el umbral (U), mediante esta fórmula:  $U = F(ab) / P(ab)$ .

Así, por ejemplo, si un estudiante escribe “\*La contaminación es la problema fundamental”, el programa detectará que hay un error ortográfico en el sustantivo “contaminación”. Una vez corregido, el siguiente paso será dividir la oración en pares como muestra la Tabla 16-1:

**Tabla 16-1. Los cinco bigramas de la oración “\*La contaminación es la problema fundamental”.**

1	La contaminación
2	contaminación es
3	es la
4	la problema
5	problema fundamental

En la Tabla 16-2 se muestran los resultados del análisis de cada uno de los bigramas de la oración propuesta.

**Tabla 16-2. Análisis de la oración “\*La contaminación es la problema fundamental”.**

Par	F par	F palabra 1	F palabra 2	P par	Umbral
la contaminación	194	3333903	375	12.5	15.52
contaminación es	3	375	544084	2.04	1.47
es la	30430	544084	3333903	18139.23	1.68
la problema	0	3333903	14002	466.81	0
problema fundamental	20	14002	1828	0.26	76.92

Comprobamos que el par “la contaminación” aparece 194 veces en el corpus de cien millones de palabras; “la”, 3333903 y el sustantivo “contaminación”, 375. Si usáramos las palabras al azar, teniendo en cuenta la frecuencia de las dos palabras, el par “la contaminación” aparecería 12.5 veces, pero, como acabamos de indicar, realmente lo encontramos 194; es

decir, es 15.52 veces más frecuente de lo esperado (ese es el umbral). Esta última cifra confirma nuestra hipótesis sobre si los hablantes nativos normalmente utilizan juntas o no dos palabras. Si el umbral es mayor de 1, podemos inferir que dos palabras suelen utilizarse juntas muchas veces.

Por otro lado, en el bigrama “\*la problema” encontramos cifras muy diferentes. En primer lugar, observamos que esa combinación no se registra nunca en el corpus, mientras que “la” lo hace más de tres millones de veces (3333903) y el sustantivo “problema”, 14002. De manera que la probabilidad de que esas dos palabras aparezcan juntas es de 466.81. El hecho de que esta combinación no se dé nunca, aunque la probabilidad sea alta, quiere decir que estas dos palabras se rechazan entre sí, y, por lo tanto, nuestra hipótesis sobre su posible combinación no se confirma.

El estudiante antes de tener acceso a esta información detallada sobre la frecuencia de las palabras recibe su texto con las combinaciones menos frecuentes, y cuyos umbrales son también bajos, resaltadas en diferentes colores.

El color rojo indica que existe un 90% de posibilidades de que haya un error.

El color naranja significa que hay un 50% de posibilidades de que ese par de palabras contenga un error.

El color amarillo revela que existe un 25% de posibilidades de error.

Así, por ejemplo, la oración con la que estamos trabajando aparecería marcada en rojo en la Figura 16-2.

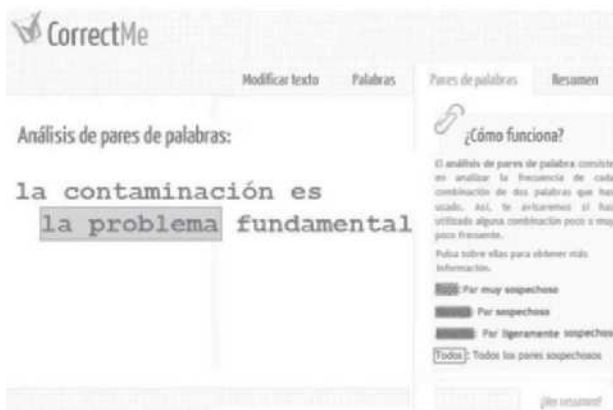


Figura 16-2. Presentación de los bigramas menos frecuentes.

Si se pulsa sobre el bigrama en cuestión (“\*la problema”), la herramienta indica que esa es una combinación muy poco frecuente: no se ha registrado nunca cuando estadísticamente tendría que aparecer al menos 466.81 veces en un corpus de cien millones de palabras.

### Corrección y tipos de errores

Detectar el error no es el objetivo final, sino solo el primer paso para corregirlo. El programa aporta la información que puede ayudar al estudiante de español como L2, pero no le da la solución directamente; de ahí que el usuario deba contar con, al menos, un nivel B1, según el *MCER* (Consejo de Europa, 2001), que le permita reflexionar sobre el funcionamiento de la lengua que estudia y llevar a cabo las inferencias correctas.

Uno de los errores más frecuentes que comete el aprendiz de español como L2 tiene que ver con la asignación de género a los sustantivos y la consecuente concordancia con adjetivos, artículos, demostrativos, etc. Así, por ejemplo, si escribe “\*la problema de toda la viaje”, el corrector indicará que los pares “la problema” y “la viaje” no aparecen en el corpus y ofrecerá la posibilidad de consultar las palabras que con más frecuencia aparecen junto a las que forman esos bigramas. De esta manera, podrá comprobar que son “el” y “un” en ambos casos (véanse las Tablas 16-3 y 16-4), con lo cual, tendrá argumentos sólidos que demuestran que les corresponde género masculino.

**Tabla 16-3. Palabras que preceden a “problema” por orden de frecuencia.**

Combinación	Frecuencia
el <i>problema</i>	5042
un <i>problema</i>	2407
ningún <i>problema</i>	792
del <i>problema</i>	593
algún <i>problema</i>	498

**Tabla 16-4. Palabras que preceden a “viaje” por orden de frecuencia.**

Combinación	Frecuencia
<i>el viaje</i>	4532
<i>un viaje</i>	2062
<i>de viaje</i>	2000
<i>su viaje</i>	1366
<i>del viaje</i>	1318
<i>este viaje</i>	649

También son muy comunes los errores relacionados con el empleo de las preposiciones. Si la preposición viene regida por un verbo o un sustantivo, la información de la frecuencia será determinante para detectar el error y corregirlo. Por ejemplo, si escribimos “\*La economía depende en el turismo”, el corrector nos indicará que el par “depende en” no es muy frecuente y si consultamos las combinaciones más frecuentes (véase la Tabla 16-5) comprobaremos que son: “depende de” y “depende del”, y “depende en” no aparece.

**Tabla 16-5. Palabras que suceden a “depende” por orden de frecuencia.**

Combinación	Frecuencia
<i>depende de</i>	1631
<i>depende del</i>	267

Otro error detectable es la omisión de una palabra, p. ej., si escribimos “\*Ayer por la noche estuvieron jugando ajedrez”, descubriremos que el bigrama “jugando ajedrez” no se registra en el corpus; y si consultamos las combinaciones más frecuentes (véanse las Tablas 16-6 y 16-7) en las que aparecen estos dos vocablos, comprobaremos que la única que coincide es “al”, con lo cual, es una buena pista para detectar la palabra omitida.

**Tabla 16-6. Palabras que suceden a “jugando” por orden de frecuencia.**

Combinación	Frecuencia
<i>jugando con</i>	659
<i>jugando a</i>	456
<i>jugando al</i>	296
<i>jugando en</i>	117



**Tabla 16-7. Palabras que preceden a “ajedrez” por orden de frecuencia.**

Combinación	Frecuencia
de <i>ajedrez</i>	550
al <i>ajedrez</i>	229
el <i>ajedrez</i>	126
del <i>ajedrez</i>	47
un <i>ajedrez</i>	18

En otras ocasiones lo que ocurre es que el estudiante incluye una palabra innecesaria. Un ejemplo lo tenemos en el caso de los anglohablantes que convierten *another* en “\*un otro”, y escriben “\*Se fue de compras y vino con un otro bolso nuevo”. En las tablas de combinaciones 16-8 y 16-9 comprobamos que el artículo determinado (“el”) podría preceder al adjetivo “otro” (si es que el destinatario conociera la referencia), pero no el indeterminado (“un”).

**Tabla 16-8. Palabras que suceden a “un” ordenadas por orden de frecuencia.**

Combinación	Frecuencia
<i>un</i> hombre	40989
<i>un</i> poco	36583
<i>un</i> momento	24183
<i>un</i> gran	12905
<i>un</i> día	12307
<i>un</i> lado	11615

**Tabla 16-9. Palabras que suceden a “un” ordenadas por orden de frecuencia.**

Combinación	Frecuencia
el <i>otro</i>	15825
al <i>otro</i>	11142
a <i>otro</i>	9943
de <i>otro</i>	6605
en <i>otro</i>	4853
del <i>otro</i>	4844

La gran ventaja de la herramienta es que le muestra al aprendiz cómo tienden a combinarse en español las palabras; de ahí que, no solo corregirá sus errores, sino que se expresará de manera más natural. Por ejemplo, podrá saber qué adjetivos caracterizan generalmente a un sustantivo; p. ej., los bigramas más frecuentes que incluyen “hombre” en primera posición son los que aparecen en la Tabla 16-10.

**Tabla 16-10. Adjetivos que caracterizan a “hombre” ordenados por orden de frecuencia.**

Combinación	Frecuencia
<i>hombre alto</i>	991
<i>hombre joven</i>	414
<i>hombre honrado</i>	339
<i>hombre corpulento</i>	303
<i>hombre rico</i>	273
<i>hombre gordo</i>	239
<i>hombre bueno</i>	233
<i>hombre bajo</i>	225
<i>hombre fuerte</i>	207
<i>hombre mayor</i>	207
<i>hombre viejo</i>	206
<i>hombre capaz</i>	204
<i>hombre delgado</i>	195

Si el aprendiz tiene en cuenta estos datos, no utilizará, p. ej., la combinación “\*hombre largo”, pues no es la que los nativos prefieren, y el análisis, probablemente, le llevará a intuir que ese adjetivo no es el correcto, pudiendo elegir de entre la lista propuesta el que mejor se adecue a lo que quiere expresar.

Además de los errores mencionados hasta ahora, el corrector también señala otros que no son exclusivos del estudiante de L2 (véase un análisis de la eficacia del corrector en San-Mateo-Valdehita, 2016), como estos que simplemente enumeramos a continuación.

- (1) Errores gramaticales u ortográficos que el corrector del procesador de textos no detecta, como la tilde en la conjunción disyuntiva *o*.
- (2) Confusiones entre palabras parónimas, bien homófonas, como *ha/a*, *ira/ ir a*, *tubo/tuvo*, *vaya/valla*; bien una con tilde diacrítica como, p. ej., *té/te*, *tú/tu*; o palabras con diferente sílaba tónica, como

*hacia/hacia, ingles/inglés, ira/irá, llegara/llegará, mama/mamá, papa/papá, paro/paró, perdida/pérdida, porque/porqué.*

- (3) Erratas provocadas por la rapidez de la escritura como, p. ej., la alteración del orden de las letras (*pero/peor, se/es, seto/esto*); o la confusión entre dos letras cuyas teclas están próximas (*piso/puso, rana/rama, rosa/risa, rusa/risa, su/si, ti/tu, tono/tomo, tonta/tinta, tubo/tuvo, vello/bello, vienes/bienes, votar/botar*); o sin estarlo: *celo/pelo, coma/cama, cuerpo/cuerno, esos/osos, especial/espacial, hoy/hay, imponente/impotente, lata/pata, mejor/menor, mor/mar, nota/rota, palo/malo, par/mar, par/por, pase/base, peso/puso/poso, pino/piño, ser/sed, tal/tan, tímpano/témpano, toca/roca, todo/topo, tubo/hubo, vagón/varón, vez/voz, vocaciones/vacaciones.*
- (4) Olvido e inclusión de alguna letra de forma incorrecta, como, p. ej., en *la/las, lee/leer, levantaos/levantaros, lo/los, redactor/reactor, tiempo/tempo, toldo/todo, veinte/vente.*

## Limitaciones

La limitación principal del programa se deriva de su propio funcionamiento: solo se analizan pares de palabras adyacentes, por lo que, no se obtiene información sobre ningún fenómeno que no se circunscribe a dicho par. De ahí que, cuando los elementos que deben concordar, ya sea el sujeto y el verbo, o un sustantivo y un adjetivo, no sean contiguos, el corrector no siempre detectará la falta de concordancia, como la de “\*No vemos a la gente que están esperando”; ni tampoco errores relacionados con la correlación verbal, como “\*Nos deseó que tengamos suerte al día siguiente”, o con los tiempos verbales: “\*Mañana Celia no fue a clase”.

Estamos ante casos que en estadística se conocen como errores de tipo II o falsos negativos; es decir, existe un error, pero el programa no lo detecta, ya que la combinación errónea está más allá de los límites del par que él analiza. Esta misma limitación ha sido señalada por Burston (1996) en su análisis de varios correctores de francés, y por Chen (2009) en el de correctores de inglés.

Por otro lado, también es posible el caso contrario: no existe el error, pero la herramienta señala una combinación porque no aparece en el corpus. Es un falso positivo o error de tipo I (también llamado “falsa alarma”). Esto ocurre, p. ej., si una de las dos palabras es de baja frecuencia.

En definitiva, el estudiante debe tener en cuenta que el programa solo ofrece información sobre dos palabras contiguas, y debe estudiar la

información sobre la frecuencia de cada vocablo para así no sacar conclusiones que lo lleven a cometer errores.

## Conclusiones

Este corrector ortográfico y gramatical, que emplea información sobre la frecuencia de las palabras y de las combinaciones de dos palabras proveniente de un corpus de cien millones de vocablos, al igual que otras propuestas destinadas a revisar textos escritos en inglés, ayuda al estudiante de español como L2 a corregir sus propios ejercicios escritos, y además le ofrece información sobre cuáles son las combinaciones de palabras más frecuentes en español. No es una herramienta que sustituya al corrector del procesador de textos (que, en vez de utilizar la estadística, emplea un sistema de análisis y etiquetado de la lengua), sino que ambos programas son complementarios, ya que detectan diferentes tipos de errores.

La ventaja de este programa es que es una herramienta de trabajo y de aprendizaje para el estudiante de L2, de ahí su utilidad didáctica. El aprendiz puede confirmar sus hipótesis lingüísticas y contrastarlas con las combinaciones de palabras que efectivamente han utilizado los hablantes nativos de español. En ningún momento el *software* le proporcionará una solución, sino que tendrá que reflexionar por sí mismo, analizar la información proporcionada y llegar a su propia conclusión. Pensamos que este proceso lo ayudará a evitar cometer el error en el futuro y no solo a corregirlo en ese momento.

Por otro lado, hemos comprobado que el corrector no es infalible y que podría mejorarse de varias formas. Una de ellas, sería ampliando la extensión de la combinación de palabras analizadas; es decir, que en vez de ser solo pares, fueran grupos de tres, cuatro o, incluso, cinco palabras; para poder así detectar algunos de los errores que no pueden ser descubiertos solo con el análisis de dos palabras adyacentes (Wu & Su, 2006). También se podría ampliar el tamaño del corpus de referencia, para así reducir los falsos positivos debidos a palabras de baja frecuencia.

Sin duda, la mejora que nos parece más útil sería que el programa proporcionara algún tipo de retroalimentación que orientara al estudiante para corregir, al menos, los errores más comunes; si bien somos conscientes de la dificultad que conlleva.

Este *software* todavía está en fase de prueba y, en definitiva, hay que corroborar con los usuarios, aprendices de español L2 de nivel intermedio (B1, al menos), que realmente es una herramienta útil, que les permite desarrollar y practicar la expresión escrita tanto como deseen (sin estar

asociada necesariamente a las actividades propuestas por el docente, que luego deberá evaluar) y ser ellos mismos sus propios correctores, lo cual conllevará una mejora y un mayor progreso en el conocimiento de la lengua.

### Referencias bibliográficas

- Ariza García, A. & Tapia Poyato, A. M. (1997-1998). El corrector ortográfico y la presentación del texto escrito. *Cauce. Revista de Filología y su Didáctica*, 20-21, 375-412.
- Athanaselis, T., Bakamidis, S. & Dologlou, I. (2006). An automatic method for revising ill-formed sentences based on *N-grams*. *Speech Prosody. ISCA Archive*. Recuperado de: [http://www.isca-speech.org/archive/sp2006/papers/sp06\\_080.pdf](http://www.isca-speech.org/archive/sp2006/papers/sp06_080.pdf).
- Briscoe, T., Medlock, B., & Andersen, Ø. (2010). Automated assessment of ESOL free text examinations. *Technical Report*. University of Cambridge Computer Laboratory, 790. Recuperado de: <http://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-790.pdf>.
- Burston, J. (1996). A comparative evaluation of French grammar checkers. *CALICO Journal*, 13(2-3), 104-111.
- Chen, H.-J. H. (2009). Evaluating two web-based grammar checkers - Microsoft ESL Assistant and NTNU Statistical Grammar Checker. *Computational Linguistics and Chinese Language Processing*, 14(2), 161-180.
- Chodorow, M. & Leacock, C. (2000). An unsupervised method for detecting grammatical errors. *Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference*, 140-147. Association for Computational Linguistics. Recuperado de: <http://acl-arc.comp.nus.edu.sg/archives/acl-arc-090501d3/data/pdf/anthology-PDF/A/A00/A00-2019.pdf>.
- Consejo de Europa (2001). *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza y evaluación (MCER)*. Estrasburgo. Recuperado de: <http://www.cvc.cervantes.es/obref/marco/indice.htm>.
- Gamon, M., Leacock, C., Brockett, C., Dolan, W. B., Gao, J., Belenko, D., & Klementiev, A. (2009). Using statistical techniques and web search to correct ESL errors. *CALICO Journal*, 26(3), 491-511.
- García-Heras Muñoz, A. (2007). Programas informáticos de corrección gramatical en el aprendizaje de una lengua extranjera (inglés): expresión escrita. *Docencia e Investigación: Revista de la Escuela Universitaria de Magisterio de Toledo*, 17, 71-101.
- Hernández García, F. (2012). Palabras problemáticas y frases incorrectas:

- una solución autónoma para detectar lo indetectable. *Revista electrónica de lingüística aplicada*, 1, 41-55.
- Islam, A. & Inkpen, D. (2011). Correcting different types of errors in texts. En C. Butz & P. Lingras (Eds.), *Advances in Artificial Intelligence* (pp. 192-203). Berlín-Heidelberg: Springer.
- Lawley, J. (2003). The development of a grammar checker for Spanish secondary students of English as a foreign language. *Revista española de lingüística aplicada*, 16, 127-138.
- . (2015). New software to help EFL students self-correct their writing. *Language Learning & Technology*, 19(1), 23-33.
- Lawley, J. & Martin, R. (2006). Corrector de gramática para estudiantes autodidactas de inglés como lengua extranjera. *Revista de Educación*, 340, 1171-1191.
- Lin, N. Y., Soe, K. M. & Thein, N. L. (2011). Developing a chunk-based grammar checker for translated English sentences. *25th Pacific Asia Conference on Language, Information and Computation*, 245-254. Recuperado de:  
<http://www.aclweb.org/anthology/Y11-1026>.
- Moré, J. (2006). A grammar checker based on web searching. *Digithum*, 8, 1-5. Recuperado de: <http://www.uoc.edu/digithum/8/dt/eng/more.pdf>.
- Nadasdi, T. & Sinclair, S. (2001-2015), *SpanishChecker.com. Corrector de ortografía y gramática*. Nadaclair Language Technologies. Recuperado de: <http://spanishchecker.com/>.
- Naber, D. (2003). *A rule-based style and grammar checker*. Tesis doctoral, Universidad de Bielefeld, Bielefeld, Alemania. Recuperado de:  
[http://www.danielnaber.de/language-tool/download/style\\_and\\_grammar\\_checker.pdf](http://www.danielnaber.de/language-tool/download/style_and_grammar_checker.pdf).
- 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. Recuperado de: <http://www.aclweb.org/anthology/W12-0304>.
- San-Mateo-Valdehita, A. (2016). Un corpus de bigramas utilizado como corrector ortográfico y gramatical destinado a hablantes nativos de español. *Revista Signos. Estudios de Lingüística*, 49(90), 94-118.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Villena, J., González, B., González, J.C., & Muriel, M. (2002). STILUS: Sistema de revisión lingüística de textos en castellano. *Iberamia. VIII Conferencia Iberoamericana de Inteligencia Artificial*. Recuperado de: <http://www.lsi.us.es/iberamia2002/confman/SUBMISSIONS/185-ulaeddlle.PDF>.

- Wu, S.-H. & Su, C.-Y. (2006). An evaluation of adopting language model as the checker of preposition usage. *Proceedings of the 18th Conference on Computational Linguistics and Speech Processing, ROCLING 2006, Taiwan*. Association for Computational Linguistics and Chinese Language Processing (ACLCLP). Recuperado de: <http://aclweb.org/anthology/O/O06/O06-1023.pdf>.
- Yannakoudakis, H. (2013). Automated assessment of English-learner writing. *Technical Report*. University of Cambridge Computer Laboratory, 842. Recuperado de: <http://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-842.pdf>.