

The Development and Evaluation of a Corpus-based Spanish Collocation Error Detection and Revision Suggestion Tool

Desarrollo y evaluación de una herramienta basada en corpus para identificar errores y realizar propuestas de corrección en las colocaciones del español

Hui-Chuan Lu
National Cheng Kung University
huichuanlu1@gmail.com

An Chung Cheng
University of Toledo
ACHeng@utnet.utoledo.edu

Shujuan Wang
University of Toledo
shujuan.wang@rockets.utoledo.edu

Resumen

Durante las últimas tres décadas, el estudio de las colocaciones ha sido uno de los polos de atención por parte de las personas interesadas en el léxico, tanto en la lingüística teórica como en la aplicada. Nuestro equipo de investigación ha desarrollado y evaluado una herramienta basada en corpus para la asistencia al aprendizaje de las colocaciones en español. Basada en dos corpus (CEATE y CPEIC) y en una herramienta de extracción de colocaciones en español (HCE), esta herramienta de aprendizaje asistido por ordenador, SpColEDRS, facilita la identificación de errores en las colocaciones del español y, además, sugiere correcciones. Los resultados de la evaluación indican que la herramienta desarrollada en esta investigación puede ayudar eficazmente a los estudiantes, en especial, a los principiantes. Así mismo, los resultados de la encuesta de satisfacción confirman la utilidad de esta herramienta en el aprendizaje asistido de las colocaciones en español. Por último, este estudio arroja luz sobre las aplicaciones pedagógicas de los corpus elaborados y el aprendizaje de colocaciones en español con un enfoque basado en corpus y en un entorno de adquisición multilingüe.

Keywords

colocaciones, identificación de errores, sugerencias de corrección, español

Abstract

The topic of collocation has drawn attention for the past three decades in the lexical area of theoretical and applied linguistics. Our research team developed and evaluated a corpus-based assisted tool for collocation learning in Spanish. Based on the two constructed corpora (CEATE and CPEIC) and a Spanish collo-

cation extraction tool (HCE), this computer-assisted learning tool, SpColEDRS, is easy to use to detect errors and also suggests revisions for Spanish collocations. Based on the evaluation, the research results indicated that the tool developed in this research can assist learners effectively, especially in the case of beginners. In addition, the results of the satisfaction survey provided positive confirmation of the effectiveness of this tool in assisting the learning of Spanish collocations. Finally, this study shed light on pedagogical applications of the constructed corpora and the learning of Spanish collocation with a corpus-based approach in a multilingual acquisition setting.

Keywords

collocation, error, detection, revision, suggestion, Spanish

1 Introduction

The topic of collocation has drawn special attention for the past thirty years in the lexical area of theoretical and applied linguistics. Wray (2000), Nattinger & DeCarrico (1992), Sinclair (1991), and Firth (1957) all indicated the importance of collocation for learning foreign languages. Many researchers have tried to define and describe collocations, but there is no one simple, precise definition of collocations. In corpus linguistics, “collocation” is defined as a group of words that occur more frequently than would be expected by chance (McKeown & Radev, 2000). For example, they comprise word combinations such as “boiling hot,” that is, phrases that are more restricted than free combinations (“very hot”) and less restricted than idioms (“get hot under the



collar”). Correct uses of collocations could be an indication of a learner’s knowledge of phrases or common combinations in the target language L2 learners because beginning learners often are not aware of the important role of collocation since they tend to focus on the learning of new words and grammatical points. Correct usage of collocation can be an obstacle even for advanced learners (Källkvist, 1995; Granger, 1998; Lorenz, 1999; Nesselhauf, 2003, 2005). In the field, compared with English collocation learning and teaching, there are fewer available tools intended to assist with learning Spanish collocations than there are for learning English (Weisser, 2016). Therefore, the purpose of this study is to further the application of previously constructed corpora and tools by developing a tool intended to assist with learning Spanish collocation. Based on a general detection and revision tool (System of Error Detection and Revision Suggestion, SEDRS) developed in 2013 (Lu et al., 2013), the corpora and the tool contained in this research comprise a learners’ corpus CEATE (Corpus Escrito de Aprendices Taiwanese de Español / Taiwanese Learners’ Written Corpus of Spanish), a parallel trilingual corpus CPEIC (Corpus Paralelo de Español, Inglés y Chino / Parallel Corpus of Spanish, English and Chinese), and a Spanish collocation extraction tool, HCE (Herramienta de Colocaciones Españolas / Spanish Collocation Tool).

The primary tasks of this study include two areas of focus. The first one deals with the development of a corpus-based tool for assisting with learning Spanish collocation (a system of Spanish collocation error detection and revision suggestions), which can identify collocation errors and make correction suggestions. The second area is an empirical evaluation of the effectiveness of the developed Spanish collocation learning tool. The following are the research questions that guided the assessment of the functions of the assisted learning tool and system. (1) Are there any significant differences between the experimental and control groups after a pedagogical intervention using the corpus-based learning tool? (2) Are there any significant differences between beginning and intermediate learners in the post-test after using the learning tool?

2 Previous research

2.1 Collocation assisted learning tools

In order to obtain a general view of computer-assisted collocation learning tool, we evaluated

eight existing tools used to learn English collocations and five tools used to learn Spanish collocations before we developed our computer-assisted collocation learning tool for Spanish. A summary of the features and disadvantages of each tool is provided below.

With regard to English assisted learning tools, a POS (part of speech) search is not available in the Hong Kong Polytechnic Web Concor-dancer (Greaves, 1999), while in TANGO (Jian et al., 2004), the POS of a keyword can be defined, and example and frequency are also provided, but types of POS are limited. In WebCollocate (Chen, 2011), a POS search is available, and search results are sorted by frequency, with a user-friendly interface and the provision of related sentences, but it is not currently available for public use. In addition, there are also collocation assisted learning tools for bilingual uses such as TOTALrecall (Wu et al., 2003), which can be searched in both Chinese and English. In Writing Assistant (Chang et al., 2008), user mistakes can be revised with the correct collocation based on Chinese-English translations. Furthermore, English collocation assisted learning tools provide customized search functions such as the Corpora and NLP for Digital Learning of English, CANDLE (Liou et al., 2006), which consists of three sub-systems tailored to different user levels; the Writing-Collocation Checker, which can automatically detect “verb + noun” collocations and provide correct collocations for users, and Linggle (Boisson et al., 2013), which has selective preference and synonym group functions and provides different types of arguments based on the predicate.

With respect to Spanish collocation assisted learning tools, CrossLexica Española (Bolshakov & Miranda-Jiménez, 2004) is a Spanish collocation assisted learning tool with a POS search function, and it is probability-based, with a grammatical function and semantic classification available, but it is not available for public use. The Corpus del Español, CdE (Davies, 2012) is more advanced, with lemmas functions, and is user friendly, but it provides too many examples and may be difficult for beginners to use. Diccionario de Colocaciones del Español, DiCE (Alonso Ramos et al., 2010) is free for users and provides general and advanced functions for searching for collocations through lexical lemma entities on specific themes, such as emotion nouns (for example, alegría “joy” and estima “esteem”) with semantic “feeling” and “mental actions” features such as “sentir una gran alegría” or “alta estima”. Syntactical structures, meaning identi-

fication, explanations, and examples associated with a list of lexical units are included to illustrate the searched collocations. The DiCE is a powerful online dictionary in terms of providing lexical information, but only a Spanish interface is available, so its high-level collocation might be difficult to understand for learners with limited proficiency in Spanish.

In addition, Sketch Engine (Kilgarriff et al., 2014) is a Spanish collocation assisted learning tool with multilingual search functions. It can select different statistical methods according to language features, but the statistical results are relatively complicated. Finally, EuroWordNet (González-Agirre & Rigau, 2013) includes a variety of European languages such as English, Spanish, and Italian, but the search results are research-oriented and might be too advanced and complicated for foreign language learners to understand and apply, especially in the case of those who are at the beginning and intermediate levels.

2.2 Evaluation of collocation assisted learning tools

In a review of the studies related to an evaluation of the developed assisted tools, it was found that Chen (2011) investigated the relative effectiveness of several computer assisted English collocation tools focusing on two groups of users, learners and teachers of English. Students from two similar classes used different tools to translate sentences from Chinese into English, whereas English teachers assessed four English collocation learning tools. The results showed that students who used WebCollocate (the developed assisted learning tool in English by Chen (2011)) performed better than those who used the other tool, Hong Kong Polytechnic Web Concordancer. Language teachers reported that using WebCollocate was less time consuming and that it was easier to search for collocations and to find many collocation examples because of the large database in the corpus.

With respect to Spanish collocation tools, Vincze et al. (2011) extended a series of collocation-related analyses based on DiCE (Diccionario de Colocaciones del Español) to studies of computer-assisted language learning; the authors utilized CEDEL2, an L1 English-L2 Spanish learner corpus (Lozano, 2009), to develop a computer-assisted learning tool for Spanish collocations. Alonso Ramos et al. (2010) annotated both the correct and incorrect collocations in the learner corpus to find collocations undetected by auto-correction tools with an analysis of er-

ror features, so as to improve the error-detection function of the collocation learning tool. Their analysis of collocation errors included recognizing collocations, correction judgment and interpretation of errors. Ferraro et al. (2014) pointed out that there are only a few tools that provide users with high accuracy and proper corrections, and most tools only offer a list of collocation options for users to choose from. For the detection of incorrect collocations, Ferraro et al. (2014) employed frequency-based techniques and attempted to provide users with proper corrections rather than simply listing all the possible corrections. They argued that although ordered lists might be helpful for advanced learners, the tool would not be as beneficial for learners at the elementary and intermediate levels, especially when the suggested lists include words with subtle semantic differences that are difficult to distinguish one from the other.

2.3 Acquisition of Spanish collocation

Among the available research on the acquisition of Spanish collocation, Laufer & Waldman (2011) found that learners at different proficiency levels used fewer collocations than native speakers. Previous studies also showed that collocation causes various degrees of difficulty for learners from beginning to advanced levels in the lexical learning process. With regard to different types of collocations, previous research (Laufer & Waldman, 2011; Nesselhauf, 2003; Alfahadi et al., 2014) has concentrated more on the adjective-noun (AdjN) and the verb-noun (VN) constructions, which are considered more problematic for learners. Going one step further, Lu & Cheng (2016) compared and contrasted four different essential types of Spanish combinations, VN, AdjN, NAdj, and VP in learner and parallel corpora. The results showed a sequence of development from NAdj, VN, to AdjN combinations. The results also suggested that most learner errors were related to the learners' L1 (Chinese) and L2 (English). Furthermore, lexical errors might be associated with the form-meaning transfer from the previous languages of learners.

As in the aforementioned learning tools for Spanish collocations intended to extend related studies, in this research, built upon previously constructed corpora, a computer-assisted learning system was developed with two major functions, error detection and revision suggestions for Spanish collocation, and an experiment was conducted in order to evaluate its effectiveness in terms of learning.

3 Research method

The methodology involved in this study included two major parts. The first one was the development of a corpus-based learning tool for Spanish collocation, and the second part was an evaluation of the developed learning tool. Based on the previous development experience using SEDRS (System of Error Detection and Revision Suggestion), the construction of the Spanish Collocation Error Detection and Revision Suggestion tool (SpColEDRS) involved the employment of data sources from two corpora (the Corpus Escrito de Aprendices Españoles / Learners' Written Corpus of Spanish, CEATE and the Corpus Paralelo de Español, Inglés y Chino / Parallel Corpus of Spanish, English and Chinese, CPEIC) and a data analysis and collocation extraction tool (Herramienta de Colocación Española / Spanish Collocation Tool, HCE). After developing the computer-assisted learning tool, SpColEDRS, with two major functions (error detection and revision suggestions), an experiment was conducted and a questionnaire was used to evaluate its effectiveness for checking Spanish collocations from the perspective of learners.

3.1 The development of a computer-assisted learning tool: SpColEDRS

The first part of this section addressed the development of the assisted learning tool, the Spanish Collocation Error Detection and Revision Suggestion (SpColEDRS). Texts were analyzed and processed using the POS tagging system, and then collocations were calculated and extracted as outputs through the Spanish collocation tool (HCE) search functions. To extract collocations from the data source, a statistical method was employed. It was defined so as to test whether the probability of two co-occurring elements in a combination was under the confidence level. Based on a highly-cited study by Manning et al. (1999), χ^2 (or Chi-squared) was determined as the statistical method for the extraction of collocations used to develop the assisted learning tool, SpColEDRS. The training data (9,807 words) for the developed tool, comprised the fairy tales¹ from the Spanish sub-corpus of the CPEIC trilingual parallel corpus and revised texts from the CEATE learners' corpus. The database of Spanish collocations was generated with machine learning and processed through data processing, collocation extraction,

¹Data sources included International Children's Library <http://en.childrenslibrary.org/> and <http://itunes.apple.com/hk/app/id440153337?mt=8>.

and manual modification. This database served as a reference to carry out collocation checking by detecting learner errors and providing possible suggestions for learners to use to correct their errors. TreeTagger was used for POS-tagging data, and PHP, AJAX, and MySQL were used as the development tools for error detection and revision suggestions. The SpColEDRS tool was designed with two main functions: error detection and revision suggestions for Spanish collocation for learning purposes.

3.2 The evaluation of the computer-assisted learning tool for Spanish collocations

To evaluate the practical effectiveness of the developed tool from the user perspective, we conducted an experiment consisting of a pretest, a video tutorial, a post-test, followed by a user questionnaire. The collected information was analyzed to examine whether the SpColEDRS tool was able to assist learners with improving their learning by comparing learning outcomes from two groups of Spanish learners, experimental and contrastive groups.

3.2.1 Participants

Thirty three (33) Spanish learners from National Cheng Kung University participated in the evaluation. Their mother language was Mandarin-Chinese; their first foreign language (L2) was English, and their second foreign language (L3) was Spanish, in which they had 180–360 instructed hours. The participants did not have much contact with the L3 Spanish outside of the classroom since Mandarin Chinese is the predominant language in Taiwan. Prior to the pretest, all participants took the Wisconsin Placement Test to assess their Spanish proficiency in general. According to their scores on the Wisconsin Placement Test, they were grouped into two proficiency levels of Spanish: 11 at beginning-high level (457–517 points) and 22 at the intermediate-low levels (535–653 points). Then, they were randomly assigned to two groups, 17 to the experimental and 16 to the contrastive groups.

3.2.2 Procedure

Both the on-line pre and post-tests contained 40 sentences with one element of the combination left blank to be filled in by the participants according to the correspondent translation in Chinese (Appendix 5). The tested combinations were four different types, including Verb-

Noun, Adjective-Noun, Noun-Adjective, and Verb-Preposition.

One week after the pretest was conducted, the participants in the experimental group were directed to view a video tutorial (two-minutes) to learn how to use the SpColEDRS computer-assisted learning tool. The participants in the control group did not receive any treatment. The video tutorial provided participants with basic instructions for using the assisted learning tool.

Then, the participants in both groups completed the post-test engaging in the same task as that used in the pretest. In the post-test, the participants from the experimental group were required to fill in a blank to complete the combined elements of the collocation before using the assisted tool, and then on another line, they were asked to indicate whether they modified the answer after using the provided tool and to explain what they had changed if this was the case (Appendix 5).

After the post-test, the participants from the experimental group were required to complete the questionnaire (Appendix 5). The questionnaire included two subsections; one was used to collect the users' levels of satisfaction with the interface on a Likert-scale, and the other involved open-ended questions regarding the usefulness of the system as well as suggestions for further modifications.

4 Results and discussion

4.1 Development of SpColEDRS

The developed computer-assisted learning tool for Spanish collocation provides a checking functionality with error detection and revision suggestions for Spanish collocation, as shown in Figure 1. If the key-in collocation exists in our database, the system responds with a confirmation, as shown in Figure 2. However, when a possible error is entered, the system responds immediately, and users can then select an appropriate revision from the provided suggestion list, as shown in Figure 3.

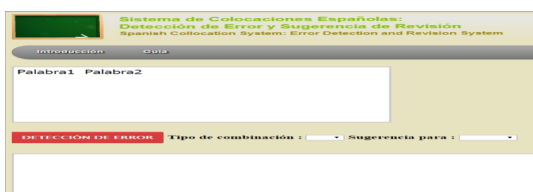


Figure 1: Spanish Collocation Error Detection and Revision Suggestion tool: User's interface.

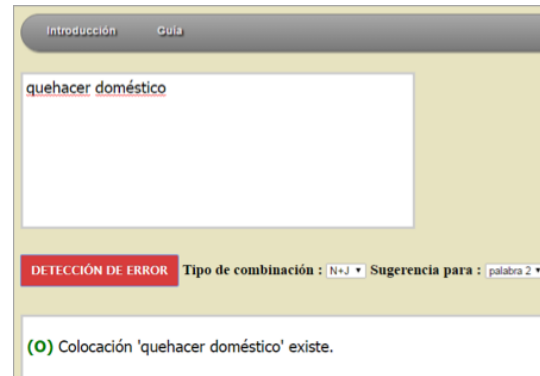


Figure 2: Spanish Collocation Error Detection and Revision Suggestion tool: Confirmation of correct use.

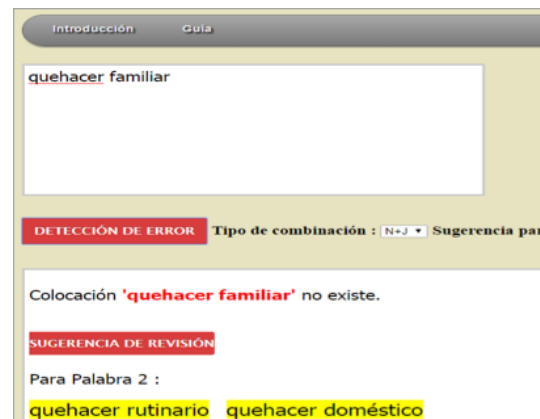


Figure 3: Spanish Collocation Error Detection and Revision Suggestion tool: Provision of correction suggestion.

4.2 Evaluation of developed tool

4.2.1 Data Analysis Methods

According to the research questions, a one-way ANCOVA was selected for the purpose of determining (1) if there were any significant differences between the experimental and control groups after the pedagogical intervention using the developed corpus-based learning tool, and (2) if there were any significant differences between learners at the beginning and intermediate levels in the post-test after using the assisted learning tool.

Prior to the analysis of research question 1, the pretest, post-test, and group variables were examined using SPSS programs to check for the accuracy of data entry, missing values, the linearity between the covariate (pretest) and dependent variables (post-test), and the assumptions of the homogeneity of the regression slopes, normality, homoscedasticity, homogeneity of variance, and outliers.

There were no missing values in the data set. Pairwise linearity was checked using within-

group scatterplots and was found to be satisfactory. There were no cases detected as outliers based on an examination of the z scores on the post-test. There was homogeneity of the regression slopes because the interaction term was not statistically significant, $F(1, 29) = 0.000$, $p = 0.982$. Because the variable post-test was severely skewed, a “reflect and logarithmic” transformation was applied, which means that the new post-test variable (PostTotal-log10-ref) was equal to the LG10 (“the highest score on the post-test plus 1” — post-test scores”). With the transformed variable in the variable set, standardized residuals for the post-test and for the overall model were normally distributed, as assessed with the Shapiro-Wilk’s test ($p > 0.05$). Also, there was homoscedasticity, as assessed by visual inspection of the standardized residuals plotted against the predicted values. The assumption of homogeneity of variances was met, as assessed by Levene’s test of homogeneity of variance ($p = 0.246$). There was no outlier in the data, which was assessed by determining that there were no cases with standardized residuals greater than ± 3 standard deviations.

4.2.2 Effectiveness

A one-way ANCOVA was run to determine the effect of the pedagogical intervention treatment using the corpus-based learning tool developed for this study on the post-test after controlling for the pretest. As shown in Table 1, after adjustment for the pretest, there was a statistically significant between-group difference in the post-test for the experimental group and the control group, $F(1, 30) = 100.768$, $p < 0.001$, partial $n^2 = 0.771$.

Source	df	MS	F	p	n^2
PreTotal	1	0.219	7.189	0.012	0.193
Group	1	3.069	100.768	0.000	0.771
Error	30	0.030			

Table 1: Analysis of covariance for the post-test with the pretest as a covariate.

The post hoc analysis was performed with a Bonferroni adjustment. The post-test scores were statistically significantly better in the experimental group than in the control group, as shown in Tables 2 and 3, because the post-test scores were transformed by a “reflect and logarithmic” transformation as explained above. Therefore, the developed assisted learning tool had a positive effect on the students’ learning of Spanish collocations.

	Unadjusted			Adjusted	
	N	M	SD	M	SE
Control	16	14.062	4.3277	13.8424	0.8132
Experiment	17	24.118	2.5952	24.3248	0.7887

Table 2: Adjusted and unadjusted experimental means and variability for the post-test with the pretest as a covariate before post-test transformation.²

Group (I)	Group (J)	Mean Difference (I-J)		Sig. ^(b)	95% C.I. ^(b)	
		SE			Lower Bound	Upper Bound
C	E	0.616 ^(a)	0.061	0.000	0.491	0.742
E	C	-0.616 ^(a)	0.061	0.000	-0.742	-0.491

Table 3: Pairwise comparison.³

In addition, a one-way ANCOVA was also selected to answer research question 2: Are there any significant differences between the different levels of learner proficiency in the post-test after using the learning tool? The same statistical analysis procedures used for research question 1 were conducted. The variable post-test was also transformed using a “reflect and logarithmic” transformation as explained above because the post-test variable has a serious skewness. With the transformation, all the assumptions for the one-way ANCOVA were satisfied.

The results of the one-way ANCOVA test shown in Table 4 show that there was no significant difference between the beginning level and intermediate level in the post-test after using the learning tool by controlling the effect of pretest, $F(1, 30) = 0.088$, $p = 0.769$.

Source	df	MS	F	p	n^2
PreTotal	1	0.016	0.120	0.732	0.004
Level	1	0.012	0.088	0.769	0.003
Error	30	0.132			

Table 4: Analysis of covariance for the post-test with the pretest as a covariate.

However, the results of the independent *t*-test shows that there was a significant difference between the beginning level and intermediate level in the pretest before using the learning tool, $p < 0.001$. After using the learning tool, the

²*N* is the number of participants, *M* the Mean, *SD* the Standard Deviation and *SE* the Standard Error.

³Dependent Variable: PostTotal-log10-ref; Based on estimated marginal means; ^(a) The mean difference is significant at the .05 level.; ^(b) Adjustment for multiple comparisons: Bonferroni.

beginning group increased their test scores from 7.909 to 18 (see Table 5). Also, the intermediate group increased their test scores from 13.0909 to 19.8636 (see Table 5). Therefore, the learning tool had a positive effect on both the beginning group and the intermediate group, but had a greater positive effect on the beginning group.

		Pretest		Post-test	
		M	SD	M	SD
Beginning	11	7.909	3.113	18	7.4027
Intermediate	22	13.0909	4.0344	19.8636	5.5574

Table 5: Means and variability for the pretest, and the post-test with the pretest as a covariate.⁴

4.2.3 Questionnaire

The results of the satisfaction survey for the interface interaction between the users and the developed tool showed that most participants were satisfied (over 3.8 on a scale of 5) with the SpColEDRS in terms of identifying collocation errors and the suggestion lists provided to them for correction, as shown in Table 6. According to the user responses, the assisted learning tool was easy and simple to use, and the reaction times for error detection and correction suggestions for Spanish collocations were appropriate. This developed tool was recommended for self-learning although users at different proficiency levels might benefit from it to a greater or lesser degree. In summary, the Spanish collocation error detection and correction suggestion functions for the lexical features included in the database were found to be useful.

Interface interaction	Q1	Q2	Q3	Q4	Q5
Beginning	4.5	3.8	4.8	5	4.5
Intermediate	4.5	4.3	4.1	4.5	4.5

Table 6: Results indicating satisfaction with the Spanish Collocation Error Detection and Revision Suggestion tool⁵.

According to the participants' responses to the open-ended questions in the survey, the advantages of this collocation learning tool included immediate feedback and ease and simplicity of the search process. However, the tool had several disadvantages. For example, users had to

⁴*N* is the number of participants, *M* the Mean and *SD* the Standard Deviation.

⁵Q1: Identifying lexical errors; Q2: Provision of suggestion list for correction; Q3: Easy and simple to use; Q4: Appropriate reaction time; Q5: Recommended for self-learning).

know at least one word of the two combined elements in order to make it possible to use the tool. It was difficult to choose the appropriate one from more than one possible correction suggestion. The users suggested future modifications such as to provide English or Chinese translations of the searched collocations to facilitate understanding of the meaning of the collocations. The participants also suggested providing examples of collocation usage to help distinguish subtle differences among the collocations offered in the feedback.

4.3 Limitations and future work

The user evaluation of the SpColEDRS was, in general, positive and suggested that the users were satisfied. However, the training data for our developed tool from the two corpora (learners' corpus CEATE and trilingual parallel corpus CPEIC) was relatively small. Therefore, the identification and detection of errors were limited to collocations within a fixed and limited range. Also, the context and the current experiment were conducted within searchable combinations. A larger amount of training data from a greater variety of text types should be included for training in the future in order to obtain better results in terms of error detection and correction suggestions, which would strengthen the applicability of this assisted corpus-based tool for teaching and learning Spanish collocations.

As has been suggested by users, translations of L1 Chinese or L2 English should be provided to assist learners with their understanding of Spanish collocations, especially in the case of beginning learners. In addition, examples of collocation uses in sentences in meaningful contexts should be listed as an option to illustrate the differences among the suggested collocations.

5 Conclusions

In this study, a corpus-based assisted tool for collocation learning in Spanish was developed and evaluated. Based on the training data compiled in two created corpora (CEATE and CPEIC) and a Spanish collocation extraction tool (HCE), this computer-assisted learning tool is easy to operate and has two major functions: error detection and revision suggestions. SpColEDRS can detect inappropriate uses of Spanish collocations and provides suggestion lists for learners to choose from for the purpose of correcting their collocation errors.

To ensure the effectiveness of and user satisfaction with the SpCoEDRS, the developed tool was evaluated using two tests and a questionnaire. The research results showed that the SpCoEDRS could assist learners effectively based on the progress of the experimental group from the pretest to the post-test, especially in the case of the beginning learners. Furthermore, the results of the satisfaction survey assessing the students' opinions of the interface and usefulness of the tool indicated that most of the participants positively confirmed that the tool was effective for assisting them with their practice with Spanish collocations. Finally, to optimize the use of the existing corpora (CEATE and CPEIC) and tool (HCE), this study extended our previous outcomes of the created corpora and tool for the advancement of studying effective learning of Spanish collocation in Taiwan and further shed light on pedagogical applications of the created corpora and on the learning of Spanish collocation with a corpus-based approach in a multilingual acquisition setting.

Acknowledgments

We wish to extend our sincere gratitude to the Ministry of Science and Technology of Taiwan for their generous support with project grant number 103-2410-H-006-059-MY2, and our appreciation for the technical support provided by the Computer Science and Information Engineering team at National Cheng Kung University in Taiwan and the research assistants involved in this project.

References

- Alfahadi, Abdulrahman M., Said Ahmed Zohairy, Mowaffaq Mohammed Momani & Mansour H. Wahby. 2014. Promoting awareness of teaching collocations techniques to beginners (adjective-noun collocations). *European Scientific Journal* 10(10). 389–396.
- Alonso Ramos, Margarita, Alfonso Nishikawa & Orsolya Vincze. 2010. DiCE in the web: An online spanish collocation dictionary. In *ELexicography in the 21st Century: New Challenges, New Applications (ELex)*, 369–374.
- Boisson, Joanne, Ting-Hui Kao, Jian-Cheng Wu, Tzu-Hsi Yen & Jason S. Chang. 2013. Linggle: a web-scale linguistic search engine for words in context. In *51st Annual Meeting of the Association for Computational Linguistics*, 139–144.
- Bolshakov, Igor A. & Sabino Miranda-Jiménez. 2004. A small system storing Spanish collocations. In *International Conference on Intelligent Text Processing and Computational Linguistics*, 248–252.
- Chang, Yu-Chia, Jason S. Chang, Hao-Jan Chen & Hsien-Chin Liou. 2008. An automatic collocation writing assistant for Taiwanese EFL learners: A case of corpus-based NLP technology. *Computer Assisted Language Learning* 21(3). 283–299. doi 10.1080/09588220802090337.
- Chen, Hao-Jan Howard. 2011. Developing and evaluating a web-based collocation retrieval tool for EFL students and teachers. *Computer Assisted Language Learning* 24(1). 59–76. doi 10.1080/09588221.2010.526945.
- Davies, Mark. 2012. Corpus del español (100 million words, 1200s-1900s). [online] <http://www.corpusdelespanol.org>.
- Ferraro, Gabriela, Rogelio Nazar, Margarita Alonso Ramos & Leo Wanner. 2014. Towards advanced collocation error correction in Spanish learner corpora. *Language Resources and Evaluation* 48(1). 45–64. doi 10.1007/s10579-013-9242-3.
- Firth, John. 1957. Modes of meaning. *Papers in Linguistics* 5. 190–215.
- González-Agirre, Aitor & German Rigau. 2013. Construcción de una base de conocimiento léxico multilingüe de amplia cobertura: Multilingual central repository. *LinguaMática* 5(1). 13–28.
- Granger, Sylviane. 1998. Prefabricated patterns in advanced EFL writing: Collocations and formulae. In *Phraseology: Theory, analysis, and applications*, 145–160. Oxford University Press.
- Greaves, Christopher. 1999. Virtual language centre study guide. [online] <http://vlc.polyu.edu.hk/>.
- Jian, Jia-Yan, Yu-Chia Chang & Jason S. Chang. 2004. TANGO: bilingual collocational concordancer. In *Association for Computational Linguistics (ACL)*, doi 10.3115/1219044.1219063.
- Källkvist, Marie. 1995. Lexical errors among verbs: A pilot-study of the written language of advanced Swedish learners of English. *Working Papers in English and Applied Linguistics* 103–115.

- Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý & Vít Suchomel. 2014. The Sketch Engine: ten years on. *Lexicography* 1(1). 7–36. doi 10.1007/s40607-014-0009-9.
- Laufer, Batia & Tina Waldman. 2011. Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning* 61(2). 647–672. doi 10.1111/j.1467-9922.2010.00621.x.
- Liou, Hsien-Chin, Jason S Chang, Hao-Jan Chen, Chih-Cheng Lin, Meei-Ling Liaw, Zhao-ming Gao, Jyh-Shing Roger Jang, Yuli Yeh, Thomas C. Chuang & Geeng-Neng You. 2006. Corpora processing and computational scaffolding for a web-based English learning environment: The CANDLER project. *CALICO journal* 24(1). 77–95.
- Lorenz, Gunter R. 1999. *Adjective intensification: learners versus native speakers: a corpus study of argumentative writing*, vol. 27. Rodopi.
- Lozano, Cristóbal. 2009. CEDEL2: Corpus escrito del español L2. In Bretones Callejas (ed.), *Applied Linguistics Now: Understanding Language and Mind / La Lingüística Aplicada Hoy: Comprendiendo el Lenguaje y la Mente.*, 197–212. Universidad de Almería.
- Lu, Hui-Chuan & An Chung Cheng. 2016. Acquisition of L3 Spanish combinations: Development in bilingual and multilingual contexts. In *8th International Conference of Language Acquisition*, n.pp.
- Lu, Hui-Chuan, Yu-Hsin Chu & Cheng-Yu Chang. 2013. A corpus-based system of error detection and revision suggestion for Spanish learners in Taiwan: A case study. *JALT CALL Journal* 9(2). 115–130.
- Manning, Christopher D, Christopher D Manning & Hinrich Schütze. 1999. *Foundations of statistical natural language processing*. MIT press.
- McKeown, Kathleen R. & Dragomir R. Radev. 2000. Collocations. In *Handbook of Natural Language Processing*, 1–23. CRC Press.
- Nattinger, James R. & Jeanette S. DeCarrico. 1992. *Lexical phrases and language teaching*. Oxford University Press.
- Nesselhauf, Nadja. 2003. The use of collocations by advanced learners of English and some implications for teaching. *Applied linguistics* 24(2). 223–242. doi 10.1093/applin/24.2.223.
- Nesselhauf, Nadja. 2005. *Collocations in a learner corpus*, vol. 14. John Benjamins Publishing.
- Sinclair, John. 1991. *Corpus, concordance, collocation*. Oxford University Press.
- Vincze, Orsolya, Margarita Alonso Ramos, Estela Mosqueira Suárez & Sabela Prieto González. 2011. Exploiting a learner corpus for the development of a CALL environment for learning Spanish collocations. In *Electronic lexicography in the 21st century: New applications for new users (eLex)*, 280–285.
- Weisser, Martin. 2016. Corpus-based linguistics links. [online] http://martinweisser.org/corpora_site/CBLLinks.html.
- Wray, Alison. 2000. Formulaic sequences in second language teaching: Principle and practice. *Applied linguistics* 21(4). 463–489. doi 10.1093/applin/21.4.463.
- Wu, Jian-Cheng, Kevin C. Yeh, Thomas C. Chuang, Wen-Chi Shei & Jason S. Chang. 2003. TOTALrecall: A bilingual concordance for computer assisted translation and language learning. In *41st Annual Meeting on Association for Computational Linguistics*, 201–204. doi 10.3115/1075178.1075216.

Appendices

Appendix 1-Control and experimental group pretest and control group post-test.

- | | |
|---|--|
| <p>1 我現在很累，想睡覺。
Estoy muy cansado. ____sueño.
那時我們看著101的煙火，在美國餐廳吃晚餐。
2 Vimos los fuegos ____ del 101 y cenamos en un restaurante americano.
她留有一頭長且直的金髮。
3 Tiene el pelo ____ , largo y liso.
在我空閒的時間裡，我喜歡聽音樂。
4 En mi ____libre, me gusta escuchar la música.
現在我和我的父母關係很好，因為我試著和他們溝通。
5 Ahora tengo unas buenas relaciones con mis padres porque yo trato ____ comunicarme con ellos.
他們生氣不是沒有道理的，您是應該專心讀書的。
6 Ellos no están enojados sin razón, usted tiene que ____atención al estudiar.
第二天，我們參觀一間很大的博物館。
7 Al día ____ , visitamos un museo muy grande.
他突然因為撞到了一塊石頭而跌倒了。
8 De pronto tropezó ____ una piedra y se cayó.
你們必須知道父母將不會對子女造成傷害。
9 Necesitáis saber que los padres no van a ____daño a sus hijos.</p> <p>10 他不僅編新故事，還找藉口。
Inventaba nuevas historias y ____excusas.
11 哥斯大黎加的傳統經濟依賴農業。
La economía tradicional de Costa Rica depende ____ la agricultura.
12 星期一到五下課後，我習慣悠遊於網際網路之中。
De lunes a viernes, suelo navegar ____ internet después de terminar la clase.
通常夏天都是很好的天氣，可是有時會下很大的雨，而且有風。
13 Generalmente, hace buen tiempo, pero a veces llueve mucho y ____viento en verano.
女人為了漂亮，有很多方法可以減重。
14 Para que las mujeres sean bonitas hay muchos métodos para ____peso.
他無法支撐起他的家庭，而且還有一大堆的問題。
15 No puede sostener ____ su familia y tiene una cantidad de problemas.
我有可能環遊全世界、認識各式各樣的人。
16 Tengo la posibilidad para viajar en el mundo ____ y conozco toda clase de personas.
他是很用功的學生，上課總會問問題。
17 Es un estudiante muy trabajador. Siempre ____preguntas en clase.
因為想家，我剛才打電話給我母親。
18 Debido a la nostalgia, acabo ____ llamar a mi madre por teléfono.
為了買到夢想中的房子，他努力省錢。
19 Para comprar la casa de sueño, hace un esfuerzo por ____dinero.
20 我的弟弟今年將要上小學。
Mi hermano menor va a ir a la escuela ____ este año.</p> | <p>21 如果他想看電視，需要先完成作業。
Si quiere ver la televisión, ____falta terminar la tarea primero.
他沒有兄弟姊妹，是獨生子。
22 No tiene hermanos, es hijo ____
這個故事發生在一個小村莊。
23 La historia ____lugar en un pequeño pueblo.
我們家有四間臥室和二間廁所。
24 Nuestra casa cuenta ____ cuatro dormitorios y dos baños.
然後，那些今年畢業生排隊離開？
25 Luego los estudiantes que gaduarán este año ____cola y salieron del auditorium.
如果您的小孩犯錯的話，不要馬上處罰他們。
26 Si sus hijos ____ errores, no les castigue inmediatamente.
夜市中有很多美食。
27 Hay muchas comidas deliciosas en el mercado nocturno ____
你必須小心不要踩到草皮，不然會被罰款。
28 Tienes que ____ cuidado de no pisar el césped, te ponen una multa.
我一直讀一直讀，終於覺得那些文字開始合乎邏輯了。
29 Continué leyendo y leyendo, y finalmente las palabras comenzaron a ____sentido.
他不喜歡幫助別人，也從不向人求助。
30 No le gustaba ayudar a otros y nunca ____ayuda a otros.
跟我談談你下一年度的計畫吧！
31 Dime algo sobre tus planes para el ____año.
32 這個節目提供了一些安全騎機車的資訊和建議。
Este programa proporciona las informaciones y consejos para andar ____ motocicleta con seguridad.
33 為了獲取好成績，我很用功讀書。
Para ____nota buena, estudio mucho en la biblioteca.
我很欣賞他作詩的才華。
34 Le admira mucho por su capacidad para ____versos
他喜歡開玩笑，因此有很多朋友。
35 Le gusta ____broma, por eso, tiene mucho amigos.
36 我想要跟那位電影明星一起合照。
Quiero que ____foto con esta estrella de cine.
37 我想要享受一個美好的退休生活。
Quisiera disfrutar ____ una vida maravillosa después de la jubilación.
38 因為我會餓，所以我想快點結束。
Quiero acabar rápido porque ____hambre.
39 因為他奇怪的笑容讓我們害怕，所以我們開始像瘋子一樣跑了起來。
Empezamos a correr como locas porque ____miedo su extraña sonrisa.
40 我認為我們必須注意/理會醫生的建議。
Creo que deberíamos ____caso a los consejos del doctor.</p> |
|---|--|

Appendix 2-Pretest of control and experiment groups, and posttest of control group.

- 1 我現在很累，想睡覺。
 - a. Estoy muy cansado. _____sueño.
 - b. 使用EDRS後未修正。修正為：_____sueño
- 2 那時我們看著101的煙火，在美國餐廳吃晚餐。
 - a. Vimos los fuegos _____ del 101 y cenamos en un restaurante americano.
 - b. 使用EDRS後未修正。修正為：fuegos _____
- 3 她留有一頭長且直的金髮。
 - a. Tiene el pelo _____, largo y liso.
 - b. 使用EDRS後未修正。修正為：pelo _____
- 4 在我空閒的時間裡，我喜歡聽音樂。
 - a. En mi _____ libre, me gusta escuchar la música.
 - b. 使用EDRS後未修正。修正為：_____ libre
- 5 現在我和我的父母關係很好，因為我試著和他們溝通。
 - a. Ahora tengo unas buenas relaciones con mis padres porque yo trato _____ comunicarme con ellos.
 - b. 使用EDRS後未修正。修正為：trato _____
- 6 ~ 40 Chinese...
 - a. Spanish sentences...
 - c. 使用EDRS後未修正。修正為：...

Appendix 3-Questionnaire

SpColEDRS: Questionnaire

「工具」使用意見調查

Please indicate if you agree or disagree with the following statements by using the scale from 5 (strongly agree) to 1 (strongly disagree.)

5 : 非常同意、 4 : 同意、 3 : 普通、 2 : 不同意、 1 : 非常不同意

5: Strongly agree, 4: Agree, 3: Neutral, 2: Disagree, 1: Strongly disagree

1. System interface

一、系統之操作介面互動

● SpColEDRS can detect errors in lexical usage?

能偵測出詞彙使用錯誤

5 4 3 2 1

● SpColEDRS can provide helpful suggestions for revision?

所建議的清單對修正有幫助

5 4 3 2 1

● SpColEDRS is easy to use?

容易操作

5 4 3 2 1

● SpColEDRS can detect errors within an acceptable response time?

系統偵測錯誤所需的時間適當

5 4 3 2 1

● SpColEDRS can provide suggestions for revision within an acceptable response time?

系統給予建議修正所需的時間適當

5 4 3 2 1

● In general, SpColEDRS facilitates the self-learning of Spanish in writing?

整體而言，對西語表達的自學有幫助

5 4 3 2 1

2. Advantages and disadvantages of SpColEDRS

二、「錯誤偵測、修正建議系統」的優點及缺點

Based on your user experience, please evaluate the SpColEDRS regarding the advantages and disadvantages of the following features.

(1) Error detection

(一)「錯誤偵測」部分：

Advantages 優點：

Disadvantages 缺點：

(2) Revision suggestion

(二)「修正建議」部分：

Advantages 優點：

Disadvantages 缺點：

3. Feedback and suggestions for SpColEDRS

三、對「錯誤偵測、修正建議系統」之建議及回饋

Based on your user experience with the SpColEDRS, please provide your feedback and suggestions for improvement in the following areas:

(1) Error detection

(一)「錯誤偵測」部分：

(2) Revision suggestion

(二)「修正建議」部分：

Thank you for your cooperation!

謝謝配合！