Chacón Beltrán, R. y R. Echitchi. (2022). Improving lexical errors in EFL writing by using software-mediated corrective feedback. *Porta Linguarum*, 37, 275-290.

<https://doi.org/10.30827/portalin.vi37.20847>

**Abstract:** This research aims at assessing the efficiency of software when it comes to providing learners with corrective feedback on their pieces of writing. This article in particular delved into the effect of software-generated feedback on students’ ability to correct and avoid lexical errors. The study, based on a series of essays written by a total of 33 undergraduate students enrolled in the degree programme in English at the National Distance Learning University in Spain contributed to confirming the assumption that technology can be an extremely useful tool in the teaching and learning process. More specifically, this work demonstrated that students could reduce significantly the number of lexical errors in their essays through the autonomous use of error-correction software and that, over time, the students improved on their ability to avoid such errors. Nevertheless, the work also demonstrated that software can in no way completely replace teachers, as computer programming is quite limited there are errors that only proficient language users can detect and correct.

**Los errores léxicos en la expresión escrita en inglés como lengua extranjera: La eficacia del feedback correctivo automatizado**

**Keywords:** CALL, feedback, self-correction, lexical errors, Grammar Checker

**Resumen:** El objetivo de este trabajo de investigación es evaluar la eficacia de las herramientas informáticas para proporcionar feedback sobre los errores de alumnos y mejorar su escritura. En este artículo se hace hincapíe en los errores léxicos, tratando de ver en qué medida el feedback proporcionado por herramientas informáticas puede ayudar a los estudiantes a corregir y evitar esos errores. Este estudio, basado en una serie de composiciones elaboradas por 33 estudiantes del Grado en estudios ingleses por la Universidad Nacional de Educación a Distancia de España, nos llevó a la conclusión de que la tecnología puede ser una herramienta de gran utilidad en el proceso de enseñanza y aprendizaje en contextos de enseñanza a distancia. Este trabajo demostró que los estudiantes pudieron reducir significativamente el número de errores léxicos en sus ensayos mediante el uso autónomo de herramientas tecnológicas de corrección de errores y que, a lo largo del tiempo, los estudiantes mejoraron su capacidad para evitar tales errores. Sin embargo, el trabajo también demostró que ninguna herramienta tecnológica puede sustituir completamente a los profesores, ya que la programación informática es de momento limitada y hay errores que sólo los usuarios con conocimientos avanzados en el idioma pueden detectar y corregir.

**Palabras clave:** ELAO, feedback, corrección autónoma, errores léxicos, Grammar Checker

**1. Introduction**

There is no doubt that humankind is experiencing times of rapid changes which have often been underpinned by the development of Information Technology (IT) that has revolutionised not only information and communication, but also teaching (Brill & Galloway, 2007; Morrison et al., 2010). The spread of IT has led to the birth of new pedagogical approaches that differ substantially from more traditional approaches characterised by the prominent role of teachers who were involved in every stage of learning, whereas learners were rather passive stakeholders in the process and had to rely on the physical presence of a teacher to achieve their learning goals.

The Internet has provided avenues for the rapid spread of information and the consolidation of the so-called “global village” where distance is no longer is a barrier and people from all over the world can interact with each other without having to relocate. In addition, nowadays’ global village has provided computer literate individuals with more autonomy when it comes to managing their daily lives. Therefore, people can manage their bank accounts or purchase train tickets from the comfort of their homes with no assistance whatsoever. This tendency towards autonomy has also had a great impact on learning, especially when it comes to error feedback and correction (Chacón-Beltrán, 2018).

**1.1 Lexical errors and language learning**

Vocabulary is one of the most important aspects of language, and this is exactly why its study very often marks the beginning of language learning. As concerns lexical errors, Agustín-Llach (2011) defines them as “deviation[s] in form and/or meaning of a target-language lexical word” (p. 75). In other words, lexical errors not only include spelling, but other erroneous constructions involving collocations and false friends, as discussed later in this paper. Agustín-Llach (2011) goes on to indicate that the study of lexical errors can say much about second language acquisition and can help teachers design materials that definitely address learner needs.

Lexical errors have thus been studied from various perspectives, taking into account different research populations. Apart from studies which involved primary or/and secondary school students in Spain (Agustín-Llach, 2011; 2015), other researchers such as Carrió-Pastor and Mestre-Mestre (2013) focused on the errors made by non-native English researchers in scientific literature.

As far as first language (L1) influence is concerned, most researchers agree that no matter their proficiency level, learners cannot completely avoid errors resulting from L1 transfer. Research works on L1 transfer have often aimed at addressing the claim that less proficient learners are therefore more likely to provide evidence of L1 transfer, though there has been no consensus on how true this assumption is. As a matter of fact, what can be retained from earlier publications is that researchers who are interested in studying lexical errors in a second language should go beyond the assumption that those errors would decrease in number as proficiency increases. In fact, as Agustín-Llach’s (2015) study reveals, some specific lexical errors resulting from literal translation and semantic confusion might indeed disappear as the learner becomes more proficient, but instances of misselection and coinage might become more numerous amongst high-proficiency learners. Agustín-Llach’s conclusion parallels claims made earlier by Cenoz (2003), as well as García Lecumberi and Gallardo (2003).

**1.2 Research on corrective feedback**

Many research works have been devoted to studying corrective feedback (Chodorow, Gamon & Tetreault, 2010; Ellis, et al., 2008; Makino, 1993; Senra-Silva, 2010) as it has always played a fundamental part in the teaching and learning process. Nevertheless, as Chandler (2003) indicated, there has been great controversy over whether or not corrective feedback could actually contribute to improving students’ knowledge of a specific language. In other words, while some researchers indicated that correcting errors might actually backfire, others insisted on its importance and suitability. One of the researchers who stood firmly against error correction is Truscott (1996), who indicated that it has a negative effect on learning and should be avoided. Truscott built those claims on previous works (Kepner, 1991; Sheppard, 1992), which pointed to the fact that corrective feedback contributed very little to improving learning. Reasons for the unsuitability of error correction often hover around the claim that it generally tends to make students more self-conscious and thus dampens their desire to learn (Fazio, 2001). Nevertheless, more recent research has proved that when done the right way, error correction and feedback could definitely contribute positively to learning. At this juncture, it is worth going back to Chandler (2003), who studied two groups of students over a period of 10 weeks. Students in both groups wrote a series of essays, with the difference being that members of the experimental group were required to correct errors in one essay before submitting the following one, whereas members of the control group would correct all their essays at the end of the research period. The study concluded that error feedback contributed significantly to improving writing amongst members of the experimental group, as they were required to build on such feedback to correct their essays and avoid similar errors in subsequent pieces of writing. Findings similar to Chandler’s were obtained more recently by Bitchener and Ferris (2012) and Lee (2013).

Nevertheless, Chandler as well as all those researchers who emphasized the suitability of corrective feedback insisted that the latter should be provided the right way, otherwise it would lose its efficiency. In fact, one of the two main purposes of Chandler’s (2003) work was to correlate achievement with the type of corrective feedback students receive. To achieve that, she compared four different types of response, namely *direct correction*, *underline and describe*, *describe* and *underline.* Here again, findings indicated that errors would reduce significantly when clearer pieces of feedback were provided. Therefore, Chandler insisted that the type of response that yielded the lowest number of mistakes was *direct correction,* while meredescription would not shield students from making some mistakes.

Many researchers have come up with different categories to describe the different types of corrective feedback. Generally speaking, there has been a distinction between direct and indirect feedback, with direct feedback referring to situations where the teacher would provide a clear correction for the error the student made, while indirect feedback would involve teachers’ highlighting errors, mostly by means of underlining or coding (also known as description), but “leav[ing] it to the student writer to solve the problem and correct the error.” (Ferris 2002, p. 19). Indirect feedback would roughly encompass Chandler’s *underline and describe*, *describe* and *underline.* Overall, there have been conflicting views on whether or not direct or indirect feedback should be preferred. While Chandler’s (2003) findings clearly advocate direct feedback, other researchers obtained different results. For instance, Ferris and Roberts (2001) studied the impact of corrective feedback by means of two experimental groups (A and B) and a control group (C). After underlining and describing errors made by students in Group A while merely underling errors amongst Group B, they found that there were no statistically significant differences between Groups A and B, thus leading the authors to conclude that it is not clear whether more direct feedback affects accuracy positively. Though there is much controversy over which type of corrective feedback should be pursued, one thing that most researchers share is that no matter its type, feedback is better than no feedback (Ene and Upton, 2014), and it is an effective tool teachers should use in order to improve accuracy in student writing.

**1.3 Computer-generated scoring and feedback**

As we mentioned earlier, we are now living in a world where technology and computers are increasingly becoming essential in every aspect of our lives, including teaching. There has thus been increasing scholarly interest in finding out the extent to which computer programs can be used to detect errors and provide corrective feedback. When discussing teachers’ use of computers in writing courses, Ware (2011) insists that we should make a difference between what she refers to as computer-generated feedback, on the one hand, and computer-generated scoring, on the other hand. According to her, the use of software programs to grade writing assignments (i.e. computer-generated scoring) was found to be unsuitable by many scholars, that is why most research works, as this present one, have rather focused on the use of computers to help students identify and correct their essays (computer-generated feedback). The use of computer programs and applications can certainly be a valuable tool in the classroom, since it could help teachers avoid manual feedback, which is indeed a time-consuming activity. In fact, El Ebyary and Windeatt (2010) indicate that one of the hurdles faced by teachers who deal with manual feedback is that “they may not have time to give individualized, immediate, content-related feedback to multiple drafts” (p. 122). Chandler’s (2003), Warschauer’s (2010) and Guichon et al.’s (2012) conclusions, which indicate that delayed corrective feedback has very little effect on learning, actually consolidate the belief that computer-generated feedback, which is always timely, has the potential of helping students gain fluency and accuracy by actually learning from their errors as confirmed by Attali and Burnstein (2006), as well as Lee, Gentile and Kantor (2010). Nevertheless, while the latter found that students did better at writing after receiving computer-generated feedback the incorporation of computer-generated feedback into the classroom has sometimes witnessed some resistance from teachers.

Ware (2011) reveals that teachers may be reluctant to implement computer-generated feedback as they believe this would make the whole teaching activity “mechanistic and formulaic, divorced from real-world contexts” (p. 771). To solve this problem, she indicates that researchers should now move beyond assessing the efficiency of computer-generated feedback to address issues such as ways into which it can be integrated into traditional learning. While this is being done, what remains clear is that computers should never be used as replacements, but rather as complements to traditional writing instruction (Ware, 2011; Warschauer & Ware, 2006; Hernández Puertas, 2018).

This paper will therefore rely on qualitative and quantitative data to attempt to provide answers to the following questions: 1) How effective are computer programs when it comes to detecting and helping students correct and avoid lexical errors? 2) To what extent can automated writing correction complement the teaching activity in distance learning environments? 3) Does computer-assisted correction have positive effects on accuracy in writing?

**2. Current study**

**2.1 Participants**

This study was completed thanks to the participation of university students enrolled at UNED which is Spain’s largest distance learning university, given its student population which, as of September 2020, is estimated at over 156,000 in number. Once the research plan was completed, one of the authors of this paper made use of UNED’s online learning platform to post a message which aimed at encouraging students to participate. The students who agreed to volunteer in the project were reassured that it would not be graded, as grading the activity could have discouraged some students, or at least might have added unnecessary stress on them. The researchers also made sure they encouraged the students by telling them that by participating in this research they would not only have the opportunity to practice their writing skills, but they would also be given extra marks. Before starting the research work proper, the students were asked to provide more information about themselves by completing an online questionnaire. All in all, a total of thirty-three (33) students including 28 female and 5 male participants aged between 21 and 72 (the mean age was over 40 years old) expressed their interest in participating in the study. The informants were all undergraduate students enrolled in UNED’s degree programme in English. Though there is no clear indication regarding their proficiency level, one can say with a high degree of certainty that as university students enrolled in a programme whose sole medium of instruction is English, most of them would be upper intermediate learners, that is to say, B2 according to the common European Framework of Reference for Languages. Finally, as far as mother tongue is concerned, all but two (2) informants indicated that their native language was either (or both) Spanish or Catalan. Based the researchers’ experience as teachers at UNED, it is clear that these pieces of data related to gender, age and mother tongue are representative of the student population in UNED’s English Studies undergraduate programme. In fact, distance learning tends to appeal to people who, due to family and professional responsibilities cannot study in traditional universities. In the aforementioned degree programme, the student population mostly consists of mid-age individuals who are looking for ways to upgrade their English skills and climb the career ladder. Finally, UNED students are most often people who somehow have ties to Spain/Spanish, i.e, Spanish citizens, mother tongue speakers of Spanish and foreigners who have lived in Spain or a Spanish speaking country for some time. Therefore, the findings of this research work can certainly be applied to a good number of English learners in Spain and Spanish-speaking countries.

**2.2 On Grammar Checker**

Before describing data collection and analysis it is important to describe the computer tool that was used, i.e. *Grammar Checker*.

*Grammar Checker* is a a software programme specifically designed to help Spanish-speaking learners of English gain accuracy autonomously. More specifically, it is equipped with a word processor that allows students to type/paste their essays and obtain corrective feedback.

Lawley (2015), Chacón-Beltrán (2017) and Harvey-Scholes (2017), give specific technical details on how *Grammar Checker* works. In fact, the software package is built on a system known as n-grams, whose initial layer includes the so-called unigram, an error correction algorithm based on individual words. This system is supported by a database, which in the case of unigrams, could just be a list of words, as they appear in the dictionary. The unigram model can be used only for spelling detection and correction since it does not go beyond the word level. In addition, this model also has a few loopholes, as it often highlights names as errors and fails to distinguish between homophones (Harvey-Scholes, 2017). It is therefore to deal with the limitations and loopholes of the unigram system that computational linguistics came up with other error-detection layers, which would include word pairs (bigrams) or even sequences of three, four and up to five words in what would be referred to as trigrams, four-grams and five-grams, respectively.

As far as its interface is concerned, *Grammar Checker* is equipped with four distinct filters, namely “Spelling”, “Incorrect Sequences”, “Problem Words” and “Pairs”. The “spelling” filter analyses the frequency of the words that appear in the text and draw conclusions on whether the words are spelt incorrectly or very infrequent. As concerns “Incorrect Sequences”, the filter compares the text being processed with some of the most frequent incorrect phrases in English and provides the user with personalised feedback that they may use to improve their text.

The “Problem Words” filter identifies and lists some of the words in the text that (Spanish) learners of English tend to use incorrectly and, the “Pairs” filter analyses the frequency of sequences of two words used in the text. If those sequences are not that common in English the software may classify them as “very suspicious”, “suspicious” and “slightly suspicious”. While the “Problem Words” filter can help learners avoid false friends, for instance, the “Pairs” filter is meant to ease the identification of incorrect collocations.

In cases where sequences of words are used in order to increase error-detection and correction, it is important for the software package to include a database of texts which represent a normative use of the language, since that database is going to make it possible for the software to gauge whether or not a sequence of words is correct by correlating it with how often it appears in the normative corpus. This is exactly why Grammar Checker is backed by various normative corpora. Nevertheless, Lawley (2015), Chacón-Beltrán (2017) and Harvey-Scholes (2017) agree that though *Grammar Checker* makes use sophisticated error-detection algorithms it fails to identify some specific types of errors, as this paper will confirm.

**2.3 Research Outline**

The findings described in this paper are part of a wider research project based on data collected between November 2018 and February 2019. Data collection consisted of three different stages with each stage yielding an essay. The overall aim of this study was to have a thorough understanding of intermediate students’ writing skills and the extent to which software (*Grammar Checker*, in this case) could help them gain accuracy in writing. The students were provided with a list of topics which they could use for their essays. In every stage of the research, the students had to first of all write an essay by hand, scan and send it to the researchers who would then transcribe verbatim the essays and return them to the students. The latter would correct the essays with the help of *Grammar Checker* and return a final version with information on how long it took them to correct the essay and the difficulties they faced when going about that. The researchers would proceed to compare the handwritten and software-corrected versions and send participants overall feedback on their essay. This feedback included the identification of the errors that the students had failed to correct and some useful tips that they might use to avoid making the same types of errors in subsequent essays. It is important to note that, before the start of the essay writing and correction process, each participant was sent a tutorial on how to use *Grammar Checker* effectively. In the end, over 100 essays of 200 to 500 words were obtained and analysed by the researchers.

The researchers then went on to identify and code every error in both the first and software-corrected versions of the essays with a view to comparing both versions and coming up with information about whether or not software-assisted error detection and correction contributes to improving learning. To avoid complications that might have a negative effect on analysis, the errors were divided into three (2) broad categories, namely lexis and grammar. The findings presented in this article are limited to lexical errors which involved not only spelling, but also the use of false friends and erroneous collocations, which might have resulted from calque, semantic confusion, etc.

Various assumptions underpinned data collection and analysis. First of all, this research was based on the supposition that lexical errors would be fewer in number, given the fact that participants were quite familiar with free form writing assignments. In addition, this was a planned writing activity where participants had enough time to write their essays. Therefore, they were expected to avoid the use of unfamiliar words or at least check their meanings before actually proceeding to using them. This work’s second postulate is the idea that software correction would not be fully efficient at detecting and correcting certain words because of its technical limitations.

**2.3 Findings**

Given the fact that this work was first of all meant to be quantitative, we started with the statistical analysis of the lexical errors that were identified and described. Later on, statistical hypothesis tests that were carried out with a view to supporting conclusions related to the significance of the statistics that had been obtained.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Table 1: Lexical errors: stage crosstabulation* | | | | | | | | |
|  |  | Stage 1 | | Stage 2 | | Stage 3 | |  |
|  |  | Pre | Post | Pre | Post | Pre | Post |  |
| Number of participants | | 33 | 33 | 33 | 33 | 32 | 29 | total |
| Number of essays containing… | No error | 7 | 9 | 7 | 8 | 7 | 3 | 41 |
| 1 error | 4 | 6 | 10 | 11 | 8 | 10 | 49 |
| 2 errors | 9 | 7 | 7 | 11 | 8 | 10 | 42 |
| 3 errors | 2 | 4 | 4 | 2 | 5 | 7 | 24 |
| 4 errors | 6 | 3 | 2 | 0 | 2 | 1 | 14 |
| 5 errors | 0 | 0 | 1 | 0 | 5 | 2 | 8 |
| 6 errors | 2 | 1 | 2 | 1 | 1 | 1 | 8 |
| 7 errors | 1 | 1 | 0 | 0 | 0 | 1 | 3 |
| 8 errors | 2 | 2 | 0 | 0 | 0 | 0 | 4 |
| Total number of errors | | 87 | 73 | 61 | 45 | 70 | 66 | 402 |

**Stage 1.** This stage represents the very first essays the informants produced, i.e, the beginning of the research work. The first set of essays, which was referred to as PRE 1, contained a total of 87 errors that ranged from zero to eight (8). As Table 1 illustrates, seven (7) essays, i.e. 21.2% of the 33 essays received, included no lexical errors, while nine (9), i.e. 27.3 % contained 2 (two) lexical errors. Furthermore, it is important to note that the mean number of lexical errors for this first phase was 2.64 for a standard deviation of 2.32 (see Table 2). Therefore, one can state categorically that lexical errors were not that pervasive in the first set of essays.

After their transcription, the initial essays were returned to the students and the latter went on to apply software-assisted correction and produced revised versions which were labelled as POST 1. The fact that they applied automated correction to the essays contributed to reducing the number of errors, as this time around, 73 errors were registered. In other words, computer-assisted correction led to the correction of 16.09 % of the lexical errors that were initially made.

A careful analysis of the scores in POST 1 reveals the absence of lexical errors in (9) essays, as opposed to seven (7), in the PRE-1 phase. Furthermore, whereas only four (4) essays in PRE 1 (four in number) contained one (1) lexical error, their percentage rose to six (6) students in POST 1. When compared to the PRE 1 phase, the mean (2.21) is lower, which suggests that software again contributed to reducing the frequency of errors in various essays. Overall, there was a general tendency towards a lower number of errors, as found in Table 2. To sum up, this first stage of research confirmed *Grammar Checker*’s potential in helping students reduce some lexical errors, though it also stressed the software’s limitations as only 16.09 % of errors that the students made in their initial essays were corrected.

**Stage 2***.* To complete Stage 2 the participants chose a different essay topic and followed the same steps as in Stage 1. The analysis of the handwritten essays, which is labelled in this paper as PRE 2 yielded the identification of a total of 61 lexical errors. The first thing noticed here is the drop in the frequency of errors, as opposed to stage 1. Another significant finding is the fact that in this stage the number of errors ranged from zero to six (6), i.e. as opposed to Stage 1 no student made seven (7) or eight (8) lexical errors in Stage 2. This therefore suggests that overtime, students might have gained skills on how to avoid lexical errors in their essays.

When it comes to specific frequencies, Table 1 indicates that most essays in the PRE 2 phase contained very few lexical errors, with one (1), two (2) or no lexical errors, being the most prominent categories. As Table 2 indicates, the mean value for the number of errors was 1.85 for a standard deviation of 1.67, which may be seen as further evidence that lexical errors are not a serious issue for Spanish-speaking learners of English at the university level.

After the students subjected their essays to computer-assisted correction, there was some more decline in the number of errors, which went from 61 to 45, thus representing a 26.23% decrease. Finally, no essays at all contained four (4) or five (5) lexical errors in POST 2, as opposed to PRE 2. Overall, the tendency towards a lower number of errors remained consistent as reflected in the low mean (1.36) and standard deviation (1.22) (see Table 2).

**Stage 3.**This final stage led to the identification of seventy (70) errors in the handwritten essays received in what constituted the PRE 3 phase. That number represented a 15% increase compared to what was obtained in PRE 2, but was still 19% lower than the number of errors recorded in PRE 1 (87 errors). Here again, most students recorded no lexical errors or only one (1) error. Furthermore, as opposed to the two phases under stage 1 where errors ranged from zero to eight (8), there were no cases of students’ recording seven (7) or eight (8) errors here. This confirms the assumption that over time, students were able to write better essays and make less errors.

After the participants completed the last submission, the number of errors in the essays thus submitted were by and large consistent with what was observed in previous stages. Nevertheless, there were some slight changes which can be interpreted as negative, with regard to student’s ability to take full advantage of computer-assisted error correction. In fact, in POST 3, fewer students recorded zero lexical errors, as compared to PRE 3. In addition, while the errors made in PRE 3 ranged from zero to six (6), this number rose to seven (7) errors in POST 3. In the end, a total of 66 errors was obtained. Nevertheless, a close look at the mean and standard deviation here reveals that though a lower number of errors was registered, the proportion of errors in POST 3 is actually slightly higher than in PRE 3. This change might be explained the fact that when trying to correct the errors that are detected but not corrected by the software, students might have ended up making more errors than in earlier versions of their papers.

Finally, after comparing the mean scores obtained in POST 3 to the ones registered in POST 1 (see Table 2) it goes without saying that overall, the students performed slightly better over time, though the difference was not statistically significant, as we shall see later.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Table 2: Lexical errors: mean, standard deviation and range* | | | | |
| Stage | Mean | Std. Deviation | Minimum | Maximum |
| Pre 1 | 2.64 | 2.329 | 0 | 8 |
| Post 1 | 2.21 | 2.302 | 0 | 8 |
| Pre 2 | 1.85 | 1.679 | 0 | 6 |
| Post 2 | 1.36 | 1.220 | 0 | 6 |
| Pre 3 | 2.19 | 1.874 | 0 | 6 |
| Post 3 | 2.28 | 1.791 | 0 | 7 |
| Total | 2.08 | 1.924 | 0 | 8 |

**2.4 Statistical significance tests**

After obtaining the findings described earlier, it was necessary to test the differences between the PRE and POST phases in each stage of the research work in order to find out whether those findings were statistically significant. To decide on which test to use, distribution had to be taken into account, since this would determine whether parametric or non-parametric tests should be used. As the number of participants was less than 50, the figures obtained were subjected to the Shapiro-Wilk test, and the p-values that represented each stage of the research were below 0.05 as evidenced in Table 3. This led to the logical conclusion that lexical errors in the essays did not have a normal distribution and that a non-parametric test had to be used in order to obtain information about the statistical significance of our findings.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Table 3: Shapiro-Wilk tests of normality* | | | | | |
|  | Statistic | | df | | Sig. |
| Lexical errors-pre1 | | .887 | | 33 | .003 |
| Lexical errors-post1 | | .838 | | 33 | .000 |
| Lexical errors-pre2 | | .871 | | 33 | .001 |
| Lexical errors-post2 | | .809 | | 33 | .000 |
| Lexical errors-pre3 | | .892 | | 32 | .004 |
| Lexical errors-post3 | | .887 | | 29 | .005 |

Given the abnormal distribution of the errors, the Wilcoxon Signed Ranks test was selected in order to find out whether the differences between the numbers of errors obtained in the PRE and POST phases were statistically significant. Here again, if the p-values were less than 0.05 there would be a strong indication that the statistical differences could be relied on. Table 4 indicates that there were indeed statistically significant differences between POST 1 and PRE 1 as well as between POST 2 and PRE 2, whereas the difference between POST 3 and PRE 3 was not totally reliable.

|  |  |  |
| --- | --- | --- |
| *Table 4:**Wilcoxon Test results* | | |
|  | Z | Sig. |
| Lexical errors-post1 – Lexical errors-pre1 | -2.648a | .008 |
| Lexical errors-post2 – Lexical errors-pre2 | -2.654a | .008 |
| Lexical errors-post3 – Lexical errors-PRE3 | -.551a | .582 |
| *a. Based on positive ranges* | | |

**3. Discussion**

The findings presented above led to various conclusions regarding lexical errors in free form writing, with a special look at undergraduate students in the degree in English. First of all, this work’s first assumption was validated, since the study confirmed that overall, the number of lexical errors made in any stage of this research work was quite low, which might be explained by two main reasons. First of all, the participants were all university students who majored in English. Therefore, though there is no clear data as to what their proficiency level was, it can be assumed that they would find no major difficulty writing short essays on general topics. Furthermore, the students had unlimited time (within the research timeline, of course) to plan their essays and they were also allowed to use dictionaries and other online resources if they wished. It is therefore not surprising that overall, they were fairly good at avoiding lexical errors.

As concerns the errors that were made, most of them were so intricate that it would be difficult for software to detect and correct them. The foregoing was a confirmation of the second assumption formulated in this work. Therefore, the findings of this paper seemed to go hand in glove with previous research works (Lawley, 2015; Chacón-Beltrán, 2017; Harvey-Scholes, 2017) which indicated that lexical errors in the form of false friends, faulty collocations, and other unusual constructions might go undetected by the software.

**3.1 False friends**

We talk of false friends when words with identical or similar forms have different meanings from one language to another. False friends between Spanish and English have been studied extensively and make up one of the most important aspects of vocabulary teaching. For instance, Chacón-Beltrán (2006) proposed a typological classification of false friends which takes into account graphic and phonetic considerations. Some of Chacón-Beltrán’s findings have been incorporated to Grammar Checker’s database of problem words, which explains why false friends such as “career” (carrera) were effectively identified and corrected by learners. Nevertheless, given that it is very difficult and virtually impossible to build a comprehensive database of all false friends, many still went undetected by the software, as exemplified in the following examples, which were selected at random.

[1] The efforts and **illusions** we put to work when we try to reach a target are far more satisfactory than the joy we have once we reach it.

[2] There was this big ape sitting behind a **crystal** wall

[3] It is a question of **moral**

[4] Without university degrees there would not be **investigators** to evolve new projects

[5] So you could die or your house could burn into ashes if traffic prevents an ambulance or a fire truck to **assist** the emergency on time

[6] There are very **sensible** people that can be very affected, specially those who want to be in fashion or look like the actors that make the advertisements.

A careful analysis of the above sentences leads one to the logical conclusion that the meanings and uses of the Spanish cognates of the words in bold might have led to lexical errors which, of course, failed to be detected by software, because, not only are they very intricate instances of false friends, but they are not part of the software’s database of problem words. To be more specific, the student’s use of “illusion” in Example [1] seems to be incorrect. In fact, while the most common meaning of Spanish “ilusión” is “hope”, “excitement” or “anticipation”, its English cognate meaning tends to refer to a deceptive idea or a false impression. In fact, a broad understanding of the above sentence makes it clear that the learner was not talking about illusions as they are understood in the English language but rather about hope, excitement and anticipation. Therefore, the students’ first language might again have contributed to use of this false friend, which the student could not identify nor correct with the help of *Grammar Checker*. This example is very similar to those in sentences [2] and [3]. To begin with example [2], though “crystal” is a very common word in English, it is normally not used to describe walls, so it was assumed that the student who wrote the sentence most probably wanted to talk about “glass”, which is also translated as “cristal” in Spanish. In addition, “moral” in sentence [3] would as a noun refer to the message one can get from a story or tale. Nevertheless, immediate context in the essay makes us believe that the idea being referred to is that of morals instead. In example [4], the use of “investigators” for “investigadores” might have worked in some situations, but it was found to be less idiomatic than “researchers” in this specific context.

As concerns sentences [5] and [6], Grammar Checker does list “assist” and “sensible” as “problem words” and therefore advises users to pay special attention to how they them. Nevertheless, this is just a recommendation that the student might reject, and this is exactly what happened.

The above examples thus illustrate the extent to which the identification of false friends often depends on a careful analysis of context and collocation. The latter, which is defined as the is the tendency of some words to go together or co-occur, can further complicate computer-assisted error detection and correction. This is exactly why Szpila (2020, pp. 234-235) states that “sometimes collocations provide further specialisation of meaning of false friends (sic), indicating the range of their use with the specific meaning…”

**3.2 Collocations**

As mentioned earlier, the study of collocations is very close to that of false friends. Nevertheless, while the link between form and meaning is emphasized in false friends, collocation involves syntagmatic relations between words, i.e, those lexical items that tend to co-occur in natural language use (Béjoint, 2000). Since collocations are mostly based on use rather than rules which may change over time, it has been very difficult for computational linguists to come up with algorithms which would contribute to the detection and correction of collocation errors. This is why a large number of the lexical errors that were left uncorrected by the participants in this study belonged to collocation.

[7] Today, we give out our privacy to all kinds of social **networks**

[8] therefore, if the traffic is **big**, it will be a really big problem to do it.

[9] Regarding our health, poor air quality affects our **breathing** system.

[10] For example, it takes you five or six more time to cover the same **path** between your house and your job when you drive on early hours of the day or last ones.

[11] in Catalonia for instance, we have fallen into the **gap** of xenophobia

[12] advertising was born as a means of spreading information to the **widest** amount of people

All the examples above provide evidence for situations where native-speaker intuition is the only way to actually gauge whether or not a using some words together is acceptable. Therefore, in natural IT jargon, “social” will certainly not collocate with “network”, but rather with “media” and neither would “traffic” collocate with “big” but rather with words such as “heavy” or “congested” to a lesser extent. Furthermore, though “breathing system” perfectly makes sense, “respiratory system” is the formula used in human biology. What’s more, you would definitely use “cover” with “distance” rather than “path” as in Example [9]. Finally the idiomatic phrases “fall into a trap” and “large amount” should have been used by the authors of sentences [11] and [12].

**3.3 (Near-)homographs/homophones and compound words**

Another set of lexical errors that Grammar Checker could not detect included words that are so similar in spelling and/or pronunciation (examples [13] to [18]) that the students might use them incorrectly. In addition we realised that the software could not detect errors involving closed compound words written as separate words (examples [19], [20] and [21].

[13] Zoo animals usually live under precarious **live** conditions

[14] in museums **were** stuffed animals were represented

[15] areas applying techniques or **practises** that attract public and private investment

[16] They **safe** endangered species from final extinction

[17] they get stressed as a result of **leaving** in small limited

[18] cars are the **principle mean** of transport

[19] In the **mean time**, a thousand ads will display their charms in front of our eyes.

[20] you **can not** find a right place in the middle

[21] It happens **every where**, all western countries are affected

To end this section, it is worth noting that in some cases, it was simply impossible to place some errors into some definite categories as their interpretation was indeed challenging. For instance the use of “point” in “There have been a lot of good people in my life, and most of them usually have had a positive **point** in my daily life” is definitely incorrect but context did not help find out what the student meant and this complicated the classification of the error. Finally, it is important to admit that due to space constraints, only the most prominent types of undetected lexical errors were presented. Nevertheless we can say with a high degree of certainty that the examples presented in this paper are representative of the 402 lexical errors identified in the essays.

**3.4 Semantics, collocation and idiomaticity against computer assisted correction**

This paper has provided further evidence that though there has been great progress with regard to the design of digital writing assistants, the automated detection and correction of lexical errors remains one of the aspects that is not fully mastered, due to issues related to semantics, collocation and idiomaticity. One would realise, for instance, that most (if not all) words in English and Spanish are polysemic. What’s more, languages keep evolving and words take up new meanings on a daily basis. The examples presented in this paper have also illustrated the close link between collocation and semantics, as those cognates which in most situations are semantic equivalents may end up becoming false friends because of meanings they take up when paired up with some other words. Finally, it goes without saying that at times, literal meaning might not help when trying to guess which words should be used in a specific construction. The designers of *Grammar Checker* thought it wise to develop a database which included problem words, common false friends and other erroneous constructions, but this research work has proved that automated correction cannot guarantee the identification and correction of all lexical errors. Nevertheless, one cannot deny the potential of automated error detection and correction so long as it is not used as a replacement, but rather as a complement to traditional teaching (Ware, 2014; Warschauer & Ware, 2006). In fact, students will completely fail to achieve advanced writing skills if they do not master collocations and false friends, and this research is further confirmation that in most cases only teachers can help learners achieve that. Combining the traditional teaching of writing with computer software can therefore make the teaching activity more complete while engaging learners and making them more autonomous. For instance, when providing feedback on free-form writing, teachers may emphasize aspects such as structure, false friends, collocation, etc. while letting students correct other own errors with the help of computer programs. This would, for sure, contribute to a better and more efficient educational experience for both teachers and learners. Furthermore, this paper has proved with solid quantitative data that computer-assisted self-correction can ultimately contribute to helping students gain better writing skills. As a matter of fact, though the automated detection of lexical errors is still weak, students were still able to make less errors by the time the research was completed. This is in line with Chandler (2003), who emphasizes the educational benefits of self-correction when it comes to gaining accuracy in writing.

**4. Final considerations**

Before ending this paper, it is worth mentioning that this study was limited to a sample of essays produced by university students in Spain. Therefore, the findings of this paper are certainly limited to the Spanish-speaking context and might not apply to other first-language contexts. Nevertheless, the limitations of software, especially when dealing with the detection and correction of lexical errors would certainly be the same in other contexts and might even be more drastic when the languages involved are distant from English, thus making it more difficult to build algorithms based on cognates. Therefore, it would be a good idea to carry out similar studies in other contexts in order to come up with knowledge that could contribute to making computer-assisted error detection and correction a possible resource for a greater number of learners. Moreover, this study has also proven beyond reasonable doubt that software-assisted error detection aimed at Spanish learners of English should be improved. It might therefore be a good idea to actually carry out research aimed at obtaining a wider inventory of lexical errors which could then be used to upgrade *Grammar Checker* or any other correction software package.

**5. References**

Agustín-Llach, M.P. (2011). *Lexical errors and accuracy in foreign Language writing*. Multilingual Matters.

Agustín-Llach, M.P. (2014). Lexical errors in writing at the end of Primary and Secondary Education: Description and pedagogical implications. *Porta Linguarum 23*, 109-124

Attali, Y., & Burstein, J. (2006). Automated essay scoring with e-rater v.2 .Technology, Learning, and Assessment, *4*(3), 1-30.

Bejoint, H. (2000). *Modern lexicography. An introduction*. Oxford University Press

Bitchener, J. & Ferris, D. R. (2012). *Written corrective feedback in second language acquisition and writing*. Routledge.

Brill, J. M., & Galloway, C. (2007) Perils and promises: University instructors’ integration of technology in classroom-based practices. *British Journal of Educational Technology* *38*(1), 95-105.

Carrió-Pastor, M. L., & Mestre-Mestre, E. M. (2013). Lexical errors in second language scientific writing: Some conceptual implications. *International Journal of English Studies*, *14*(1), 97-108. https://doi.org/10.6018/ijes/14/1/154361

Cenoz, J. (2003) The influence and age on the acquisition of English: General proficiency, attitudes and code-mixing. In M.P. García-Mayo, & M.L. García Lecumberri (Eds.) *Age and the acquisition of a foreign language* (pp. 77-93). Multilingual Matters.

Chacón-Beltrán, R. (2006). Towards a tyological classification of false friends (Spanish-English). *RESLA*, *19*, 29-39.

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>

Chacón-Beltrán, R. (2018). Vocabulary learning strategies outside the classroom context: What adults learn in a technology-based learner-centred environment. *The Language Learning Journal*, *46*(5), 583-593. <https://doi.org/10.1080/09571736.2018.1503135>

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*, 267-296.

Chodorow, M., Gamon, M., & Tetreault, J. (2010). The utility of article and preposition error correction systems for English language learners: Feedback and assessment. *Language Testing*, *27*(3), 419-436.

El Ebyary, K. & Windeatt, S. (2010). The impact of Computer-based feedback on student s’ written work. *International Journal of English Studies*, *10*(2), 121-142. DOI: 10.6018/ijes.10.2.119231

Ellis, R., Sheen, Y., Murakami, M., & Takashima, H. (2008). The effects of focused and unfocused written corrective feedback in an English as a foreign language context. *System*, *36*(3), 353-371.

Ene, E. & Upton, T. A. (2014). Learner uptake of teacher electronic feedback in ESL composition. *System*, *46*, 80-95.  <https://doi.org/10.1016/j.system.2014.07.011>

Fazio, L. (2001). The effect of corrections and commentaries on the journal writing accuracy of minority-and majority-language students. *Journal of Second Language Writing*, *10* (4), 235-249.

Ferris, D. R. (2002). *Treatment of error in second language student writing*. University of Michigan Press.

Ferris, D. R. & Roberts, B. (2001). Error feedback in L2 writing classes. How explicit does it need to be? *Journal of Second Language Writing*, *10*, 161-184.

García-Lecumberri, M.L. & Gallardo, F. (2003). English FL sounds in school learners of different ages. In M.P. García-Mayo, & M.L. García Lecumberri (Eds.) *Age and the acquisition of a foreign language* (pp. 115-135). Multilingual Matters.

Guichon, N., Betrancourt, M. & Prié, Y. (2012). Managing written and oral negative feedback in a synchronous online teaching situation. *Computer Assisted Language Learning*, *25*(2), 181-197.

Harvey-Scholes, C. (2017). Computer-assisted detection of 90% of EFL student errors. *Computer Assisted Language Learning*. http://dx.doi.org/10.1080/09588221.2017.1392322

Hernandez Puertas, T. (2018). Teachers’ feedback vs. Computer-generated feedback: A focus on articles. *Language Value*, *10*(1), 67-88. <http://dx.doi.org/10.6035/LanguageV.2018.10.5>

Kepner, C. G. (1991). An experiment in the relationship of types of written feedback to the development of second-language writing skills. *The Modern Language Journal*, *75*, 305–313.

Lawley, J. (2015). New software to help EFL students self-correct their writing. *Language Learning & Technology*, *19*(1), 23-33.

Lee, I. (2013). Research into practice: Written corrective feedback. *Language Teaching*, *46*(1), 108-119. doi: 10.1017/S0261444812000390

Lee, Y-W., Gentile, C., & Kantor, R. (2010) Toward automated multi-trait scoring of essays: Investigating links among holistic, analytic, and text feature scores. *Applied Linguistics*, *31*(3), 391-417. http://dx.doi.org/10.1093/applin/amp040.

Makino, T. Y. (1993). Learner self-correction in EFL written compositions. *ELT journal*, *47*(4), 337-341.

Morrison, G.R., Ross, S.M., Kemp, J.E:, & Kalman, H. (2010). *Designing effective instruction. Applications of instructional design* (6th edition). Wiley.

Senra-Silva, I. (2010). Designing computer-generated pedagogical feedback for Spanish students of EFL. *Revista Española de Lingüística Aplicada*, *23*, 281-296.

Sheppard, K. (1992). Two feedback types: Do they make a difference? *RELC Journal*, *23*, 103–110.

Szpila, G. (2020). Collocational aspects of false friends. In J. Lesniewska, & E. Witalisz, (Eds.) *English and American Studies in the age of Globalization*. Jagiellonian University Press.

Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning*, *46*(2), 327–369.

Ware, P. (2011). Computer-generated feedback on student writing. *TESOL Quarterly*, *45*(4), 769-774. <https://doi.org/10.5054/tq.2011.272525>

Warschauer, M. (2010). Invited commentary: New tools for teaching writing. *Language Learning & Technology*, *14* (1), 3–8.

Warschauer, M., & Ware, P. (2006). Automated writing evaluation: Defining the classroom research agenda. *Language Teaching Research*, *10*(2), 1-24.