

Improving lexical errors in EFL writing by using software-mediated corrective feedback

RUBÉN CHACÓN BELTRÁN
ORCID: 0000-0002-3055-0682

RAYMOND ECHITCHI
ORCID: 0000-0003-4154-7961
National University of Distance Education

Received: 26 April 2021 / Accepted: 23 January 2022
DOI: <https://doi.org/10.30827/portalin.vi37.20847>
ISSN paper edition: 1697-7467, ISSN digital edition: 2695-8244

ABSTRACT: This research aims at revisiting the role of software when it comes to providing learners with corrective feedback on their pieces of writing. The study, based on the analysis of handwritten and software-corrected versions of essays written by 33 undergraduate students enrolled in the undergraduate degree programme in English Studies at a Spanish University contributed to confirming the assumption that technology can indeed be a useful tool in the teaching and learning process. More specifically, this study 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 can improve on their ability to avoid such errors. Nevertheless, the study has also confirmed that software can in no way completely replace teachers, as computer programming is quite limited and there are errors that only proficient language users can detect and correct.

Key words: CALL, feedback, self-correction, lexical errors, Grammar Checker

La mejora en los errores léxicos de la expresión escrita en inglés como lengua extranjera mediante el uso de feedback correctivo automatizado

RESUMEN: El objetivo de este trabajo de investigación es reconsiderar el papel de las herramientas informáticas para proporcionar feedback sobre los errores de alumnos y mejorar su escritura. Este estudio, basado en el análisis de las versiones escritas y corregidas por programa de varias composiciones elaboradas por 33 estudiantes del Grado en estudios ingleses por una universidad española, 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. En este trabajo se vió que los estudiantes pudieron reducir significativamente el número de errores léxicos en sus composiciones 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

Feedback on errors has always been a very important topic in academic circles, especially amongst scholars interested in second language acquisition. The efficiency of error feedback became a very controversial topic after Truscott (1996) wrote a paper that challenged the positive effects of feedback on learning. This sparked a debate between Truscott and Ferris (Ferris 1999; Truscott 1999), which eventually spread to the academic community as a whole. While Truscott has always challenged the effectiveness of feedback -something he reiterates in his recent outings (Mohebbi, 2021)-, the overwhelming majority of researchers insist that feedback is necessary because it makes students aware of their errors and may eventually help them seek ways to avoid those errors (Bitchener, 2012; Ferris, 2015; Lee, 2013, 2016). Therefore, the research we carried out was built on the belief that error feedback is indeed a very important component of the teaching activity and that the most important thing is to find out ways in which teachers can effectively detect, correct and provide additional on student errors without affecting the overall academic achievement those students.

The research presented in this paper was also motivated by present-day technological advances, which have led to growing interest in the use of technology in the classroom. In fact, the use of technology in teaching activities could contribute to fostering autonomous learning, especially when it comes to error feedback and correction (Chacón-Beltrán, 2018). This article will therefore attempt to provide answers to the following questions: 1) To what extent can automated error-detection contribute to reducing the number of errors in student writing? 2) What types of errors may software find it difficult to detect?

2. THEORETICAL BACKGROUND

2.1. Lexical errors and crosslinguistic influence

Agustín-Llach (2011) defines lexical errors as “deviation[s] in form and/or meaning of a target-language lexical word” (p. 75), which include spelling mistakes as well as other incorrect sequences like erroneous collocations and false friends. The author then 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. As concerns English in the Spanish context, apart from studies which involved primary or/and secondary school students (Agustín-Llach, 2011; 2015), other researchers such as Carrió-Pastor and Mestre-Mestre (2013) focused on the errors made by adult learners as a whole.

With regard to crosslinguistic influence, while most studies suggested that less proficient learners are more likely to make errors related to first language (L1) influence (see Verspoor et al.'s 2012 and Olsen's 1999 studies of L2 English amongst Dutch and Norwegian learners of English, respectively), some studies yielded results contrary to this. For instance, Mukattash (1986) found crosslinguistic influence to be more frequent amongst native Arabic speakers with advanced English proficiency. Far beyond this controversy what remains clear

is that no matter their proficiency level, learners cannot completely avoid errors resulting from crosslinguistic influence (Olsen, 1999). Furthermore, it is important for researchers who are interested in studying lexical errors in a second language to move beyond the assumption that those errors would decrease in number as proficiency increases. In fact, as Agustín-Llach's (2015) study reveals, while some specific lexical errors resulting from literal translation and semantic confusion might indeed disappear as the learner becomes more proficient, cases of misselection and coinage might become more frequent amongst high-proficiency learners. Agustín-Llach's conclusion parallels claims made earlier by Cenoz (2003), as well as García Lecumberi and Gallardo (2003) who carried out similar studies amongst Spanish learners of English.

2.2. Corrective feedback

Corrective feedback has attracted the attention of many researchers (Chodorow, Gamon & Tetreault, 2010; Ellis, et al., 2008; Makino, 1993; Senra-Silva, 2010), since it has always played a fundamental part in the teaching and learning process. Nevertheless, as we indicated in the introduction, there has been a big 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 stated that it has a negative effect on learning and should be completely avoided. Truscott built those claims on earlier publications (Kepner, 1991; Sheppard, 1992), which suggested that corrective feedback contributed very little to improving learning. Reasons for the unsuitability of error correction often revolve 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 publications (Lee, 2013; 2016; Bitchener, 2012; Ferris, 2015;) have proved that when done the right way, error correction and feedback could definitely contribute positively to learning, but they have also been criticized by Truscott, who insists that no researcher can prove beyond a reasonable doubt that error correction has positive effects on learning (Mohebbi, 2021). As this debate is nowhere close to ending, we decided to build this research on the assumption that error feedback is indeed necessary, as supported by most researchers.

One study worth mentioning here is that of Chandler (2003), who studied two groups of Asian (from Korea, Japan, China and Taiwan) learners of English 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) paper

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*.

2.3. Computer-generated scoring and feedback

In today's world technology and computers are 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 on those errors. As concerns the use of computer programs to teach writing, Ware (2011) insists that it is important to differentiate between what she refers to as computer-generated feedback, on the one hand, and computer-generated scoring, on the other hand. Despite the technological advances that have been achieved since 2011, Ware's claim that it is very difficult to design software programs that could grade writing assignments is still valid. This is definitely why this paper rather focuses on the use of computers to help students identify and correct their essays (computer-generated feedback). Computer programs and applications can be valuable tools in the classroom, since they would 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 is that "they may not have time to give individualized, immediate, content-related feedback to multiple drafts" (p. 122). Therefore, teachers are often left with no choice but to provide students with delayed feedback, which many researchers found to have very little effect on learning (Chandler 2003; Warschauer, 2010; Guichon et al., 2012).

Despite the large number of publications that make a case for the use computer-generated feedback in the classroom (Attali and Burnstein 2006; Lee, Gentile and Kantor, 2010), its incorporation has been shunned by many educational stakeholders. For instance, Ware (2011) reveals that teachers may be reluctant to resort to computer-generated feedback because many of them 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 move beyond assessing the efficiency of computer-generated feedback and discuss ways in 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).

3. METHOD

3.1. Participants

This project was completed thanks to the participation of university students enrolled at a Spanish distance learning university. Once the research was designed, the authors of this article made use of the university's online learning platform to post a call for participation.

The latter reassured students that their essays would not be graded, as grading the activity could have discouraged, or at least might have added unnecessary stress on them. The researchers also made sure they encouraged the students by telling them that this research would not only give them the opportunity to practice their writing skills, but also earn them extra marks (all prospective participants were undergraduate students enrolled in the degree programme in English Studies and took modules taught by the researchers). All in all, a total of thirty-three (33) students expressed their interest in participating in the project, and they were asked to provide more information about themselves by completing an online questionnaire. The participants included 28 women and 5 men aged between 21 and 72. The age brackets with the largest number of informants were 41-50 (n=12) and 31-40 (n=10), and the mean age was over 40 years old. 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 on the researchers' experience as lecturers, it is clear that these pieces of data on gender, age and mother tongue are representative of the student population in the undergraduate programme in English Studies at the aforementioned university. In fact, distance learning tends to appeal to people who cannot study in traditional universities because of family and professional obligations. The English Studies student population mostly consists of middle aged individuals who are looking for ways to improve their English skills and climb the career ladder. Furthermore, most of them somehow have ties to Spain or the Spanish language, in other words, they are Spanish citizens, are mother tongue speakers of Spanish or foreigners who have lived in Spain or a Spanish-speaking country for some time.

3.2. Grammar Checker

Before describing data collection and analysis it is important to describe the computer tool that was used. *Grammar Checker*. *Grammar Checker* is an error-detection program that was developed by researchers belonging to the linguistics and computer science departments at two universities based in Spain and in the United Kingdom to help Spanish-speaking learners of English gain writing accuracy autonomously. More specifically, it is equipped with a word processor that allows students to type/paste their essays and obtain corrective feedback.

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 draws conclusions on whether the words are spelt incorrectly or very infrequent. In relation to "Incorrect Sequences", it compares the text being processed with some of the most frequent incorrect phrases in English and provides users 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.

As an error-detection program, *Grammar Checker* does not actually correct errors, but rather points at different types of errors (or possible errors in the case of the “Problem Words” filter) by highlighting them. Therefore, the program guarantees that students remain actively involved in the learning process, since after the software detects errors, the students are supposed to find ways to correct them.

3.3. Procedure

The findings described here 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 find out more about 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 write their essays on. 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 type the essays verbatim 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 essays and the difficulties they faced when going about that. It is important to note that, while the participants were allowed to use as many resources as possible to draft their handwritten essays, they were instructed to use only *Grammar Checker* to correct their essays.

After they had received both versions of the essays, the researchers would proceed to compare the handwritten and software-corrected versions and send participants overall feedback. This feedback consisted in highlighting errors that the students had failed to correct and providing them with some useful tips on how to avoid 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 handwritten and software-corrected versions of the essays with a view to comparing both versions and thus finding out 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 two (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.

This research was based on the assumption that lexical errors would be fewer in number, given 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 study’s second assumption is that software correction would not be fully efficient at detecting certain words because of its technical limitations.

4. FINDINGS

As our research was first of all meant to be quantitative, we started with the statistical analysis of the lexical errors that had been identified. Later on, statistical hypothesis tests were carried out in order to determine how significant our results were.

Table 1: Lexical errors: stage crosstabulation

		STAGE 1		STAGE 2		STAGE 3		
		PRE	POST	PRE	POST	PRE	POST	TOTAL
NUMBER OF PARTICIPANTS		33	33	33	33	32	29	
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

4.1. Stage 1

This stage represents the very first essays the informants produced. 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, 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. After computer-assisted correction the number of errors was reduced, and this time around, 73 errors were registered. In other words, computer-assisted correction contributed 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 smaller number of errors, as found in Table 2.

4.2. 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 that in this stage the number of errors ranged from zero to six (6), i.e., no student made seven (7) or eight (8) lexical errors in Stage 2 (as opposed to Stage 1). 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 our informants.

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 smaller number of errors remained consistent as reflected in the low mean (1.36) and standard deviation (1.22) (see Table 2).

4.3. 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% less 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 essays containing seven (7) or eight (8) errors here. This confirms the assumption that over time, students were able to make fewer errors in their essays.

After the final 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 students' 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 could reveal that when trying to correct the errors that were detected 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

4.4. 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 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

	POPULATION	P-VALUE
LEXICAL ERRORS-PRE1	33	.003
LEXICAL ERRORS-POST1	33	.000
LEXICAL ERRORS-PRE2	33	.001
LEXICAL ERRORS-POST2	33	.000
LEXICAL ERRORS-PRE3	32	.004
LEXICAL ERRORS-POST3	29	.005

The Wilcoxon signed-rank test was therefore 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 statistically significant.

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		

5. DISCUSSION

Before addressing the research questions formulated in this paper, it is important to note that this study confirmed our initial belief that the number of lexical errors made throughout the project would be quite low because the participants were all university students who majored in English. Though there are no clear data as to what their proficiency level was, university students in a degree programme taught exclusively in English would have little difficulty writing short essays on general topics. Furthermore, the students had unlimited time (within the research timeline, of course) to plan and write their essays and they were also free 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 our first research question, it seems software indeed contributed to further reducing the number of errors in the essays. The fact that the differences between handwritten and revised essays were statistically significant confirms that software did contribute to students' eliminating some of those errors. The limitations of software could therefore explain why some errors the students made initially were left uncorrected. In fact, a close look at those errors seems to confirm Lawley's (2015), Chacón-Beltrán's (2017) and Harvey-Scholes's (2017) belief that software might often fail to detect lexical errors in the form of false friends, faulty collocations, and other unusual constructions. Therefore, the following sections will discuss the types of errors that *Grammar Checker* could not detect, in an attempt to answer our second research question.

5.1. False friends

We talk of false friends when words with identical or similar forms have different meanings when moving 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, as exemplified in the following fragments, 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 *Grammar Checker*, 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. 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 they 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 use them. Nevertheless, this is just a recommendation that the student might reject, and this is exactly what happened.

5.2. 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 ease the detection

and correction of collocation errors. This is why a large number of 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 definitely 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 biology. What’s more, one would definitely use “cover” with “distance” rather than “path” as in Example [9]. Finally the idiomatic phrases “fall into a trap” and “large number” should have been used by the authors of sentences [11] and [12].

5.3. (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 within 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.

5.4. 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 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 here 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 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 study 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.

6. LIMITATIONS

Before ending this article it is worth mentioning some issues that may affect the validity of our findings. First of all, it would be impossible state for sure that it is only thanks to *Grammar Checker* that the errors the participants made in their essays were indeed corrected. Mere proofreading, skills they might have acquired over time, and the use of resources other than *Grammar Checker* (though participants were instructed no to do this) could have helped students eliminate some of those errors. Therefore, we cannot say that this study really tested how efficient *Grammar Checker* is. For such a study to be carried out, researchers should make sure no other resources are used and that students write and correct their essays with no delay. Unfortunately, it is very complicated to carry out such research in distance education contexts, which is why we could not do it.

The second limitation worth mentioning here is the sample itself, which is quite reduced and complicates any attempts at generalising our results. Therefore, this study can only be used to build assumptions and motivate similar studies which should be carried out amongst a larger number of informants in order to be generalised.

7. 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. (2012). Written corrective feedback for L2 development: Current knowledge and future research. *TESOL Quarterly*, 46(4), 855–867. <https://doi.org/10.1002/tesq.62>
- 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. <https://doi.org/10.1111/j.1467-8535.2006.00601.x>
- 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. [https://doi.org/10.1016/S1060-3743\(03\)00038-9](https://doi.org/10.1016/S1060-3743(03)00038-9)
- 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. <https://doi.org/10.1177/0265532210364391>
- 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. <https://doi.org/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. <https://doi.org/10.1016/j.system.2008.02.001>

- 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. [https://doi.org/10.1016/S1060-3743\(01\)00042-X](https://doi.org/10.1016/S1060-3743(01)00042-X)
- Ferris, D. R. (1999). The case for grammar correction in L2 writing classes: A response to Truscott (1996). *Journal of Second Language Writing*, 8(1), 1–11. [https://doi.org/10.1016/S1060-3743\(99\)80110-6](https://doi.org/10.1016/S1060-3743(99)80110-6)
- Ferris, D. R. (2015). Written corrective feedback in L2 writing: Connors & Lunsford (1988); Lunsford & Lunsford (2008); Lalande (1982). *Language Teaching*, 48(4), 531–544. <https://doi.org/10.1017/S0261444815000257>
- 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. <https://doi.org/10.1080/09588221.2011.636054>
- 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. <https://doi.org/10.2307/328724>
- 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. <https://doi.org/10.1017/S0261444812000390>
- Lee, I. (2016). Teacher education on feedback in EFL writing: Issues, challenges, and future directions. *TESOL Quarterly*, 50(2), 518–527. <https://doi.org/10.1002/tesq.303>
- 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. <https://doi.org/10.1093/elt/47.4.337>
- Mohebbi, H. (2021). 25 years on, the written error correction debate continues: an interview with John Truscott. *Asian-Pacific Journal of Second and Foreign Language Education*, 6(3). <https://sfl.education.springeropen.com/articles/10.1186/s40862-021-00110-9>
- Morrison, G.R., Ross, S.M., Kemp, J.E, & Kalman, H. (2010). *Designing effective instruction. Applications of instructional design* (6th edition). Wiley.
- Mukattash, L. (1986). Persistence of Fossilization. *IRAL*, 24: 187-203. <https://doi.org/10.1515/iral.1986.24.1-4.187>
- 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. <https://doi.org/10.1177/003368829202300107>
- Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning*, 46(2), 327–369. <https://doi.org/10.1111/j.1467-1770.1996.tb01238.x>
- Truscott, J. (1999). The case for “the case against grammar correction in L2 writing classes”: A response to Ferris. *Journal of Second Language Writing*, 8(2), 111–122. [https://doi.org/10.1016/S1060-3743\(99\)80124-6](https://doi.org/10.1016/S1060-3743(99)80124-6)
- Olsen, S. (1999). Errors and compensatory strategies: a study of grammar and vocabulary in texts written by Norwegian learners of English. *System*, 27, 191-205. [https://doi.org/10.1016/S0346-251X\(99\)00016-0](https://doi.org/10.1016/S0346-251X(99)00016-0)
- Verspoor, M., Schmid, M.S. & Xu, X. (2012). A dynamic usage based perspective on L2 writing. *Journal of Second Language Writing*, 21(3), 239-263. <https://doi.org/10.1016/j.jslw.2012.03.007>
- 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. <https://doi.org/10.1191/1362168806lr1900a>