

The IALLT Journal

A publication of the International Association for Language Learning Technology

Measuring the Impact of Online Translation on FL Writing Scores

Errol M. O'Neill
University of Memphis

ABSTRACT

Online translation (OT) sites, which automatically convert text from one language to another, have been around for nearly 20 years. While foreign language students and teachers have long been aware of their existence, and debates about the accuracy and usefulness of OT are well known, surprisingly little research has been done to analyze the actual effects of online translator usage on student writing. The current study compares the scores of two composition tasks by third- and fourth-semester university students of French who used an online translator, with or without prior training, to the scores of students who did not use OT. Students using an online translator did not perform significantly worse than those not using the translator on either task. In fact, students who received prior training in OT outscored the control group overall on the second writing task. Additionally, students using the online translator received higher subscores on one or both writing tasks for features such as comprehensibility, spelling, content, and grammar. The results of the current study are discussed in detail; implications for the foreign language classroom are presented; and avenues for future research are proposed.

INTRODUCTION

As the world becomes increasingly connected through the use of the Internet and related technologies, foreign language (FL) students also have access to the same resources available to the general public. Some of these resources (e.g., electronic dictionaries, eBooks) resemble and fulfill similar roles to those that have been used for decades in the FL classroom (e.g., paper dictionaries, textbooks). The computer age has also brought about new tools that may help or hinder second language (L2) acquisition. One such technology, online translation (OT), offers near-instantaneous translation of inputted text from one language into another. OT sites, which are free and available to the general public, are used to translate over 500 million words daily (Koehn, 2010). Through OT, it is possible to write partially or entirely in one's native language, enter what one has written into the online translator, and receive a computer-generated text in the target language.

OT has existed for nearly two decades, with some authors referring to it as Web-Based Machine Translation, or WBMT. Yang and Lange (1998) noted that within a year of the 1997 launch of the first online translator, Babel Fish, language students had already begun using OT to try to write in their L2. Students are still well aware of the existence of this technology, with many availing themselves of OT even when explicitly prohibited from doing so by their instructor (White and Heidrich, 2013). Given OT's longevity, as well as student and teacher awareness of its existence, it is surprising that there has been relatively little research done into online translators. In particular, no quantitative studies have been found that compare the writing of students using an online translator to those who do not use OT sites. The present study aims to address this issue by comparing the scores of students writing compositions with the aid of an online translator, either with or without prior training on the potential strengths and pitfalls of OT, to those of students who did not use a translator.

REVIEW OF LITERATURE

Perceptions about, and recommendations concerning, online translator usage have been addressed in the literature. A spectrum of viewpoints can be found, ranging from complete rejection of this technology (because of errors believed to be made by online translators, or due to the view that the computer, and not the student, may be the one doing the work), to embracing the use of OT for different

purposes in the classroom. One approach has been to focus on the errors produced by online translators.

Luton (2003) described a number of error types encountered while using OT, including those related to mistranslated idioms, proper names, colloquial language, and words misspelled in the L1 that were not processed by the translator. These errors were highlighted to aid instructors in detecting prohibited use of OT in compositions. McCarthy (2004) recommended the limited use of OT in the classroom specifically in order to discourage students studying translation from using online translators. He offered 12 suggestions to teachers who received texts with content produced via OT, including giving students regular lessons on the deficiencies of online translators, having them submit a first draft of all assignments in class to avoid access to OT sites, or imposing academic penalties on those caught using an online translator. Aiken and Wong (2006) found a low level of accuracy for translating 20 sentences from a beginning-level Spanish textbook into English, with grammatical and lexical accuracy ranging from 55% to 75% as judged by missing or extraneous words, improper lexical choices, and total number of correct words. Williams (2006) found that online translators “produce inaccurate, unacceptable translations” (p. 567), in particular for prepositions, adjectives, nouns, and verb phrases in French. Stapleton (2007), while discussing various online resources available to L2 learners, singles out OT when citing the “unethical and damaging nature of translations and electronic ‘lifting’.” (p. 187).

Steding (2009) similarly presented online translator use as cheating or plagiarism and described the role of instructors as threefold: detection, reaction, and prevention. The first of these involves learning to identify OT use through knowing one’s students well, being familiar with typical mistakes for a given level, and understanding the limits of the technology. Reaction to OT, according to Steding, should involve testing out suspected cases of online translator use online and consulting with colleagues. Lastly, instructors can try to prevent OT use by creating clear sanctions, making assignments that would be difficult to complete using an online translator (e.g., those that ask students to use specific words or structures), requiring students to sign statements saying they did not use an online translator or other unauthorized help, and demonstrating to students clear or funny errors produced by an online translator. While many of these suggestions may prove helpful, they generally assume that output from online translators will be poor or at least recognizable and place the “burden of proof” for suspected use of OT on the student (p. 186). Fountain and Fountain (2009) acknowledged that while “substantial progress” has likely been made in

translation technology since the 1990s — no detailed investigation into this point was provided or found — they nonetheless recommended against general online translator use by students. Instead, the authors suggested using “carefully selected examples” as a way to warn students of the technology’s limitations and decrease the desire of students to use online translators.

Other early and more recent authors have highlighted possible benefits to using OT. Cribb (2000) found that while Babel Fish produced some lexical errors translating between English and French or German, it successfully translated the majority of semantic content, in particular denotative meanings. For this reason he concluded that “despite the current weaknesses of MT [machine translation] output, current capabilities warrant its serious consideration as a bridge across the global language gap” (p. 565). Abraham (2009) highlighted some limitations of OT, including difficulties with translating polysemic words, false cognates, mass versus count nouns, compound nouns, possession, and TAM (tense, aspect, mood). Despite these limitations, Abrahams found OT sites could still be beneficial as they “allowed learners to become aware of grammatical, lexical, and orthographic problems” (p. 75) while increasing students’ awareness of how languages work through comparing differences between English (their native language) and Spanish (their L2). In his study of third-year students of Spanish, pairs of participants were able to identify and resolve 63% of errors in the OT output, leading Abrahams to recommend using OT to foster linguistic awareness and discussion among students. While Somers (2003) expressed caution at using OT beyond its originally intended use — obtaining a translation of a text written in a language that is not one’s own native tongue in order to get a gist of the meaning — she similarly suggested that learning the weaknesses and strengths of machine translators (including OT) should be a part of foreign language curricula to help students understand differences between their native and target languages.

Another avenue for analysis has been comparing the production of online translators with that of language learners. Ablanedo et al. (2007) painted a positive picture overall for OT. Of ten English-language text samples translated by Babel Fish into Spanish, all were judged to be usable and understandable, with 70% accuracy for OT as compared to an intermediate-level human translator who achieved 80% accuracy. Niño (2008) compared the errors made by two groups of advanced learners of Spanish: an experimental group who post-edited the raw output of an English-to-Spanish translation produced by an online translator, and a control group who performed a translation into Spanish of the same English text without access to the output from the online translator. Niño found that the post-editing group, whose participants used the text produced by

OT as the starting point for their translation, had a lower number of errors in three of four categories (lexical, grammatical, and spelling) than students not using the online translator output. The control group had a slightly lower mean than the post-editing group for discursive errors, but this did not reach the level of statistical significance. Since the study found that the post-editing group generally made similar types of errors to the control group while making fewer of them overall, Niño advocated the use of output from online translators for teacher-selected texts in order to help advanced language learners become aware of differences between their L1 and L2 and to assist them in identifying and correcting errors made in the target language.

In a later article, Niño (2009) presented a number of additional arguments in favor of allowing students to use OT, including the fact that online translators provide immediate results for students' language needs, the success OT can have with simpler sentence structures, the opportunity OT provides to beginning and intermediate students to gist more advanced texts in the L2 that might otherwise be inaccessible to them, and the ability to consider both successful and failed online translator output in order to highlight differences between languages.

Three more recent articles have also looked at the issue of OT in the context of classroom learning. O'Neill (2013) found that instructors can generally detect whether OT had been used for French compositions written by students. While this result reached statistical significance ($p < 0.001$), the raw numbers showed that the instructors who rated the compositions correctly judged OT use only 70.9% of the time, meaning that in nearly 30% of cases instructors either thought a translator had been used when it had not (12.9% of all cases) or did not suspect OT usage even when one had been used (16.4% of judgments). Although representing the results of one study, this finding may cast some doubt on the efficacy of prohibiting OT and penalizing suspected cases of its use since a number of students may be using OT without getting caught, while others may be incorrectly suspected or punished.

Larson-Guenette (2013) surveyed second- through sixth-semester students of German to find out their self-reported frequency of use for web-based technologies, including online translators. Out of 71 respondents, 68% indicated using OT sites within the past semester. Despite most of them using OT, students had mixed impressions of its effects and accuracy. Some students in follow-up interviews questioned the accuracy of online translators and warned against becoming dependent on them, while others spoke positively of OT as a way to look up unknown words or to check their sentences after trying first on their own.

Larson-Guenette's recommendations included having a workshop with students concerning OT sites and other online resources, as well as acknowledging and creating a dialogue with students about the use of such tools.

White and Heidrich (2013) had the most comprehensive look to date concerning OT in the context of FL writing. In addition to finding that 12 of the 18 participants reported using OT in spite of it being prohibited on their syllabus, White and Heidrich's exploratory study focused on German L2 students' writing when using an online translator. Errors in output using Google Translate were overwhelmingly related to structure (phrasing, complexity, word order, and auxiliaries), resulting in 109 out of 215 errors. Additional errors included contextual (word choice and typos) and nominal (declension), each accounting for 46 mistakes, as well as 14 verbal errors (tense and conjugation). Looking at specific features, phrasing alone accounted for 64 errors, representing 29.8% of all errors, followed by declension (21.4%), word choice (20%), and word order (15.8%). Students were able to correct these errors in OT output in some cases — for example, structural errors decreased from 109 to 86 after editing — while in other cases, students either did not correct errors or actually introduced new errors not present in the online translator output (e.g., nominal errors increased from 46 in the raw output to 57 in the edited texts submitted by students).

While providing detailed analysis concerning the types of errors in raw and edited online translator output, as well as enlightening descriptions of students' beliefs and attitudes towards online translators (outside the scope of the current discussion), their study did not attempt to analyze writing through assigning overall composition scores, or comparing the writing of students who used OT to those who did not. These facets are explored in the current research, which attempts to address the desire of a growing number of researchers (e.g., Williams, 2006; Niño, 2009; White and Heidrich, 2013) to look beyond trying to discourage or prohibit the use of OT — a technology that is freely available and widely used — and investigate more closely any effects online translator use may have on FL writing.

A study was conducted to investigate three aspects of the issue of OT in the classroom.

- **Research question 1a:** Does the use of online translators by L2 writers of French result in quantifiably different global scores on compositions written in the L2?

- **Research question 1b:** Of students who use an online translator, do those who have been trained in the use of online translators achieve higher scores on L2 compositions than those who have not received such training?
- **Research question 2:** Does online translator usage affect perceived performance on any specific features of L2 learners' writing, including comprehensibility, vocabulary, syntax, grammar, spelling, and content?

METHODS

Participants

Thirty-four university students participated in a six-week study investigating the effects of online translator usage on L2 writing. All participants were enrolled in one of three French courses at the time of completing the study: an accelerated course covering second- and third-semester French (having reached material from the third semester at the time of the study), the third-semester course, or the fourth-semester course. Two participants were eliminated from consideration: one due to indicating a native language other than English on a background questionnaire (to avoid a possible confounding variable), and another due to not completing the final experimental task (written posttest), leaving 32 participants in all whose results are considered.

Procedures

All participants were randomly assigned to one of three conditions. Those in Group A (control group, 10 participants) had no translator access during the writing tasks and no translator training, both of which are described below. Students in Group B (11 participants) were allowed access to an online translator but had no training in OT prior to using it. Participants in Group C (11 participants) were allowed access to the online translator and received training in its use prior to completing the writing tasks. Each participant completed the following:

- **a self-report background questionnaire**, which was given to all participants to collect data concerning biographical information and previous experiences. Besides the participant excluded due to having a native language other than English, no one else was determined to warrant exclusion from analysis.

- **a reading pretest**, which consisted of an excerpt from the practice version of the College-Level Examination Program (CLEP) for French. This pretest was used for the purposes of evaluating student reading level at the onset of the study.

- **a written pretest** (Appendix A), a three-paragraph composition written a week after the reading pretest. The written pretest was administered to gauge student writing level prior to the tasks described below. Students had thirty minutes to write, and were not allowed to use OT on the written pretest in order to measure their writing abilities without the use of an online translator. The prompt of this composition, as well as those described below, followed a format adapted from Scott (1996) and were written after consulting the 2001 ACTFL written proficiency guidelines (ACTFL, 2001) as well as the subject matter and vocabulary covered in the participants' textbooks.

- **one of two instructional sessions**, given the week after the written pretest and according to group assignment. Group C participated in an online translator training session developed in part based on the recommendations of Burton (2003) and the discussion of Luton (2003) highlighting potential strengths and shortcomings inherent in OT. An excerpt of the training is found in Appendix B. The session was designed to raise awareness among participants about what online translators are, what they can do, and what ways OT might be beneficial or detrimental to their writing. Only participants in Group C attended this training in order to see whether implicit and explicit instruction related to OT could improve student writing when using a translator; the other two groups attended a cultural session about the field of translation designed as a control to the training session.

Participants in the training first translated from French (their L2) into English (their L1) in order to see from a native speaker's perspective that the output of OT can be correct or understandable; in the sentence "*Ma voisine est sympa, mais elle est trop curieuse*" *ma voisine* was correctly translated as *my neighbor*. On the other hand, students saw that online translators can give output that is incorrect or may lead to confusion (e.g., *sympa* and *elle* in the same sentence being mistranslated as *sympa* and *it* instead of *nice* and *she*, respectively).

Next, students were given a series of sentences to translate from English to French that highlighted cases where an online translator might give a correct or understandable translation in their L2 (for example, *I don't like your house* correctly translated as *Je n'aime pas vraiment votre maison*). These were contrasted with translations that would be incorrect or incomprehensible to a

native French speaker, e.g., *she wrote it down* ~ *elle l'a écrit avalent* with the extraneous verb “*avalent*” (meaning *they down* in the sense of *they swallow*) mistranslating the phrasal verb particle “*down*”.

Lastly, students were given an overview, based in part on the sample translations they had just done, of 14 possible strengths and weaknesses of online translators. The training was summarized in part by the conclusion, “If you know what translators generally are good at doing or what they’re bad at doing, it can help you to use them more effectively in a given situation” (Appendix B).

- **two writing tasks** (Tasks One and Two, Appendices C and D), given a week apart, which instructed participants to write a three-paragraph composition. The writing prompts followed the same format as that of the pretest, but each was on a different topic. The instructions given to participants were identical to those in the pretest, except that Groups B and C were told to use the OT site Free Translation (<http://freetranslation.com>; SDL, 2012) when writing their compositions. This way, the results of Group A (no online translator use) could be compared with the scores of Group B (which did not receive training in OT before writing with the translator) and Group C (which received prior training). Participants were free to use the online translator as little or as much as they wanted to aid in their writing. Although participants in the online translator training group practiced with both Free Translation and Babel Fish during training, Free Translation was chosen for the writing tasks since the researcher found that Babel Fish presented errors with processing certain types of apostrophes: a straight apostrophe ' (e.g., in the French expression *C'est*) would process output as expected, while the smart or “curly” apostrophe ’ (e.g., *C'est*) would leave the word containing the apostrophe untranslated.

- a **written posttest**, which was equivalent in form and procedure to the pretest but on a different topic (Appendix E). The written posttest was given to detect any change in writing level among the three groups at the end of the study, once again without the use of OT.

- finally, a **self-report exit questionnaire** that was administered the same day as the written posttest to collect participants’ feedback on their performance. The results of this questionnaire fall outside the scope of the current discussion, which focuses on student writing scores.

Analyses

In all, 128 compositions were collected: 32 each for the Written Pretest, Task One, Task Two, and Written Posttest. Given the large amount of writing to score, it was not feasible to have all compositions scored by all raters. Each composition was evaluated by two raters out of a group of eight French-speaking instructors. Raters were all native English speakers with experience teaching beginning and intermediate French language courses in the United States. They used a rubric to evaluate each composition on six linguistic features mentioned in the literature as important to FL writing and that may have been affected by using OT — content (Luoma & Tarnana, 2003), comprehensibility (Leffa, 1994), syntax (Watters & Patel, 2000), vocabulary (Scott, 1996) and spelling (Iwai, 1999), as well as remaining grammar. The working definitions for these features are explored in the discussion section. For each of these six categories, raters assigned a score on a scale from 1 (lowest) to 5 (highest). The overall scores for each composition were obtained by adding the six subscores together, yielding an overall total out of 30 possible points for each composition. Those scoring the essays also indicated whether or not they believed the student who wrote the paper had access to a translator while writing; the results of this portion of the study are described by O’Neill (2013).

RESULTS

The results for the current discussion are broken down into four sections: Reading Pretest, interrater reliability, overall composition scores, and component subscores.

Reading Pretest

The number of correct answers on the reading pretest for each participant was tabulated and the mean scores for the three groups analyzed using the GLM procedure. A one-way ANOVA performed for the results of Groups A (-training, -translator), B (-training, +translator) and C (+training, +translator) showed there was no statistically significant difference among the three groups ($F(2, 29) = 0.06, p = 0.9381$), indicating that groups were similar in reading level in French at the start of the study. Groups B and C had identical means ($\bar{x} = 15.182$ out of 35 possible points) while Group A’s mean was slightly but not significantly higher ($\bar{x} = 15.800$). The results support the assumption that the groups began the study at a similar enough level to allow for further comparison.

Interrater reliability

The remaining items (Written Pretest, Tasks One and Two, Written Posttest) were compositions. Each was scored on the six different features mentioned above and assigned an overall total by adding the subscores. The 128 compositions were scored independently by two raters each, resulting in 256 ratings. Kendall's tau-b and Spearman's rho tests were run to assess the correlation between the pairs of ratings for all tasks. A significant correlation was found, $r_s = .700$, p (two-tailed) < 0.01 and $\tau = .556$, p (two-tailed) < 0.01 . These results (Table 1) show a high correlation between raters' scores, indicating the results are reliable.

Table 1. Interrater reliability on composition scores

			Rater 1	Rater 2
Kendall's tau_b	Rater1	Correlation Coefficient	1.000	.556**
		Sig. (2-tailed)	.	.000
		N	128	128
	Rater2	Correlation Coefficient	.556**	1.000
		Sig. (2-tailed)	.000	.
		N	128	128
Spearman's rho	Rater1	Correlation Coefficient	1.000	.700**
		Sig. (2-tailed)	.	.000
		N	128	128
	Rater2	Correlation Coefficient	.700**	1.000
		Sig. (2-tailed)	.000	.
		N	128	128

** . Correlation is significant at the 0.01 level (2-tailed).

Overall composition scores

Written Pretest

The Written Pretest measured participants' writing level prior to Tasks One and Two. The average of the two raters' scores for each composition was calculated and a one-way ANOVA run to compare group means. The results showed no significant difference in scores among groups ($F(2, 29) = 0.41$, $p = 0.6690$). While Group A's mean ($\bar{x} = 22.250$ out of 30 possible points) was somewhat higher than those of Group B ($\bar{x} = 20.955$) and C ($\bar{x} = 20.864$), the difference was not significant at the $p < 0.05$ level. As with the Reading Pretest,

this result suggests participants' level was similar enough among groups prior to the experimental tasks for further comparison.

Task One

After completing the online translator training session (Group C) or a cultural lesson about the field of translation (Groups A and B), each participant wrote a three-paragraph composition. Participants wrote either without (Group A) or with (Groups B and C) the aid of OT. A one-way ANOVA on mean scores found no significant difference among the groups ($F(2, 29) = 2.88, p = 0.0725$). The mean score of Group A ($\bar{x} = 20.700$), the control group, was notably lower than those of Groups B ($\bar{x} = 23.909$) and C ($\bar{x} = 23.500$), but this result only approached significance ($p = 0.0725$). The lack of statistical difference, despite a raw difference of approximately three points out of 30 on this task, may be due in part to the low sample sizes. Nonetheless, no difference can be confirmed between the control group (Group A) and the two groups using an online translator (Groups B and C) for overall scores on the first task.

Task Two

As with Task One, participants in Group A (-training -translator), Group B (-training +translator) and Group C (+training +translator) wrote a composition that was scored by two raters. A one-way ANOVA found a significantly high level of difference among groups ($F(2, 29) = 7.05, p = 0.0032$). The means of the three groups varied, with Group C scoring highest ($\bar{x} = 25.364$), followed by Groups B ($\bar{x} = 23.681$) and A ($\bar{x} = 20.650$). In fact, as Figure 1 shows, the difference between the two groups was enough that Group C's lower quartile of scores was higher than the upper quartile of Group A.

Bonferroni (Dunn) t Tests for overall scores were run to confirm which specific comparisons yielded a significant difference. The pairwise comparison between Group A's (-training -translator) and Group C's (+training +translator) means showed they differed significantly at the $p < 0.05$ level (Table 2). The differences between Groups A and B and Groups B and C were also high, but neither reached this threshold. These results indicate that the overall performance of Group C, which was trained in the use of OT and used an online translator to write, was higher than that of the control group.

Figure 1. Task Two score distribution



Table 2. Bonferroni (Dunn) t Test results for Task Two scores

Group Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
Group C - Group B	0.2273	-0.5939	1.0484	
Group C - Group A	0.9045	0.0631	1.7460	*
Group B - Group C	-0.2273	-1.0484	0.5939	
Group B - Group A	0.6773	-0.1642	1.5187	
Group A - Group C	-0.9045	-1.7460	-0.0631	*
Group A - Group B	-0.6773	-1.5187	0.1642	
Group A: Control group; Group B: OT without training; Group C: OT with training				
* Difference between A and C significant at 0.05 level				

Written Posttest

Participants wrote a final three-paragraph composition to compare their level at the end of the study. As with the Written Pretest, no online translators were allowed. Group means analyzed via a one-way ANOVA were found not to be significantly different among groups ($F(2, 29) = 0.27, p = 0.7668$). The means for Groups A ($\bar{x} = 21.600$) and B ($\bar{x} = 21.682$) were nearly identical. Group C's mean, while over a point higher ($\bar{x} = 22.727$), was not significantly different from

the others. After having used an online translator for Tasks One and Two, Groups B and C scored similarly to Group A again at the end of the study when writing without the aid of OT. A summary of overall means for the four compositions can be found in Table 3, with the shaded cells indicating the significant difference found in Task Two.

Table 3. Mean overall scores on compositions

	Mean scores (\bar{x}) and Standard Deviations (SD)			
	Written Pretest	Task One	Task Two	Written Posttest
Group A: Control group (n = 10)	\bar{x} = 22.250 SD=2.552	\bar{x} = 20.700 SD=3.048	\bar{x} = 20.700* SD=3.859	\bar{x} = 21.600 SD=2.706
Group B: OT without training (n = 11)	\bar{x} = 20.955 SD=3.650	\bar{x} = 23.909 SD=3.277	\bar{x} = 23.681 SD=2.513	\bar{x} = 21.682 SD=3.649
Group C: OT with training (n = 11)	\bar{x} = 20.864 SD=4.995	\bar{x} = 23.500 SD=3.557	\bar{x} = 25.364* SD=2.169	\bar{x} = 22.727 SD=5.120

* Difference between A and C significant at 0.05 level

Component subscores

Written Pretest

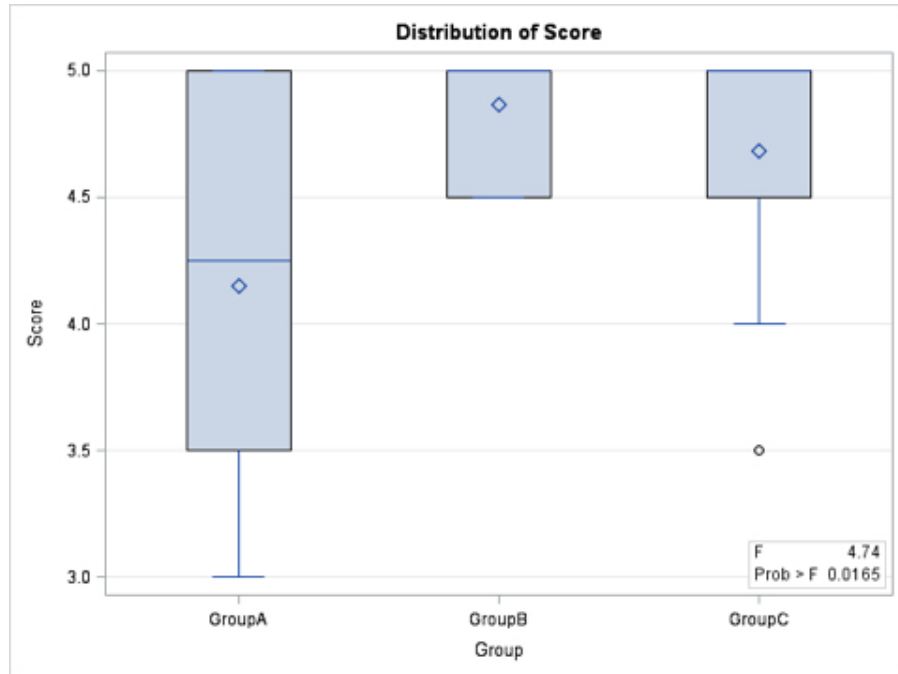
In addition to analyzing the overall scores of compositions, one-way ANOVA tests were also run on the six component subscores for each composition task to determine whether or not raters' evaluations of specific linguistic features differed between the three groups. For the Written Pretest, there were no significant differences among the groups for any of the six subcategories at the $p < 0.05$ level: Overall Comprehensibility ($F(2, 29) = 0.02$, $p = 0.9762$), Content ($F(2, 29) = 2.40$, $p = 0.1085$), Spelling and Accents ($F(2, 29) = 0.67$, $p = 0.5206$), Syntax ($F(2, 29) = 0.76$, $p = 0.4756$), Remaining Grammar ($F(2, 29) = 0.20$, $p = 0.8235$), Vocabulary ($F(2, 29) = 1.36$, $p = 0.2714$). These results suggest the groups performed similarly in each of these six features prior to the experimental tasks.

Task One

For Task One, significant differences were found among groups for Content ($F(2, 29) = 4.74$, $p = 0.0165$) and Remaining Grammar ($F(2, 29) = 3.59$, $p = 0.0403$). Further analysis was done to determine which specific pairwise comparisons yielded significant results. Bonferroni (Dunn) t Tests for Content show that Group A's (-training -translator) mean content subscore was significantly lower than that of Group B (-training +translator). Group A's mean ($\bar{x} = 4.150$ out of five points possible) differs by over a half point compared to

those of Groups B ($\bar{x} = 4.864$) and C ($\bar{x} = 4.682$); however, only the first two groups differed to the level of statistical significance. One participant in Group C (+training +translator) scored over a point below the mean: the outlier in Figure 2. This score was not excluded because of the small n size and the fact that the score falls within the midrange of Group A's scores.

Figure 2. Task One Content subscore distribution



The subscores for Remaining Grammar also varied, with the mean for Group A ($\bar{x} = 2.900$ out of 5) again over a half point lower than those of Group B ($\bar{x} = 3.591$) and C ($\bar{x} = 3.409$). As with Content, a pairwise comparison of the first two groups showed the means differed significantly, while Group C — which again had one score well below the mean — did not have a mean that differed significantly. On Task One, raters gave higher scores on Content and Remaining Grammar to Group B, which used OT to write their compositions, than the control group, which did not use a translator.

The four remaining features showed no pairwise differences among groups. While the difference in means for Spelling and Accents (Group A, $\bar{x} = 3.550$; Groups B and C, $\bar{x} = 4.227$) and Syntax (Group A, $\bar{x} = 2.900$; Group B, $\bar{x} = 3.590$; Group C, $\bar{x} = 3.409$) closely approached significance ($p = 0.0536$ and $p = 0.0638$ respectively), these subscores and those of Overall Comprehensibility ($p =$

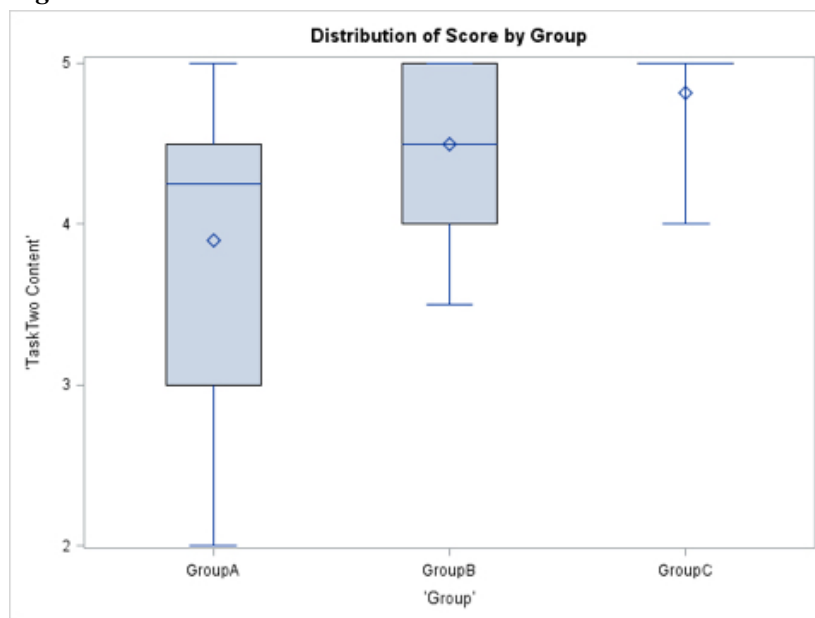
0.6937) and Vocabulary ($p = 0.8662$) were not statistically different among groups. This result indicates that there was no statistical difference among groups for those features on Task One.

Task Two

There were significant or highly significant differences discovered among groups in four of the six features for this task: as with Task One, Content ($F(2, 29) = 5.23, p = 0.0115$) and Remaining Grammar ($F(2, 29) = 6.94, p = 0.0034$) differed significantly among groups. In addition, Overall Comprehensibility ($F(2, 29) = 3.77, p = 0.0350$) and Spelling and Accents ($F(2, 29) = 7.60, p = 0.0022$) had significant differences. None was found for Syntax ($F(2, 29) = 0.31, p = 0.7375$) or Vocabulary ($F(2, 29) = 1.13, p = 0.3371$).

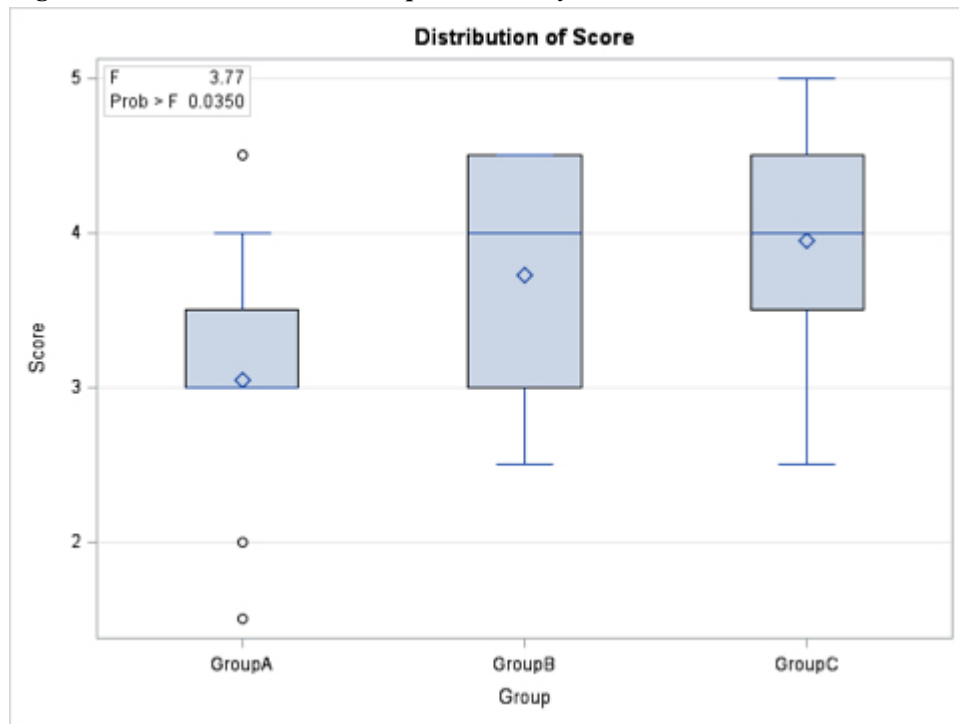
Bonferroni (Dunn) t Tests showed that the pairwise comparison between Groups A and C reached the level of statistical significance for Content. Group A's mean ($\bar{x} = 3.900$) was the lowest and significantly different from that of Group C ($\bar{x} = 4.818$), but Group B's mean ($\bar{x} = 4.500$) did not differ significantly. Figure 3 does not display an upper and lower quartile for Group C because all ratings were either a 4 or 5. Based on these results, Group C, which had been trained in OT and used a translator for this task, outperformed the control group on Content at the $p < 0.05$ level.

Figure 3. Task Two Content subscore distribution



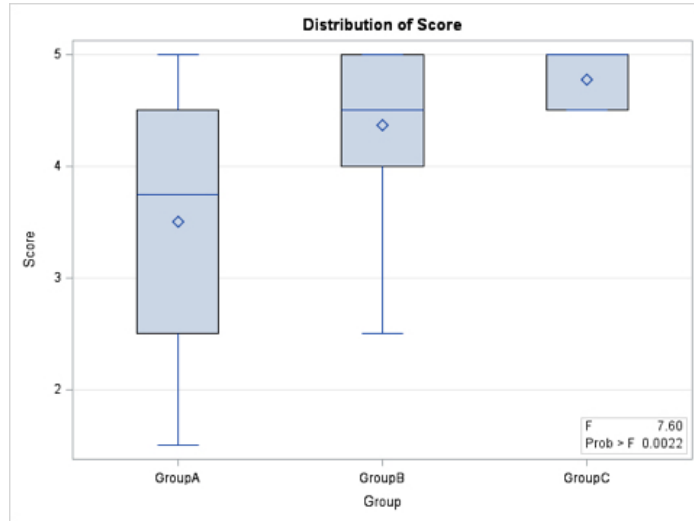
Bonferroni (Dunn) t Tests also show a difference between Groups A and C for Overall Comprehensibility at the $p < 0.05$ level. Group A's mean ($\bar{x} = 3.050$) was almost a full point below that of Group C ($\bar{x} = 3.955$). Notably, the lower quartile for Group C (3.5) is equal to the upper quartile of Group A (Figure 4). Group B's mean ($\bar{x} = 3.727$), while also greater than that of Group A, was not high enough to differ from it significantly. These data show that Group A (-training -translator) received lower scores on comprehensibility than Groups B (-training +translator) and C (+training +translator), with only the latter comparison reaching the level of statistical significance.

Figure 4. Task Two Overall Comprehensibility score distribution



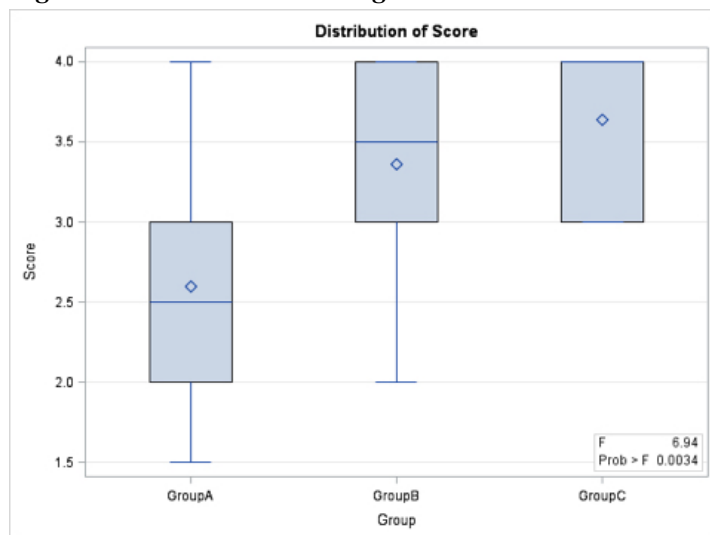
The means for Spelling and Accents also differed significantly, approaching the $p < 0.001$ level as determined by the results of Bonferroni (Dunn) t Tests. A pairwise analysis again found a difference between Groups A and C, but this time between Groups A and B as well, both at the $p < 0.05$ level. Group A's mean ($\bar{x} = 3.500$ out of 5) was over a full point lower than Group C's ($\bar{x} = 4.773$) mean and about four-fifths of a point lower than that of Group B ($\bar{x} = 4.364$). As seen on Figure 5, the upper quartile of Group A was equal to the median of Group B and the lower quartile of Group C. These results show that both groups using the online translator clearly outperformed Group A for Spelling and Accents.

Figure 5. Task Two Spelling and Accents distribution



Lastly, pairwise analyses found a significant difference ($p < 0.05$) between Groups A and B, as well as between Groups A and C, for Remaining Grammar. The means for the experimental groups ($\bar{x} = 3.364$ for Group B, $\bar{x} = 3.636$ for Group C) were greater than Group A's mean ($\bar{x} = 2.600$). As shown in Figure 6, Group A's upper quartile (3.0) was also equal to the lower quartile of Groups B and C. These results indicate that the control group did not do as well on grammar as the groups that used OT.

Figure 6. Task Two Remaining Grammar distribution



Posttest

One-way ANOVA tests were also conducted for each feature for the Written Posttest. As with the Written Pretest, no significant differences were found among the groups for any of the subscores: Overall Comprehensibility ($F(2, 29) = 0.51, p = 0.6062$), Content ($F(2, 29) = 0.08, p = 0.6062$), Spelling and Accents ($F(2, 29) = 0.03, p = 0.9746$), Syntax ($F(2, 29) = 0.66, p = 0.5232$), Remaining Grammar ($F(2, 29) = 0.46, p = 0.6345$), Vocabulary ($F(2, 29) = 0.20, p = 0.8224$). These data suggest that just as overall performance was again similar among groups once OT was no longer used, individual features also were not significantly different among groups in the Posttest.

The results for the component scores are given in Tables 4 through 7, with the shaded cells indicating cases for which a significant difference was found.

Table 4. Mean component scores for Content

	Mean scores (\bar{x}) and Standard Deviations (SD)			
	Written Pretest	Task One	Task Two	Written Posttest
Group A: Control group (n = 10)	$\bar{x} = 4.400$ SD=0.516	$\bar{x} = 4.150^*$ SD=0.783	$\bar{x} = 3.900^{**}$ SD=0.937	$\bar{x} = 4.450$ SD=0.599
Group B: OT without training (n = 11)	$\bar{x} = 4.500$ SD=0.500	$\bar{x} = 4.864^*$ SD=0.234	$\bar{x} = 4.818$ SD=0.548	$\bar{x} = 4.364$ SD=0.951
Group C: OT with training (n = 11)	$\bar{x} = 3.955$ SD=0.789	$\bar{x} = 4.682$ SD=0.513	$\bar{x} = 4.500^{**}$ SD=0.405	$\bar{x} = 4.500$ SD=0.806
* Difference between A and B significant at 0.05 level				
** Difference between A and C significant at 0.05 level				

Table 5. Mean component scores for Remaining Grammar

	Mean scores (\bar{x}) and Standard Deviations (SD)			
	Written Pretest	Task One	Task Two	Written Posttest
Group A: Control group (n = 10)	$\bar{x} = 3.000$ SD=0.667	$\bar{x} = 2.900^*$ SD=0.516	$\bar{x} = 2.600^{**}$ SD=0.775	$\bar{x} = 2.550$ SD=0.599
Group B: OT without training (n = 11)	$\bar{x} = 2.864$ SD=0.745	$\bar{x} = 3.591^*$ SD=0.539	$\bar{x} = 3.364^{**}$ SD=0.710	$\bar{x} = 2.682$ SD=0.643
Group C: OT with training (n = 11)	$\bar{x} = 2.773$ SD=1.034	$\bar{x} = 3.409$ SD=0.735	$\bar{x} = 3.636^{**}$ SD=0.452	$\bar{x} = 2.864$ SD=0.951
* Difference between A and B significant at 0.05 level				
** Differences between A and B and between A and C significant at 0.05 level				

Table 6. Mean component scores for Overall Comprehensibility

	Mean scores (\bar{x}) and Standard Deviations (SD)			
	Written Pretest	Task One	Task Two	Written Posttest
Group A: Control group (n = 10)	$\bar{x} = 3.250$ SD=0.717	$\bar{x} = 3.250$ SD=0.825	$\bar{x} = 3.050^*$ SD=0.864	$\bar{x} = 3.350$ SD=0.626
Group B: OT without training (n = 11)	$\bar{x} = 3.227$ SD=1.009	$\bar{x} = 3.591$ SD=1.068	$\bar{x} = 3.727$ SD=0.754	$\bar{x} = 3.591$ SD=0.861
Group C: OT with training (n = 11)	$\bar{x} = 3.318$ SD=1.210	$\bar{x} = 3.409$ SD=0.801	$\bar{x} = 3.955^*$ SD=0.723	$\bar{x} = 3.727$ SD=1.034
* Difference between A and C significant at 0.05 level				

Table 7. Mean component scores for Spelling and Accents

	Mean scores (\bar{x}) and Standard Deviations (SD)			
	Written Pretest	Task One	Task Two	Written Posttest
Group A: Control group (n = 10)	\bar{x} = 3.650 SD=0.747	\bar{x} = 3.550 SD=0.669	\bar{x} = 3.500* SD=1.106	\bar{x} = 3.800 SD=0.675
Group B: OT without training (n = 11)	\bar{x} = 3.227 SD=0.847	\bar{x} = 4.227 SD=0.735	\bar{x} = 4.364* SD=0.710	\bar{x} = 3.727 SD=0.817
Group C: OT with training (n = 11)	\bar{x} = 3.455 SD=0.907	\bar{x} = 4.227 SD=0.924	\bar{x} = 4.773* SD=0.261	\bar{x} = 3.818 SD=1.328

* Differences between A and B and between A and C significant at 0.05 level

DISCUSSION

Overall scores

The results of the present study address the two research questions (RQs) posed earlier. RQ 1 has two related parts (the effect of online translator use on overall scores and of training on those using OT) that will be addressed together. The data offer a mixed picture for this question. In Task One, there was no significant difference for overall composition scores among Groups A (-training, -translator), B (-training, +translator), and C (+training, +translator). Use of OT did not appear to aid or hinder overall scores for the first task. On Task Two, however, Group C's scores were significantly higher (nearly five points out of 30 possible) than those of the control group; Group B's scores were also higher than Group A's (by over three points), but not significantly so. While students having prior training in online translators scored somewhat higher, particularly on the second task, as compared to the students using OT who had not received training, the results did not reach significance. For these reasons, the two parts of RQ 1 as stated cannot be answered affirmatively.

Nonetheless, the results are still meaningful and warrant further discussion. Not only did online translator use not have a negative effect on writers' overall performance, but the group that had prior training before using the online translator received better scores on one task than the control group. One interpretation of these results would be that OT may sometimes have a positive effect on students' written production. This finding would not fully address the fact, however, that Group B did not score significantly higher than the control group on either task, and Group C only did so on the second of two tasks.

It is possible that online translator training, coupled with practice writing with the aid of the translator in Task One, allowed participants in Group C to become more skilled at using OT. This might explain why Group C scored

significantly higher than the control group on Task Two, while Group B (with no prior training) did not. If correct, this interpretation would be consistent with Somers (2003), Williams (2006), and White and Heidrich (2013), who suggest that explicit discussion with students about online translators might be useful towards helping them understand how beneficial or detrimental OT can be. Support for this interpretation might come from a study by Bishop (2001) related to training in dictionary usage, which found that students scored higher on final composition scores when trained in using a dictionary than those who were not trained in dictionary usage. Students who were trained in online translator usage in the current study were exposed to correct and incorrect translations produced by OT. Training in potential strengths and weaknesses of online translators and exposure to both successful and failed translations may have helped students make decisions to use the online translator more effectively.

The fact that students using OT did not perform worse than the control group goes against what some in the literature (e.g., Luton, 2003; Williams, 2006) have emphasized in regard to online translators. The focus on the negative aspects of OT for FL student writing, as well as discussions on how to detect and penalize online translator use — such as McCarthy's (2004) suggestion to grade such compositions on their merits as one proposed sanction — appear to assume that OT-aided texts will be recognizably worse than others. At least in terms of overall composition scores, the findings of this study do not support the assumption that students using OT will produce compositions that are poorer in quality. While the online translator groups did not fare worse than the control group, further research is needed to confirm the importance training may have played in Group C's significantly higher result on Task Two.

Component scores

The second RQ — “Does online translator usage affect perceived performance on any specific features of L2 learner writing, including comprehensibility, vocabulary, syntax, grammar, spelling, and content?” — can be answered affirmatively. While results are again mixed, participants using an online translator received higher subscores than the control group for four of the six features on one or both tasks.

The instructors rating composition tasks were asked to give a subscore on spelling and accents based on the following question: “Are the letters and diacritical marks in words written as would be expected in standard or colloquial written French?”, with the clarification that this “does not include conjugation or

other grammatical usage.” For Task One, the difference among groups very closely approached significance ($p = 0.0536$), with the mean for Group C over 0.6 points higher out of five possible points than the mean for Group A. On Task Two, both online translator groups significantly outperformed Group A on Spelling and Accents.

These results are not entirely unexpected when looking outside the context of OT. While Hurman and Tall (2002) found that participants did not use dictionaries to check for spelling, Myers (2000) found that the spelling of students who used a pocket electronic dictionary improved. Additionally, while no examination was found of orthographical or diacritical errors made by online translators, one might expect that online translators, if programmed correctly, would not make spelling or accent mistakes, unlike FL learners. The results favoring OT for performance on spelling and accents, which approached significance in Task One and reached it for Task Two, argue against a negative effect and suggest a possibly positive one for online translator usage.

Raters were also asked to judge writers’ content, described to raters as “whether or not the ideas used are sufficiently creative and appropriate to the subtopics outlined in the task.” On Task One, Group B (-training +translator) scored significantly higher on Content than Group A (-training –translator). Group C’s (+training +translator) scores were higher than those of the control group, but not at a level reaching significance. On Task Two, the situation was the opposite: Group C had significantly better Content scores than the control group, while Group B’s mean score was higher. Based in part on the writing model put forward by Luoma and Tarnanen (2003), which define the three main writing priorities as meaning, form, and expression, it was expected that using OT might improve content scores. Since the online translator could be used by participants for the form or surface features of their text, writers might have more time to work on the meaning or substance of what they were writing. Since the results are mixed on this feature, however, it is only possible to say that online translator use did not have a negative effect on Content subscores.

Group C performed significantly better on Task Two than the control group on Overall Comprehensibility, or as stated in the rubric, the ability to “be understood by a native French speaker with little or no knowledge of English.” Participants trained in online translator usage were judged to be more understandable than those writing without OT. This finding is somewhat surprising and not anticipated. Some support for machine or online translators being as understandable as human writing can be found in Leffa (1994), whose

study found high-school students understood a science text translated by a professional human translator as well as the same text translated by a machine translator, as judged by a series of comprehension questions. In addition, Ablanedo et al. (2007) found that Babel Fish translations performed on sentences from a textbook were all understandable, although less accurate, than human translations.

A possible explanation for the finding that Group C outperformed the control group on the second task is that Group A had only their own knowledge of French to make themselves understood, while Group C had access both to their own knowledge as well as the output of the online translator to form comprehensible output. Unlike Group B, which did not score significantly better than the control group, Group C had received prior training in how to use OT effectively and an explicit warning that at times “someone reading the translation wouldn’t be able to make heads or tails of what comes out” (Appendix B). Participants in this group may have been more wary about whether or not the output of the online translator could be understood. However, since there was only a significant difference on the second of the two experimental tasks, and only between Group C and the control group, it is only clear that in this study online translator usage did not have a negative effect on comprehensibility.

The last subscore that had a statistically interesting result was Remaining Grammar. Raters were told that this category included “tense/mood, subject-verb agreement, number/gender agreements, articles, negation, etc.” For Task One, Group B had a higher mean than Group A ($p < 0.01$). Group C’s mean score was also greater than that of the control group, but this difference did not reach significance. On the second experimental writing task, both translator groups scored significantly higher than Group A. Online translator use was not harmful to the scores of participants, and the results strongly suggest it was in fact helpful to them.

The literature is divided on the issue of grammar with online translators. Authors such as McCarthy (2004) and Fountain and Fountain (2009) highlight problems with grammatical accuracy, while Williams (2006) found for French that adjective placement and agreement were usually correct when using OT. Similar as to what was suggested above concerning comprehensibility, the fact that students who use OT have access to additional grammatical information besides what they have already learned might give them an advantage over students who only have access to what they have acquired and can produce on

their own. The results of the current study indicate that writers' scores were not harmed by, and most likely were aided by, having access to OT.

The only two features for which no difference was found on either experimental task were syntax and vocabulary. Syntax had been defined for the purposes of this study as “word order [...] appropriate to standard or colloquial written French.” The lack of significant difference among groups was not completely surprising. Watters and Patel (2000) found that online translators made errors in translation between German and English, with words sometimes being placed into the sentence “often without regard to syntax” (p. 156). In McCarthy's (2004) lessons on online translator usage, he pointed out examples of syntactical errors between French and English. Since students using OT did not do significantly worse than the control group, it is possible that OT sites make the same types or numbers of errors, but this supposition would merit further investigation.

Finally, scorers were asked to judge whether or not the vocabulary used in each composition was “accurate and effective in standard or colloquial written French, [...] not includ[ing] spelling, accents, or grammatical usage.” Groups B and C did not do statistically better or worse on vocabulary than the control group. This result is somewhat surprising. Yates (2006) found that 15 out of 20 translations on law texts by Babel Fish were failed, due to lexical and structural errors. While Cribb (2000) found that OT produced generally understandable output, lexical errors — in particular for denotative meanings — were noted as a weakness. One possibility is that student performance with vocabulary for OT might depend on the topic or specific items they are translating. Burton (2003) mentioned that one potential use of translators is as a dictionary for learners to check or explore vocabulary. White and Heidrich (2013) found students were able to edit output from an online translator to make fewer errors in word choice in their final composition (37 errors) than the online translator made on its own (43 errors). This finding, along with similar ones in their study investigating how students edit OT output, does not directly address how students using OT might differ from those not using it in the first place. Their research does suggest, however, that some lexical errors made by the online translator in the current study may have been compensated to some extent by students correcting them or not including them in their compositions. It is also possible that other correct words that were suggested by the translator, but which the student would not have otherwise known, acted to counterbalance any OT mistakes that participants did not catch. Additional research is needed to explore this issue further.

To return to the second research question, it can be noted that for two features, syntax and vocabulary, OT did not significantly affect students' scores. In contrast, online translator usage was found to have a positive effect at least some of the time on items such as comprehensibility, grammar, spelling, and content. Although no clear pattern emerges concerning which of the features are affected by prior training in OT, it is important to stress that all results that were significant pointed towards a positive effect for those using an online translator. The groups using an online translator did not perform worse than Group A, and in a number of cases outperformed them significantly.

LIMITATIONS

There are several limitations to consider. The sample size for each group was relatively small and not uniform (10 for the control group, 11 for each of the experimental groups). Participants in this study were in third- or fourth-semester courses and completing their final semester; it is possible that students representing different levels, continuing in French, or using OT to write in other languages might perform differently on tasks. Additionally, it was not feasible to use video, screen capture, or key logging software for this study to track how much or little participants were using the translator — some may have use OT sparingly, while others may have pasted all or most of their composition into the translator.

While students were proctored for all writing sessions and given explicit instructions, there is a small chance that OT may have been accessed by someone who was not authorized to do so. No participants reported doing so in their exit questionnaires, and the analysis of compositions found no suspicious cases of OT use in compositions where there should have been none. Lastly, a fourth condition would have been included under ideal circumstances: a group with participants who would receive training in OT but not use the translator. This condition was eliminated due to the small number of participants and the expectation that online translator training was not likely to have an effect on participants who subsequently would not use OT.

Implications

Unlike much of the previous literature on OT, which has often focused on issues such as the errors that online translators make, reasons that OT might not be advisable for FL learners, or ways of penalizing students who use online

translators for academic work, this study approached the issue from a different perspective. Since a number of students do avail themselves of online translators (White and Heidrich, 2013) and their effects cannot always be detected (O'Neill, 2013), it may be useful to focus on what the effects of OT are on student compositions as compared to writing done without a translator.

The results from this study do not support the assumption by some in the field that online translator usage necessarily leads to poorly-written compositions. In fact, in the only cases where a significant difference was found between students using an online translator and those not using one, it was always one or both of the translator groups that outperformed students not using OT.

At the same time, this preliminary finding on its own does not suggest that OT should be encouraged, allowed, or even tolerated in the FL classroom. More in-depth testing, such as delayed reading or writing posttests, is needed to gauge any lasting effects online translator usage might have on L2 acquisition as opposed to composition scores: it is possible students can achieve higher scores while using a translator, but not learn as much from the experience as they would without the help of an online translator. Using OT might be similar to copying and pasting a term paper off the Internet, with little to no learning occurring. On the other hand, writing with the aid of an online translator might be more like using an electronic calculator, once banned from some math classrooms but now embraced by many to allow students to gain knowledge and practice with more advanced material.

In light of the results of this study, research and discussion are called for on several fronts. Further investigation is needed to help determine how and how much students are already using OT in their language learning, both for classroom use and on their own. Surveys of current FL students might shed more light on student perceptions and use of OT. In addition to analyzing how overall scores and individual features are impacted by OT, using equipment or software to track students as they type can expand on White and Heidrich's (2013) research and examine on what features, at what parts of the writing process, and to what extent, students use OT when given the option to do so.

More research into whether training or instruction related to online translators would improve student performance in writing may also prove important. Since prior training about the strengths and weaknesses of OT appears to have been a factor to some extent in the current study — as suggested most

notably on the second writing task, for which only the students receiving prior training in OT significantly outscored the control group — further exploration of whether or how to train students in online translator usage could give instructors the tools they need to deal effectively with the topic of online translators.

Longitudinal research is needed to track any impacts the use of OT might have on students' L2 acquisition. The question of whether or not online translators could be used to improve learners' proficiency in the target language, in addition to helping them write in their L2, should be investigated. Another promising area of investigation is a comparison between online translators and other electronic tools, such as online dictionaries or parallel corpora search engines (as suggested by Niño, 2009), to see whether access to any electronic material improves student writing, or if certain tools are more useful and appropriate than others.

CONCLUSIONS

This study's findings suggest that in spite of the negative impression that some have concerning online translators, OT does not have a negative impact on student writing, as measured by scores earned on the compositions participants completed. While caution is called for at this stage, an analysis of the data indicates that online translator use may actually improve student performance on certain features (comprehensibility, content, grammar, spelling), or overall with prior training. In spite of what appears to be widespread use of online translators among FL learners, relatively little research has been found directly addressing the impact of OT on student writing. It is hoped this study, and further research, will open up discussion on this issue.

In the current age of widespread technology, students, instructors, and administrators must make informed decisions on what tools are useful and appropriate for academic contexts. A variety of approaches are possible as regards online translators. OT might be viewed as a technology that is unethical or unuseful for learning, presented briefly in class by instructors highlighting errors that would discourage its use, or allowed for limited use on certain types of tasks. If the results of this study are confirmed, OT might be embraced more fully as an option that, with proper guidance, can provide students an additional tool to help improve their writing and explore the target language.

ABOUT THE AUTHOR

Errol M. O'Neill is an Assistant Professor in the Department of Foreign Languages and Literatures, as well as Course Coordinator and Supervisor of Graduate Teaching Assistants in French at the University of Memphis. He received his Ph.D. in French SLATE from the University of Illinois at Urbana-Champaign. His research interests include investigating the use of online translators, dictionaries, videos, and other related technologies, as well as their impact on foreign language learning and instruction.

REFERENCES

- Ablanedo, J., Aiken, M., & Vanjani, M. (2007). Efficacy of English to Spanish automatic translation. *International Journal of Information and Operations Management Education*, 2, 194-210.
- Abraham, L.B. (2009). Web-based translation for promoting language awareness: Evidence from Spanish. In L.B. Abraham & L. Williams (Eds.). *Electronic discourse in language learning and language teaching*. (pp. 65-83). Amsterdam: John Benjamins.
- Aiken, M., & Wong, Z. (2006). Spanish-to-English translation using the Web. *Southwest Decision Sciences Institute*. Available from <http://www.swdsi.org/swdsi06/Proceedings06/Papers/IBT04.pdf>
- American Council on the Teaching of Foreign Languages. (2001). *ACTFL proficiency guidelines — Writing*. Alexandria, VA. Available from <http://www.actfl.org/files/public/ACTFLProficiencyGuidelines-Writing2001.pdf>
- Burton, C. (2003). IDEA: The online translator: implementing National Standard 4.1. *Hispania*, 86, 320- 321.
- Cribb, V. M. (2000). Machine translation: The alternative for the 21st century? *TESOL Quarterly*, 34, 560-569.
- Hurman, J. & Tall, G. (2002). Quantitative and qualitative effects of dictionary use on written examination scores. *Language Learning Journal*, 25, 21-26.
- Iwai, C. (1999). Computer-assisted writing in CALL. *IALL Journal of Language Learning Technologies*, 31, 25-33.
- Koehn, P. (2010). *Statistical Machine Translation*. New York: Cambridge University Press.
- Larson - Guenette, J. (2013). “It's just reflex now”: German language learners' use of online resources. *Die Unterrichtspraxis*, 46(1), 62-74.
- Leffa, V. (1994). Machine-translated text: Is it comprehensible to proficient readers? *System*, 22.3, 391-399.

- Luoma, S. & Tarnanen, M. (2003). Creating a self-rating instrument for second language writing: from idea to implementation. *Language Testing*, 20(4), 440-465.
- Luton, L. (2003). If the computer did my homework, how come I didn't get an "A"? *French Review*, 76, 766-770.
- McCarthy, B. (2004). Does online machine translation spell the end of take-home translation assignments? *CALL-EJ online*, 6.1. Available from <http://www.clec.ritsumei.ac.jp/english/callejonline/9-1/mccarthy.html>
- Niño, A. (2008). Evaluating the use of machine translation post-editing in the foreign language class. *Computer Assisted Language Learning*, 21(1), 29-49.
- Niño, A. (2009). Machine translation in foreign language learning: language learners' and tutors' perceptions of its advantages and disadvantages. *ReCALL*, 21(2), 241-258.
- O'Neill, Errol (2013). Online translator usage in foreign language writing. In P. Swanson and K. Hoyt (Eds.) *Dimension 2013. World Language Learning: Setting the Global Standard*, 74-88. Decatur, GA: Southern Conference on Language Teaching.
- Scott, V. (1996). *Rethinking foreign language writing*. Boston, MA: Heinle & Heinle.
- SDL. (2012). *Free Translation*. Available from <http://www.freetranslation.com>.
- Steding, S. (2009). Machine translation in the German classroom: Detection, reaction, prevention. *Die Unterrichtspraxis*, 42(2), 178-189.
- Stapleton, P. (2005). Using the Web as a Research Source: Implications for L2 Academic Writing. *The Modern Language Journal*, 89, 177-189.
- Watters, P., & Patel, M. (2000). Semantic processing performance of Internet machine translation systems. *Internet Research*, 9(2), 153-160.
- White, K. D. and Heidrich, E. (2013). Our Policies, Their Text: German Language Students' Strategies with and Beliefs about Web-Based Machine Translation. *Die Unterrichtspraxis*, 46.2, 230-250.

O'Neill

- Williams, L. (2006). Web-based machine translation: A tool for promoting electronic literacy and language awareness. *Foreign Language Annals* 39, 565-578.
- Yang, J., & Lange, E. D. (1998). SYSTRAN on Alta Vista: A user study on real-time machine translation on the Internet. In Farwell, D. et al. (eds.), *Proceedings of the Association for Machine Translation in the Americas* (pp. 275-285) Berlin: Springer Verlag.
- Yates, S. (2006). Scaling the tower of Babel Fish: An analysis of the machine translation of legal information. *Law Library Journal* 98(3), 481-500.

APPENDIX A — WRITTEN PRETEST PROMPT

Write a 3-paragraph composition on the following topic. Please follow the directions as stated below. You will have a total of **30 minutes** to complete your composition.

Topic:

You are writing a letter to Corine, a student from Paris, France who is writing an article about summer vacation at American universities. (3 paragraphs, minimum of 4-5 sentences each)

1) Greet Corine, then give 2 reasons why summer vacation can be interesting/fun for American students and 2 reasons why it can be boring/annoying.

2) Talk about some interesting activities you did this past summer, either for work or with your family and friends.

3) Convince Corine why it is important for students to have a long summer vacation or why it would be better for there to be a shorter (or no) summer vacation.

Important Note: Remember, you are **not** allowed to use any outside help to assist you in your writing, either online or from anyone else (including the proctor). You should type your response in Microsoft Word and save the file to your desktop. Please raise your hand to let the proctor know once you have finished or if you experience technical problems.

APPENDIX B — EXCERPTS FROM ONLINE TRANSLATOR TRAINING SESSION

II) The good, the bad, and the unknown [...]

B) Native-language judgments

- 1) Easier to understand and spot errors in your native language than in a foreign language.
- 2) Go to <http://freetranslation.com>
- 3) Set translator to “French to English”
- 4) Enter **Ma voisine est sympa, mais elle est trop curieuse.**
- 5) What is the result you get in English? Do you see any mistakes?

6) Sympa: very common word in French, but informal (sympa = sympathique). Apparently not in Freetranslation’s dictionary.

7) Would you be able to understand this sentence in English if you didn’t know a foreign language?

8) Go to <http://babelfish.yahoo.com/>

9) Set translator to “French to English”.

10) Enter **Ma voisine est sympa, mais elle est trop curieuse.**

11) What is the result you get in English? Do you see any mistakes?

12) Elle: can refer to people or things. Babel Fish didn’t figure out from context that “elle” refers to “voisine.”

13) Would you be able to understand this sentence in English? How is it different from Freetranslation’s? Are either of these translations satisfactory to you? Why or why not?

C) Foreign-language judgments

1) As a non-native speaker of French, you may be able to notice some mistakes or be able to understand even if there is a mistake, but harder than in your native language.

2) Go to <http://freetranslation.com>

3) Set translator to “English to French”

4) Enter **What a beautiful car!**

5) What is the result you get in French? Do you see any mistakes?

6) “What” at the beginning of a sentence often indicates a question (“What are you doing?” could use “que” / “qu’est-ce que”), the translator didn’t get that this is an exclamation (context).

7) If your teacher got this sentence, do you think he/she would understand it?

8) Go to <http://babelfish.yahoo.com/>

9) Set translator to “English to French”

10) Enter **What a beautiful car!**

11) What is the result you get in French? Do you see any mistakes?

12) Babel Fish translated the right idea in a grammatically-correct, colloquial manner.

Online translators are NOT always wrong.

III) Example translations from English to French. [...]

A) Enter the following sentences into Babel Fish (<http://babelfish.yahoo.com>), selecting “English to French”.

1) **I don’t really like your house.**

- Correct / Could be correct / Incorrect?
- Word order in English and French
- Corrected sentence/Other options?

2) **I really don’t like your house.**

- Correct / Could be correct / Incorrect?
- Word order in English and French
- Corrected sentence/Other options?

3) **She wrote it down.**

- Correct / Could be correct / Incorrect?
- Phrasal verbs in English (e.g. to wake up, to sit down, etc.) don’t exist in French
- Other options?

4) **She woke up.**

- Correct / Could be correct / Incorrect?
- Phrasal verbs in English (e.g. to wake up, to sit down, etc.) don’t exist in French
- Corrected sentence/Other options?

5) **What’s up?**

- Correct / Could be correct / Incorrect?
- Familiar/colloquial expressions in English and French
- Corrected sentence/Other options?

6) **What a drag!**

- Correct / Could be correct / Incorrect?
- Familiar/colloquial expressions in English and French
- Corrected sentence/Other options?

B) Now, let’s look at FreeTranslation. Enter the following sentences into FreeTranslation (<http://freetranslation.com>), selecting “English to French”.

1) **I gotta go.**

- Correct / Could be correct / Incorrect?
- “Incorrect” grammar in English
- Corrected sentence/Other options?

2) **She don't understand.**

- Correct / Could be correct / Incorrect?
- "Incorrect" grammar in English
- Corrected sentence/Other options?

3) **She's written a new book.**

- Correct / Could be correct / Incorrect?
- Contractions in English
- Corrected sentence/Other options?

4) **It's written in English.**

- Correct / Could be correct / Incorrect?
- Contractions in English
- Corrected sentence/Other options?

5) **I have lived here for 4 years.**

- Correct / Could be correct / Incorrect?
- Different tenses in French and English
- Corrected sentence/Other options?

6) **When I was young, I would always go there.**

- Correct / Could be correct / Incorrect?
- Different tenses in French and English
- Corrected sentence/Other options?

7) **I'm 21.**

- Correct / Could be correct / Incorrect?
- Different verbs/expressions in French and English
- Corrected sentence/Other options?

8) **I'm never wrong.**

- Correct / Could be correct / Incorrect?
- Different verbs/expressions in French and English
- Corrected sentence/Other options?

9) **a television**

- Correct / Could be correct / Incorrect?
- Isolated words in French and English
- Corrected sentence/Other options?

10) **set**

- Correct / Could be correct / Incorrect?
- Isolated words in French and English
- Corrected sentence/Other options?

IV) Online Translation

Some possible strengths and weaknesses

A. What online translators often can do.

1. Translate words that have one meaning in isolation

(for example, *the television*=*la télévision*)

2. Translate some words in context that have a small number of common, clear/distinct meanings
(for example: *I see the man*=*Je vois l'homme*)
3. Translate some common, straightforward expressions correctly or with the right general idea
(for example, *What a bummer!*=*Quelle déception!*)
4. Make basic agreement between nouns and nearby adjectives and verbs
(for example, *She went*=*elle est allée* *We're tired*=*Nous sommes fatigués*)
5. Translate tenses correctly when they're the same in French and English
(for example, *I like the Internet*=*j'aime l'Internet*)
6. Give a general idea or "gist" of the meaning of a sentence or longer passage

B. What online translators often cannot do or what they have problems doing.

1. Translate words that have a lot of meanings or uses
(for example, *set*=*ensemble* in only some cases)
2. Translate many specialized or technical words in context
(for example: *the escape key* on a keyboard is not *la clef d'évasion*)
3. Translate most expressions and phrasal verbs.
(for example: *She wrote it down.* is not *Elle l'a écrit avalent.* *What a drag!* is not *Quelle drague!*)
4. Make correct agreement between nouns/adjectives and subjects/verbs, especially if they aren't nearby (for example: *She was, as you know, very intelligent.* should have fem. adjective agreement in French)
5. Translate tenses when the tenses are different between French and English
(for example: *I've been doing...since*)
6. Recognize misspelled or misused words
(for example: *their* vs. *they're* vs. *there* vs. *thier*)
7. Deal with proper nouns (people's names, place names, etc.) effectively
(for example: *Bill's house* ≠ *la maison de la facture*)
8. Give an accurate, polished translation

C. Based on what we just saw, online translators have mixed results. Sometimes they give a completely correct translation, other times there are little mistakes here and there, and other times someone reading the translation wouldn't be able to make heads or tails of what comes out. If you know what translators generally are good at doing or what they're bad at doing, it can help you to use them more effectively in a given situation.

APPENDIX C — TASK ONE WRITING PROMPT

Write a 3-paragraph composition on the following topic. Please follow the directions as stated below. You will have a total of **30 minutes** to complete your composition.

Topic:

You are writing a letter to Marie-Claire, a student from Avignon, France who is considering studying at your university in the United States. (3 paragraphs, minimum of 4-5 sentences each)

1) Greet Marie-Claire, then give 2 reasons why your university is better/more desirable than other schools and 2 reasons why it is worse/less desirable than other schools.

2) Talk about some interesting activities you did last semester while attending your university, either at school for class or in town with friends.

3) Convince Marie-Claire either why it is important for her to come to your university or why she might consider another school.

Important Note: Remember, you are to use the translator located at <http://freetranslation.com> to help you in writing your composition. You may **not** use any other programs, websites or other help to assist you in your writing, either online or from anyone else (including the proctor). You should type your response in Microsoft Word and save the file to your desktop. Please raise your hand to let the proctor know once you have finished or if you experience technical problems.

APPENDIX D — TASK TWO WRITING PROMPT

Write a 3-paragraph composition on the following topic. Please follow the directions as stated below. You will have a total of **30 minutes** to complete your composition.

Topic:

You are writing a letter to Jean-Pierre, a student from Lyon, France who is considering applying for a job in your field in the United States. (3 paragraphs, minimum of 4-5 sentences each)

1) Greet Jean-Pierre, then describe 2 things that are potentially good about jobs in your field and 2 things that are potentially negative about jobs in your field.

2) Talk about some important things you did, either at school or elsewhere, which helped you to prepare for having a job in this field.

3) Convince Jean-Pierre either why it is important for him to apply for your field in the United States or why he might consider another field (either related or completely different) instead.

APPENDIX E — WRITTEN POSTTEST WRITING PROMPT

Write a 3-paragraph composition on the following topic. Please follow the directions as stated below. You will have a total of **30 minutes** to complete your composition.

Topic:

You are writing a letter to Bernard, a student from Cannes, France who is writing an article about winter vacation at American universities. (3 paragraphs, minimum of 4-5 sentences each)

1) Greet Bernard, then give 2 reasons why winter vacation can be interesting/fun for American students and 2 reasons why it can be boring/annoying.

2) Talk about some interesting activities you did last winter, either for work or with your family and friends.

3) Convince Bernard why it is important for students to have a long winter vacation or why it would be better for there to be a shorter (or no) winter vacation.