INTRODUCING MACHINE TRANSLATION IN TRANSLATOR TRAINING: COMPARING "INFORMATION MINING" WITH POST-EDITING

Jeff A. Killman
University of North Carolina at Charlotte (Estados Unidos)

ABSTRACT

This paper presents the results of an experiment to introduce translation students to statistical machine translation (SMT) by comparing post-editing it with using traditional documentation methods ("information mining"). At the beginning of an introductory course on translation technologies, the students translated a text in English or Spanish and annotated instances where the use of a particular electronic reference tool that they previously learned about helped them hand-pick a suitable translation solution. At the end of the course, the students revisited the same text and selected passages they had spent considerable time and effort translating, as evidenced by their annotations. They fed these passages into the open-domain SMT system Google Translate (GT) and then post-edited the output. This paper compares the students' annotated hand-picked translation solutions with the corresponding unedited GT ones (to assess differences in quality) and with the post-edited renditions of GT's solutions (to analyze whether the students' decisions to accept or reject the MT solutions resulted in quality gains or losses). According to the results, the quality of the unedited MT solutions was on average just below that of the students' own solutions, and post-editing resulted in a slight average increase in quality when compared with information mining.

KEY WORDS: machine translation, translator training, translator competence, translation technologies, documentation.

RESUMEN

En este trabajo se presentan los resultados de un experimento para introducir a alumnos de traducción la traducción automática estadística, comparando su postedición con el empleo de métodos de documentación tradicionales. Al inicio de una asignatura introductoria de las tecnologías de traducción, los estudiantes tradujeron un texto en español o inglés y anotaron los casos en los que el empleo de una herramienta de referencia electrónica, sobre la que habían aprendido anteriormente, les había ayudado a elegir manualmente una resolución de traducción adecuada. Al final de la asignatura los estudiantes volvieron a replantearse el mismo texto y eligieron partes de él, cuya traducción les había requerido una cantidad de tiempo y esfuerzo considerable, como ponen de manifiesto sus anotaciones. Las partes del texto que habían elegido las introdujeron en el gratuito sistema abierto de traducción automática estadística de Google Translate (GT) y, después, posteditaron los resultados de este.

En este trabajo se comparan las traducciones que habían elegido manualmente los estudiantes con las producidas por GT (para determinar diferencias de calidad) y con las postediciones de estas últimas (con el fin de analizar si las decisiones de los estudiantes de aceptar o rechazar las traducciones automáticas dieron como resultado ganancias o pérdidas cualitativas). Según los resultados, la calidad de las traducciones automáticas no editadas era de media muy por debajo de la de las producidas por los estudiantes, y la postedición resultó en un ligero aumento medio de calidad al compararse con los métodos de documentación tradicionales.

PALABRAS CLAVE: traducción automática, formación de traductores, competencia traductora, tecnologías de traducción, documentación.

FECHA DE RECEPCIÓN: 24/03/2015
FECHA DE ACEPTACIÓN: 25/06/2015
PÁGINAS: 179-193

1. INTRODUCTION

Free online statistical machine translation (SMT) is becoming commonplace. For example, Google Translate (GT) is often used to get the gist of a website or document in a foreign language. Moreover, it is now being used in some of the leading-edge technologies of translation memories (such as SDL Trados Studio and Déjà Vu X) to offer translators the option to post-edit machine generated translations when it is not possible to draw on translations from the translation memory databases.

This paper asks whether using an SMT system like GT might, in theory, reduce the amount of lookups one might have to do in order to come up with translation resolutions. One can end up spending a good deal of time manually looking items up in reference resources, such as dictionaries, termbases, parallel or related documents, or corpora, and still make mistakes. Little time, however, is invested in using MT; it produces word and phrase translations instantly, which if correct can favor quality while potentially saving time. The paper looks at the translation solutions that master's students manually looked up in electronic reference resourceswhen working on a translation they prepared in English or Spanish in a course on translation technologies. It compares the quality of these "handpicked" solutions with that of the translations suggested by GT for the same items, to test whether this SMT system has the potential to suggest good translation solutions students might otherwise have to look up themselves. The paper then analyzes whether the students decided to use the acceptable and unacceptable GT suggestions when they were later tasked with post-editing them. It also analyzes how their decisions to accept or not accept the suggestions influenced the quality of their post-edited versions when compared to that of their human versions.

Recent studies (Pym 2009, Garcia 2010 and 2011, Lee and Liao 2011, Şahin and Dungan 2014) indicate that GT now can be used by university students to attain a level of productivity and quality comparable to that attained by human translation. Pym (2009) looks at GT machine translations produced between English and Chinese, French, and Korean; Garcia (2010 and 2011) and Lee and Liao (2011) look at GT machine translations produced between English and Chinese; and Şahin and Dungan (2014) look at GT machine translations produced from English into Turkish. Though none of these studies tested the quality of SMT between English and Spanish, it is reasonable to expect that similar results might emerge as Spanish is part of GT's first stage of supported languages that could be translated into and from English.

Three of these studies (Pym 2009, Lee and Liao 2011, and Şahin and Dungan 2014) report on GT's accuracy. One of the groups of participants in Pym's study reported appreciating some of GT's terminology suggestions

(2009: 141), and the participants in Şahin and Dungan's study reported that some of the advantages of using GT included help with vocabulary and word choice (2014: 79). Lee and Liao (2011) include considerably more information about vocabulary accuracy in their study. In the case of the linguistically weaker of the two student participant groups that participated in their study, they found that "the more words from the MT text a student uses, using [the] sentence as a unit, the less likely a student would make a mistake in translating that particular sentence [positing] that the students recognize they can use the MT directly if the meaning is intact, and they would only have to do a little tweaking" (Lee and Liao 2011: 128). They also note several instances where GT was more accurate than students when it came to the contextually appropriate meaning of an ambiguous piece of language (Lee and Liao 2011: 136-137) and the contextual meanings of adjectives and nouns depending on the collocations they were used in (Lee and Liao 2011: 137-138). As regards the actual wording of items in the translation, they also found there were register advantages to using MT (Lee and Liao 2011: 133).

These studies provide some evidence that GT might be a viable source of vocabulary translation support in the case of students. This paper seeks to further assess if this is indeed the case (in the English-Spanish language pair).

2. MACHINE TRANSLATION

MT as a technology is not new, with the first proposals for actual machine translation coming about in the early 20th century. In fact, "The first public demonstration of a machine translation system was the Russian-English Georgetown University System, a collaborative effort between IBM and Georgetown University, carried out in 1954" (Quah 2006: 60). There was strong optimism that high-quality translation could be produced by machines, and large-scale funding was also implemented in other countries such as the USSR.However, the late 1950s yielded to pessimism as problems with early research were being identified. Language was considered too complex and the task of translation, very human and not easily simulated by a computer. The expectations for MT were overoptimistic and naïve. Because the machines were equipped only with dictionaries and, in some cases, grammars, they were often unable to translate words according to the context they occurred in. As a result, the first half of the 1970s was a quiet period for MT, especially in the US. Interest was renewed during the 1980s as hardware and software improved and computational linguistics developed. This period marks the shift from research-driven to needs-driven MT development.

The currently dominant paradigm in MT research is SMT, with "a growing share of the MT market, though prior to the 1990s attempts to use statistical methods were unsuccessful. In SMT, one says that a SL [source language] and a TL [target language] sentence are a translation of each other *with a certain probability*" (Forcada 2010: 221). SMT "learns to translate" from an enormous database of previously translated texts where millions of sentences in one language have been aligned with their human translations in the other language. That database "has been previously processed to guess at how humans have typically translated various words and how that translation depends on the immediate co-text, that is, a word or two on either side of the word in question" (Bendana and Melby 2012: 45). Though SMT may "learn longer and longer useful phrases, sometimes even memorize the translation of an entire sentence" (Koehn 2010: 128), "longer phrases are less frequent and less statistically reliable" (Koehn 2010: 141) and hence might not be chosen. Statistics for short phrases or words will be supported more often, even though short phrases and words have a higher probability of being translated in more than one way, depending on the context. For example, *acción* in Spanish might be translated as "action" in an everyday context or "share" in a business one. Thus ambiguity still remains a problem (Forcada 2010: 216), but certainly not to the same extent as before the phrase-based statistical approach.

Another problem of particular concern is whether an SMT system containing in its database different translation solutions for the same item is able to select the most suitable one. Specific translation solutions might compete with other possible ones that while semantically correct, are not lexically desirable for a number of reasons including inconsistency with a client's unique preferences or the geographical or social variety of the language involved. For example, "legal person" is a *persona jurídica* in Spain and a *persona moral* in Mexico. Whatever the case may be, whether or not a particular SMT system selects a particular word or phrase depends on if it can be searched in the database or if the statistics for it are reliable enough so that it may be chosen.

3. TEST DESIGN

This study compares the quality of translation solutions that were handpicked from electronic reference resources with that ofthe translation solutions that were suggested by GT for the same source items, to test whether this free online SMT system might help lighten the load of documentation work one might need to do.It also analyzes whether the GT suggestions were

accepted in a post-editing exercise and how whether selecting them resulted in quality gains or losses.

The participants in the study were enrolled in a translation technologies course that is offered in a master's degree program in Spanish, with a specialization track in translation. The course is a conceptual and selective hands-on introduction to MT, translation memory tools, terminology management, term banks, (bilingual) corpora, and corpus-analysis tools. The class comprised 11 students for whom it was the first time having taken a course of this sort. Some had previously taken translation practice courses in the master's program (e.g. legal translation, literary translation, technical translation, and so on), while for some it was indeed their first time ever having taken any type of translation course.

At the beginning of the course the students learned about tools such as corpora (web-crawled, online, or ad hoc), termbases, online document repositories, etc. Training activities were assimilated so the students could learn how to use these tools in order to gain productivity. For instance, in some translation assignments they were instructed to use a particular termbase, glossary, or corpus for specific purposes. They also learned how to create their own disposable corpora with BootCat, "a tool which, based on user input, automatically selects and downloads potentially relevant webpages, thus making it possible to gather ad hoc corpora literally within minutes" (Ferraresi 2009: 5). In turn, the students learned how to use a concordancer to search the corpus and retrieve examples of phrases or terms they consulted. They were also introduced to web-crawled multilingual corpora such as WeBiText (Désilets et al. 2008) or Linguee (both of which can be thought of as a sort of 'Google of translated text').

After such training, the students were asked to submit a proposal for a non-literary text they wished to translate, either in English or Spanish (according to abilities). Once their proposal was approved, the students were then asked to translate the text on their own time but with the assistance of various reference tools. The idea at this point was that they would be able to practice using various reference resources to translate a non-literary text. In particular, they were asked to do the following:

1. Compile two BootCat corpora specifically for their project (one in source language and one in the target language)
2. Consult web-crawled bilingual corpora such as WeBiText or Linguee
3. Consult online termbases such as IATE (inter-active terminology for Europe) and TERMIUM Plus (Translation Bureau of Canada)
4. Consult (ready-made) corpora already available online, such as the Corpus del español (Mark Davies), the Corpus diacrónico del español (CORDE), Corpus de referencia del español actual (CREA), the

American National Corpus, the Leeds corpus, the Collins Cobuild corpus, and so on.

The students were then asked to annotate their translation in instances when they felt the need to translate with the aid of any of these resources. Overall, the goal was to develop the students' documentation skills or ability to "mine" information with the assistance of technology, while having them evaluate and reflect on the tools critically and methodically so that they could be efficiently served by them and not simply using them for technology's sake.

Towards the end of the course (about a month and a half later and after they had completed the course unit on translation memory tools), the students were asked to revisit their annotated translation on their own time and choose a few paragraphs (some 500 words) that they particularly needed help with and where they made a good number of annotations. This time they were instructed to machine translate these paragraphs with GT and then post-edit the output in a Microsoft Word file. Before proceeding to post-edit, they were instructed not to look at the translations they carried out previously with annotations and were asked to activate 'Track Changes' and scroll to 'Final' so that the red changes would not distract them. Though there was no way of preventing them from referring back to their previous work, the idea was simply to encourage them to approach the machine translations as objectively as possible (to be open to the possibility of different translations, either equal to or better than their own), to favor quality while potentially saving time.

While post-editing, the students had the advantage of previous knowledge about the source text and how an acceptable translation of it might take form, as they had already translated the same text on a previous occasion but in a different way (i.e. they should have a sort of déjà vu feeling in a good number of instances). Knowing how a translation might acceptably take form is necessary in order to determine if MT output is worthwhile or not; had the students translated the text the first time, it might have been especially difficult for them to know what to trust or mistrust, without doing extensive research in some cases. Moreover, the students would now be better able to see if what they spent considerable time researching could be resolved instantaneously by GT and learn about how free online SMT might or might not help them, depending on their needs.

All but three of the texts selected by the 11 students were in Spanish. Two students' texts (both in Spanish) have been excluded from the study because it turns out they used MT during the annotation assignment at the beginning of the semester despite the instructions they were given not to. Overall, the nine students' texts that are under study varied in terms of language, subject matter, and genre, as can be seen in table 1.

Spanish	English
1. Business journal article	1. Business journal article
2. Geology journal article	2. Funeral industry website
3. Linguistics journal article	3. University financial aid website
4. Linguistics journal article	
5. Mexican immigration form	
6. Public health journal article	

Table 1. Text types and languages

Like Lee and Liao (2011) this study uses Pym's (1992) classification based on binary and non-binary errors, to assess the quality of the handpicked translation solutions and the translation solutions suggested by GT for the same source items. As defined by Pym (1992: 282), a binary error means "It's wrong!", while a non-binary error means "It's correct, but…". Table 2 summarizes the specific criteria we used to apply the two types of errors.

Type		**Definition**
Binary error	Misunderstanding of ST	Misinterpretation of words, meanings
	Faulty rendition of TT	Syntactic errors, omissions
Non-binary error	Improper word use	Improper collocation, register mismatch, vague expression, inability to convey the message
	Insufficient transfer competence	Over-long modifier, logic inconsistency, over-use of pronouns, superfluous words

Table 2. Error Classification taken from Lee and Liao (2011: 113)

Basically, a binary error is attributed to a lack of language proficiency, whereas a non-binary error is the result of poor translation ability. A binary error deviates from the true meaning, whereas a non-binary error deviates from readability (i.e. the target text is acceptable in terms of meaning but could indeed be worded better). With only two categories, it is easier to classify errors. Of course, as Pym points out:

> it is relatively easy to produce a terminological system of three or seven or perhaps twenty odd types of translation error and then find examples to illustrate the phenomenal level and presumed causality of each [but] it is quite a different matter to classify errors as they actually appear in translated texts, where elements

> of different types are perpetually mixed and numerous cases straddle the presupposed distinctions. Such classifications will always have either too few or too many terms, at least for as long as there is no clear awareness of why translation errors should be classified in the first place. (1992: 282)

Pym (1992: 287) also mentions that as translation students advance they should be making less binary errors, with a greater proportion of non-binary errors, constituting a yardstick against which gains in translation abilities can be measured.

A total of two points could be given for every term or phrase translation that the students documented. If the translated term or phrase was correct, then two points were given. While one point was given in the case of non-binary errors, none were given in that of binary errors. The same metrics were applied to GT's translations of the same source items. Table 3 shows an example of how points were calculated in the case of one student, who researched and annotated a total of ten items. It also displays whether or not the student accepted GT's suggestions when post-editing. Taking stock of each student's decisions to accept or reject correct or incorrect MT suggestions is a way to test their abilities to recognize good output or bad output. The information obtained can also answer whether MT helped them improve their translations, made no difference, or was disadvantageous. It might also suggest who might have the most to gain from post-editing. In table 3 we can see that in every instance GT's suggestions were accepted, even where in instance 7, GT's solution was incorrect and the student's was not. In any event, the student clearly had more to gain than to lose by simply accepting all of GT's suggestions.

Researched source items	**Student solution 13/20 points**	**GT suggestion (same or different) 15/20 points**	**If different (worse, equal, or better)**	**Did student use GT suggestion? 15/20**
1. *principal vector*	"main carrier" (2)	"main vector" (2)	equal	yes (2)
2. *incidencia*	"incidence" (2)	same (2)		yes (2)
3. *hasta alcanzar la cifra máxima de*	"and reached the maximum figure of" (1)	"to reach a peak of" (2)	better	yes (2)
4. *la carga real de la enfermedad*	"the real burden of the disease" (2)	same (2)		yes (2)
5. *mantiene una tendencia ascendente*	"maintained its ascending tendency" (1)	"maintains an upward trend" (2)	better	yes (2)

6. *determinantes antigénicos o tropismo alterados*	"antigenic determinants or altered tropism" (0) [needs to be "altered antigenic determinants or tropism"]	same (0)		yes (0)
7. *los signos de alarma clínicos*	"clinical warning signs" (2)	"warning signs" (0)	worse	yes (0)
8. *los estudios ultrasonográficos*	"ultrasonographic studies" (2)	same (2)		yes (2)
9. *sistema fagocítico mononuclear*	"phagocyte mononuclear system" (0)	"mononuclear phagocyte system" (2)	better	yes (2)
10. *amplificación dependiente de anticuerpos*	"antibody-dependent amplification" (1) [should be "antibody-dependent enhancement"]	same (1)		yes (1)

Table 3. Example of how points were calculated

4. RESULTS

Table 4 compares scores of the handpicked suggestions ("human translations"), GT's suggestions ("machine translations"), and the post-edited solutions. The cases under study are numbered from lowest to highest in terms of human translation scores and will be referred to accordingly. The machine translations (MTs) were on average 91% as good as the human translations (HTs). It can thus be said that free online SMT indeed has the potential to lighten the load of documentation work one might need to do, especially in the case of translators in training. In cases 2 and 6 GT yielded a higher score than the student, in case 4 there was a tie, and in the remaining six cases (1, 3, 5, 7, 8, and 9) the student's researched solutions earned more points. The scores obtained by post-editing were on average 7% better than the HT scores. In cases 2, 4, 5, and 6 quality improved, while in cases 3, 8, and 9 it did not and in cases 1 and 7, it decreased.

Case	**Human translations 101/130 (78%)**	**Machine translations 92/130 (71%)**	**Post-editing 109/130 (84%)**
1	5/10 (50%)	4/10 (40%)	4/10 (40%)
2	13/20 (65%)	15/20 (75%)	15/20 (75%)
3	4/6 (67%)	2/6 (33%)	4/6 (67%)
4	11/16 (69%)	11/16 (69%)	13/16 (81%)
5	16/22 (73%)	14/22 (64%)	20/22 (91%)
6	9/12 (75%)	10/12 (83%)	11/12 (92%)
7	17/18 (94%)	16/18 (89%)	16/18 (89%)

8	10/10 (100%)	7/10 (70%)	10/10 (100%)
9	16/16 (100%)	13/16 (81%)	16/16 (100%)

Table 4. General results. Figures with highest scores are highlighted

Of the MTs produced by GT for the 65 total items under study, 50 were accepted and 15, rejected. Table 5 categorizes the 50 MTs that were accepted in each case, detailing whether they were correct (2 pts.), partially correct (1 pt.), or incorrect (0 pts.), as well as how they were the same as or different from the HTs in each case. 80% (40/50) of the accepted MTs were correct, 10% (5/50) were partially correct, and 10% (5/50) were incorrect. 77.5% (31/40) of the correct MTs had correct corresponding HTs, although more than half (17/31) were different. 22.5% (9/40) of the correct MTs constituted an improvement; the corresponding HTs were in five instances partially correct (two items in case 2 and one item in cases 4, 5, and 7) and in four, incorrect (one item in case 2 and case 5 and two items in case 4). Only correct MTs were accepted in cases 3 (one instance), 6 (three instances), 7 (eight instances), 8 (three instances), and 9 (6 instances).

Of the five partially correct MTs, one was the same as the HT (case 2), two were worse (in cases 4 and 5 the corresponding HTs were correct), one was different but equal in quality (in case 1 the corresponding HT was different but also partially correct), and one was better (in case 1 the corresponding HT was incorrect). Counting this partially correct MT that was beneficial and the nine correct MTs that were also beneficial, it can be said that accepting MTs improved quality in 20% (10/50) of the instances, as seen in cases: 1 (in one instance where a partially correct MT was better than the incorrect HT), 2 (where two correct MTs were better than two partially correct HTs, and one correct MT was better than an incorrect HT), 4 (where two correct MTs were better than two incorrect HTs, and one correct MT was better than a partially correct HT), 5 (where one correct MT was better than an incorrect HT, and one correct MT was better than a partially correct HT), and 7 (where one correct MT was better than a partially correct HT). Of course, there was no room for improvement in cases 8 and 9, whose human translations were error free. In case 2, because the MTs yielded two more points than the corresponding HTs, simply accepting all of them resulted in a benefit of two points. In case 1, however, simply accepting all the MTs resulted in a loss of a point. Had the students in cases 1 and 2 avoided accepting an incorrect MT, whose corresponding HT was indeed correct, post-editing would have received the highest score in each case. Had the student in case 4 not accepted a partially correct MT and an incorrect MT, whose corresponding HTs in each instance were also correct,case 4's post-editing score would have been flawless. And had the student in case 5 not accepted a partially correct MT, whose corresponding

HT was indeed correct, post-editing in this case would have been just one point shy of flawless.

Case	1	2	3	4	5	6	7	8	9	Total
Correct MTs	1	7	1	5	6	3	8	3	6	40
same: ht	0	3	0	1	2	2	1	1	4	14
diff: cht	1	1	1	1	2	1	6	2	2	17
diff: pcht	0	2	0	1	1	0	1	0	0	5
diff: iht	0	1	0	2	1	0	0	0	0	4
Partially correct MTs	2	1	0	1	1	0	0	0	0	5
same: ht	0	1	0	0	0	0	0	0	0	1
diff: cht	0	0	0	1	1	0	0	0	0	2
diff: pcht	1	0	0	0	0	0	0	0	0	1
diff: iht	1	0	0	0	0	0	0	0	0	1
Incorrect MTs	2	2	0	1	0	0	0	0	0	5
same: ht	0	1	0	0	0	0	0	0	0	1
diff: cht	1	1	0	1	0	0	0	0	0	3
diff: pcht	0	0	0	0	0	0	0	0	0	0
diff: iht	1	0	0	0	0	0	0	0	0	1
Total items accepted	**5/5**	**10/10**	**1/3**	**7/8**	**7/11**	**3/6**	**8/9**	**3/5**	**6/8**	**50/65**

Table 5. Correct, partially correct, and incorrect MTs that were accepted, which were either the same as ("same:") the corresponding human translation ("ht") or different from ("diff:") the corresponding: correct human translation ("cht"), partially correct human translation ("pcht"), or incorrect human translation ("iht").

A total of 15 MTs were post-edited (see table 6). Ten of these MTs were incorrect, while just three were partially correct and two, correct. On this basis 2 out of the 15 (13%) did not need to be improved upon. Eight out of the ten incorrect MTs were post-edited correctly; 6 of the post-edits were the same as the corresponding HT (one instance in case 3, one instance in case 4, two instances in case 5, one instance in case 8, and one instance in case 9), while 2 were different from the corresponding HT that was incorrect (one instance in case 5 and another in case 6). The remaining two incorrect MTs were post-edited incorrectly; the post-edit in case 3 was the same as the corresponding HT, while that in case 7 was different from the corresponding HT that was correct. Had this correct HT been resorted to, the post-editing score (in case 7) would have been flawless. Two out of the three partially correct MTs were post-edited correctly with the corresponding HT in cases 8 and 9, while in case 5 the remaining partially correct MT was post-edited with a partially correct solution other than the partially correct corresponding HT. The two correct MTs were post-edited in case 6. In one instance, the change made was the same as the corresponding HT that was partially correct, while in the other, the

change was the same as the corresponding HT that was completely correct. Had this partially correct change not been introduced, the post-editing score in case 6 would have been flawless.

Case	3	4	5	6	7	8	9	Total
Incorrect MTs	**2**	**1**	**3**	**1**	**1**	**1**	**1**	**10**
Correct changes	1	1	3	1	0	1	1	8
same: ht	1	1	2	0	0	1	1	6
diff: iht	0	0	1	1	0	0	0	2
Incorrect changes	1	0	0	0	1	0	0	2
same: ht	1	0	0	0	0	0	0	1
diff: cht	0	0	0	0	1	0	0	1
Partially correct MTs	**0**	**0**	**1**	**0**	**0**	**1**	**1**	**3**
Correct changes	0	0	0	0	0	1	1	2
same: ht	0	0	0	0	0	1	1	2
Partially correct changes	0	0	1	0	0	0	0	1
diff: pcht	0	0	1	0	0	0	0	0
Correct MTs	**0**	**0**	**0**	**2**	**0**	**0**	**0**	**2**
Correct changes	0	0	0	1	0	0	0	1
same: ht	0	0	0	1	0	0	0	1
Partially correct changes	0	0	0	1	0	0	0	1
same: ht	0	0	0	1	0	0	0	1
Total changes made	**2/3**	**1/8**	**4/11**	**3/6**	**1/9**	**2/5**	**2/8**	**15/65**

Table 6. Incorrect, partially correct, and correct MTs that were changed. Changes made were correct, partially correct, or incorrect and were either the same as ("same:") the corresponding human translation ("ht") in each case or different from ("diff:") the corresponding: incorrect human translation ("iht"), correct human translation ("cht") or partially correct human translation ("pcht").

5. CONCLUSION

The results of this study show that GT could instantly provide translation solutions that were 91% as good as the human translation solutions that had to be researched in each case. This is a free online SMT tool which may very well favor quality while potentially saving time, especially if it can consistently save translators from having to do research time and time again. As mentioned above, it is included in some of today's leading-edge translation memory tools so translators might not have to look elsewhere for solutions.

The post-editing scores were on average 7% better than the HT scores. Especially in the cases in which there were more human errors, it seems

translators might have much to gain in terms of quality from MT output. However, further investigation is needed to corroborate these findings. The number of items that were annotated as having been researched in each case were somewhat limited and ranged considerably from case to case. Moreover, anecdotal evidence suggests that there were a good number of other items, in some cases more difficult than the ones annotated, that GT was able to help the students translate more successfully than they could without the tool. In any event, there may have been instances where the students should have taken it upon themselves to do research or did not take care to annotate what they did indeed do research on, for reasons of time. I propose investigating the remaining non-annotated parts of the translated texts, which will comprise a sample of around 500 words in each student's case. The non-annotated parts will be compared with the corresponding MTs (to further assess differences in quality) and with the post-edited renditions of these (to analyze further whether the students'decisions to accept or reject the MT solutions resulted in quality gains or losses). This research may also give a clearer picture of the profiles of users who might gain the most and who might gain the least from post-editing MT output, as well as yield findings as to how they might or might not in each case.

6. BIBLIOGRAPHY

Bendana, L. y A. Melby, (2012). *Almost Everything You Ever Wanted to Know about Translation*. Toronto (Canadá), Multi-Languages Corporation.

Désilets, A. et al. (2008). "WeBiText: Building Large Heterogeneous Translation Memories from Parallel Web Content" en las Proceedings of Translating and the Computer (30). Londres, 23-28 de noviembre de 2008.

Ferraresi, A. (2009). "Google and Beyond: Web-As-Corpus Methodologies for Translators", *Revista Tradumática*, 7.

Forcada, M. (2010). "Machine Translation Today" en Gambier, Y. y L. V. Doorslaer (eds.): *Handbook of Translation Studies Volume 1*, 215-223. Ámsterdam/Filadelfia, John Benjamins.

Garcia, I. (2010). "Is Machine Translation Ready Yet?", *Target*, 22(1), 7-21.

Garcia, I. (2011). "Translating by Post-Editing: Is it the Way Forward?", *Machine Translation*, 25, 217-237.

Koehn, P. (2010). *Statistical Machine Translation*. Nueva York, Cambridge University Press.

Lee, J. y P. Liao (2011). "A Comparative Study of Human Translation and Machine Translation with Post-Editing", *Compilation and Translation Review*, 4(2), 105-149.

Pym, A. (1992). "Translation Error Analysis and the Interface with Language Teaching" en Dollerup, C. y A. Loddegaard (eds.): *Teaching Translation and Interpreting: Training, Talent and Experience*, 279-288. Ámsterdam/Filadelfia, John Benjamins.

Pym, A. (2009). "Using Process Studies in Translator Training: Self-Discovery through Lousy Experiments" en Göpferich, S, Alves, F. y I. Mees (eds.): *Methodology, Technology and Innovation in Translation Process Research*, 135-156. Copenhague, Samfundslitteratur.

Quah, C. (2006). *Translation and Technology*. Nueva York, Palgrave Macmillan.

Şahin, M. y N. Dungan (2014). "Translation Testing and Evaluation: a Study on Methods and Needs", *Translation & Interpreting*, 6(2), 67-90.