

Translation technology explored: Has a three-year maturation period done Google Translate any good?

Susan Lotz and Alta van Rensburg

Language Centre, Stellenbosch University, South Africa
Email: slotz@sun.ac.za; avrens@sun.ac.za

Abstract

Language users in multilingual environments who are trying to make sense of the linguistic challenges they face may well regard the advent of online machine translation (MT) applications as a welcome intervention. Such applications have made it possible for virtually anyone to try their hand at translation – with minimum effort, at that. However, the usefulness of the output of these translation applications varies. The empirical research described in this article is a continuation of an investigation into the usefulness of MT in a higher education context. In 2010, Afrikaans and English translations generated by Google Translate and two human translators, based on the same set of source texts, were evaluated by a panel of raters by means of a holistic assessment tool. In 2011 and 2012, the same set of source texts was translated again with Google Translate, and those translations have since been evaluated in exactly the same manner. The results show that the quality of Google Translate's output has improved over the three years. Subsequently, an error analysis was performed on the translation set of one text type by means of a second assessment tool. Despite an overall improvement in quality, we found that the 2012 translation contained unexpected new errors. In addition, the error analysis showed that mistranslation posed the largest risk when using this MT application. Users of MT should, therefore, understand the risks of their choice and that some text types and contexts are better suited to MT than others. Armed with this knowledge, translators and multilingual communities can make informed decisions regarding MT and translation technology in general.

Keywords: error analysis, Google Translate, higher education, machine translation, multilingualism, translation quality

1. Introduction

In multilingual environments, language users need all the tools they can possibly use to keep up linguistically. We have to function in a world in which the limits have been moved much further back than those initially imposed on us by our first language. There is no longer a

correlation between “[t]he limits of my language” and “the limits of my world”¹, as Wittgenstein (1922) would have it, and we have to keep up with these shifting boundaries.

Language users in multilingual environments who are trying to make sense of linguistic challenges may very well regard the advent of online machine translation (MT) applications as a welcome technological intervention. Such applications have made it possible for virtually anyone to try their hand at translation – with minimum effort, at that. The output of these translation applications varies in usefulness, however.

Sager (1994:261) describes translation as “a mediating activity”, the particular form of which is determined by the text as well as the specific communicative circumstances, for example, the purpose of the communication. An MT system should not be expected to render a translation similar in quality to that which a professional translator can achieve. Sager (1994:262) asserts that MT “has a proper place beside human translation as an alternative technique for achieving different communicative objectives”. The variety of communicative objectives of translation means that the demand for translation also varies, and, along with it, expectations regarding quality. The communicative objectives can be divided into three main groups, namely (i) dissemination, where quality is most important (for example, the translation of a manuscript for publication), (ii) assimilation, where speed is more important than quality (for example, the online translation of a foreign-language newspaper article to get the gist of it), and (iii) interpersonal communication, where real-time communication, such as social-network messages or blogs, is translated (Hutchins 2001, Bennett and Gerber 2003:180-181, Quah 2006:89-90).

In our experience at the Stellenbosch University Language Centre, clients do not necessarily make the above distinctions with regard to MT. In the next section, the context of the study reported on in this article will be further explained. Section 3 will shed more light on Google Translate, after which the empirical study will be described in section 4. Section 5 concludes with a summary and suggestions for further research.

2. The context of this study

We find ourselves in a multilingual environment at Stellenbosch University, since its language plan requires that study material, such as examination papers and class notes, be available in Afrikaans as well as English, whenever possible. The language plan also specifies that documents relating to service conditions should be made available in Afrikaans, English and isiXhosa, depending on the requirements of staff (Stellenbosch University 2010). Over the past few years we have had numerous enquiries about the use of MT by University clients, often particularly with regard to using the online MT application Google Translate to save money and time in the translation process.

Subsequently, we decided to explore what Google Translate could offer our clients. This resulted in a study of which the first phase will be mentioned (see Van Rensburg, Snyman and Lotz (2012) for a detailed discussion hereof), and the second and third phases will be described in this article. In addition to our own research objective of investigating the usefulness of MT

¹ Proposition 5.6 from Wittgenstein’s 1922 publication *Tractatus Logico-Philosophicus*: “The limits of my language are the limits of my world”. Original German: “Die Grenzen meiner Sprache bedeuten die Grenzen meiner Welt”.

in a higher education context, we chose to incorporate a client perspective in our research, where possible, and compare and consider factors that we observed were important to clients.

3. What is Google Translate and how does it do what it does?

Google Translate is a free online application, offered by Google Inc., that allows users to have words, sentences, documents and even websites translated in an instant.

The translations are generated by Google Translate computer systems and are based on patterns found in large amounts of text, rather than sets of rules for a particular language. Although human language users cope with rules and their exceptions, those exceptions have proved to be problematic for rule-based MT. One way around the problems that exceptions to language rules pose for systems aiming to translate is to let the systems discover the rules for themselves – a principle of statistical MT. A system “learns” by analysing the source texts and target texts of documents that have already been translated by human translators.

According to Google (Google n.d.(a)), the texts or corpora from which its system “learns” come from books, organisations such as the UN, and the Internet. The Google Translate system scans the texts it harvests, searching for patterns between the target texts and source texts that are unlikely to have occurred by chance. Once such a statistically significant pattern is identified, this pattern can be used to translate similar texts in future. Google Translate repeats this process of pattern recognition continuously with new texts, and subsequently a database of a vast amount of patterns gets established for the system from which it can draw translations.

Google Translate thus works on the principle of statistical MT. In Schulz (2013), Franz Och, principal scientist and head of MT at Google Inc. at the time, explains how Google Translate incorporates statistical MT as follows: “[...] what the system is basically doing (is) correlating existing translations and learning more or less on its own how to do that with billions and billions of words of text. In the end, we compute probabilities of translation”.

The more data available, the better the Google Translate system works. Therefore the Internet, being a platform that contains an enormous amount of data, including an abundance of already existing translations from which the Google Translate system can learn, is a crucial component of Google Translate. Since Google is, among other things, a powerful internet search engine with a mission “to organize the world’s information and make it universally accessible and useful” (Google n.d.(b)), Google Translate is positioned excellently for access to corpora on the world-wide web. According to Och, Google Translate’s “[...] current quality improvement curve is still pretty steep” (Helft 2010). It follows that Google Translate’s output should improve over time as even more data – mostly through the Internet – become available to the system in different language pairs.

Although Google Translate currently supports 80 languages (Google n.d.(a)), there is not an equal amount of data available for those 80 languages. In the case of language pairs that have few documents available on platforms where Google Translate typically harvests translations, fewer patterns will have been detected for those language pairs by the Google Translate system. The quality of translations by Google Translate will thus be lower for those language pairs than for language pairs for which an extensive database of patterns has already been established. For example, French and English are prominent world languages and a prominent language

combination, whereas a language combination such as Afrikaans and English has a much smaller user base. Consequently, much more French–English document pairs than Afrikaans–English document pairs would be available on the Internet, and therefore available for Google Translate to learn from. It would therefore be more likely that Google Translate would produce better French–English output than Afrikaans–English output.

There are also other factors that come into play with regard to translation quality and MT in general. Some languages just seem to be more suited to MT than others. For example, translators working with highly inflected languages such as Greek and Polish indicated that MT underperformed and was thus of little use (Lagoudaki 2008:267). The fact that the quality of MT output depends on the language combination is confirmed by others such as Brasler and Zetzsche (2013), who call the Scandinavian languages (except Finnish) in combination with English “a sort of poster child for output quality”. Other languages, such as German (which differs greatly from English on a syntactical level), did not produce the same usable results. Another language pair that is considered “more advanced than others” is English–Spanish; French–Russian output, on the other hand, is of lower quality (DGT 2011).

In a bid to involve users to help improve the quality of translations into their languages, Google Inc. has created the Google Translator Toolkit. This online translation software incorporates translation memory technology, with which translators (or anyone who is interested) can create translations by using translation memories online or uploading their own translation memories to be available to themselves online – and to Google Translate – to draw from. By means of this application, one could gain access to other shared translation memories and glossaries, but the quality of these shared resources is questionable (Bowker and Fisher 2010). Google has also established other ways in which volunteers could contribute to improving some languages; one can sign up as a volunteer, and it seems that a Google Community Translation feature is under way in this regard (Google Translate Blog 2013).

In the past, the quality of the output of Google Translate has been investigated and compared to other MT systems by means of a few translated sentences (Aiken and Balan 2011), and Austerlühl (2011) has compared the quality of Google Translate’s output of 60 business letters translated from English into German with that of MS Bing Translator. However, we have not found any studies that compare the quality of Google Translate’s output from one year to that of the next, for a few consecutive years. Given all the above, the authors of this article wanted to investigate Google Translate’s performance in a less prominent language combination – one that was relevant to our immediate environment – for a few consecutive years. Subsequently, our study, containing several phases, investigates this in different ways. The first three phases of the study will be discussed in section 4.

4. Empirical study into the quality of Google Translate's output over three consecutive years

4.1 Investigation into quality improvement in Google Translate's output

In order to determine whether the quality of output has indeed improved for the language pairs that we at the Stellenbosch University Language Centre work with most, our first research question in this article is: Has the quality of Google Translate's output in the language pair Afrikaans–English (AF–EN) and vice versa improved over the period 2010 to 2012?

4.1.1 Phase 1: Quality of Google Translate's output in 2010 measured against that of a student translator and a professional translator

At the Stellenbosch University Language Centre, we often receive enquiries about using Google Translate, ranging from why translation is so expensive “if you can just use Google Translate”, to whether we would edit (as editing is cheaper) very suspect translations that usually turn out to be the work of Google Translate.

Such enquiries prompted us to conduct the first phase of our study (Van Rensburg et al. 2012) in which the quality of translation products created in 2010 by Google Translate, a translation student and a professional translator were assessed and compared. Six different text types were translated, one document each from Afrikaans into English and another from English into Afrikaans. Five raters assessed the quality of the translation products. The raters all held language-related degrees and worked as language practitioners (years of experience at the time varied from 6 to 29 years). The 36 translation products were assessed using a holistic assessment tool developed by Colina (2009) for the evaluation of the quality of translations. We adjusted the tool to be effective in our context. For the purposes of this article, we will refer to that assessment tool as the “first assessment tool”. Four weighted categories were evaluated in the first assessment tool, namely (i) Target Language, (ii) Functional and Textual Adequacy, (iii) Non-Specialised Content, and (iv) Specialised Content and Terminology. Each category contained four descriptive sentences, ranging from positive to negative statements. The raters had to choose the most suitable descriptive sentence in each category.²

As expected, the results of the first phase of our study showed that the quality of the translations by a professional translator would not require a client to spend much time correcting such translation products. The student translator's work was of lesser quality than that of the professional translator, but better than the Google Translate output. The results further showed that the translations by Google Translate needed substantial improvement regarding quality. It was important to show that the professional translator's work was acceptable and useful as it was delivered, whereas the student's work needed a fair amount of revision and Google Translate's output needed extensive post-editing before it would be useful for professional purposes. We wanted to illustrate to clients that Google Translate output could not be used as it was delivered by the system, and that although it may cost them next to nothing to obtain such translations, they would still have to pay for post-editing or try to post-edit the texts themselves to make the translations useful for professional purposes. This may seem obvious to the reader, but in our experience it is not that obvious to some clients.

² Please refer to Van Rensburg et al. (2012) for more information on the raters, the first assessment tool, and how the latter was adjusted.

Another finding was that, of the six text types translated by Google Translate in 2010, slide-show texts yielded the best results, with an average of 46% assigned by our raters. Figure 1 shows the average scores that the different translation entities achieved for the different text types in our evaluation during the first phase of the study.

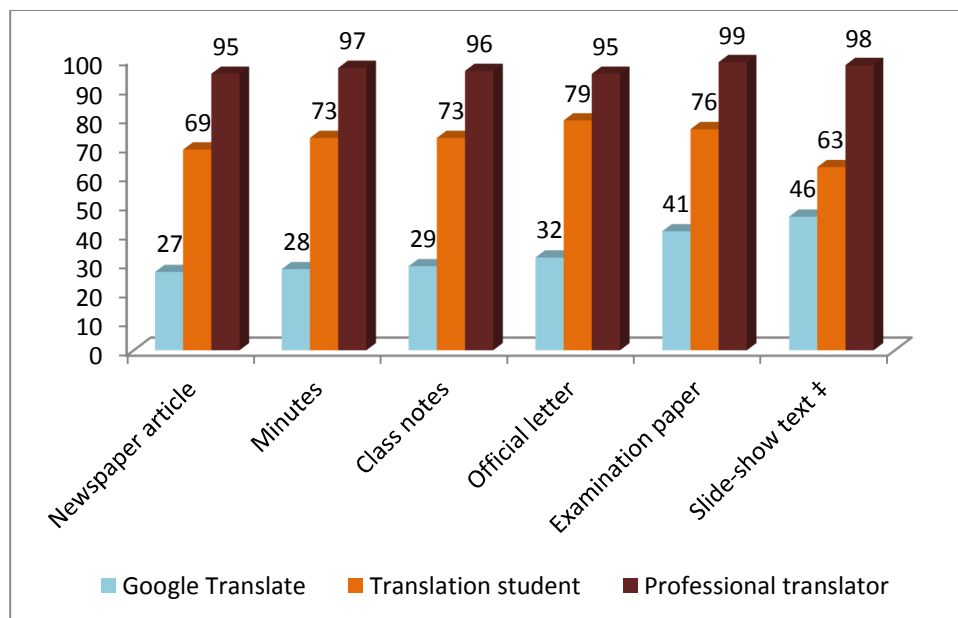


Figure 1. Performance per text type by a professional translator, a translation student and Google Translate in 2010

‡In the original figure (Van Rensburg et al. 2012), this text type was labelled “PowerPoint slides”. As explained in section 4.2.1.1, we have since decided to call it “slide-show text”.

4.1.2 Phase 2: Quality of Google Translate’s output in 2011 and 2012

In order to answer our first research question, relating to whether Google Translate’s output improves over time, additional Google Translate translation sets of the initial six Afrikaans and six English source texts were generated in 2011 and 2012.³ In the next part of our study, those sets were evaluated in exactly the same manner as the 2010 translation sets. We used the same five raters and exactly the same assessment tool. The raters received the 2011 texts early in 2012 for assessment, and the 2012 texts a year later – early in 2013. When we compared each year’s combined results to those of the other two years, the results showed a steady improvement in the quality of Google Translate target texts over the three years, in both language pairs (AF–EN and EN–AF), as illustrated in the combined results in Figure 2.

³ The 2010 target texts were not made public in any way, so they would not have been available on a platform where Google Translate could have harvested them and improved its translations in this manner.

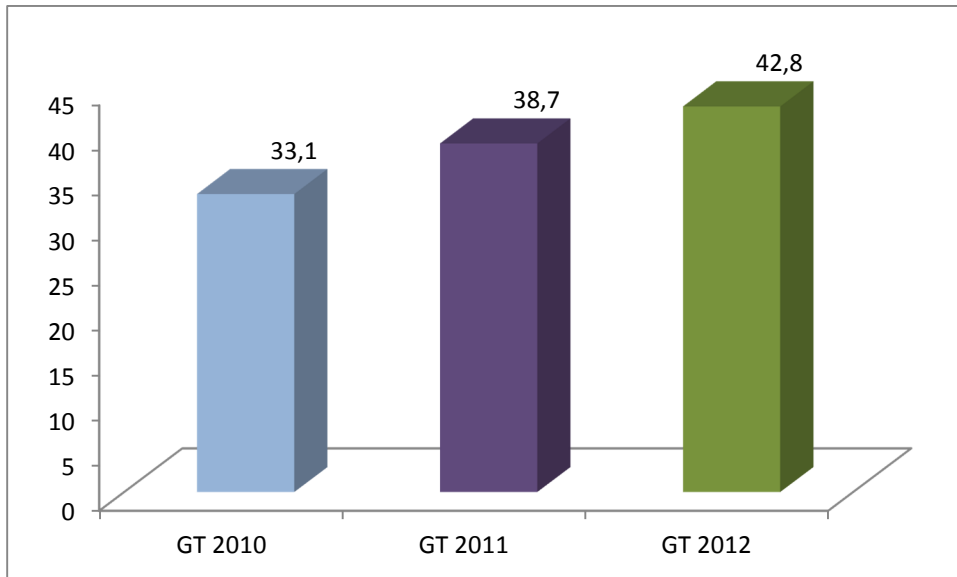


Figure 2. Overall improvement in quality of Google Translate output over three years according to the first assessment tool (AF–EN and EN–AF combined)

Therefore, the answer to our first research question is in the affirmative: Google Translate’s average output for the combined text types has indeed improved over the period 2010 to 2012, in both language pairs (AF–EN and EN–AF).

As previously mentioned, our initial investigation showed that slide-show texts seemed to yield the best results when translated by Google Translate. However, when we had a closer look at the results reflecting Google Translate’s performance per text type in 2010, 2011 *and* 2012, it emerged that the slide-show texts translated in 2011 scored significantly higher marks than those translated in 2012 – an unexpected deviation in the general pattern of improvement that emerged from the results for the six text types combined. Figure 3 shows this deviation.

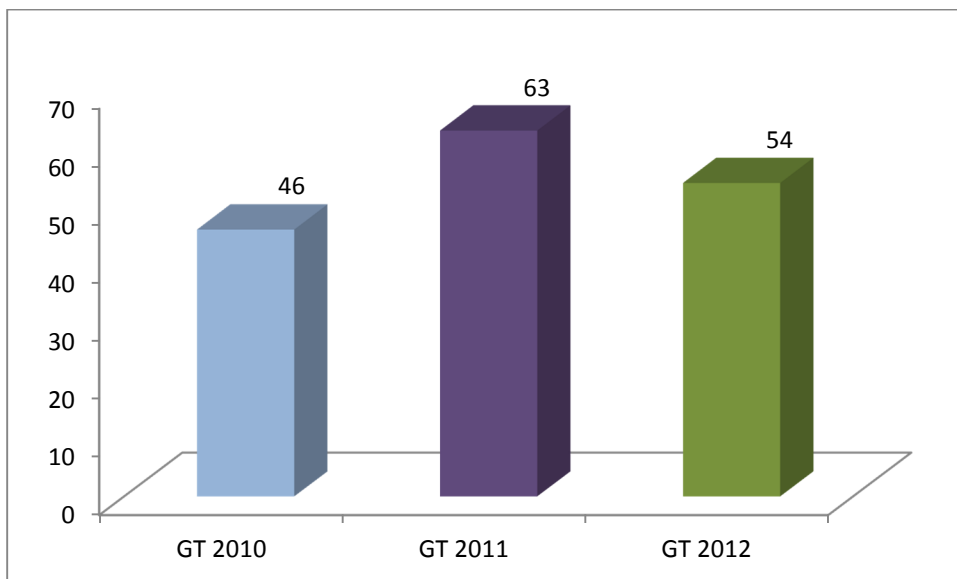


Figure 3. Unexpected results: 2011 translation of slide-show texts scored higher than the 2012 translation

4.2 Phase 3: Error analysis of slide-show texts translated by Google Translate (2010-2012)

Due to the unexpected results of Google Translate's performance over the three years with regard to the slide-show texts, we decided to use a second assessment tool to conduct an error analysis of the 2010 to 2012 slide-show texts generated in the language pair AF–EN. We wanted to determine why the results of the slide-show texts deviated from the expected pattern by (i) verifying the raters' evaluations of that text, and (ii) identifying the most frequent types of errors made in that text type. We chose to work with the AF–EN language pair after we considered the two language directions' results for all texts separately. It transpired that, for all text types over the three years, the AF–EN results were poorer.

Therefore, our second research question is: To what extent do the raters' evaluations of the slide-show texts translated by Google Translate in the language pair AF–EN in 2010, 2011 and 2012, conducted by means of the first assessment tool, correlate with the error analysis performed on the same texts by means of a second assessment tool?

4.2.1 Error analysis

In this article, we use the term “error analysis” to refer to the identification and classification of individual errors that occur in a translated text. According to Stymne and Ahrenberg (2012:1785), an error analysis of the output of an MT system gives an indication of the “specific strengths and problem areas” of that system. This information is difficult to obtain from standard automatic evaluation metrics such as BLEU (Papineni, Roukos, Ward and Zhu 2001) or by having humans rank sentences (Callison-Burch, Fordyce, Koehn, Monz and Schroeder 2001, in Stymne and Ahrenberg 2012:1785).

We also wanted to record what typical translation errors Google Translate made so as to know what to expect from translations by this application, and how to advise prospective users who turn to us for guidance. The evaluation of the quality of MT in general is a subjective process by nature, because – as Flanagan (1994) and Fiederer and O'Brien (2009) put it – the process is based on human judgement. According to Flanagan, the classification of errors may help these judgements “to be made in a more consistent and systematic manner” (1994:71). Different types of errors require different remedies. Therefore, an error analysis could give insight into the post-editing effort that would be necessary to make a machine-translated text useful (Gaspari, Toral and Naskar 2011:14).

Pym (1992, 2010) distinguishes between binary and non-binary errors in translation. In the case of binary errors, there is a clear right and wrong translation option. Such errors seem to be related to grammar and the rules of the target language. Non-binary errors are harder to judge; they involve at least two or more acceptable translation options and still more unacceptable options.

In the same vein, scholars in translation studies seem to have distinguished between language errors in translations and translation errors (cf. Koby and Champe 2013). A language error entails an “error in the mechanics of target language usage”, whereas a translation error marks an “error that is one of transfer of meaning” (Koby and Champe 2013:165). While language errors could be common to any form of written communication, translation errors can only

occur when a source text and a target text stand in a specific relation to each other (Conde 2013:98, Hansen 2010).

Despite this distinction, both kinds of error could significantly influence the quality of a target text. Therefore, we took both kinds of error into account in our error analysis. Since the significance of errors in a target text could vary (Koby and Champe 2013:165, Hansen 2010), we decided to distinguish in our analysis between errors that would have a negative impact on the transfer of meaning and other errors that are clearly wrong but that would not influence the reader's understanding of the text, although the quality of the text would be affected (an example of this would be language errors). We kept it simple – severe errors (with regard to their effect on the meaning of the text) were assigned a weight of 2, and less serious errors were weighted as 1. Consequently, the higher the score, the lower the quality of the translation.

4.2.1.1 Source text

As stated earlier, we decided to analyse the AF–EN slide-show text translation products that had been evaluated in the first part of the study. The source text was originally created as a Microsoft PowerPoint presentation for a lecture in social anthropology and consisted of 312 words, forming 10 slides. Since Google Translate cannot translate text in PowerPoint format, the text had to be extracted, hence our reference to “slide-show text” rather than “PowerPoint slides” to denote this text type. Also, slide-show texts may occur in a variety of presentation software applications other than PowerPoint, such as Prezi or Apple's Keynote. The translations by Google Translate of this source text in three consecutive years – namely 2010, 2011 and 2012 – were analysed to identify translation errors as an indication of quality.

4.2.1.2 Framework for evaluation

The Framework for Standardized Error Marking of the American Translators Association (ATA; henceforth referred to as “the ATA Framework”) was adapted to perform the error analyses, thus becoming our second assessment tool. Among other evaluation tools, evaluators employ the ATA Framework to evaluate translations submitted for the ATA certification examination (ATA 2009, 2013). The ATA Framework specifies errors by type, and Doyle (2003:21) finds that it provides “a ready-made, standardised, time-tested, and professionally recognized model for conducting theory-based, systematic, coherent, and consistent evaluations of [...] translations”. The ATA Framework was therefore a suitable basis for the error analysis we wished to perform.

We adapted the ATA Framework used by Doyle (2003:22-23) slightly to provide for errors that may occur in machine-translated text and that usually do not occur in translations by humans (see the appendix for our adapted framework, the second assessment tool in this study). We added two categories, namely Non-Translation: Insertion of Word from Source Language, and Switched Elements; both additions are discussed in section 4.2.2.3. The categories of Illegibility and Unfinished Passage were inapplicable and therefore excluded. Since the ATA certification examination is a handwritten exercise taken within a set time frame, legibility and the ability to finish translating a passage are relevant factors in the evaluation of candidates taking this exam. In our case, however, where we assessed translations generated by a computer, neither legibility nor time constraints were factors influencing the performance we wished to measure. Like Doyle (2003), we coded the criteria numerically, from 1 to 22, and inserted comments

containing the corresponding numbers in the analysed texts to indicate to which category each error in the texts belonged.

The first author performed the error analysis, after which the second author authenticated it. When we disagreed on an error category, we discussed the case until we reached a consensus. At the time of the error analysis, the first author had 12 years of experience as a language practitioner and her highest qualification was an MA in General Linguistics. The second author had 10 years of experience and her highest qualification was an MPhil in Translation Studies. She was also working on her PhD in Translation Studies at the time.

4.2.2 Results

4.2.2.1 Number of errors

The error analysis of the three translations yielded 71 errors for the Google Translate (GT) 2010 translation, 60 errors for the GT 2011 translation, and 48 errors for the GT 2012 translation. This yields intervals of improvement of 11 and 12, respectively. The interval of improvement between the GT 2010 and GT 2012 error count is 23. The results are illustrated in Figure 4.

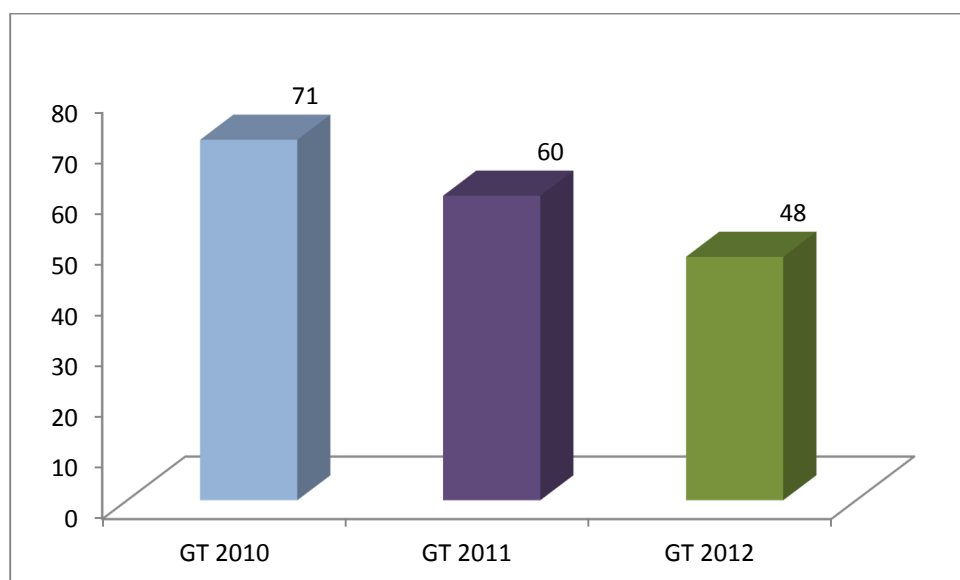


Figure 4. Number of errors in 2010, 2011 and 2012 slide-show text translations by Google Translate

There is a definite decrease in errors over the three years in question, which indicates that the quality of Google Translate's output has indeed improved since 2010. This confirms Och's assertion that Google Translate's output quality for particular language combinations improves over time, since more text in those language combinations gradually becomes available to Google Translate to "learn" from.

4.2.2.2 Weighted error analysis scores

The weighted error analysis scores were 107 for the GT 2010 translation, 92 for the GT 2011 translation, and 69 for the GT 2012 translation. See Figure 5 in this regard (recall that, as mentioned in section 4.2.1, the higher the score, the lower the quality).

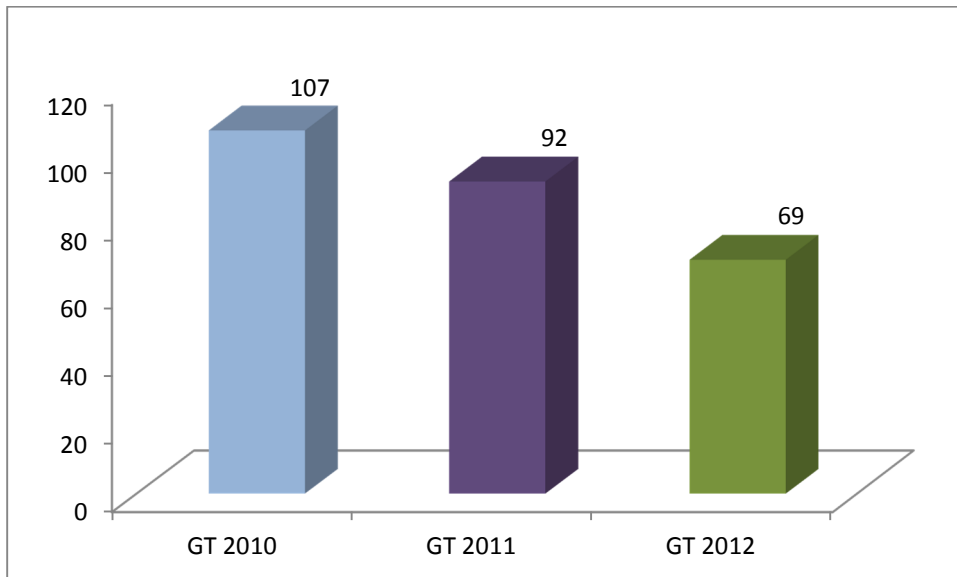


Figure 5. Weighted error analysis scores of the 2010, 2011 and 2012 slide-show text translations by Google Translate

There is a steady decrease in the weighted error analysis scores over the three years, with an interval of improvement of 15 between the GT 2010 and GT 2011 translations, and an interval of improvement of 23 between the GT 2011 and GT 2012 translations. There is a combined interval of improvement of 38 between 2010 and 2012. Table 1 shows the number of errors and the weighted error analysis scores of the 2010, 2011 and 2012 translations for easy comparison.

Table 1. Number of errors and weighted error analysis scores of the 2010, 2011 and 2012 slide-show text translations by Google Translate

	Number of errors	Weighted error analysis score
GT 2010	71	107
GT 2011	60	92
GT 2012	48	69

The error scores and the weighted error analysis scores both indicate a significant improvement in Google Translate’s output over the three years in question. This once again confirms our finding concerning our first research question. However, the results of the holistic evaluation conducted by means of the first assessment tool for the same set of texts differ from this analytical evaluation. As mentioned earlier, the first assessment results showed that the quality of the GT 2011 slide-show texts was perceived to be higher than that of the 2012 version.

Consequently, the answer to our second research question is that, due to unexpected results in the GT 2011 slide-show texts obtained in the first assessment, the first assessment results do not correspond fully with the results from the error analysis. The error analysis indicates a steady improvement over all three years in question, whereas the first assessment results indicate an improvement only if the 2010 results are compared with the 2012 results, without taking the 2011 results into account.

Upon closer investigation of the individual raters' scores in the first assessment, it became clear that the sharp increase in the score of the GT 2011 slide-show translation was due mainly to two of the five raters perceiving the GT 2011 slide-show translation to be of a much higher quality than the 2012 version. One should take into account that new errors – that is, errors that did not occur in the 2010 or 2011 translations – had been introduced in the GT 2012 text. For example, in the 2010 and 2011 texts, “voorouers” was correctly translated to “ancestors”. The 2012 translation used “parents” as a translation of “voorouers”, which constitutes a mistranslation according to the second evaluation tool. Another new, prominent mistranslation in the 2012 text was “foreign affairs” for “buite-egtelike verhoudings” (‘extra-marital affairs’). Although many errors that occurred in 2010 and 2011 have been improved on in the 2012 text, the errors that were newly introduced in 2012 simply may have borne more negative weight in the opinion of the two raters concerned, and may have tipped the scale for those raters to assign the 2012 text a lower score.⁴ This is possible particularly since the first assessment tool is a holistic evaluation tool. The discussion of the distribution of errors in the last section of this article sheds more light on such newly introduced errors.

This brings us to our third and last research question, namely: What was the distribution of errors in the Google Translate AF–EN translations of the slide-show texts in 2010, 2011 and 2012?

4.2.2.3 Distribution of errors

In Figures 6 to 8 below, the errors that have been identified in the error analysis of the translations by Google Translate in the different years are arranged in descending order of frequency. We followed the reasoning in Doyle (2003) that, when the data are presented in this manner, problem areas are revealed more clearly.

⁴ Keep in mind that the raters did not assess the texts comparatively – the assessments took place in yearly intervals.

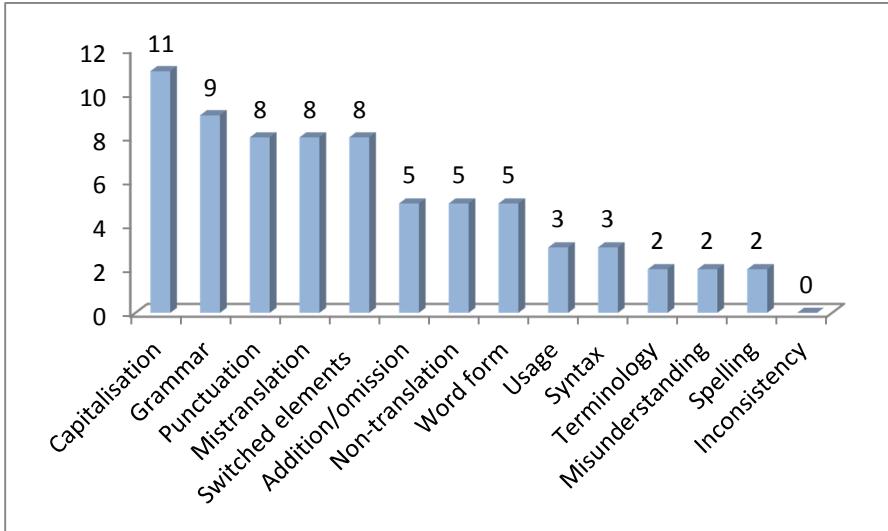


Figure 6. Error distribution in GT 2010 translation

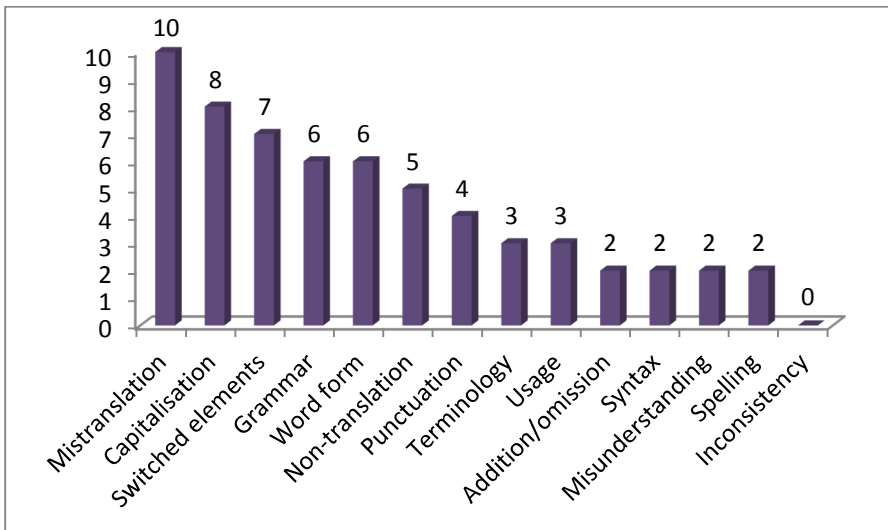


Figure 7. Error distribution in GT 2011 translation

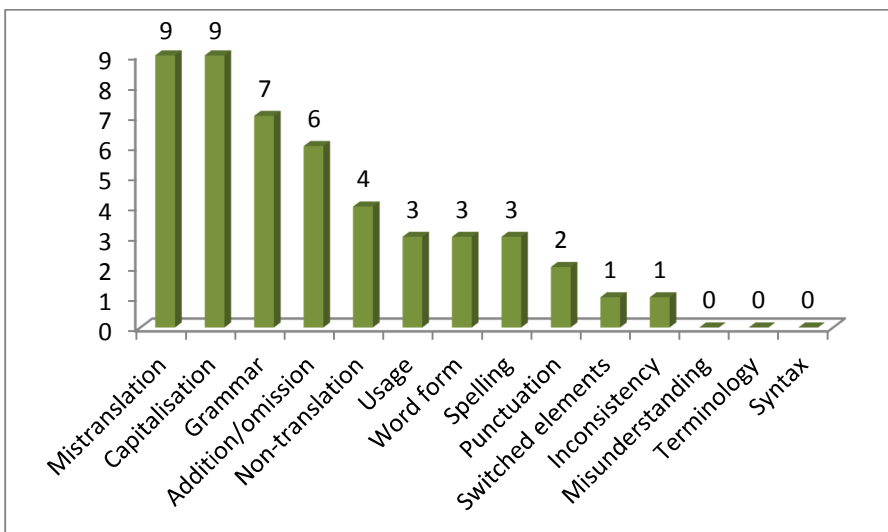


Figure 8. Error distribution in GT 2012 translation

Since the 2012 GT translation is the most recent translation – and therefore the most relevant – as well as the best of the Google Translate target texts, we will begin by discussing that text and Figure 8, mentioning the earlier two translations when relevant.

(i) *Mistranslation*

Nine mistranslation errors were recorded in the analysis of the 2012 translation by Google Translate. Table 2 contains examples of mistranslations in the 2012 translation:

Table 2. Mistranslations in GT 2012 translation

Source text	Target text	Appropriate translation
Rituele mistifiseer	Ritual *demystify	Ritual mystifies/Rituals mystify
Afhanklikes ontvang slegs 30% daarvan	*Dependents receive only 30% of *them	Dependants receive only 30% of that
verkry invloed	*possess influence	obtain influence

* denotes an error

The most entertaining mistranslated phrase occurred with the translation of “bier, handel en **buite-egtelike verhoudings**”, which was rendered in the different years as follows:

2010: beer, trading and extra-marital ***relations**

2011: beer, ***marketing** and extra-marital ***relationships**

2012: beer, trade and ***foreign** affairs

“Extra-marital relations” (2010) and “relationships” (2011) were marked as word choice/terminology errors, since the unmarked form of this would be “extra-marital affairs”. This was a borderline case, since the translation may be acceptable, but is not 100% correct. This brings to mind Pym’s (1992, 2010) non-binary errors mentioned earlier in this article. With “marketing” that occurred in the 2011 translation of “handel”, another mistranslation was introduced, and “extra-marital relationships” was once again marked as a word choice/terminology error. The 2012 translation containing the mistranslation “foreign affairs” for “buite-egtelike verhoudings” could indeed result in serious misunderstanding.

As mentioned in the previous section, “voorouers” was mistranslated a few times as “parents”, instead of “ancestors”, in the three translations. It is significant that the correct translation equivalent, “ancestors”, was indeed used in the Google Translate 2010 translation (albeit only once), and both times where it occurred in the 2011 translation. In 2012, “voorouers” was again translated as “parents”.

This regression could be an indication of the dynamic state of Google Translate’s translation memory system or database from which it draws its translation choices. Due to its crowdsourcing and harvesting of translations wherever possible, anyone could influence the Google Translate database, especially in a language combination such as AF–EN, for which less data exist than for a more dominant language pair such as French–English. Precisely

because there is no control over what is fed into the database, good translations could be replaced by less fortunate ones just as easily as the other way round. Google Translate is usually quick to rectify mistranslations when it becomes aware of users playing the system for so-called “Easter eggs” (amusing translation options)⁵ in more prominent languages.

For the 2011 translation, as in the case of the 2012 translation, Mistranslation was the error category containing the highest number of errors – a total of ten. The 2010 translation had eight errors in this category.

Two other recent studies on the evaluation of MT, those of Valotkaite and Asadullah (2012) (Portuguese–English) and Avramidis, Burchardt, Federman, Popović, Tscherwinka and Vilar (2012:1128) (Spanish–English, German–English, English–German), found “wrong lexical choices” to be the most frequent error type that they encountered. “Wrong lexical choices” and “mistranslation” denote broadly the same kind of error in this case. Gaspari et al. (2011) highlight the mistranslation of compounds in particular as a prominent error in German–English MT.

(ii) Capitalisation

There were also nine errors with regard to capitalisation in the 2012 text. Capitalisation concerns whether upper or lower case is used appropriately in the target text to reflect how it has been used in the source text. The data suggest that there has been almost no decrease in errors in this category over the three years in question, as can be seen when comparing Figures 6, 7 and 8. It seems that Google Translate has not yet devised a way of dealing with transferring source-text capitalisation to the target text appropriately.

(iii) Grammar

With regard to Grammar, the category reflecting the second highest number of errors for the 2012 translation, errors were generally rooted in subject-verb agreement or the form of the verb. For example, “Women [...] *does not control the resources”, and “ritual *hide the gender inequality”.

As can be seen in Figures 6, 7 and 8, Grammar was a prominent error category in all three years.

(iv) Addition or omission

In the three translations by Google Translate, elements were omitted rather than added. In addition to some words that have been omitted in the translated texts, the graphic element that the source text contained was also omitted in all three translations, since Google Translate does not insert such elements in its generated translations. Post-editing in this respect is definitely required.

Izwaini (2006:147) also identified “addition and deletion problems” as a specific concern in a study that evaluated Google Translate’s beta English–Arabic/Arabic–English output.

⁵ For example, Hauksson (2011) recorded that when one translated “Lalalalala” from English to Icelandic, Google Translate translated it as “Björk”, the famous Icelandic singer. Being Icelandic himself, Hauksson emphatically states that this was an incorrect translation. It is humorous, however. When the authors tried this, “Lalalalala” translated to “Umm” – it has clearly been changed.

(v) Non-translation

The Non-translation category was added to our framework to provide for a common Google Translate error, namely that if the application does not find a match for a source-text word or combination of words, it simply inserts the source-text word or combination of words into the target text, almost as a placeholder. There were five instances of non-translation in the Google 2010 and 2011 translations, whereas the 2012 translation contained four instances. “Mistifikasie”, “mistifiseer”, “trekarbeiderinkomste” and “saameet” (the contraction of the latter was a spelling error in the source text) baffled the application in all three years.

Valotkaite and Asadullah (2012) identified “untranslated words” as a prominent error category in their study of the quality of two rule-based MT systems’ output in Portuguese–English.

A way of ensuring optimum MT results is by editing source texts carefully to avoid unnecessary mismatches or non-translation, as in the case of the contraction “saameet” in our source text. A further step is to apply controlled language rules to the source text, by which constructions or elements known to cause errors in MT are edited or removed beforehand. Problematic features include long sentences, gerunds, long noun strings, and ambiguous anaphoric referents (Fiederer and O’Brien 2009:52).

(vi) Spelling

Three spelling errors were recorded for the 2012 translation, as well as two spelling errors for each of the 2011 and 2010 translations. In all three translations, “dependants” (*“dependents”) was spelled wrong consistently. There was an additional issue with spelling, namely that we were not able to change the language setting to differentiate between American English and British English in Google Translate – only one language option was given, namely “English”. Therefore, in each of the translations, one error was counted for the fact that American English rather than British or South African English was used (for example, “labor” instead of “labour”), since the implicit translation brief for all documents at Stellenbosch University is that British English be used, and it would have been necessary to correct those words had the texts been post-edited. We felt that counting an error for each occurrence of a word with American spelling would have been unreasonable and that it would have skewed the results, since the American forms are not intrinsically wrong.

Spelling is a Google Translate problem that Izwaini (2006:147) highlights as well, albeit with regard to a beta language combination at the time.

Scores for the categories Word Form and Usage did not vary significantly over the three years.

(vii) Punctuation

The incidence of punctuation errors decreased considerably over the three years. Only two errors in this category were identified in the 2012 translation, while four punctuation errors occurred in the 2011 translation – half the number of errors logged for the 2010 translation. Punctuation errors consisted mostly of colons replaced by commas, which influenced the transfer of meaning since the colon in those cases indicated that an explanation of whatever preceded the colon would follow, whereas a comma usually indicates a pause or serves as a mechanism to facilitate a list of items. In a few cases, the comma in the source text was simply omitted in the target text.

(viii) Switched elements and syntax errors

The Switched Elements category was also added to our framework to provide for a frequent error that occurs in Google Translate translations. “Elements” may refer to words or phrases. This category involves two adjacent elements having been translated correctly in the target text, but appear to be switched around, in comparison to the position of those elements in the source text (see the appendix with our evaluation framework, the second assessment tool). Simply marking such an error as a syntax error is one way of dealing with this, since the natural word order of the target language has indeed not been followed if elements have been switched. However, it does not explain what has happened in the text and in the translation process.

Consider the following example:

Source text: [...] in die vorm van rituele om goeie verhoudings te bou

Target text: *[...] in the form of rituals to good building relationships

In the target text, “building” and “good” have been switched. The Switched Elements category allows for differentiation between two elements that have merely been switched, on the one hand, and a problematic or nonsensical arrangement of more than two words or other elements of a sentence, on the other hand, which is identified by means of the Syntax category. Due to this differentiation, the 2012 translation does not contain any syntax errors and only one switched element, the 2011 translation contains two syntax errors and seven switched elements, and the 2010 translation contains three syntax errors and eight switched elements. From this it is clear that switched elements as well as syntax errors decreased steadily over the three years.

In a study by Avramidis et al. (2012) in which the quality of the output of three MT systems in the language combinations German–English, German–Spanish and Spanish–German was investigated, syntax is listed as the next frequent error after wrong lexical choices.

(ix) Inconsistency

In the 2012 translation, one inconsistency was recorded. In this year, “voorouers”, which occurred twice in the source text, was translated once as “parents” and once as “ancestors”. No inconsistencies were recorded for the previous two years’ translations. The Inconsistency error category is an example of a category in which the most recent translation performed worse than those of the previous years.

(x) Misunderstanding

Errors with regard to misunderstanding decreased over the three years, with no errors in this category recorded for the 2012 translation. The proper noun “Frederick Engels” occurred in the source text; in the 2010 and 2011 translations the surname “Engels” was translated as “English”, which we regarded as misunderstanding, since the surname was mistaken for the name of a language and translated as such. However, the surname was translated correctly in the 2012 translation when used in combination with the name Frederick. This indicates that this combination must have been added to the database from which Google Translate draws its options for translation from Afrikaans to English during the time frame between our 2011 and 2012 translations.

Studies by Gaspari et al. (2011), concerning the language combinations of German–English and Dutch–English, and Chang-Meadows (2008, in Temizöz 2012), who worked in the combination

of Chinese–English, also identified the translation of proper nouns as problematic for MT systems.

(xi) Terminology

Terminology errors decreased over the three years in question, with no errors recorded for the 2012 translation in this category.

Terminology errors and Mistranslation errors (the highest ranking category for the 2011 and 2012 translations, as discussed earlier in this section) are closely related categories in that both are concerned with the best translation equivalent in the target language. A terminology error in particular occurs when a term specific to a specialised subject field is not used even though the corresponding term is used in the source text; a mistranslation error is a more general indication that a segment of the source text has not been conveyed properly in the target language (see the appendix for our evaluation framework, the second assessment tool).

Van Rensburg et al. (2012) (AF–EN and EN–AF) and Zuo (2010, in Temizöz 2012) (English–Chinese), among other studies on the performance of MT systems, have found that MT performs well terminology-wise if the database from which the translation options are drawn contains the correct equivalents.

4.2.3 Final word on the error analysis

The error analysis has pointed out that mistranslation posed the largest risk associated with using Google Translate in the language pair we investigated. Grammatical errors and the non-translation and omission of elements are other likely risks that could impact on the quality of the target text when Google Translate is used to generate translations. These findings correspond with findings of other studies on the evaluation of the output and performance of MT systems.

The error analysis has also shown that even the best translation in our case study, namely the 2012 translation, showed unexpected new errors due to the very same dynamism of the databases that generally helps to improve Google Translate's output. We suspect that these unexpected new errors also led two of our raters to perceive the quality of the 2012 Google Translate translation of slide-show texts more negatively than was actually the case, in comparison to their perceptions of the translations from the previous two years.

5. Summary, conclusion and recommendations for further studies

Online MT has put MT on the map for the general public who uses the Internet. Over the past few years, we at the Stellenbosch University Language Centre have experienced that more and more clients view online translation services, particularly those that are free of charge, as a viable option to meet their translation needs.

The empirical study described in this article focused on the quality of the output of the free online MT application Google Translate. We investigated three research questions in this regard. Our first research question was: Has the quality of Google Translate's output in the language combination AF–EN and EN–AF improved over the period 2010 to 2012? The verdict is a resounding yes, in accordance with Och and Austermühl's assertions regarding the

improvement of MT over time. Our second research question investigated to what extent the raters' evaluation of the slide-show texts translated by Google Translate in the language pair AF–EN in 2010, 2011 and 2012, conducted by means of our first assessment tool, correlated with the error analysis of the same texts, performed by means of our second assessment tool. The first assessment results did not correspond fully with the results of the error analysis. The first assessment results showed that the raters awarded the GT 2011 slide-show translation higher marks than the 2012 translation. We suspect that new errors in the 2012 translation may have had a large impact on the raters' perception of the quality of the 2012 translation. Our third research question explored the distribution of errors in the Google Translate AF–EN translations of the slide-show texts in 2010, 2011 and 2012. We found mistranslation to be the largest risk of using Google Translate. Errors regarding capitalisation, grammatical errors and the non-translation or omission of elements – followed by spelling errors and punctuation errors – are other likely risks users of this application must be prepared to deal with.

The raw texts that were generated by Google Translate for this study were not usable for the purpose for which the source text had been created, namely for a presentation in a lecture. The translations would require thorough post-editing before being able to accommodate multilingual information exchange in the classroom. However, the fact that the raw Google Translate output was not usable without being post-edited does not render Google Translate useless. Google Translate is a useful tool in certain contexts – it depends on what the translations will be used for and whether they are post-edited adequately. We agree with Fiederer and O'Brien (2009)'s view that, "when used intelligently, MT does not have to be synonymous with poor quality in translation".

A possibility for further study is an investigation on how long it would take to post-edit the texts that have been generated by Google Translate in this research. This will tackle a question of particular interest to clients, namely how much they will have to spend on post-editing if they decide to harness MT. In addition, since Google Translate added isiZulu, among other languages spoken in Africa, to its oeuvre in December 2013 (Google Translate Blog 2013), research on the quality of its translations involving isiZulu could also now be undertaken.

References

- Aiken, M. and S. Balan. 2011. An analysis of Google Translate accuracy. *Translation Journal* 16(2). Available online: <http://translationjournal.net/journal/56google.htm> (Accessed 13 February 2012).
- ATA (American Translators Association). 2009. ATA Certification Program. Available online: https://www.atanet.org/certification/aboutexams_overview.php (Accessed 21 June 2013).
- ATA (American Translators Association). 2013. ATA Certification Program: Framework for Standardised Error Marking. Available online: https://www.atanet.org/certification/aboutexams_error.php (Accessed 25 February 2013).
- Austermühl, F. 2011. On clouds and crowds: Current developments in translation technology. *T21N – Translation in Transition 2011-09*. Available online: www.t21n.com (Accessed 3 October 2013).

Avramidis, E., A. Burchardt, C. Federman, M. Popović, C. Tschewinka and D. Vilar. 2012. Involving language professionals in the evaluation of machine translation. Proceedings of the Language Resources and Evaluation Conference, 21-27 May, Istanbul, Turkey. Available online: http://www.lrec-conf.org/proceedings/lrec2012/pdf/294_Paper.pdf (Accessed 15 April 2013).

Bennett, S. and L. Gerber. 2003. Inside commercial machine translation. In H. Somers (ed.) *Computers and translation*. Amsterdam and Philadelphia: John Benjamins. pp. 175-190.

Bowker, L. and D. Fisher. 2010. Computer-aided translation. In Y. Gambier and L. van Doorslaer (eds.) *Handbook of translation studies*. Available online: www.benjamins.com/online/hts (Accessed 30 August 2011).

Brasler, C. and J. Zetzsche. 2013. Machine translation: 50 shades of grey. *The ATA Chronicle* February 2013. Available online: http://www.atanet.org/chronicle/feature_article_february_2013.php (Accessed 18 September 2013).

Colina, S. 2009. Further evidence for a functionalist approach to translation quality evaluation. *Target* 21(2): 235-264.

Conde, T. 2013. Translation versus language errors in translation evaluation. In D. Tsagar and R. van Deemter (eds.) *Language testing and evaluation*. Frankfurt: Peter Lang. pp. 97-112.

DGT (Directorate-General for Translation). 2011. Translator profile: What skills for a changing market? Brussels: European Commission. Available online: http://ec.europa.eu/dgs/translation/profframmes/languageindustry/platform/documents/translator_profile_conclusions_en.pdf (Accessed 19 September 2013).

Doyle, M. 2003. Translation pedagogy and assessment: Adopting ATA's Framework for Standard Error Marking. *The ATA Chronicle* November/December: 21-29.

Fiederer, R. and S. O'Brien. 2009. Quality and machine translation: A realistic objective? *Journal of Specialised Translation* 11: 52-74.

Flanagan, M. 1994. Error classification for MT evaluation. In *Technology Partnerships for Crossing the Language Barrier*. Proceedings of the First Conference of the Association for Machine Translation in the Americas, 5-8 October, Columbia, USA. Available online: <http://mt-archive.info/AMTA-1994-Flanagan.pdf> (Accessed 10 June 2013).

Gaspari, F., A. Toral and S. Naskar. 2011. User-focused task-oriented MT evaluation for wikis: A case study. Proceedings of the Third Joint EM+/CNGL Workshop "Bringing MT to the User: Research Meets Translators". 14 October 2011, European Commission, Luxembourg.

Google. N.d.(a). Google Translate. Available online: http://translate.google.com/about/intl/en_ALL/ (Accessed 12 February 2014).

Google. N.d.(b). About Google. Available online: <http://www.google.com/intl/en/about/> (Accessed 12 February 2014).

Google Translate Blog. 2013. Google Translate – now in 80 languages. 10 December. Available online: <http://googletranslate.blogspot.it/> (Accessed 24 March 2014).

Hansen, G. 2010. Translation errors. In Y. Gambier and L. van Doorslaer (eds.) *Handbook of translation studies*. Available online: <http://benjamins.com/online/hts/> (Accessed 28 August 2011).

Hauksson, C.M. 2011. *Google Translate turns Lalalalala into Björk*. Available online: <http://www.optimizeyourweb.com/2011/01/10/google-translate-turns-lalalalala-into-bjork/> (Accessed 30 October 2013).

Helft, M. 2010. Google's computing power refines translation tool. *The New York Times*, 9 March. Available online: <http://www.nytimes.com/2010/03/09/technology/09translate.html> (Accessed 4 October 2013).

Hutchins, J. 2001. Machine translation and human translation: In competition or in complementation? *International Journal of Translation* 13(1-2): 5-20. Available online: <http://www.hutchinsweb.me.uk/IJT-2001.pdf> (Accessed 4 November 2013).

Izwaini, S. 2006. Problems of Arabic machine translation: Evaluation of three systems. Proceedings of the International Conference on the challenge of Arabic for NLP/MT, 23 October, the British Computer Society (BSC), London, UK. Available online: <http://mt-archive.info/BCS-2006-Izwaini.pdf> (Accessed 15 August 2013).

Koby, G. and G. Champe. 2013. Welcome to the real world: Professional-level translator certification. *Translation & Interpreting* 5(1): 156-172.

Lagoudaki, E. 2008. The value of machine translation for the professional translator. In *AMTA-2008: MT at work*. Proceedings of the Eighth Conference of the Association for Machine Translation in the Americas, Waikiki, Hawaii. Available online: <http://www.mt-archive.info/AMTA-2008-Lagoudaki.pdf> (Accessed 15 August 2013).

Papineni, K., S. Roukos, T. Ward and W-J. Zhu. 2001. BLEU: A method for automatic evaluation of machine translation. Technical report RC22176. Yorktown Heights, NY: IBM Research Division, Thomas J. Watson Research Center. Available online: <http://www1.cs.columbia.edu/nlp/sgd/bleu.pdf> (Accessed 11 August 2013).

Pym, A. 1992. Translation error analysis and the interface with language teaching. Available online: http://usuaris.tinet.cat/apym/on-line/training/1992_error.pdf (Accessed 7 August 2013).

Pym, A. 2010. Text and risk in translation. Available online: http://usuaris.tinet.cat/apym/on-line/translation/risk_analysis.pdf (Accessed 12 October 2013).

Quah, C. 2006. *Translation and technology*. New York: Palgrave Macmillan.

Sager, J. 1994. *Language engineering and translation: Consequences of automation*. Amsterdam: John Benjamins.

Schulz, T. 2013. Translate this: Google's quest to end the language barrier. *Der Spiegel*, 9 September. Available online: <http://www.spiegel.de/international/europe/google-translate-has-ambitious-goals-for-machine-translation-a-921646.html> (Accessed 25 September 2013).

Stellenbosch University. 2010. Language plan of Stellenbosch University. Available online: http://www.sun.ac.za/english/Documents/Language/english_language/Taalplan2010Eng.pdf#search=language%20plan (Accessed 28 October 2013).

Stymne, S. and L. Ahrenberg. 2012. On the practice of error analysis for machine translation evaluation. Proceedings of the Language Resources and Evaluation Conference, 21-27 May, Istanbul, Turkey. Available online: http://www.lrec-conf.org/proceedings/lrec2012/pdf/717_Paper.pdf (Accessed 15 April 2013).

Temizöz, O. 2012. Machine translation and postediting. European Society for Translation Studies Research Committee: State of the Art Research Reports. Available online: <http://www.est-translationstudies.org/intranet/research/MT.pdf> (Accessed 10 August 2013).

Valotkaite, J. and M. Asadullah. 2012. Error detection for post-editing rule-based machine translation. Proceedings of the Association for Machine Translation in the Americas 2012, Workshop on Post-editing Technology and Practice (WPTP 2012), 28 October-1 November, San Diego, USA. Available online: http://amta2012.amtaweb.org/AMTA2012files/html6/6_paper.pdf (Accessed 13 April 2013).

Van Rensburg, A., C. Snyman and S. Lotz. 2012. Applying Google Translate in a higher education environment: Translation products assessed. *Southern African Linguistics and Applied Language Studies* 30(4): 511-524.

Wittgenstein, L. 1922. *Tractatus logico-philosophicus*. London: Kegan Paul.

Appendix: Second assessment tool – Framework for error marking (Adapted ATA Framework for Standardized Error Marking)

Code	Error category	Criteria and description of each error category
1	Misunderstanding of original text	This category applies when the evaluator can see – usually by back-translating the target text – that the error arises from misreading a word, for example, or misinterpreting the syntax of a sentence. In other words, the result is wrong because the translation was based on a misunderstood source text.
2	Mistranslation into target language	The meaning of a segment of the source text is not conveyed properly in the target language. This category applies particularly when any other category in this framework relating to mistranslation would be too forgiving.
3	Addition or omission	Addition: Something is inserted that is not clearly expressed in the source text; when clarifying material is added. Omission: Elements essential to the meaning are left out.
4	Non-translation: Insertion of word from source language	The insertion of a source-language word in the target text when the translator cannot find an equivalent term in the target language.
5	Switched elements	Two adjacent elements have been switched around in the target text. The elements have been translated correctly, but appear to be switched around in comparison to the position of those elements in the source text.
6	Too freely translated	Translators are asked to translate the meaning and intent of the source text, not to rewrite or improve on it. The evaluator will carefully compare the target text to the source text. If a 'creative' rendition changes the meaning, an error will be marked. If recasting a sentence – i.e. altering the order of its major elements – destroys the flow, changes the emphasis, or obscures the author's intent, an error may be marked.
7	Literalness	A literalness error occurs when a translation that follows the source text word for word results in awkward, unidiomatic, or incorrect renditions.
8	Style	If the source text is characterised by a distinctive manner of expression – flowery, staccato, conversational, instructional – this should be reflected in the target text. Awkward or clumsy renditions that obscure the meaning may also be penalised.
9	Usage	Correct and idiomatic usage of the target language is expected. Errors include the use of the wrong preposition or misuse of a grammatical form. Examples: <i>take vs make</i> a walk, married <i>to</i> vs married <i>with</i> , etc.

10	Register	Language level and degree of formality should be preserved in the target text (<i>u/jy</i> in Afrikaans); examples of errors include making a legal document sound journalistic and using anachronisms or culturally inappropriate expressions.
11	Terminology, word choice	A terminology error occurs when a term specific to a special subject field is not used when the corresponding term is used in the source text. This category often involves terms used in various technical, legal, and financial contexts, where words often have very specific meanings. In more general texts, the candidate might not have selected the most appropriate word among several that have similar (but not identical) meanings.
12	Inconsistency, same term translated differently	In general, a term that is used consistently in the source text should be translated consistently into the target language. Conversely, if the source text uses different words for the same idea interchangeably, the translator should try to come up with a similar variety in the target language.
13	False friends	In some language pairs, this is the most common type of error. Examples: English–Afrikaans: <i>Indian</i> ('of India') translated as <i>Indiaan</i> ('of Latin America') instead of <i>Indiër</i> ; Dutch–Afrikaans: <i>aardig</i> ('friendly') translated as <i>aardig</i> ('strange'); <i>vaak</i> ('often') translated as <i>vaak</i> ('sleepy').
14	Ambiguity	If the meaning is clear in the source text but ambiguous in the target text, an error may be marked. The reader should not have to puzzle out the meaning.
15	Indecision – giving more than one option	Translators sometimes give more than one option for the translation of a word. Even if both options are correct, an error will be marked. The use of asterisks, footnotes, brackets, or other hedging devices are not acceptable. Clarifications are not acceptable unless readers from the target language will surely miss the meaning without them.
16	Grammar	A grammar error occurs when a sentence in the translation violates the grammatical rules of the target language. Grammatical errors include lack of agreement between subject and verb, incorrect verb forms, incorrect case of nouns, pronouns, or adjectives, and use of an adjective where an adverb is needed.
17	Syntax (phrase/clause/sentence structure)	The arrangement of words or other elements of a sentence should conform to the rules of the target language. Errors in this category include sentence fragments, improper modification, lack of parallelism, and unnatural word order. If incorrect syntax changes or obscures the meaning, the error is more serious.
18	Word form	The root of the word is correct, but the wrong form is used. Example in English: <i>The product has been tampered with and is no longer safety</i> . This category also includes incorrect plural or singular forms of words.

19	Spelling	A spelling error occurs when a word or character in the translation is spelled/used incorrectly according to target-language conventions. Spelling errors can cause confusion about the intended meaning (e.g. <i>principle/principal, systemic/systematic, peddle/pedal, dear/deer, sight/site</i>). Context is a factor as well. If a word has alternate acceptable spellings, the specific word should be spelled in the same way throughout the passage concerned.
20	Punctuation	The function of the punctuation in the source text should be reflected adequately in the target text. The conventions of the target language with regard to punctuation should be followed, including those governing the use of quotation marks, commas, semicolons, and colons. Incorrect or unclear paragraphing is counted as an error.
21	Accents and other diacritical marks	The conventions of the target language should be followed consistently. If incorrect or missing diacritical marks obscure the meaning, the error is more serious.
22	Capitalisation (upper/lower case)	The conventions of the target language (and, where applicable, of the target text itself as adopted from the source text) should be followed. Examples: There is Mr Lee and Mrs Johnson vs <i>Daar is mnr Lee en mev Johnson</i> .