Google Translation Service (GTS): A friend of a Foe of the Algerian EFL Learners?

MOSTARI Hind Amel
Université Djillali Liabés
Sidi Bel Abbés

Résumé

No one can deny the great importance of Google translation service translation (GTS) in daily communication, whether it is written or oral. As learners of a foreign language, the Algerian students use GTS for textual translation and even for listening to correct phonetic transcription of words. However, while using GTS, the following questions may arise:

How often do Algerian EFL learners use GTS?

What types of words /texts (literary, scientific, religious, idioms, proverbs) do they translate?

What are the common errors and disadvantages of GTS? .

1. Introduction

With the emergence of Internet and WebPages, scientists were eager to set up a new system of machine translator that might go beyond a mere textual translation. Thus, over the last decade, Google invested much research and development in a machine translation tool that does not follow the original model of context, grammar and word-based literal translation but instead uses a statistical approach.

On behalf of Google, the scientist Franz- Josef Och developed a concept that applies statistical analysis to a bilingual, parallel text body of at least of 1 million words in each language and two monolingual bodies of at least 1 billion translations for

new texts between these two languages without the need to understand the individual languages or their specific rules (Och, 2006).

In comparison with other MT applications, Google Translate indeed proved to be the most powerful and accurate of any of the readily available machine translation tools, thereby pushing MT into the public spotlight.

2. Machine Translation

Machine translation is not a recent addition to the vast array of ICT technologies and has been a subject of research for over a half of a century. It was instigated, as Berner (2003:5) reports, by the desire to "remove the language barriers that hinder scientific communication and international understanding". The challenge was and is still to produce translations as good as those produced by humans. By definition, "machine translation involves the use of computer programmes to translate texts from one natural language into another automatically" (Baker and Saldanha 2008:165). Similar to translation done by human translators, MT does not simply involve substituting words in one language for another, but applies complex linguistic knowledge to the text and/or selects the most probable words and sentence sequences out of huge corpora of already existing translations.

Classifying along the lines of MT systems' architecture, MT can be broadly categorized as *rule-based* or *corpus-based*. Rule-based MT systems, the best known among which is *Systran*, for the most part retrieve data from gigantic bilingual dictionaries and grammars which are then supplemented by sophisticated linguistic and other rules. Based on the identified word and sentence structure, the software first parses text and then creates a transitional representation, from which the text in the target language is generated. This process requires

extensive lexicons with morphological, syntactic, and semantic information, and large sets of rules[1]. This rule-based approach was prevalent to the end of the 1980s.

Its dominance was broken in the beginning of the 1990s by the emergence of the corpus-based machine translation technology (further divided into statistical and example-based MT), made possible by the development of computer technology (especially CPU and memory capacities).

Contemporary statistical MT systems rely on a large amount of human-engineered translations which are utilised to automatically infer a statistical model of translation**.** The underlying premise is that for every source language element there are a number of possible translations, and the most adequate translation is assigned the highest probability by the system (Veritas 2009). The statistical MT systems differ from example-based MT in the way that they are (or rather were) based primarily on word frequency and word combinations, whereas example-based relies on the extraction and combination of phrases (or other short segments of texts) (Hutchins 2005). There are several freely available web-based MT systems, including the following:

- Google Translate (http://translate.google.com/)
- SDL Automated Translation Solutions
- (http://www.freetranslation.com/)
- Bing_Translator (http://www.microsofttranslator.com/)
- Yahoo! Babel Fish (http://babelfish.yahoo.com/)

3. Google Translation Service (GTS)

Using software originally developed in the 1980s by researchers at IBM, Google has created an automatic translation tool that is unlike all others. It is not based on the

---

[1] http://www.systran.co.uk

intellectual presuppositions of early machine translation efforts – it isn't an algorithm designed only to extract the meaning of an expression from its syntax and vocabulary. In fact, at bottom, it doesn't deal with meaning at all. Instead of taking a linguistic expression as something that requires decoding, Google Translate (GT) takes it as something that has probably been said before. It uses vast computing power to scour the internet in the blink of an eye, looking for the expression in some text that exists alongside its paired translation.

The corpus it can scan includes all the paper put out since 1957 by the EU in two dozen languages, everything the UN and its agencies have ever done in writing in six official languages, and huge amounts of other material, from the records of international tribunals to company reports and all the articles and books in bilingual form that have been put up on the web by individuals, libraries, booksellers, authors and academic departments. Drawing on the already established patterns of matches between these millions of paired documents, Google Translate uses statistical methods to pick out the most probable acceptable version of what's been submitted to it. Much of the time, it works. It's quite stunning.

And it is largely responsible for the new mood of optimism about the prospects for "fully automated high-quality machine translation". Google Translate could not work without a very large pre-existing corpus of translations. It is built upon the millions of hours of labour of human translators who produced the texts that GT scours. Google's own promotional video doesn't dwell on this at all. At present it offers twoway translation between 58 languages; that is 3,306 separate translation services, more than have ever existed in all human history to date.

Virtually, many online translation services emerged from 1980's in order to facilitate the internet users of several

websites. One of the most powerful machine came in 2006 when Google launched its pioneering Google translate which offers translation services in and out of more than 70 languages, ranging from translating individual words, sentences to whole   documents and websites .

The purpose is therefore to fulfil the needs of its users with great efficiency and popularity. Actually, GTS allows translating whole documents, for example, in the form of PDF, TXT, DOC, PPT, XLS or RTF, by just clicking the "translate a document" link and submitting a file without the need for copying and pasting large blocks of text .

4. GTS New Approach to Machine Translation

Over the last decade, Google invested much research and development in a machine translation tool that does not follow the original model of context, grammar and word-based literal translation but instead uses a statistical approach. On behalf of Google, the scientist Franz-Josef Och developed a concept that applies statistical analysis to a bilingual, parallel text body of at least of 1 million words in each language and two monolingual bodies of at least 1 billion words.

Statistical models and algorithms that are derived from that data are then used to compile translations for new texts between these two languages without the need to understand the individual languages or their specific rules (Och, 2006). In comparison with other MT applications, Google Translate indeed proved to be the most powerful and accurate of any of the readily available machine translation tools, thereby pushing GTS translation into the public spotlight.

Google began to use a quite different translation machine called Statistical Machine Translation (SMT) approach. This starts by examining and comparing massive corpora of texts on the Web that have already been translated by human beings. It

looks for matches between source and target texts and works out which translations are likely to be the most accurate. As more and more corpora are added to the Web this means that Google Translate will keep improving until it reaches a point where it will be very difficult to tell that a machine has done the translation.

Besides using a much more sophisticated and accurate translation engine, GTS also offers the possibility of interaction. When the translated text appears, the internet user can hover his/her mouse over the text and ask Google Translate to suggest alternative renderings if the user doesn't accept what it offers as the first choice. These may be different vocabulary items, different tenses, different case endings, etc. S/he can also rearrange the word order. Thus s/he can edit the text until you are satisfied with it – and then s/he can copy and paste the text into *Microsoft Word* and edit it further using the inbuilt foreign-language spell checkers, grammar checkers and thesauruses.

This machine translation process is based on following steps:

1- Analyze – source language text based on vocabulary, morphological and syntactical analysis
2. Conversion (translation of source text to target text) and
3. Synthesis – creation of text for target language based on syntactical and morphological appearance of text.
All these steps in machine translation system may be interrelated closely and/or may be absent. The machine translation algorithm is usually performed by computer using application software. The text translated, by the machine translation may be then edited by users, to avoid ambiguities and mistakes.

5. Research questions

As learners of a foreign language, the Algerian students use GTS for textual translation and even for listening to correct phonetic transcription of words. However, while using GTS, the following questions may arise:

- How often do Algerian EFL learners use GTS?
- What types of words /texts (literary, scientific, religious, idioms, proverbs) do they translate?
- What are the common errors and disadvantages of GTS?

For this purpose, a preliminary investigation was led in the department of English – Faculty of Letters, languages and Arts – Djillali Liabés University of Sidi Bel Abbés , on a sample 30 students from 3rd year licence and Master 1 students of linguistics and language dynamics . The outcomes of such investigation will serve answer the above questions.

6. Data collection

Thirty (30) EFL Students (3rd year and Master 1 Linguistics) have been asked to respond the following questions:

Do you use Google Translation Service?	Yes 81%	No 9%			
How often	Always 25%	Often 35%	Rarely 22.50 %	Never 17.5%	
For which Purpose Do you use GTS ?	Translation of word 80%	Idiom 2%	Sentence 15.50%	page (web) 32.5%	Pronunciation 39%
How do you find GTS?	Effective 72%	Ineffective 28%			
Which problems have you encountered while using GTS?	Missing words 46 %	Word order 35%	Incorrect words 51%	Unknown words 61%	Punctuation errors 15%
What is the language pair used in translation	English – Arabic	Arabic– English	English – French	French – English	

107

7. Data analysis

Following the above table's findings, about 35% of EFL students use GTS very often in order to translate basically words (81%), sentences and checking their pronunciation. The fact that GTS does not take into account the cultural context in which the words have been used, may have led to wrong translation like in the case of idioms (2%). Besides, our respondents state they have faced some semantic problems, in addition to some types of lexical or syntactic ambiguities, missing words (46%), word order (35%), incorrect words (51%), unknown words (61%) and punctuation errors (15%)

Because of their bilingual background and due to 12 years of arabized education, EFL students have tendency to translate more from Arabic to English and vice versa.

Let us use GTS for the English translation of this idiom:

'He is an ache in the neck' (an American idiom meaning a difficult person) , GTS gives us the following translation from English to French : '*Il est une douleur dans le cou*' (literary meaning he is a pain in the neck) which has no sense in English .

From Arabic to English,

إنفطر قلبي meaning in Arabic: my heart is broken ,
GTS gives the following translation in French: mon cœur
anaftr: my heart + anaftr (unknown word)

There is no doubt GTS online translators can give some idea about the meaning of the words and sentences, but their various flaws prove that they are far from replacing professional translators. As shown above, GTS creates a false sense of security, because it may not detect or accurately convey linguistic and certainly not cultural nuances. It can also

misinterpret grammatically complex structures and thus provides translations that may not be accurate and precise enough for the task at hand while the user may not be aware of any potential errors and inadequacies. Needless to mention that the EFL students use other online translators such as PROMT, REVERSO and COLLINS, among others.

8. Merits and demerits of GTS

When time is a crucial factor, machine translation can save the day. You don't have to spend hours poring over dictionaries to translate the words. Instead, the software can translate the content quickly and provide a quality output to the user in no time at all. The next benefit of <u>machine translation</u> is that it is comparatively cheap. Initially, it might look like a unnecessary investment but in the long run it is a very small cost considering the return it provides. This is because if you use the expertise of a professional translator, he will charge you on a per page basis which is going to be extremely costly while this will be cheap.

Confidentiality is another matter which makes machine translation favourable. Giving sensitive data to a translator might be risky while with machine translation your information is protected..A machine translator usually translates text which is in any language so there is no such major concern while a professional translator specializes in one particular field. Today, Google Translate is available as a free online application and can be run on top of third party websites to offer visitors an automated translation of the content in any of the available languages. The tool is fast and easy to use and it claims to provide adequate general content translation for over 70 languages. For informational purposes, it provides the user with a powerful tool to gather data and find information on sites that were previously inaccessible due to the language barrier. Examples where Google Translate proofs to be almost

99% accurate involve short Subject-Verb-Object sentences of an unambiguous nature, such as driving directions, simple user instructions or registration forms.

Nevertheless, since the tool is free and can be used by anyone with fast-speed access to the internet, the use of the system is not based on any user agreement between the user and Google and there is therefore no consumer protection or any type of liability obligation in place on Google's side. Any use is therefore at the sole risk of the user and the accuracy of the translations and therefore any potential damage due to such an inaccuracy is also at the sole risk of the user. Also, Accuracy is not offered by the machine translation on a consistent basis.

You can get the gist of the draft or documents but machine translation only does word to word translation without comprehending the information which might have to be corrected manually later on. Systematic and formal rules are followed by machine translation so it cannot concentrate on a context and solve ambiguity and neither makes use of experience or mental outlook like a human translator can.

Among other disadvantages of *GT* and statistical machine translation systems in general, authors also emphasise the users' complete lack of control over the translation input (the controversial issue of confidentiality of data fed into the system) and output, often caused by the generic, surrogate, unverified and often unhelpful data used to train machine translation systems (resulting in unpredictable and inconsistent translations); poor training of the models on general language corpora, meaning text other than the specified domain; significant hardware and software infrastructure which is needed to build, manage and maintain large translation models, and last but not least, human assistance before, during and particularly after the machine translation process.

On the other hand, advantages put forward by several authors include improved quality due to greater terminological and phraseological consistency, enhanced productivity and speed of translations, cost reduction, lack of bias and general availability of the systems. Both lists are by no means exhaustive.

9. Some recommendations for correct translation

Computer code as a universal language may at one point be able to transcend human language structure and software programs may become intelligent enough to truly understand language. However, humans are somewhat illogical beings and language is an adaptable, ever-changing, living concept that reflects the human psyche which may never be entirely captured in its essence by a machine.

Consequently, authors are suggesting few recommendations for users and for these service providers to improve these translation issues or to avoid any confusion if any occur.

a. User should not blindly trust these translations and should use their common sense, which may resolve translation ambiguities sometimes.
b. For religious/ cultural/ social or/and humorous works translation, use authentic human-translated Website, rather than automated Websites.
c. These translation services may also refer authentic translations of specific texts and improve their database to adopt those words, which are more appropriate and acceptable in such translations
d. These translation services may adopt different languages vowels to avoid any cultural ambiguities
e. These translation services should focus on grammatical structure of several languages, otherwise they will be treated like dictionaries

g. These translation services also need to cooperate with linguistic professional of all languages to cope .

Conclusion

Computers and powerful software programs such as GTS will play an ever increasing role in inter-language communication and will become more effective in bridging language gaps. Going forward, it will remain important to be cognoscente of the limitations GTS.

The present research conducted on EFL students genuinely reflects the use of GTS as an important tool for learning a foreign language equally, among licence and master students. It is clear that the cultural and social gap that exists between languages may enhance students and GTS users to limit their translations to word and pronunciation checking.

Certainly, foreign language translations do not run the risk of turning into a commodity and the translation profession is still living its golden age.

References

Ahmed Abdel Azim ELSHIEKH, 2012. *Google translate service: transfer of meaning, distortion or simply a new creation? an investigation into the translation process & problems at google, english language and literature studies*, Vol. 2, No. 1, March 2012.
BAKER, M., & SALDANHA, G. 2008. *Routledge encyclopedia of translation studies*. London and New York: Routledge.
BELL, Roger, 1992, *Translation and translating*. Longman Group Ltd, London, UK
BERNER, S. 2003: *Lost in translation: cross-lingual communication, and virtual academic communities*, Proceedings of 5th Annual Conference on World Wide Web Applications, Durban, South Africa, 2003 [accessed on 1 September 2011]. Available from http://general.rau.ac.za/infosci/www2003/Papers/Berner,%20S%20Lost%20in%20Translation.pdf
HATIM, B., & MASON, J, 2003, 4th edition, *Discourse and the translator*, Longman, Singapore, "Hello, World"

Wired,http://www.bbc.co.uk/religion/religions/islam/texts/quran_1.shtml, Retrieved December 2012,

HUTCHINS, J. (2003, March). *Machine translation and computer-based translation tools: what's available and how it's used*. Presentation at the University of Valladolid (Spain). Retrieved from
http://hutchinsweb.me.uk/Valladolid-2004.pdf

HUTCHINS, J. 2005. *Recent advances in example-based machine translation*. Edited by Michael Carl and Andy Way. Dordrecht: Kluwer Academic Publishers, 482pp. (Text, Speech and Language Technology, vol. 21)

OCH, F. J. 2006. *The alignment template approach to statistical machine translation. mit press journals*, *30*(4), pp. 417-449.

ULITKIN, I. 2011. *Computer-assisted translation tools: a brief review. Translation Journal*, *15*(1). From
http://translationjournal.net/journal//55computers.htm

VERITAS. 2009. *Statistical machine translation and example-based machine translation* [accessed on 1 September 2011].
http://www.proz.com/translation-articles/articles/2483/1/Statistical-Machine-Translation-and-Example-based-machine-Translation.