Machine versus Human:

Will Google Translate Replace Professional Translators?

Friderike Butler

IT 103-055

George Mason University

Professor Lord

February 20, 2011

"By placing this statement on my webpage, I certify that I have read and understand the GMU Honor Code on http://academicintegrity.gmu.edu/honorcode/ . I am fully aware of the following sections of the Honor Code: Extent of the Honor Code, Responsibility of the Student and Penalty. In addition, I have received permission from the copyright holder for any copyrighted material that is displayed on my site. This includes quoting extensive amounts of text, any material copied directly from a web page and graphics/pictures that are copyrighted. This project or subject material has not been used in another class by me or any other student. Finally, I certify that this site is not for commercial purposes, which is a violation of the George Mason Responsible Use of Computing (RUC) Policy posted on http://universitypolicy.gmu.edu/1301gen.html web site."

Introduction

Language barriers have hindered and complicated communication between humans since people began to express themselves through distinguishable words with specific meanings. While an exact definition of differentiating linguistic factors can be disputed, it is safe to say that 5000 to 6000 different languages are spoken in the world today. Human interpreters have always been a critical link in the communication process between people who spoke different languages. The birth of the written word created the need for text-based translation and laid the foundation for the translation profession which many linguists see as a form of art rather than a process of simply exchanging the words of one language with the words of another (Burwick, 2007).

As the world became more and connected and an ever increasing stream of written and verbal communication in different languages demanded to be considered in politics, business, literary circles, and everyday life, the need for foreign language skills increased and beckoned for tools to facilitate and speed up the translation process. People who did not speak a particular foreign language were looking for ways to overcome language barriers without having to take the time to actually learn a foreign language. Specialized dictionaries and phrase books were compiled and published in print and later in digitalized form. As powerful software programs were developed to mine vast amounts of data in order to locate specific information, the computerization of the translation process was a logical application of the ability to search for, harvest, and process data. Many attempts have been undertaken to make the Star Trek vision of a Universal Translator a reality and many software companies have claimed that they have found just the right tool to rebuild the Tower of Babel to new heights. How far have those efforts really come and will it one day become obsolete for a person to know a foreign language by heart and to translate between different languages?

The emergence of computer-aided tools in the linguistic sector

In 1948, Dr. Andrew D. Booth, Birkbeck College, London, and his assistant Dr. Richens were among the first to work on a design for a computer-searchable dictionary. They focused on devising a computerized word-by-word translation tool and greatly wrestled with the problem of words carrying multiple meanings or changing their meaning depending on prepositions or as they were used as compound words. It was quickly evident that such an approach would not provide an acceptable solution for the translation of a narrative text. Inspired by their work and his experiences as a cryptography specialist in World War II, the mathematician Dr. Warren Weaver wrote an influential paper in 1949 proposing computerized context-based translation based on logic. His work spurred a magnitude of further research and development focusing on the issue of intelligent translation (Weaver, 1955, p. 19-20).

After much excitement in the industry and press, a committee of the major sponsors of the MT (machine translation) project issued the disappointing verdict in 1966 that the quality of the produced outputs was simply too poor and that future efforts should rather concentrate on the development of computer-aided tools to support human translation with the help of customizable databases and reusable translation segments (Hutchins, 2003).

IBM and other software companies indeed focused on the development of computer-aided translation tools and in the 90s the first commercially available software programs hit the market. The programs essentially merged word processing tools with powerful databases which allow translators to search electronic dictionaries, apply translation phrases and save new translation segments for future use. Additional functions and capabilities such as multi-lingual spell-checking or format copying were integrated as personal computers and word-processing

software further developed, thus providing human translators with powerful tools to provide faster and higher-quality output (Ulitkin, 2011).

Google's 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 SMT translation into the public spotlight.

Benefits and disadvantages of machine translations through Google Translate

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 50 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. Due to its limitations, however, it 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 potential inadequacies. Here is a brief example of the potential loss of information in the SMT process when translating from a source language (English) to a target language (German) back to the original source language:

English: A new assignment was given to the class to highlight the advantages of an intelligent application built with state-of-the-art technology.

German: *Eine neue Aufgabe war es, die Klasse gegeben, die Vorteile einer intelligenten Anwendung gebaut Highlight state-of-the-art Technologie.*

English: A new task, the class where the benefits of an intelligent application was built highlight state-of-the-art technology.

It is clear that the back-translation is not intelligible by itself. For someone who does not speak German it would be impossible to detect that the translation into German already contains major flaws which are then exacerbated in the back-translation into English. If the proper German translation is run through Google Translate, the results are not grammatically correct either, but at least the intent of the message can be adequately guessed:

German: *Der Klasse wurde eine neue Aufgabe gestellt, um die Vorzüge einer intelligenten mit hoch-moderner Technologie gebauten Anwendung herauszustreichen.*

English: The class was asked a new task to emphasize the benefits of an intelligent high-technology modern-built application.

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. The SMT translation of scientific texts written in simple sentences also provides satisfying results, and for a mere informational perspective Google

Translate provides intelligible translations for websites that at least allow the visitor to ascertain the general nature the site and the gist of its pages.

Legal and ethical considerations

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. It is clear that this application is not intended to be used commercially. If a translation deals with confidential subject matters or requires the highest possible degree of accuracy, Google Translation will not be able to replace the human translator. Professional translators may even argue that the editing of a complex translation that was run through an SMT application can take longer than if the source text was translated with the help of powerful, state-of-the art computer-aided translation tools.

Another area where the application's usefulness is somewhat questionable is in the social media arena. Especially short-lived colloquialisms, new word creations, sudden shifts in meaning or humorous plays on words are unlikely to be translated adequately and the composition of texts for social media communication into other, unfamiliar languages can therefore bear dangerous pitfalls. Legal drafts, safety applications, medical equipment, political texts, negotiations, etc. are likewise ill-suited for unchecked machine translations and should not be used without a quality check by an editor who is well-versed in the respective language pair.

Social and linguistic concerns

Highly publicized MT or SMT translations ultimately promise more than they can hold and often produce outputs of mediocre quality. However, the underlying message that is communicated to the public is that foreign language skills are overrated because a computer can do the job much faster and that learning a foreign language is antiquated. Foreign language translations run the risk of turning into a commodity and the translation profession is being devalued because it seems that translation technology is at the brink of making the human translator obsolete as it reduces the human involvement in the translation process to a mere editing or proofreading job.

Furthermore, if MT or SMT translation tools are to be used successfully, the input needs to be simplified to a level where language losses its luster and beauty and the message itself turns into an unmemorable static expression, dummied down to accommodate the limitations of machine language converters: Enter the text. Push the button. Read the result. Ultimately, the statistical machine translation thereby is at risk to promote a loss of culture since it doesn't truly understand the language and the tradition of a language anymore but simply draws from a pool of words and samples if used as a stand-alone process.

Impact on the translation industry

For obvious reasons, the translation industry has been rejecting MT and also SMT translations for the longest time, however, as the technology further improves, it is also clear that the it is here to stay and that pointing out the flaws will also ultimately lead to improvements. In a shift of position, the translation industry as represented through the American Translators Association has indicated that the wide-spread availability and use of MT translations calls for a standardization of the human-machine-human translation process. In a speech during the 2010 American Machine Translation Association Conference, the president of ATA, Nicholas

Hartmann, indicated the need for a standardization of human post-editing processes and MT integration into standardized computer-aided translation tools (Grunwald, 2010).

The future of machine and human 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. Computers and powerful software programs will play an every 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 of software programs and machines, work around the flaws, embrace the benefits and add the human touch where necessary, so that one is not at risk to "step under floating burdens" which is Google translate for "stand under suspended loads" (Kakase, 2011, p. 2). You never know when the tower may be tumbling down and then, everyone will benefit from speaking more than just one's own language.

Bibliography

Burwick, F. (2007). Coleridge's art of translation. *Wordsworth Circle, 38*(3), 108-105.

This article explores the works of English romantic poet Samuel Taylor Coleridge as a translator and commentator of German literature and philosophy. The application of his rules and principles is further discussed in light of translation as a creative process and a craft as it requires the translator to intimately understand the source text and to wield the ability to recreate the same meaning, emotion, and message in the target text.

Grunwald, D. (2010, November 1). GTS Blog. *Industry embraces machine translation at AMTA-2010 conference | GTS Blog*. Retrieved March 2, 2011, from http://blog.gts-translation.com/2010/11/01/industry-embraces-machine-translation-at-amta-2010/

The author provides a synopsis of the conference agenda, individual topics and gives an overview of the development within the machine translation industry.

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

The presentation gives an overview of machine translation and translation tools and their different applications as well as weaknesses and strengths.

Kakase, K. (2011, February 21). Google, Yahoo! BabelFish use math principles to translate documents online. *Science News*. Online newspaper, . Retrieved March 2, 2011, from http://www.washingtonpost.com/wp-dyn/content/article/2011/02/21/AR2011022102191.html?referrer=emailarticle

The article discusses a new machine translation approach similar to or following Google's statistical machine translation approach. It also includes anecdotal use of Google

Translate.

Och, F. J. (2006). The alignment template approach to statistical machine translation. *MIT Press Journals*, *30*(4), 417-449.

The article describes the mechanics and the logic of the phrase-based statistical machine translation concept which incorporates a multi-lateral approach to the use of words. This model differs greatly from logic and rule-based models which require the entire mapping of a language. Here, the language is analyzed and seen in the context of a linguistic body of 1 billion words in a source-channel approach.

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

The article provides a brief historical and technical overview of a range of computer-aided translation tools. It explains the need for such tools as well as the concept and application.

Weaver, W. (1955). Translation. In W. Locke & A. Booth (Eds.), *Machine translation of languages: fourteen essays* (pp. 15-23). Cambridge, MA: Technology Press of M.I.T.

In July of 1949, Warren Weaver presented this paper proposing the use of computers for translation. Interesting read to see how a non-linguist approaches the problem of linguistic ambiguity.

Online tool used for this paper:

Google Translate©2010 http://translate.google.com accessed on March 1, 2011