

Machine translation status and its effect on business

Mohammed M. Sakre

Dr. Mohammed Mahmoud Sakre
Professor of Information Systems,
High Institute of Computers & Information Technology,
Shorouk Academy,
Shorouk City, Cairo, EGYPT.
E-mail: dr.mohamed.sakr@sha.edu.eg

Abstract

The globe has become like a town which is inhabited by people speaking more than two thousands languages. The role of Machine Translation (MT) became essential. For global corporations, MT is providing a viable answer to their needs by supplying better control over translation volume, speed and brand fidelity. This article provides an overview for the state of the art of MT systems, the languages covered by MT systems and the used technologies. Furthermore, Pros and Cons of using MT in business and the current challenges are explained. The main MT systems involving Arabic language are presented and the need for post-editing is recognized. MT evaluation methods, both human based or Automated, are discussed. Speech-to-Speech translation technology appears to have finally become a reality; this article outlines the current state, challenges and opportunities of speech translation.

Keywords: Machine translation, Speech translation, Business of translation, Artificial Intelligence, Arabic Machine translation.

Machine Translation Status and its Effect on Business

For many centuries ago, scientists were trying to find out one common language to be spoken and understood by everyone. They were hoping that a common language could resolve global problems that produce conflicts. Esperanto, which achieved some success, was proposed. Today, English is considered as the universal language and it is the mostly used language over the internet. There are hundreds of other languages in use over the internet. (*Internet World Stats*, 2017) presents an estimation for the most common languages used on the internet in June 2017 for a total number of 3.885 billion users as shown in figure A1 (in Appendix A). Due to the recent advances in the area of computer technology, both hardware and software, MT is recognized by the majority of the internet users. MT can be defined as the use of intelligent software in translating large amount of data between different languages.

Today, machine translation systems are becoming a fact of business life. Rough translation, For consumers and small business, may be good enough while for large global enterprises whose success depends on communicating with millions of customers, employees and partners in their native language more accurate translation is required.

Machine Translation engines translate to and from English for most of the major languages such as French, German, Arabic, Portuguese, Spanish, Greek, Italian, Swedish, Norwegian, Danish, Finnish, Hungarian, Romanian, Russian, Hebrew, Chinese, Japanese, Korean, and more. Increasingly, Machine Translation engines are extending their capabilities to include new mixed pairs such as Russian–Chinese.

Machine Translation systems are employed by global companies that need to translate high volumes of business content such as sales guarantees, user manuals or websites between many languages. Companies undertaking a localization process may use Machine Translation of content to be effectively localized into multiple languages. Companies managing customer support through online channels such as chat and popular social media sites use Machine Translation to offer a more personalized customer understanding.

Machine Translation is also used by Language Service Providers as part of their integrated translation solution for global businesses.

Categories of MT in business

(Safaba Translation Solutions, 2017) roughly classified the MT systems into three main categories as following:

- Generic Machine Translation systems that describe a general solution and refer to engines such as Google-type engines that translate written text from one language to another. Used by users for informal translations of short texts, “Generic MT is exposed for grammar and syntax mistakes and is less accurate than “customized” machine translation.
- Customized Machine Translation systems that involve “training” or adaptation of the translation software to recognize language belonging to a specific domain, industry or organization. These systems offer business a higher level of accuracy with high volume.
- Enterprise Machine Translation systems represent the next generation of machine translation engines that employ sophisticated technology to reproduce style, format and terminology more faithfully than other solutions. “Enterprise” Machine Translation answers the need of today’s global business for high volume and high velocity content localization and real-time multi-lingual communication.

Technical Approaches of MT

The present advances in MT have been gradually achieved during the last seven decades. It started by word to word translation between English and Russian during the cold war in fifties of the last century. The science of Computational Linguistics appeared and attention was given to syntax for about three decades to improve the translation. Artificial intelligence techniques and Knowledge representation methods emerged to resolve the semantics problems in the area of computational linguistics. In the late nineties, Statistical Machine Translation appeared to get better translation.

Now neural networks, which convey the concept of machine learning, is considered to be the recent technology in the area. During the evolving process of MT mixes of the previous technologies are used to enhance the translation accuracy. In the following lines brief descriptions, for MT approaches, are presented.

Rule-Based Machine Translation systems (RBMT):

RBMT includes as much information as it is possible about the linguistics of the source and target languages. The morphological and syntactic rules in accompany of the semantic analysis of both languages are used. The basic approach involves the use of a parser and an analyzer for the source language, a generator for the target language, and a bilingual lexicon for the actual translation. RBMT's main problems are those of ambiguity, idiomatic expressions. Ambiguity can be on any of the three levels: lexical; syntax or semantic. Fairly good automated translations, with predictable results, are delivered with the support of human factor. However, due to significant manual labor, rule-based systems can be quite costly, time consuming to implement and maintain. This approach is known as Transfer-based machine translation as well (*Jurafsky, D.; Martin, S. H. , 2009*).

Example-Based MT (EBMT):

EBMT introduced the idea of translation by analogy which represents the rejection of the idea that people translate sentences by doing deep linguistic analysis. It is based on the concept that people translate by first decomposing a sentence into phrases, then translating these phrases into the target language, and finally composing these fragments into one sentence. This concept of translation is encoded to present the EBMT systems through the examples of translation used for training (Makoto Nagao, 1984), (Carl, M.; Way, A., 2003)

More recent approaches to machine translation, including statistical machine translation, also use bilingual parallel corpora to learn the process of translation.

Statistical Machine Translation (SMT):

SMT is related to other data-driven methods in machine translation, such as the earlier work on EBMT. Contrast this to systems that are based

on hand-crafted rules (*Philip, K., 2009*). SMT uses computer algorithms to extract a translation that looks best statistically from millions of sentences. Statistical models consist of words and phrases learned automatically from bilingual parallel sentences.

Machine learning methodologies are used in building the SMT systems, leading to rapid turnaround time and low cost of processing power required for constructing and operating these statistical models. However, the level of translation is enhanced but still far from the human translation. Syntactic information has helped significantly to improve SMT. However; the use of syntactic information may have a negative impact on the speed of translation because of the large number of rules, especially when syntax labels are projected from a parser in syntax-augmented machine translation (Hideya M., 2014).

Neural Machine Translation (NMT)

Artificial neural networks (ANNs) are computing systems imitate the biological neural networks that constitute animal brains. These systems learn to do tasks by considering examples and gradually improve performance. Neural network software applies ANNs together with a deep learning based approach to simulate and develop some artificial intelligence applications like natural language processing and machine translation. Neural Machine Translation (NMT) has made rapid progress in recent years. Google has announced its translation services are now using this technology in preference to its previous statistical methods (Albat; Thomas Fritz, 2017). Also other MT systems providers like Kantan MT (Slator, 2017), Omniscien Technologies (Slator, 2018) and SDL (Rowe, S., 2017) have announced the deployment of neural machine translation technology in 2017.

The effect of MT on the Global Business

Most Machine Translation engines can be integrated into a translation workflow with minimal engineering effort. Regardless of the delivery model (SaaS vs. Onsite), translation management systems call on Machine Translation to translate a segment of text. The text is converted into a format that Machine Translation can read and comes back translated in a format that the workflow can continue to process. The same applies for real-time

communications (chat), customer engagement (user reviews) and other applications requiring automated translation.

Machine Translation offers global enterprises significant translation efficiencies, especially when translation volumes are high and content is variable (Safaba Translation Solutions, 2017):

- It speeds up time-to-market by greatly reducing the length of the translation cycle.
- It increases knowledge sharing internally by making it easier for staff in local markets to access company documents (such as sales presentations) in their native language.
- It enhances customer experience and staff effectiveness alike by making ‘real time and on-demand translation’ a regular part of their business process.

Machine translation Pros and cons

Machine Translation has the advantages of speed of translation and lower cost. The dark side of it is that the translation can be inaccurate, incomprehensible or possibly dangerous. Generally speaking the user satisfaction depends on the purpose of the required translation.

The advantages of MT can be concluded in the following points:

- Availability of many free systems on the internet (e.g. Google Translate, etc.)
- Quick response time.
- The ability to translate between multiple languages using one system.

The disadvantages of MT can be concluded in the following points:

- The translation accuracy can be low.
- Varying accuracies across different languages when using the same multilingual MTS.
- Machines can’t translate context.

- The most important issue to consider with any kind of translation is the cost of mistakes in translation. For example translating instructions for medical devices and legal documents require full accuracy. In such examples, mistakes can cost lives, money and lasting damage to enterprise's image.

So, human translation is the best option when accuracy is important. MT is used when a large bulk of content is to be translated and the general meaning is enough. It can be concluded that both machine and human translation deserve their place in the industry, even if they play very different roles. Choosing the right service is the main issue. Also, it is important to remember that technology advances every year and the quality of translation it offers is gradually growing.

Speech to speech translation

Figure A2 (in Appendix A) shows that speech translation system would typically integrate the following three software technologies: Automatic Speech Recognition (ASR), machine translation (MT) and voice synthesis (Text To Speech). Currently free speech translation technology is available as products to translate for multi-lingual conversations. Speaker pronunciation is an important issue to deal with during speech recognition module execution. This module must recognize acoustic noise forms as well (Satoshi, N., 2009).

MT systems working in business

In this section the most well-known translation applications and their main features are presented.

Google Translate:

Google Translate supports more languages than other competitors, and its comprehensive feature set makes it especially well-suited for travelers.

Supported Languages: Google Translate offers varying degrees of support for 103 languages ("Google Translate", 2017):

- Type to translate: 103 languages
- Offline support: 52 languages

Machine translation status and its effect on business

- Real Time Video translation: 30 languages
- Camera Mode: 37 Languages
- Speech-to-speech translation: 32 languages
- Handwriting translation: 93 languages

Cool Tricks: It translates signs, menus and other written content using your phone's camera. It offers offline support for some languages and integration with the Android operating system for translating text messages and websites.

Adding Neural Machine Translations (NMT) to Google translate improved accuracy of translation to many languages.

Microsoft Translator:

Microsoft's translator represent a worthy competitor with unique strengths of its own to Google translate.

Cool Tricks: Microsoft Translator has free applications for Windows, iOs, and Android. The application can translate speech, text, and images.

Microsoft Translator lacks Google Translator's real time video translation feature. However, Microsoft wins a major point over Google with the superior design of its real-time conversation mode. This feature makes it easier to have natural conversations with the people you meet on your travels.

Supported Languages: Microsoft Translator supports only 60 languages with different features for different languages. For example, speech recognition and conversation mode are only available for 10 languages ("Microsoft Translator", 2017). Skype, which has been purchased by Microsoft in 2011, has unveiled a new feature for Windows users – a real time translator for instant messages. Users can translate voice-to-voice in seven languages and text-to-text ("Skype ", 2017).

The Translation Application: iTranslate Voice

iTranslate-Voice provides text-to-speech and voice-to-voice translation on both iOs and Android smartphone devices.

Cool Tricks: Translates what you say right after you say it. Some offline capabilities are available for specific languages. Save your most commonly used phrases in a personalized “phrasebook” for easy access.

Supported Languages: iTranslate Voice supports 44 languages and dialects, but not all to the same degree ("iTranslate Voice ", 2017).

A reviewer compared this translator and Google translator, during a volunteer stint at a refugee camp. In his experience, iTranslate had better voice input and output.

The Translation Application: SayHi

SayHi is a “universal translator” for the iPhone and Kindle, offering speech-to-speech translation in 90 languages and dialects.

Cool tricks: You speak, SayHi translates. The app claims 95% accuracy for voice recognition. According to LapTop Magazine (Alba, D. ,2012).

Supported Languages: SayHi features 90 languages and dialects, including an impressive list of Arabic dialects (SayHi, 2017).

There are many other Speech to Speech translation applications. Few of these systems are mentioned briefly in the following lines:

WayGo instantly translates Chinese, Japanese and Korean characters with no data connection required.

TripLingo can connect you to a real live human translator (for an additional fee, of course).

Speak & Translate focuses on speech-to-speech translations. It’s thoroughly integrated with iCloud, so the user can view his translation history ("Speak & Translate", 2017).

Papago is a translation application that focuses exclusively on translating the “big 3” Asian languages (Korean, Japanese and Mandarin) to and from English.

ABBYY Textgrabber is an Optical Character Recognition application that instantly captures text from live video stream or photo and offers translations ("ABBYY Textgrabber ", 2017).

MT systems involving Arabic language

Many MT systems are involving Arabic language in both directions of translation. Some of these MT systems are mentioned with the used technologies in appendix B. These MT applications are translating not only to and from Arabic language but also to and from many other languages. A portion of the text of this article is used to examine the output of translation into Arabic language of the mentioned above systems. The Arabic language output of the Google translator for English language text is shown in figure A3 (in Appendix A), It is clear from the English-Arabic Google's translation that post-editing is needed. Post-editing is rapidly becoming a popular solution for the translated content. Language service providers experience considerable pressure to adopt post-editing for internal efficiency or to respond to client demands.

Translation evaluation

There are various means for evaluating the output quality of machine translation systems. The oldest is the use of human translators to assess a translation's quality. Even though human evaluation is time-consuming, it is still the most reliable method to compare different systems such as rule-based and statistical systems (Anderson, 1995).. Automated means of evaluation include BLEU, NIST, METEOR (Han, 2012).

Since pure human translations may contain errors it must be reviewed to get 100% correct translation. Also, MT output should be reviewed by a human to get a free of error translation. During this revision, post-editing may take place. Pre-editing, also, is necessary in order to provide input for machine-translation software such that the output will not be meaningless (Han, 2012).

Many non-speaking English users commented about many MT systems, such as Google translate and Reverso, that they got strong help in translation, in particular from English to their native languages. Those users use the MT applications on smart phones as well. Other users commented that the results are quick but the accuracy is always an issue in all applications. They recommend using professional Translation service for 100 % accuracy (Alison, 2016).

The best-in-breed Enterprise Machine Translation solutions employ smart technologies that raise output quality, lowering the human factor across the spectrum of applications, and thus improving cost-effectiveness. Augmented and Hyped technologies Machine Translation are capable of delivering exceptional levels of brand fidelity, while at the same time reducing the need for human post editing, by aligning translations with enterprise's brand language prior to editing.

Conclusion

The world population exceeds 7 billion people; about four billion of them are internet users with different languages. Machine Translation systems are employed by global companies that need to translate to portion of those users in their native languages. This paper provided an overview for the state of the art of MT systems, the languages covered by MT systems and the used technologies. The forms of using MT in business have been described. Pros and Cons of using MT in business and the current challenges are explained. The main MT systems involving Arabic language are presented and the need for post-editing is recognized. Many MT software systems working in business, in both text and speech, has been presented and analyzed. MT evaluation methods, both human based or Automated, are discussed.

References

- ABBYY Textgrabber , (2017, December 1), Retrieved from <http://www.textgrabber.pro/en/>.
- Alba, D. (2012, November 13), SayHi Translate Review, Retrieved from <https://www.laptopmag.com/reviews/apps/sayhi-translate#sthash.BCJMGyxm.dpuf>
- Albat, Thomas Fritz (2017, December 1). "Systems and Methods for Automatically Estimating a Translation Time." US Patent 0185235, Retrieved from https://wikivisually.com/wiki/Translation_made_by_Machine
- Alison K. , (2016, July), Comments in Language Blog, Retrieved on 7 December 2017 from: <http://www.k-international.com/blog/translation-apps-2016/>
- Anderson, D.D. (1995). Machine translation as a tool in second language learning. CALICO Journal. 13(1). 68–96.
- Carl, Michael; Way, Andy (2003). Recent Advances in Example-Based Machine Translation. Netherlands: Springer. ISBN 978-1-4020-1400-0.
- Google Translate, (2017, December 1). Retrieved from http://translate.google.com/about/intl/en_ALL/languages/
- Han, (2012). LEPOR: A Robust Evaluation Metric for Machine Translation with Augmented Factors, in Proceedings of the 24th International Conference on Computational Linguistics (COLING 2012): Posters, pages 441–450, Mumbai, India.
- Hideya Mino (2014). "Syntax-Augmented Machine Translation using Syntax-Label Clustering", Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 165–171, Doha, Qatar. Association for Computational Linguistics.
- Internet World Stats. (2017, November 30). Retrieved from <http://www.internetworldstats.com/stats7.htm>

- iTranslate Voice, (2017, December 1). Retrieved from <http://itranslatevoice.com/>
- Jurafsky, Daniel; Martin, James H. (2009). *Speech and Language Processing*. Pearson. pp. 906–908.
- Makoto Nagao (1984). "A framework of a mechanical translation between Japanese and English by analogy principle". In A. Elithorn and R. Banerji. "Artificial and Human Intelligence". Elsevier Science Publishers.
- Microsoft Translator, (2017, December 1). Retrieved from <https://www.microsoft.com/en-us/translator/languages.aspx>
- Philipp Koehn (2009). *Statistical Machine Translation*. Cambridge University Press. p. 27. ISBN 0521874157.
- Rowe, Sam Del (2017, June). "SDL Adds Neural Machine Translation to Its Enterprise Translation Server". CRM Magazine. Retrieved from <http://www.destinationcrm.com/Articles/CRM-News/CRM-Featured-Articles/SDL-Adds-Neural-Machine-Translation-to-Its-Enterprise-Translation-Server-118775.aspx>
- Safaba Translation Solution (2017, November 30). Retrieved from <http://www.machinetranslation.net>
- Satoshi, Nakamura (2009), "Overcoming the Language Barrier with Speech Translation Technology", *Science & Technology Trends - Quarterly Review* No.31 April 2009.
- SayHi, (2017, December 1), SayHi Translate, Retrieved from <https://itunes.apple.com/us/app/sayhi-translate-use-your-voice/id437818260?mt=8>
- Skype , (2017, December 1). Retrieved from <https://www.csmonitor.com/Technology/2016/0115/Skype-s-universal-translator-bridges-the-language-gap-mostly>
- Slator, (2017, November 1), "Kantan MT Users Can Now Customize and Deploy Neural Machine Translation Engines ". Retrieved From <https://slator.com/press-releases/kantanmt-users-can-now-customise-and-deploy-neural-machine-translation-engines/>

Machine translation status and its effect on business

- Slator, (2018, January 1), "Omniscien Technologies Announces Release of Language Studio™ with Next-Generation NMT Technology", Retrieved From
<https://slator.com/?s=%22Omniscien+Technologies+Announces+Release+of+Language+Studio%E2%84%A2+with+Next-Generation+NMT+Technology>
- Speak & Translate, (2017, December 1), Retrieved from
http://www.apalon.com/speak_and_translate.html

Appendix A

List of figures

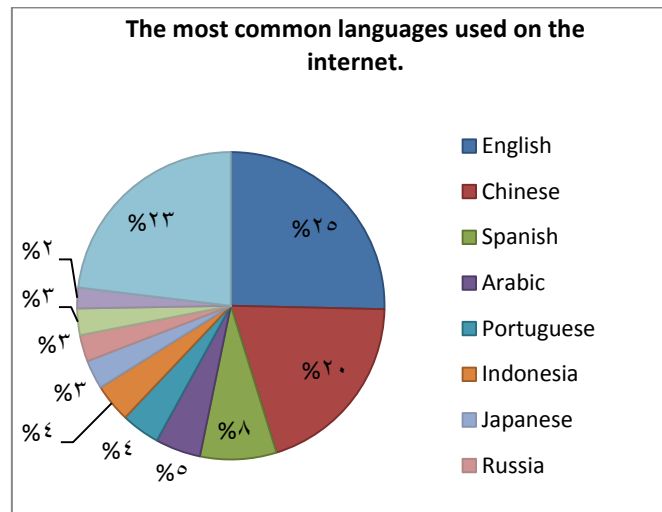


Figure A1: The most common languages used on the internet (*Internet World Stats, 2017*).

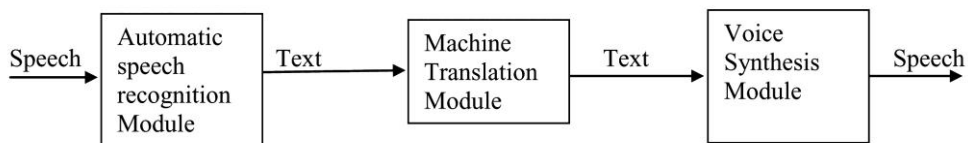


Figure A2: Speech translation system's Modules

Machine translation status and its effect on business

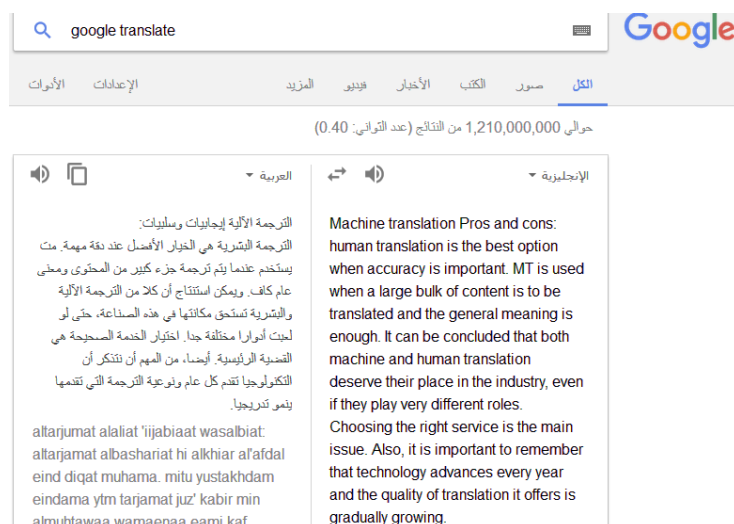


Figure A3: A sample of the Arabic language output of the Google translator for English language text.

Appendix B MT systems involving Arabic language

The MT application	The used technology
Google Translate	Statistical and neural machine translation
Omniscien Technologies Language Studio	Hybrid Neural Machine Translation, Hybrid Deep Neural Machine Translation, syntax-based, rules-based and Statistical Machine Translation. Provides a server platform and a set of desktop tools to aid in translation. Omniscien also offers Machine learning, data cleaning and language processing tools.
Bing Translator	Microsoft's linguistically informed statistical MT system
WorldLingo	undeclared
SYSTRAN	Hybrid rules-based and SMT