Trustpoint.One

How Artificial Intelligence is Transforming Machine Translation and the Global Language Business

Contents

Executive Summary

Artificial intelligence (AI) is changing the way we collect, interpret and analyze data and, in the process, the way we conduct business. This includes the global translation industry. AI, in the form of artificial neural networks and neural machine translation (NMT), are at the forefront of advancing translation technology. In this white paper we will explore a brief history of MT and attempt to show how it is being impacted by AI and what that might mean for businesses who have a need for professional translation services.

Introduction

With increasing global access to technology, there has never been more content in the world. Recent research estimates that more than 2.5 quintillion bytes of data are created every day. Combined with globalization and international eCommerce, this means businesses have a surplus of content—marketing materials, product descriptions and internal documents—that need to be translated.

Advances in machine translation (MT) via artificial intelligence (AI) and deep learning offer the potential for an efficient, cost-effective solution to this growing need. In this paper, we will examine the evolution of machine translation from the early days of rule-based operations, the effect of AI on MT and the future possibilities.

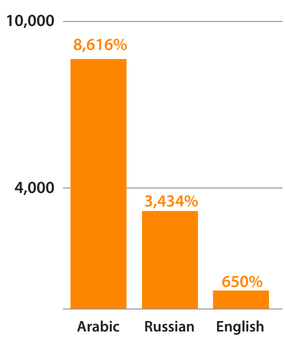
[1] *Domo Data Never Sleeps 5.0* https://www.domo.com/learn/data-never-sleeps-5

Background

In places like North America and Europe, many people are used to English being the lingua-franca. However, two regions of the world are not representative of the whole—and certainly not representative of the internet.

As of 2018, more than 4.2 billion people use the internet.[2] That is more than half the world's population, and not all of those users speak English. In fact, growth in internet usage by non-English speakers has far outpaced English speakers.

Percentage Growth in Internet Usage by Language

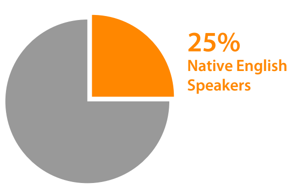


From 2000 to 2018, Arabic has seen an **8,616%** increase and Russian has seen a **3,434%** increase. Meanwhile, in comparison, English has seen a **650%** increase.

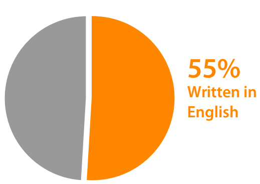
While this is partially attributable to the fact that English was likely the first language used on the internet[3], it shows that other languages are catching up.

These other languages have not quite closed the gap, though.

Internet Users



Online Content



Native English speakers make up **25%** of internet users, but it is estimated that around **55%** of online content is written in English.[4]

This reveals a fundamental asymmetry in the current ecosystem, and it is a problem for businesses that need to communicate effectively in a global economy.

This highlights the continued need for translation services. Still, businesses in multilingual markets—such as healthcare and pharmaceuticals—have noted several consistent translation challenges: quality, process management and deadlines. Current MT technology—in tandem with human translators—offers increasing potential for businesses in a variety of industries.

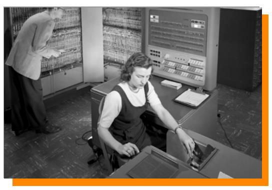
[2] *Internet World Stats* https://www.internetworldstats.com/stats.htm
[3] *The Guardian The Digital Language Divide* http://labs.theguardian.com/digital-language-divide/
[4] *Unbabel Top Languages of the Internet, Today and Tomorrow* https://unbabel.com/blog/top-languages-of-the-internet/

The History of Artificial Intelligence and Machine Translation

AI and MT sound like modern concepts, but in fact, their history spans decades.

From Rule-Based Machine Translation to Neural Machine Translation[5,6]:

1933:
Petr Smirnov Troyanskii develops the idea for, and patents a device that, transforms word-root sequences into their other-language equivalents. He suggests a three-stage process involving editors.



1939:
Bell Labs demonstrates the Voder, the first electronic speech-synthesizing device at the New York World's Fair. The Smithsonian's Science News Letter describes it as the, "first device that actually creates human speech."

1954:
Georgetown University and IBM partner for the first public demonstration of machine translation. More than 60 Russian sentences are translated to English with six grammar rules and a vocabulary totaling 250 words.

1966:
MT is dealt a blow from the Automatic Language Processing Advisory Committee (ALPAC). The group releases a report concluding that research up to that point had not produced useful results. It results in a freeze in federal funding for MT research and development.

1967:
L. E. Baum and colleagues at the Institute for Defense Analyses (IDA) in Princeton, New Jersey, develop hidden Markov models, the mathematical basis of continuous-speech recognition.

1968:
Peter Toma, a member of the Georgetown-IBM MT project, founds SYSTRAN despite lack of funding from the government. Soon after, the company contracts with the United States Air Force to provide Russian to English MT during the Cold War.

1983:
The Automated Language Processing System (ALPS) is the first MT software for a microcomputer.

1988:
IBM's Thomas J. Watson Research Center starts working on statistical MT methods that equate parallel texts, then calculate the probabilities that words in one version will correspond to words in another.

1989:
Knowledge-Based Machine Translation, or KANT, is founded at Carnegie Mellon University to develop scalable, practical MT systems for technical documents.

1990:
Dragon Systems releases DragonDictate, the first real consumer speech-to-text system product for PCs. Initially, it comes with a $9,000 price tag.

[5] *Wire Machine Translation's Past and Future* https://www.wired.com/2000/05/timeline/
[6] *William John Hutchins Machine Translation: A Brief History* http://hutchinsweb.me.uk/ConcHistoryLangSci-1995.pdf

1993:

BBN Technologies demonstrates the first off-the-shelf MT workstation for real-time, large-vocabulary (20,000 words), speaker-independent continuous-speech-recognition software.

1997:

AltaVista's Babel Fish offers real-time SYSTRAN translation on the Web. The same year, Dragon Systems' NaturallySpeaking and IBM's ViaVoice are the first large-vocabulary continuous-speech-recognition products for PCs.

1999:

IBM releases ViaVoice for the Macintosh, the first continuous-speech-recognition Mac software.

2004:

The Translation Automation User Society (TAUS) is formed at a roundtable meeting in San Francisco as a forum and community for businesses to exchange ideas, share experiences and drive innovation.

2006:

Google launches Google Translate run by a statistical machine translation (SMT) engine.

2009:

Safaba Translation Solutions is founded by Alon Lavie, a research professor at Carnegie Mellon University, and Robert Olszewski, the principal software engineer at Carnegie Speech.

2014:

The first academic papers proposing NMT and the use of neural networks are published.

2015:

Amazon purchases Safaba Translation Solutions to help develop the company’s MT technology.

2016:

The Google Brain team introduces Google Neural Machine Translation to Google Translate, leveraging an artificial neural network capable of deep learning. OpenNMT, an open-source initiative for NMT and neural sequence modeling, debuts the same year.

2015:

A team from the University of Montreal wins first and third prizes at the Workshop on Statistical Machine Translation using NMT. Baidu, a Chinese technology company, authors a paper about using NMT for multiple language translation.

2018:

The Facebook AI Research (FAIR) team introduces grants for proposals on NMT and low-resource languages. FAIR also introduces research on unsupervised MT using a model based on the combination of NMT and phrase-based MT.

[5] *Wire Machine Translation’s Past and Future* https://www.wired.com/2000/05/timeline/
[6] *William John Hutchins Machine Translation: A Brief History* http://hutchinsweb.me.uk/ConcHistoryLangSci-1995.pdf

The Beginnings of Machine Translation

As we can see, MT has a long history, but the technology really started to advance in the '80s. Dr. John Tinsley, CEO and co-founder of Iconic Translation Machines, says that rule-based MT (RBMT) was the predominant approach until the mid-to-late '80s. However, Tinsley notes that the technology did not scale effectively and never lived up to its initial promise.

This is due to the nature of how RBMT works. It is reliant on grammatical rule sets and vocabularies for a source language and a target language. By applying these rule sets, conversion from one language to another is possible.

RBMT can effectively translate basic content, as long as a quality bilingual dictionary is available. Historically, the availability of these has been insufficient, and creating one is expensive. There is also difficulty in dealing with rule interactions in large systems and with idiomatic expressions.[7]

Additionally, adding new content or adjusting rule sets becomes time-consuming, burdensome and, once again, expensive. Furthermore, a study on Catalan-Spanish translations concluded that RBMT fared fine orthographically and morphologically, but a newer form of MT, statistical machine translation (SMT), performed better semantically.[8]

Soon, the tides turned to statistical machine translation (SMT).

SMT was facilitated by increasingly powerful computers that could handle heavier processing. SMT was the first instance of using examples of previous versions so that the machine would learn how to produce new translations. The web's earliest translation tools such as Babel Fish and Google Translate were based on this model.

Despite these advances and the commercial viability of SMT, there are limits to the technology. The constraints are related to building training data for the translation engine. A large database is required, which takes time and resources.

SMT also faces challenges translating into grammatically-rich or morphologically-rich languages. This can be attributed to an increased variety of "grammatical features expressed with morphology."[9]

[7] *International Journal of Computer Science Issues Machine Translation Approaches: Issues and Challenges* https://www.ijcsi.org/papers/IJCSI-11-5-2-159-165.pdf
[8] *Computing and Informatics Study and Comparison of Rule-Based and Statistical Catalan-Spanish Machine Translation Systems* http://www.cai.sk/ojs/index.php/cai/article/view/940
[9] *Language Technologies Institute Translating into Morphologically Rich Languages with Synthetic Phrases* http://www.aclweb.org/anthology/D13-1174

To put it simply, grammatically-rich and morphologically-rich languages often have complex word structures and flexible word orders. For instance, in Hungarian, word order can be variable, making it confusing to non-native speakers and MT engines.

The most important part of a sentence in Hungarian is a verb (or predicate). A sentence is built around the verb and the other words' relation to the verb. That is quite different than the standard subject-verb-object construction seen in English.

Eventually, from a research perspective, that technology hit the limits of its capabilities. Then in 2014, the first academic papers proposing the use of neural networks in MT appeared. Within two years, it was a reality.

Neural networks are not necessarily a new concept, but applying them to translation is. Tinsley notes that people started testing the concept of NMT several years ago. "It had really good results out of the box," he said.

"People kind of picked up that ball and started to run with it to the extent that it is now by far the most predominant approach."

Dr. Joss Moorkens, assistant professor at the School of Applied Language and Intercultural Studies at Dublin City University, agrees that the biggest efforts are being put toward NMT. Moorkens says NMT is what the majority of research is focusing on currently. He notes the initial jumps from SMT to NMT were significant, and right now, improvements are incremental with high potential in the future.

The learning involved in NMT is quite different than past models—primarily because neural networks are so much bigger and more powerful than other models. Because of that, they were not practical until computer hardware was up to the task.

Now, graphics processing units (GPU) allow for more complicated machine learning. GPUs are widely used for video and computer games, and the ever-advancing video game consoles and gaming PCs created a competitive market. This drove down hardware prices, making cheap, multiprocessor GPUs widely available.[10]

These GPUs excel at operations necessary for NMT—like fast matrix and vector multiplications—which accelerate learning greatly.

[10] *Jurgen Schmidhuber Deep Learning in Neural Networks: An Overview* https://arxiv.org/pdf/1404.7828.pdf

Current Innovations

For those in the translation industry, the useful part is that NMT takes more context into account while translating. Tinsley says that previous models were good at reproducing things that they had seen over and over—they were good at memorizing when it came to looking at 'windows' of words in a sentence.

SMT systems look at one window, translate it and move onto the next window. NMT acts much more like a human brain—learning and looking for the whole context of a particular translation. When a NMT engine translates the last word in a sentence, it is inherently considering all of the previous words in the sentence.

Indeed, today NMT is showing gains on previous models. When Macduff Hughes, director of Google Translate, and Jeff Dean, senior fellow of Google Brain, teamed up to tackle NMT around 2015, others within the company were unconvinced of the practical improvements NMT could achieve.

In their first real test, a NMT system and a SMT system were run side-by-side translating English to French. The NMT system showed an improvement over the old system of seven points on the BLEU score. BLEU, or bilingual evaluation understudy, scores are used to compare MT with average human translations. At the time, an improvement of even one point in English-French translation was considered very good.[11]

To ensure it was not chance occurrence, human translators did a side-by-side comparison of the translations. Based on their scores, a distinct improvement was noted. A more recent study, conducted by Antonio Toral and Andy Way, pitted a NMT system against a phrase-based SMT system to translate literature from English to Catalan. Literature, text with artistic aspects and nuance, is widely thought to be one of the biggest challenges for any MT system. Yet their results showed NMT fared "significantly better" than the SMT system, with an 11 percent improvement overall.[12]

How is this possible?

[11] *The New York Times Magazine The Great A.I. Awakening* https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html
[12] *Antonio Toral and Andy Way What Level of Quality can Neural Machine Translation Attain on Literary Text?* https://arxiv.org/pdf/1801.04962.pdf

AI, Deep Learning and MT

The short answer is artificial intelligence (AI)—but specifically deep learning and artificial neural networks.

Like AI, there are a number of definitions that describe deep learning. Li Deng and Dong Yu, authors of *Deep Learning: Methods and Applications*,[13] note that there are two common threads among these definitions:

1. Models consisting of multiple layers or stages of non-linear information processing.
2. Methods for supervised or unsupervised learning of feature representation at successively higher, more abstract layers.

At a high level, deep learning involves machine learning algorithms that process multiple layers of data for feature extraction and to model relationships among the data. This involves a process by which successive layers depend on input from previous layers.

Basically, higher-level concepts are, in part, being defined by lower-level concepts. Through this process, multiple abstractions are identified and a hierarchy of concepts is created—unsupervised or supervised by humans. This approach differs from using task-orientated algorithms in that humans are not defining what the algorithms need to look for; they find patterns and relationships organically.

The goal is to mimic neural networks of the human brain. In his paper *Deep Learning in Neural Networks: An Overview*, Jurgen Schmidhuber breaks down how neural networks in the human mind work:

"A standard neural network (NN) consists of many simple, connected processors called neurons, each producing a sequence of real-valued activations. Input neurons get activated through sensors perceiving the environment, other neurons get activated through weighted connections from previously active neurons. Some neurons may influence the environment by triggering actions."

To put it in layman's terms, the human brain contains billions of nerves cells, or neurons, which process and send information via electrochemical signals. These neurons are connected by trillions of synaptic connections, forming a neural network. Every time you experience something new or learn from a new event or fact, your brain adapts and rewires itself to account for this information.

From this standpoint, we can see how AI is working to replicate these processes through non-linear learning and hierarchical concept building.

[13] *Li Deng, Dong Yu Deep Learning: Methods and Applications* https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/Deep-Learning-NowPublishing-Vol7-SIG-039.pdf

The Efficiencies of Neural Translation

In practice, these advances mean greater efficiency for translators and, subsequently, greater efficiency for their clients. Tinsley and Moorkens note however, that efficiencies depend on how NMT is used.

Moorkens says it is typical for translators to sort tasks into "premium" and "bulk" work. For instance, premium work would be documents that require a lot of subject matter expertise where there would be serious repercussions for incorrect translations.

However, Moorkens believes bulk translations are where NMT is making inroads. He points to content that is very straightforward and less creative as primed for NMT.

Tinsley points to two use cases for MT: raw translations and MT as a productivity tool. A raw MT translation means there is no editing whatsoever. It could be reading a single document, solving a support query or translating an email.

Hundreds of millions of words are machine-translated every day, and the vast majority are not edited. This includes situations where the content is not valuable enough or the turnaround time is too quick to warrant human translation at all.

For instance, imagine you were conducting research and your search results keep bringing up articles in a foreign language. With something like Google Translate, you can hit a button and get a fairly good idea of what they are saying in a matter of seconds. From that standpoint, the efficiencies involved are immense.

Consider then how MT can also produce efficiencies for translators. It can be used to get a suggested translation, which serves as the foundation for a final translation.

Jim Moore, senior vice president at Trustpoint.One, recalls the case of a law firm that had roughly 10,000 pages of content to be translated from Spanish to English. That content was machine translated and then Trustpoint.One provided post-editing of the MT content.

"In this particular case, the quality of the machine translation output was good enough where we could find translators who were willing and able to go through and clean it up, if you will, and deliver to the client a product they found to be acceptable," Moore said.

Even though some translation service providers are using machine translation and post-editing with a translator, Tinsley cautions that the efficiencies are variable based on language, the MT system being used and individual translators.

John Romanski, a translator with 20 years of experience, estimates his productivity improves by about 30 to 40 percent on projects with a good quality MT output.

As Moorkens mentioned, less creative, less nuanced texts are much easier for MT systems to work with. Romanski gives the example of a standard operation manual with simple directions—"press one," "press two," "go here"—as a prime candidate for MT.

Because of the large amount of training data needed for MT, particularly NMT, some languages are better served than others by the technology. Those would be languages that are more widely spoken, such as English, Spanish and Chinese.

"You know, of the 6,000 to 7,000 languages that are existent, only roughly 2 percent of them have any sort of MT availability and that would be the most major languages," Moorkens said.

Tinsley echoes this, observing that languages with many speakers in large markets—English, Chinese, etc.—are where MT is most effective.

We are already seeing the benefits of MT, but where will things go from here?

The Future of Machine Translation

Predicting the future is always difficult. After all, according to futurists from past decades we are supposed to be using jetpacks to travel to our three-day-a-week jobs. That being said, development in the field of translation has been moving quite swiftly.

Key Players in AI

Of course, it is difficult to discuss the future of technology without casting an eye toward current tech giants such as Facebook, Microsoft, Amazon and Google.

Tinsley feels that what they are doing now is the essence of what they will continue to do. Their interests seem to be geared more toward providing functionality and features within their platforms and use cases rather than separate services.

"Google and Microsoft are the two more direct and longer-standing players in the game, and they are doing a lot of leading development in this area," Tinsley said.

In the case of Google and Microsoft, users are uploading their own data to the platforms and can adapt them to their own specific needs. This what Tinsley calls "light customization." This is possible through something like Google's paid services, which is separate from Google Translate.

Via services such as Google Cloud Translation and its new Cloud AutoML (which is currently in beta), Tinsley says users could, for example, adapt it specifically for eCommerce product descriptions. "You can upload some of your data to Google's system and have it kind of learn from that for your content," he said.

Google's former AI chief scientist Fei-Fei Li echoed this when Cloud AutoML was announced. She said, "And with AutoML Translation you can upload translated language pairs to train your own custom translation model."[14]

Tinsley explains that the organizations Iconic works with have a variety of translation requirements, which necessitate different workflows for different content. Google Translate can cover some of those translations, but not all. Iconic focuses on the areas where Microsoft or Google's solutions are not viable for one reason or another.

For example, when a large language service provider needed an MT system to translate opinions about patent applications, it came to Iconic. This is the type of specialized task that would be nearly impossible by simply using Google Translate.

It required a system that could deal with the complexity of patents—no small task—but also subjective language in relation to the patents. With a small amount of additional data from the client and a tweak to an MT engine built specifically for intellectual property content, Iconic produced an adapted version of the system for the client's specific need.

Still, Moorkens feels that these big companies are at the forefront of NMT and that they will most likely make the big breakthroughs in the future.

In fact, this year Facebook developed[15] an unsupervised MT system—relying on a combination of NMT and phrase-based MT—for fast, accurate translations for more languages. This model showed BLEU score improvements of more than 10 points.

Not only that, but this unsupervised system also makes strides in improving translations when there are little training resources for a language. Regarding the system, Facebook wrote:

"For low-resource languages, there is now a way to learn to translate between, say, Urdu and English by having access only to text in English and completely unrelated text in Urdu – without having any of the respective translations."

Improving translations for low-resource languages is one area it appears the tech giants will have an active hand in advancing. Of course, it benefits their business interests but internet users across the world will benefit, as well.

---

[14] *Google Empowering businesses and developers to do more with AI* https://www.blog.google/products/google-cloud/empowering-businesses-and-developers-to-do-more-with-ai/

[15] *Facebook Unsupervised machine translation: A novel approach to provide fast, accurate translations for more languages* https://code.fb.com/ai-research/unsupervised-machine-translation-a-novel-approach-to-provide-fast-accurate-translations-for-more-languages/

Looking Ahead

Tinsley admits that he cannot say for sure where NMT will end up, but there are three areas with potential.

1. Domain adaptation:

One area that NMT has not quite touched on yet is more specialized domain adaptations. Because it is so new, most people have been working on what Tinsley calls the "general use case," or the "Google Translate use case." This means trying to translate any type of content in any language. It also means that people have not been working to apply NMT to a specific use case. Tinsley says if you use this broad kind of tool but you also have a very specific need or special requirements for the translation output, the results might not meet your expectations. Once the general research starts to top out, more research will be put into domain adaptations for specific content types.

2. MT for long-tail languages:

Another area that is ideal for advancement is in "low-resource" or "long-tail" languages. These are languages where there is little data available to train translation engines. Tinsley says, generally speaking, most commercial developments and localization happen where there are big markets and demand—languages like English, Spanish, French, Chinese, Japanese, etc. However, there is increasing consumer demand online for translation of low-resource languages. In his opinion, it will be a large area of research in the coming years.

3. Increased processing power:

Computer hardware and software often drive new developments, and Tinsley believes this will continue. Tinsley hopes this will lead to NMT systems that will eventually be able to look at document-level context. So, rather than looking at a document or email sentence by sentence, the whole document would be taken into account.

Another issue that is often debated regarding the future of MT is how accurate it will become. Can it ever truly reach 100 percent accuracy? Currently, there are some issues tied to terminology that Moorkens thinks will soon be solved.

There is a researcher at his university trying to insert a separate neural network within the NMT process to guarantee accuracy in terminology. There are others attempting to use a process called "forced decoding" in NMT to force systems to always use the specified terminology as well.

As to whether NMT will reach 100 percent accuracy, Moorkens feels there are still many creative texts that require real-world knowledge and that NMT would have to make strides to properly translate them. Tinsley says the accuracy, once again, depends on the use case.

For something like Google Translate where the input could be any type of content, in any language, he does not think it will reach a point where a machine will be able to guarantee 100 percent accuracy.

That being said, specific use case accuracy could improve significantly. When organizations have more control over the content input for translation, MT could reach extremely high qualities.

There are others who are decidedly more optimistic. Sundar Pichai, CEO of Google, touted Google Translate's advances in 2016. Pichai noted Google's Chinese to English translation accuracy[16] was only a step away from human-level quality.

Google also covered these claims in a 2016 paper[17] in which the tech company concluded:

"Using human-rated side-by-side comparison as a metric, we show that our Google Neural Machine Translation system approaches the accuracy achieved by average bilingual human translators on some of our test sets. In particular, compared to the previous phrase-based production system, this GNMT system delivers roughly a 60 percent reduction in translation errors on several popular language pairs."

Naturally, organizations such as the American Translators Association (ATA) push back against this assertion. The ATA states:

> "While both MT and human translation play an important role in the translation industry, human translation by skilled professionals will always produce a more accurate, precise, and true-to-nature result than MT."[18]

Whether or not MT ever achieves 100 percent accuracy, it is already having an effect on those in the industry.

[16] *The Verge Apple boasts about sales; Google boasts about how good its AI is* https://www.theverge.com/2016/10/4/13122406/google-phone-event-stats
[17] *Google Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation* https://arxiv.org/pdf/1609.08144.pdf
[18] *American Translators Association Machine Translation vs. Human Translation* https://www.atanet.org/governance/advocacy_day_2017_handout_myth.pdf

The Impact on Professional Translators

As we have moved into the 21st Century, the effect of automation on human labor has been a concern in many industries—including translation.

Tinsley notes there are many use cases—think about using Google Translate for a personal translation—where there is already no need for human translation. There are also use cases where MT is not applicable. This could be the need for highly-specific, highly-accurate translations where it just makes more sense for a human to handle the translation.

There is a middle ground developing rapidly between machine-only or human-only translations. This primarily involves review or post-editing of MT. "I think that's where the role of the translator is going to change," said Tinsley.

In the future, there will likely be a machine somewhere in the loop, changing the nature of how translators work. As Tinsley noted, direct translation work will not be as common as reviewing and quality assurance work.

Moore agrees, pointing toward other areas in the economy such as the auto industry. He says there has been concern about robotics and automation in automotive manufacturing, but there has not been a concurrent loss of employment. Rather, there have been shifts in the roles of people assembling cars to quality control roles.

"I think the role will be modified, but there will still be a demand," Moore said.

Moorkens also agrees that the work of human translators will likely include the use of MT going forward, but warns there are repercussions for translators to using this model, though.

Translators are already at the very end of the production network without access to earlier stages. Practically, this means they usually do not know the provenance of the content they are post-editing—what MT system was used or if the quality expectations are different from the last MT content they were dealt. It may exacerbate the issues already at play.

Then there is the actual work to consider. Moorkens says that translators are finding that they are more exhausted from post-editing after smaller lengths of time. They are able to translate from scratch for a longer time without feeling burned out, even if the overall progress of translation is slower.

Romanski has also seen a shift in translation work. Over the years he has seen an increase in pressure when it comes to productivity and deadlines due to MT. Twenty years ago, he would receive a document and have about a week to translate it. Now, most the projects he takes on need to be done by the next day.

On the other hand, Romanski sees MT as a sign of progress, particularly when a translation is needed for informational purposes only.

“Taking a text, let’s say in French, and rendering it into English...it’s tedious manual work. There is no need for people to spend hours doing that if the computer can do the job,” Romanski said.

Ultimately, MT technology is changing the way in which translators work. As with other industries changed by technology, they continue to adapt to new paradigms and persist.

Conclusion

Considering AI technology is fundamentally changing services across every industry, companies and translators should prepare for the same with translation services. Not every detail of the industry’s future is completely clear, but there are trends that seem quite certain:

- NMT translation will be the standard for MT moving forward.
- MT with post-editing will become more common for standard translations, while specialized content will still need the involvement of expert human translation.
- The future of MT is bright—especially for low-resource languages, more specialized domain adaptations and the opportunities increased computing presents.

Appendix: Key Terms

Artificial Intelligence (AI):

While the definition of artificial intelligence is malleable, it is generally defined as technologies and processes that augment human knowledge and capabilities. By augmenting these capabilities, AI collects and analyzes high volumes of data more efficiently and accurately than the human mind.

BLEU Score:

The bilingual evaluation understudy is an algorithm that scores machine translations against average human translations. It is the standard barometer for translations.

Domain Adaptation:

Using a machine translation system to recognize and apply information and patterns from a source language (or domain) to a target language (another domain) for translation.

Low-Resource / Long-Tail Languages:

In terms of machine translation, a low-resource or long-tail language generally means one of two things. It either refers to a language with a limited amount of human speakers, or a language with less content available for training machine translation engines.

Machine Translation (MT):

Machine translation is technology that translates text from one language to another without human involvement. Historically, there have been several underlying approaches to MT.

Neural Machine Translation (NMT):

Currently, it is the dominant form of MT. Instead of the previous models, neural machine translation is the process of building and training an artificial neural network that continually improves upon itself with new data.

Rule-Based Machine Translation (RBMT):

This is the classical approach to MT. The first commercial MT solutions were founded on rule-based systems developed in the 1970s. These systems rely on numerous inputed linguistic rules and bilingual dictionaries. Morphological, syntactic and semantic markers are then used to create a translation.

Statistical Machine Translation (SMT):

An approach to MT that supplanted RBMT in the 1980s. Statistical Machine Translation uses statistical models derived from analysis of monolingual and bilingual training data. Models are built with translated text files and uploaded to train a translation engine.

About Trustpoint.One

Trustpoint.One offers translation solutions for a wide variety of industry verticals including: advertising, agriculture, banking, chemical, manufacturing, marketing, legal, life sciences, education, energy, entertainment, finance, government, hospitals and healthcare, information technology, insurance, pharmaceuticals and transportation.

We will consult with you in regards to your specific translation needs and suggest the most suitable translation solution for your particular needs.

Whether you need translations for your website, ethics and compliance content, training materials, technical manuals, software, advertising, medical IFUs or legal documents, Trustpoint.One is here to be your trusted translation partner.

Learn more about our business and translation solutions here.

www.trustpoint.one
translate@trustpoint.one
412.261.1101