SNS COLLEGE OF ENGINEERING

Kurumbapalayam (Po), Coimbatore – 641 107

AN AUTONOMOUS INSTITUTION

Accredited by NBA – AICTE and Accredited by NAAC – UGC with 'A' Grade
Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai

CS8691-ARTIFICIAL INTELLIGENCE

Machine Translation

An Introduction to Machine Translation

This is the first in a series of Machine Translation (MT) blogs by Globalization Partners International. The MT technology discussed in this series will be of interest to both new and sophisticated translation users, especially those who face the daunting task of translating tremendous amounts of content with very tight budgets and time constraints.

We will start the series with a brief introduction to MT, discussing its history, explaining how the technology works and providing an overview of its use in business today.

History of Machine Translation

The modern history of automated translation begins primarily in the post WW2 & Cold War era, when the race for information and technology motivated researchers and scholars to find a way to quickly translate information.

In 1947, an American mathematician and scientist named Warren Weaver published a memorandum to his peers outlining his beliefs about a computer's capability to render one language into another using logic, cryptography, frequencies of letter combinations, and linguistic patterns. Fueled by the exciting potential of this concept, universities began research programs for MT, which eventually founded the Computational Linguistics field of study.

In the 1950's, one such research program at Georgetown University teamed up with IBM for a MT experiment. In January 1954, they demonstrated their technology to a keenly interested public. Even though the machine translated only a few dozen phrases from Russian into English, the project was hailed a success, as it showcased the possibilities of fully automated MT and provoked interest in and funding of MT research worldwide.

The optimism of the Georgetown-IBM experiment was replaced by a feeling of pessimism in the 1960's, when researchers and scholars became frustrated at the lack of progress made in the field of computational linguistics despite huge funding. In 1966, a special committee formed by the

United States government reported that MT could not and would not ever be comparable to human translation and therefore was an expensive venture that would never yield any useable results.
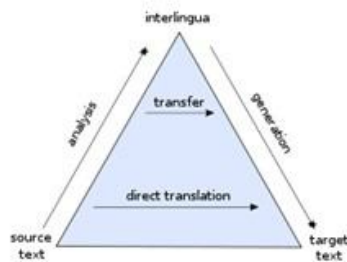
In the 1970's and 80's, researchers changed their focus to develop tools that would facilitate the translation process rather than replace human translators, leading to the development of Translation Memory (TM) and other Computer Assisted Translation (CAT) tools that are still an integral part of the localization process today.

In the 1990's, globalization, the proliferation of the internet, access to cheap & powerful computers and advancements in speech recognition software were a few factors that fueled the progress of MT.

Today, despite advancements in the field, MT is still seen by most professionals and firms requiring accurate translations as an inadequate substitution for human translation teams. However, many companies are embracing MT and are applying the technology in their localization work flows, allowing them to get more out of their translation budgets, if post editing can be completed in a cost-effective manner (more on this in later blogs). In fact, if companies use MT in the right way, quality can actually be improved compared to pure Human Translation.

How does Machine Translation work?

Machine Translation is software that automatically renders text from one language into another. There are many different types of MT software, but for the sake of brevity, we will cover only Rule-Based Machine Translation (RBMT) and Statistical Machine Translation (SMT), two commonly used and very different approaches.

Rule-based Machine Translation is built on the premise that a language is based on sets of grammatical and syntactic rules. In order to obtain the translation of a segment, the software would need to have a robust bilingual dictionary for the specified language pair, a carefully outlined set of linguistic rules for the sentence structure of each language and a set of rules to link the two sentence structures together. These requirements can be time-consuming and expensive to create and must be done each time a new language pair is required. A key benefit of RBMT over other MT approaches is that it can produce better quality for language pairs with very different word orders (for example, English to Japanese).

Statistical Machine Translation is built on the premise of probabilities. For each segment of source text, there are a number of possible target segments with a varying degree of probability of being the correct translation. The software will select the segment with the highest statistical probability. SMT is generating the most interest in the field today, as it can be applied to a wide range of languages and is not as resource intensive as RBMT.

Machine Translation in Business Today

Companies these days are being forced to do more with less. Organizations are desperately trying to handle huge amounts of content in an era where shrinking budgets and instant gratification cause tremendous pressure. It is simply too cost prohibitive and time consuming to tackle these mountains of content using the standard human translation process, so it's no surprise that translation buyers are looking to MT for the answers.

Yes, Machine Translation quality is imperfect and it may never be perfect in our lifetime; but it is continuously improving. When in experienced and informed hands, it is a useful linguistic tool to translate large volumes of the right content.

Some of the world's largest companies and most recognizable brand names are putting MT technology to use in their localization workflows, in conjunction with Translation Memory, glossaries, style guides and human translators. This process allows a leading e-commerce company to make it simple for customers to buy & sell across borders. It enables a large global equipment manufacturer to send translated operations manuals to their technicians out in the field. It gives a multinational computer technology corporation the means to provide online support to foreign customers in dozens of languages.

Conclusion

This Machine Translation series will continue with more in-depth blogs on the technology, implementation, use, benefits, and limitations of Machine Translation. We will also discuss business situations where Machine Translation would and would not be a practical solution (with supporting case studies) and will explain how Machine Translation can be built into a translation/localization workflow in a way that helps organizations achieve their translation goals.