



## WHAT'S MACHINE TRANSLATION?

**Machine Translation (MT) is gaining exposure as a translation tool. It is used most successfully in domains where the text is technical, instructional, well structured, and repetitive.**

MT is the use of computer software to translate one language into another. It's been around since the 1950s, with several ups and downs in terms of research and development. Fictional stories abound of the early days of MT with phrases such as "*out of sight, out of mind*" being translated as "*blind fool*," and "*the spirit is willing but the body is weak*" becoming "*the vodka is strong but the meat is rotten*."

Today, MT is gaining public exposure as the Web has become an essential part of global commercial communication. MT "engines" are being incorporated in search engines and other sites. Improvements in MT's output quality have helped in this resurgence, both for general purposes and for use by professional language service providers (LSPs).

So, how does MT work and will all translation be processed by machine one day?

### How it Works

MT systems first came about as a combination of some computational linguists' attempts to understand human language through computer models and the US Department of Defense's need to translate Russian documents into English during the Cold War. Since then, a number of different approaches to MT have emerged:

- ➔ **Rules-based.** In this approach, just like you might have learned to diagram the grammatical structure of sentences in school, the software attempts to interpret the grammar of the input document to build a grammatical model of each sentence. The grammatical model of the input language is then mapped to the grammatical model of the output language.
- ➔ **Statistically-based.** Here, the MT engine is trained based on large volumes of existing content and its translation known as "bilingual text corpora." The MT engine uses the large volumes of data to create statistical correspondence tables. This statistical information determines the appropriate selection based on the probability that given a certain word, phrase, or sentence in one language, a

particular word, phrase, or sentence is the correct translation in the target language. While this approach is not language specific, large volumes of electronic text of similar content are required to get the best quality output from the MT engine.

- ➔ **Example-based.** Similar to the "statistically-based" MT approach, a bilingual text corpus is required. However, in the example-based approach, the corpus is used as a knowledge base to derive translations directly from matching instances of parallel structures of source text and translation found in the corpus.

### Examples

Although there are frequent, heated debates about which MT philosophy is more effective, when it comes to MT for commercial translations, it's probably not worth getting too hung up on which approach is used.

LSPs such as Lionbridge see MT as only one part of a service solution—meaning MT is only one element of the total process. For example, Lionbridge combines MT with the more "traditional" Translation Memory (TM) to provide a hybrid MTM approach. Text is first matched against a TM. If a solid match is not found, the text is sent for MT.

To get an MT process up and running, you need to be sure you have the content volumes required to make the upfront investment worthwhile (usually 200,000/year of new words). Initial language work is required to "tune" or customize the MT engine with the specific terminology and language rules for the languages to be translated and for the domain of the content.

For example, "trim" is likely to be translated differently in different domains: automotive, hairstyling, yachting, fashion and design, health and fitness.

Translations are run through the MT engine and the quality of the "raw" (i.e. unedited) output is assessed. Results are then used to further tune the engine as needed.

The more closely related the source and target languages are to each other, the higher the quality that can be achieved. For example, the quality of MT between variants of Spanish is higher than between Spanish and Portuguese, higher than between Spanish and other Latin languages such as French and Italian, and higher than Latin to Germanic languages. When you get to non-related

languages such as Semitic, Indian, and Far-Eastern languages, MT has to work much harder to get a quality output.

Only when raw output of decent quality has been achieved is it worth using the engine in a business environment. Raw output on its own can be used for “gisting” purposes (i.e., getting an idea of what a piece of text is about) and only needs to be of a certain quality level. Examples of this can be found commercially in Research & Development departments where they gist translation patents filed by competitors, and then ask for full human translations of patent portions that look interesting.

Raw output is also sometimes being used by companies as a way of publishing translated versions of large-volume, low-visibility content such as portions of knowledge bases for which a full-quality translation is cost-prohibitive.

Most often, however, raw output needs to be “post-edited” by a linguist to be correct and comprehensible. Post-editing poor raw output can be more time-consuming than translating from scratch, since the MT errors have to be understood and corrected before being reshaped into a correct translation. Most MT used as part of a professional service takes this approach. It is most frequently used in domains where the text is technical, instructional, well structured, and repetitive: for example, repair and maintenance procedures in automotive and IT sectors.

Sometimes source text is “pre-edited” to ensure that any ambiguities are removed according to tightly controlled language rules for terminology use and sentence structure. This ensures the MT engine produces the highest possible quality output, potentially with little or no post editing. This approach is mostly used with in-house solutions, where the individuals running the process have control over the source text and the translation process.

Lionbridge has invested in MT over the years and has experienced linguists on staff to ensure that any MT-based solution we implement achieves its objective; namely to provide an appropriate translation to meet our customers’ requirements for quality, timeliness, and cost.

## What you Should Know

- ➔ MT is the use of computer applications to translate from one language to another.

- ➔ MT is used by translation service providers as one element of a complete quality translation process.
- ➔ For good quality output, the MT engine requires customization by skilled linguists.
- ➔ Human intervention (post-editing) after MT is always required to achieve final-quality results.
- ➔ Pre-editing or writing in a Controlled Language specifically with MT in mind is also used to optimize the source language content for translation.
- ➔ Because of the customization required, larger projects are more suited to MT. Small sporadic projects unrelated to any previous customization are not good candidates. Not all language “pairs” are supported. Either they have not been developed for MT or they do not provide adequate output quality for professional purposes.
- ➔ Because it’s a machine doing the translation, MT is most suited to instructional or documentation-type material. Marketing and colloquial style projects require the kind of creative translation that only a human translator can provide.

---

## About Lionbridge

Lionbridge (Nasdaq: LIOX) is the leading provider of translation, localization, and testing services. Organizations in all industries rely on Lionbridge language and testing services to increase international market share, speed adoption of products and content, and ensure the integrity of their global brands. Based in Waltham, Mass., Lionbridge operates across 26 countries, and provides services under the Lionbridge and VeriTest® brands.

## Corporate Headquarters

Lionbridge  
1050 Winter Street  
Waltham, MA 02451  
USA

[www.lionbridge.com](http://www.lionbridge.com)

Contact Us: [marketing@lionbridge.com](mailto:marketing@lionbridge.com)