

Machine Translation: Saving Time, Lowering Costs, and Improving Service

Executive Summary

Commercial interest in machine translation (MT) is experiencing a significant rebound, though not necessarily because the quality of raw MT output has improved dramatically. Instead, a number of business, technology, and non-linguistic factors are combining to produce a resurgence of interest in MT.

During the last decade, product cycles have shortened across many industries, the volume of information has increased, and the pace of change has accelerated. At the same time, the rapid growth of informal electronic communications such as chat and email have changed expectations about the quality of textual matter, vast amounts of enterprise content have been made available in standard file formats, and MT systems have become more widely available. Integrated translation environments are now available that allow MT to be integrated via workflow automation with translation memory, terminology management, and human translation, significantly improving output quality over MT alone. Finally, involving content authoring teams early in the content development process ensures the highest quality output, while reducing translation times by as much as 50% and producing up to 30% cost savings in some cases.

While MT output is not equivalent to human translation, it does add a range of options for saving time, reducing costs, and improving service to customers, partners, and employees. In recent years, a number of successes have been reported by enterprises applying MT in commercial operations.

This topic brief explains how factors increasing the utility and applicability of MT have opened the door for market adoption. It explores recent applications of MT, and outlines specific criteria to use when evaluating enterprise MT initiatives. Understanding why, when, and where MT is most applicable helps ensure successful deployments.

MT Momentum

In recent years, MT systems have become more accessible to a range of users and systems. In addition, the wholesale digitization of enterprise documents means that content is now available in machine-understandable formats, which are increasingly standardized.

A number of other factors, including shorter product cycles, rapidly increasing information volumes, an expanding list of languages and locales to reach, Internet penetration, and changing expectations about the quality of text have also contributed to elevating the utility of MT and heightened its applicability in enterprise environments. These factors, and a growing trend toward globalization, have opened the way for broader adoption of MT solutions.

MT technologies

The technologies of most MT applications have their roots in government- and university-funded projects from the 1950s, '60s and '70s. These "rule based" systems process translation requests using linguistic rules — rules that are necessarily complex and inter-

reliant. Rule-based MT engines are particularly useful in situations where the input adheres to strict style and terminology criteria.

Newer approaches that use statistics (Statistical MT, or SMT) and examples of existing translations (Example Based Machine Translation, or EBMT) have provided improvements. These systems are “trained” using large collections of human translated documents (corpora), which the translation engine uses to search for parallel phrases across languages and ranks them by probability; the larger the corpus, the better the results. MT vendors are also developing hybrid approaches that use both linguistic rules and statistical methods to deliver higher quality translations.

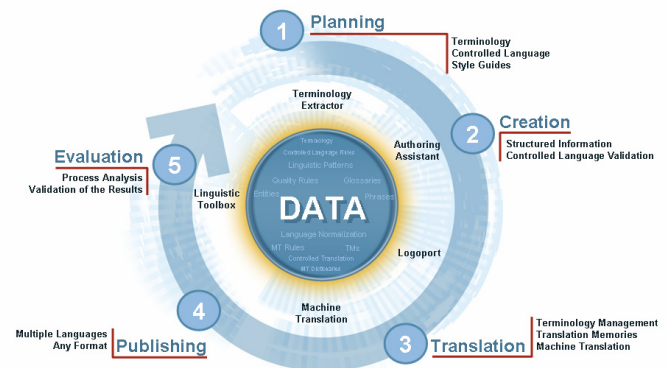
Combining translation memory technology with MT can improve the quality of the MT output; however, translation memory relies on human translators. MT output quality is also improved by customizing terminology and building pre- and post-processing scripts to handle known issues with terminology, usage, and grammar. Enterprises seeking optimal quality and efficiency run MT in conjunction with ongoing human translation projects, incorporating translation memory, terminology management, and pre- and post-processing scripts when possible.

Integrated language management environments, such as Lionbridge’s global translation management platform Freeway,™ support the integration of MT with terminology management and translation memory programs. They also support integration with workflow management systems to automate the movement of content from one application to another. This integrated approach has a significant and positive impact on the utility and applicability of MT in a wide range of commercial situations.

Time-to-market

Time-to-market expectations are changing, not only in the hardware and software markets, but also in industries such as automotive and consumer electronics. The window of commercial value for product-related content is shrinking rapidly as product delivery cycles get shorter.

As an example, consider an office products manufacturer producing new laser printer models each year. For each model, the software, online help system, and user documentation must be localized prior to the printer’s release. Although unedited MT output would not be acceptable in this instance, using MT to shorten the translation cycle is advantageous in speeding time-to-market, even if the total cost (with post-editing) is equivalent to human translation.



The Lionbridge Full Content Lifecycle includes MT, translation memories, and terminology management in conjunction with human translation

Volume and change

The volume of text, particularly in the area of customer and technical support, has grown to such an extent that companies are forced to use a selective approach to translation. As content production has become more efficient, the pace at which new content is added and existing content is modified has accelerated. Due to cost

considerations, most companies supporting products internationally can translate only a subset of this content into only a subset of the languages desired. In these cases, the availability of MT is the difference between having localized customer information available, or not.

Changing quality expectations

During the last decade, as the volume of information has increased and the pace of exchange has accelerated, user expectations about the quality of textual matter have changed. Email, chat, and other informal, text-based communications tend to contain typographic errors, abbreviations, omissions, non-standard grammar, etc. Communications of this quality sent by letter or fax would have been considered offensive and unacceptable only a few years ago. Today, however, many users are willing to concede some quality in exchange for translations that provide the gist of the message.

Globalization

An accelerating trend toward globalization is also contributing to the growing appeal of MT. Continued expansion of the Internet into the non-English speaking world is a good example. Free MT translation engines, such as Google's language tools and Yahoo! Babel Fish, have sprung up to meet the need for instant translation of Web content that would be too costly or take too long to translate manually. Many corporate Web sites include links to one or another of these free translation tools. MT availability is now expanding beyond just Web pages—Google also translates search queries, and Google Chat can now machine translate instant messages. Taking this a step further, social networking sites such as Facebook are beginning to actively employ a tactic called “crowdsourcing,” which uses people from the user community to improve the translation output. Lionbridge also uses the community approach to post-edit MT output.

Not only do a wide range of companies sell globally, but many manufacture their products globally as well. Ford

Motor Company writes fresh assembly instructions for every new vehicle model and updates these instructions as many as four times during each model year. The instructions are written in English and fed into their translation systems, which use a combination of MT and some human input to convert them into Spanish, German, Portuguese, Dutch, and Turkish. Ford plans to extend the use of MT into more languages in the near future.¹

These are just some of the many examples showing a globalization trend that is driving broader adoption of MT; a trend that we predict will only grow over the next 3-5 years, independent of any gains in translation quality. New factors will add traction to this trend, including integrated translation environments that allow MT to be easily and effectively integrated via workflow automation with human translation, terminology management, and translation memory technologies.

Applying MT

A general misconception persists that MT is meant to replace human translators. MT output is not equivalent to human translation, and MT does not replace the value of a human translator. But it does add a range of options that enable you to:

Reduce cost when the expense of human translation is prohibitive, but rough translation is better than no translation

Save time when local language content is required but human translation would take too long to meet market needs

Improve service to your customers, partners, or employees when human translation cannot meet the demands of content volume or rate of change

¹ Peter Loftus, “How do you say...” *Wall Street Journal*, October 22, 2007

Reducing Cost

To date, no company has reliably produced raw machine translation output of the same quality as professional human translation and we don't see this changing in the foreseeable future. When post-editing is used to bring MT quality into parity with the gold standard of human translation, the total cost can rise to the point of rough equivalence. So, although you can obtain savings of 20 to 40% using MT, depending on the type of text, the increased fixed costs associated with implementation and maintenance can reduce or even eliminate these savings.

This means that cost savings alone may not justify the exclusive use of MT when quality requirements are high. However, you can achieve cost savings when quality is not the primary consideration. Three approaches to MT are examined below, each of which can bring you different levels of cost savings.

MT with minimal customization

MT “out of the box” or with a limited amount of customization will provide you with the greatest cost savings. Customization involves linguistic coding of industry and company-specific terminology, as well as building “scripts” to clean up both the input and the output.

Using MT with minimal customization means that output will be uneven in quality, producing results that vary from near perfect to unintelligible. This approach is best suited to applications involving internal communications, research projects, or sifting through volumes of information for content worthy of human translation.

MT with customization and ongoing maintenance

MT can become moderately reliable over time within a specific subject domain. To achieve reliability, however, you will need significant linguistic resources to

customize and maintain or “train” the system on an ongoing basis.

With sufficient volume of content, you can achieve substantial savings with this approach. Nevertheless, rogue errors and anomalies are unavoidable without human post-editing. Using this approach will produce results that are adequate for general audiences, especially if translated content is needed quickly and the alternative would be to have no translation at all.

MT with customization, ongoing maintenance, and post-editing

Post-editing MT output to match the quality of human translation will reduce your cost savings. Fortunately, you can still achieve levels of quality ranging from satisfactory to optimal while preserving a cost benefit, assuming sufficient volume. Generally speaking, these levels of quality can be characterized by the following post-editing tasks:

Correcting word choice: you can achieve satisfactory quality by removing the obvious MT blunders to ensure the absence of misinformation.

Correcting grammar: you can achieve better quality by smoothing out grammatical anomalies and further improving word choice.

Correcting style: you can achieve optimal quality by editing to improve the flow and the voice of the writing, ultimately reaching the polish of professional writing. At this final level, however, the translation cost can easily meet or exceed the cost of human translation.

Saving Time

In any of the MT approaches discussed above, the time required to produce results is less than human translation – in some cases significantly less. When saving time is the primary objective, using an MT approach is a good bet.

MT for short-lived content

For material that is valuable for a brief time, but is worthless within an hour or a day, an MT approach without post-editing is not only appropriate, it might be the only way to make translated content available. Examples of this content type include near-real time communication, news reports, and financial data.

MT to improve time-to-market

In cases where time-to-market is crucial due to shortened product life cycles, the time saved by MT adds “shelf life,” thereby increasing total revenue capture per product release.

While working on customer projects, Lionbridge has gained important experience assisting product companies in accelerating time-to-market. Some examples of applications where MT has been particularly useful include:

Information Technology: technical articles, help systems, documentation

Automotive: technical drawings, instructions for manufacturing

Aerospace: aircraft maintenance manuals

Consumer electronics: product documentation

These projects and others like them have shown that MT with customization, ongoing maintenance, and post-editing can significantly accelerate time-to-market. Cost savings may be less than expected in some cases. Even though post-editing of MT output is less expensive than human translation, the costs of implementing and maintaining the MT system can reduce overall cost savings. But when time-to-market issues outweigh cost concerns, MT can play a key role.

You can achieve further advantages by integrating MT with other tools and processes. In many MT projects, there is minimal connection between content creation and translation. Involving your content authoring teams

THE WALL STREET JOURNAL

MT in Action - Microsoft

In what may be one of the largest commercial MT deployments in the IT industry, Lionbridge worked with Microsoft Corp. on a project that used MT, together with translation memory, workflow automation, and human translators, to translate approximately 15 million words of documentation connected to Microsoft’s 2005 Visual Studio product suite. The project involved highly technical content, including an electronic user manual and other content kept on Microsoft’s Web site or included on CDs.

About 60% of the material was handled by translation memory software because it had appeared in earlier translated versions of Visual Studio documentation. The rest of the material was run through machine translation by Lionbridge, and verified by Lionbridge’s human translators.

This was the first time machine translation was added to Microsoft’s translation process—previous translations had been handled through a combination of translation memory and human translation. Microsoft opted to try this new process to drive productivity, drive down costs, and come out with a quality product in the end.

Not only was Microsoft able to meet its time-to-market requirements, it also achieved the necessary high quality and saved an estimated 6 to 8% overall on this multi-million-dollar project by making human translators more productive.

Source: Peter Loftus, “How do you say...” *Wall Street Journal*, October 22, 2007

early in the process, identifying translatability issues early, translating and post-editing frequently used text segments up front, and building feedback loops into an iterative translation process can significantly improve the quality of the resulting translation. Using this advanced model, Lionbridge has found that translation

time can be reduced by 25% to 50%, while producing high-quality translation output, with up to 30% cost savings in many cases.

Improving Service

Often, the sheer volume and change rate of content are too high to allow human translation, since neither the budget nor the resources are available. In this instance, there are two types of automated process for producing MT output that can deliver the message, even if it is not perfect:

Batch-mode MT: where the content is translated wholesale and made available to the users without further post-editing. In this case, we recommend providing a warning to users, indicating that the content has been processed using machine translation. This will help set appropriate expectations and avert damage to the corporate brand.

On-demand MT: enables users to request a translation online, and get an immediate result. When users need to access content, but find it is in a language they do not understand, they can make their own choice to use the MT service.

Conclusion

A number of factors are driving renewed interest in machine translation, including time-to-market pressures, increasing volumes of machine-readable content, accelerated change rates, and a growing trend toward globalization. These factors are driving an insatiable demand for translated content that simply cannot be met by human translation alone.

Machine translation does not replace human translation. Instead, MT creates a range of options for achieving different business goals, including:

- reducing costs
- accelerating time-to-market
- improving services to customers

Businesses need to evaluate their translation needs and goals to ensure that MT – in conjunction with other tools and processes, or alone – is applied where it can provide the most benefit.

Lionbridge's experience has shown that MT in combination with other technologies, such as translation memory, terminology management, workflow automation, and human pre- and post-editing, can significantly improve on the benefits of MT.

Furthermore, successful large-scale MT projects have shown that the most significant benefits can be achieved by making MT part of an integrated, iterative translation process – one that involves content authors from beginning to end and enables continuous improvement of the results.

About Lionbridge

Lionbridge Technologies, Inc. (Nasdaq: LIOX) is the leading provider of globalization and testing services. Lionbridge combines global resources with proven program management methodologies to serve as an outsource partner throughout a client's product and content lifecycle – from development to globalization, testing, and maintenance. Global organizations in all industries rely on Lionbridge services to increase international market share, speed adoption of global products and content, and enhance their return on enterprise applications and IT system investments. Based in Waltham, Mass., Lionbridge maintains solution centers in 26 countries and provides services under the Lionbridge® and VeriTest® brands.

Corporate Headquarters

Lionbridge
1050 Winter Street
Waltham, MA 02451
USA
www.lionbridge.com