Introduction to Machine Translation and the Post-Editing Paradigm Shift
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>03</td>
</tr>
<tr>
<td>What is machine translation?</td>
<td>05</td>
</tr>
<tr>
<td>Primer on MT plus post-editing</td>
<td>08</td>
</tr>
<tr>
<td>Post-editing best practices</td>
<td>10</td>
</tr>
<tr>
<td>Profile of a post-editing translator</td>
<td>12</td>
</tr>
<tr>
<td>The bottom-line benefits of machine translation</td>
<td>14</td>
</tr>
<tr>
<td>MT plus post-editing adoption success stories</td>
<td>16</td>
</tr>
<tr>
<td>Getting started with MT</td>
<td>20</td>
</tr>
<tr>
<td>How to define your MT business requirements</td>
<td>21</td>
</tr>
<tr>
<td>Using MT for unconventional content types</td>
<td>23</td>
</tr>
</tbody>
</table>
In 2013 we are all part of the global economy. The last 20 years have seen huge advances in international trade meaning it is now far easier to do business in new markets. Today, customers want to buy products across numerous countries, but via the same identical purchasing process. This change has been accelerated by the proliferation of the internet, ecommerce and mobile commerce. Every year we sell more and more online.

US mCommerce* sales are expected to hit $33bn in 2015

* M-commerce (mobile commerce) is the buying and selling of goods and services through wireless handheld devices such as cellular telephones or tablets.
Introduction

Andy Reid, SDL

However, with new opportunities come new responsibilities. Selling could be just the start of the relationship with a customer and they will expect post-sales support, customer service and a continued presence in their market to maintain credibility. These are some of the drivers for the creation of new content in multiple languages, and it’s not a trend that will slow down any time soon.

Customers expect online and social interactions to occur in their native languages and this expectation will continue to increase. At SDL, we call this a Big Language problem.

‘Going Global’ is not just about translating marketing collateral, it’s about ensuring that all content is accessible to as many customers as possible, in a language they instantly understand. That includes web content, editorial, user-generated content and reviews, and knowledge bases. It is estimated that 2.5 quintillion (10^18) bytes of data are created every day, meaning we need more translated content than ever before -- and there simply aren’t enough translators in the world to keep up with demand.

This truly is a Big Language problem that can’t be helped with human resources alone. We have to look to technology to help.

*Machine translation* (also known as *automated translation* or *computer-generated translation*) is the key to keeping up with digitally-driven demand and staying ahead in the global game.

In 2012, only 0.000000067% of all content generated was translated.

Common Sense Advisory, 2011
What is machine translation?

The desire to get a machine to perform this complex human activity has a long history that has been spurred on at various stages during its development. In times of war, machine translation has been used to accelerate the translation of large volumes of monitored content. More recently the growth of the Internet and globalization has driven the need for more cost effective methods of translation.

As a result of continued development, new paradigms and an increase in computing power, machine translation now achieves a performance that enables it to be deployed commercially as part of a high-quality translation process.

For some, their only interaction with machine translation will be an application on a mobile. Others may experiment further with instant translation of chat sessions on social networks, or conversion of foreign web pages into a different language.

While these are common uses of what we call machine translation, it’s far more powerful in the hands of experts and its power reaches beyond personal use and into the corporate domain.
What is machine translation?

There are two main types of application for MT

1. **Where the MT output is used directly to get an understandable translation.**
   Typically this is short-life content where the meaning of the original can be understood from the MT output.

2. **As a technology tool that improves the efficiency of high-quality human translation.**
   The MT helps the human translator achieve a high quality output by providing an initial translation. The process is referred to as *MT plus post-editing.*

In both of these cases the performance of the MT is enhanced by training the engine for the content that is being translated so that the terminology and style are appropriate. This is especially important for technical documentation where the correct use of terminology is very important for understanding and for achieving high-quality publishable content. This additional training over and above the basic language training for MT engines is known as *customization* and is a big differentiator from the free online systems where one size fits all.

For many types of content, raw MT output is sufficient for purpose. It is capable of producing fluid, accurate and naturalistic output and for the purpose of ‘understanding’ this is more than sufficient. It is a cheap, fast and efficient way to access information that would otherwise remain outside your reach. But for some content types, the requirement will be for polished, publishable content that retains brand voice.
So the choice is simple - companies either choose MT or human translation to localize their content?

**A**

No. Today companies can also opt for a third option: a combination of the two known as **post-edited machine translation**.
Post-editing is the process of combining machine translation with skilled specialist human post-editors to produce high-quality translation and high publishing standards. Typically post-editors are linguists trained in the art of post-editing. Because they have detailed knowledge of the behavior of machine translation, they are trained to identify areas for improvement and/or correction in the raw output that improves the final quality of the documentation.

Post-edited translation can be achieved faster than human translation (HT) alone. As with any translation job, productivity depends on the source content. Different content types will naturally lend themselves to different translation approaches. If we take the example of consumer product, the marketing collateral for that product can be filled with emotion, historic context and cultural references. Often HT is the best approach to achieve the same degree of creativity in the target text. However, the accompanying specification and technical documentation lends itself well to MT with post-editing.
Primer on MT plus post-editing

Using a MT engine that has been customized for the content that is being translated makes a big difference to the efficiency of post-editing. This customization step is critical for optimizing performance for terminology-rich content like technical documentation.

Then you’ll have customer created product reviews; useful for other customers to read, but expensive to translate. These represent a potentially uncapped volume of information that has not originated from within your own organization. You are keen to know what is being said about your product for your own benefit and that of potential purchasers but the priority is ‘understanding’ not ‘publishing’. So for this third content type, raw MT output will often suffice.

It is important to apply MT and post-editing to the right content types for the best results.

$675MILLION

Post-editing MT now accounts for $675m of work for translation agencies.

Common Sense Advisory, Trends in Machine Translation, 2011
Post-editing best practices

The trained MT plus post-editing process has been in production for 10 years and is now in the mainstream of high-quality translation. It is becoming as ubiquitous as the use of translation memory and is a great example of how the combination of technology and human skills help to solve the demand for more and more global content.

But to improve the efficiency of high-quality translation, reduce costs and decrease time to market, a MT plus post-editing solution needs to be deployed in the right way. Here are a few things companies should consider when thinking of incorporating MT and post-editing into an existing localization process.

Integrated translation technology
For optimum performance it is important to take full advantage of other highly developed translation tools and add MT and post-editing to this mix. Translation tools have become very sophisticated and have many functions that ensure high productivity and quality. Translation memory, terminology management and QA tools are good examples.

MT plus post-editing works best if it is tightly integrated with a translation environment. In this way the benefits of MT and post-editing are added to an already efficient process. In a typical solution the new content for translation
Post-editing best practices

is leveraged against existing translation memories and any sentences not found in the memory, the “new words,” are automatically sent to the machine translation engine. The draft MT translations are presented in the translation interface along with the translation memory matches. All of the source language sentences now have a target translation ready for the post-editor to craft into full publishable quality text. This process is much faster than translating from scratch and the post-editor is supported by all the tools in the translation environment that ensure a high quality translation.

**MT engine training**
Content for translation can come from all sorts of business sectors that have distinct terminology. To ensure that the post-editors work is as efficient as possible it is important to train the MT engine for the content that is being translated. A one-size-fits-all approach to the MT engine does not deliver the best efficiency. There are two ways that this is accomplished; either through a vertical engine that is trained for a specific domain like Automotive or IT or by training a custom engine for a specific customer workflow.

New tools recently announced allow customers to take a straightforward do-it-yourself approach to engine training or they can call in the experts to build and fine tune a custom MT engine.
Profile of a post-editing translator

Today’s most common requirement is for high quality post-edited content, by which we mean the end product should transfer all information from the source correctly, using the correct spelling, grammar, syntax, and terminology. It should also be fluent as well as naturalistic and adhering to all project guidelines. This does not mean however that the post-edited translation will be exactly the same as a human translation, as the style may be slightly less refined in order to make the most efficient use of the MT output. To achieve all of this requires a particular type of person, one that has learned the art of post-editing.

Necessary ingredients
Creating high quality post-edited content is a skill, one that requires a few distinct traits; namely a positive attitude and a willingness to work with MT output from the outset.

Good overall knowledge of the project in question, the domain and terminology and CAT tools are also necessary skills that will help boost productivity.

And it is essential that everyone has realistic expectations - as the quality of the output will vary from project to project, file to file and even segment to segment. There will be near perfect translations and others which require considerably more post-editing effort, but it should all balance out over time.
The value of experience
With experience, post-editors will learn the MT output behaviors to watch out for and correct. By honing their skills, they will bring post-editing to the right level, ensuring they aren’t under-editing by making too few modifications or over-editing where they have made too many preferential changes, which will of course stop them being productive. You cannot teach post-editing as such (just as you cannot teach translating) - everyone needs to find out what method works best for them and use their own individual approach, but learning the basics and getting some tips initially will simplify matters and set you off on the right track.

It may be slower at the beginning as you learn the ropes, but will get faster over time and it isn’t uncommon to achieve throughput of 15,000 words a day. Post-editing is a skill that needs to be practiced in order to achieve proficiency so it really is a case of practice makes perfect.

The impact of post-edited MT
Using a highly trained system in combination with experienced post-editors, customers have seen a 140% productivity increase.
The bottom-line benefits of machine translation

By introducing MT into your translation processes four things can be achieved, almost instantly:

1. Ability to handle increased volumes of content
2. Affordable solution to translate previously inaccessible content
3. Reduced translation costs
4. Faster turnaround times
The bottom-line benefits of machine translation

The process of translation involves a common flow of work. Whether you use a translation agency or your own in-house team, eventually the document will land on the desk of a translator who will convert the text into the new language, i.e. the ‘target language’. The length of this task is directly related to the number of new words that need to be translated into the target language. Very often translation is charged per word for this reason.

MT gives you the opportunity to pre-translate some or all of your documents before sending for translation. This should give your company a head start as it allows the translator to return the final documents faster and therefore gives you scope to negotiate a lower per word price. However, care should be taken with this approach and it is important to discuss this with your agency first. A good partner will want to work with you to find the best solution for both of you. To do this, it is likely they will ask for a sample of your pre-translated files and then offer you a new rate for translating these texts. But they will of course need to be confident there is sufficient productivity gain with the MT output to allow them to post edit faster than human translation.

With a sensible and open approach, it is very likely that MT can be integrated into your workflow to quickly realize these benefits.

Opting to choose machine translation and post-editing instead of human translation when working with a translation agency could mean translating 150-200% more words for the same money.

Common Sense Advisory, Trends in Machine Translation, 2011
MT plus post-editing success stories

**Business requirements**
- Drive up self-service levels for dealer support
- Drive down cost and time-to-localize technical support knowledge
- Translate over 4 million words into 6 languages
- Integrate with Knova Service Resolution Management solution

**Results**
- Dealer success rate with self-service increased from less than 20% to over 80%
- Up to three-fold increase in global content throughput
- GIM costs reduced by 40%
- Integrated system for translation of different types of support content

Through our partnership with SDL, we provide support to our global dealers at 40% lower costs with significantly faster turnaround times. This provides solutions to their technical issues every hour of every day of the year, whilst significantly reducing the workload on support centres.

Joe Pstrak, Director of Dealer e-Business Services, CNH
MT plus post-editing success stories

**Business requirements**

- Improve quality and consistency of owner manuals for new European models
- Shorten production times
- Significantly lower translation costs
- Integrate with Tweddle Group’s systems, such as XyEnterprise for content management

**Results**

- Exceeded expectations with over 35% reduction in costs
- Achieved higher quality levels than had been achieved with traditional human translations
- Consistent and scalable process for service information distribution

"The Tweddle/SDL solution has exceeded our expectations for translation quality control, delivery and cost reduction. We now have a consistent and scalable process for the delivery of service documentation."

Paul Mansfield, Manager Service Information Distribution, Chrysler Group
MT plus post-editing success stories

**Business requirement**
- Provide large product range in local language
- Improve efficiency of providing global content
- Significantly lower translation costs
- Translate over 1 million words of content
- Integrate within an XML environment

**Results**
- Ability to simultaneously ship in global markets
- 27% cost savings
- 25% productivity improvement
- 35% reduction in time-to-market
- Automated, centralized process improves efficiency
- Improved consistency of global brand

RS Components – Products Catalogue

“SDL has provided us with a 25% improvement in productivity, as well as reduction in costs of 27%. It meets our tight deadlines and provides us with high-quality multilingual content, which enables us to support our new business strategy.”

Rob Fifield, Head of Media Production, RS Components
MT plus post-editing success stories

Best Western – Hotels Website

Business requirements
- Enhance customer experience by enabling local language interaction (web, bookings) for all 4,100+ hotels
- Reduce costs
- Integrate with reservations system
- Outsource complete business process
- Regular updates

Results
- All hotel websites synchronized in 7 languages
- Delivery time down from 24 months estimate to 8 months
- Saved $2m costs in year 1
- Increased online revenues

“The ability for our customers to learn about properties or book a reservation in their native language is an absolute necessity for our continued global expansion and growth.”

David Kong, President & CEO, Best Western
Getting started with MT

It’s easier than ever to get access to good MT solutions, but think carefully before taking your first steps. Here are your top three things to consider:

1. Identify your content types
   Look at what you need to get translated and into which languages. Estimate how many words there are. Work out how often the content will need updating.

2. Decide quality levels
   Think about each piece of content and decide whether it needs to be a literal translation, a localized version or simply a gist of the original.

3. Choose a MT solution
   Explore which solution best suits your needs. How important is data security to you? What integrates best with your existing technologies? Would you prefer to have access to the technology directly, or through a partner than can help you develop your use of MT?
How to define your MT business requirements

One thing is certain in modern machine translation solutions, it takes a huge amount of computing power. That’s why most MT solutions are available as cloud software. When choosing your MT partner, think about the following things.

Security
Cloud solutions are, by definition, Internet based and very often have many organizations using the technology at the same time. When you place your data, or your customer’s data, into a cloud-based MT environment, make sure you understand how protected that information is. What are the information policies of the MT provider? What claims does it make to the content that passes through its solution?

Customization
When you get started with MT, you will select the languages that suit you. At this point the translation engines are set up to give the best possible results across a broad range of needs and subject areas. These are known as baselines. However, to get the best results, you should be able to customize the results to suit your organization, customer, or subject area. To achieve this you will need what is known as a corpus of bilingual data. This generally is previously translated content that can be used to train the translation engine to give better results. In the past, this has been done by the MT partner. However, today’s
leading solutions offer self-training capability. With a little guidance, this enables your own team to customize a language to suit you.

**Terminology**
Protecting brand and maintaining the correct corporate tone is essential for consistent global communication. Choose an MT solution that offers *brand* and *term dictionaries* to ensure that specific words, such as a company name or marketing tagline are protected from automatic translation.

**Languages**
Whichever solution you choose, make sure that you can have a good open dialog with your technology provider. Machine translation is an exciting breakthrough and the most effective relationships will be built when you work together.
Four ways to leverage MT for unconventional content types

MT can be used in many ways across your organization. Language specialists will immediately see the benefits from the perspective of bringing down the costs and time it takes to translate content. However, it is also possible to build a platform for broader use.

Many of the Big Language challenges enterprises face today can be solved by applying MT to the following translation scenarios to build a more modern translation strategy.

Some other uses of MT:

User-generated content
If you have a website or a community that enables your users to post product reviews or forum posts, it’s quite likely that you have a wealth of information building up. Using MT, you can translate this content in regular batches, making it available in the native language of other users.
Four ways to leverage MT for unconventional content types

**Live chat**
That global reach of a contact center, be it customer care or sales, is often limited by the languages spoken by the individual agents. However, when you integrate instant translation of live chat into your customer care operation, your agents will be able to get the gist of a problem without knowing the language of the customer.

**Secure ad-hoc translation**
How many times a day do your colleagues translate confidential corporate information using online translation engines that have not been approved for use? By selecting a robust commercial MT solution, you can offer employees a secure URL to visit where private content can be translated without the risk of posting important intellectual property through uncontrolled internet solutions.

**Global internal communication**
If your organization is spread across many different regions, cultures and languages, it may be necessary to introduce MT to help your teams collaborate. With a dedicated solution focused on your business, you can ensure that communication is tailored to your corporate language and secure.
Want to learn more about how you can leverage machine translation within your organization?

Read our industry white paper, **Top 5 Big Language Problems Solved by Machine Translation**, to find out how you can quickly and affordably apply machine translation to the following areas of your business:

- Post-editing for improved translation capacity and turnaround time
- Global customer service and support
- Localized user-generated content, reviews and comments
- International employee collaboration, intranets, and wikis
- Translation security and protection of intellectual property
- Secure ad-hoc translation

Top 5 Big Language Business Problems Solved by Machine Translation

Introduction

“Anytime, anywhere” access to information is now the norm. Fueled by the volume, velocity and variety of this information, global companies now have a Big Language™ problem to solve in order to meet customer expectations and deliver a consistent customer experience across languages.

Case in point: Language barriers between companies and their global customers are stifling economic growth. In fact, forty-nine percent of executives say a language barrier has stood in the way of a major international business deal. Nearly two-thirds (64 percent) of those same executives said language barriers are making it difficult to gain a foothold in international markets. Whether inside or outside your company, your global audiences prefer to read in their native languages. It speeds efficiency, increases receptivity, and allows for easier processing of concepts.

To solve the Big Language™ problem and expand your global and multilingual footprint, there are opportunities to leverage both human translation and machine translation. There are ideal business uses for each, as well as opportunities to use them in combination, to maximize and balance efficiency and quality based on target audience, purpose of the content, and budget.

**Human translation** is best for content that is legally binding and high value, branded content. However, human translation can be costly, can take weeks (or even months) to complete, and can’t address all of the real-time needs of your business to serve multilingual prospects, partners and customers.

**Machine translation** is fast becoming an essential complement to human translation efforts. It is well suited for use as part of a human translation process, but also solves high-volume and real-time Big Language™ challenges that human translation cannot on its own — including the five that are the focus of this white paper.
About SDL
SDL enables global businesses to enrich their customers’ experience through the entire customer journey. SDL’s technology and services help brands to predict what their customers want and engage with them across multiple languages, cultures, channels and devices. SDL has over 1,500 enterprise customers, 400 partners and a global infrastructure of 70 offices in 38 countries. 42 out of the top 50 brands work with SDL. For more information, visit www.sdl.com.