

---

# When does the machine pay off?

Technical editing routinely commissions high volumes of translations. The perception that often prevails: Translating technical documents is a cost driver. Good reason to switch to machine translation?

---

Text by Rachel Herwitz



The amount of content that must be translated has increased considerably. While it used to be sufficient to translate technical documents into a few core languages, today they must be provided in the language of each target country. The number of languages companies have to translate into just increased with the Product Liability Act coming into effect. Corporate websites, too, are now expected to be presented in the language of the visitor's country. Moreover, content like FAQs, blogs, and chatbots must also be translated – ideally in real time. A human translator will often not be able to meet time and budget requirements. This is why machine translation and post-editing have become more important than ever. Prior optimization of machine translation

can reduce the post-editing workload, by carefully selecting the translation engines depending on the language combination, by customizing the system, and by integrating terminology.

## Economic considerations

Studies show that Neural Machine Translation (NMT) allows companies to reduce costs by 40 to 60 percent – whether it is used as an alternative to or complement of a translation memory system (TMS). While a human translator processes about 2,000 words a day, the recurring use of translations in the TMS allows for an output of up

to 5,000 words per day. By interfacing with a machine translation system, output can be increased to 7,000 words per day. However, post-editing machine-translated texts is not very popular with translators, as it usually pays less. Conceived as an “augmented translation” – i.e. human translation supported by a machine, translation memories, terminology databases, and quality control functions – translating becomes a task that requires extensive use of technology and the translator's full concentration in the process – a kind of “translation plus”. Moreover, customizable machine translation engines are not always easy to train. The training material is either missing completely or must undergo a complex

MT systems: a comparison				
NMT system	DeepL Pro (advanced version)	Google AutoML Translation	eTranslation (EU Commission)	Language Weaver (RWS)
Generic	Yes (Linguee stock)	Yes	Yes (EU texts)	Yes
Customizable	No	Yes	No	With corporate license only
Bilingual terminology can be integrated during translation	2,000 glossaries with 5,000 entries each	No (in Google AutoML as additional training material only) Yes (in Translation API Advanced)	No	Yes
Language combinations	24 source languages can be combined with 26 target languages	50 language pairs; English must be target or source language	From and into all EU languages	58 languages and language varieties
Costs	19.99 EUR per month/user/20 documents	45 USD per hour of training (300 USD max per training session; up to 500,000 characters per user/month/account for free; 80 USD for every additional million characters (up to 250 million, then tiered pricing), document translation: 0.25 USD per page	Free for registered users	Up to 500,000 characters per user/month/account for free for translators using Trados Studio (generic engines)
DSGVO	Headquartered in Cologne, Germany; no transfer of ownership; no further use of data; no data transfer to third-party countries	Headquartered in the USA; no transfer of ownership; no further use of data; no data transfer to third-party countries	Note: Do not upload confidential data; no data transfer	Comprehensive data protection with company license only

Table 1: Comparing different types of NMT systems

Source: Rachel Herwartz

cleaning process first. The training effort thus partly neutralizes the anticipated cost reductions.

## A system and its costs

How much a company must invest in machine translation varies widely. Table 1 shows four types of systems, each representing a different approach.

### Generic systems (example: DeepL):

DeepL is sold by the company of the same name, which has also developed the free translation database “Linguee”. DeepL was trained with bilingual sentence pairs stored in this database. DeepL is a generic system: It cannot be customized to a specific corporate language. A good feature is the new glossary function, as it allows company-specific terminology listed in a glossary to be used in a translation. However, the feature is only available in the paid version, DeepL Pro. The free version provides a single glossary, which is really only suitable for testing purposes.

DeepL is headquartered in Germany and subject to the German Data Protection Act (DSGVO). Texts and glossaries remain in the possession of their owners. They are neither used for training the system nor transferred to a third-party country. The paid versions provided usually offer better data protection than those freely available.

### Customizable systems (example:

**Google AutoML):** Unlike DeepL, AutoML can be trained with company-specific data. A data inventory of millions of records – the so-called “baseline” – is permanently available. Company-specific data in the form of bilingual sentence pairs from a translation memory can be incorporated into the training. Depending on the provider and the license, terminology can only be used as training material. Companies headquartered in the United States can also commit to renouncing ownership of uploaded texts, not adopting them as training material, and not imparting them to third parties. However, being U.S.-based, English must

always be either the source or target language.

The option to train the system with proprietary data may involve additional expenses, such as costs for training a single MT engine (one source language into one target language), hosting the system, and translating with the trained engine.

### Free MT systems (example: eTranslation):

The MT system of the European Commission can be used for free. It is a generic system, trained with texts from the EU Commission, which machine-translates from and into all EU languages. It is particularly suitable for legal documents.

### MT in translation memory systems

(example: **Language Weaver**): Developers of translation memory systems also provide MT solutions. The advantage for translators: Machine translation is integrated into the TMS environment and can be activated for free. Interfaces to external MT systems must usually be paid for. Also, not every producer of translation memory systems supports every MT system.

## Post-editing costs

If machine translation is used, the translation process needs to be redesigned. This is an issue that provokes discussions. Some

consulting agencies think it is possible to machine-translate technical documentation without post-editing. Translation agencies disapprove of this option, as the output quality varies significantly between different types of text as well as between the selected target and source language. Table 2 shows how different service providers work with machine translation. Some service providers offer an initial check of the input material. They determine if it is suitable for machine translation, or if the costs of full post-editing exceed the economic benefit of MT (A and B). The younger generation of translation agencies often favors machine translation combined with post-editing (C and D).

Other agencies offer a combination of translations from a quality-controlled translation memory system and translations generated by MT engines. The latter are rearranged for every new job (E), or trained once only with the translation memory of a specific industry, e.g., mechanical engineering (F).

## Calculation methods

The compensation of post-editing services causes disagreement between agencies and translators. For the latter, post-editing work is like copy-editing or proofreading an already translated text, which is usually paid

<p><b>Provider A:</b></p> <ul style="list-style-type: none"> <li>Optionally checks if MT can be applied (check by human translator)</li> </ul>	<p><b>Provider C:</b></p> <ul style="list-style-type: none"> <li>Always with MT</li> <li>Proprietary MT engine</li> <li>Proprietary UI for post-editing</li> </ul>	<p><b>Provider E:</b></p> <ul style="list-style-type: none"> <li>Uses TMS and MT</li> <li>Selects commercial MT engine for each translation project</li> </ul>
<p><b>Provider B:</b></p> <ul style="list-style-type: none"> <li>Checks, for every translation job, if MT can be applied (check with automated score)</li> </ul>	<p><b>Provider D:</b></p> <ul style="list-style-type: none"> <li>Always with MT</li> <li>Customizes MT engine for each client</li> </ul>	<p><b>Provider F:</b></p> <ul style="list-style-type: none"> <li>Uses TMS and MT</li> <li>Proprietary MT engine, trained with anonymized, industry-specific client data</li> </ul>

Table 2: Six different approaches to machine translation

Source: Rachel Herwartz

at an hourly rate. They claim that this type of compensation matches the particular attention required to detect errors in the fluently readable output of neural machine translations.

For agencies using a billing system, it is much easier to treat the post-editing of a sentence translated by a machine the same as the post-editing of translation memory hits. The TMS produces hits based on matching existing input sentences. These range from “no match” – requiring a new translation – to a “fuzzy 70 percent match” – with 30 percent of the sentence to be adapted – to a “100 percent match” – where the complete sentence may be adopted.

If translators adapt a 70 percent match from a translation memory that they created and quality-managed themselves, a compensation of 30 percent of the word price may seem appropriate. However, if it is standard practice to treat a translation suggested by a machine the same as a 70-percent match from a translation memory – and thus only compensate 30 percent –, the translator will not be happy. This issue is only amplified when the MT system used as the basis is a grab bag of language combinations, industries, and text types, where the translator can never be sure what will come out of it. Attempts to solve this conflict focus on measuring the post-editing distance and the time spent. To calculate the post-editing distance, the machine-translated sentence is automatically checked against the post-edited sentence to determine how many characters the translator moved, deleted, or added. If he or she adopted the sentence without changing it, however, there may still have been work involved in checking beforehand. This post-editing effort can be measured using a plugin that is integrated into the translation system and records the time – which brings us back to hourly rates for post-editing services.

## Complete check vs partial check

What type of work was performed, respective to what degree of completion, also affects post-editing costs. Technical writing

Text type	Suited for MT/MT post-editing	Note
Marketing material (print/online)	Full post-editing	Commercial statements must be completely transcreated.
Technical documentation	Full post-editing	The engine may only be trained with consistent material to ensure that specific wordings are adhered to.
Catalogs	Full post-editing	Integrating terminology is of particular importance here.
Legal documents	Full post-editing	Ideally performed with engines that were trained with legal documents.
Blogs/FAQs	With/without light post-editing	
Chats	Without/with light post-editing	

Table 3: Text types and their suitability for machine translation

Source: Rachel Herwartz

in particular calls for full post-editing to meet legal requirements. The translation is expected to have the quality of a human translation. “Good enough” results from light post-editing, let alone the complete adoption of a machine translation, are not acceptable for safety-related sections of a manual. Table 3 shows which combination is suited for which context.

## Depending on service providers

Transparent communication between client and service provider is very important. If good translation memories are available, it does not make sense to have everything retranslated by a machine and checked again. Machine translation can complement the process here, used for segments that have not been translated before.

The question that arises, however, is whether existing translation memories have consistent quality in all language combinations. If they do not, they do not qualify as training material.

When integrating a machine into the translation process, one should always keep in mind that this entails a commitment to the respective service provider to some extent. Changing the service provider means that the engines trained by that provider, as well as the adaptations and optimizations made, cannot be taken along as easily as, for example, translation memories exported in TMX format.

Also, some translation memory systems only support a limited number of MT engines. It really depends on the tools the translation service provider uses and the translators’ expertise and experience with those tools – particularly when it comes to post-editing.

## Target group terminology

A further cause of disagreement between client and translator or agency is terminology work – mostly unpaid and therefore not much loved. Clients expect translators to take care of the terminology “along the way”. After all, the search for equivalents in the target language is part of every translation. However, this is not automated. Unlike sentence pairs (translation units) that the TMS automatically saves to the translation memory during or after the translation or post-editing process, the terminology – including all necessary data categories – does not automatically end up in an underlying database. Maintaining such a database is time-consuming – and must be billable.

## Economies of the terminology database

Creating new data records of languages and their “sub-languages” (e.g., en-GB and en-US), as is done in translation memories, produces a multitude of duplicates in the terminology database. Most of the attributes must be maintained in parallel. This is a time-consuming task, which

Translation with TM	Translation with MT
70% cost reduction in technical documentation	Plus 20% cost reduction in technical documentation
10% cost reduction in marketing	Plus 60% cost reduction in marketing

Table 4: Cost benefits (example)

Source: Rachel Herwartz

some providers of language services take on just to make the terminology control feature available in a translation memory system. This time, however, would be better invested in the maintenance of relevant attributes. It is thus worthwhile for translators to engage in enriching terminological entries with definitions and discipline-specific meanings to account for homonyms, entering additional usage information, and gathering the creative neologisms produced by the engine. Moreover, the effort will pay off in future translation projects. Hopefully, translation memory systems will one day be able to handle attributes such as country code for quality control purposes.

## Costs and benefits

What should technical editors contracting language services pay attention to in order to be able to assess the cost-reducing potential of machine translation? Initially, costs and benefits can only be estimated. Whether the economies can be attributed to the MT or the translation memory really depends on the quality of the latter. In Table 4 it is assumed that, in technical documentation, the 70 percent cost reductions already incurred through very well-maintained translation memories in combination with standardized text production could actually be increased to 90 percent if machine translation was used. In marketing, on the other hand, where text is typically not reused and translation memory technology hardly ever applied, 60 percent savings can be achieved using MT and PE.

It must be considered carefully – and tested – whether a generic system with a glossary function or a customized system makes more sense. The test results, for that matter, always only relate to the selected MT system, the text type (and target group) tested, and the language combination. No general conclusions can be drawn from the results.

When estimating the costs, it is important to distinguish between one-time (Table 5) and recurring costs (Table 6). A constant reselection of the engine or frequent retraining, however, will have to be treated as recurring costs.

If a well-stocked termbase and pre-edited texts are available, the costs for further terminology work and pre-editing are lower than if a company starts from scratch.

**ACP TRADUCTERA**  
 Translation - Localization - Subtitling  
 tekomp-Messe 2022  
 Stand 2A30  
[www.traductera.com](http://www.traductera.com)  
 Your healthy partner for your success in  
 Central & Eastern Europe



One-time costs	
<b>Cleaning TMs</b>	Automated using scripts or tools (e.g., for removing tags, enriching data)
<b>Training engines</b>	Billed based on training time in case of customization
<b>Compiling training materials</b>	Trial texts, used to assess the quality of an MT engine (billed based on hourly rate)
<b>Interface between MT and TMS/MT and terminology database</b>	Billed based on hours used or one-time payment per interface
<b>Tools for measuring editing distance and duration</b>	Fixed price for plugins or proprietary scripts (programming)

Table 5: One-time costs

Source: Rachel Herwartz

In this case, the benefits of the terminology work and the standardization of text modules are not limited to translation but benefit other departments as well.

## Transparency for improved communication

With the growing demand for translations, MT has become a topic of conversation in technical editing departments. Still, there is no real dialogue between clients, agencies, and translators. An overview of the different costs involved in a translation process integrating machine translation can make cost discussions more transparent – and thus improve the communication between the parties involved. One should, however, not expect miracles. As with any technology, the success of MT depends on it being properly applied and wisely integrated into the overall process. As the market for providers of machine translation services and producers of TMS tools is very active, this requires cautious testing and thorough training of those involved in the process.

Recurring costs	
<b>Terminology work</b>	Billed on an hourly basis; to add new terminology or complete existing entries; can reduce the post-editing costs in all languages; benefits all company departments, not just translations
<b>Pre-editing</b>	Billed per hour or per document; can reduce the post-editing costs in all languages
<b>Post-editing</b>	Billed per hour, or based on PE distance, or treated like a fuzzy match
<b>Quality control</b>	Billed per hour; can be combined with quality control tools
<b>Continuous feedback</b>	Billed per hour; typical mistakes from the MT must be reported back to the training

Table 6: Recurring costs = costs per translation (pre-editing: costs per document to be translated)

Source: Rachel Herwartz

### ABOUT THE AUTHOR

#### Prof. Rachel Herwartz

advises clients on designing translation and terminology processes. She is CEO of TermSolutions GmbH, a provider of terminology software and consulting services. Prof. Herwartz is also head of the Translation Management program at the International University SDI Munich and is in charge of the university's online certificates in post-editing and machine translation.



@ rachel.herwartz@termsolutions.de  
 www.termsolutions.com