MT | Success Guide

Iconic
Translation Machines

We help the world's largest translation companies, information providers, and government organisations to adopt specialised, secure machine translation solutions of superior quality, tailored with subject matter expertise.

Our mission is to make machine translation practical for new industries and technical content types that are not traditionally accessible to language technology through our Ensemble Architecture™.

We are committed to providing reliable solutions that help our partners to achieve their strategic objectives.

Grow with us.

CONTENTS

MT Success Guide Introduction

Machine Translation (MT) is undoubtedly a complex technology. It's rare to have software that can work so well in some cases, and yet struggle in others. This obviously presents a challenge to potential end users and buyers of MT who will naturally be asking "how likely is it that MT will fulfil my particular needs?". It's a question that we've had to answer on countless occasions and now we would like to share some of our knowledge with you.

Our MT expertise has helped us to successfully assess the suitability of translation projects for MT. We have distilled this process down to 8 key factors that should ideally be taken into consideration before starting any MT project.

8 Key Factors for MT Success

- ⊕ Language
- 📄 Content Type
- 📊 Volume
- 🎖 Quality Expectations
- ⚙ Integration Requirements
- 🔤 Training Data
- 🗄 Translation Memory Leverage
- 🤝 Buyer Experience

Inside this guide, we'll introduce you to each of these 8 factors, what roles they play, what effects they have on the MT process, on each other, and on the end product.

LANGUAGE

How does language affect the feasibility of a MT project?

It's true that "not all languages are created equal". Some language pairs are more suitable for MT than others. Broadly speaking, the closer two languages are in terms of word order and grammatical structure, the more they lend themselves to MT.

What languages are suitable for MT?

Assuming English as a source or target language, we can separate language into 4 categories based on their suitability for MT (see sample chart below).

1. These languages lend themselves well to MT. They have moderate syntactic differences and ample resources, including corpora, syntactic parsers, and other linguistic tools.

2. These languages are linguistically challenging. They require additional specialised processing. However, they represent large markets, are well studied, and also have ample resources. Acceptable quality levels can be reached given sufficient time.

3. These languages are more linguistically challenging. They are less in demand, less studied, and have less resources available. MT success may depend on the other 7 factors in this guide.

4. These languages do not lend themselves well to MT. They often have extreme levels of morphology, no known origin from which to draw experience, or are rare and obscure. Successful MT projects are the exception rather than the rule.

++	+	-	--
Category 1	**Category 2**	**Category 3**	**Category 4**
French	German	Turkish	Finnish
Spanish	Chinese	Korean	Hungarian
Portuguese	Japanese	Thai	Basque

"Not all languages are created equal"

CONTENT TYPE

What is a content type?

Think of a content type as a genre of information, a broad classification of data, each with its own characteristics (e.g. Legal, Life Sciences, eDiscovery). Content types affect MT projects in a similar way to language, with certain content types being more suited to MT than others.

What makes one content type more suitable for MT than another?

Content that is relatively simple, with text written in a clear and consistent manner without complex terminology is ideal for MT (e.g. weather reporting). Technical and complex content types however are more challenging. Take patents with very long sentences, or user-generated content with ungrammatical phrases, for example.

How can I overcome the challenge of different content types?

It's vital to have different MT engines that are adapted not only to whatever the languages are but also to the specifics of the content type and the style in which it is written. This can require more investment in training data, engine build time, and especially additional customisations and expertise to address the various complexities presented by the content.

"Certain content types are more suited to MT than others"

1,2-oxy(ethane)

Hey, whatsup?

Technical Specification:

VOLUME
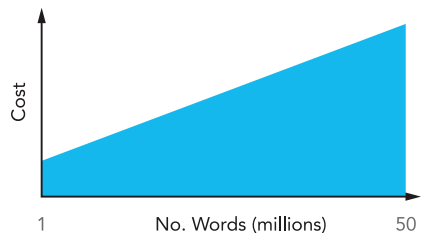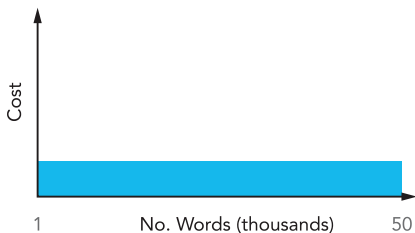
How is volume relevant to MT?

Volume refers to the number of words being translated. The capacity to translate many words quickly may be the number one reason for using MT. Although it's not strictly a like-for-like comparison, the average human translator can translate anywhere from 2,500 to 8,000 words a day, while MT can handle hundreds of thousands, if not millions of words. There are certain high-volume use cases where MT is simply the only option.

How does volume affect a MT project?

Volume can impact the cost and economic viability of any MT project. MT business models are typically driven by compute time. This means there might be no effective difference in cost between a 1,000 word translation job and a 25,000 word translation job because they may both be completed in the same compute time. Cost differences usually only start to kick-in as translation volume scales, say from 1 million words to 5 million words and upwards. Costs can also potentially vary depending on the speed requirements of the end user.

QUALITY EXPECTATIONS

Why is quality expectation management important?

The lack of expectation management has plagued MT for over a decade. The maxim "if it sounds too good to be true it probably is" applies here. If quality requirements are set in stone before assessing all of the factors of a MT project, it's a recipe for failure before you've even begun. Having realistic quality expectations is crucial to the successful deployment of any MT project.

What are realistic machine translation quality expectations?

Applying the other 7 factors in this guide will help you to explain, set and manage quality requirements. It's vital to work with your MT vendor to assess the quality requirements of your MT project in order to determine feasibility. Designing a suitable implementation plan, quality evaluation plan, and process for continuous improvement is key.

Remember MT doesn't aim to replicate human translation quality, but rather produce output that is fit for a particular purpose. Don't be misled by the initial output of any MT engine either, quality can be greatly increased over the course of a project by way of continuous training and enhancement.

Expectation	Fully automated human-quality MT output.
Reality	MT can produce fully automated output that is fit for a particular purpose and can be used as is.

Expectation	300% increase in translator productivity when post-editing MT output.
Reality	MT can deliver significant gains in productivity that can lead to cost savings and quicker project turnaround.

INTEGRATION REQUIREMENTS

How does integration affect a MT project?

Having high-quality MT engines is great, but you will not feel the full benefit unless they are efficiently integrated into your workflow. Fortunately, integration is an engineering task that can be achieved with sufficient time and skill. Integration doesn't affect MT feasibility, but the more complex the integration, the greater the lead-in time and cost in deploying your MT system.

What integration options exist for a MT project?

The most standard integration for Language Service Providers (LSPs) is via Computer-aided Translation (CAT) tool plugins, which come with most MT solutions and offer a fast, low-barrier to entry. For enterprise and other users, integration into a Content Management System or other software is usually through a MT provider's Application Programming Interface (API) for MT cloud services, or for an on-premise installation. Other options include web interfaces and FTP server integration.

Ultimately, integration will depend on how fast you need your MT service to be, how often you need to access it, and other considerations such as information confidentiality.

"The more complex the integration, the greater the lead-in time and cost"

	What Factors Impact on Cost/Lead Time	
Factor	Cost	Project Lead Time
Integration Requirements	Y	Y
Deployment Option	Y	Y
Increased Speed	Y	N

TRAINING DATA

What is training data?

Training data is any human translation of existing content. Translation Memories (TMs), bilingual glossaries, terminology lists, and dictionaries are all examples of training data. We cannot build MT engines without training data.
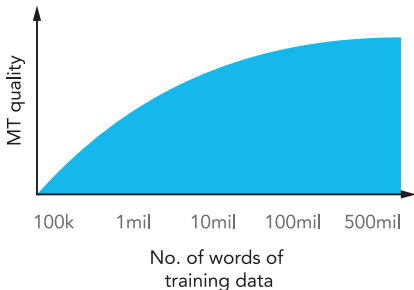
How much training data do you need to build an effective MT engine?

The more broad and complex the content type, the greater the need for training data. However, the law of diminishing returns applies when it comes to adding more training data to an engine. More does not always equal better. Data needs to be relevant and we need to know how to manipulate it. While MT providers often have "stock" engines for technical areas, this is never optimal. Project specific training data is crucial to increase chances of a successful MT project.

How does the level of training data affect the quality of MT output?

We can broadly rank the preferred scenarios of available training data into 4 levels:

1. The best case scenario, with lots of relevant project-specific training data.
2. Some project-specific data is available and supplemented with external data.
3. No data is available. Investment in data creation and content-specific development is required (and supplemented with external data).
4. No possibility to have project-specific data. "Stock" engines must be used. Your MT project may succeed in some cases but generally the quality will not be optimal.

TRANSLATION MEMORY LEVERAGE

What is a translation memory?
A Translation Memory (TM) is an example of training data and is a database that stores "segments", which can be sentences, paragraphs or sentence-like units (headings, titles or elements in a list) that have previously been translated, in order to aid human translators.

How does translation memory leverage affect a MT project?
TMs can greatly speed up the translation process by identifying commonly used phrases and suggesting translations through the use of Computer-aided Translation (CAT) tools. Language service providers often integrate MT into their CAT tools alongside TMs and terminology lists to speed up an end-to-end translation project.

If translators are getting high TM matches, then only a small percentage of the content is being sent to MT, and oftentimes this content is difficult for MT. For example, if only 8% of your content is going to MT, it might not be viable to deploy MT, i.e. the more leverage there is from the TM, the lower the effectiveness of the MT. Therefore, it is vital to assess the TM leverage of any content in advance of developing and integrating MT.

High TM
Leverage

Low MT
Effectiveness

"The more leverage there is from the TM, the lower the effectiveness of the MT!"

BUYER EXPERIENCE

Why is MT buyer experience important?

MT is a rapidly developing, disruptive technology. Integrating it into your workflow will always be a challenge, from pilot to production and beyond. Recognising where MT will be most effective is something that requires experience. The more experience you have in adopting MT — from evaluation, to integration, to training — the easier the process, and the greater the likelihood of success.

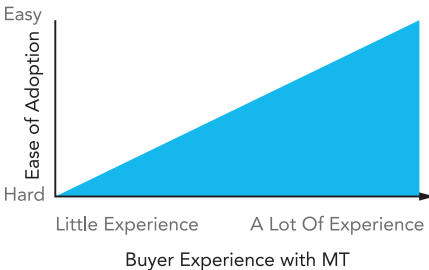## What impact can MT buyer experience have on a MT project?

For a first-time buyer the process will take more time. It may be challenging to get buy-in from users, so ensure they understand the types of errors that can be addressed and to what degree. Keeping users informed ensures easier adoption and increased acceptance. Well-designed evaluations are also vital to ensuring the quality is fit for purpose.

What if I have no MT experience?

MT is complex. Your MT vendor's support is key to success. At all stages, vendors should explain what the process is, from engine development, to evaluation, to deployment. It's also crucial to manage expectations and avoid promises of unrealistic quality levels.

Buyer Experience with MT

This guide is the first step to your MT success.

MT is a force multiplier for growth, enabling you to stay ahead of your competition.

Let us guide you along the way.

iconictranslation.com

Why choose Iconic?

We created the world's first patent-specific MT engine, and deliver custom solutions for specialist industries, including legal, life sciences, and financial services. It's MT with subject matter expertise.

Exceptional MT talent
Our expert team of MT PhDs and specialist engineers constantly innovate to deliver cutting edge MT software solutions.

Personalised service
Our clients enjoy working with us. We combine friendliness with a level of professionalism and responsiveness that goes above and beyond.

Better quality MT
Our proprietary Ensemble Architecture™ enables superior MT engines with a mix of statistics, rules, and linguistic engineering techniques adapted to suit each content type and language.

Expert MT guidance
Our experience can guide you on how best to adopt MT, from best practice and tailoring of engines to translation evaluations and technical support.

Long term, reliable partner
Our company is committed to providing reliable solutions in the long run. Grow with us.

For a more comprehensive guide, check out our MT Success blog series online now: www.iconictranslation.com/news

welocalize
doing things differently

"Iconic is a highly personable company. Their MT expertise is unquestionable. But what really makes them stand out is their capability to flexibly adapt solutions to make them fit just right, ensuring better quality MT."

Olga Beregovaya
Vice President, Language Tools, Welocalize

Iconic
Translation Machines

Need some expert MT guidance? Get in touch.

solutions@iconictranslation.com
www.iconictranslation.com