



## **Predictions on the Future of the Translation Industry**

## **Contents**

1	Introduction.....	3
2	Improving Translation Quality.....	3
2.1	Use of Dictionaries.....	3
2.2	Do Not Translate (DNT).....	3
2.3	Improving the translatability of the source language text .....	3
2.4	Not Translating already translated text.....	3
2.5	Handling of Unrecognized Words.....	4
3.1	User Dictionaries .....	4
3.2	Share Dictionaries .....	4
3.3	“Do Not Translate” lists.....	4
3.4	Managing Already Translated Text.....	4
3.5	Handling “Not Found Words”.....	4
3.6	Tips on improving the translatability of the source text .....	5
4	Measuring Improvements in Translation Quality .....	5
5	What of the Future?.....	5
6	Will Human Translators Become Redundant? .....	6
7	Conclusions.....	6

## 1 Introduction

Most people's experience of using machine translation (MT) to translate text from one language into another is likely to be by using either Google ([www.google.com](http://www.google.com)) to translate a foreign-language web page, or a service like Babel Fish ([www.babelfish.altavista.com](http://www.babelfish.altavista.com)) to translate a block of text from one language.

Using MT is becoming an increasingly common experience: today there are over three billion pages translated by machine from one language to another on the Internet every month. This is likely to increase as the number of people with English as their first language as a proportion on internet users continues to drop.

But if you have used one of these services, you'll know that the results are often far from perfect. These services generally convey the meaning of the message, but are rarely fluent and you certainly wouldn't want to use them to translate an important document from your language into another!

So if there's so much of it being done, why is the translation quality poor?

The reality is that neither Google nor Babel Fish are good examples of what MT can really achieve. There is nothing wrong with the underlying translation technology, but these implementations are simply doing a sentence by sentence translation without any additional processing.

There are a number of ways in which the quality of an automatic machine translation can be dramatically improved today. These include:

- Use of Dictionaries
- Identification of words and phrases that should not be translated
- Improving the translatability of the source language text
- Not translating already translated text.
- Handling of unrecognized words

## 2 Improving Translation Quality

### 2.1 Use of Dictionaries

These may also be known as glossaries or term bases. Dictionaries provide a list of terms, technical terms and other words and phrases, together with the appropriate translation. Establishing a dictionary and updating it over time has been shown to improve translation quality dramatically.

### 2.2 Do Not Translate (DNT)

One of the major problems of simply running an MT engine on a block of text is that there are some words that shouldn't be translated. For example: people's names and addresses. Indeed translating such words or phrases can alter the entire meaning of the sentence.

### 2.3 Improving the translatability of the source language text

In order to achieve optimum results using Machine Translation (MT), any document or email should be written using clear, concise, grammatically correct language.

### 2.4 Not Translating already translated text

One technique that is often applied is that of translating back into the original language. This is never a good idea because errors multiply. If the original translation has any errors, then re-translating in the other direction will multiply the errors and substantially reduce the quality of the second translation. This technique can never provide any indication of the quality of the original translation.

Unfortunately free web based translation services have been designed in such a way that this is very easy to do, as you can simply cut and paste the already translated text and press the translation button.

## **2.5 Handling of Unrecognized Words**

In almost any text translation there are going to be words that the MT engine doesn't recognize. They could be misspellings, or they could be particularly esoteric or uncommon words that haven't been included in the MT engine's dictionary. Without further processing these words simply do not get translated. Because MT engines translate a complete sentence, not translating words in the middle of a piece of text must detract from the quality and comprehensibility of the translated piece.

## **3 Translation Quality Tools**

Translution includes a number of tools in its machine translation software products designed to significantly improve translation quality. These tools are intended to be used by non linguists and give all users the ability to improve translation quality, specifically addressing the issues explored above.

### **3.1 User Dictionaries**

Everyone uses words and phrases that, to them, have particular meanings. For example: parts lists, where a specific translation is required for an individual part description. Translution users can easily set up their own user dictionaries, which control how the translation is handled.

### **3.2 Share Dictionaries**

Translution users can also share dictionaries across their whole organization. These dictionaries are available via the web, so that they are available to all users at any time, meaning that translation quality can be improved across your entire organization.

### **3.3 "Do Not Translate" lists**

Translution provides facilities for authors to mark up text within a specific document that they do not want translated. It also provides Do Not Translate lists i.e. words which the author or the organisation would never want translated.

Research undertaken by Professor Tony Hartley of the University of Leeds, shows that implementing Do Not Translate lists improves local translation accuracy by, on average, 23%. (Ref: Improving Machine Translation Quality with Automatic Named Entity Recognition, EAMT Workshop at EACL 2003 Budapest 13 April 2003).

### **3.4 Managing Already Translated Text**

It is essential that already translated text is not re-translated as this reduces overall translation quality. This is important, particularly when replying to or forwarding emails across language barriers. Translution ensures this does not happen by always including the original text with the translated text, and by identifying already translated text in email replies and forwarded messages.

### **3.5 Handling "Not Found Words"**

Machine Translation (MT) engines are usually supplied with dictionaries of the most common words used in a language. However no MT engine could possibly include all the words in a language, particularly "rich" languages such as English, French, German and Chinese. Until these words are translated and added to the dictionaries of MT engines held centrally, they will not be recognised by the translation engines and will simply pass through un-translated.

As part of its ongoing commitment to improve translation quality, Translution collects and analyses all Not Found Words generated by all users and, after researching them, continuously adds selected new words

and phrases to its dictionaries. In this way overall quality of translation will be improved for all users of Translution products.

### **3.6 Tips on improving the translatability of the source text**

Translution have gained huge experience and understanding of the most common issues that lead to reductions in translation quality. It publishes a “primer” – “translation tips” which identifies the most common issues. We also publish a white paper “Improving the Translatability of Text” which gives a more in depth analysis and guidance on how to improve the translatability of text. Finally we also offer a training course in “how to write for translation”

## **4 Measuring Improvements in Translation Quality**

It is easy to assert that a particular technique improves translation quality, but it is quite another thing to actually demonstrate and measure the degree of improvement.

To address this issue Translution worked with Professor Tony Hartley and his team at the Centre for Linguistic Studies at the University of Leeds. They established a means both comparing the quality of output of MT engines and comparing the quality of translations that result after other techniques have been applied. The evaluation method developed at Leeds University has provided Translution with an independent and objective means of systematically measuring improvements in translation quality as additional quality tools, techniques and content are provided.

Out of this work has come Translution’s 3 essential techniques for improving machine translation quality which is proving time after time with its customers that machine translation can deliver useful translation, even for technical documents.

The 3 techniques are simple:

- a) Examine your source document for translatability, and re-write if necessary.
- b) Mark up any text that you do not want translated such as names, places (i.e. Proper nouns).
- c) Create a dictionary “tuned” to your specific requirements.

## **5 What of the Future?**

MT technology will improve over the next few years as new MT engines are developed and released. Also the number of languages offered will increase.

This trend is already happening as we have seen with the recent developments from Google and the emergence of “open source” technologies such as Moses.

Because we have designed Translution to be “agnostic” to the machine translation engine used, Translution will be at the forefront of these developments.

Translution’s current technology relies on a “rules based” approach to translation (RBMT), whereas the newer technologies use an example based approach – broadly known as statistical machine translation (SMT).

Whilst SMT generally provides a better “gist” translation than RBMT, SMT engines can only be improved by training them with large quantities of human created “translation memory” specific to the “domain” of the translation. Most individuals and businesses simply don’t have access to these linguistic resources.

However, RBMT can easily be trained as all that is required is access to mono lingual data in order to build a dictionary. Once a dictionary has been created for an organization, RBMT will usually be much more accurate than “gist” or untrained SMT.

So our predictions are as follows:

- a) SMT or hybrid systems will continue to emerge from companies like Google and Asia online. Over time these will be the systems that will largely be used by individuals to understand the vast amounts of content in other languages on the internet.
- b) RBMT systems (with dictionaries and translation memory) will continue to be used by most businesses. Individuals will also continue to use RBMT systems (with dictionaries) for day to day communication.
- c) Global enterprises with access to large amounts of translation memory will adopt and customise SMT systems for specific purposes such as the translation of “knowledge bases.”

## **6 Will Human Translators Become Redundant?**

With all of this technological development, it has been suggested that one day we won't need human translators

We are confident that there will always be a role for the human translator; because humans are extremely adept at understanding context and extracting meaning from sentences. Machines will continue to struggle with this aspect of translation for some time yet.

While machine translations may achieve much improved translation accuracy, this may still never be good enough for some purposes such as translating legal documents and marketing literature. They will certainly never be good enough for works of literature!

Translation will continue to require human intervention – although the development of improved machine translations such as is provide by “trained” RBMT systems and customised SMT systems - will remove much of the “grunt” work.

The will mean that skilled translators will focus on post-editing and fine tuning the final translated work, rather than translating from scratch.

This doesn't devalue the work of translators, what it means is that the amount of human intervention to achieve a perfect translation will be reduced, so it is feasible, both because of reduced time and reduced cost, to translate many more items than are translated today. This actually increases the amount of opportunity for translators.

## **7 Conclusions**

As the quality of machine translation gets more accurate and closer to that of human translation, both the time to perform translations and the costs of doing so will fall significantly. This doesn't mean that the days of the human translator are numbered, what it means is that things that were never considered as candidates for translation in the past, such as emails and instant messages, can and will be capable of being automatically translated.

What it also means is that the types of work that translators do will change.



## Predictions on the Future of the Translation Industry

---

Today, email, and many previously untranslated documents, while not perfectly translated, can be delivered in the reader's native language and are immediately comprehensible and largely context-accurate. In many instances this will be good enough and will deliver significant business and cost benefits.

So the improvement in quality of automated translations can add substantially to the quality and value of both business and personal communications. Translution is at the forefront of this revolution in language communications.