



Improving the Quality of Machine Translations

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) to translate a foreign-language web page, or a service like Babel Fish (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 provide a general understanding, but are rarely fluent, often inaccurate 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 today's MT technologies can really achieve. There is nothing wrong with the underlying MT technology – and Translution use the same MT engines as these services, - but these implementations are simply doing a literal sentence by sentence translation. They do not provide an ability to add any additional information that, when used, can dramatically improve the translation quality.

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
- Using good grammar and spelling
- Not translating already translated text.
- Handling of unrecognized words
- Using Human Translations

Let's consider each of these in turn:

Use of Dictionaries

These may also be known as glossaries. Dictionaries provide a list of unusual 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

ⁱ Babel Fish is from the eponymous fish in the "Hitch Hiker's Guide to the Galaxy" by Douglas Adams. The Babel fish is placed in your ear and translates from any language by feeding on brainwave energy from those around you and excreting a telepathic matrix in a form you understand. The Babel fish is itself named after the biblical Tower of Babel where God intervened to prevent the people of the earth becoming more powerful than him. He scattered them across the earth and forced them to speak many different languages.



quality dramatically depending on the underlying MT engine and the number of terms included in the dictionary.

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, company names, and addresses. Indeed translating such words or phrases can alter the entire meaning of the sentence.

Using good grammar and spelling

In order to achieve optimum results using Machine Translation (MT), any document or email should be written using clear, concise, grammatically correct language. There is little difference between writing a document for MT and writing a document for human translation.

Not Translating already translated text

One technique that is often applied when people play with free web based translation services is that of translating back into the original language. This is never a good idea because errors multiply. If the original translation is 90% accurate then re-translating in the other direction will multiply the errors and substantially reduce the quality of the second translation. This technique cannot provide any indication of the quality of the original translation, due to the nature and complexity of language. 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.

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.

Using Translation Memory (TM)

Translation Memory (TM) is source and target human translations stored and aligned into segments (usually sentences) in a database. Using TM for both machine and human translation ensures consistent and perfect translations of "standard phrases". This approach can ensure a very high standard of translation such as proposals and support communication where "standard phrases" are used.

Storing, managing and using Translation Memory can also dramatically reduce the cost of translation by Translution's human translation teams

Translation Quality Tools

Translution includes a number of tools in its 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.



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. For organisations, Translution have developed a semi-automated approach which quickly and cost effectively identifies the majority of the specialist terms used within the organisation. Once translated, (for Translution Business users), these terms can be made available to the whole organisation to provide a consistent and accurate translation.

Domain and Corporate Dictionaries

Translution Corporate users can also develop their own Domain dictionaries (for specialist subject areas or departments) and Corporate Dictionaries (for the organisation as a whole) which determine how specific words are translated and which can then be shared throughout the organisation. These dictionaries are available via the web, so that they are available and can be worked on anywhere.

Translution's dictionaries are designed to be multi-lingual rather than bi-lingual. Using bi-lingual dictionaries, with just five languages, users would have to provide 20 entries (one for each language pair, in each direction). With multi-lingual dictionaries, only five entries need be provided. This considerably reduces the amount of coding that is required to develop dictionaries that can be used across a number of languages.

“Do Not Translate” lists

Translution provides a facility for the user to mark up text within a specific document or e-mail that they do not want translated. It also provides Do Not Translate lists i.e. words which the user or the organisation would never want translated. Translution's dictionary creation services automatically capture terms that should not be translated from existing contact databases.

Research undertaken by Professor Tony Hartley, Director of the Centre for Translation Studies at 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).

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. Translution also ensures already translated web pages and documents are not re-translated.

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 and German. 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.

Use of Translation Memory (TM)

Translution Business and Translution Corporate use stored Translation Memory (aligned human translations) to improve translation accuracy. Using TM for both machine and human translation ensures consistent and perfect translations of “standard phrases”..

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 have established a means both of comparing the quality of output of MT engines and comparing the quality of translations that result after other techniques have been applied. The evaluations proved that, properly used, MT provides a highly understandable and accurate translation which can be used for many communication purposes.

The evaluation method developed at Leeds University also provided Translution with an independent and objective means of systematically measuring improvements in translation quality as customers implement projects.

Translution believes that MT is “fit for purpose”, where properly applied. To demonstrate this, Translution has developed a service which – for a small returnable fee - takes a selection of customer’s documents and emails and demonstrates the quality before and after the use of the tools discussed above

What of the Future?

MT technology will improve dramatically over the next few years and Translution will be at the forefront of these developments.

Current technology relies on a “rules based” approach to translation. Whilst this provides translation quality which is “fit for purpose” in that “meaning” is largely communicated accurately, it lacks fluency in comparison to human translation quality. It therefore cannot replace human translation where a high quality of translation is required such as for web sites, legal documents and marketing material.

Translution has already evaluated a new approach to machine translation - statistical MT - as it holds the promise that dramatic improvements in the fluency of machine translation can be achieved when the system has been trained with customer’s Translation Memory (TM). However, today, for general purposes, this new technology does not deliver any improvement in translation over the rules based systems. Translution are however continuing to work closely with statistical MT suppliers, and when the benefits can be demonstrated, will integrate this new technology.



Translution believe that with the use of Translution's dictionary and TM technologies, one day we will be able to deliver consistent accuracy and improved fluency from statistical MT and even match first draft human translation for some purposes.

Because Translution has designed its systems to be able to take advantage of new emerging technologies, Translution's translation service will remain at the forefront of automated translation solutions.

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. And anyway the new statistical MT systems rely on Translation Memory – which can only be produced by qualified translators.

While machine translations may achieve very good translation accuracy, this may still never be good enough for many business, legal or contractual purposes. 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 will remove much of the “grunt” work, leaving skilled translators to focus on post-editing and fine tuning the final translated work. 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.

Conclusion

As the quality of machine translation gets ever 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 and accurately translated.

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.

The consequences of Translution's suite of products being available to individual consumers and businesses alike are yet to be fully identified but it is likely that:

- individuals will benefit from improved communications when arranging foreign travel and commercial transactions.
- businesses are likely to be able to do business with areas of the world, suppliers and customers which up to now unavailable to them due to language constraints.



- multi-national businesses will gain considerable benefit through enabling all employees and their wider community to collaborate and communicate without language barriers.