Machine translation versus human translation – An overview (long version)

PART 1 — MACHINE TRANSLATION

Most of us nowadays use an online translator like "Google Translate" when a translation is needed, but not everyone is aware of the fact that online systems, like Google's, do not translate as a human would. Usually they just "stitch together" small snippets of text from their massive database of parallel texts, such as professionally translated multilingual EU and UN documents, and apply some corrective measures to improve the result. That method has its weaknesses users have to be aware of. Please read on in order to get best results from online translators.

A valid new alternative to "Google Translate" for translating between the most common European languages is the new arrival "DeepL" from Germany. In most cases, this free online translator will yield better results than "Google Translate". It avoids some of the pitfalls present in Google, doing considerably better with complex text written in a good and clear style. However, just like Google, it translates certain expressions too literally, a fact that can render a translation useless in the worst case, and sometimes it will unfortunately "screw up" a sentence completely.

When comparing translations from identical inputs made by the above systems, it looks as if "Deep L" often uses a translation base very similar to Google's and improves it by applying superior vocabulary and syntax control. But that's speculation on my side, I have to confess. BING's translator, that of the Russian search engine "Yandex" and others in my experience produce less convincing results than the two above-mentioned systems.

So, knowing the do's and don'ts, especially "Deep L" sometimes yields perfect, even flawless results (with the stress on "sometimes"). But if you do things you'd better not do, the systems might not work as they should and produce, in the worst-case scenario, only a "salad" of poorly connected words. Take notice of the fact that the more similar your text is to a standard template or model text that others have already created before you, the higher the chances the system will translate it correctly! That even goes for current expressions in commercial correspondence, - so whenever you leave the "beaten path", you may be faced with undesired results.

In what concerns "Google Translate" and similar systems, they perform at their worst when running out of parallel material and having to enter a kind of "raw or direct mode". Fortunately, today that happens less often, but the risk is always there! Below I'll show you by means of a special test (*link on the Google and Co. page*) what might happen, when automatic translation loses its mind, so to speak.

But before here are some useful hints of how you should handle "Google Translate" or other online translation systems:

Use a straightforward and simple style, no contorted phraseology please. Whenever possible use the most common words describing a given object, situation, etc. The system will always look for material matching your own, and whenever it can't find any, it will improvise, - and it's doing very poorly at that compared to a human! Take my introductory paragraphs, they are a little different in their respective languages, because I wanted to introduce a little variation, but otherwise "Google Translate" or "DeepL" translate them nearly flawlessly! The main reason seems to be, that although I haven't copied the text from anywhere, my version is practically identical to what others had composed for the same purpose.

If however, you have zero knowledge of the language you want the system to translate into, use only the most simple and clear sentences you can think of! Whenever possible, avoid idiomatic phrases and words you might find a more straightforward synonym for. In my opinion, it is an undisputed fact that users themselves are partly to blame when translation systems deliver useless results!

Since we're talking about machine translation here, it will hardly surprise anyone that I fed this very text into the afore-mentioned systems to obtain the German and Italian versions, just to see how they're doing. That sometimes worked very well for certain paragraphs, fairly well for others, but many had to be completely rewritten, whenever the original text was more complex.

So please heed my advice about keeping your input material as clear and simple as possible in order to obtain good or at least acceptable results from any type of machine translation! Also avoid words with a double meaning, and if possible, replace them with synonyms of an unambiguous nature. Up to now, systems are often unable to decode the meaning of a phrase, let alone that of a whole paragraph. But even then "Google Translate", "DeepL" or similar systems may not find the right equivalent for a word or a group of words, or even "screw up" certain sections altogether, although no "error" has been made on the user side! (There might be data "residue" from similar but not identical texts, for instance.)

How badly Google performs when acting as an online dictionary many of you may have already experienced. Google search usually displays a translation preview window on top of search results, whenever you input an English word followed by "French", for instance. (Leave a little space between the words.) Now most of the time the displayed results are totally useless!

So the instant success rate of Google here in my opinion is around **20%**, compared to **80%** + for good online dictionaries (most are free anyway). Mainly that has to do with the fact that Google, besides having to detect the language of the entered word automatically, is hell-bent on offering us a <u>single choice</u> per entered word, a simplistic method that is bound to produce errors, however good the data-base may be (and the one Google uses leaves to be desired anyway!) Ample room for improvement here in my opinion for the market leader!

So, although I'm certainly not a specialist on machine translation, I devised a little trick to confuse the automatic translators in order to show you how <u>switching to raw or direct mode</u> looks in translated texts (when systems are left to their own devices, so to speak).

In order to achieve this, I used an uncommon language in which nevertheless many parallel texts exist. Additionally I expect some words of the parallel text to be marked, showing their origin! Completely out of this world? No. There is one single scenario in my opinion, making this possible.

My trick works by using **a.)** portions from the most widely read book, the bible, and more precisely from the Latin gospel of John, and **b.)** a text that is <u>very similar, but not identical</u> to the biblical narra-tive. It is also common for many translations of biblical texts to use older English forms like "thou" or "doth, sayest", etc. "Google Translate" and the Russian "Yandex" translator therefore inadvertently tag those words, <u>proving</u> <u>unequivocally</u> the systems **did** use segments of <u>ready text</u> available in their database!

The version I used as an alternative text or "agent provocateur", is a fourth or fifth century apocryphal text, somewhat incorrectly called the "Gospel of Nicodemus". Now, in its first part, which is better known as the "Acts of Pilate", it tells the story of Jesus before Rome's governor Pilate. Here, the text is often nearly identical to that found in the gospel of John, but it may suddenly switch to a version not found in the biblical narrative.

So for the identical portions Google and Yandex usually find a ready English translation complete with "tagged" words (sometimes this hinges on some ridiculously small but system-relevant detail); whereas for the divergent parts the systems have to enter "direct mode" which one might quite fittingly call "<u>disaster mode</u>", especially for Latin. My apologies to the people who programmed those systems!

Now you might rightly object and say: <u>Who really cares about translating from Latin</u> into English in this day and age?! Although that is certainly true, this little stratagem shows the inherent weaknesses of the systems very clearly, and it's precisely those weaknesses that can badly <u>mar your own trans-lation between modern tongues</u>. Also remember that Google and Russian Yandex do support Latin, whereas other online translators do not.

The examples I have used in the second part of my test, constitute very straightforward and simple short Latin texts (mostly dialogues) that have been taken from the same apocryphal text, employing a language already very close to our modern tongues. These portions <u>do not</u> contain any special Christian terminology. It will be seen though that most of the time the tested systems (Google and Yandex) do not even recognize the most simple forms of word declension and conjugation, so <u>whenever there are no parallel texts available</u>, they produce only a maze of garbled words.

As a control I added a moderately complex phrase from René Descartes' "Rules on the direction of the mind", which Google and Yandex totally fail to interpret correctly, although they do <u>infinitely better</u> in converting a previously translated version, be it from English into German, or vice versa, even when no parallel version had been found.

Something similar happens whenever Google or Yandex have to translate <u>from the Russian</u> into another Western-European language. That is because even modern Russian has a grammar that is very similar to Latin (at least with respect to case endings playing a great part in determining the sense of a given phrase.) The systems will therefore encounter <u>similar problems</u> they did when translating Latin texts, sometimes obliterating or inverting the meaning of a sentence completely. But recently "Google Translate", joined now by "DeepL", have made great progress when translating from the Russian.

So in concluding, my best advice to you is: Do <u>NOT</u> use automatic translation, if a) for whatever reason you're unable to check the result, or get it checked, and if b) there is a risk that a poor quality translation <u>may have a negative effect on your business or reputation</u>, or if c) your text uses <u>specialist or uncommon</u> terminology and/or phrasing.

When writing to friends or acquaintances, a few errors are not serious, since everybody uses automatic translation today. It might in the worst case cause some merry laughter, but that's not exactly what you would like your business partners to indulge in, is it?

PART 2 — HUMAN TRANSLATION

Now let's see how a human translator works.

1. He or she should understand the original text perfectly before translating it. Having to rely on external help to get to the meaning of a sentence is not a good start. There are some exceptions to this rule, of course, like the more complex sentence from René Descartes, mentioned before. Although having had a good German translation at hand, I had to look quite hard in order to produce a sufficiently accurate English version for my test.

2. But if a text is straightforward, an experienced translator can produce a raw translation quickly, by either typing it, or dictating it into a voice/text converter, or even by using one of the mentioned systems, whenever the quality and clarity of the original text permit it. One would initially leave out some terms or parts of the text causing difficulties, and check for a solution later.

3. The translator will then use the Internet for two scenarios: 1.) looking up synonyms, and 2.) finding a translation for difficult words or terms <u>in context</u>. Websites offering parallel texts for the purpose of searching for certain terms or expressions exist, but there is no absolute guarantee that the displayed results are reliable. The final decision therefore rests with the translator at all times.

4. Sometimes, an answer in a specialized translator's forum will help, but there is always a risk involved with that as well. I remember finding a very convincing translation of a rare legal term in an Italian /English forum, it did sound right but I had second thoughts. And yes, as it turned out later (luckily in time!), it was plain wrong, as Italian speakers have a natural tendency to equate an Italian word of Latin origin with its English counterpart, if such a word exists. Sometimes that easy-going method works, but most of the time it produces a completely wrong solution!

5. So, if a complicated term unknown to the translator cannot be found in context, the translation is risky. This can happen for instance with medical or legal terminology if the translator is not an expert in the field. A responsible general translator therefore must know his or her limits, just as general practitioners in medicine should know and act correspondingly, whenever a condition appears to be outside their diagnostic capabilities. But as we all know, that is not always the case, sometimes even putting the life of a patient at risk.

6. The raw text is then reread and gradually improved while going along. Once convinced that it has been well converted, it will be put aside for some time. (The human brain will get "blind" when concentrating on a specific thing or object for too long, a phenomenon well known to painters, sculptors and artists in general.)

7. Later the translator will check the text again, controlling consistency and hunting down smaller errors that might have crept in during the translation process. German is especially nasty and unforgiving in this respect, as a tiny change in a word triggers subsequent changes in word endings in other places. Sometimes such an erroneous ending is overlooked even when German is the translator's mother tongue. A remedy for that is to read the **text aloud to yourself** at least twice, a method that yields good results for English as well.

8. If the schedule is not too tight, one reads the text again the next day. Sometimes you notice things you haven't before. Unfortunately, reality shows that sometimes one or two smaller errors have been overlooked and make it into the final version. The "four eyes" system is certainly a fix for this, but that only works if texts are very carefully reviewed, which takes up a lot of time, and has to be ultimately paid for by the customer. In any case, bad style is worse than a typo or a faulty ending, in my opinion.

9. In conclusion, it might be said that while translating, the human brain works largely by association and by using ready "templates", and also seems to have an innate talent for syntax which machines do <u>not yet</u> possess. In addition to that, a trained human brain will even re-cognize and automatically reassemble partially mutilated or very poor quality text when there are just enough elements available for recognition, something that machine intelligence strug-gles heavily with at present. This also goes for electronic text recognition of some older printed texts, with OCR breaking down constantly in certain places, where a human used to that type of print would just continue reading without encountering any problems.

10. Probably that has to do with the fact that human beings possess superior pattern processing in order to identify threats or enemies when our ancestors still had to survive in a hostile environment. Enhancing that ability with drugs can produce hallucinations, because pattern processing and recognition work already at such an elevated level in humans. Surely, "Google Translate" or similar systems do not hallucinate, but they can loose their bearings altogether! If they were "honest", they would refuse to translate certain portions of texts they do not understand. (Though "understanding" is no part of an IT solution yet.) Furthermore, who would like to see some clearly marked "white" spaces in a translated text?

To conclude: I am in no way against using machine translations, but you have to be aware of the limitations and risks involved. <u>If a bad translation has the potential to damage your business, your academic standing or your reputation in general</u>, it's better to pass it on to a professional translator, even if that may be expensive at times.

In thanking you for your kind attention, I hope my advice may help you get better results when using automatic translation.

The link to the PDF *"Testing Google Translate and Yandex"* is here