

Useful Machine Translations of Japanese Patents Have Become a Reality

By Steve Vlasta Vitek
Magister of Arts,
A freelance technical translator
from Japanese, German, Czech, Slovak,
Russian, Polish and French into English
USA

Computers may in the future weigh no more than 1.5 tons.
Popular Mechanics, 1949.

The number of transistors on a microprocessor will double approximately every 18 months.
Gordon Moore, 1965.

Some two years ago, I wrote an article entitled "Reflections of a Human Translator on Machine Translation" in which I was considering the phenomenon of machine translation from the viewpoint of a human translator of Japanese patents. I offered it for publishing online to Gabe Bokor's Translation Journal mostly because I was interested in the potential of response from translators and non-translators alike who may stumble upon my article on the democratic medium of the Internet which is easily accessible to just about anyone. The response to my article was nothing short of amazing. I received several e-mails in response to my article from translators in several countries. The article was later reprinted on paper in the *Translorial*, the magazine of the Northern California Translators Association (NCTA), and in the *Capitol Translator*, the magazine of the National Capital Area Chapter of the American Translators Association (NCATA). As I found out later when I was running a search on Google and other search engines, the article in its entirety or excerpts from it were also published online and/or discussed on websites in several languages, some of which I cannot read: the entire article is published in English on a Chinese website, parts of the article are discussed in Arabic, Hungarian, Spanish and other languages on other websites, and of course, there is a mention of my article (including a short biographic note about yours truly commenting on my three amazing trilingual dogs who understand Japanese, English and Czech) on a website in Japan that lists resources for machine translation. It was also picked up by a Canadian journalist who interviewed me by phone for his own article on machine translation and I also received requests for permission to use my article from students who cite it in their PhD theses. It is listed in the bibliography of several online articles written by researchers analyzing machine translation problems and there are links to it on websites specializing in machine translation and other hi-tech issues. A translator of Hebrew has a link to my article on her website, etc. (The biggest booster for my ego is in fact when individual translators put links to my articles on their websites because to me this means that I am saying things that in their opinion need to be said).

The gist of the article was that in spite of the incredible speed with which computers can now process information, machine translation will never amount to anything more than a useful tool for translation of words, but not really meaningful sentences, because it is impossible for machines to understand the concept of meaning, as it is impossible to translate from one language into another without a clear understanding of the meaning in the original language. I still think that unlike humans, dogs and chimpanzees, machines will never understand the meaning of anything. If robots could understand the meaning of things, they could in fact become dangerous to humans because they would probably eventually figure out that they don't really need those slow-thinking and lazy humans for anything. This concept, by the way, was for the first time introduced in early thirties of the last century by the Czech writer Karl Capek in his play "R.U.R" who introduced the word "robot" into English (from the Czech word "robota" meaning "hard labor").

[Click On DETAILS For Machine Translation Of Recent Japanese Patents on the JPO Website](#)

But I must say, the service of the Japanese Patent Office website, which offers free online translation of relatively recent Japanese patents into English, comes pretty close to translation of the real meaning of the original Japanese sentence. If you want to see for yourself how it works, click on the following link:

<http://www1.ipdl.jpo.go.jp/PA1/cgi-bin/PA1INIT?999434042140>, type a technical term in the search field when **TEXT SEARCH** is highlighted, or a number of a relatively recent Japanese Patent Application when the **NUMBER SEARCH** is highlighted to display an English summary of related Japanese patents. Once you have a list of Japanese patents displayed, click on **DETAILS** for machine translations of individual recent patents. Or, if you are reading this on paper, go to my website at www.PatentTranslators.com and click on the link **JPO (English Instructions)** on the main page of my website. One of the problems with the JPO website is that most searches for technical terms will result in too many hits in which case no patent will be displayed. You can resolve this problem by going to the European Patent Office website and by clicking on the following link: <http://ep.espacenet.com> Or, if you are reading this on paper, go to my site and then click on my link to **FIND FOREIGN PATENTS (EPO)**. The EPO site displays the first 500 patents in several languages, including Japanese, regardless of the number of hits, which makes much more sense from the translator's viewpoint. Once you discover the right Japanese patent application number on the EPO website, you can go back to the JPO website to have it machine-translated. Only patent applications going back about six years or so can be translated in this way.

[I Did Not Know That Ricoh, Ltd., Had A Patent on "Supporting Education"](#)

One of the patents displayed as a result of my search for the words "Japanese patent translation system" resulted in JP 10-283364, DOCUMENT INFORMATION MANAGEMENT SYSTEM FOR SUPPORTING TRANSLATION AND THE SAME SYSTEM FOR MANAGING

PATENT INFORMATION AND THE SAME SYSTEM FOR SUPPORTING EDUCATION, filed by Ricoh Co., Ltd., on August 4, 1997. (So now we know that Ricoh, Ltd., has a patent on "supporting education," and here I thought Bill and Melinda Gates owned this patent). This patent, like so many patents, is really based on one simple idea that is basically common sense applied to technology: linking documents through hypertext to make them easily accessible and combining this accessibility with machine translation. Nothing new is really invented here, of course. The interesting thing is that the machine translation of the patent almost makes sense, as the sentences are relatively short and the underlying idea is simple enough for anyone to understand it to a point where inferring the real meaning from usual machine mistranslations is not terribly difficult. Paragraph (0010) for example explains hypertext. (0010 - MT (machine translation): *Specifically, a hypertext is the meeting of the text (document processed electronically) linked (relating), and is the fundamental concept of multimedia software which enabled it to refer to each text which relates hierarchical and pluralistically through link structure and was carried out in arbitrary sequence. Therefore, the information (namely, information as a hypertext) for forming the link structure other than the information visually offered to a user is included in the document of this hypertext. However, if the document of a hypertext is outputted as a document of paper, since the information as this hypertext will once be lost, the document of the outputted paper stops being the document of a hypertext any longer.* Now let's see how a human translator (yours truly) would translate the same paragraph: (0010 - HT (human translation): Specifically, hypertext is a basic concept in multimedia software indicating an organization of texts (electronically processed documents) through links (attached relationships) attached to these texts. This organization makes it possible to reference in any order any texts through a structure of links which can be attached in a hierarchical order on multiple levels. Therefore, these hypertext documents contain in addition to visual information that is presented to a user also information forming the structure of the links (that is to say information in the form of hypertext). Incidentally, because this hypertext information will be lost once a hypertext document is output in the form of a paper document, a paper document will no longer be a hypertext document.

Basically all the information that is in the text translated by a human is contained in the above text translated by a machine. I even used a term from the machine-translated part (kaisoteki = hierarchical) because I had forgotten the English translation of this word. Had I been working on my own, I would have to look the word up in a dictionary. I remember that I did encounter this word about two years ago in another patent about machine translation which I was translating for a patent law firm at that time, but the association of the Japanese word "kaisoteki" with the English word "hierarchical" was temporarily lost in my sometimes malfunctioning brain and its sometimes volatile memory. Computers, on the other hand, don't forget anything, provided that the software and hardware are working properly. Once somebody feeds information into the memory of a silicon translator, it should stay there indefinitely. Once a Japanese programmer inputs into the MT memory of the JPO's translation device that "barikyappu" is an abbreviation used by Japanese engineers that means "a

variable capacitor," and that "henkei korupitsukei hasshin kairo" means a "variable Colpitts oscillator circuit," the machine never forgets, unlike a human translator.

Meaning - The Final Frontier In Machine Translation - Will Never Be Conquered By Machines

The problem with machines, of course, is that when they translate a text, they are unable to look for any relationships that have not been input by a human programmer in advance into the memory. Every translation is in fact an interpretation of the original text which must be reinterpreted in a different language. Some things are left out from the original, and some new things are added in the target language in most translations, unless it is a translation of a very simple sentence between very similar languages. In a translation from Japanese, which unlike for instance German or French is very different from English, important grammatical categories such as the topic "wadai" can be translated as a subject, or as an adverb which would be the closest grammatical equivalent in English except that it is almost never possible to translate "wadai" in this manner into English, or it can be sometime ignored, depending on the meaning of the sentence because these categories have no equivalents in English. On the other hand, unlike in Japanese, you usually have to have a subject in English, your verb has to have past, present or future tense, your nouns must be in singular or plural, etc. Technical Japanese usually does not specify the tense of the verbs or whether the nouns are in singular or plural. A human translator makes usually dozens of inferences and interpretations every minute. These interpretations are based on his or her knowledge of the languages involved, but also on the lifelong experience and mysterious chemical reactions occurring in the brain of the translator that we know next to nothing about. The result of these interpretations is in the end a translation of the real meaning of the original text. I believe that meaning is a human category that cannot be programmed into a machine. The machine-translated text above may lead a human reader to somewhat different inferences about the original meaning of the Japanese text than my translation, but the meaning is really inferred by a human reader. Human translators translate meaning, but meaning is not put into the translated text by a translating machine, except where it has been input in advance into the machine's memory by a programmer who was using his or her own human sense of what things really mean. Meaning is the final frontier in machine translation. Human programmers may be able to come closer to the real meaning of certain texts when they are programming their translating machines. After all, patents deal with technological fields that have been divided into precise categories and various terms do have a specific meaning in various contexts. This can be programmed ahead of time into permanent machine memory, which is why machine translation can in some cases come very close to the actual meaning of the original sentence, especially if the machine-translated product is read by a human reader who can make the proper inference based on his or her experience. But will a machine ever understand that a perfect fit cannot be achieved between a square peg and a round hole? A human reader knows this information from experience. But machines have no experience. All they have is a storage medium combined with a processor. They will try to fit the peg

into the next available hole, regardless of its shape, unless we tell them otherwise first.

As many customers who were somewhat stupefied by the results of machine translation must have discovered to their dismay, MT in fact often stands for Mad Translation, and the results of MT can be even more hilarious than the famous imaginary translation examples described by Stephen Budiansky in the December 1998 issue of the Atlantic Monthly Magazine

<http://www.theatlantic.com/issues/98dec/computer.htm>. The article starts with a reference to the British comedy series *Monty Python*: a foreign-looking tourist clad in an outmoded leather trenchcoat appears at the entrance of a London shop, marches up to the man behind the counter, solemnly consults a phrase book, and declares in a thick Middle European accent: "My hovercraft ... is full of *ee/s!*" The scene then shifts to the Old Bailey courthouse, where the prisoner at the bar stands accused of intent to cause a breach of the peace for having published an English-Hungarian phrase book full of spurious translations. For example, the Hungarian phrase "Can you direct me to the railway station" is translated as "Please fondle my buttocks."

It Is Impossible To Resurrect A Relationship That Has Died

Few things are as fascinating as watching machines trying to make sense out of things around them. I remember how my children were fascinated when I bought for them a few years ago gigapet toys that were made in Japan, of course, as a Christmas present. These gigapets were little square device provided with a tiny screen and a couple of buttons that kids can use to program the devices. The screen displays tiny pictures indicating the mood of their electronic toys: "I am hungry, feed me!" or "I am bored, play with me!" You can feed them or play with them with the programming button, but unless you remember to do it on a regular basis, your gigapet, no longer loved and cared for, will eventually lose its will to live and a scary grave complete with a cross on it will be displayed on the tiny LCD screen. That is what happened to our gigapets anyway, after only a few days when my children finally got fed up with their new toys. I think that they lost interest in their toys because they realized that those pictures don't really have a meaning, except when meaning is simulated in short bursts of preprogrammed sequences. I hope that my children learned from this experience how difficult it is to resurrect a relationship that has died. But if not, it is a safe bet that they will have plenty of opportunity to learn this later in life.

It is very frustrating that machines don't understand the category of meaning because they lack human experience. If I have a software problem and try to use the "help" feature of just about any software package, it is almost always hopeless because I cannot communicate with the machine. The machine will lead me into various directions, but none of them will be related to the problem that I am experiencing at the moment because, unlike a dog or a chimpanzee, a machine is unable to make logical inferences from questions that I am answering on the screen. It simply throws preprogrammed lines of computer codes at me. This is even more true about machine

translation. Translations without logical inferences that are based on human experience can only make sense if the readers can infer the real meaning from the machine-translated product based on their own human experience.

A Machine Does Not Care Whether It Lives Or Dies

If I was a religious person, I would have to say that meaning is preprogrammed into organic life forms by God or at least some sort of a divine force on a higher level than the one accessible to our human understanding. Even though I am not really a religious person, I believe that this is in fact the case. Even trees and plants understand that some things are bad for them - extreme heat or cold for instance, and they shirk from extremes in order to survive, or learn how to cope with extremes in successive generations. But if you have a bad sector on your hard disk, it will eventually destroy your machine, unless your software, developed by human programmers based on human experience catches the problem and isolates it. Unlike a tree, a plant, a dog, or a human, the machine does not care whether it lives or dies. If we could teach machines the category of "meaning," we would be godlike. To put it in another way, if machines were able to understand things around them, this would represent positive proof that there is no God. Living things, for some reason, understand that some things make sense and some things don't. Inanimate objects, for some reason, cannot even conceive of this category called "understanding."

It is very dangerous to make predictions, especially absolute predictions, such as the one from "Popular Mechanics" in the beginning of my article. But I will go ahead and make the following prediction: Because machine translation will be always hampered by the fact that machines don't understand the meaning of anything, machine translation will never really make sense. Under strictly defined conditions, it may come quite close to a simulation of meaning, as does the machine translation site maintained by the Japanese Patent Office. But meaning is the final frontier that will never be crossed by machines. One day soon we may be able to program our car to drive us to work or to the supermarket while we read our paper and drink our coffee. After all, it is always the same route and most cars already have the "cruise control" feature built in. What do I know, it may be even possible some day for people to travel through space by having their bodies disassembled into atoms that will be beamed to different galaxies and then reassembled again on a different planet. Perhaps people will be able to travel through time one day. It is really annoying that travel in time is possible only in one direction. But the fact that meaning has to be inferred by humans from reality surrounding them is the Achilles heel of machine translation. This is the reason why machine translation will never really become what the general public expects it to be one day soon.

I am also predicting that the fact that websites of the Japanese Patent Office and European Patent Office make so much information available to translators online will make life increasingly easier for translators. Possibly the most difficult problem encountered by translators from Japanese are foreign words (foreign in Japanese) which are transliterated into katakana, one of two native Japanese alphabets used in

addition to kanji characters. Because the phonologic structure of Japanese contains relatively few vowels and consonants when compared to other languages and the Japanese alphabet is really a syllabary, the pronunciation of a foreign word can be very different from the word in the original language. On top of that, foreign words are often shortened in Japanese and thus basically mutilated almost beyond recognition (as in my example of "barikyappu" = "variable capacitor"). Translators can now use the machine translation feature of the Japanese Patent Office to "guess" the correct spelling of a foreign word in Japanese because it may be a term that has been input already into memory by a human translator. However, if it is a recent or unusual word, the machine will display a series of asterisks, which is what is displayed when the machine does not "understand" something. (I wish I could do that too!!) Another thing that translators can try with the JPO or EPO website is to throw a few key terms from the patent into the search line of the website to display similar patents that may contain the term in question, provided that it is a key term that has been already translated by somebody else. When all of the above fails, I have to try to guess from the mutilated katakana transcription whether the original word is in English, French, German, Polish, Russian or another language based on my knowledge of these languages and the context and then try to find this word in another search engine such as Google. It really helps when you can get the correct spelling from a machine with a few keystrokes.

I am also predicting that my customers who usually don't read any Japanese will be able to use the machine translation feature of the JPO website to look for patents that are relevant to their customers, eliminate patents that are not very relevant without the help of a human translator, which tends to be expensive, and identify patents that need to be translated by a human translator in this manner. I think that the net result will be more work for human translators who understand the category of meaning when unnecessary work is eliminated and when, with the help of machine translation, it is easier to discover facts that might have been hidden before.

Machines can be very helpful, but no matter how many transistors you can cram into the width of a human hair, the thing about machines is, they don't really understand anything at all. In the end, you do need a human to make sense out of things.