MACHINE TRANSLATION: SUCCESS OR FAILURE
USING MT IN AN IT RESEARCH AND DEVELOPMENT ENVIRONMENT

Ursula Bernhard
GMD - Forschungszentrum Informationstechnik GmbH, D-53754 Sankt Augustin

The present paper discusses the use of raw machine translations in an IT research and development environment. Researchers use machine translation as a drafting tool for scientific papers. The language pairs are German -> English and English -> German. The raw machine translations are produced on an experimental basis by means of the MT systems LOGOS, METAL and Globalink Power Translator Professional. The experiments are to show whether the systems are suitable for this purpose or not. Since the use of raw machine translations is considered to be crucial to the future of MT, the paper discusses success or failure of the technology against this background.

INTRODUCTION

The history of machine translation (MT) is characterized by a strange mixture of euphoria and disappointment. Even after more than forty years of research and development in this field, machine translation is by no means a technology which is generally accepted and used. It is even still open whether machine translation will be a success or a failure.

Against this background, the present paper describes experiments conducted to find out whether machine translation is suitable for purposes outside the traditional translation business, e.g. for drafting scientific papers in the information-technology-oriented research and development environment of GMD.

GMD - German National Research Centre for Information Technology (IT) conducts research aiming at the development of innovative methods and applications. It cooperates closely with industry and users, thereby increasing the competitiveness of the German and European economies. GMD has four main research areas: system design technology, communication and cooperation, intelligent multimedia systems and parallel computing. GMD's headquarters and five of its nine research institutes are located at Birlinghoven Castle in Sankt Augustin, near Bonn. Two research institutes each are located in Berlin and Darmstadt, and there is a liaison office in Tokyo. GMD has a staff of 1200. The total annual budget of GMD is DM 170 million. A basic funding of 130 million is provided by the German Federal Ministry for Education, Science, Research and Technology (90%) and by the Federal States of North Rhine-Westfalia, Hesse and Berlin, an additional 40 million comes from project funds and industrial cooperation.

GMD researchers work at international level, their working language is English and most publications are written in English without producing a German version.

SOME HISTORY

Let us first have a look at the history of machine translation in GMD.

Introducing Machine Translation

Since 1986 GMD has been using machine translation for the everyday work of its Translation Service which is a central service unit belonging to the Library and Publication Service. In 1986, the translation volume was about 3000 pages per year and showed an increasing tendency. Main language pair was German into English followed by English into German, French into German and Russian into German. The reason for introducing machine translation was primarily not a quantity problem, but rather a quality problem. GMD did and does research and development in completely new fields with new terminologies. These terminologies are still developing and they are often not known to commercial translation services which are oriented to the requirements of trade and industry. Therefore, GMD had to find a means to do such highly specialized translations inhouse in close cooperation with the experts knowing the terminology better than a non-expert translator.

The machine translation system introduced in spring 1986 was LOGOS for the language pair German into English on an IBM mainframe. At that time this platform fitted best in GMD's environment. The decision for LOGOS was relatively easy as LOGOS was the only translation system for German into English offering dictionary extension by the user in the mid-80s. After having solved a lot of mainly technical problems, for example, missing converters for word processing systems, the actual production of translations by means of LOGOS began in autumn 1987. And in contrast to other machine translation applications at that time, using LOGOS turned out to be successful since the time required for handling a translation job was reduced by 50%.

Changing from LOGOS to METAL

In 1989, GMD decided to change to METAL which had just been marketed by Siemens. In Germany, GMD is the only user who has ever directly changed from one machine translation system to another. The METAL system seemed to be more advanced and promising and the Munich-based development centre for German as source language made us hope for a better and quicker response to bug reports than it was the case at that time with the US-based LOGOS development centre.

In December 1989, METAL was installed in GMD. The GMD terminology entered into the LOGOS dictionary was converted to the METAL format and could thus be reused which saved a lot of time and money. The METAL installation was as successful as the LOGOS installation, perhaps even a little bit more successful.

Translated Material

The material translated by METAL and formerly by LOGOS were software manuals, project reports and scientific papers. These are texts of a relatively transparent style and of a relatively homogeneous terminology since this is a prerequisite for machine translation. Legal or political texts and any material of a more colloquial style are not suitable for machine translation.

CURRENT SITUATION

In the meantime, it has become GMD's corporate philosophy to publish preferably in English. GMD's researchers are to be able to write their scientific papers in English language. This has changed the working environment of the Translation Service dramatically. The overall demand for translations has decreased considerably though there is an increase for translations from English into German since many texts originally written in English have to be translated into German if needed in this language, e.g. for German leaflets or for the German funding authorities.

There is also another change. People are no longer willing to wait for a translation, they want to have it immediately. GMD's researchers work in an environment which is characterized by most innovative technology. Worldwide access to any information is secured by Internet. Within seconds, the whole world can be displayed on the screen. GMD's researchers want to do almost everything by key stroke or mouse click without leaving their office and, if any possible, they want to do it themselves; they want to be independent.

INTRODUCING A RAW MACHINE TRANSLATION SERVICE

Though having a solid knowledge of English language, some researchers asked for a support for writing their papers in English or for producing a German version of a paper originally written in English. They did not want to be confronted with an empty screen. They wanted to have a tool producing an English or German text for using it as a basis, thus facilitating and accelerating the process of writing. Therefore, based on earlier experiences with raw machine translation requesters and their acceptance of this type of translation, a raw machine translation service has been introduced on an experimental basis. The fact that GMD's Translation Service is also involved in research and teaching secured a most interesting test environment.

Environment

Three machine translation systems are used for the tests in the experimental phase. Globalink, version 4.0 for Apple Macintosh, translating the language pairs English-German and German-English, LOGOS, version 7.7, running on a Sun SparcStation 10, equally for English-German and German-English and METAL, version 3.03, running on the same machine, for the language pair German-English only. Source texts are received on floppy, via file transfer or e-mail using GMD's network which interconnects inhouse computer systems and provides access to outside networks.

Procedure

The source texts are first scanned for unknown words. In a second step, the source texts are preedited by correcting typos or by protecting material not to be translated by specific characters as far as this material has been listed as unknown words. The actually unknown words are added to the dictionary of the MT system or if necessary the requester is sent a list of unknown words with suggested translations asking him or her for revision. After having added the unknown words to the system, the system's translation process is activated. The result, i.e. the raw machine translation is returned to the requester via e-mail, file transfer or on floppy. Usually, translation is done overnight and the raw machine translation is returned in the morning of the following day.

In the experimental phase, the texts to be translated from German into English are processed by means of Globalink, LOGOS and METAL, translations from English into German can only be done by Globalink and LOGOS since the METAL version available in GMD is only for German into English. The requester is usually delivered the translation done by the MT system which returned the shortest list of unknown words since this secures fast handling. The other translations are examined by the Translation Service or by requesters who are actively involved in the test. The examination is done by comparing the translations and by marking the best version and the worst version of each sentence and by identifying those sentences which are completely correct, i.e. which show a near to human translation quality. The criteria underlying this classification are grammatical correctness and understandability.

FIRST RESULTS

As expected, GMD's researchers happily accepted the idea of a raw machine translation service and used it intensively. Though, it is to be noted that the results presented in this paper are only of a provisional nature. The service was introduced in July so that many people were on summer holidays. The actual use of the service started only after their return in September. Therefore only a small part of the raw machine translations have been examined so far. More reliable results will probably be added on the very presentation of the paper.

It is to be noted that the tests of the experimental phase do not aim at evaluating MT systems. They are simply to show whether the raw machine translations produced by the three systems in the information technology domain are usable for the intended purpose. May be that tests in another domain will lead to results which are completely different from those presented here.

The following results concern source text problems, problems with the general handling of MT systems and dictionary expansion and last but not least translation quality.

Source Text Format

Though the source texts should be ASCII or WORD files without syllabification at the end of a line and without foreign format control characters, the input files often contained such material which required a lot of boring and time-consuming preediting. As already told before, the requesters were asked to provide terminology. Sometimes they did so by including the specific target-language terms in the very source text. Since an MT system does not understand the target language in the analysis phase, these target-language words had to be removed or protected against translation which required again a lot of tedious preediting.

Source Text Correctness and Style

Experienced users of machine translation systems know that the quality of a raw machine translation considerably depends on the quality of the source text. Even in an academic research environment as in GMD, the source texts are far from being perfect. Though spelling checkers are available to everybody, the source texts contain typos, the use of grammar and punctuation is not always correct and the sentences are often most complicated and lengthy. This applies in particular to German source texts. Therefore, a certain amount of bad raw translation quality is to be attributed to poor source text quality.

Source Text Ambiguity

Most people underestimate the ambiguity of language. They are often not aware that the terms they use have also another meaning in some other domain and that certain phrases can simply be misunderstood. Using MT systems for producing translations mostly means to select subject-specific dictionaries and to exclude general dictionaries or other subject dictionaries. For example, in a technical translation the English phrase

"everything will turn out all right"

was translated into German as

"alles schaltet alles Recht aus"

The verb "turn out" was available only in its technical meaning. The same applied to "all right" which was translated literally.

MT System Handling and Dictionary Expansion

The general handling of the MT system and the ease of dictionary expansion are factors which are of great importance to the usability of MT systems for raw machine translation since they influence the speed of translation delivery.

Globalink. The Globalink Power Translation version used for the present tests does not include a facility enabling the user to scan a text first for unknown words. One has to activate the translation process for doing so, it delivers a list of words not found in the source text. Therefore, machine translation with Globalink requires two complete translation processes, one before dictionary expansion, the other after dictionary expansion. Globalink for German->English does not deliver a list of recognized compounds with suggested translations. Therefore, there is no chance to correct strange combinations or to expand the dictionary by the required subject-specific compound translations.

Dictionary expansion is somewhat tedious since the menus are not self-explaining. One has to remember specific codes for teaching the system the inflection or the usage of a word. Probably, naive users will master the menu better than somebody with linguistic education since the latter might expect a specific logical structure of the dictionary menus. The facilities of teaching the system non-standard translations are restricted.

LOGOS. LOGOS provides the facility of a so-called new word search. This search produces a file of unknown words and suggests translations for compounds based on the known constituents. The file can be used as input file for dictionary expansion, thus enabling an adaptation of the dictionary to the specific source text. Dictionary expansion which is done by means of a tool named ALEX is rather easy and quick. The fact that new verbs cannot be entered by the user was of minor importance to the tests since the source texts contained hardly any new verb. In addition, LOGOS is rather robust with respect to unknown words. In most cases, sentence analysis is successful and the unknown word is in the correct position and need only be replaced with the target language term. The ALEX menus are self-explaining though some of them are somewhat intransparent. A tool called SEMANTHA enables the user to teach LOGOS non-standard translations within a restricted number of templates.

METAL. Among the three systems, METAL is the only one providing a so-called pattern matcher though the forthcoming LOGOS version will also include such a facility. The pattern matcher allows global modifications to be done on source text or target text level. It is a valuable help for protecting proper names against translation.

METAL offers a dictionary lookup facility producing three lists: unknown words, compounds with suggested translations on the basis of the known constituents and known words with subject area, these are words already included in the dictionary in probably another subject area as required. These three lists are used for adapting the dictionary to the specific source text. Especially the list of known words is a rather valuable help for avoiding wrong translations on account of ambiguity. METAL provides the most sophisticated facilities of teaching the system non-standard translations though using these facilities takes a lot of time and requires linguistic skills.

Translation Quality

Translation quality is of course the most crucial factor if discussing the usability of MT systems. It is difficult to avoid subjectivity when evaluating translation quality by using understandability as criterion. This also applies to the present paper though the domain expert status of the evaluators has helped to avoid too much bias.

Globalink. As for the translation quality, Globalink showed the highest percentage of worst versions though its percentage of sentences with near to human translation quality was almost equal to that shown by LOGOS. Especially more complex German sentences requiring a word reordering in English caused a lot of problems. When translating from English into German, the agreement between grammatically linked items was often missing which led to sentences which could hardly be understood.

LOGOS. LOGOS showed the highest percentage of best versions. In particular, it showed a certain robustness with respect to grammatical errors in the source text, e.g. missing agreement between linked items, and it mastered very well the impersonal constructions which are most popular in German scientific papers. From English into German, it showed sometimes surprisingly good results in multi-word phrase analysis.

METAL. As for translation quality, METAL showed a lower percentage of best versions than LOGOS, but the highest percentage of sentences with near to human translation quality which is due to its sophisticated facilities of learning non-standard translations. A serious shortage is missing robustness. In the case of complex sentence structure or too many words in a sentence (usually more than 30 words) METAL stops analysis after firing 5000 rules and delivers a so-called phrasal analysis, that is a translation of the recognized phrases without interrelating them.

CONCLUSIONS

The following conclusions are only of a preliminary nature since they are based on tests which have not yet been completed. The test phase will be continued to obtain

more reliable results. The present conclusions deal with the general suitability of raw machine translations for the intended purpose and the specific suitability of the various systems for producing such translations. Finally, success or failure is discussed against the background of the results obtained so far though these results are of a restricted applicability since they consider only a specific domain. Nevertheless the findings may point to a trend or a chance securing the success of a technology whose failure has often been predicted by experts.

Suitability

Though being sometimes scared or amused at the quality of raw machine translations, the majority of GMD's researchers participating in the tests think that machine translation could be used as a drafting tool for scientific papers. Being the authors of the source texts they easily recognize mistranslations. Having all the required words displayed on the screen, it is easy for them to put the words into the correct order and to find a better version if necessary. In addition, they can be sure that the specific terminology of their domain is used consistently. The researchers participating in the tests agree that the raw machine translations facilitate and accelerate the drafting of scientific papers.

As to the suitability of the various systems, it is still too early to select a specific system as the most suitable one. In the translations examined so far, LOGOS and METAL showed almost equally good or equally bad results for German into English. For the language pair, English into German, LOGOS did better than Globalink, but with respect to Globalink it should not be forgotten that this system is very cheap. If compared with rather expensive systems such as LOGOS and METAL, Globalink is not so bad. In addition, LOGOS and METAL have been longer in use at GMD and are therefore more tailored to the specific needs which apply in particular to METAL. Some GMD researchers decided to purchase Globalink for their local PC or Macintosh to be independent. Since it is cheap, they accept mistranslations more easily than in the case of much more expensive MT systems.

Success or Failure

Though the results obtained so far are not reliable, they indicate a promising use of machine translation, namely the use of machine translation as drafting tool. The traditional use of machine translation in the translation business requires that the raw machine translation be postedited by a human translator. This method leads to problems with cost-benefit analysis and acceptance by posteditors. The greatest advantage of machine translation, namely speed, is wrecked by time-consuming postediting though speed is the only argument to justify the cost of this technology. In addition, at least in Germany, human translation is rather cheap due to unemployment in this sector, therefore fully postedited machine translation cannot compete. To survive, machine translation has to look for application areas using its actual advantage, namely speed and it is speed which is required for the information transfer on the data highways of the global village.

However, as evolution teaches, survival requires change. Machine translation systems have to be improved to be successful. This improvement should not be restricted to user interfaces. It should also include robustness and it should implement the state of the art of computational linguistics since available machine translation systems are far behind it. The user securing the success of machine translation will be the end user and not the translator and what the end user needs are easily operable, robust and fault-tolerant machine translation systems enabling him or her to overcome the language barriers in global communication. With such end-user-friendly systems, machine translation will be a success and not a failure.