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.