

**THE MINISTRY OF HIGHER AND SECONDARY SPECIAL
EDUCATION OF THE REPUBLIC OF UZBEKISTAN
Uzbekistan State World Languages University**

**Translation/Interpretation Faculty
Translation Theory and
Practice Department**

COURSE PAPER

**“Machine Translation: Past, Present and
Future”**

Written by: Bekchanova D, student of group № 411

Scientific adviser: teacher M. Shoumarova

TASHKENT -2011

Contents

1. Preface
2. Machine Translation: The First 40 Years, 1949-1989
3. Machine Translation in 1990s
4. Machine Translation Quality
5. Machine Translation and Internet
6. Machine and Human Translation
7. Concluding remarks
8. Literature used

Preface

Now it is time to analyze what has happened in the 50 years since machine translation began, review the present situation, and speculate on what the future may bring. Progress in the basic processes of computerized translation has not been as striking as developments in computer technology and software. There is still much scope for the improvement of the linguistic quality of machine translation output, which hopefully developments in both rule-based and corpus-based methods can bring. Greater impact on the future machine translation scenario will probably come from the expected huge increase in demand for on-line real-time communication in many languages, where quality may be less important than accessibility and usability.

Machine Translation: The First 40 Years, 1949-1989

About fifty years ago, Warren Weaver, a former director of the division of natural sciences at the Rockefeller Institute (1932-55), wrote his famous memorandum which had launched research on machine translation at first primarily in the United States but before the end of the 1950s throughout the world.

In those early days and for many years afterwards, computers were quite different from those that we have today. They were very expensive machines disposed in large rooms with reinforced flooring and ventilation systems to reduce excess heat. They required a huge number of maintenance engineers and a dedicated staff of operators and programmers. Most of the work was mathematical in fact, either directly for military institutions or for university departments of physics and applied mathematics with strong links to the armed forces. It was perhaps natural in these circumstances that much of the earliest work on machine translation was supported by military or intelligence funds directly or indirectly, and was destined for usage by such organizations – hence the emphasis in the United States on Russian-to-English translation, and in the Soviet Union on English-to-Russian translation.

Although machine translation attracted a great deal of funding in the 1950s and 1960s, particularly when the arms and space races began in earnest after the launch of the first satellite in 1957, and the first space flight by Gagarin in 1961, the results of this period of activity were disappointing. US was even going to close the research after the publication of the shattering ALPAC (Automatic Language Processing Advisory Committee) report (1966) which concluded that the United States had no need of machine translation even if the prospect of reasonable translations were realistic – which then seemed unlikely. The authors of the report had compared unfavourably the quality of the output produced by current systems with the artificially high quality of the first public demonstration of machine translation in 1954 – the Russian-English program developed jointly by IBM and

Georgetown University. The linguistic problems encountered by machine translation researchers had proved to be much greater than anticipated, and that progress had been painfully slow. It should be mentioned that just over five years earlier Joshua Bar-Hillel, one of the first enthusiasts for machine translation who had been disabused of his work, had published his critical review of machine translation research in which he had rejected the implicit aim of fully automatic high quality translation (FAHQT). Indeed he provided a proof of its "non-feasibility". The writers of the ALPAC report agreed with this diagnosis and recommended that research on fully automatic systems should stop and that attention should be directed to lower-level aids for translators.

For some years after ALPAC, research continued on a much-reduced financing. By the mid 1970s, some success could be shown: in 1970 the US Air Force began to use the Systran system for Russian-English translations, in 1976 the Canadians began public use of weather reports translated by the Meteo sublanguage machine translation system, and the Commission of the European Communities applied the English-French version of Systran for helping it with its heavy translation burden – which soon was followed by the development of systems for other European languages. In the 1980s, machine translation rose from its post-ALPAC low spirits: activity began again all over the world – most notably in Japan – with new ideas for research (particularly on knowledge-based and interlingua-based systems), new sources of financial support (the European Union, computer companies), and in particular with the appearance of the first commercial machine translation systems on the market.

Initially, however, attention to the renewed activity was still almost focuses on automatic translation with human assistance, both before (pre-editing), during (interactive solution of problems) and after (post-editing) the translation process itself. The development of computer-based aids or tools for use by human translators was still relatively neglected – despite the explicit requests of translators.

Nearly all research activities in the 1980s were devoted to the exploration of methods of linguistic analysis in order to create generation of programs based on traditional rule-based transfer and interlingua (AI-type knowledge bases representing the more innovative tendency). The needs of translators were left to commercial interests: software for terminology management became available and ALPNET produced a series of translator tools during the 1980s – among them it may be noted was an early version of a program "Translation Memory" (a bilingual database).

Machine Translation in 1990s

The real emergence of translator aids came in the early 1990s with the "translator workstation", among them were such programs as "Trados Translator Workbench", "IBM Translation Manager 2", "STAR Transit", "Eurolang Optimizer", which combined sophisticated text processing and publishing software, terminology management and translation memories.

In the early 1990s, research on machine translation was reinforced by the coming of corpus-based methods, especially by the introduction of statistical methods ("IBM Candide") and of example-based translation. Statistical (stochastic) techniques have brought a relief from the increasingly evident limitations and inadequacies of previous exclusively rule-based (often syntax-oriented) approaches. Problems of disambiguation, refraining from repetition and more idiomatic generation have become more solvable with corpus-based techniques. On their own, statistical methods are no more the answer in contrast to rule-based methods, but there are now prospects of improved output quality which did not seem reachable 15 years ago. As many observers have indicated, the most promising approaches will probably integrate rule-based and corpus-based methods. Even outside research environments integration is already evident: many commercial machine translation systems now incorporate translation memories, and many translation memory systems are being enriched by machine translation methods.

The main feature of the 1990s has been the rapid increase in the use of machine translation and translation tools. The globalization of commerce and information is placing increasing demands upon the provision of translations. It means not only continuing (maybe even accelerating) growth of the use by multinational companies and translation services of systems to assist in the production of good quality documentation in many languages – by the use of machine translation and translation memory systems or by multilingual document authoring systems, or by combinations of both. Until recent times, the production

of translations has been seen essentially as a self-contained activity. For large users, the appearance of translation systems has stimulated the integration of translation and documentation (technical writing and publishing) processes. Translation is now seen as one stage in the processes of communication and getting information. Future products for such kind will not be separate independent machine translation systems, translator workstations or translation tools, but multilingual documentation software complexes combining document creation, translation and revision, document archiving, information analysis, restoration and extraction, etc. in order to satisfy the specific needs of companies.

Machine Translation Quality

Despite the prospects for the future, it has to be said that the new approaches of the present have not yet resulted notable improvements in the quality of the raw output by translation systems. These improvements may come in the future, but overall it has to be said that at present the actual translations produced do not represent major advances on those made by the machine translation systems of the 1970s. We still see the same errors: wrong pronouns, wrong prepositions, anomalous syntax, incorrect choice of terms, plurals instead of singulars, wrong tenses, etc. – errors that no human translators would ever commit. Unfortunately, this situation probably won't change in the near future. There is little sign that basic generalpurpose machine translation programs are soon going to show significant advances in translation quality. And I think that if producers of machine translating systems are still to continue sating market with software of low quality (as in present) the whole machine translation industry may be condemned for ever by the general public as producers of essentially poor-quality software, that could possibly cause damaging of the research and development or even its closure.

In order not to be unsubstantiated I would like to present examples of translation by the programs of machine translation which are the most widely distributed in Ukraine – "Promt" and "Magic Gooddy" (same producer), "Pragma", "Socrat" and one web-resource which provides on-line real-time translation. Their work will be presented on the basis of translation of the extract from the British newspaper article:

The Sunday Times:

Egypt has been training British MI5 and MI6 agents in how to combat Islamic terrorists, underlining Cairo's growing importance to the war against terror and the Middle East peace process.

A senior Middle Eastern military intelligence official revealed last week that British officers had undergone the training as part of a co-operation programme with Egypt that began after the September 11 attacks on America in 2001 and continued until last year.

Details have not been revealed, but it is believed to have included instruction in specialised interrogation techniques and in the terminology used by terrorists, which will enable agents to understand monitored telephone conversations.

Prompt XT (Magic Gooddy):

Египет обучил британский MI5 и MI6 агентов при том, как сразиться с Исламскими террористами, подчеркивая важность роста Каира к войне против ужаса{террора} и ближневосточного мирного процесса.

Старшее Ближневосточное военное должностное лицо сведений{интеллекта} показало на прошлой неделе, что британские чиновники{офицеры} подверглись обучению как часть программы сотрудничества с Египтом, который начал после 11 сентября нападения на Америку в 2001 и продолжался до прошлого года.

Детали не были показаны, но это, как полагают, включило инструкцию в специализированные методы допроса и в терминологию, используемую террористами, которые позволят агентам понять проверенные телефонные беседы.

Socrat:

Египет готовил British MI5 и агентов MI6 в как, чтобы бороться террористов Islamic, подчеркивающих Каир растущего значения в войну против террора и мирный процесс Среднего Востока.

Старший Средний Восточный военный чиновник разведки обнаружил прошлую неделю, которую Британские должностные лица подверглись подготовку как часть программы сотрудничества с Египтом, который

начинался после Сентябрьские 11 атак в Америке в 2001 и продолженной до прошлого года.

Детали не обнаружены, но поверено включенная инструкция в specialised методах опроса и в терминологии использованной террористами, которые позволяют, чтобы понимать проверенную телефонную беседу.

Pragma:

Египет обучает Британский MI5 и агенты MI6 в том, как бороться с Мусульманскими террористами, подчеркивая Cairo's, достигающий важность война против террора и процесса мира Ближнего Востока.

Ответственный Средний Восточный военный работник сведений, показанный на последней неделе, что Британские офицеры имели undergone обучение, так как часть программы сотрудничества с Египтом, который начался после 11 сентября, нападает на Америке в 2001 и длительный пока прошлый год.

Детали не были показаны, но это, как полагается, включало инструкцию в specialised методах допроса и в терминологии, используемой террористами, которые предоставят возможность агентам понять контролируемые телефонные беседы.

Internet machine translator www.tranexp.com:

Египетский вспомогательный глагол для образования сложных времен быть воспитывать Британский MI5 и MI6 действующая сила в как к бой Мусульманский террорист, подчеркивать Пирамида из камней рост значительность к грамматический определенный член война от ужас и грамматический определенный член Середина Восток мир процесс. ВЫСШАЯ ОТМЕТКА ЗА КЛАССНУЮ РАБОТУ старший Середина Восточный военный ум служебный открывать последний неделя тот Британский чиновник вспомогательный глагол для образования сложных

времен испытывать грамматический определенный член воспитывать как часть яние) от высшая отметка за классную работу со - действие программа с Египетский тот начинать за грамматический определенный член Сентябрь 11 атаковать на Американский в 2001 и непрерывный до прошлый год. Подробность вспомогательный глагол для образования сложных времен не быть открывать, только он быть верить к вспомогательный глагол для образования сложных времен заключать обучение в специализация вопрос техника и в грамматический определенный член терминология употребление у террорист, который воля давать возможность или право действующая сила к понимать наставник телефон разговор.

Literary translation:

Египет обучал агентов пятого и шестого отделов Британской военной разведки методам борьбы с исламскими террористами, тем самым, подчеркнув растущую значимость Каира в мирном процессе на Ближнем Востоке и борьбе с террором.

Старшее должностное лицо Ближневосточной военной разведки обнародовал секретные данные о том, что Британские офицеры прошли курс подготовки в качестве части программы сотрудничества с Египтом, которая началась вскоре после атак на Америку 11 сентября 2001 года и продолжалась до прошлого года.

Детали не разглашались, однако считается, что они прошли курс обучения специальным техникам допроса и терминологии используемой террористами, который позволит агентам расшифровывать перехваченные телефонные разговоры.

No doubt that the most appropriate translation was made by "Promt", but still its producer Russian company "ПРОЕКТ МТ" shouldn't stop on achieved.

Machine Translation and Internet

The impact of the Internet has been significant in recent years. We are already seeing an accelerating growth of real-time on-line translation on the Internet itself. In recent years, we have seen many systems designed specifically for the translation of Web pages ("Pop-Up Dictionary", "Site Translator") and of electronic mail ("SKIIN"). The demand for immediate translations will surely continue to grow rapidly, but at the same time users are also going to want better results. There is clearly an urgent need for translation systems developed specifically to deal with the kind of colloquial (often wrongly formed and badly spelled) messages found on the Internet. The old linguistics rule-based approaches are probably not equal to the task on their own, and corpusbased methods making use of the massive data available on the Internet itself are obviously appropriate. But as yet there has been little research on such systems. At the same time as we are seeing this growing demand for "crummy" translations, the Internet is also providing the means for more rapid delivery of quality translation to individuals and to small companies. A number of machine translation systems on the sale are already offering translation services, usually "adding value" by human post-editing. More will surely appear as the years go by.

However, the Internet is having further profound impacts that will surely change the future prospects for machine translation. There are predictions that the stand-alone PC with its array of software for word-processing, databases and games will be replaced by Network Computers which would download systems and programs from the Internet at any time as required. In this scenario, the one-off purchase of individually packaged machine translation software or dictionaries would be replaced by remote stores of machine translation programs, dictionaries, grammars, translation archives or specialized glossaries which would obviously be paid for according to usage. It should be said, that such a change would have profound effect on the way in which machine translation systems are developed.

Another profound impact of the Internet will concern the nature of the software itself. What users of Internet services are seeking is information in whatever language it may have been written or stored. Users will want a seamless integration of information retrieval, extraction and summarization systems with translation

In fact, it is possible that in next years there will be fewer "pure" machine translation systems (commercial or on-line) and many more computer-based tools and applications in which automatic translation is just one component. As a first step, it will surely not be long before all word-processing software includes translation as an in-built option. Integrated language software will be the norm not only for the multinational companies but also available and accessible for anyone from their own computer (desktop, laptop, notebook or network-based server) and for any device like television or mobile telephone which interfacing with computer networks.

Spoken Language Translation

The most widely anticipated development of the next decade must be that of speech translation. When current research projects (ATR, C-STAR, JANUS, Verbmobil) were begun in the late 1980s and early 1990s, it was known that practical applications were unlikely before the next century. The limitation of these systems to small domains has clearly been essential for any progress, such are the complexities of the task; but these limitations mean that, when practical demonstrations are made, observers will want to know when broader coverage will be realizable. There is a danger here that the mistakes of the 1950s and 1960s might be repeated; then, it was assumed that once basic principles and methods had been successfully demonstrated on small-scale research systems it would be merely a question of finance and engineering to create large practical systems. The truth was otherwise; large-scale machine translation systems have to be designed as such from the beginning, and that requires many man-years of effort. It is still true to say that the best written-language machine translation systems of today are the outcome of decades of research and development.

Whatever the high expectations, it is surely unlikely that we will see practical speech translation of significantly large domains for commercial exploitation for another twenty years or more. Far more likely, and in line with general trends within the field of written language machine translation, is that there will be numerous applications of spoken language translation as components of small-domain natural language applications, e.g. interrogation of databases (particularly financial and stockmarket data), interactions in business negotiations or intra-company communication.

Machine and Human Translation

In the past there has often been tension between the translation profession and those who advocate and research computer-based translation tools. But now at the beginning of the 21-st century it is already apparent that machine translation and human translation can and will co-exist in relative harmony. Those skills which the human translator can contribute will always be in demand.

Where translation has to be of "publishable" quality, both human translation and machine translation perform their roles. Machine translation is demonstrably cost-effective for large scale and/or rapid translation of (boring) technical documentation, (highly repetitive) software localization manuals, and many other situations where the costs of machine translation plus essential human preparation and revision or the costs of using computerized translation tools are significantly less than those of traditional human translation with no computer aids. By contrast, the human translator is (and will remain) unrivalled for non-repetitive linguistically sophisticated texts (in literature or law), and even for one-off texts in specific highly-specialized technical subjects.

For the translation of texts where the quality of output is much less important, machine translation is often an ideal solution. For example, to produce "rough" translations of scientific and technical documents that may be read by only one person who wants to find out only the general content and information and is unconcerned whether everything is intelligible or not, and who is certainly not discouraged by stylistic awkwardness or grammatical errors, machine translation will increasingly be the only appropriate decision. In general, human translators are not prepared (and may resent being asked) to produce such "rough" translations. In such a case the only alternative to machine translation is no translation at all.

However, as I have already mentioned, greater familiarity with "crummy" translations will inevitably stimulate demand for the kind of good quality translations which only human translators can satisfy.

For the one-to-one interchange of information, there will probably always be a role for the human translator, that is for the translation of business correspondence (particularly if the content is sensitive or legally binding). But for the translation of personal letters, machine translation systems are likely to be increasingly used; and, for e-mail and for the extraction of information from Web pages and computer-based information services, machine translation is the only feasible solution.

As for spoken translation, there must surely always be a place for the human translator. There can be no prospect of automatic translation replacing the interpreter of diplomatic negotiations.

Finally, machine translation systems are opening up new areas where human translation has never featured: the production of draft versions for authors writing in a foreign language, who need assistance in producing an original text; the real-time on-line translation of television subtitles; the translation of information from databases; and, no doubt, more such new applications will appear in the future as the global communication networks expand and as the realistic usability of machine translation (however poor in quality compared with human translation) becomes familiar to a wider public.

Concluding remarks

Different electronic devices have become common nowadays. Taking information from foreign languages with the help of different electronic devices represents quite a new approach in modern translation practice. Due to the fundamental research in the systems of algorithms and in the establishment of lexical equivalence in different strata of lexicon, machine translation has made considerable progress in recent years. Nevertheless, its usage remains restricted in scientific, technological, lexicographic realms. That is because machine translation can be performed only on the basis of programmes worked out by linguistically trained operators. Besides, the process of preparing programmes for any matter is inseparably connected with great difficulties and takes much time, whereas the quality of translation is far from being satisfactory even at the lexical level, which have direct equivalent lexemes in the target language. Considerably greater difficulties, which are insurmountable for machine translation programs, present morphological elements like prefixes, suffixes, endings, etc. Syntactic units (word combinations, sentences) with various means of connection between their components are also great obstacles for machine translation. Moreover, modern electronic devices which perform translation do not possess the necessary lexical, grammatical and stylistic memory to provide the required standard of correct literary translation. Hence, the frequent violations of syntactic agreement and government between the parts of the sentence in machine translated texts. Very often the machine translation program can not select in its memory the correct order of words in word-combinations and sentences in the target language. And as a result of it, any machine translation requires a thorough proof reading and editing and this takes no less time and efforts and may be as tiresome as the usual hand-made translation of the passage.

Literature

1. Weaver Warren - "Translation". Cambridge, Mass.: Technology Press of M.I.T., 1955.
2. Hutchins W.J. - "Machine Translation: Past, Present, Future". "Wiley", Chichester, Ellis Horwood, N.Y. etc., 1986.
3. Materials from Machine Translation Summit VII, 13th-17th September 1999, Kent Ridge Labs, Singapore.
4. "New Scientist Magazine" (www.newscientist.com):
 - "Device translates spoken Japanese and English" - 07/10/2004
 - "I think it thinks" - 06/10/2001
 - "Technology: Machine minds your language" - 26/10/1996
5. Беляева Л.Н., Откупщикова М.И. - "Прикладное языкознание" (Раздел - Автоматический (машинный) перевод). Изд-во Санкт-Петербургского ун-та, СПб., 2001.
6. Журнал "Вопросы языкознания" - Шаляпина З.М. - "Автоматический перевод: эволюция и современные тенденции", 1996, № 2.
7. Баранов А.Н. - "Введение в прикладную лингвистику" (Раздел - Машинный перевод). УРСС, М., 2001.
8. Леонтьева Н.Н. - "К теории автоматического понимания естественных текстов". Издательство Московского университета, М., 2000.
9. Бакулов А.Д., Леонтьева Н.Н. - "Теоретические аспекты машинного перевода". Радио и связь, М., 1990.
10. Нелюбин Л.Л. - "Компьютерная лингвистика и машинный перевод". ВЦП, М., 1991.