English/Arabic/English Machine Translation: A Historical Perspective

MUHAMMAD RAJI ZUGHOUL
Yarmouk University, Irbid, Jordan
murazug@hotmail.com

AWATEF MIZ'IL ABU-ALSHAAR
Al Al-Bayt University, Mafraq, Jordan

RÉSUMÉ

Cet article examine l'histoire et l'évolution des applications de la traduction automatique (TA) en langue arabe, dans le contexte de l'histoire de la TA en général. Il commence par décrire les débuts de la TA aux États-Unis et son déclin dû à l'épuisement du financement ; ensuite, son renouveau suscité par la mondialisation, le développement des technologies de l'information et les besoins croissants de lever les barrières linguistiques. Finalement, il aborde les progrès vertigineux réalisés grâce à l'informatique. L'article étudie aussi les principales approches de la TA dans une perspective historique. Le cas de l'arabe est traité dans cette perspective, compte tenu des travaux effectués par les instituts de recherche occidentaux et quelques sociétés privées occidentales. Un accent particulier est mis sur les recherches de la société arabe Sakr, fondée dès 1982, qui a mis au point plusieurs logiciels de traitement de langues naturelles pour l'arabe. Ces divers logiciels de TA arabe-anglais-arabe ainsi que des applications associées sont présentés dans un cadre historique.

ABSTRACT

This paper examines the history and development of Machine Translation (MT) applications for the Arabic language in the context of the history and machine translation in general. It starts with a discussion of the beginnings of MT in the US and then, depending on the work of MT historians, surveys the decline of the work on MT and drying up of funding; then the revival with globalization, development of information technology and the rising needs for breaking the language barriers in the world; and last on the dramatic developments that came with the advances in computer technology. The paper also examined some of the major approaches for MT within a historical perspective. The case of Arabic is treated along the same lines focusing on the work that was done on Arabic by Western research institutes and Western profit motivated companies. Special attention is given to the work of the one Arab company, Sakr of Al-Alamiyya Group, which was established in 1982 and has seriously since then worked on developing software applications for Arabic under the umbrella of natural language processing for the Arabic language. Major available software applications for Arabic/English Arabic MT as well as MT related software were surveyed within a historical framework.

MOTS-CLÉS/KEYWORDS

Arabic, historical perspective, Machine Translation (MT), Sakr

"Give me enough parallel data, and you can have a translation system in hours" A famous boast in the history of Engineering, echoing Archimedes' historic boast after providing a mathematical explanation for the lever, made by Franz Joseph Och after his

software scored highest among 23 Arabic and Chinese to English translation systems tested in a recently concluded US Department of Commerce trial. (Mankin 2003)

1. WHAT IS MACHINE TRANSLATION?

Machine Translation (henceforth, MT), sometimes called Automatic Translation (AT) has been defined as "the process that utilizes computer software to translate text from one natural language to another" (Systran 2004). This definition involves accounting for the grammatical structure of each language and using rules and assumptions to transfer the grammatical structure of the source language (text to be translated) into the target language (translated text). This definition also stresses the fact that machine translation is not simply substituting words for other words, but like human translation it involves the application of complex linguistic rules especially in morphology, syntax and semantics. This definition was widely accepted till the more dramatic developments in the history of MT took place recently when the statistical approaches to MT have started to gain ground as will be shown later in the paper.

The European Association for Machine Translation puts it simply as "the application of computers to the task of translating texts from one natural language to another." The association adds that "One of the very earliest pursuits in computer science, MT has proved to be an elusive goal, but today a number of systems are available which produce output which, if not perfect, is of sufficient quality to be useful in a number of specific domains." (European Association For Machine Translation 2004). Doug Arnold et al. (1994), in their book entitled *Machine Translation: Introductory Guide,* define it as "the attempt to automate all, or part of the process of translating from one human language to another." Technology Review in the *Christian Science Monitor* (April 22, 2004) asserts that "Universal translation is one of 10 emerging technologies that will affect our lives and work in revolutionary ways within a decade." This translation can be "unidirectional" translating in one direction as in the case of English into Arabic, "bi-directional" translating in both directions as from English into Arabic and from Arabic into English or even multidirectional translation back and forth between more than two languages or language pairs.

Another aspect can be added to this definition which is the presence of a computer system as an initiative for translation. Hahn (2004) distinguishes between "Autonomous" (Unassisted MT) initiative with the computer system where the target is what has been termed in the literature as Fully Automatic High Quality Translation (FAHQT) on one hand and Machine Aided Translation (MAT) where the user is asked to perform post editing and answer disambiguation/clarification questions on the other. Machine Aided Translation is human translation supported by a lot of help from computer systems. This help includes translation memory, lexical data, domain information and organizational support. The human cleans up after the translation in order to get better results. The text in Unassisted Machine translation results in what has been termed in the literature as "gisting" which is the gist of the source unpolished. (see Napier 2000). Another distinction is made between Human Aided Machine Translation (HAMT) or Computer Aided Translation (CAT) where the machine uses human help and Machine Aided Human Translation (MAHT) where the human uses machine help.

2. THE EARLY DAYS

W. John Hutchins (1995), an authority and widely published historian on MT, traces the beginnings of machine translation to the 17th century when the use of mechanical dictionaries to overcome the barriers of language was first suggested. But it was not until the twentieth century that concrete proposals were made with patents issued independently by George Artsouni in France in 1933 for a storage device on paper tape which could be used to find the equivalent of any word in another language and by Peter Smirnov-Troyanskii in Russia in 1937 who envisioned three stages of editing for mechanical translation; an editor knowing the source language to undertake the analysis of words, a machine to transform sequences into equivalents in another language and another editor to do analysis in the target language.

The inspiration for a machine that translates from one language to another, maintains Napier (2000), stemmed originally from code cracking of the second world war. Napier quotes a memo drafted by Warren Weaver who was the vice president of the Rockefeller Foundation and director of its natural science division and he was also an exceptional mathematician and coauthor of a text entitled "The Mathematical Theory of Communication." The memo which ushers in the beginnings of Machine Translation reads as follows:

> I have a text in front of me which is written in Russian but I am going to pretend that it is really written in English and that it has been coded in some strange symbols. All I need do is strip off the code in order to retrieve the information contained in the text.

Weaver, making use of the developments in computing and Shannon's information theory, outlined the prospects of machine translation and suggested different methods. This prompted research in some American universities a few years later, especially at the University of Washington at Seattle, the University of California at Los Angeles, and the Massachusetts Institute of Technology. The first Machine Translation Conference was convened in 1952, and according to *The Economist Technology Quarterly (2002)*, the first public demonstration of a MT system which came as a result of collaboration between IBM and Georgetown University took place in 1954. This early system which was based on "a simple bilingual dictionary with a few rules to determine word order, caused a surge of enthusiasm and funding." According to Wikipedia (2004) the demonstration was widely reported in the newspapers even though it was no more than "what today would be called a toy system." With 250 words and translating "49 carefully selected Russian sentences into English in the field of Chemistry," that demonstration pushed the perception of the imminence of MT and more importantly it stimulated research funding not only in the US but worldwide.

2.1. Optimism, High Expectations and Disillusion

It was a time of optimism and high expectation for machine translation as most MT historians put it. MT research tried also to overcome some of the basic problems they were facing in MT, chief among which is the problem of the "limitations of simple dictionary-based systems" using more developed approaches involving analysis of source texts using grammatical rules. The Economist (2002) cites the *Atlantic Monthly* in 1959 declaring that "Today, the computer, or the electronic brain, is well along toward picking up the burden of machine translation." However, the optimism

was turned into "disillusion" in the middle of the 1960s. The government sponsors of MT in the US formed a committee to evaluate and examine the prospects of this field. The Automatic Processing Advisory Committee (ALPAC) submitted its famous report in 1966 and the report concluded, as Hutchins (1995) states, that "Machine Translation was slower, less accurate, and twice as expensive as human translation and that there is no immediate or predictable prospect of useful machine translation." The report added that there was no need for further investment in MT research. Instead, the report recommended that research should focus on the development of systems to assist human translators rather than to replace them. The report had a profound influence on MT and funding for pure MT research "dried up." MT was for a decade after that perceived as a "complete failure."

It is revealing at this point to compare the findings of the ALPAC report to some of the claims made by Systran, a known machine translation software developer, a few years later. Translation and Interpretation (2004), a website, interestingly reports, based on information from Systran, that MT is much faster than human translation. Systran estimates that humans can translate 2000-3000 words a day while Systran's MT software is estimated to translate 3700 words per minute. Systran also claims that its software has a better memory than human translators in that it can store documents that have been translated and use phrases that have already been translated. Systran also claims that if MT is used to provide a first draft to be polished by a human translator, MT can save both time and money.

The decade after the ALPAC report was called by Hutchins "the quiet decade." Work continued in the US, apparently for political reasons related to the perception of the Soviet danger, on English translation of Russian scientific and technical materials. In Canada, there was a big demand on English-French translation. Systran sold a Russian to English translation system to the US Air force in 1970 and the same system was later adopted by the European Commission. It was in this decade when, according to the *Economist* (2002), that demand for translation systems started to emerge in business communities.

2.2. Revival

New developments worldwide in the fields of technology, political and socioeconomic trends starting in the 1980s contributed to a revival of Machine Translation advancement and research. These developments include the strides made in information technology, a rapid fall in the cost of computing power, globalization and increasing demand from multinational companies and governments for translation. These developments are by no means the prime mover of research and development behind MT; they just helped increase the pace of development. Translation and Interpretation (2004) asserts that research and development of MT has been going on since the 1950s "engaging some of the best minds in computing, linguistics and artificial intelligence," and cites an often quoted statement by Steve Silberman from an article of his entitled "Hello, World" which appeared in the publication Wired (May 2000) in which he says:

> The dream of translation by computer is older than the high tech industry itself. Before email, before word processing, before command-line interfaces, machine translation – or MT – was one of the first two computer applications designed to act upon words

instead of numbers (the other was code breaking)… But it turns out that really good MT is so hard to pull off that the task exhausted the top-end computing resources of every generation attempting it. Regardless, machine translation R&D is going stronger than ever, fired up by the globalization of the Net. Today, all over the world, software designers, programmers, hardware engineers, neural-network experts, AI specialists, linguists, and cognitive scientists are enlisted in the effort to teach computers how to port words and ideas from language to language." ("Hello, World," *Wired*, May 2000)

Since the 1980s, many new operational MT systems appeared, were expanded and were driven by the commercial markets. These systems have included: The Georgetown system developed in the mid 1960s; the French multilingual system TITUS; the Chinese English CULT system; the Spanish-English SPANAM by the Pan American Health Organization and the tailor made systems developed by the New York based Smart Corporation; the Systran Russian-English system which was adopted by the US Air Force and the European Community; the System of Logos Corporation. In Europe, the Commission of the European Communities (CEC) supported a lot of work on the English-French version of the Systran. Another system is that developed in Germany and called SUSY (Saarbrucker Ubersetzungssystem), the French German System (ASCOF) and (SEMSYN) for the translation of Japanese scientific articles into German. A more ambitious and reputable system developed in this era is the EUROTRA project of the European Communities. This project aimed the development multilingual transfer system for translating among all the Community languages. In the 1980s, according to Hutchins (1995), Japan maintained the greatest commercial activity where most computer companies developed software for computer aided translation mainly for the Japanese English Japanese markets. According to WTEC Hyper Librarian (1994), machine translation in Japan is viewed as an "important strategic technology that is expected to lay a key role in Japan's increasing participation in the world economy." Examples of Japanese MT industry include systems from Oki (PENSEE), Mitsubishi (MELTRAN), Sanyo, Toshiba (ASHITACI, HICATS) and Fujitsu (ATLAS). The most sophisticated system commercially available was METAL, a German-English system originated from research at the University of Texas at Austin with the support of Siemens which obtained commercial rights for marketing it. (Lehmann 2000: 162)

New factors in the 1980s caused serious revival of research and interest in MT to the point where the spirit of the age in translation is captured by Steve Silberman (AAAI Report 2005) when he strongly asserts that "a renewed international effort is gearing up to design computers and software that smash language barriers and create a borderless global marketplace." Most important of these factors referred to are the sharp falls in the cost of computing power, greater demand from governments and multinational corporations and the growing spread of globalization. Instead of work on rule manipulation of the syntax of different languages, according to the *Economist* (2002) the focus shifted to the development of interlingua systems which depended on translating the source text into an intermediate language or symbolic representation from which it could be translated into any of other several languages.

2.3. Dramatic Developments

The most dramatic development in MT took place in the 1990s as computers became more powerful and storage capacities much larger and cheaper. The new development shifts from grammar based approaches to what has been called "statistical approaches" emanating from the study of "corpus linguistics." Statistical translation systems do not depend on underlying grammatical rules any longer. Put very simply, as Farah (2003) put it in an article for the *New York Times* (reprinted in the *International Herald Tribune*), traditional MT relied heavily on bilingual programmers to enter the vast wealth of information on the lexicon and syntax that is needed by the computer to translate from one language to another. A team from International Business Machines (IBM) in the 1990s tried feeding a computer an English text and its translation in a different language, then by the use of statistical analysis, the computer learns the second language. The example given by Farah (2003) is revealing:

> Compare two simple phrases in Arabic: "rajl kabir" and "rajl tawil." If a computer knows that the first phrase means "big man" and the second means "tall man," the machine can compare the two and deduce that rajl means "man," while kabir and tawil mean "big" and "tall," respectively. Phrases like these, called N-grams (with "N" representing the number of terms in a given phrase), are the basic building blocks of statistical machine translation.

Researchers on MT are focusing more on the quality of MT especially after the introduction of Kevin Knight's newly developed software package called Egypt/Giza at the Information Science Institute at the University of Southern California. The package made statistical translation accessible to researchers across the US. Farah summarizes the impact of the new development saying:

> Today, researchers are racing to improve the quality and accuracy of the translations. The final translations generally give an average reader a solid understanding of the original meaning but are far from grammatically correct. While not perfect, statistics-based technology is also allowing scientists to crack scores of languages in a fraction of the time, and at a fraction of the cost, that traditional methods involved. A team of computer scientists at Johns Hopkins led by David Yarowsky is developing machine translations of such languages as Uzbek, Bengali, Nepali – and one from "Star Trek."

Mackin (2003), in an article interestingly entitled "Romancing the Rosetta Stone," reports on work on translation using statistical approaches, which is being pursued at the University of Southern California. Mackin quotes the USC computer scientist Franz Joseph Och boasting: "Give me enough parallel data, and you can have a translation system in hours.' This boast comes after Och's software was judged best in head to head tests against seven Arabic translation systems (five research and two commercial off the shelf products) and 14 Chinese systems (nine research and four commercial) by the US Commerce Department's National Institute of Standards and Technology. Makin asserts that Och's boast is a reminder of the Greek scientist Archimedes' historic boast when he said "Give me a place to stand on, and I will move the world" after offering a mathematical explanation for the lever. The new approach for translation uses huge volumes of "matched bilingual texts" which are the encoded equivalents to gigabites and gigabites of the Rosetta Stone inscriptions. Elaborating on this model in a workshop at Johns Hopkins University, Och (Makin

2003) asserts that the new approach uses statistical models to find "the most likely translation for a given input." Unlike the older approaches used for developing MT commercial systems largely depending on encoding the grammar and lexicon of a foreign language then analyzing and consequently producing English sentences based on hard rules, the new approach tries based on a statistical model to find the English sentence that is the most likely translation of the foreign input sentence. The new approach ignores or "rather rolls' over explicit rules of grammar and traditional dictionary lists of the lexicon in order to have the computer itself match up patterns between Arabic texts and English translations. Och's work (Makin 2003) is an improvement on earlier work on the statistical approach that started back in the late 1980s and early 1990s by Peter F. Brown and his colleagues at IBM's Watson Research Center. Much of the work also, as Makin (2003) states, was expanded originally at Aachen University of Technology (Rheinisch-Westfaelisschen Hochschule Aachen) where Och himself did post-doctoral work.

Up to now, it can be generalized that the quality of machine translation leaves much to be desired. It has been improving and it has been being improved by MT research and commercial enterprise because of its immense commercial prospects. But, especially in the case of Arabic, more work is still needed in the area of semantic representation systems without which it is difficult to achieve high quality translation as Aljlayl (2004) concludes while referring to the work of Aljlayl and Frieder (2001), Beesely (1998), and El-Dessouki et al (1998). Evidence reported in MT sources generally indicates that MT has scored real success in two major fields. The first of which is the field of "restricted language" or restricted subject matter where the syntax of the discourse is simplified, vocabulary is predictable and words tend to have one meaning because of the technical nature of this restricted variety. This is the language of technical documents, manuals, maintenance manuals and weather reports. Translation and Interpretation (2004) gives the example of the Meteo system which was developed in Montreal for translating Canadian Bulletins between English and French on a daily basis since 1977. The other field is what has been referred to in the literature as machine aided translation (MAT), computer aided translation (CAT) or "interactive translation" (IT). This kind of translation is intended for professional translators fluent in two languages and they seek a draft which would save time. Translation.net (1994-2003) lists the tools of CAT to include translation memory, terminology management, data banks and data bases.

3. OTHER IMPORTANT MT RELATED SOFTWARE DEVELOPMENT

Whenever the notion of Machine Translation is used, it is directly associated with the translation of the written language. The question in order here is whether it is really the case that Machine Translation is restricted to the written variety of language(s) targeted. The answer is negative, because MT work and research have been concerned with so many other aspects than the written language. AAAI (2005) reports on a number of MT related items that could have dramatic impact on the lives of the target populations. These included Speech to speech technology developed by NEC and tested in Papero (Partner –Type Personal Robot) which can translate "verbally between two languages in colloquial tongue (see News 2004); E-translators for universal translation (see Lamb 2004); Speech recognition (see Rupley 2004), Robotalk and pocket

translators (see Twist 2004); the "Verbmobile" a computer that translates between German, English and Japanese developed in Germany and it brings humanity "one step closer to the concept of artificial intelligence" (see Dw-world.de 2001); Computerized language translators for the military in several languages (see Associated Press 2002), Paula, a virtual interpreter, a digital English to American Sign Language (ASL) translator (see Associated Press/ National Geographic 2002), Avatars converting spoken English into text taken by avatars and the meaning signed on a screen (see Wurtzel, BBC 2002) Babylon and one way phrase translation system developed by the army, among many other developments.

4. THE CASE OF ARABIC

It must be made very clear at the outset that Arabic is one of the major languages on which some experimentation in MT was made in the very early days of MT and specifically in the US. When we bear in mind that one of the earliest motives behind MT was breaking the codes and monitoring various fronts of the enemy lines, especially those related to scientific, technical and military literature, we can explain why there was primary focus on Russian in the field of MT in its earliest days. Scanty evidence reported in the literature of the history of MT shows that Arabic ranked very high on the list of languages for which MT tools were to be researched and developed in the US fifty or so years ago. Muriel Vasconcellos (2000) narrated the story of the development of the Georgetown MT project and detailed the role of Leon Dostert in it. The project which depended on what they called General Analysis Technique (GAT) prevailed after the Pentagon demonstration changed and the initials reinterpreted it to stand for "Georgetown Automatic Translation." Georgetown got the largest grant to be given for Machine Translation from the CIA using the National Science Foundation as a front for this grant which accounted for 93.5% of the research money made available for MT research and development. That was in the late 1950s. Vasconcellos (2000: 92-930) writes:

> Even though the CIA (three words which one dared not speak aloud, Vasconcellos wrote earlier) grant was for research on Russian, Dostert believed that insights could be gained from linguists specialized in other languages. With this thought in mind, he invited A F. R. "Tony" Brown, then a professor of Semitic linguistics at the institute, to consider how he might address the MT task. Brown, using a French Corpus, wrote a program which he called "Simulated Linguistic Computer." Eventually he took over the programming of GAT. **Arabic, which had been designated a priority language by the US government was taken by Nancy Kennedy, a graduate student at the institute** [emphasis mine].... All these people came together at the Friday morning seminar which was the highlight of the week. The various groups and individuals would take turns presenting their work and answering questions from colleagues. Sometimes discussions got rather heated. Yngve (2004) reports that Arabic was one of the languages besides English, German and French which were the subjects of the COMIT project in the late 1950s. Attempts at parsing algorithms in a three step scheme were made.

Even though certain aspects of Arabic were researched in the early days of machine translation, the language has always been considered "due to its morphological, syntactic, phonetic and phonologic properties one of the most difficult languages for written and spoken language processing" (Boualem 2003). The same author continues to state in a recent call for papers for a major conference on Arabic Language Processing:

Research on written Arabic language processing started in the 1970s, even before the problems of Arabic text editing were completely solved. The first studies focused primarily on lexicons and morphology. In the past ten years, the internationalization of the WWW and the proliferation of communication tools in Arabic have led to the need for a large number of Arabic NLP applications. As a result, research activity has extended to address more general areas of Arabic language processing, including syntactic analysis, machine translation, document indexing, information retrieval, etc.

Research on Arabic speech processing has made significant progress due to more improved signal processing technologies, and to recent advances in the knowledge of the prosodic and the segmental characteristics of Arabic and the acoustic modeling of Arab schemes. These results should make it possible to further progress in more innovative areas, such as Arabic speech recognition and synthesis, speech translation and automatic identification of a speaker and his/her geographic origin discrimination, etc.

In a very interesting paper entitled "Toward Corpus Based Machine Translation of Arabic," Guidere (2002) distinguishes between two approaches for the study of machine processing of Arabic at large. These are the "particularistic" approaches which delineate the linguistic idiosyncrasies of Arabic and use them for a local processing approach specific to the internal linguistic system of Arabic. Such approaches have been concerned with the morphological and semantic aspects of Arabic language, especially the trilateral root system. Sakhr, the only and foremost Arabic speaking group working systematically on Arabic, put the particular aspects of Arabic at the top of its priorities in developing software solutions. Sakr (2004) asserts:

The Arabic language differs tremendously in terms of its characters, <u>morphology</u> and <u>diacritization</u> from other languages, and to claim otherwise would be a mistake. Furthermore, to import solutions from these other languages would only be at the expense of the unique features of the Arabic language.

Systran, a company that has been developing software applications for different languages in the world including Arabic specified the following points which they called "facts that help in translating Arabic" (TranslationSoftware4u.com 2004):

Facts that help in Translation of Arabic

- Arabic is written from right to left in a horizontal form.
- Arabic writing sits on the line.
- There are no capital letters in Arabic.
- Punctuation is similar to English except for comas which sit on the line instead of under the line.
- Arabic uses gender for all known nouns, no neutral ones.
- Space is left between words in a sentence.
- Some letters change shape depending on whether they are at the start, in the middle or at the end of the word.
- There are 29 letters in Arabic – with 3 letter sounds which do not even exist in the English language.
- Arabic does not distinguish between vowels and consonants; the use of a small sign on the top or under the letter indicates the pronunciation

The other approach which is viewed as complementary to the "particularist" approach is called the "universalist" approach which explores possibilities of applications of

methods already tried on other languages like English or French with or without adaptation. These latter approaches focus on the syntactic aspects of the linguistic system in general. Regardless of which approach has been used, Guidere (2002) interestingly concludes that the few systems of machine translation available to and from Arabic primarily concern the Arabic English pair and "in reality constitute improved versions of electronic dictionaries." Guidere continues to assert that other available applications by well known companies have a restricted coverage of Arabic linguistic phenomena and they are essentially based on specialized dictionaries. These applications are technical translation aids rather than machine translation software packages. As this discussion shows, there is some kind of polarization between work based on the "particular approach" as claimed by Sakhr on one hand and the "universalist" approach already followed by most of the companies producing software application for Arabic.

5. MODERN WORK ON ENGLISH/ARABIC/ENGLISH MACHINE TRANSLATION

It is to be pointed out here at the beginning of this section that apart from one company in the Arab world, there has been no serious work on the development of software packages of machine translation or other applications of natural language processing for Arabic in its areas of influence, i.e., where it is a first language. The Arab universities, educational institutions, research centers and other relevant organizations have seriously lagged behind in the area of Information Technology research and Arabic language related natural language processing in particular. They left this kind of research and development to the profit driven markets in the West which benefit greatly from research work at universities, research centers and educational institutions (see Makin 2003, referred to earlier, and the outstanding work of Och at USC on a statistically based translation system of English/Arabic). There has been individual work, of course, but it has never been coordinated and the individual work has also been done by Arabs working in Western educational institutions (see Aljlayl et al 2004, Aljlayl et al 2001, El-Dessouki et al 1989). The one company referred to earlier is Sakr Software, a branch of al-Alamiah Group and it was established in Kuwait in 1982 and then relocated after the Iraqi occupation of Kuwait in 1990 to Egypt. The company boasts in its literature of appointing no less than 100 specialists in information technology to work on Arabic language software applications. On their website, Muhammad Al-Sharekh, Chairman of the Board of Sakhr Software, summarizes their objectives and research work as follows:

Sakhr Software has directed its every effort over the past twenty years towards enlisting computer and communication technology to serve the unique needs of the Arabic language, and not vice versa. The Arabic language differs tremendously in terms of its characters, morphology and diacritization from other languages, and to claim otherwise would be a mistake. Furthermore, to import solutions from these other languages would only be at the expense of the unique features of the Arabic language. This was apparent to us from the start, so we undertook the development of the Automatic Morphologizer, the Automatic Diacritizer and the electronic dictionaries. Our ardent dedication to the research and development of these tools formed the cornerstone of all future Sakhr products and Internet solutions, including Automatic

Speech Recognition, Machine Translation and Electronic Publishing solutions. Our investment in Natural Language Processing, or NLP, has yielded outstanding results, and is expected to yield even greater results as the demand for Natural Language Processing in the exchange of information grows bigger.

Indeed, it is worth stating here that Sakhr, whose products go under other names internationally (See Armedia 2004), has been successful in the creation of a number of "CORE" software which makes the processing of Arabic possible and easier. It makes the task of developing software packages for various applications such as machine translation, publishing, and development of electronic dictionaries possible. Sakhr has done to Arabic the basic work that was done earlier to European languages and mainly to English. Following is a list of the software developed by Sakhr to make Natural Language Processing of Arabic feasible. The following list was reported by Armedia (2004).

1.   Arabic Optical Character Recognition (A-OCR). Work on developing OCR started in 1993.

2.   Multi-Mode Morphological Processor (MMMP) This package is claimed to be a morphological analyzer-synthesizer of Arabic. The analyzer identifies all possible stem forms of a word, i.e., extracting its basic form stripped from affixes. Unlike the English Stemmer, the MMMP analyzer does not stop at the stem level but proceeds to extract the root and the Morphological Pattern (MP) of the word. Decomposing Arabic words into their morphological primitives is a basic requirement for full text indexing, search, dictionary organization and look up, as well as for spelling and grammatical checking. Even more important, the MMMP enables deeper processing of Arabic at the syntax and semantic levels. The MMMP synthesizer works in a reverse mode to generate linguistically-correct final word forms. The synthesizer is a key tool for generating the required output in machine translation systems and other text generation applications, such as summarizers and style checkers.

3.   Multi-mode Syntactic Processor (MMSP) MMSP parses the Arabic sentence into its constituents Verb, Subject, Object, Adverb, Predicate …etc.

4.   Arabic Automatic Diacritizer (AAD). AAD handles unvowelized Arabic texts. In other words it provides the Arabic diacritical markings which indicate syntactic functions of the word in the sentence. It is claimed to stimulate "the mental process exercised by Arabic native speakers in interpreting undiacritized text and substituting missing vowels. The Automatic Diacritizer provides different options for diacritization: full, mandatory, or case ending diacritics. The AAD is the entry point for rendering written Arabic text suitable for serious computation."

5.   Arabic Text Fragmenter (ATF) ATF automatically divides the text into sentences. It serves as a basic front-end processor, which prepares narratives for sentence-based processors such as parsers and for machine translation.

6.   Arabic Automatic Indexer (AAI) examines the content of a document to identify key words and phrases. It enables the creation of book indices for Arabic books. AAI has different levels of indexing and has an HTML version for the Internet.

7. Arabic Text to Speech (TTS) and Automatic Speech Recognition (ASR) engines. The TTS engine converts any Arabic computer readable text into a human sounding synthetic voice. The ASR engine recognizes Arabic utterances and commands from different speakers and different accents.

8. The Summarizer. It is used for summarizing Arabic and English documents. It extracts the main ideas, based on linguistic analysis of the document, to make it possible for the user to preview these ideas instead of reading the whole document,

9. Johaina. A news search engine which translates news from different sources into Arabic. It has both a navigation and monitoring service. (see Armedia 2004-Johaina)

10. IBSAR. An integrated bilingual solution for the blind or visually impaired, in the Arab speaking countries, Ibsar for Windows works with any PC to provide access to most of today's software applications and the Internet. With its Arabic/English Text To Speech (TTS) engine and the computer's soundcard, information is read aloud, providing access to a wide variety of information. IBSAR also reads documents, printed books with Sakhr's Optical Character Recognizer and enables users to print on normal and Braille printers. (see Armedia 2004-Ibsar). According to the company's website, IBSAR has the following features:

    • Provides self-learning for the blind using a computer.
    • Maintains the user's privacy and independence.
    • Reads the screen's output and any pressed key on the keyboard.
    • Spells every word in a program.
    • Integrated Spell checker.
    • Reads and navigates any Web page in IE directly.
    • Collects links within web pages and then uses the keyboard to select the desired link.
    • Searches for any (Arabic/English) information on the Web.
    • Writes, reads and sends email messages using Microsoft Outlook or popular webmail services like Hotmail and Yahoo.
    • New Integrated Arabic. Speech synthesizer to speed up reading.
    • Integrated OCR 7.1.
    • New enhanced Integrated Arabic diacritizer.
    • Reads all the details within the active dialogs and windows.
    • Supports Microsoft Excel and Microsoft Word.
    • Reads tables in Microsoft Word.
    • Supports fast printing using a Braille printer (full control in Word, Notepad, etc).
    • Integrated Dictionary.

11. IDRISI general search engine. It is claimed to have the following features. (see ArmediA 2004-idrisi)

    • Integrates with INSO filters that support all Microsoft Office formats and other standard formats such as HTML, TXT, RTF and WRI.
    • Supports multiple code pages, whether from different platforms (such as Macintosh) or different languages (such as French).
    • Complies with Windows NT security schemes, with control of access to data.
    • Is capable of handling, indexing and updating data automatically allowing end-users to get the latest and most accurate information without manually rebuilding and re-indexing collections.
    • Is a fully bilingual search engine, supporting both Arabic and English features.

- Provides users with customizable search and result templates. These templates cover all the options that an end user may need, from simple or compact results to advanced detailed information about the number of hits.

12. Arab DOX. An English Arabic French Document management system. (see Aramedia 2004-ArabDOX)

13. Sakhr Corrector. The corrector automatically detects and corrects Arabic spelling mistakes as well as grammatical mistakes. (see AramediA 2004-corrector)

14. Sakr Categorization Engine. Sakhr Categorization Engine and organizes valuable information into a topic, tree or taxonomy. (see AramediA 2004-categorization)

6. ACCELERATION OF WORK ON ARABIC

Apart from Sakhr, the interest in developing translation related software packages for Arabic has been sharply growing for three major reasons. The first is related to globalization, developments in information technology and the international giant strides in communication. The second is commercial and is represented by the many companies in the industrial world which were led by the tremendous prospects for business in the Arab world to invest in the development of different translation related software applications. As will be shown in this part of the paper, most of these countries are either American or Japanese. Systran was founded in 1968 in San Diego in the US. On its home page (Systran 2004-company), Systran maintains that it is "the leading provider of the world's most scalable and modular translation architecture. Its core technology powers revolutionary translation solutions for the Internet, PCs and network infrastructures that facilitate communication in 36 language pairs and in 20 specialized domains. Their expertise, they assert, "spans over 30 years of building customized translation solutions for large corporations, portals, ISPs, governments and public administrations through open and robust architectures." They have done work for major corporations in America including Ford Motor Company, Cisco Systems, NCR, DaimlerChrysler Corporation, PricewaterhouseCoopers, Dow Corning Corporation, and others. They also did work for Google, AOL, Altavista, Apple's Sherlock Internet Search, CompuServe, Lycos, OracleMobile.com, and others. They contracted the European Commission and the US Intelligence Community. ATA (ATA Software 2004) is another major company producing software targeting Arabic English Machine Translation and it considers itself as a world leader in English Arabic Machine Translation. The company asserts, on its home page (ATA software 2004-company), that it is a London based company specializing in Arabic Software production. It was established in 1992 by programmers and specialists whose experience goes back to the 1980s. In 1995 it released Al-mutarjim Al-arabey. Another known company, especially in the area of electronic dictionaries, notably handheld electronic translators and pocket multilingual talking dictionaries, is ECTACO, an American based company established in 1990. ECTACO asserts, on its homepage (About ECTACO-2004-company), that it is "the world leader in the development and production of electronic handheld dictionaries. LingvoSoft™ is a registered trademark and an ECTACO division delivering translation software." Since its foundation in 1990, the company maintains that it has produced 7 generations of electronic dictionaries of **Language Teacher**® and **Partner**® brands for over 45 languages. Another company

with products in the Arabic market is CIMOS, a French based company that was established in 1997. The company, according to its profile (CIMOS 2004-profile), is interested in Translation and Localization Multilingual Processing Development of linguistic Software. Their products have included Automatic Translator Software (English, French, Arabic), Universal Semantic Analyser General and Specialized dictionaries (English, French, Arabic), Topic and thesaurus dictionaries (English, French, Arabic), and Tools for NLP (Natural Language Processing), Tools for NLU (Natural Language Understanding), Translation and Localization Services. CIMOS produced morphological Analyzer, grammatical analyzer, automatic vocalizer and a number of electronic dictionaries.

It is worth stating that the bulk of the work has so far focused on Arabic/English/Arabic. More recently, and specifically after the events of Sept the 11th, the demand for translating Arabic increased. Many companies "immediately made Arabic translation a priority. The Newsletter of Systran Translation Solutions (TranslationSoftware4u.com) (2004) quotes Everett Jordan, Director of the National Virtual Translation Center, an organization jointly sponsored by the FBI and CIA under the USA Patriot Act. When talking about their English/Arabic/English translation software they say:

Linguistics technology is beginning to play an increasingly important role when it comes to ensuring national security. Because of the enormous volume of multilingual intelligence information that must be analyzed with limited human resources, technologies that can assist in sifting, sorting, and finding critical information are essential in ensuring that threats are detected as quickly as possible. Whereas the US Government cannot endorse any one product over another, we are pleased to see that companies are responding to the government's call for solutions to these difficult issues.

7. SOME AVAILABLE SOFTWARE APPLICATIONS FOR ARABIC

The translation related software applications available can be classified in the following major categories:

7.1. Unidirectional, bi-directional and multidirectional general machine translation systems.
7.2. Translational systems directed to Web translation.
7.3. Computer aided translation systems.
7.4. Unidirectional, bi-directional and multidirectional electronic dictionaries.
7.5. Other translation related software packages.

7.1. Unidirectional, bi-directional and multidirectional general machine translation systems:

–   **Tarjim**: A bi-directional English/Arabic English Machine translation system developed by Sakr and available on Ajeeb website (http://tarjim.ajeeb.com/ajeeb/). It also translates web sites and web pages from English into Arabic and vice versa.

–   **Al Mutarjim Al Araby V.2.0, V3.0** A machine translation system For English/Arabic developed by ATA Software Technology.

–   **AL-Wafi Translator V.2.0** (discontinued), a smaller version of Al Mutarjim Al Arabi.

– **Al-Wafi v4.00.** English/Arabic translation system developed by ATA Software.

– **Al-Misbar**. ATA Software English/Arabic uni-directional translation system as well as a website and internet (URL) translation system. It is an online free service. (see ATA Software –2004)

– **MutarjimNet**. A translation system for companies and institutions that provides a translation network for their employees. (ATA Products 2004-Products).

– **Al-Nakel El-Arabi**: Machine translation system for English Arabic. General dictionaries with over 100,000 words and phrases. Special dictionaries on banking, commerce, computers, law, petroleum, gas production and trade. Available are English/Arabic bi-directional Al-Nakel, Single directional English Arabic, Single directional Arabic/English and bi-directional French and Arabic. (see Aramedia 2004-nakel) advertised also by CIMOS (2004-index).

– **TranslateNet.** Arabic/English/French, also translates web sites. Developed by CIMOS.

– **English to Arabic Translation/Arabic to English Translation**. A bi-directional translation system developed by Systran.

– **Weinder WCC.** Machine translation system for English/Arabic, developed in the USA, commercially available since 1980; Micro CAT for PCs and Macro CAT for Minicomputers.

– **Multilingual Machine Translation**, developed for a dozen pairs of languages. It goes under the name of Apptek products. (see Aramedia 2004-nakel)

– **SYSTRAN 5.0**. A Machine translation system developed by Systran for bi-directional machine translation of a number of world language pairs including Arabic/English and English/Arabic. (Systran 2004-Systransoft)
SYSTRAN'S features include (SYSTRAN 2004-mainentry):
•   Translates directly from your Office desktop with Microsoft® Office plugging for Word, Outlook, Excel, and PowerPoint files.
•   Real–time translation of Web pages with plugging for Internet Explorer™.
•   PDF plugging allows you to easily translate business–critical PDF documents into Word.
•   Includes 5 specialized dictionaries – Business, Colloquial, Industries, Sciences, Life, covering 20 domains.
•   Integrated Translation Memory – supports TMX standard.

7. 2. Translational system directed to Web translation:

– **Sakhr Enterprise Translation (SET)** (see Armedia 2004-Set) Web based English Arabic bi-directional Machine Translation Solution. It is targeted for companies and institutes that have a high flow of information to be translated. It translates documents and web pages and it can create memory databases and glossaries to assure consistency of translation. SET is claimed by Armedia to have the following features:

1. Automatic translation from English ⇔ Arabic.
2. Ability to save your translated sentences in a database called translation memory for future reuse. This will enhance the translation quality especially when you translate files in the same domain.
3. Ability to add your own glossaries to enhance and customize translation output.
4. High-speed translation for the common file formats: html, rtf, and txt.
5. Web pages translation, of any size.
6. Lookup for meaning of words in a bi-directional dictionary, Arabic ⇔ English.
7. Integrated bilingual spell checker to correct spelling mistakes before translation.
8. Administrator has full control on the account and can define access rights for each user.
9. Ability to submit any document for high-quality human translation. Our professionals translate this document in a timely manner with a competitive price.
10. Easy access from anywhere, just type your login name and password.
11. Simple, user-friendly web interface with a detailed help section.
12. Benefit from daily updates and ongoing developments on our engines and glossaries.
Sakhr's Easy Lingo Instant Translator. English Arabic. (see Aramedia 2004-homeinter)

7.3. Computer aided translation systems:

–   **CAT Translator Workbench**, CAT Translator Enterprise. A Translator Workbench developed by Sakhr and meant to be a Computer-Aided Translation system that supports bi-directional, bilingual translation between English and Arabic. Sakhr Translator Workbench System is said to apply Natural Language Processing (NLP) technologies to both Arabic and English as source languages. (see Armedia 2004-catrans)

7.4. Unidirectional, bi-directional and multidirectional electronic dictionaries.

–   **Al-Wafi School Dictionary V1.00.** English /Arabic/English.
–   **Al-Wafi English/Arabic**/English developed by Ata Software.
–   Pocket Electronic Translators:
–   **ECTACO English/Arabic/English Talking Partner for Pocket PC.**
–   **ECTACO language Teacher English/Arabic/English**
–   **ECTACO Police Speech Guard PD-4**
–   **ECTACO Military Speech Guard GI-4a Arabic, English**
–   **ECTACO Medical Speech Guard MD-4**
–   **ECTACO Partner EAF430T Arabic – French – English**
–   **Bidirectional Trilingual Talking Arabic/English/French Dictionary. Talks English and French.** It has medical, technical, legal, business terms, as well as slang, idioms, and general expressions. (see Aramedia 2004-talkdic2)
–   **Sakhr's Bilingual Dictionary Al-Qamoos**. Arabic /English/Arabic Dictionary with Arabic Synonym Dictionary, English Synonym Dictionary, Arabic Antonym Dictionary, English Antonym Dictionary. (see Aramedia 2004-diction)
–   **E. W. LANE Arabic English Lexicon:** The special feature of this lexicon is the fact that entry in Arabic is by root. The different derivatives of the root are translated into English. (see Armedia 2004-ewlane)

- **Sakhr's Al-Qamoos Multilingual Dictionary.** Arabic/English/French/German/
  Turkish (see Aramedia 2004-dictionaries)
- **World Translator™** It provides the most advanced Multi-Language Translating
  Dictionary in over 100 languages. (Multimedia 2004-world trans dictionary)
- World Translator™ Multi-Language Translating Dictionary Packages Contents:
  Arabic ⇔ Arabic Dictionaries
  Arabic ⇔ Arabic, English & French Dictionaries
  English ⇔ Farsi Dictionary
  English ⇔ French, German, Spanish, Italian Brazilian Portuguese – Standard
  English ⇔ French, German, Spanish, Italian Brazilian Portuguese – Professional
  English ⇔ French, German, Spanish, Italian Brazilian Portuguese – Advanced
  Multi-language European Dictionaries Package (25 Languages, 30 Dictionaries)
  Bosnian, Bulgarian, Croatian, Czech, Danish, Dutch, Finnish, French, German, Greek,
  Hungarian, Icelandic, Italian, Latin, Norwegian, Polish, Polish (law), Polish (Business),
  Portuguese (Brazilian), Portuguese (Portugal), Romanian, Russian, Russian (Aero).
  Russian (Business), Russian (Mine), Serbian (Cyrillic), Serbian (Latin), Slovenian,
  Spanish, Swedish.
- **Sakhr's Multilingual Islamic Dictionary**. A dictionary intended for Muslim
  scholars and those interested in Islamic religion. It has the following features
  (see Aramedia 2004-dictionary):
  The program includes more than 55,000 items and meanings, translated into six
  languages: English, Malay, Indonesian, Turkish, French and German. It is outstanding
  in its simple display, fast search and accurate meaning of items included in the fields of
  Faith, Jurisprudence, Economy, History, Islamic art and Architecture, in addition to
  sections tackling many other fields supported by multimedia technology. It displays an
  indexed dictionary of 35,000 words from the Holy Qur'an with their meaning. It also
  cross-references these words to verses where they are found with recitation. It includes
  the meaning of more than 2,500 difficult words mentioned in the Prophetic Hadiths,
  linking them to the Hadiths containing them. It also includes terms on the Science of
  Hadith and their meanings. It presents a brief explanation of many juristic terms
  including terms of Zakah, transaction, Islamic economy in addition to an Economic
  Subject Tree linked to the Qur'anic Verses. The history section includes historical
  terminology and the most important battles, conquests, treaties and historical places. It
  provides an Islamic World Atlas focusing on the significant information on each
  Islamic country.

7.5. Other translation related software packages.

- Sakhr software, for example, has been working in Arabic NLP for about 12 years,
  and in English NLP for about 4 years now. 3 years ago, Sakhr initiated their Bidi
  English ⇔ Arabic MT project (Translate Model). Sakhr has now produced ver-
  sion 1 of the MT engine that has been used to develop Tarjim.com site, which is
  the first Web on-line translation English to Arabic on the Internet of which the
  average accuracy is ~ 60%.

8. CONCLUSION

Work on Machine translation is a long chapter in the history of the collaboration
between linguists and computer scientist and it has gone through ups and downs
depending on many non-linguistic and non-computer related variables. It has

depended on the political atmosphere prevailing at certain times, on economic factors, on availability of funding, on how convinced the decision makers in both governments and the private sectors are concerning the feasibility of this project. More recently, with the revolutions in information technology, the strides in computer development at both levels of hardware and software, and with the political atmosphere overwhelmed by globalization and concern for security, MT has been given real boosts and with the collaboration of research institutes and industry, attemps are being made to perfect MT. Moreover, MT has prospects for dramatic impact on the lives of the deaf and the blind. Undoubtedly, the bulk of the work has been on the translation of English into other languages and the translation of some languages of interest to the US in particular into English. Pashtu, for example, does not rank high among world languages in terms of number of speakers, but it is a language of concern for the US being one of the languages of Afghanistan.

Arabic is an important language in all aspects mentioned. It is the sixth language in the world as to the number of speakers, it is of high concern for the US at the political level, and it is potentially a tremendous market. Despite these factors, work on Arabic lagged behind work on major languages in the world. Work on Arabic started effectively in the 1980s and accelerated, along with MT work on other languages, in the 1990s. It has been shown to work in English/Arabic/English. Work on MT for Arabic started in the West by research institutes at American and some European universities and by Western profit driven companies. The establishment of the one Arab company, Sakr – Al-Alamiyyah Group, with serious interest in developing software applications for Arabic, has had a significant bearing on MT as well as on developing different programs for the natural processing of Arabic for different purposes. Major software applications for unidirectional, bi-directional and multidirectional translation developed by Western companies as well as Sakhr were listed. The listing remains partial because of the expanding market for some of these applications. If we take electronic dictionaries, for example, we will find there is a multiplicity of these dictionaries developed by Western companies, and a growing number of these being developed by Japanese companies.

REFERENCES

AAAI (2005): "Machine Translation"
    <http://www.aaai.org/AITopics/html/machtr.html>
About ECTACO (2004): "About ECTACO)
    <http://www.ectaco.com/company/index.php3?refid=831>
    <http://arabic.dictionaries.ectaco.com/>
ALJLAYL, M., FRIEDER, O. and D. GROSSMAN (2004): "On Arabic-English Cross Language Information Retrieval: A Machine Translation Approach.
ALJLAYL, M. and O. FRIEDER (2001): "Effective Arabic-English Cross-Language Information Retrieval via Machine Readable Dictionaries and Machine Translation," ACM Tenth Conference on Information and Knowledge Management, Atlanta, Georgia.
Aramedia (2004): "Arabic Software, Desk Top Publishing, Machine Translation, Document management, NLP, OCR, ASR, TTS and Multimedia"
    <http://www.aramedia.com/nlp.htm>
    <http://aramedia.com/speech.htm>
    <http://aramedia.com/set.htm>
    <http://aramedia.com/johaina.htm>

<http://aramedia.com/ibsar.htm>
<http://www.aramedia.com/arabdox.htm>
<http://aramedia.com/correction.htm>
<http://aramedia.com/categorization.htm>
<http://aramedia.com/catrans.htm>
<http://www.aramedia.com/talkdict2.htm>
<http://www.aramedia.com/worldtransdictionary.htm>
<http://aramedia.com/diction.htm>
<http://aramedia.com/dictionary.htm>
<http://aramedia.com/homeinter.htm>
<http://www.aramedia.com/dictionaries.htm>
<http://aramedia.com/ewlane.htm>

AScribe Newswire (2002): "Computer Program Translates Spoken English into Sign Language" *National Geographic*, August 12.

Associated Press(2002): "U.S. Soldiers Get Talking translators" October 7.

ATA Software (2004): "ATA Products"
<http://www.atasoft.com/products.htm>
<http://www.almisbar.com/about_trans.html>
<http://www.atasoft.com/company.htm>

Automatic Translation (2004):
<http://www.uni-leipzig.de/~xlatio/software/soft-axlo.htm>

Beesley, K. (1998): "Arabic Morphological Analysis on the Internet" Proceedings of the 6th International Conference and Exhibition on Multi-Lingual Computing. Cambridge.

Boualem, M. (2003): "CFP – JEP – TALN – Special Session on Arabic Language Processing"
<http://www.mail-archive.com/mt-list@eamt.org/msg00612.html>

CIMOS (2004): "Company Profile"
<http://www.lang-tech.org/exhibition/descriptions/cimos#compinfo>
<http://www.cimos.com/index.asp?src=produits>
<http://www.cimos.com/index.asp?src=produits>

DW-WORLD.DE (2001): "Another Step Closer To Artificial Intelligence" December 1.

Economist. (2002): "Tongues of the Web." Vol. 362, Issue 8264, p26, 39, 2c.

EBSCOHOST (2004): Academic Search Premier, Accession No. 6347997.

European Association for Machine Translation (2004): "What Is Machine Translation?"
<http://www.eamt.org/mt.html>

Farah, C. J. (2003): "Big Leap in Machine Language Translation" Written for the *New York Times*, Reprinted in the *International Herald Tribune*, Thursday, July 31, 2003.
<http:/www.iht.com/articles/104781.html>

Guidère, M. (2002): "Toward Corpus-Based Machine Translation for Standard Arabic" Translation Journal, 6, 1.
<http://accurapid.com/journal/19mt.htm>

Hahn, W. (2004): "Machine Translation"
<http://www.racai.ro/awd/awd16/hahn.html>

Hutchins, W. J. (1995): "Machine Translation: A Brief History" In E. Koerner and R. Asher (eds.), *Concise History of the Language Sciences: From the Sumerians to the Cognitivisits*, Oxford: Pergamon Press.
<http://ourworld.compuserve.com/homepages/WJH Hutchins/Conchist.htm>

Lamb, G. M. (2004): "E-Translators – The More you Say The Better" *The Christian Science Monitor*, April 22.

Lehman, W. P. (2000) "My Early Years in Machine Translation" In W. John Hutchins (ed.) *Early Years in Machine Translation*. Amsterdam: John Benjamins, 2000, p. 147-165.

Mackin, E. (2003): "Romancing the Rosetta Stone."
<http://www.eurekalert.org/pub_releases/2003-07/uosc-rtr072403.php>

Markowitz, J. (2003): "I See What You Are Saying" Speech Technology Magazine. September/ October.
Napier, M. (2000): "The Soldiers are in the Coffee: An Introduction to machine translation" <http://www.cultivate-int.org/issue2/mt/>
Rupely, S. (2004): "Scaling The Language Barrier," July 13.
Sakhr Software (2004): "About Us" <http://www.sakhr.com/Sakhr_e/Technology/MT.htm?Index=5&Main=Technology&Sub= Machine>
Systran. (2004): "Questions and Answers" <http://www.alarmsbc.com/sysfaq.htm> <http://www.systransoft.com/company/index.html> <http://www.systransoft.com/> <http://www.digitalriver.com/dr/v2/ec_MAIN.Entry10?V1=641910&PN=2&SP=10023& xid=28102>
"Technology Review" (2004) in the *Christian Science Monitor* (April 22, 2004)
Translation and Interpreting (2004): "Machine Translation" <http://www.diplomacy.edu/Language/Translation/machine.htm>
Translation.net (1994-2003): "The Universal Translation Frequently Asked Questions –FAQ" <http://www.translation.net/tsfaq.html>
TranslationSoftware4u.com (2004): "English to Arabic Translation; Arabic to English Translation" <http://translationsoftware4u.com/arabic-translation.htm>
Twist, J. (2004): "Robo-talk Helps Pocket Translator," BBC News, March 4.
Vasconcellos, M. (2000): "The Georgetown Project and Leon Dostert Recollections of a Young Assistant." In W. J. Hutchins (ed.) *Early Years in Machine Translation.* Amsterdam: John Benjamins, p. 87-96.
Wurtzel, J. (2002): "Digital 'Characters' Talk to the Deaf," BBC. March 2.
Wikipedia (2004): "Machine Translation" <http://en.wikipedia.org/wiki/Machine_translation>
WTEC Hyper Librarian (1994) "Machine Translation in Japan," March Issue File//A: /Machine Translation in Japan.htm
Yngve, V. (2000): "Early Research at MIT in Research of Adequate Theory." In W. J. Hutchins (ed.) *Early Years in Machine Translation*, Amsterdam: John Benjamins, p. 39-72.