

# The Language Professional's Guide to... **The ABC of Machine Translation — E-M**

by W. John Hutchins and Geoffrey Kingscott

*This continues the alphabetical guide to Machine Translation published in Language International issue 2.3 (1990). For readers who wish to keep these informative pages together, the MT Guide is published in the centre of each issue and can be taken out and filed together. MT-ppf refers to John Hutchins's book, Machine translation: past, present, future (Ellis Horwood, 1986) and ND-MT refers to his updating paper "Recent Developments in Machine Translation" (in New Directions in Machine Translation, Foris, 1988) where further information can be obtained.*

## E

**ECHA** A system being developed in the People's Republic of China for translation from Chinese to Esperanto and English.

**EDR** Electronic Dictionary Research Institute, a collaborative project in Japan to compile dictionaries for use in MT systems. (*ND-MT*, pages 44-45)

**Electrotechnical Laboratory** An institution in Tokyo, Japan, which was a pioneer in work on English to Japanese MT.

**ENGSPAN** As a result of the success of using its SPANAM (q.v.) system for translation from Spanish to English, the Pan-American Health Organisation in Washington DC began work in 1982 on developing an English to Spanish system, ENGSPAN. The original basic dictionary was created by reversing the existing SPANAM dictionary, then amending and adding as necessary. ENGSPAN is a batch processing system with post-editing. (A detailed description by Dr Muriel Vasconcellos, who helped to develop the system, and now oversees its application, can be found in Slocum: *Machine Translation Systems*, Cambridge University Press, 1998 pages 187-235.)

**ENTRA** A project in the mid-1980s to use the existing English to German Weidner program as a starting point for the development of an English to Norwegian system, sponsored jointly by the World Communications Corporation (the then Weidner organisation) and the University of Bergen; no recent information is available on this project. (*ND-MT*, page 27)

**ETAP-2** An English to Russian MT system for electrical engineering texts being developed in the USSR; little recent information is available. (*ND-MT*, page 77)

**Eurotra** One of the world's largest current MT projects, Eurotra is an attempt by the European Commission to produce a system, using the latest computational and linguistic techniques, which would be fully multilingual between the working languages of the Community. First discussions were held in February 1978, and development began the following year. Researchers in each Member State of the Community have been working on development, with coordination in the hands of a small team under Serge Perschke in Luxembourg. Languages involved in Eurotra are Danish, Dutch, English, French, German, Greek, Italian, Portuguese and Spanish. It is hoped to produce shortly a prototype capable of translating some texts in the limited domains of

information technology and official Community documents.

(*ND-MT*, pages 31-33; see also for more detailed information on how Eurotra came into being the interview with Serge Perschke which appeared in issue 2.5 of *Language International* and *Danzin report is out — new assessment of Eurotra*, in issue 2.4.)

## F

**FAHQT** Acronym for Fully Automatic High Quality Translation, a term devised by Yehoshua Bar-Hillel in 1959 to denote the ultimate but perhaps impossible aim of MT research. (*MT-ppf*, page 154)

**Fifth Generation** A term used to denote intelligent computers, i.e. that can learn, and make inferences. In October 1981 the Japanese Ministry of International Trade and Industry launched its Fifth Generation Computer Systems project to stimulate research and development. (*MT-ppf*, page 329)

**FRAP** A French to Russian MT system being developed at the All-Union Centre for Translation of Scientific and Technical Literature and Documentation, Moscow. (*MT-ppf*, pages 311-312, *ND-MT*, page 77)

**Fujitsu** Japanese company which has developed the ATLAS (q.v.) system for translation from Japanese to English.

## G

**Gachot** The French engineering firm Gachot acquired, in a series of moves in the late 1980s, rights in the private sector in most parts of the world to the Systran (q.v.) system. The firm now develops and markets MT from its headquarters near Paris, and from the World Translation Center at La Jolla, California, the latter being headed by Denis Gachot, son of the company's founder. One of the Gachot initiatives was to make Systran available over the French Minitel system, the first MT had been made available to a mass audience. The acquisitions by Gachot of different parts of the previously somewhat fragmented Systran rights, and close collaboration with the EC Commission, who own certain Systran rights in the public sector, have in recent years facilitated much more co-ordinated development of Systran. (*ND-MT*, pages 22-24. See also interview with Jean Gachot in *Language International* 1.6.)

**GAT** Acronym for Georgetown Automatic Translation (the same system had previously been termed General Analysis Technique). See

also *Georgetown University*. GAT Russian to English systems were installed at the Oak Ridge National Laboratory in the USA, for the US Atomic Energy Commission, and at the Euratom centre in Ispra, Italy, in 1963/64, and thereby became one of the first operational systems. (*MT-ppf*, pages 70-78)

**Georgetown University** This University in Washington, DC, played a major part in the early history of MT. An MT research team was established there under Leon Dostert in 1952, and a prototype Russian to English system, produced in collaboration with IBM, was demonstrated in January 1954, an event which created considerable public interest, and stimulated major US research and government funding in the following decade. At Georgetown itself the GAT (*q.v.*) system was developed, Peter Toma, who was later to develop his own SYSTRAN (*q.v.*) system, being one of those who worked on the development. The main research project ended in 1963. (*MT-ppf*, pages 70-78)

**GETA** Acronym for Groupe d'Etudes pour la Traduction Automatique, the principal French MT activity, a research team based at the University of Grenoble, which succeeded CETA (*q.v.*) in 1971. The main system is based on Ariane (*q.v.*). (*MT-ppf*, pages 239-248, *ND-MT*, page 34)

**Gigatext** This company was formed in Saskatchewan, Canada, in 1987 with the support of the provincial government, with the initial task of translating all provincial laws from English into French, using a system created specially for it and never described in any published document, has been closed down this year. No translations were ever produced, and the episode is now considered one of MT's fiascos.

**Globalink** A low price MT system marketed by the company of that name, sometimes known as GTS (Globalink Translation System). Can work in either interactive (using a split screen) or batch mode. The system was designed by Bedrich Chaloupa, one of the members of the original Georgetown University team. Language pairs now available are French to English, Spanish to English, German to English, English to French, and English to Spanish, with several other pairs said to be under development. A review of the system by Claude Bédard appeared in the September 1989 issue of *Electric Word*.

**GPSG** Abbreviation for Generalised Phrase Structure Grammar.

**GRADE** A programming environment for grammar writing developed from research at Kyoto University, for translation between Japanese and English. (*ND-MT*, page 44)

**GTS** Abbreviation for Globalink (*q.v.*) Translation System.

## H

**HICATS** The acronym stands for Hitachi Computer Aided Translation System, and is marketed in Japan for the language pair Japanese to English by the Hitachi company. It was previously known as ATHENE/N. The language pair English to Japanese has been developed but is not yet on the market. (*ND-MT*, page 47)

**Hitachi** Large Japanese company marketing the HICATS (*q.v.*) system. They recently also announced the completion of a Japanese to Korean system developed in collaboration with the Korean Institute of Economics and Technology, and there are plans to set up a translation centre in Seoul and to develop a Korean to Japanese system.

## I

**IBM** IBM were very active at the start of MT research and development, and collaborated closely with Leon Dostert's 1952-54 project at Georgetown University, which resulted in the January 1954 public demonstration of a Russian to English translation system, an event which caused considerable stir. It seemed to show that MT was a feasible objective. IBM was never again to be at the forefront of MT development, though they have never totally forsaken the field (see *LMT*, below). There have been reports of IBM research centres in Spain and Israel looking at MT systems for English to Spanish and English to Hebrew translation, but there has been no recent confirmation. IBM Japan has a prototype English to Japanese system for translating IBM computer manuals. (*MT-ppf* page 37.)

**IFTT 89** The abbreviation stands for International Forum on Translation Technology, a major MT conference held in Tokyo in April 1989.

**Institut Textile de France** Developers and operators of the TITUS system (*q.v.*).

**International Association for Machine Translation** This association does not yet exist, but the suggestions for such an organisation were discussed at the conferences IFTT 89 and MT Summit II, also held in 1989, and such an association will almost certainly be constituted at MT Summit III, a conference which is provisionally scheduled for the United States in 1991. A prospectus for the proposed association, drawn up by Dr Muriel Vasconcellos, was published in the April 1990 issue of the *ATA Chronicle*.

**IONA** Japanese company, headed by Sadao Kawasaki, which owns the Systran Corporation of Japan and rights to the Systran Japanese programs.

**ISSCO** Acronym for (Dalle Molle) Institute for Semantic and Cognitive Studies, an organisation based in Geneva, and headed by Margaret King. It has carried out experimental MT work on its own account, and also was closely involved in the early stages of the Eurotra (*q.v.*) project, acting at one time as a sort of secretariat.

**ISTIC-1** A system being developed by the Institute of Scientific and Technical Information (ISTIC) of China for the translation from English to Chinese of titles in electronics and computer documents.

**Japanese Electronic Dictionary Project** This collaborative venture is funded by eight Japanese commercial companies, many of whom are active in machine translation development, and the Japanese government, and is intended to save the participating companies much individual dictionary-building effort.

**JETR** Abbreviation for Japanese English Translation, a small-scale knowledge-based experimental MT system at the University of California at Irvine, based on a corpus of Japanese cooking recipes and instructions for digital watches. *NDMT*, pages 30-31.

**JICST** Abbreviation for Japan Information Centre for Science and Technology, a Japanese-government supported machine translation research project (also known as Mu II) which succeeded the *Mu* project, (*q.v.*). Test operation of a system produced under this project is said to be imminent.

**JFY-IV** An experimental MT system which has been under development in China for the past 14 years. Most of the work has been done on translation from Chinese to English, but tests have also been carried out with translation from Chinese into French, German and Russian. An experimental model is also being established for translation from Esperanto to Chinese.

**K**

**KBMT** Abbreviation for Knowledge Based Machine Translation

**KIT** Acronym for Künstliche Intelligenz und Textverstehen, the name of a research programme being carried out at the Technical University of Berlin on AI approaches to MT transfer.

(*NDMT*, page 38.)

**KY-1** This is a post-edited batch transfer MT system under development in China, with facilities for the compilation of English to Chinese dictionaries and for statistical analyses of English texts.

(*NDMT*, page 50.)

**L**

**LAMB** This is an experimental system for Japanese to English translation being developed by Canon in Japan.

**Language Engineering Laboratory (LEL).** This unit, part of the China Software Technique Corporation (CSTC), is responsible for conducting and organising the MT projects approved by the Chinese government under its current five-year plan.

**Latsec Inc.** A company founded in 1968 by Dr Peter Toma to develop Systran (*q.v.*) in the United States; now part of the Gachot (*q.v.*) organisation.

**LIDIA** Acronym for Large Internationalisation of the Documents by Interacting with their Authors. This is a *GETA* (*q.v.*) project for producing a prototype for what the team call "personal MT". Work is going forward on translation from French into at least Russian and German (reversing previous systems), using a small corpus from the Ariane-G5 user interface (containing some on-line documentation) in HyperCard form. A paper describing the project was given by Christian Boitet at the COLING c# in Helsinki in August 1990.

**Linguistic Products** A company established in Houston, Texas, by Ralph Dessau and George Mallard to market low-price computer-assisted translation packages, such as *PC-Translator* (*q.v.*).

**LINTRAN** A generating system created by the VCP organisation in the Soviet Union for the automatic construction of MT systems to meet the specific information requirements of end-users.

**LFG** Abbreviation for Lexical Functional Grammar

**LMT** Abbreviation for Logic-programming-based Machine Translation, an IBM research system. Prototype versions have been developed for translation from English to German, English to Danish, Danish to English, English to French, German to English and English to Spanish.

**Logos** The Logos Development Corporation was founded in 1969 by Bernard E. Scott in the United States, to work on an English to Vietnamese automatic translation system for military aircraft maintenance manuals, for the US Air Force (it was the time of the Vietnam war). Although there was testing and evaluation of a system (LOGOS I), the end of the war brought with it the end of the demand for translation into Vietnamese. The company turned to work on systems for other languages, but the next major break-through came in 1982 with the launch of a commercial product, the "Logos Intelligent Translation System", for German to English translation. English to German was added in 1984, and other language pairs now offered include German to French, English to French, English to German and English to Spanish. Major users include Nixdorf and IBM in Germany, and Lexi-Tech and the Canadian Government in Canada.

(*MT-ppf*, pages 218-219, 255-257; *NDMT*, pages 25-26.)

**LUTE** Acronym for Language Understander, Translator & Editor, an experimental system for Japanese to English and English to Japanese translation being developed by the Nippon Telegraph and Telephone Corporation in Japan. The approach is a sophisticated one, with researchers aiming more at providing a research environment for MT rather than producing an early prototype.

(*MT-ppf*, pages 316-317; *NDMT*, page 49.)

**M**

**Machine Translation Summit** The first in this conference series was held in Tokyo in September 1987, the second in Munich in August 1989. The third in the series will take place in Washington in July 1991.

**MacroCAT** The mainframe version of the Weidner (*q.v.*) MT system.

**MARIS** Acronym for Multilinguale Anwendung von Referenz-Informationssystemen. This is a project developed at the University of Saarbrücken to create a multilingual information retrieval system to meet in particular the needs of German-speaking users of documentation in English. For this purpose the team are developing a computer-assisted translation system, STS (*q.v.*) based on other Saarbrücken research (see *SUSY*).

**MELFRAN** This is a transfer system, for use in interactive mode, for translation from Japanese to English, being developed by Mitsubishi.

**METAL** Acronym for Mechanical Translation and Analysis of Languages. This system, now being marketed by Siemens GmbH in the Federal Republic of Germany, was originally developed by the Linguistics Research Center of the University of Texas. Siemens became involved in providing funds for the system in 1979, and became sole sponsors from 1980. Research and development is continuing at Siemens Munich, at the University of Texas, at the University of Louvain in Belgium (French to Dutch and vice versa), in Barcelona (German to Spanish), and at Kolding in Denmark (German to Danish). The most mature language pair available is German to English, and in August 1989 there were said to be more than 12 installations.

(*MT-ppf* 248-254, *NDMT* 26.)

**METEO** METEO, also known as TAUM-METEO, emerged from the TAUM (*q.v.*) group at the University of Montreal. In May 1975 the group was contracted to develop a system for translating public weather forecasts from English into French. A working system was brought into operation in 1977, and has continued to provide the near-automatic translation of Canadian government weather forecasts ever since. Improvements led to METEO2 being introduced in 1983. With a relatively small dictionary of some 2,000 words and standard expressions, the system translates hourly weather forecasts for Environment Canada's meteorological stations in various centres across Canada. A French to English system was introduced in 1989 for Montreal. METEO handles an average of 30,000 words a day.

(*MT-ppf*, pages 224-225, 228-231.)

**MicroCAT** The PC version of the Weidner (*q.v.*) MT system).

**MT-VAN** Abbreviation for Machine Translation — Value Added Network, a term used by Fujitsu et al. to denote links between MT and human translation and revision. A paper describing the concept was given at IFTT '89 by Hozumi Tanaka, of the Tokyo Institute of Technology

**Mu** The national machine translation project of the Japanese government, overseen by Professor Makoto Nagao, of Kyoto University, one of the world's leading figures in MT, ran from 1982 to 1986. It was then succeeded by the *JICST* project (*q.v.*).