

The Language Professional's Guide to...

The ABC of Machine Translation I

compiled by W. John Hutchins and Geoffrey Kingscott_____

One of the most widely praised features of *Language International* has been our "Guides to...". For some time now readers have been asking for a guide to machine translation.

This is now a very large subject indeed, and one that cannot be done in a simple summary. We have therefore decided to carry the guide as a feature running through several issues of *Language International*, in a way that can be detached and filed for reference.

The way we have tackled this problem is to take as our basis the nearest thing to a definitive survey of machine translation to date, John Hutchins's 1986 book, *Machine translation: past, present, future*, (referred to in the text as *MT-ppf*, published by Ellis Horwood) together with his 1988 updating paper, "Recent Developments in Machine Translation", published as part of *New Directions in Machine Translation* (Foris, Dordrecht), referred to in the text as *ND-MT*. A further valuable collection with articles describing the major current systems and a good bibliography was edited by Jonathan Slocum:

Machine Translation Systems (Cambridge University Press, 1988).

The directory contains entries for systems and projects (usually under their acronyms), and for organisations and institutions involved in MT development. Only particularly important older systems or organisations are included. It includes also a few common terms encountered in descriptions of MT systems, but not the terminology of linguistic theory or computational linguistics, such as case grammar, dependency grammar, etc. Definitions for these terms may be found in dictionaries of linguistics.

Language International editor Geoffrey Kingscott, who in 1989 wrote a report, *Applications of Machine Translation*, for the Commission of the European Communities, and John Hutchins are collaborating in the compilation of this new Directory.

It is emphasised that this "ABC of Machine Translation" is an indicative guide, for initial information only, not a detailed study of each system mentioned.

"The Language Professional Guide to..." is a supplement to
Language International
The magazine for the language professions
© John Benjamins Publishing Co. 1990
P.O. Box 52519
1007 HA Amsterdam
Netherlands

The Language Professional's Guide to...

The ABC of Machine Translation — A-D

A

AI (Artificial Intelligence) A general term covering the use of computers to undertake tasks which are normally assumed to require human intelligence. In the field of machine translation it is normally applied to systems which make use of extra-linguistic information (i.e. information which is not accessible directly from the texts being handled or from general knowledge of the grammar and vocabulary of the language), in order to make "intelligent" inferences about meanings and particularly to resolve ambiguities. See also KBMT (Knowledge Based Machine Translation).

AIDTRANS Name of a research project at Sheffield University on a system for Japanese-to-English translation derived from a method of teaching people with little linguistic talent how to 'decode' Japanese texts using a grammar-dictionary in an 'automaton-like' fashion. The system is based on the direct translation model of linear predictive analysis. The intention of the project is to computerise the almost 'mechanical' procedures of the human learners ('decoders') and to produce a type of machine-aided system of translation. (ND-MT, page 41)

ALPAC (Automatic Language Processing Advisory Committee) A committee set up in April 1964 by the USA National Academy of Sciences to study the demand, supply and cost of translation; the demand for, and availability of translators; and the role of machine translation.

Reporting in 1966, and looking almost exclusively at Russian to English translation, it found that the supply of translators greatly exceeded the demand, that there were no significant amounts of Soviet technical literature to be translated that was not being translated, that while more work needed to be done on aids for translators, particularly termbanks, machine translation was not itself a solution to the translation problem, since progress had been "uniformly discouraging", and that little had been achieved despite the expenditure of a great deal of, mainly public, money.

As a consequence of the report, the funding of MT research almost ceased in the United States and diminished substantially elsewhere. Its effects were to last some ten years. (MT-ppf, pages 164-167)

ALPS (Automated Language Processing Systems Inc.) A company formed in 1980 to develop research begun at Brigham Young University (q.v.). The ALPS system can function on three different levels, the higher levels encompassing all the features of the lower levels. The highest level (Transactive), the machine translation level, operates in interactive mode, offering a proposed translation to the translator sentence by sentence, with questions to elucidate ambiguities. (MT-ppf, pages 302-303; ND-MT, page 27)

AMPAR (Avtomatizirovannyi Mashinnyi Pervod s Angliiskogo yazyka na Russkii) The "direct translation" system for English-to-Russian translation developed at the All-Union Centre for Translation (Moscow) and in operational use since 1979. The English to Russian context dictionary was compiled on the basis of an analysis of English newspaper texts on political, economic and scientific topics. Later integrated with NEPRA (q.v.) as the AMPAR system (q.v.) and now part of the SPRINT complex of systems (q.v.). (MT-ppf, pages 308-310; ND-MT, pages 43, 76)

ANRAP (Angliisko- i Nemetsko-Russkii Avtomaticheskii Pervod) A modular system developed at the All-Union Centre for Translation for translation from English and German into Russian, based on the earlier AMPAR and NEPRA systems (q.v.). Now being integrated into the SPRINT "family" of systems (q.v.).

APAC (Automaticky Preklad z Anglictiny do Cestiny) A system for English-to-Czech translation based on dependency grammar and implemented in Q-systems (developed originally by TAUM (q.v.)). Initial experiments began in the late 1970s, the latest version APAC32 was tested from 1981 to 1986 as a production-oriented project. Research continues at Charles University's Faculty of Mathematics and Physics. (See Z. Kirschner and A. Rosen "APAC - an experiment in machine translation" *Machine Translation* 4(3), 1989, 177-193.)

ArchTran A system for English-to-Chinese translation under the joint development of the National Tsing Hua University and the Behaviour Tech Computer Corp (BTC) since May 1985. Research is based at the Science-based Industrial Park at Hsinchu (Taiwan). Archtran is a transfer system intended initially for the domain of computer manuals.

ARAP An English-to-Russian machine translation system being developed in the Soviet Union. (ND-MT, page 80)

Ariane The "engine" of a machine translation system developed by the GETA (q.v.) team in Grenoble, essentially a transfer system with morphological and syntactic analysis, transfer, and syntactic and morphological synthesis. Ariane-78 forms the basis of the French national computer-aided translation project Calliope (q.v.) and has been the basis for many projects outside France, notably joint projects in Malaysia and Thailand. The latest version is Ariane-85. See also *Calliope*. (MT-ppf, pages 239-248; ND-MT, pages 33-35)

ASCOF (Analyse und Synthese des Französischen mit Comskee) A research project at the University of the Saar in Saarbrücken for translation from French into German using the programming tool Comskee *Computing and String Keeping Language*. It is a transfer system incorporating AI-style "knowledge-based" semantic analysis, aiming for high-quality batch translation. The research emphasis has been on problems of French analysis. (ND-MT, pages 36-37. For a detailed description of ASCOF see also *Machine Translation Systems*, edited by Jonathan Slocum, Cambridge University Press, 1988.)

Aslib Aslib is an information organisation in the British Isles which launched in 1978, in collaboration with the then Translators' Guild of the Institute of Linguists, and later with the Institute of Translation and Interpreting, the international Translating and the Computer conference, held in London every November, and sometimes called the Aslib conference.

AS-TRANSAC (ASseries Translation Accelerator) An English to Japanese machine translation system, of the transfer type, produced by the Japanese company Toshiba, which has been available commercially since 1986 (a previous experimental stage was known as TAURUS). A Japanese to English system is still under development. (ND-MT, page 47)

ATAMIRI (Automata Traductor Algoritmico Multilingue Interactivo Recursivo Inteligente) A machine translation system developed in Bolivia by Ivan Guzman de Rojas, using the Amerindian language Aymara as an interlingua. An English-Spanish system is used by the Panama Canal Commission.
(*ND-MT*, pages 123-128)

ATEF (Analyse des Textes en Etats Finis) A finite state program module in the Ariane system (*q.v.*) used for morphological analysis.

ATHENE (Automatic Translation of Hitachi from English into Nihongo with Editing Support) Name of the research prototypes at Hitachi Corporation for Japanese-to-English and English-to-Japanese translation, subsequently marketed as HICATS (*q.v.*).
(*MT-ppf*, page 315; *ND-MT*, page 47)

ATLAS The name of two systems from Fujitsu. ATLAS/I is a "direct translation" system for English-to-Japanese, available commercially since 1984; ATLAS/II for Japanese-to-English is based on the transfer approach. ATLAS/II has been the basis for an experimental Japanese-to-German system (jointly with the University of Stuttgart (see also SEMSYN), for a Japanese-to-Korean system (jointly with the Korean Advanced Institute of Science and Technology), and for the Japan-Info project of the European Communities for translating Japanese abstracts.
(*MT-ppf*, page 316; *ND-MT*, page 46)

ATN (Augmented Transition Network) A parser developed by William Woods in the early 1970s, widely used in computation linguistics and artificial intelligence.

ATR (Advanced Telecommunications Research) A Japanese project involved in research on the automatic translation of telephone conversations, based on advanced AI techniques and methods developed in Japanese MT research.
(*ND-MT*, page 45)

ATTP (Automatic Translation Typing Phone) A research project by the Toshiba Corporation of Japan aiming to allow an English message to be typed into a work station, translated into Japanese, and transmitted via satellite, and vice versa.
(*ND-MT*, page 47)

Autoterm The medium level of the ALPS system (*q.v.*) allowing automatic consultation of dictionaries and terminology sources.

B

batch mode The implementation of a machine translation system in which whole texts are translated without the intervention of operators. For the improvement of quality, human assistance is provided by the editing of texts before input (pre-editing *q.v.*) or after output of the rough or "raw" translation (post-editing *q.v.*). Batch mode is contrasted with interactive mode (*q.v.*) where the operator may intervene during the translation process.

BKB (Bilingual Knowledge Bank) A component of the DLT system (*q.v.*) comprising texts in two languages, original and translated, as a source for contextually appropriate equivalents during the generation of translation.
(For a fuller treatment see the paper by Claude Bédard: *The BKB, a promising new approach to machine translation*, in proceedings of the American Translators Association conference 1989.)

Bravis Inc. This Japanese company (the spelling "Bravice" is also found) purchased a majority holding in the company holding the rights to the Weidner machine translation system in 1984, and was particularly successful in the mid-1980s with the sale of its MicroPack Japanese to English system.
(An interview with Takehiko Yamamoto, head of Bravis, was published in *Language International* 1/3, 1989.)

Brigham Young University (Provo, Utah) Research on MT began in the early 1970s under Eldon G. Lytle, funded by the Church of Jesus Christ of the Latter Day Saints (Mormon church). An interactive system for translating English texts into a number of languages was based on "junction grammar", a modified version of transformational grammar. The research provided the basis for the ALPS and Weidner systems (*q.v.*) and for the subsequent work at BYU by Alan Melby on translator workstations and terminology management. (See also *Termex* and *Linguatex*.)
(*MT-ppf*, pages 299-231)

British Telecom Undertaking research at Martlesham Heath, near Ipswich, on a speech translation system based on the recognition of a limited set of standard business phrases from a small number of "key" words. A prototype English-to-French system was demonstrated in 1987 and received much publicity.

BSO (Buro voor Systeemontwikkeling, Utrecht, Netherlands) Software company, developers of the DLT system (*q.v.*).

B'VITAL (Bernard Vauquois Informatisée et Traitement Automatique des Langues) A company set up in 1984 in Grenoble to market the products of GETA (*q.v.*). In January 1990 the company was purchased by the French documentation company SFTE.
(See report in *Language International* 2/2, 1990.)

C

CADA (Computer Assisted Dialect Adaptation) A method which has been developed at the Summer Institute of Linguistics in the United States for translation between closely related languages and dialects. It has been used with some success for South American Indian languages.
(*ND-MT*, page 29)

Calliope This was a project designed to make use of the *GETA Ariane* development in practical applications. There have been two applications, Calliope-Aero, a French to English system for aeronautics documentation, and Calliope-Info, an English to French system for data processing. Also being developed is a translator's work-station, known as Calliope-Révision.
(*ND-MT*, page 35. See also, Odile Vaissade, *Le projet national de traduction aidée par ordinateur et le système Calliope*, in the proceedings of the *Journées européennes de la traduction professionnelle*, *Encrages*, issue 17 (Université Paris VIII Vincennes à Saint-Denis).)

Canon The Canon company of Japan is developing a Japanese to English machine translation system, LAMB.

Cap Sogeti Innovation French company engaged in research on computer-based natural language products, including the development of a system for translating German military telex messages into French.
(See L. Cosserat et al. in Abbou, A. (ed.) *Traduction assistée par ordinateur*.)

Carnegie-Mellon University An important centre for MT research in the United States, using AI techniques and unification grammar formalism, for the development of a knowledge-based MT system (KBMT) initially from English into Japanese, later into other languages. (For details see *Machine Translation*, vol. 4, issues 1 and 2, 1989.) The Centre is also involved in research on interactive dialogue systems for MT and joint projects with a number of US and Japanese computer companies.
(*MT-ppf*, page 306; *ND-MT*, pages 29, 30)

CAT Acronym for Computer-aided Translation.

CAT2 An experimental system developed at Saarbrücken within the EUROTRA project (*q.v.*). The name refers to the three kinds of formal objects in the formalism; C(onstructors), A(toms), and T(ranslator) rules. The multilingual transfer design, implemented in versions of Prolog, is being developed for translations between English, German, Spanish and French.

CCL (Centre for Computational Linguistics), a department at the University of Manchester Institute of Science and Technology which studies problems of machine translation.

CETA (Centre d'Etudes de la Traduction Automatique) This research body was established in December 1959 at Paris and Grenoble, but from 1963 work was concentrated at Grenoble, under the leadership of one of the major pioneers of machine translation, the late Bernard Vauquois. CETA was renamed GETA (*Groupe d'Etudes pour la Traduction Automatique*) in 1971. See also Ariane, B'VITAL, and Calliope.

(*MT-ppf*, pages 128, 129 and 175, 176. The contribution to machine translation of Bernard Vauquois, who died in 1985, was commemorated in 1988 with the publication of *Analectes, Bernard Vauquois et la TAO*, a selection of his writings, edited by the present head of GETA, Professor Christian Boitet.)

Charles University, Prague. An important centre for research on computational linguistics and MT for many years (since 1957); see the APAC system (*q.v.*).
(*MT-ppf*, pages 142, 143 and 254, 255)

China Software Corporation Beijing company involved in the development of TRANSTAR system (*q.v.*) and collaborating in the multinational CICC project (*q.v.*).

CICC (Center of the International Cooperation for Computerization) Japanese organisation seeking to promote cooperation between Japan and other Asian countries, particularly in the field of machine translation and terminology. The main collaborative project seeks to develop an interlingua-based system for multilingual translation on behalf of the Japanese Overseas Development Agency (ODA), between Japanese, Thai, Malay, Indonesian and Chinese.
(*ND-MT*, page 45)

CLRU (Cambridge Language Research Unit) Body which played a major role in MT research in the 1950s and early 1960s. Its director was the late Margaret Masterman.
(*MT-ppf*, pages 155-122)

Colgate University Location for research during the 1980s on the AI-based TRANSLATOR project, subsequently continued at Carnegie-Mellon University (*q.v.*).
(*MT-ppf*, pages 282-283)

Commission of the European Communities The Commission has been closely involved with machine translation since the mid-1970s, when it began investigating the possible use of machine translation to meet the

likely huge increase in translation demand. At about this time it acquired *Systran*, which it has continued to develop at Luxembourg in collaboration with the various Systran companies (*q.v.*).
(*MT-ppf*, pages 258-271)

computational linguistics The application of the concepts and techniques of computer science to the analysis of natural language; hence the basis for much work in machine translation. Latest work in computational linguistics is usually described at the international COLING conferences, held every two years. The 1990 conference will be held in Helsinki in August (see *Calendar of Events* in this issue).

CONTRAST (Context Translation System) A small experimental Japanese to English machine translation system using contextual information, being carried out at the Electrotechnical Laboratory (*q.v.*) in connection with the CICC project (*q.v.*).
(*ND-MT*, pages 45)

CoRTeXT A recently announced microcomputer-based system from Universal MicroSystems (Vermont, USA) for English-to-Portuguese translation, with other languages under development. (See *Electric Word* 18, March/April 1990, p. 9).

CRITER A knowledge-based sublanguage system for translation of agricultural reports either from English to French or from French to English. Research is sponsored by the Centre Canadien de Recherche sur l'Information du Travail.
(Further information is given in a paper by P. Isabelle, M. Dymetman and E. Macklovitch in the proceedings of COLING-88.)

CSK Research Institute This organisation in Japan is reported to be using an in-house English-to-Japanese system for translating economics texts.
(*ND-MT*, page 49)

CULT (Chinese University Language Translator) A Chinese to English machine translation system developed at the Chinese University of Hong Kong. Research began in 1969, and from 1975 the system has been used regularly for the cover-to-cover translation of the Beijing mathematics journal *Acta Mathematica Sinica*. Experimental work has also been done on an English to Chinese version.
(*MT-ppf*, page 298)

D

direct translation systems Machine translation systems specifically designed to operate between a particular pair of languages. Many of the early machine translation systems were of this type, while later systems tend to be either of the *interlingua* or *transfer* type.
(*ND-MT*, page 9)

DLT (Distributed Language Translation) A system being developed by BSO in the Netherlands using Esperanto as an "interlingua". The aim is a multilingual system for monolingual users wishing to transmit texts in languages not their own, incorporating on-line interactive clarification of ambiguities in the user's own language. A prototype demonstrating English-to-French translation was demonstrated in December 1987. Later developments incorporate the BKB (*q.v.*) concept for semantic processing.
(*MT-ppf*, pages 288-290; *ND-MT*, pages 39-40)

DUET An English-to-Japanese system from Sharp Corporation (Nara, Japan), based on the transfer approach and case grammar rules. The commercial system, launched in 1988 includes facilities for essential pre-editing and post-editing.

To be continued in following issues of *Language International*.