

Current Machine Translation Systems in Japan

Summary

The machine translation systems listed in the following pages are those that are gathered from Machine Translation Summit participants.

The content states at May 1st 1988.

As for the terminologies in the lists, we have not changed original key words usage because of different meanings in the context.

The followings are summary of the main items in the systems.

[1] Current status of the systems

Seven English to Japanese systems and six Japanese to English systems have been already commercialized. Four Japanese-to-English systems and three English-to-Japanese systems are under the research and development, in which one will be brought to the market in 1988 and the another one will be serviced in 1990.

[2] Machine Usage and programming language of the systems

The development of the first stage machine translation system has been started on the mainframe computers. Now it is known that the machine translation systems can be implemented also in the minicomputer, work station for artificial intelligence, general purpose (UNIX) work station and personal computers. Those on the work station for artificial intelligence are limited to the research and development in this lists.

Concerning the programming language of the machine translation systems, there are used PL/I, C, LISP and Assembler in the main frame computers, Prolog and LISP in the work stations for artificial intelligence, and C in the general purpose work station and the personal computers. The commercialized systems are almost described by C language.

[3] Grammar for the systems

There exist three kinds of structure in the language analysis, that is, syntactic tree level, deep case or dependency tree level and interlingual level. In the analysis of Japanese-to-English systems, the structure level is almost deep case, although in the analysis of English-to-Japanese systems, there are two levels taking syntactic tree or deep case.

[4] Dictionaries

The dictionary components in the systems consists of basic term dictionary, technical term dictionary and user term dictionary. One system has only basic term dictionaries (Japanese dictionary and English dictionary) and another system creates the dictionaries from the neutral dictionaries. Word size in the norm of the dictionaries are followed:

basic term dictionary	50,000
technical term dictionary	250,000
technical term per fields	20,000 ~ 30,000
user term dictionary	30,000 ~ 50,000

1) Name of system	LAMB
2) Name of organization and contact address	CANON INC. Information Systems Research Center YOICHI KAWABATA 9-4, Shimomaruko 2-Chome, Ohta-ku, Tokyo 146, Japan Phone: Tokyo (03) 758-2101
3) Current status of the system	For research
4) Feature of the system	Knowledge-based machine translation
5) Translated language	Japanese to English
6) Strategies on translation <ul style="list-style-type: none"> Type of grammar for analysis and generation Translation processes 	Phrase structure grammar Tree-to-tree transformations
7) Dictionaries <ul style="list-style-type: none"> Structure Size 	Japanese analysis dictionary Transfer dictionary English generation dictionary Domain knowledge dictionary 2,000 words
8) Equipment <ul style="list-style-type: none"> Implementation language Operating system Type of CPU 	Symbolics Common Lisp Symbolics OS release 6.1 Symbolics 3620
9) Performance <ul style="list-style-type: none"> Translation speed 	1,000 words/hour
10) Facilities	Dictionary and grammar development environment Testing tool

1) Name of system	ATLAS-I (Automatic Translation System-I)	ATLAS II (Automatic Translation System II)
2) Name of organization and contact address	FUJITSU, LTD. 1-6-1, Marunouchi, Chiyoda-ku, Tokyo 100	
3) Current status of the system	Commercial (Program product)	
4) Feature of the system	Syntax direct approach using semantic features	Interlingua transfer approach using world model and language model
5) Translated language	English to Japanese	Japanese to English
6) Strategies on translation <ul style="list-style-type: none"> Type of grammar Analysis Generation grammar Translation process 	Phrase structure grammar Phrase structure grammar (1) Morphological analysis (2) Syntax analysis and generation (3) Morphological synthesis	Dependency grammar + phrase structure grammar Phrase structure grammar (1) Morphological analysis (2) Syntax and semantic analysis (3) Conceptual structure transfer (4) Syntax and morphological synthesis

7) Dictionaries <ul style="list-style-type: none"> • Structure • Size 	English-Japanese dictionary Basic words -----53,000 words Technical term -----250,000 words	Japanese dictionary Basic words -----50,000 words Technical term dictionary -----250,000 words English dictionary Basic words -----50,000 words Technical term dictionary -----250,000 words
8) Equipment <ul style="list-style-type: none"> • Implementation language • Operating system • Type of CPU 	Assembler OSIV/F4/MSP, OVIS/S, OSIV/F4/FSP, FACOM M series computer FACOM S-3000 series computer	C OSIV/F4/MSP FACOM M series computer
9) Performance <ul style="list-style-type: none"> • Translation speed 	60,000 words/hour (CPU: FACOM M380)	60,000 words/hour (CPU: FACOM M380)
10) Facilities	<ul style="list-style-type: none"> • Bilingual editor • Dictionary editor 	<ul style="list-style-type: none"> • Bilingual editor • Dictionary editor

1) Name of system	HICATS/JE (Hitachi Computer Aided Translation System/Japanese to English)	HICATS/EJ (Hitachi Computer Aided Translation System/English to Japanese)
2) Name of organization and contact address	Hitachi, Ltd. Computer Group, Hitachi, Ltd. Hitachi Omori 2nd Bldg., 6-27-18, Minami-Oi, Shinagawa-ku, Tokyo 140, Japan	
3) Current status of the system	Commercial (Program product)	
4) Feature of the system	Semantic transfer based on Conceptual Dependency Diagram	Syntactic transfer using co- occurrence relations
5) Translated language	Japanese to English	English to Japanese
6) Strategies on translation <ul style="list-style-type: none"> • Analysis grammar • Generation grammar • Translation processes 	Dependency grammar Phrase structure grammar (1) Morphological analysis (2) Syntactic/semantic analysis (3) Transformation of Conceptual Dependency Diagram (4) Syntactic generation (5) Morphological synthesis	Phrase structure grammar Case phrase structure grammar (1) Morphological analysis (2) Syntactic analysis (3) Co-occurrence relation checking (4) Syntactic transfer (5) Case phrase structure generation (6) Morphological synthesis
• Number of rules	approx. 6,000	approx. 2,500
7) Dictionaries	<ul style="list-style-type: none"> • Basic dictionary --- 50,000 words • Technical term dictionary (option) --- 250,000 words • User dictionary 	

8) Equipment	GDL (Grammar Description Language) and PL/I	
<ul style="list-style-type: none"> • Implementation language • Operating system • Type of CPU 	VOS3 (Virtual-storage Operating System 3) HITAC M series computers	
9) Performance	40,000-80,000 words/hour (CPU: HITAC M-680) depends on operational environments	30,000-60,000 words/hour (CPU: HITAC M-680) depends on operational environments
<ul style="list-style-type: none"> • Translation speed • Translation cost 		
10) Facilities	<ul style="list-style-type: none"> • Dependency specification • Selection of alternative sentence structure • Alternative word selection • Conversational editor • Batch maintenance program 	<ul style="list-style-type: none"> • Conversational editors (for expert and nonexpert) • Batch maintenance program
<ul style="list-style-type: none"> • Pre - editing • Post - editing • Dictionary development 		

1) Name of system	PAROLE
2) Name of organization and contact address	Wireless Research Laboratory Matsushita Electric Industrial Co., Ltd. 1006, Kadoma, Kadoma-shi, Osaka 571 Japan Phone (06) 908-1291
3) Current status of the system	Has been conducted as a research project
4) Feature of the system	Use of semantic transfer method System expandability (Grammar Rules, Dictionaries, Facilities, etc.)
5) Translated language	Japanese to English
6) Strategies on translation	<ul style="list-style-type: none"> • Type of grammar for analysis and generation • Translation processes • Rule size for analysis, transfer and generation (Aug. 1987)
<ul style="list-style-type: none"> • Case Grammar and Tree Transducer for analysis • Phrase Structure Grammar and Tree Transducer for generation • Japanese analysis using Case Frame • Japanese to English transfer using Dependency Structure • English generation using Phrase Structure • Analysis 200 rules (approx.) • Transfer 100 rules (approx.) • Generation 200 rules (approx.) 	
7) Dictionaries	<ul style="list-style-type: none"> • Structure • Size
<ul style="list-style-type: none"> • Represented by Lisp S-expression • Semantically classified • Including morphological, syntactic and semantic information • Basic 5,000 words (Aug. 1987) 	
8) Equipment	Prolog and Lisp
<ul style="list-style-type: none"> • Implementation language • Operating system • Type of CPU 	Genera Symbolics
9) Performance	Not measured

10) Facilities	Grammar development environment Language for describing tree-to-tree-transducing rules Debugging utilities for rule developments
1) Name of system	Mu Project
2) Name of organization and contact address	Nagao Laboratory, Dept. of Electrical Eng., Kyoto Univ. Address: Dept. of Electrical Eng., Kyoto Univ., Kyoto, Japan Electrotechnical Laboratory, MITI Japan Information Center for Science and Technology
3) Current status of system	The Mu project started at April, 1982 and completed at March, 1986. 10,000 Japanese sentences and 3,000 English sentences were translated and the results were evaluated by professional translators. A new project was started at April, 1986, which follows the basic principles of the Mu systems.
4) Feature of the system	The systems do not expect any pre-and post-editings. The basic approach is the transfer approach. The systems are characterized by Lexicon Driven Processing, Neutral Dictionaries and Heuristically Guided Processing.
5) Translated Language	English to Japanese, Japanese to English
6) Strategies on translation	A software called GRADE provides flexible pattern matching facilities to treat complicated linguistic phenomena. Rules can also be defined in the lexicon to treat word specific linguistic phenomena. Analysis grammar produces deep case interpretations of input sentences, which are annotated by various levels of information, such as typo-graphical, morphological, syntactic information, semantic marker, etc. The transfer grammar is divided into three sub-grammars, Pre-Transfer-Loop, Main Transfer, and Post-Transfer-Loop.
7) Dictionaries • Structure • Size	Japanese and English dictionaries are developed as neutral dictionaries which are independent of processing modes. Dictionaries for a translation system are created from the neutral dictionaries. About 80,000 items. The dictionaries for the new project will be extended to cover 300,000 items.
8) Equipment	• Uti-LISP on FACOM M382, M780 (FACOM OS/IV F4 ESP) • Zeta-LISP on Symbolics LISP Machines
9) Performance	Translation speed is about 4,000 words/hour (CPU: M780).
10) Facilities	An integrated translation environment was developed, which includes editors for texts, grammars and dictionaries. The new project will develop a new environment appropriate for translation services at JICST.
1) Name of system	MELTRAN-J/E (Melcom TRANslation system-Japanese/English)
2) Name of organization and contact address	T:DASAI, Information Systems & Electronics Development Lab., Mitsubishi Electric Corp., 5-1-1, Ofuna, Kamakura-shi, Kanagawa 247, Japan
3) Current status of the system	Under commercial development
4) Feature of the system	Transfer system based on logic programming
5) Translated language	Japanese to English

6) Strategies on translation <ul style="list-style-type: none"> Type of grammar for analysis and generation Translation processes Rule size for analysis, transfer and generation 	Phrase structure grammar Tree transducer about 1,000
7) Dictionaries <ul style="list-style-type: none"> Structure Size 	Basic lexicon/Technical Terminology/User dictionary Basic-50,000/Technical-30,000 for information processing
8) Equipment <ul style="list-style-type: none"> Implementation language Operating system Type of CPU 	ESP (Extended Self-Contained Prolog) SIMPOS MELCOM PSI
9) Performance <ul style="list-style-type: none"> Translation speed Translation cost 	5,000 words/hour (CPU) (evaluating)
10) Facilities	Pre, Post-editing, interactive/batch translation, and grammar and dictionary development environment

1) Name of system	PIVOT
2) Name of organization and contact address	NEC Corporation 33-1, Shiba 5-chome, Minato-ku, Tokyo 108 TEL.(03)454-1111
3) Current status of the system	Released for closed users
4) Feature of the system	Interlingual Approach
5) Translated language	English-to-Japanese, Japanese-to-English
6) Strategies on translation	1) dependency grammar, augmented shift-reduced parser tree to tree transducer 2) a1: Morphological Analysis a2: Grammatical and Semantic Analysis a3: Semantic Extraction g1: Conceptual Wordng g2: Grammatical Generation g3: Morphological Generation 3) analysis more than 3,000 generation about 2,500
7) Dictionaries <ul style="list-style-type: none"> System Core Dictionary Term Dictionary 	Japanese 40,000 entries English 53,000 entries covers more than 20 domain each of which includes more than 20,000 entries, maximum of 100,000 entries
8) Equipment	1) C-language 2) Acos-4, 6, EWS-UNIX 3) Acos Main-Framc, EWS-4800
9) Performance	1) 60,000 2) 1,500 yen/A4 (double space) (tentative)

10) Facilities	Batch and Interactive operation Modes Batch Tools: Text processing and Management Dictionary updation and management Text transfer between terminal and host machine Unknown word detection Interactive: Bi-lingual and mono-lingual text processing and management Functions: Unknown word detection Dictionary updation etc.
----------------	--

1) Name of system	Machine Translation System at NIPPON · DATA GENERAL CORPORATION
2) Name of organization and contact address	1st R&D Group NIPPON · DATA GENERAL CORPORATION 2165, Mochida, Gyoda-shi, Saitama, 361 Japan
3) Current status of the system	Prototype system has been implemented. And we are upgrading it for our own company's use.
4) Feature of the system	Component of totally comprehensive electric office system Easy knowledge acquisition capability
5) Translated language	English → Japanese
6) Strategies on translation	Phrase structure and deep case analysis using tree-transducer Transfer approach Morphological analysis ↓ Syntactic & Semantic analysis (about 1,000 rules) ↓ Transfer (600 rules) ↓ Synthesis (400 rules) ↓ Morphological synthesis Basic softwares are written by C. Main computer is ECLIPSE MV series.
7) Dictionaries	English-Japanese and Japanese dictionaries are on computer.

1) Name of system	PENSÉE
2) Name of organization and contact address	OKI Electric Industry Co., Ltd. Osaka Gas Information System Research Institute
3) Current status of the system	Released in Autumn 1986 (Japanese → English) Announced in May 1988(English → Japanese)
4) Feature of the system	1) High-quality translation on a small-size UNIX-based super personal computer 2) Simultaneous semantic processing and analysis of sentence structure 3) Integration of Japanese-English and English-Japanese translation
5) Translated language	Translation of Japanese into English Translation of English into Japanese

6) Strategies on translation <ul style="list-style-type: none"> Type of grammar for analysis and generation Translation processes 	Case grammar with deep structure The translation system is made up of Japanese morphological analysis, interactive translation with syntactic and semantic analysis, and English morphological generation.									
7) Dictionaries <ul style="list-style-type: none"> Structure Size 	<table border="1" data-bbox="813 459 1157 728"> <thead> <tr> <th></th> <th>Japanese-English dictionary</th> <th>English-Japanese dictionary</th> </tr> </thead> <tbody> <tr> <td>System dictionary</td> <td>90,000 words</td> <td>60,000 words</td> </tr> <tr> <td>User's dictionary</td> <td>60,000 words-capacity</td> <td>40,000 words-capacity</td> </tr> </tbody> </table>		Japanese-English dictionary	English-Japanese dictionary	System dictionary	90,000 words	60,000 words	User's dictionary	60,000 words-capacity	40,000 words-capacity
	Japanese-English dictionary	English-Japanese dictionary								
System dictionary	90,000 words	60,000 words								
User's dictionary	60,000 words-capacity	40,000 words-capacity								
8) Equipment <ul style="list-style-type: none"> Implementation language Operating system Type of CPU 	C language UNIPLUS+ (UNIX System V base) MC68020									
9) Performance <ul style="list-style-type: none"> Translation speed 	6,000 words/CPU-hour									
10) Facilities	<ol style="list-style-type: none"> A bilingual editing system for Japanese and English A dictionary editing system which allows the registration of words into the user's dictionary A preprocessing system which shows the result of Japanese morphological analysis 									

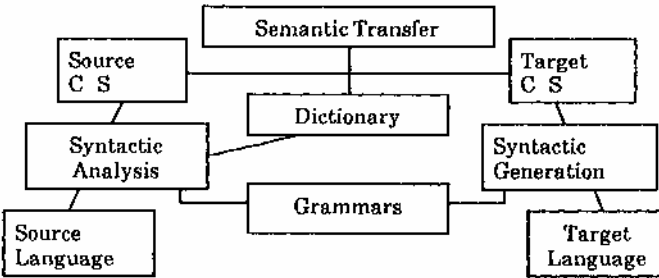
1) Name of system	Ricoh English-Japanese Machine Translation System (RMT)
2) Name of organization and contact address	Ricoh Company Limited Research and Development Center 4686, Nippa-cho, Kouhoku-ku, Yokohama-shi, Kanagawa 223 Japan Phone: 045-593-3411
3) Current status of the system	Now, under development. (Will be brought to market in 1988)
4) Feature of the system	Structure transfer referring to semantic features of words RMT offers one or more possible solutions to a source sentence. You choose your most favorite one.
5) Translated language	English to Japanese
6) Strategies on translation <ul style="list-style-type: none"> Morphological Analyses Syntactic Analyses Plausibility Evaluation Transfer 	Use some features of sentences. (200 rules) Augmented context free grammar. (2,200 rules) Makes dependency tree using semantic features. (60 features) Makes Japanese tree with extended cases. (300 rules)

7) Dictionaries	Main Dictionary (30,000 words) User Dictionary (30,000 words at most) Industry- or company-specific dictionaries (50,000 words)
8) Equipment <ul style="list-style-type: none"> • Implementation language • Operating system • Type of CPU 	C UNIX system V 3B2 (AT&T)
9) Performance <ul style="list-style-type: none"> • Translation speed 	4,500 words/h
10) Facilities	Editing Software <ul style="list-style-type: none"> * Split-screen displays source and target text simultaneously. * Misspelled or unknown words are checked. * Editing tasks such as word swap can be performed easily. Entering Source Text <ul style="list-style-type: none"> * Source text may also be entered through OCR. Dictionary development utilities <ul style="list-style-type: none"> * You can enter new terms in User Dictionary by a menu provided.

1) Name of system	Translation Word Processor SWP-7800
2) Name of organization and contact address	Sanyo Electric Co., Ltd. Dept. of Information Systems 2-7-25, Edobori, Nisi-ku, Osaka 550 (06) 443-5144
3) Current status of the system	SWP-7800 has been on the market since April, 1987.
4) Feature of the system	The system uses the Transfer method of translation. By placing a translation system in a word processor, the efficiency of the translation work is improved.
5) Translated language	Japanese to English translation only
6) Strategies on translation	1) Augmented CFG, in conjunction with case grammar, is used to analyze the Japanese text and also to determine the dependency structure. 2) Translation processes are, in order, Japanese morphological analysis, Japanese syntactic analysis, Japanese to English transfer, and English generation.
7) Dictionaries	The basic dictionary consists of four specific purpose dictionary files. They are 1) Japanese morphological dictionary, 2) Japanese analysis dictionary (syntactic and semantic), 3) Transfer dictionary, and 4) English generation dictionary (syntactic and morphological). In addition to the basic dictionary consisting of 55,000 words, memory area for another 55,000 words has been allocated for a user-oriented dictionary.
8) Equipment	The translation system is written in C programming language, and it runs on iRMX operating system. Type of CPU used is 80186.
9) Performance	The speed of translation is 3,500 words/hour.
10) Facilities	All the editing functions of our top class word processor are available for pre- and post-editing of the text, and a dictionary editor assists the editing of the user dictionary.

1) Name of system	Sharp English-Japanese Machine Translation System
2) Name of organization and contact address	Sharp Corp., Information Systems Group, Information Systems Laboratories Address: 492, Minosho-cho, Yamatokoriyama-shi, Nara, 639-11, JAPAN Phone: 07435-3-5521 Facsimile: 07435-3-0792
3) Current status of the system	Commercial base
4) Feature of the system	High speed translation on desktop computer
5) Translated language	English to Japanese
6) Strategies on translation <ul style="list-style-type: none"> Type of grammar for analysis and generation Translation processes Rule size for analysis, transfer and generation 	Augmented Context-Free Grammar (for syntactic analysis) Case Grammar (for semantic analysis) Generation process is built by the combination of rules and procedures. Employs advanced natural language processing by means of semantic analysis Around 1,000 rules (for syntactic analysis) Transfer and generation processes are built by the combination of rules and procedures.
7) Dictionaries <ul style="list-style-type: none"> Structure and size 	- Basic dictionary (around 60,000 words) - Technical term dictionary - up to 40,000 words/field - covering 4 fields such as Economics, Information Processing, Electronics, and Mechanical Engineering - User dictionary (up to 40,000 words per dictionary)
8) Equipment <ul style="list-style-type: none"> Implementation language Operating system Type of CPU Hardware 	Language C UNIX System V MC68010, 68020 OA-110WS, OA-210, OA-310, IX-7
9) Performance <ul style="list-style-type: none"> Translation speed 	5,000 words/hour (MC68010)
10) Facilities <ul style="list-style-type: none"> Pre-editing Post-editing 	Users can specify (a) the correct part-of-speech of syntactically ambiguous words; and (b) the phrase boundaries. Interactive Syntactical and Lexical Disambiguation, and Learning.

1) Name of system	TAURAS (for a research model only)
2) Name of organization and contact address	Information Systems Lab., Toshiba R&D Center 1, Komukai Toshiba-Cho, Saiwai-ku, Kawasaki 210, Japan
3) Current status of the system	Commercial
4) Feature of the system	Implemented on Engineering Workstation Transfer approach
5) Translated language	English-Japanese

<p>6) Strategies on translation</p> <ul style="list-style-type: none"> • Type of grammar for analysis and generation • Translation Process (figure) 	<p>ATN + Lexical Grammars</p> <p style="text-align: center;"><TRANSLATION PROCESS></p>  <pre> graph TD SL[Source Language] --> SA[Syntactic Analysis] SA --> ST[Semantic Transfer] SA --> G[Grammars] ST --> DS[Dictionary] DS --> SG[Syntactic Generation] SG --> G G --> TL[Target Language] </pre>
<p>7) Dictionaries</p> <ul style="list-style-type: none"> • Structure • Size 	<p>for each entry: syntactic category; syntactic properties; semantic properties; lexical rules; Max 130,000 words (general 50,000; technical 50,000; user-defined up to 30,000)</p>
<p>8) Equipment</p> <ul style="list-style-type: none"> • Implementation language • Operating system • Type of CPU 	<p>C UNIX MC68020</p>
<p>9) Performance</p> <ul style="list-style-type: none"> • Translation speed 	<p>7,000 words/hour</p>
<p>10) Facilities</p>	<p>Pre- and Post-editor, Bi-lingual editor, Dictionary maintenance system, Word processors</p>
<p>1) Name of system</p>	<p>SYSTRAN</p>
<p>2) Name of organization and contact address</p>	<p>SYSTRAN Corporation Nittochi Bldg, 1-4-1, Kasumigaseki, Chiyoda-ku, Tokyo 100, Japan (03) 504-1003</p>
<p>3) Current status of the system</p>	<p>Commercial</p>
<p>4) Features of the system</p>	<p>Multi-lingual machine translation system which combines qualities of speed, reasonable cost, accuracy, accessibility, and flexibility</p>
<p>5) Translated language</p>	<p>E/J, J/E, E/F, F/E, E/G, G/E*, E/R, R/E*, E/S, S/E, E/D, F/D, E/A*, E/P, F/G, E/I E (English), J (Japanese), F (French), G (German), R (Russian) S (Spanish), D (Dutch), A (Arabic), P (Portuguese), I (Italian)</p>
<p>6) Strategies on translation</p> <ul style="list-style-type: none"> • Type of grammar for analysis and generation • Translation process 	<p>Systran-specific grammar which incorporates the best/most useful concepts from current theoretical linguistics Source Language Morphological Analysis Semosyntactic Analysis (These source language analysis processes are totally independent from any given target languages.) Target Language Synthesis</p>

7) Dictionaries	Basic dictionary (80,000 for E/J) Scientific and technical terminology dictionary (250,000 for E/J) User specific dictionaries
8) Equipment <ul style="list-style-type: none"> • Implementation language • Operating system • Type of CPU 	Assembly and Systran Macro IBM MVS, MVS/XA and their compatibles IBM compatibles
9) Performance <ul style="list-style-type: none"> • Translation speed 	2,000,000 words per CPU Hour (E/J on Facom M 780)
10) Facilities	Bilingual editors for texts, rules and dictionaries Linguistic development tools (logic and dictionary) Batch/on-line Translation Not found word scan

*Certain restrictions may apply.