Current machine translation systems in Japan

1. Name of project
   LAMB

2. Name of organization and contact address
   1) Name of organization
      CANON INC. Information Systems Research Center
   2) Contact address
      YOICHI KAWABATA
      9-1, 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. Strategy on translation
   1) Type of grammar for analysis and generation
      Phrase structure grammar
   2) Translation processes
      Tree-to-tree transformations

7. Dictionaries
   1) Structure
      - Japanese analysis dictionary
      - Transfer dictionary
      - English generation dictionary
      - Domain knowledge dictionary
   2) Size
      2000 words

8. Equipment
   1) Implementation language : Symbolics Common Lisp
   2) Operating system : Symbolics OS release 6.1
   3) Type of CPU : Symbolics 3620

9. Performance
   Translation speed : 1000 words/hour

10. Facilities
    1) Dictionary and grammar development environment
    2) Testing tool
<table>
<thead>
<tr>
<th></th>
<th>ATLAS-I</th>
<th>ATLAS-II</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1) Name of system</strong></td>
<td>(Automatic Translation System-I)</td>
<td>(Automatic Translation System-II)</td>
</tr>
<tr>
<td><strong>2) Name of Organization</strong></td>
<td>FUJITSU, LTD.</td>
<td></td>
</tr>
<tr>
<td><strong>(Contact address)</strong></td>
<td>1-6-1, Marunouchi, Chiyoda-ku, Tokyo 100</td>
<td></td>
</tr>
<tr>
<td><strong>3) Current status</strong></td>
<td>Commercial (Program product)</td>
<td></td>
</tr>
<tr>
<td><strong>4) Features of the system</strong></td>
<td>Syntax direct approach using semantic features</td>
<td>Semantic transfer approach using world model and language model</td>
</tr>
<tr>
<td><strong>5) Language translated</strong></td>
<td>English to Japanese</td>
<td>Japanese to English</td>
</tr>
<tr>
<td><strong>6) Translation strategies</strong></td>
<td>Phase structure grammar</td>
<td>Dependency grammar</td>
</tr>
<tr>
<td><strong>· Type of grammar for analysis and generation</strong></td>
<td>(1) Morphological analysis</td>
<td>(1) Morphological analysis</td>
</tr>
<tr>
<td><strong>· Translation process</strong></td>
<td>(2) Applied grammar</td>
<td>(2) Syntax and semantic analysis</td>
</tr>
<tr>
<td></td>
<td>(3) Morphological generation</td>
<td>(3) Conceptual transfer</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(4) Syntax and morphological generation</td>
</tr>
<tr>
<td><strong>7) Dictionaries</strong></td>
<td>English-Japanese dictionary</td>
<td>Japanese dictionary</td>
</tr>
<tr>
<td><strong>· Structure</strong></td>
<td>Basic dictionary</td>
<td>English dictionary</td>
</tr>
<tr>
<td><strong>· Size</strong></td>
<td>--- 53,000 words</td>
<td>--- 50,000 words</td>
</tr>
<tr>
<td></td>
<td>Technical term dictionary</td>
<td>Technical term dictionary</td>
</tr>
<tr>
<td></td>
<td>--- 250,000 words</td>
<td>--- 250,000 words</td>
</tr>
<tr>
<td><strong>8) Equipment</strong></td>
<td>Assembler</td>
<td>C</td>
</tr>
<tr>
<td><strong>· Implementation language</strong></td>
<td>OSIV/F4/MSP, OVIS/S,</td>
<td>OSIV/F4/MSP</td>
</tr>
<tr>
<td><strong>· Operating system</strong></td>
<td>OSIV/F4/FSP,</td>
<td></td>
</tr>
<tr>
<td><strong>· Type of CPU</strong></td>
<td>FACOM M series computer</td>
<td>FACOM M series computer</td>
</tr>
<tr>
<td></td>
<td>FACOM S-3000 series computer</td>
<td></td>
</tr>
<tr>
<td><strong>9) Performance</strong></td>
<td>60,000 words/hour</td>
<td>60,000 words/hour</td>
</tr>
<tr>
<td><strong>· Translation speed</strong></td>
<td>(CPU: FACOM M380)</td>
<td>(CPU: FACOM M380)</td>
</tr>
<tr>
<td><strong>10) Facilities</strong></td>
<td>Bilingual editor</td>
<td>Bilingual editor</td>
</tr>
<tr>
<td></td>
<td>Dictionary editor</td>
<td>Dictionary editor</td>
</tr>
<tr>
<td>1) Name of system</td>
<td>HICATS / JE</td>
<td>HICATS / EJ</td>
</tr>
<tr>
<td>-------------------</td>
<td>-------------</td>
<td>--------------</td>
</tr>
<tr>
<td>(Hitachi Computer Aided Translation System / Japanese to English)</td>
<td>(Hitachi Computer Aided Translation System / English to Japanese)</td>
<td></td>
</tr>
</tbody>
</table>

| 2) Name of Organization (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 | 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 | Dependency grammar Phrase structure grammar Case phrase structure grammar (1) Morphological analysis (2) Syntactic/semantic analysis (3) Transformation of Conceptual Dependency Diagram (4) Syntactic generation (5) Morphological synthesis approx. 5,000 | 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 approx. 2,000 |

<table>
<thead>
<tr>
<th>- Analysis grammar</th>
<th>- Generation grammar</th>
<th>- Translation processes</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Number of rules</td>
<td>(1) Morphological analysis (2) Syntactic/semantic analysis (3) Transformation of Conceptual Dependency Diagram (4) Syntactic generation (5) Morphological synthesis approx. 5,000</td>
<td></td>
</tr>
</tbody>
</table>

| 7) Dictionaries | Basic dictionary --- 50,000 words Technical term dictionary (option) --- 250,000 words User dictionary |

| 8) Equipment | GDL (Grammar Description Language) and PL/I VOS3 (Virtual-storage Operating System 3) HITAC M series computers |

<table>
<thead>
<tr>
<th>9) Performance</th>
<th>Translation speed</th>
<th>Translation cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>20,000-60,000 words/hour (CPU: HITAC M-680) depends on operational environments</td>
<td>30,000-60,000 words/hour (CPU: HITAC M-680) depends on operational environments</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>10) Facilities</th>
<th>Dependency specification Selection of alternative sentence structure Alternative word selection</th>
<th>Conversational editors (for expert and nonexpert) Batch maintenance program</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Pre-editing</td>
<td>- Post-editing</td>
<td>Conversation editor development</td>
</tr>
<tr>
<td>- Dictionary development</td>
<td></td>
<td>Batch maintenance program</td>
</tr>
</tbody>
</table>

--- 149 ---
Name of project
PAROLE

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

Current status of the system
Has been conducted as a research project

Feature of the system
Use of semantic transfer method
System expandability (Grammar Rules, Dictionaries, Facilities, etc.)

Translated language
Japanese to English

Strategies on translation
1) Type of grammar for analysis and generation
   Case Grammar and Tree Transducer for analysis
   Phrase Structure Grammar and Tree Transducer for generation
2) Translation processes
   Japanese analysis using Case Frame
   Japanese to English transfer using Dependency Structure
   English generation using Phrase Structure
3) Rule size for analysis, transfer and generation (Aug. 1987)
   Analysis 200 rules (approx.)
   Transfer 100 rules (approx.)
   Generation 200 rules (approx.)

Dictaonaries
1) Structure
   Represented by Lisp S-expression
   Semantically classified
   Including morphological, syntactic and semantic information
2) Size
   Basic 5000 words (Aug. 1987)

Equipment
1) Implementation language
   Prolog and Lisp
2) Operating system
   Genera
3) Type of CPU
   Symbolics

Performance
Not measured

Facilities
Grammar development environment
Language for describing tree-to-tree-transducing rules
Debugging utilities for rule developments
1. Name of system or project: Mu Project.

2. Name of organisation and contact address:
   Nagao Laboratory, Dept. of Electrical Eng., Kyoto Univ.
   Address: Dept. of Electrical Eng., Kyoto Univ., Kyoto, Japan

3. Current status of the system:
   The Mu project started at April, 1982 and completed at March, 1986. Two systems, Japanese to English MT system and English to Japanese MT system were developed. 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. The project aims to develop a system which will be used at JICST (Japan Information Center for Science and Technology) for actual translation services from April, 1990.

4. Features of the system:
   The Mu systems were developed to show the technical feasibilities of practical MT systems which translate texts of certain restricted document types and subject fields (abstracts of scientific and technological papers, especially electrical engineering fields). The systems do not expect any pre- and post-editions.

   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 in translation:
   A special software called GRADE is developed for the project. 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. Special cares are taken to treat large gaps of the two languages, English and Japanese. 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. The Pre-Transfer-Loop transforms source language oriented representations into more neutral ones, and the Post-Transfer-Loop transforms neutral representations into more target oriented ones. Because of these two loops, the main transfer can work on deep semantic representations.

7. Dictionaries:
   (a) Structure: 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.
   (b) Size: About 80,000 items. The dictionaries for the new project will be extended to cover 300,000 items.

8. Equipment:
   - Uri-LISP on FACOM M332, M780 (FACOM OS/IV P4 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 or project.
   HELTRAN-J/E(Welcom TRANslatin system-Japanese/English)

2. Name of organization and contact address.
   T: DASAI, Information Systems & Electronics Development Lab., Mitsubishi Electric
   Corp., 5-1-1 Ofuna, Kanagawa 247, Japan.

   Under commercial development.

4. Feature of the system.
   Transfer system based on logic programming.

5. Translated languages.
   Japanese to English.

   1) Type of grammar for analysis and generation --- Phrase structure grammar.
   2) Translation processes --- Tree transducer.
   3) Rule size for analysis, transfer and generation --- about 1000.

7. Dictionaries
   1) Structure --- Basic lexicon / Technical terminology / User dictionary.
   2) Size --- Basic-50,000 / Technical-30,000 for information processing.

8. Equipment
   1) Implementation language --- ESP(Extended Self-Contained Prolog)
   2) Operating system --- SIMPOS
   3) Type of CPU --- WELCOM PSI

9. Performance
   1) Translation speed --- 5000 words/hour(CPU:)
   2) Translation cost --- (evaluating)

10. Facilities
    Pre, Post-editing, interactive/batch translation, and grammar and dictionary
    development environment.
1. PIVOT
2. NEC Corp., Minato-ku Mita, Tokyo TEL. (03) 454-1111
3. Released for closed users
4. Interlingual Approach
5. English-to-Japanese, Japanese-to-English
6. 1) dependency grammar, augmented shift-reduced parser
tree to tree transducer
2) ai:Morphological Analysis
   a2:Gramatical and Semantic Analysis
   a3:Semantic Extraction
   g1:Conceptual Wording
   g2:Grammatical Generation
   g3:Morphological Generation
3) analysis more than 3000
generation about 2500
7. System Core Dictionary
   Japanese 40000 entries
   English 53000 entries
Term Dictionary
   covers more than 20 domain
   each of which includes less than 20000 entries
8. 1) C-language
   2) Acos-4
   3) Acos Main-Frame
9. 1) 60000
   2) 1500yen/A4 (double space) (tentative)
10. Bach and Interactive operation Modes
    Bach Tools:  Text Processing and Management
                Dictionary updation and management
                Text transfer between terminal and host machine
    Interactive: Bi-lingual and mono-lingual text processing and management
                Functions Unknown word detection
                Dictionary updation
                etc.
Machine Translation System
at NIPPON • DATA GENERAL CORPORATION

1) Address

1st R&D Group  NIPPON • DATA GENERAL CORPORATION
2165, Nochida, Gyoda-shi, Saitama, 361 Japan

2) Current state

Prototype system has been implemented.
And we are upgrading it for our own company's use.

3) Goals

Component of totally comprehensive electric office system.
Easy knowledge acquisition capability.

4) Languages

English ---) Japanese

5) Approaches

Phrase structure and deep case analysis using tree-transducer

Transfer approach
Morphological analysis
↓
Syntactic & Semantic analysis (about 1000 rules)
↓
Transfer (600 rules)
↓
Synthesis (400 rules)
↓
Morphological synthesis

Basic softwares are written by C.
Main computer is ECLIPSE MV series.

6) Dictionary

English-Japanese and Japanese dictionaries are on computer.
1. Name of system or project.
   PENSEE.

2. Name of organization and contact address.
   OKI Industry Co., Ltd.

   Released in Autumn 1986.

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.

5. Translated language.
   Translation of Japanese into English.
   Developing Translation of English into Japanese.

   1) Type of grammar for analysis and generation.
      Case grammar with deep structure.
   2) Translation processes.
      The translation system is made up of Japanese morphological analysis, interactive translation with syntactic and semantic analysis, and English morphological generation.

7. Dictionaries.
   1) Structure 2) Size
      System dictionary 60,000 words
      User's dictionary 40,000 word-capacity

8. Equipment.
   1) Implementation language.
      C language.
   2) Operating system
      UNIPLUS+ (UNIX System V base).
   3) Type of CPU
      MC68010, MC68020.

   1) Translation speed
      4,000 words / CPU-hour (MC68010 10MHz)

10. Facilities.
    1) A bilingual editing system for Japanese and English.
    2) A dictionary editing system which allows the registration of words into the user's dictionary.
    3) A preprocessing system which shows the result of Japanese morphological analysis.
1. System name
   Ricoh English-Japanese Machine Translation System (RMT)

2. Address
   Ricoh Company Limited
   Research and Development Center
   4686 Nippa-cho
   Kouhoku-ku, Yokohama
   223 Japan
   Phone:045-593-3411

3. Current status
   Now, under development. (Will be brought to market in 1988)

4. Features
   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. Language Pair
   English to Japanese

6. Strategies
   Morphological Analyses
   Use some features of sentences. (200 rules)
   Syntactic Analyses
   Augmented context free grammar. (2,200 rules)
   Plausibility Evaluation
   Makes dependency tree using semantic features.
   (60 features)
   Transfer
   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
   1) Implementation language : C
   2) Operating system : UNIX System V
   3) Type of CPU : 3B2(AT&T)

9. Performance
   Translation speed : 4,500 words/h

10. Facilities
    Editing Software
    *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
    *Source text may also be entered through OCR.

    Dictionary development utilities
    *You can enter new terms in User Dictionary by a menu provided.
1. Name of the system
   Translation Word Processor SWP-7800

2. Contact address
   Sanyo Electric Co., Ltd.
   Dept. of Information Systems
   2-7-25 Edobori
   Nishi-ku, Osaka 550
   (06) 443-5144

3. Current status of the system
   SWP-7800 has been on the market since April, 1987.

4. Characteristics
   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. Languages subject to translation
   Japanese to English translation only

6. 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. Dictionary
   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 3500 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: Commercial base

4. Feature of the system: High speed translation on desktop computer

5. Translated language: English to Japanese

6. Strategies of translation:
   1) Type of grammar for analysis 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.
   2) Translation processes:
      Employs advanced natural language processing by means of
      semantic analysis.
   3) Rule size for analysis, transfer and generation:
      Around 1,000 rules (for syntactic analysis).
      Transfer and generation processes are built by the
      combination of rules and procedures.

7. Dictionaries:
   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. 1) Implementation language: Language C
    2) Operating system: UNIX System V
    3) Type of CPU: MC68010, 68020
    4) Hardware: OA-110WS, OA-210, OA-310, IX-7

9. Translation speed: 5,000 words/hour (MC68010)

10. Facilities:
    1) Pre-editing:
        Users can specify (a) the correct part-of-speech of
        syntactically ambiguous words; and (b) the phrase boundaries.
    2) Post-editing:
        Interactive Syntactic and Lexical Disambiguation, and Learning.
1. Name of system or project

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-Ro, Kawasaki 210, Japan

3. Current status of the system

Commercial

4. Features of the system

Implemented on Engineering Workstation
Transfer approach

5. Languages translated

English-Japanese

< TRANSLATION PROCESS >

6. Strategies of translation

1) Type of grammar for analysis and generation
   ATN + Lexical Grammars

2) Translation Process (figure)

7. Dictionaries

1) Structure
   for each entry: syntactic category; syntactic properties;
   semantic properties; lexical rules;

2) Size
   Max 130,000 words (general 50,000; technical 50,000;
   user-defined up to 30,000)

8. Equipment

1) Implementation language : C
2) Operating system : UNIX
3) Type of CPU : MC68020

9. Performance

Translation speed : 7,000 words/hour

10. Facilities

Pre- and Post-editor, Bi-lingual editor, Dictionary
    maintenance system, Word processors