Science and Technology Agency's Mu Machine Translation Project

Makoto NAGAO, Jun'ichi TSUJII and
Jun'ichi NAKAMURA
Dept. of Electrical Engineering, Kyoto University,
Yoshida-honcho, Sakyo-ku, Kyoto-shi 606, Japan

This paper describes the current status of a machine translation project that aims to develop Japanese-to-English and English-to-Japanese translation systems for abstracts of scientific and technical articles. The paper describes the development of Japanse-to-English translation systems, and discusses methodologies for evaluating translation results. Particular emphasis is placed on techniques necessary for translation between Japanese and English – languages having fundamentally different structures.

1. Outline of the Project

The Mu project intends to develop a Japanese-to-English translation system for abstracts of scientific and technical papers written in Japanese, primarily those published by Japan Information Center for Science and Technology (JICST). An English-to-Japanese system will translate abstracts published by INSPEC *.

About two and a half years have elapsed (as of the fall, 1984) since the inauguration of the project. The first prototype of the Japanese-to-English system has been completed and is currently being improved. The system has the following features:

(1) Based on a transfer scheme which performs translation through three steps: analysis, transfer, and generation. This approach permits the system to be used for multilingual translation, makes it easier to locate the causes of translation errors, and facilitates system development by dividing the whole translation process into three conceptually different phases.

(2) A programming language called GRADE was specially designed for grammar rule specification, in which (human) grammar writers can specify arbitrary tree-to-tree transformation rules. Each node of a tree can be annotated with any number of semantic features and other information. The same programming language is used to specify the three phases of translation – analysis, transfer and generation.

(3) In GRADE, the grammar is organized in the form of sub-grammar networks which break grammatical rules into a number of packages for application. This speeds up analysis, allows several researchers to develop grammati-

---

* INSPEC is a document database which covers fields such as physics, electrical engineering, electronics, computers & control, and is published by the Institute of Electrical Engineers in Great Britain.

cal rules in parallel, and helps uncover the causes of translation errors.

(4). Grammatical rules specific to individual words can be written in dictionaries. This assists in dealing with linguistic phenomena that are dependent on specific lexical items.

(5) Even if an input sentence contains unknown words, the system can produce partial translations. Unknown words in the source language appear directly (i.e., untranslated) in the translated sentence, this approach significantly outperforms methods that reject sentences with unknown words.

2. Analysis of Japanese Sentences

An essential requirement for machine translation systems is the ability to handle as many linguistic phenomena as possible. Systems should be evaluated as to whether they satisfy this demand. To this end, our systems use the various techniques described below.

2.1. Overview of the Analysis Process

The Japanese analysis process receives the results of the morphological analysis, determines the phrase structure of an input sentence, and converts the structure into a dependency structure which expresses the semantic relationships among the words in the sentence. Our analysis grammar has the following features:

(1) Instead of outputting all possible interpretations, it provides a single interpretation which seems most appropriate both syntactically and semantically. When local ambiguity exists, the grammar tries several times to analyze that portion, compares the results of the analyses to find the one seemingly best for that stage, and proceeds with the subsequent stage of analysis.

(2) Rules specific to individual words can be stored in the lexical entry of a dictionary as lexical rules. This allows the grammar to handle word-specific phenomena without the necessity of significantly modifying the entire grammar system.

(3) The current grammar is domain-specific, i.e., it has been specifically developed to translate abstracts of scientific and technical papers published by JICST. This grammar is highly modular, consisting of several subgrammars, each of which handles a particular linguistic phenomenon. This configuration permits the grammar to be applied to a new field only by adding a subgrammar related to that field.

2.2. Analysis Flow

The processing flow in the analysis process is shown in Fig. 1.

2.3. Examples of Japanese Analysis

(1) *Disambiguation of words that can be assigned to multiple parts-of-speech categories.* Words interpreted as belonging to multiple parts-of-speech categories appear in Japanese sentences, although their frequency is lower than in English. For example, the word " で (de)" can be interpreted as either the renyo-form of the auxiliary verb " だ (da)" or as postpositional case particle. Also, the renyo-form of some adjectives and adverbs are identical to certain words belonging to different parts-of-speech categories. These words are disambiguated in the first stage of the analysis.

▶ この方法はR. Bellman の理論から導いたもの で、解析幾何理論の応用できる。

This method is one derived from theories of R. Bellman, and this method is an application of geometric diffraction theories.

▶ 本論は、エネルギー管理制御システムについての 現状を報告することが目的で、New England 電話 会社の例を挙げている。

As for this theory, aims are to report current conditions about energy management and control systems, and this theory has given examples of New England Telephone Company.

▶ 計算機制御方式で、測定インタフェース、リレー群、 プリンタを装備。

Measurement interfaces, relays and printers are equipped by computer control types.

▶ この回路はわずかの修正で、他の計算機にも接続 できる。

This circuit can also be connected with other computers by slight correction.
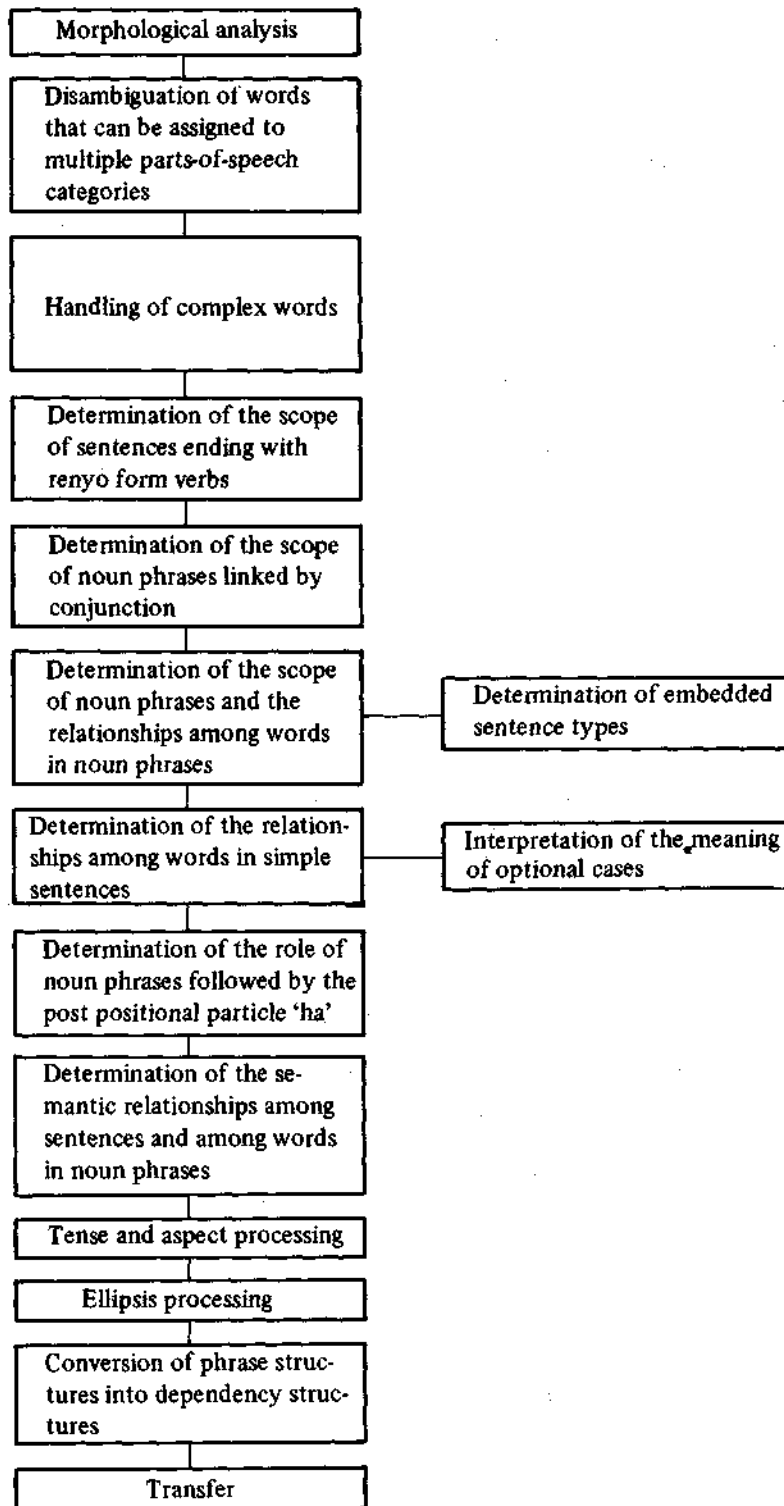
```
┌─────────────────────────────┐
│   Morphological analysis    │
└─────────────────────────────┘
              │
┌─────────────────────────────┐
│ Disambiguation of words     │
│ that can be assigned to     │
│ multiple parts-of-speech    │
│ categories                  │
└─────────────────────────────┘
              │
┌─────────────────────────────┐
│                             │
│ Handling of complex words   │
│                             │
└─────────────────────────────┘
              │
┌─────────────────────────────┐
│ Determination of the scope  │
│ of sentences ending with    │
│ renyo form verbs            │
└─────────────────────────────┘
              │
┌─────────────────────────────┐
│ Determination of the scope  │
│ of noun phrases linked by   │
│ conjunction                 │
└─────────────────────────────┘
              │
┌─────────────────────────────┐         ┌──────────────────────────┐
│ Determination of the scope  │         │ Determination of embedded│
│ of noun phrases and the     │─────────│ sentence types           │
│ relationships among words   │         └──────────────────────────┘
│ in noun phrases             │
└─────────────────────────────┘
              │
┌─────────────────────────────┐         ┌──────────────────────────┐
│ Determination of the relation-│       │ Interpretation of the meaning│
│ ships among words in simple │─────────│ of optional cases        │
│ sentences                   │         └──────────────────────────┘
└─────────────────────────────┘
              │
┌─────────────────────────────┐
│ Determination of the role of│
│ noun phrases followed by the│
│ post positional particle 'ha'│
└─────────────────────────────┘
              │
┌─────────────────────────────┐
│ Determination of the se-    │
│ mantic relationships among  │
│ sentences and among words   │
│ in noun phrases             │
└─────────────────────────────┘
              │
┌─────────────────────────────┐
│ Tense and aspect processing │
└─────────────────────────────┘
              │
┌─────────────────────────────┐
│    Ellipsis processing      │
└─────────────────────────────┘
              │
┌─────────────────────────────┐
│ Conversion of phrase struc- │
│ tures into dependency struc-│
│ tures                       │
└─────────────────────────────┘
              │
┌─────────────────────────────┐
│         Transfer            │
└─────────────────────────────┘
```

Fig. 1. Processing Flow in the Analysis Process.

▶ この方法では伝搬方向を指定してそれを分離でき，一次元方程式を多次元の場合に容易に拡張できる。

It can be separated by designating directions of propagation by this method, and first-order equations can be easily expanded to multidimensional cases.

▶ Hg, I, Na, Sc, Sc⁺, ScI からの自己吸収放射の正味放射係数を実験と理論から個々におよび総合的に導いた。

Net radiation coefficients of the self-absorption radiation from Hg, the I, Na, Sc, Sc⁺and ScI are derived from experiments and theories individually and comprehensively.

▶ プリント配線の集積度が向上し，導体間隔がせまくなり，基板の絶縁抵抗が重要になった。

The integration of the printed wiring is improved, the conductor spacing becomes narrow, and the insulation resistance of substrates becomes important.

▶ 界面における粘度比は初期温度の変化によって小さくできることも判明。

It is also confirmed that the specific viscosity in interfaces can be decreased by the variation of initial temperatures.

(2) *Determination of the scope of sentences ending with 'renyo' form verbs.* Abstracts of scientific and technical papers in Japanese include many sentences terminating in renyo-form verbs, which indicate that the sentence is related to the subsequent sentence. The analysis stage determines the scope of sentence connection by referring to the semantic category of the preciates (intensional, extensional, descriptive verb, etc.).

▶ 計算機で基板のパターン設計を行い，部品の手配から製品までの各工程を制御するシステムを検討。

The systems are studied which carry out the pattern design of substrates by computers, and control each process from arrangements of parts up to products.

▶ 計算機で基板のパターン設計を行い，部品の手配から製品までの各工程に与える効果を検討。

The pattern design of substrates is carried out by computers, and effects on each process from arrangements of parts up to products are studied.

▶ 電力用半導体の標記パッケージ法について電気的，熱的，機械的性質について検討し，実際の応用分野での性能を述べ，本方法の利点についてのべた。

Subject packaging methods of power semiconductors are studied about electrical and thermal-mechanical properties, the performance in application fields the practice is described, and advantages of this method are described.

▶ プリント配線を使用目的で分類して各々の要求品質水準を示し，品質管理上着目すべき個所を列記。

Each required quality level is shown by classifying the printed wiring by use purposes, and the points are enumerated which should be noticed in respect of the quality control.

(3) *Determination of the semantic relationships among sentences.* A single postpostional particle that connects two independent clauses indicates several different semantic relationships between the clauses; different relationships must be expressed by different conjunctions like 'because' or 'in order that' in English ** Therefore, whether the correct relationship can be established directly affects the translation.

▶ 強度がゆらぎを示すため，観測が困難である。

Observations are difficult because the intensity shows fluctuations.

▶ 人間による翻訳がゆらぎを示すため，管理が困難である。

The control is difficult because 翻訳 by 人間 shows fluctuations.

▶ 計算機による翻訳の可能性を示すため，文献抄録に注目した。

文献抄録 was noticed in order to show the possibility of 翻訳 by computers.

** Note that the postpositional particle "NODE" can express two different relationships, which should be translated as either 'because' or 'in order that'.

▶ コンデンサに生じる電流, 電圧による検出器を述べ, 人体への作用との関係を示す。

Detectors by the current voltage formed in capacitors are described, and effects and relationships to human bodies are shown.

▶ 励振条件を変えて表面電流絶対値分布を示す。

The surface current absolute value distribution is shown by changing excitation conditions.

(4) *Determination of the scope of noun phrases linked by conjunction.* Abstracts of scientific and technological paper contain a huge number of conjuncted noun phrases. It is therefore crucial to exactly determine the scope of the conjunction. This can be done with considerable precision by referring to the semantic categories of nouns, as well as by referring to particular nouns and function words.

▶ システム工学, 情報工学, ソフトウエア工学など新しい技術の急速な成長により, IEE および電気工学教育の全パターンでの重点が移動した。

Rapid growth of new technology such as systems engineering, information engineering and software engineering moved the emphasis of IEE and electrical engineering education in all pattern.

▶ 国際電気委員会, ソ連国家規格, 衛生規則による人体に対する許容接触電圧および電流の規定, および主要な防御系を述べる。

The regulation for permissible contact voltages and currents by International Electrotechnical Commission Soviet national regulation health regulation to human bodies and major protection systems are described.

▶ ほとんどのスイッチドキャパシタ (SC) 回路の出力雑音はサンプル・ホールド雑音と広帯域雑音の二つのクラスに分けることができる。

The output noise of almost all switched capacitor (SC) circuits can be divided into two classes of sample hold noise and broadband noise.

▶ 集中／分布回路網の時間および周波数領域における状態空間解析。

State space analyses in time and frequency domains of lumped/distributed networks.

▶ この新設計システム方法論に基づく, 幾何学的構造機能に関する会話形設計プログラムである CAD-AM の利用と, 生産性, 設計の正確さ, ドキュメンテーション品質の改善効果を示す。

The utilization of the CADAM which is interactive design programs about the geometric construction function, the productivity, the accuracy of the design and improvement effects of the documentation quality based on this new design system methodology are shown.

(5) *Determination of embedded clause types.* Embedded clauses in Japanese do not always directly correspond to relative clauses in English. For example, such clauses must often be translated using infinitive phrases or that-clauses in apposition, depending on the context. This requires the analysis phase to correctly distinguish the semantic relationship between an embedded clause and the noun modified by that clause.

▶ ベネフィットの総和についての購買者の効用を測定する 1 つの方法を示す。

A method for measuring the utility about totals of the benefit of buyers is shown.

▶ 初期に加えられた磁束を圧縮することによって高磁場を発生する方法を提案する。

Methods of generating high magnetic fields by compressing the magnetic flux applied to the initial stage are proposed.

▶ 真空中の誘電体板に入射する平面波の回折により生じる表面波の複素振幅を積分方程式の数値解により見出す。

Complex amplitudes of the surface waves formed by the diffraction of the plane waves emitted to dielectric plates in vacuums are found by numerical solutions of integral equations.

▶ 接着剤の選択における設計上考慮すべきことを検討。

What should be considered in respect of the
(that節ではなく what節になっていることに注目)
design in the selection of the adhesive is studied.

(6) *Interpretation of the meaning of the optional cases* * A Japanese postpositional case particle represents various deep case relationships. Therefore, it must be translated into different English prepositions depending upon the relationship represented by the particle.

▶ 相-電位回路のパラメータは測定結果の数学的確率論的方法による処理<u>で</u>得られる当該電気設備の設計電圧に基づいて選択すべきであると述べた。

It is mentioned that parameters of phase-voltage circuits should be selected on the basis of design voltages of当該電気設備which can be obtained <u>by</u> the processing of measurement results <u>by</u> mathematical statistical methods.

▶ 電気回路の計算機によるレイアウト<u>で</u>は力のモデルが有用である。

Models of the force are useful <u>in</u> layouts by computers of electric circuits.

▶ 主として現在市場<u>で</u>活躍している数値制御高速穴あけ機について説明。

The numerically-controlled superhigh speed drilling machines are explained which are active mainly now <u>in</u> markets.

▶ 逆制動放射によるレーザ吸収項について二つの取り扱い方<u>で</u>問題を解き, 数値解を求めた。

Problems are solved by two handling methods about laser absorption terms <u>by</u> the inverse damping radiation, and numerical solutions are obtained.

▶ LSIの発達<u>で</u>プリント配線基板.1枚当りの集積度が増大し, 確実な試験法が重要になった。

<u>Advances of LSI</u> increase the integration of printed circuit boards per one piece, and reliable testing methods become important.

*** The case element which is strictly governed by a verb can be translated appropriately by referring to the verb dictionary. The verb dictionary contains the information on how to instantiate the deep case relations using English prepositions. Such case elements constitute the obligatory case. On the other hand, other case elements are collocated with almost all verbs, and their omission does not yield ungrammatical sentences. Such case elements constitute the "optional case".

(7) *Determination of the predicate linked by "は" (Ha).* A noun phrase followed by the postpositional particle "は" (Ha) is connected to the predicate most remotely located in the sentence.

▶ 荷電粒子の運動は回折放射バーストを<u>ともなう</u>ことを確認した。

It is confirmed that the motion of charged particles accompanies diffractive radiation bursts.

▶ 振幅スクリーンでの伝搬は位相スクリーンと質的に共通の法則性を<u>もつ</u>ことを示す。

It is shown that the propagation by amplitude screens has the law common qualitatively with phase screens.

▶ 放電のスペクトル出力の調整はその放電の加熱を用いることにより<u>可能である</u>。

The adjustment of the spectrum output of the discharge is possible by using the heating of the discharge.

(8) *Determination of the syntactic structure of noun phrase and semantic relationships between words in them.*

▶ まず熱力学第二法則とその物理的解釈を述べ, 次に Schutz の理論を解説する。

First, second law of thermodynamics and the physical interpretation are described, and <u>Schutz's theory</u> is explained next.

▶ 加速された<u>荷電場</u>のブラデオン部分とタキオン部分への<u>分解</u>。

<u>The analysis of accelerated charged fields</u> to bradyonic and tachyonic parts.

▶ 気体の火花破壊のストリーマ理論の困難を解決するために, 粒子ビームとプラズマ振動の相互作用を考慮し, 準中性プラズマの温度, 電離度, 伝導度に与える効果を考察した。

Interactions <u>between</u> particle beams and plasma vibration were considered, and effects on temperatures, ionization degrees and conductivity of the quasi-neutral plasma were studied in order to solve the difficulty of streamer's theories of the spark breakdown of the gas.

▶ 静止空気中の熱せられた円筒のコロナによる冷
却。

° The cooling of heated cylinders in the still air
by coronas.

▶ NbSi の電子ビーム蒸着中に酸素を添加すると A
15 相の安定性が増大し, 6 K から 8.3 K の超電導性
転移の発生が増す。

If the oxygen is added during the electron beam
evaporation of the NbSi, the stability of A 15
phases will increase, and the generation of the
superconductive transfer from 6 K to 8.3 K incre-
ases.

(9) Handling of 'fuku' postpositional particles.

▶ 真空中で加速された電荷の場をブラデオン部分と
タキオン部分とに分解し, 後者のみが電荷によって
放出される全エネルギーに寄与することを示した。

Fields of the charges accelerated in vacuums
are decomposed with bradyonic and tachyonic
parts, and it is shown that the contribution is
made to the total energy which only the latter
release by charges.

▶ 本方法は広範囲の集積ミクロ回路のマクロモデル
を構成する場合にも適用できる。

This method can also be applied to the cases in
which macromodels of integrated microcircuits
of wide areas are formed.

▶ 正負イオンのアンバランスの度合いの効果, 平均
正イオン密度の効果, 極性反転電圧の効果, 拡散の
効果, ギャップ長の効果などについて考察。

Effects of degrees of the imbalance of positive
and negative ions, effects of the mean positive ion
density, effects of polarity inversion voltages,
effects from the diffusion, effects of the gap lengt-
h, and so on are studied.

▶ 3 種類の接着システムについてせん断強さ, はく
離強さ及び付着耐久度などの機械的性質について
比較試験を実施。

Comparative tests are carried out about me-
chanical properties such as shearing strength,
peel strength and adhesion durability about three
types of adhesion systems.

*(10) Handling of special expressions, such as
parenthetical text.* Abstracts often include, paren-
thetical text, algebraic expressions, units of mea-
sure, itemizations, etc. In practical MT systems,
we have to treat such 'non-linguistic' symbols as
parentheses, symbols in mathematical formulas,
etc. Parenthetical expressions for example, can be
used for different purposes such as additional
explanation, abbreviations, apposition, etc. Each
of these must be translated differently.

▶ 標題静磁界解析につき解析手法と解析例(リフテ
ィングマグネット, 電磁クラッチ/ブレーキ, ソレ
ノイド)を紹介した。

Methods for analyzing and analysis sample
(lifting magnets, solenoid clutch/brakes andsol-
enoid) are introduced about the subject analysis
of static magnetic field.

▶ 非線形時変 n ポートのエネルギーに関連した概
念：受動性と無損失性。

The concepts related to the energy of the
nonlinear time-varying n-port : passivity and
losslessness.

▶ 4) 熱挙動設計には OF のスネーク熱挙動計算式
が適用できる：ワイヤしゃへい層送電損失(内部自
己インダクタンスの理論式と測定値, 回路損失率の
計算方法, うず電流損失), 熱挙動特性(基礎検討,
延線工法, スネーク取り工法, 熱挙動実験)。

4) snake thermal behavior calculation formulas
of OF can be applied to the thermal behavior
design: wire shielding layer transmission loss (the-
oretical formulas, measured values, calculation
methods and eddy current loss) and thermal be-
havior characteristic (basic study, wire extension
techniques, snake catching techniques and thermal
behavior tests).

3. Japanese-to-English Transfer Process

3.1. Overview of Transfer

The Japanese analysis described above pro-
duces a dependency structure which represents the
semantic relationship between the words in an
input Japanese sentence. The transfer process uses

the dependency structure to construct its English equivalent. The transfer process performs the following tasks:

° (1) An appropriate English equivalent of a Japanese word by examining the syntactic and semantic environments in which the word is used. For verbs, for example, word selection is conducted by referring to semantic information (e..g, a noun of a particular semantic category is filled in a certain specific case slot of the verb.) and idiomatic information (e.g., the verb is collocated with a specific noun.) Such information is stored in dictionaries for the Japanese-to-English conversion.

(2) For a pair of languages, such as Japanese and English, each belonging to quite different language family, translated sentences often appear to be odd, unless the syntactic structures of input sentences are substantially transformed. The transfer process includes such structural transformation, depending on the language pair involved.

The transfer process must be redesigned for each new language pair. This requires the system to be as modular as possible so that it can be expanded in the future into a multilingual version.

3.2. Transfer Flow

The transfer process is outlined in Fig. 2. As the figure illustrates, transfer takes place in three stages.

(1) *Pre-transfer loop.* If a sentence structure is unique to Japanese and thus has no direct English equivalent, the pretransfer converts it to a more neutral one.

(Ex." "(X be many)

→ " "(Many Xs exist or There

are many Xs)
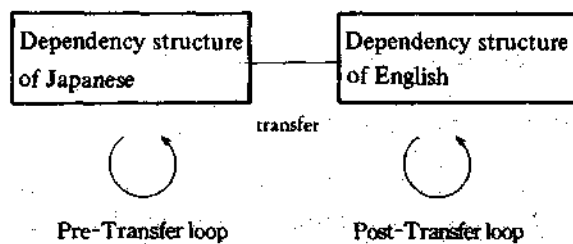
Fig. 2. The Transfer Process.

(2) *Transfer.* This stage performs word selection and local structural conversion accompanying such selection as described step-by-step below, transfer starts with the top node of the Japanese dependency structure, and recursively activates conversion of the lower nodes.

a. Input sentence: " オペレータのプラント運転による負担を計算機導入によって軽減することが|できる " (The burden on overworked plant operators can be reduced by introducing a computer.) This sentence is analyzed to have the following dependency structure:
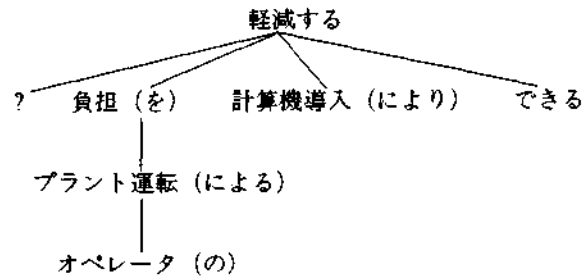

b. Checking of collocation of predicate and case elements. (This step handles idiomatic relationships between predicates and nouns. It does not apply to this example.)

c. Dictionary look-up of the predicate. (This step determines the appropriate English equivalent for the predicate by referring to the semantic categories of case elements linked to the predicate.)
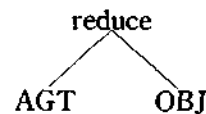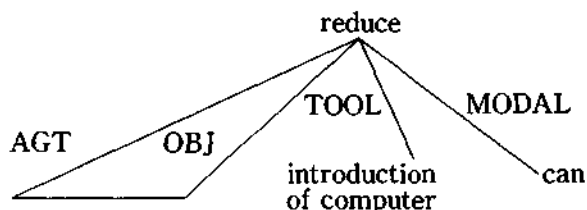

d. Conversion of obligatory cases
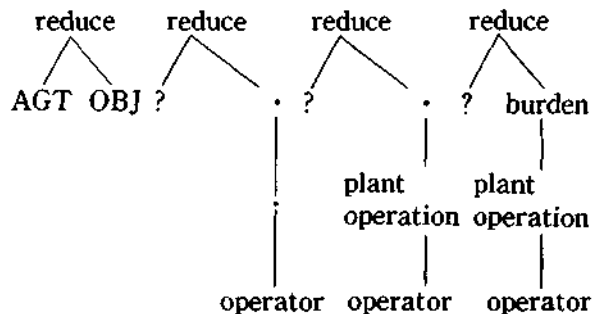
e. Conversion of the optional cases. (This step handles not only the optional cases of noun phrases, but also adverbs and subordinate sentences used as case elements.)

When a case element in the sentence is modified by an embedded clause, the embedded clause is converted into a corresponding English dependency structure by recursively calling the above steps.

(3) *Post-transfer loop.* English dependency structure are modified to produce more English-like expressions. (Ex.) " A を回転させる " (A is rotated.)
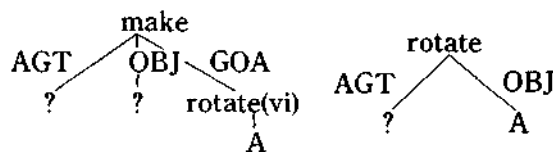
3.3. Examples of Transfer

(1) *Translation based on collocation of predicate and case elements.* Translation between Japanese and English must deal with the correspondences between words occurring in pairs. Particularly, ideas expressed by a paired noun and verb (or adjective) in Japanese are often represented by a single verb (or adjective) in English (e.g., " 影響 (influence)を 与える(to give)" (to influence)). Such phenomena can be dealt with by word-specific rules, such lexical rules are stored in the transfer-dictionary entry for each noun.

▶ 米国ボストンの New England 電話会社におい て, マイクロコンピュータをベースとしたエネルギ 一管理制御システムを設備し, エネルギー原価の低 減に成功した事例を紹介。

The cases are introduced in which the success is made in the reduction of energy costs by equipping the energy management and control systems based on microcomputers in New England Telephone Company of the United States Boston.

▶ 実験と比較し, よい一致を見た。

The comparison is made with experiments, and good coincidence is obtained.

(2) *Translation of verbs (adjectives) with wide ranges of usages.* Japanese verbs such as " する " (suru) (to make, do, perform,...) " なる " (naru (to consist of, provide, ...) and " みる " (miru) (to look, regard, try, ...) have an extensive range of usage, and therefore must be translated into different English verbs depending on context. A dedicated subgrammar provided for each of these verbs is activated during the transfer to select the appropriate translation by examining various sources of information.

▶ 試験法にはイン・サーキット・テストと独立機能 テストがある。

Testing methods include in-circuit tests and independent functional tests.

▶ 誘導体厚が薄いので背後電極上に電界が集中し, 沿面線上ストリーマと背後電極間の充電電流が大 きい点に特長がある。

Characteristics are that electric fields concentrate on the back electrode because the dielectric thickness is thin and the charging current between the streamer on surface wire and the back electrode is large.

▶ 標記の基板は通常のプリント基板やセラミック基 板に比べ 10%以上安く, 機械強度, 熱伝導も優れて いるが, ボイド, Na や K イオンの存在などの欠点 がある。

Above-mentioned substrates are inexpensive in comparison with the ordinary printed circuit boards and ceramic boards above 10 %, the mechanical strength and the thermal conduction are also excellent, and it has shortcomings such as voids, Na and the existence of K ions.

▶ 成形法としては, トランスファ成形と射出成形が あり, それぞれ金型, 成形機, 成形条件について比 較, 説明。

The transfer formation and the injection molding provide molding techniques, the comparison is

respectively made about metal molds, molding machines and molding conditions, and the explanation is made.

(3) *Translation of ordinary verbs (adjectives).* Japanese verbs not handled by rules (1) and (2) can be translated into their English equivalents by referring to the semantic categories of nouns linked to the verbs as case elements (e.g., The Japanese verb " 生じる " should be translated into 'to form', 'to occur', 'to produce', 'to happen' etc., depending on the semantic categories of nouns in the 'OBJECT' case slot.)

▶ 与えられた正規双直交系を含む正規双直交基底の構成法に関して，具体的な構成法を示すことにより補題を証明。

Lemmas are verified by showing specific constitution methods about constitution methods of the normal double orthogonal bases which include given normal double orthogonal systems.

▶ ナトリウムおよびよう化スカンジウムを含む鉛直水銀アークの放射および対流模型。

The radiation and convection models of the vertical mercury arcs which contain the sodium and the scandium iodide.

▶ この厳密なアプローチは，導電材料中の渦電流分布の正確な決定に導く。

This strict approach results in accurate determination of the eddy current distribution in conductive materials.

▶ Hg, I, Na, Sc, Sc⁺, ScI からの自己吸収放射の正味放射係数を実験と理論から個々におよび総合的に導いた。

Net radiation coefficients of the self-absorption radiation from Hg, the I, Na, Sc, Sc⁺, and ScI, are derived from experiments and theories individually and comprehensively.

(4) *Modification of source language dependency structures.* The main flow of the transfer process assumes that the dependency structure of Japanese closely parallels that of English. Word selection is recursively performed stating with the top node. Therefore, when the dependency structure differs between the two languages, the Japanese structure

should be modified in the pre-transfer loop. The expression " X は～する 傾向にある ." (X tends to do something) for example, is interpreted by the Japanese analysis grammar as having " ある " as the main verb, "X" as the subject, etc. The sentence is translated into English using the auxiliary-verb-like expression 'tends to".

▶ セラミックチップキャリアの実装方法が今後のLSI高密度実装のキーテクノロジーとなる傾向にあることを示した。

It is shown that packaging methods of ceramic chip carriers tend to become the key technology of LSI high-density packagings.

▶ また貯蔵単位，輸送単位ごとに個数をまとめる目的もある。

Then, aims are also to rearrange numbers per storage units and the transportation unit.

▶ はんだメッキは，プリント配線板の金属エッチングレジスト，部品リードのはんだ付けを良くする目的で使用されている。

The solder plating is used in order to improve the soldering of the metal etching resist and part leads of printed circuit boards.

(5) *Handling of affixes.* The major parts of speech in Japanese, such as verb and noun, are expressed by English words of similar parts of speech. Affixes, such as " 用 " (for), " 中 " (in), and " 全体 " (whole), and nouns of special nouns, however, must be translated into words of various parts of speech in English, such as prepositions, determiners, reflective pronouns, and adjectives. This requires that a number of rules be written in advance for individual words.

▶ 真空中の誘電体板に入射する平面波の回折により生じる表面波の複素振幅と等価表面電流分布を積分方程式の数値解により見出す。

Complex amplitudes and equivalent surface current distribution of the surface waves formed by the diffraction of the plane waves emitted to dielectric plates in vacuums are found by numerical solutions of integral equations.

▶ 1981 年度の船舶用電気技術。

The electrical technology for marine vessel in 1981.

▶ システム全体を左右するアートワークシステム，産業用基板に重要な穴あけ・外形加工，多層基板を製作する新しい方式などを紹介。

The drilling and the outline machining and new systems for manufacturing multilayer boards which are important for the artwork systems affecting the whole system and industrial substrates are introduced.

(6) *Conversion of auxiliary-verb-like expressions into verbs.* Auxiliary-verb-like expressions in Japanese, such as " ようとする " (to attempt) and " たがる " (to want) are converted into English expressions using verbs.

▶ 最近有線伝送の漏話を対数正規分布で解析しようとする動きがある。

There is movement to recently attempt to analyze the crosstalk of the wired transmission by the logarithmic normal distribution.

(7) *Conversion of causative expressions into transitive verbs.* The Japanese causative expresssion " させる " (to make) of often used only to enable intransitive verbs to function as transitive ones; when literally translated into English, the expression is likely to be redundant. Such expressions should be expressed simply using English transitive verbs.

▶ 結合のスロットを回転させて放射による粒子エネルギー損失を連続制御できる。

The particle energy loss by the radiation can be continuously controlled by rotating coupled slots.

▶ 生産性を向上させるレーザマイクロソルダリング。

The laser microsoldering improving the productivity.

▶ ポッケルス素子を回転させることにより，素子に付着する電荷の影響を除去しつつ，空間電界を直接測定する方法を提案し，直流正コロナが生じている双曲針対平板電極間の電界を測定。

Methods of performing the direct measurment while eliminating effects of the charges attached to elements by rotating Pockels elements are proposed, and electric fields between the double curved needle flat plate electrode between which DC positive coronas have been formed are measured space field.

4. English Generation Process

4.1. Overview of the Generation Process

The generation process converts an English dependency structure passed by the transfer stage into an English phrase structure expression. To produce grammatically correct English sentences, information about grammatical number is added to the expression, tense/aspect/modal expressions with auxiliary verbs are generated, concord processing (i.e., number agreement of subject and verb) is performed, and verb tense are adjusted. When necessary, the generation process also refines the style of the sentence using 'It-that', 'There', and other structures to make it easier to understand.

4.2. Generation Flow

Generation is accomplished in three stages.

(1) *Pre-generation.* This stage performs any structural transformations that can be done before English generation.

(2) *Generation.* In this stage, dependency structures are converted into phrase structures. Generation subgrammars are applied recursively to the dependency structure, from the top node down to all the lower nodes. The processing flow is briefly described below.

a. Generation of delimiters
b. Dictionary consultation
c. Generates the obligatory cases and determines their positions in the phrase structure
d. Generates the optimal cases and determines their positions in the phrase structure
e. Deletes the delimiter and generates an S-node.
f. Performs passivization or uses a substitution verb. *
g. Handles verb phrases

* 実 行 し た (to execute). if we cannot find both the agent and the object of this verb in the sentence, we generate an English sentence such as "execution was performed." In this sentence, the predicate "to execute" is nominalized and the substitution verb (to perform) is inserted.

(3) *Post-generation.* English sentences generated by the preceding stages have a passable English structure but may sometimes be extremely hard to read. The post-generation stage therefore, converts the sentence into more readable English by applying the following rules:

a. Introduction of expletives, 'it' and 'there'.
b. Reduction of conjuncted noun phrases **
c. Determination of number
d. Determination of person
e. Adjustment of comparative, tense, etc.
f. Pre-transformation of prepositional phrases
g. Determination of articles
h. Placement of a relative clause after the verb
i. Generation of adverbs to modify noun phrases
j. Deletion of dummy nodes
k. Reduction of co-ordinate conjunctions of verbs

4.3. Sample of Generation Processing

(1) *Transfer of case slots, such as instruments and causes, to the subject position.* To produce more readable English expressions, a phrase of the tool or cause case should be moved to the subject position. The result is a sentence with a non-living thing.

▶ 電気計測法, データ処理, 自動化機器の進歩で自動化船が増加した。

Advances in electrical measurement, dataprocessing and automatic equipment increased the number of automated ships.

▶ LSIの発達でプリント配線基板1枚当りの集積度が増大し, 確実な試験法が重要になった。

Advances of LSI increase the integration of printed circuit boards per one piece, and reliable testing methods become important.

▶ システム工学, 情報工学, ソフトウエア工学など新しい技術の急速な成長により, IEEおよび電気工学教育の全パターンでの重点が移動した。

Rapid growth of new technology such as systems engineering, information engineering and

** Direct translation of conjuncted Japanese NPs often leads to redundant expressions such as 'development of A and development of B'. Therefore, such expressions are reduced.

software engineering moved the emphasis of IEE and electrical engineering education in all pattern.

(2) *Placement of relative clause after the verb.* A relative clause qualifying the subject tends to make the subject phrase very long, thereby causing the sentence to be hard to read. To prevent this, the relative clause should be placed after the verb.

▶ スクリーンが集束能力をもつパラメタ範囲を決定。

The parameter range is determined in which screens have the converging capability.

▶ 長い直線状導体に, 方形の電流パルスを流した場合に生ずる表皮効果を研究。

The skin effects are studied which are produced in long linear conductors when rectangular current pulses were passed.

▶ ゆらぎのある散逸性媒質内での反射波を一方向伝搬する波の Green 関数で表わす関係式を得た。

The relational expressions were obtained which were shown by Green functions of the waves propagating reflected waves in scattering mediums with fluctuations in one direction.

(3) *Reduction of conjunct noun and verb phrases.* This process removes duplicate modifying phrases in conjunct noun phrases. It also eliminates or pronomializes duplicate case elements in compound sentences.

▶ 等価電流は一様成分と非一様成分からなる。

Equivalent currents consist of uniform and nonuniform components.

▶ 沿面線上ストリーマの伸展過程を駒撮り写真から検討し, 従来の沿面放電の進展過程と比較。

Elongation processes of the streamer on surface wire are studied according to high-speed motion pictures, and compared with conventional development processes of the surface discharge.

▶ プリント回路に対する部品の自動取付機は軸端子形から始まり, 平行端子形, トランジスタ, IC, チップの順に発達した。

Automatic mounting machines of parts to printed circuits start from the axial terminal type,

and they are developed in the order of the parallel terminal type, transistors, IC and chips.

(4) *Nominalization of the main verb and substitution of verb.* When a Japanese sentence lacks both agent and object, even passive transformation cannot produce a grammatically correct English sentence. In this case, the main verb is nominalized as the subject, and a verb phrase like 'is made' or 'is performed' is added as the main predicate.

▶ ヘリウムで調べ次のことがわかった。

The examination is made by the helium, and following facts are found.

▶ 米国ボストンの New England 電話会社において，マイクロコンピュータをベースとしたエネルギー管理制御システムを設備し，エネルギー原価の低減に成功した事例を紹介。

The cases are introduced in which the success is made in the reduction of energy costs by equipping the energy management and control systems based on microcomputers in New England Telephone Company of the United States Boston.

▶ 標記コンデンサは信頼性，長寿命性及び限界温度耐性が優れているため工業用に多用されているが，より高価格なタンタル電解コンデンサと比べて容積が大きいという欠点を持っていた。

Subject capacitors are abundantly used for the industry because they are excellent in reliability, 長寿命性 and resistance to the limit temperature, and they have shortcomings, the volume is large in comparison with more expensive tantalum electrolytic capacitors.

(5) *Introduction of the preparatory subject 'it'.* When the subject phrase is too long, the preparatory subject 'it' is introduced to make the English sentence easier to read.

▶ 伝導率の緩和に起因する容量性の渦電流は，円すい状試料について測定した標記のフェライトの複素磁気分散曲線の初期転回の特性を決定づける本質的な要因であろうと結論した。

It is concluded that capacitive eddy currents will be essential factors deciding characteristics of the initial rotation of the complex magnetic dispersion curve of above-mentioned ferrite measured in the relaxation of the conductivity about conical test pieces.

▶ セラミックチップキャリアの実装方法が今後のLSI高密度実装のキーテクノロジーとなる傾向にあることを示した。

It is shown that packaging methods of ceramic chip carriers tend to become the key technology of LSI high-density packagings.

▶ プリント配線板生産もシステマチックな装置産業に変ぼうすることを指摘。

It is pointed out that printed circuit board productions also change to systematic equipment industries.

(6) *Determination of the adverb position.* The position of adverbs in English sentences should be determined depending on their syntactic subcategorization.

▶ 同様にして，円筒の内側と外側の空気の領域における磁束密度分布も決定できる。

In the same way, the magnetic flux density distribution of inside and the outside of cylinders in regions of the air can also be determined.

▶ まず熱力学第二法則とその物理的解釈を述べ，次に Schutz の理論を解説する。

First, second law of thermodynamics and the physical interpretation are described, and Schutz's theory is explained next.

▶ 複素伝導率の緩和を Grant の表式で採り入れて，複素透磁率の初期部分を理論的に再現した。

The relaxation of the complex conductivity is incorporated by the formula by Grant, and initial parts of the complex permeability are theoretically reproduced.

▶ 音波周波数の交流電流放電が収縮状態へ転移する条件を実験的に研究した。

The conditions in which the AC current discharge of sound wave frequencies transfers to the contraction state are studied experimentally.

▶ この磁界構造においては，後進波も伝搬できる。

Backward waves can also be propagated in this magnetic field construction.

(7) *Selection of negation expression.* English differs from Japanese in that the negative word 'no' can be used to negate noun phrases. This technique is used when applicable.

▶ 集中定数と分布定数の混合回路網では，正実条件が必要十分とはならない。

No positive real conditions become necessary and sufficient in mixed networks of lumped and distributed constants.

▶ 帯電していない双極子の運動の安定性について。

The stability of the motion of the dipoles which are not charged.

5. Evaluation of Machine Translation Systems

The Mu project is evaluating the translated results using ease of understanding and fidelity. For ease of understanding, native speakers of the target language read the translated texts and rate the degree of their understanding using a five-level grade. For fidelity, checkers familiar with both source and target languages verify whether the content of the input sentence is transferred to the output, and evaluate each sentence using a seven-level grade.

Sentences evaluated to date total 1983. Sentences 1 to 683 (Group A) were used for the development of the system; the other 1,300 sentences (Group B) have been subjected to new experiment only once. All the sentences were taken in the order of their occurrence from abstracts of scientific and technical papers in the electrical en-

Table 1
Evaluation Results of Group A sentences

Ease of understanding	Fidelity									Total	Percentage
	0	1	2	3	4	5	6	7	Imperfect		
1	229	0	17	8	11	1	1	0	1	268	39.2
2	1	167	31	12	6	5	0	0	0	222	32.5
3	0	0	71	35	27	4	4	0	2	143	20.9
4	0	0	6	15	22	3	2	0	1	49	7.2
5	0	0	0	0	0	0	1	0	0	1	0.1
Total	230	167	125	70	66	13	8	0	4	683	
Percentage	33.7	24.5	18.3	10.2	9.7	1.9	1.2		0.6		

Table 2
Evaluation Results of Group B Sentences

Ease of understanding	Fidelity									Total	Persentage
	0	1	2	3	4	5	6	7	Imperfect		
1	256	1	34	13	6	6	2	0	0	318	24.5
2	1	274	36	31	26	11	1	0	1	381	29.3
3	0	0	157	87	79	26	29	0	6	384	29.5
4	0	0	19	39	60	16	24	1	3	162	12.5
5	0	0	2	3	18	2	10	0	0	35	2.7
?				1	1				18	20	1.5
Total	257	275	248	174	190	61	66	1	28	1,300	
Percentage	19.8	21.2	19.1	13.4	14.6	4.7	5.1	0.1	2.2		

gineering field published by JICST. No pre-editing was performed. Algebraic expressions and chemical reaction equations were eliminated, because subgrammars to handle such expressions had not been completed when the experiment started. Every sentence was collected from abstracts in order of occurrence.

Tables 1 and 2 summarize the evaluation results for Group A and B, respectively. The overall evaluation results are less favorable for Group B, because these sentences were used for the first time in these experiments and therefore contain words unknown to the system. These unrecognized words caused errors in morphological analysis, inadequate lexical information led to additional errors.

As shown by these tables, ease of understanding decreased with fidelity. While few sentences had higher fidelity and lower ease of understanding, those with lower fidelity but higher ease of understanding accounted for a fairly large number.

6. Conclusion

Considerably sophisticated machine translation systems are still unable to handle all sentences. This makes the idea of restricted language essential. To perform translations as smoothly and correctly as possible, some restriction rules must be imposed on input sentences or pre-editing must be performed on them. Particularly for abstracts, some restriction on style can facilitate sentence analysis and help produce better translations.

Work on machine translation systems not only helps implement practical systems, but also technological innovations resulting from such research also contribute to progresses in a number of other research fields. From this standpoint, it seems well worth the effort to complete the project.

References

[1] Jun-ichi Nakamura, Jun-ichi Tsujii, Makoto Nagao: Grammar Writing System (GRADE) of Mu-Machine Translation Project and its characteristics, Journal of Information Processing, vol. 8, ao. 2, pp. 93-100 (1985)

[2] Jun-ichi Tsujii, Jun-ichi Nakamura, Makoto Nagao: Analysis Grammar of Japanese in Mu-Project, Proceedings of COLING 84 (1984)

[3] Makoto Nagao: Toyoaki Nishida, Jun-ichi Tsujii: Dealing with the Incompleteness of Linguistic Knowledge in Language Translation, Proceedings of COLING 84, 1984

[4] Makoto Nagao, Jun-ichi Tsujii, Jun-ichi Nakamura: The Japanese Government Project for Machine Translation, Computational Linguistics, vol. 11, no. 2-3, 1985