科学技術論文表題の英和機械翻訳システム†

長尾 真†† 辻井潤一††
矢田光治††† 柿元俊博††††

科学技術論文の表題文の構造は比較的簡単で，その種類もそれほど多くないと推定される．これを約一万文献の表題文について調べたところ，約1,000程度のパターンとなった．そこでこれをさらにいくつかの処理によって縮約して18の基本パターンを決定した．そしてそれらすべてのパターンに対して，対応する日本語の訳語の語順を与えた．これは，それ以上パターンの詳しい構造解析をしなくても，18パターンすべてを調べるという簡単な方法で，文構造の解析と合成を行ったのと同等な効果がえられ，英語から日本語への翻訳が効率よく行われるからである．しかし構造のあいまいとなる場合もいくつかあり，動詞の—ing形が形容詞的に働くのか，目的語をとる動詞として働くのかに関する判断には意味による決定を導入しなければならなかった．

このシステムは，開発後データベースシステムと結合され，INSPECなどの論文表題を日本語に翻訳し，これをデータベース化して検索するシステムにおいて試用されている．翻訳の精度は平均的には80%程度であるが，INSPECという特定のデータベースの表題文では93%程度の翻訳率が得られた．

1. はじめに

機械翻訳システムを作る場合の中心的課題は，与えられた文の構文解析と意味解析である．構文解析の手法としては，これまでに非常に多くの方法が提案されてきた．しかし構文解析のための個々の書き換え規則としてどのようなものを取るのがよいかについては，あまり研究されてきていない．具体的な文法法則，あるいは書き換え規則は対象とするテキストの性質によると考えられ，モントリオール大学のTAUM METEOやTAUM AVIATIONなどでは，“サブグラマ”という考え方を提案している．これはある言語全体をおおう文法ではないが，対象とするテキストを十分解析できる文法，またそのテキストの“くせ”に合わせた文法という概念である．

科学技術論文の表題を英語から日本語に翻訳する場合にも，同様の考え方をとることができる．すなわち論文表題にはある種の型（パターン）があり，多くの場合名詞句となっていて，あまり多くの異なった動詞は現れない，使われる単語は限られた分野の専門用語がほとんどであり，また文の長さもそれほど長くないといった特徴である．そこで“形容詞＋名詞”や“名詞＋名詞”といった局所的にまとめられることが明らかである部分をまとめて1つの名詞と考えれば，英語の論文表題の文の構造は，名詞と動詞，前置詞の組み合せで表現できる．そしてこの意味での論文表題のパターンが同じになるものが多く，異なったパターンの数はそれほど多くないと考えられる．このように異なったパターンの数が少ない場合には，それ以上くわしいめんどうな構文解析をしなくても，英語表題文のそれぞれのパターンに日本語のパターンを対応させることが可能となる．

このような考え方にたって，処理内容ができるだけ簡単で，処理速度が早く，実用になることを目ざした翻訳システムを作った．当初対象としたのは，日本科学技術情報センターの文献速報，電気工学編の中の英語論文表題1万件である．このように限定された分野の文のパターンの種類はそれほど多くなく，構文解析をくわしくしなくても，表形式で英語のパターンに対して日本語のパターンを対応させることによって翻訳ができるのではないか，ということを実験的に確かめてみることも目的であったので，英語の形態素解析の部分は省略し，単語の変化形もすべて辞書に入れ記憶した．辞書の語数は約9,400語となった．

表題文の解析の第一段階で一万件から約1,000の異なるパターンを抽出したが，さらにいくつかの縮約操作をへてその数をへらし，最終的に18の基本パターン

---

† An English-Japanese Machine Translation System of the Titles of Scientific and Technical Papers by MAKOTO NAGAO, JUN-ICHI TSUJII (Faculty of Engineering, Kyoto University), KOJI YADA (Software Division, Electrotechnical Laboratory) and TOSHIHIRO KAKIMOTO (Scientific Computer Systems Development Dept., Fujitsu Co.).

†† 京都大学工学部
††† 電子技術総合研究所ソフトウェア部
†††† 富士通(株)科学システム開発部

をとり出した．これらの 18 の基本パターンに日本語訳の基本パターン（訳文の語順）を与え，日本語の合成を行った．英語の構文解析において，あいまいさの生じる場合がいくつかある．並列句の処理はその1つであるが，最も簡単な並列の場合だけを取り扱った．

動詞の —ing 形が形容詞として働くのか，目的語をとる動詞として働くのかに関する判断の部分については，意味パラメータを導入して行わねばならなかった．

この翻訳システムは，1978 年から 1979 年にかけて京都大学で作成され，種々のテストが行われた．その後これは 1980 年に工業技術院筑波研究センターの計算機システム RIPS の上に移植され，システムの周辺部分が整備されるとともに，データベースシステムと結合して数カ月間の試用が行われて現在に至っている．

このシステムの翻訳の能力は，対象とするデータによって異なるが，平均的には約 80% である．INSPEC（イギリスの電気学会が作っている物理，電気，電子，計算機，制御関係の文献サービス・データベース）のデータに対してはうまく働き，約 93% の翻訳率が得られた．

2. 機械翻訳システム構成の基本的考え方

論文表題を機械翻訳システムの対象としてみると，次のような特徴を持っている．

（1） 表題文中の単語は専門用語がほとんどで，多義語の問題をさけることができる．

専門用語であっても専門分野によって別の訳語がつけられている場合もあるが，各訳語ごとに分野コードを指定する等，簡単な方法で多義語の問題をさけることができる．ただし，訳語と分野コードの指定は当初のシステムでは行っていない．

（2） 特殊単語列の変換

文中には特殊な単語列が現れ，普通の書き換え規則でまとめるのが困難なものがある．そのような単語列をイディオムとして，1つの単語相当として扱うために，イディオム中のある単語の辞書項目の内容の欄にそのイディオムを登録し，それを認識でき，その後の処理に支障のないようにした．time varying（時間とともに変化する）など，各分野でよく使われる固有の表現（イディオム的な表現）のほかに，英語一般にみられるイディオム的表現(based on, prependicular to 等）もまた辞書の中に記述され，辞書引きの結果1つの単語として取り扱われる．その例を**表1**に示す．

表 1 イディオムの特殊例

Table 1 Examples of special idioms.

	品詞列の置換	例
(1)	adj+prep→prep	consistent with, sensitive to, important to
(2)	—ing+adv→—ing	narrowing down, based
	—ed+adv→—ed	merely on
(3)	—ing→prep	surrounding, affecting
(4)	from+NUM+to+NUM→NUM	from 170 to 138
(5)	the+NUM+th→NUM	the 29th

（3） 専門用語は必ずしも1語単位ではなく，多数の語で複合語をなすことが多い．

たとえば，Large Scale Integrated Circuit（大規模集積回路）のように，専門用語では複数個の単語で複合語を構成し，1つの概念を表わすことが多い．この場合，訳出の過程で個々の単語の訳語から全体の訳語を合成していたのでは，ほとんど理解不可能な訳文が生じる．そこでこれらの単語列を1つの単語とみなして辞書に登録した．この場合もイディオムの登録と同様で，ある単語（主として最後の単語）の辞書項目の内容の部分に複合語を記憶している．

一般に，複合語の単語数を多くとればとるほど，辞書の規模は大きくなるが，逆に後での処理は簡単になり，訳文の質も向上する．特にある特定の専門分野に対象を固定した場合には，この手法は有効である．この複合語の専門用語を選定する過程で，学会発行の学術用語集等を参考にしたが，用語数が少なく，ほとんど役に立たなかった．現実的な機械翻訳システム作成のためには，現在使われている専門用語を網羅的に収集した terminology Data Bank 的なものを用意する必要がある．

（4） 表題文においては，長い形容詞と名詞の連続によって名詞句を構成する電文調の表現が多くあらわれる．

通常の文体では前置詞等を使って結合される名詞群が，表題文においては，単に隣接しておかれるだけで結合された名詞句を構成することが多い（たとえば High Resolution Bragg Reflection Method, Pulse Compression Distance Measuring Equipment System など）．これらの名詞間，形容詞—名詞間の結合関係を識別し，訳文合成の際に適切な語順変換と助詞の挿入を行うことはもちろん理想的ではあるが，これを行うためには，かなり深い意味的な解析を必要とし，現時点ではほとんど不可能である．

一方，日本語においても，表題文においてはかなり長い漢字連続で英語と同様な表現をとることが可能であ

る（上記の例の場合には,「高分解ブラッグ反射手法」,「パルス圧縮距離測定装置システム」となる）．このような場合，日本語の語順はほとんど英語の語順と同じである．したがって，現在の我々のシステムにおいては，このような長い形容詞と名詞の連続内での単語間の詳細な係り受けの関係は解析せず，単にそのような長い名詞句があったということだけを簡単な遷移網文法 (Transition Network Grammar, 以下 TN と略す) で認識することにした．現在のシステムでは日本語と英語の語順が変化するのは，前置詞，—ed 形，—ing 形等の単語があらわれる場合だけで，形容詞を含む名詞連続では語順がかわらないと考えている．したがって，ここで使われる TN はこのような語順の切れ目を示す語があらわれるまでを１つの名詞句としてまとめてゆく．ただし，$\begin{Bmatrix}\text{冠詞}\\ \text{形容詞}\end{Bmatrix}+\begin{Bmatrix}\text{—ed 形}\\ \text{—ing 形}\end{Bmatrix}$ のように，明らかに形容詞的に使われることがわかる—ing 形，—ed 形の場合には，語順を変化する必要がないので，TN によって処理し，名詞にかかるものとしてまとめる．

（5） 表題文にあらわれる文構造は，書き換え規則で処理する必要があるほど多様ではない．

言語解析において，書き換え規則に代表される規則の体系を使用する大きな理由は，言語表現の持つ「無限」の可能性を，有限個の規則群で処理したいためである．しかしながら，表題文においては，表題文全体の長さが制限されていること，埋込み文等の複雑な文体があまりあらわれないこと等の理由によって，少なくとも品詞並びに還元してみる限り，規則で処理しなければならないほどの無限の可能性があるとは考えられない．しかも，上記の(3)の項で述べたように，複数個の単語群で表現される専門用語を１つの辞書項目と考え，また，(4)で述べたように長い名詞連続を TN によって１つの名詞に縮約する操作を行うと，可能な品詞の並びはさらに局限される．

そこで我々は，表題文の品詞列をこのような方法で縮退させれば，その結果残ったものは名詞と動詞，前置詞の組み合さった構造で，その種類は有限となるだろうと推定した．このような表題文の縮退した品詞列のことを文型パターンと呼んでいる．そして，この文型パターンの種類が有限であれば，これを書き換え規則などでそれ以上解析しなくても，それぞれの文型パターンに対して目的言語（日本語）の同様な文型パターンをそれぞれ対応させればよい．

翻訳すべき表題文が与えられると，まず，イディオムの検出を行い，その部分を１語とみる．次に上記の縮約を行い，それ以上縮約できない品詞列をうる．これを表題文の骨格パターンと呼んでいる．そこで，骨格パターンと一致する文型パターンを文型パターン辞書中に発見し，その文型パターンに対応する日本語の文型パターンを取り出し，そのパターンに応じて訳語を与えて訳出を行うというステップをとる．このように，比較的長い品詞列の文型パターンをとり，それらが訳文においてどのような語順をとるかを決める方が，長さの短い書き換え規則を何回も使用して深いトリーの形に解析し，また合成するステップを実行するよりも，大局的な構造の決定をしやすく，翻訳結果もよいと考えた．

3. 翻訳システムの構成

表題文の翻訳は**図１**に示す各ステップによって行われる．以下に各ステップでの処理の概要を述べる．

3.1 辞書引きおよびイディオムの処理(ステップ1)

現在のシステムでは，英語の形態素処理を全く行わないために，単語の形態素変形もすべて辞書項目として含まれている．もちろん，実用化の際には規則化可

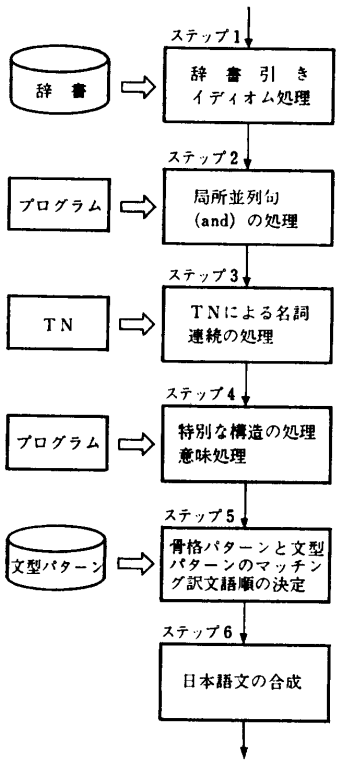

図 1 翻訳の流れ
Fig. 1 Flow of translation.

能な部分を規則化することによって，辞書項目数を減少させる必要がある．

ただし，同じ 'v+ing' の形式で作られた語であっても，accounting, bonding, scattering, engineering 等はほとんどいつも名詞的に，superconducting は形容詞的に，using はもっぱら目的語をとってその直前の名詞を修飾して前置詞的に，determining は他動詞的に使われる，などの単語による個別的特徴がある．一般に，'v+ing' 形はこれらのすべての使われ方をする可能性があるが，特定の語については，分野を固定するとほとんどある一定の使われ方をする．したがって，辞書記述の中にこの固定的な使われ方をするということを記述しておくことにより，後の段階での複雑な処理をさけることができる．また time varying, based on のように，イディオムに組み込まれているものもある．

3.2 局所的な並列句の処理（ステップ2）

ここでは and だけでなく or や but のような等位接続詞によって作られる並列句一般の処理を行う．実際には，対象とした表題文中には，等位接続詞としては and だけしか現われなかった（1万表題中，and を含む表題722件，内2つ以上の and を含むもの 51件）．

等位接続詞を含む句の解釈には，よく知られているように範囲に関するあいまいさがある．たとえば，Controllability and Stability of the system を「システムの可制御性と安定性」と訳すべきか，「可制御性とシステムの安定性」と訳すべきかは，結局 Controllability, Stability, System の3語間の意味的関係によって決定せざるを得ず，意味処理の問題と密接に関連する．これを「システムの安定性と可制御性」のように，並列されている句の順序を入れ替えて，日本語においても同じ曖昧さを持つ表現に置き換えることもできるが，本質的な解決にはならない．

我々のシステムでは，文全体の大域的な構造を参照せず，局所的な品詞情報から決定できる場合だけを扱った．and の含まれた構文を調べてみると，多くは次のような形をしている．

（1） adj+and+adj, v+and+v, —ing+and+—ing, adv+and+adv, n+and+n

（2） prep+n+and+prep+n

（3） n+prep+n+and+n+prep+n

これらのうち(1)は安全にまとまるものとして優先的にまとめるが，n+and+n でも（3）のような場合がありうるので注意が必要となる．（2）は and の直後に前置詞が存在することで比較的容易に発見できる．(3)の形式はより広い範囲の語順と語群の意味的関係を調べねば決定できない．この形はあまり多く現われなかったので取り扱っていない．文と文が並列する場合も現在は扱えない．

3.3 名詞連続の処理（ステップ3）

ここまでのステップでつくられた品詞列は，イディオムと簡単な構造の並列句を処理しただけで，入力表題文の品詞列とほぼ同じものである．したがって，このままでは入力表題文の数だけの異なった品詞列があり得る．このステップ3では，語順の変換を必要としない 'n+n' やそれに形容詞が挿入されたような品詞列を，1つの名詞句に縮約する．これを行うための TN を**図2**に示す．

このステップによって，次のような句が1つの単語に縮約される．

（例） This film waveguides→waveguides
High quality phosphor screens→screens
An automated general purpose test system →system

3.4 その他の特別な構造の処理（ステップ4）

これまでの処理で，かなりの表題文が同じ構造となるが，まだある部分構造で縮約した方がよいものがあり，また特別な構造でこの段階で処理すべきものがある．

（1） n_1+of+n_2

"n_1 of n_2" はよく現れる形であり，ほとんどの場合，"n_2 の n_1" という形に翻訳されるものである．そこで "n_1 of n_2→n_1" とし，訳語の順は "n_2 の n_1" とすることにした．

（2） 文頭の prep+n

"on pattern recognition" というように，前置詞で始まる表題がある．この場合には "prep+n→n" として，全体を n として扱った．それ以後に修飾句がある場合もこれで外見上は矛盾なく訳せることがわかった．日本語訳としては "n+prep" の順とする．

（3） —ed+prep, —ed+adv+prep

動詞の過去分詞が形容詞として働くときは，すでにステップ3で処理されている．残るものは "—ed+prep" の形である．これは一括して "—ed+prep→prep" として置きかえることにした．日本語に訳すときは，"prep+—ed+される"，あるいは "prep+adv+—ed+される" とする．

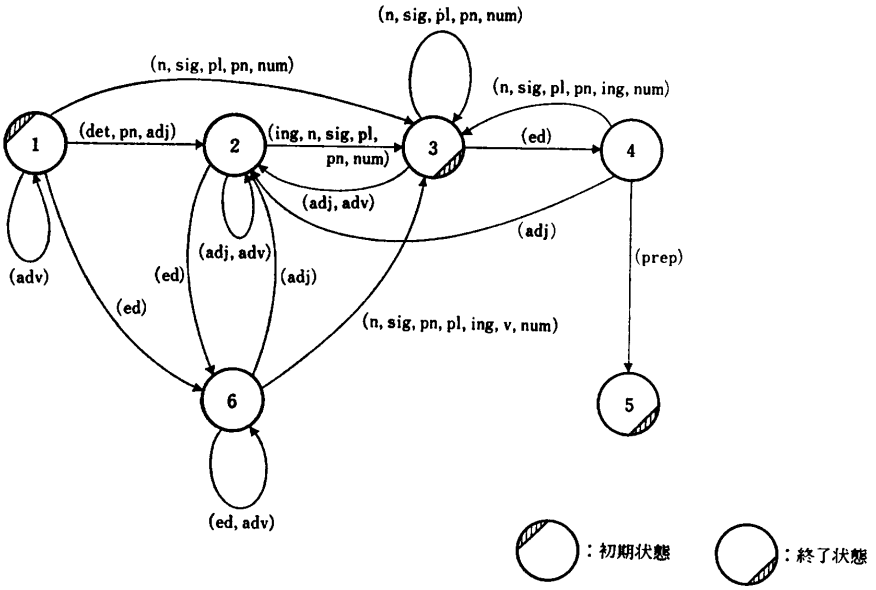

図 2 遷移網文法 (TN)
Fig. 2 Transition network (TN).

(4) 文頭の —ing+prep

"concerning……" のような場合であり，—ing は名詞として扱い，"—ing+prep=n+prep" に置換してその後の処理にわたす．

(5) n+—ing+prep

この場合の —ing の機能を決定するのはむずかしいが，表題文の場合には，—ing を名詞とする場合の方が多く，"n+—ing+prep=n+n+prep" と置換した．

3.5 意味処理

1つの文型パターンに対して複数個の訳出順序が存在する場合が生じる．特に —ing 形の存在する場合にそれが生じる．たとえば，measuring device や measuring temperature のように動詞と名詞の意味的関係によって，—ing が形容詞的に働いたり，動詞として目的語を取る形となったりする．この区別を知るためには，単語相互間の意味的関係をしらべることがどうしても必要となる．我々のシステムでは，処理をできるだけ簡単にするために，主語と述語の関係をしらべる SUBJ という意味チェック関数と，述語と目的語（補語）の関係をしらべる OBJ という意味チェック関数の2つだけを用意した．SUB (—ing, n) は動詞 (—ing) の格スロットの主語部分に書かれている意味カテゴリの1つが名詞 (n) の意味カテゴリの1つと一致するかどうかをチェックする関数である．OBJ (—ing, n) は動詞の格スロットの目的語部分の意味カテゴリと名詞の意味カテゴリの一致を調べるものである．

骨格パターン中に —ing+n という部分があると，これら2つの意味チェック関数を働かせ，その構造をきめる．SUBJ が成立するときには訳出語順を変更せず "—ing+n→n" とし，OBJ が成立する場合には訳出語順を入れかえるが，全体はやはりnとして先へ進む．—ing を含むパターンの処理については次節に詳しく述べる．

このような意味のチェックを行うために，意味カテゴリとして，観点，道具，理論，単位，物体，抽象の6つをとり，各名詞にこれらの意味カテゴリを付与した．1つの名詞が2つの意味カテゴリをとることもある．その例を**表2**に示す．そして個々の動詞が主語・目的語にどのような意味カテゴリの名詞をとるかを記述した．たとえば measure は主語として理論や道具をとり，目的語として物体や観点をとるといったことを各動詞に対して記述しておく．このような意味分類は非常に困難で，恣意的となるのはやむをえないと思われるが，対象分野を限定することによってある程度可能となる．

3.6 —ing を含む句の処理

—ing には動名詞の場合や名詞化したものもある．名詞化したものは辞書に名詞として登録してあるが，その他の —ing についてはその現われ方によって次のような処理を行った．

(1) n+—ing. ing の後に動詞が来るか，文の終

表 2 名詞のカテゴリ分類の1部分
Table 2 A Part of the categorization of nouns.

道具	probe, instrument, equipment system, set, machine unit
理論	technique, principle, method approach
観点	velocity, displacement acceleration, position, energy inductance, registance capacitance, conductance temperature mobility, permittivity intensity level parameter coefficient, time, angle
物体	laser, beam, light solenoid, coil film, tape, disc metal MOS, device, waveguide diode, transistor oil, water, liquid, solid air, earth star, stone

っている場合には動名詞とする．ing の次に prep が来る場合には意味チェック関数によって構文を決める．それで決まらないときは，ing を動名詞とし，n が ing を修飾するという解釈をとっている．これはどちらの用法が多いかを調べた結果によっている．

(2) —ing+n. n が冠詞で始まる場合には n は目的語となっている．そうでないときは，意味チェック関数によって決める．それで決まらないときは，用法の多い現在分詞の形の方をとっている．

(3) n_1+ing+n_2. ing が n_2 を目的語としてとるときには，ing は現在分詞として n_1 を修飾するが，ing が n_1 を目的語にとり n_2 を修飾する場合もある．したがって，意味チェック関数で調べるが，それでも決まらない場合には，用法の多い ing が n_1 を目的語にとる方を優先した．

(4) prep+ing+prep. この場合の ing は動名詞と解釈する．

以上のような訳出語順の決定は，骨格パターン中の係り受け関係の決定によって行われていることになるが，この過程で問題になるのは，前置詞句の修飾先である．英語では，前置詞句はそれが修飾する語よりも後に置かれ，日本語では全くその逆になるので，英語の語順を完全に逆転させておけば，英語において前置詞句の修飾先が曖昧である場合に，日本語においても全く同様な曖昧な表現が得られる．したがって意味処理をくわしく行わない本システムではこれでがまんすることにした．

3.7 骨格パターンと文型パターンとのマッチング（ステップ5）

以上のステップを経て，表題文の骨格パターンが抽出される．我々の仮定は，この骨格パターンの種類は有限であり，それらをすべて文型パターンとして列挙すれば，構文解析をせずに目的言語の文型（語順）を生成することができるということである．日本科学技術情報センターの文献速報の英文表題一万件をこのように処理して文型パターンを取り出したら，約 1,000 の異なったパターンがあった．これはステップ4の縮約操作の前の数であるので，ステップ4を経ることによって文型パターンは大幅にへった．これらの文型パターンのうちのほとんどは，ただ一回しか現われないものであり，多くの表題文は比較的小数の型になっている．そこで最終的に**表3**に示すような 18 個の文型パターンを取ることにした（最初は 13 パターンであったが，後述の INSPEC などを扱う時に5個ふやした）．そして，それぞれの文型パターンに対して訳文の語順情報を与えた．

これらの文型パターンは，対象とする論文の分野，あるいは論文の性格によって，かなり異なるようにみえ

表 3 文型パターンと INSPEC における使用頻度
Table 3 Sentential patterns and the frequency of their usage in INSPEC translation.

英語文型パターン	日本語語順パターン	INSPEC における使用頻度
(1) ing・n	ing・n	0
(2) n	n	466
(3) n・ing	n・ing	0
(4) n_1・prep・n_2	n_2・prep・n_1	536
(5) n_1・prep・n_2・ing	n_2・ing・prep・n_1	0
(6) n_1・$prep_1$・n_2・ing・$prep_2$・n_3	n_3・$prep_2$・n_2・ing・$prep_1$・n_1	0
(7) n_1・$prep_1$・n_2・$prep_2$・n_3	n_3・$prep_2$・n_2・$prep_1$・n_1	147
(8) n_1・$prep_1$・n_2・$prep_2$・n_3・$prep_3$・n_4	n_4・$prep_3$・n_3・$prep_2$・n_2・$prep_1$・n_1	32
(9) n_1・$perp_1$・n_2・$prep_2$・n_3・$prep_3$・n_4・$prep_4$・n_5	n_5・$prep_4$・n_4・$prep_3$・n_3・$prep_2$・n_2・$prep_1$・n_1	2
(10) n_1・$prep_1$・n_2・$prep_2$・n_3・$prep_3$・n_4・$prep_4$・n_5・$prep_5$・n_6	n_6・$prep_5$・n_5・$prep_4$・n_4・$prep_3$・n_3・$prep_2$・n_2・$prep_1$・n_1	0
(11) n_1・prep・n_2・v・n_3	n_2・prep・n_1・は・n_3・を・v	1
(12) n・v・adj	n・は・adj・v	0
(13) n_1・v・n_2	n_1・は・n_2・を・v	1
(14) n_1・v・n_2・prep・n_3	n_1・は・n_3・prep・n_2・v	1
(15) n_1・v・prep・n_2	n_1・は・n_2・prep・v	0
(16) v・n	n・v	2
(17) v・n_1・n_2	n_1・は・n_2・v・か	1
(18) v・n_1・prep・n_2	n_2・prep・n_1・を・v	1

表 4 前置詞の訳
Table 4 Translation of prepositions.

of	の	to	への (n•to•n)
by	による	on	についての
with	による	in	での
at	における	about	について
for	のための		

る．表3の文型パターンのうち最初の13パターンは日本科学技術情報センターの文献速報・電気工学編一万文献から決めたものであるが，表3のパターンがどのように使われているかを INSPEC の 1,000 文献でテストした結果は表3の右端の欄の数字となった(1,000をこえるのは1つの表題が2つの文からなるものがあったため)．実際に使われた文型パターンは11個で，かなり使われているものは，それらのうちでただ4個であるというおもしろい結果をえた．

3.8 日本語文の合成（ステップ6）

以上のようにして日本語文の語順がきまると，文の生成は簡単である．名詞は辞書から日本語の訳語を取り出し，主語となるとき“は”を，目的語となるとき“を”をつける．前置詞に対しては**表4**の訳語が対応させてあり，そのまま出す．定冠詞，不定冠詞は訳出しない．動詞はほとんどの場合，終止形か，次の名詞にかかる連体形かであるが，幸いなことに終止形と連体形は同じ形なので区別せず，活用のことを考えずに出すことができる．このようにして出力した例を**表5**に示す．前置詞の訳語を固定していて読みにくい場合があるが，これを場合に応じて最適な形の訳にすることは非常にむつかしい．未定義語は名詞とみなして，そのままの位置で，そのままの形で出力する．

4. 外国文献データベースとの結合

このシステムの開発は1978年から1979年にかけて行われ，種々のテストが行われた後，1980年に工業技術院筑波研究センターの計算機システム RIPS に移植された．その際実用化のために次のような機能の追加が行われた．

（1） 翻訳対象の分野に依存しない部分と依存する部分をできるだけ分離する．前者は汎用部分としてプログラムに入れ，後者は辞書データセットに格納する．現在，単語・イディオムなど本来の辞書データと，文型パターンが辞書データセットに入っており，その他はプログラムに入っている．

（2） 英文中の特殊記号は，構文上重要な役割を担うものがあるので，なるべく英文に忠実な日本文になるように，その処理を行う．

（3） 英和翻訳プログラムそのものへの機能追加に関しては，翻訳実験により汎用化できるもののみ考慮する．すなわち，入力文の骨格パターンが文型パターンにマッチングせず，骨格パターン中に，たとえば n_1+n_2, adj, $prep_1+prep_2$ という品詞列の部分パターンがある場合に，n_2, adj, $prep_2$ に対応する単語が他の品詞をも持っている場合（辞書引きでわかっている），その品詞に変更して，イディオム処理から文型パターンマッチングまでのステップを再度実行してみる．また，辞書に未登録な語があった場合，規則的な語尾変化の処理を行い，再度辞書引きする処理を行う．

このシステムは，市販の英語の文献データベース中の表題や，研究者が個別に収集した英論文の表題を日本語に翻訳し，他の文献データ（たとえば，著者，抄録など）とともに情報検索システム(FAIRS-I)を使っ

表 5 翻 訳 例
Table 5 Examples of English-Japanese translation.

SEQUENCE CONTROLLERS WITH STANDARD HARDWARE AND CUSTOM FIRMWARE
標準ハードウエアとカスタムファームウェアによるシーケンス制御装置

AN ARCHITECTURAL COMPARISON OF CONTEMPORARY 16-BIT MICROPROCESSORS
現代16ビットマイクロプロセッサのアーキテクチャ上の比較

INK JET PRINTING OF JAPANESE KANJI CHARACTERS
日本語漢字文字のインクジェットプリンティング

LEVEL-INDEPENDENT NOTATION FOR MICROCOMPUTER PROGRAMS
マイクロコンピュータプログラムのためのレベル独立記法

A VLSI ARCHITECTURE FOR SOFTWARE STRUCTURE : THE INTEL 8086
ソフトウエア構造のためのVLSIアーキテクチャ：インテル社8086

A PROPOSED STANDARD FOR EXTENDING HIGH-LEVEL LANGUAGES FOR MICROPROCESSORS
マイクロプロセッサのためのハイレベル言語拡張のための提案標準

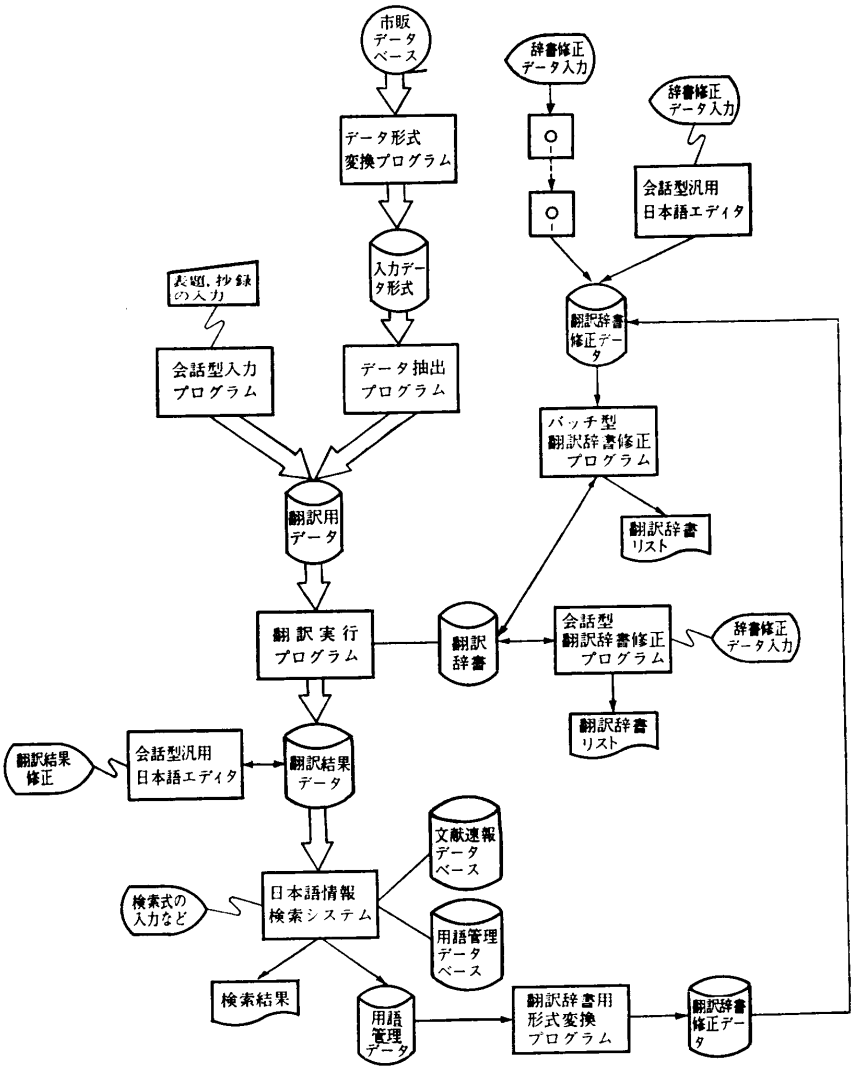

図 3　機械翻訳を組み込んだ文献検索システム

Fig. 3　Document retrieval system with machine translation facility.

てデータベースに格納し，文献速報サービスを行うものであり，**図 3** のような内容になっている．このシステムの特徴は，翻訳結果を左右する辞書のサポートツールをいろいろ備えていることである．これらには，大量の辞書データを処理するためのバッチ型翻訳辞書修正プログラムがある．さらに電総研では，研究者が翻訳を行う場合の用語統一のために利用する用語管理データベースが別途作られているので，そのデータを翻訳辞書に加えるためのプログラムを備えている．

現在このシステムは INSPEC や MICRO, COMPUTER GRAPHICS という雑誌の表題の翻訳に利用されている．

5. 翻訳実験と検討

上述の文献翻訳速報システムを利用して，英和翻訳プログラムの性能評価と辞書データの収集を目的として以下のような実験を行った．翻訳対象は INSPEC の中の物理・数学関係の論文 3,000 文献とし，1,000 文献ごとに翻訳し，辞書に単語，イディオムの登録や修正を行った．最初の 1,000 文献で翻訳結果の詳細な分析とプログラムの修正を行った．翻訳実験の開始前は，辞書登録語数 9,410，文型パターン数 13 であったが，1,000 文献翻訳の結果，追加を行い現在は辞書 11,272 語，文型パターン数 18 である．

翻訳実験の結果は**表 6** のようになった．最初の

表 6 翻訳実験結果

Table 6 Result of machine translation experiments.

文献 No.	計算機処理時間	未登録語数	処理不可能数	登録後処理可能数	処理不可能数
1～1000	100.16秒	816 語	49 文	38 文	11 文
1001～2000	107.54秒	567 語	39 文	23 文	16 文
2001～3000	115.4 秒	479 語	29 文	14 文	15 文

計算機処理時間：M 200 CPU 時間
処理不可能数：入力文の骨格パターンが文型パターンとマッチングしなかった数
登録後処理可能数：熟語，単語の登録により処理可能になった文の数

1,000 文献の詳細な分析の結果，816単語，79イディオムの登録とプログラムの若干の修正を行い，再度実験したら完全な誤訳と処理不可能な文が合わせて 11 文あった．そのおもなものは，and の範囲の処理をまちがったもの（4文），疑問文（4文）が扱えないなどである．3,000文献までの処理不能の数は全部で42文で1.4%である．翻訳できた文のうち約5%は修正を必要とし，残りはまず満足できるか，がまんできるものであった．

この翻訳実験の結果，翻訳の質の向上や翻訳不能数の減少に最も効果的なのは，イディオムの登録であることがわかった．

6. おわりに

現在，分野別の辞書を収集するために，いろいろな応用分野の検討を行っている．将来は翻訳プログラムのレベルアップにより抄録の翻訳や，英語以外の外国語と日本語の間の翻訳も考えている．

謝辞 本研究の主要部分のプログラムは，大学院学生建部周二君(現在東芝勤務)によって作られ，その後学部学生日樫英孝君（現在セイコーシステム勤務）によって改良された．ここに記して感謝の意を表する．また本研究の一部は文部省科学研究費補助金によった．

参考文献

1) 長尾，辻井，建部：技術論文表題の英和自動翻訳の試み，情報処理学会計算言語学研究会資料 19-2 (1979. 9. 21).
2) 長尾 真：計算機による日本語文章の解析に関する研究，昭和53年度文部省科学研究費特定研究（1）報告書（昭和54年2月）.

(昭和56年6月15日受付)
(昭和56年11月18日採録)