機械翻訳セミナ論文紹介*

7. 機械と言語学

E.D. Pendergraft: Hardware for Mechanical Translation: Relation between Hardware and Linguistics〔p. 20〕

MT のためには入力言語の分析，入力言語と出力言語の対応，出力言語の合成の理論を考える必要がある．それには一般理論と具体的言語の記述分析の2本建でゆき，こういった仕事が出来上がり，普通の電子計算機にかけて，うまくいったら始めて MT 専用の電子機械を作ればよい．

今日みられる MT の各派の異なった行き方は，相反するものではなく，相補い合うものであるから，それらの間の相互関係を考える必要がある．それから翻訳の問題は，やさしい場合から始め，だんだんむずかしい場合へと研究を進めてゆくのがよい．

構成要素による翻訳　いちばん単純なもので，これを実践するには，豊富な資料による研究が必要であり，また行なわれつつある．これには次の4通りがある．

語彙項目間の翻訳　初期の MT で考えられたことで，構文脈による意味決定，出力言語における語順変更などから，構文翻訳へと進んだ．

構文翻訳　これには入力言語の構文分析，入力言語と出力言語のそれぞれの構文記述の対応，出力言語の構文記述より出力言語への合成，この3段階がある．構文記述はメタ言語によるから，中間の対応過程はメタ言語どうしの交替ということになる．

意義翻訳　中間において2カ国語間の構文なり，語彙の記述が対応されうるのは，それらが意味のうえで同じか，似ているからである．つまり2カ国語間の意味記述どうしの交換を行なう方法である．

状況翻訳　意義翻訳より高次のもので，異なった表現が同一状況に使われ，また異なった状況が同一と考えられる事実を翻訳過程におり込もうというもので，現在ではたいへん複雑な仕事であろう．

変換翻訳　これまで述べた構成要素による翻訳に対し，同一の表現を示すものどうしの関係を記述したものを利用し，つまりA言語の言い変えた表現と，B言語の言い変えた表現によって翻訳する場合を指す．

さらに文体，主題，態度といった面もA言語について分析し，B言語で合成して表現することもできるかもしれない．

具体的作業　記述に先立ち仮説の構築が必須で，多くの研究者の協力が必要である．

使う機械は，言語のデータ集積，言語の認知，記憶といった過程がそもそも計算機的だから，汎用の大型計算機でよい．　（小笠原林樹）

8. プログラミング言語

Arnold C. Satterthwait: Hardware for Mechanical Translation: Problem-Oriented Programming Languages.〔p. 18〕

一般にプログラミング言語は，機械向き（m）言語，または問題向き（p）言語，または両者の組合せである．

p言語は，その計算機が備えている翻訳ルーチンにより，m言語に変換される．

m言語は，p言語に比べて非常に融通性があり，計算機の利用時間も少なくてすむという利点がある．

p言語は，習得するのに容易であるし，プログラムを書く時間も少なくてすむ．また，言語学，意味論，論理学のような分野の問題には，このような言語は重要である．p言語の中に，自動的に誤まりを発見する操作を備えれば，プログラムを入れて最後にチェックするまでの時間は，m言語よりずっと少なくてすむ．

また1つの m 言語に対し，p 言語はいくつかあるが，いちばん融通性のある言語をみつければよい．しかし，1つの言語に対し，あまり広い一般性を与えることは危険でもある．

ここでは，機械翻訳に必要な記号操作のために，特に開発された5つのp言語を検討する．ランドで開発中のものも含めると，これらの体系は，3つの異なった思想を表わしているようである．1つは，ランドの Hays が他の論文で述べているから，ここでは省略する．1つは，自然言語の構造の研究における強力な道具として，プログラミング言語の研究をする．最後

---

* 本誌 5-5 に引続いて機械翻訳セミナ（本誌 5-3, p. 178）の提出論文の全部を紹介する．日本側発表論文はすべて Preprints for Seminar on Mechanical Translation〔April 1964, U.S.-Japan Committee on Scientific Cooperation, p. 146〕からとった．米国側のものはすべて pre-publication copy の形をとっているので，通常の文献引用の形式を守っていない．

は，習得するのにやさしく，プログラマでない人にもわかるようなプログラミング言語を研究する．

COMIT：MIT で開発され，最も広く使われている．データは，プラス記号によって構成要素がつながれた形で，計算機の workspace に含まれる．プログラムは，一連のルールの集まりからできている．

SNOBOL：ベル電話研究所で機械翻訳，プログラムの翻訳などの研究のために計画され，記号列の操作に合うように作られている．データ構造は，文字と空白の列である．列の形成，パターンの比較，おきかえの演算ができる．

LRS：テキサス大学で開発された．9つの異なった領域にわけられ，これらは，中央のコントロールプログラムによって統合される．

Interpretive System：ウェイン州立大学において，ロシア語の構文分析の研究のために用意された．入出力翻訳ルーチンと解釈ルーチンからできている．

MIMIC：ランドで備えた．命令，条件，定義の3種類の陳述が許されている．挿入，削除，おきかえの演算が用意されているが，それ以外は定義によってつけ加えることができる．（込山敏子）

9. 入力資料の読取り

E.N. Adams: Input Reading for Language Processing System. 〔pp. 20～27〕

これは，IBM の研究所で開発している光学的文字読取り機についての報告である．どこの研究所でも大同小異であって，秘密に属することではあるが，みんな同じような方法で，同じような問題で苦しんでおり，その意味で，この報告は一般性を持っているといえる．確かに，この報告自身 IBM の機械のすばらしい点や，うまい方法を述べているわけでもなく，おそらく専門家には周知のことばかりであろう．しかしながら，この報告で文字読取り機が根の深い，困難なテーマであること，そして，その困難性のありかを具体的に知ることができる．

まず全体の構成は，document handling → scanning → data reduction logic → decision logic，それらを支配するシステム・コントロール，それに計算機が付随している．document handling には移動式と固定式の2方式が考えられる．前者は資料のほうが動き，後者はその反対である．前者だと，読み取りの始めと終りの資料の加速度はできるだけ大きくなければならず，1,000 文字/秒という要求には，かなり厳しい問題になる．いずれにせよ，資料の平面を精確に所定の場所に置くということが，きつく要求される．次に，像を電気的パルスに変換するのには，(1) flying spot scanner によるものと，(2) storage tube と image dissection を用いるものと考えられる．

(1) の場合，分解能と信号の間には，厳しい trade-off problem がある．分解能を増すには，多くの点を走査しなければならず，そうすると，一つ一つの点にはいる光の量は減少し，信号が弱まる．しかるに光源の輝度は，その大きさにあまりよらないので，分解能を増すことは非常に高くつく．(2) の場合，資料の全面積から同時刻に信号がはいるのは，非常な利点であるが，まだ，この分野で十分な tube が開発されていない．

次の問題は，パターン認識の中心問題で，煩雑な生の信号から本質的な信号を求めることである．不可欠な処理として，正規化の問題が取り上げられている．これには種々の方式があるが，著者は digital normalization strategy という，資料の種類により digital に正規化する基準を考えている．次に，困難な問題は文字の分離であるが，これには，いくつの文字が走査されたかによる強制分離の方法が良いとしている．

最後の決定を下す所は，古典的な決定関数の方法を用いているが，この場合，無理に読んでしまうよりも読めないということが重要で，決定の信用度を測定し，その結果によっては読み直し，決め直しをさせる．そうでない場合，検出されなかった誤りよりはわからなかったほうがましである．決定のさい，測定空間の分離の表面は，data reduction がうまくいけば平面がいちばん簡単で良いわけで，なるべくそうなるのが望ましいと主張している．その他，文字だけでなく，広い意味のパターン記載の問題の困難さが議論されている．（森 俊二）

10. 文章生成のための動詞の型

Osamu Watanabe: Verb Patterns for the Basis of Sentence Generation 〔pp. 39～58〕

日英翻訳のためには，両国語の対応表をつくり，日本語の語彙と文法に英語の有する諸特徴を対応させなければならない．文法的には，日本語の構造を形のうえに現われた相互関係から確かめれば，かかる構造の記述は文を構成する形式的法則にせまるものとなる．これは，橋本進吉著国文法体系論の連文節論を要約

ン，いくらかの修正を加えたものである．

語は自立語 α と付属語 β とに分けられる．$\alpha+n\beta$(n=0, 1, 2, ……, i) は文節である．1つの文節の前後にどんなものがくるかはきまっていて，それといっしょになって大きなまとまりをつくる．前のものをウケル関係 b は，文節の自立語 α の品詞で区別される．次へのカカル関係 e は付属語 β，ときには自立語 α などで示される．

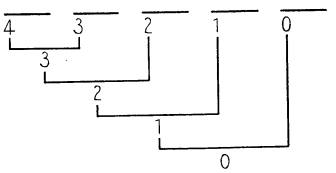
$\alpha+n\beta \rightarrow [{}_{b}A_{e}] \rightarrow A+B$ ($[{}_{b}A_{e}]$ は文節，A，B はそれぞれ b, e の表示者，とくにAは α)

文節の連続は互いに何らかの関係で結合してつながる．b, e が前後との相互関係を示す．

I．文節は1つの語 A と一定の関係で結合するものとする．

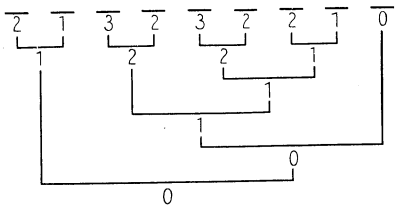
1つのAのほかに結合するウケル関係がただ1つであって，種々の文節につながらないとき，この文節群は連文節となる．この連文節全体のウケル・カカル関係は，先頭の文節のAと最後の文節のBとで示される．すなわち，連文節は文節 $[{}_{b}A_{e}]$ と同様な形式をもつ．

最後にきた文節 $[{}_{bn}A_{nen}]$ が断止文節であればこれは文となる．この文における各文節の次元と群化の手順は



1つのAあるいは文節がいくつかの文節（群）をうけることがある．このとき1つの文節はいくつかの文節を距てた後のAに連なり，この文節群がまとまって1つの連文節となる．

この文における各文節の次元と群化の手順は次図のとおりである．



A_i が用言のとき，それに連なる $[{}_{b_1}A_{1e_1}] \rightarrow A_1+B_1$, $[{}_{b_2}A_{2}A_{2}] \rightarrow A_2+B_2$, ……はそれぞれ主語，客語……といういろいろの特徴 $a_1, a_2, \cdots\cdots$ をもち，その関係は $B_1, B_2, \cdots\cdots$ に表わされる．ある用言は特別な a_i を要求するので，この性質によって動詞を類別できる．

文は連文節に変換することができ，より大きな構造の一部分となる．

II．1つの文節が〔A(用言)+B〕のBと一定の関係で結合することがある．

例．モハヤ，決シテ，トウテイ．……

III．(A_i+B_i) と (A_k+B_k) とが結合することがある．そのうち，ハを伴った題目語や「幸ニ，アイニク」や連結語などの構造を含む文は，二文の連接，列挙から生じたものと考えることができる．

このように用言を中心として文を構成する手順を，日英両語について独立に調べ，文の講造，構成の手順の対応表をつくるとき，精確な日英対訳の実現に近づくこととなろう．　　（渡辺　修）

11．予備編集された日本語の機械翻訳

Akihiro Nozaki: Mechanical Translation of Pre-edited Japanese〔pp. 83～92〕

予備編集された日本語を，能率よく英語に翻訳することを目標に，実験を試みた．ここには採用された翻訳手順と，その文法的基礎的考察とが述べられている．

1． 英文法について．

英文には，動詞によって決定される5つの型がある．修飾句の挿入を除いて，すべての英文は，そのどれかの変形である．

変形とは，次の作業を含んでいる．

1) 副詞“not”または“do not”の挿入
2) 助動詞（句）の挿入
3) 動詞変化
4) 動詞・助動詞および疑問詞の位置変更

日本文の解析にさいして，これらに関する情報に注目しなければならない．

2． 国文法について．

日本語においては，名詞・代名詞などの文法的機能は，それに伴う助詞によって決定されると考えられる．しかし，同一の助詞が多様な意味で使用されることに注意しなければならない．

これを区別する手がかりとして考えられるのは，その文（あるいは節）で使用されている述語であろう．例によって示されるように，助詞の用法は，述語によって，著しく制約される．

私ハ横浜ニ行ク ⟶ I go to Yokohama

私ハ横浜ニ生マレタ ⟶ I was born in Yokohama

このように，述語は意味の上ばかりでなく，文法上も重要な情報をになっている．端的にいえば，述語によって対応する英文の構造が定められ，助詞の用法（したがって名詞の機能）も確定する．

われわれのプログラムは，この事実を利用する．

3． 翻訳手順

わかち書きのすんだ入力の日本語は，ローマ字書きになおされ，紙テープを通して記憶装置の中に文字列として格納される．

プログラムは，その文字列を，逆順に解析してゆく．まず，文（節）の述語が処理される．述語（になり得る語句）の文法的性質（対応英文の型，助詞に対する指示など）は，すべて辞書に記載されている．この段階で，対応英文の**原型**がリストの形で作られる．

その他の語句も，やはり逆順に，1語ずつ調べられ，すでに得られている原型に順次あてはめられる．

あてはめが終るまでに，**1**に述べた‘変形’のための情報が収集される．それから，次のルーチンに進み，動詞変化などに必要な変形が行なわれる．そして最終的英文が作られ，印刷される．

（実験は弥永氏の論文にある文例（1），（2），（3）について行なわれ，所期の目的を達した．翻訳時間は入出力を除き，1文あたり 10 ms 以下であった）

（野崎昭弘）

12． 日本の MT 用機械の調査

Tuneo Tamati: A Survey of MT Oriented Hardware in Japan〔pp. 93～108〕

現在わが国では，電気試験所，九州大学，京都大学，防衛庁などで機械翻訳に関する実験が進められている．そのうち，前二者は特殊機械を用いており，その他では，たとえば京都大学では KDC-1，防衛庁では NEAC-1103 などの汎用電子計算機によって実験が行なわれている．ここでは汎用電子計算機についての説明を除き，前二者についてその概略を紹介した．

九州大学では 1955 年から MT の可能性を検討し始め，1960 年に実験機 KT-1 を作った．電気試験所では 1958 年に研究を開始し，1959 年に翻訳機“やまと”を完成した．この二者は，ほぼ同じ頃に完成したものであるが，種々の点で興味ある対照をなしている．

KT-1 は任意の2カ国語間の翻訳に必要な一般的な過程を対象とし，探索方法や文字演算などに特に便利な機能を持つ2進法直並列型計算機である．記憶装置は 10 万ビットの磁気ドラムのほかに，100 万ビットの磁気ドラムも最近併置された．使用文字は数字，アルファベットおよびカタカナより成る．現在この機械を用いて日本語，ドイツ語および英語の3カ国語間の翻訳や機械辞書の自動作成の基礎実験などが行なわれており，さらに大型の機械 KT-2 の製作が計画されている．KT-1 で行なわれた翻訳のプロセスは，表を用いた直接構成要素法の改良で，その構文処理は，すべて同一形式の法則を繰り返し適用する操作によっている．

“やまと”は特定国語間，すなわち英語から日本語への翻訳を基礎として設計製作されたもので，機構の簡単化，特殊な機能の高速化などに特色を持つ2進法直列型計算機である．記憶装置は 82 万ビットの磁気ドラムで，内部は辞書，テーブルおよびプログラムの三部分に分かれ，辞書は可変語長方式である．“やまと”で行なわれた翻訳の方法は，まず原文の構造を単純化して基本文型を知り，これを日本語構文に変換してゆくもので，その構文変換はプログラム法によって詳細に記述される．辞書は word, idiom, syntax および訳語の四種から成り，そのおのおのに，対応する情報を収容した3つのテーブル（訳語については不要）が付属する．辞書には約 1,000 語の英単語が収容された．

日本語と印欧語の間の翻訳では，印欧語相互間の翻訳に比べて

1） 相当長い情報を一時記憶に貯えなければならない．

2） 構文の処理は複雑で一般に時間が長くかかる．

3） 入出力で数多くの文字とその処理方法を必要とする．

などの問題がある．現在わが国の MT 研究は，まだ糸口についたばかりで，今後の発展にまつ所が大きいが，それに伴って機械に関するいろいろな問題が解決されるものと思われる．（田町常夫）

13． 機械翻訳の模型と方式

Toshiyuki Sakai: Models and Strategies for MT〔pp. 109～123〕

（1） Preface: 機械翻訳において，対象となる2つの言語が同一言語系に属していないときには，語順

の変更・字句の挿入や削除などを行なう必要があり，複雑となる．英語・日本語の場合は同系ではないが，日本語は英語に比して冗長度が低いから，英語から日本語への直接の翻訳は，比較的容易にいくと思われる．翻訳では従来の品詞の考え方にとらわれず，機械的な対応関係を中心にして，品詞などを決めるのがよい．翻訳の評価は，文の種類・辞書の大きさ・処理時間・訳文の質などで決まる．機械翻訳の手法としては，flow chart 法や look-up 法が用いられるが，後者のほうが機械にとって処理しやすい．

（2） Model: 直接構成，予測，従属，合成に基づく解析などの解析法が発表されているが，いずれも本質的には大差はない．Context free なときと，context sensitive なときを別々の方法で取り扱う方式に対し，適用する規則に rank をつけ形式的には同一の方法を用いるより機械的な処理が考えられる．後者は単純マッチングを行なうだけであり，rank はパターン表を区切るだけでよい．

（3） 辞書: Alphabet，カナ，漢字を入出力として扱うときには，辞書の形式・code について熟慮する必要がある．辞書の“見出し”として check sum の方法をとると，混乱の問題が残る．

（4） Look up method: 現在われわれの所で実験中の方法について詳述する．その特徴は（i）隣接結合，（ii）3連・2連のパターン，（iii）解析と合成が同時，（iv）右方→左方 scan である．3連・2連のパターンにはさらに A·B 2つの rank がつけられている．品詞は，英語と日本語を含めた系に関して定める．すなわち form class words は名詞・動詞・形容詞・副詞に相当し，おのおのはさらに細分される．Structure class words はそれ以外の単語で，構造上たいせつな働きをするので，品詞を1つだけ持つように定める．なお，この2つのほかに句，節に相当する cluster 記号があり，全部で品詞記号は96個である．翻訳の手続上最も重要なパターンの構成は，Noun cluster, verb cluster の規則的構造より，その一般的な型が抽出できる．現在は約 800 のパターンがある．パターンの rank A は，強い結合で主として，名詞，Bは弱い結合で主として動詞を含むものである．翻訳手順は，英文読込み・語尾処理・熟語処理・単語辞書・パターン・マッチング・日本語訳の順で文末から文頭へと scan しながら縮退させ，同時にそのパターンに対応する日本語訳を合成していく．その規則を示す．（i）A: 3連・CDE→F，2連・GH→I；B: 3連・C′D′E′→F′，2連・G′H′→I′．（ii）Aは Bに優先し，3連は2連に優先する．（iii）Aに一致パターンがないときのみBを用いる．

（5） 最後に，この方法の問題点をいくつか挙げ，処理した例文をパターンとともに示した．（坂井利之）

14. 英和翻訳公式

Bumpei Koro: English-Japanese Translation Formulas〔pp. 31～38〕

英語・ドイツ語・フランス語などのインドゲルマン語と日本語との間には，外面的言語形態の非常な相違にもかかわらず，文章構造の内面的一致（諸言語の普遍的原型）が存在している．この原型があるからこそ翻訳が可能になるのであって，これが諸言語間の翻訳公式 translation formula となって現われてくる．

英語の有客体動詞 objective verbs が，文章論的には，対格客体部 accusative objects を支配するか，もしくは対格客体部と前置詞付客体部 prepositional objects とを同時に支配するように，日本語の有客体動詞は，次の 12 の類型の後置詞付客体部 postpositional objects のみを支配する．

ε型		○を	読む	=read	○	
π型		○と	争う	=contend		with○
ρ型		○に	対応する	=correspond		to ○
ρ′型		○へ	行く	=go		to ○
σ型		○で	旅行する	=travel		by ○
τ型		○から	来る	=come		from○
α型	○を	○と	結ぶ	=combine	○	with○
β型	○を	○に	与える	=give	○	to ○
β′型	○を	○へ	運ぶ	=carry	○	to ○
γ型	○を	○で	洗う	=bathe	○	with○
δ型	○を	○から	抜く	=draw	○	from○
ω型	○の	○を	奪う	=deprive	○	of ○

英和翻訳公式と和英翻訳公式との主要部分を開発し，そのうちの「○ with ○」と「○ from ○」とを報告してある．（紅露文平）

15. 日本の MT の現状と問題点

Hiroshi Wada: Status and Problems of Mechanical Translation in Japan［pp. 3～12］

日本における機械翻訳の研究経過が記されている．すなわち 1955 年に九州大学が，1956 年に教育大と電気試験所が始めてから，学会の研究委員会，文部省の助成金による丹羽委員会が一段と研究を促進したこと

である．

つづいて日本語の概説がある．漢字，仮名文字の由来，漢字には同音で異形異義のものが多いことを指摘している．ローマ字では発音だけしか示せぬこと，漢字と仮名が混じていると分かち書きの役目を一応果していることを述べてある．構文では関係詞のないこと，単語の間にスペースのないこと，文は人間が主体となる思想で表わされることが多く，しかもその主語が省略できることなど，日本人が英語に苦しむ理由を英文とその訳文とを例示して述べてある．したがって日本語と英語との間の翻訳では単語辞典の集積に努めたアメリカの露英翻訳と違って，少くとも単文を単位とした構文分析法に立脚して語順を変更し，テニヲハを加えることが必要なことを示唆している．

わが国の翻訳は主に英和であるが，単語辞典，熟語辞典には，それ故に，構文の分析に要るデータが特に意を用いて含められていることを電気試験所のプログラムについて説いてある．

大容量の記憶装置を使うほど大規模な研究がないので，実用化にはほど遠いが，独得の装置として漢字テレタイプを紹介し，約2,500の文字が取扱えるので日本人が穿孔する際はあまり労せず使える筈だが，実際にこれを用いた研究はない．

日本文の英訳を企てる研究も始まったが，字引の見出し語の作成法，分かち書きの処理法，構文分析に有効な文法などを確立するために言語学者の蹶起を要請している．

そして，日英の協力で研究が有効に促進できる方途を見出したいと結んでいる．（和田 弘）

16. 自然言語の一般理論の試み

Makoto Nagao: An Approach to the General Theory of Natural Language [pp. 59～73]

機械翻訳の立場から文章をみたときの二つの大きな問題は syntax と semantics である．syntax については種々の方法が開発されているが，semantics の分野での研究はほとんどない．現在の MT の直面している最も困難な問題の一つは多義語の処理等の意味の問題である．よって semantics に関する一般的かつ強力な方法を確立することが望まれるわけである．syntax と semantics の両面は，実際の文章となって現われるとき，有機的に影響しあい，結合しあって現われるものであるから，syntax と semantics とを切離して完全な研究はなしえない．

この論文で提案しているものは，このような立場からの言語の一つの有力な研究方法である．

ここでは言語を文章の生成（sentence generation）という面からとらえ，意味と構造との間に存在する関係を明らかにしようとする．生成された文章の評価としては「文法，用語等の点で完全ではないが，具体的に理解しうる観念を伝えている文章で，一般常識のあるものが容易に正しい文章に訂正可能（corrigible）である」ということを目標とした．

まず文章構造は句構造としてとらえるが，このとき，その句の中心的概念を構成している要素を主構成要素とする．次にこれらの句を結合して一つ上のレベルでの句を構成するが，このとき各々の句の主構成要素間で意味的な結合関係が矛盾しないようにする．たとえば，文章＝名詞句＋動詞句 の各句の主構成要素，名詞と動詞が文章というレベルで結びつきうるかどうかをテストするわけである．このようなテストを一つの文章の句構造表現のすべての個所で行なうが，実際にはすでに定まった構成要素を基準として，このようなテストに合格するような単語を他の構成要素に対して選ぶ．

このようなことができるためには，各単語に対して種々の特性が与えられていなければならない．たとえば，ある動詞は動詞自身の分類では身体の運動に関するものであるというような意味的なものから，その動詞の主語になりうる名詞はどのようなカテゴリーのものか，目的語，補語等にはどのようなカテゴリーのものがなりうるか，またその動詞を修飾しうる副詞はどんなものか等である．つまり単語同志の結びつきの可能性という観点から単語のもつ諸性質を明らかにし，これによって各句構造内での単語選択がなされる．しかしすべての句構造に対して，このようなテストはできないので，ここでは主要な句に対してのみ意味的一致の条件をもちこんでいる．この実験で明らかになって来た問題点は，各単語に対して上記のような特性が適切に矛盾なく与えうるかということ，一つの単語の選択が文章中のその句構造をこえた遠くの部分にまで影響をもつ場合がしばしばあるということ，逆に前置詞句（従属節）とそれがかかってゆく主句（主節）との間の意味的な結合関係をとらえることが非常にむずかしいこと，などである．

この方法で生成された文章に次のようなものがある．

Often my friend always enjoy many good voyage to Kyoto and sea.

He find the Mt. Fuji but the lake Biwa dark.
To heve them lead, the god think Japan bad.
The north look cold.　　（長尾　真）

17. 日本の MT の心理，言語学的研究

Seiichiro Ohnishi: Psych-Linguistic Studies of Mechanical Translation in Japan [pp. 13～22]

日本における MT に関する心理，言語学的研究は，二つの部面からまとめることができる．

第一は，日本語の情報論的研究についてであり，蒲生秀也，田町常夫，伊沢秀而，筆者等によって行なわれた．これらの研究は，日本文および単語における文字を成素とした情報量の測定，ならびに日本文における単語を成素とした情報量の測定とにわけることができる．そして，日本語において F_N の値は，N の値の増大にともなって単調減少してゆき，ことに F_2, F_3 のあたりで急に減少することが示された．これは，日本語の文章構造上の一つの特質と考えることができる．

また，単語について文字を成素とした情報量についてみると，F_2 以下急に低い値を示す．個々の単語についてみると，引きつづき使用される文字が先行文字に規定される程度はかなり高く，ここにまた単語の一般的特質がみられる．

第二には，翻訳に関連づけられた心理，言語学献研究についてであるが，ここでは，言語の統計的研究と文型ならびに翻訳に関連した動詞，前置詞についての研究を中心としてみた．

まず，単語についていうと，書物の性質によって語の使用頻度がかなり異なっていることが明らかにされた．たとえば，小保内虎夫，金子隆芳は，数学書と一般科学書とを材料として調査を行なったが，数学書では 150 個の単語で全体の 80% の用をたすが，一般科学書では同じ量の単語で全体の 60% をおおうにすぎない．しかし一般的には，異なり語についてその使用頻度順位と使用頻度についての Zipf の法則はかなり妥当することが明らかにされた．

次に筆者は，日本文を品詞別にわけてその結合状況を調査したが，1000 個の 5 重語は 294 種の型にわかれ，もっとも多く出現する型は，名詞—助詞—名詞—助詞—名詞であり，140 回出現している．そして 5 回以上出現している結合の型は 641 個であり，結合の型には，かなりのかたよりがあることが明らかにされた．

次に，田口孝之は，文章の中で動詞がどんな形で使用されているかを調べ，とくに“する”という動詞をつけることによって作られている日本語の特質について考察した．

さらに金子，大坪は，中学校用リーダーを資料として前置詞を調べている．それらは，形容詞句的に使われる場合と，副詞句的に使われる場合とがあり，日本語の“の”にどのように訳されうるかを考察した．また“名詞—前置詞—名詞”の形であらわれる場合の前置詞の翻訳をどのようにきめるかの問題を考察した．

以上ここにあげたものは，MT と密着した研究とはいいえないものも含まれている．しかし，このような心理，主語学的な研究が，さらに今後研究の推進のために役立ちうるならば幸いである．　　（大西誠一郎）

18. 特殊計算機 KT-2 の方式

Toshihiko Kurihara: Design of the Special Purpose Computer KT-2 [pp. 137～143]

翻訳の機械化を行なうに必要な辞書の構成法，および文法操作の考え方を述べ，金物の見地からまとめてその概略を述べてある．

単語，文法辞書の探索および構文操作

翻訳過程を機械化する場合，必ず必要となる単語辞書探索についてはつぎの七つの場合に分けた．（1）原語が変化していない単語．（2）語尾が変化した単語．（3）went のような完全に違った変化形の単語．（4）単語が未知である場合．（5）続いた熟語．（6）合成語．（7）離れた熟語．各々について以後の処理に必要な情報を取り出す．この情報をもとにして文法辞書の探索が行なえるように語類の配列を作る．

文法辞書の探索においても，未知語のないとき，未知語あるいは多語類のあるとき，のそれぞれについて考察を行なった．この場合も前述の単語処理の場合と同じく，一致したシンタックスパタンの内容がすぐ取り出せるように必要な情報を取り出す．

つぎに構文操作に移るわけだが，構文操作により取り出された文法辞書の内容をもとにして，まとまった語順をセクエンシャルにたどれるようにする．これと同時に多義語指定，語尾挿入および文法情報の欄に新しい情報を加える．さらに一語を二つ以上に分けて訳す場合の処理を行なう．

機械翻訳における構文辞書の構成の方法はつぎの点を考慮して変換辞書を作成している．（1）構文辞書の構成については完全にくくれることが確認される程

度の長さの配列をとり，その一部をくくる．（2）構文配列をくくるたびに得られる情報を附加する．（3）できるだけ分岐を避けるように辞書を作るが，文法的に一意に定まらないときは特殊の符号⊗をつけて現在の情報を記憶する．単語辞書，構文辞書の内容は紙面の都合で省略する．

入力言語について

情報処理用機械では通常仮名タイプが用いられているので，これを漢字まじり文に変換することを要求されることがある．これも翻訳の一種と考えられる．この変換を行なうには同音異語の問題を解決しなければならない．また，実用的な見地から仮名文は，分かち書きしてあるものとする．この変換用の辞書を作り，紙上で実用文について試みたが，約 90% 程度変換可能であった．この操作も上記の探索および構文操作の一部として取り上げることができる．（栗原俊彦）

19. 機械による分かち書き

Hirohiko Nisimura: Automatic Segmentation of Japanese Text [p. 125～135]

機械翻訳の第一段階は自然言語の字の形で表現されている単語を，辞書を引いて適当な機械語の情報におきかえることである．英文においてはスペースが単語を浮きださせているので，語尾変化の問題はあるにしても辞書引きは容易であろう．和文は字が隙間なしに連なっているので困難を生ずる．字の連鎖のなかから単語を見つけだす手続を分かち書きとよぶ．

まず辞書に登録されている連鎖のみを単語とみなすことにする（見出原則）．つぎに内包された多重の単語については一番長いものを最尤解とする（最長一致原則）．最後に字の連鎖は左から右へ生長させるものとする（右方向原則）．

この三原則を厳密にしかも効率よく適用するには辞書の見出を Iverson 流の相続順位リストにして機械に入れるとよい．こうすると索長は Booth の二分法にほぼみあうぐらい速くできる．また1字／1機械語に割りつけても記憶容量に損失はない．つまり普通のやりかたでは可変長の自然語を固定長の機械語に入れるため，手続が面倒なうえ，右側に余白ができる．見出語の左側の重複部分も冗長である．これらがカバーされるのである．

その他，原文があらかじめスペースで区切られていても特別の処置を要しないこと，リスト言語やバッカス記法によく調和していることなどの利点がある．欠点としては外部記憶装置には不向きなこと，登録されていない単語が原文に含まれていたらお手上げなことなどがある．

普通の形式でパンチされた辞書をリストに組みたてるルーチン，およびこの辞書を引いて原文を字の連鎖から語の連鎖に組みかえるルーチンが Fortran でコーディングされ IBM 7090 でテスト中である．

（西村恕彦）

20. 日本文法の第0近似

Isao Imai: Japanese Grammar of the Zeroth Approximation [p. 23～30]

日本語の文法を，その本質を生かしながら，できるだけ簡単化する．そして日本語文法の第0近似とも呼ぶべき，日本語の骨格をえた．

日本語の解析にあたって，体言Sと，用言dとを区別するのが便利である．また，隣接する語の機能上の結合関係を示すために，記号＊，×，・，：を導入しよう．これらの記号によれば，任意の文 σ は，次の基本形によって表現できる．

$$\sigma = S*d$$

名詞Nと副詞Aとは，体言である．動詞 v，形容詞 a は用言である（用言の語尾変化を表の形で示しておいたが，それは日本文法の第1近似を与える）．

S, d は，単独な語だけでなく，次の公式から導かれる語群でもよい．

$$d \cdot N = N$$

$$A \times a = a,\ A \times v = v,\ A \times A = A$$

$$\sigma = a$$

第0近似においては，助詞は無視された．助詞 p を加えて公式を拡張すれば，第1近似の日本文法ができる．しかし，体言の相互関係が明らかなばあいには，助詞を省いてもよいことは注意しておこう．たとえば，次の二つの例文を比べてみよう．

キミ　コノ　ホン　ヨンダ？

キミ　ハ　コノ　ホンヲ　ヨンダカ？

話し言葉としては，第1の例の方が実は自然である．なお，記号：は，助詞ハを介する結合を示すために用いられる．

日本語文の基本原則は，要するに，曖昧さの起らない限り，冗長度をなくす点にある．助詞や格語尾などの省略は，この原則に基づいている．結論として，日本語文法は，少なくとも第1近似では，西欧語に比べて著しく簡単であるといえる．このことは，他国語か

う日本語への機械翻訳を，簡単にしてくれるであろう．（編集部）

21. 日本語の予備編集の方法

Shokichi Iyanaga: Procedure for pre-editing apanese [pp. 75～81]

機械翻訳の究極の目標は，翻訳の全過程の機械化にある．しかし，そのことは，構造の全く異なった2国語間の翻訳のばあいには，困難な問題を含んでいる．そこでいちおう人間の介入を許し，将来は人手による作業の機械化を考えることにした．

翻訳の全過程は，次の3段階にわけられる．1）予備編集．2）機械翻訳．3）再編集．ここでは和文英訳のための予備編集について述べた．

予備編集者は，日本語を理解し，また多少の文法知識を有するものとする．しかし英語と機械については，予備知識は仮定しない．

入力用の文字としては，ローマ字を用いることにした．そこで，予備編集者は，原文をローマ字書きになおすことから始めることになる．分かち書きも，のちに述べる注意の下で，行なわれる．

予備編集者は，さらに次の仕事を行なう．

（1） 文を単文ごとに括弧で区切り，主語を補う．

（2） 助詞および接続詞について，それが支配する語の範囲を明示し，また“標準化”を行なう（標準化は，他のむずかしい語法についても行なわれる）．

（3） 分かち書きを，辞書にあうように訂正する．

デ　アル→デアル，ドウヨウ　ニ→ドウヨウニ．

（4） 名詞を修飾する形容詞節の中での，修飾される名詞の役割を，斜線／つきの指示代名詞ソレによって，明示する．

このような作業を行なうために必要な，簡単な“文法”も要約して述べてある．最後に，11の文例について，原文と，予備編集の結果とを，予想される翻訳結果（英文）と併せて掲げた．（編集部）