

## 機械翻訳セミナ論文紹介

### 90. 開会の挨拶

Yasujiro Niwa: Welcome Address [pp. 1~2]

科学協力に関する日米委員会は 1961 年の池田総理とケネディー大統領との会談の結果生まれたもので、その第一回会合において 5 項目を選び、両国政府に勧告した。その中に科学情報と資料との交換の促進があるが、このために両国にそれぞれのパネルが置かれ、密接に連絡して仕事をすすめている。米国側パネルの主査は、最初は IBM 社の Dr. E.R. Piore であったが、今は NBS の Dr. A.V. Astin である。日本側は丹羽がつとめている。科学情報と資料のパネルでは、論文の抄録、目録索引に関する専門家の会議や一次出版物編集者の会合を企画した。この度の機械翻訳に関するセミナーは日米両国における機械翻訳の研究の現状を検討し、今後の協力につき協議する目的で、第 2 回の日米科学委員会で提案せられた。

日本における機械翻訳の研究は、今まで物理学者、数学者、電子学者、言語学者、心理学者などの数グループにより、大学や、研究所などで行なわれていた。これらの研究の連絡、討議のため、1962 年に機械翻訳研究委員会が発足し、日本における機械翻訳の研究者を網羅する態勢ができ、今までに 10 回のシンポジウムを催した。今回の会合についても、この委員会に負うところが少なくない。

日本の国語、国学、国文は欧米のそれに比し、非常に異なっている。この事実は論文の和欧または欧和の機械翻訳の研究をいっそうむつかしいものにするが、一方それだけ機械翻訳研究の国際協力の必要性を強くするものといえよう。今回の会合が日米両国における機械翻訳の研究と両国の協力の推進に寄与することを望んでやまないものである。(丹羽保次郎)

### 91. 米国における自動翻訳概観

Franz L. Alt: A Survey of Automatic Translation in the United States [p. 10]

機械翻訳の方法とその問題点を歴史的に述べ、米国の現状を明らかにする。

機械翻訳の考えは 1946 年に討論され、原文に対応する訳文を打出す辞書のようなものを考えていた。単語の対応が 1 対 1 でないの、あまりうまくいかない。

多義語について 1949 年に Weaver が、前後の文脈で意味がきまるはずであると述べたが、実用的には不可能である。しかし、ごく単純化して要約や検索に用いようとする試みはある。もう 1 つの多品詞語については、1951 年カリホルニア大学の Oswald と Fletcher が独英翻訳のさいに取り上げ、今日の研究基礎を築いた。

1951 年ごろから機械翻訳は広く興味もたれるようになり、逐語訳では不十分、意味のあいまいさの解決は時期尚早、構文分析合成のためのコード研究が必要などの点が認められた。

一般に逐語訳では不十分と思われているにもかかわらず、1954 年にジョージタウン大学、続いてハーバード大学が実験、その後も IBM 社などで研究が進められている。

関連した研究も発展し、ハーバードやワシントン大学の機械辞書、Rand 社の頻度調査、TRW や MIT の機械言語などがある。翻訳のためのプログラミング言語が、重要なことはよく知られている。インジアナ大学では翻訳結果の評価方法を検討している。

文法は理論的、伝統的、経験的な流派に大別される。

理論派では句構造言語の理論が最大の収穫である。この派から標準言語理論、変換理論、層化理論の 3 つの論文を、このセミナーで報告する。MIT、カリホルニア大学、テキサス大学が、この派に属する。なお分析と合成をつなぐのに元の情報全部を含む一種の言語(中間言語)をおく考えがあるが、著者としては自然言語

\* 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、通常の文献引用の形式を守っていない。

は、前後の語や文などで意味を規定しているの見込みはないと思っている。

経験派は、少数の規則から始め、テキストで試し、不足があれば新規則を加えていくものである。この方法は、まとまりのない規則のかたまりができる欠点がある。ジョージタウン大学の研究が現在進行困難なのは、このためと思われる。

伝統的文法を利用する方法は、ジョージタウン大学、TRW、1958年にNBS、ハーバード大学で採用したが、大成功に至っていない。

米国は10数年の研究にもかかわらず、原文から訳文を打出すまでに至っていない。しっかりした理論的基礎をもった研究は着実に進んでいるから、むずかしい研究ではあるが、望みはあると思う。ことに分析のほうはだいたい目安がつきそうで、ただ意味などと独立に処理できるかどうかが問題である。(玉井陽子)

## 92. 米国側寄稿の要約

D.G. Hays: Summary of U.S. contributions [p. 11]

計算機の発展につれて機械翻訳を始めとし、いろいろな言語学的の研究が発足した。それらをまとめて近時計算言語学 (computational linguistics) という。

計算言語学においては

(1) 言語学の理論として伝統的な文法を拠点とする学者は少なくない。つまり文法は散漫な嫌いはあるが、これを数学的に記述すると内容が膨大で、どうにもならないとの立場をとるものだ。P. Garvin, H.H. Josselson, F. Alt などはこの派に属する。

(2) これに対し N. Chomsky はすべてを形式化 (formalization) することを企て、i) 言語学の理論はどんな資料をも完全に説明できなくてはならないとの主義に立脚して、ii) immediate constituent 論に通ずる phrase structure grammar を確立し、iii) transformation のモデルを示して、多くの信奉者を作った。W. Wang はこの派に属する。

(3) S. Lamb は strata 説である。stratum を認める学者は少なくないが、彼のは徹底したものである。

(4) Hays の dependency 論は、predictive model その他多くの説とともに、文法のある一面を眺めたモデルにすぎない。

(5) E.D. Pendergraft のメタ言語を用いる方法は、Chomsky, Lamb に似た面がある。

以上の諸説の優劣は実験結果に待つ以外あるまい。それには文章を計算機用に記述しなくてはならないが、これがめんどうな仕事なのに、できた資料の交換、集中配布などがなかなかうまくゆかない。いっそうの協力が必要と思う。ことにその言語を母国語とするひとびとと。

だからなまのテキスト、注釈をつけたテキストは提供する用意がある。もっとも注釈の内容に問題はあるが。

機械については本体よりも typesetting のようなものに期待したい。記憶装置への要求は際限ないけれども、file は次第に大きくなり、parsing のルーチンなどには、大きな容量のコアがある。専用計算機については、まず理論を建てて、それを試験してみないと何も評価できない。

なお自分のいる RAND では、テキストを磁気テープに入れ、format 変換のプログラム、テキストを修正編集するプログラム、テキストの一部をコアメモリに読み込ますプログラムなどが作られている。

(和田 弘)

## 93. 機械翻訳の言語学的基礎

H.H. Josselson: Linguistic Basis of mechanical Translation. Contributions of Standard Linguistic Theory [pp. 39~60]

米国の MT のほとんどのグループは、米国言語学界で今日開発されつつある諸理論を採用している。MT に使う文法は、伝統的な文法よりもっと、人間が翻訳をする際のいろいろな因子を形式化してこれを組みこみ、電子計算機にかけられる形にする必要がある。

MT における言語学者の第一の仕事は、当該2カ国語の単語や接辞の変化表の辞書を編集することである。次に文法を作るが、米国の MT の文法理論は、大勢は成分構造 (Constituent Structure) が依存関係構造 (Dependency Structure) かである。前者は文を主部と述部、主部はさらに名詞句とその他、述部は動詞句+名詞句というように、次々に2分してゆく方法であり、後者の理論は、文は述部動詞が独立位で、これに主部とか目的部、副詞修飾部がついている依存の構成であるとして分析してゆく立場である。

文法に載せる語類は、語形論的分類と構文論的分類とがある。文の構文構造をみつける blocking 法というのがある。これは文の語線の中から名詞群、前置詞

群、修飾部群などと次々にみつけてゆく方法を、機械に教えるものである。また予想法というのは、ある形式の次にどういう形式がくるかを、それまでの文法上の諸特性から予測する方法である。

入力言語と出力言語の文法構造上の差異は、困難を生ずる一原因で、2カ国語の比較分析が重要になる。2カ国語の間に一対一の対応が欠けているときや、一つの形式が別の形式の中に入り込んでいる（nesting, embedding）場合はやっかいである。

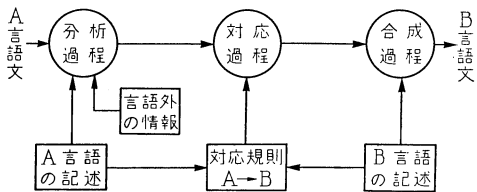
今日、米国の言語学界で勢力をもちつつある言語理論に、変換（Transformation）理論と層化（Stratification）理論がある。多くの MT グループは、成分構造理論によっているが、変換理論や層化理論によっているところもある。

MT がもっと人間の翻訳過程に似てくるためには、2カ国語の文法研究がもっと進まなければならない。意味の研究は少しずつ進められているが、言語学が進み、この方面のことがもっとわかれば MT に貢献するところが多いはずである。（小笠原林樹）

#### 94. 機械翻訳への変換理論的アプローチ

W. S-Y. Wang: The Linguistic Basis of Translation: a Transformational Approach.

人間の翻訳過程は、次図のように考えられる。



したがってまず、A, B の両言語記述ができていないといけない。逆に MT は、ある言語の記述が妥当かどうかの一つの試練ともなる。

言語の記述は、辞書と文法規則、意味規則からなる。辞書は語の文法的特性、意味的特性、訳語を記し、その語のいろいろな用法を説明できるものでなければならない。これらの諸特性のどれが使われてどの意味が実現しているかをはっきりさせる必要がある。

意味特性についてはほとんど研究されていないので、この方面の研究を進めるべきで、Katz-Fodor の研究などは注目してよい。

上述の諸特性の間には従属関係（Dependency）がある。これには同類の単語のもつ諸特性間の内的従属

と、文中の他の語の諸特性間にみられる外的従属とがある。屈折語尾形式の従属、選択関係の従属は後者の例である。

以前の言語学は単位の設定の問題にのみかかわっていたが、近年は文を作るルールを描くことを重視するようになり、したがって成分構造の形式の本質、依存選択関係の研究が行なわれた。

文法形式の研究は近年、抽象オートマトンの理論をモデルにして、文法はある言語の文を作り出すのに必要で十分な一連の規則である。各文の形成に必要な規則を構造の記述と呼ぼう。するとできあがった文の形が同じであっても、その形成の規則が違うような場合、つまり、構造の記述が違うような場合の説明もつく文法の構築が必要である。そして文法は文法的な文のみを作り、非文法的文は排除するような文法でなくてはならぬ。このような文法の作成が MT においては特に必要である。

文法には、いわゆる句構造文法があるが、これだけでは等位構文や AA'BB' のような構造をうまく記述できないので、いわゆる変換理論をたてる必要がある。

変換理論によれば、句構造が全く異なる文の間関係もつけられ、また句構造が全く同じ文どうしもさらに区別できる。変換には1つの句構造に対して適用される単一変換と、2つの句構造に同時に適用される一般の変換とがあり、また必須の変換と随意的変換とがある。変換規則どうし間の適用順序も問題であり、Fillmore (1963) はこの点を研究したものである。

A, B 両言語間の対応規則を作るには、それぞれの言語記述が等質的にできていなければならないし、構文の一点と語形論上の一点や構文の一点と語彙論上の一点がズレて対応するというようなこともある。

分析過程、合成過程の研究はあまり行なわれていない。Matthews (1961) は「総合による分析」という方法を唱えているが、問題点が多い。このように MT の基礎となる言語理論自体が、開発途上でまだまだ問題点が多いのである。（小笠原林樹）

#### 95. 機械翻訳の基礎としての言語層化理論

S.M. Lamb: Stratificational Linguistics as a Basis for Mechanical Translation [p. 41]

言語の記述は、簡潔かつ一般的なものでなくてはならぬが、それは繰り返し現われる類似部分をまとめる

ことによって達せられる。MTにおいても、この原則により、(1) 言語記述とプログラムを区別分離する。(2) 語単位でなく lex 単位の辞書による、(3) 翻訳過程を言語層化理論の4つの階層に応じた形にする。これらによって、ずっと効率的になると信ずる。

言語層化理論を説明すると、自然言語においては、文字なり音声と意味がいきなり結びついているのではなくて、その間にいくつかの媒体があり、各媒体は別別な型をなしているのので、この間の関係をつける必要がある。意味が形に、あるいは形が意味に具現されるのは、これらの各階層に換算しなおされつつ形を備えるに至ると考えられる。この階層ならびに各層における単位形式を示そう。

階層	構成子	構成素	構成素連続体
Sememic	Semon	Sememe	Discourse Block
Lexemic	Lexon	Lexeme	Sentence
Morphemic	Morphon	Morpheme	Word
Phonemic	Phonon	Phoneme	Syllable

例を挙げて説明すると、3つの phonon/閉鎖性/両唇性/鼻音性/が組み合わさって phoneme/m/になり、/æ/n/と組んで syllable/mæn/となる。これは morphon/m/æ/n/に匹敵し、これが集まり morpheme/mæn/になり、これはまた word でもある。これはさらに次の次元では lexon/man/になり、そのまま lexeme/man/でもある。しかし lexemeの中には、2つあるいはそれ以上の lexon からなっているものもある。/farmer/は /farm/, /er/ という2つの lexon からなっている。また morpheme と lexeme の違いを better について示すと、/better/と /good/er/のようにそれぞれ表記できるから、これを比較すればわかる(いわゆる語形論の一面と構文論は lexemic 層の lexotactic rule を示すことにな

る)。一方 sememe, たとえば/人/は3つの semon/物体性/生物性/人間性/が組み合わさったものであり、lexeme/human being/person/man/ はいずれも/人/という sememe を具現したものである。

言語層化理論による言語記述は、これら各層の構成ルール、具現ルールを含むものとする。多分 lexemic rule は Hays らの依存的樹枝状図によって示されるであろうし、sememic 層は、網状図によって示されよう。

MT に利用する場合には、MT はふつう文字言語によっているから、phonemic の代わりに graphemic 層をもってすればよい。

MT には層化理論による言語記述のすべてが必要でない。意義子の網状図を合成する必要もないからであり、また lexon のあるものは機械に入れられないであろう。層化理論による MT の過程を示せば、A 言語のテキスト (grapheme の連続体) が機械に入れられると、それが morphon 連続体 → lex (lexeme の形態素層における対応体) 連続体 → lexemic tree のように順次変えられる。そして、これが sememic network に変えられることになるが、この辺を機械にかけるには、まだ数年の研究が必要であろう。かくてA言語の意義子網状図は、B言語のそれに変えられ、以下逆のコースをとることになる。A言語とB言語の意味構造が似ていれば、意義子網状図は共通のものでよい。

層化理論に似た考えは、多くの言語学者に多少ともうかがうことができようが、これを組織的、顕現的に示し、層の数もこのように多く設け、意味面の層化も含めた研究は、これまでなかったように思う。

(小笠原林樹)