

## 解説



## 機械翻訳の概要と歴史†

田町 常夫††

## 1. はじめに

言葉を機械で翻訳することは40年前、1946年 ENIAC が電子計算機として登場して以来、人間の長年の夢であった。最近、社会の高度情報化とともに国際間の情報流通もはげしくなり、機械翻訳 (MT, Machine Translation) の必要性が急速に増している。

しかし計算機技術の著しい発展にもかかわらず、機械翻訳の進展は遅く、40年前の基本的問題点でも未解決のものが多い。もちろんハード、ソフトの道具立ての進展により構文規則や意味規則の精密化と処理の高速度化などでは著しい進展があり、現在ではかなり楽観的な見方を持っている向きが多い。

この40年間の大きな出来事として、1955年頃の公開実験の華やかな時期、1966年の ALPAC レポートによる MT からの撤退、1970年頃から人工知能分野 (AI) での言語理解への急激な進展、そして現時点での楽観論の4つがあげられよう。1945年から1985年までを10年で区切って時間的に見るとおおむね上の出来事が節目になっている。

1946~1955	草創期
1956~1965	試行期 (第1世代)
1966~1975	再建期 (第2世代)
1966~1970	(再建検討期)
1976~1985	発展期
1986~	(実現期?)

各10年期に名前をつけるとすれば上の右欄のようなことになるかもしれない。そして ALPAC レポート (1966) が出されたのは歴史のちょうど中間点であったことになる。

研究内容から見ると1940年代は提案期、1950年代

は単語レベルの翻訳、1960年代は構文レベルの翻訳、1970年代は意味レベルの処理、1980年代は知識レベルの処理ということになるであろう。また方式で見ると ALPAC レポートを境にしてそれ以前が第1世代、以後が第2世代の翻訳方式として大別される。ここでは時代経過を主にして眺める立場から一応10年の単位で区切ることにした。

MT の歴史と展望については各時点で多くの卓越したサーベイがあり、本稿もそれらに負う所が大きい。とくに W. J. Hutchins<sup>1)</sup> のすぐれたサーベイについては、すでに本誌でも各所で紹介されてきたが、本稿においても MT システムや機械支援 MT システムの紹介についてこれに負う所が大きい。なお過去10年位から以降の著しい展開は本号の各分野でのべられると思うので主としてそれ以前 (1975年頃以前) を概観することとしたい。

## 2. 草創期 (1955 以前)

MT の最初の提案はロシア人 Smirnov-Troyanskii とフランス人 Artsouni によって独立になされていた。Troyanskii は1933年に『一つの言語から他の一つまたは同時に多数の言語への翻訳に際し、単語を選別しかつ印刷する機械』の特許をとった<sup>2)</sup>。その考えはロシア語を電信を通じて同時に数カ国語に翻訳しようというものであった。彼は1944年頃までその支持を求めて回ったが不発に終わった<sup>3)</sup>。

最初の電子計算機が出現したのは1945年、ENIAC とそれに続いて EDVAC などが完成した頃であるが、MT の問題はすでに、電子計算機の一般的利用可能性の論議の中で起こっている。1946年 Birkbeck College の A. D. Booth と Rockfeller 財団の副会長である W. Weaver の議論の中で Booth は電子辞書の構想をのべ、Weaver は暗号解読の技術の延長として翻訳の可能性を示唆した。当時 Booth は構文処理のことは考えていなかったし、またすべての単語を訳すことも考えてはいなかった。文の意味は訳されたキーワー

† A Historical Overview of Machine Translation by Tuneo TAMATI (Department of Information Systems, Interdisciplinary Graduate School of Engineering Sciences Kyushu University).

†† 九州大学大学院総合理工学研究科

ドから推察すればよいとしていた。そして実際単語対単語の翻訳でもかなり役にたつ場合が多いことを示した。彼は日→英も試みている（筆者が例文を見たところでは、とても英国人には分かるまいと思われた）。

1949年 W. Weaver の“Translation”のメモが出され、これが刺激となって急速に MT の関心が高まり<sup>4)</sup>、human communication の共通のベースは何か、論理的不変量、統計的不変量、意味的不変量などが考えられるかなどの議論が盛んになった。先駆的な MT の研究は Booth のほか、E. Reifler (Washington State Univ.), V. A. Oswald Jr, S. L. Fletcher (UCLA), Y. Bar-Hillel (1951年から MIT) らに始まった。

1952年 MIT で英、米の言語学と計算機の専門家18名を集め、Bar-Hillel が組織し Rockefeller の援助で第1回の MT Conference が開かれた。MT の問題と目標について次の合意を得た。（この会議の歴史的意義は大きい。）『Step 1: およそ12カ月後に自動辞書を作る。Step 2: 構文解析の方法について研究を実施する。機械の設計はその後に行う。さらに単向の翻訳を実施、その結果により多国語間 MT へ、研究としては構文の論理、万能文法、意味分類など中間言語へ向けた研究を行う。また MT の研究現状について最新の成果を刊行する』これに基づいて“Machine Translation of Language” (Locke & Booth eds., 1955) が刊行された<sup>5)</sup>。

1954年、Georgetown 大学の L. Dostert 及び IBM の Sheridan らによりロシア語から英語への翻訳実験が公開された。単語250語と6つの構文規則を用いた IBM 701 による実験である。一般の反響があまりに大きく、問題点などについてまともに受け取られなかったくらいがある。しかしこれが政府、軍関係の多額の資金援助の契機となった。さらにこの年雑誌“MT”が発刊された。またこの年、ソ連でも MT の研究が開始された。D. Panov が BESM 計算機で実験<sup>6)</sup>、MT は可能との結論に達している。

1955年 Booth の仏→英 MT 実験が BBC テレビを通じて公開された。

### 3. 試行期 (1956~1965年頃)

#### 3.1 概要

1956年 MIT で第2回の MT に関する国際会議が開催 (International Conf. としては第1回) された。Panov も参加、ソ連の方法は構文処理に及んでい

た。活動の中心は米、英、ソであり、問題は可能性でなく、最高能率の問題であること、バラバラの研究を一定の目標に統一することなどの合意があった。

1956年には中国も研究を開始している。

我が国では1955年より九大が研究を開始し、1957年から文部省科学研究費の交付をうけて、MT 専用機 KT-1 を試作 (1959年)、日、英、独、三カ国語相互間の翻訳実験を行った (1960年)。電気試験所も1956年より研究を開始、1958年、英日 MT 専用機『やまと』の製作を開始し、1959年2月実験に成功した。いずれも当時の我が国には汎用計算機がなかったため専用機的设计に頼らざるを得なかった。

1957年モスクワで科学技術会議 (独、英、ハンガリー、中国など参加56件) が開催され、言語解析の数学的拡張、言語学研究の方向付けの必要性が認められた。また1958、1959年は高等教育省による数理言語学、MT などの会議ラッシュであった。

1959年までに米国では MIT, Georgetown Univ., Univ. Michigan, Washington State Univ., UCLA, RAND Corp. など12以上のグループが NSF (National Science Foundation) の助成で研究を行っている。英国では Birkbeck College (Booth ほか)、CLRU (Cambridge Language Research Unit) (M. Masterman), ソ連では科学アカデミー、レニングラード大学と Steklom 数学研究所 (O. S. Kulagina ほか) の言語の基礎研究グループ、実験派では Panov らの BESM による実験などがある<sup>7)</sup>。

1959年6月、UNESCO で International Conference on Information Processing (Paris) が開催され、米、ソ連や電気試験所『やまと』の研究などが発表された<sup>8)</sup>。なお IFIP は実質的にこのとき発足した (この会議は第0回 IFIP ともいわれている)。

以上1950年代には方式についてのアイデアが相次いで出された。それ以後は言語データの蓄積、整備関係が多くなり、用語は IBM や UCLA では基本形で10万程度、当時としてはかなりのものとされる。対象言語は、特定言語間 (露→英、中→英など) が多い。中間言語については全般には関心がうすい。生成文法の考え方に関心が深いのが特色である。その他辞書が大型化したため、辞書編集の自動化にかなり関心が高かった。なお語義的多義や曖昧な文脈などの問題があるが、短期間には解決しそうでないとの認識である。

1962年6月に Association for Machine Translation and Computational Linguistics (AMTCL) とい

う国際学会が発足した。雑誌「MT」がこの学会の機関紙として継続されることになった。ついでながら我が国では1960年4月に情報処理学会が創立され、創刊号には和田氏の翻訳機の話がのっており、意義深い。

1964年4月、日米科学協力委員会でMTに大変熱心だった故丹羽保次郎氏の尽力により、日米機械翻訳セミナー(U.S.-JAPAN Seminar on Mechanical Translation)が東京で開催された<sup>6), 6)</sup>。この頃、国内の実験グループは、電気試験所、九大工、京大工、東大理、名大工、防衛庁技研、電機大、などで、主に英→日MTを対象としていた。東大、電試では日→英も試みた。言語学のグループは国語研、新潟大、京大、名大、教育大、紅露研などであった<sup>7)</sup>。1965年New Yorkで第1回のInternational Conference on Computational Linguistics (ICCL)が開催され、この直後、同じ場所で第2回目の日米機械翻訳セミナーが開かれた。しかしこの頃にはアメリカではかなり状況が変わってきていた。これまでに2000万ドルを投じてきたがなかなかものにならないことに対する反省である。

1966年、National Science Foundationが組織したAutomatic Language Processing Advisory Committee (ALPAC)がNational Academy of ScienceにいわゆるALPACレポート<sup>8)</sup>を提出した。当時のMTの研究状況と周囲の状況から見て、近い将来MTが有用であると予測できる展望がないという結論である。つまり必要性、経済性その他の理由から人間の行う翻訳に比べて利点がないということである。その代わりに自動辞書のような翻訳者の機械的補助手段、計算言語学(CL, computational linguistics)の一般分野に対して継続的な援助を勧告した。

これに対し当然鋭い批判が出た。限られた範囲のMTの研究のみから結論を急いでいる。MTの関係する全般の分野に目をつぶっており、偏狭で近視眼的だということである。

この時代、試行期に積上げてきた成果はもちろん大きい。当時実用に近いと目された翻訳は特定二カ国語間の翻訳(直接翻訳)で、この方式を現今では第1世代のMTと称している。

### 3.2 試行期の研究の流れ

#### 1) MTシステム<sup>1)</sup>

前半はおもに単語対単語、後半は構文解析中心である。MTのシステムとして前項でのべた初期のものを除く代表的なものを二、三列挙する。

Mark II システム：アメリカ空軍外国技術部に1964

年設置され、1970年まで単語対単語の露→英直接翻訳を行っている。システムはE. Reifler (Washington 大学)の初期のシステムを改良したもので、post-editionを前提とする。辞書として光学記憶装置を開発した。このシステムの開発状況がALPACに影響したといわれている。

Georgetown 大学 (Leon Dostert): 1960年から構文解析へのアプローチをいくつかのグループで進めていた。(SLCという基本プログラムなど)。1964年に露→英の直接翻訳システムをOak Ridge National LaboratoryとイタリーIspraのEURATOMに供与した。グループは当時アメリカで最大のものであった。

日本では<sup>9), 6)</sup>日→英翻訳はまだ本格的には行われていない。野崎(東大)は和英翻訳を、原文をpre-editして入力する方法で行って、日本語の問題点を指摘した。実際の日本語文に挑戦したのはこれが最初である。その他、京大、九大、電気試験所の研究がある。

#### 2) 研究方法

##### i) 単語辞書

逐語訳が盛んなときなので辞書探索法、単語に関するものが多く、多義語、多品詞語(意味の問題)、辞書構造、見出し語の圧縮、単語の分解照合などの問題が多く研究されている<sup>11)</sup>。

##### ii) 構文解析

英語、日本語などについてまずどんな文法を使うかが問題であった。そこで経験派、伝統派、理論派の三派が出てくる<sup>7)</sup>。理論派はおもに生成文法理論を根拠とするものである。文解析法としては直接構成要素法、依存構造解析、予測分析法、ストリング解析法などが提示されている。直接構成要素法ではY. Bar-Hillel, Univ. Michigan, V. H. Yngve, 九大(KT-1)<sup>10)</sup>など、依存構造解析ではD. G. Hays (RAND Corp.), 予測分析法はS. Kuno (Harvard Univ.), ストリング解析法ではZ. S. Harris (Pennsylvania Univ.)などがおもなものであった。各分析法で考えている範囲は能力で文脈自由型の範囲である。V. H. Yngveのようにトップダウンのものもある。予測分析法はNBS (National Bureau of Standards)のIda Rhodesが露→英MT研究で用いているが、S. Kunoは、プッシュダウン・ストアを用いた巧みな手法を考えた<sup>11), 12)</sup>。

##### iii) 中間言語

CLRU (前述)のThesaurus method<sup>11)</sup>はRogetのシソーラスのような方法によって意味を分類し、意味の依存性によって語義的なあいまいさを解決した同

じ内容をもった質のよい表現を求めようとしたもので、中間言語を目標としている。Grenoble 大学の CETA は基本文形に対する表層的な中間言語構造 (pivot language) を考え、1961 年から MT の研究を始めている<sup>11)</sup>。その他 S. M. Lamb (Univ. Calif.) の層化理論 (stratification theory) もある<sup>5)</sup>。

iv) 記述方式として RAND の MIMIC, Georgetown の SLC, MIT の COMIT などの研究がある<sup>7), 11)</sup>。

v) 辞書及び構文規則の作成の自動化。

学習過程を含むアルゴリズムで、一例をあげると、システムが概念形成過程に相当する推論法則を持ち、対訳文章を次々に与えるとき、要素の抽出と概念化の 2 つの規則で単語、熟語、構文の対応を求めるものである。これは無辞書の状態から始めるともばら経験だけに依存するため、入力対訳文の与え方、順序により、普通の文法にはないような対応がでてくる。たとえば “fall to me.” などの “me” は、“おちてくる” の “てくる” に対応するようなことになる。問題は形成された概念の分解、統合の過程をうまく作ることである。この考え方は後に動詞の意味の内包を求めるときにも用いている<sup>13), 20)</sup>。

#### 4. 再建期前半 (1966 ~ 1970 年頃) (再建検討期)

##### 4.1 研究の概況

ALPAC レポート以後、アメリカでは MT を直接研究テーマとして行っている所はなくなり、限られた範囲での MT 研究が続けられるか、基礎的研究または他の言語処理へ目標が変更されるなどして、いくつかの方向へ向うことになった。カナダや EC 各国など翻訳に関して切実な問題をかかえる各国にも、この影響が及んだ。しかし MT の研究はあきらめられたわけではなく、基礎技術や機械支援システムの中で着実に研究が進められた。

現在考えられている構文解析の色々な手法がほとんどこの時期にあらわれてきている。またこの時期の終り頃 (1970 年前後) には人工知能関係 (以下 AI と略記する) からの言語処理研究が急速に進展してきたため、この前後で研究の流れが大きく変わることになった。

1965 年に発足した ICCL (計算言語学国際会議) (前述) や 1969 年に発足した IJCAI (人工知能国際会議) は以後 2 年ごと (6th ICCL, 1976 のみ 3 年目) に

開催され発展を続けている。

日米間では前に日米科学協力委員会の仕事として MT の共同研究が行われてきたが、基礎的な研究に重点をおく方がよいとして計算言語学のテーマでサーベイセミナー (U.S.-Japan Survey Seminar on Computational Linguistics, Hawaii, 1968-03) が開催された<sup>14), 15)</sup>。CL を日米科学協力の計画に含め、1971 年までにカナ、漢字用の入出力装置の開発、日米の構文・意味の研究と調査を行うという結論が勧告文として NSF 及び日本学術振興会に提出された。

##### 4.2 再建期前半の研究の流れ

###### 1) MT のシステム<sup>11)</sup>

この時期に出現した二、三のおもなものを掲げる。LOGOS I: 20 年の研究歴をもち、英→ベトナム語の直接翻訳で、米空軍、海軍のマニュアルの翻訳などを行った。EC の英→仏 MT の設計のためにも使用された。modularity の原理 (解析と合成を分離) を用いている。LOGOS III は英→露、英→仏も可能である。

CETA: グルノーブルのシステム (前述) で、依存木の中間語表現を介する露→仏の翻訳を開発した。この中間表現は語の意味分解まではやっていない。1967 ~ 1971 年の間に 40 万語の文でテストされた。中間言語方式は原言語の情報が落ちるため適訳が得にくいことを指摘している。1971 年終了。

Lehmann らの Texas 大学のシステム (METALS): CETA と同じ頃開始、中間言語方式で、独→英翻訳であるが他の言語間にも適用できる。CETA と同様、構文の中間語を使用する。

SLC-II: CETIS (EURATOM の研究センタ) で表層構文構造表現の間の変換としてトランスファ方式を導入したもの。最近では自動インデックスシステムとしての実験に集中している。

TAUM プロジェクト: トランスファ方式の典型である。1962 年以来、Montreal 大学がカナダ政府の援助で英→仏翻訳の自動化を行っているもので最も実用に近い第 2 世代システムである。翻訳の各段階は Q-system の文法で構成されている。

KULTS<sup>16)</sup>: 九大が 1965 年から始めたもの。表層依存構造 (D-tree) の変換を経由する英→日 MT 実験システムで、我が国最初のトランスファ方式による翻訳である。

その他: 京大<sup>17)</sup> でも英→日、日→英の実験が行われた。また CLRU の M. Masterman のように会話的にやる方式もある。後編集の困難を避けるため、十

分な前編集を対話でやる。対話により訳語構文に対応した形に直す。

## 2) 解析の方式

この時期に特に関心を持たれたものは生成変形理論、格構造、意味表現などである。特に意味の問題が取りあげられ始めたのはこの時期である。Fillmore<sup>18)</sup>の格の考えは、元来依存構造の関係からきている。依存関係は意味的关系を表し、ある動詞を中心にして各格がもつべき役割が決まれば文の意味が決まる。

次に意味の把握の仕方については多くの考え方が出てきたが、大別して二通りある。1つは内包的な捉え方で端的に言えば単語の意味を成分に分解することにあたる。画像の認識などのときこの要素的情報を認識すればその単語の概念を把握したことになるので図形記述には都合がよい。この考えで動詞そのものの意味を分類することもできる<sup>34)</sup>。その意味は格関係の意味と内容的意味からなっている。Chomskyは動詞そのものの意味は深く考えてない。つぎに意味の捉え方の他の1つは外延的な捉え方で、ある語の意味は他の概念との結合関係の全体であるとするものである。Quillian<sup>19)</sup>の意味ネットワークの考え方もこれに属する。九大(栗原)もこの関係を徹底させた<sup>33)</sup>。

自然言語による対話の立場では文の解析よりも直接内容の把握をめざすので、意味情報を取り入れた解析が考えられた。この時期の代表的なものとして拡大遷移網(ATN)(W. Woods<sup>22)</sup>、Q-system(Colmerauer<sup>21)</sup>)がある。

## 3) 機械支援型翻訳<sup>1)</sup>

1965年来、ドイツ連邦翻訳局は翻訳者のための辞書システムLEXISの作成に着手、またBrusselの石炭鉄鋼連盟では1963年以来、データバンクDICAUTOMを作成、いずれも分からない単語だけ線をひいて機械に照会する方式で、人が辞書をひいて翻訳するのと比べ、50%以上の能率をあげている。

## 5. 再建期後半(1971~1975年頃)(再建拡充期)

本格的に意味処理の時代に入ったといえる。前半期のものの多くはさらに進展を続けているが、ここでは特に注目されるものだけに留める。

### 5.1 MTシステム概要<sup>1)</sup>

#### 1) MTシステム

SYSTRAN: P. TomaがGeorgetownのアプローチの大幅な改良版として設計したもので、方式は直接

翻訳(第1世代)である。露→英の翻訳で、米国空軍外国技術部は1970年にMark IIからこれにおきかえた。NASAでアポロソユーズ計画の間使用された。EURATOMでも1976年にこれにおきかえた。

TAUMプロジェクト: 前半期に引続き活発に進められている。また応用として1974年からTAUM METEO(天気予報のMT)を開発し、1976年実働に入った。完全自動化した最初の実用第2世代システムである。文法はQ-systemを採用している。また1978年完成予定でAVIATION計画(航空術のMT)を始めた。

GETA: グルノーブルで1971年計算機システム変更を契機として新たに開発したシステムである。CE-TAの欠点であった、テキストの意味表現への変換の不利な点、目的言語の合成に必要な表層情報の脱落、多言語・プログラム両面の汎用性の問題を解決したトランスファ方式である。

Leibnizグループ: 多言語間MTシステムを開発する国際協力グループとして1974年設立されたもの。GETAのアルゴリズムを中心としている。

国内: KULTS(九大)では中学校の英語テキスト(一種類全部)と高校英語を翻訳した<sup>23),24)</sup>。九大(栗原グループ)では意味的基本構文の関係を用いて構文対構文の変換を行う日→英MT実験を行っている<sup>35)</sup>。

#### 2) 意味論を基礎とする翻訳

Stanford大学ではWilksのMTモデルを用いたまったくの意味的アプローチによる英→仏のMTを考えている<sup>25),26)</sup>。これは言語理解の過程を加味したものであるから中間語的MTである。このモデルの特色は、意味的特徴全体に操作が及ぶ意味的パーサを用いること、優先意味論(preference semantics)、常識的推論を行うルールの使用、操作が文の境界をこえて及ぶ解析などである。文脈解析や知識の利用を含むことは望ましいが、MTの幅広い領域での知識との協調は困難との批判もある。

#### 3) 会話的MTシステム

人が支援してMTを行うシステムは有望である。たとえばCULT(Chinese Univ., Hongkong)は1975年以来、中国の数学雑誌の英語翻訳で実働している。会話型MTの唯一の実用システムと思われる<sup>1)</sup>。直接翻訳方式で、人間が冠詞の挿入、Voice、時制の決定などを行い、あいまいさを除去する。間違った解析に対しては人がその1文を訳す。MINDシステムはRAND Corp.のM. Kayが開発した汎用言語処理用の研

研究室システムで、会話型あいまいさ除去機能を持つ<sup>27)</sup>。

#### 4) 翻訳者のための機械支援<sup>1)</sup>

実働システムが多く継続的に専門用語データベースの充実が進められている。EURODICAUTOM (EC), TERMIUM (カナダ), LEXIS (西ドイツ連邦言語局), TEAM (Siemens 社), NORMATERM (AF-NOR フランス語標準化協会) などがある。

#### 5.2 意味解析

最近の能率よい柔軟な構文解析のパーサとして二つのタイプがある。いずれも能力においては変形文法に等価である。1つは非制限パーサで、これは TAUM の Q-system のように柔軟な書き替え規則に基づいたもので、構文解析、形態素解析、合成の各部分で使用され、GETA ではこのタイプのものが構文解析、意味解析、合成過程で用いられている。その他に MIND システムの powerful parser とか、Winograd のパーサ (PROGRAMMAR) などがある。MT に有利な点はいつでもなんらかの解析を生ずること、受理器でなく変換器として動くということである。

他の1つは拡大遷移網 (ATN) に基づくパーサで、これは前にのべた W. A. Woods<sup>22)</sup> によって開発されたもの。(詳細は本会誌に優れた解説がある<sup>4)</sup>のでそれに委ねる。) その柔軟性のため多くのシステムがこの方式を採用している。TAUM AVIATION では構文と形態素の解析に、Simmons<sup>28)</sup> は構文と意味の解析に用い、他の所でも多く構文解析に使用している。PLATON (京大) は Q-system と ATN の考え方を結合したものである<sup>30)</sup>。

Schank の方法<sup>29)</sup> は概念依存理論に基づく深層レベルの意味表現を得るための解析である。解析が必ずしもアルゴリズムとして明確でなく、これが MT の中間語として適当かどうかは分からない。筆者のこれに近いモデルでの経験では多くの場合、深入りしすぎる傾向となる。

Montague 文法は範疇文法により構文解析された文の構造と対応づけた述語論理表現により構文の意味を与えることになる<sup>30)</sup>。まだ問題が多い。

日本語の意味解析に関して九大 (栗原, 吉田) では入力文をいったん標準形に分解し<sup>37)</sup>、これを意味的基本構文でチェックし、さらにそのときの環境条件 (スクリプトなど) のもとで文の意味をきめるという方法を開発した。意味的なかかりうけの調査、助詞の意味、語の意味の調査資料は膨大である<sup>33)</sup>。岡田らはこれと同様の意味的基本構文を得る動詞の内包の意味分

類を行っている<sup>34), 35)</sup>。

#### 5.3 知識・文脈情報の導入

T. Winograd は、限られた世界ではあるが、文脈情報、語用論的情報をはじめて明白に文章解析に取り入れた文脈解析システムを作った<sup>31), 32)</sup>。

京大の日本語解析システム<sup>39), 40)</sup> は文脈情報を短期記憶として取り入れて、照応表現 (指示詞) や省略の解決をはかり、また次に生ずるものを推論する知識を導入して次の文の解析に役立てている。

Schank のスクリプトや Minsky のフレームは文脈解析のための常識領域の知識として、ある典型的な物語の流れを表現するデータ構造である。九大 (栗原, 吉田) は Schank より早く環境・行動をとり入れた解析を提案していた<sup>33)</sup>。

### 6. 発展期 (1976 年以後) の展望と今後

最近の概況は本号の各題目について、それぞれの分野で触れられることになるだろうからここでは総括的なことをのべるにとどめる。

ALPAC レポートから 10 年後の 1976 年、アメリカで FBIS Seminar on MT が開かれたが、MT についてはまだ実用化の前に基礎研究を積重ねる必要があるという全般的な合意があったという<sup>42)</sup>。一方で同じ年、EC の委員会の報告では、委員会が実際に MT システムを導入し、かつ MT に関する研究プロジェクトを援助する意図のあることを明らかにした。現実の技術的背景は近い将来、実用可能の状況にあるとの見通しに立ったものと言えよう。

#### 6.1 MT システム<sup>1)</sup>

最近の多くの MT システムの中で前にのべた、SYSTRAN, TAUM, 及び GETA はその後もすぐれたシステムとして発展している。

日本語処理の研究は他の国語に比べて遅れていたが、最近文部省特定研究などで急速に進展している。特に 1980 年代になって第 5 世代コンピュータの関連もあって大学、研究機関の他、多くの企業が積極的に取り組んでいることは大きな変革といえよう。

#### 6.2 文の解析について

文の解析の各レベルは最近では意味・文脈の導入、質や能率の向上を目指して独立でなくなり、単語辞書と構文処理、構文解析と意味解析、意味解析と文脈解析などのレベルの区別がしにくくなっている。

1) 形態素解析: 言語処理においては何をやるにもまず必要な段階であり、それぞれのシステムの中で同

じょうなことをやるのが実情だろうが、その中で TAUM と GETA のアプローチは現在典型的な方法と言えるであろう。

2) 構文解析: 日本語については意味的な考慮が必要なのでまだ手のつかない問題が多い。欧米語については解析の戦略と文法的モデルは次第に能率のよいいくつかのパーサにしばられてきた傾向がある。最近のパーサの一般的特徴は言語データとアルゴリズムが明確に分離されていることであろう。日本語の解析には格構造を解析する格文法を用いる方が適しており、語順の影響を受ける ATN はあまり有効でない<sup>43)</sup>。

3) 意味解析: MT で重要なことはあいまいさの解決である。しかし、一般に必要な意味解析の範囲は不明確である。SYSTRAN や TAUM のようにかなり実用に近いものでもある程度以上の意味的特徴は用いていない。複合語、並列句、訳語の選択などの問題に意味がどの程度必要なのか、またどの程度有効なのかまだ明確でない。

4) 文脈解析: MT では照応表現(代名詞、不定冠詞など)の解明などに少なくとも文脈の知識が必要であり、文間の意味的關係を解釈するためにフレーム、またはスクリプトを用いることは MT の場合にも有用なことが期待されるが、これも有効性についてはよく分からない。

### 6.3 今後の展望

最近の 10 年間は発展のさなかにあって、この間にかなりの変革が起こっている。すなわち

- ① MT の社会的要請の高まり
- ② 処理機械の進歩
- ③ ソフトウェアの進歩
- ④ 知識工学の発展
- ⑤ 研究者の増加
- ⑥ 日本語処理の進歩

などである。処理アルゴリズムとしては前述のように形態素解析は大体方法が固まってきたという認識であるし、構文解析も問題点が明確になったといえる。意味処理についてもその必要性が明らかにされ、その戦略に見通しを立てることが出来る段階にあるといえるだろう。知識の導入については機械翻訳の立場と目的によってその必要性と範囲が異なるであろうが、分野を限定するなどすれば導入の範囲をきめることができるであろう。総じて質と目的によっては、現状でも充分実用可能という場合もあろう。現実に SYSTRAN や TAUM METEO はそれぞれまだ不十分ながら第

1 世代、第 2 世代の代表的な実用システムであるといえる。もし十分な質を望むなら機械支援システムの体制でやれるということも言える。

今後特に展開の必要性が感じられる問題としては、

- 1) 専門用語辞書の充実、2) 日本語解析の精密化、3) 意味・文脈処理、4) 強力な解析ソフトウェア、5) 入出力、能率のよいハードウェア、などがある。また開発促進のための施策として、1) ある程度方法の固まったものの標準化、2) 開発体制(国家的、または分野の協力、国際協力など)、などを必要に応じてとるようなことが望まれる。

今後の 10 年は実現期といえるのかどうか。それを期待したい。

最後に機械翻訳の基本問題として

- ① 機械翻訳の工学的目標は何か?
- ② それに伴って人工知能との関係は?

という疑問がある。つまり MT は知識の人間の代役なのか、人間的(情緒的)人間の代役なのかという問題である。翻訳の場合、機械はあくまでも客観的でなければならないのか、あるいは人間らしい出力を求められるのかということである。MT の社会的要請も研究者の態度もその点をはっきりしていないように見受けられる。つきに知識の利用について言語処理の立場と人工知能の立場の違いについてである。われわれは言語を通じて知能処理をやっているのであり、言語処理はコミュニケーションとしての必要な知的処理をやっているのである。知能処理は自然言語を用いなくてもよい。コミュニケーションの道具としての言語は知識を学習することができる。

本稿は紙面の制約でおもな史実を羅列するにとどまったが、正確でない点もあると思われる。大方のご叱正を乞う次第である。参考文献についてはおもにサーベイを引用し、詳細はそれぞれの中の文献を参照して頂くようにした。

### 参考文献

- 1) Hutchins, W. J.: Progress in Documentation, Journal of Documentation, Vol. 34, No. 2 (1978).
- 2) 玉木英彦, 喜安善市編: 自動翻訳, みすず書房(1960).
- 3) Delavenay, E.: An Introduction to Machine Translation, Thames and Hudson, London (1960).
- 4) Weaver, W.: Translation, in 9)
- 5) 情報処理, Vol. 5, No. 5 (1964).

- 6) 情報処理, Vol. 6, No. 1 (1965).
- 7) 和田 弘: 機械翻訳の現状と将来, 情報処理, Vol. 6, No. 3 (1965).
- 8) Languages and Machines: Computers in Translation and Linguistics, A Report by the Automatic Language Processing Advisory Committee, Division of Behavioral Sciences, National Academy of Sciences, National Research Council, Washington D.C. (1966).
- 9) Locke, W.N. and Booth, A.D.: Machine Translation of Language, MIT Press (1955).
- 10) Tamati, T., Kurihara, T. and Yoshimura, A.: The Translation Process and the Design of an Automatic Translator as an Information Processing Machine, Information Storage and Retrieval, Vol. 1, No. 1 (1963).
- 11) 田町常夫: 言語の機械翻訳(II), 信学誌, Vol. 46, No. 11 (1963).
- 12) Kuno, S. and Oettinger, A.G.: Multiple-path Syntactic Analyzer, in M. Poplewell: Proc. of the IFIP Congress 62, North-Holland Pub. (1963).
- 13) 田町常夫: 機械辞書の自動作成の一方法, 情報と制御の研究, No. 6, 電子通信学会 (1964).
- 14) 情報処理, Vol. 9, No. 3 (1968).
- 15) 情報処理, Vol. 9, No. 5 (1968).
- 16) 田町常夫: 機械翻訳の現状, エレクトロニクス, Vol. 13, No. 4 (1968).
- 17) 坂井利之, 杉田繁治: 機械による英和翻訳, 信学誌, Vol. 49, No. 2 (1966).
- 18) Fillmore, C.: The Case for Case, in Bach and Harms eds.: Universals in Linguistic Theory, Holt, Rinehart and Winston, N.Y. (1968).
- 19) Quillian, M.R.: Semantic Memory, in M. Minsky ed.: Semantic Information Processing, MIT Press (1968).
- 20) 岡田直之: 自然語の意味情報とその抽出および分類, 信学論, Vol. 52-C, No. 10 (1969).
- 21) Colmerauer, A.: Les Systemes-Q ou un Formalisme pour Analyser et Synthetiser des Phrases sur Ordinateur, in TAUM 71.
- 22) Woods, W.A.: Transition Network Grammars for Natural Language Analysis, CACM, 13 (1970).
- 23) 石原好宏, 田町常夫: D-tree モデルとそれに基づく英日機械翻訳のための言語分析について, 信学論, Vol. 57-D, No. 7 (1974).
- 24) 石原好宏, 田町常夫: D-tree モデルに基づく一つの英日機械翻訳システムおよび実験, 信学論, Vol. 57-D, No. 7 (1974).
- 25) Wilks, Y.: The Stanford Machine Translation Project. in Rustin, R. ed.: Natural Language Processing, Algorithmics Press, N.Y. (1973).
- 26) Wilks, Y.: An Artificial Intelligence Approach to Machine Translation, in Schank, R.C. and Colby, K.M. eds.: Computer Models of thought and Language, Freeman, San Francisco (1973).
- 27) Kay, M.: The MIND System, in Rustin, R. ed.: Natural Language Processing, Algorithmics Press, N.Y. (1973).
- 28) Simmons, R.F.: Semantic Networks: Their Computation and Use for Understanding English Sentences, in Schank, R.C. and Colby, K.M. eds.: Computer Models of Thought and Language, Freeman, San Francisco (1973).
- 29) Schank, R.C.: Conceptual Information Processing, North-Holland Pub. (1975).
- 30) Hobbs, J.R. and Rosenschein, S.J.: Making Computational Sense of Montague's Intentional Logic, Artificial Intelligence, 9 (1978).
- 31) Winograd, T.: Procedures as a Representation for Data in a Computer Program for Understanding Natural Language, MIT Report, MAC TR-48 (1971).
- 32) Winograd, T.: Understanding Natural Language, Academic Press, N.Y. (1972).
- 33) 栗原俊彦: 自然言語の機械処理, 情報処理, Vol. 14, No. 4 (1973).
- 34) 岡田直之, 田町常夫: 自然語及び図形解釈のための単純事象概念の分析及び分類, 信学誌, Vol. 56-D, No. 9 (1973).
- 35) 岡田直之, 田町常夫: 自然語及び図形解釈のための非単純事象概念の分析及び分類, 信学誌, Vol. 56-D, No. 10 (1973).
- 36) 首藤公昭: 専門分野を対象とした日英機械翻訳について, 情報処理, Vol. 14, No. 9 (1974).
- 37) 藤田 毅, 鶴丸弘昭, 吉田 将: 日本語の機械処理——日本語文の標準形分解, 信学論, Vol. 58-D, No. 7 (1975).
- 38) 長尾 真, 辻井潤一: 自然語処理のためのプログラミング言語 PLATON, 情報処理, Vol. 15, No. 9 (1974).
- 39) 長尾 真, 辻井潤一, 田中一敏: 意味および文脈情報を用いた日本語文の解析——名詞句・単文の処理, 情報処理, Vol. 17, No. 1 (1976).
- 40) 長尾 真, 辻井潤一, 田中一敏: 意味および文脈情報を用いた日本語文の解析——文脈を考慮した処理, 情報処理, Vol. 17, No. 1 (1976).
- 41) 長尾 真, 辻井潤一: 自然言語処理プログラム, 情報処理, Vol. 18, No. 1 (1977).
- 42) 長尾 真: 言語情報処理の過去・現在・将来, 情報処理, Vol. 19, No. 2 (1978).
- 43) 長尾 真: 機械翻訳, 信学誌, Vol. 65, No. 4 (1982).

(昭和60年9月10日受付)