

機械翻訳論の試み

石原好宏*

A Step toward a Theory of Machine Translation of Natural Language

Yoshihiro ISHIHARA

Abstract

Machine translation of natural language (MT) was one of the main subjects of study for those who were interested in the applications of the electronic computer just after it was put to practical use. Various kinds of researches and experiments had been performed up until 1966, when the ALPAC report was published in the U.S.A. . Then the main interests of the related researchers shifted from MT to subjects such as machine understanding of natural language, question-answering systems, or more basic researches of natural language.

But recently, interest in MT seems to have returned. The author believes at this juncture that a thorough examination and reconsideration of the results and the controversial points of the past researches should be performed at the outset, and that new guiding principles in MT research should be established.

This paper, therefore, outlines the underlying theories and methodology of MT used by various researchers (including myself) in the past. The intrinsic problems of translation of natural language, the differences of conditions and environments between translation performed by human beings and those performed by machines are also discussed. Finally we briefly discuss goals for the future of MT, and problems to be solved in order to approach those goals.

1. ま え が き

自然語の機械翻訳の研究は、電子計算機が実用化された1940年代後半にその応用研究の一環としてアメリカとイギリスを中心に始められ、1960年代半ばまでは世界各地で様々な試みが行われた^{1),3),4)}。しかし、期待されたほどの成果が得られないまま ALPAC レポートがまとめられ⁴⁾、それを契機として特にアメリカにおける組織的な研究はほとんど中止されてしまった。その後はごく一部で小規模で地味な研究が続けられたものの、大方の言語学者・計算機学者・情報科学者達の関心は薄らいでいった³⁾。こうして自然語の機械処理に関する研究者の関心の中心は機械翻訳から自然語の意味理解や質問応答系などの問題に移っていった。そしてその方面では T. Winograd の SHRDLU システムを始めとする種々の実験システムが試作され、多くの経験や実験データが蓄積されていった¹⁷⁾。このよ

うに地道な研究の積み重ねが一方にあり、他方では政治的経済的活動の広域化や国内の複雑な言語事情などに起因する切実な要請もあって、特にヨーロッパやカナダなどで実用的な機械翻訳システムを実現することが最近再び大きな関心呼び始めている^{3),8),20)}。

一方、日本における機械翻訳の研究は、1955年に九州大学で、1958年に電気試験所（現在の電子技術総合研究所）で開始されたことに始まる。欧米に比べて研究の開始がいくらか遅れたこともあって、日本では逐語訳の時代は経験せず、いきなり構文翻訳に入った。しかし、当時の日本はまだ電子計算機がほとんど普及していない状況であった。このため、機械翻訳向けの自然語の文法・翻訳アルゴリズム・機械辞書の形式などの研究と同時に、機械翻訳専用の電子計算機的设计・製作も並行して行われた。その初期の成果が九州大学においては自動翻訳機 KT-1⁷⁾ と英独日3カ国語間相互同時翻訳¹⁵⁾ であり、電気試験所においては翻訳専用機ヤマトと英文和訳のプログラム¹⁶⁾ である。その後京都大学などでも英文和訳、和文英訳等の研究が開

* 工業短期大学部 情報処理工学科

始され、汎用の電子計算機を使った実験も試みられはじめた。しかしやがて海外の動向に合わせるかのように、大方の関心は自然語理解その他に移っていった。ところが最近日本でも再び機械翻訳についての具体的な試みが取り上げられるようになってきた¹⁸⁾。

機械翻訳に関するこのような状況を、失敗の繰返しに終らせず健全な発展に導くためには、過去の研究成果と問題点をこの際十分に検討し反省した上で、今後の研究の指針を改めて設定しておくことが肝要であろうと思われる。その核は機械翻訳の課題の本質を見極めることであり、その線に沿った幾つかのサブゴールを設定しそれらを実現するための具体的な方策を検討する必要がある。

そこで本稿では、最初に、従来の機械翻訳研究において多くの研究者が準拠した翻訳論と機械翻訳の方法論を、筆者の KULTS-I^{13),14)} の場合と合せて概観する。続いて自然語の性質と特徴を、E. Nida の分析⁵⁾ を参考にして、翻訳の観点から概括する。更に、翻訳の課題、それを人間が遂行する場合と機械に遂行させる場合の条件や環境、などについて簡単に述べる。最後に、今後の機械翻訳の目標とそこへ近づくために解決すべき問題点について若干の考察を試みる。

2. 従来の機械翻訳研究の背景に流れる翻訳論

新しい試みを始めようとする場合、何を考慮に入れ何を無視するかということとは実際にやってみなければ分らない面が多いことは確かである。しかし同時に、できる限り周到な基礎調査とその結果の分析および検討を事前に行い、ある程度の予測を立てておくことが必要なことも確かである。ところで、このような事前の基礎調査や分析が従来の機械翻訳研究の中でどの程度行われてきたのであろうか。多くの場合、当面利用する言語理論や言語処理技術の適用可能範囲、適用した結果や問題点、などはその都度調べられ検討されてきた。しかし、そのような当面の調査・分析が、機械翻訳研究の究極の目標に対してどのような関連性と意義をもつのかといった点に関しては必ずしも十分な検討や議論が行われてきたとは言いがたい。このような状況では、個々の研究結果の真の意義が見落され、又、今後の研究の展開の方向も見失われることになりかねない。過去の機械翻訳研究において研究内容の理解や評価が必ずしも十分にそして適切には行われていない原因の一つにはこのようなことが考えられる。

そこで本章では、従来の機械翻訳研究の背景に見られる考え方の傾向やその変遷の概略をたどり^{3),19)}、機

械翻訳の目標を改めて設定するための一助としたい。

2.1 機械翻訳の目標と方法論の変遷³⁾

機械翻訳の試みが開始された当初は、暗号解読術などから連想される比較的単純な方式の翻訳が試みられたようである²⁾。しかしそのように単純でしかも人手を全く介入させない翻訳方式では品質のよい訳文はとも望めないことが問もなく知られるようになり、改めて二つの相反する目標が設定されたようである¹⁾。その一つは完全な自動翻訳を目指すものであり、もう一つは高品質の翻訳を目指し、必要ならば人手の介入も許すというものであった。前者は文体的には不適切であったり受入れ難い訳文であっても、原文の意図した意味がくみとれるならばそれでよしとする。これに対して後者は、人間と機械が互いに助け合って品質の高い訳文を作り出そうとするもので、人間の翻訳者を支援する言語データバンクとそれを会話的に利用しながら翻訳作業を進めるマン・マシン・システムを作ることが主題となる。前者を狭義の機械翻訳 (Machine Translation; MT)、後者を機械援助翻訳 (Machine-aided Translation; MAT) と呼ぶ。但し、今後本稿の中で単に機械翻訳といえば狭義の機械翻訳を意味する。

機械翻訳の設計思想は1966年前後を境にして大きく変化している。それ以前の機械翻訳ではほとんどの場合、特定の言語対ごとに ad hoc な処理を考えていたと言っても過言ではない。その場合原始言語をどの程度どのように分析するのかが目標言語および訳文に対する要求に応じて勘案されたと言ってもよい。この方式を“直接翻訳法” (Direct Translation Approach)、あるいは“第一世代の機械翻訳”と呼んでいる。これに対してその後の機械翻訳では原始言語の分析と目標言語の合成を切り離して考えることが多くなってきた。この場合原始言語の文章は、共通の中間言語を介して、あるいは原始言語と目標言語のそれぞれの深層構文構造又は意味表現の間で意味的に等価なものを対応づける移し換え操作 (Transfer Operation) を行って、間接的に二カ国語間の翻訳を行う。この方式を“間接翻訳法” (Indirect Translation Approach)、あるいは“第二世代の機械翻訳”と呼んでいる。

2.1.1 第一世代の機械翻訳 (直接翻訳法)

機械翻訳研究のごく初期には対訳辞書を用いた“語対語”翻訳がまず試みられ、やがて構文情報を利用した翻訳 (構文翻訳) が主流となった。しかしいづれにしても、言語分析が基本的には試行錯誤的であったこと、しっかりとした理論的裏づけが弱かったこと、後編集 (post-editing) に強く依存していたこと、など多

くの問題点を含んでいた。又、意味や構文の問題点を解決しようとするが大抵辞書内容が膨張し、しかも複雑になるので、結局一度作られた翻訳システムの改良はほとんどできなかった。従って、意味処理は事実上あまり行われず、又、多義処理などの困難に直面してもその処理は直接には行われず、考えられる解を全部出力して人間の判断に委ねるといった処置がとられることが多かった。しかし、このような状況を来たした原因は自然語に対する人間の理解の不十分さであることが次第に認識されるようになり、当面の問題は自然語そのものをより深く知ることであると考えられるようになった。

この方法論による機械翻訳システムとしては、アメリカ空軍の海外技術部門 (FTD) に1964年に設置され1970年まで稼動した Mark II (露英翻訳)、アメリカのオークリッジ国立研究所原子力エネルギー委員会とイタリア・イスプラの EURATOM に1964年に設置され最近まで利用されたジョージタウン大学の露英翻訳システム、その改良版である SYSTRAN (1970年に FTD が Mark II から、又1976年に EURATOM がジョージタウン大学システムから切り換えて使用。又 NASA もアポロ計画で利用)、更に SYSTRAN の英仏翻訳版 (EC が試用) などがあった。

2.1.2 第二世代の機械翻訳 (間接翻訳法)

たとえば言語系統が同じか近いものでも文法構造や意味の表現法は自然語ごとに微妙に異なり一様でない以上、異なる言語の表現形式の間に直接の対応を求めようとした第一世代の翻訳法には本質的な無理があったと言えよう。従って、個々の自然語の枠内で十分な分析と合成を行うことが機械翻訳のためには必要であると考えられるようになったことは当然であったと言えよう。その場合最大の課題は異なる自然語の間の媒介となる中間言語を設定することであり、更にそれに附随して各自然語とその中間言語を対応づけるための各自然語の分析および合成が問題となる。これが第二世代の機械翻訳における主な研究テーマとなる。ところでこのような中間言語を介した機械翻訳の考え方は既に1949年の W. Weaver のメモの中に見られると言う³⁾。しかし今日なお真の中間言語を介した機械翻訳の試みはないようである。ただ、このことと深い関連があると思われる二つの方式が試みられている。一つは特定の自然語に共通な表現形式を設定しそれを中間言語と考える。その場合翻訳処理は、原始言語の表現をまず中間言語へ移し換え、それを更に目標言語の表現に移し換える二段階方式となる。もう一つは原始言語の枠内の文法構造または意味構造から目標言語の

それらへの移し換え操作で自然語間の対応づけを試みる。その場合翻訳処理は原始言語からその文法構造または意味構造を抽出する段階、それら为目标言語の文法構造または意味構造へ移し換える段階、更にその結果を使って目標言語の文章を生成する段階の三段階から成る。

いずれにしても、直接翻訳法と比べた場合の最大の特徴は、それぞれの自然語がもつ性質や特徴をできる限り個別に考慮した分析や合成をしようとしていることである。以下に間接翻訳法として試みられてきた二つの具体的方式とその特徴を概観する。

〔1〕 中間言語を用いた翻訳

“pivot 言語”を中間言語とするフランス・グルノーブル大学自動翻訳研究センター (CETA) の露仏翻訳システム、N. Chomsky の提唱する深層構造を中間言語とするアメリカ・テキサス大学の独英翻訳システム METALS が代表的な実例と考えられる。このうち、pivot 言語はその構文が自然語の構文の共通の基礎となるように設計されているが、語彙に関しては対応する二カ国語の語彙項目を結びつけるだけである。更に、同じ意味をもつ複数の文から共通なただ一つの pivot 言語表現を得るようには設計されていない。一方、テキサス・グループの深層構造による中間言語も、CETA の pivot 言語の場合と同様構文に関しては一種の中間言語となっているが、語彙の意味に関してはそのような配慮がなされていない。しかも構文分析では文脈自由文法を使っているため構文の多義性に対する処理が十分にはできない。

以上の二例はいずれも中間言語を設定するに当たって構文と意味を分離して考え、しかも後者に対する具体策は提示していない。これに対して、機械翻訳を本来の目的として考案されたものではないが、R.C. Schank の Conceptual Dependency Theory (CD 理論) は中間言語による自然語の機械翻訳の一つのひな型を示していると言える⁶⁾。彼のモデルでは、原文を概念分析してその意味内容を概念依存表現として抽出する。続いて概念依存表現からそれを表す自然語文を生成する。その際自然語が原文のそれと同じであれば単なる言い換えであり、異なっていれば機械翻訳となるわけである。ここで概念依存分析と文章生成のための処理アルゴリズム及び機械辞書は自然語ごとに一種類ずつ準備すればよい。従って、概念依存表現は特定の自然語からは独立した一つの概念構造記述言語すなわち中間言語としての役割を果す。ただここで実用的な翻訳システムの実現性の観点からこのモデルを観ると、概念依存に関する機械辞書を実用に耐えるように準備しよ

うとすれば膨大なものになることが予想され、しかもその内容を確定することはかなり困難であると考えられる。従って、R. C. Schank のモデルもそのまま実用的なシステムに拡大して適用することは困難であろう。しかし興味深いモデルの一つであることは確かである。

〔2〕 移し換え操作を用いた翻訳

この方式による機械翻訳の最初のもつた報告は V. Yngve が1957年に行っており、その中では移し換えを構文の表層構造表現の間の操作と考えていたと言う³⁾。しかもこの考え方は、イタリア・イスプラの EURATOM 研究センターが1975年まで実験的な開発を続けた機械翻訳システムの基本概念でもあったようである。しかし、表層構造より深い分析を全く行わない機械翻訳は一般には不十分であり、見当違いの訳を出しかねない。従って最近では、移し換えは深層構造あるいは意味構造のレベルで行うべきだとする考え方が多く見られるようになってきた。しかし他方では、大がかりな意味分析は必要でなく又望ましくもないとする考え方も、特に構文や語彙の類似する言語の間で機械翻訳を考える場合には依然として存在しているように思われる。

カナダ・モントリオール大学の TAUM プロジェクトは移し換え方式を用いた典型的な例と言える。しかもほぼ完全に実用化された最初の第二世代の機械翻訳システムであって、1976年以来毎日1,500~2,000の短い天気予報文の英仏翻訳を一般向けに行っているという (TAUM-METEO)³⁾。

移し換え方式の特徴としては、処理段階が明確に分離されていること、機械辞書も原文分析用、移し換え用、訳文合成用がそれぞれ独立していること、原文分析は深層構造より深くは行わず、必要がなければ深層構造にさえ達しないこと、意味分析は木構造の変換に必要な範囲に限って行うこと、などである。

移し換え方式を用いた例としては上記の外に、アメリカ・カリフォルニア大学バークレー校の中英機械翻訳システム POLA⁹⁾ などもある。

この外に注意すべきこととしては、フランス・グルノーブル大学が1971年に電子計算機システムを変更したのを機会に、翻訳方式も中間言語方式から移し換え方式に切り換えたことがある。切り換えた理由は、中間言語方式の場合原文からその意味表現を作るときに問題があるのと、訳文を生成するとき有用な原文の表層構造の情報が相当程度失われることが分ったからであるという。もっとも、この場合の真の原因が中間言語方式の不備にあったのか、それともその際に設定

された中間言語そのものの不備にあったのかという点については検討してみる余地が残されているように思われる。特に、十分な実用性が未だ保障されていないとは言え、CD 理論を軸とした R. C. Schank らの一連の試みなどと比較照合してみると、グルノーブル大学の中間言語方式の欠陥の原因は、方式そのものよりもむしろ中間言語の設定の仕方により大きなものがあったのではないかとさえ考えられる。ともあれ、GETA (フランス・グルノーブル大学機械翻訳研究グループで CETA を改称したもの) は、移し換え方式を採り、又処理アルゴリズムの作成に当っては当面の問題処理に必要な最少限の情報だけしか利用しないとす立場から、着実に研究データを蓄積してきた。その結果、研究はまだ初期の段階にあるとは言え、将来の機械翻訳研究、中でもヨーロッパが当面する多言語間翻訳の協同開発のため一つの基礎を築いたと考えられている^{3), 8), 20)}。その例証として次の二つの動きをあげることができる。一つは1974年にグルノーブル大学を中心にヨーロッパに国際的なライプニッツ・グループが組織され機械翻訳の研究活動が続けられてきたこと、もう一つは1979年以降ヨーロッパ共同体の中で実用的な多言語間機械翻訳システムを開発するための研究プロジェクトを発足させる動きがあることである^{3), 20)}。

2.1.3 意味情報に基づく機械翻訳

2.1.1と2.1.2で述べた機械翻訳は、2.1.2の〔2〕で言及した R. C. Schank の例を除いて、いずれも基本的には構文情報に基づいた処理を行っている。中には原文の意味解釈を含む翻訳処理を考えた例も見られるが、その場合でも、原文に対して直接に意味解釈の処理を行うのではなく、まず構文分析をある程度まで行い、その結果に対して意味解釈を行おうとしている。又、その場合の処理対象としては高々一つの文の範囲を限度と考えており、文の枠を越えた意味的特徴などはほとんど考慮されていない。文の枠を越えて処理が行われた例としてはわずかに代名詞の照応問題などがあるが、これとて必ずしも意味的特徴を十分に考慮したものとは言い難い。

このような中であって、前述の R. C. Schank⁶⁾ や Y. Wilks¹⁰⁾ の意味論的立場からのアプローチには極めて興味深いものがある。この二人は互いに独立に仕事を行っており、自然語の意味の捕え方やその記述の仕方にはそれぞれ独自のものがあり、かなりの相違も見られる。しかし、自然語の処理における基本的考え方には共通する点も幾つか見られる。その根本にある考え方の一つは、“自然語文を分析して求めるべきものは

その構文ではなくて、それらが表現しようとしている意味内容である”ということであるように思われる。そしてこの考え方を具体化する方法論の段階では、以下に述べるような注目すべき共通点をもっている。すなわち、両者とも、人間の言語理解の機構をできる限り反映させた言語処理のモデルを考えようとする人工知能論的接近を試みていること、この立場から基本的には構文情報に頼らず自然語文から直接にその意味内容を抽出しようとしていること、自然語処理に付き物の多義処理には、従来よく試みられた単純な組合せ論的な発想ではなく、意味的特徴に着目した文脈情報やいわゆる一般常識に基づく推論を行う処理アルゴリズムを導入しようとしていること、などである。更に、両者共、文の枠をそれほど意識することなく、必要に応じて単語・節・文章などの意味内容とその構造を特定の自然語には依存しない中間言語で記述することを志向していることも注目してよい共通点であると言える。この両者の試みの基礎にある意味論は R. C. Schank の場合が CD 理論⁹⁾、Y. Wilks の場合が Preference Semantics¹¹⁾ である。これらは完成した理論ではなく、それぞれの立場から更に種々の検討が加えられ、理論の枠組の修正・強化・発展などがその後も続けられている。

ところで、意味処理を機械翻訳処理の中に導入しようとするれば、R. C. Schank や Y. Wilks のモデルでも試みられたように、単なる semantics ばかりではなく、外界情報・一般常識などを必要に応じて活用するための pragmatics をいかに具現化するかも大きな問題となる。この点に関して従来試みられてきた方法論は、限られた小さな世界あるいは少量のデータを対象としている限りは有効であるかも知れない。しかし、対象とする世界を広げたりデータ量を増していくと、処理アルゴリズムやデータの保守管理が難しくなって来るために、いわゆる実用化が困難ではなからうかという心配もなくはない¹⁾。従って今後意味情報に基づく機械翻訳の研究を進めるについては、意味処理に関する基本的なメカニズムを解明していくのと同時に、実用レベルでの実現可能性という点についても十分に留意していくことが肝要であろうと思われる。

2.2 KULTS-I における翻訳論

KULTS-I (Kyushu University Language Translation System, the First Version) は筆者らがかつて作成した英日機械翻訳のための実験システム^{13), 14)} である。KULTS-I を作成する準備としての言語分析、KULTS-I の具体的なシステム構成、実験結果などに

ついては文献 (13) と (14) を参照していただくとして、ここではこの実験システムを設計・製作したときに筆者が念頭に置いていた基本的考え方に焦点を絞って、以下の各問題点についてのまとめと考察を試みる。

- (1) 機械翻訳をどのようなものと考えたのか。
- (2) その立場から機械翻訳を行うために、自然語のどのような性質・特徴を利用しようとしたのか。
- (3) それらを利用すべく設計した実験システムではどのような考え方に基づいて処理アルゴリズムを構成したのか。
- (4) それらがどのような効果ないし影響をもたらし、同時にどのような問題を残しているのか。

2.2.1 基本的考え方と具体的方策

何よりもまず明示しておくべきことは、自然語の翻訳および機械翻訳について、少なくとも KULTS-I の設計の時点で筆者がどのように考えていたかという問題であろう。筆者は、自然語の翻訳とは、一つの自然語(原始言語)で表現されている意味内容を、原則として短縮ないし圧縮(要約の場合には必要)や展開(敷衍や説明などの場合には必要)を行わずに、原始言語とは異なる自然語(目標言語)で表現するための移し換えの作業である、と考えた。そして、この作業を電子計算機に内蔵した言語データと翻訳処理アルゴリズムによって行うのが機械翻訳である、と考えた。この立場に立って機械翻訳システムを作成しようとする場合最初に行うべきことは、次の各点についての我々の方針を明示することであろうと思われる。

- (1) 自然語あるいは自然語文の表す意味内容をどのようなものとして捕えるのか。
- (2) 機械翻訳を目的とした場合それをどの程度まで、又どのような方法で取り扱えばよいのか。

これらの問題に対処するための我々の基本方針が決まれば、着目し利用すべき言語的特徴や性質はおのずから決まるであろうし、電子計算機に内蔵すべき言語データ及び翻訳処理アルゴリズムの大綱もほとんど決ってしまうものと思われる。

そこで我々が KULTS-I を設計するに当たってまず仮定したことは、日本の中学校の英語教科書で取り扱う程度の基本的な英語文とその対訳日本語文の範囲では、自然語文の意味構造(意味内容がもつ構造)を D-tree (Dependency tree)¹³⁾ で近似することができるということである。D-tree は、自然語文中の単語を構文的特徴によってまず“語群”(伝統文法でいう句、または文節—phrase—to ほぼ相当)にまとめ、更に述語語群を中心とした語群間の依存関係をこれも基本的に

は構文的特徴に基づいて調べ、その結果を木構造で表現したものである。すなわち、D-tree は、その葉が語群の品詞情報と単語列から、その枝が依存関係から、その枝の種類が支配語群に対する従属語群の機能要素¹⁴⁾から、その根が述語語群（通常の動詞文、形容詞文、形容動詞文の場合）又は名詞語群（名詞文の場合）の品詞情報、機能要素および変形情報（述語語尾あるいは助動詞の接続で表現）から、それぞれ構成されている。又、D-tree の基本構造は、文の述語語群の核（動詞、形容詞、形容動詞）が必須的に要求する機能要素の組合せであって、機能パターンと名付けている¹³⁾。

以上のことから D-tree は、直接には自然語文の構文構造を示していると考えられる。しかし、日本の中学校の英語文とその対訳日本語文について実際にそれぞれの D-tree を作って構造を対比させると、ごく粗い近似としてではあるが、両者がほぼ同型をなすと見なしてよいことが分かる¹³⁾。このことは、同じ意味内容を表現する英語文と日本語文との間に存在すると思われる共通の意味構造を D-tree が表していると解釈してよいことを意味する。このことから KULTS-I では D-tree を日英両国語における一種の中間言語として利用している。

D-tree がこのように自然語文の意味構造を近似すると仮定した背景には、自然語文の性質に関して、更に次のような細部の仮定を含んでいる。

[D-1] 機械翻訳処理に必要な程度の自然語文の意味内容は、単語および単語間の構文的関係によって表現することができる。

[D-2] 構文的特徴に基づいて設定される語群は、構文上の構成単位であると同時に、意味構造上の構成単位ともなっている。

[D-3] 述語語群を形成する助動詞類、それらの変化語尾類、及びそれらの接続関係が表現している意味は、その語群の主要素である動詞を修飾する働きをもつ語群変形情報として取り出される。しかし、語群間の依存関係の決定に関与することは特でない。

[D-4] 自然語文の意味上の切れ目は構文的特徴による文の切れ目（例えば接続詞、関係詞の前または後の位置）ともなっている。

機械翻訳を主目的として上記のような特徴に着目する自然語の捕え方を我々は D-tree モデルと称している¹³⁾。D-tree モデルに基づく自然語データの中で、語群の形成に関する部分は一部に変形概念を含めた句構造文法で、又語群間および部分 D-tree¹³⁾ 間の依存

関係の形成に関する部分は依存文法で、それぞれ記述することができる。

次に、このような D-tree モデルに基づいて KULTS-I のシステム設計および各部の処理アルゴリズムの作成を行ったときに我々が採った基本方針あるいは設けた仮定のうち主なものをあげておく。

[P-1] 翻訳方式は D-tree を中間言語として介在させる一種の移し換え方式を採用する。なお翻訳過程は、原文から D-tree を抽出する原文分析の段階と、D-tree から訳文を合成する訳文合成の段階の 2 段階方式で構成する。

[P-2] 処理対象の最大限は文とし、文の枠を越えないものとする。その限りでは、複文、重文なども処理対象に含める。

[P-3] 原文分析の際の処理アルゴリズムは、ほぼ完全な bottom-up 方式である。しかも、機械辞書が許す多義は原則として組合せ論的方法によりすべてチェックする¹²⁾。

[P-4] 原文分析の際、原文の各単語・各語群が表す“もの”又は“こと”に関する意味特徴は全く顧慮されない。もちろん、代名詞の指示内容についても全く考慮されない。

[P-5] 日本語の訳語は以下に示すように構文的特徴で区別できる範囲で一意的なものを準備する。すなわち、(i) 述語で自立語の場合；機能パターン毎に一つの訳語；(ii) 述語で附属語の場合；語群変形情報の変形要素²¹⁾ごとに一つずつ；(iii) 非述語で自立語の場合；原始言語の一つの単語品詞ごとに一つの訳語；(iv) 非述語で附属語の場合；D-tree の各枝の機能要素に一つずつ；をそれぞれ与える。

[P-6] 日本語訳文の語順を決定する要因には次の二つがある。一つは語群の訳出順であり、これは D-tree が示す依存関係と各機能要素、更に原文における語群の位置関係を勘案して決定される。もう一つは語群内の構成単語の訳出順であって、これは原文における語群の構文的特徴に基づいて決められる^{14), 21)}。

以上を要するに KULTS-I においては、捕えたい対象は意味内容でありながら、それを直接に捕えるのではなくて、終始自然語文の終端表現に構文的特徴として現れているものを介して間接的に捕えようとしていると言える。このような立場に立った試みが実際にはどのような結果をもたらし、どのような課題を残しているのかについて次節で論ずることにする。

2.2.2 検討および今後の課題

D-tree モデルは、構文的特徴という間接的ではあるが従来の経験から取り扱いの比較的容易な手がかりを通じて、機械翻訳に必要な自然語の意味情報を捕えるための第一近似的な手段である。従って、D-tree モデルに基づく機械翻訳から直ちに実用レベルの結果を得ることは期待できないであろうし、すべきでもないと思われる。しかし、このようなモデルに基づいて実動する実験システムを作ることによって、具体的に直面する種々の事柄を分析・整理し、その中から将来の研究の展開に有用なデータを見出し、活用していくことができるならば、それが最大の成果となろう。

このような立場からここでは、前節に示した D-tree に関する細部の仮定 D-1~D-4 については一括して、又 KULTS-I の作成に関する基本方針・仮定 P-1~P-6 については逐条的に簡単な検討と考察を行う。

〔1〕 D-1~D-4 に関連して

これらの仮定の良否は、自然語文が表す意味構造と構文的特徴とがどの程度並行あるいは対応しているのかに依存する。もし完全な並行・対応関係があるのなら、どれほど複雑な意味構造をもった文の場合でも支障なく構文的特徴によって意味構造が捕えられるはずである。ところが一般的な自然語文をよく観察してみると、そのような完全並行・完全対応の関係はないと考えた方がよさそうである。例えば次の一つの例について考えてみよう。

それは述語語群が幾つかの語群変形要素を伴った場合である。このとき日本語では語群変形要素を助動詞または補助動詞の語幹と語尾を一定の規則で接続して表す。しかも D-tree モデルの基本的な仮定によってこれらの連鎖は一つの述語語群にまとめられる。そしてそれが支配する文中の主語は述語語群の主要素のものとならされる。ところが現実には、以下の例文に見るように、文中の主語が述語語群の主要素のものとは限らない。

(例 1) KARE HA HEYA HE K(・O)・RARE
(・A)・TA(・KU)・NA(・I)・YOU(・A)・
DES(・U)・*

→彼は部屋へ来られたくないようです。

上例の述語語群(~~~~の部分)の分析:

(i) K(「来る」の語幹)の主語;文中には

* 下線部が語群のまとまりを、括弧内は活用語尾を、A は零記号、・は接続を、=は下線部が次行へ続くことをそれぞれ表す。

含まれない。一般に第三者と解される。

(ii) RARE(「られる」の語幹), TA(「たい」の語幹)の主語; KARE(「彼」).

(iii) NA(「ない」の語幹);主語の概念は考えられない。否定を表す。

(iv) YOU(「ようだ」の語幹), DES(「です」の語幹)の主語;この文の作者。ただし日本語文ではそれを陽に文中に表記することは一般にない。

(例 2) KARE HA INU NI ESA WO TABE
(・A)・SASE(・A)・TA(・A)・

→彼は犬に餌を食べさせた。

上例の述語語群の分析:

(i) TABE(「食べる」の語幹)の主語; INU(「犬」).

(ii) SASE(「させる」の語幹)の主語; KARE(「彼」).

(iii) TA(「た」の語幹);主語の概念は考えられない。時の概念を表す。

(例 3) KARA HA DOKOKA HE IT(・A)・TES
IMAT(・A)・TA(・A)・YOU(・A)・DES(・U)・

→彼はどこかへ行ってしまったようです。

上例の述語語群の分析:

(i) IK(「行く」の語幹)の主語; KARE

(ii) TESIMAW(「てしまう」の語幹);主語の概念は考えられない。

以上の例文中“KARE HA”が述語語群の主要素(ここでは動詞)の行動主体を意味する主語となっているのは例3だけである。その外の例では附加された変形要素の意味を表す助動詞・補助動詞の発意主体(変形要素が意味する判断や配慮の主体)となっている。しかも例1の場合の動詞「K」(来る)の行動主体は文中には全く示されてさえない。

このように行動主体や変形要素の発意主体の表現の仕方は日本語の場合だけを考えてもそれほど単純ではない。従って D-tree モデルの仮定だけではこれらの変化を十分に捕えることはできない。こうなると当然これらを含む自然語文の翻訳は難しくなる。ただ実際上は、機械翻訳の対象をどの範囲に設定するので、本質的な欠陥が表面に顕現する程度は違って来ると思われる。しかし範囲を制限し過ぎると、翻訳システムの実用性・有用性の低下は免れない。従って単純に範囲の制限だけを考えることは本質的でない。意味的特徴についても本質的なものは直接に処理できるような

工夫をすることが必要であろう。

〔2〕 P-1 に関連して

D-tree モデルが幾つかの欠陥をもっていることは本節の他の個所でも指摘している通りである。ただしその意図するところは、特定の自然語に依らず原文の意味内容を記述する中間の構造を設定し、それを中継した翻訳過程を構成することである。これは言語構造の違いが特に大きな言語間の翻訳では必要であろうと思われる。D-tree はそのようなものの第一近似ではあるが、これによって、処理過程の構成が原文分析と訳文合成にかなり明確に分割できた。このため翻訳システムの内部構成の見通しも比較的好くなり、又、構文的特徴や文の構成要素が意味表現に果す役割も比較的好く捕えられるようになったと言えよう。

〔3〕 P-2 に関連して

個々の文の枠内が D-tree モデルで処理できる理論的限界とは考えられない。このことは複文、重文を一応処理できていることから明らかであろう。P-2 の仮定は当面の実験システムの作成の便宜を考慮したからに過ぎない。ところで実用的な自然語文では、一つの文の枠内で完結した意味内容を常に表現しているとは限らない。前後する一連の文章あるいはそれらの文章の置かれている環境も一つ一つの文の意味内容を決めるための大きな要因となることが少なくない。従って、そのような機能がある程度備えることも今後の機械翻訳システムに望まれることの一つであろうと思われる。

〔4〕 P-3 に関連して

組合せ論的な総当り方式、bottom-up 方式をとる利点は処理アルゴリズムが比較的単純で、漏れのないチェックができることであろう。しかしこの方法は、単語や語群ごとの多義が増すにつれ、そして文の長さが長くなるにつれ、組合せの爆発を招き無駄なチェックに大半の時間を費すことになる。そこで我々も幾つかの改良を試みた¹²⁾。しかし、その構文的特徴による語群処理・依存構造処理という部分には本質的変更を加えなかった。

本質的な改良を行うためには、必要に応じて意味の特徴を考慮すること、top-down 的なチェック機能も必要に応じて行えること、作業の進行につれて機械辞書には存在しなかったような新しい言語データが原文に則して形成されていくような処理機能の導入などが必要であろうと思われる。

〔5〕 P-4 に関連して

D-tree モデルが取り扱う言語情報で意味構造に直接的な関連をもつと思われるのは述語の機能パターン

情報である。ところが、述語語群を中心とした語群間の依存関係を求める際に、述語が支配する語群に関しては構文的特徴（語幹の品詞、語尾情報、語の位置関係など）しか利用しない。その外実際には品詞、機能パターンが同じであっても、述語そのものや支配される語群の意味内容の特徴によっては、依存構造が違ったり、訳語が違って来ることがあり得る。又、品詞は名詞であっても機能的には述語的な役割を文中で果す語（例えば英語の trial, support など；日本語のいわゆるサ変名詞など）、述語的役割を果す前置詞などがある。このような現象は単なる表層的な構文情報だけで捕え切れるものではない。代名詞の処理についても然りである。

以上列挙したような点は、柔軟な翻訳システムを目指す意味からも、できる限り本質に根差した処理方法を工夫する必要があると思われる。

〔6〕 P-5, P-6 に関連して

KULTS-I における訳語割当ての基本的考え方は、原単語の終端記号あるいはその構文的特徴で区別できる限りにおいて異なる訳語を与えるというものである。しかし現実には、原単語の綴りや品詞などが違い又単語の連鎖が違って同じ意味を表し、従って同じ訳語になる場合があり、逆に単語の綴り・品詞・機能パターンなどが同じでも、文脈によっては異なる訳語を与えるべき場合もある。ところが KULTS-I ではそのような柔軟な処理はほとんど行っていない。更に、訳出の語順、訳語と訳語の接続部分の訳出の場合などにも、原文における構文的特徴よりも、表現すべき意味構造を反映する目標言語の構文的特徴に影響を受ける面が実際には多いと思われる。

このような事情から、訳文合成に関する部分の今後の課題は、原文の構文的特徴ではなくその意味的特徴に基づいた柔軟な訳文合成ができるようにすることであると言えよう。

〔7〕 その他（システムの構成と記述言語）

翻訳システムの論理的構成については 2.2.1 の P-1 と 2.2.2 の〔2〕で少し議論した。ここではその物理的構成と記述言語について簡単に述べる。

システムの記述言語としては大部分を FORTRAN に、ごく一部分を FASP（アセンブリー言語の一種）に頼った。これは必ずしもそれらが最適の記述言語であると考えたからではない。当時最も手軽に利用できたこと、それにもかかわらずある程度満足できるシステム記述ができると判断したためである¹⁴⁾。その際特に心がけたことは、できる限りシステム構成に関して見通しの良さを保つこと、システムの修正・変更が比

較的容易に行えることなどである。そのために、一つのプログラム単位には原則として論理的なまとまりをもった一つの機能だけをもたせるようにした。その結果100個近いサブルーチン副プログラムから成るモジュール構成となった。このようにして出来上がったシステムは、どちらかと言えばやはり管理上いろいろ不都合があると言わざるを得ない。その主な原因は、本来ストリング処理向きでないFORTRANに無理にそれをやらせているため、表層文とその意味(処理内容)とが簡単には対応づけし難いことによると思われる。

以上のようなことから、この項目に関する今後の課題としては、ストリング処理を意識して設計された記述言語によるシステム記述と、その環境の下での然るべきシステム構成の検討が重要となるであろう。

3. 機械翻訳の新たな展開への準備

2.では、従来各方面で実際に試みられた多くの機械翻訳を振り返り、それらの背景にあった翻訳論や機械翻訳に関する方法論を整理し、又今後に残された問題点の検討を行った。ところで、一つの研究の展開を図ろうとする場合、過去の研究の延長線上に成果が期待できる部分と、新たな見方・考え方が必要な部分とがある。機械翻訳研究の場合、翻訳の立場から自然語の性質や特徴を見直してみることで、人間による翻訳と機械による翻訳の条件や環境を対比してみると、その上で機械翻訳の目標を改めて探り、そこへ到達するために解決すべき問題点を検討してみると、などがこの際必要であるように思われる。そこで本章では、これらの3点に関して従来の機械翻訳研究が見過してきたと思われる事柄を改めて簡単に指摘し、今後の研究の新たな展開への手がかりとする。

3.1 考慮すべき自然語の性質・特徴と翻訳論

人間の歴史の中で同じ内容が最も多くの異なる自然語で表現された例はキリスト教の聖書であるという。言語学者のE. Nidaは、みづからがその聖書の翻訳に参加した経験を踏まえて、自然語の翻訳操作、及びその観点からの自然語の性質や特徴の分析と整理を試みている⁵⁾。なおそれらの作業は主として人間が行う翻訳を念頭に置いている。ところで過去の機械翻訳研究を振り返ってみた場合、E. Nidaのような目で機械翻訳と自然語を見詰めた例は、筆者の知る限り、ない。従来の機械翻訳の場合、2.1で見たように、どちらかと言えば表層的な言語記号とそれらの相互関係を主な手がかりにして何とか翻訳してしまおうとする考え

が強かったように思われる。そしてその背後に次のような考え方があったと思われる。すなわち自然語の意味内容を解釈して理解するのは結局のところ人間である。従って、翻訳処理に伴う自然語の解釈であっても、それは翻訳文が出来上がった後で人間がゆっくり行えばよいのであって、翻訳処理の途中で機械に無理して行わせる必要はない、とする考え方である。この考え方の底には更に次のような考え方も潜在しているように思われる。すなわち；

(1) 自然語による意味表現の仕方——意味内容と意味構造の終端表現への写像の仕方——は、自然語の種類に関係なく互いによく対応している。

(2) ある意図や意味内容を表現する場合、終端表現の構成要素によって陽に表すべき部分および表現の仕方は自然語の種類に依らず共通である。

もしこのようなことが想定されていないならば、翻訳処理の途中で機械が原文を解釈する部分を持たない翻訳過程は考えられないはずだからである。ところが現実には、上記の(1)、(2)に反する事実と直面することが少なくない。これは、従来の機械翻訳が考慮に入れてきた自然語の性質や特徴がかなり一面的であったことを示していると考えられる。従って今後は、E. Nidaも指摘している⁵⁾ように、以下のような自然語の側面にも十分注意を払って、より本質的な機械翻訳の実現を図る必要があると考えられる。

(1) 自然語による意味表現の違いはそれぞれの自然語の恣意性に基づくことが少なくない。——例えば単語や単語列、内心構造による表現と外心構造による表現⁵⁾、言語記号による表現と言語外の表現など——

(2) 特別な標識を文に置かないままで異なる表現意図(比喩、反語、メタ表現など)のための構成部分を一つの文中に混在させることがある。

(3) 自然語文の品詞列が同じであって一見構文が全く同じようであっても、全く異なる意味構造を持つ自然語文であることが少なくない。

(4) 自然語文の意味はそれが置かれた文脈で決まるよう意図されることが少なくない。しかもその文脈は、自然語文そのもので明示している場合と、それらの文が発話される周囲の状況で暗に示している場合とがある。

(5) 自然語文の意味形成に寄与する因子として次のような事項も欠かせない。

- a) 発話内容の生起、発話内容に対する発話者の判断、発話行為などの“時”の取り扱い。
- b) 発話内容への発話者の関与の仕方——発話内容への着眼点、発話者の立場、発話内容への発

話者の評価・判断など——

- c) 発話内容に関与するもの——発話者、聴き手、発話内容の事象に関与するもの、などの取り扱い

以上のような自然語そのものについての性質・特徴の見通しの外に、どうすることが翻訳であるのかについても考え直してみる必要があるであろう。それは自然語の表す意味が多岐に亘っており、必ずしも一様な取り扱いを許さないからである。

一般には、形式的等価をねらう翻訳と、同等効果をもくろみ動的等価をねらう翻訳があると考えられている⁵⁾。必ずしも即断は許されないが、機械翻訳の場合、原文と訳文の意味内容の同一性ないし等価性をねらう立場から、動的等価の実現が基本になると考えられる。しかし、これらの内容の肉付けは今後の検討を待たねばならない。

3.2 人間による翻訳、機械翻訳の条件と環境

従来の機械翻訳では、機械にとって取り扱いができるだけ簡単な言語データをうまく利用して、できるだけ有用な結果を出そうとする傾向がなかったとは言えない。しかしこれではどうしても言語のもつ性質・特徴の一面だけしか見ないことになりがちである。人間による翻訳であればそのような場合の軌道修正は比較的簡単で、例えば意味情報、文脈情報の追加も本質的な困難は生じない。しかし機械の場合は新たな情報や機能の導入はシステムの全くの作り替えを意味することにもなりかねない。従って融通性に富む、そして汎用性の高い機械翻訳を実現することは容易でない。しかし、環境を固定すればある程度柔軟性のある機械翻訳を考えうる段階には来ているように見える。一方、翻訳という言語行為は、言語学、心理学を始めとする多くの言語関連科学に関係がある。従って機械翻訳には、それらの研究分野の個別の成果を総合化した場合の評価を客観的に行う手段としての役割を与えることも可能である。

以上のように、機械翻訳は、単に人間による翻訳の代替という目的だけでなく、関連する諸科学の有機的つながりを具体的にチェックする道具建てとしての積極的面も持っている。従って、人間による翻訳と機械翻訳の対比を試みる場合、単に人間と機械の物理的な違いだけでなく、そのような実用面からの役割の違いにも目を向ける必要があると思われる。

3.3 今後の目標と問題点

今後の機械翻訳で更に深い検討を要する課題・問題

Vol. 31 No. 1 (1980)

点などは、2.2以降折に触れてかなり詳細に議論してきた。その中に一貫していた問題意識は、機械翻訳において問題となる意味情報とは何か、ということであった。これに対する明解な回答はまだない。しかし、少しでも明確な回答を得るためには、本文中の各個所で指摘した個々の問題の解明が必要であろう。それと同時に、それらを個別の問題としてではなく、全体を組織化し、一つの機械翻訳論といったものにまとめていく努力も必要であると思われる。

4. むすび

本稿では、従来の機械翻訳の試みの流れをまず追跡し、その中に見られる基本的考え方や問題点などの洗い出しと整理を試み、今後の方向を探った。人間にとってさえ必ずしも容易でない翻訳を機械がすべての面で取って代ることはあるいは無理かも知れない。しかし、翻訳にはどのような側面があり、その中のどの部分がどのような方法で実現可能なのか、あるいは不可能なのかを知ることは人間自身の言語生活のためにも重要である。従って、2.2.2と3.3で述べたように、機械翻訳のための意味処理を探究し、機械翻訳論を模索し、合せて何らかの実用性を備えた機械翻訳システムの具体化を考えていくことが今後の課題と言える。

参 考 文 献

- 1) Bar-Hillel, Y.: The Present Status of Automatic Translation of Languages, in Alt, F. L. (ed.) *Advances in Computers*, **1**, Academic Press (1960) p. 91-163
- 2) Delavenay, É.: *La Machine a Traduire*, Collection *Que seisc-je?* No. 834 (別所, 沢辺訳, 翻訳機械, 白水社, (1964))
- 3) Hutchines, W.J.: *Machine Translation and Machine-aided Translation*, *Journal of Documentation*, **34**, 2, 119-159 (1978)
- 4) Josselson, H.H.: *Automatic Translation of Languages Since 1960: A Linguist's View*, in Alt, F.L. and Rubinoff, M. (eds.) *Advances in Computers*, **11**, Academic Press (1971) p. 1-58
- 5) Nida, E. A.: *Toward a Science of Translating*, (成瀬武史訳: 翻訳学序説, 開文社 (1972)) p. 406
- 6) Schank, R. C.: *Conceptual Information Processing*, North-Holland (1975)
- 7) 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*, **1**, 1, 51-68 (1963)
- 8) Vauquois, B.: ヨーロッパにおける自動翻訳計画, 情

- 報処理学会 CL 研究会, **CL 18-1** (1979)
- 9) Wang, W. S-Y., Chan, S. W. and T'sou, B. K.: Chinese Linguistics and the Computer, *Linguistics*, **118**, 89-117 (1973)
 - 10) Wilks, Y.: An Artificial Intelligence Approach to Machine Translation, Stanford AI-Memo 161 (1972)
 - 11) Wilks, Y.: Preference Semantics, in Keenan, E. L. (ed.) *Formal Semantics of Natural Language*, Cambridge Univ. Press, 329-348 (1975)
 - 12) 石原好宏, 田町常夫: D-tree モデルに基づく英日機械翻訳における単語・語群の構文的多義処理について, 昭48電気四学会九支連大, 341
 - 13) 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, 電子通信学会論文誌, **57-D**, 435-442 (1974)
 - 14) 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, 電子通信学会論文誌, **57-D**, 443-450 (1974)
 - 15) 穂富和夫: KT-1 による英独日 3ヶ国語間の相互同時翻訳の実験 (第1報), 九州大学工学集報, **36**, 20 (1963) (この外関連報告として第2報が九大工学集報, **36**, 3 (1963) に, 第3報が同じく **36**, 4 (1964) にある)
 - 16) 夢沼良一: 電子計算機による英文和訳の研究, 電気試験所研究報告, No. 624, 71 (1961) (この外関連報告として第2報が電試研報, No. 631 (1962) に, 第3報が同じく No. 633 (1962) にある)
 - 17) 長尾 真: 言語情報処理の過去・現在・将来, *情報処理*, **19**, 106-112 (1978)
 - 18) 長尾 真, 辻井潤一, 建部周一: 技術論文標題の英和自動翻訳の試み, *情報処理学会 CL 研究会資料* **19-2** (1979)
 - 19) 長尾 真: 機械翻訳, *情報処理*, **20**, 896-902 (1979)
 - 20) 長尾 真: ヨーロッパにおける機械翻訳の現状, *情報処理学会 CL 研究会*, **CL 20-1** (1979)
 - 21) 横田将生, 石原好宏, 田町常夫: D-tree モデルに基づく英日機械翻訳における述語語群訳語の合成について, 昭48電気四学会九支連大, 342

(昭和55年4月15日 受理)