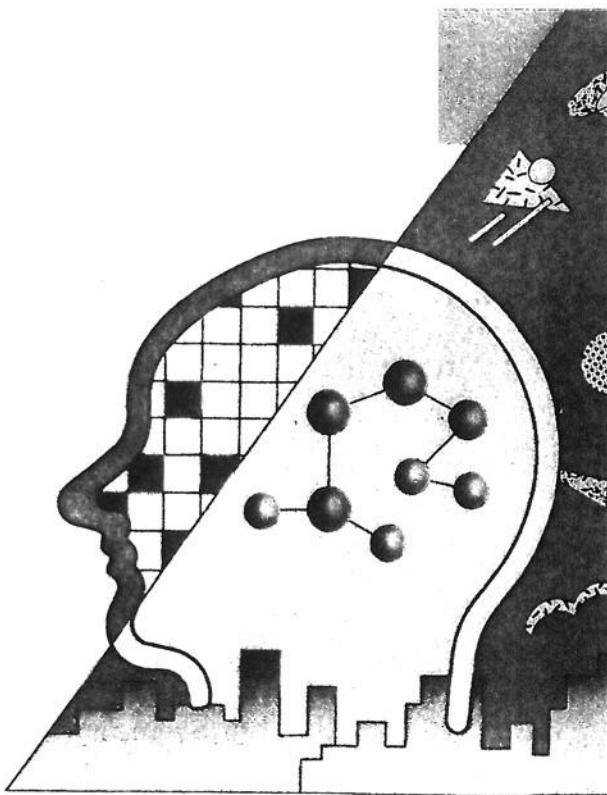


# Prologによる機械翻訳システム

その1

高野 真

今回から数回にわたり、Prologによる機械翻訳システムを紹介する。今回は今までの研究用、実用機械翻訳システムで用いられてきた翻訳方式についての簡単な解説を行ない、次回以降の指針を示す。



## AI 機械翻訳とはなにか

1946年のコンピュータの誕生からわずか6年後の1952年には、第1回の機械翻訳国際会議が催されている。このように機械翻訳は実用システムへの高いニーズに支えられて、30年以上にわたり研究開発が続けられ、現在に至っている。

しかし、その間には研究の挫折を何度も繰り返しており、一時期のアメリカでは、実用化には時期尚早という理由で「機械翻訳」と名の付く研究には1セントの予算も出ない時期があった。

現在、機械翻訳は一部のパソコンにも乗るくらいに普及してきている。普及すると同時に薄まりつつあるのは、機械翻訳の出現によって翻訳者が失業する、あるいは人間がいなくても機械だけで翻訳を行なってくれるという誤解である。

現在の機械翻訳では、機械にかける前に、機械の処理し易い文に直すプリエディットと、機械にかけたあとに人間の読み易い文に直すポストエディットが欠かせない。さらに、直訳はともかく意訳が必要な文（例えば小説）を対象とした機械翻訳は全く使い物にならない。機械翻訳に関するあるシンポジウムで聞いた話であるが、オフコン用の機械翻訳システムを売り出したあるメーカーのシステムは、いち早くパソコン用機械翻訳システムを出した翻訳会社のシステムに対抗するため、完成時期を2年も繰り上げて市場に出したということである。

こういった話はどこも似たりよったりで、メーカー側のこうした姿勢に泣かされるユーザーも多いようである。こうしてみると、機械翻訳はまだ発展途上のシステムであるといえる。

現在の機械翻訳システムは、翻訳のスピードよりも翻訳の質をいかに上げていくかが課題である。その解決のためには、意味の処理をがどうしても必要となってくる。意味の問題を導入するとなると、翻訳の対象を「特定分野の技術文献」「天気予報」「新聞の株式欄」というように限定することが必要となってくる。

今後のシステムは、翻訳の質を上げると同時に、より適用領域を特殊化する方向に進んで行くことが予想される。

## AI 機械翻訳方式

機械翻訳の方式は大きく分けて、ダイレクト方式、トランスファー方式、中間言語方式の3つがある。ユーザーの立場からすれば、翻訳方式などどうでもよい話であるが、ここでは機械翻訳システムを実現する側に立っているので簡単に触ることにする。

### ①ダイレクト方式

3つのうちこの方式だけは、要注意といってようだろう。ダイレクト方式は単語対単語変換を基本とした物で、語彙の入れ替えで翻訳を行なおうという単純な発想に基づいている。

1965年以前の機械翻訳システムやパソコン用の低価格システムではこの方式が用いられている。

双方の言語の構造が類似しており語彙の入れ替えですむ場合が多いときに有利な翻訳方式と言われているが、実際は日本語と韓国語のように文法体系が類似した言語を対象とした翻訳でも、ダイレクト方式での翻訳では本格的なものは作れないと言われている。

図1にダイレクト方式での翻訳の流れを示す。

### ②中間言語方式

中間言語方式は、原文を2つの言語に共通な構文表現あるいは意味表現に落とす翻訳方式である。図2に中間言語方式の概念図を示す。多言語に共通の中間言語を想定するものはピボット方式と言われることもある。

多言語間翻訳システムを開発する側からすれば各言語について、中間言語への解析と、中間言語からの生成だけ考えればよく、手間がかからないが、各言語固有の表面的情報まで落として両言語の共通部分だけ取り出すので、翻訳の質が落ちてしまうという欠点がある。

翻訳方式として中間言語方式の方が次に述べるトランスファー方式よりも優れているという見方があるが、これはあくまで開発する立場での意見で、逆に翻訳の質ではトランスファー方式の方が優れているようである。

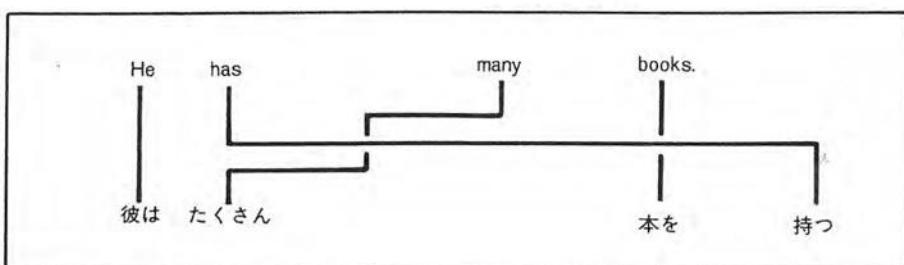


図1 ダイレクト方式での翻訳の流れ

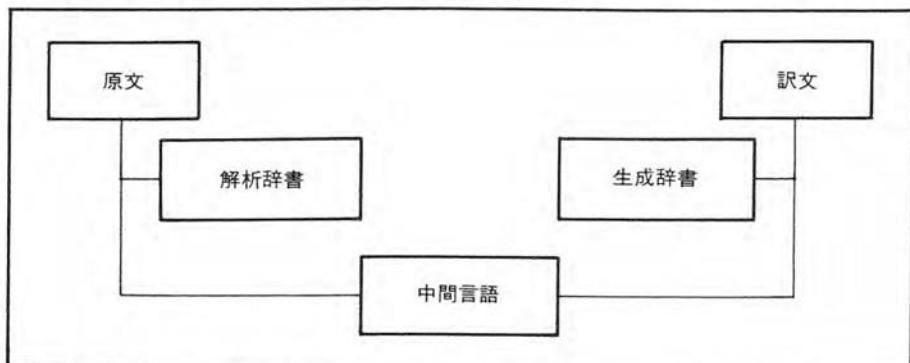


図2 中間言語方式による翻訳の概念図

フランスで作成された CETA という機械翻訳システムは、ピボット方式を用いていたが、GETA という後継システムではトランスファー方式に切り替わっている。また、特殊な中間言語方式としては、エスペラント語やドイツ語などを多言語共通の中間言語と考える方式もある。

### ③トランスファー方式

トランスファー方式は、実用、研究を問わず現在最も多く用いられている方式である。英日翻訳システムを例にとって説明すると、トランスファー方式では、まず英文を英語独自の中間表現に表わす。その後に英語の中間表現を日本語の中間表現に変換する。さらに日本語の中間表現から日本語を生成し翻訳を完成する。図3にトランスファー方式による翻訳の流れを示す。

まず、最初の解析の部分では、辞書引き、構文解析、意味解析が行なわれる。ただしこの例では、意味解析は行なわれていない。最初の辞書引きの部分では、文頭の大文字、三人称単数現在の S、過去形、過去分詞形、名詞の複数形や所有格についての形態素処理を施して標準形に直した後に、計算機内の単語辞書が参照される。次は句構造や語句の係受けを抽出する過程に進むが、このときに多品詞語や多義語があるとその決定がむずかしくなる。

このようにして、英語の中間表現が生成された後は、英語の中間表現を日本語の中間表現に変換する作業が必要になる。トランスファー方式の名はここに由来している。

具体的には、英単語を日本語単語に直したり、(主語、中心動詞、目的語) を (主語、目的語、中心動詞) といった語順に変えるという作業が行なわれる。

最後は、日本語の中間表現から生成辞書を参照して日本語を合成する。3段階のうち最も重要なのは最初の解析の部分である。ここで失敗するとその後の処理が全て無駄になってしまうからである。

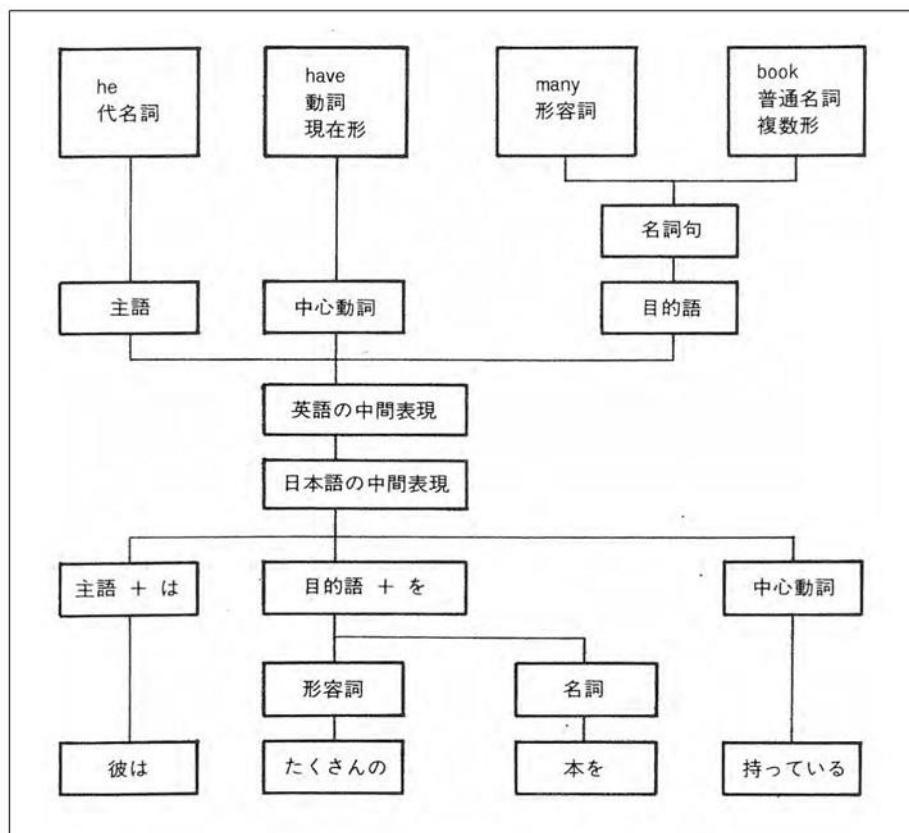


図3 トランスファー方式による翻訳の流れ

## AI 翻訳の質を高める方法

こうした翻訳に用いられる解析、変換、生成の3つの辞書はいずれも、複雑な経験則の膨大な集合である。言語のような複雑な対象を少数の規則で表現することなど不可能である。したがって、言語処理に用いるデータは、膨大な文献を調べることにより、何千何万という言語使用の規則を経験的に抽出せざるを得ないのである。もし、こうした経験則を当てはまらない文章に出くわしたときは、当然その文章の処理は失敗することになるのである。経験則自体があいまいさを含んだ、「大体においてあてはまる規則」であるため、誤りの無い機械翻訳システムはあり得ないのである。

機械翻訳の質の向上は、経験則の質を向上させ、量を増やすことで達成できる。あるいは機械翻訳を適用する領域を限定して、その中で十分な量の経験則を細部にわたって記述することも考えられる。いずれのアプローチをとるにせよ意味処理の問題を取り込んでいく必要がある。

意味処理が必要な文章の例としては、

Time flies like an arrow.

を、時蠅は矢を好む

と訳してしまう有名な例があるが、もっと身近な例でも

write with pen ペンで書く

go with him 彼と一緒にいく

go with a stick ステッキを持っていく

といったように、with の訳は前後の動詞や名詞の意味によって変わってくる。

こうした文章を正しく解釈するには、「write with の後には道具がくる」「ペンは道具である」「go with の後が人ならば『一緒に』と訳し、物ならば『持って』と訳す」といった判断を行なうための意味情報を用いて文を解釈しなければならない。

翻訳の質を高めるには、上記のような経験則を準備して置けば良いのだが、領域を限定しない場合、こういった経験則はほとんど無数と行って良いくらい存在する。さらに、どこまで詳しく意味を記述すれば翻訳できるかも、場合に応じて様々な基準がある。したがって、翻訳の質を高めるには翻訳の対照領域をうまく絞るやり方が現実的になってくる。

## AI インプリメントするシステム

連載で取り上げるシステムは、英日翻訳を対象とする。対象領域は中学校英語の参考書をもとに、文法的にしっかりした英文を扱うことにする。参考書としては「中学英語公式集」(樋口 忠彦著:むさし書房刊)を選択する。従って、語いについてはこの参考書に出てくる単語のみを扱うこととする。

また、「中学英語公式集」では、「This is a pen.」から関係詞まで100近く項目があるが、基礎的なものから順次サポートしていくように進める。従って後から文法や語いを追加し易いように辞書や文法、中間表現を汎用性の高いものにする。

翻訳方式については、トランスファー方式を採用する。中間言語としては格文法（英日共通）を用いて、英日変換規則を単語の変換規則だけにとどめるようにする予定である。中間表現は変更があるかも知れないが次回までには決定する。

次回は、英文を解析する手順について、Prolog プログラムとその実行例を用いて解説を行なう。