

## 特集「機械翻訳」にあたって

田中穂積\*

芸術的な文学作品や詩歌ではなく、科学技術や新聞記事などといった比較的きちんとしたスタイルで書かれた文書に対して、現在の機械翻訳システムは、予め翻訳すべき文書を解析しやすい文体に直す前編集や、機械翻訳の過程で人間から助言を得たり、翻訳結果を修正する後編集技術を組み合わせて、部分的であるとはいえ実用化の段階に入ってきている。人間からの助言や、後編集が必要であるということは、機械翻訳が完成した技術でないことを示している。

機械翻訳システムの使用者も、最近ではむしろ完全な翻訳結果を機械翻訳システムに求めるといふより、不完全な翻訳結果を出力する現在の機械翻訳システムとどうつき合うかという観点が重要であると認識するようになってきた。不完全な翻訳結果しか得られなくても、翻訳の対象となる文書の範囲を制限し、機械翻訳で最も困難な問題を人間からの助言という形でバイパスすることにより、機械翻訳システムの実用化が可能になってきたと言ってよいだろう。しかし、さらに高度な機械翻訳システムを生み出す努力を今後も粘り強く続けなければならない。本特集号は、機械翻訳システムの基本技術を解説すると共に、機械翻訳の将来を展望することを目的に編集されている。

ここで機械翻訳と人工知能の研究の関連を述べておきたい。機械翻訳の研究が本格化するのは1960年代に入ってからである。アメリカでは、MIT、ハーバード大学、IBMなどで研究が行われていた。当時の機械翻訳の研究は、コンピュータ・パワーの制約、未熟な自然言語処理技術、言語理論の問題など、本格的な展開を図るに十分な基盤があったとは到底言えない。それにもかかわらず当時の機械翻訳の研究者の多くは、統語論(シンタックス)を中心とした言語解析で十分であると考えていた。しかし、この考え方に限界があることが1960年代の初頭で明らかになる。特にアメリカでの機械翻訳の研究は、National Academy of Sciences(アメリカ)のDivision of Behavioral Sci-

encesに設置されたAutomatic Language Processing Advisory Committeeが1966年に出したALPAC報告書により致命的な打撃を受ける。ALPAC報告書は、これまでの機械翻訳の研究の延長上には未来はないとし、機械翻訳の実現のためには計算言語学などの基礎的な研究を行うべきであると結論し、機械翻訳に対する研究資金の打ち切りを勧告した。その影響は、アメリカにとどまらずカナダ、フランスを除く世界各国に及ぶことになる。1960年代の後半から1970年代全般にわたり、機械翻訳の研究が目ざましい成果を挙げることなく停滞するのは、このような技術的、歴史的事情があった。

このころ、統語論中心の考え方ではなく意味を中心に据えた情報処理の重要性を説く人工知能の研究者らがいた。1960年代の後半から1970年代にかけて、人工知能の研究から自然言語処理に関する研究として興味あるものが数多く生まれた。意味の問題が困難なものであることは十分予想されることである。彼らは意味を扱うにあたって、処理の対象となる文の範囲を非常に限定した。したがって、扱うことのできる文の文型も限られることになり、語彙の大きさも極めて限定されることになる。人工知能の研究者は、自然言語の文を概念の世界に写像しコンピュータに自然言語を理解させるシステム、言い換えると自然言語理解システムの実現に努力した。機械翻訳の研究者が、自然言語のかなり広範な言語現象を扱い、解析は浅いレベルにとどめようとしたのに対して、人工知能の研究者は、自然言語の極めて狭い言語現象を深く追求しようとしたのである。このような事情があったためであろうか、アメリカでの人工知能の研究には、機械翻訳の研究がほとんど含まれていない。

考えてみればこれは不幸なことである。そもそも機械翻訳は、多量の言語的知識と一般常識、翻訳過程では言語解析や生成のためのさまざまな推論を必要とする。知識情報処理の典型と見なすこともできるシステムである。アメリカでの人工知能の研究者が機械翻訳の研究にこれまで積極的に取り組むことがなかったの

\* 東京工業大学工学部

は、上述した理由があるとはいえ、不幸なことであった。幸いにして我が国では、機械翻訳は人工知能の典型的な研究課題であると考えられている。知識を扱うことは人工知能の中核をなす研究課題であるから、このことはむしろ当然と言ってよい。

これに関連して、ヨーロッパでの機械翻訳の研究状況を観察すると、彼らは依然として統語論の範疇にとどまっているように思う。一方、我が国の機械翻訳の研究者は、機械翻訳の質をいま一段向上させるためには、辞書に含まれる情報の量と質の向上、言い換えると知識の量と質が問題であると考えている。第五世代コンピュータ計画の一環として初期に計画された大規模な辞書の構築を国が支援しているのも、他に類を見ないことである。

機械翻訳は、基本的には原言語（翻訳の対象となる言語；source language）を目標言語（翻訳結果の言語；target language）に変換（写像）することである。その意味で典型的な記号処理であるといえるが、写像の過程で、原言語の解析と目標言語の生成とが主たる自然言語処理技術となる。原言語の解析で最も重要な問題は、原言語の文に含まれる曖昧性の除去であ

る。生成では、訳語の選択などが問題になる。これらについては、本特集号の野村先生の「自然言語解析」、辻井先生の「機械翻訳における文章の生成」、石崎先生らの「文脈と対象世界モデルを利用した機械翻訳へ向けて」を参照されたい。辞書の問題については内田先生の「機械翻訳と電子化辞書」を参照されたい。

機械翻訳には大別して二つの方式がある。一つはトランスファ方式であり、もう一つは中間言語方式である。いずれもフランスのグルノーブル大学から生まれた考えであるが、両方式の利害得失が最近問題になっている。この問題は座談会「機械翻訳における中間言語方式をめぐって」で、第一線の研究者にお集まりいただき御議論いただいた。

最後に、我が国の機械翻訳の技術を育ててこられた長尾先生に「機械翻訳の新たな展開に向けて」と題して、今後を展望していただいた。

本特集号を読まれた読者が、機械翻訳の基本原則を理解し、そこに含まれる問題をもう一度再考する機会になれば幸いである。なお、お忙しい中、本特集号のために御執筆、あるいは座談会に御参加下さいました皆様に感謝いたします。