

言語と知識

田中穂積

(東京工業大学工学部)

コンピュータに自然言語を理解させる(自然)言語理解システムの研究が近年盛んになってきている。これは、人間とコンピュータとの間の使用言語の相違に基づくセマンティック・ギャップを解消して、人間とコンピュータとの間に新しいインタフェースを確立しようとするものである。言語理解システムは、第5世代コンピュータ計画では、知的インタフェースの研究の一部として含まれている。

自然言語を理解したり、自然言語の文を作るために、我々が知識を使っていることは明らかだろう。それと同様に高度な自然言語理解システムには、自然言語を理解するための知識とそれを使うメカニズムが必要である。

言語理解のための知識として、我々は言語学的な知識(文法等)とそれ以外の知識(常識等)とがある事を知っている。前者については言語学者が様々な角度から研究している。ところが後者については、それがどのようなものであるかについて、心理学者、言語学者、哲学者等によって断片的な研究はなされてはいるものの、体系的な研究がなされているとは言い難い。言語理解や文を作り出す過程で使われている知識がどのような形、姿をしているかについて我々はまだ十分な知見を得ているわけではない。

頭脳の中から知識を取り出して眺めることができれば、我々は知識の姿について、なにがしかコンセンサスが得られるかもしれない。そのためにたとえば、頭脳の機構を神経細胞のレベルから解明することが考えられる。知識の正体に対するいわばボトムアップ的な迫り方である。しかし、微細な神経細胞に関するこれまでの研究成果は、必ずしも言語理解システムの作成に役立つレベルにはない。

一方、言語理解における知識の研究をトップダウン的に進めることもできる。言語理解システムの作成者が採用しているアプローチがそれである。ひとまず我々は、頭脳の神経細胞レベルでのミクロな研究を離れて、言語理解システムの作成を試み、システムがどのような機能から構成さ

れるべきかを研究し、その中で知識がどのような形式である事が望ましいか、またそれらがどのように使われるかを研究するのである。この場合、システム内部の知識は、とにかく何等かの実体(記号の集合)として表現される。この実体を我々は具体的に手に取り出して眺めることもできる。しかもシステムの挙動は頭脳と異なりシステムに適当なブロープを差込むことで比較的詳細に追跡することもできる。

このような実験により言語理解に関する新しい知見が得られれば、それを我々の頭脳が行っている言語理解の機構に対する一つのモデルないしは予測として用い、トップダウンとは逆に、予測にガイドされたボトムアップ的な神経細胞レベルの研究を効率良く進めることができる。今世紀最後の科学であるといわれる認知科学は、コンピュータ上にこのようなシステムを実現し、それを実験道具として我々の認知の機構に迫る事が一つの眼目であったと思われる。

ここで言語理解システムを作成する上での望ましい知識表現とはどのようなものであるかを考えてみたい。それには次の3つの側面を考える必要がある。

- 1) 知識をどのような形式で表現するか(知識表現形式)。
- 2) 知識をどのように使うか(推論方式)。
- 3) 知識をどのようにして獲得するか(知識獲得)。

1)の知識表現形式については、人工知能の研究者によって様々な研究が行われて来た。代表的なものにセマンティック・ネットワークやフレームがある。筆者は、人工知能の研究者がいうほど両者に違いがあるとは考えていない。むしろ両者は相互に等価な表現形式であると考えている。セマンティック・ネットワークやフレーム形式に拡張を施すことにより、述語論理による表現形式と等価な知識表現形式を得ることができるとも知られている。この拡張は、主として述語論理における限量記号をセマンティック・ネットワークやフレーム形式に取り入れることである。

この様に考えると、セマンティック・ネットワークや

フレームによる知識表現形式と述語論理による知識表現形式とは、同一視して良いと思うかもしれない。しかし両者は2)の問題に対する考え方が基本的に異なる。

セマンティック・ネットワークやフレーム形式による知識表現では、表現された知識をどのようなものとして使うかは、セマンティック・ネットワークやフレーム形式の使用者個々の作成するインタプリタがどのようなものであるかに依存している。これらの知識表現形式には、本来的に推論機構の裏付けがないから、我々が知識表現形式としてセマンティック・ネットワークやフレーム形式を選択すれば、それを解釈するためのインタプリタを別途(必ず)作成しなければならなかった。

筆者は知識は推論と背中合わせの関係にあるものだと考えている。極論すれば、推論機構の裏付けのない知識は知識とは言えないと考えている。知識に対するこのような見方をすれば、形式的な推論機構を背景にした述語論理による知識表現形式は、望ましい知識表現形式の一つであるといえる。特に最近のロジック・プログラミングの枠内で知識の問題を考えることは意味がある。なぜなら、一定の形式の述語論理で表現された知識に関する推論は、全てロジック・プログラミング・システムに組み込みの機能を利用して行うことが可能になるからである。

筆者等の最近の研究によれば、セマンティック・ネットワークを自在に辿る機能、上位にある知識を下位からアクセスする機能(知識継承機能)、上位のメソッドを使う機能等は全て、ロジック・プログラミングの枠内のユニフィケーション機構で代用可能なことが明らかになっている。言語学的な知識である文法規則についても、これをロジック・プログラムの形式に変換し、このプログラムに文を与えて構文解析する事も可能なことが知られている。文の意味解析を行うために必要な知識もロジック・プログラミングの枠内で自然に記述できる。

パターン照合が推論の基本にあることは多くの人の認めるところであろう。ところでユニフィケーションは、パターン照合の最も純粋で基本的な形態である。ロジック・プログラミングがユニフィケーションを基本計算機構としていることを考えると、述語論理に基づく知識表現は心理

学的にも検討に価すると思われるがどうだろうか。筆者のこれまでの経験によれば、ロジック・プログラミング言語の典型であるPrologは、Lispと同じ位面白い言語である。その上で言語理解のシステムを作成することは、認知科学の観点からも意味のあることだと思う。

3)については研究が緒についた段階にある。学習の問題とも関連する研究課題である。知識の獲得には抽象化、一般化が必要になる。そのために最も大切なことは、ある知識とある知識が似ていることを我々はどのようにして認識しているかを知ることであろう。相似な知識の認識を通じて、我々は多くの事を一般化し学んでいると思われるからである。筆者は知識獲得の問題を相似の問題に還元させながら解くことが、最終的には学習の問題を解くことにつながると考えている。

最後に今後の研究課題として、文脈をどのように処理するかと言う問題があげられる。文脈を理解するためには、文を処理した結果としての知識の形式をどのようにしたら良いかを深く考えなければならない。文脈理解の過程で様々な推論が行われることは良く知られている。したがって文脈理解についても、推論機構の裏付けのある知識を文から抽出しておくことは重要であろう。また比喩や隠喩の問題も、推論と深く関係している。この種の問題の解決には長期を要するだろうが、解決に向けての努力をすべき時期に来ていると思われる。それにより言語と知識との関連がより一層明確になるだろうし、知識表現の形式についてもよりよい考え方が生まれるように思われる。