



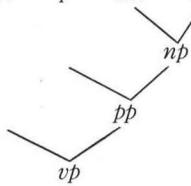
言語理解システムの完成を目指して ——田中研究室～情報工学科——



田中穂積教授

図1.

Time flies like an arrow.
 n vi p det n



日常生活において我々が何の気なしに用いている言葉は自然言語といわれている。一方、コンピュータで使われているのは人工言語、プログラミング言語である。そのため、我々がコンピュータを扱うには、それが理解できる人工言語を学ばねばならない。つまり、使用言語の相違が

両者を隔てる最大の障壁となっているのだ。もしコンピュータがこの障壁を打ち破り、自然言語を理解できるようになったら、コンピュータはずっと身近な存在になるだろう。今回は、そのような夢を抱きつつ研究に取り組まれている田中穂積教授を訪問し、お話を伺った。



パーザーを使っての言語解析

6年前、筑波にある通産省の電子技術総合研究所から移ってこられた田中教授は、まず最初に良い研究のための良い道具づくりとして、言語解析をするための道具である、パーザーづくりを手がけられた。

パーザー(<parse≡解剖する)とは、自然言語の文を解析して主部、述部を決め、さらにその主部、述部がどのような語から成っているか、というところまで分析するためのプログラムである。言語解析では、このパーザーを使って言葉の一次元の列を解剖し、どこに何があるかを明らかにした上で、それが文法にあつていてあるかをチェックする。そこでコンピュータに文法を教え込む必要が出てくるわけだが、文法だけを使って解析すると、膨大な数の解析木ができてしまう。

例として“Time flies like an arrow.”という有名な諺を解析してみよう。まず本来の意味に従ってこの文を解析すると、図1のようになる。すると「時は矢のように飛ぶ」くらいには翻訳することができる。ところが、flies to fly (ハエ) の複

数形と考えると、図2のようにも解析される。また time には「乗する」という意味の動詞もあるから、命令形として図3のようにもなる。これらはどれも文法的には正しいが、意味を考えると、「時バエは矢を好む」、「矢のようにハエを掛け算せよ」のような意味の通じない文になってしまふのである。

上記のような簡単な文にでも、図1～3以外に3つ、計6個の解析木が存在する。また、30語の日本文で1億個の解析結果が出たという例もある。もちろん人間は1億個もの解析結果を出すことはできないから、コンピュータにやらせるのだが、時間がかかり過ぎるのは困る。そこで解析結果が1億個あっても平気で使える、高速パーザーの開発が必須となるのだ。

田中研ではまずこの問題に取り組み、今では6年前と比べて、解析時間が約30分の1に短縮できるようになった。そして現在、田中研にあるパーザーのスピードが世界最高速といわれている。



常識をコンピュータに教える

図2.

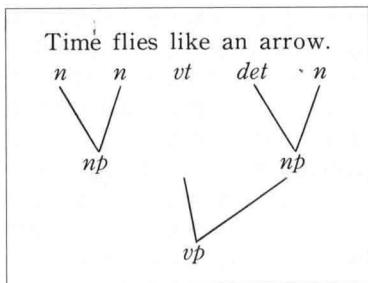
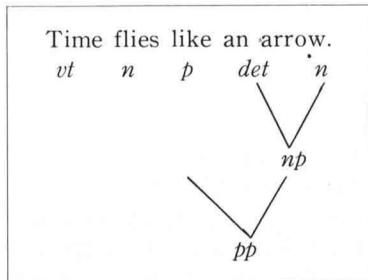


図3.



以上のようなパーザーの研究が2

年間続き、スピードに関しては充分な性能が出せるようになったので、次に意味を考える研究が始まった。

意味を考えるにあたり難問として表面化してくるのは、人間にある常識がコンピュータにはないということだ。例として「さんまを焼く母親が振り向いた」、「さんまを焼く煙がたちこめる」という2つの文を考えてみよう。

まず前者の意味解析では「母親」という語が「焼く」に対してどんな役割を果たしているかを分析しなければならない。そこで「焼く」を調べると、「焼く」には動作主と焼く対象が必要となっており、「さんまを焼く」というと動作主が欠けていることがわかる。動作主は人間でなければならず、母親は人間だから、「母親」は動作主であると決定され

ることになる。

ところが後者の場合、煙は動作主でも焼く対象でもない。煙がさんまを焼くのでもなく、煙がさんまを焼くのでも煙にさんまを焼くのでもない。よってこのような文を理解するには、推論のチェーンを持った常識が必要なのだ。人間は、さんまを焼けば煙が出てることを常識として知っているから、難なくこの文を理解することができる。だが、コンピュータには常識がないから、我々が当然と思っている常識をも教え込まねばならない。そして教え込むにしても「さんまを焼けば煙が出て」いう知識ではなく、もっと応用範囲の広い一般的な知識にしなければならない。こういった常識と意味の関係の研究も進み、今では前出のさんまの文程度のものであれば、完全に理解できるようになっている。

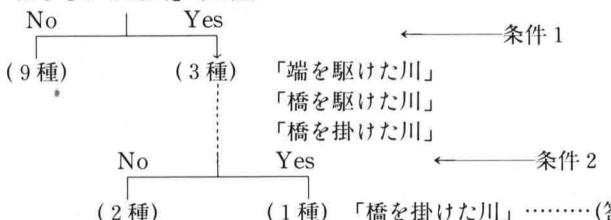


曖昧さの解消に欠かせない Decision Tree

前にも述べたように、文法のみを用いて解析すると、多くの解析結果が出るのだが、これは言葉に曖昧さがあるためと考えられる。しかしこの曖昧さは、意味を考えることにより絞ることができる。

例えば「はしをかけた川」という言葉があったとする。「はし」には端、箸、橋の3種、「かける」には賭ける、欠ける、掛ける、駆けるの4種があるので、「はしをかけた」という文は、12の可能性を持つことになる。これに1つずつ条件を加えていき、だんだんと可能性をせばめていく、最後に辿り着いたものが答えとなる。(図4参照)図4のように、Yes, Noで絞りこんでいく樹形図はDecision Treeと呼ばれる。現在では、教授の言葉を借りれば、「増進的曖昧性の解消」に

「はしをかけた川」(12種)



条件1……「はしをかけた」の意味が通るか
条件2……川に関係しているか

図4. Decision Tree

これが使われている。

しかし Decision Tree にも欠点があり、後で判断するものが先に出てきたり、省略語があったりすると、いつまでたっても答えに辿り着かなくなってしまう。この問題の解決が最近の課題であったが、漸くこの解法も見つかり、Decision Tree を用いた研究がさらに進んでいる。

また、比喩に関する研究がなされている。「男は狼だ」という比喩を考えてみよう。一般に狼は獰猛な野獣というイメージを持たれているが、アイヌでは神聖な動物と考えられているので、先の比喩の意味するところとは全く別のものになってし

まう。つまり、比喩はその土地の文化によって異なるのだ。その上比喩は普通の常識では理解できない。男は人間であって狼ではない。そこが比喩の難しいところでもあり、興味深いところでもある。

この問題を解くにあたって、まず狼という観点から男を見てみると、狼の全ての性質が男に移っているわけではないことから、どんな性質が移るのかをいかにして調べるかが問題であることがわかる。そこで取り上げられたのがエントロピーを用いる方法である。初めに、狼の特徴として最も確率の高いようなものを持ってくる。次に、狼と同等の

動物の性質を見る。そして、なぜ象ではなく狼が選ばれたのか、なぜ狼でなくてはいけなかったのか等を考え、他の動物にはない特徴をエントロピーを用いて取り出すのだ。比喩は、意味的に正常な文と意味の全く通じない文とのちょうど中間に位置するから、狼の性質の確率密度関数を変化させて、中間にくるように狼の特徴を選んでやればよい。これを実行するには膨大な量の知識が必要になるが、この比喩に関する言語理解システムの完成も近い将来のこととして期待できるであろう。



多言語間翻訳の実現と今後の課題

田中研では、これまでに紹介した研究をベースとした応用システム、機械翻訳についての考察もなされている。機械翻訳には2通りの方法があり、その一つは、普段我々が翻訳する時のように二言語間で考えるものである。日本語を英語に訳すとすると、まず日本語を解析して構造を明らかにし、それを英語の構造に直してから英文にするという3経路を辿る。

もう一つの方法は、language universal な構造を考えて、日本文からそれに直し、そこから英文に直すというものである。こうすると言語のペアを考えなくて済み、多言語間翻訳が容易にできる。即ち、前者の方

法では n 言語間翻訳の際に nC_2 個のシステムが必要となるが、後者では n 個のシステムでよいのだ。

ただここで問題となるのは、language universal なるものが果たして存在するのかということである。絶対に存在しないと主張する学者もいるが、田中研では存在する方に賭けているそうだ。

「すぐに見つかるとは思わないけれど、このアプローチの方が将来性があっておもしろいからね。」と田中教授は微笑まれた。

今のコンピュータシステムというのは、わかったようなふりをして実はわかっていないというのが非常に多い。換言すると、コンピュータは

人間の意図がわからないために、全く融通が利かないのだ。

例えば、「東工大の音楽科出身で有名な人はいますか?」と聞かれたら、我々であれば「東工大に音楽科なんてないよ。」と答えるだろう。ところがコンピュータは、データベースを探しても実際にそういう学生はないから「いません。」と答える。すると、東工大にも音楽科はあるんだな、という誤解を招いてしまう。このような誤解の検出方法も、基礎的大が大変重要な問題である。田中研では、この問題に関する研究も始まっている。

「自然言語を理解するコンピュータの開発なんて不可能だ、と言う学者がいるんだよ。鍊金術みたいなものだ、とね。でもたとえ不可能だと証明されたとしても、漸近的に可能に近づく努力が非常に重要なんだ。少しでも近づければ、それはそれで

意味がある。だからやらなくてはならないんだ。」という田中教授の言葉は、何事にもあてはまるのではないだろうか。可能でなくとも一歩でも近づこうという精神を我々も身に付けてみたいのだ。

最後に、お忙しい中、快く取材を

お引き受け下さった田中教授をはじめ、取材時にご協力頂いた助手の徳永さん、D1の岩山さんに心からお礼申し上げます。

(大谷)