

研究室紹介

東京工業大学工学部情報工学科田中研究室

東京工業大学工学部情報工学科田中研究室は、昭和58年度から自然言語処理を中心に研究を行っている。研究室としての歴史はまだ浅いが、幸いにして学生の努力により、いくつかの興味ある研究成果が生まれてきている。自然言語は、これまで主として言語学や心理学、哲学の研究対象であった。その意味で自然言語処理研究のアプローチは、計算機科学にとどまらず多岐にわたる。田中研究室では、主として人工知能と計算言語学の立場から自然言語処理の研究を行っているが、言語学、心理学、哲学などの学問から得られた知見をも参考にしたいと考えている。

自然言語処理の研究は本質的に長期を要するものである。その点を考慮し田中研究室では、自然言語処理のための基礎的でしかも新しい技術の開発を目標に研究を進めている。

研究課題は、次の四つに大別される。

- ① 自然言語処理のためのソフトウェアツールの研究
- ② 知識表現形式の研究
- ③ 自然言語理解システムの研究
- ④ 多言語間翻訳の研究

①の研究は、今後の研究の基礎になるものであり、主として日本語や英語などの自然言語処理を行うためのソフトウェア環境(ソフトウェアツール)の構築をProlog上に行う。これは機能の拡張と速度向上の両面から研究をし、現在、機能的、速度的にはほぼ満足できるツール(LangLABと呼ぶ)が構築されている。LangLABは、関係代名詞節に見られる左外置の現象を扱う機能を付加したことにより、文法記述が容易になっている。

②では、DCKRと呼ばれる知識表現形式を開発している。これは現在、③、④の研究で用いられており、その有効性が確認されつつある。現在、検索の高速化のための作業を行っている。また意味解析のベースとなる上位/下位関係シソーラス ISAMAPの作成を進めている。およそ2000の概念について(角川新類義語辞典に含まれる主要な概念1000を含む)上位/下位関係についてのシソーラスを作成したが、現在視点を考慮に入れたシソーラスを開発中である。視点の導入

にともない視点付きの概念を提案し、それをを用いたシソーラス検索用の機能を開発した。これは③、④の研究で利用されている。

③では、自然言語理解の中核に位置する意味解析の研究を現在行っている。文を読み進むにつれて増進的に意味的曖昧さを解消する計算モデルを研究している。これは制約プログラミングの考え方を意味解析に応用したと見なすことができるが、ユニフィケーションを拡張して実現している。その有効性について現在検討を進めており、談話理解にも応用可能なことが明らかにされている。談話理解の研究に取り組む準備ができてきたといえるであろう。そのほかに、比喩理解の研究を開始したところである。類推の問題との関連も研究している。

④の機械翻訳の研究では、多言語間翻訳の問題に二つの方向から取り組んでいる。一つは、ある言語の解析用と生成用の文法をできるだけ共用化して重複を避ける研究である。もう一つは既存の対訳辞書から、多言語間翻訳で必要となる中間概念を自動的に抽出するアルゴリズムの研究を行っている。これは知識獲得の問題ととらえて研究を進める予定である。以上の研究は、多言語間機械翻訳で理想とされていた中間言語方式に貢献できるものと考えている。

研究室の構成は、教授田中穂積、助手徳永健伸、奥村学、博士課程学生4名、修士課程学生6名、学部学生3名、研究生2名の計18名、主要計算機は、Sun3/260, Sun3/60(×5台), (Sun2/120(×2台), 三菱社製プサイI, II, 立石社製LUNA, 沖if1000, IBM PC, マッキントッシュである。使用プログラム言語はすべてPrologを中心とした論理型プログラム言語である。学生はPrologの後でLispを教えても、Lispに馴染もうとはしない。Lispもよい言語なので、最近では教える順序を逆にしようかと思っている。

最近のワークステーションの高性能化と低価格化により、大学でも計算機環境の面ではさほど不具合がなくなってきた。その意味で大学でかなりの研究成果が挙げられうる状況になってきている。そうしたこともあってか、学生が驚くほどのパワーを発揮しびっくりすることがあるのはうれしいことである。

〔田中 穂積〕