

報告

第6回計数言語学国際会議に出席して*

長尾 真** 辻井 潤 一**

第6回計数言語学国際会議 (International Conference on Computational Linguistics 略して ICCL) は、1976年6月28日～7月2日の5日間カナダのオタワ大学で開かれた。一般講演56件、招待講演7件で、講演時間は質疑応答を含めて各50分とかなり余裕のあるプログラムであった。バイリングアリズムの問題から生じたカナダ航空のストライキのため、参加できない人もいて、参加者は約200名と主催者の予想を若干下回ったようである。日本からは、長尾、辻井(京大)、有川節夫氏、武谷峻一氏(九大)の4名が参加した。一般講演は4つの session で並行して行われ、会議の公式言語は英語と仏語で、しかもオタワで開催されたということもあって、約1/3の講演が仏語でなされた。

一時期の機械翻訳のような中心的課題がないために、言語処理研究といっても多様化し、人工知能研究の一環としての自然言語理解システムから、言語の統計的な研究まで、種々のものが含まれていた。今回は論文数を増さず、活発な討論を中心として会が運営されたことが1つの特色であろう。またこれまでの会議とくらべると、人工知能研究における手法をとりいれた論文が圧倒的に多かったということが、今回の会議の最大の特徴であるといえる。目立った傾向を整理すると、ほぼ次のようになる。

(1) 知識の表現

MITのMinskyのFrame理論は、人工知能全般に対する方法論として提唱されたが、自然言語の理解とも深く関係している。この会議でも、多くの研究者がこの理論との関連の中で、自分の研究を位置づけていた。特にエール大学のR.C. SchankのグループのScript(筋書き方式)や、エディンバラ大学のY. WilksのFrame理論に関する講演には多くの人が関心を持

ったようである。しかしながら、エール大学のScriptにしても、まだ統一的に把握されるまでには至らず、個々の例題に応じて、その例題用のScriptが作られているのにすぎず、まだまだ多くの問題を含んでいる。また、言語理解の過程で、膨大な知識が関与していることは、誰もが認めるけれども、その関与の構造をScriptやFrameといったもので表現するのが適切かどうか、あるいは、この種の抽象的な議論が現実の成果に結びつくかどうか等のかなり厳しい批判も出された。

(2) 質問応答システムとデータベース・システム
質問応答システム(Q.A.システム)は、機械翻訳に代わる言語処理研究のテーマとして、多くの研究がなされてきた。今回の会議でも、我々の発表以外にも多くのグループから、このテーマに関連する発表が行われた。さらに注目すべき傾向としては、データベース・システムの研究と、Q.A.システムの研究が非常に近づきつつあることである。この方向の研究として、オランダのフィリップス社、Bolt Beranek and Newman (BBN)、IBMからの発表があった。特に、relational data baseの枠組の中で、従来からのsemantic networkに、より厳密な論理学的基礎を与えようとする試みが、BBN、SRI (Stanford Research Institute)等から発表された。

(3) 音声認識、音声合成

音声認識ではBBN、SRI、CMU(カーネギー・メロン大学)、九大(武谷氏)から発表があった。いずれも音声認識における言語情報処理に焦点をあてたものであった。米国のspeech understanding systemは、今年の秋でプロジェクトを解散するところが多く、これまでの各グループの研究の総合報告的な発表も多かった。音声合成では、MITのグループが形態素の発音辞書を使った合成システムを発表した。

(4) 情報検索、統計処理

コーネル大のG. Saltonが自動インデックスの手法

* 6th International Conference on Computational Linguistics
Report by Makoto NAGAO and Jun-ichi TSUJII (Department
of Electrical Engineering, Kyoto University)

** 京都大学工学部

について述べた。基本的には頻度の大きい語は句に、頻度の小さい語は上位語に変換することにより、呼出率、適合率の良い key word, key phrase の組を得ようというもので、説得力のある発表であった。情報検索では、九大(有川氏)からの発表があった。

(5) 機械翻訳

グルノーブル大、モントリオール大を中心に発表があった。モントリオール大のシステムは、以前の system-Q を中心としたシステムから、ATN-(Augmented Transition Network) モデルを基礎にした REZO というプログラム言語に移行しつつあり、かなり系統的に、長期の計画に従って研究をすすめているようである(ATN については BBN から ATN のコンパイラを作ったとの報告もあった)。モントリオール大のシステムは、技術論文の翻訳に関しては、かなり上質の結果を得ているようである。またエール大は、英語から中国語の翻訳システムを発表したが、これは実用システムを目指したものではなく、彼等の conceptual dependency structure から任意の言語を生成できるかをテストするための仕事との印象をうけ

た。

会議全体を通じて、招待講演の C. J. Fillmore, H. Thompson 等の言語学者が再び計算機による言語処理に興味を持ち出していること、また計算機科学の側からも単に既成の言語理論を計算機で実行するだけでなく、積極的に独自の言語モデルを構築しようとしていることなどが目立った。言語処理(というよりも言語科学)の将来は、心理学・言語学・人工知能といった多様な分野の間での交流を今後ますます盛んにすることによって発展してゆくと思われる。一方、現在までの成果を整理して体系化し、データベースその他へ応用してゆこうという動きも注目されるだろう。機械翻訳の研究がいくつかの伝統あるセンターで現在も地味ながら着々と研究され成果をあげつつあるのが印象的である。機械翻訳をあらためて見なおそうという動きも感じられる時代となって来た。いずれにしても計算機のできる範囲と、非常にむずかしい部分がかなり明確になって来て、ICCL はある意味で1つの曲り角にさしかかっているといえることができるだろう。

(昭和51年8月26日受付)

文献紹介

76-36 可変長命令方式による計算機デザイン

Eric C. R. Hehner: Computer Design to Minimize Memory Requirements

(IEEE Computer, Vol. 9, No. 8, pp. 65~70 (August 1976))

Key: language-directed machine design, iterative pairing, conditional coding, variable length binary sign-and-magnitude number scheme

現在の計算機の命令セットが、コンパイラやオペレーティング・システムの記述には、不適合であるといった指摘は従来から行われているが、本論文では、現実の命令セットから出発し、これを新しい命令セットに改善し、かつ使用頻度に基づく可変長符号を用い、大幅な記憶容量の削減を実現する方法について述べている。適切な可変長符号表現により、コンパイラも目的プログラムも小型化され、時間の嵩む記憶階層間の情報転送の高密度化が図られ、オーバーフローの危惧

から生ずる記憶域浪費から解放されるという。

可変長符号化方式でまず問題となるのは、命令域のアドレス単位の選択である。単位が小さい程、使用頻度分布に密着させられるが、分岐アドレスの表現には不利となる。またアドレスや即値データの項目は、最適符号が考え難い。若干の議論により、著者はアドレス単位は1ビットに近い程有利と結論し、この仮説をIBM 360の機械語60kバイトのXPLコンパイラで検証する。

命令列の出現依存性を利用し、命令対の1命令化、先行命令の種類によって解釈を変える条件符号化の採用、あるいは可変長2進数を利用した分岐アドレスの相対表現、XPLのブロック構造を利用したデータ・アドレス表現の採用により命令は47種から178種へ増加するが、命令域全体の容量は元の25%になったという。ちなみに、整数、データ・アドレス、分岐アドレスの平均長は、各々5.5, 7.3, 14ビットであったという。