



会議レポート

第 19 回計算言語学国際会議

COLING 2002 (The 19th International Conference on Computational Linguistics) が 8 月 24 日から 9 月 1 日の 9 日間、台北で開催された。ICCL (International Committee on Computational Linguistics) が主催する COLING は、自然言語処理の分野で最も歴史のある会議である。自然言語処理全体をカバーすること、ヨーロッパ、アメリカ、アジアの各国から多数の研究者が参加することが特徴であり、この分野で最も重要な国際会議となっている。アジアでの開催は 1980 年の東京、1994 年の京都に次いで 3 回目であった。台北が開催地に選ばれたことは、東アジア地域における自然言語処理のめざましい発展を象徴している。

8 月 26 ~ 30 日に福華国際文教会館で開かれた本会議には 32 カ国から 540 名が参加した。最近の COLING とほぼ同じ参加者数であったが、欧米からの参加者がやや減少した。開催国である台湾が多いのは当然であるが、日本からの参加者が目立った。

本会議では 3 件の基調講演があったが、COLING のスコープの広さを反映してバラエティに富んだ内容であった。最初の講演者は香港城市大学の William S.-Y. Wang 教授で、“Computer modeling of language evolution” と題して、言語の発生・進化に関する興味深い研究を紹介した。これまでは語彙や音韻に関する研究が中心であったが、言語の特徴である階層構造や曖昧性がどのようにして生まれたかの解明が期待されている。このような学際研究において計算言語学の果たす役割は大きい。2 番目の講演は、ザールランド大学 / DFKI (ドイツ人工知能研究センター) の Hans Uszkoreit 教授による “New chances for deep

linguistic processing” であった。HPSG (Head-driven Phrase Structure Grammar) の大規模な文法・パーザの研究開発プロジェクトの成果をふまえて、自然言語処理のタスクに深い処理を取り入れる時期にきていると主張した。このプロジェクトはスタンフォード大学、東京大学と協力して進められたもので、国際協力の重要性も力説された。最後の講演者は、格文法の創始者として有名な Charles J. Fillmore 教授 (カリフォルニア大学バークレー校) であった。“FrameNet and the linking between semantics and syntactic relations” と題して、教授がこの数年間率いてきた FrameNet プロジェクトを紹介した。語の概念構造をフレームとして記述するとともに、British National Corpus などの例文に対するアノテーションを行っている。構文と意味をつなぐ大規模な言語リソースとして、自然言語処理の今後の発展に資するものと期待される。

本会議で発表された一般論文は 198 編 (うち 28 編はプロジェクトノート) で、前回よりやや増加した。投稿数は 435 であり、45% の採択率は過去の COLING よりやや高かった。著者の所属機関の国別では、日本 46、米国 42、ドイツ 18、英国 14、フランス 13、韓国 10、中国 9、台湾 8、スペイン 5、カナダ 5 と続き、東アジアの躍進が目立った。日本が最多であったのは 1996 年のコペンハーゲン大会以来である。一般論文のセッションは 5 並列であった。いつもの COLING と同様に 1 論文に 30 分が割り当てられ、密度の濃い討論が行われた。

全体的な傾向としては、統計あるいは確率的なアプローチの論文が過半数を占めていることが目立った。この傾向は過去数回の COLING に共通である。コーパス (言語データ) に基づく自然言語処理という潮流は今後も続くと思われる。アジアの言語を扱った論文が多かったのも今回の特徴である。韓国、中国・香港、台湾、タイなどにおける活発な研究を反映している。

分野別の傾向を見ると、構文やパーズングの関連では、深い処理の効率向上方法、深い処理と浅い処理の統合、コーパスのアノテーションなどに関する論文が多かった。英語以外の言語の Treebank (構文情報が付与されたコーパス) も充実してきている。アノテーションの半自動化、語彙知識の獲得、ドメインへの適応手法などが今後の研究方向である。Treebank などのリソースやツール、評価データの共有も重要な課題である。

意味の関連では、語義の曖昧性解消、類義語などの概念獲得に関する論文が多かった。語義の曖昧性解消は、多くの自然言語処理タスクに必要であり、きわめて活発に研究されている。教師あり学習方法を用いた曖昧性解消の研究が多いが、タグ付けのコストが大きな問題であり、教師なしの方法に対する期待が大きい。概念の獲

得では、コーパスからドメイン固有の概念を抽出して、WordNet など既存の概念体系の中に位置づけるというアプローチが多い。

機械翻訳、情報検索、テキスト分類、情報抽出、質問応答など、自然言語処理の応用をセッション名としたセッションも多かった。報告者の記憶では、機械翻訳以外の応用が COLING のセッション名に使われたことは従来あまりなかった。自然言語処理の応用が広がってきたことを物語っている。論文の数や内容から、特にオープンドメインの質問応答の研究が急速に進展していることがうかがえる。質問応答は、情報検索と情報抽出の結合を超えて、さまざまな要素技術、深い処理を必要とするアプリケーションである。今後の発展が期待される。

パネル討論のテーマは、次世代 WWW として最近話題になっているセマンティック Web と言語処理技術の関係であった。オントロジーの構築における言語処理の役割、情報抽出などのアプリケーションにおけるオントロジーの利用、言語リソースとしてのセマンティック Web などが議論されたが、結論には至らなかった。

本会議前の 8 月 24、25 日に中央研究院で行われたチュートリアルも盛況であった。「バイオインフォマティクスと自然言語処理」、「オープンドメイン質問応答」

など、自然言語処理研究者の関心が高いテーマについて、最先端の研究者による興味深い講演が行われた。本会議後の 8 月 31 日と 9 月 1 日には、同じく中央研究院を会場として、11 のワークショップが開催された。1 日間のワークショップが 9 つ、2 日間のワークショップが 2 つであった。「アジアの言語リソースと国際標準化 (第 3 回)」、「アジアの機械翻訳」、「中国語処理 (ACL SIGHAN, 第 1 回)」の 3 つのワークショップは台北ならではのテーマであり、アジア地域の多数の研究者が活発に議論した。その他のワークショップのテーマは、最近の研究動向を反映して「意味ネットワーク」、「ターミノロジー」、「多言語要約・質問応答」、「文法工学」、「自然言語の学習」などであった。

COLING 2002 は、日本以外のアジア諸国の自然言語処理を世界に知らしめた意義深い会議であった。アジアにおいてもヨーロッパと同様に国際的な研究協力が進む契機になると思われる。アジアの先頭ランナーとして日本のイニシアチブが期待される。なお、次回の COLING 2004 はジュネーブで開催される。

(梶 博行 / 日立製作所中央研究所)