

書 評



田中穂積 (監修)

自然言語処理 一基礎と応用一

(社) 電子情報通信学会, 1999年, 5,400円 (税別)

ISBN4-88552-160-2

本書は、自然言語処理技術の基礎から応用まで広い範囲をカバーした入門書である。基礎技術と応用技術の2部構成になっており、基礎技術を取り扱った第I部は以下の6章からなる。

- 第1章 形態素、構文解析
- 第2章 意味解析
- 第3章 談話解析
- 第4章 文章生成
- 第5章 辞書
- 第6章 コーパスベースの技術

第1章では、古典的な解析アルゴリズムに加えて、統計的言語モデルに基づく手法について詳しく説明している。第2章では、意味構造の表現形式、解析に用いる知識、解析手法という観点でまとめている。機械翻訳やクロス言語情報検索における訳語選択への応用にも触れている。

第1, 2章は文単位の解析であったが、第3章は文の纏まりである談話を分析対象とする。談話構造の決定、照応や省略等の処理、プラン認識に基づく対話の理解について詳述している。

第4章では、第3章までのテキスト理解とは逆に、話し手の意図からテキストを生成する技術を紹介している。話し手の意図から伝達内容を決める深層生成、伝達内容をテキストに変換する表層生成という流れで説明している。

自然言語処理システムは、辞書やルールの質と量が性能を大きく左右する。第5章では、各種解析処理で用いる辞書の内容について詳細かつ具体的に論じている。構築支援技術や実装法は割愛されている。

第6章では、コーパスからの言語知識の抽出法と、抽出した知識の解析処理への利用という2つの側面から研究動向をまとめている。

応用技術を取り扱った第II部は、以下の5章からなる。

- 第7章 文書処理
- 第8章 機械翻訳
- 第9章 対話システム
- 第10章 音声対話システム
- 第11章 マルチモーダル対話システム

第7章では、校正支援と文字認識後処理を取り上げている。仮名漢字変換技術については、形態素解析技術の応用として第1章で触れている。

第8章では、機械翻訳の3つの基本的な処理方式と、新しい枠組みについて詳しく説明している。

第9章では、自然言語を用いて人間と対話を行う対話システムに関する技術全般についてキーボードベースのシステムを例に説明している。

第10章では、音声認識技術を説明した後(音声合成は割愛)、音声言語・音声対話の特殊性について論じている。

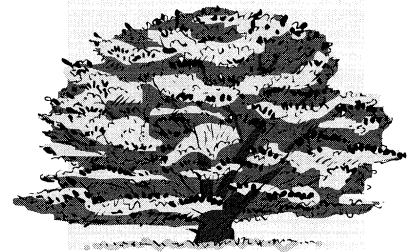
第11章は、言語に加えて動作や視覚的な情報を用いるマルチモーダル対話システムである。複数のメディアの有機的な連携を強調しながら、さまざまなシステムを紹介している。

情報検索や情報抽出、要約等が特に取り上げられていない一方で、このところ研究が活発化しつつある対話システムに3つの章を割いて大きく取り上げているところが本書の特徴といえよう。本誌4月号でも特集が組まれる等、近年自然言語処理技術はその応用範囲を急速に広げており、1冊の入門書でそれらを網羅的に取り上げるのは困難になりつつある。本書のように思いきって対象を絞り込んでみるのも面白い編集だと思う。

本書は12人の執筆陣による分担で書かれている。第一線でご活躍中の第一人者ばかりである。特にコーパス関連の技術等ここ数年で大きく進展した分野では、豊富な参考文献とともに研究動向が簡潔にまとめられており、初学者以外にも大いに役立つ内容となっている。また、語り口に執筆者ごとの個性がよく出ていて、そういう点も楽しみながら読ませていただいた。

約350ページの分量に盛りだくさんの内容なので、所々説明が駆け足にならざるを得ない箇所もあるが、自然言語処理技術の現状を概観するには最適な書物である。本書を通じてより多くの人が自然言語処理技術への関心と理解を深めてくれることを期待する。

[出羽達也/東芝 研究開発センター]



会議レポート

TMI 99参加報告

第8回目のTMI (International Conference on Theoretical and Methodological Issues in Machine Translation) ¹⁾ が1999年8月22日から26日の5日間にわたって英国チェスターのチェスター大学で開催された。TMIは機械翻訳を主眼とした国際会議であり、1985年第1回会議が米国で開催されて以来2年に1度開かれている。機械翻訳に関する代表的な会議としては、TMIの他にMT summitがあるが、MT summitは研究者だけでなく機械翻訳システム販売者や翻訳家などの利用者也参加するビジネスの色合いの濃い会議であるのに対し、TMIは機械翻訳の研究、開発に携わる人々が学術論文を発表する会議である。テーマが機械翻訳ということから開催地に偏りができないよう配慮され、北米、ヨーロッパ、アジアから順に選ばれている。また運営母体となる組織はなく、各会議ごとに機械翻訳研究者から運営委員会が組織される。

今回のTMI 99は、初日がチュートリアルセッション、中3日間が本会議、最終日が英独機械翻訳に関するワークショップという構成だった。参加者数は最近のTMIと同程度の70余名で、国際会議としてはこぢんまりとまとまっている印象を受けた。チュートリアルは、2つのテーマが並行して行われた。1つは“Acquisition of Knowledge about a Low-Density Language for

Use in MT”で、電子化された言語リソースの量が少ない言語との翻訳知識を獲得する問題が取り上げられた。もう一つは“Introduction to MT and Translation Tools”で localisation や、TRADOS 社の翻訳支援ツール、EUでの翻訳サービスについての話題であった。本会議は招待講演2件と総投稿数43件から選ばれた一般講演22件からなっていた。講演はシングルセッションで進行していくため、聴講がたいへんしやすかった。ヨーロッパでの開催であったが、アジア、北米からが多かったのが今回の特徴である。対象となる言語も、日本語、英語、中国語、韓国語、スペイン語、ロシア語など多様である。

機械翻訳専門の国際会議であるが、今回は機械翻訳の方法論そのものよりも、自然言語処理一般の研究でも主流のテーマであるコーパスからの知識獲得に関するものが目立った。異なる言語で対訳となる表現を獲得する手法や、用言に意味属性を自動的に付与する手法、翻訳対象となる言語表現の語義の曖昧性を解消する問題を扱ったものなど興味深いものが多かった。WWWで流通するテキストを始めとして機械処理に利用できる電子化された文書量の増加に伴って、このテーマは今後機械翻訳においても1つの主要な流れとして続いていくと感じた。翻訳手法に関するものでは、用例を扱ったものが多く、用例と語彙知識を組み合わせるもの、時制、様相（「だろ

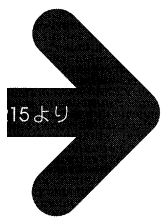
う」「かもしれない」など話者の認識、判断を表す表現）、アスペクト（「ている」「てしまう」など動作の継続、完了などを表す表現）を用例の利用により翻訳するものがあった。本会議ならではの少し異色の論文としては機械翻訳システムが訳出した結果に信頼度をつけ、信頼性の低い文を除いてユーザに提示する方法について述べた“A Confidence Index for Machine Translation” (Arendse Bernth)があった。他に翻訳プロジェクトにおいて、さまざまなリソース（機械翻訳システムや各種ツール、言語リソース、人間の翻訳家など）を組み合わせた翻訳過程を決定するために行うプロジェクトのプロファイリングについて述べた“Profiling Translation Projects” (Nancy L. Underwood et al.) も興味深かった。

次回TMI 2001はアジアが開催地となる順番であるが、京都のNTTコミュニケーション科学基礎研究所で行われる予定である。これを機会に日本の機械翻訳研究が一層活性化することを期待したい。

参考資料

1) <URL:<http://www.ccl.umist.ac.uk/events/tmi99/>>

(田中貴秋/NTTコミュニケーション科学基礎研究所)



これら一連の裁判で明らかになったのは、わいせつ図画は、ぼかしがかかっているとしても、それが簡単にはずせるものなら、公然と（ホームページ）に陳列してはいけないということである。しかしリンクの是非についてはまだ裁判所の判断は出ていない。牧野氏はK氏は無罪だと主張している。リンク集はサーチエンジンでも自動的にできてしまうくらいだから、この事件でもまさか有罪にはならないだろうと私も思う。

ところで、この裁判の直前には、大阪地裁にほど近い岡村（久道）法律事務所でも牧野氏らと打合せを行った。聞けば、岡村氏は弁護士界では恐らく随一の（いい意味での）ハッカーらしく、事務所においてあるサーバには自分でUNIXをインストールし、NTTのOCN専用線でインターネットとつないでいるというからすごい。岡村氏のドメイン名はなんとwww.law.co.jpである。law（法律）をいち早くとったとは流石である。そのホームページには法律（特にインターネット関係の）に関するさまざまな情報やリンク集があって、参考になる。

また、この打合せで分かったのは、K氏の弁護に、牧

野先生をはじめとする約20人の弁護士達が手弁当で協力しているということである。彼らと話していて、インターネットをなるべくお上からの制約のない形で、自己責任で使ってゆきたいという熱意が感じられて共感を覚えた。

最後に、この種の裁判が、なぜ大阪にあって、東京ではないのかということ、風俗犯罪は、大阪が本場で、先進的だからという。新しい風俗は大阪で発明され、東上する。今度の証言行為はそんなことも学んだ貴重な体験だった。

