

特集

# フィールドを広げる自然言語処理

中野 幹生 NTTコミュニケーション科学基礎研究所  
那須川哲哉 日本アイ・ビー・エム (株)

自然言語処理とは、日本語や英語のように人間が日常のコミュニケーションに用いている言語を、計算機で理解したり生成したりする技術である。1990年代の初め頃まで、自然言語処理技術の応用の中心的なものはコンピュータによる自動翻訳（機械翻訳）だった。また、基礎研究として、構文解析（文の文法構造の解析）など、特定のアプリケーションによらない汎用的な要素技術の研究が行われていた。

数年前から自然言語処理研究をとりまく環境は大きく変化してきている。1つには、ワープロをはじめとする入力環境の整備に加え、電子メール、グループウェア等により文書の電子化が急速に進み、さらには、WWWの普及によってアクセス可能な文書データが膨大な量に及んだことがあげられる。実際、膨大なメールや文書データの山に悩まされている読者も多いのではないだろうか。さらに、音声認識やOCRなどのキーボード以外の入力技術の性能が向上したことがあげられる。特に音声認識に関しては不特定話者の連続音声認識（単語と単語の間にポーズをおかなくても認識できること）が可能になった。

これらのことは、自然言語処理が対象としてきた、言語で表される情報が増大したことを意味する。さらに、翻訳だけでなく、さまざまな形で処理することが求められてきており、多様なアプリケーションが研究の対象になってきた。たとえば膨大な文書の山から簡単に欲しい文書を見つけたり、文書の山そのものを知識源として有用な情報を発見・抽出する技術の需要が高まっている。また、カーナビなどでも音声で対話できる機能が取り入れられ始めている。これに応じて、形態素解析（テキストの単語分割）や構文解析などの要素技術の研究も、多様なアプリケーションに対応できるように変化してきている。本特集の解説は、以上のような状況で、現在活発に研究が行われている（または行われ始めている）分野

に関するものである。

以下に、各解説を簡単に紹介する。第1編では、大量の文書を自動分類し、興味のある文書を検索することを可能にする技術が解説されている。第2編では、自然言語処理技術とデータマイニング技術との融合により、大量の文書データにおいて記述されている多様な事象の情報を抽出し、各事象の特徴や傾向を分析する技術が紹介されている。第3編では、音声処理技術と言語処理技術との融合による音声対話システムの実現法が解説されている。第4編では、大量の文書データから、企業の人事異動のような定型的情報を抽出する技術が解説されている。第5編では、対訳になっている文章から機械翻訳システムの構築や英作文支援に使えるような辞書データの自動構築法が解説されている。第6編では、大量の文章の文法的な解析を高速に行うことができる技術が解説されている。なお、これらの解説は、エディタの主観により、应用到に近い（と思われる）ものから順に並べた。

言うまでもなく、本特集で取り上げたものが自然言語処理研究の最近の話題のすべてではない。たとえば、文書データや対話データのタグ付けとそれに基づく分析、高速な形態素解析、マルチモーダル対話システム、テキスト読み上げ技術、多言語情報処理、文章要約など、多くの技術をあげることができる。これらに関しては当会誌や他の学会の学会誌などに解説記事が掲載されたものも少なくないので、それらを参照いただきたい。本特集では比較的新しい分野のものを取り上げたが、以前から研究されているアプリケーションも着実に進歩している。そのようなアプリケーションの1つである機械翻訳システムの最近の進展については、大山芳史氏による本号の「道しるべ」をご覧いただきたい。本特集がきっかけとなって自然言語処理の新しい技術に興味を持ってくださる方が増えれば幸いである。

(平成11年3月19日)