

日本語を科学する

Quantitative Research and Analysis on Japanese Text Data Base

— View Point of Psycholinguistics and Cognitive Science —

野 崎 浩 成

Hironari NOZAKI

情報科学系大講座

Information Science

Email : hnozaki@cc.nagooya-cu.ac.jp

1. はじめに

本稿は、1997年12月16日に計算センター会議室で開催された情報科学系大講座主催の談話会において、「日本語を科学する」というテーマで講演した内容をまとめたものであり、筆者の研究パラダイムを紹介することを目的とする。なお、当日の講演では、参加者を被験者とする模擬心理学実験を実施し、心理学の研究法論を具体的に論じたが、今回は紙面の都合で模擬実験の解説は割愛する。

2. 研究対象領域 —— 言語情報学とは何か。

筆者が研究対象領域とする言語情報学は、我々が日常生活で用いる日本語や英語など、いわゆる自然言語を科学的に取り扱おうとする学問分野である。その研究成果は、ワープロの漢字変換効率の向上や自動翻訳システムの開発、郵便番号の自動読み取り装置にも応用されているOCRの文字認識率の向上など、私たちの日常生活場面と密接に関連している場合が多い。その中でも、私が最も興味を持って取り組んでいる課題は、日本語の語彙シソーラスの研究である。シソーラス (Thesaurus) とは、電子化された辞書の一種であり、情報検索を実現するための語彙を意味の上から分類して配列したもので、コンピュータに格納する知識ベースである。自動翻訳システムのようにコンピュータを用いて日本語を翻訳するには電子化された辞書 (シソーラス) が必要であり、電子化辞書の設計が翻訳システムの性能を左右するといっても過言ではない。また、コンピュータにどのような知識ベース (辞書) を与えれば人間のような知的な言語活動が実現できるのか、さらに発展させると、人間はどのようにして言語を獲得するのか、母語と第2言語ではその習得過程が異なるのか、人間が言語を導出するメカニズムは如何にすれば解明できるのか等々、興味の尽きない話題は少なくない。

3. 研究方略 —— 認知科学的研究アプローチ

筆者の研究方略は、認知科学的なアプローチを用いている。認知科学とは、人間が行う「知的な情報処理過程」を解明する学問である。ここでいう知的な情報処理とは、外界の状況を知ること (知覚)、経験したことを覚えたり思い出したりすること (記憶)、問題を理解したり解決したりすること (学習・思考) などである。周知の通り、記憶および学習・思考は、人間が行う言語活動によって支えられている場合が多く、認知科学が目指す「人間の知的な情報処理過程を解明する」ためには、言語を研究対象領域とすることは必要不可欠である。このような理由から、筆者は前節で述べた通り、自然言語を研究対象領域とした。

認知科学のもう一つの特徴は、情報処理モデル (Information Processing Model) を扱うことにある (図1)。すなわち、人間を情報処理システムとみなして、その情報処理過程を記述するモデルを構築した上で、そのシステムの内部では何が起きているのかを解明することである。その成立には、次に示す2つの研究アプローチがある。その第1番目の立場は、情報処理システムがどのようなハードウェア的な仕組みを持っているのかを明らかにすることである。これは情報処理システムへの入出力関係からシステム内部の機能を推察するものである。第2番目は、情報処理を行うためのプログラムやデータ構造は何かを明らかにする立場で、人間の内的な知識表現を推察するものである。これらをコンピュータにたとえるならば、前者はハードウェア、後者はソフトウェアに該当する。従って、認知科学的な研究アプローチとは、コンピュータをアナロジーとして、人間の知的な活動を情報处理的観点から研究することといえよう。

4. まとめ

本稿では、筆者の研究対象領域と研究方略を簡潔に紹介した。自然言語を対象に、科学的な分析を行うという

のが、筆者の研究パラダイムである。今日、印刷物として公刊されている書籍や辞書は、CD-ROM化されるようになった。また、インターネットを始めとする情報通信基盤のインフラ整備が進み、膨大な日本語テキストデータが電子メディアを介して世界中に流通し始めている。このような時代の流れの中で、電子メディアにおける日本語を科学的に分析する必要性を感じている。

なお、研究の詳細については、学術雑誌に掲載された筆者の研究論文をご参照いただきたい。

