

# 言語学におけるコーパスの位置づけ

## —コーパス言語学の現状と今後の展開—

富 谷 玲 子

今回の講演では、現代の言語学において重要な役割を持つ「コーパス」（大規模な言語データベース）をめぐって、言語学におけるコーパスの位置づけや、コーパスを使った分析の具体例などが紹介された。講師には、現在コーパスの構築作業に携わっている、国立国語研究所の丸山岳彦氏をお招きした。講演の主なテーマは、(1)「言語学におけるコーパスの位置づけ」、(2)『『日本語話し言葉コーパス』の紹介」、(3)「KOTONOHAプロジェクトの概要」であった。

「コーパス」とは、言語研究に用いることを目的として編纂された、大規模な言語データベースのことである。「コーパスに基づく言語研究」という考え方自体には、すでに50年以上の歴史があるが、特に1990年代半ば以降は、コンピュータの高性能化と爆発的な普及を契機として、大量の言語データから一定の傾向や規則性を見出そうとする研究が盛んに行われるようになった。「コーパス言語学」と呼ばれるこの研究分野は、コーパスが早期に整備されたイギリスによって主導されてきた。日本ではコーパスの整備が遅れたことから、日本語を対象としたコーパス言語学は完全に立ち遅れている状況にある。講演では、まず、20世紀における言語学の動向と、その中でのコーパスの位置づけについて、概説が行われた。

次に、国立国語研究所が2004年に公開した『日本語話し言葉コーパス』が紹介された。これは、661時間もの自発的な話し言葉を録音し、テキスト（文字）に書き起こした上で、品詞情報など様々

な研究用情報を付与したものである。講演では、『日本語話し言葉コーパス』を用いた音声現象の分析例や、話し言葉の構造をどのように記述・解明していけばよいかなどについて、「ら抜き言葉」の使用実態、「～ですね」の使用実態に関する分析など、具体的で興味深い研究成果の紹介があった。

最後に、現在国立国語研究所が進めている「KOTONOHAプロジェクト」についての概説があった。この研究計画は、現代日本語の書き言葉をバランスよく集めた「均衡コーパス」を作るという、日本では初めての試みである。研究期間は2006年から5年間で、コーパスの規模は1億語を目指すという。

実際にコーパス構築に関わっている現役の研究者を招聘することにより、コーパスがどのようなものか、どのように構築し、どのように使うのか、などといった具体的な情報が豊富に示され、今後のコーパス研究への期待を高ませる講演となった。当日は学生・教員合わせて約50名もの出席があった。特に学生にとっては未知の新領域に関する高度な内容であったが、具体的で興味深い研究成果が次々と紹介されたため、最後まで熱心に聞き入っていた。講演終了後も内容の濃い質疑応答が続き、盛況な講演会となった。講師の丸山氏は本学の卒業生である。この講演会では研究者として第一線で活躍中の先輩から、研究の最前線について直接お話を聞くことができたが、これは言語に関心を持つ学生にとって非常に大きな刺激になったものと思われる。