
言語研究センター共同研究

学習者コーパスを活用した上級日本語教育の教材開発

富谷 玲子・高木南欧子

2004年に公開された『日本語話し言葉コーパス』(国立国語研究所・情報通信研究機構・東京工業大学)、2011年に完成した『現代日本語書き言葉均衡コーパス』(特定領域研究「日本語コーパス」)に代表されるように、近年、現代日本語のコーパスの開発と公開がすすめられている。日本語教育においてもコーパスの存在は広く知られるようになり、さまざまな研究グループによって開発、公開がなされ、研究に利用され始めている。日本語教育では、これらの基礎資料を統語構造の分析への応用や、教材評価、コースデザインの開発など多岐に渡る応用がなされるようになったと同時に、比較参照するための学習者コーパスの必要性も求められるようになった。学習者コーパスとは、第2言語学習者の作文や発話データを大量に収集、分析に必要となる情報を付加し、検索が可能な形に電子化したものを指す。対照言語研究や誤用分析、教室活動の分析などから得る知見を教室活動に還元する際、学習者コーパスを参照することにより数量的な分析が可能になる一方、従来とは異なった知見を得られる可能性もある。本研究では、学部留学生に対する日本語教育への還元、初年度教育としての日本人学生に対する日本語教育への還元を可能にする基盤構築を目的としている。現在、一般に公開されている日本語の学習者コーパスは「日本語学習者による日本語作文と、その母語訳との対訳データベース(作文対訳DB)オンライン版」(国立国語研究所)、「KYコーパス」(OPIテープを文字化した言語資料)などがあり、進行中のプロジェクトなども多々見受けられる。これらの学習者コーパスを構築という点でみた場合に

共通してみられる問題点は、学習者の母語や学習歴などがさまざまである上に、書き起こされたテキストや発話データに対するタグの付与に非常に時間がかかること、また、言語運用場面や目的が多岐に渡ることである。しかし一方で、細かい分類と膨大な基礎データの支えがなければ、期待する抽出結果も得られにくくなり、コーパスとしての有用性が低下する。解決策の一つとして、抽出結果を上記にあげた『現代日本語書き言葉均衡コーパス』とつなげ、参照を可能にし、付与される情報の客観性、信頼性を向上させる試みや、特定分野の専門辞書を組み込むことによって精度を向上させる、辞書の更新・増設を可能にするシステムを採用するなどの試みがみられる。これらの試みは、学習者コーパスの設計には非常に参考になると思われる。

