

修士論文要旨

学籍番号 22GH201

第 号

氏名

尾野 海琴

人文社会科学専攻 (コース: 現代共生)

論文題目

Creation of Tsugaru Dialect Dictionary Database Based on Tsugaru Dialect Corpus and Proposal of Usage as Educational Tool

2022年6月時点で、日本国内には2,961,969人の外国人が居住している。一方で青森県内における外国人居住者数は6,306人まで落ち込む。他県と比較しても広い領域を持つ青森県内に少ない外国人居住者が散在する現状は、外国人日本語学習者への言語的サポートを困難にしていると考えられる。このような状況下において、地方在住の外国人、外国人留学生や外国人労働者といった日本語非母語話者にとって日本語はコミュニケーションの大きな壁になる。特に、共通語を学びつつある日本語学習者（非母語話者）にとって方言はコミュニケーションの大きな障害になっている。この要因としては既存の方言集・方言辞典が日本語学習者にとってわかりにくいものである、という可能性が考えられる。また、津軽方言の語彙そのものに注目している研究、その中でもコーパスに基づいて方言の辞書データベースを作成している研究が少ないことが挙げられる。

そこで本稿はこのような日本語学習者を取り巻く状況を少しでも改善し、この学習ツールが学習者の理解の一助になることを目指し、その達成のために津軽方言の民話データにより構成されている方言コーパスに結び付けられた方言辞書データベースを作成し、このコーパスと辞書データベースを学習ツールとして利用する方法を提案した。

本研究では、方言コーパスとしてNPCMJプロジェクト(NINJASL Parsed Corpus of Modern Japanese)にて作成されたThe Matsunoki Treebankを使用した。この方言コーパスは、弘前市周辺地域で活動している語り部グループである、「和の昔コ」による民話の語りを録音し、それをローマ字に書き起こしたデータを使用している。ローマ字による書き起こしには、ヘボン式ローマ字を改訂したものをを用いることで、非日本語母語話者の学習者にとって実際の津軽方言の発音がわかりやすいよう工夫を施した。

津軽方言辞書データベースはひな形に沿って作成した。ひな形は主に見出し語の大まかな意味を表す単語と数字で構成された"key"、単純化された品詞の情報、見出し語の定義、発音のバリエーション、同義語、方言コーパスから抽出された例文によって構成され、品詞ごとに追加の情報が付与される。名詞には、写真と名詞の特性を表した"ENTITY"という情報が、動詞には日本語の動詞の言い換え関係を記述することを目的とするシソーラスのフレームIDが追加されている。辞書定義はコウビルド形式の定義を参考にし、英文中に見出し語を埋め込む方法により記述した。

また、辞書データベースの作成と同時並行的に、方言コーパスに津軽方言民話スクリプトやコーパスの解析に基づいた英語訳を付け加えることによって、方言コーパス自体の拡充・改良も進めた。

今後はコーパス中に登場する全単語への定義づけを行うことで、学習者がより利用しやすいツール作りを実現していく。また、新たな見出し語とその定義、津軽方言の地域差や年代差の情報や、津軽方言辞書データベースへ適切な例文を付与することも検討している。

本研究で作成した津軽方言に関するコンテンツを用いることで、日本語学習者がより容易に津軽方言に関する情報を得ることが可能になるのではないかと考える。