

古典作品翻訳のデータ構造化と利活用に関する研究 －『源氏物語』を対象として－

池田 ひかり

近年、国立国会図書館デジタルコレクションや青空文庫などにより古典文学作品のデジタル化が進んでいる。しかし、翻訳については、訳者や訳出範囲といった詳細な情報を既存の書誌情報から十分に把握できないという課題が存在する。本研究では、日本古典文学作品の翻訳に関する情報の構造化と利活用を目的とし、『源氏物語』の英訳を対象として、LOD (Linked Open Data) に基づいてデータモデルの設計とデータセットの構築・公開を行った。

本研究では、特に訳出範囲や訳者の情報に注目し、出版物・訳者・訳出範囲の三種のリソースから成るデータモデルを設計した。語彙には国立国会図書館ダブリンコアメタデータ記述 (DC-NDL) を中心に、Dublin Core、FOAF 等の既存語彙を採用し、出版物と訳者、訳出範囲の対応関係を記述できる構造とした。データセットの構築にあたっては、国立国会図書館サーチや CiNii Books を用いて『源氏物語』の英訳の書誌情報を収集したほか、出版物の複写やデジタル版等を活用して訳出範囲の詳細な調査を行った。

データセット構築の結果、出版物リソース 71 件、訳者リソース 9 件、訳出範囲リソース 54 件、合計 3455 トリプルからなる LOD データセットを作成できた。さらに、Web 上でデータセットの公開を行い、出版物・訳者・訳出範囲を相互に参照可能な形で提示した。また、想定したユーザーシナリオとそこから考えられるユーザーのニーズに基づき、必要な機能と SPARQL による検索・利活用方法を検証し、特定の訳者や訳出範囲、出版年に基づく情報の取得が可能であることを示した。

本研究は、『源氏物語』の英訳を構造的に整理し、翻訳研究や学習、読書支援などへの応用可能性を示した。また、今後『源氏物語』以外の作品や英語以外の出版物も対象としてデータセットを構築することができれば、より多くのユーザーのニーズを満たすことができるようになると考えられる。一方で、データ収集の多くが人手での作業に依存している点や、データセットを利活用するための応用アプリケーションの実装などが今後の課題として挙げられる。

(指導教員 高久 雅生)