

唐詩の構造化に関する研究  
— Linked Data と TEI マークアップ  
を用いて —

A Study on Structuring Tang Poems with  
Linked Data and TEI Markup

学籍番号: 201721707

氏名: 叢 艶

CONG Yan

唐詩は中国古典文化資源の一部として、現在までは千年以上の歴史がある。唐詩は中国の文化遺産として、中国古典文学研究に欠かせない基本的な文献であり、日本の文学にも大きな影響を与えた。

唐詩作品は中国では常に学ぶ作品である一方、唐詩作品に関わる情報を利用できるデータベースがばらばらに各所で作られ、研究者や教員、学生などが使えないという問題があり、学習のための利便性も低いと考える。また、日本の中学校と高等学校の教科書に、含まれる同一の作品であっても訓読が違うものがあり、詩の意味の捉え方に大きな影響を与える点や、文法が一致しない点が指摘されてきた。そのため、教育ニーズに応じて、唐詩情報は基本的な情報の習得に加えて、本文フルテキストにおいて、違う訓読情報やルビ情報がつけられていても、唐詩作品に関わる情報を統一的に扱える学習環境が望ましい。

このような唐詩情報の共有や学習のための課題を解決するためには、唐詩作品の構造化が望ましい。唐詩作品の構造化については、唐詩における多様なニーズに応じて、多様な関連情報を簡潔に扱え、自由に利用できることが必要である。

本研究では文化資源に注目し、唐詩作品を研究対象として、唐詩の構造化を目指す。教育学習のニーズに応じて、異なる表現を統一的に扱う学習環境を提供したいと考える。研究手法として、唐詩の Linked Open Data (LOD) 化と TEI マークアップを用いる。

具体的には、平成 28 年度使用の日本の中学校と高等学校で掲載された教科書に、含まれる唐詩作品異なり 53 首、延べ 374 首があり、作者は異なり 21 名の対象データを利用した。これらのデータセットを用いて、唐詩作品のタイトル、詩体などの情報や、唐詩を掲載される教科書の関連関係と DBpedia、Wikidata、CTEXT などの外部リソースと繋げて、LOD 化することを試みた。LOD 化の構築にあたって、BIBFRAME Model に基づいて、新たなメタデータモデルを構築した。それらの一連のデータセットを公開用のウェブサイト自由に利用できるように提供した。

唐詩作品の本文フルテキストに関わる訓読文、書き下し文を中心に、訓点情報やルビ情報などの要素を対象として、TEI マークアップを用いて、標準的にマークアップする手法を提案した。LOD 化したデータセットの中から、6 首を選んで、マークアップできることを示した。

今後の課題としては、教育ニーズに基づき、唐詩作品の応用アプリケーションを設計したいと考える。また、唐詩作品の対象データを拡大し、多くの対象データを公開できるようにしたい。そのほか、外部リソースの関連や、可視化などの機能も追加したいと考える。

研究指導教員: 高久 雅生  
副研究指導教員: 宇陀 則彦