

## 唐詩情報の Linked Open Data 化とその利活用の試み

### A Prototype of Linked Open Data Application for Tang Poems

叢 艶<sup>1</sup>, 江草 由佳<sup>2</sup>, 高久 雅生<sup>3</sup>

Yan Cong<sup>1</sup>, Yuka Egusa<sup>2</sup>, Masao Takaku<sup>3</sup>

<sup>1</sup>筑波大学大学院図書館情報メディア研究科

University of Tsukuba

<sup>2</sup>国立教育政策研究所研究企画部教育研究情報推進室

National Institute for Educational Policy Research

<sup>3</sup>筑波大学図書館情報メディア系

University of Tsukuba

**Abstract:** In recent years, a linked open data technology has emerged and many cultural institutions used it for making their information open to the public. It allows people to consume and reuse their information freely. This paper focuses on information of cultural resources with tang poems which are written in Tang Dynasty of China from 7th to 10th centuries. This paper proposes a data model and its prototype for tang poems as linked open data. The target of this research is the tang poems which are used in the textbooks of Japanese junior high schools and high schools. The proposed model covers metadata on tang poems, authors, poem styles, textbooks and their relationships. The data model utilizes several standard vocabularies such as BIBFRAME, Schema.org, and Dublin Core as well as its own vocabulary. We collected a data set of 374 tang poems, which consisted of 59 unique poems, from the 25 textbooks used for junior high schools and high schools in Japan as of 2016.

## 1 はじめに

近年、情報技術が発達するとともに、提供される情報も段々と増加している。これに対して、大量の情報の中からデータを正確に抽出できる Linked Open Data (LOD) [1] と呼ばれる方法が注目を集めている。

また、これを用いた行政や文化機関による市史や文化遺産などのデジタル化も進んでいる。例えば、横浜市は美術館所蔵品データ [2] の Linked Open Data 化を公開し、利活用されている。2016年2月から中国の上海図書館は所蔵デジタル資料を共有して、上海図書館公開サイト [3] 上で提供している。それらはオープン化と再利用によるデータの価値向上を目指すもので、誰でも使える情報が提供されている。

中国には、唐詩という伝統的な詩の形態がある。唐の時代(618-907)に、李白や杜甫をはじめとする多くの詩人が現れ、優れた詩が数多く作られた。それが唐詩である。唐詩作品は中国古典文化資源の一部として、現在まで千年以上の歴史がある。それは日本の文学にも大きな影響

を与えた。

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、日本の中学校と高等学校で学習する唐詩作品、それらを掲載する教科書の関連関係を Linked Open Data 化することを試みる。Linked Open Data 化の構築にあたって、BIBFRAME Model に基づいて、唐詩作品を研究対象とするモデル化も行い、これらを報告する。

## 2 対象と方法

本研究では唐詩作品を研究対象とする。唐詩作品は中国の文化資源として、現在まで千年以上の歴史がある。唐詩は中国古典文学研究に欠かせない基本的な文献である。本研究では中国の文化資源に注目して、唐詩作品を研究対象として、日本の中学校と高等学校の国語と古典の現行教科書に掲載される唐詩作品、唐詩作品を含む教科書とその関連付けを Linked Open Data 化する。

本研究では、BIBFRAME (Bibliographic

Framework) Modelの応用も目指す。BIBFRAME Modelは2012年に米国議会図書館が公表した書誌情報のためのデータモデルとボキャブラリであり、最新版は2016年4月に公開されたBIBFRAME 2.0 Model[4]である。これはLinked Dataのモデルに沿って、新たな文献目録の仕組みを作成する基準である。2016年4月のBIBFRAME 2.0の概要によって、このモデルはWork、InstanceとItemの3つの核心的なエンティティから構成される。

本研究で扱うBIBFRAME Modelを唐詩作品に適用した例を図1に示す。本研究では唐詩作品を創作作品Workとして、その関連関係を対象としてBIBFRAME Modelで構築する。具体的に教科書に掲載された唐詩をInstanceとして使っている。また、Instanceと関連する教科書の情報を付けている。唐詩作品の作者はWorkと関連すると考える。例えば、唐詩作品「送元二使安西」に関する基本的な枠組みを図1に示す。唐詩作品「送元二使安西」は2015年発行された中学校で使用する国語教科書に「元二の安西に使ひするを送る」というタイトルで含まれるという関係で、抽象的な唐詩作品「送元二使安西」のクラスはWorkとして使用し、それと唐詩作品「元二の安西に使ひするを送る」をInstanceとして関連付けている。そのほか、唐詩作品のInstanceはまた掲載される教科書と関連付ける。

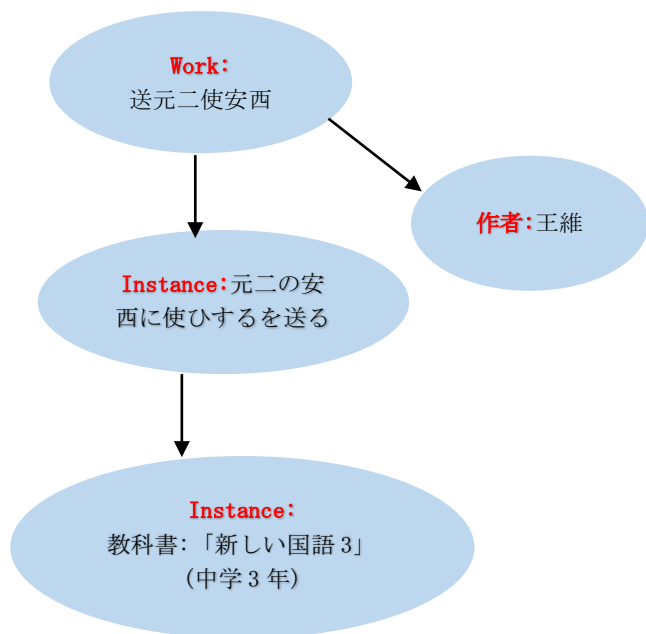


図1:唐詩作品の BIBFRAME Model

### 3 唐詩作品の Linked Open Data 化

#### 3.1 研究範囲の選択

唐詩作品を主な研究リソースとして、応用アプリケーション構築のための研究対象とするため、唐詩作品の調査範囲の確認は重要な問題である。どの唐詩作品を調査対象とするのが問題になる。

唐詩作品の使用状況を確認するために筑波大学附属中央図書館に所蔵されている中学校と高等学校の国語と古典の現行教科書を調べた。教科書の内容は改訂されることがあるため、唐詩作品の利用状況も変化している。そのため、本研究では、平成28年度に使用する教科書[5]を研究範囲とする。

平成28年度日本の中学校と高等学校で使用する教科書における唐詩作品の利用状況を調べた。研究対象について掲載されている唐詩作品を現行教科書に含む数は表1に示す。中学校の部分から見ると、唐詩作品を含む教科書は5冊、唐詩作品は延べ12首、異なり6首が含まれている。高等学校の教科書から調べると、国語の国語総合の教科書23冊、古典A教科書は6冊、古典B教科書は19冊を調査して、唐詩作品を含む数は延べ362首があり、異なり53首がある。それぞれに対応する唐詩作品の作者は25名であった。つまり、平成28年度使用の現行教科書53冊に唐詩作品は延べ374首、異なり59首があり、作者は25名に対応することが分かった。

表1: 現行教科書に掲載されている唐詩作品数

調査対象	唐詩作品		教科書 (冊数)	作者 (異なり数)
	延べ数	異なり数		
中学校	12首	6首	5冊	4名
高等学校	362首	53首	48冊	21名
合計	374首	59首	53冊	25名

### 3.2 唐詩作品の表現の違いについて

唐詩作品の編集は同時代の人々によって集行が行われていたが、後は現在に至るまで改訂が多くの利用者によって積み重ねられている。とりわけ唐詩作品の違いを補うことに見るべきものがある。中国に既存の唐詩作品のみならず日本に伝わる資料及び文献の違いは日本と中国における漢詩の読解方法の文化の違いに起因すると考えられる。

日本の中学校と高等学校が利用する教科書 53 冊に、掲載される唐詩作品を調査したところ、唐詩作品中に表記の異なるものがあることが分かった。

同じ唐詩作品でも、違う漢字を使用するものがある。例えば、「送元二使安西」という唐詩作品を利用する時、中学校の教科書ではそのタイトルは訓読文で「元二の安西に使ひするを送る」を使用したり、「静夜思」は「静夜の思ひ」を用いている。

また、翻訳する時、唐詩作品のタイトルが変わることもある。例えば、「秋夜寄丘二十二員外」という唐詩作品は「秋夜寄丘員外」を採用したり、「山中对酌」は「山中与幽人对酌」を使うこともあると分かった。これらの例を表 2 に示す。

表 2: 唐詩作品の表現の相違点 (抜粋)

問題	唐詩作品の相違点		含む教科書	
	中国語 @zh	日本語 @ja		
書き下し文	静夜思	静夜の思ひ	「新編古典 B」 東京書籍	「中学校国語 3」 学校図書
内容を増補	山中对酌	山中与幽人对酌	「古典 B 漢文編」 数研出版	「新編国語総合」 大修館書店

そして、唐詩作品延べ 374 首をまとめる時、唐詩作品のタイトルと教科書の中に使われているタイトルを区別して扱う。唐詩作品の現行教科書におけるタイトルは日本語で表現されたものとして扱う。唐詩作品異なり 59 首のタ

イトルは一般的に使う唐詩作品のタイトルを使って、中国語の表現だと考える。

### 3.3 唐詩作品の情報の枠組み

2 節に基づいて、唐詩作品を主な研究リソースの対象として、Linked Open Data 化を試みる。具体的に説明すると、唐詩作品を主体として、基本的な唐詩作品のタイトル、詩体、詩体の定義の解説、現行教科書に含まれるページ数、作者の DBpedia への外部リンク、唐詩作品の作者と現行教科書の関連付けを行う。その中、作者の情報は詳細化すると、また作者氏名、性別、字(あざな)、生年月日、没年月日と作者の紹介などがある。これらの情報について、唐詩作品を主語として、唐詩作品は詩体や作者などの情報は重要なリンクとしている。BIBFRAME Model に基づいて、それらのリソースの枠組みを図 2 に示す。

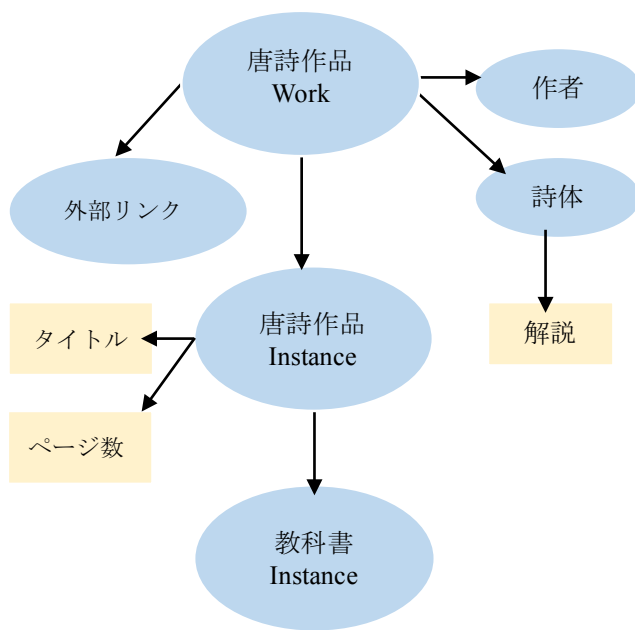


図 2: 唐詩作品情報の枠組み

### 3.4 唐詩作品リソースの URI の設計

本研究では唐詩作品を研究リソースとして、唐詩作品の基本的な情報と現行教科書の関連関係の枠組みを考え、Linked Open Data 化する研究を行う。本研究における Linked Open Data 化とは優先的にウェブ上の URI リソース同士を関連付け、RDF (Resource Description Framework) 枠組みを表現して、



DBpedia への記事<<http://ja.dbpedia.org/resource/作者名>>を用いる。

### 3.5 唐詩作品の語彙の選択

本研究で構築した唐詩作品の Linked Open Data モデルを図 3 に示す。3.4 節の URI の設計に基づいて、情報の検索を向上できるように RDF 枠組みの属性を選択する。それは RDF 枠組みの検索が詳細化できるようにするために付与する。唐詩作品の語彙は LOD ボキャブラリには既存のものを採用し、BIBFRAME ボキャブラリ、Dublin Core、Schema.org、RDF Schema と FOAF ボキャブラリを用いて、唐詩作品の独自ボキャブラリも作成する。

Dublin Core の語彙から説明すると、唐詩作品の詩体などの解説は dc:description で採用する。

RDF Schema のボキャブラリは唐詩作品のタイトルのラベル rdfs:label、外部へリンクする URI として具体的な DBpedia への記事 rdfs:seeAlso を用いる。

Schema.org Vocabulary の利用は各リソースの種別として、唐詩作品のクラス schema:CreativeWork、作者のクラスは schema:Person を用いる。作者の生年月日 schema:birthDate/deathDate を採用、作者の性別 schema:gender とする。唐詩作品を含む教科書のページ数が schema:pageStart を採用する。

唐詩作品 Work と Instance の言語表現は中国語と日本語を区別して、schema:inLanguage の zh と ja で扱う。唐詩作品の詩体などの解説の言語表現は文字列に日本語言語タグ@ja を付与する。

また、唐詩作品の作者 25 名について、字は FOAF Vocabulary から foaf:nick を用いる。唐詩作品の独自ボキャブラリは特に唐詩作品の詩体で使って、tangpoem:style として表現する。

2 節の BIBFRAME 2.0 Vocabulary について、唐詩作品を創作作品として、Work のクラスは bf:Work で、Instance は bf:Instance で表現する。それに基づき、詳しい属性プロパティを与えると、唐詩作品 Work のタイトルは bf:workTitle、Instance のタイトルは bf:instanceTitle で採用する。Work から Instance への関連付けは bf:hasInstance とし

て定義する。また、Instance として唐詩作品と教科書の関係は bf:partOf で扱う。

## 4 結果と考察

唐詩作品の情報の枠組みは3節における日本の中学校と高等学校で学んでいる唐詩作品延べ374首を創作作品のInstanceとして、基本的なモデルを設計し、各関係を有向矢印で結んで、構築されている。その属性プロパティは独自ボキャブラリのほか、Dublin Core、Schema.org、FOAFなどを採用して、BIBFRAME Modelの語彙も扱う。

そして、唐詩作品の関連関係の枠組みは Linked Open Data 化を可能として、BIBFRAME Modelに関わる唐詩作品の仕組みの構築もできると考える。

## 5 おわりに

中国で唐詩作品は古典文化資源の一部とし、中国古典文学についての研究に欠かせない文献として、重要な役割を果たしている。そのため、本研究では唐詩作品を研究対象として、日本の中学校と高等学校で学ぶ唐詩作品と掲載教科書の関連関係の日本語版を Linked Open Data 化する研究を行った。また、BIBFRAME Model による唐詩情報のモデル化も構築できた。ただ、本研究では現在のところ、BIBFRAME 2.0 Model ボキャブラリの利用は限定的なため、それらは今後の課題として検討する。

また、唐詩作品のモデルに基づいて、唐詩作品を増やすこと、多言語化することも今後の課題である。例えば、中国の全唐詩庫[6]にあるような多くの唐詩作品を関連付けて、Linked Open Data 化できるようにして、それらに基づいて、検索や閲覧ができるアプリケーションを構築したい。

## 謝辞

本研究の一部はJSPS科研費16H02913の助成によるものである。

## 参考文献

[1] トム・ヒース, クリスチャン・バイツァー著. 武田英明, 大向一輝, 加藤文彦, 嘉村哲郎, 亀田堯宙, 小出誠二, 深見嘉明, 松村冬子,

南佳孝訳, Linked Data: Web をグローバルなデータ空間にする仕組み, 近代科学社, 2013, 139p. ISBN 978-4-7649-0427-9.

[2] “横浜の芸術文化情報のオープンデータ”, <http://yan.yafjp.org/lod>, (accessed 2016-08-20).

[3] 上海図書館. “上海デジタル公開サイト”. <http://wr2016.library.sh.cn>, (accessed 2016-08-20).

[4] “Overview of the BIBFRAME 2.0 Model”. <http://www.loc.gov/bibframe/docs/bibframe2-model.html>, 2016-04-21, (accessed 2016-08-20).

[5] “教科書目録 (平成 27 年 4 月)”, [http://www.mext.go.jp/a\\_menu/shotou/kyoukasho/mokuroku/27/1357046.html](http://www.mext.go.jp/a_menu/shotou/kyoukasho/mokuroku/27/1357046.html), (accessed 2016-08-20).

[6] 鄭州大学管理中心. 全唐詩庫. (中国語), <http://www16.zzu.edu.cn/qts/>, (accessed 2016-08-20).