唐詩作品の本文フルテキストに対するTEI マークアップ手法の提案

doi: 10.2964/jsik_2018_017

情報知識学会第 26 回年次大会

唐詩作品の本文フルテキストに対する TEI マークアップ手法の提案

A Proposal of TEI Markup for the Content of Tang Poems

叢 艷[1*], 高久 雅生[2]
Yan CONG[1*], Masao TAKAKU[2]

1 筑波大学大学院 図書館情報メディア研究科
Graduate School of Library Information and Media Studies. University of Tsukuba
〒305-8550 つくば市春日1-2
E-mail: cong.y@slis.tsukuba.ac.jp
2 筑波大学 図書館情報メディア系
Faculty of Library, Information and Media Science. University of Tsukuba
〒305-8550 つくば市春日1-2
E-mail: masao@slis.tsukuba.ac.jp
*連絡先著者 Corresponding Author

本研究では，平成 28 年度使用の中学校と高等学校の現行教科書に含まれる唐詩作品を研究対象とし，現行教科書の原版面を尊重したデジタル化として，唐詩作品の本文フルテキストの TEI マークアップ手法を提案する．唐詩作品の訓読文における返り点や送り仮名などの訓点情報やルビを表現する方法を検討し，現行教科書の原版面を尊重した上で，テキスト化することを目指す．

In this paper, we focus a TEI markup for the content of Tang poems. Based on textbooks of Japanese high schools and junior high schools in 2016, we represent fulltext of Tang poems along with the kunten information on punctuated text (kundokubun;訓読文). As a result, we show the results of markuped text which are made as the same as the original fulltext in textbooks.

キーワード: 唐詩, LOD, TEI マークアップ，教科書，本文フルテキスト
Tang Poem, Linked Open Data, TEI markup, Textbook, Fulltext Content

1 はじめに

近年，ウェブ上で膨大なデータの中から関連情報を正確に抽出し，リンクで直接繋がるLinked Open Data (LOD)という技術に注目が集まる．筆者らは文化資源に注目し，中学校と高等学校の国語と古典の現行教科書の教育学習ニーズに向けて，学校の生徒が唐詩作品をより簡単に理解できることを目指す．そのため，現行教科書に含まれる唐詩作品および作者の情報，詩体などの有益な情報を関連づけて，Linked Open Data化することを試してきた[1][2][3]．本論文では，それらの研究成果を拡張して唐詩作品の本文フルテキストのデジタル化を試みる．

唐詩作品の本文フルテキストのデジタル化は重要な役割がある．まず，唐詩作品の本文フルテキストの情報を公開・利用できれば，対象とする唐詩作品の情報テキストを検索できて，学習者が漢文や唐詩作品などの情報を理解しやすくなると考える．また，教科書における唐詩作品のデジタル化に当たっては，学習に必要な教科書の原版面を尊重し，訓読文と書き下し文に基づくデジタル化する．訓読文の特有の訓点情報を機械可読のテキストにできれば，有益だと考える．従って，テキスト検索や訓点情報の表現ができれば，唐詩作品のテキストとその関連情報は，多くの人が共有でき，簡単に使える利便性がある．

古典籍を機械可読化して共有するTEIマークアップ[4]が利用されている．TEIマークアップは汎用的なXMLと組み合わせて，記述できる．その特徴としては，TEI:P5 ガイドライン[5]にタイトルや本文フルテキストなどの特定要素がある．そのため，本研究では，唐詩作品を研究リソースとして，TEIマークアップを用いる．

現在，国語や古典の教育学習ニーズに応じて，生徒がより簡便に古典籍を読めるよう，現行教科書に含まれる漢詩などの本文フルテキストは，書き下し文にルビで漢字の解釈などを付与され，訓読文に関わる訓点情報も注釈として付けられる．本文フルテキストの内容に対する言葉の解説などの注釈も，番号や小さな符号などで，その語の周りに付している．これらの訓点資料や，ルビ情報，注釈などの要素をテキスト化，ウェブサイト作成などをするとき，基本的な作成基準が存在しない[6]．すなわち，古典籍や現行教科書などの資料の原版面を尊重してデジタル化する上で，訓点情報およびルビ情報や注釈などのテキスト化は難しい研究課題である．

そのため，本研究では，平成28年度使用の中学校と高等学校の現行教科書に含まれる唐詩作品を研究対象とし，唐詩作品の本文フルテキストのTEI マークアップ手法を提案する．唐詩作品の訓読文における送り仮名や返り点などの訓点情報や，書き下し文におけるルビ情報の表現する方法を検討し，現行教科書の原版面を尊重した上で，テキスト化することを目指す．

2 研究対象

本研究では，唐詩作品の本文フルテキストを研究リソースとして，TEI マークアップを試みる．

まず，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つつ，唐詩作品の利用状況も変化するため，本研究では，平成 28 年度に使用する現行教科書を研究対象とする．

平成 28 年度の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て，掲載される唐詩作品の数は表 1 に示す．中学校の現行教科書で掲載されたものは 5 冊，唐詩作品は延べ 12 首，異なり 6 首が含まれる．高等学校の現行教科書は，国語の国語総合の現行教科書 23 冊，古典 A 教科書 6 冊，古典 B 教科書 19 冊を調査して，唐詩作品を含む数は延べ 362 首，異なり 53 首があり，作者は異なり 25 名がいた[2]．

表 1 現行教科書に掲載される唐詩作品の数

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 唐詩作品の本文フルテキスト

唐詩作品の本文フルテキストをデジタル化する際の観点として，(1)テキスト表現の文体，(2)訓点情報やルビ情報の表現をテキスト内の要素点の2点を挙げる．以下でそれぞれについて述べる．

本研究では，平成28年度の中学校と高等学校に利用する現行教科書に含まれる唐詩作品の本文フルテキストを研究対象として，現行教科書に含まれる唐詩作品の利用状況の調査を行った．調査したところ，唐詩作品の本文フルテキストは4つの文体に分類できる．これら4つの文体は「白文」，「訓読文」，「書き下し文」，「翻訳文」と呼ぶ[3]．この分類は唐詩作品の本文フルテキストにおける表現の違いに起因するものである．

中学校の教科書では，生徒の学習をより容易にするため，多数の作品が書き下し文も訓読文も両方が同時に掲載される．一方，高等学校の現行教科書に含まれる唐詩作品はすべて訓読文のみで掲載されている．翻訳文は唐詩作品の本文フルテキストの全文の意味の解釈として，中学校の現行教科書に掲載される．

唐詩作品などの漢文における作品の読み順は，常に縦書きで教材の上から下まで1行，右から左の順番で読んでいる．詳しくは以下の節に紹介する．

3.1 白文の概要

白文は唐詩作品の本文フルテキストのみ，返り点，送り仮名などの訓点情報が付いていないそのままの原文資料である．この文体は主に中国の古典籍および漢文のコンテンツに利用するが，そのまま読んでも日本語の文法と合わずに読解が難しくなるため，日本の中学校と高等学校の現行教科書には白文が含まれない．

ただし，訓読文や書き下し文などの文体は白文のような原文資料に基づいて，訓点情報や，音読などのルビ情報，注釈要素を提示して，日本語の語順に合わせて利用しているため，白文は古典籍および漢文の原文資料として，重要な役割を果たす．

3.2 訓読文の概要

訓読文は漢文を日本語の語順で読めるようにしたものである．訓読文は白文の原文資料に基づき，古典籍および漢文などの白文を日本語の文法に従って，返り点や送り仮名などの訓点情報を白文内の各文字の四隅に書き入れたものである．したがって，日本語話者の生徒や研究者などの利用者がその白文の意味を理解しやすくなる．

訓点情報は広い意味で言えば，返り点，送り仮名と句読点である．平成28年度の教科書における唐詩作品の本文フルテキストはほぼ送り仮名と返り点だけを含んだものだが，高校教科書のタイトルは「八月十五日夜，禁中独直，対月憶元九」の唐詩作品1首のみはタイトルに句読点がある．送り仮名は文章の単語を解説する場合，その漢字や単語の読み方を表すために，その四周に加える小さな文字である．返り点は縦書きの場合，漢字や単語の左下のところに付けて，下から上へ戻って読むことを示す記号である．中学校と高等学校の教科書に扱う返り点の記号種類は主にレ点，一・二・三点である．

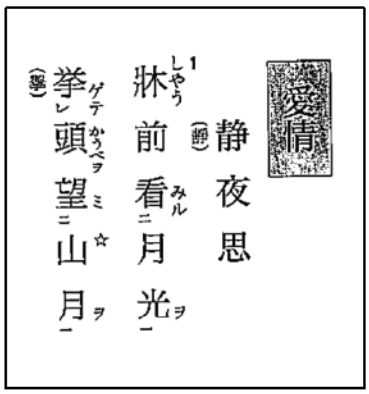


図1 訓読文の文体の事例[7]

図1の訓読文の例は第一学習社から出版された国語の現行教科書「高等学校国語総合」に掲載された李白の唐詩作品「静夜思」である[7]．このような訓読文を読むとき，返り点や送り仮名などの訓点情報の読み方に従って，助詞，助動詞などを補い，日本語の文法に合わせることができる．

そのため，訓点情報の表記は古典籍および漢文における生徒の読解力を上げる利益をもたらす．

3.3 書き下し文の概要

書き下し文は訓読文に関わる返り点や送り仮名などの訓点情報の要素に基づき，漢文を日本語の語順に合わせて，書き直した文章である．図2は中学校の現行教科書に掲載の「静夜の思ひ」[8]という唐詩作品の訓読文(上)と書き下し文(下)を一緒に掲載した本文フルテキストである．

書き下し文は訓読文に表記する訓点のルールに基づいて，日本語の語順で読みやすい文で読解し，訓読文を日本語として読み取りやすい利益がある．

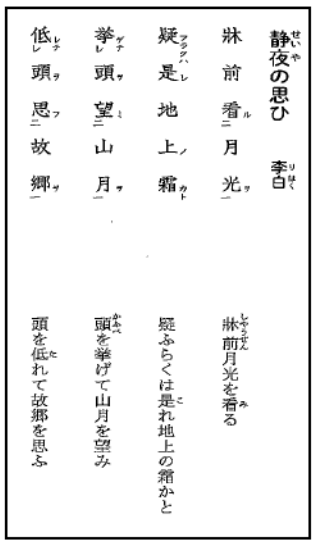


図2 訓読文(上)と書き下し文(下)を一緒に掲載された事例[8]

3.4 ルビの概要および利用状況

ルビ[9][10][11]は現代文や古典籍の文章のある文字に対し，振り仮名や文字の説明，音読，異なる読み方などを原文資料の親文字より小さな文字で付与されるものである．それは日本語の文章では，漢字の発音や，漢字の説明などに役に立つ．ルビはモノルビ，グループルビと熟語ルビの3つ[12]の種類がある．

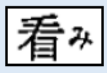
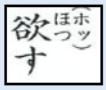
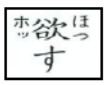
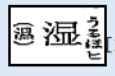
現代文や古典籍に文章は縦書きする場合，音読や注釈などのルビの位置は基本的に原文資料の任意の漢字の右側に仮名で付与する．教科書におけるルビ情報では，通常のルビに加えて，両側ルビを配置する場合，親文字の左側にカタカナでこの親文字の振り仮名や，文字の説明，音読，異なる読み方などの情報を小さな文字で表記する．

中学校の教科書には，書き下し文も訓読文も掲載されているため，ルビ情報は，書き下し文だけで付けられている．高等学校の教科書に全ての唐詩作品は訓読文だけで掲載されるため，ルビ情報は，振り仮名や，文字の説明，音読，異なる読み方などの注釈情報をそれぞれの語の周りに対応して利用する．なお，縦書きの場合，ルビ文字列の位置に基づいて，ルビの処理は片側にのみ配置することと両側に配置する場合の処理として議論する．

平成28年度の現行教科書における唐詩作品の書き下し文および訓読文におけるルビの利用状況を表2でまとめた．教科書に含むルビの利用状況は，ルビの位置によって，書き下し文と訓読文に関わる教科書に含むルビ付けの事例である．

表2において，第1番から5番までが片側ルビである．第1番と2番は，ルビの位置における基本的な片側ルビとして，親文字の右側に平仮名や左側にカタカナで表記する．第3番と4番は親文字の片側に括弧で音読や，漢字の解説を示す事例である．第5番は基本的な片側ルビと違って，親文字の片側に，2列並べて置いている事例である．また，5番目の親文字から2列目のルビは，括弧で囲んでいる．

表2 現行教科書におけるルビの利用状況

NO.	ルビの位置	教科書に含む事例	文体
1	片側ルビ	 [8]	書き下し文
2		 [13]	訓読文
3		 [14]	書き下し文
4		 [7]	訓読文
5		 [14]	書き下し文
6	両側ルビ	 [13]	書き下し文
7		 [15]	訓読文

次に，6番目と7番目は両側に配するルビの事例である．6番目の事例は基本的な両側ルビとして配置するが，5番目事例の唐詩作品と同一であり，教科書の違いによって，表現も違うことが分かる．7番目の事例は両側ルビに基づき，左側のルビが括弧で囲む場合もある．

3.5 注釈情報

注釈とは既存の文章における任意の言葉の解説である．それは読者が文章を理解できるように，文書の出典や，言葉の意味などを示す．それは，日本および中国の書物でも付与する．唐詩作品の注釈は本文内容の言葉や，作者の紹介，本文フルテキストに関連する背景などの内容を補足する．中学校の教科書では，注釈番号は本文に付けられず，まとめて同じページの下部に置いている．高等学校の教科書に唐詩作品は全て訓読文で掲載され，番号や，符号が本文に付けられ，対応する説明は同じページの下部に置いている．例えば，表3の例として注釈情報の事例を説明する．

教科書に掲載される唐詩作品の注釈は番号や符号で付ける．番号を付ける場合は，この言葉の解釈などを示す．例えば，表3の1番目の注釈事例は，番号で表記して，唐詩作品の本文フルテキストに関わる地名情報を解説する．また，符号を付する場合は表3の2番目の事例であり，教育学習ニーズに応じて，学習者により唐詩作品の本文フルテキストを理解しやすくなるために，読解などの問題も示す．

表3 注釈情報の事例

NO.	注釈の事例	対応する注釈内容
1	今夜鄜州ノ月[1ふ] [15]	1鄜州　今の陝西省富県。当時、杜甫の妻子は、安史の乱(七五五―五六三)を逃れて、そこに避難していた(→三三九ページ注3)。 [15]
2	低レ頭思フ☆故郷ヲ[たレテヲ] [7]	問「山月」と「故郷」とは、どのように関係しているか。 [7]

4 唐詩作品のマークアップ手法

本研究では，平成28年使用の中学校と高等学校の現行教科書に含まれる唐詩作品の本文フルテキストを研究対象として，TEI マークアップ手法を用い，唐詩作品のデジタル化を試みる．訓読文における唐詩作品の特有の要素として，送り仮名や返り点などの情報を表現する方法を検討する．

TEI (The Text Encoding Initiative)[4]は人文学系の文章を対象とするデジタル化を促進するために基本方針を定める組織である．TEI ガイドラインは1994年から図書館や博物館など，または個人的な利用者が研究，教育，保存を進むために，幅広く利用されてきた．

本研究の研究手法では，TEI マークアップの標準的なTEI:P5 ガイドライン[5]の基準を利用し，平成28年度利用する中学校と高等学校の現行教科書に掲載される唐詩作品の本文フルテキストを研究データとして，唐詩作品の本文フルテキストのTEI マークアップを試みる．

4.1 マークアップ対象

3章の記述する通りに，唐詩作品の本文フルテキストのデジタル化は唐詩作品に関わるタイトル，作者や本文フルテキスト

を一括でTEI マークアップする．TEI マークアップにおいて，詳細に扱う唐詩情報は以下の通りになる．

1) 唐詩作品のタイトルおよび作者名．書き下し文と訓読文の両方の文体を含む場合はそれら両方に対応するメタデータになる．
2) 本文フルテキストに関するコンテンツの内容．主に訓読文と書き下し文を中心とする．訓読文は，親文字を中心として，四周に付ける返り点や送り仮名などの訓点情報を付く．書き下し文は唐詩作品の内容をそのままマークアップし，ルビなどの情報も付くことがある．
3) 本文フルテキストに関わる文体．
4) 唐詩作品の本文フルテキストに関わる行番号．

4.2 TEI マークアップ手法

唐詩作品の本文フルテキストのXMLに基づくTEI マークアップを行う．

まず，唐詩作品のタイトルは書き下し文の形のタイトルと訓読文の形のタイトルの2種類がある．唐詩作品のタイトルの表現が要素<title>を用いる．

次に，唐詩作品に関わる作者名に扱うタグは要素<author>を用いる．この要素<author>は書誌情報における著作者(個人・団体)の名前を示す要素である．

唐詩作品の本文フルテキストの全文は要素<lg>を用い，要素<lg>はline groupの省略であり，全文の詩節などのまとまりを示す要素である．唐詩作品の全文の特徴として，属性値はverse(韻文)としてそれぞれの唐詩作品<lg>要素の属性に対応する．

唐詩作品の本文フルテキストにおける4つの文体は「白文」，「訓読文」，「書き下し文」，「翻訳文」と呼ぶ．白文はunpunctuated text，訓読文はpunctuated text，書き下し文は reading text，翻訳文はtranslation textで文体を区別するための名前として利用する[3]．唐詩作品の本文フルテキストにおける文体の種類は独自の属性を用いて表現する．この属性はprefix:tpとして，名前空間URIはhttps://w3id.org/tangpoemを用いる[1][2]．その属性名はtp:fulltextTypeを用いる．属性値は，4つの文体の英文名を利用して，全唐詩作品の本文フルテキストに関わる文体の種類を表現できる．

また，本文フルテキストの詩節の各行は要素<l>を用い，行番号はその属性n=1,2,…8のようにする．<l>要素はverse lineであり，詩節全体の<lg>要素に含まれる．つまり，4句の唐詩作品は4つの要素<l>で，8句の唐詩作品8つの要素<l>を組み合わせて，1つの要素<lg>に含まれる．

```
<title>静夜の思ひ</title>
<author>李白</author>
<lg type="verse"
tp:fulltextType="reading">
   <l n="1">牀前月光を看る</l>
   <l n="2">疑ふらくは是れ地上の霜かと
     </l>
   <l n="3">頭を挙げて山月を望み</l>
   <l n="4">頭を低れて故郷を思ふ</l>
</lg>
```

図3 「静夜の思ひ」[8]に関わる書き方

これらに基づいて，図3に「静夜の思ひ」[8]という唐詩作品の基本的なマークアップ例を示す．これは作者が李白の「静夜の思ひ」の唐詩作品が任意のverse(韻文)とし，

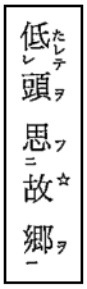
書き下し文(reading)の文体を用い，本文フルテキストの第1行目の内容は「牀前月光を看る」という意味である．

訓読文と書き下し文における本文フルテキストのTEIマークアップは全体とする唐詩作品の詩の全体要素<lg>や，各行の要素<l>を利用する．訓読文にある任意の漢字は訓点を付ける場合，送り仮名や返り点などの訓点情報は本文の注釈として見なし，要素を用いる．要素はテキスト部分に解釈的注釈を関連づけるという意味を持つTEI要素である．

TEI ガイドラインでは，訓読文の四隅に付与される訓点を説明できる要素は含まれないので，ここでは独自の属性値として，訓点の種類を日本語の読み方のローマ字で命名することとした．つまり，送り仮名の要素の属性値は送り仮名の読み方okurigana，返り点の要素は返り点の読み方kaeritenとして使い，訓読文用のTEIマークアップを行った．表4は高等学校に含む訓読文「静夜思」[7]を事例として，要素の使い方である．

表4では，唐詩作品「静夜思」[7]の第4行目の詩句を例として，訓読文のTEIマークアップを行った．詩句の冒頭として，要素<l>を入れる．第4行目の説明とする属性値を要素<l>に書き入れ，親文字「低」の右下の送り仮名「レテ」と左下の返り点「レ点」を表記すると，前者がレテとマークアップして，後者はレとマークアップする．送り仮名と返り点の表記順番は先に送り仮名，次は返り点という順を用いる．この書き方に基づき，訓読文にそれぞれの親文字および訓点情報を対応づける．

表4 送り仮名と返り点の事例の書き方

原文資料	 [7]
使い方	<l n="4"> 低レテレ 頭ヲ 思フ二 故郷ヲ一 </l>

4.3 ルビ情報のマークアップ

3.4節の通り，本研究では唐詩作品の本文フルテキストは現行教科書の原版面を尊重するため，教科書に掲載された唐詩作品と同様の縦書きの形でルビ情報を表現する必要がある．

まず，ルビの要素は次のHTMLの基本的な要素<ruby>，<rb>，<rt>，<rtc>と<rp>の5つの要素[9][10][11]があり，これを採用する．ルビの種類はモノルビ，グループルビと熟語の3つ[12]がある．要素<ruby>はルビと親文字のまとまりを表す要素であり，それは全体の親文字と小文字を表す要素である．<rb>は親文字だけを表し，<rt>はルビを表すタグである．そのほか，<rtc>は1つの親文字に対して複数のルビ情報を付ける場合に，ルビを<rtc>要素に入れて使えることができる．要素<rp>はルビをサポートし

ないブラウザ向けのフォールバックを提供できる要素である．筆者らは，ウェブ上におけるルビのマークアップの実装[11][16]を参考にして，ルビ要素タグを組み合わせて，書き下し文におけるルビのマークアップを行う．ルビ情報のマークアップは基本的に，文章に文字や言葉などを単位として，マークアップすると考える．

表5 ルビのマークアップの書き方

NO.	原文資料	種類	書き方
1	 [6]	片側ルビ	<ruby>看 <rt>み </rt>る </ruby>
2	 [6]	片側ルビ	<ruby>牀前 <rt>しゃうぜん</rt> </ruby>
3	 [10]	両側ルビ	<ruby>烽 <rt>ほう </rt> </ruby> <ruby>火 <rt>くわ <rtc>カ </rtc> </rt></ruby> <ruby>三月 <rt>さんげつ </rt></ruby> に連なり

書き下し文「静夜の思ひ」[8]という唐詩作品をシンプルな事例として，基本的なルビのマークアップの方法は表5に示す．表5に示す1番目の「看る」のような単一の文字を片側ルビで表記する場合に，モノルビとしてマークアップする．表記する際に，ルビ情報を付ける親文字の冒頭に要素<ruby>を書き入れ，ルビは要素<rt>内に表記し，終了タグ</rt>と</ruby>で終わる．

親文字の数が複数の場合はグループルビとし，全体の読みを一緒にルビとして表記する．ルビ情報が片側に配置する2番目の場合，表記方法が1番目の事例の書き方と同じく，冒頭に要素<ruby>を入れ，片側ルビを表示する要素<rt>要素を書き込んで，終了タグで囲んで書く．

グループルビを両側ルビに配置する場合は，一般論としてはグループルビとして，要素<rt>と<rtc>を用いて，単語として表記するが，本研究では，簡便にマークアップするために，単一の文字ごとに表記してマークアップすると考える．その書き方は表5の3番目の事例である．

5 研究結果と考察

TEI:P5 ガイドライン[5]を研究基準とし，主に平成28年の中学校と高等学校に利用する唐詩作品の本文フルテキストを着目して，TEI マークアップを行った．

本研究では，主に訓読文と書き下し文における訓点情報やルビ情報をマークアップする方法を中心として実装した．実装したデータとしては平成28年度用の中学校の教科書に唐詩作品は異なり6首があることに基づき，任意の高等学校の教科書に含む同一の唐詩作品ごとに各1首を選択した．表6に示すリストは唐詩作品におけるTEIマークアップを行ったリストである．唐詩作品の欄に，唐詩作品のタイトルは書き下し文および訓読の両方を一緒に表記した．掲載された本文フルテキストは中学校と高等学校用の教科書に基づいて，選択した．このリストの情報に基づいて，TEIマークアップを行った．

表6 TEI マークアップを行った唐詩作品リスト

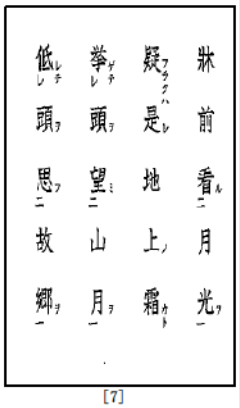
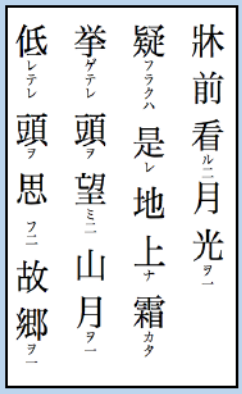
NO	唐詩作品	掲載教科書	
		中学校	高等学校
1	黄鶴楼送孟浩然之広陵（黄鶴楼にて孟浩然の広陵に之くを送る）	現代の国語2[17]	精選国語総合[18]
2	春望	現代の国語2[13]	精選国語総合[19]
3	春暁	現代の国語2[20]	精選国語総合[21]
4	絶句	国語2[22]	精選国語総合[23]
5	送元二使安西（元二の安西に使ひするを送る）	中学校国語3[24]	精選国語総合[25]
6	静夜思（静夜の思ひ）	中学校国語3[8]	高等学校国語総合[7]

唐詩作品の本文フルテキストの表現に，訓読文，書き下し文などの4つの文体がある．漢文および唐詩作品に関わる本文フルテキストは現行教科書に掲載されるとき，文体によってテキストの表現が異なる．これらの文体の役割は利用者が原文資料を理解しやすくなるために，日本語の文法に合わせて，それぞれの訓点情報や注釈などの要素が付与された．そのため，各文体は要素を付与される状況の違いに従って，教科書の原版面を尊重し，中学校と高等学校の現行教科書における唐詩作品の本文フルテキストの訓点情報などを確認した上で，デジタル化した．

実装環境は，MacBook Proで主にSafariブラウザ バージョン11.1を使い，マークアップの内容も試みた．また，訓点情報およびルビはXHTMLとCSS[26]を組み合わせて利用した．

訓読文における「静夜思」[7]のマークアップの表示は表7に示す．原文ページの欄に原文資料を参照して，右側のデジタル化の表示欄にブラウザ表示のスクリーンショットがある．スクリーンショットを示すように大きな文字は親文字であり，送り仮名とレ点，一・二点などの訓点情報も全てを書き入れた．ただし，原文資料と比べると，現時点で返り点や送り仮名の位置は行中央に小さな文字で表示され，全て親文字の下に並べた状態としている．

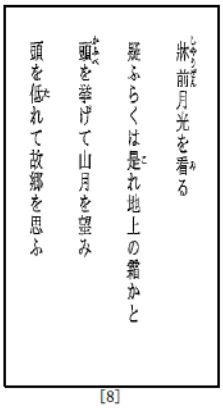
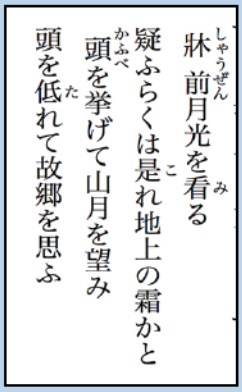
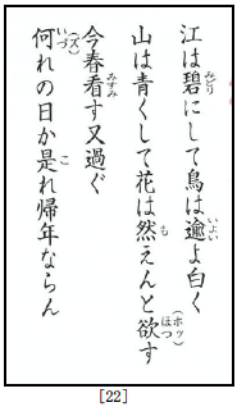
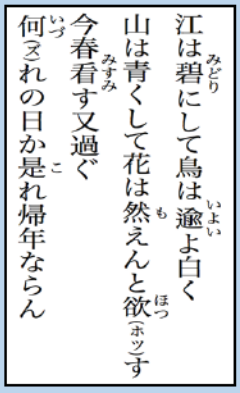
表7 訓読文におけるデジタル化の表示

NO	原文ページ	デジタル化の表示
1	 [7]	

書き下し文におけるルビのマークアップは表8に示す．表8は書き下し文の2つの事例として，白色の背景で教科書に掲載された原文ページであり，青色の部分はマークアップの結果である．表8の1番目は原文資料の書き下し文のルビの表現は基本的な片側で付すため，デジタル化は教科書の原文ページに従って同様にした．2番目は

書き下し文における唐詩作品「絶句」[22]の本文フルテキストであり，原文資料の書き下し文のルビの表現によって，親文字に基づいて，片側ルビは2列並べる表現もあり，括弧の利用も付けている．現時点で基本的な片側ルビが表示できるが，両列に並べるなどの表示はできていない．

表8 書き下し文におけるデジタル化の表示

NO	原文ページ	デジタル化の表示
1		
2		

6 おわりに

本研究では，平成28年度使用の中学校と高等学校の現行教科書に掲載された唐詩作品を研究リソースとして，唐詩作品の本文フルテキストのTEI マークアップ手法を提案した．そのうち，現行教科書に含まれる唐詩作品の原版面を尊重した上で，唐詩作品の本文フルテキストに基づくマークアップを行い，ブラウザの表示を試みた．

今後の課題としては，(1) TEIマークアップを行った時，返り点や送り仮名などの訓点情報を共有するためには標準化を検討する必要がある．(2) ルビの要素は基本的な要素を用いたが，デジタル化の表示にはブラウザ対応の改善やCSSの調整が必要である．(3)本文フルテキストに含まれる注釈情報もマークアップし，データの一部としたい．

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] 叢艶；高久雅生：「唐詩情報のLinked Data化の試み」．第15回情報メディア学会研究大会，2016，pp. 17-20.

[2] 叢艶；江草由佳；高久雅生：「唐詩情報Linked Open Data化とその利活用の試み」．人工知能学会セマンティックウェブとオントロジー(SWO)．第39回研究会，2016. SIG-SWO-039-07.

[3] Yan CONG; Masao TAKAKU:“Prototype of Linked Open Data Model for Tang Poems”. Japanese Association for Digital Humanities Conference 2017 (JADH2017), Kyoto, Japan, pp. 50-52 (2017-09).

[4] Text Encoding Initiative: “The Text Encoding Initiative”,

http://www.tei-c.org/index.xml, (accessed 2018-04-13).
[5] The Text Encoding initiative: “TEI: P5 Guidelines”. http://www.teic.org/Guidelines/P5/, (accessed 2018-04-13).
[6] 高橋晃一:「論理構造と物理構造が混在するテキストのXMLによるマークアップに関する考察」, 研究報告人文科学とコンピュータ(CH), 情報処理学会, Vol. 98, No. 6, pp. 1-5, 2013.
[7] 「静夜思」, 高等学校 国語総合, 東郷克美;伊井春樹 ほか28名編, 第一学習社, 2012, p. 336.
[8] 「静夜の思ひ」, 中学校国語3, 野地潤家;新井満 ほか28名編, 学校図書, 2015, p. 176.
[9] 日本規格協会:「日本語文書の組版方法 JIS X 4051」. 日本規格協会, 2004, 206p.
[10] 阿南康宏 ほか編: “日本語組版処理の要件(日本語版)”, 2012, https://www.w3.org/TR/jlreq/ja/, (accessed 2018-04-13).
[11] Marcin SAWICKI; Michel SUIGNARD; Masayasu ISHIKAWA; Martin DÜRST: Ruby Annotation, https://www.w3.org/TR/ruby/, (accessed 2018-04-05).
[12] Robin Berjon; “ W3C HTML Ruby Markup Extensions”. https://www.w3.org/TR/html-ruby-extensions/, (accessed 2018-04-13).
[13] 「春望」, 現代の国語2, 中洌正堯 ほか39名編, 三省堂, 2015, p. 142.
[14] 「春望」, 国語2, 甲斐睦朗 ほか27名編, 光村, 2015, p. 152.
[15] 「月夜」, 高等学校 国語総合, 東郷克美;伊井春樹 ほか28名編, 第一学習社, 2012, p. 336.
[16] 縦書きweb普及委員会: “ルビの解説とマークアップ方法”. https://tategaki.github.io, (accessed 2018-04-13).
[17] 「黄鶴楼にて孟浩然の広陵に之くを送る」, 現代の国語2, 中洌正堯 ほか39名編, 三省堂, 2015, p. 123.
[18] 「黄鶴楼送孟浩然之広陵」, 精選国語総合, 三角洋一;池内輝雄;小町谷照彦 ほか27名編. 東京書籍, 2012, p. 345.
[19] 「春望」, 精選国語総合, 三角洋一;池内輝雄;小町谷照彦 ほか27名編, 東京書籍, 2012, p. 348.
[20] 「春暁」, 現代の国語2, 中洌正堯 ほか39名編, 三省堂. 2015, p. 122.
[21] 「春暁」, 精選国語総合, 三角洋一;池内輝雄;小町谷照彦 ほか27名編, 東京書籍, 2012, p. 343.
[22] 「絶句」, 国語2, 甲斐睦朗 ほか27名編, 光村, 2015, p. 148.
[23] 「絶句」, 精選国語総合, 北原保雄 ほか21名編, 2012, p. 304.
[24] 「元二の安西に使ひするを送る」, 中学校国語3, 野地潤家;新井満 ほか28名編, 学校図書, 2015, p. 174.
[25] 「送元二使安西」, 精選国語総合, 三角洋一;池内輝雄;小町谷照彦 ほか27名編, 2012, p. 306.
[26] Elika J. Etemad; Koji Ishii: “CSS Ruby Layout Module Level ”, https://www.w3.org/TR/css-ruby-1/, (accessed 2018-04-13).