唐詩の構造化に関する研究：テキストの差異提示機能の検討

叢 艷
筑波大学大学院　図書館情報メディア研究科

高久 雅生
筑波大学 図書館情報メディア系

筆者らは唐詩作品のLOD化および本文フルテキストのTEIマークアップに関わる研究を行ってきた. 本研究では, それらの研究成果を踏まえて, 同一の唐詩作品Workにおける唐詩作品のインスタンス同士の異なる表現を比べ合って, 同じ唐詩作品Workに属するInstanceの間の差異を提示する機能を検討する. 唐詩作品の校合対象については唐詩作品の(1)タイトル, (2)訓点情報, (3)漢字, (4)漢字の順序がある. 本研究の比較方法は, LCS (Longest Common Subsequence)アルゴリズムによる差分検出であり, 差分の結果の表現は, HTMLおよびCSSと組み合わせて, 利用者が直接見られるようにする. 結果としては, 唐詩作品のインスタンス同士に存在する差異をそれぞれ表し, 校合テキストを結果として, 提示する.

A Prototype of Collation for Tang Poems

Yan CONG
Graduate School of Library, Information and Media Studies, University of Tsukuba

Masao TAKAKU
Faculty of Library, Information and Media Science, University of Tsukuba

We did the researches on the Linked Open Data of Tang Poems and the TEI Markup for the Content of Tang Poems in the past. Based on the previous research outcomes, we aim at developing a prototype of collation for Tang Poems to let people find the difference between the tang poem's instances intuitively. At first, we compare the differences of Tang Poem's Instances which are belong the same Tang Poem's Work. Then we show the results of differences. The targets of collation for Tang poems are as follows, (1) Title, (2) *Kunten* information, (3) *Kanji* Characters, (4) the order of the characters in a sentence. We use the method of longest common subsequence algorithms to collate the poems, and we also show the results of differences with HTML and CSS.

１．はじめに

日本における歴代法案の校訂や档案史料などの古典籍を校合する研究[1]がある. それらは現在まで古典籍を発行された歩みを振り返る記事であり, 現在における貴重な古典の学習資料として, 多くの編集者を利用し, 現在まで各種の校訂本は数多く発行された. これらの校訂本は元々の底本と比べて, 細部にわたって相違点が存在する可能性が高い. したがって, 詳細な調査, 考察や残される底本との関連性を明らかにする研究によって, 底本との関係調査を繰り返す作業などの必要があるため, これらの校訂本と底本の間の相違点が記録されれば, 便利になる. また, このような解明のための研究にはかなり労働力がかかると同時に, 紙媒体の校訂本の間の相違点の状況を精査する時間もかかる問題がある.

近年, 出版物のデジタル化が盛んとなり, 古典作品のデジタル化資料も積極的に活かされている.

このような古典作品の一部に唐詩がある. 唐詩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を持つ. 唐詩は中国の貴重な古

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．

唐詩作品をデジタル化したデータがあれば，機械可読処理が可能となり，さらに，標準化されたデジタル化手法により，様々な応用のための相互運用性を高められ，利用も便利になる．同時に，唐詩作品の構造化における活用の機能を提供できれば，唐詩作品の学習ももっと簡単にできると考える．

ただし，唐詩は古文の校訂により伝わってきて，出典由来の違いなどの問題により，同じ唐詩作品であっても差異が存在することがある．これらの相違点をいつでも確認しやすい形式で記録されてないため，生徒や研究者が利用する際に，差異点によって唐詩のコンテンツの誤解や遺漏などの問題も起きると考える．そのため，唐詩作品における差異も表記できれば，調査などに対する研究を進む時間や労力を節約できるだろうと考える．

そのため，本研究では文化資源に注目し，唐詩作品の構造化されたデジタルテキストを利用し，テキストの差異の校合を支援する．唐詩作品の LOD 化[2][3][4]，TEI マークアップ手法での標準化された方法[5][6]の研究成果を踏まえて，唐詩作品の異文に関わる差異を提示し，その差異を校合する手法を提案する．つまり，同一の唐詩作品 Instance 同士に，それぞれの異文の差異を比べ合わせて，その差異を出力し，差異提示機能を実現する．

このような差異を照合することを校合(きょうごう；collation；collating)と呼ぶ．図書館情報学用語辞典[7]によれば，校合は，写本の本文の異同，既出の刊本の本文の異同，そして特に初期刊本の同一版内での本文の異同を比較，記述する作業である．

唐詩作品の構造化として，(1)唐詩の LOD 化は同じの唐詩作品を，唐詩作品 Work と唐詩作品 Instance に分けた時，同じ唐詩作品 Work に属する Instance に差異もあることを把握する．(2)唐詩作品の本文フルテキストの TEI マークアップによって，本文フルテキストに付随される訓点，ルビ情報などを符号化する．これらの構造化の研究成果を踏まえて，同一の唐詩作品に対する複数の異文がある場合，テキストの差異を提示する機能が整備できれば，唐詩作品の差異の区別などに有用となると考え，同じ唐詩作品に関わる唐詩作品のインスタンスの相違点を利用者に閲覧できるようになる利益がある．

2. 関連研究

2.1 古典籍の校合研究

日本の古典籍に関わる古文書など紙媒体を校合する研究が多数存在する．それらは，底本に採用された古文と底本以外で重要な写本と比べて，異なる部分を調べて，適用な方法で新たな校訂本を作るというものである．ただし，これらの研究を進んでくる時，様々な問題点がある．例えば，荊木ら[8]は新道大系『肥前国風土記』を事例として，古典の校訂方法に関する考察を行った．また，古文書を復元などの研究を進むために，底本か校訂本などをどのような差異があるのをわかるようにすれば，時間や，労働力も節約できると考える．例えば，田中[9]は古典籍を校訂する場合は，写本は必ず底本か校合本に使われる必要がある．ただ，現在まで発行された校訂本は何本もあり，それぞれ別の校訂者によって，本文の復元が行われている．これらに基づく本文の確定が難しくなると考える．また，諸本を確認するだけでも労働力や，時間がかかるため，これらの校訂本の特色や異同をわかりやすく示すことの重要性があると指摘した[9]．次に，守[1]は「歴代法案」という外交文書の校合を中心として，現存する写本による校訂には誤写や文字の欠落など大きな問題があったと述べ，校合によって復元が可能だと判断された上で，復元まで大きな労働力や，時間がかかるという問題を述べた[1]．

日本の古典籍だけでは無く，中国の古典籍を日本で利用する時，その資料の校合研究もある．例えば，張[10]は四書の一つである「大学」を原典とした「大学書」を取り上げて，それぞれの資料に相違する箇所のあることとは言及されたが，詳しく内容を明らかにされていない実態を述べた[10]．また，校合の学問意義としては，講者の講述に対して，その内容を意図的に改変するのではなく，聞書抄というものを講者の学説を反映できるものにするための調整作業ともなって，同時に後世へと流布できるための土台作りともなっていると言えるのではないかと指摘した[10]．

2.2 デジタル化による校合研究

近年，膨大な情報量に対して，デジタル世界に，デジタル書籍の校合手法もある．石田[11]は蔵訳「賢愚経」第２章のテキストを生物系統学の系統推定手法で系統樹を作成して，テクストの校訂の「テキストの系統分類」と「校合のための底本の選択」の作業を処理する方法を試みた．それは，多数の底本と他のテキストと校合し読みを選択する手法をまだ検討中だと指摘された．生物系統学で系統推定手法を系統推定で各テキストの位置づけが可視化されて，詳しい差分がどこにあるのが提示してない問題があるとわかった[11]．安形[12]はデジタル画像を用いた刊本の校合の手法を提案した．それ

は画像の静的重ね合わせと画像の動的重ね合わせという 2 つの手法を用いる.校合の手法を提案した.画像の静的重ね合わせとは画像処理ソフトウェアを使って,2 つの画像の解像度を合わせ,傾きなどを調整した一方で,片方の画像を半透明にし,二枚の画像を重ね合わせる,そうすれば,異なる部分はぶれたようにぼやけて見える.この手法に必要とされるのは,複数のレイヤーを表示することができるソフトウェアである.また,画像の動的重ね合わせとは高速で複数の画像を切り替えながら表示して,その違いを見られるようにする.この手法の限界としては,画像データの入手が難しく,最終的な判断を作業者の肉眼に頼っている点である[12].また,大内[13]は源氏物語の異文校合の自動処理を提案した.源氏物語を研究する際の必要な前提としては本文の吟味である.それで,近年まで膨大な数の異本が存在する作品の本文を整理するや,一覧するデータに対応して,あるフォーマットにしたがって,異文の様子を簡単に確認できるものであるツール「Kogetsu」を紹介した.このツールを使用する前提としては,データがすでに,CSV 形式で提供されている必要がある.大内は紙媒体を電子化したデータの整備が望ましいと指摘した[13].

デジタル化による校合研究における Susan[14]は Versioning machine とは,TEI マークアップを行った上で,二つのファイルの差異を比較し・差異を表示するツールであると指摘した.Versioning machine5.0 を掲載されるウェブサイト[15]は Versioning Machine による校合テキストの表示が,同行の二つのコンテンツに一つをクリックして,差異がある場合,結果表示については,行単位で比較しつつ,行内の単語ごとに比較する必要がある.また,結果を表現する方法は,その差異がある単語をクリックすると,結果が出てくるように表示している[15].

これらの研究は,紙媒体の資料の校訂する方法の紹介や,校合テキストの考察,デジタル化に関わる画像などのデジタル化に関わる古典籍などの校合する研究がある.それぞれの研究に扱う資料の復元や,校訂などに基づいて,多少の差異を比較して,校合テキストで直観的に利用者に提供できれば,時間や手間が節約でき,望ましい.

そのため,本研究では,唐詩作品を研究対象として,唐詩作品のデジタルテキストの差異提示機能を検討する.

3. 研究対象

3.1 唐詩作品

本研究では, 唐詩作品を対象として, 同一の唐詩作品 Work における唐詩作品のインスタンス同士に存在する差異提示機能を検討する.

本研究の研究対象としては,唐詩作品LOD化を対象データとする[2][3][4].平成28年度使用の中学校と高等学校の国語と古典の教科書に含まれる唐詩作品を研究対象として, 唐詩作品の利用状況が含まれる.その上で,同一の唐詩作品Workにおける唐詩作品のインスタンス同士に差異が存在する状況もわかった.

具体的には,日本国内で使用されていた平成28 年度の中学校と高等学校の教科書を用いて,それらの教科書における唐詩作品の利用状況を調べた.唐詩作品 work における異なり 59 首があり,それらに関わる唐詩作品のインスタンスを含む数は延べ 362 首である.

本研究では,これらの唐詩作品のインスタンスを校合する対象データする.唐詩作品 LOD における唐詩作品のインスタンス同士に差異があるかどうかを比べ合わせて,唐詩作品のインスタンスの間に相違点がある場合,それらの差異を比較した上で,校合・提示する方法を検討したい.

3.2 校合テキスト

1 節で述べた通り,本研究で扱う校合(きょうごう; collation; collating)とは,図書館情報学用語辞典[7]に定義を用いて,写本の本文の異同, 既出の刊本の本文の異同,そして特に初期刊本の同一版内での本文の異同を比較,記述する作業である.具体的には,唐詩作品のインスタンス同士に存在する差異を比較・記述・提示することである.

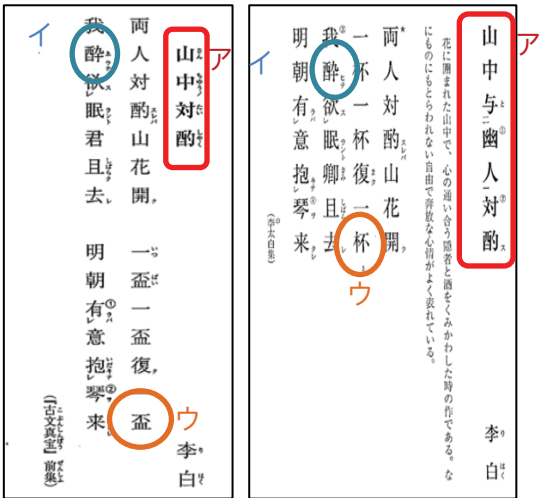
校合テキストは,本文の異同を照合・比較し,記述する作業であると考えつつ,本研究で扱う唐詩作品の校合テキストはここまでの唐詩の LOD 化[2][3][4]と TEI マークアップ手法[5][6]の研究成果に基づいて, 平成 28 年度の教科書に掲載された唐詩作品のインスタンス同士を比べ合わせた後, その差異を示す出力テキストである.この唐詩作品の差異提示機能は,同一の唐詩作品のインスタンス同士を比べ合わせた上で,異同を詳細に提示して,表現する機能であると考える.

また,本研究で扱う唐詩作品の校合対象としては唐詩作品の(1)タイトル, (2)訓点情報, (3)漢字,(4)漢字の順序がある.

例えば, 唐詩作品「山中対酌」を事例として唐詩作品の校合対象を説明する. 具体例を図1に示す.

唐詩作品LODにおけるWorkのtangpoem:18「山

中対酌」は，唐詩作品tangpoem:instance/102「山中与幽人対酌」[17]と唐詩作品tangpoem:instance/313「山中対酌」[16]という2首の唐詩作品インスタンスが含まれる．この唐詩作品インスタンス同士を比較すると，(1) まず，唐詩作品インスタンスのタイトルに差異が存在する．図1(ア)に示されるように，唐詩作品のインスタンスtangpoem:instance/102[17]は唐詩作品のインスタンスtangpoem:instance/313[16]より，タイトルに人物「与幽人」を加わっていることがわかる．(2) 次に，(イ)のように，唐詩作品の本文中の漢字に付随する訓点情報の違いもあり，唐詩作品のインスタンスtangpoem:instance/102「山中与幽人対酌」[17]では送り仮名“ヒテ”が付く一方，唐詩作品のインスタンスtangpoem:instance/313「山中対酌」[16]では“ウテ”が付随された．また，(3) 本文フルテキストに漢字の利用が違うことを(ウ)のところに表示されて，唐詩作品のインスタンスtangpoem:instance/102「山中与幽人対酌」[17]が漢字“一杯一杯”を利用する一方，唐詩作品のインスタンスtangpoem:instance/313「山中対酌」[16]が“一盃一盃”を使うことがわかる．



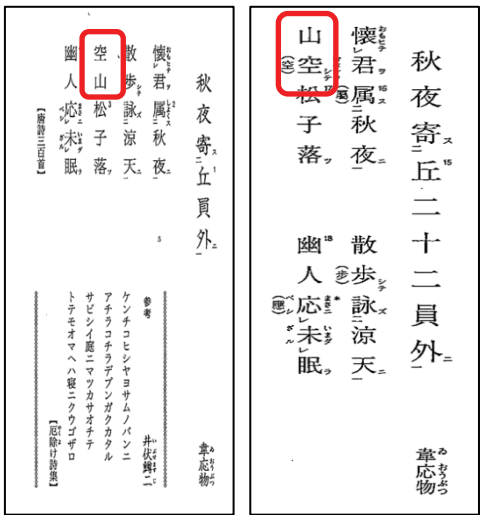
[16] [17]

*ア　タイトルの差異　　*イ　訓点情報の差異
*ウ　漢字利用の差異

図 1 唐詩作品インスタンス同士における差異の事例

そのほか，(4)漢字の順序の相違点である．例えば，唐詩作品Work「秋夜寄丘二十二員外」を事例として，唐詩作品の相違点を説明する．具体例は図2に示す．

図2の事例は唐詩作品のインスタンスtangpoem:instance/252[17]と唐詩作品のインスタンスtangpoem:instance/156[19]である．これらの唐詩作品のインスタンスの相違点が図中に示したが，同じの唐詩作品Workにおける二つの唐詩作品のインスタンスの間に本文内の漢字の順序が変わっている．漢字の順序の変更の問題によって，その本文の意味も変わっている可能性がある．



[18] [19]

図 2 唐詩作品インスタンス同士における コンテンツの差異の事例

4. 提案手法

本研究では，唐詩作品を対象として，同一の唐詩作品Workにおける異なる唐詩作品のインスタンス同士に存在する差異を校合する機能を検討したい．

まず，本研究では唐詩の構造化における唐詩作品のLOD化[2][3][4]と唐詩の本文フルテキストのTEIマークアップ手法の提案[5][6]を中心とした研究成果を踏まえた．それらの研究成果に基づいて，唐詩作品の校合テキストを求める．

唐詩作品の校合テキストは唐詩作品のインスタンス同士に，本文フルテキストの異同を照合・比較し，異同が存在する場合，その差分を記述・表現する作業である．

研究手法としては，LCS (Longest Common Subsequence) アルゴリズム[20]による差分検出であり，差分の結果はターミナルの画面上で表示するだけではなく，唐詩のLOD化[2][3][4]に基づく公開データを利用した上で，HTMLおよびCSS[21]と組み合わせて，校合テキストを記述し，利用者が直接見られるようにする方法で

ある.

本研究で適用する比較方法の全体像としては,同一の唐詩作品のWorkにおける複数の唐詩作品インスタンスがある場合, 任意の２つの唐詩作品のインスタンス同士で,TEI マークアップされた唐詩作品のインスタンスの本文フルテキストのファイルを選択して,比較する.具体的には,まず(1)比較:唐詩作品のインスタンスの二つのコンテンツファイルを探索してみる.(2)唐詩作品のコンテンツに差異が存在しない場合,唐詩作品のインスタンスが変化無く,原文のままで表示される.(3)唐詩作品のコンテンツに差異が存在する場合,唐詩作品のインスタンスの同士に,任意の二つの唐詩作品のインスタンスを比較して,共通の部分が原文のままで表示され,相違点によって,一つのファイルにのみを存在するコンテンツを削除するとみなし,もう片方のコンテンツに新たな内容が追加されたとみなし,その差分における校合テキストをユーザーに見られるように提供する.

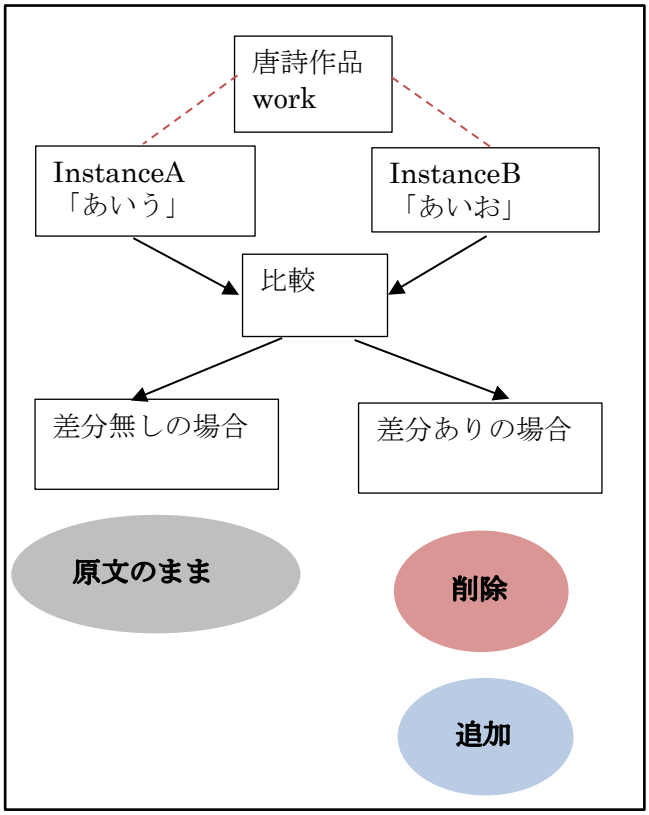


図3 本研究の比較する手法の全体像

本研究の比較する手法の全体像のイメージを図 3 に示す.例えば,図 3 に示すように,ある唐詩作品の Work における唐詩作品のインスタンスは,二つファイル InstanceA と InstanceB であり,それぞれの唐詩作品のインスタンスのコンテンツ内容としては,「あいう」と「あいお」である.二つのファイルを比較する場合,まず,二つのファイルを比べて,コンテンツの相違点を探索する.共通点があるコンテンツ「あい」がそのままで表示されて,差異を見つけた場合,唐詩作品の InstanceA のみ存在するコンテンツの差異を削除とし,InstanceA に存在しなく,InstanceB のコンテンツに存在する新たな内容を InstanceA への追加とする.

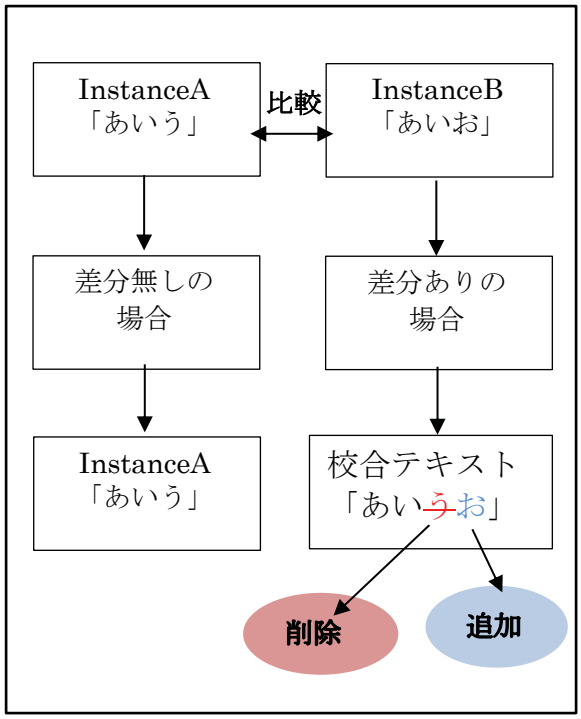


図4 校合テキストを提示する方法

さらに,本研究に扱う校合テキストは差分の提示結果として,図4を示す.本研究で扱う差分の結果はターミナルの画面上で表示するだけではなく,唐詩のLOD化[2][3][4]に基づく公開データを利用した上で,HTML および CSS と組み合わせて,校合テキストを記述し,利用者が直接見られるようにする.例えば,図4のように示して,唐詩作品の InstanceA と InstanceB の共通点が原文のままで表示される.差異が存在する時,表示の方法は削除の箇所は色つけの取消線で差異を表示する.また, 新たな内容の追加は唐詩作品のInstanceBに存在するコンテンツが傍線を色付けで立て, その内容を追加して,表す.

本研究の取消線では赤色で設定し表示されており,新たな内容の追加は青色で縦棒を色付けて表記すると設定した.

5. 結果

この研究における実装は, LCS アルゴリズム[20]による差分検出であり,Ruby プログラム

[21]を利用し，HTML および CSS[21]と組み合わせて，出力することにより実現する．具体的に出力は唐詩作品のインスタンスの間に存在する差異をそれぞれ表し，提示する．

第 3 節の図 1 の事例を利用して，唐詩作品 Work の tangpoem:18「山中対酌」において，唐詩作品のインスタンス tangpoem:instance/102「山中与幽人対酌」[17]と唐詩作品 tangpoem:instance/313「山中対酌」[16]という 2 首の唐詩作品インスタンスを比較して，校合テキストを出力する結果として説明する．

図 1 の通り，唐詩作品 Work の tangpoem:18「山中対酌」を事例として唐詩作品の校合テキストを説明する．(1)唐詩作品「山中対酌」というインスタンスのタイトルに差異が存在する．次に，(2)本文フルテキストに漢字の利用が違う．また，(3)唐詩作品の本文フルテキストに漢字に付随する訓点情報の違いもあるとわかった．これらの唐詩作品のインスタンス同士に比較して，その結果を説明する．

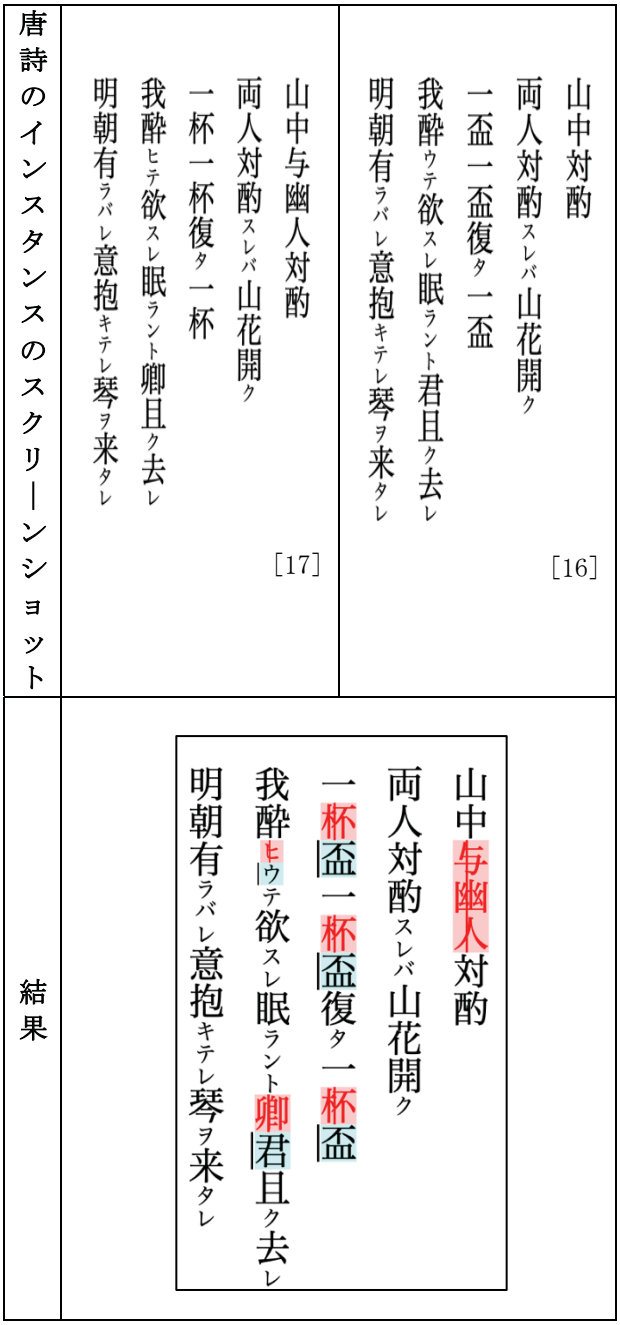
校合テキストの出力結果は表 1 のスクリーンショットになる．表示環境は，MacBook Pro で主に Safari ブラウザ バージョン 11.1 を使い，唐詩作品の校合テキストを表示した．

唐詩作品 Work「山中対酌」は唐詩作品インスタンス tangpoem:instance/102[17] と tangpoem:instance/313[16]の 2 つを持つ．校合テキストの結果としては，唐詩作品インスタンスの同士に(1)タイトル (2)訓点情報 (3)漢字の相違点があると表示された．唐詩作品の校合テキスのスクリーンショットの事例を表 1 に示す．

表 1 の上部に示したスクリーンショットは唐詩作品の TEI マークアップ手法の提案[5][6]に関わるデータ公開するために作ったスクリーンショットであり，そのスクリーンショットに扱う内容は唐詩作品の TEI マークアップ手法の提案[5][6]に基づくマークアップされたファイルである．それらのファイルの内容としては，本文フルテキスの漢字に付随する訓点情報やルビ情報も含まれている．それらの表現はブラウザにより，縦書きで表示されている．それらの研究成果を踏まえて，今回の唐詩作品のインスタンスの差異を校合する結果を示す．

まず，(1) 唐詩作品のインスタンス tangpoem:instance/102 のタイトルは唐詩作品のインスタンス tangpoem:instance/313 のタイトルと比べて，「与幽人」が加わっている．(2) 漢字は唐詩作品のインスタンス tangpoem:instance/102[17]における「杯」の代わりに，唐詩作品のインスタンスの tangpoem:instance/313[16]では「盃」を使用．また，(3) 唐詩作品のインスタンス tangpoem:instance/102[17]に「酔」という漢字に付随する送り仮名が「ヒテ」の代わりに，唐詩作品のインスタンス tangpoem:instance/313[16]に送り仮名とした「ウテ」で付随されている．

表 1 唐詩作品の校合テキストのスクリーンショットの事例



これらの三つの差異が，唐詩作品の校合テキストの結果として表 1 の下部に表示されている．

表 1 の下部に表示されているスクリーンショットによって，唐詩作品のインスタンス同士に差異がある場合，任意の 2 つの唐詩作品のイン

スタンス同士を比較した後,その差異をRubyプログラムにより,HTMLおよびCSS[21]で出力することにより実現できた.校合テキストは唐詩作品のインスタンスの間に存在する差異をそれぞれ表し,提示する機能である.

また,出力テキストを提示する方法は同じの唐詩作品のインスタンス tangpoem:instance/102 と唐詩作品のインスタンス tangpoem:instance/313 を比較した場合,校合テキストの結果を表示する方法は唐詩作品のインスタンス tangpoem:instance/102 にのみ存在する.コンテンツが赤色の取消線で描く.また,唐詩作品のインスタンス tangpoem:instance313 のみ存在するコンテンツの左側に青色で縦棒を立て,その内容を追加し,その結果を表した.

6. 考察

本研究では唐詩作品を研究対象として差異の提示機能という研究を行なったが,研究手法としては,LCS アルゴリズム[20]による差分検出であり,差分の結果はターミナルの画面上で表示するだけではなく,HTML および CSS[21]と組み合わせて表す方法である.

本研究では,唐詩作品の差異を提示する方法が,2 つのファイル同士を比べて,差異が存在するかどうかを明らかにする.唐詩の LOD 化によって,唐詩作品のインスタンスが 2 つよりも多い場合に校合する方法を考える必要がある.例えば,今回利用したデータセットの中に,唐詩作品 Instance が最大 20 回以上存在する唐詩作品 Work があって,それぞれの唐詩作品インスタンスを比較する場合,それぞれの唐詩作品の Instance を選択して,どの二つのインスタンスペアを比較するのかわかりやすく示すのは難しいと考える.そのため,複数のファイルを比較する場合の選択方法を検討する必要があるかと考えている.

また,今回の研究対象データとしては,唐詩作品のLOD化に関わるメタデータ同士を用いて,校合テキストの研究を行なったが,その比較方法は,対象の本文を比較するだけではなく,出典情報の比較も必要になると考える.ただし,今回利用した対象データにおける元データは,各種の出典データを含まれてないため,出典を比較する作業を行なってない.出典データを追加して,出典と本文データの間にどのような相違点があるのを検討する.

本研究では,唐詩の LOD 化と TEI マークアップ提案手法の研究を揃えた上で,唐詩作品のインスタンスの同士に差分の提示機能を検討した.唐詩作品のLOD化に基づいて,唐詩作品のインスタンスに関わる対象データが多くなれば,差異がもっと簡単に表示できると考える.また,古典籍における唐詩作品に扱う訓点情報などで標準化されて,研究を進んで来たら,詳細な作品に関わる差異の比較を校合する作業が簡単になることが期待できる.

7. おわりに

本研究では,唐詩作品を対象として,同一の唐詩作品が異なる出版物に掲載される場合の差異を提示する機能を検討した.

具体的には平成28年度使用の中学校と高等学校の国語と古典の教科書に含まれる唐詩作品を対象として,同一の唐詩作品の間の差異の状況を述べた上で,校合テキストの差異を提示するLCSアルゴリズムを用いて,作成し,比較する方法を議論した.提示する方法は,表示画面で,縦棒や色などのフォントで区別して,差分を提示したことである.本研究では,唐詩のLOD化とTEIマークアップに基づいて,唐詩作品のインスタンスのデジタル化フルテキストの同士を容易に比較可能な環境を構築した.

今後の予定としては,ここまでの研究結果を踏まえて,このテキストの提示機能をどうやって広くほかのテキスト等の対象にも適用可能になるかを検討したいと考える.また,唐詩作品同士の差異を校合するだけではなく,唐詩作品の出典を追加し,出典に掲載される唐詩作品と校訂本を比べて,差異が存在するかどうかを比較したいと考える.

謝辞

本研究成果の一部はJSPS科研費 JP19H04420, JP17K00449 の助成を受けたものです.

参考文献

[1] 守赤嶺. 「歴代宝案の校訂と档案史料 : 国立故宮博物院収蔵档案史料との校合を中心に」. 琉球アジア文化論集 : 琉球大学法文学部紀要,vol.4, pp.1-14,2018-03.

[2] 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).

[3] 叢艶, 江草由佳, 高久雅生. 唐詩情報の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(SWO) 第39回研究会, 東京, 2016年09月05日.

[4] 叢艶, 高久雅生. 唐詩情報のLinked Data化の試み. 情報メディア学会第15回研究大会, つくば, 2016年06月25日, p.17-20.

[5] 叢艶, 高久雅生. 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, 情報知識学会第26回年次大会. 東京, 2018年05月27日. 情報知識学会誌, Vol.28, No.2, p.174-185.
[6] Yan CONG, Masao TAKAKU. “A TEI Markup for the Contents of Tang Poems”. Japanese Association for Digital Humanities Conference 2018 (JADH2018), Tokyo, Japan, pp.80-81. (2018-09-11).
[7] 日本図書館情報学会 用語辞典編集委員会.「図書館情報学用語辞典」(第4版).丸善出版. p.51.
[8] 荊木治恵,林崎治恵.「古典の校訂方法に関する-考察:新道大系「肥前国風土記」を事例として」. 四條畷学園短期大学紀要,vol.48, pp.30-36, 2015.
[9] 田中卓.「古典校訂に関する再検討と新提案」.神道古典研究所紀要(3).29-52,1997-03.
[10] 張 硯君.「林宗和聞書抄『大学抄』の生成とその価値 ―講述聞書における校合の実態をめぐって―」.国際日本文学研究集会会議録.no.41, pp.114-93, 2018-03-28.
[11] 石田勝世.「生物系統学の系統推定手法を利用した蔵訳『賢愚経』テキスト校訂の試み」.印度学佛教学研究, vol.67, No.3, pp.1210-1215, 2019.
[12] 安形麻理.「デジタル画像を用いた刊本の校合の手法」.三田図書館・情報学会.no.53, pp.1-17,2005.
[13] 大内英範. 「源氏物語の異文校合--自動処理の活用」.漢字文献情報処理研究.vol.3, pp.46-50,2002-10.
[14] Susan Schreibman. Versioning Machine 5. Literary and Linguistic Computing, Vol.18, no.1, April 2003, pp.101-107, https://doi.org/10.1093/llc/18.1.101.
[15] Susan Schreibman. Versioning Machine. http://v-machine.org/samples/, (accessed 2019-10-25).
[16]「山中対酌」, 古典B漢文編, 木下資一 ほか14名. 数研出版,p.16.2013.
[17]「山中与幽人対酌」, 新編国語総合, 北原保雄 ほか21名.大修館書店,p.316.2013.
[18]「秋夜寄丘員外」, 精選古典B 漢文編, 三角洋一 池内輝雄 小町谷照彦 ほか27名, 東京書籍, p.21,2013.
[19] 「秋夜寄丘二十二員外」,国語総合, 紅野謙介 鈴木日出男 ほか9名. 筑摩書房, p.373, 2012.
[20] James W. Hunt, Thomas G. Szymanski. “A Fast Algorithm for Computing Longest Common Subsequences”, Communications of ACM, vol.20, no.5, pp.350-353, May 1977.
[21] Elika J. Etemad; Koji Ishii: “CSS Ruby Layout Module Level”, https://www.w3.org/TR/css-ruby-1/, (accessed 2018-04-13).

