

『日本語歴史コーパス』への TEI 適用に基づく諸本比較 ——『万葉集』における「読添えのモ」を事例として——

小池俊希¹ 大向一輝² 鴻野知暁³ 永崎研宣⁴

概要: 『万葉集』は、現存する最古の歌集であり、漢字のみで記された資料である。多くの写本や注釈書には、漢字列である「本文」と、その本文から再現された「訓」が並記されるが、訓は後世に推定されたものであるため異同が生じることが少なくない。日本古典文学作品における最も代表的なコーパスである『日本語歴史コーパス』には、2020年3月に『万葉集校本データベース』へのリンクが追加されたものの、諸本の比較はいまだ困難である。本研究では、『万葉集』の「読添えのモ」を事例として、写本や注釈書のテキストを TEI P5 ガイドラインに準拠してマークアップすることにより、諸本比較を可能とする手法の構築を試みる。

A Comparison of Manuscripts Based on Application of TEI to the Corpus of Historical Japanese ——The Case of “Mo” of Yomisoe in *Man'yōshū*——

TOSHIKI KOIKE^{†1} IKKI OHMUKAI^{†2}
KONO TOMOAKI^{†3} KIYONORI NAGASAKI^{†4}

Abstract: *Man'yōshū* is the oldest anthology of Japanese poet in Japan, composed by Chinese characters. There have been many variants among reading texts of the poets in its witnesses which consist of many manuscripts and its commentaries, because the reading was supposed in latter days. However, *the Corpus of Historical Japanese*, which is a representative of a series of corpora of Japanese classics, has not yet had a convenient function to compare the variants, while it has been linked to *the Man'yōshū Kohon Database*, which provides a view to list fragments of images of the witnesses. In this paper, we will present a method to compare witnesses by marking-up fragments of texts including “mo” of yomisoe in *Man'yōshū* according to the TEI P5 Guidelines.

1. はじめに

本研究で取り扱う『万葉集』は、8世紀後半に成立したとされる、日本に現存する最古の歌集である。やや先行して成立した『古事記』『日本書紀』と共に、上代語（＝奈良時代の日本語）を研究する上での第一級資料とされ、特に『万葉集』は、約4500首を収録し、その語数は約100000語に上るなど、他資料に比して言語量が非常に豊富であるため、上代語研究には不可欠の資料である。

小学館『新編日本古典文学全集』や岩波書店『新日本古典文学大系』、塙書房『万葉集』など、『万葉集』の紙媒体のテキストは枚挙に暇がなく、『新編国歌大観』や前述の塙書房本は、電子テキストとしても刊行されている。また、コーパスに関しては、『日本語歴史コーパス (CHJ)』や『オックスフォード・NINJAL 上代語コーパス (ONCOJ)』に収録されており、それらのコーパスを活用した研究もすでに発表されている[1]。

『日本語歴史コーパス』[2]は、国立国語研究所によって

開発されているコーパスであり、奈良時代から明治・大正に至るまでの日本語資料を収録している。そのうち、『万葉集』は、奈良時代編に収録されており、『新編日本古典文学全集』を底本にしたテキストを、品詞・活用形などの形態素情報が付与された形で分析することができる。

しかしながら、『万葉集』は、後述のように表現形式が多様で、推定や補読を必要とする箇所が少なくないことから、現代のテキストにさえ訓の異同が生じる。そのような場合には、諸本を参照して蓋然性の高い訓を検討する必要があるが、既存のコーパスは、ひとつのテキストに依拠しているため、それらを用いて諸本の比較を行うことは難しい。

『日本語歴史コーパス』は、2020年3月に『万葉集校本データベース』[a]へのリンクを追加したものの、特定の単語などについて諸本を横断して検索することはいまだ不可能である。本研究では、『日本語歴史コーパス』が収録する『新編日本古典文学全集』の『万葉集』（以降、単に『新編全集』と呼称する）を事例として、コーパスが不得手とする諸本の比較について、その解決方法の一案を提示する。

1 東京大学大学院人文社会系研究科
Graduate School of Humanities and Sociology, University of Tokyo
t-koike@g.ecc.u-tokyo.ac.jp

2 東京大学大学院人文社会系研究科
Graduate School of Humanities and Sociology, University of Tokyo

3 東京大学大学院人文社会系研究科

Graduate School of Humanities and Sociology, University of Tokyo

4 一般財団法人人文情報学研究所

International Institute for Digital Humanities

a) https://www.manyou.gr.jp/SMAN_1/ (参照 2020-05-12)

具体的には、諸本を翻刻して作成したテキストに対して、TEI ガイドラインに準拠してマークアップを行う。その上で、機械可読な形で参照番号を明示することにより、将来的に他の研究者が作成したデータと本研究とが相互に参照できることを目指す。

2. 『万葉集』の概要

2.1 表記法と読添え

『万葉集』は、平仮名や片仮名の成立以前の資料であるため、漢字のみによって記される。日本語が漢字によって表記される際、書記者が想定する訓がひとつであるとしても、その資料を解釈する後世の人々が訓をひとつに断定できるとは限らない。漢字に対する訓というものは、不確定な要素であり、漢字に対してどのような訓を与えたかという資料は、日本語史研究の研究対象たり得るのである。

さて、『万葉集』の表記法には、大きく分けて以下の3つの形式があり、例えば、「ヒサカタノ」という和語を表すには次のような表記が考えられる。

1. 訓字主体表記（万葉仮名による補助のないもの）
(例) 久方 / 久堅
2. 訓字主体表記（万葉仮名による補助のあるもの）
(例) 久方乃 / 久方之 / 久堅乃 …
3. 万葉仮名主体表記
(例) 比左迦多能 / 比佐可多能 …

日本語史研究における有用性の高低は、1~3で同一ではない。本文（漢字列）からその訓がほぼ確実に再現できる3の価値が最も高く、次いで補読や推定の必要な2、1の順になる。特に、1で例示した「久方 / 久堅（ヒサカタノ）」における「ノ」は、その訓に対応する漢字が存在しないため、「読添え」と呼称される。「読添え」は、蜂矢宣朗氏[3]によってその傾向がある程度明らかにされているものの、やはり対応する漢字がないことから、写本間はもちろん、現代のテキスト間においても異同が少なくない。本研究では、異同が比較的多い「読添え」の現象に着目し、『新編全集』において、「読添え」られる助詞モ（以降、「読添えのモ」と呼称する）を含む句を事例として、『万葉集』の諸本比較を試行する。

2.2 本研究が分析対象とする資料

『万葉集』の原本は現存せず、写本や写本を元にした版

b) 951年の詔による梨壺における訓読が「古点（本）」、鎌倉中期の仙覚による訓読が「新点（本）」とされる。そして、古点本以前に本文のみで記されていたと推定される本が「無訓本」であり、古点本から新点本までの間の訓読が「次点（本）」である。なお、無訓本は現存せず、古点本も一部の写本にその痕跡を残すのみである。

c) 1~4については、『万葉集校本データベース』（前掲注 a）を参照してテキストを作成した。なお、「2.広瀬本」については、後掲注 d のデジタルア

本が残るのみである。『校本万葉集』[4]の分類に従うと、写本や版本は、無訓本・古点本・次点本・新点本に分類される[b]。本研究では、比較的多くの巻数を現代に伝える次点本と新点本を分析の対象とする。写本の他に、江戸時代に刊行された版本や注釈書、そして現代における校訂テキストも分析の対象とする。取り扱う資料と、『万葉事始』[5]による簡潔な説明は以下の通りである[c]。なお、底本関係などの説明を適宜加えた。

1. 紀州本
…前10巻は鎌倉末期写の次点本系統であり、後10巻は室町末期写の新点本系統。
2. 広瀬本[d]
…江戸中期写の次点本系統。次点本系統唯一の完本。
3. 西本願寺本
…鎌倉後期写の新点本系統。最古の完本。「6.旧大系」「7.新編全集」「8.塙書房」の底本であり、近年の校訂テキストの多くが底本に採用する。
4. 寛永版本
…江戸初期刊の流布本。『校本万葉集』や「5.万葉代匠記」の底本に採用される。
5. 万葉代匠記[6][e]
…契沖による江戸初期の注釈書。全歌に対する注釈書としては非常に早いもので、後世に与えた影響も大きい。初校本（初）と精選本（精）に分かれる。
6. 日本古典文学大系[7]
…1957-62刊。『オックスフォード・NINJAL 上代語コーパス』に採録される『万葉集』の参照テキスト。
7. 新編日本古典文学全集[8]
…1994-96刊。『日本語歴史コーパス』に採録される『万葉集』の底本。本文中では『新編全集』と呼称する。
8. 塙書房『万葉集』本文篇（補訂版）[9]
…1998刊。1963刊の版に対して新出資料である広瀬本を校合に加えた版。

2.3 訓の異同

次に、『万葉集』の訓の異同の実例として、『万葉集』巻12・2959番歌の2句目、「言絶有」に対する諸本の訓を示す。なお、『新編全集』の訓は「コトモタエタリ」であり、3文字目の「モ」が「読添えのモ」にあたる。

1. 紀州本 コトタエタレヤ

ーカイクも併せて参照した。

d) <https://www.lib.kansai-u.ac.jp/webopac/DM50702470> (参照 2020-05-12)

e) 『万葉代匠記』は本文などを提示しないため、契沖が参照したとされる寛永版本に対して注釈を参照して適宜改訂する形でテキストを作成した。

- | | |
|-----------|---------|
| 2. 広瀬本 | コトタエレヤ |
| 3. 西本願寺本 | コトタエタレヤ |
| 4. 寛永版本 | コトタエタレヤ |
| 5a.代匠記(初) | コトタエタルヲ |
| 5b.代匠記(精) | コトタエタルヲ |
| 6. 旧大系 | コトハタエタリ |
| 7. 新編全集 | コトモタエタリ |
| 8. 塙書房 | コトタエニタリ |

諸本共に、活用による差異はあるものの、「言」に「コト」、「絶」に「タク」、「有」に「(タ)リ」という訓を与える点は共通している。下線を付した箇所が読添えの語ということになるが、1~4の写本・版本は終助詞ヤ、『万葉代匠記』は終助詞ヲを補い、現代のテキストは係助詞ハ・モ、助動詞ヌを読添えており、異同が大きいさまを見て取ることができる。

なお、訓の異同は読添え以外の箇所にも起こり得る。次に示す例は、『万葉集』巻16・3863番歌の5句目、「不楽有哉」に対する諸本の訓である。

- | | |
|-----------|----------|
| 1. 紀州本 | カナシクモアレヤ |
| 2. 広瀬本 | タノシカラスヤ |
| 3. 西本願寺本 | カナシクモアルカ |
| 4. 寛永版本 | カナシクモアルカ |
| 5a.代匠記(初) | サヒシクモアレヤ |
| 5b.代匠記(精) | サヒシクモアルカ |
| 6. 旧大系 | サブシクモアルカ |
| 7. 新編全集 | サブシクモアルカ |
| 8. 塙書房 | サブシクモアルカ |

広瀬本のみ「モ」を読添えないが、残る諸本の読添えは同一である。この例では、「不楽」に対する訓に「カナシ」・「タノシ(カラ)ズ」・「サビシ」・「サブシ」の異同が認められる。また、「哉」の訓にも「ヤ」・「カ」の異同が認められる。このように、正訓字に対する訓にも異同が生じやすい。

3. 諸本比較機能の設計

2.2節で示した『万葉集』の諸本について、『新編全集』のテキストに依拠しつつ、テキストの比較を可能とするデータ構造を検討する。これにあたっては、欧米の人文学資料において広く用いられる TEI (Text Encoding Initiative) P5 ガイドライン[10] (以下、TEI-G) に準拠したマークアップ手法を設計する。特に、校本の異同を記述する手法に関しては、TEI-G の第12章 Critical Apparatus において詳述されており、(1) location-referenced method, (2) double-end-point-attached method, (3) parallel segmentation method の3種が

提示されている。ここでは、対応するアプリケーションの多さと記述内容の詳細化とのバランスから、(3) parallel segmentation method を採用する。

3.1 『万葉集』のテキスト構造

『万葉集』のテキストは、漢字列である「本文」、その本文から再現される「訓」、そして、漢文体で記される「題詞／左注」の3つの要素から構成される。本研究では、このうちの「本文」と「訓」を分析対象とする。「題詞／左注」も『万葉集』を構成する要素のひとつであるが、これらは漢文体であり、純粋な日本語資料とは言いがたいため、『日本語歴史コーパス』と同様に分析対象とするテキストからは除外する。

また、写本を分析対象とする都合上、「本文」と「訓」以外に、「書き入れ」(図1の場合は「ミセケチ」)による校訂なども反映させ、校訂などが施されている場合は、校訂前のテキストと校訂後のテキストを並記する形でテキストを作成する。

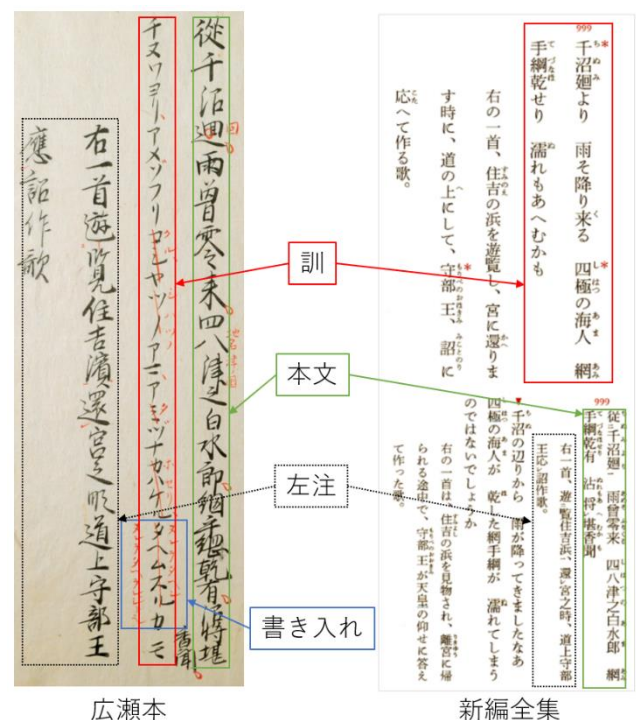


図1 『万葉集』のテキスト構造

3.2 TEI 適用の方法

以上の検討を踏まえつつ、TEI-G が提示する構造とすりあわせた結果、現時点では、以下のような構造とすることとした。

『新編全集』の歌ごとに<I>を付与し、xml:id として巻数・歌番号を付す。そして、本文もしくは訓に対して異文が存在する場合には、<app>を用いて、参照した各本に割り当てた xml:id を@wit 属性で参照し、『新編全集』のテキストは

<lem>を使用して、それ以外のものは<rdg>を使用して記す。また、書き入れに関しては、補入は<add>、別訓は訓と同様にマークアップしつつその位置を@rend で記述して区別できるようにし、青墨や朱書きは文字色を@style で記述、ミセケチは で表記する。

ここまでは TEI -G が提示する典型的な記述手法で済んでいるが、少し工夫を要する点として、『万葉集』の特徴として、本文に対して訓が平行に存在する点が挙げられる。本研究ではそのうちの訓を主要な研究対象とするため、それぞれの異同を示しつつそれらを一覧しやすくすることが必要となる。今回の試行では、本文に対する訓をルビのような扱いとして記述し、本文と訓のいずれか、もしくは双方が異なる場合には、本文に<app>として付することとした。ルビに関しては、現在の TEI-G では対応するタグが用意されていないため、将来、対応した際に変更することを見据えて、<seg>と@type を用いて記述した。ただし、通常のルビとは異なる処理が必要となることも考えられるため、@type の値は、ruby/rb/rt のようなものではなく、暫定的に manyo/honmon/kun とした。これは、いずれ『万葉集』研究における TEI マークアップ手法が整備されていくなかで、それに合わせて変更していくことを企図している。

以上のような設計に基づいて、まずは、訓のみが異なる例のマークアップを以下に示す。

```
<| xml:id="sp11-2360">
  <seg type="manyo">
    <seg type="honmon">人祖未通女兒居守山辺柄朝と通公 <app>
      <lem wit="#旧 #新 #稿">
        <seg type="manyo">
          <seg type="honmon">不來哀</seg>
          <seg type="kun">不來哀</seg>
        </seg>
      </lem>
      <rdg wit="#紀 #広 #西 #寛 #初 #精"> 訓の異同
        <seg type="manyo">
          <seg type="honmon">不來哀</seg>
          <seg type="kun">不來哀</seg>
        </seg>
      </rdg>
    </app>
  </seg>
  <seg type="kun">人の親の娘子供居ゑ守山辺から朝な朝な
    通ひし君が来ねば悲しも</seg>
</|>
```

図 2 訓のみが異なる場合のマークアップ例

次に、訓と本文がそれぞれ異なる例のマークアップを以下に示す。

```
<| xml:id="sp10-1895">
  <seg type="manyo">
    <seg type="honmon">春去先三枝幸命在 <app>
      <lem wit="#精 #旧 #新 #稿">
        <seg type="manyo">
          <seg type="honmon">後相</seg>
          <seg type="kun">ノチモアヒム</seg>
        </seg>
      </lem>
      <rdg wit="#紀">
        <seg type="manyo">
          <seg type="honmon">後相武</seg>
          <seg type="kun">ノチモアヒム</seg>
        </seg>
      </rdg>
      <rdg wit="#広 #寛 #初">
        <seg type="manyo">
          <seg type="honmon">後相</seg>
          <seg type="kun">ノチモアヒム</seg>
        </seg>
      </rdg>
      <rdg wit="#西">
        <seg type="manyo">
          <seg type="honmon">後相</seg>
          <seg type="kun">ノチモアヒム</seg>
        </seg>
      </rdg>
    </app> 莫恋吾妹</seg>
    <seg type="kun">春さればまづ三枝の幸くあらば
      後にも逢はむな恋ひそ我妹</seg>
  </seg>
</|>
```

図 3 訓と本文がそれぞれ異なる場合のマークアップ例

このようなマークアップにより、『新編全集』のテキストを基礎として、諸本における異同箇所を、文脈も含めて容易に確認できるようになる。ただし、研究資料として広く共有しようとする場合、権利関係の問題から『新編全集』のテキストそのものを共有することはできない。この場合には、『万葉集』における巻数・歌番号などによる参照番号を、<|>における xml:id のように機械可読な形で明示することで、研究者が作成したデータを相互に活用しやすくするという方法があり得るであろう。

4. 諸本比較の一例

以上のようなマークアップ作業により、例えば、諸本を視覚的に対照させることができるようになる。そうすることで、異同の有無の調査や、歴史的な訓の変遷の調査を行う際に、これまで以上に効率的に作業を進めることが可能となる。

4.1 テキストの異同の調査

TEI-G の parallel segmentation method に従って作成したデータに TEI Critical Apparatus Toolbox[11]を用いることで、異同がある箇所を橙色で視覚的に対照することができる(逆に異同がない箇所は白色で示される)。テキストの異同の調査の実例として、『万葉集』巻3・335 番歌の5句目を取り上げ、諸本比較の一部を図示する。

<p>Text according to 紀</p> <p>吾行者久者不有夢乃和太湍者不成而 淵有毛 フチニアリ トモ！我が行きは久にはあらし夢のわだ瀬にはならず</p>
<p>Text according to 初</p> <p>吾行者久者不有夢乃和太湍者不成而 淵有毛 フチニアレ ヤモ！我が行きは久にはあらし夢のわだ瀬にはならず</p>
<p>Text according to 新</p> <p>吾行者久者不有夢乃和太湍者不成而 淵有毛 フチニモア リコソ！我が行きは久にはあらし夢のわだ瀬にはなら</p>
<p>Text according to 埜</p> <p>吾行者久者不有夢乃和太湍者不成而 淵有毛 フチニシヤ リコソ！我が行きは久にはあらし夢のわだ瀬にはなら</p>

図 4 卷 3・335 番歌 5 句目の異同

この例では、『新編全集』(新)の訓に対して、写本や版本は元より、比較的近時に刊行された埜書房本『万葉集』(埜)さえも異なる訓を採用していることが見て取れる。現代においても訓に異同が存するという事は、各テキスト間で本文から再現される最も蓋然性が高い訓に相違が存するという事を示している。このような場合、ひとつのテキストの訓をもって、確実な上代語資料とは言いがたく、用例の集計に用いる際などには注意を要する。

4.2 訓の変遷の調査

次に、『万葉集』の訓の変遷について取り上げる。訓の変遷が顕著な例として、希求表現「～モ～ヌカ」を取り上げる。「～モ～ヌカ」に「読添えのモ」が用いられる例は7例見られるが、このうちの卷 11・2366 番歌の 3 句目について、TEI Critical Apparatus Toolbox を用いて諸本比較の一部を示す。

Text according to 広

真十鏡見之賀登念妹相可聞イモニアハムカモ！玉緒之絶
 有恋之繁比者まそ鏡見しかと思ふ妹も逢はぬかも玉の

Text according to 精

真十鏡見之賀登念妹相可聞イモニアハムカモ！玉緒之絶
 有恋之繁比者まそ鏡見しかと思ふ妹も逢はぬかも玉の

Text according to 旧

真十鏡見之賀登念妹相可聞イモモアハヌカモ！玉緒之絶
 有恋之繁比者まそ鏡見しかと思ふ妹も逢はぬかも玉の

Text according to 新

真十鏡見之賀登念妹相可聞イモモアハヌカモ！玉緒之絶
 有恋之繁比者まそ鏡見しかと思ふ妹も逢はぬかも玉の

図 5 卷 11・2366 番歌 3 句目の異同

この例では、『万葉代匠記』(精)までの「イモニアハムカモ」という訓が、『日本古典文学大系』(旧)以降は「イモモアハヌカモ」に読み換えられている。また、残る 6 例についても、近世までのテキストには「読添えのモ」がほとんど見られず、それとは対照的に現代のテキストでは、「～モ～ヌカ」の訓が多く採用されている。これは、『万葉集』中の他の確実な例から「～モ～ヌカ」という強固な呼応関係が指摘され、また、田島光平氏[12]などにより、希求表現の「～モ～ヌカ」には「不」字が用いられないことが指摘されるなど、『万葉集』研究の進展を反映した結果であると解釈される。

5. 本研究の応用可能性

最後に、本研究の応用可能性について検討する。本研究では、『万葉集』の「読添えのモ」を事例として、諸本の校異情報に対して TEI-G に準拠したマークアップを行った。データを TEI-G の parallel segmentation method に従って作成したことにより、4.1 節・4.2 節に示したように、TEI Critical Apparatus Toolbox のようなツールを用いてデータを活用することができるようになるという利点が、まず生まれる。

また、『新編全集』の巻数や歌番号を、<1>における xml:id によって示すことにより、他のデータとの相互参照が可能となり、将来的には、『日本語歴史コーパス』が付与した形態素情報をあわせて参照することが可能になるものと考え

られる。そうすることで、諸本を比較したデータに対して、品詞情報などを使用した分析が可能になる。

本研究が取り上げた『万葉集』は、漢字に対する訓の与え方が問題となる資料の代表例である。この問題は、『古事記』『日本書紀』などを含む上代語資料や訓点資料などを日本語史研究に活用する際に、それらの資料が根本的に抱えている問題でもある。本研究が示したマークアップの手法をそれらの資料に対して用いることで、諸本の校異をより容易に比較できるようになることが期待される。

謝辞 本研究は、国立国語研究所の共同研究プロジェクト「通時コーパスの構築と日本語史研究の新展開」による研究成果を報告したものである。また、JSPS 科研費 JP19H05477 の助成を受けている。

参考文献

- [1] 鴻野知暁：上代日本語の複合動詞の項構造について—二つの内項を取る場合を中心に—，言語・情報・テキスト（東京大学大学院総合文化研究科言語情報科学専攻紀要）25，pp.41-50（2018）。
- [2] 国立国語研究所：日本語歴史コーパス ver. 2020.03（2020）<https://chunagon.ninjal.ac.jp/>（参照 2020-05-12）
- [3] 蜂矢宣朗：読添へる助詞と読添へぬ助詞，山辺道 10，pp. 1-23（1964）。
- [4] 佐佐木信綱ほか[編]：校本万葉集（新增補追補版），岩波書店（1994-1995）。
- [5] 坂本信幸，毛利正守[編]：万葉事始，和泉書院（1995）。
- [6] 佐佐木信綱ほか[編]：契沖全集 1-4，朝日新聞社（1926）。
- [7] 高木市之助，五味智英，大野晋[校注]：日本古典文学大系 4-7，岩波書店（1957-1962）。
- [8] 小島憲之，木下正俊，東野治之[校注]：新編日本古典文学全集 6-9，小学館（1994-1996）。
- [9] 佐竹昭広，木下正俊，小島憲之[校注]：補訂版 万葉集 本文篇，塙書房（1998）。
- [10] Text Encoding Initiative Consortium.. *P5: Guidelines for Electronic Text Encoding and Interchange*. 2020, Version 4.0.0. <https://tei-c.org/release/doc/tei-p5-doc/en/html/index.html>（参照 2020-05-12）
- [11] Marjorie Burghart.. The TEI Critical Apparatus Toolbox: Empowering Textual Scholars through Display, Control, and Comparison Features. *Journal of the Text Encoding Initiative* [Online], 2016, Issue 10. <http://journals.openedition.org/jtei/1520>（参照 2020-05-12）
- [12] 田島光平：万葉集に於ける「ず」の表記の特色とそれより導かれる種々の問題，国語と国文学 27-3，pp.22-38（1950）。