

広領域連携型基幹研究プロジェクト

「異分野融合による総合書物学の拡張的研究」3ユニット合同シンポジウム

# 『延喜式』の校異注入力と 一括マークアップ

国立歴史民俗博物館「延喜式のデジタル技術による汎用化」

渡邊美紗子 戸村美月

# はじめに

---

## デジタル延喜式

- ・TEIを適用した『延喜式』の校訂文、現代語訳、英訳を公開

校訂文

校異注

現代語訳

ルビ

英訳

英訳注、イタリック表記

## 作業

- ・TEIを適用したXMLデータの枠組み作成
- ・マークアップ、マークアップの確認、既存データとの調整
- ・XMLデータの充実をはかる

# 『延喜式』の校異注入力

## ・校異注のマークアップ方法（●=該当文字）

### テキスト部分（<p>タグ内）

```
<anchor xml:id="app0000"/>●<anchor xml:id="app0000e"/>
```

### テキスト外（<p>タグ外、<div>タグ内）

```
<app from="#app0000" to="#app0000e">  
  <lem wit="#写本版本名">●</lem>  
  <rdg wit="#写本版本名 #写本版本名">○</rdg>  
  <note>● 校異注本文</note>  
</app>
```

## ・内容により、基本形では対応できない校異注がある

訂正書・傍書等を参照した文字校訂  
傍書・頭書等の提示  
『延喜式』以外の関連史料へ言及

## ・デジタル延喜式のビューアとの兼ね合いが悪い

該当文字の範囲が広い  
条文全体に関わる解説  
語句の配列

# マークアップの調整

---

## ・該当文字の範囲が広い、条文全体に関わる解説、語句の配列等

### 【問題点】

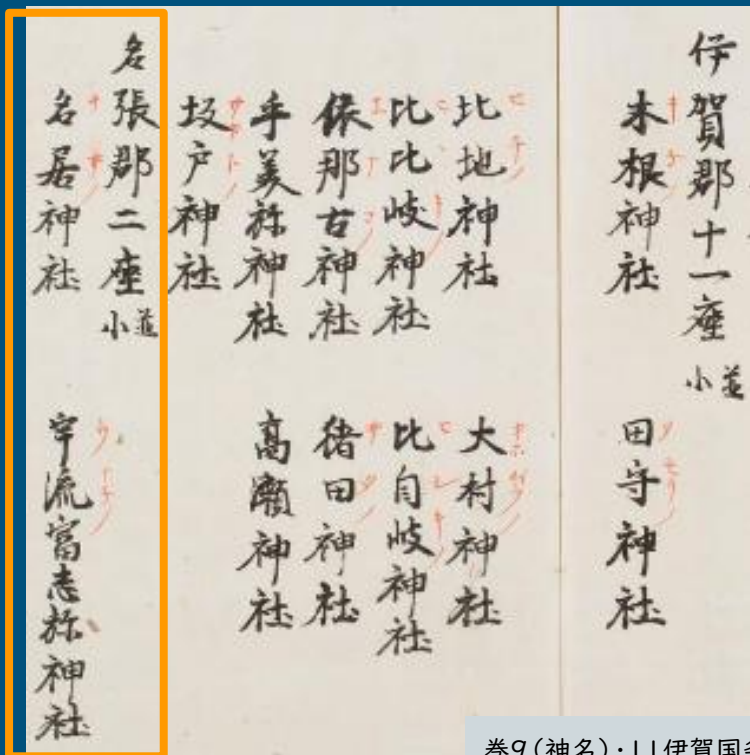
範囲が広いと赤字範囲が多く、ビューアが見にくい

一つの校異注の中に、別の校異注が入り込んだ入子状態になる→入子部分の校異注が選択できない  
段組みを表現するための<table>タグと相性が悪く、ビューア上で文字が消える

### 【対応】

冒頭(単語)にのみ<anchor/>を付しnote(解説)に「冒頭(単語)……文末(単語)」と範囲を提示  
条文全体に関わる解説かつ校異がない場合はnote(解説)のみ記述  
配列の場合は、該当部分(単語)それぞれに校異注を付し、note(解説)のみ記述する

# 事例 デジタル延喜式のビューアと校異注入力



## ・校異注

- ①吉本は名居神社の前に宇流富志弥神社を配列する。
- ②吉本「井」、永本左傍書「井 官」。土本・永本・九本・京博本「居」。

## ・段組み表示にしたい

## 当初のマークアップ

```
<table>
<row>
  <cell rendition="#width300"><anchor xml:id="app0910110501"/>名<anchor
    xml:id="app0910110502"/>居<anchor xml:id="app0910110502e"/>神社</cell>
  <cell rendition="#width50"/>
  <cell rendition="#width300">宇流富志弥神社<anchor xml:id="app0910110501e"/>
</cell>
  <cell rendition="#width50"/>
  <cell rendition="#width300"/>
</row>
</table>
<table>
```

・表形式(<table>)を使用

・校異注の範囲

①「名居神社 宇流富志弥神社」

②「居」

## 起こった不具合

・ビューアから「名居神社 宇流富志弥神社」の表示が消える

原因:<anchor/>と<anchor/>が二つの<cell>に跨って設定されていること

・ビューアから「居」に関する校異注がクリックできない

原因:<app>の<lem>の値をビューアの該当箇所当てはめて表示をしていた

## 修正後のマークアップ

```
<table>
<row>
<cell rendition="#width300">
<anchor xml:id="app0910110501"/> 名 <anchor xml:id="app0910110502"/> 居 <anchor
xml:id="app0910110502e"/> 神社 <anchor xml:id="app0910110501e"/>
</cell>
<cell rendition="#width50"/>
<cell rendition="#width300">
<anchor xml:id="app0910110503"/> 宇流富志弥神社 <anchor xml:id="app0910110503e"/>
</cell>
<cell rendition="#width50"/>
<cell rendition="#width300"/>
</row>
</table>
```

## デジタル延喜式上の表示

[09101105]  
名張郡二座（並小、）  
名居神社

宇流富志弥神社

- ・<anchor/>は<cell>をまたがないよう設定
- ・校異注の範囲変更 ①'名居神社 ②'居 ③'宇流富志弥神社
- ・<p>タグに入力されたテキストをビューアに提示
- ・校異注の入子部分は文字色を青にして表示

# マークアップ作業のPythonによる効率化

---

## 情報学メンバーの協力

- ・ 枠組み生成
- ・ 校異注IDの付与
- ・ 画像URLの一括生成
- ・ ヘッダー部分の一括追加・修正
- ・ データベース公開時に使用するID等の一括調整

# 歴史学作業者のPythonコード活用

---

## 作業者の基盤

- ・基礎的なPython知識の習得
- ・ChatGPTの支援を活用

## 事例紹介

- ・マークアップ支援コードの具体例紹介

(①ルビのチェック ②地名・官司名の一括マークアップ)

# 事例① ルビのチェック作業

---

## 目標

- ・すでに入力されたルビの間違いを抽出
- ・一括修正ではなく、個別判断を行う

## 課題

- ・前後の語句などによってルビ対応が変わる語句の存在

## 対応

- ・ルビ一覧に前後の語句欄を作成し、XMLデータからのルビ抽出データとは対象語句のみ、XMLデータとは結合語句として、それぞれ比較する。

# 事例② 地名・官司名の一括マークアップ

---

## 目標

- ・既に公開中ものを含めたデータに、新たに地名・官司名のマークアップを行う

## 課題

- ・短縮記述とタグにタグを付与してしまうエラーの発生

## 対応

- ・地名・官司名とそれに対応するタグのリストを複数準備し、それぞれデータを置き換える

# 事例①・②に残る課題

---

## 事例①

- ・同じ意味で異なる読み方をする語句への対応 (例:「齋宮」「さいくう」「さいぐう」)
- <taxonomy> タグの充実によって解決可能か

## 事例②

- ・同じ文字列の誤タグ付け (例:「紀伊(国)」「紀伊(郡)」)
  - ・極端な省略表記の扱い (例:「(式部)省」「式(部省)・兵(部省)」)
- 目視と手作業による対応が間違いが起きないか

# まとめと今後の展望

---

## 【まとめ】

- ・校異注入力方法の確立
- ・Python による作業効率の向上

## 【今後の展望】

- ・マークアップ対象の拡大（物品名・度量衡）
- ・<taxonomy> と <standOff> の充実
- ・デジタル『延喜式』活用の補助機能を拡充（検索語句候補の提示）

# (参考資料)

|            | マークアップ内容  | テキスト内 (<p>タグ内) |              |             | テキスト外 (<p>タグ外)   |                          |                               |  |
|------------|-----------|----------------|--------------|-------------|--|--------------------------|-------------------------------|--|
|            |           | タグ1            | タグ2          | @属性         | タグ1  | タグ2                      | @属性                           |  |
| 校訂文        | 校異注       | <anchor>       | —            | @from @to   | <app>  | <lem><br><rdg><br><note> | @wit @cause<br>@wit<br>@place |  |
|            |           | 物品名(数量・単位等も含む) | <measure>    | <measure>   | @ana @commodity<br>@quantity @type<br>@atLeast @sameAs |                          |                               |  |
|            |           |                |              | <unit>      | @unitRef   |                          |                               |  |
|            |           |                | <note>       | @target     |  |                          |                               |  |
|            | 国名        | <placeName>    | —            | @sameAs     |  |                          |                               |  |
| 官司名        | <orgName> | —              | @sameAs      |             |  |                          |                               |  |
| 現代語訳       | ルビ        | <ruby>         | <rb><br><rt> | —<br>@place |  |                          |                               |  |
| 英訳         | 英訳注       | <seg>          | —            | @xml:id     | <note>   | —                        | @target @type                 |  |
|            | イタリック表記   | <seg>          | —            | @rend       | <seg>  | —                        | @rend                         |  |
| 校訂文・現代語訳共通 | 段組み       | <table>        | <row>        | @rendition  |  |                          |                               |  |
|            |           |                | <cell>       | @rendition  |  |                          |                               |  |
|            | インデント     | <s>            | —            | @rendition  |  |                          |                               |  |
|            | 文字の色指定    | —              | —            | —           | <seg>  | —                        | @style                        |  |

※タグ2 はタグ1 内に記述される。