

TEIガイドラインにもとづく 古辞書の電子テキスト化とそ の意義

研究集会 「古辞書データ共有と拡張」

2023/1/21 オンライン

はじめに

- 本発表は岡田「日本古辞書のTEI符号化」（石田友梨ほか（編）『人文学のためのテキストデータ構築入門 TEIガイドラインに準拠した取り組みにむけて』文学通信、2022）をもとに、
- 「異分野融合による総合書物学の拡張的研究」国語研ユニット「古辞書類に基づく語彙資源の拡張と語彙・表記の史的変遷」の成果を取り込んで発表するものです

TEIガイドラインを東アジア資料に適用する意義

- 人文学テキストデータ符号化のための国際的な取り決めのひとつ
- TEIコンソーシアムが維持発展させている
- 欧米圏の資料への適用ばかりが進んでいたが、近年、東アジア資料への適用も試みられるようになった
- 東アジア古典籍資料の構造が現状どのていどTEIによって表現可能か、また、できないとすれば、どのような理由によってできないかを検討することは、単なる一取り決めの適用に留まらないデジタル人文学的課題を含み、かつ、来るべき人文学データ共有のありかたにもかかわる

日本古辞書

- 漢文と和文という2言語（あるいは梵語をくわえて3言語）の交点
- 中国における辞書・字書や注解の伝統を受容するなかで成立
 - 『説文解字』や『干禄字書』、『一切経音義』など
- 日本現存最古の字書が中国の字書『玉篇』を空海が抄した『篆隸万象名義』であるのはたんなる偶然ではあるまい
- のちには『色葉字類抄』のような訓を重視した字書も現れる

類聚名義抄・図書寮本

- ・ 「六帖か。うち一帖のみ存。宮内庁書陵部蔵。平安末期写。11世紀末ごろ成立。著者は法相宗の僧か。古辞書、仏書、古訓点など多数の先行文献を忠実に引用し、国語史学上重要な資料である。」（築島裕「類聚名義抄」『ニッポニカ』）
- ・ 改編本系に完本の観智院旧蔵本などがあるが、先行文献の引用は削除されている
- ・ HDICに両者とも収載

古辞書の項目構造

- 『篆隸万象名義』については李媛氏がつぎのように構造を記述している
(「TEI P5 Dictionariesモジュールに基づく古辞書の構造化記述の試み
—篆隸万象名義を中心に—」 『情報処理学会研究報告』2018-CH-117)
- 項目
 - 掲出字
 - 注文
 - 反切
 - 義注
 - 字体注

古辞書の項目構造

- 劉冠偉氏は、観智院本の構造につき以下のように記述する
（「部首分類体日本古辞書の項目構造の多様性に対応したマークアップ・ツールの開発」 『じんもんこん2017論文集』）
- 注文
 - 音注 - 音級、発音
 - 意義注 - 義級、漢文
 - 字体注 - 字級、異体字
 - 和訓 - 仮名

TEIガイドライン辞書モジュール

- TEIガイドラインの辞書モジュールは、辞書項目の符号化のための要素や属性を提供する
- 辞書項目は大きく以下に分かれ、構造的・非構造的かどうかで、XML構造の制約の度合いが大きく異なる
 - 1. 構造的項目 (<entry>)
 - 2. 非構造的項目 (<entryFree>)
 - 3. 断片 (<dictScrap>)

TEIガイドライン辞書モジュール

- `<entry>`
 - `<form>`
 - `<orth>competitor</orth>`
 - `<hyph>com|peti|tor</hyph>`
 - `<pron>k@m"petit@(r)</pron>`
 - `</form>`
 - `<gramGrp>`
 - `<pos>n</pos>`
 - `</gramGrp>`
 - `<def>person who competes.</def>`
 - `</entry>`
- `<entry>`の下位の記述内容としてはform, gramGrp, def, cit, usg, xr, etym, re, note以外は入れることができない；その他はこのどれかに属す
 - つづりと発音が分離しているようなデータはそのまま構造的項目とすることができない

古辞書に現代の辞書符号化を適用する際の問題点

- 古辞書と現代の辞書とは目的が異なるため、同じ構造を持つと記述してよいのかは問題である
 - 日本古辞書は字と附訓を中心、辞としての記述は概略
 - 現代の辞書では字や附訓は「表記」や「語義」の一部
- 李（2018）では、構造は異なるとして<entryFree>を基本的に採用する

古辞書における字と音の構造

- じっさいには、古辞書も、音の弁別をしたうえで字義などの説明を行う
- 詳 宋云本音祥・慈云安者徐也、一 [詳] 者審也。或以章反一 [詳] 狂也。今從初・中云探玄記云一 [安] 一 [詳] 審諦之状。捷作。養也。之言詳也亦通・弘云審也論也議也也詐也。・
又音与羊同 アキラカ詩 イツハル記 ツハヒラカニ選 眞云
シヤウ
- <entry>の下位には<entry>（やそれに類するもの）を入れることならのできるのので、見出し字ごとに音を別とする語の説明がされていると考えると、構造化記述がまったくできないわけではない

古辞書における引用という構造

- 古辞書における引用は、おおく、求める情報によって分かれる
- したがって、（まったく問題なしとはしないが）引用の目的に応じて辞書項目の説明分類にあわせた符号化をすることができる
- それでもTEIの制約に合わない項目であってはじめて<entryFree>などの非構造的表現を考えるべきではないか

図書寮本の符号化

- ```
<entry xml:id="Z07422">
 <form>
 <orth><oRef target="#Z07372-1"/><!-- 訶 -->梨怛鷄</orth>
 </form>
 <cit>
 <bibl>
 <name>應</name>云</bibl>
 <def>舊云呵梨勒。翻爲天主持來。此果堪爲藥以<choice>
 <sic>切</sic>
 <corr>功</corr>
 </choice>用極多如此土人參石斛等。</def>
 </cit>
 </entry>
```

# 図書寮本符号化テキストへの情報 与

- 既存の語彙資源との連携を図ることによって、注釈テキストとしての性質を富化することができる
- 漢語は思い当たらないが、仏教語や日本語などはjapanナレッジへのリンクが考えられる

# 図書寮本符号化テキストへの情報附与

- `<cit type="japanese"><!-- 不「計」 -->`  
    `<quote ana="https://japanknowledge.com/psnl/display/?lid=200203568fbd4v2YwDT0">ハ<metamark ana="L"/>力<metamark ana="L"/>ル<metamark ana="H"/>`  
    `<bibl><name><add place="bottom">月</add></name></bibl>`  
    `</quote>`  
    `<quote ana="https://japanknowledge.com/psnl/display/?lid=200200d5830dH89dL1JC"> 力<metamark ana="L"/>ソ<metamark ana="L" "/>フ<metamark ana="H"/>`  
    `<bibl><name>律</name></bibl>`  
    `</quote>`  
    `</cit>`

# 図書寮本符号化テキストへの情報附与

- ・ 日本国語大辞典 「かぞえる」

## 表記

【筭】色葉・和玉・易林・書言・ヘボン

【計】色葉・名義・易林・書言

【数】色葉・名義・和玉・言海

【道】色葉・名義・和玉

【量】色葉・名義・書言

【乘・散・笑・勝・原】色葉・名義

【員】色葉・和玉

【往・諷】名義・和玉

【押】色葉

【和・測・科・麗】名義

【呀・樞・道・黠】和玉

## 同訓異字

### かぞえる【数・算・計】

【数】(スウ・ス) かずをかぞえる。かず。「数值」「数量」いくらかの、いくつかの。「数人」「数日」《古かず・かぞふ・あまた・あまたたび・しばしば》

【算】(サン) かずをかぞえる。かずの中に入れる。かず。「算数」「予算」転じて、おしはかる。はかりごと。「算段」「誤算」《古かず・かぞふ・はかる・はかりこと》

【計】(ケイ) かずをかぞえる。数量をはかる。「計算」「計量」「合計」転じて、おしはかる。くわだてる。「早計」「計画」《古かぞふ・かず・はかる・はかりこと・はかりみる・かむがふ》