

# 玉篇系字書の構造化記述に関する TEI マークアップについて

李 媛<sup>1</sup>

## 1 はじめに<sup>2</sup>

古写本古辞書の文献資料に膨大な数の漢字を含み、そのため、従来の研究成果公表は手書きで公刊されることが多く、このような古写本の多漢字文献資料の電子化も遅れる状況にある。2000 年以來、Unicode の普及や進展につれ、パソコンで処理できる漢字数が増加し<sup>3</sup>、古写本古辞書の文献資料は、現代の辞書と比べると、整然とする凡例および項目構造を持つものが少ない。電子化テキストは、言語資料として活用するために、より効率的な検索機能を実装する必要がある。また、オープンソース化の際には、ほかの研究グループや機関との互換性も求められる。これらを実現する手段の一つは、プレーンテキストに翻刻者・辞書研究に関連する附加情報を付け加えること、すなわちタグ付けを行うことである。

TEI P5<sup>[3]</sup>は人文科学学資料のマークアップのためのルールを定めている最新のガイドラインである。元來欧州および北米の大学や研究機構を中心に 1990 年に公表されたもので、対象文献も欧米の文献資料がメインであった。しかし、近年、日本や台湾における TEI P5 の日本語訳<sup>[4][5]</sup>・中国語訳<sup>[6]</sup>の進展の進展、そして TEI を適応した研究プロジェクトの増加から、東アジアの文献資料も対象となりつつある。

TEI P5 では様々なジャンルの文献に対応するモジュールがあり、584 種のタグ（元素）と 268 種の属性等が定義されている（Version 4.4.0, 2022.4.19 改訂）。辞書内容記述に対応する部分は、辞書資料（Dictionaries）モジュール、中核（Core）モジュールおよび写本記述（Manuscript Description）の一部である。TEI P5 に準拠した日本古辞書の構造化記述に関する研究は多くないが、岡田「日本平安期古辞書の符号化モデ

---

<sup>1</sup> 関西大学東西学術研究所 アジア・オープン・リサーチセンター「KU-ORCAS」

<sup>2</sup> 古辞書の構造化記述に関連する研究背景については、2018 年発表者が執筆した論考<sup>[1]</sup>に依拠しつつ述べてみたい。

<sup>3</sup> 永崎「東アジア人文情報学の可能性についての試論」は、2010 年代以降の状況について、次のように指摘している。「この時期には、Unicode で使用可能な文字が飛躍的に増加しただけでなく、Unicode では同定されてしまう微細な漢字の字形差についてもテキストデータのレベルで記述できるようにする IVS（Ideographic Variation Sequence）が普及し、10 万種類以上の漢字字形が Unicode で利用可能となった。」（p.253）<sup>[2]</sup>

ル」<sup>47)</sup>は TEI をもとにした日本古辞書の効率的な符号化モデルについての論考であり、参考するところが大きい。

本稿は日本古辞書編纂に多大な影響を及ぼす玉篇系字書（原本『玉篇』、『篆隸万象名義』、宋本『玉篇』）を対象に、岡田「日本平安期古辞書の符号化モデル」にて提示されたモデルを参照し、玉篇系字書の構造化記述に関する TEI マークアップを試みるものである。

## 2 対象資料

本研究は、玉篇系字書（原本『玉篇』残巻、『篆隸万象名義』、宋本『玉篇』）を中心資料とする。原本『玉篇』、そして原本『玉篇』と同じ部首分類体系を持つ『篆隸万象名義』、宋本『玉篇』、これら三書を合わせて「玉篇系字書」と呼ぶ。

池田「平安時代漢字字書総合データベース構築の方法と課題」<sup>48)</sup>では、『玉篇』について、次のように「系統 (genealogy)」と「体系 (system)」とそれぞれの意味用法を示し、さらに切韻系韻書と比べながら、玉篇系字書について説明している。

原本系『玉篇』と「系」を付けて系統 (genealogy) の意味で使用することが最近は多くなっており、それでもよい。ただし、中国では「系」は体系 (system) の意味で使用するので、『玉篇』系字書のように表現する。顧野王『玉篇』と宋本『玉篇』とは、同じ部首分類の体系を持つと判断されるので、『玉篇』系字書と呼ぶのである。『切韻』系韻書と言った場合にも、陸法言『切韻』の系譜の韻書という意味ではなく、陸法言が立てた韻の分類体系を基本とする韻書という意味であり、『広韻』まで含めて『切韻』系韻書と呼ぶのである。(pp.363-364)

上記の論文では、『篆隸万象名義』には触れられていないが、上の「同じ部首分類の体系を持つ」という規準に従えば、『篆隸万象名義』も玉篇系字書と呼ぶことができる。逆に言えば、古辞書に限定する場合、玉篇系字書は原本『玉篇』、『篆隸万象名義』、宋本『玉篇』の三書を指すと言える。

以下、本研究で利用する玉篇系字書三書の資料についての概要を簡単に述べる。

原本『玉篇』は中国南北朝時代、梁大同 9 年 (543) に成立した、典拠や用例が充実した部首別の漢字字書であり、後世の字書に大きな影響を及ぼした。原本『玉篇』の字数は、唐・封演の『封氏聞見記』の「梁朝顧野王撰玉篇三十卷、凡一萬六千九百一十七

<sup>4</sup> 石田ほか編『人文学のためのテキストデータ構築入門』2022 に改訂のうえ再録される草稿も検討させていただき機会を得た。石田ほか編『人文学のためのテキストデータ構築入門』2022：一般財団法人人文情報学研究所（監修）『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 年 7 月下旬に刊行予定。

字」という記述に従うならば、16,917字である。原本『玉篇』は中国では逸書となり、約八分の一の残巻が日本に存する。原本『玉篇』を受け継ぐものとして今に伝わっているのが、前述の日本側の『篆隸万象名義』と中国側の宋本『玉篇』である。両書の編纂の方針は類似しているが、相違点も少なくない。

平安初期（830以降）に弘法大師空海によって編纂された『篆隸万象名義』は現存する日本最古の漢字字書である。六帖構成をとり、第一帖から第四帖までは空海撰述であるが、第五帖と第六帖は別人の手によるものである。永久二年（1114）に書写された高山寺本が唯一の古写本である。明治三十二年（1899）に国宝（旧国宝）に指定され、昭和二十七（1952）年に文化財保護法により国宝に指定され、現在は国立京都博物館に寄託されている。他の転写本はこの伝本の系統を引くものであるとされる。内容は、約16,000字の掲出字に対して、字音・字義・字体の記述を収録したものであり、中国・梁の顧野王の原本『玉篇』から音注や義注を簡略化して示した字書である。原本『玉篇』は早く散逸し、日本の零本しか残っていない。それゆえ、完本である『篆隸万象名義』は原本『玉篇』の元の姿を窺う資料として価値が高い。

宋本『玉篇』は、原本『玉篇』を唐の孫強が高宗の上元元年（674）に修訂・増字したものを、宋の大中祥符六年（1013）、陳彭年等が勅命により重修したものである。注文内容は原本『玉篇』にあった長文の訓注を略し、掲出字を増補したもので、約22,000字の掲出字を収録する。

### 3 玉篇系字書の全文テキスト

筆者の所属していた北海道大学大学院文学研究院言語科学講座池田研究室では「平安時代漢字字書総合データベース（HDIC）」を構築するプロジェクトを推進していた<sup>5</sup>。

HDICプロジェクトの代表者の池田証壽は、日本古辞書の研究について、次のように主張している。

日本古辞書の研究では、その土台となった中国字書、特に玉篇系字書である原本『玉篇』、『篆隸万象名義』、宋本『玉篇』との対照作業が不可欠となっている。これまで

---

<sup>5</sup> 平安時代漢字字書総合データベース（Integrated Database of Hanzi Dictionaries in Early Japan, 略称HDIC）は、7万字を超える漢字の処理ができるUnicodeを用いて、平安時代を代表する部首分類体字書である『篆隸万象名義』『新撰字鏡』『類聚名義抄』を総合した全文テキストデータベースを整備・構築し、日本古辞書の研究基盤の確立を目指す。これらの漢字字書は、語彙・音韻・漢字字体等の日本語史研究の重要な資料である。これらの漢字字書に見える掲出字と注文とを電子テキスト化し、掲出字のリレーションシップを設定することによって、自在に検索できるシステムを構築し、その公開によって学術研究の発展に資することを意図してきた。https://hdic.jp

の古辞書データベース化では、玉篇系字書のデータベース化を疎かにしていた。HDICの全体は玉篇系字書データベースを基礎とした計画である。

HDICが対象とした主たる日本・中国の漢字字書、及びそのリレーションシップの設定を次の図1<sup>6</sup>で示す。黒い枠で示す三つの字書は玉篇系字書である。

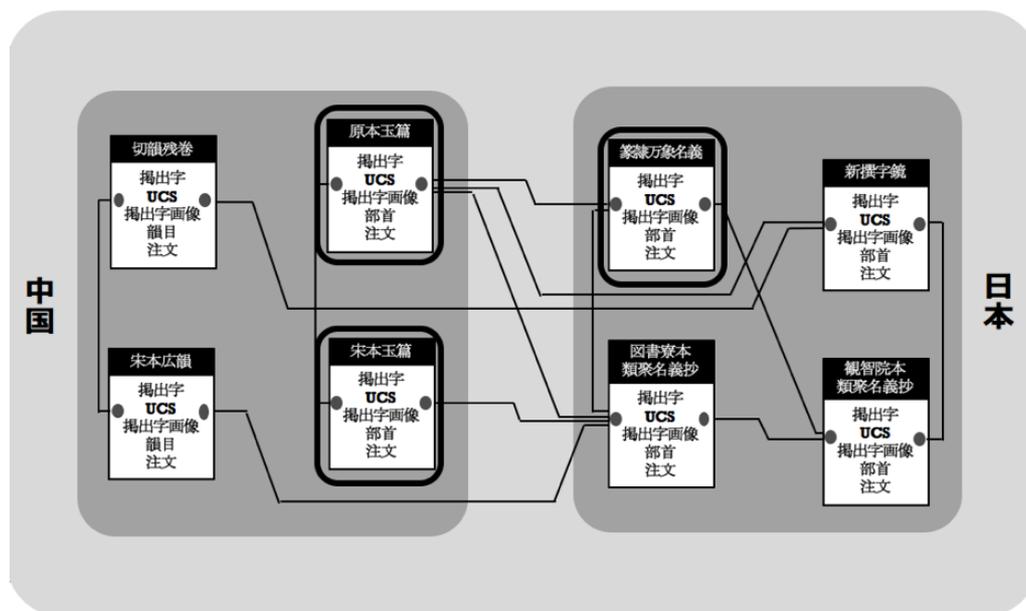


図1 HDICで主に対象とする漢字字書及びそのリレーションシップの設定

漢字字書の掲出字は、コンピュータで入力処理が可能な文字と、原本画像から一文字ごとに切り出した画像を対象としてデータベース化を進めた。字音・字義・字形を説明する注文の部分は、コンピュータで入力処理が可能な文字を入力した。

HDICは古辞書翻刻・入力作業の効率化をはかるため、まず古版本である宋本『玉篇』をデータ化し、これを土台に高山寺本『篆隸万象名義』の入力作業も完成した。2016年4月に宋本『玉篇』、9月に『篆隸万象名義』の全文テキストデータを順次に公開した<sup>7[10]</sup>。原本『玉篇』残巻の全文テキストも構築済みであるが、校正等の関係で内部利用に留まっている。

原本『玉篇』の残存する7巻、63部首について、データベースに基づき調査し、現存字数は2,087字である。この原本『玉篇』残巻にある2,087字に対応する玉篇系字書三書のプレーンテキストを構造化記述の基礎データとなる。

<sup>6</sup> これは池田証壽・李媛・申雄哲・賈智・斎木正直、「平安時代漢字字書のリレーションシップ」<sup>[9]</sup>（日本語学会2014年秋季大会，2014.10.18-19，北海道大学）ブース発表のポスターにあるイメージ図を転載したものである。作成は共著者の申雄哲による。

<sup>7</sup> <https://github.com/shikeda/HDIC>

#### 4 構造化記述における課題

上記の玉篇系字書三書のプレーンテキストを基礎にマークアップ作業を進める。その構造化記述における課題は次の4点が挙げられる。この4点の内容を TEI P5 に準拠したタグ付けを通じて研究資料を整理する必要がある。

- ・書誌学情報
- ・体裁・分巻・分部に関する構成情報
- ・項目構造の内容
- ・古辞書研究の内容

次節は、玉篇系字書三書の項目構造に関する構造化記述を詳しくみていく。

#### 5 項目構造に関する構造化記述

##### 5.1 玉篇系字書の項目構造

玉篇系字書三書は第3節で述べたとおり、異なる地域・時代に成立した同源の三つの漢字字書である。『篆隸万象名義』と宋本『玉篇』は、いずれも原本『玉篇』に依拠し、簡略化したものである。次の図2は「詭」の例で、三書の項目構造をみる。

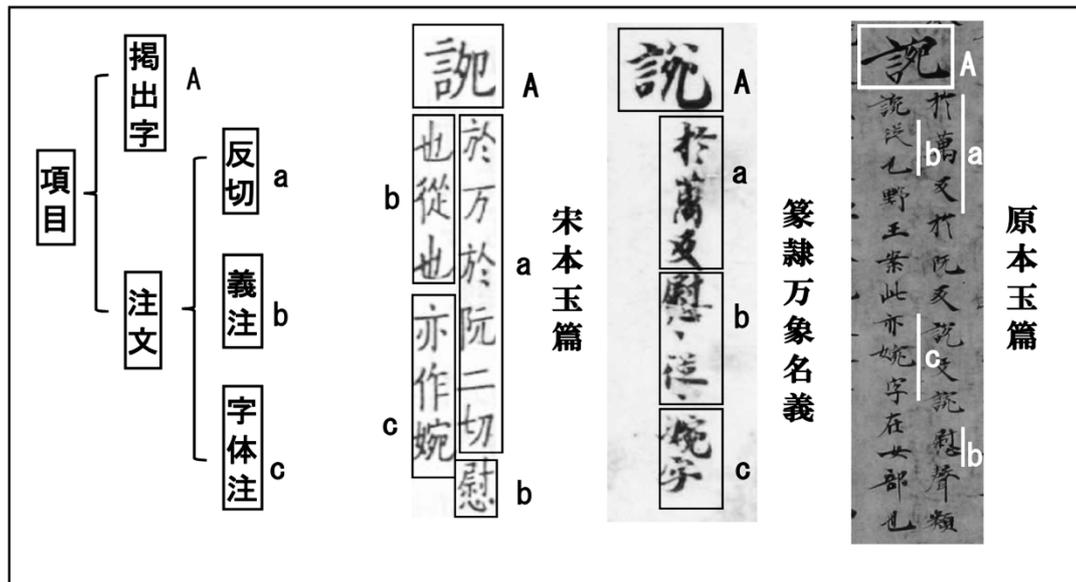


図2 「詭」原本玉篇・篆隸万象名義・宋本玉篇の項目構造

詭

於萬反。慰也、從也。婉字。(『篆隸万象名義』第3帖17丁裏)

於萬反、於阮反。『説文』：詭慰。『聲類』：詭從也。野王案：此亦婉字、在女部也。

(原本『玉篇』残卷 言部)

於万於阮二切。慰也、從也。亦作婉。(宋本『玉篇』上85丁表)

『篆隸万象名義』と宋本『玉篇』の項目構造は類似する。掲出字、反切、意味注記は、それぞれ形、音、義を解釈する役割を担う。字体注記は形、直音注は音の説明を補助する。原本『玉篇』は、『篆隸万象名義』と宋本『玉篇』より、さらに複雑な注文の構成となっている。

次の「記」(図3)を例に、その構造を示す。まず反切で字音を示し、次いで多くの儒家經典の原文と注釈書の解釈を載せる。さらに小学書による訓詁積義を引用し、著者顧野王の案語を付す。異体字注記や、部首の違う異体字を参照する注記を加えることもある。前述のように原本『玉篇』は出典を明記する字書であり、經典名・注釈家名・小学書名を詳細に記載する。

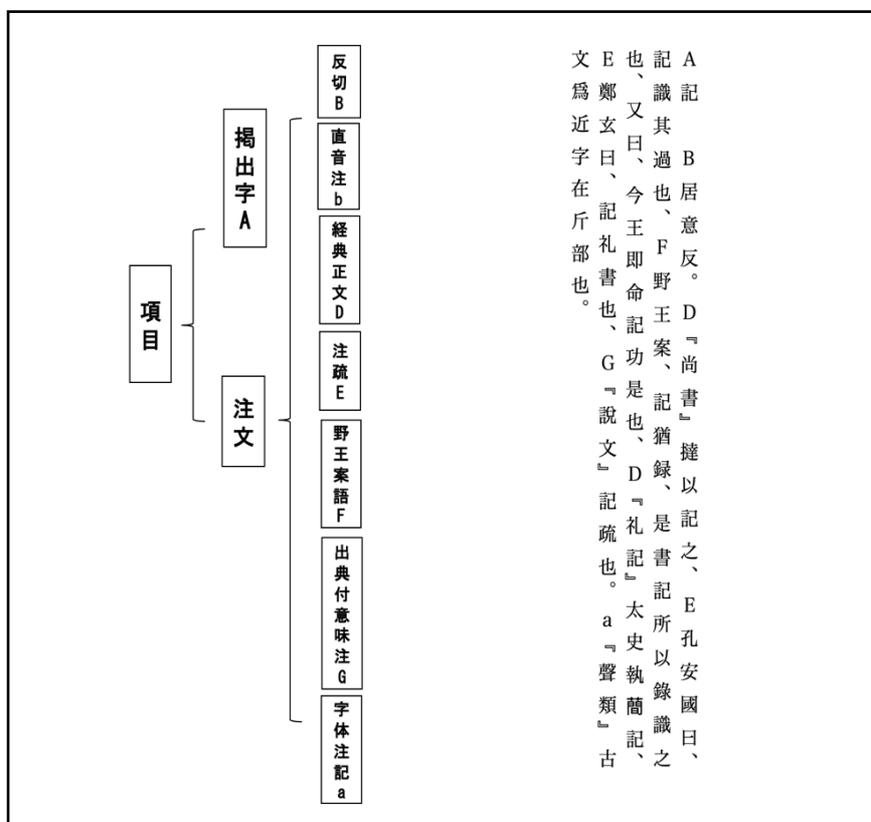


図3 「記」原本玉篇の項目構造

### 記

居意反。疏也、録也。(『篆隸万象名義』第3帖11丁裏)

居意反。『尚書』：撻以記之，孔安國曰：記識其過也、野王案：記猶録、是書記所以錄識之也、又曰：今王即命記功是也、『礼記』：太史執簡記、鄭玄曰：記礼書也、『說文』：記疏也。『聲類』：古文爲近字在斤部也。(原本『玉篇』殘卷 言部)

居意切。録也、識也。(宋本『玉篇』上83丁裏)

## 5.2 「詠」の TEI 文書

次に、5.1 で例にした玉篇系字書三書の「詠」について、TEI 文書を書いてみる。

「詠」原本『玉篇』

```
<?xml version="1.0" encoding="UTF-8"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>玉篇</title>
        <editor>顧野王</editor>
        <respStmt>
          <resp>TEI Encoding</resp>
          <persName>Yuan LI</persName>
        </respStmt>
      </titleStmt>
      <publicationStmt>
        <p>[Omit]</p>
      </publicationStmt>
      <sourceDesc>
        <bibl>
          <title>玉篇</title>
          <editor>顧野王</editor>
          <orgName>古典籍総合データベース (早稲田大学図書館)</orgName>
          <note>
            <lb/>請求記号 Call No. ホ 04 02555
            <lb/>国宝
            <lb/>26.9×1625.9cm(外寸 27.7×1728.4cm)
            <lb/>紙背:金剛界私記(治安元年写)
            <lb/>昭和 29 年修補
            <lb/>朱・墨書入あり
            <lb/>欠落あり 欠損あり
            <lb/>卷子装
            <lb/>銭屋総四郎,磯淳,川田剛,田中光顕旧蔵
            <lb/>言部-幸部
          </note>
        </bibl>
      </sourceDesc>
    </fileDesc>
  </teiHeader>
</TEI>
```

```
</bibl>
</sourceDesc>
</fileDesc>
</teiHeader>
<text>
  <body>
    <milestone unit="volume" n="9"/>
    <entry xml:id="Y09a-12-13-1">
      <form><orth>詵</orth></form>
      <form><pron type="fanqie">於萬反、</pron></form>
      <form><pron type="fanqie">於阮反。</pron></form>
      <sense>
        <cit>
          <bibl>
            <name>《說文》</name></bibl>
            <def>詵、慰。</def>
          </cit>
          <cit>
            <bibl>
              <name>《聲類》</name></bibl>
              <def>詵、從也。</def>
            </cit>
            <note type="variant">野王案:此亦<seg corresp="#1_072_B41">婉字
</seg>、在女部也。</note>
          </sense>
        </entry>
      </body>
    </text>
  </TEI>
```

「詭」『篆隸万象名義』

```
<?xml version="1.0" encoding="UTF-8"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>篆隸万象名義</title>
        <editor>東大寺沙門大僧都空海撰</editor>
        <respStmt>
          <resp>TEI Encoding</resp>
          <persName>Yuan LI</persName>
        </respStmt>
      </titleStmt>
      <publicationStmt>
        <p>[Omit] </p>
      </publicationStmt>
      <sourceDesc>
        <p>北海道大学大学院言語科学講座所蔵焼付写真(石塚晴通名誉教授)</p>
      </sourceDesc>
    </fileDesc>
    <encodingDesc>
      <charDecl>
        <glyph xml:id="u46f7-a">
          <desc>言八𠂔</desc>
          <figure>[1]
            <graphic url="https://glyphwiki.org/wiki/hdic_u46c4-var-001.png"/>
          </figure>
          <mapping type="standard">詭</mapping>
        </glyph>
      </charDecl>
    </encodingDesc>
  </teiHeader>
  <text>
    <body>
      <milestone unit="fascicle" n="3"/>
```

```

<div type="volume" n="G29">
<entry xml:id="T3_017_B12">
  <form><orth><g ref="#u46f7-a">詭</g></orth>
  </form>
  <hom xml:id="T3_017_B12_01">
    <form><pron type="fanqie"><choice>
      <sic>於</sic>
      <corr>於</corr>
    </choice>萬反。</pron></form>
    <sense><def>慰也、從也。</def>
      <usg type="variant">婉字也。</usg>
    </sense>
  </hom>
</entry>
</div>
</body>
</text>
</TEI>

```

「詭」宋本『玉篇』

```

<?xml version="1.0" encoding="UTF-8"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>大庑益会玉篇</title>
        <editor>顧野王</editor>
        <editor>陳彭年等</editor>
      <respStmt>
        <resp>TEI Encoding</resp>
        <persName>Yuan LI</persName>
      </respStmt>
    </titleStmt>
    <publicationStmt>
      <p>[Omit]</p>
    </publicationStmt>

```

```
<sourceDesc>
  <bibl>
    <title>大廣益會玉篇-古代字書輯刊</title>
    <editor>顧野王</editor>
    <publisher>中華書局</publisher>
  </bibl>
  <p>Information about the source</p>
</sourceDesc>
</fileDesc>
</teiHeader>
<text>
  <body>
    <milestone unit="volume" n="9"/>
    <entry xml:id="a085a053">
      <form><orth>
        <choice>
          <orig>諛</orig>
          <reg>詭</reg>
        </choice>
      </orth>
      </form>
      <hom xml:id="a085a053">
        <form><pron type="fanqie">於万於阮二切。</pron></form>
        <sense>慰也、從也。</sense>
        <dicscrap>
          <form type="variant"><note>亦作</note>
            <orth>婉。</orth></form>
          </dicscrap>
        </hom>
      </entry>
    </body>
  </text>
</TEI>
```

### 5.3 「記」の TEI 文書

次に、5.1 で例にした玉篇系字書三書の「記」について、TEI 文書を書いてみる (<teiHeader>の部分を省略する)。

「記」原本『玉篇』

```
<?xml version="1.0" encoding="UTF-8"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0">
<text>
  <body>
    <milestone unit="volume" n="9"/>
    <entry xml:id="Y09a-03-11-1">
      <form><orth>記</orth></form>
      <form><pron type="fanqie">居意反。</pron></form>
      <sense>
        <cit>
          <bibl>
            <title ref="https://viaf.org/viaf/181848476/">『尚書』:</title>
          </bibl>
          <def>撻以記之,</def>
          <bibl>
            <name>孔安國</name>曰:</bibl>
          <def>記識其過也,</def>
        </cit>
        <note type="meaning">野王案:記猶録,是書記所以録識之也,又曰:
        今王即命記功是也,</note>
        <cit>
          <bibl>
            <title>『礼記』:</title></bibl>
          <def>太史執蒿記,</def>
          <bibl>
            <name>鄭玄</name>曰:</bibl>
          <def>記礼書也,</def>
        </cit>
        <cit>
          <bibl>
            <name>『説文』:</name></bibl>
```

```

        <def>記疏也。</def>
    </cit>
    <cit>
        <bibl>
            <name>『聲類』:</name></bibl>
            <usg type="variant">古文爲近字,</usg> 在斤部也。
        </cit>
    </sense>
</entry>
</body>
<back>
</back>
</text>
</TEI>

```

「記」『篆隸万象名義』

```

<?xml version="1.0" encoding="UTF-8"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0">
    <text>
        <body>
            <milestone unit="fascicle" n="3"/>
            <div type="volume" n="G29">
                <entry xml:id="T3_011_B42">
                    <form><orth>記</orth>
                    </form>
                    <hom xml:id="T3_011_B42_01">
                        <form><pron type="fanqie">居意反。</pron></form>
                        <sense><def>疏也、録也。</def>
                        </sense>
                    </hom>
                </entry>
            </div>
        </body>
    </text>
</TEI>

```

## 「記」 宋本『玉篇』

```
<?xml version="1.0" encoding="UTF-8"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <text>
    <body>
      <milestone unit="volume" n="9"/>
      <entry xml:id="a083b062">
        <form><orth>記</orth></form>
        <hom xml:id="a083b062">
          <form><pron type="fanqie">居意切。</pron></form>
          <sense>居意切。</sense>
        </hom>
      </entry>
    </body>
  </text>
</TEI>
```

## 6 おわりに

以上、本稿は岡田「日本平安期古辞書の符号化モデル」にて提示された日本古辞書の符号化モデルを参照し、玉篇系字書（原本『玉篇』、『篆隸万象名義』、宋本『玉篇』）三書を対象に、その構造化記述に関する TEI マークアップを試みた。

『篆隸万象名義』と宋本『玉篇』は項目内容の構成要素が少なく、比較的単純な項目構造を持つが、原本『玉篇』の場合は、経典引用の例文や出典注記等が複雑になる。このような玉篇系字書に対して、共通の符号化が一定程度実現できることを確認した。なお、多音字注記、異体字記述等に関しては、さらに議論する余地がある。また、書誌学情報、体裁・分巻・分部に関する構成情報、古辞書研究の内容も含まれる、玉篇系字書の全体の構造化記述に関する TEI マークアップが今後の課題になる。

附記：本稿の一部は「TEI 研究会」（TEI-C 東アジア/日本語分科会、2022 年 6 月 30 日、Zoom）で取り上げられ、参加者の先生方からいろいろな視点でご意見をいただいた。本稿の準備にあたって岡田一祐氏（北海学園大学）からご意見を頂戴している。また、本研究を進めるうえで、守岡知彦先生（京都大学）、岡田一祐氏、申雄哲氏（檀国大学校）との議論に触発されるところが大きかった。

## 参考文献

- [1] 李媛 「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、no.5 (2018年5月)、pp.1-8.
- [2] 永崎研宣 「東アジア人文情報学の可能性についての試論：デジタルテキスト構造化の動向を中心として」、『KU-ORCAS が開くデジタル化時代の東アジア文化研究: オープン・プラットフォームで浮かび上がる、新たな東アジアの姿』、pp.243-256、2022年3月
- [3] TEI Consortium. TEI P5: Guidelines for Electronic Text Encoding and Interchange, Version 4.4.0, revision ff9cc28b0 (April 2022) , <https://tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf>
- [4] TEI P5 Guidelines 日本語版:<https://docsci.infon.org/stack/P5JA/JAja.html> (accessed 2022.7.12)
- [5] TEI ガイドライン日本語訳 (関西大学 Ku-orcas 東アジア DH ポータル) <https://www.dh.ku-orcas.kansai-u.ac.jp/?cat=9> (accessed 2022.7.12)
- [6] TEI P5 Guidelines 中国語版: <http://www.tei-c.org/Support/Learn/TEI-ChinLoc-2ndPrintEd.pdf> (accessed 2022.7.12)
- [7] 岡田一祐 「日本平安期古辞書の符号化モデル: TEI をもとにした符号化」、『デジタル・ヒューマニティーズ』vol.2、pp.26-54、2020年11月
- [8] 池田証壽 「平安時代漢字字書総合データベース構築の方法と課題：『類聚名義抄』を中心にして」、『漢字字体史研究』二、勉誠出版、pp.360-276、2016年11月
- [9] 池田 証壽, 李 媛, 申 雄哲, 賈 智, 斎木 正直 「平安時代漢字字書のリレーションシップ」、『日本語の研究』12巻2号、pp.68-75、2016年4月
- [10] 李媛, 池田証壽 「篆隸万象名義の全文テキストと公開システムについて」人文科学とコンピュータシンポジウム「じんもんこん2016」論文集、pp.95-102、2016年12月
- [11] 劉 冠偉, 中村 覚, 山田 太造 「部品と画数で漢字を検索するためのUnicode入力支援ツール」、『情報処理学会研究報告』2022-CH-128、no.2 (2022年2月)、pp.1-4.