



人文学のためのテキストデータ構造化のチュートリアル

第 11 章

TEI と辞書

——東アジア古辞書を例として——

李 媛
岡田一祐

version 1.0
2026.3.21 作成

本資料は、文部科学省委託事業「人文学・社会科学のDX化に向けた研究開発推進事業」(JPMXP1624)において、学校法人慶應義塾が、大学共同利用機関法人人間文化研究機構から再委託を受けて作成したものです。本資料の利用にあたっては、出典を必ず記載するなど、「文部科学省ウェブサイト利用規約」を準用（ただし、商用利用は不可とする。）してください。

1. はじめに

本章では TEI で辞書を符号化していく実際を東アジアの古辞書を事例としつつ見ていく。

TEI では、当初より辞書の符号化が課題であり、TEI P1 (1990) にはすでに辞書の章がある。むしろ、それらは欧米の辞書を符号化することから出発したのであって、東洋の辞書の伝統に即して作られたということはない。これを用いて東アジア古辞書を符号化するには、TEI の辞書符号化の考え方と東アジア古辞書の構造との双方の理解が欠かせない。本章ではその導入とすべく、TEI の辞書モジュールとそれをコンピューター処理用に整理した TEI Lex-0 (Tasovac et al. 2018) について説明し、東アジアの古辞書から実例をとって符号化の考え方について説明する。なお、本章での TEI P5 の説明は Version 4.10.2, revision bcfa98f42 (TEI Consortium 2025) にもとづき、TEI Lex-0 の説明は Version 0.9.4 (2024) にもとづく。

2. TEI における辞書の符号化

TEI においては、辞書は独立したモジュールが割り当てられており、重要な対象のひとつである。モジュールとは、符号化プロジェクトにおうじた符号化の道具立てのカスタマイズを許容すべく、用途ごとに要素などをまとめた加除可能な構成部品である (“iv. About these guidelines”)。辞書モジュールはそのなかで、語彙資源の符号化に関係する要素やクラスをまとめたもので、その解説はガイドラインの第 10 章「辞書」が宛てられている。

語彙とはなんらかのかたちで単語を集成したものであり、辞書モジュールはそのような語彙についてのあらゆる資源の符号化に資する要素などを準備している。したがって、辞書と名のつくものばかりではなく、単語リストや用語集などの符号化にも使えるし、コンピューター処理を容易にする道具立ても用意されている。

資料からどのような構造を読み取るかはさまざまな条件に左右される。語彙資源としてコンピューターでの可用性を重視するならば原資料の構成を分解してでも語彙についての情報を扱いやすく整理することも考えうるし、資料のありようを重視して、そのなかで情報の構造化に取り組むこともまた一案である。前者のたちばを推し進めたものに TEI Lex-0 があり、それを意識しながらも古辞書のありようを尊重して符号化を試みたものに岡田一祐 (2022) がある。

ここでは、より基本的な点から辞書の符号化について見ていきたい。まず、現代の漢和辞書 (『角川新字源』改訂新版、2017 年にもとづく物書堂版 ver 1.0.3) の「涼」を例にしてみよう (図 1)。

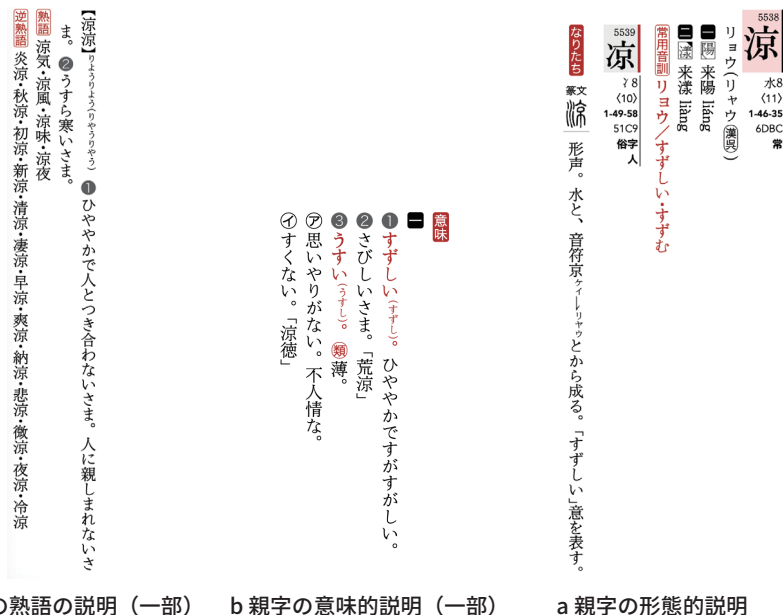


図1 現代漢和辞典の「涼」より抜粋 (『角川新字源』改訂新版、2017年にもとづく物書堂版 ver 1.0.3)

語彙資源は、一般に、見出し語について情報を記述するものである。『新字源』では、説明されるべき字(「親字」という、ここでは「涼」)をその番号(「検字番号」、ここでは「5538」とともに掲出し、その下に部首や画数などについての情報が示される(図1a。なお、箇々の情報については『新字源』の凡例や付録を参照のこと)。このような構成は、この辞書において、説明されるべきもの(「見出し(字・語)」)が部首と画数から探されるべきものであることによる。これは、現行のほとんどの漢和辞書で大筋において共通するものである。

それにつづいてその字の日本における伝統的なかたち(「字音」、ここでは漢呉音ともに「リョウ(リャウ)」、と中国におけるかたち(中古音¹⁾の「韻目」・「声母・韻母」と「現代中国音」。発音は意味によって異なり、韻目はそれぞれ「陽(平声)」と「漾(去声)」²⁾、「声母・韻母」は、それぞれ「来陽」および「来漾」、普通話で代表される「現代中国音」はここでは「liáng」と「liàng」によって発音が示され、あわせて常用漢字の音訓が附記される。

さらに、旧字・異体字欄では「涼」という俗字形があることが示され、親字同様に部首や画数などの情報が示される。『新字源』では、本字・古字・別体・俗字・誤字と異体字の分類を行っている。

最後に、「なりたち」の標識以下に古代文字欄となりたち欄とが置かれ、それぞれ楷書以前の古代文字の提示と漢字の構造の説明とがされている。以上が見出しの特定や異形にかんする情報で、ここでは形態的説明と呼んでおく。

そのあと、意味欄として見出しの使われ方についての情報が与えられる。ここでは意味的説明と呼ぶ(図1b)。意味は発音の別と対応して記述されている。図2のように、参考情報を示す参

考欄が意味欄のあとに設けられることがある。さらに、熟語欄が設けられ (図 1c)、熟合したばあいの意味や用法が示されている。

① 片仮名の「三」は三による。
② 証書・証券類では三のかわりに参(參)を用いることがある。

図 2 『角川新字源』改訂新版「三」より抜粋 (『角川新字源』改訂新版、2017 年にもとづく物書堂版 ver 1.0.3)

箇々の情報の提示手法には箇々の辞書の工夫があるが、現代の漢和辞書においてこのような 3 部構成は非常に一般的である。このような情報の符号化をどのように TEI で考えているか、以下では代表的なアプローチを紹介する。

2-1. TEI 辞書モジュール

TEI の辞書モジュールは、先述のとおり、紙に書かれてきた伝統的辞書の符号化が焦点であるが、同時に、語彙資源は計算機においても重要な問題であったため、電子的な資源についても考慮がされてきた。後者の側面については次節でとくに触れることになるので、ここでは最低限に留める。なお、辞書に特化した内容であるとは、辞書の語彙についての記述に特化しているということでもある。すなわち、ほかの文献資料にもあるような標題紙などの符号化は共通のモジュールにまかされており、辞書モジュールの範疇にはないということはいちおう注意しておきたい。

TEI の辞書モジュールの特徴は、項目を扱ううえで、その構造内の規則性を重視した <entry> 要素と、それを前提としない <entryFree> 要素に分けて考えることである。そのうえで、上位項目を扱う <superEntry>、同綴り語を扱う <hom> や関連語を扱う <re> などの要素が用意されている (<superEntry> と <re> は廃止予定)。

<entry> 要素は、かぎられた上位要素を立て、すべての情報はそのなかに整理されるべきだとする。具体的には、<form> (形態)、<gramGrp> (文法的情報)、<sense> (意味) などの要素と、<hom> などの内部項目にかんする要素である。<entry> 要素で <form> に入るべきとされる <orth> (綴り)、<pron> (発音) などの要素も、<entryFree> 要素ではまともずに直接用いることになる。

<entry> の構造についての理解はより専門的な資料に譲り、図1の文字をきょくりよく残して <entryFree> によって符号化すればおおよそ次のようになるだろう（篆文の再現は省く）。

```
<entryFree xml:id="Shinjigen_5538">
  <form>
    <orth>涼</orth>
    <idno type=" 検字番号 ">5538</idno>
    <lbl type=" 部首 ">水 8</lbl>
    <lbl type=" 総画数 ">〈11〉</lbl>
    <idno type="JIS">1-46-35</idno>
    <idno type="Unicode">62C5</idno>
    <lbl type=" 漢字種別 ">常</lbl>
    <lb/><pron notation=" 字音 ">リョウ</pron> (<pron notation=" 字音仮名づかい
">リャウ</pron><lbl type=" 字音種別 ">漢呉</lbl>)
    <lb/><hi rend=" 白抜き ">一</hi><pron type=" 韻目 ">陽<metamark function="
声調 " ana=" 平 "/></pron><pron type=" 韻母声母 ">来陽</pron><pron type=" 現代中
国音 ">liáng</pron>
    <lb/><hi rend=" 白抜き ">二</hi><pron type=" 韻目 ">漾<metamark function="
声調 " ana=" 去 "/></pron><pron type=" 韻母声母 ">来漾</pron><pron type=" 現代中
国音 ">liàng</pron>
    <lb/><label>常用音訓</label><pron type=" 常用音訓 ">リョウ</pron> / <pron
type=" 常用音訓 "><emph>すず</emph> しい</pron> ・ <pron type=" 常用音訓 "><emph>
すず</emph> む</pron>
  </form>
  <form>
    <lb/><orth type=" 異体字 " xml:id="e_5539">涼</orth>
    <idno type=" 検字番号 ">5539</idno>
    <lbl type=" 部首 ">冫 8</lbl>
    <lbl type=" 総画数 ">〈10〉</lbl>
    <idno type="JIS">1-49-58</idno>
    <idno type="Unicode">51C9</idno>
    <lbl type=" 異体字 ">俗字</lbl>
    <lbl type=" 漢字種別 ">人</lbl>
  </form>
  <lb/><label>なりたち</label><lbl>篆文</lbl> <etym> 形声。水と、音符京<pron
```

```

rend=" 字下げ ">ケイ→リヤウ </pron> とから成る。「すずしい」意を表す。 </etym>
<lb/><label> 意味 </label>
<lb/><sense><hi rend=" 白抜き ">一 </hi>
  <lb/><sense>① <def><hi rend="red"><emph>すず </emph>しい (すずし)。 </hi> ひ
  ややかですがすがしい。 </def></sense>
  <lb/><sense>② <def>さびしいさま </def>。 <quote>「荒涼」 </quote></sense>
  <lb/><sense>③ <def><hi rend="red"><emph>うす </emph>い (うすし)。 </hi></
  def><xr><lbl>類 </lbl><ref>薄 </ref>。 </xr>
  <lb/><sense>㊦ <def>思いやりがない。不人情な。 </def></sense>
  <lb/><sense>① <def>すくない。 </def><quote>「涼徳」 </quote></sense>
</sense>
<lb/><sense>④ <xr><lbl>→ </lbl><ref>涼陰 <hi rend=" 白抜き ">二 </hi><note>
りようあん </note></ref></xr></sense>
</sense>
<entryFree type=" 熟語 "><form>【<orth>涼涼 </orth>】 <pron rend=" 字下げ ">りよ
うりよう (りやうりやう) </pron></form><sense n="1">① <def>ひややかで人につき合わ
ないさま。人に親しまれないさま。 </def></sense><sense n="2">② <def>うすら寒いさま。
</def></sense></entryFree>
<lb/><label> 熟語 </label><xr><ref>涼気 </ref>・ <ref>涼風 </ref>・ <ref>涼味 </
ref>・ <ref>涼夜 </ref></xr>
<lb/><label> 逆熟語 </label><xr><ref>炎涼 </ref>・ <ref>秋涼 </ref>・ <ref>初涼
</ref>・ <ref>新涼 </ref>・ <ref>清涼 </ref>・ <ref>淒涼 </ref>・ <ref>早涼 </ref>・
<ref>爽涼 </ref>・ <ref>納涼 </ref>・ <ref>悲涼 </ref>・ <ref>微涼 </ref>・ <ref>夜
涼 </ref>・ <ref>冷涼 </ref></xr>
</entryFree>

```

字づらを残して、かつ、構造もあるていど取り出せるようになり、いったんはこれで完成と言える。

ただし、ここでは属性をもちいて情報を書き加えているが、総画数を示す「< >」の括弧は、それじたいに意味はなく、属性に与えられた情報によって不要となっている。原資料の字づらを再現するという目的からするとこれらは保存されるべきものであるが、語彙資源としての活用を考えると、処理のさいに括弧をいちいち除去せねばならず、むしろ阻碍要因となっているという考え方もある。また、漢字の種別の「常・人」なども、なんの略か分かりにくく不便ということも考えられる。そのような活用を促進する符号化方針がつぎに紹介する TEI Lex-0 プロジェクトである。

2-2. TEI Lex-0 プロジェクト

TEI Lex-0 プロジェクトは、TEI で符号化された語彙資源の計算的活用を推し進めるべく、TEI 辞書モジュールの非規則的要素を極力抑え、互換性の高い符号化の実践を模索するためのプロジェクトである。したがって、TEI Lex-0 プロジェクトはあくまで TEI 辞書モジュールの一活用であってそれを置き換えることを目指していないことを強調している。

すでにある語彙資源を TEI Lex-0 に沿って符号化することは、印刷のために高度に慣習化した辞書の字づらを意味的に分解することからはじまる。たとえば、図 1a の韻目「陽（平声）」の括弧内の情報は、版面上は「陽」の字の左隅を白く示すことで示されるのみであるが、その要素を判読して保存することも一案であるが、その意味を取って符号化することも一案であろうし、機械的には、後者のほうが使いやすい。これは TEI 辞書モジュールですでに指摘されていることで、オリジナルに忠実であることをよしとする立場、語彙情報の処理の容易性をよしとする立場のどちらも成り立つのである（両立させる道もありうるだろう）。このことは、TEI ガイドライン第 10.5 節でも論じられているのでくわしくはそちらを見られたい。

TEI Lex-0 プロジェクトでは <entry> を用いるため、<entryFree> のように情報をしいて符号化せずに残すようなことができないが、これによって処理上の例外が減るという利点が齎される。また、前述の例では語義説明の番号を文字で表しているが、これもこのままでは単なる説明なのか語義番号なのか分からない。そのため、文字面としては消してしまって、要素のほうに情報を持たせるといふ符号化が考えられる。また、「(すずし)・(うすし)」などの訓読のときのかたちを含む語義項目は、言い換え訳とも考えられ、たんなる語義 <def> では足りないとも考えられる。そうすると、熟語の例「荒涼」などとの区別が必要になり、<cit> 要素で引用情報を括り、そこに type 属性をあたえて種別を分けるという考え方が生まれる。それにしたがって整理したものが以下である。

```
<lb/><sense n="1.1"><cit type="translation" rend="red"><quote><emph>すず</emph>しい(すずし)。</quote></cit><def>ひややかですすがすがしい</def>.</sense>
<lb/><sense n="1.2"><def>さびしいさま</def>.<cit type="example"><quote>「荒涼」</quote></cit></sense>
<lb/><sense n="1.3"><cit type="translation" rend="red"><quote><emph>うす</emph>い(うすし)。</quote></cit><xr><lb1>類</lb1><ref>薄</ref>.</xr>
```

どれが正解ということではなく、用途に見合ったものを選ぶことが重要である。

3. 東アジア古辞書への応用

3-1. 東アジア古辞書の世界——中国・日本古辞書を例として

漢字は中国から日本へ伝来したものであり、日本古辞書の基礎もまた中国の辞書にある。中国の辞書は、漢字を構成する3要素である形・音・義のいずれに重きを置いて整理されているかによって分類することができる。すなわち、部首分類（字形）によって体系化された字書、韻分類（字音）に基づく韻書、そして意義分類（字義）によって整理された義書の3種類である。

後漢の許慎による『説文解字』は、漢字の本義を記録した最初の本格的な部首分類体の字書であり、9,353字を収録している。原本『玉篇』は、中国南北朝時代の梁に成立した部首別字書で、豊富な典拠や用例を備え、約16,000字を収録する（中国では早く散逸し、日本では約8分の1にあたる7巻の残巻が伝存）。隋の陸法言『切韻』や、宋の陳彭年らによる『広韻』は、韻書を代表する辞書である。また、前漢時代に成立した『爾雅』は、現存する最古の義書として知られる。

中国辞書史に関する研究は数多く存在するが、概説書としては、大島正二『辞書の発明：中国言語学史入門』（1997）、同著者『中国言語学史』（1997）、および『漢字と中国人：文化史をよみとく』（2003）が基本文献として広く参照されている。

日本古辞書研究についても多くの著作、論考があるが、概論としては、吉田金彦「辞書の歴史」（1971）や、山田健三「奈良・平安時代の辞書」（1995）が挙げられる。さらに近年の重要な成果として、大槻信『平安時代辞書論考：辞書と材料』（2019）、および池田証壽『日本辞書史研究：草創と形成』（2024）がある。

中国および日本の古辞書の分類については、山田俊雄『日本のことばと古辞書』（2003: 93-94）に示された次の整理が参考になる。

日本の辞書は、中国の辞書を土台にして、それに和訓を万葉仮名や片仮名で補ふという形で発生したから、日本語の辞書を考へるときは先づ中国の辞書についての種類などを考へておくことが必要である。それぞれの種類の下に生じたものを示しながら、次に概念図のやうな図式を示しておく。

(A) 字書

① 小学教科書——千字文・急就篇——（往来物）〔語彙連想〕

字形分類の字書——説文・玉篇——倭玉篇・名義抄〔部首別〕

② 韻による字書——韻書——切韻・広韻——聚分韻略・いろは韻〔四声別・反切〕

(B) 辞書

③ 意義分類体の字書・辞書（類書）——积名・爾雅——字類抄・節用集・いろは集〔いろは別・部門別〕

④ 仏教經典の巻末要語解 glossary（用語索引 concordance）——大般若經音義・一切經音義・

日本靈異記注〔語の総索引〕

- ⑤ 歌語辞書 語釈・用語集・かなづかひ書
 連歌俳諧辞書 作法書・方言辞書・諺集——物類称呼・俚言集覧
 物語その他の用語を加へて 和語辞書——雅言集覧・倭訓栞

本章において TEI による構造化記述の対象とする古辞書は、以上の分類における字形分類の字書に該当するものであり、具体的には、中国の『説文解字』、原本『玉篇』および宋本『玉篇』（『大広益会玉篇』とも、北宋陳彭年等重修、1013年）、日本の『篆隸万象名義』（空海撰、9世紀初頭）、『新撰字鏡』（昌住撰、9世紀末）、図書寮本『類聚名義抄』（法相宗の僧侶の撰、1100年前後）、観智院本『類聚名義抄』（真言宗の僧侶の撰、12世紀前半）を対象とする。

また、古辞書には、諸種の文献を忠実に引用する字書が存在する。これは「本文」を権威とする学問形態に関わるもので、本文主義³⁾とも呼ばれる。原本『玉篇』および図書寮本『類聚名義抄』は、文献名や典拠原文を詳細に引用する字書の典型である。一方で、文献名や典拠原文をほぼ省略し、漢字の字音・字義・和訓を中心に記すことで検索性を高めた字書も存在する。『篆隸万象名義』、宋本『玉篇』、『新撰字鏡』、および観智院本『類聚名義抄』は、その代表例である。

次節では、具体的な事例を示しながら、TEIによる古辞書の符号化の実際について述べる。

3-2. 東アジア古辞書の符号化

本節では、さきにも述べた「涼」を例として取り上げ、『説文解字』・原本『玉篇』・宋本『玉篇』・『篆隸万象名義』・『新撰字鏡』・図書寮本『類聚名義抄』・および観智院本『類聚名義抄』における各辞書項目に即して、TEIによる符号化例を示す。

古辞書の符号化においては、とくに掲出字について、写本に見える異体字を本文の形として尊重しつつ、現行字形との対応関係を明示することで利便性が増す。そのため、本節では、TEIの<orth>要素内に<choice>を用い、写本字形（例:<orig>涼</orig>）と正規化した字形（例:<reg>涼</reg>）とを併記して記述している。

先に述べた岡田一祐（2022）では、古辞書共通の構造化をより重視して<entry>による符号化を最終段階としているが、まずは古辞書項目の記述の多様性を重視して、岡田一祐（2022）で第一段階としている<entryFree>による符号化について検討する。また、これにくわえ、岡田一祐（2022）では行っていない辞書同士の項目の連携についても検討する⁴⁾。なお、これ以降の段階として、岡田一祐（2022）のように<entry>による高度な構造化を目指すかはさまざまな考慮条件がある。箇々の古辞書の構造の多様性、文献学的な慎重さを重視するならば、この段階をより精緻にするほうがよいという考えもじゅうぶんあり得るものである。


```

    <def> 薄也 </def>
  </sense>

  <etym>
    <seg type="semantic">
      从水
    </seg>
    <seg type="phonetic">
      京聲
    </seg>
  </etym>
</entryFree>

```

②原本『玉篇』

涼 力漿反。《毛詩》：“職涼善背”。《傳》曰：“涼、薄也”。野王案：《左氏傳》“號多涼德”是也。又音力將反。《韓詩》：“(孔)北風其涼，雨雪其雱。涼、寒貌也”。野王案：今謂薄寒為涼。《禮記》：“孟秋涼風至”、《楚辭》：“秋之為氣也、薄寒之中人”是也。《爾雅》：“北風謂之涼風”。《說文》：“為颯字。在風部。不善、悲涼為涼字、在无部。飲漿 名涼為醪字。在西部”。(卷19 水部)⁵⁾

原本『玉篇』は、先に述べた典拠内容を忠実に引用する本文主義を徹底しており、掲出字を示したのち、まず反切によって字音を明らかにし、その後、先行文献からの引用を通して字義や用法を提示する。本項でも複数の反切が示され、《毛詩》《左氏傳》《韓詩》などの引用が連続して配されている。これらの引用には用例的なものと解説的なものがあり、途中に「野王案」として編者顧野王自身の見解が挿入される。以下の符号化では、このような本文中心の記述構造と、掲出字・反切・引用・編者注という段階的な構成を区別して記述することを重視したため、ヘッダー内の<sourceDesc>に典拠の情報を整理した。

又曰其葉滑子傳曰滑盛白也又其葉滑之傳曰艾葉映疎不相此也說文曰浼也蟬蒼蒼色也廣雅滑盛也滑之流也

酒 也野之葉毛詩天不酒尔以酒是也韓詩飲酒閉門不出

容曰或為酒 湯昆及不雅感在中曰酒

字在晉部 **涇** 灘說文食邑而後吐之也

古言今之飲漿也野之葉毛詩不以其酒或以其

漿是也又曰漿人當乎飲水漿醴涼漿也 **冰** 漿字也

力漿及毛詩釅涼善於傷曰涼薄也野之葉左氏傳幹多涼德是也

也又音力將及韓詩孔北風其涼雨雪其垂涼寒白也野之葉今謂薄寒為涼孔北風秋涼風至楚辭秋之為氣也薄寒之中人是也

字惟北風謂之涼風說文為飄字在風部不善悲涼為燥字在

无部能漿名涼

為無字在酉部 **淡** 徒敢反記礼君子之道淡而不飲鄭全曰淡味似薄也野之葉說文淡薄味也故礼記君子淡以成人甘以壞老子道之公言淡乎其云味也而用之不可

孰是也史記故苦食淡如得日食名淡莊曰淡安恬之使無使字在

人 **澆** 公堯及楚辭波澆澆而揚澆王逸曰澆波為澆澆南澆天

部 下之澆許外音澆薄也說文澆波也野之葉澆澆

可以壞故澆料者也廣雅澆澆也又音五昂

及寒澆之子名澆左氏傳以寒澆澆是也

堯石及楚辭吸飛泉之微澆也 **澆** 公娥坦雅二反

懷琬琰之華英說文澆澆也 **澆** 說文多澆也

二及說文 **潘** 充甚反左氏傳猶拾潘之杜預曰

者豆汁也 **潘** 潘汁也野之葉礼記為榆潘是也

灑 餘莫反

主篇水部

テキスト画像：古代字書輯刊 原本玉篇 中華書局 1985

<sourceDesc>

<!-- 典拠：書名 -->

<ab>

<listBibl>

<bibl xml:id="Maoshi"><title>毛詩</title></bibl>

<bibl xml:id="Zhuan"><title>傳</title></bibl>

<bibl xml:id="Zuoshizhuan"><title>左氏傳</title></bibl>

<bibl xml:id="Hanshi"><title>韓詩</title></bibl>

<bibl xml:id="Liji"><title>礼記</title></bibl>

<bibl xml:id="Chuci"><title>楚辭</title></bibl>

<bibl xml:id="Erya"><title>爾雅</title></bibl>

<bibl xml:id="Shuowen"><title>說文</title></bibl>

</listBibl>

<!-- 典拠：人名 -->

<listPerson>

<person xml:id="GuYeWang">

<persName>顧野王</persName>

```

        <note> 编者（顧野王）が自身の見解・注記を述べる時に現れる。
</note>

        </person>
    </listPerson>
</ab>
</sourceDesc>

[...]

<entryFree xml:id="Y19300124-2">
    <form type="lemma">
        <orth>
            <choice>
                <orig> 涼 </orig>
                <reg> 涼 </reg>
            </choice>
        </orth>
    </form>

    <!-- 音注：力漿反 -->
    <form type="pronunciation" xml:id="Y19300124-2_pron1">
        <pron type="fanqie"> 力漿反 </pron>
    </form>

    <!-- 第一義：薄 -->
    <sense n="1">
        <cit type="quotation">
            <bibl><title ref="#Maoshi">《毛詩》</title></bibl>
            <quote>「職涼善背」</quote>
        </cit>
        <cit type="commentary">
            <bibl><title ref="#Zhuan">《傳》</title> 曰 </bibl>
            <quote>「涼、薄也。」</quote>
        </cit>
        <note type="commentary" resp="#GuYeWang">

```

野王案：《左氏傳》“號多涼德”是也。

</note>

</sense>

<!-- 又音：力將反 -->

<form type="pronunciation" xml:id="Y19300124-2_pron2">

又音

<pron>力將反</pron>

</form>

<!-- 第二義：寒貌 -->

<sense n="2">

<cit type="quotation">

<bibl><title ref="#Hanshi">《韓詩》</title></bibl>

<quote>「北風其涼，雨雪其雱。」</quote>

</cit>

<def>涼、寒貌也。</def>

<note type="commentary" resp="#GuYewang">

野王案：今謂薄寒為涼。

</note>

<cit type="quotation">

<bibl><title ref="#Liji">《禮記》</title></bibl>

<quote>「孟秋涼風至」</quote>

</cit>

<cit type="quotation">

<bibl><title ref="#Chuci">《楚辭》</title></bibl>

<quote>「秋之為氣也、薄寒之中人」</quote>

</cit>

<cit type="quotation">

<bibl><title ref="#Erya">《爾雅》</title></bibl>

<quote>「北風謂之涼風」</quote>

</cit>

</sense>


```

<entryFree xml:id="TB_5_098_A31">
  <form type="lemma">
    <orth>
      <choice>
        <orig>涼</orig>
        <reg>涼</reg>
      </choice>
    </orth>
  </form>

  <form type=" pronunciation">
    <pron type="fanqie">力醬反</pron>
  </form>

  <sense n="1">
    <def>薄也</def>
  </sense>

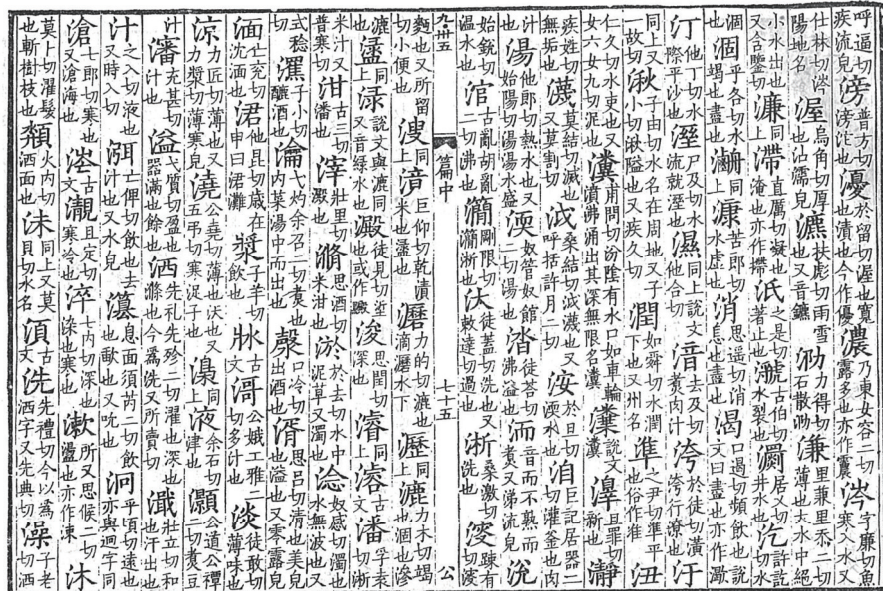
  <sense n="2">
    <def>冷也</def>
  </sense>
</entryFree>

```

④宋本『玉篇』

涼 力匠切。薄也。又力漿切。薄寒兒。(中篇 75 丁裏 水部)

宋本『玉篇』では、掲出字の後に反切と字義を対応させて示す、整理された記述形式が採られている。「力匠切」に対して「薄也」、「又力漿切」に対して「薄寒兒」と、異なる読みに応じて字義が簡潔に区別されている。原本『玉篇』に見られる多数の本文引用や「野王案」は省略され、情報は要約された形で提示されている。このように宋本『玉篇』は、本文主義的な原本『玉篇』の内容を整理、再編した字書であり、TEIによる符号化においても、反切と字義の対応関係を明示する構造として記述することができる。



テキスト画像：古代字書輯刊 大廣益會玉篇 中華書局 1987

```

<entryFree xml:id="SY_b075b061">
  <form type="lemma">
    <orth>涼</orth>
  </form>
  <!-- 音1：力匠切 → 義：薄也 -->
  <form type="pronunciation" xml:id="SY_b075b061_pron1">
    <pron type="fanqie">力匠切</pron>
  </form>
  <sense n="1">
    <def>薄也</def>
  </sense>
  <!-- 音2：又力漿切 → 義：薄寒兒 -->
  <form type="pronunciation" xml:id="Sb075b061_pron2">
    <pron type="fanqie">力漿切</pron>
  </form>
  <sense n="2">
    <def>薄寒兒</def>
  </sense>
</entryFree>

```

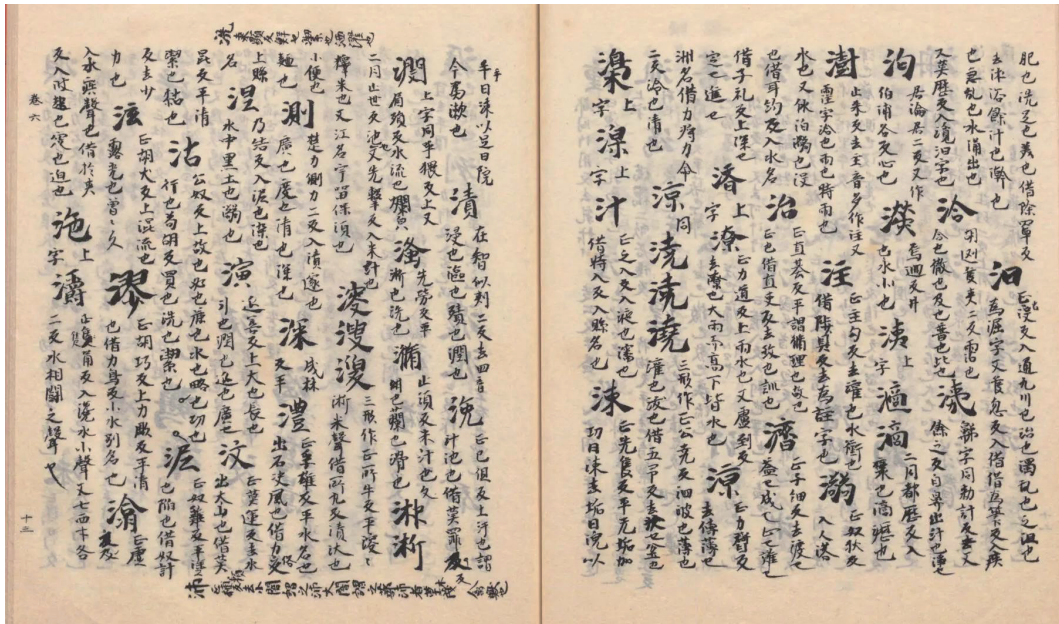
</sense>

</entryFree>

⑤天治本『新撰字鏡』

涼 正：力醬反。去：『傳』：薄也、洲名。借：力將力命二反。冷也、清也。（卷6 12 丁裏 水部）

天治本『新撰字鏡』では、掲出字を示したのち、字音に関する情報を「正」・「去」・「借」といった区分によって整理して示す。当該項目では、「正」によって基本的な読みを、「去」によって別の声調・用法を、「借」によって文脈依存の読みを区別して記している。各字音には反切が与えられ、それぞれに対応する字義や用法が簡潔に示されている。原本『玉篇』の本文主義的な記述に比べると、『新撰字鏡』は字音を中心に情報を再編した辞書であり、TEIによる符号化においても、正・去・借といった字音区分と、それに対応する反切・字義とを明示的に区別して記述した。



テキスト画像：国立国会図書館デジタルコレクション 新撰字鏡 六合館 大正 5年 info:ndljp/pid/3438197

<entryFree xml:id="SJ_s0612b603">
<form type="lemma">
<orth>
<choice>
<orig>涼</orig>

```
<reg> 涼 </reg>
</choice>
</orth>
</form>

<!-- 正：力醬反 -->

<lbl type="polyphonic_variation"> 正 </lbl>

<form type="reading">
  <pron type="fanqie"> 力醬反 </pron>
</form>

<!-- 去：『傳』：薄也、洲名 -->
<lbl type="polyphonic_variation"> 去 </lbl>

<cit type="source">
  <bibl>
    <title> 傳 </title>
  </bibl>
</cit>
<sense n="1">
  <def> 薄也 </def>
</sense>

<sense n="2">
  <def> 洲名 </def>
</sense>

<!-- 借：力將・力命二反。冷也、清也 -->
<lbl type="polyphonic_variation"> 借 </lbl>

<form type="pronunciation">
  <pron type="fanqie"> 力將反 </pron>
  <pron type="fanqie"> 力命反 </pron>
```

```

</form>

<sense n="3">
  <def> 冷也 </def>
</sense>

<sense n="4">
  <def> 清也 </def>
</sense>
</entryFree>

```

⑥ 図書寮本『類聚名義抄』

清涼 音良(L)真云薄冷也。玉云力醬反薄也。又力将反寒也北風謂之一(涼)風。タスク(LLH)佐也 ー (清涼) ホシサカサシム(LHHH^ˆH○○)(法上 水部)⁶⁾

図書寮本『類聚名義抄』は、単字のみならず熟語を掲出語として収録している点に特徴がある。たとえば「清涼」を見出しとして掲げ、その後に字音を示し、さらに「真云」「玉云」として『大般若経音訓』や『玉篇』(項目によっては『篆隸万象名義』)などの先行字書の説明を引いて字義を整理している(出典略注については池田証壽 2000 を参照)。

これらの引用は、字義説明や用例の補足として機能しており、複数の読みや意味が並置されている。また、項目末には仮名による和訓や、その声調を示す声点注記が付されており、日本語による理解を強く意識した構成となっている。TEI による符号化にあたっては、熟語見出し、字音注記、先行書からの引用、和訓といった異なる情報層を区別して記述することを重視した。声点は現行の TEI では入力が高いため、訓点語学で一般的な記法で代替した。


```

        <orig> 𤇗 涼 </orig>
        <reg> 清涼 </reg>
    </choice>
</orth>
</form>

<!-- 音注：音良 (L) -->
<form type="reading">
    音 <pron type="on" ana="#tone_L"> 良 </pron>
    <note type="tone"> (L) </note>
</form>

<!-- 真云：薄冷也 -->
    <bibl>
        <title> 真 </title>
    </bibl>
    <sense n="1">
        <def> 薄冷也 </def>
    </sense>

<!-- 玉云：力醬反…又力將反…北風謂之涼風 -->
    <bibl>
        <title> 玉 </title>
    </bibl>

    <form type="pronunciation" xml:id="Z011740-2_pron1">
        <pron notation="fanqie"> 力醬反 </pron>
    </form>
    <sense n="2">
        <def> 薄也 </def>
    </sense>

    <form type="reading" xml:id="Z011740-2_pron2">
        <pron type="fanqie"> 力將反 </pron>

```

```

</form>
<sense n="3">
  <def> 寒也 </def>

  <cit type="quotation">
    <quote> 北風謂之涼風 </quote>
    <bibl>
      <title> 尔雅 </title>
    </bibl>
  </cit>
</sense>

<!-- 和訓・注記：タスク (LLH) 佐也 / 清涼ホシサカサシム (LHHH` H〇〇)
-->
<form type="gloss" xml:id="ruiju_meigisho_seiryo_kun1">
  <pron notation="kana"> タスク </pron>
  <note type="accent"> (LLH) </note>
  <note type="commentary"> 佐也 </note>
</form>

<form type="gloss" xml:id="ruiju_meigisho_seiryo_kun2">
  <pron type="kana"> ホシサカサシム </pron>
  <note type="accent"> (LHHH` H〇〇) </note>
  <note type="commentary"> 一一 (清涼) </note>
</form>
</entryFree>

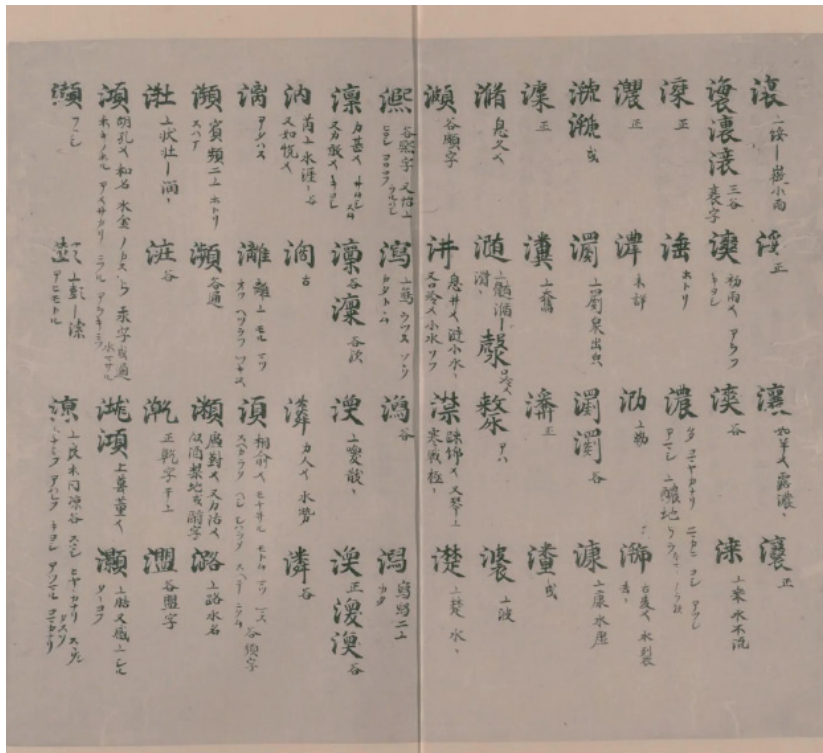
```

⑦ 観智院本『類聚名義抄』

涼 音良 (L) 禾同 涼谷 ス、シ ヒヤ、カナリ ス、ク-シ (H) カナシフ アハレフ キヨシ アツマル コマカナリ タスク (LLH) (法上 20 丁表 水部)

観智院本『類聚名義抄』では、単字を掲出語とし、その後に字音、字形・同系字の注記、字義、和訓を連続的に示す構成をとる。本項では、まず「音良」として字音が示され、続いて「禾同」「涼谷」などの字形・語義に関する注記が置かれている。その後、「ス、シ」「ヒヤ、カナリ」など複数の和訓が列挙され、語義の広がりや日本語によって詳細に補われている。ところどころ声調を

示す声点注記が付されている。図書寮本に比べると、観智院本は和訓情報を大幅に増補し、日本語による意味理解を強く志向した字書であることが分かる。TEI による符号化においては、字音注記、漢文による説明、和訓を層別して記述することを重んじた。声点の符号化は図書寮本と同様である。



テキスト画像：国立国会図書館デジタルコレクション 類聚名義抄 昭和 12
info.ndljp/pid/2586895

```
<entry xml:id="KR_K0707223">
  <form type="lemma">
    <orth>
      <choice>
        <orig>涼</orig>
        <reg>涼</reg>
      </choice>
    </orth>
  </form>
```

```
<!-- 音注：音良 (L) -->
<form type="reading">
  <pron type="on" ana="#tone_L"> 音良 </pron>
  <note type="tone"> (L) </note>
</form>

<!-- 字形注：禾同／涼谷（異体・字形関係として保持） -->
<note type="graph">
  <list>
    <item> 禾同 </item>
    <item> 涼谷 </item>
  </list>
</note>

<!-- 和訓・語釈（列挙） -->
<form type="gloss" xml:id="rm_kanchiin_liang_kun1">
  <pron type="kana"> ス、シ </pron>
</form>
<form type="gloss" xml:id="rm_kanchiin_liang_kun2">
  <pron type="kana"> ヒヤ、カナリ </pron>
</form>
<form type="gloss" xml:id="rm_kanchiin_liang_kun3">
  <pron type="kana"> ス、ク - シ </pron>
  <note type="accent"> (H) </note>
</form>
<form type="gloss" xml:id="rm_kanchiin_liang_kun4">
  <pron type="kana"> カナシフ </pron>
</form>
<form type="gloss" xml:id="rm_kanchiin_liang_kun5">
  <pron type="kana"> アハレフ </pron>
</form>
<form type="gloss" xml:id="rm_kanchiin_liang_kun6">
  <pron type="kana"> キヨシ </pron>
</form>
<form type="gloss" xml:id="rm_kanchiin_liang_kun7">
```

```
<pron type="kana">アツマル </pron>
</form>
<form type="gloss" xml:id="rm_kanchiin_liang_kun8">
  <pron type="kana">コマカナリ </pron>
</form>
<form type="gloss" xml:id="rm_kanchiin_liang_kun9">
  <pron type="kana">タスク </pron>
  <note type="accent"> (LLH) </note>
</form>
</entry>
```

4. 符号化の活用と限界

本節では、前節で示した符号化実践を踏まえ、TEIによる構造化記述が古辞書研究にもたらす意義と課題を総括する。符号化の利点とともに、資料の性格に由来する制約についても検討する。

4-1. 構造化による比較可能性の拡張

TEIによる符号化は、個々の古辞書の記述を忠実に保持しつつ、異なる時代・地域、系統の字書項目を横断的に比較するための共通基盤を提供する。本章で採用した<entryFree>による段階的符号化は、本文主義的な原本『玉篇』と、検索性を重視する日本古辞書とを、無理に同一構造へ還元することなく併置することを可能にした。このような構造化は、字音、字義、典拠、和訓の取捨や配置の差異を明示的に記述できる点で有効であり、古辞書研究における比較の視座を拡張するものといえる。

4-2. 符号化が示す編集思想の差異

一方、符号化の過程は、各辞書が前提とする編集方針や知的背景の違いを不可避的に顕在化させる。例えば、原本『玉篇』における詳細な引書や編者注と、宋本『玉篇』や『篆隸万象名義』に見られる簡潔な語釈とは、単なる情報量の差ではなく、辞書の用途や想定読者の相違を反映している。TEI符号化は、こうした差異を「統一」するための手段ではなく、むしろ差異そのものを構造として保持・提示するための枠組みとして機能しうる点である点を、本章の事例は示している。

4-3. 訓点、校訂差をめぐる限界と今後の課題

しかしながら、明示的な符号化には限界も存在する。たとえば、和訓、声調、訓点が重層的に記される資料や、異なる校訂本間の異同を横断的に扱う場合、単一の<entry>構造では十分に

対応できない場面が生じうる。どの粒度で情報を対応づけ、どこまでを同一項目として扱うかは、符号化規則だけでなく研究目的に依存する問題である。したがって、TEI 符号化は万能の解決策ではなく、その適用範囲と限界を自覚した上で運用する必要がある。この点は、今後の統合的古辞書データ構築に向けた重要な検討課題として残される。

附記

本稿は主要な部分の素稿を李が執筆し、岡田が補筆を行い、両者で全体を検討して成稿を見た。本章では、HDIC が公開しているテキストデータを活用した（HDIC 公開データ：<https://github.com/shikeda/HDIC>, 2025年11月）。

注

- 1 中国語中古音（4～11世紀の中国語音、とくに601年になった韻書『切韻』のそれ）は、声母と韻母のふたつの漢字の音を合成して発音を示す手法によって示されてきた。韻母におうじて韻目が整理される。ここではくわしく立ち入らない。
- 2 括弧内の平声、去声は中国語中古音の声調の種類で、漢字を取り囲む四角のどの隅に印があるかで読み解かれる。
- 3 本文権威に関わる学問形態については池田源太（1969）に詳しく、古辞書の文脈および本文主義については大槻信（2019）が重要な先行研究として挙げられる。
- 4 李媛 2025 で部分的に検討している。
- 5 原本『玉篇』残巻のデータは、校勘作業等の事情により現在は HDIC 内部でのみ公開しており、本研究では当該データに基づき、胡吉宣（1989）、臧克和（2008）、呂浩（2018）、姚永銘（2023）などの先行研究を参照した。
- 6 HDIC 公開データを参照。合わせて池田証壽（2000）、池田証壽（2005）を参照した。

参考文献

- ・池田源太 1969 「平安期に於ける「本文」を権威とする学問形態と有識故実」『延喜天曆時代の研究』吉川弘文館
- ・池田証壽（2000）「図書寮本類聚名義抄出典略注」『古辞書と JIS 漢字』3
- ・池田証壽（2005）『図書寮本類聚名義抄の翻字本文及び注解の作成に関する基礎的研究』平成 14～16 年度化学研究費補助金基盤研究 (C)(2) 研究成果報告書
- ・池田証壽 2024 『日本辞書史研究：草創と形成』汲古書院
- ・大島正二 1997 『辞書の発明：中国言語学史入門』三省堂
- ・大島正二 1997 『中国言語学史』汲古書院 増訂版 1998
- ・大島正二 2003 『漢字と中国人：文化史をよみとく』岩波新書 岩波書店
- ・大槻信 2019 『平安時代辞書論考：辞書と材料』吉川弘文館
- ・岡田一祐 2022 「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）
- ・山田健三 1995 「奈良・平安時代の辞書」西崎亨編『日本古辞書を学ぶ人のために』世界思想社
- ・山田俊雄 2003 『日本のことばと古辞書』三省堂

-
- ・ 吉田金彦 1971 「辞書の歴史」 阪倉篤義編 『講座国語史 3 語彙史』 大修館書店
 - ・ 李媛 2025 「TEI による篆隸万象名義と原本玉篇残巻の対応部分の項目構造記述：原本玉篇再構に向けて」 『じんもんこん 2025 論文集』
 - ・ TEI Consortium. eds. 2025. *Guidelines for Electronic Text Encoding and Interchange*. 4th September 2025. <http://www.tei-c.org/P5/>
 - ・ Toma Tasovac, Laurent Romary *et al.* 2018. TEI Lex-0: A baseline encoding for lexicographic data. Version 0.9.4. DARIAH Working Group on Lexical Resources. <https://dariah-eric.github.io/lexicalresources/pages/TEILex0/TEILex0.html>.
 - ・ 胡吉宣 (1989) 『玉篇校釋』 上海古籍出版社
 - ・ 呂浩 (2018) 『玉篇文獻考述』 上海人民出版
 - ・ 姚永銘 (2023) 『原本玉篇殘卷校証』 浙江古籍出版社
 - ・ 臧克和 (2008) 『中古漢字流變』 華東師範大学出版社