

LLM を用いた原本『玉篇』残巻の典拠情報整理 ―巻第九を中心に―

李 媛¹

1. はじめに

原本『玉篇』は、中国南北朝梁の顧野王によって編纂された字書であり、後世の漢字字書編纂に大きな影響を与えた。しかし、原本『玉篇』は中国では早くに散逸し、日本では約八分の一の残巻七巻が伝存している。これらの残巻には、『大広益会玉篇』（宋本『玉篇』とも称される）や『篆隸万象名義』では抜粋、改編されたと考えられる典拠情報が多く保存されており、原本『玉篇』の編纂方針や注釈体系を考察するうえで重要である。

筆者は李（2026）において、原本『玉篇』残巻巻第九の「食部」を対象とし、典拠書名および注釈家等人名の整理を行った。そこでは、食部 144 項目に見られる典拠情報を抽出し、典拠書名 39 種延べ 328 回、注釈家等人名 13 種延べ 97 回を確認した。さらに、これらの情報を TEI によって記述する方法を検討し、典拠書名を <listBibl>、注釈家等人名を <listPerson> に整理したうえで、本文中の引用、注釈、案語を構造化して表現するモデルを提示した。

しかし、食部は巻九の一部にすぎない。巻九全体には 27 部首、693 項目が含まれており、その典拠情報を網羅的に整理するためには、より大規模な抽出と確認の作業が必要となる。すべてを手で TEI 化することは、精度の面では望ましい一方で、作業量の面では大きな負担を伴う。そこで本稿では、前稿で作成した食部の人手 TEI データを規準データとして利用しつつ、LLM を用いた典拠情報抽出の可能性を検討する。なお、LLM の抽出結果はそのまま最終データとはせず、人手による確認・修正を組み込んだ Human-in-the-Loop の手法を採用する。

2. 先行研究

2.1 原本『玉篇』残巻の本文研究

胡吉宣『玉篇校釈』1989 は、原本『玉篇』残巻研究における重要な基礎的成果である。宋本『玉篇』を規準に原本『玉篇』残巻を取り込み、関連典籍と比較しながら校訂し、字形、音義、佚文の検討を行ったものであり、その後の研究に大きな影響を与えた。臧克和『中古漢字流變』2008 も、原本『玉篇』残巻を中古漢字研究の重要資料として位置づけ、『説文解字』や『篆隸万象名義』などとの比較を通じて、文字形体および字書本文の変遷を考察している。

また、呂浩『〈玉篇〉文献考述』2018 は、『玉篇』に関わる文献学的問題を総合的に整理した研究である。さらに、姚永銘『原本玉篇残巻校証』2023 は、残巻本文を精密に校訂し、関連文献との照合を通じて内容復元を試みた近年の代表的成果である。これらの本文研究は、主として個々の掲出字項目の校勘や、字形・音義に関する問題を中心とするものである。こうした研究の

¹ 京都大学人文科学研究所附属人文情報学創新センター

蓄積により、原本『玉篇』残巻の本文解説および校勘は大きく進展してきた。

2.2 原本『玉篇』の引書・典拠研究

原本『玉篇』残巻の大きな特徴の一つは、経書・史書・諸子・訓詁書・小学書など、多様な文献からの豊富な引用を含む点にある。これらの引書は、単に字義や用例を補うための材料にとどまらず、南北朝期における典籍受容のあり方や、当時の字書編纂に利用された文献の範囲を考察するうえで重要な手がかりとなる。

引書全体を扱った代表的研究として、胡吉宣「《玉篇》引書考異」(1982)が挙げられる。同論文は、原本『玉篇』に見られる引書を広く整理し、『周易』『毛詩』『尚書』『国語』『左伝』など約40種の典籍を対象として、引用文の文字異同や出典関係を考察したものである。この研究は、残巻に保存された典拠資料の文献学的価値を総合的に示した点で重要である。

その後は、特定の典籍に焦点を当てた個別研究が多く蓄積されてきた。朱葆華(1998)は原本『玉篇』残巻に見られる『説文』引用を諸本と比較し、その異同を整理した。蘇芑(2006)は『左氏伝』引用を通じて残巻の資料的価値を論じ、徐前師(2007)は唐写本『玉篇』に見える「字書」の性格と出所を考証した。郭萬青(2009)は『国語』引用例を校訂し、用字や注釈の継承関係を検討している。

さらに、蘇芑(2011)は『史記』引用と裴駰『史記集解』との関係を考察し、陳錦春・蓋翠傑(2014)は『詩経』引用の特徴と文献学的意義を論じた。蘇芑および胡智雄(2017)は、それぞれ『儀礼』『礼記』の引用例を通じて南朝期の経伝テキストの様相を検討した。蘇芑(2018)は『春秋公羊伝』引用からその伝本系統を考察し、李慶彬(2019)と劉杰陽(2021)は『尚書』引用の異文と伝本性格を分析している。また、劉新楠(2021)は、複数の典籍引用に見られる字詞異文の資料的価値を指摘した。

以上の研究は、原本『玉篇』残巻に含まれる引書の文献学的価値を具体的に示し、個別典籍の伝本研究に大きく寄与してきた。一方で、これらの研究は特定の典籍や引用例を対象とするものが多く、巻ごと、あるいは部首ごとに、典拠文献の出現頻度や分布を計量的に比較する研究は、なお十分とはいえない。とりわけ、同一巻内の複数部首を比較し、各部首にどの典拠が集中するのか、また小学書、経書、史書、諸子等がどのように分布するのかを把握することは、従来の個別引書研究とは異なる視点を提供する。本稿は李(2026)を踏まえ、巻九全体を対象として典拠情報を抽出し、その分布と特徴を統計的に分析するものである。

3. 研究方法

3.1 TEI による構造化

近年、古典籍研究においては、本文を単に電子テキスト化するだけでなく、文献構造・注釈構造・引用関係を機械可読な形で記述することの重要性が高まっている。その代表的な枠組みの一つが TEI である。TEI を用いることにより、書誌情報、本文構造、異文、注釈、引用、人物名、典拠文献などを体系的に記述することが可能となる。

李 (2026) において、原本『玉篇』残巻卷九「食部」を対象とし、典拠情報を TEI によって記述する方法を検討した。具体的には、典拠書名を <listBibl> に、注釈家等人名を <listPerson> に整理し、本文中の引用を<cit type="quotation">、注釈を<cit type="commentary">、顧野王自身の案語を <note type="editor">として区別するモデルを提示した。この方法により、原本『玉篇』残巻に見られる複雑な引用、注釈構造を、再利用可能かつ検索可能な形で表現できることを示した。

3.2 LLM を用いた典拠情報抽出

最近、LLM は自然言語処理のさまざまな領域で利用され、古典籍や歴史資料の整理においても、固有名抽出、分類、要約、表記揺れの整理、注釈支援などへの応用可能性が注目されている。とくに、原本『玉篇』残巻のように、すでに電子テキスト化されているが、典拠情報や注釈構造が十分に構造化されていない資料に対しては、LLM を用いた半自動抽出が有効であると考えられる。

本研究において LLM に期待される役割は、書符号《》で明示された典拠書名を抽出することである。書符号を伴わないが文脈上典拠と判断される候補を発見することである。注釈家名や「野王案」のような案語を抽出し、典拠書名との関係を整理することである。説文/説文、礼記/禮記、尔雅/爾雅、埤倉/埤蒼のような表記揺れを正規化することである。

もっとも、LLM の出力をそのまま研究データとして採用することはできない。古辞書本文には、略称、異体字、残缺字、校勘注、引用と注文の連続など、判断の難しい要素が多く含まれる。たとえば、ある人名を注釈家として数えるべきか、ある書名が本当に典拠として機能しているのかといった問題は、文献学的判断を要する。したがって、LLM はあくまで候補抽出と初期整理を担い、その後には人手による確認と修正を行う必要がある。

この点で、人手による確認、修正を組み込んだ Human-in-the-Loop の手法を採用する。すなわち、まず LLM によって典拠情報を抽出し、ついで食部の人手 TEI データを規準として精度を評価し、差異を分析する。そのうえで、未標識典拠候補や表記揺れを手で確認し、確認済みデータとして第九巻全体の統計分析に用いる。この方法により、LLM の効率性と人手校訂の信頼性を組み合わせることができる。

3.3 食部 TEI データ

・TEI による典拠情報の構造化

食部の典拠情報を TEI によって記述するために、まず典拠書名と注釈家等人名を `teiHeader` 内に整理する方法を採用した。具体的には、典拠書名を <listBibl> にまとめ、それぞれの文獻に `xml:id` を付与した。また、注釈家等人名については <listPerson> を用い、鄭玄、郭璞、杜預、孔安國などの注釈家に加えて、編纂者である顧野王の案語を示す「野王案」も記した。

このように `teiHeader` に典拠情報を集約することにより、本文中では書名や人名を直接繰り返し記述するだけでなく、対応する ID を参照する形で統一的に扱うことが可能となる。たと

えば、本文中に現れる『説文』は #Shuowen、鄭玄は #ZhengXuan のように参照される。この方法は、表記揺れの整理、典拠情報の一貫性の確保、後続の検索、集計に有効である。

本文部分では、各掲出字項目を <entry> として記述し、その内部に字形、反切、語義、引用、注釈、案語を配置した。字頭は <orth>、反切は <pron> によって示し、語義ごとのまとまりは <sense> によって区分した。文献からの引用は <cit type="quotation">、注釈家による説明は <cit type="commentary"> とし、顧野王自身の案語は <note type="editor"> として区別した。

このような記述により、原本『玉篇』残巻に見られる複雑な本文構造を、機械可読かつ再利用可能な形で表現することができる。とくに、原典引用、注釈、編纂者案語を区別して記述できる点は、後の典拠情報分析にとって重要である。

・「食」項目の TEI 記述例

李 (2026) では、食部の代表例として「食」項目を取り上げ、TEI による具体的な記述例を示した。「食」項目には、『尚書』『論語』『左氏傳』『周禮』『世本』『禮記』『爾雅』『史記』など多くの文献が引用されており、孔安國、鄭玄、杜預らの注釈も含まれる。また、顧野王自身による案語も複数見られる。

この項目では、まず「是力反」という反切を <pron> に記述し、続いて複数の語義を <sense> に分けて整理した。それぞれの語義のなかに、文献引用を <cit type="quotation"> として置き、注釈家の説明を <cit type="commentary"> として記述した。さらに、「野王案」による説明は <note type="editor" resp="#GuYeWang"> として記述し、通常の注釈家による注文とは区別した。

また、「又音慈吏反」のような異読を伴う部分については、主項目とは別に下位項目として扱い、異なる音読とそれに対応する用例を区別して記述した。この処理により、同一字形に複数の音義が存在する場合でも、それぞれの引用・注釈関係を明確に表現することができる。

「食」項目は、原本『玉篇』残巻における典拠情報の複雑さをよく示す例である。複数文献からの引用、注釈家による解釈、顧野王自身の判断、異読・異義が一つの項目内に重層的に含まれており、TEI による階層的記述の有効性を確認することができる。

4. 原本玉篇残巻

4.1 残巻とそのデータベース

原本『玉篇』は中国で早く散逸し日本では約八分の一の残巻7巻が伝存している。原本『玉篇』残巻とその所蔵状況について、岡井 (1933) と池田 (2014) とを参照すると、詳細は次の通りである。

表 1 原本玉篇残巻とその所蔵状況

巻	八	九			十八	十九	二十二	二十四	二十七	
収録 部首	心部	言部～ 冊部	冊部～ 欠部	欠部～ 牽部	放部～ 方部	水部	山部～ 宀部	魚部	糸部 前半	糸部～ 索部
所蔵 機関	大東 急記 文庫	早稲田 大学図 書館蔵 ＊語部 不明	京都国立 博物館蔵 (福井崇 蘭館旧 蔵)	早稲田 大学図 書館蔵	藤田家 (東大寺 尊勝院旧 蔵)＊車 部の「軛 ～輯」・ 「軛～ 輦」不 明	藤田家 (東大寺 尊勝院旧 蔵)＊ 「冷～ 湛」安田 文庫旧 蔵、現蔵 未詳	神宮文庫 蔵	大福光 寺蔵	高山寺 蔵(糸 ～縵)	石山寺 蔵(經 ～縵)

筆者が所属した北海道大学文学研究院言語科学講座池田証壽研究室では、平安時代漢字字書を対象とする「平安時代漢字字書総合データベース (HDIC)」プロジェクトを推進している。プロジェクトの一環として、原本『玉篇』残巻の内容を収録したサブデータベースである「原本『玉篇』残巻データベース」が構築された。データベースの構築にあたっては、東方文化学院『東方文化叢書第六』所収の複製本を基礎資料とし、あわせて既刊の諸複製本およびウェブ上で公開されている写真画像を参照した。複製本および写真画像が確認できない箇所については、翻刻本文が存在する場合に限り、それを参照して補った²。本研究で用いるテキストは、HDIC によるものである。

4.2 玉篇残巻の巻・部首・字数

原本『玉篇』の残存する 7 巻、63 部首³について、データベースに基づき調査し、その存欠状況及び分巻の詳細を、次のようにまとめた (表 2)。

表 2 原本『玉篇』残巻の存欠の詳細⁴

名義分帖	三									
名義分卷	G25	G26	G27	G27	G28	G29				
玉篇分卷	八			九						
所属部首	心部			言部			語部	日部	乃部	ㄎ部
部首番号	88			91			92	93	94	95
残存状況	首尾欠			首欠			全存	全存	全存	全存
残存字数	7			312			6	11	5	4
名義分帖	三									
名義分卷	G29			G30						
玉篇分卷	九									
所属部首	可部	兮部	号部	于部	云部	音部	告部	ㄥ部	ㄩ部	品部
部首番号	96	97	98	99	100	101	102	103	104	105

² 池田証壽「平安時代漢字字書総合データベース：現状と課題 2014 夏」(『漢デジ 2014—デジタル翻刻の未来』、京都大学人文科学研究所附属東アジア人文情報学研究センター、2014)

³ 李媛「篆隸万象名義の本文データベースの構築 — 玉篇残巻対応部分を中心に —」(『国語国文研究』146、2015)

⁴ 李 (2026) 掲載の「表 2 原本『玉篇』残巻の存欠の詳細」には一部入力ミスがあったため、本表をもって訂正する。

残存状況	全存	全存	首欠	全存	全存	全存	全存	全存	全存	全存
残存字数	4	6	1	6	2	16	2	1	13	4
名義分帖	三									
名義分巻	G30							G31		
玉篇分巻	九									
所属部首	梟部	兪部	冊部	𠂔部	只部	尙部	欠部	食部	甘部	旨部
部首番号	106	107	108	109	110	111	112	113	114	115
残存状況	全存	全存	全存	全存	全存	全存	中欠	全存	中欠	中欠
残存字数	3	9	4	9	2	6	101	144	10	3
名義分帖	三		五							
名義分巻	G31		Z18							
玉篇分巻	九		十八							
所属部首	次部	𠂔部	放部	𠂔部	左部	工部	卜部	兆部	用部	爻部
部首番号	116	117	271	272	273	274	277	278	279	280
残存状況	中尾欠	中尾欠	全存	全存	全存	尾欠	首欠	全存	全存	全存
残存字数	5	4	3	12	3	3	8	2	7	3
名義分帖	五					六				
名義分巻	Z18				Z19	Z22				
玉篇分巻	十八				十九	二十二				
所属部首	𠂔部	車部	舟部	方部	水部	山部	𠂔部	鬼部	尸部	尸部
部首番号	281	282	283	284	285	343	344	345	346	347
残存状況	全存	中欠	中欠	全存	中尾欠	全存	全存	全存	全存	ほぼ全存
残存字数	4	102	28	4	145	143	2	2	10	96
名義分帖	六									
名義分巻	Z22									Z24
玉篇分巻	二十二									二十四
所属部首	厂部	高部	危部	石部	磬部	自部	阜部	𠂔部	𠂔部	魚部
部首番号	348	349	350	351	352	353	354	355	356	397
残存状況	全存	全存	全存	全存	全存	全存	全存	全存	全存	全存
残存字数	40	7	4	152	8	2	139	5	4	26
名義分帖	六							計 2,087 字		
名義分巻	Z27									
玉篇分巻	二十七									
所属部首	糸部	糸部	素部	絲部	𦃟部	率部	索部			
部首番号	425	426	427	428	429	430	431			
残存状況	全存	全存	全存	全存	全存	全存	全存			
残存字数	392	5	8	7	7	1	3			

5. LLM による典拠情報抽出

5.1 テキスト

本研究で用いるテキストは、前述した HDIC の原本『玉篇』残巻巻九の電子テキストである⁵。巻九全体のデータは TSV 形式で整理されており、各行が一つの字項目に対応する。データには、YYID (所在情報)、YY_vol_radical (部首番号)、Entry (掲出字)、Lv_page, Yao_page (諸家解説)、YY_def (項目注文)、YY_remarks (備考注記)、TBID, SYID (関連所在)、YY_img_filename (画像番号) などが含まれる。

巻九全体の TSV データには、27 部首、693 項目が収録されている。部首は、言部、誼部、日部、乃部、𠂔部、可部、兮部、号部、于部、云部、音部、告部、𠂔部、𠂔部、品部、梟部、兪部

⁵ <https://github.com/shikeda/HDIC/blob/master/YYP.tsv>

部、冊部、𠂔部、只部、肉部、欠部、食部、甘部、旨部、次部、𠂔部である。

入力データの各項目には、たとえば次のような情報が含まれる。

# YYID	Item ID. Begins with `Y`, followed by volume and location-based numbering as assigned in the working data.
# YY_vol_radical	Volume and radical number, e.g. `v08#088`, `v27#425`.
# Entry	Headword.
# Lv_page	Page or location reference used in the Lü Hao (Lv Hao) 呂浩 reference framework.
# Yao_page	Page or location reference used in the Yao Yongming 姚永銘 reference framework; available only for a small portion of the dataset.
# YY_def	Definition of pronunciation, meaning, variants, and citations.
# YY_remarks	Editor's notes and structural notes in the working data.
# TBID	Cross-reference ID to `KTb.tsv` when available.
# SYID	Cross-reference ID to `SYP.tsv` when available; `△` indicates no direct correspondence was assigned in the current data.
# YY_img_filename	Image filename corresponding to the entry in the current image set; this refers to an internal working image set, not publicly released.

このうち、本研究で典拠情報抽出の主対象としたのは#YY_defである。ただし、後続の確認作業や出典位置の追跡のため、#YYID、# YY_vol_radical、# Entry 等も抽出結果に保持した。

5.2 抽出対象

本研究で抽出対象とした典拠情報は、大きく三種類に分けられる。

- ・書名符号《》によって明示された典拠書名である。原本『玉篇』残巻巻九の電子テキストでは、多くの引用文献名が《説文》《廣雅》《方言》《毛詩》《尚書》のように書名符号で示されている。このような場合、書名符号内の文字列を典拠書名として抽出した。

- ・書名符号を伴わない未標識典拠候補である。本文中には、文献名と考えられる語が書名符号なしで現れる例がある。たとえば、食部の「食」項目に見える「世本、黃帝作火食」、あるいは「周易鼎象也」のような例である。これらは書名符号で明示されていないため、単純な形式的抽出では漏れるが、文脈上は典拠文献名として扱うべき可能性が高い。LLM によってこのような未標識典拠候補を抽出し、その後、人手で確認した。

- ・注釈家等人名である。本文中には、鄭玄、郭璞、孔安國、杜預、王逸、賈逵などの注釈家名が頻出する。また、「野王案」は顧野王自身による案語を示すものであり、一般の注釈家とは性格を異にするが、典拠情報の構造を把握するうえで重要である。そのため、本研究では注釈家名とともに「野王案」も抽出対象とした。

以上の抽出対象を整理すると、次のようになる。

種類	内容	例
明示的典拠書名	《》で示された書名	《説文》《廣雅》《毛詩》
未標識典拠候補	書名符号なしで現れる文献名	世本、周易、淮南
注釈家等人名	注釈家名・案語	鄭玄、郭璞、杜預、野王案

5.3 表記揺れの正規化

原本『玉篇』残巻の電子テキストには、書名や人名の表記揺れが存在する。たとえば、『説文』は「説文」と「說文」の両形で現れ、『禮記』は「礼記」と書かれる場合がある。また、『爾雅』には「爾雅」「尔雅」「□(尔)雅」のような表記が見られる。さらに、『埤蒼』と『埤倉』のように、異なる字形で現れるものもある。

このような表記揺れをそのまま集計すると、同一文献が別々の典拠として数えられてしまう。そこで本研究では、抽出時に正規化表を用意し、異なる表記を統一的な名称にまとめた。たとえば、次のような処理を行った。

原表記	正規化後
説文・說文	説文
礼記・禮記	禮記
周礼・周禮	周禮
儀礼・儀禮	儀禮
尔雅・爾雅・□(尔)雅	爾雅
埤倉・埤蒼	埤蒼
楚辞・楚辭	楚辭

人名についても同様に、表記揺れを一定の形に統一した。たとえば、「孔安国」は「孔安國」に統一し、「何休」「何休」は同一人物として扱った。また、「野王案」は顧野王による案語として、「顧野王」に対応させた。

5.4 抽出手順

抽出手順を図式化すると、次のようになる。

TSV データ

↓

条目単位に分割

↓

《》内典拠書名の抽出

↓

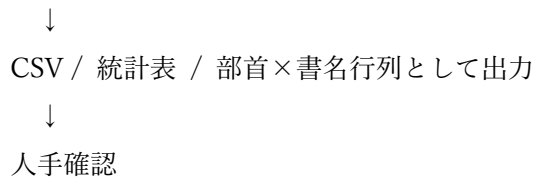
未標識典拠候補の抽出

↓

注釈家名・野王案の抽出

↓

表記揺れの正規化



5.5 Human-in-the-loop による確認

上述の LLM の抽出結果をそのまま最終結果とはせず、人手確認を組み込んだ Human-in-the-loop 型の方法を採用した。具体的には、まず食部について、LLM による抽出結果を前述の人手 TEI データと比較し、書名抽出、注釈家名抽出、未標識典拠抽出の精度を評価した。その結果を踏まえて、表記揺れの正規化表や未標識典拠候補の抽出条件を調整した。

次に、第九卷全体に対して LLM 抽出を行った。抽出された未標識典拠候補については、人手で確認し、典拠情報として採用すべきかどうかを判断した。その結果、第九卷全体で抽出された 13 件の未標識典拠候補は、すべて典拠情報として確認された。これにより、第九卷の典拠書名出現数は、《》内の明示的典拠 1819 件に、人手確認済みの未標識典拠 13 件を加え、合計 1832 件となった。

このように、本研究における LLM は、典拠情報を機械的に確定するための道具ではなく、人手確認を効率化するための補助的手段として位置づけられる。LLM によって候補を広く抽出し、人手によって確認、補正を行うことで、大規模な典拠情報整理と信頼性の両立を図った。

6. 食部を用いた検証と評価

本節では、食部の TEI データを規準として、LLM による典拠情報抽出結果を検証する。食部はすでに人手による典拠書名、注釈家等人名の整理が行われているため、LLM 抽出結果の精度を評価するための規準データとして利用できる。本稿では、まず食部 144 項目を対象として LLM による典拠情報の抽出を行い、その結果を前稿における人手による整理結果と比較した。

6.1 食部における LLM 抽出結果

LLM による抽出では、まず本文中に書名符号《》で示された典拠書名を抽出した。その結果、食部 144 項目から、《》内に明示された典拠書名として 322 件が抽出された。表記揺れを正規化したうえで整理すると、35 種の典拠書名が確認された。

また、書名符号を伴わないが文脈上典拠と判断される候補も抽出した。その結果、5 件の未標識典拠候補が得られた。具体的には、次のような例である。

掲出字	未標識典拠候補
食	世本
飪	周易
飭	淮南
餅	墨子
飭	大戴禮

これらはいずれも書名符号《》では示されていないが、文脈上は典拠文献名として機能していると判断できるものである。たとえば、「世本、黃帝作火食」や「周易鼎象也」のような箇所は、形式的には書名符号を欠くものの、内容上は明らかに文献引用として扱うべきである。

注釈家等人名については、102 件が抽出された。抽出された人名には、顧野王、鄭玄、郭璞、杜預、孔安國、王逸、何休などが含まれる。

6.2 TEI データとの比較

前述した食部の TEI データを規準データとして、LLM による典拠情報抽出結果を評価した。ここでいう規準データとは、食部 144 項目について人手で典拠書名および注釈家等人名を整理したものであり、典拠書名 39 種延べ 328 回、注釈家等人名 13 種延べ 97 回を含む。LLM 抽出結果との一致度を検討した。

評価にあたっては、二つの単位を区別する必要がある。一つは、典拠書名や人名の異なり数を対象とする type-level evaluation であり、もう一つは、本文中に実際に出現する各箇所を一回ずつ数える mention-level evaluation である。本稿で示す Precision、Recall、F1 は、主として mention-level、すなわち各出現箇所を単位として算出したものである。これは、典拠情報の整理においては、どの文献が存在するかだけでなく、それが何回、どの部首・どの項目に現れるかが重要であるためである。ただし、典拠書名の種類数についても、表記揺れの正規化後に別途確認した。

一致判定においては、原則として人手 TEI データにおける正規化後の典拠名、人名と一致する場合を正解とした。たとえば、「説文」と「説文」、「礼記」と「禮記」、「周礼」と「周禮」、「尔雅」と「爾雅」、「埤倉」と「埤蒼」は、表記揺れとして同一典拠に正規化した。一方、「毛詩」と「毛詩傳」は区別した。本文中に《傳》として現れる例は、卷九では文脈上『毛詩』引用の直後に現れ、『毛詩』の伝を指すものと判断できるため、本稿では『毛詩』に合算せず、「毛詩傳」として独立項目に立てた。

注釈家等人名については、人名そのものが抽出され、かつ人手規準において注釈家・案語として扱われている場合を正解とした。たとえば、「鄭玄曰」を「鄭玄」と抽出した場合は正解としたが、「鄭玄曰」全体を人名として抽出した場合や、直後の「曰」までを含めて境界を誤った場合は、境界誤りとして別に確認した。また、「野王案」は一般の注釈家名とは異なるが、顧野王自身の案語として独立して扱った。評価結果は次の通りである。

表 3 LLM 抽出結果の評価

評価対象	Precision	Recall	F1
書名：《》内の明示的典拠のみ	0.9783	0.9604	0.9692
書名：未標識典拠候補を含む	0.9786	0.9756	0.9771
注釈家等人名	0.9216	0.9691	0.9447

この結果から、書名符号《》で明示された典拠書名については、LLM が比較的安定して抽出できることが分かる。とくに注目されるのは、未標識典拠候補を加えた場合、Precision がほぼ

低下しない一方で、Recall が 0.9604 から 0.9756 へ向上した点である。これは、LLM が単に《》内の文字列を抽出するだけでなく、書名符号を伴わない文脈上の典拠候補を発見するうえでも有効であることを示している。

実際に、食部では「世本」「周易」「淮南」「墨子」「大戴禮」の 5 件が、書名符号を伴わない典拠候補として抽出された。たとえば、「世本、黃帝作火食」や「周易鼎象也」のような箇所は、形式的には《》を欠くため、書名符号のみに基づく抽出では漏れてしまう。しかし、文脈上は明らかに典拠名として機能している。このような例は、LLM が既存の形式的標識に依存するだけでなく、人手整理において見落とされやすい潜在的典拠を発見する補助となりうることを示している。

一方で、LLM 抽出結果には誤抽出・漏抽出も見られた。書名については、「字書」「説文」「方言」などで人手規準より多く数えられる例があり、本文中の校注・字体注・重複表現をどこまで典拠として数えるかという境界判断が問題となった。また、「韓詩」や「毛詩傳」のように、前後の引用関係を考慮しなければ正しく判断しにくい例もある。注釈家等人名についても、人名の直後に「曰」が現れる場合であっても、それが常に独立した注釈家典拠として数えられるとは限らず、抽出単位の判断に差が生じた。

以上の評価は、LLM の出力が人手規準データとどの程度一致するかを示すものである。ただし、この数値は規準データの作成方針に依存する。本稿で用いた規準データは、前稿において筆者が食部本文を対象に作成した TEI データに基づく。典拠書名・人名の表記揺れ、未標識典拠、注釈家名の扱いについては、文脈に即して判断した。

また、LLM 抽出の再現性を確保するためには、使用したモデル、入力単位、プロンプト、出力形式、temperature などの条件を明示することが重要である。本研究では、各字項目の本文を入力単位とし、典拠書名、未標識典拠候補、注釈家等人名を構造化して抽出する形式を採用した。今後は、同一条件での複数回実行や、異なるモデル間の比較を行い、抽出結果の安定性をさらに検証する必要がある。

したがって、本節の意義は、単に F1 値の高さを示す点にあるのではない。むしろ重要なのは、LLM がどのようなテキスト現象に対して有効であり、どのような文献学的判断において限界を持つかを明らかにする点にある。食部を用いた検証から、《》で明示された書名の抽出、未標識典拠候補の発見、表記揺れの正規化、注釈家等人名の確認という各段階で、LLM と人手確認を組み合わせる必要性が確認された。以上を踏まえ、次節ではこの方法を巻九 27 部首、693 項目へ適用した結果を述べる。

7. 第九巻全体へ

7.1 巻九の構成

原本『玉篇』残巻巻九には、部首番号 091 から 117 まで、すなわち言部から牽部までの 27 部首が含まれる。具体的には、091 言部、092 誥部、093 日部、094 乃部、095 亏部、096 可部、097 兮部、098 号部、099 于部、100 云部、101 音部、102 告部、103 口部、104 𠂔部、105 品部、

106 臬部、107 龠部、108 冊部、109 𪛗部、110 只部、111 尙部、112 欠部、113 食部、114 甘部、115 旨部、116 次部、117 牽部であり、項目数は全 693 項目に及ぶ。

このうち、言部、食部、欠部は項目数が多く、典拠情報も豊富に含まれる。一方で、凵部、号部、只部などのように、項目数の少ない部首も含まれている。そのため、巻九全体を対象とすることで、部首ごとの典拠密度や引用傾向の差異を比較することが可能となる。

7.2 巻九全体の抽出結果

巻九全体 693 項目に対して LLM による典拠情報抽出を行った結果、書名符号《》で明示された典拠書名は 1819 件抽出された。また、書名符号を伴わない未標識典拠候補として 13 件が抽出された。これら 13 件については人手で確認を行い、すべて典拠情報として採用した。

その結果、巻九全体における典拠書名の出現数は、明示的典拠 1819 件に人手確認済み未標識典拠 13 件を加え、合計 1832 件となった。正規化後の典拠書名の種類は 85 種である。また、注釈家等人名は 618 件、15 種が確認された。

表 4 玉篇残巻巻九典拠情報に関する抽出結果

項目	数量
条目数	693
部首数	27
《》内の典拠書名	1,819
人手確認済み未標識典拠	13
典拠書名総数	1,832
典拠書名種数	85
注釈家等人名	618
注釈家等人名種数	15

この結果から、巻九には非常に多くの典拠情報が含まれていることが分かる。単純に平均すると、一項目あたり約 2.64 件の典拠書名が現れる計算になる。ただし、実際には部首や項目によって典拠密度には大きな差があるため、単純平均だけでは巻九内部の特徴を十分に把握することはできない。

7.3 未標識典拠の人手確認

巻九全体への拡張において重要であったのは、未標識典拠候補の扱いである。LLM によって抽出された 13 件の未標識典拠候補は、すべて人手確認を経て典拠情報として採用した。人手確認済みの未標識典拠を整理すると、次のようになる。

典拠書名	件数
大戴禮	2
廣倉	2
淮南	2
左氏傳	1
世本	1

周易	1
墨子	1
莊子	1
國語	1
韓詩	1

これらの例は、書名符号《》を伴わないため、形式的な抽出だけでは見落とされやすい。しかし、前後の文脈を見ると、いずれも典拠文献名として機能していることが確認できる。たとえば、食部の「世本」や「周易」、欠部に見える「大戴禮」「廣倉」などは、本文の説明に直接関わる文献名として用いられている。

このように、LLM による未標識候補の抽出と人手確認を組み合わせることで、書名符号に依存した抽出では漏れる典拠情報を補うことができた。

7.4 部首別の典拠書名

卷九全体の典拠書名を部首別に集計すると、部首によって出現数に大きな差が見られる。もっとも多いのは言部であり、典拠書名出現数は 818 件であった。次いで、食部が 327 件、欠部が 229 件である。これら三部首だけで、卷九全体の典拠書名出現数の大部分を占めている。部首別の上位を示すと、次の通りである。

表4 部首別典拠情報出現数（上位 10 部首）

部首	典拠書名出現数
言部	818
食部	327
欠部	229
音部	50
𠂔部	49
于部	36
日部	33
𠂔部	30
甘部	28
誥部	24

この結果は、項目数の多い部首ほど典拠書名出現数も多くなる傾向を示している。言部は 312 項目を有し、卷九のなかで最大の部首であるため、典拠書名出現数も突出して多い。また、食部は 144 項目、欠部は 101 項目を有し、いずれも比較的大きな部首である。

しかし、典拠書名出現数は単に項目数に比例するだけではない。部首ごとの字義の性質や、引用される文献の種類によっても差が生じる。たとえば、食部では飲食・穀物・調理に関わる語義が多く、『方言』『廣雅』『字書』などの訓詁書、小学書が比較的多く利用される。一方、言部では言語行為や徳目、政教に関わる語が多く、経書、伝注類の引用も多く見られる。

7.5 注釈家等人名

巻九全体では、注釈家等人名として 618 件が抽出された。もっとも多く現れるのは「野王案」、すなわち顧野王自身の案語であり、219 件を数える。次いで鄭玄 126 件、郭璞 73 件、孔安國 53 件、杜預 50 件が多い。主な注釈家等人名の出現数は、次の通りである。

表 5 主な注釈家等人名の出現数

人名	出現数
顧野王	219
鄭玄	126
郭璞	73
孔安國	53
杜預	50
賈逵	27
王逸	19
何休	13
鄭衆	12
王弼	6
孔子	6
劉熙	6
孟康	4
馬融	3
蔡謨	1

この結果から、巻九の本文構成において、顧野王自身の案語が非常に重要な役割を果たしていることが分かる。また、鄭玄、孔安國、杜預、王逸、何休など、經書・伝注に関わる注釈家が多く現れることから、原本『玉篇』が単に字義を列挙するだけでなく、既存の注釈伝統を積極的に取り込んでいたことがうかがえる。

一方、郭璞の出現数が多いことは、『爾雅』注との関係を示すものと考えられる。また、賈逵、孟康、蔡謨などの名が見えることは、史書、伝注、古注類の利用状況を考えるうえで注目される。

8. 第九巻における典拠分布分析

8.1 高頻度典拠書名の分布

巻九全体において、もっとも多く出現する典拠書名は『説文』であり、386 回確認された。次いで、『字書』148 回、『廣雅』129 回、『毛詩』95 回、『禮記』85 回、『蒼頡篇』78 回、『聲類』77 回、『方言』71 回、『左氏傳』70 回、『尚書』67 回が続く。高頻度典拠書名の上位を示すと、次の通りである。

表 6 主な注釈家等人名の出現数

典拠書名	出現数
説文	386
字書	148

廣雅	129
毛詩	95
禮記	85
蒼頡篇	78
聲類	77
方言	71
左氏傳	70
尚書	67
傳	62
爾雅	57
埤蒼	55
周禮	51
國語	36
漢書	33
論語	33
楚辭	26
韓詩	20
周易	19

この結果から、卷九においては小学書、訓詁書の比重が非常に高いことが分かる。『説文』は突出して多く、卷九全体の字義説明において中心的な典拠として機能している。『字書』『廣雅』『蒼頡篇』『聲類』『方言』『爾雅』『埤蒼』なども高頻度で現れており、原本『玉篇』が既存の小学書・訓詁書を広く参照していたことがうかがえる。

一方で、『毛詩』『禮記』『左氏傳』『尚書』『周禮』『國語』『論語』『楚辭』『韓詩』『周易』など、經書、史書、諸子類も多く引用されている。これは、原本『玉篇』が字義を説明する際、単に小学書、訓詁書に依拠するだけでなく、經伝本文や古注における用例を重視していたことを示している。

8.2 注釈家等人名の分布

卷九全体では、注釈家等人名が 618 件確認された。もっとも多いのは「野王案」であり、顧野王自身の案語として 219 件確認された。次いで鄭玄 126 件、郭璞 73 件、孔安國 53 件、杜預 50 件が多い。主な注釈家等人名の出現数は、次の通りである（表 5 再掲）。

表 5 主な注釈家等人名の出現数

人名	出現数
顧野王	219
鄭玄	126
郭璞	73
孔安國	53
杜預	50
賈逵	27
王逸	19

何休	13
鄭衆	12
王弼	6
孔子	6
劉熙	6
孟康	4
馬融	3
蔡謨	1

この結果から、巻九の本文において顧野王自身の案語が大きな位置を占めていることが分かる。「野王案」は、単なる典拠引用ではなく、引用された文献や語義に対する編纂者自身の判断を示すものである。そのため、典拠書名とは異なる性質を持つが、原本『玉篇』の注釈体系を理解するうえで重要である。

鄭玄の出現数が多いことは、『周禮』『禮記』などの礼書引用と深く関わる。孔安國は主に『尚書』、杜預は『左氏傳』、郭璞は『爾雅』、王逸は『楚辭』と結びつく場合が多い。つまり、注釈家名の分布は、単なる人名の頻度ではなく、原本『玉篇』がどのような注釈伝統を媒介として典拠文献を利用していたかを示す指標となる。

また、賈逵、孟康、蔡謨などの人名が見えることは、『國語』『漢書』などの史書・古注資料との関係を考えるうえで注目される。これらの人名は出現数こそ多くないが、引用文献の伝本や古注の系統を検討する際に重要な手がかりとなる。

8.3 部首別の典拠密度

次に、部首別に典拠書名の出現数を確認する。巻九において典拠書名出現数をもっとも多いのは言部であり、818 件である。次いで、食部が 327 件、欠部が 229 件である。この三部首が巻九全体の典拠情報の大部分を占めている。部首別の典拠書名出現数の上位を示すと、次の通りである。

表 7 巻九部首・項目数・典拠書名出現数・平均典拠引用数（上位 10 部首）

部首	項目数	典拠書名出現数	平均典拠引用数
言部	312	818	2.62
食部	144	327	2.27
欠部	101	229	2.27
音部	16	50	3.13
𠂔部	13	49	3.77
于部	6	36	6.00
日部	11	33	3.00
𠂔部	9	30	3.33
甘部	10	28	2.80
諱部	6	24	4.00

この表から明らかなように、典拠書名出現数はおおむね項目数の多い部首に集中する。言部は

312 項目を有する卷九最大の部首であり、典拠書名出現数も 818 件と突出している。食部は 144 項目 327 件、欠部は 101 項目 229 件であり、いずれも項目数の多さに対応して典拠情報が多く現れている。

ただし、部首ごとの特徴を把握するには、総数だけでなく、一項目あたりの平均典拠引用数を見る必要がある。干部は 6 項目に対して 36 件、平均 6.00 と最も高く、言部も平均 4.00、𠂔部は 3.77、𠂔部は 3.33 を示す。これらは項目数こそ少ないが、各項目に複数の典拠が集中的に付されている部首である。

一方、言部は総数では最大であるものの、平均典拠引用数は 2.62 にとどまる。食部と欠部も平均 2.27 であり、出現数の多さは主として項目数に由来する。したがって、卷九の典拠分布を分析する際には、典拠書名出現数と平均典拠引用数を併せて見ることで、各部首の典拠密度や記述の性格をより適切に把握できる。

8.4 部首ごとの典拠傾向

部首ごとの典拠傾向を見ると、言部、食部、欠部にはそれぞれ異なる特徴が認められる。言部では、『説文』をはじめ、『廣雅』『毛詩』『禮記』『尚書』『左氏傳』『國語』『論語』など、多様な典拠が集中している。言部の字は、言語行為、政教、徳目、言説に関わるものが多いため、経書・史書・諸子類からの用例が豊富に利用されたと考えられる。また、字義説明の基礎として、『説文』や『廣雅』のような小学・訓詁系文献も多く参照されている。

食部では、『説文』『字書』『方言』『廣雅』『毛詩』『禮記』『周禮』などが目立つ。食部の字は、飲食、穀物、調理、供饌など生活文化に関わるものが多く、具体的な物名や動作を説明する必要がある。そのため、『方言』『廣雅』『字書』などの訓詁書に加え、礼制や祭祀と関係する『禮記』『周禮』も重要な典拠となっている。

欠部では、『説文』『字書』『廣雅』『方言』『爾雅』などの小学・訓詁系文献が多く見られる。欠部の字は、口を開く動作、声、呼吸、感情表現などに関わるため、語義の微細な差異を説明するうえで、訓詁書類が多く利用されたと考えられる。

このように、部首ごとの典拠傾向は、各部首に属する字の意味領域と密接に関係している。『説文』は主要部首に広く見られるが、特に言部・食部・欠部で多い。一方、『方言』や『廣雅』は具体的な語義説明を必要とする部首で目立ち、『禮記』『周禮』は礼制・儀礼・飲食に関わる字項目で多く現れる。また、『毛詩』『尚書』『左氏傳』『國語』などの経伝、史書類は言部に集中する傾向がある。これは、言部の字が言語行為や政教的概念と関わり、経書、史書の用例が字義説明に利用されやすいことを示している。

8.5 典拠分布から見た卷九の特徴

以上の分析から、卷九における典拠分布にはいくつかの特徴が認められる。

- ・『説文』の突出した利用である。『説文』は 386 回出現し、他の文献を大きく上回る。これは、原本『玉篇』が字形・字義説明において『説文』を重要な基礎資料として利用していた

ことを示す。

- ・小学書・訓詁書の比重の高さである。『字書』『廣雅』『蒼頡篇』『聲類』『方言』『爾雅』『埤蒼』などが高頻度で現れることから、巻九の字義説明は既存の小学書、訓詁書を広く参照して構成されていると考えられる。
- ・経伝・史書類の利用も少なくないことである。『毛詩』『禮記』『左氏傳』『尚書』『周禮』『國語』『論語』などは、語義を具体的な用例によって説明する際に利用されている。特に言部では、これらの文献が多く引用される傾向がある。
- ・部首ごとの意味領域と典拠分布との関係である。食部では飲食・礼制に関わる文献が目立ち、言部では経伝、史書類が多く、欠部では訓詁系文献が比較的多い。このことは、原本『玉篇』が部首ごとの語彙領域に応じて、異なる典拠群を選択的に利用していた可能性を示している。

また、「野王案」は 219 件確認され、編纂者である顧野王自身の判断が巻九の字義説明に重要な役割を果たしていることが分かる。以上のように、巻九の典拠分布分析は、原本『玉篇』が引用文献と編纂者の判断を組み合わせることで字義を構成していたこと、また部首ごとの語彙領域に応じて異なる典拠群を利用していた可能性を示すものである。

9. 考察

9.1 LLM 抽出の有効性

本研究の結果から、LLM を用いた典拠情報抽出は、古辞書の電子テキストを整理するうえで有効に機能することが確認された。とくに、書名符号《》によって明示された典拠書名については、比較的安定した抽出が可能であった。食部を対象とした検証では、《》内に示された書名の抽出結果が人手 TEI データと高い割合で一致し、Precision と Recall のいずれにおいても良好な値を示した。これは、電子テキスト中に一定の標識が存在する場合、LLM がその形式を手がかりとして典拠書名を抽出できることを示している。

また、LLM は書名符号を伴わない未標識典拠の発見にも有効であった。食部では、「世本」「周易」「淮南」「墨子」「大戴禮」などが候補として抽出され、人手確認の結果、典拠情報として扱うべきものと判断された。さらに、巻九全体でも 13 件の未標識典拠候補が確認され、すべて典拠情報として採用された。この点は、単に《》内の文字列を取り出すだけでは得られない成果である。未標識典拠を発見するには、前後の文脈を読み取り、その語が文献名として機能しているかを判断する必要がある。LLM はこの文脈判断を支援し、人手確認の対象を効率的に提示する役割を果たした。

さらに、LLM は大規模データへの拡張にも適している。食部 144 項目であれば人手による TEI 記述も可能であるが、巻九全体は 27 部首、693 項目に及び、典拠書名の出現数も人手確認後で 1832 件に達する。LLM を用いることで、初期抽出、候補提示、正規化、部首別集計を効率的に進めることができた。したがって、LLM は古辞書の典拠情報整理において、人手作業を置き換えるものではなく、大量の電子テキストから研究上有用な候補を取り出し、統計分析可

能な形へ整えるための補助的手段として有効である。

9.2 LLM 抽出の限界

一方で、LLM による抽出には明確な限界も存在する。特に問題となるのは、注釈家等人名の計数における境界の設定である。人名の直後に「曰」が続く場合は、注釈家による注文として比較的判断しやすい。しかし、本文中には人名が単なる言及として現れる場合や、引用文献名と注釈家名が連続して現れる場合もある。そのため、文字列として人名を抽出することと、それを典拠情報として数えることは同一ではない。また、「野王案」は編纂者である顧野王自身の案語であり、一般の注釈家名とは性格が異なる。これを注釈家等人名に含めるか、あるいは編纂者案語として別に扱うかは、研究目的に応じて明確に定義する必要がある。食部の検証においても、注釈家等人名の抽出では書名抽出に比べて差異が大きく、計数基準の設定が課題となった。

また、LLM は異体字、残缺字、校勘符号を含む本文に影響を受けやすい。原本『玉篇』残巻の電子テキストには、異体字、俗字、残缺を示す記号、補訂、校勘注などが含まれる。たとえば、「説文／説文」「爾雅／爾雅」「礼記／禮記」「埤倉／埤蒼」のような表記揺れは、正規化しなければ別々の文献として集計されてしまう。さらに、虫損や欠字を示す記号が書名や引用文の内部に含まれる場合、書名の境界や引用範囲の判断に影響を及ぼす可能性がある。したがって、抽出後には表記の正規化と原文確認が不可欠である。

さらに重要なのは、LLM の抽出結果が文献学的判断そのものを代替するものではないという点である。LLM は典拠候補を提示することはできるが、その候補を典拠情報として採用できるか、どの文献に帰属させるべきか、どの表記に正規化すべきかについては、最終的に研究者の判断が必要である。特に、略称、古注、異文、校勘注を含む箇所では、単なる自動処理だけでは十分ではない。このため、本研究では LLM の抽出結果をそのまま最終データとはせず、人手確認を組み込んだ。食部では人手 TEI データを基準として抽出結果を検証し、巻九全体では未標識典拠候補を人手で確認した。このような確認作業を経ることによって、LLM の効率性を活用しつつ、研究データとしての信頼性を確保することができる。

9.3 TEI と LLM の関係

本研究の目的は、TEI による構造化を LLM に置き換えることではない。むしろ、TEI と LLM は相互補完的な関係にある。

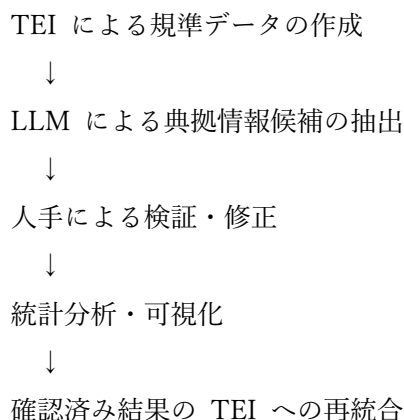
TEI は、典拠情報を精密かつ規範的に記述するための枠組みを提供する。前稿で示したように、典拠書名は <listBibl>、注釈家等人名は <listPerson> に整理することができる。また、本文中の引用は <cit type="quotation">、注釈家による注文は <cit type="commentary">、顧野王自身の案語は <note type="editor"> として区別できる。このような構造化によって、典拠情報を検索・集計・再利用しやすい形で保持することが可能となる。

一方、LLM は、大量の電子テキストから典拠情報の候補を効率的に抽出する役割を担う。巻九全体のように 693 項目、1832 件の典拠書名を含むデータでは、すべてを最初から人手で TEI

化するには大きな負担がかかる。LLM を用いることで、明示的な書名、未標識典拠候補、注釈家名を一括して抽出し、研究者が確認すべき箇所を整理することができる。

このとき重要なのが Human-in-the-loop の手順である。LLM が抽出した結果を、食部の人手 TEI データや原文と照合し、研究者が確認・修正を行うことにより、文献学的信頼性を確保することができる。特に、未標識典拠、注釈家名の計数、異体字の正規化などは、人手による判断が不可欠である。

したがって、本研究における理想的な流れは、次のように整理できる。



この流れにおいて、TEI は構造と規範を提供し、LLM は抽出と拡張を支援し、Human-in-the-loop は研究データとしての信頼性を担保する。三者は代替関係ではなく、相互に補い合う関係にある。最終的には、LLM によって抽出され、人手確認を経た典拠情報を、再び TEI データへ戻すことが可能である。たとえば、確認済みの典拠書名は `<listBibl>` に追加し、本文中の引用箇所は `<cit>` や `<ref>` として記述できる。注釈家名は `<listPerson>` と対応させ、顧野王の案語は `<note type="editor">` として整理できる。このように、LLM による抽出結果は、TEI データを拡充するための中間成果として位置づけることができる。

以上の考察から、LLM は古辞書研究における典拠情報整理を効率化する有力な手段であるが、その成果を信頼できる研究データとするためには、TEI による構造化と人手による確認が不可欠である。

10. おわりに

本稿では、原本『玉篇』残巻巻九を対象として、LLM を用いた典拠情報抽出とその検証、さらに部首別分布の分析を行った。前稿で作成した食部の人手 TEI データを規準データとして利用し、LLM の抽出結果を評価したうえで、同じ手順を巻九全体 27 部首、693 項目へ拡張した。その結果、巻九には典拠書名が延べ 1832 回、85 種、注釈家等人名が延べ 618 回、15 種確認された。

食部を用いた検証では、『《』で明示された典拠書名について、LLM が高い精度で抽出できることを確認した。また、未標識典拠候補を含めることで再現率が向上し、LLM が人手による形式

的整理では見落とされやすい典拠候補を発見する補助的手段となりうることも示された。これは、LLM を単なる自動化の道具としてではなく、研究者の確認作業を支援する Human-in-the-loop 型の方法として位置づけるうえで重要である。

さらに、整理した結果、典拠情報は言部・食部・欠部に集中する傾向があること、また『説文』『字書』『廣雅』『毛詩』『禮記』などの主要典拠が部首ごとに異なる分布を示すことを確認した。もっとも、注釈家名の計数規準、異体字・闕字・校勘符号を含む本文への対応等には、なお人手による文献学的判断が必要である。

今後は、本稿の方法を残卷 7 卷 63 部首へ拡張し、『篆隸万象名義』や宋本『玉篇』との比較、典拠情報データベースの構築、確認済みデータの TEI 化を進めたい。

参考文献

- 池田証壽「平安時代漢字字書総合データベース：現状と課題 2014 夏」『漢デジ 2014—デジタル翻刻の未来』、京都大学人文科学研究所附属東アジア人文情報学研究センター 2014.
- 陳錦春、蓋翠傑「原本『玉篇』引『詩』考論」『図書館雑誌』上海図書館 2014 (33-03) .
- 郭萬青「《原本玉篇殘卷》引《國語》例辨正」『学灯』国家図書館 2009 (02) .
- 胡吉宣『玉篇校釋』上海古籍出版社 1989.
- 胡吉宣「《玉篇》引書考異」『語言文字研究專輯』上『中華文史論叢』增刊 上海古籍出版社 1982.
- 胡智雄「原本《玉篇》殘卷引《礼記》管窺」『安徽文学』安徽省文聯 2017 (11) .
- 李慶彬「原本《玉篇》引《尚書》考論」『歷史文献研究』中国歷史文献研究会 2019 (01) .
- 李媛「篆隸万象名義の本文データベースの構築 —玉篇殘卷対応部分を中心にして」『国語国文研究』146、2015.
- 李媛「食部を例に原本玉篇の典拠情報整理について」、『東洋学へのコンピュータ利用 第 40 回研究セミナー』予稿集、2026 年.
- 劉杰陽「原本《玉篇》引《尚書》傳本性質試析」『古典文献研究』南京大学古典文献研究所 2021 (03) .
- 劉新楠「原本《玉篇》引書字詞異文價值举隅」『忻州師範学院学报』2021 (04) .
- 呂浩『玉篇文獻考述』上海人民出版社 (2018) .
- 徐前師「唐写本《玉篇》中的《字書》」『湖南科技大学学报』2007 (04) .
- 蘇芃「試論《原本玉篇殘卷》引書材料的文献学價值：以引《左氏伝》為例」『図書館雑誌』上海図書館 2006 (12) .
- 蘇芃「原本《玉篇》引《史記》及関連古注材料考論：裴駰《史記集解》南朝梁代傳本之發現」『文史哲』山東大学 2011 (06) .
- 蘇芃「《玉篇》殘卷征引『儀礼』考」『古籍整理研究学刊』東北師範大学文学院古籍整理研究所 2017 (05) .
- 蘇芃「南朝蕭梁時期《春秋公羊傳》經注文本探微：以原本《玉篇》引書為例」『歷史文献研究』中国歷史文献研究会 2018 (01) .

姚永銘『原本玉篇殘卷校証』浙江古籍出版社 2023.

臧克和『中古漢字流變』華東師範大學出版社 2008.

朱葆華「原本《玉篇》引《說文》考略(一)」『中文自學指導』全國高等教育自學考試指導委員會 1998 (02) .

TEI Consortium, TEI P5: Guidelines for Electronic Text Encoding and Interchange, <https://tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf> (2026 年 7 月 12 日閱覽) .

TEI 協會東アジア／日本語分科会, <https://tei-c.org/activities/sig/eastasian/> (2026 年 7 月 12 日閱覽) .