

本草和名と古活字版和名類聚抄の全文テキストデータ（附：和名索引）

劉 冠偉・武 倩・申 雄哲・韓 一・藤本 灯

本稿では、平安時代に日本で成立した古辞書『本草和名』、『和名類聚抄』の内容を Unicode によって符号化したテキストデータ、および両辞書の和名を片仮名にした和名索引データを紹介する。本データは、日本語史研究や学術データベースの構築に応用できるものと期待される。

キーワード

古辞書, 平安時代, 和名, 日本語史

1. 概要

1.1 本データについて

本データは、平安時代に日本で成立した古辞書である『本草和名』（底本：享和2年・版本）と『和名類聚抄』（底本：古活字版）に掲出された内容を Unicode によって符号化したテキストデータである。両辞書とも漢語を見出しとし、注文はすべて漢字、漢文（一部変体漢文）で書かれており、見出しの漢語に対応する和名¹⁾も漢字の音訓を借用した万葉仮名による。

筆者らは数年にわたり、両辞書のデータ化に取り組んできた^{1), 2)}。今回は、全文テキストデータのほか、本文データへのアクセスの利便性を向上させるため、万葉仮名を片仮名に変換したリストも併せて作成し、両辞書の見出しと所在を含める索引として公開した。なお、すでに『和名類聚抄』のデータは国立国語研究所の「日本語史研究用テキストデータ集」にて公開済みであるが、本文のテキストのみであったため、IDを付与し、索引として利用するために、本文中にある和名語彙³⁾の抽出を行った。

両辞書は成立年代と表記体系（漢字と万葉仮名の併用）が近く、いずれも日本語史における重要な研究資源である。日本語名詞の和訓を追究するために、両者を同時に参照する場面も多い。しかしこれまでに作成したデータは、同じ語彙に対する両辞書の表記および符号化方針に相違があったため、横断的な利用が困難であった。そこで、両辞書の語彙を揃え、万葉仮名を片仮名に変換した索引を作成し、一つの統合されたデータセットとして提供することを目指すこととした。

本データは、筆者らが構築・公開している日本語古辞書データベース『辞書語彙データベース』⁴⁾内で検索イン

タフェースも提供している³⁾。

1.2 作成者

本草和名本文：武倩，劉冠偉

古活字版和名類聚抄本文：藤本灯，申雄哲，韓一

本草和名・古活字版和名類聚抄 和名索引：劉冠偉，藤本灯，武倩

2. 標準

2.1 データフォーマット

UTF-8 でエンコードした Comma-Separated Values (CSV) プレインテキストを採用した。

2.2 フォーマットを採用した理由

本データセットは二つの古辞書に掲出された内容を整理して研究データとして利用するためのものである。見出し語、和名、所在などの対応関係を記録しやすい表形式を採用した。

また、Berners-Lee (2006) では、データを公開する際に、専用のソフトウェアに縛られず、独占的ではないフォーマットが有益である⁴⁾ため、UTF-8 の CSV プレインテキストデータを採用した。

3. データ量

(1) 本草和名：1,033 レコード（約 172KB）

(2) 古活字版和名類聚抄：8,511 レコード（約 978KB）

(3) 和名索引：3,115 レコード（約 207KB）

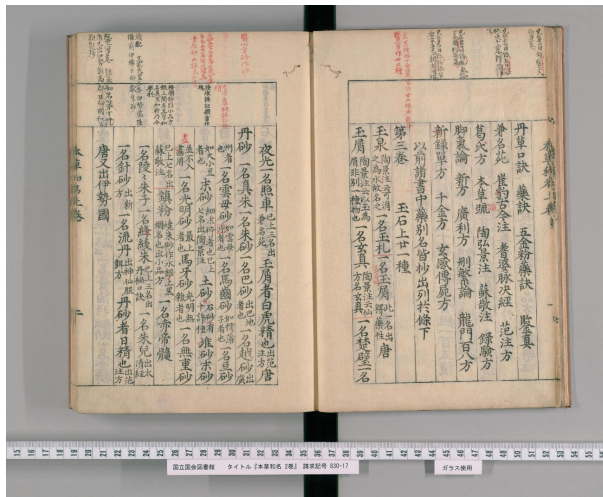


Fig. 1. 『本草和名』の一例
An example of Honzō Wamyō

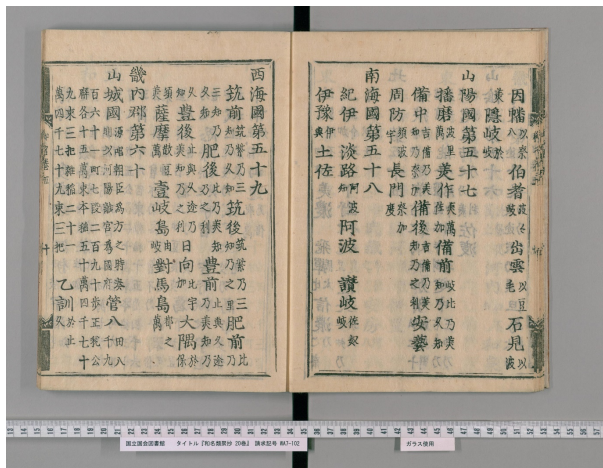


Fig. 2. 『和名類聚抄』の一例
An example of Wamyō Ruijushō

4. 構造

4.1 底本の概要と構成

底本とした文献の概要と構成については次のとおりである (藤本 2021⁵⁾ に加筆修正)。

本草和名

〔概要〕 918 (延喜 18) 年ごろ、醍醐天皇に侍医として仕えた深根輔仁 (生没年未詳) が著した、日本現存最古の本草書。「輔仁本草」「和名本草」とも。「本草」とは医療に用いる動植物を指し、「本草学」は中国における不老長寿のための医薬の学問、「本草書」はそのための書物である。日本でも奈良時代以降、医学生らが中国の本草書を学んだ記録があるが、その一つである『新修本草』(唐の蘇敬らによる) を主体とし、中国名 (漢語・漢名) に相当する日本名 (和語・和名) をあて、日本における産地 (和産地) などを記したものが本書である (Fig. 1.)。

今日の学問水準に照らせば必ずしも正しくない内容を含むものの、丹波康頼撰『医心方』(984 年) 等、多くの医学書に引用され、後世の日本の医学に影響を与えたとされる。『本草和名』の書名や内容は『和名類聚抄』や『本朝書籍目録』(鎌倉後期成立か) 等に引用されて知られていたものの、長らく古写本の所在は不明となっていたが、江戸幕府の医家で、丹波康頼の末裔である多紀元簡 (1755-1810) が幕府の書庫である紅葉山文庫より古写本を発見し、校訂した結果を 1802 年 (享和二年) に版行したため、再び世に広まった。古写本の行方は再び不明となったが、福山藩医であった森立之 (1807-1885) が古写本を弟子に影写させたものが台湾の国立故宮博物院に現存する。また版本に森立之・約之親子による校注が付されたものが日本古典全集や「松本書屋貴書叢刊」(1993) 等から出版されており、現在もよく利用されている。

〔構成〕 上下 2 冊。版本の巻首の記述によれば、『新修本草』から 850 種、『諸家食経』から 105 種、『本草拾遺』その他から 70 種の計 1025 品目を収載。収録された項目は玉石 3 巻・草 6 巻・木 3 巻・獣禽 1 巻・虫魚 1 巻・菓 1 巻・菜 1 巻・米穀 1 巻・有名無用 1 巻・本草外薬の 10 種に意義分類され、この順序は『新修本草』に倣う。本書は、『新修本草』全 20 巻のうち、最初の 2 巻 (序例) を省略し、「第三巻」(玉石) から「第二十巻」(有名無用) までを引用したものに加え、『新修本草』以外の書より抜粋した「本草外薬」から成る。

和名類聚抄

〔概要〕 源順 (911-983) が、醍醐天皇第 4 皇女の勤子内親王 (904-938) の命を受けて撰進した、漢文体による意義分類体辞書。934 (承平 4) 年ごろ完成。基本的には漢語 (多く、物の名) を見出し項目として、注文には出典・用例・意義注・音注・字体注・和訓等を載せる (Fig. 2.)。和漢の書物を多く引用し、中には『楊氏漢語抄』『弁色立成』(序文には『和名本草』『日本紀私記』等とともに『漢語抄』と称される) のように、本邦古辞書と目される書名も見え、逸書復元の手かがりともなっている。序文に編纂の動機や本書構成に関わる事項が詳述される。

〔構成〕 大きく 20 巻本系統と 10 巻本系統に分かれる。「部」による意義分類の下に更に細かく「類」「具」等を設ける (例えば 20 巻本冒頭の天部の下には景宿類・雲雨類・風雪類がある)。20 巻本 (元和古活字本) は 32 部 249 類約 3350 語、10 巻本 (箋注倭名類聚抄) は 24 部 128 類約 2600 語からなる。20 巻本にのみ見える部には「職官」「国郡」等がある。

4.2 作成の方針

Unicode によって底本に掲出された字体を可能な限りに忠実に符号化した。Unicode に未収の字体⁶ に対して、本テキストデータセットは IDS⁷ 方式 (例：■ [□穴忠]) で表す。

(1) 本草和名

項目	内容	例
ID	見出し語の ID、本草和名の略称 HW の次に、巻 (A が上、B が下)、丁数と表裏 (a が表、b が裏)、所在行を置く、	HW_A_01b_8
所在	見出し語の所在	巻上三、1 丁裏、8 行
部類	巻首にある部類情報	第三巻 玉石上廿一種
見出し	見出しの漢名	玉泉
本文	本文の内容	[陶景注云可消之為水故名之] 一名玉札一名玉屑 [此一名出積薬性] 唐

(2) 古活字版和名類聚抄

項目	内容	例
ID	見出し語の ID、古活字版和名類聚抄の略称 KWRS の次に、巻数、丁数と表裏 (a が表、b が裏)、所在行を置く、	KWRS-05-10a-07
所在	見出し語の所在、巻、丁、表裏、出現行からなる	巻 5、10 丁表、7 行
部	見出し語が所属する部門	国郡部第 12
類・具・国郡	見出し語が所属する分類	畿内郡第 60
分類 (他)	見出し語が所属する下位分類	山城国
見出し語	見出し語の表記	山城国 [源唱朝臣為方之時奏明以河陽離宮為国府] 管八 [田八千九百六十一町七段二百九十歩正税公廩各十五万束本稻五十万四千七十九束三把雑稻二十一万四千七十九束三把]
注文	見出し語の語釈	

両辞書に掲載された和名は、漢字の音または訓を用いて日本語の音韻を記述する万葉仮名で表記されるが、同じ音韻に対して複数の万葉仮名がある場合、漢文字符に揺れが生じることがある。検索・対照上の便利を図るため、索引では片仮名に変換した。

両辞書の本文には、双行で記載された割注が存在している。本データセットでは、割注を「[]」で括弧。また、本草和名では、「一名」の後ろにある別名は「[]」で、松本本の上欄にある頭注に対応する漢字は「〈〉」で括弧。

4.3 項目

5. メタデータ

(3) 和名索引

項目	内容	例
和名 (片仮名表記)	片仮名に変換した和名	アイツツシ
和名 (原文表記)	辞書に出現したままの表記	阿伊豆々之
見出し語	和名が出現した項目の見出し語 ¹⁾	山榴
見出し語 ID	見出し語の ID、本草和名は HZWM から、和名類聚抄は KWRS から始まる	KWRS-20-26b-09
所在	見出し語の所在	和名類聚抄：巻 20、26 丁裏、9 行

データ名 (Name)	本草和名と古活字版和名類聚抄の全文テキストデータ (附：和名索引)
配布者 (Contributor)	劉冠偉
作成者 (Creator)	劉冠偉、武倩、申雄哲、韓一、藤本灯
キーワード (Keywords)	古辞書；平安時代；
URL	https://dataset.kojisho.com/hzwm-kwrs-wamyo-dataset
対象時期 (Temporal)	日本平安時代
対象地域 (Spatial)	日本
データフォーマット (Encoding Format)	text/csv
データの言語 (In Language)	日本語、中国語
公開日 (Date Published)	2024-05-07
研究資金 (Funding)	JSPS-24K16080; JSPS-23K20465; 2023THZWJC31
ライセンス (License)	CC BY-NC-SA 4.0

6. ライセンス

クリエイティブ・コモンズ・ライセンスの「表示-非営利-継承 4.0」(CC-BY-NC-SA 4.0) の下で配布している。

7. リンク

<https://dataset.kojisho.com/hzwm-kwrs-wamyo-dataset>

8. データの展望

本データセットに附した索引は、『本草和名』と『和名類聚抄』に掲載された日本語 (和名) の部分のみをそれぞれ抽出したものであった。しかし、各書における注の構造は複雑であり、和名に対応する漢語が、見出し・見出しの一部・注のいずれに当たるのかといった情報についても、さらに綿密な語学的検討を経て紐づけていく必要があるであろう⁵⁾。またこの 2 書には濁音専用の万葉仮名がほとんど用いられていないことから、本来の

語形を復元するためには他の文献等と照合しながら検討しなければならない。『本草和名』から『和名類聚抄』へ、また『和名類聚抄』から他の辞書に引用された項目も多い。語形を確定させるためにも、『辞書語彙データベース』に収録されるその他の辞書間の関係を参照させるような情報を付与していくことが、目下の課題である。

参考文献

- 武倩, and 劉冠偉. 2018. “『本草和名』データベースの構築に向けて.” In 第23回公開シンポジウム「人文科学とデータベース」発表論文集, 37–42. <https://www.jinbun-db.com/symposium/archives/23>.
- 藤本灯, 韓一, and 高田智和. “古辞書の構造化記述の試み：『和名類聚抄』を例に.” 国立国語研究所論集, no. 21 (July 2021): 85–94. <https://doi.org/10.15084/00003438>.
- 藤本灯, 劉冠偉, 久保根子, and 大島英之. “古辞書データベースの開発.” In じんもんこん 2023 論文集, 2023:67–72, 2023.
- Tim Berners-Lee. 2006. “Linked Data.” Accessed May 3, 2024. <https://www.w3.org/DesignIssues/LinkedData.html>.
- 藤本灯. “古辞書の語彙.” In “古代の語彙 一大陸人・貴族の時代— (シリーズ「日本語の語彙」2).” Edited by 飛田良文 and 佐藤武義, 38–48. 朝倉書店, 2021.
- Consortium, The Unicode. *The Unicode Standard*, Version 15.0.0. Mountain View, CA: The Unicode Consortium, 2022.

End notes

- ^a 西暦 1802 年.
- ^b 漢語の意味に対応する日本語.
- ^c 翻字本文も藤本・申・韓が作成したものであり、国立国語研究所『日本語史研究用テキストデータ集』「二十

巻本和名類聚抄 [古活字版]」(https://www2.ninjal.ac.jp/textdb_dataset/kwrs/) で公開されている。当該データは人間文化研究機構広領域連携基幹研究プロジェクト「異分野融合による総合書物学」の国立国語研究所ユニット「表記情報と書誌形態情報を加えた日本語歴史コーパスの精緻化」(代表：高田智和) において作成されたものである。

- ^d 厳密には「碑磔：謝古 (シヤコ)」のような二字以上の漢語の語形を示す類音注や「箏：象乃古止 (シヤウノコト)」のように漢字音と訓で二度重ねて読む文選読みも含める。
- ^e <https://jisho-goi.kojisho.com/>
- ^f 国立国会図書館デジタルコレクション <https://dl.ndl.go.jp/pid/2538099> (参照 2024-05-07)
- ^g 国立国会図書館デジタルコレクション <https://dl.ndl.go.jp/pid/2544218> (参照 2024-05-07)
- ^h 本草和名は Unicode 6.0 (拡張漢字 D), 和名類聚抄は Unicode 3.1 (拡張漢字 B) まで対応している。
- ⁱ Ideographic Description Sequence (漢字構成記述文字), Unicode に収録されていない漢字などを, IDC (Ideographic Description Characters) で構造を, Unicode に既存の漢字符号で構成部品を表す方法である⁶⁾.
- ^j 必ずしも和名に対応する漢字表記ではない場合もある。
- ^k <https://creativecommons.org/licenses/by-nc-sa/4.0/>

謝辞

本研究は JSPS 科研費 24K16080, 23K20465, 清口大学大学自主科研計画文科専門経費・基礎研究専門 2023THZWJC31 の助成を受けたものです。

Full Text Data of Honzō Wamyō and Wamyō Ruijushō (with Wamyō Index)

Guanwei Liu, Qian Wu, Woongchul Shin, Yi Han and Akari Fujimoto

This manuscript presents a comprehensive overview of text data encoded in Unicode, delineating the contents of two dictionaries from Japan's Heian period: Honzō Wamyō and Wamyō Ruijushō. Additionally, it includes an index of Japanese names transcribed into Katakana. The utilization of Unicode facilitates the preservation and dissemination of these historical linguistic resources, contributing significantly to scholarly endeavors in Japanese language studies. The encoded text data not only preserves the linguistic heritage encapsulated within these dictionaries but also opens avenues for further research in Japanese historical linguistics.

Key Words: old dictionary, Hei-an period, Wamyō, Japanese historical linguistics.