

# 『日本語歴史コーパス 室町時代編Ⅲ抄物』Ver.0.8 テキストの凡例と『中納言』表示項目について

2026年3月31日 村山実和子・渡辺由貴・北崎勇帆

## はじめに

『日本語歴史コーパス 室町時代編Ⅲ抄物』を構築するにあたっては、まず、最も善本とされる古活字版 18 行本を底本とし、亀井孝『語学資料としての中華若木詩抄(校本)』(清文堂出版)所収の「校異の部」などを参考に、写本 6 種、古活字版 2 種、整版 3 種と校合して独自の校訂本文テキストを作成している(古田 2023、青木 2024)。コーパス化に際して、そのテキストを一部変換し、そこに様々な情報を付加することで XML データを構築した(北崎・古田2023)。その多様な情報が反映されたものが中納言版の本コーパスである。本稿では中納言上に表示される各種情報について、テキストの校訂状況や表示情報に関する概要を述べる。

なお、本コーパスでは、研究上必要と思われる情報をできるだけ底本の状況に即して記述するよう努めたが、完全に反映できているわけではない。そのため、適宜「ページ番号」に表示された丁数を基に底本の本文を確認されることを推奨する。

## 1. テキストの凡例

### 1.1. テキストに使用する文字

本文テキストの文字入力は JISX0213 の文字集合に準拠している。字形・用法の近い同音・同訓の JIS 内の文字がある場合は代用し、または適切なものがないと判断した場合は、外字として「■」(げた記号)に置き換えて入力した。コーパス本文・原文文字列ともに「■」表示のものと、コーパス本文では代用した文字を表示し原文文字列では「■」を表示するものがある。いずれも、底本の本文、参考リンクとして用意した版本画像等を適宜参照されたい。

#### 【例】

- |                   |              |                 |
|-------------------|--------------|-----------------|
| (1) 〈入力〉 何人観瀑鬢■然  | 〈原文〉何人観瀑鬢■然  | 白+番(幡)、U+76A4   |
| (2) 〈入力〉 自■岩下漣漪碧  | 〈原文〉自■岩下漣漪碧  | 大+目+斗(鬲)、U+3ABA |
| (3) 〈入力〉 いかにも愚痴に… | 〈原文〉イカニモ愚■ニ… | 疒+和             |
| (4) 〈入力〉 砧をうつぞ    | 〈原文〉■ヲウツソ    | 石+召             |

### 1.2. テキストの校訂

#### 1.2.1. カタカナ

底本本文中カタカナで表記された箇所は、平仮名で表示した。変換前の文字列は「原文文字列」列に表示される。

#### 【例】

〈原文〉 ソコテ、曹操カツヲル処テ、蜀劉備ハ恐レテ、江陵ヘニクルソ。曹操カヲツカケタソ。(卷之一・三ウ)

〈入力〉そこで、曹操がつをる処で、蜀の劉備は恐れて、江陵へにぐるぞ。曹操がをつかけたぞ。

### 1.2.2. 濁点

濁音が期待される箇所には濁点が付されていない場合、校合を踏まえ、必要な箇所は濁点を補った。濁点を補う前の文字列は「原文文字列」列に表示される。

【例】

〈原文〉 辛勞シテナカンズルヨリハ、口ヲツクンテ啼カスシテ残春ヲ賞シテ可也。(卷之一・一オ)

〈入力〉 辛勞してなかんずるよりは、口をつぐんで啼かずして残春を賞じて可也。

### 1.2.3. 踊り字

仮名一字分の踊り字(ゝ)は、想定される仮名に置き換えた。変換前の文字列は「原文文字列」列に表示される。なお、漢字の繰り返しや二字分以上に相当するくの字点等は置換しない。

【例】

〈原文〉 ナニト語ソナレハ、ヨクノゝ此楓ヲ御覽セヨ、春ニ三月ノコロ百花ノサキ乱ルゝ中ニモ、カヤウナ花ハアルマイソ。(卷之一・三六オ)

〈入力〉 なにと語るぞなれば、よくノゝ此楓を御覽ぜよ、春二三月のころ百花のさき乱るる中にも、かやうな花はあるまいぞ。

### 1.2.4. 誤字・脱字・衍字

本文中の誤字・脱字・衍字と思しい箇所について、想定される文字列となるよう本文テキストに反映させた。修正される前の文字列は「原文文字列」列に表示される。

(1) 〈原文〉 一二ノ句、皇天ノコトナレバ、天地モコガルハヤウ也。(卷之一・三八ウ ※誤字の例)

〈入力〉 一二の句、皇天のことなれば、天地もこがるるやう也。

(2) 〈原文〉 ミタリニ抹過スヘカラ\_\_。(卷之一・四五ウ ※脱字の例)

〈入力〉 みだりに抹過すべからず。

〈原文〉 二ノ句、サアルホトニ、イマタ年モヨラネトモ、鬢カラ衰テ行タソ。(卷之一・二九ウ ※衍字の例)

〈入力〉 二の句、さあるほどに、いまだ年もよらねども、鬢から衰へて行きたぞ。

### 1.2.5. テキストの補い

形態論情報の付与にあたり、期待される送り仮名・活用語尾・助動詞等が、明示されない場合、また表記上分割不可能な場合に、タグ付きで本文を補った。「原文文字列」列には補う前の文字列が表示される。

(1) 〈原文〉 孫権モ名ヲ好ム者ナレハ、悦テヤカテ同心ス。(卷之一・三ウ)

〈入力〉 孫権も名を好む者なれば、悦うでやがて同心す。

(2) 〈原文〉 船十艘コシラヘテ、人ヲハノセスシテ、イカニモホシスマシタル柴、燥荻ナントヲ其船

ニ積テ(卷之一・四ウ)

〈入力〉 船十艘こしらへて、人をばのせずして、いかにもほしすましたる柴、燥ける荻などを其船に積んで

#### 1.2.6. 漢文

返り点が付される等しており、文脈上、読み下すことが適当だと思われる箇所については、訓読した形を本文とした。その他の漢文相当箇所は「未知語」として扱い、品詞「漢文」とした(別紙『日本語歴史コーパス 室町時代編Ⅲ抄物』形態論情報の概要)も参照)。

【例】

〈原文〉 半ハ記、半ハ忘トテ、夢不熟。サレトモ、江上春風ニモ不驚シテ依旧テモトノ白鷗ノ前ニア  
ルソトニノ句ニ相応スルソ。(卷之一・六オ)

〈入力〉 半ば記し、半ば忘るとて、夢熟せず。されども、江上の春風にも驚かずして旧に依りてもとの白鷗の前にあるぞと二の句に相ひ応ずるぞ。

## 2. 中納言における表示項目と内容

『日本語歴史コーパス 室町時代編Ⅲ抄物』の本文には様々なタグ(本稿末参考表)や単語情報(後述)が付されており、その情報は、Web上のコーパス検索ツール「中納言」上に、検索結果として表示される(図1・図2)。以下では、中納言上の主な表示項目とその内容に関して概説する。

図1 検索画面

4 件の検索結果が見つかりました。  
 検索対象語数: 20,427 記号・補助記号・空白を除いた検索対象語数: 17,775 検索対象サンプル数: 1

サンプルID	開始位置	連番	コア	前文脈	キ	後文脈	語彙素ID	語彙素読み	語彙素	語形	品詞	活用型	活用形	原文文字列
40-抄物 1520_01001	16310	10400	1	に わ れ て 、 鼎 の 三 足 の や う に し て と り や う 也 。#  カ は 曹 操 が  天下三ツツコレテ、鼎ノ三足ノヤウニシテトリアウ也。# カハ曹操カ	つをい	也。# 曹操、 劉 表 を 伐 つ 。#  劉 表 が 子 、 荆 州 を 持 ち な が ら 曹 操 に 降 参 す る  也。# 曹操、 劉 表 ヲ 伐 。#  劉 表 カ 子 、 荆 州 ヲ 持 ナ カ ラ 曹 操 ニ 降 参 ス ル 。	24791	ツヨイ	強い	ツオイ	形容詞-一般	形容詞	連体形-一般	ツライ
40-抄物 1520_01001	159810	105480	1	間 、 諸 侯 王 は 知 行 を そ ろ り と 削 つ て 天 子 の 協 た へ と り て 、 天 子 を  、 諸 侯 王 ヲ ハ 知 行 ヲ ソ ロ リ ト 削 ツ テ 天 子 ノ カ タ ヘ ト リ テ 、 天 子 ヲ	つをく	し て 諸 侯 は は め ん に と 思 ふ 也 。#  こ の 謀 り こ と は 、 天 子 の 万 世 ま で 也  シテ諸侯ヲヨハメント思フ也。# コノ謀リコトハ、天子ノ万世マテモ安穩ニアルヘキ利也。	24791	ツヨイ	強い	ツオイ	形容詞-一般	形容詞	連用形-一般	ツラク
40-抄物 1520_01001	117630	76660	1	又 面 白 き 也 。#  何 れ に し て も 苦 し か ら ぬ 也 。#  二 の 句 、 春 雨 な れ ば  詩人ノ上ニ云モ又面白也。# 何レニシテモ不苦也。# 二ノ句、春雨ナレハ	つよく	は は ら ず し て い かに こ も 細 そ か と し て 塵 な ん ど の 朝 日 に み ゆる や う な 也  ハフラスシテイカニモ細ソカトシテ塵ナントノ朝日ニミユルヤウナ也	24791	ツヨイ	強い	ツヨイ	形容詞-一般	形容詞	連用形-一般	ツヨク
40-抄物 1520_01001	152440	100370	1	及 ば ず し て 乱 は し づ ま る べ し と 云 ふ ぞ 。#  こ の 七 ヶ 国 は 、 う で の  左様ニアラハ、合戦ニ及ハスシテ乱ハシツマルベシト云フ。# コノ七ヶ国ハ、ウデノ	つよき	国 也 。#  当 今 と 御 一 族 也 。#  袁 盎 が 申 し こ と を 聞 こ し 召 し て 、 し ば ら く 物 を 持  国也。# 当今ト御一族也。# 袁盎ガ申しこを聞こし召して、しばらく物を持	24791	ツヨイ	強い	ツヨシ	形容詞-一般	文語形容詞-ク	連体形-一般	ツヨキ

図2 検索結果の表示イメージ

## 2.1. 形態論情報

基本的にBCCWJや『室町時代編Ⅰ狂言』『室町時代編Ⅱキリシタン資料』と同様であり、小椋他(2011)、国立国語研究所言語変化研究領域編(2019)等を参照されたい。中納言において表示される形態論情報(短単位)は、UniDicの見出しに対応している。以下には利用に際して注意すべき点を幾つか挙げる。

### 〔1〕語彙素・語彙素読み

「語彙素」は単語の各種語形・活用形・書字形(表記)を統合した辞書の見出しレベルの階層であり、一般的な漢字・仮名で表記される。「語彙素読み」はその読みをカタカナ表記したものである。語彙素で検索することで、同語彙素内の各種語形・活用形・書字形等の異なるものを一括して取得することができる(図2は語彙素読み「ツヨイ(形容詞「強い」)」の検索結果である)。

### 〔2〕語形

「語形」は、異語形を区別するレベルであり、例えば「ツヨイ」「ツライ」のように音形の異なるものは、いずれも語彙素「強い」の語形として認定される(図3参照)。また、活用語の場合には、語彙素「強い」の語形に「ツヨシ」(文語形容詞・ク活用)、「ツヨイ」(口語形容詞)があるように、口語活用・文語活用の区別もこれに含まれる。

### 〔3〕書字形

「書字形」は異表記を区別するレベルである。同語形でありながら、活用語尾を除いた箇所別の文字符号が与えられる場合、それぞれ別の書字形となる。

### 〔4〕品詞

学校文法における形容動詞は、語幹は「形状詞」、語尾は「助動詞」に分割されている。

### 〔5〕活用例

文語活用として処理されているものには「文語下一段」のように「文語」が表示されるが、口語活用には「下一段」のように「口語」は表示されない。形容詞も同様に、文語活用のものは「文語形容詞」、口語のものは「形容詞」と表示される。

### 〔6〕活用形

活用語の場合、その活用形が表示される。活用形の小分類には、「一般」「～音便(ウ音便、撥音便、促音便等)」「補助」「省略」「融合」がある。本コーパスに見られる「融合」は、「ではない」「である」等の「で」のみである。これは『室町時代編Ⅰ狂言』『室町時代編Ⅱキリシタン資料』の処理と同じく、断定の助動詞「なり」の連用形「に」に助詞「て」の付いた「にて」の変化したものと見て、語彙素「なり」(助動詞・文語助動詞-ナリ-断定・連用形-融合)としている。

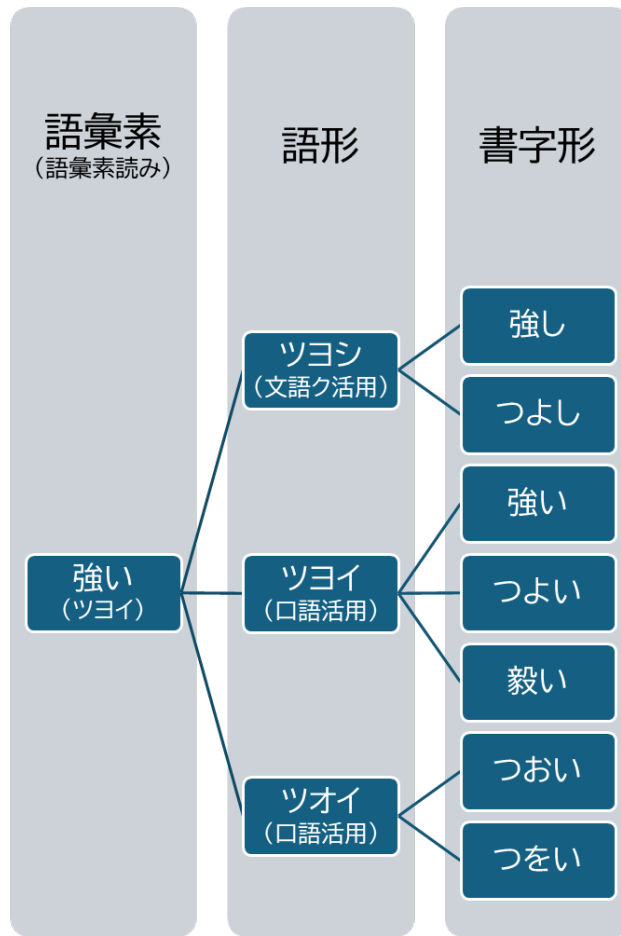


図3 語彙素「強い(ツヨイ)」の語彙素・語形・書字形

〔7〕 原文文字列

カタカナ・濁点・踊り字・誤字等を校訂する前の本文は「原文文字列」および「原文 KWIC」列に表示される(図4)。

サンプル ID	開始位置	連番	コア	前文脈	キー	後文脈	語彙素 ID	語彙素読み	語彙素	語形	品詞	活用型	活用形	原文文字列
40-抄物 1520_01001	265020	173530	1	也。# 喜雨 # 義堂 # 天地如蕉草木黄 [雲霓 日々幾相望 [愁辺喜聴今宵雨 [滴在芭蕉也 不妨 # 一二の句]。[早天のことなれ]。[天 地も]	こがる る	[やう]也。# 草木も[こと]ノ く[枯れ]て[黄色]になる[ぞ]。# 日々[空]を[望]み[で]、いかなる	12487	コガレ ル	焦がれ る	コガレ ル	動詞- 一般	文語下 二段-ラ 行	連体 形-一 般	コカレハ
				自然ニ題ニ出スコトモアレハ、其時ノタスケニナ ルヘキ也。# 喜雨 # 義堂 # 天地如蕉草木黄 雲霓日々幾相望 愁辺喜聴今宵雨 滴在芭蕉 也不妨 # 一二の句、早天ノコトナレハ、天地モ	コカ ルハ	ヤウ也。# 草木モコトノ 枯テ黄色ニナルゾ。# 日々空 ヲ望テ、イカナル雲モアリテ雨 トナレカシ、又、雲モフケカ								

図4 本文テキスト(出現書字形)と原文文字列

キーとして表示された「こがるる」が、原文文字列ではカタカナ表記であり、原文では濁点無表記であったこと、踊り字が含まれていたことが示される。また、前後文脈の原文KWIC表示を見る

と、「早天(語彙素:「干天)」」が、原文では「早天」となっており、コーパス本文では修正されていること、マ行四段動詞のウ音便形「望うで」が、原文では「望テ」であり、コーパス化に際してテキストが補われていること等が確認できる。ただし、底本の詳しい状況を確認する場合は、ページ数を元に底本を直接ご参照いただきたい。

## 2.2. 本文情報

次に、本文の漢文・抄文の別等の表示方法について述べる。

### 〔1〕 本文種別

性質の異なる本文を区別し、「本文種別」として表示している。具体的な分類は下記のとおりである。

表 2 本文種別の表示

本文種別	下位分類	説明
漢文	詩題	個別の漢詩の詩題を表す
	作者	個別の漢詩の作者を表す
	漢詩	個別の漢詩のまとまりを表す
(明示なし)		先行する漢詩に対する注釈、抄文に相当する
発話引用		抄文中に見られる、発話や思考内容の引用相当箇所を表す
その他	タイトル	作品(巻)の題(尾題含む)を表す
		抄者等を表す

### 〔2〕 発話引用

漢詩に対する抄文の中で、発話(心内文含む)の引用に相当する範囲について、<speech>タグを付与し、中納言上での本文種別が「発話引用」として表示されるようにしている。

#### 【例】

- (1) 孫権が劉備に同心すると云ふことを聞いて、曹操いかりて孫権が処へ案内を云ふことは、「水軍八十万人を以て、船にて参じて、一合戦仕るべし。呉国にて参会申すべし」と云ふぞ。  
(巻之一・三ウ)
- (2) 「春もはや秋になりたるよ。此の如く時節のうつりやすきにつけても学をせいでは」と思ふて、  
ふつとをきて灯を吹いて漢書をよむ也。(巻之一・十四オ)

## 2.3. 作品情報

### 〔1〕 ジャンル

「ジャンル」には、一律に「抄物」と表示した。

## 〔2〕 作品名

「作品名」には、一律に「中華若木詩抄」と表示される。

## 〔3〕 成立年

「成立年」には、『中華若木詩抄』の成立年として推定される「1520(年)」が表示される(如月寿印の没年は1533年である)。サンプルIDにもこれが反映される。

## 〔4〕 巻名等

「巻名等」には「巻之一」のように巻の別が表示される。

## 2.4. 底本情報

### 〔1〕 底本

「底本」項目には、「古活字版 18 行本中華若木詩抄」と表示した。厳密には、古活字版十八行本を底本とし、写本 6 種、古活字版 2 種、整版 3 種と校合して、校訂本文を作成している。

### 〔2〕 ページ番号

「ページ番号」には、対応する底本の丁数と表裏の別(オ/ウで区別)が表示される。

## 2.5. 外部リンク

『中華若木詩抄』について、底本と同じく古活字版18行本の、東京大学附属図書館青洲文庫蔵本の公開画像への参照用リンクを用意した(<https://da.dl.itc.u-tokyo.ac.jp/portal/collection/0059ade7-1028-113b-6a35-8cb225a12a32>)。

検索結果に応じて、その参照先が表示される(表示名:UT-Seishu)。底本とは異なる「参照」用であり、コーパス用に校訂された本文とは必ずしも一致しない点を含みうることに注意されたい。

### 附録 本コーパスのタグセット

要素(タグ)名	説明
<text>	一卷全体
<front>	前付相当の箇所
<body>	主本文相当の箇所
<article>	一記事の範囲(『中華若木詩抄』においては漢詩+抄文の1セット)
<titleBlock>	記事とは認められない、<text>直下レベルでの表題周り
<p>	段落を表す。抄物ではタイトルや漢詩部分を除く主本文(『中華若木詩抄』においては抄文に相当)
<block>	記事中のタイトル等、主本文とは切り分けたい段落要素(『中華若木詩抄』においては、詩題・作者・漢詩の区別にも使用)

<speech>	発話・思惟内容の引用相当箇所。話者情報は含まれない。
<s>	文
<corr>	濁点無表記、踊り字、漢文の返読、誤字・脱字・衍字等の本文の修正
<char>	一字を表す単位、抄物ではカタカナ表記箇所に使用
<pb><lb>	底本の改ページ位置・改行位置

## 参考文献

- 青木博史(2024)「抄物コーパス」の構築をめざして『文献探究』62, pp.57-56
- 小椋秀樹・小磯花絵・富士池優美・宮内佐夜香・小西光・原裕(2011)『現代日本語書き言葉均衡コーパス』形態論情報規定集第4版(下)平成22年度研究成果報告書 国立国語研究所
- 亀井孝『語学資料としての中華若木詩抄(校本)』清文堂出版
- 北崎勇帆・古田龍啓(2024)「ひまわり版抄物コーパス」(中華若木詩抄)構築報告「抄物コーパス」の構築とコーパスを応用した日本語史研究 2023年度第2回研究会 発表資料
- 国立国語研究所言語変化研究領域(片山久留美)編(2019)『日本語歴史コーパス 室町時代編』形態論情報規程集 Ver.1.0([https://clrd.ninjal.ac.jp/chj/morph\\_muromachi\\_v1\\_0.pdf.pdf](https://clrd.ninjal.ac.jp/chj/morph_muromachi_v1_0.pdf.pdf)) (2026年3月10日最終閲覧)
- 古田龍啓(2023)「抄物コーパスの構築へ向けて」通時コーパスシンポジウム 2023 発表資料