

『日本語歴史コーパス 奈良時代編Ⅱ宣命』 テキストの凡例と「中納言」表示項目について

2020年3月31日 吳寧真

1. はじめに

『日本語歴史コーパス 奈良時代編Ⅱ宣命』は、池田・須永（2013, 2015）にて構築された「五国史」宣命コーパスを『日本語歴史コーパス』に合わせて調整したものである。「五国史」宣命コーパスは北川和秀（1982）『続日本紀宣命 校本・総索引』を底本として作成したコーパスであるため、本コーパスも『続日本紀宣命 校本・総索引』を底本としている。底本をコーパス化するにあたっては一部テキストを校訂し、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。

本コーパスは、漢字と万葉仮名によって表記されている原文を、独自の漢字仮名交じりテキストに変換して形態論情報を付与している。これは漢字が連続する文字列や訓読を前提とした漢文的な語順のままでは UniDic による形態素解析が困難であること、また『日本語歴史コーパス』の他のサブコーパスの検索結果と並べて表示した際に視認性が下がることなどの理由から、漢字仮名交じりテキストが必要となるためである。『日本語歴史コーパス』のサブコーパスの一つとして他のサブコーパスとの整合性を保ちつつ、極力原文に即した漢字仮名交じりテキストの作成を目指した。

一方で原文の表記は、日本語研究上欠かすことのできない情報であるため、本コーパスでは漢字仮名交じりテキストと同時に原文の表記も参照できるようにした。

この文書では本コーパスの漢字仮名交じりテキストの性質と、原文文字列の概要、検索アプリケーション「中納言」における表示項目について例を挙げながら述べる。

2. テキストの凡例

2. 1 漢字仮名交じりテキスト

本コーパスでは、漢字と万葉仮名表記の原文を漢字仮名交じり表記へと変換したものを本文テキストとして使用している。

宣命の原文は、基本は和文体だが、部分的に変体漢文や漢文が含まれる混淆文である。変体漢文や漢文の部分は、底本の振り仮名に従って読み下し、漢字仮名交じりテキストを作成した。¹

本文テキストを作成する際、2. 1. 1 から 2. 1. 7 までの作業を行った。

¹ なお、底本にて振り仮名情報がない漢文が少数存在するが、この部分については読み下さず、品詞情報「漢文」のみを付与した。

2. 1. 1 万葉仮名の処理

原文が万葉仮名である場合、平仮名に変換する。（〔 〕内は原文表記、以下同。）

【例】掛けまくも畏き天皇〔掛母畏支天皇〕

【例】中今に至るまでに〔中今至麻弓尔〕

2. 1. 2 補読の処理

底本では、原文に表示されない送り仮名・助詞・助動詞・敬語などを振り仮名で補読している。本コーパスの本文テキストはそれに従い平仮名を補っている。

【例】官々に仕へ奉る韓人部一二人に〔官々仕奉韓人部一二人尔〕

【例】自然に作成れる和銅〔自然作成和銅〕

【例】諸聞き食へと詔りたまふ〔諸聞食止詔〕

また、人名・地名に後接する格助詞「の」は、同名、同地でありながら、万葉仮名があるかどうかによって、「の」の表記がある場合とない場合がある。

【例】近江大津宮に〔近江大津宮〕

近江大津の宮に〔近江大津乃宮尔〕

【例】藤原永手朝臣〔藤原永手朝臣〕

物部浄之の朝臣〔物部浄之乃朝臣〕

2. 1. 3 返読の処理

原文が漢文的な語順で、訓読する際には字順を転倒させて読む場合、本文テキストはすべて返読処理をした。

【例】位有る人等に位一階給ふ〔有位人等給位一階〕

【例】神ながら所念し行す〔随神所念行〕

なお、本コーパスには 1 例だけ再読文字が見られたが、これについても同様に返読処理を行った。

【例】人の心未だ定らず〔人心未定〕

2. 1. 4 漢文的な語順による衍字の処理

漢文を和文に読み下す場合、一部の文字が衍字になることがある。特に次の【例】のような「動詞連体形＋格助詞＋名詞」の構造が散見される。この場合、助詞「之」は衍字で

あり、本文テキストでは表記しない。

【例】諂ひ欺く心無く忠に赤き誠を以て〔無諂欺^之心以忠赤^之誠〕

【例】天の授けぬを得て〔天乃不授^所乎得天〕

2. 1. 5 漢文的な語順における万葉仮名の衍字の処理

漢文を和文に読み下す場合、原文の「打消の「不」＋動詞」を「動詞＋打消の助動詞」に返読するが、打消の助動詞に当たる万葉仮名が同時に存在する構造が散見される。この場合、本文テキストでは打消にあたる要素は一回しか表記しない。

【例】進みも知らに退きも知らに〔進母^不知^尔退母^不知^尔〕

【例】荷重きは堪へじかと〔荷重波^不堪^自加止〕

2. 1. 6 踊り字の処理

原文には「々」「ゝ」の2種類の踊り字がある。

「々」については、漢字一字の繰り返しで、かつ前接文字と合わせて一語をなす場合（「人々」など）はそのまま「々」を用いて、それ以外は適宜繰り返されると想定される文字に置き換えた。なお、万葉仮名は平仮名に変換した。

【例】人々好からぬ謀を〔人々不好謀乎〕

【例】御称称りて緩ひ怠る事無く〔御称々而緩怠事無久〕

【例】天皇が御世御世聞し看し来る〔天皇御世々々聞看来〕

【例】よさし奉りしまにまに〔与佐斯奉志麻尔々々〕

「ゝ」は4例ある。これについてはすべて前接する万葉仮名に従って平仮名に変換した。

【例】思ほしつゝ大坐坐す〔思保之^ツゝ大坐坐〕

【例】所念しし位となも〔所念之^ゝ位止奈毛〕

2. 1. 7 句読点の処理

底本では、漢文には句点・読点の双方があり、和文には句点だけが使用されている。本コーパスの本文テキストでは、漢文では底本の句読点をそのまま使用し、和文では句の切れ目になる助詞、接続詞、動詞連用形などに読点を付与した。句点の付与は底本に従った。

【例】諸国々郡司、加位一階。〔諸国々郡司、加位一階。〕（※漢文）

【例】加以、元来風の病に苦びつつ、身体安らず。〔加以元来風病尔苦都々身体不安。〕

【例】受け賜り歡び、受け賜り貴び。〔受賜利歡受賜利貴〕

2. 2 原文文字列

2.1.1、2.1.2 で述べたように、本文テキストが平仮名である場合、原文が万葉仮名である可能性も、底本による補読である可能性もあるため、原文の確認が必要である。そこで、本コーパスでは漢字仮名交じりテキストと共に、漢字と万葉仮名表記の「原文文字列」を公開している。「中納言」上では原文 KWIC 機能を使って、キーだけでなく前後文脈の原文文字列情報を同時に参照することができる（詳しくは 3.5 に後述）。

底本には、基本以下の 6 種類の表記が見られる。

① 漢字

【例】現御神止大八嶋国所知天皇大命良麻止

② 漢字（小字）

【例】我皇太上天皇大前尔恐古土物（かしこじもの）

③ 万葉仮名

【例】天皇御子之阿礼坐牟弥繼繼尔

④ 万葉仮名（小字）

【例】現御神止大八嶋国所知天皇大命良麻止

⑤ 漢字表記の機能語

【例】天皇御子之（の）阿礼坐牟弥繼繼尔

⑥ 漢字表記の機能語（小字）

【例】頂伎恐美供奉生（つつ）

原文文字列では、大字の漢字・万葉仮名・漢字表記の機能語を底本どおりの表記にした。ただし、「中納言」の表示上、文字のサイズを調整することが難しいこと、また、サイズを変更しても、「中納言」の検索結果をコピーして貼り付ける場合も、検索結果のダウンロードをした EXCEL 上でも、文字サイズの大小が反映されず、実用性が低いことから、小字は再現していない。

3. 「中納言」における表示項目と内容

本コーパスの本文にはさまざまなタグや単語情報が付与されており、言語研究を目的とした利用に資するものとなっている。これらの情報は「中納言」上に検索結果として表示される。以下では「中納言」上の主な表示項目とその内容について説明する。

155 件の検索結果が見つかりました。

検索対象語数:20,017 記号・補助記号・空白を除いた検索対象語数:18,429

サンプル ID	開始位置	連番	コア	前文脈	キ	後文脈	語彙素読み	語彙素	語形	品詞	活用型	活用形	原文文字列	振り仮名	本文種別	話者	ジャンル	作品名	成立年	巻名等	作者	生年	底本	ページ番号	底本リンク	参考リンク
10-宣命 0797_01001	120	60	1	現御神と大八嶋国所知	天皇	が[大命らまと詔りたまふ[大命]を]、操り侍る[皇子[等]王[降]百官]の人[等]天	スメラ	皇	スメラ	名詞・普通名詞・一般			天皇	スメラ			宣命	続日本紀	797	第一詔 文武元年八月庚辰(巻一)文武天皇			続日本紀宣命校本・総索引	3		
10-宣命 0797_01001	860	540	1	と詔りたまふ。#高天原[こ]事[始めて]、[遠]天皇祖の[御世]中今に[至る]まで[に]、[大命良麻止詔大命乎集侍皇子等王等百官人等天下公民諸國食止詔。#高天原[事]始而遠天皇祖御世中今至麻呂尔	天皇	が[御子の]あわれ[坐]さむ[跡]継[継]に、[大八嶋]国[知ら]は[た]、[次]と、[天]つ	スメラ	皇	スメラ	名詞・普通名詞・一般			天皇	スメラ			宣命	続日本紀	797	第一詔 文武元年八月庚辰(巻一)文武天皇			続日本紀宣命校本・総索引	3		

【「中納言」 検索結果イメージ】

3. 1 形態論情報

「中納言」に表示される形態論情報は UniDic の見出しと対応するものである。基本的には『現代日本語書き言葉均衡コーパス』や『日本語歴史コーパス』の他のサブコーパスと同様の情報であるが、以下に利用に際しての注意点を挙げる。

3. 1. 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。語彙素の表記がわからない場合は語彙素読みで検索することも有用である。

3. 1. 2 語形

「語形」は、異語形を区別するレベルである。例えば語彙素「皇」に対する語形「スメラギ」と「スメロキ」との違いや、語彙素「忘れる」の語形に「ワスレル」（下二段 - ラ行）、「ワスル」（文語下二段 - ラ行）があるといった口語活用と文語活用の区別などがある。語形はカタカナによって表記される。

3. 1. 3 品詞

UniDic の設計に基づいた品詞が付与される。学校文法における「形容動詞」は、語幹に当たる部分が「形状詞」、活用語尾は「助動詞」に分割されている点に注意が必要である。

3. 1. 4 活用型

活用型には文語活用と口語活用があるが、本コーパスは口語活用がない。文語活用とし

て処理されているものには「文語上一段」のように「文語」が表示される。形容詞も同様に、文語活用のは「文語形容詞」と表示される。

3. 1. 5 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」と「補助」（形容詞等の補助活用）のほか「省略」（「持たり」のように活用語尾が省略されたもの）がある。

3. 1. 6 原文文字列

本コーパスでは第2節で詳述した原文情報が表示される。また「中納言」上で「原文 KWIC」欄を表示することで、キーに対応する原文文字列だけでなく、前後文脈も合わせて参照できる。

また「中納言」の「文字列検索」で対象の文字列を「原文」とすることで、原文の漢字と万葉仮名からの検索も可能となっている（下図）。

文字列検索

検索フォームで検索 検索条件式で検索 履歴で検索

検索文字列 検索する文字列を入力してください。 対象文字列 校訂本文 原文 結果表示単位 短単位 長単位 キー表示形式 語 文字列全体

検索文字列では、特別な意味を持つメタ文字として以下に挙げる文字を使用できます。

x	任意の文字列	例: 国立 x 所 → 「国立国語研究所」「国立社会保障・人口問題研究所」など
-	任意の1文字	例: 日 - 友好 → 「日米友好」「日中友好」「日韓友好」など
[]	文字クラス	例: 日本 [人国語] → 「日本人」「日本国」「日本語」
[]	文字クラスの否定	例: 日本 [人国語] → 「日本漢」「日本史」「日本は」「日本の」など

3. 1. 7 振り仮名

底本の振り仮名を、仮名遣いを含めてそのまま採用した。ただし、万葉仮名と補読の部分は平仮名に変換する処理を行ったため、漢字の部分にのみ振り仮名を付与した。

3. 2 本文情報

● 本文種別と話者

宣命は天皇による口頭で発布する文であるが、聞き手と会話するものではなく、勅命を読み上げるものであるため、一般的な会話文ではない。従って、本文種別情報（会話、引用など）を付与しない。また、話し手は天皇と想定されるが、通常の会話ではないため、話者情報の付与も行っていない。

3. 3 作品情報

● 巻名等

詔の冒頭に示されている「詔の番号」「発布年月日」「巻の番号」「発布した天皇」を表示した。

● 部

『続日本紀』における巻（巻一～巻四十）を表示した。宣命の発布がない巻（巻二、巻五など）はコーパスの本文テキストに入っていない。

3. 4 底本情報

● 底本・ページ番号

底本列には底本である『続日本紀宣命 校本・総索引』を表示し、ページ番号列には底本における当該箇所のページ番号を表示している。

3. 5 原文 KWIC 表示

原文 KWIC とは、キーと前後文脈の直下に、対応する原文文字列を表示することができる「中納言」の機能である。形態論情報欄の「原文文字列」列ではキーに対応する原文しか見られなかったのに対し、前後文脈を合わせて同時に参照できる点で有用なものである。

原文 KWIC は下図の以下の箇所にを入れることで表示／非表示を選択することができる。

また検索結果をダウンロードした際も、前後文脈を含めた原文文字列がダウンロード結果に含まれる。

155 件の検索結果が見つかりました。
検索対象語数: 20,017 記号・補助記号・空白を除いた検索対象語数: 18,429

サンプル ID	開始位置	連番	コア	前文脈	キー	後文脈	語彙読み	語彙	語形	品詞	活用型	活用形	原文文字列	振り仮名	本文種別	話者	ジャンル	作品名	成立年	巻名等	作者	生年	底本	ページ番号	底本リンク	参考リンク
10_宣命 0797_01001	120	60	1	現御神止大八嶋國所知す	天皇	ゆが大命らまとは別たまふ(大命を)し(應)り侍る皇子(等)王(等)百官(の)人(等)天	スメラ	皇	スメラ	名詞 普通名詞 一般			天皇	スメラ		宣命	続日本紀	797	第一詔 文武元年 八月庚辰 (巻一)文武天皇				続日本紀 宣命 校本・ 総索引	3		

3. 6 副本文表示

本コーパスは、漢文由来の熟語表記、地名と数詞に、主本文としての訓読と、副本文としての音読を付与した。検索する際、下図のように、「副本文を検索対象に含む／副本文を検索対象に含まない」を選択することができる。

「副本文を検索対象に含む」を選択すると、直下部の「列の表示」「コーパス情報」において「層」「主本文」「多重化情報」のチェックボックスにチェックが入る。

検索結果の「主本文」列に「1」と表示される場合、検索結果は主本文のものであり、「0」と表示される場合は副本文のものである。

検索対象 [?] 設定を隠す

検索対象を選択 検索対象をクリア

全て

検索動作 [?] 設定を隠す

文脈中の区切り記号 [?] | 文脈中の文区切り記号 [?] # 前後文脈の語数 [?] 20

副本文 [?] 副本文を検索対象に含む 共起条件の範囲 [?] 文境界をまたがない

▼ 副本文を検索対象に含む

▼ 副本文を検索対象に含まない

列の表示 [?] 設定を隠す

コーパス情報

時代名 サブコーパス名 サンプル ID 開始位置 連番 コア 層 層内連番 主本文 多重化種別

なお、主本文には複数の短単位（または長単位）から構成されている場合があるが、「中納言」での検索画面においては、複数の単位のあるうちの最後の語の形態論情報のみが代表して表示される。そのため、当該箇所すべての語を確認したい場合は、検索結果の画面のサンプル ID をクリックし、詳細な文脈情報を表示させることで確認できる。

サンプル ID	開始位置	連番	コア	層	主本文	多重化種別	前文脈	キー	後文脈	語彙	語彙読み	語形	品詞	活用型	活用形	原文文字列	振り仮名	本文種別	話者	ジャンル	作品名	成立年	巻名等	作者	生年	底本	ページ番号	底本リンク	参考リンク
10-宣命 0797_04003	1070	700	1	0		0	天皇、[丁酉]の八月、[此の]此の[天]の[下]の[御]を、[旧]並所知皇太子の	嬪子	[今]御宇[此]の[天]の[下]を	チャクシ	嬪子	チャクシ	名詞-普通名詞-一般			嬪子	ムカヒメバラニコ	主本文	宣命	続日本紀		797	第三詔 慶雲四年七月壬子(巻四)元明天皇			続日本紀 宣命 校本・総索引	6		



日本語辞典コーパス (H.I.)

詳細な文脈情報

サンプル ID	開始位置	連番	コア	層	主本文	多重化種別	前文脈	キー	後文脈	語彙	語彙読み	語形	品詞	活用型	活用形	原文文字列	振り仮名	本文種別	話者	ジャンル	作品名	成立年	巻名等	作者	生年	底本	ページ番号	底本リンク	参考リンク
10-宣命 0797_04003	670	1					日知所知	ヒナメシ	ヒナメシ	名詞-固有名詞-人名-一般										日知所知	ヒナメシ								
10-宣命 0797_04003	680	1					皇太子	ミコ	御子	名詞-普通名詞-一般										和	皇太子	ミコ/ミコト							
10-宣命 0797_04003	680	1					皇太子	ノ	の	助詞-格助詞										和	皇太子	ミコ/ミコト							
10-宣命 0797_04003	680	1					皇太子	ミコト	尊	名詞-普通名詞-一般										和	皇太子	ミコ/ミコト							
10-宣命 0797_04003	680	1					の	ノ	の	助詞-格助詞										和	之								
10-宣命 0797_04003	700	1					嬪子	ムカヒメ	嬪妻	名詞-普通名詞-一般										和	嬪子	ムカヒメ/バラニコ							
10-宣命 0797_04003	700	1					嬪子	ハラ	嬪	名詞-普通名詞-一般										和	嬪子	ムカヒメ/バラニコ							
10-宣命 0797_04003	700	1					嬪子	ノ	の	助詞-格助詞										和	嬪子	ムカヒメ/バラニコ							
10-宣命 0797_04003	700	1					嬪子	ミコ	御子	名詞-普通名詞-一般										和	嬪子	ムカヒメ/バラニコ							
10-宣命 0797_04003	710	1					,	,	,	補助記号-読点																			
10-宣命 0797_04003	720	1					今	イマ	今	名詞-普通名詞-副詞可能										和	今	イマ							
10-宣命 0797_04003	730	1					御宇	アメ	天	名詞-普通名詞-一般										和	御宇	アメ/シタラシメ							

閉じる

Copyright © National Institute for Japanese Language and Linguistics.

参考文献

- 池田幸恵・須永哲矢（2013）「『五国史』 宣命のコーパス化」『第4回コーパス日本語学ワークショップ予稿集』 pp. 187-194
- 池田幸恵・須永哲矢（2015）「『五国史』 宣命コーパスの設計とその利用」『訓点語と訓点資料』 134、訓点語学会、pp. 80-98
- 小木曾智信（2016）「多重の読みをもつテキストのコーパス化」『言語資源活用ワークショップ 2016 発表論文集』 pp. 159-162
- 小木曾智信・岡照晃・中村壮範・八木豊（2017）「『日本語歴史コーパス』における原文 KWIC 表示機能の実装」『言語資源活用ワークショップ 2017 発表論文集』国立国語研究所、pp. 252-257