

『日本語歴史コーパス 奈良時代篇Ⅲ祝詞』(短単位データ 1.0 / 長単位データ 1.0)

テキストの凡例と「中納言」表示項目について

2022年3月31日 間淵洋子・近藤明日子

1. はじめに

『日本語歴史コーパス 奈良時代編Ⅲ祝詞』は、10世紀に編纂された律令の施行細則をまとめた法典(「格式」)の一つである『延喜式』(927年成立、967年施行)全50巻のうちの、巻第八に収められた27の「祝詞」(以下「延喜式祝詞」)を収録したものである。

延喜式祝詞は現存する最古の祝詞で、奈良朝以前から平安初期までに成立したとされている詞章が伝承されることで、文体、表記、語彙、語法など様々な面で古代の日本語を残す。残存する和文資料の少ない上代・中古初期の日本語を知るための、貴重な資料の一つと考えられている。

延喜式祝詞の原文は、表音仮名として機能する万葉仮名を含め、全編が漢字で構成されているが、本コーパスの本文は、『日本語歴史コーパス(CHJ)』の他のサブコーパスと統合的な形態論情報(=形態素解析辞書 UniDic の体系に基づく形態論情報)を付与するために、先行して公開した『CHJ 奈良時代編 I 万葉集』『同 II 宣命』と同様、漢字仮名混じり文に改めている。この形態素解析用の本文テキストは、沖森卓也編(1995)『東京国立博物館蔵本 延喜式祝詞総索引』(古典研究会)の「訓讀文」を底本とし、ほぼそれをそのままの形で利用した。一方、訓讀前の原文の表記についても、日本語研究上欠かすことのできない情報であることから、コーパス本文テキストと共に参照できる形で提供することとした。

この文書では、コーパス本文の成り立ちと、検索アプリケーション「中納言」における検索結果の表示項目について、例示しながらその概要を示す。

2. テキストの凡例

2.1 テキストに使用する文字

本コーパスの本文テキストに使用した文字の範囲は、JIS 第1水準から第2水準までの漢字を原則とし、JIS 第2水準までで入力できない文字は、Unicode で表現できる範囲で底本を忠実に文字化した(図1参照¹⁾)。

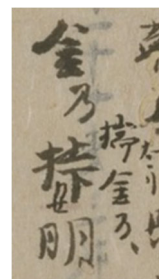


図1 例1 原資料画像

- 例1) 金の櫛・金の栳、
(10-祝詞 0927_00004, 7210, 龍田風神祭)²
※「櫛(U+692F)」は JIS 外字

¹ 以下、底本原資料画像は「e 国宝」(<https://emuseum.nich.go.jp>)による。

² 以下、出典表示は「サンプル ID, 開始位置, 祝詞名」のように示す。

2.2 原文－底本間の校訂

コーパスの本文テキストの底本とした沖森（1995）の「訓讀文」それ自体の原資料は、『延喜式』の現存する最古の写本とされる平安中期書写、東京国立博物館蔵『延喜式』（九條家本）である。この写本には墨点による訓点（ヲコト点、仮名点等）が付されているが、沖森（1995）の「訓讀文」は、これらの訓点を生かしつつも、「できるだけ、奈良時代風に訓讀することを旨とし」て作成されており、必ずしも訓点の読みとは一致しない。そこで、本コーパスでは、原資料の訓点を除く本行本文を、コーパス本文テキストの原文と位置付けて、原文から底本への校訂を詳細に情報付けした。以下に、原文－底本間でなされた校訂を示すことで、コーパス本文と原文との関係を示す。

2.2.1 万葉仮名の展開

原文が万葉仮名である場合、現行の平仮名に展開している。

例 2) 本文：天皇が朝庭にいかしやぐはえの如く仕へ奉り、さかえしめ賜へと
原文：天皇我朝庭爾伊加志夜久波叡乃如久仕奉利佐加叡志米賜止

(10-祝詞 0927_00002, 4310, 春日祭)

2.2.2 返読と漢字の展開

原文が漢文的字順である場合、訓読に合わせて順番を入れ替えている（返読）。また、訓読時に助動詞等機能語に相当する助字は、平仮名に展開している。

例 3) 本文：番繩の類之を綱根と謂ふ。

原文：語番鼠之類謂之綱根

(10-祝詞 0927_00008, 4310, 大殿祭)

例 4) 本文：豊かに榮えしめ給ひ、

原文：豊爾令榮給比

(10-祝詞 0927_00019, 2200, 伊勢大神宮 六月月次祭)

2.2.3 補読

原文には表示されていない機能語、送り仮名などを補読している。

例 5) 本文：凡そ祭祀の祝詞は

原文：凡祭祀祝詞

(10-祝詞 0927_00000, 160, 延喜式祝詞)

2.2.4 踊り字の展開・置換

原文の踊り字が語頭に位置する場合や、語（文字列）全体を繰り返すのに用いられている場合は、繰り返されると想定される文字に展開している。その際、万葉仮名は平仮名に改める。また、前接文字と併せて1語をなす場合は、現行の繰り返し符号に改めている。

例 6) 本文：天の下の公民の作る作り物を

- 原文：天下乃公民乃作_レ物乎 (10-祝詞 0927_00004, 2580, 龍田風神祭)
- 例 7) 本文：品々の幣帛
- 原文：品_レ幣帛 (10-祝詞 0927_00004, 4260, 龍田風神祭)

2.2.5 誤りの訂正

文脈上、原文の誤りが明らかである場合（誤字、脱字、衍字、転倒）は改めている。なお、万葉仮名の誤りである場合には、本来当てるべき変体仮名を補記した形で原文を表記している。

- 例 8) 本文：鎮火祭
- 原文：鎮火祭祭 (10-祝詞 0927_00012, 30, 鎮火祭)
- 例 9) 本文：奥つ御年を、八束穂の
- 原文：奥都御年牟_レ八束穂乃 (10-祝詞 0927_00001, 16920, 祈念祭)

2.3 底本－本文テキスト間の校訂

本文テキストは底本である沖森（1995）の「訓讀文」をほぼそのままの形で用いているが、一部に、形態論情報を付与するにあたって解釈、認定が困難になるために底本を改変した箇所が1箇所ある³。以下に該当例を示す（抹消線で見せ消ちにした文字が沖森本文、下線で示したのがコーパスの校訂本文）。

- 例 10) 本文：ち別きにち別~~ち~~きて (10-祝詞 0927_00010, 3280, 六月晦大原祓)

2.4 原文文字列

本コーパスでは、コーパス本文テキストに対する原文文字列の情報を提供する。

2.2 節に示した通り、コーパス本文は原文文字列からのさまざまな校訂を経て形成されているが、上代の日本語を分析する上では、万葉仮名や漢文的字順の使用など、表記が重要な研究対象となる。そこで、2.2 節に示した校訂前の原文文字列を、コーパス検索アプリケーション「中納言」で利用する際に「原文文字列」「原文 KWIC」（小木曾・岡・中村ほか 2018）という項目で確認できるよう整備した。

³ この改変については、コーパス本文－原文間の校訂にあたらなため、「中納言」での検索結果のいずれにも現れず、原文での検索にも用いることができない。

ただし、延喜式祝詞の特徴でもある宣命書き（活用語尾や機能語などを小書きする表記スタイル）については、文字サイズの情報を検索結果に含めて提供することが難しく、また実用性にも乏しいため、再現しないこととした。そのため、以下のような原文での差異は、コーパスの原文情報からは得られない。

- 例 11) 原文：宇豆乃幣帛
 本文：宇豆乃幣帛 (10-祝詞 0927_00001, 10060, 祈念祭)
- 例 12) 原文：宇豆(改行)乃幣帛
 本文：宇豆乃幣帛 (10-祝詞 0927_00002, 5440, 祈念祭)

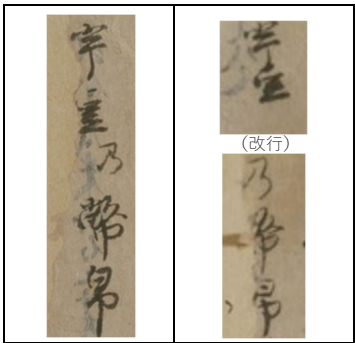
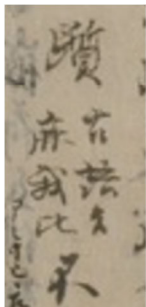


図 2 例 11,12 原資料画像

2.5 その他

底本原資料では、2.4 節に示したいわゆる「宣命書き」のほかに、割書き（本行中に、1 行を 2 行分に分割して小書き文字で注などを書き入れる表記法）がある。本コーパス内での割書きは、概して古言や異文に関する注記であり、祝詞本文と区別されるべきテキストであるため、「割書き」であることをマークアップした。この情報は、「中納言」で利用する際に参照することができる。



- 例 13) 手躓足躓古語云麻我比不令爲互
 (10-祝詞 0927_00008, 10060, 大殿祭)

図 3 例 13 原資料画像

3. テキストの範囲とサンプル

延喜式卷第八は、内題、序、各祝詞（27 編）、尾題から構成される。本コーパスには、これらの文字列が漏れなく収録されている。中心となる各祝詞は、それぞれ成立時期なども異なる個別の文章であるため、それぞれ個別の「サンプル」とした。

「サンプル」は、サンプル ID という個別に識別される ID を持つ。サンプル ID は、本コーパスにおいては以下の通り構成される。各祝詞の別は、このうちサンプル ID の下 5 桁（11-15 桁目）に、先頭から順に「00001」「00002」と連番を振って示した。

表 1 本コーパスにおけるサンプル ID の構成

1-2 桁		3 桁		4-5 桁	6-9 桁	10 桁	11-15 桁
時代通し番号		ジャンル		作品 ID	成立時期	区切り記号	作品内での通し番号
10	奈良	-	なし	祝詞	0927	-	祝詞連番 5 桁

一方、内部の祝詞以外の部分、すなわち、内題・序・尾題は、巻全体の骨格に相当する部分として、全体をひとまとめにして 1 サンプルとし、サンプル ID 下 5 桁を「00000」とした。これらの各要素とサンプルの関係を示すと以下ようになる。「00000」で終わるファイルについては、各祝詞本文とは異なる属性を持つため、留意が必要である。

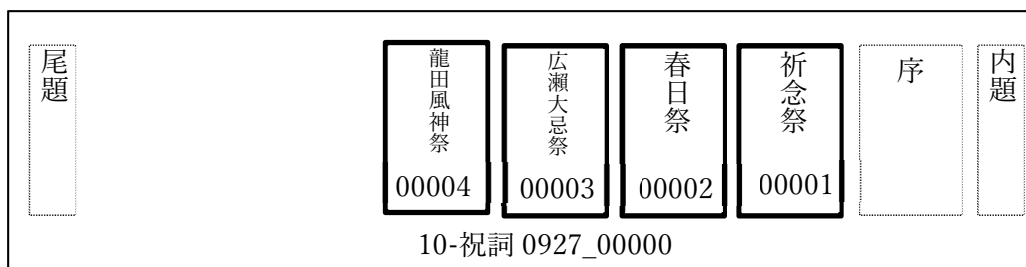


図4 延喜式卷第八の構成とサンプルの関係

なお、『日本語歴史コーパス』の各コーパスのサンプルは、人手により形態論情報を確認・修正することで、一定精度を担保する「コア」サンプルと、十分には修正が行われていない「非コア」サンプルの2種に分けられるが、本コーパスは、全て人手による修正を経た「コア」サンプルである⁴。

4. 「中納言」での検索方法と検索結果の表示内容

4.1 検索対象の設定

「中納言」では「検索対象」欄で「検索対象を選択」ボタンをクリックすると、「検索対象の選択」画面が開き、サブコーパスや本文種別等で検索対象を絞り込むことができる（図5）。

本コーパスは、全て「コア」サンプルであるため、検索対象には「コア」のみを選択できる。また、作品は「延喜式祝詞」のみである。

本文種別は、2.5節に示した通り「割書き」を弁別しているため、「割書き」とそれ以外の「地の文ほか」を選択して検索することができる。



図5 検索対象の選択画面

4.2 副本文の検索

本コーパスは、先行公開された『CHJ 奈良時代編Ⅱ 宣命』に倣い、単一の本文に対して短単位形態論情報を複数重複して付与した箇所がある（形態論情報の多重化については、小木曾 2017 を参照）。

まず、図6上段「諸王」の用例で示すように、底本原資料の漢語に相当する漢字列に対して、底本で振り仮名等により複数短単位からなる和語の読みを指定している場合、本文の読みに対応する和語を「主本文」、本行の漢語を「副本文」と位置付けて、二種類の形態論情報を付与している。

次に、図6下段「晦」の用例で示すように、底本原資料の漢字列が漢語には相当しないが、底本で振り仮名等により複数短単位からなる読みを指定している場合、読みの短単位を「主本文」と位置づけ形態論情報を付与している。この場合、コーパスの仕様上「副本文」を設定する必要があり、品詞「対象語無し」の未知語として形態論情報を付与している。

⁴ サンプリングによる短単位情報の精度（適合率）は99.9%である。

原資料	底本	本文種	短単位数	コーパスの形態論情報抜粋 (語彙素/語彙素読み/語種/品詞大分類)	
				大君/オオキミ/和/名詞-普通名詞-一般	達/タチ/和/接尾辞-名詞的-一般
諸王	おほきみたち	主本文	2	諸王/ショオウ/漢/名詞-普通名詞-一般	
		副本文	1		
晦	ついでりの	主本文	2	晦/ツゴモリ/和/名詞-普通名詞-一般	の/ノ/和/助詞-格助詞
		副本文	1	///対象語無し ⁵	

図 6 形態論情報の多重化例

本コーパスを「中納言」で利用する際、初期状態ではこのうちの「主本文」のみが検索対象となるが、「検索動作」の「副本文」項の設定により、副本文の形態論情報を含めて検索対象とすることができる(図 7)。ただし、「副本文を検索対象に含む」を選択した場合には、前方・後方共起条件を指定しての検索ができない点に注意が必要である。

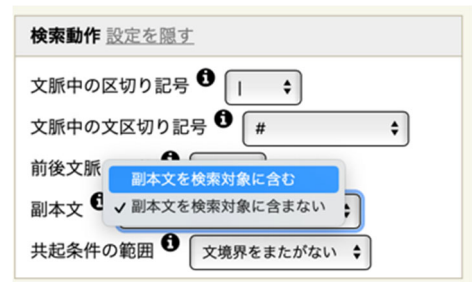


図 7 検索動作の設定画面

なお、長単位は短単位の主本文に対して形態論情報を付与しており、形態論情報の多重化は行っていない。

4.3 検索結果表示の内容

本コーパスは、言語研究を目的とした利用に資する、さまざまな情報を有している。これらの情報は、コーパス検索アプリケーション「中納言」の検索結果表示画面において参照可能な形で提供される(図 8、9)。

サンプルID	開始位置	連番	コア	前文脈	キー	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	語者	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	参照リンク
10-祝詞 0927_00001	1120	750		1 つ社と稱へ御意へ奉る(皇神等の)前に(白さく)、(今年)の(二月)の(御年初め)	福は	(む)と(して)、(皇)御(孫)の(命)の(う)つ(の)幣(角)を、(一)朝(日)の(露)逆(登)り(に)。	タマウ	給う	タマフ	動詞-非自立可能	文語-四段-八行	未然形-一般	福	たま		祝詞	延喜式祝詞	927	祈念祭				延喜式祝詞総索引	12	中納言	
10-祝詞 0927_00002	750	450		1 天之子八根命・比賣神、(四柱)の(皇神等)の(廣)前に(白)さ	福ひ	(の)任(に)、(春)日(の)三(重)の(山)の(下)つ(石)根(に)宮(柱)廣(知)り(立)て、(高)天(の)原(に)。	タマウ	給う	タマフ	動詞-非自立可能	文語-四段-八行	連用形-一般	福比	たま		祝詞	延喜式祝詞	927	春日祭				延喜式祝詞総索引	79	中納言	

図 8 「中納言」の検索結果表示画面(初期状態)

⁵ 品詞「対象語無し」の短単位は語彙素・語彙素読み・語種は空値となる。

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	語彙集	語彙集	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	参考文献リンク
10-祝詞 0927_00019		3090	2050	1	10	1	振り仮名	、横山の如く(置き)足成して、(大)中臣(大)玉串に(應)れ侍りて、(今)年の(六)月(四)日	十 と ま か り 十	トオ	+	トオ	名詞-数詞		十	とをかま			祝詞	延喜式祝詞	927	伊勢大神宮 六月月次祭			延喜式祝詞総索引	436	西国宝	
10-祝詞 0927_00019		3090	2050	1	0	0	振り仮名	、横山の如く(置き)足成して、(大)中臣(大)玉串に(應)れ侍りて、(今)年の(六)月(四)日	十 と ま か り 十	ジュウ	+	ジュウ	名詞-数詞		十	とをかま			祝詞	延喜式祝詞	927	伊勢大神宮 六月月次祭			延喜式祝詞総索引	436	西国宝	

図 9 「中納言」の検索結果表示画面（副本文を含めた検索時）

表 2 に、「中納言」の検索画面で参照できる付加情報のうち、初期設定で表示される項目について、その内容を示す⁶。

表 2 「中納言」検索結果の主な表示項目

情報種別	項目名	内容
コーパス情報	サンプル ID	検索対象が含まれるサンプルの ID (→3 節)。
	開始位置	「キー」の本行テキストの先頭文字の、サンプル内における位置を表す ID。10 きざみの連番。
	連番	「キー」に対応する主本文の短単位の、サンプル内における位置を表す ID。10 きざみの連番。
	コア	検索対象の含まれるサンプルが「コア」(→3 節)であることを表す。「1」がコア、「0」が非コアを表す。本コーパスは全て「1」(コア)である。
	主本文*	検索結果として表示されている形態論情報の付与されている本文の主副の別 (→4.2 節)。「1」が主本文、「0」が副本文を表す。
	多重化種別*	形態論情報の多重化 (→4.2 節)を行っている箇所、その多重化の種別を表す。本コーパスでは全て「振り仮名」である。
形態論情報 ⁷	前文脈	検索対象の前方文脈。コーパス本文の下に、
	キー	検索対象の書字形出現形 (表記形)。
	後文脈	検索対象の後方文脈。
	原文 KWIC	上記項目「前文脈」「キー」「後文脈」に対する、校訂前の原文に近い形のテキスト (→2.4 節)。ただし、ルビは校訂後のテキストが表示される。

⁶ 副本文を含めた検索時に表示される項目には「*」を付した。なお、この際表示される「層」の列は、データ構築上の便宜的なラベルであり、コーパスの内容に関わらないため、表 2 には含めない。

⁷ 形態論情報の個々の項目の内容は、「前文脈」「キー」「後文脈」「原文 KWIC」「原文文字列」「振り仮名」を除き、UniDic の見出しに対応している。形態論情報の詳細については、国立国語研究所言語変化研究領域(呉寧真・近藤明日子)編 (2022) を参照されたい。

情報種別	項目名	内容
	語彙素読み	検索対象の語彙素（下記項目「語彙素」参照）の読み。片仮名表記である。
	語彙素	検索対象の語彙素の表記。語彙素は、単語の様々なバリエーション（語形、活用形、表記形など）を統合した辞書の見出しに相当するもので、一般の和語・漢語は漢字平仮名表記、外来語・人名・地名は片仮名表記である。
	語形	検索対象の語形。語形は、語彙素では統合される、語形の別（例：語彙素「矢張り」に対する「ヤハリ」「ヤッパリ」など）や活用形の別（例：語彙素「読む」に対する「ヨム（五段-マ行）」「ヨム（文語四段-マ行）」「ヨメル（下一段-マ行；可能動詞形）」など）等を区別した語の個々の形に相当するもので、片仮名表記である。
	品詞	検索対象の品詞で、UniDicの体系に基づく。学校文法における「形容動詞」は、語幹が「形状詞」、活用語尾が「助動詞」に分割される点に注意が必要である。 UniDicの体系に基づかない品詞として「対象語無し」（→4.2節）がある。「対象語無し」の付与された短単位は、形態論情報に関する項目のうち、「前文脈」「キー」「後文脈」「原文 KWIC」「品詞」「原文文字列」「振り仮名」以外は空欄となっている。
	活用型	検索対象活用語の活用の型。口語活用は活用の型と行で「五段-サ行」のように示されるが、文語活用は「文語」が加わり「文語四段-サ行」のように示される。本コーパスでは全て文語活用である。
	活用形	検索対象活用語の活用形。
	原文文字列	検索対象の校訂前本文（原文）の文字列（→2.4節）。
	振り仮名	検索対象に付された振り仮名の文字列。
本文情報	本文種別	検索対象の含まれる文が「地の文」以外の場合の、その種別。本コーパスでは「割書き」（→2.5節）のみを表示する。
	話者	本コーパスでは用いない。
作品情報	ジャンル	検索対象が含まれるサンプルのジャンル。全て「祝詞」とした。
	作品名	全て「延喜式祝詞」とした。
	成立年	『延喜式』自体の成立年「927」年とした。
	巻名等	祝詞名を示す。
作者情報	作者	作者不明であるため空欄である。
	生年	全て空欄である。

情報種別	項目名	内容
底本情報	底本	沖森卓也編（1995）『東京国立博物館蔵本 延喜式祝詞総索引』（古典研究会）を略記し「延喜式祝詞総索引」とした。
	ページ番号	底本原資料の東京国立博物館蔵『延喜式』（九條家本）は卷子本であるためページ情報の代わりに行番号を記載した。
その他	底本リンク	検索対象の原資料画像へのリンク。国立文化財機構文化財高精細画像公開システム「e国宝」の東京国立博物館蔵『延喜式』へのリンクを「 e国宝 」ボタンで示す。
	参照リンク	検索対象の底本以外の参照本画像へのリンク。国立歴史民俗博物館所蔵の土御門家本『延喜式』へのリンクを「 歴博 」ボタンで示す。

謝辞

本コーパスは、国立国語研究所基幹研究プロジェクト「通時コーパスの構築と日本語史研究の新展開」、および、人間文化研究機構広領域連携型基幹研究プロジェクト「異分野融合による「総合書物学」の構築」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」による成果の一部である。

参考文献

- 小木曾智信(2017)「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』国立国語研究所、pp.159-162
- 小木曾智信・岡照晃・中村壮範・八木豊(2018)『『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
- 沖森卓也編（1995）『東京国立博物館蔵本 延喜式祝詞総索引』古典研究会
- 国立国語研究所言語変化研究領域(呉寧真・近藤明日子)編（2022）『『日本語歴史コーパス奈良時代編』形態論情報規程集 Ver.2.0』https://clrd.ninjal.ac.jp/chj/doc/morph_nara_v2_0.pdf

参考 URL

- UniDic <https://unidic.ninjal.ac.jp/>
- コーパス検索アプリケーション「中納言」 <https://chunagon.ninjal.ac.jp/>
- 『日本語歴史コーパス』 <https://clrd.ninjal.ac.jp/chj/>