

「日本語歴史コーパス 奈良時代編Ⅲ祝詞」の公開

まぶち ようこ 間淵 洋子（和洋女子大学・元国立国語研究所）・おぎ そ としのぶ 小木曾 智信・たかだ ともかず 高田 智和（国立国語研究所）

1. はじめに

国立国語研究所（以下、国語研）では、既に『日本語歴史コーパス 奈良時代編』として公開している『万葉集』[1]（以下「奈良時代編Ⅰ万葉集」）および『続日本紀』「宣命」[2]（以下「奈良時代編Ⅱ宣命」）に続く資料として、『延喜式』収録の「祝詞」（以下「延喜式祝詞」）を加え、2021年3月、コーパス検索アプリケーション「中納言」により公開した。本稿では、この『日本語歴史コーパス 奈良時代編Ⅲ祝詞』（以下「奈良時代編Ⅲ祝詞」）について、資料的意義、底本、データ内容等について報告・解説する。

2. 「延喜式祝詞」の位置付け

『延喜式』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」）で、延喜5（905）年8月、醍醐天皇より命を受けた藤原時平により編纂が開始された。延喜に撰修を始めたことに『延喜式』の名前の由来があるが、その完成は22年後の延長5（927）年、施行は康保4（967）年である。全50巻中、巻第一から第十までが神祇祭祀に関する規定を収めたもので、そのうちの巻第八が「祝詞」の巻となっている[3]。

「祝詞」は神前に奏上する際の独特の文体を持つ言葉であるが、「延喜式祝詞」は現存する最古の祝詞とされ、27の祝詞を所載する。それぞれの祝詞の成立は祭祀の開始時期により奈良朝以前から平安初期までとされるが諸説あり[4][5]、安易に上代日本語資料として扱うことには注意が必要であるものの、祭祀儀礼に関わるその性質上、古い形が伝承され、文体、表記、語彙、語法など様々な面で上代の日本語を残している。表記法として、いわゆる「宣命書き」（体言や用言の語幹を通常の文字大で、助詞・助動詞や活用語尾などを万葉仮名の小書き右寄せ、または2行の割書で表記する独特の表記法）が用いられている点などが特徴的である。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、貴重な資料の一つであると言える。

そこで、先行して公開している「奈良時代編Ⅰ万葉集」「奈良時代編Ⅱ宣命」との比較対照なども企図し、「延喜式祝詞」を『日本語歴史コーパス 奈良時代編』として追加することとした。

3. コーパスの概要

3.1 コーパス本文の底本

本コーパスの本文には、漢字と万葉仮名からなる原文を読み下して漢字仮名交じり文に改め、句読点を付与し、振り仮名により読みを指定した沖森（1995）の「訓讀文」を用いた。『延喜式』の古

写本は多く、九條家本、卜部兼永自筆本、卜部兼右自筆本などが知られるが、沖森（1995）が底本とした東京国立博物館蔵『延喜式』（九條家本）は、平安中期書写とされ現存する最古の写本であり、巻第八祝詞を含め、『延喜式』全 50 巻中 27 巻が残存する。鎌倉時代初期の加點と見られる仮名点、ヲコト点¹が施されているが、沖森（1995）の訓読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」て作成されている。よって、『日本語歴史コーパス 奈良時代編』として構築するコーパスの本文として最適な訓読文としてこれを選定した。

なお、東京国立博物館蔵『延喜式』は、国立文化財機構が提供する Web コンテンツ「e 国宝」（<https://emuseum.nich.go.jp>）により高精細画像を閲覧することができ、原典への参照性の高さにも特長がある。

3.2 コーパスの形態論情報

「奈良時代編Ⅲ祝詞」の語（短単位[6][7]）の分割や語法については、例(1)格助詞「つ」、例(2)尊敬の助動詞「す」を認めるなど、奈良時代編の先行公開資料「奈良時代編Ⅰ万葉集」「奈良時代編Ⅱ宣命」に準じた。

(1) 天つ社・國つ社と稱へ辭竟へ奉る皇神等の前に白さく、(10-祝詞 0927_00001, 780、祈念祭)¹

(2) 八束穂のいかし穂に皇神等の依さし奉らば、(10-祝詞 0927_00001, 2270、祈念祭)

『日本語歴史コーパス』の構築にあたっては、時代・資料に適した形態素解析辞書を開発し解析を行っているが[8][9]、「奈良時代編Ⅲ祝詞」の構築においても、「奈良時代編Ⅰ万葉集」「奈良時代編Ⅱ宣命」を学習データとして用いた上代資料用の解析辞書を作成し形態素解析を行い、その上で全編にわたって人手による修正を加えた。その際、語認定に係る解釈が困難な場合、また沖森（1995）の訓読文に疑念がある場合などには、他本の訓読、翻刻[10][11][12]を参照した。

なお、「奈良時代編Ⅱ宣命」における形態論情報付与[13]に倣い、コーパス本文の底本訓読文において、漢語とみなせる漢字列に訓読を施した箇所に対しては、形態論情報の多重化[14]の処理を行った。

原典	訓読	コーパスの形態論情報抜粋（語彙素/語彙素読み/語種）			
		本文種	単位数	形態論情報(1)	形態論情報(2)
		主本文	2	大君/オオキミ/和	達/タチ/和
			1	諸王/ショオウ/漢	

図1 原典本文、訓読文、コーパスの形態論情報の関係²

¹ 用例中の下線は著者。出典は「サンプル ID, 開始位置、巻名等 (=祝詞名)」の形式で示す。

² 原典本文画像は「e 国宝」（<https://emuseum.nich.go.jp>）、訓読文画像は沖森（1995）による。

具体的には、図1に示すように、原典本文「諸王」に対して訓読文で「おほきみたち」の振り仮名が施された箇所などに対して、振り仮名に基づく主本文「大君(オオキミ)-達(タチ)」(和語)に加えて、漢字列を字音で読んだ際の「諸王(ショオウ)」を副本文として二つの形態論情報を付与したものである。形態論情報の多重化は、漢語の訓読箇所の他に数表現(原典本文「百八十六」に対して「ももやそまりむ」の振り仮名がある場合など)に対しても行った。

3.3 データ種別とデータ量

今回公開のデータは短単位データのみで、コーパス検索アプリケーション「中納言」を介して利用できる。記号類を除く主本文の語数は、延べ9,164語、異なり1,035語、副本文の語数は、延べ34語、異なり21語である。表1に語種別・品詞別の言語量を、図2に主本文の語種比率・品詞比率(記号を除く)を示す。

表1 語種別・品詞別の言語量

① 語種別言語量

種別	主本文				副本文
語種	和語	漢語	固有名詞	記号	漢語
延べ	8,862	61	241	1,451	34
異なり	885	41	109	3	21

② 品詞別言語量

種別	主本文														副本文	
品詞	名詞	代名詞	動詞	形容詞	形状詞	副詞	感動詞	接続詞	助詞	助動詞	接頭辞	接尾辞	補助記号	接頭辞	名詞	
延べ	3,338	93	1,876	164	13	36	8	8	3,011	266	180	171	1,451	2	32	
異なり	656	14	251	27	9	12	1	3	21	17	10	14	3	1	20	

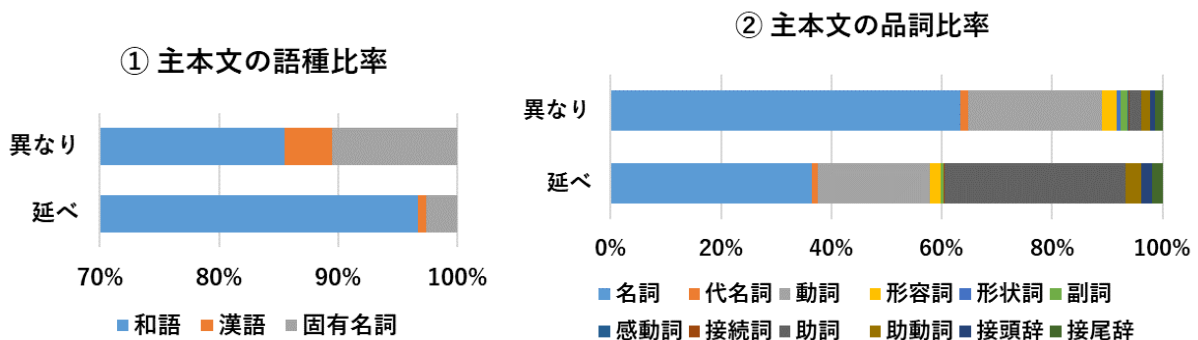


図2 主本文の語種比率・品詞比率(記号を除く)

併せて、語種比率・品詞比率について奈良時代編の他資料と比較を行うと、図3に示すように、語種比率に関しては他資料と同様だが、品詞比率に関しては、名詞比率の高さと助動詞比率の低さに特徴が見られ、文体的な特殊性が見て取れる。

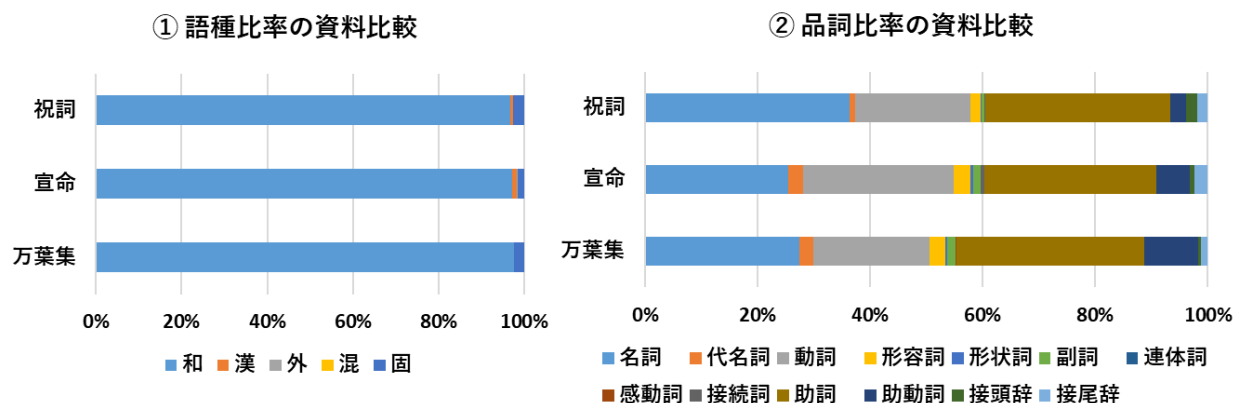


図 3 語種比率および品詞比率の資料比較

4. 検索アプリケーション「中納言」による利用

4.1 さまざまな検索方法

「奈良時代編Ⅲ祝詞」は、インターネット上の検索アプリケーション「中納言」を通じて公開する。「中納言」で提供される他のコーパスと同様、形態論情報を用いた検索（2021 年 5 月現在、短単位検索のみ）、文字列検索、位置検索が可能である。特に、形態論情報を用いた検索では、通常の出現形（「中納言」では「出現書字形」）での検索のほか、様々な表記・発音バリエーションを統一的に検索できる「語彙素」「語彙素読み」による検索、品詞や語種、活用語の活用型や活用形による検索なども可能である。

また、3.2 節に示した通り、「奈良時代編Ⅲ祝詞」は一つの文字列に対して多重に形態論情報を付与した箇所があるため、副本文を検索対象に含めて検索をすることができる。

短単位検索

検索フォーム

前方共起条件の追加

キー: -- 1 語 キーの条件を指定しない

語彙素 が 親王 短単位の条件の追加

後方共起条件の追加

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

ダウンロード 副本文を検索対象に含まない

検索

図 4 副本文を検索対象に含めた検索対象の設定

4.2 原文の参照

『日本語歴史コーパス』の「中納言」では、コーパス本文に校訂本文を用いた場合、その原文の情報として、「原文文字列」と原文 KWIC 表示を提供している。「奈良時代編Ⅲ祝詞」では、コーパス本文として用いた沖森（1995）の「訓讀文」に対して、沖森（1995）の「翻字本文」を元に、訓点を省いた漢字・万葉仮名で表記された本行本文を「原文」として認定し、原文情報を構成した。ただし、表記研究への便宜のため、万葉仮名の誤りを修正した情報に限り、沖森（1995）の「翻字本文」の情報を取り入れ記号付きで原文を示した。

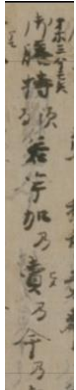
原典	翻字	訓読	コーパスの本文と原文の対応							
			本文	御膳	持ち	する	若字加の賣	の	命	と
			原文	御膳	持	須留	若字加乃賣	乃	命	乃〈止〉
	御膳持須留若字加乃賣乃命	御膳持ちする若字加の賣の命と	校訂内容		活用語尾補読	万葉仮名展開	万葉仮名展開	万葉仮名展開		万葉仮名展開・原文誤り修正

図5 コーパス本文に対する原文情報と、原典・翻字本文・訓読文の関係

4.3 底本画像へのリンク

検索結果の「底本リンク」列に示される「e 国宝」ボタンから、沖森（1995）の底本となっている東京国立博物館蔵『延喜式』（九條家本）の高精細画像へリンクする。これにより、訓点も含めた原典の様相を詳細に見ることができる。

なお、九條家本は卷子本であるため、e 国宝は丁（ページに相当する単位。半丁、見開きなど）を単位とした画像情報を持たない。よって、「奈良時代編Ⅲ祝詞」の底本リンクでは、行を単位として画像の位置情報との対応付けを行っており、検索結果からリンク先画像へ遷移した際、検索対象語の存在する行が、概ね画像中央に表示されるようになっている。

5. おわりに

本稿では、2021 年 3 月に公開した「奈良時代編Ⅲ祝詞」の資料、底本、データ内容および検索アプリケーション「中納言」による利用について報告・解説した。最後に、残された課題と今後の展開について 3 点を挙げる。

- (1) 2021 年 3 月公開データは「短単位」のみの提供である。先行する「奈良時代編Ⅰ万葉集」「奈良時代編Ⅱ宣命」では長単位についても公開しているため、「奈良時代編Ⅲ祝詞」においても、今後長単位の整備を進め、2021 年度中に公開する。
- (2) 本研究プロジェクトの連携機関である国立歴史民俗博物館により、所蔵する『延喜式』（土御門

家本、17世紀写本)の資料画像が2021年春以降順次公開される。そこで、2021年度中に「中納言」の「参考リンク」(底本以外の参照資料へのリンク)として、この画像へのリンクを提供する。

(3)「中納言」による公開は、主に形態論情報に基づく利用を想定しており、訓点を含めた詳細な表記情報について十分には情報を提示していない。今後は、日本語表記史研究に資する言語資源の構築、また分野横断的な人文学資料のアーカイブ及び共有を目指す試みとして、訓点情報等詳細な表記情報を記述したXMLデータの公開を検討している。

付記

本研究は、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」国語研ユニット「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・参照リンク

- [1] 国立国語研究所(2017)『日本語歴史コーパス 奈良時代編Ⅰ万葉集』
https://pj.ninjal.ac.jp/corpus_center/chj/nara.html#manyo
- [2] 国立国語研究所(2020)『日本語歴史コーパス 奈良時代編Ⅱ宣命』
https://pj.ninjal.ac.jp/corpus_center/chj/nara.html#senmyo
- [3] 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
- [4] 倉野憲司(1943)『日本文学史 第三卷 大和時代 下』三省堂
- [5] 沖森卓也(2000)『日本古代の表記と文体』吉川弘文館
- [6] 伝康晴他(2007)「コーパス日本語学のための言語資源:形態素解析用電子化辞書の開発とその応用」『日本語科学』22
- [7] 小椋秀樹他(2011)『現代日本語書き言葉均衡コーパス』形態論情報規程集 第4版 (下)』(文部科学省科学研究費特定領域研究「日本語コーパス」データ班)
- [8] 小木曾智信他(2013)「中古仮名文学作品の形態素解析」『日本語の研究』9(4)
- [9] 小木曾智信(2013)「歴史的日本語資料を対象とした形態素解析」『自然言語処理』20(5)
- [10] 倉野憲司・武田祐吉校注(1958)『日本古典文学大系Ⅰ 古事記 祝詞』岩波書店
- [11] 金子善光(2012)「翻刻・歴史民俗博物館所蔵「延喜式 卷八祝詞」」『神社と実務』11
- [12] 金子善光(2014)「翻刻・京都大学図書館蔵「陽明文庫本 延喜式・卷八・祝詞」」『文化史史料考證刊行委員会』編『嵐義人先生古稀記念論集 文化史史料考證』
- [13] 呉寧真他(2020)「『日本語歴史コーパス 奈良時代編Ⅱ宣命』の公開」『日本語学会2020年度春季大会発表予稿集』日本語学会
- [14] 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』