

『日本語歴史コーパス 和歌集編』のテキストと「中納言」での表示

2020年10月1日

2023年1月10日更新

松崎安子

1. はじめに

『日本語歴史コーパス 和歌集編』（以下『CHJ 和歌集編』）は、国文学研究資料館（以下 National Institute of Japanese Literature 以下、NIJL）収蔵の正保四年版本『二十一代集』の翻字データを底本とし活用した。この翻字データはCD-ROM版で刊行されているほか、web上でNIJLと人文学オープンデータ共同センター（Center for Open Data in the Humanities以下、CODH）が公開している「日本古典籍データセット」からも取得可能なものである。

底本をコーパス化するにあたっては、形態素解析を施す前処理として一部テキスト整形（本文書では以下「校訂」と呼ぶ）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。

この文書では本コーパスのテキスト校訂と、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述べる。

2. コーパス化の対象範囲

『CHJ 和歌集編』では、NIJL収蔵の正保四年版本『二十一代集』のうち八代集を対象としコーパス化した。八代集を統一的にコーパス化するにあたり、コーパスの本文としては序文、詞書、歌、左注を対象とした。このうち序文は『古今和歌集』、『後拾遺和歌集』、『千載和歌集』、『新古今和歌集』が有しているが、『古今和歌集』および『新古今和歌集』が有する真名序は対象から除いている。そのほか、次にあてはまる部分と歌についてもコーパスの本文から除いている。

- ・『古今和歌集』『金葉和歌集』の異本歌
- ・『後撰和歌集』『後拾遺和歌集』のうち、新編国歌大観番号を有しない歌

3. テキストの校訂

テキスト校訂は次に掲出する8項目にわたり行った。なお、掲出する各用例には当該例のサンプルIDおよび開始位置を併記する。

(1) 濁点付与

NIJL作成の翻字データは全て清音表記となっている。このNIJL翻字データについて濁点が期待される箇所付与した。濁点付与の参考として、新日本古典文学大系（岩波書店）の八代集分冊本も用いた。

例（底本）ちはやふる神代よりはしまりて➡(校訂)ちはやぶる神代よりはじまりて
(20W千載1188_00001,170) (20W千載1188_00001,230)

(2) 踊り字・くの字点をひらく

踊り字、くの字点は繰り返されている文字に置換した。ただし、文法や短単位としてのま

とまり、歌意等を勘案した場合、単純に踊り字直前の文字の繰り返しでは不都合が生じる場合はその限りではない。なお、NIJL翻字データにおける「々」はそのままとした。

例 (底本) 春の野々わかなもなにも → (校訂) 春の野のわかなもなにも
(20W後撰0955_01001,2970)

(底本) いよ / \ → (校訂) いよいよ (20W古今0905_17018,18950ほか)

(底本) さゝ浪よする → (校訂) さぎ浪よする (20W詞花1151_01001,390ほか)

(3) 助詞・活用語尾の省略表記を補字

NIJL翻字データには、版本での表記を全てかなに開いた「かな標準化表記」がある。その「かな標準化表記」に従い読んだ場合に必要となる助詞や活用語尾をかなで補った。

例 (底本) 賀茂臨時祭 → (校訂) 賀茂の臨時の祭
(20W後撰0955_16016,3860ほか)

(底本) 久恋といふ事を → (校訂) 久しき恋といふ事を
(20W千載1188_14015,19670ほか)

ただし、「延喜御時歌合」や「宇治前太政大臣」といった例は、「延喜の御時の歌合」や「宇治の前の太政大臣」とはしない。これらのような読み添えの「の」については国立国語研究所コーパス開発センター(池上尚)編(2016)『日本語歴史コーパス 平安時代編』形態論情報規程集』に依拠している。

(4) 欠字・誤記に対する補字・訂正

NIJL翻字データについては、コーパス化の対象としたすべてに対し形態論情報を付与するため、底本での明らかな欠字や誤記についてはNIJL翻字データの注記および「新日本古典文学大系」の注記をもとに補字、訂正を行った。

例 (底本) きのとこのといふ所に → (校訂) きの丸どこのといふ所に
(20W後拾1087_18019,24250)

(底本) 幡磨国にくたり侍けるに → (校訂) 播磨にくたり侍けるに
(20W後拾1087_19020,5310)

(底本) : 法眼つかはす → 校訂 : 法服つかはす
(30W新古1205_17018,14160)

(底本) : 朝みとり → 校訂 : 浅みどり (30W新古1205_08009,430)

(5) 短単位を跨ぐ漢字表記をひらがなへ開く

NIJL翻字データにおいて漢字表記とされていても、短単位の境界を跨ぐ場合は漢字表記をひらがなを開いた。

例 (底本) 袖やぬれ南 → (校訂) 袖やぬれなむ (20W古今0905_01002,16610)

(底本) いのちとも哉 → (校訂) いのちともがな

(20W後拾1087_12013,26840)

(6) 返読箇所への対応

NIJL翻字データにおける「かな標準化表記」に従い訓読した場合、返読の必要な箇所は訓読した日本語文の語順に入れ替えた。その際、不足・重複する形態素を補読・削除した。

例 (底本) 郭公驚夢といへる事をよめる

→ (校訂) 郭公夢を驚かすといへる事をよめる (補読の例)

(20W金葉1128_02002,9280)

(底本) 逢不逢恋といへる心

→ (校訂) 逢ひて逢はざる恋といへる心 (削除および補読の例)

(20W千載1188_14015,4730)

(7) 日本語文として訓読されていない箇所への対応

NIJL翻字データの「かな標準化表記」においても日本語文として訓読されていない漢文字列については「漢文」扱いとし、形態論情報は付与しないこととした。

例 (底本) 蓮花初開楽 (NIJLかな標準化表記「れんけしよかいらく」)

(30W新古1205_20021,9720)

(底本) 梅檀香風悦可衆心 (NIJLかな標準化表記「せんたんかうふうえつかしゆしん」)

(30W新古1205_20021,15600)

上記の校訂内容および校訂前の原文の状態は全てXMLタグに記録されている。中納言利用者は、XMLから復元した底本での表記を、検索結果の原文表記欄で確認できる。

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	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
30W新古1205_17018	14180	9090		嵐の風も心して ふけ # 少将たかみ つ横川にまかりて かしらおろし待ける に	法眼 つか はす	とて # おく山の苔の 衣にくらへみよいつ れか露の夜きまさるとも # 返し # 白露のあした	ツカ ワス	遣わす	ツカ ワス	動詞- 一般	文語 四段- サ行	終止 形- 一般	つか はす	詞書		歌集	新古今和歌集	1205	雑歌中				古典選集 本文DB 正保版本 「二十一代集」	1454	Nijl	Maruzen
				申しけ世をのかれ にしみ山辺に嵐の風 も心してふけ # 少将 たかみつ横川にまか りてかしらおろし待 けるに	法眼 つか はす	とて # おく山の苔の衣 にくらへみよいつれか 露のをきまさるとも # 返し # 白露のあした夕に おくやまの苔のころもは 風もさはらす																				

上図に示すように、「中納言」検索結果画面における「前文脈」「キー」「後文脈」各欄は上下に分かれており、青色囲みの上部が校訂本文、赤色囲みの下部が原文表記を表示している。

4. 中納言における表示項目

『CHJ 和歌集編』に付与された形態論情報や書誌的な情報は検索アプリケーション「中納言」で各欄に表示される。本書では、とくに和歌集編を検索対象とし、その結果を得た際の読み取りについて留意点を記す。

(1) 本文種別

『CHJ 和歌集編』では次のような本文種別を表示する。

- ・古注（仮名序における古注部分）
- ・古注-詞書（仮名序における古注に含まれた歌に付随する詞書部分）
- ・古注-引用歌（仮名序における古注に含まれた引用歌）
- ・詞書
- ・詞書-歌
- ・歌
- ・左注
- ・左注-歌（左注に含まれる歌）

なお、仮名序の地の文については本文種別が空欄となる。

(2) 話者

『CHJ 和歌集編』では話者欄に和歌の作者を表示する。和歌の作者はNIJL翻字データの内「和歌作者標準化表記」と『勅撰集付新葉集作者索引』により整備し、同名表記であっても、その出自や和歌集編纂の時期から一意に定めた。その一部については、作歌者名をクリックすることで「国立国会図書館典拠データ検索・提供サービス (Web NDL Authorities)」を閲覧し、作者情報を得ることが可能となっている。

(3) 作品情報

- ・成立年 和歌集によってはその編纂年に諸説あるが、犬養廉ほか『和歌大辞典』に依拠し決定した。
- ・巻名等 各和歌集の部立を表示する。

(4) 底本情報

底本情報のうち「ページ番号」には、次項目に挙げる「底本リンク」画像のコマ数を表示する。

(5) 画像リンク情報

コーパス検索アプリケーション「中納言」による和歌集編の検索結果については、外部リンクとして次の二つを用意した。

①「底本リンク」

NIJL所蔵正保版本『二十一代集』の影印画像が無償で閲覧可能な「日本古典籍ビューア」(<http://codh.rois.ac.jp/iiif/iiif-curation-viewer/>)へリンク付けを行った。

②「参考リンク」

「新日本古典文学大系」(岩波書店)の各和歌集が閲覧可能な「Maruzen eBook Library」(<https://elib.maruzen.co.jp/>)へリンク付けを行った。ただし、こちらは機関向け電子書籍配信サービス(有償)で、利用にあたっては利用者の所属機関が購読契約を結んでいる必要がある。

参考文献

犬養廉・井上宗雄・大久保正・小野寛・田中裕・橋本不美男・藤平春男編(1986)『和歌大辞典』明治書院

菊池そのみ・菅野倫匡(2022)「和歌語彙の研究に『日本語歴史コーパス』を利用する際の留意点」『筑波日本語研究』第26号

国立国語研究所コーパス開発センター(池上尚)編(2016)『『日本語歴史コーパス 平安時代編』形態論情報規程集』(https://pj.ninjal.ac.jp/corpus_center/chj/doc/morph-heian-2016.pdfにて公開)

名古屋和歌文学研究会編(1986)『勅撰集付新葉集作者索引』和泉書院

松崎安子(2019)「CHJ『和歌集編』(八代集)の構築と公開」ワークショップ「『日本語歴史コーパス』の今とこれから」『日本語学会2019年度春季大会予稿集』

松崎安子・小木曾智信・中村壮範(2019)「『日本語歴史コーパス 和歌集編』Ver.1.0の公開」『日本語学会2019年秋季大会予稿集』