

『日本語歴史コーパス 和歌集編』形態論情報の概要

2020年10月1日 松崎安子

『日本語歴史コーパス 和歌集編』（以下『CHJ 和歌集編』とする）における形態論情報は、検索の便宜・処理上の便宜を考慮して付与されているため、既存の索引類やいわゆる学校文法とは異なる尺度で付されたものも存在する。注意を要するものとして以下にそれらを取り上げる。

1. 依拠した規程と言語単位について

『CHJ 和歌集編』の内実は八代集のコーパスである。八代集の編纂時期は平安時代から鎌倉時代にまたがっているものの、コーパスの構築に際して、規程は一貫して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』を用いた。それに基づき短単位と長単位との二種類の言語単位をとり、それぞれに形態論情報を付与している。

2. データ構築の概要

2. 1 短単位データの認定

短単位データの認定は『CHJ 平安時代編』と同様とし、詳細は国立国語研究所（2014）『『日本語歴史コーパス 平安時代編』形態論情報の概要』に譲る。

2. 2 長単位データ

長単位データ作成にあたっては、人手修正済み短単位データを基としながら次のような手順に沿って行った。

1) 文節境界の付与

和歌集編の文節境界は次のようなルールをあらかじめ用意し、自動的に付与した。

- ・ 和歌集巻頭
- ・ 序、詞書、左注などの文頭
- ・ 副詞か形状詞
- ・ 感動詞
- ・ 接頭辞
- ・ 前接語名詞か代名詞で、当該語が名詞か代名詞
- ・ 前接語が名詞か代名詞で、当該語が動詞
- ・ 前接語が名詞か代名詞で、当該語が形容詞か形状詞
- ・ 前接語が動詞で、当該語が名詞か代名詞

- ・前接語が動詞で、当該語が動詞
- ・前接語が形容詞か形状詞で、当該語が名詞か代名詞
- ・前接語が形容詞で、当該語が動詞
- ・前接語が助詞か助動詞で、当該語が助詞助動詞以外
- ・前接語が副詞で、当該語が助詞か助動詞以外
- ・前接語が感動詞で、当該語が何であれ
- ・前接語が接尾辞で、当該語が助詞か助動詞以外

2) 長単位への結合

文節境界を自動付与した後、長単位結合が必要な個所について人手で結合作業を行った。すでに富士池(2012)で述べられたように、『CHJ 平安時代編』に収録されている中古和文学作品の長単位自動解析にあたっては、短単位から長単位を自動構成する解析器 Comainu ver.0.60(小澤ほか 2011)を用いているが、中古和文特有の接頭辞や接尾辞、敬語表現が関連する箇所での誤解析が報告されている。また、和歌では名詞および動詞連続のパターンが多いが、それらが全て複合名詞および複合動詞として結合されるべきものでない。そのため、和歌集については人手で長単位結合作業を行った。一部の例を表1、表2に示す。

表1 古今和歌集4番歌詞書の長単位結合例

文節	短単位	長単位	長単位語彙素	長単位語彙素読み	長単位品詞
B	二条	二条のきさき	二条の後	ニジョウノキサキ	名詞-固有名詞-人名-一般
	の				
	きさき				
	の	の	の	ノ	助詞-格助詞
B	春	春	春	ハル	名詞-普通名詞-一般
	の	の	の	ノ	助詞-格助詞
B	はじめ	はじめ	始め	ハジメ	名詞-普通名詞-一般
	の	の	の	ノ	助詞-格助詞
B	御	御歌	御歌	オオンウタ	名詞-普通名詞-一般
	歌				

表2 新古今和歌集8番歌の長単位結合例

文節	短単位	長単位	長単位語彙素	長単位語彙素読み	長単位品詞
B	岩ま	岩ま	岩間	イワマ	名詞-普通名詞-一般
B	とぢ	とぢ	閉じる	トジ	動詞-一般
	し	し	き	シ	助動詞
B	氷	氷	氷	コオリ	名詞-普通名詞-一般
	も	も	も	モ	助詞-係助詞
B	けさ	けさ	今朝	ケサ	名詞-普通名詞-一般
	は	は	は	ワ	助詞-係助詞
B	解け	解け初め	解け初める	ソケソメル	動詞-一般
	初め				
	て	て	て	テ	助詞-接続助詞

3) 連語

『CHJ 平安時代編』同様、連濁や係り受けを考慮したとき、付属語部分で切り離さずひとまとまりとするのが妥当と考えられるものをリストで管理する。

表3 『CHJ 和歌集編』連語リスト

語彙素読み	語彙素	品詞
アシノハナゲ	葦の花毛	名詞
アリアリテ	有り有りて	副詞
オモイノホカ	思いの他	名詞
キノハガクレ	木の葉隠れ	名詞
ココロト	心と	副詞
ココロノホカ	心の他	名詞
コトナシグサ	事成し草	名詞
コトナシグサ	事無し草	名詞
コトノホカ	事の他	名詞
コノモカノモ	是の面彼の面	名詞
サキサカズ	咲き咲かず	名詞
シラズガオ	知らず顔	名詞
チリチラズ	散り散らず	名詞
トワズガタリ	問わず語	名詞
ナイモノグサ	無い物草	名詞
ミマクホシイ	見まく欲しい	形容詞
ワレト	我と	副詞

2. 2. 形態論情報

1) 表記等による語彙素の判断

小椋ほか（2011）の「同語異語判別規程」に拠るなどし、『CHJ 和歌集編』では意味・機能の別と一致しない場合であっても、表4に示すように、表記等によって語彙素の判断を行った場合がある。

表4 同語異語判別

語彙素	語形	書字形	備考
尋ねる	タズヌ	たづぬ	仮名表記の場合、質問相手の存在が明確な場合は「尋ねる」、それ以外は「訪ねる」とした。漢字表記の場合、たとえば「尋ぬ」表記で「訪問」を意味する場合であっても語彙素は「尋ねる」とする。
		尋ぬ	
訪ねる	タズヌ	たづぬ	
		訪ぬ	

2) 文末の活用形について

小田（2015）によれば中古以前では主節中に主格の「の」が見られる場合、原則、文末が終止形で結ばれることがない。その指摘に従い、文末（歌末）の用言の活用形が終止形と連

体形で同形の場合で主格の「の」が直近に見られるとき、文末は連体形とした。

例) 天川もみぢをはしにわたせばやたなばたつめの秋をしもまつ
(20W 古今 0905_04005, 2380)

例) 山ざとの垣ねに春やしらるらむ霞まぬさきに鶯のなく
(20W 千載 1188_01002, 6000)

3) 詞書や歌題に見られた経典の一節について

当該箇所は「漢文」として扱うこととした。その種別は、コーパス検索アプリケーション「中納言」における「品詞」欄に、「漢文」と表示される。これらはまた、コーパス検索アプリケーション「中納言」の「文字列検索」から本文検索が可能である。

例) 三界唯一心 舍利講のつみでに、願成仏道の心を、
常在靈鷲山の心をよめる 心如月輪若在輕霧中のころを

参考文献

- 小椋秀樹・小磯花絵・富士池優美・宮内佐夜香・小西光・原裕 (2011) 『特定領域研究「日本語コーパス」平成 22 年度研究成果報告書『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)』 (https://pj.ninjal.ac.jp/corpus_center/bccwj/doc/report/JC-D-10-05-02.pdf にて公開)
- 小澤俊介・内元清貴・伝康晴 (2011) 「BCCWJ に基づく中・長単位解析ツール」特定領域「日本語コーパス」平成 22 年度公開ワークショップ予稿集 (https://pj.ninjal.ac.jp/corpus_center/bccwj/doc/workshop/JC-G-10-02.pdf にて公開)
- 小田勝 (2015) 『実例詳解 古典文法総覧』和泉書院
- 国立国語研究所 (2014) 『『日本語歴史コーパス 平安時代編』形態論情報の概要』国立国語研究所コーパス開発センター (https://pj.ninjal.ac.jp/corpus_center/chj/doc/abstract-heian-2016.pdf にて公開)
- 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)
- 国立国語研究所コーパス開発センター (近藤明日子) 編 (2016) 『近代文語 UniDic 短単位規定集 Ver.1.1』 (https://pj.ninjal.ac.jp/corpus_center/chj/doc/unidic-MLJ_rulebook_v1_1.pdf にて公開)
- 富士池優美 (2012) 「中古和文における長単位の概要」『第 2 回コーパス日本語学ワークショップ予稿集』国立国語研究所 (https://www.ninjal.ac.jp/event/specialists/project-meeting/files/JCL_Workshop_no2_papers/JCLWorkshop2012_2_web.pdf にて公開)