

『日本語歴史コーパス 江戸時代編V注釈書』 Ver.1.0 形態論情報の概要

2025年3月28日 久保 柊子

1. はじめに

『日本語歴史コーパス 江戸時代編V注釈書』 Ver.1.0（以下、『日本語歴史コーパス』をCHJと称する）の形態論情報は、先行して公開された『CHJ 江戸時代編I洒落本』『CHJ II人情本』『CHJ III近松浄瑠璃』『CHJ IV随筆・紀行』と同じ規程によっている。単位規程は原則的に『日本語歴史コーパス 江戸時代編I洒落本』『日本語歴史コーパス 江戸時代編II人情本』形態論情報の概要（村山 2019）、『日本語歴史コーパス 江戸時代編III近松浄瑠璃』形態論情報の概要（片山 2020）に準拠し、本文書では、『CHJ 江戸時代編V注釈書』 Ver.1.0（以下「V注釈書」と称する）に独自の点について述べる。

2. 短単位データの作成

「V注釈書」の短単位データの作成は、他のCHJのサブコーパスと同様に自動形態素解析と人手修正によって行われている。

形態素解析処理としては、収録資料である『古今集遠鏡』の特徴である、異なる文体によって構成されている点に対応するため、以下のように形態素解析辞書を区別している。

文の種別	解析辞書
地の文 (宣長の解説、横井千秋の注)	中古和文 UniDic
古今集の和歌の原文	和歌 UniDic
宣長による訳文	近世上方口語 UniDic

3. 活用例

洒落本・人情本コーパスでは、「会話」と「割書き」の本文種別によって「口語」と「文語」の区別を行っている。本コーパスでも、この本文種別による区別を踏襲し、原則として、『古今和歌集』の本文と本居宣長や横井千秋による解説部分を「文語」、宣長訳の箇所を「口語」とみなして処理を行なった。

宣長訳内でも『古今和歌集』本文を引用している箇所や、文語活用でしか対応できないものについては、本文の形を優先して形態論情報の付与を行っている。

【例】

- (1) ちはやぶる神代の時分には歌の文字の数もまだ定まつた事もなし
…文語形容詞-ク 終止形-一般 (52-遠鏡 1793_00102, 10260)
- (2) 春になつたれば花ぢやと思ふてやら
…文語助動詞-タリ-完了 已然形-一般 (52-遠鏡 1793_00201, 5070)
- (3) ○梅の枝を折たによつてそれで袖がにほふのでこそあれ
…文語ラ行変格 已然形-一般 (52-遠鏡 1793_00201, 34000)

4. 指定辞の判別

指定辞について、『CHJ 室町時代編 I 狂言』では「なり」「じゃ」を、『CHJ 明治・大正編』では「だ」「じゃ」をそれぞれ併用しており、洒落本・人情本コーパスでは、指定の助動詞として「なり」「じゃ」「だ」の3種類をみとめ、処理を行っている。本コーパスでは、洒落本・人情本コーパスに従って上記の3種類を採用し、本文種別によって以下の使い分けを行った。

文語文による地の文・注釈、古今集の和歌の原文→「なり」
宣長による訳文→「じゃ」「だ」

宣長訳中であっても、ナリ活用で処理することが妥当な形式については、助動詞「なり」として処理する点、仮定条件を表すもの（「ナラ」、「ナラバ」）については便宜上すべて助動詞「だ」に倒す点についても洒落本・人情本（村山 2019）に従った。

5. 特殊品詞「万葉仮名」

形態論情報の付与が困難な箇所は、品詞列に「特殊品詞」を付与している。本コーパスでは、特殊品詞「万葉仮名」を付与している。

【例】

- (4) 餘材に引る、佛足石の歌の、夜與都も是也 (52-遠鏡 1793_00219, 13290)
- (5) 万葉にあるは、いさとを寸許瀬、余名告奈にて、ここなると同じ意也
(52-遠鏡 1793_00220, 43520)

6. 記号の種別

本コーパスで「記号」としたのものには次のようなものがある。

6.1 省略した句を示すもの

『古今集遠鏡』の例言には、宣長が和歌に俚言訳を施す際、序詞などの元の句を省略する場合には訳を施さなかった句を示す印をつけることが述べられている。実際、上の句を省略する場合の「上」、一句目を省略する場合の「一」といった□に囲まれた印が本文中に多く見られる。このような印は、テキストデータに文字のみを示し、形態論情報は付与せず、記号として処理した。

【例】

(6) わがせこが衣のすそを吹かへしうらめづらしき秋の初風

○上これはノ、めづらしい秋風ぢやさても涼しいころよ

(52-遠鏡 1793_00204,2760)

(7) から衣きつつなれにしつましあればはるノ、きぬる旅をしぞおもふ

○一きつつ故郷になじんだ妻があれば別れてはるノ、と来た此旅がさころぼそう物がなしう思はるる

(52-遠鏡 1793_00209,11470)

6.2 語の一部

宣長の注釈内には、語の一部を引用してその語構成を解説したり、誤字を指摘したりする箇所が見られる。このような語の一部は、形態論情報の付与が困難であるため、記号として処理した。

【例】

(8) さうなは、さまなるといふことなるを、音便にさうといひ、るをはぶける也

(52-遠鏡 1793_00001,49070)

(9) いづもは、いでくも也、でくつづまりてづとなる

(52-遠鏡 1793_00102,14110)

(10) ○千秋云、此集のなればはみをなに写し誤れる也、みとなと似たり

(52-遠鏡 1793_00214,68580)

【参考文献】

村山 実和子 (2019) 『『日本語歴史コーパス江戸時代編 I 洒落本・II 人情本』形態論情報の概要』 <https://clrd.ninjal.ac.jp/chj/doc/morph-edo-2019.pdf>

片山 久留美 (2020) 『『日本語歴史コーパス江戸時代編III近松浄瑠璃』形態論情報の概要』 <https://clrd.ninjal.ac.jp/chj/doc/morph-chikamatsu-2020.pdf>

松崎 安子 (2023) 『日本語歴史コーパス江戸時代編IV随筆・紀行』 Ver.0.8 形態論情報の概要 <https://clrd.ninjal.ac.jp/chj/doc/morph-zuihitsu-202303.pdf>