

『日本語歴史コーパス江戸時代編IV随筆・紀行』 Ver.0.4(芭蕉の紀行文)

形態論情報の概要

2021年3月31日 松崎安子

はじめに

『日本語歴史コーパス 江戸時代編IV随筆・紀行』 Ver.0.4(芭蕉の紀行文)の形態論情報は、先行して公開された『日本語歴史コーパス 江戸時代編 I 洒落本』『同 II 人情本』『同 III 近松浄瑠璃』と同じ規定によっている。規定の詳細は村山(2018)「『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」、および、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」に譲ることとし、以下では『日本語歴史コーパス 江戸時代編 IV随筆・紀行』 Ver.0.4(芭蕉の紀行文)に独自の点について述べる。

1. 短単位データの作成

『日本語歴史コーパス 江戸時代編IV随筆・紀行』 Ver.0.4(芭蕉の紀行文)の短単位データの作成は、他の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、解析用辞書に「中世文語 UniDic」を用いているが、『日本語歴史コーパス 江戸時代編』の地の文を学習用データとして加え解析を行っている。

2. 活用型・活用形

現代語のコーパスおよび『日本語歴史コーパス』では、活用語について「文語」「口語(明示なし)」の二大別を行っている。「I 洒落本」「II 人情本」「III 近松浄瑠璃」ではそれらを本文種別によって大別し対処している。つまり、本文種別は「会話」では口語活用を、本文種別が「会話」以外であれば文語活用によっている。それに従えば、「IV随筆・紀行」は全文、文語文であるため、原則として活用は文語のものとなる。

ただし次のような例は、『室町時代編』までのように未然形+助動詞「う」(用言+助動詞)と分割せず、現代語のコーパスと同じく、ひとまとまりの「意志推量形」として扱っている。このように、本文種別が会話以外の場合でも、意志推量形については口語活用として対処した。

【例】

(1) よし野にて桜見せふぞ檜の木笠 (51-芭蕉 1691-01001, 22790)

(2) よし野にて我も見せふぞ檜の木笠 (51-芭蕉 1691-01001, 22940)

(いずれも本文種別「韻文-俳句」/「為る」動詞-非自立可能、サ行変格、意志推量形)

3. 未知語の扱い

注釈書によっても解釈が不明な箇所は「未知語」として扱い、その種別を「品詞」欄に

表示する。本コーパスでは次の2件があたる。なお文字列検索にはヒットする。

【例】

(3) 𠄎碗玉卮せいわんぎよくしの心ちせらるも所がらなり。(51-芭蕉 1689-01001, 9390)

(4) 𠄎岐 またれつる五月さつきもちかしむこちまき聾ろう (51-芭蕉 1691-02001, 22270) 本文種別：前書

【参考文献】

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://ccd.ninjal.ac.jp/chj/doc/morph-chikamatsu-2020.pdf>

村山実和子 (2018) 「『日本語歴史コーパス江戸時代編Ⅰ洒落本』『同 江戸時代編Ⅱ人情本』

形態論情報の概要」(2019年3月29日更新)

<https://ccd.ninjal.ac.jp/chj/doc/morph-edo-2019.pdf>