

『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

2021年5月25日

2023年3月31日更新

松崎安子

『日本語歴史コーパス江戸時代編Ⅳ随筆・紀行』 Ver.0.8 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.0.8 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.0.8 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、解析用辞書として芭蕉の紀行文には「中世文語 UniDic」を、随筆作品群には「近世文語 UniDic」を用い解析を行っている。

2. 解釈不明の語の扱い

参考文献等における注釈によっても解釈が不明な箇所は「解釈不明」として扱い、その種別を「品詞」欄に表示する。形態論情報は与えないが、文字列検索にはヒットする。

【例】

- (1) の心ちせらるも所がらなり。 (51-芭蕉 1689-01001, 9390)
- (2) 三岐 またれつるもちかし (51-芭蕉 1691-02001, 22270)
- (3) 「ゲール」といふものは身軀にはなし。 (52-玄白 1815-22007, 3890)

【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>