

『日本語歴史コーパス 江戸時代編Ⅲ近松浄瑠璃』テキスト・表示項目について

2020年3月31日 片山久留美

1. はじめに

『日本語歴史コーパス 江戸時代編Ⅲ近松浄瑠璃』(以下「本コーパス」と呼ぶ)は、『新編日本古典文学全集』74 近松門左衛門集(1)・75 近松門左衛門集(2)(小学館)を底本としている。電子化に際して一部テキストを校訂し、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。本稿では中納言上に表示される各種情報について、テキストの校訂状況や表示情報に関する概要を述べる。

2. テキスト校訂と本文の範囲

本コーパスのテキストは、基本的に『新編日本古典文学全集』のテキストをそのまま使用しているが、コーパス化にあたり以下の各項目について校訂を行った。

●踊り字

仮名1字分の踊り字(ゝ, ゞ, ヰ, ヱ)は、想定される仮名に置き換えた。変換前の文字列は「原文文字列」および「原文 KWIC」に表示される。2字分以上に相当するくの字点は置き換えの対象としない。

例)

(底本) このごとく踏みたゝかれ

→ (本コーパスのテキスト) このごとく踏みたたかれ

(51-近松 1703_11002, 36820¹)

(底本) アゝこれ/\、母は今、蔵から出た

→ (本コーパスのテキスト) アアこれ/\、母は今、蔵から出た

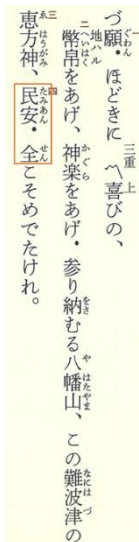
(51-近松 1704_06002, 15550)

●内題・見出し

『新編日本古典文学全集』では、各作品に対して内題や章ごとの見出しなどが付与されている。本コーパスでは、これらの内題・見出しについて、近松全集刊行会編(1985-1996)『近松全集』(岩波書店)の版本に記載のあるもののみをコーパス本文とし、版本に記載のないものはコーパス本文に含めなかった。

¹ 当該例の「中納言」におけるサンプルID、開始位置を記載する。

●文字譜・点



近松浄瑠璃のテキストには、左図のように点や文字譜が付されている。点については、「音曲的な句切りを示すもので、あながち文意にそって付されたものではない」（信田純一校注（1986）『新潮日本古典集成 近松門左衛門集』凡例 p.3）と言われており、必ずしも文や語の切れ目を表すものにはなっていない。また、「ウ」「ハル」「詞」「地ハル」などの文字譜は、音の高低や台詞回しに関するものなど、台本を音曲として実演するときに必要な情報である。

これらの情報は、演劇として舞台上で演じられることを考える際には非常に重要なものであるが、本コーパスはあくまでも言語研究への利用を第一の目的としている。そのため、これらの文字譜・点などの記号類についてはタグにより位置の情報を残すにとどめ、コーパス本文からは除外した。

3. 「中納言」における表示項目と内容

本コーパスの本文にはさまざまなタグや単語情報が付与されており、言語研究を目的とした利用に資するものとなっている。これらの情報はコーパス検索アプリケーション「中納言」上に検索結果として表示される。形態論情報の表示については他のサブコーパスと基本的に変わらないため省略し、以下では本コーパス使用に際して特に注意の必要な「本文情報」「作品情報」などの各列について表示内容の詳細を述べる。

サンプルID	開始位置	連番	コア	前文脈	キ	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
51-近松1703_11002	14890	9500		1 立て四月七日まで(に)立つと立て、(商ひ)の勘定せよ # まくり出して(大坂の地)は	踏み	せぬ # と、際(か)るる # それ(が)し(男)の我、 # オオ、ソレ、費(つ)つた # と、	フム	踏む	フム	動詞-一般	五段-マ行	未然形-一般	踏み	会話	平野屋徳兵衛・男・町人→醬油屋手代	近松浄瑠璃	曾根崎心中	1703	生玉社の場	近松門左衛門	1653	新編全集 <75>	23	JK		
51-近松1703_11002	36800	23070		1 しと町内へ(披露)して、(か)へつて(今)の(逆)ねだれ # 口(憎)しや、(無)念(や)な # この(と)とく	踏み	びた(か)り、(男)も(立)た(ず)、(身)も(立)た(ず) # (工)工(最)前(に) (つか)み(つ)き(ひ)つ(て)なりとも	フム	踏む	フム	動詞-一般	五段-マ行	連用形-一般	踏み	会話	平野屋徳兵衛・男・町人→醬油屋手代	近松浄瑠璃	曾根崎心中	1703	生玉社の場	近松門左衛門	1653	新編全集 <75>	28	JK		

3.1. 本文情報

●本文種別

本コーパスでは、本文種別を以下のように区別し、「本文種別」欄に表示している。地の文の場合は「本文種別」は空欄となる。

会話：登場人物の発話箇所

その他-内題：作品名および「上之巻」「中之巻」などの見出し・内題部分

その他-刊記：作品末尾の刊記部分

●話者

本文種別が「会話」の箇所について、その発話の話者名を表示する。『新編日本古典文学全集』の各作品の梗概を参考に、同じ発話者の会話に対しては同じ話者名を表示させている。また、それぞれの話者について、「性別」「身分」「職」の情報をわかる範囲で入力した。これらの情報を『江戸時代編 I 洒落本』の話者情報の表示形式に合わせて以下の形式で表示している。

話者名-性別-身分-(地域)-(年齢)-職

例) 笠屋お亀-女-町人---古道具屋娘 笹野権三-男-武士---表小姓

()を付した「地域」「年齢」については、本コーパスでは情報を付与していないため、何も表示されない(「-」のみが表示される)。『江戸時代編 I 洒落本』の表示形式と合わせることで、検索結果をダウンロードしたあとの集計の利便性を担保している。

以下に話者情報の各項目の表示内容について述べる。

話者名：『新編日本古典文学全集』の梗概に記された登場人物名を参考に、同一の発話者には同じ名前を付与した。発話箇所ではあるが発話者が明確でない場合「不明」、複数の人物による発話であると考えられる場合は「その他(複数)」とした。また、作品中で同じ人物が身をやつしたり、変装して別人を名乗って現れることがある。そのような場合は、話者名については「元の人物名【変装後の人物名】」の形で表し、話者情報については変装後の人物の情報を表示している。

例)『薩摩歌』の笹野三五兵衛(本当は肥後の侍だが、林という名の奥女中に変装して現れる)

本来の姿での発話：笹野三五兵衛-男-武士---

「林」としての発話：笹野三五兵衛【林】-女-武士---武家女中

性別：発話者の男女の別を表す。話者名が「不明」の場合は空欄としている。

身分：「武士」「町人」「百姓」「遊女」「子供」「その他」「不明」のいずれかの情報を表示する。「武士」には武家の奉公人、武家の女房なども含む。僧侶・医者などは「その他」に含まれる。

職：「身分」の下位分類として、具体的な職種・立場などを本文から読み取れる範囲で示した。具体的な職名などがわからない場合は空欄としている。

3.2 作品情報

●ジャンル

一律「近松浄瑠璃」と表示される。

●作品名

個別の作品名が表示される。なお角書き部分は省略している。

●成立年

各作品の初演年が表示される。初演年について複数の説がある作品もあるが、『新編日本古典文学全集』の各作品の梗概によって決定した。

●巻名等

各作品の中での「上之巻」「中之巻」「下之巻」の区別を表示する。上中下巻の区別のない『曾根崎心中』のみ場の名前が表示される。

●部

本コーパスでは空欄表示となる。

3.3 底本情報

●底本・ページ番号

底本とした『新編日本古典文学全集』の巻数と、当該例が出現する新編全集のページ番号が表示される。

3.4 外部リンク

●底本リンク

ジャパンナレッジの『新編日本古典文学全集』の該当ページへのリンクが表示される（利用にはジャパンナレッジの利用登録が必要）。

【参考文献・参考 URL】

上野左絵 (2016) 「近松浄瑠璃のコーパス化—語りのテキストをどう扱うか」『じんもんこん 2016 論文集』 pp.25-30

村山実和子 (2019) 『『日本語歴史コーパス 江戸時代編 I 洒落本』テキストの凡例と『中納言』表示項目について』

https://pj.ninjal.ac.jp/corpus_center/chj/doc/abstract-share-2019.pdf

国立国語研究所 (2020) 『日本語歴史コーパス 江戸時代編Ⅲ近松浄瑠璃』

http://pj.ninjal.ac.jp/corpus_center/chj/edo.html#chikamatsu (2020年3月31日確認)