

『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）の概要と公開

松崎 安子（国立国語研究所）

小木曾 智信（国立国語研究所）

1. はじめに

『日本語歴史コーパス江戸時代編IV随筆・紀行』は国立国語研究所の共同研究プロジェクト「通時コーパスの構築と日本語史研究の新展開」で構築中のコーパスで、2021年3月にVer.0.4が公開された。本発表ではこのコーパスの構築の概要と特徴および今後の課題を述べる。

2. データ構築の概要

2. 1 収録作品と単語の単位

『日本語歴史コーパス江戸時代編IV随筆・紀行』は、その名の通り江戸時代の随筆や紀行文を対象に幅広く収録することを意図したものであるが、今回報告するVer.0.4（以下、「本コーパス」）には、俳諧師・松尾芭蕉（1644～1694）の紀行文「鹿島詣」「野ざらし紀行」「更科紀行」「笈の小文」「嵯峨日記」「おくのほそ道」の6作品を収録している。

公開にあたり、他の『日本語歴史コーパス』収録作品と共通の規定に基づき、単語分割や品詞認定などの形態論情報の付与を行った。表1には収録作品とおおよその短単位数を示した。

表1『日本語歴史コーパス江戸時代編IV随筆・紀行』Ver.0.4の収録作品と短単位数

	作品名	成立年	短単位数
1	鹿島詣	1687	800
2	野ざらし紀行	1687	1,700
3	更科紀行	1689	700
4	笈の小文	1691	2,900
5	嵯峨日記	1691	1,700
6	おくのほそ道	1694	8,000
	計		15,800

2. 2 底本テキストとその校訂

本コーパスの底本テキストは、基本的には『新編日本古典文学全集 71 松尾芭蕉集 (2)』の「日記・紀行編」の本文を用いているが、コーパスの構築と利用に適したものとするため以下に記すような校訂を施している。

2. 2. 1 一部の片仮名の置き換え

底本テキストのうち〔1〕格助詞「ガ」の表記や、〔2〕の副助詞「のミ」、〔3〕の副詞「暫ク」、助動詞「ン」、〔4〕の動詞「音信ル」のような活用語尾の一部、〔5〕の形状詞「大キ」の一部、〔6〕「詣拜ス」といったサ変動詞などに見られる片仮名表記部分については、機械解析の際の誤解析を防ぐことや、見出し語としての統一を考慮し、平仮名へと置換した。「中納言」検索結果画面では、置換前の文字列が「原文文字列」および「原文 KWIC」に表示される。

〔1〕〈底本〉去来ガ落柿舎に到。→〈校訂後〉去来が落柿舎に到。

- [2] 〈底本〉おかしき事の^ミ多し。→〈校訂後〉おかしき事の^ミ多し。
〔51-芭蕉 1691-02001, 200〕
- [3] 〈底本〉暫^{しばら}く^{まん}学で愚を^ぐ暁^{きとら}ン→〈校訂後〉暫^{しばら}く^{しばら}暫で愚を愚^ぐ暁^{きとら}ン
〔51-芭蕉 1689-01001, 1630〕
- [4] 〈底本〉何^{なに}がしの方に音信^{おとづ}ル。→〈校訂後〉何^{なに}がしの方に音信^{おとづ}る。
〔51-芭蕉 1691-01001, 1630. 1700〕
- [5] 〈底本〉大^{おほ}キ^{なるくり}成栗の木陰を→〈校訂後〉大^{おほ}キ^{なるくり}成栗の木陰を
〔51-芭蕉 1694-01009, 16350〕
- [6] 〈底本〉御山に詣^{けいはい}拜ス。→〈校訂後〉御山に詣^{けいはい}拜す。
〔51-芭蕉 1694-01013, 27710〕
- [7] 〈底本〉紙・硯^{かみ}をか^{すずり}へて→〈校訂後〉紙・硯^{かみ}をか^{すずり}へて
〔51-芭蕉 1694-01006, 9880〕

2. 2. 2 一部の漢字の置き換え

本コーパスの文字入力は Unicode に準拠している。底本の字体のうち、現行の字体に適切なものがない「^𠄎碗」、「^𠄎岐」(51-芭蕉 1689-01001, 9390)、「^𠄎岐」(51-芭蕉 1691-02001, 22270)のように「^𠄎」へ置き換えた。一方、底本の字体のうち『新編日本古典文学全集』の頭注で当該漢字体が異体字であるという断り書きとともに現行表記による字体が記されている場合は、それをコーパスの本文表記とした(表2)。

表2 『日本語歴史コーパス江戸時代編IV随筆・紀行』
Ver.0.4における異体字置換一覧

字体			作品名
底本	校訂本文		
1	盃	盃	野ざらし紀行、更科紀行、笈の小文、おくのほそ道
2	宵	宵	笈の小文、嵯峨日記
3	騷	騷	おくのほそ道
4	装	装	おくのほそ道
5	霧	霧	おくのほそ道
6	寮	寮	おくのほそ道

2. 2. 3 踊り字の校訂

底本における仮名一字分の踊り字は、想定される仮名に置き換えた。「中納言」検索結果画面では、置換前の踊り字表記が「原文文字列」および「原文 KWIC」に表示される。

- [7] 〈底本〉紙・硯^{かみ}をか^{すずり}へて→〈校訂後〉紙・硯^{かみ}をか^{すずり}へて
〔51-芭蕉 1694-01043, 104590〕

一方、2 字分以上に相当するくの字点や、漢字一字の繰り返しを表す二の字点は置き換えの対象としていない。ただし、[8]のように、文節末の同語をくりかえし、かつ、文節の頭となるような場合は、想定される語を仮名にひらいて表示している。

- [8] 〈底本〉ともなふ人ふたり、浪客^{ろうかく}の士ひとり、^{ひとり}ハ水雲^{すいうん}の僧。
→〈校訂後〉ともなふ人ふたり、浪客^{ろうかく}の士ひとり、^{ひとり}ハ水雲^{すいうん}の僧。
〔51-芭蕉 1687-01001, 950〕

2. 2. 4 合略仮名の置き換え

底本における合略仮名「方」は平仮名にひらいた。「中納言」検索結果画面では置換前の合略仮名表記が「原文文字列」および「原文 KWIC」に表示される。

[9] 〈底本〉それ^{よりほちまんぐう まうづ}方八幡宮に詣。→〈校訂後〉それ^{よりほちまんぐう まうづ}より八幡宮に詣。

[51-芭蕉 1694-01009, 17480]

2. 2. 5 テキストへの補い

底本において助詞や助動詞等が振り仮名として記されている場合、それを元のまま保持しつつ、校訂本文にも平仮名で補ったところがある。[10] では、本行表記が全て漢字列だが、「朋友信」の右傍に「ほうゆうにしん」と振り仮名が付されていたため、格助詞「に」をルビに保持しながら、校訂本文に平仮名で「に」を補うという対応をした。

[10] 〈底本〉朋友^{ほうゆうにしんあるかな このひと}信有哉、此人。→〈校訂後〉朋友^{ほうゆうにしんあるかな このひと}に信有哉、此人。

[51-芭蕉 1687-02001, 1800]

2. 2. 6 字順転倒箇所の対応

底本テキストには、漢字列に付される返り点や振り仮名により、字順を置き換えて読むことが表示されている箇所がある。校訂本文ではそれらの指示通りの語順となるよう字順を転倒させ、漢字平仮名交じりの書き下し文として校訂した。[11] では「不申」に対しレ点の加點および「もうさず」との振り仮名があることから、字順を入れ替え、かつ、助動詞としての「ず」は平仮名に開くという校訂を行った。

[11] 〈底本〉ゆきは^{まうさずまづ}不レ申先むらさきのつくばかな

→〈校訂後〉ゆきは^{まうさずまづ}申ず先むらさきのつくばかな [51-芭蕉 1687-01001, 4460]

2. 2. 7 不読字への対応

底本テキストにおいて漢字列に傍記された振り仮名に従って読むと“不読”となる漢字文字については、コーパスのテキストにも表示しない。[12] は末尾の「而」字が不読となっていることから、不読字処理を行っている。処理前の文字列は「中納言」検索結果画面中の「原文文字列」および「原文 KWIC」に表示される。

[12] 〈底本〉将軍恵美朝臣朝獨修造^{しょうぐんえみのあそんあさかりしゆざう}而→〈校訂後〉将軍恵美朝臣朝獨修造

[51-芭蕉 1694-01021, 44850]

2. 2. 8 漢詩作品部分

漢詩部分については『日本語歴史コーパス江戸時代編』の他のコーパスと同様に、字順転倒などの処理を行わず、形態論情報を付与せず、文字列検索にのみヒットするようにした。

2. 3 掛詞情報の入力と「中納言」での表示について

本コーパスでは『日本語歴史コーパス江戸時代編』の他のコーパスと同様、本文において一つの読みに対し複数の意味解釈が可能となる掛詞の技法が用いられている箇所には形

態論情報を複数持たせている。

掛詞の箇所と範囲の認定は、底本とした『新編日本古典文学全集 71 松尾芭蕉集 (2)』および、紀行文の発句を別に収録した『新編日本古典文学全集 70 松尾芭蕉集 (1)』の頭注・注釈に従った。あわせて、国立国語研究所コーパス開発センター(鴻野知曉)編(2017)の「掛詞」に関する規程に記されているように、「原則として、掛詞の後ろの語句とのつながりで解釈する。この原則によっても意味を一つに特定できないときは、文脈全体から自然な解釈を選び、俳句の主たる文脈を作るほうを主本文とし、そうでないほうを掛詞ととった。そうしたときに掛詞は、主本文に対して「副本文」と呼ぶこととした。

[13] 蛤^{はまぐり}のふたみに^{わかれゆくあき}別行秋ぞ [51-芭蕉 1694-01049, 117870]

主本文：語彙素「蓋身」名詞-普通名詞-一般

掛詞(副本文)：語彙素「フタミ」名詞-固有名詞-地名-一般

[14] 誰が^た聲ぞ^{むこ}齒朶^{しだ}に餅^{もち}おふうしの年 [51-芭蕉 1687-02001, 20030]

主本文：語彙素「負う」動詞-一般

掛詞(副本文)：語彙素「追う」動詞-一般

[15] 桜^{さくら}より^{まつ}松は^{ふたき}二木を^{みつぎ}三月越し [51-芭蕉 1694-01019, 40210]

主本文：語彙素「松」名詞-普通名詞-一般

掛詞(副本文)：語彙素「待つ」動詞-一般/文語四段-タ行/連体形-一般

下図1は[15]における表層形「松」に掛詞の情報を期待せず「検索動作」の選択でデフォルトの「副本文を検索対象に含まない」のまま検索を試みたものである。ここで「多重化種別」を表示させると「掛詞」と表示され、当該箇所に掛詞の技法が用いられていると判るようになっている。

検索動作 設定を隠す

文脈中の区切り記号 | 文脈中の文区切り記号 # 前後文脈の語数 20

副本文 副本文を検索対象に含まない 共起条件の範囲 文境界をまたがない

列の表示 設定を隠す

☐ コーパス情報

☐ 時代名 ☐ サブコーパス名 ☒ サンプル ID ☒ 開始位置 ☐ 連番 ☒ コア ☐ 層 ☐ 層内連番 ☐ 主本文 ☒ 多重化種別

サンプル ID	開始位置	コア	多重化種別	前文脈	キー	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
51-芭蕉 1694-01019	40210	1	掛詞	#桜 より	松	は 二 木 を 三 月 越し#	マツ	松	マツ	名詞- 普通 名詞- 一般			まつ	

図1 「中納言」での掛詞箇所検出例

2. 4 外部リンク

「底本リンク」として、JapanKnowledgeの『新編日本古典文学全集』を参照できるようになっている。一方、「参考リンク」として、「笈の小文」「更科紀行」「おくのほそ道」の三作品について国文研鶴飼文庫収蔵の影印本画像をIIIF Curation Viewer経由で閲覧することを可能としている。

また、紀行文の作者・松尾芭蕉と、紀行文中の俳句・韻文の一部の読み手については、Web NDL Authoritiesへのリンクを用意した。

3. 『日本語歴史コーパス江戸時代編Ⅳ随筆・紀行』の特徴

『日本語歴史コーパス』において本コーパスと同じ時代編にある「Ⅰ洒落本」「Ⅱ人情本」「Ⅲ近松浄瑠璃」といったコーパスは、口語的性格の強い資料として、室町時代編の「Ⅰ狂言」「Ⅱキリシタン資料」からの連続した日本語の実態を捉えられるようになっていいる。それに対して本コーパスは、近世文語の様相の一端をとらえられるコーパスとして、延いては、近代文語文への連続した様相をとらえられるコーパスとして期待される。

一方、紀行文という文芸ジャンルを切り口として見た場合、すでに公開されている『日本語歴史コーパス平安時代編』に収録されている日記や随筆に相当する作品、鎌倉時代編の「Ⅰ説話・随筆」「Ⅱ日記・紀行」の作品群との比較・対照研究も考えられる。

松尾芭蕉は「笈の小文」において、「抑^{そもそも}道の日記といふものは、紀氏^{きし}・長明^{ちやうめい}・阿仏^{あぶつ}の尼^{あま}の、文^{ぶん}をふるひ情^{じやう}を盡^{つく}してより、余^よは皆^{おもかげ}倣^{おもかげ}似^{おもかげ}かよひて、その糟粕^{さうはく}を改^{あらたむ}る事あたはず」というように、紀貫之、鴨長明、阿仏尼の名を記し、それらの紀行文を称えている。紀貫之には「土佐日記」、阿仏尼には「十六日記」、そして、鴨長明には芭蕉が生きていたころは長明の手によるものと伝えられていた「海道記」がある。「東関紀行」もまた芭蕉に影響をもたらした紀行文とされている。そこで、試みに『日本語歴史コーパス』に収録されているそれらの作品および周辺作品の語種割合と、芭蕉の紀行文のそれとを比較してみれば下図2のようになる。

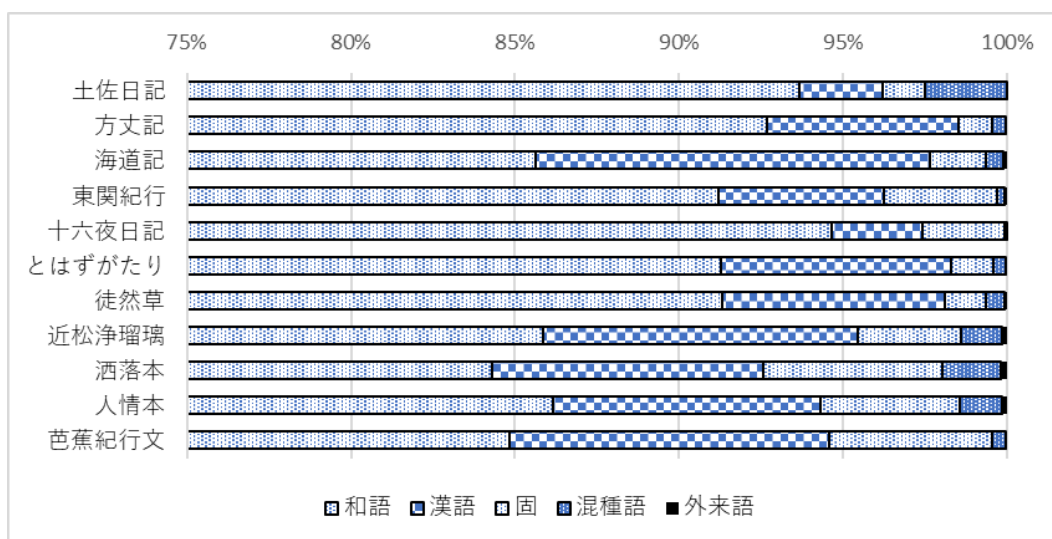


図2 芭蕉の紀行文と『日本語歴史コーパス』収録作品との語種比較

いずれの作品でも全体を占める割合の大きい和語と漢語の割合を比較した場合、芭蕉の紀行作品は今回取り上げた中古・中世の作品の中では「海道記」と、和語漢語の使用割合が似ている。また、同じ時代編に収録されている近松浄瑠璃、洒落本、人情本とも、文語・口語の文体差を超えて和語・漢語の語種割合が似ているのは興味深い。

4. 今後の課題

『日本語歴史コーパス江戸時代編Ⅳ随筆・紀行』Ver.0.4では芭蕉の紀行文を6作品収録するにとどまったが、将来的にはサブコーパス名に沿うようなかたちで、貝原益軒や本居宣長、杉田玄白、新井白石などの随筆作品を追加収録するよう計画し、データのバージョンアップを図る予定である。

【参考規程集】

池上尚（2016）「『日本語歴史コーパス 鎌倉時代編Ⅰ説話・随筆』形態論情報の概要」

https://pj.ninjal.ac.jp/corpus_center/chj/doc/morph-kamakura-2016.pdf

国立国語研究所コーパス開発センター(鴻野知暁) 編（2017）『『日本語歴史コーパス 鎌倉時代編』短単位規程集 Ver.1.0』

https://pj.ninjal.ac.jp/corpus_center/chj/doc/morph_kamakura_v1_0.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://pj.ninjal.ac.jp/corpus_center/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://pj.ninjal.ac.jp/corpus_center/chj/doc/morph-edo-2019.pdf

【参考文献】

『日本古典文学大辞典』岩波書店

『新編日本古典文学全集 70 松尾芭蕉集（1）』小学館

『新編日本古典文学全集 71 松尾芭蕉集（2）』小学館

【関連する URL】

国立国語研究所（2021）『日本語歴史コーパス』バージョン 2021.3

https://pj.ninjal.ac.jp/corpus_center/chj/

日本古典籍データセット <http://codh.rois.ac.jp/pmjt/>

IIIF Curation Viewer（人文学オープンデータ共同利用センター）

<http://codh.rois.ac.jp/software/iiif-curation-viewer/>

JapanKnowledge <https://japanknowledge.com/>

付記 本発表は国立国語研究所共同研究プロジェクト「通時コーパスの構築と日本語史研究の新展開」の成果の一部である。