

## 『日本語歴史コーパス 江戸時代編IV随筆・紀行』 Ver.0.8

## テキストの凡例と「中納言」表示項目について

2023年3月31日 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 は、Ver.0.4 として公開した芭蕉の紀行文 6 作品に貝原益軒、新井白石、室鳩巢、本居宣長、杉田玄白らの随筆 6 作品を追加したものである。随筆 6 作品のコーパス化に際しては、Ver.0.4 同様『日本語歴史コーパス (CHJ)』の他のサブコーパスと整合性のとれた形態論情報を付与するため、一部、底本を校訂している。本稿では底本に加えた校訂とそれらが検索アプリケーション「中納言」によって検索・表示される場合の各種情報についてその概要を述べる。

## 1 収録作品

『日本語歴史コーパス 江戸時代編IV随筆・紀行』 Ver.0.4 から Ver.0.8 とするにあたり追加収録した作品の一覧を表 1 に示す。短単位数は記号等を除いている。

表 1 『日本語歴史コーパス江戸時代編IV随筆・紀行』 Ver.0.8 で追加収録した随筆作品

著者	作品名 (収録部分)	作品全体の冊数 発行元など	成立年	所蔵	短単位数
貝原益軒	養生訓 (巻第一・二)	4 巻 8 冊 永田調兵衛版行	1712	九州大学医学部図書館 杏仁醫館文庫	12,998
新井白石	西洋紀聞 (上巻)	3 冊 自筆写本	1915	国立公文書館	7,943
新井白石	折りたく柴の記 (上)	3 冊 写本	1716	国立国会図書館	26,627
室鳩巢	駿台雑話 (仁集)	5 冊 前川六左衛門版行	1732	国立国会図書館	17,962
本居宣長	うひ山ぶみ (全)	1 冊 菱屋孫兵衛〈京都〉, 吉野屋仁兵衛〈京都〉他	1798	国文学研究資料館 一般	14,624
杉田玄白	蘭学事始 (上・下之巻)	2 冊 天真楼〈東京府〉	1815	早稲田大学図書館 洋学文庫	15,083
計					95,237

表 1 にあげた作品とその範囲において、原則として表紙、目録、奥付（版元の表示）を除いた全テキストを収録対象としている。序、刊記、内題、尾題の類についても収録することとし、いずれもタグ付きで要素を区別しており、それらは「中納言」検索結果の本文種別で「その他」の要素として表示される。

## 2 テキストの校訂

Ver.0.8 で追加収録した作品についても Ver.0.4 芭蕉の紀行文で施したのと同様に、一部の片仮名表記の置き換え、一部の漢字の置き換え、踊り字の校訂、合略仮名の置き換え、テキストへの補い、字順転倒箇所への校訂、不読字への対応、漢詩作品部分への対応を行った。本書では Ver.0.4 とは異なる校訂について触れる。

### 2.1 一部の漢字の置き換え

本コーパスのテキストの文字入力にあたっては JISX0213 に準拠しながら、底本での表記が現行字体と異なっている場合でも、読みが同じで字形・用法の近い字体とすることができるものは包摂・代用し表示した。

『日本語歴史コーパス』検索アプリケーション「中納言」では、検索結果における「原文 KWIC」および「原文文字列」欄で底本の表記を表示するようになっているが、「江戸時代編IV随筆・紀行」Ver.0.8 で追加収録した作品において、底本表記について文字コードを持つ文字へと正確に映すことのできなかった字体を「≡」（げた記号，U+3013）として残した。このような箇所については利用にあたり、外部リンクとして用意した底本画像を適宜参照，確認されたい。

[1] 〈コーパスの本文テキスト〉養生訓卷第一 ← 〈原文 KWIC 表示〉養生訓≡第一

(51-益軒 1712-01001,40)

[2] 〈コーパスの本文テキスト〉痞満 ← 〈原文 KWIC 表示〉痞≡

(51-益軒 1712-01015,550)

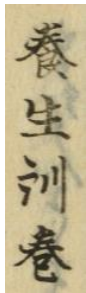


図1 例 [1] 「卷」字箇所

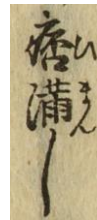


図2 例 [2] 「満」字箇所

### 2.2 字順転倒箇所への校訂

底本には、漢字文字列に対して返り点や振り仮名により字順を置き換えて読むことが指示されている箇所がある。本コーパスでは、それらの指示通りの語順となるよう字順を転倒させたり、漢字ひらがな交じりで書き下し文としたりする校訂を加えた箇所がある。校訂前の文字列は「中納言」検索画面の「原文 KWIC」および「原文文字列」に表示される。

ただし、Ver.0.4 では「原文 KWIC」および「原文文字列」において、返り点（レ，一，二点）は含まず漢字列のみ表示しているのに対し、Ver.0.8 で追加収録した作品については「原文 KWIC」および「原文文字列」に返り点（レ，一，二点）が混入し表示されていることに注意されたい。この点については、今後解決すべき課題として挙げておきたい。

## 2.3 漢字列部分

「江戸時代編IV随筆・紀行」Ver.0.8 で追加収録した作品においても、『日本語歴史コーパス 江戸時代編』に収録されている他のコーパス同様、漢字列について返り点が付されず、訓読の指示がなされていない場合はとくに形態論情報を付与することなく、品詞を「漢文」としたうえで、文字列検索にのみヒットするようにした。

[3] 三傳市虎人皆従といふ詩と (51-白石 1716-12025,1300)

## 3 「中納言」における表示項目と内容

「江戸時代編IV随筆・紀行」Ver.0.8 で追加収録した作品のコーパスの本文に付したさまざまなタグや単語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。表示については他のサブコーパスと統一を図っているため概略・省略しながら、各項目について Ver.0.4 と異なる点を中心に記す。

### 3.1 コーパス情報

#### 3.1.1 サンプル ID

『日本語歴史コーパス』に統一的なルールに従いながら、「江戸時代編IV随筆・紀行」Ver.0.8 に追加収録した全作品には表 2 に示すようにサンプル ID を付与した。

「江戸時代編IV随筆・紀行」Ver.0.8 に追加収録した作品のサンプル ID については、表 2 に示す 1-2 桁目を、すでに公開されている他の「江戸時代編」収録コーパスでの時代区分の割り振りに従い「51」（江戸時代前期）、「52」（江戸時代後期）とした。

4-5 桁目は各作品の著者名とした。

6-9 桁目の成立時期については、各書通説はまちまちだが『日本古典文学大辞典』（岩波書店）を参考としながら決定した。

11-12 桁目では、成立年が同年の作品を区別している。「江戸時代編IV随筆・紀行」Ver.0.8 では同作品の巻を区別している。途中、たとえば『養生訓 巻第二』の 11-12 桁目「02」から『西洋紀聞 上巻』の 11-12 桁目「09」へと数字のとびがある。これは、『養生訓 巻第二』以降、巻三～八があり、今後それらを追加収録する場合の空き番号としていることによる。

13-15 桁目には、作品内で出現する記事の順番に通し番号を付与している。

表2 サンプルID付与のルール

作品	1-2 桁目	3 桁目	4-5 桁目	6-9 桁目	10 桁目	11-12 桁目	13-15 桁目
	時代	区切り 記号	作品 ID	成立時期	区切り 記号	同成立年作 品の区別	同作品内の出 現順通し番号
養生訓 巻第一	51	-	益軒	1712	-	01	001~040
養生訓 巻第二	51	-	益軒	1712	-	02	001~068
西洋紀聞 上巻	51	-	白石	1715	-	09	001~017
折りたく柴の記 上	51	-	白石	1716	-	12	001~063
駿台雑話 仁集	52	-	鳩巢	1732	-	15	001~020
うひ山ぶみ	52	-	宣長	1798	-	20	001~031
蘭学事始 上之巻	52	-	玄白	1815	-	21	001~022
蘭学事始 下之巻	52	-	玄白	1815	-	22	001~025

### 3.1.2 多重化種別

「江戸時代編IV随筆・紀行」Ver0.8 追加収録作品においては、その底本で本行の漢字文字列に対し、本来の字音読みではないルビがついている場合、ルビによってテキストが複層的に読まれることが期待されていると考え、そのような箇所に対し形態論情報を複数付与している。

例 [4] では、「東方闔州」に右傍ルビとして「につぽんそうこくちう」があるところから、当該箇所には語彙素「東方」+語彙素「闔州」のほかに、語彙素「日本」+語彙素「総」（接頭辞）+語彙素「国中」の形態論情報を付与した。例 [5] では「報帖」に左ルビとして「ひきふだ」があることから、当該箇所には語彙素「報帖」のほかに、語彙素「引き札」の形態論情報を付与した。例 [6] では「行人」に右ルビとして「コウジン」、左ルビとして「ツカイ」とある。この場合、左ルビ「ツカイ」は本来の字音読みではないが、文脈上「行人」の意味を表すルビとして働いている。そのため、当該箇所には語彙素「行人」のほかに、語彙素「使い」の形態論情報を付与した。

[4] 我東方闔州（右ルビ：につぽんそうこくちう） （52-玄白 1815-22001,3150）

[5] 是は俗間にいふ報帖（左ルビ：ひきふだ）同様のもの （52-玄白 1815-22007,1290）

[6] ここに来れる、行人（右ルビ：カウジン、左ルビ：ツカヒ） （51-白石 1715-09012,430）

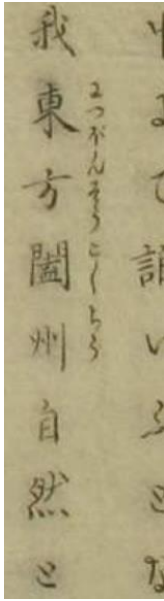


図2 例 [4]「東方闔州」箇所

右ルビ「につぽんそうこくちう」



図3 例 [6]「行人」箇所

右ルビ「カウジン」、左ルビ「ツカイ」

このように、通常の字音読みでないルビに従って付与された形態論情報は「中納言」検索結果画面における「多重化種別」列で「振り仮名」と表示される。

## 3.2 本文情報

### 3.2.1 本文種別

「江戸時代編随筆・紀行IV」Ver.0.8 で追加収録した作品では、本文種別を表3のように区別し、「中納言」検索結果の「本文種別」欄に表示する。なお、地の文は「本文種別」を空欄とした。

「江戸時代編IV随筆・紀行」Ver.0.8 の追加収録作品における本文種別とその階層表記については次の通りである。

- 「割書き」は大字表記の本文の中にあって、小字で2行に分け表記されている部分。
- 「韻文- {和歌・漢詩}」は著者の作となる和歌・漢詩。
- 「会話」は登場人物の発話箇所。
  - ・「-会話-発話引用」はさらなる会話が入れ子となっており、そこにほかの発話が引用されている箇所。
  - ・「-韻文- {和歌・漢詩}」は会話で和歌・漢詩が述べられている箇所。
  - ・「-引用」以下は、当該箇所が引用であることがわかる場合、また特に典拠がある場合には階層

を設けて区別した。

- 「引用」は地の文に対し他からの引用が行われている箇所を区別した。典拠があるとわかる場合はそれを区別した。
- 「その他」は、底本において記事本体でない序文にあたる箇所を「序」とした他、本文とは異なる要素をそれぞれ区別した。

表3 『日本語歴史コーパス江戸時代編IV随筆・紀行』 Ver0.8 に追加収録した作品の本文種別

上位階層	以下の階層	上位階層	以下の階層
割書き		引用	-その他
韻文	-和歌		-手紙
	-漢詩		-典拠・漢文
会話			-典拠・その他
	-会話-発話引用	その他	-序
	-韻文-和歌		-内題
	-韻文-漢詩		-見出し
	-引用-韻文・和歌		-日付
	-引用-典拠・和歌-韻文		-作者
	-引用-典拠・漢詩		-刊記
	-引用-典拠・漢文		-尾題
	-引用-典拠・その他		
	-引用-その他		

### 3.2.2 話者

本コーパスにおける会話文の話者名については、可能な限り本文中での呼称を用い表示している。そのため、特定の固有名詞ではない話者名が表示されることもある。なお、同時代編「Ⅰ洒落本」「Ⅱ人情本」「Ⅲ近松浄瑠璃」とは異なり、話者の性別・身分・地域・年齢といった属性は記していない。

### 3.3 作品情報

「中納言」検索結果の各欄に表示される内容について記す。

#### ●ジャンル

「ジャンル」には一律「随筆」と表示される。

#### ●作品名

「作品名」には個別の作品名が表示される。

#### ●成立年

収録作品には成立年に諸説あるものも含まれるが、『日本古典文学大辞典』（岩波書店）を参考とし成立年を定めた。

## ●巻名等

巻名等の表示については、作品ごとに次のように行う。

- 『養生訓』では、底本において行頭一字あげとすることで記事のまとまりを示している。それに基づき記事を分割し、その記事の内容がある程度把握される冒頭部分の語句を短く抜粋し表示する。
- 『西洋紀聞』については、『日本随筆大系 35 新井白石』（岩波書店）に収録される『西洋紀聞』での段落分けを参考とし、記事を分割した。その記事の冒頭部分の語句をごく短く抜粋し表示する。
- 『折りたく柴の記』については、『折りたく柴の記』（岩波文庫）を参考とし、話題のまとまりに従って記事を分割した。その記事の冒頭部分の語句を短く抜粋し表示する。
- 『駿台雑話』については、まず底本で序文と本体とを分割して表記していることからそれらを区別した。さらに、本体については「目録」に基づき記事を分割し、そのタイトル部分を巻名等の表示として用いる。
- 『うひ山ぶみ』については、冒頭の総論をひとまとまりとし「総論」とした。以後の記事は、著者自身による丸印付きのイロハ符号がつけられているのを目安に分割し、その記事の冒頭部分の語句を短く抜粋し表示する。
- 『蘭学事始』については講談社学術文庫『蘭学事始』を参考とし、話題のまとまりに従って記事を分割した。その記事の冒頭部分の語句を短く抜粋し表示する。

## 3.4 底本情報

### ●底本および底本画像リンク

「中納言」検索画面における各作品の底本表記と画像リンク欄に表示される略称は以下のようになる。

表 4 作品の底本表記

作品	底本表記	略称ロゴ
養生訓	九州大学附属図書館	九大
西洋紀聞	国立公文書館	NAJ
折りたく柴の記	国立国会図書館	NDL
駿台雑話	国立国会図書館	NDL
うひ山ぶみ	国文学研究資料館	NJIL
蘭学事始	早稲田大学図書館	Waseda

### ●ページ番号

当該例が出現する底本画像のコマ数を表示する。

## 3.5 外部リンク

### ●底本画像リンク

「中納言」検索結果の底本リンク列において略称ロゴをクリックすると用例のある画像が表示される。

#### 【参考規程・概要書】

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://clrd.ninjal.ac.jp/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://clrd.ninjal.ac.jp/chj/doc/morph\\_kamakura\\_v1\\_0.pdf](https://clrd.ninjal.ac.jp/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』テキスト・表示項目について

<https://clrd.ninjal.ac.jp/chj/doc/abstract-chikamatsu-2020.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>

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』副本文の形態論情報の概要』

<https://clrd.ninjal.ac.jp/chj/doc/morph-at-chikamatsu-2020.pdf>

松崎安子 (2021) 『『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) テキストの凡例と「中納言」表示項目について

<https://clrd.ninjal.ac.jp/chj/doc/abstract-zuihitsu-202103.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』テキストの凡例と『中納言』表示項目について』(2019年3月29日更新)

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

村山実和子 (2019) 『『日本語歴史コーパス 江戸時代編 II 人情本』テキストの凡例と『中納言』表示項目について』 <https://clrd.ninjal.ac.jp/chj/doc/abstract-ninjo-2019.pdf>

#### 【参考文献】

岩波文庫『折りたく柴の記』(岩波書店)

講談社学術文庫『本居宣長『うひ山ぶみ』全訳注』(講談社)

講談社学術文庫『養生訓』(講談社)

講談社学術文庫『蘭学事始』(講談社)

『日本古典文学大辞典』岩波書店

『日本思想大系 35 新井白石』(岩波書店)

日本随筆大成『駿台雑話』(吉川弘文館)

#### 【関連する URL】

九州大学附属図書館 <https://www.lib.kyushu-u.ac.jp/ja>

国文学研究資料館 <https://www.nijl.ac.jp/>

国立公文書館 <https://www.archives.go.jp/>



国立国語研究所 (2023) 『日本語歴史コーパス』バージョン 2023.3 <https://clrd.ninjal.ac.jp/chj/>

国立国会図書館 <https://ndlonline.ndl.go.jp/>

国書データベース <https://kokusho.nijl.ac.jp/>

人文学オープンデータ共同利用センター IIF Curation Viewer <http://codh.rois.ac.jp/software/iiif-curation-viewer/>

早稲田大学図書館 <https://www.waseda.jp/library/>