

『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

2018年3月30日 村山実和子

2019年3月29日 村山実和子 更新

1. はじめに

『日本語歴史コーパス 江戸時代編Ⅰ洒落本』は、洒落本大成編集委員会編『洒落本大成』（1978-88 中央公論社）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（詳しくは市村・河瀬・小木曾 2013, 市村 2014 参照）。その多様な情報が反映されたものが中納言版の本コーパスである。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。なお、本コーパスでは、研究上必要と思われる情報をできるだけ底本の状況に即して記述するよう努めたが、完全に反映できているわけではなく、また誤りが残存している可能性もある。そのため、適宜「ページ番号」を基に底本の本文を確認されることを推奨する。

2 テキストの凡例

2.1 外字等の処理

2.1.1 外字

本文テキストの文字入力には JISX0213 に準拠している。読みが同じで字形・用法の近い文字、または適切なものがないと判断した場合は「≡」に置き換えて入力した（表 1 参照）。

表 1 『江戸時代編Ⅰ洒落本』における外字一覧

	振り仮名	作品名	Unicode 番号	
1	≡妃	よきおんな	聖遊廓	U+36F9
2	≡	けん	聖遊廓	- さんずい+上「酉」+ 下「華」
3	≡閣		陽台遺編・姁閣秘言	U+21742
4	金≡	らん	異本郭中奇譚	- 左「糸」+右「闌」
5	≡	わけ	無論里問答	U+4C0E
6	≡	カタコト	総籙	- 左「言」+右「包」
7	≡	アラタメ	総籙	- 左「β」+右「文」
8	≡筆	はけ	総籙	- U+29B7B の「日」部分 が「夕」

9	≡籠	しんこ	仕懸文庫	-	
10	≡	つりばり	北華通情	U+28C56	
11	並≡	かどなみ	北華通情	U+95B5	
12	≡	ぞめき	北華通情	-	U+6056 の「メ」が、斜めになった「卩」
13	≡行	ぞめき	北華通情	-	同上
14	≡	ぼさつ	南遊記	-	上「卩」+下「卩」
15	≡ふ	くひちが	南遊記	-	左「齒」+右「苦」
16	≡強	つん/\	南遊記	U+602E	
17	≡		当世嘘之川	-	
18	≡とと	どつ	花街鑑	U+35E2	
19	均≡	うるはしき	花街鑑	U+74C5	
20	≡子		郭中奇譚	-	左「月」+右「表」

また、以下の形式については、他の仮名で代用させた。

- ① てつつり（箱まくら）※「つ」に白圏点（原文文字列「て≡り」）
- ② ゑつさいこらつさい（郭中奇譚）※「さ」に白圏点（原文文字列「エツ≡イ」）

2.1.2 絵文字・記号

絵文字や記号など、電子化にあたって置き換えが不可能なものについても、同様に「≡」で表示している。未知語扱いとして、「絵文字・記号等」の品詞を与えた。検索結果の「ページ番号」をもとに、『洒落本大成』の該当ページを確認するか、「外部リンク」が表示されるものについては版本画像もご参照いただきたい。

2.2 テキストの校訂

2.2.1 濁点

〔1〕濁点の付与

濁音が期待される箇所には濁点が付されていない場合は、諸資料を参考に検討の上、必要な箇所は濁点を補った。ただし、清濁両形あり判断に迷う場合には極力濁点を付与しない方針をとった（「室町時代編 狂言 I」に同じ）。濁点を補う前の文字列は、「原文文字列」に表示される。

【例】

- (1) 〈原文〉をつとおまへがたの事しやない
 〈入力〉をつとおまへがたの事じやない（当世嘘之川）

〔2〕濁点の保持

清音が期待される箇所には濁点が付与されている場合、極力その形式を保持することに努め

た。それにより形態論情報の付与が困難な形式は、品詞「解釈不明」としている。

【例】

- (2) 〈原文・入力同じ〉是は先生どふじやどづつとは入れば（当世左様候）
※格助詞「と」と見られるがママ
- (3) 〈原文・入力同じ〉今でもほうふでいふものは。しつたかさ（総籬）
※「ほうほう（方々）」と見られるがママ
- (4) 〈原文・入力同じ〉せんだいかふいへばおかしいけれども。（北華通情）
※「ぜんたい（全体）」と見られるがママ
- (5) 〈原文・入力同じ〉終にこの冊を耕といべども花に風。（北華通情）
※「いへ（言う）」と見られるがママ
- (6) 〈原文・入力同じ〉是生滅法とはしり出じ（竊潜妻）
※「はしり出し（走り出す）」と見られるがママ

2.2.2. 踊り字

仮名 1 字分の踊り字（ゝ、ゞ、ゝ、ゞ）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。対して、2 字分以上に相当するくの字点、また漢字の連続は、置き換えの対象としない。また、振り仮名内の踊り字も置換しない。

【例】

- (7) 〈原文〉鳥羽瀬のげいしやもゝへよばれるかへ
〈入力〉鳥羽瀬のげいしやもこへよばれるかへ（仕懸文庫）
- (8) 〈原文・入力同じ〉何にもいわぬ是じやゝと手を合す（竊潜妻）

2.2.3 カタカナ

底本本文においてカタカナで表記された箇所は、平仮名で表示した。変換前の文字列は「原文文字列」に表示される。ただし、振り仮名内のカタカナは置換していない。

【例】

- (9) 〈原文〉ホンニあきれんすによウ
〈入力〉ほんにあきれんすによう（郭中奇譚）

2.2.4 誤字・脱字・衍字等

〔1〕校注者による校訂

『洒落本大成』は基本的に底本に忠実なテキストであり、ほとんど校訂箇所は見られない。今回コーパス化した作品のうち、校注者による補読が行われていた箇所は以下の 3 例であり、補ったことがわかるようにタグ付きで入力している。文字を補う前の文字列は、「原文文字列」に表示される。

【例】

- (10) 隣の明家を。詠めながら愁懐と。(花街鑑)
(11) 右側の七けん軒をならぶ。(総籬)
(12) ゑん二郎おしやう。むかふの兵庫やに居るの (同上)

〔2〕コーパス担当者による校訂

コーパス本文について、なるべく原態を維持するよう努めたが、検討のうえ必要と判断された箇所については、タグ付きでテキストの訂正を行った。訂正を行った箇所は以下の通り。いずれも、「原文文字列」には変換前の文字列が表示される。

【例：脱字】

- (13) 〈原文〉 おいとし様やとあいけうあるとかたりがらそこらをたづねる
〈入力〉 おいとし様やとあいけうあるとかたりながらそこらをたづねる (当世嘘之川)
(14) 〈原文〉 きやしな手でのふれんをまくり (興斗月)
〈入力〉 きやしやな手でのふれんをまくり (興斗月)
(15) 〈原文〉 どふぞ五六をとつおくれと (風流裸人形)
〈入力〉 どふぞ五六をとつておくれと (風流裸人形)

【例：衍字】

- (16) 〈原文〉 茶よりりやあ酒を
〈入力〉 茶よりりやあ酒を (南閨雑話)
(17) 〈原文〉 なんのじつとおとおとなしうして居なされ
〈入力〉 なんのじつとおとおとなしうして居なされ (色深狭睡夢)

2.2.5 テキストの置き換え・補い

〔1〕本行文字列と振り仮名の置き換え

Ver.0.9において、処理単位上の問題により、以下のように振り仮名の文字列をタグ付きで本文と置き換えた箇所があった。

【例】(「|」は短単位の境界を示す)

- (18) 〈原文〉 ^{てがたがやける}不首尾成 → 〈入力〉 | てがた | が | やける | (南遊記)
(19) 〈原文〉 大低 ^{きのうつくしい}柔和お方 → 〈入力〉 | き | の | うつくしい | (同上)
(20) 〈原文〉 ^{てを}百折千磨しくどけども → 〈入力〉 | て | を | つくし | (誰か面影)

図1には(17)の検索結果を示した。このように仮名に開いた形式については、置換前の文字

列が「原文文字列」および、「原文 KWIC」列に表示されていた。

図1 原文文字列および原文 KWIC での表示例（『中納言』検索結果）

キー	後文脈	語彙素 読み	語彙素	語形	品詞	原文文字列	振り仮名
きめろつくしい	お方じやないはへといふたればないいやそふじやない法起 = で大の	ウツクシイ	美しい	ウツクシイ	形容詞-一般	(柔和)	きのうつくしい
柔和	お方じやないはへといふたればないいやそふじやない法起 = で大の						

この箇所について、データベースの拡張により、原文を維持したまま複数の形態論情報を付与できるようになった。それに伴い、上記の(18)~(20)のような例について、ver.1.0では以下のように更新される（詳細は別紙「アップデートの概要について」に譲る）。

【例】

- (21) <原文> ^{てがたがやける}不首尾成 → <入力> | ^{てがたがやける}不 | 首尾 | 成 | (南遊記)
- (22) <原文> 大低 ^{きのうつくしい}柔和 お方 → <入力> | ^{きのうつくしい}柔和 | お | 方 | (同上)
- (23) <原文> ^{てをまっく}百折千磨しくどけども → <入力> | ^{てをまっく}百折 | 千磨 | し | (誰か面影)

例外的な処置として、『聖遊廓』（『洒落本大成〈3〉』pp.333-334）の「かきをき」（以下、(21)の例）は、梵字に片仮名の振り仮名を付した表記であったが、振り仮名を平仮名表記した本文として入力することで対応した。原文文字列に梵字は反映させず、空欄表示としている。

(24)かきをき

うれしからぬみらいとてもながらへそいもならず。ままならぬ。あかぬしやばなれど。けふをかぎりのいのちぞと。をもひあきらめ。しぬるかくごのたびころも。こころせくまなまなかに。なれまじひとになれそめて。なばかりのこす。ふみづきのそらいつまでも。そひはふべきとをもひしに。をもひしかひも。しでのやまぢに

〔2〕振り仮名にもとづく補読

本文中、期待される送り仮名・活用語尾・助詞等が振り仮名に含まれる場合、タグ付きで本文に補った。その際、振り仮名自体は、元のまま保持するようにした。置換前の文字列は、「原文文字列」に表示される。

【例】

- (25) 〈原文〉 さればこそ 銀 猫は抛ども → 〈入力〉 | 銀 | の | 猫 | (聖遊廓)
- (26) 〈原文〉 其幸を 抛。 → 〈入力〉 | 抛 | て | (色深狹睡夢)
- (27) 〈原文〉 其音の 訛 を知しめんが為なり → 〈入力〉 | 訛 | る | を | (総籬)
- (28) 〈原文〉 毒をしつて 恐慎 には。 → 〈入力〉 | 恐慎 | ん | (仕懸文庫)

また例外的に、下記のように文字列を補った箇所がある。

【例】

- (29) 〈原文〉 里仁為美 扱不処仁焉得知
〈入力〉 里仁為美 扱不処仁焉得知 さとはじんなるをよしとすゑらんでじんにおらずん
ばなんぞちをゑん (聖遊廓)
- (30) 〈原文〉 十目所視 十手所指
〈入力〉 十目所視 十手所指 ひとのみるてまへをおもふて (同上)
- (31) 〈原文〉 会者常離 臨命終時不随者
〈入力〉 会者常離 臨命終時不随者 あふはわかれしばしのたのしみじや (同上)
- (32) 〈原文〉 悦涙閨情濃則 我意雖真大快 亦想对余人如此 怨氣如烈火然也
〈入力〉 悦涙閨情濃則 ないてきすのは我意雖真大快 しんじつま ことにうれしいけれど
も亦想对余人如此 またひとにもこうかとおもへば 怨氣如烈火然也 めつほう かいにはら
がたつ (色深狹睡夢)

〔3〕 臨時的な補い

コーパス本文において、テキストを補う際には原則的に振り仮名にもとづいているが、以下の 2 形式については処理単位上の問題から、例外的にタグ付きで仮名に置き換え／文字を補っている。「原文文字列」には変換前の文字列が表示される。

【例】

- (33) 〈原文〉 大空は座敷へ 来り → 〈入力〉 | 来 | たり | (聖遊廓)
- (34) 〈原文〉 少しはうさもはれ 候半 → 〈入力〉 | 候は | ん | (異本郭中奇譚)

2.2.7. 漢文

漢文（風）の箇所について、返り点や振り仮名などで訓読が可能な漢文等については、訓読した形を本文とした。返り点のない漢文箇所や、長大な部分（跋文等）は、「未知語」として扱い、品詞を「漢文」とした。また置き字についても、品詞「漢文」としている。

【例】

- (35) 〈原文〉御評判よろしく 奉願上候 → 〈入力〉 願上 奉候 （誰か面影）
- (36) 〈原文〉眉、如_レ翠羽_レ肌、如_レ白雪_レ… → 〈入力〉眉は翠羽の如く肌は白雪の如く
（総籙）
- (37) 〈原文〉以南鐐一片換日柄一日 → 〈入力〉ママ （短華薬葉）

〔1〕 語彙素・語彙素読み

「**語彙素**」は単語の各種語形・活用形・書字形（表記）を統合した辞書の見出しレベルの階層であり，一般的な漢字・仮名で表記される。「**語彙素読み**」はその読みを**カタカナ表記**したものである。語彙素で検索することで，同語彙素内の各種語形・活用形・書字形等の異なるもの一括して取得することができる(図2は語彙素読み「チョット（副詞「一寸」)」による検索結果である)。

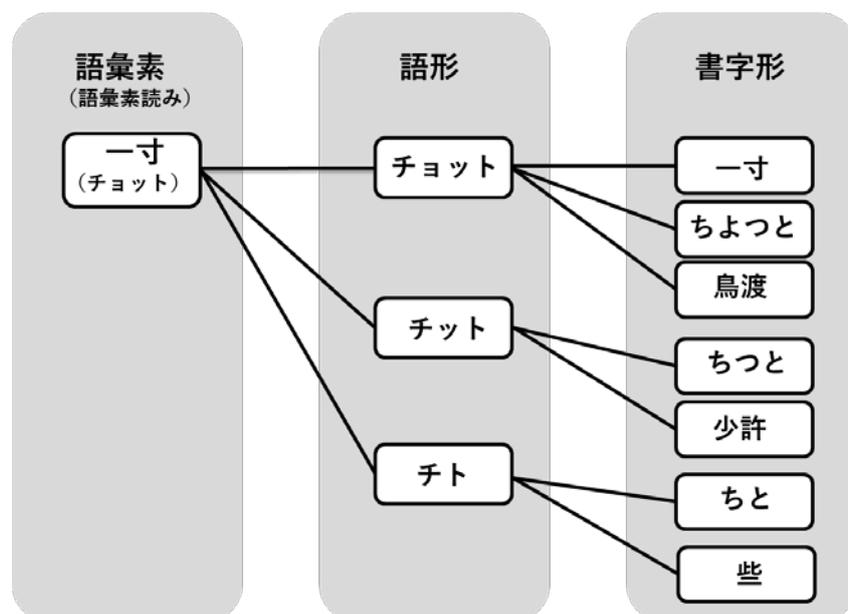
〔2〕 語形

「**語形**」は，異語形を区別するレベルであり，例えば「チット」「チョト」など音形の異なるものは，語彙素「一寸」の語形として認定される。(図3参照)

〔3〕 書字形

「**書字形**」は異表記を区別するレベルである。同語形でありながら，活用語尾を除いた箇所
に別の文字符号が与えられる場合，それぞれ別の書字形となる。(図3参照)

図3 語彙素「一寸」の語彙素・語形・書字形



〔4〕 品詞

学校文法における形容動詞は，語幹は「**形状詞**」，語尾は「**助動詞**」に分割されている。

〔5〕 活用型

文語活用として処理されているものには「文語下一段」のように「文語」が表示されるが，口語活用には「下一段」のように「口語」は表示されない（「文語四段」「五段」も同様）。

〔6〕活用形

「活用形一小分類」の「融合」は、前部の活用語に後続する助詞等が取り込まれるなどし、短単位として分割しがたいものである。例えば形容詞「悪い」の連用形「悪く」に助詞「は」の接した「悪くは」の変化した「わるかあ」は、下記のように処理されている。

【例】

(38) | わつち | が | わるかあ | あやまり | やしやう | (仕懸文庫)

→形容詞「悪い」・形容詞-一般・連用形-融合

〔7〕原文文字列

踊り字・濁点・片仮名等を校訂する前の本文は「原文文字列」および「原文 KWIC」に表示される (図 4)。

図 4 本文テキスト (出現書字形) と原文文字列

サンプル ID	開始位置	前文脈	キー	後文脈	語彙素	品詞	活用型	活用形	原文文字列	振り仮名
52-洒落 1793_01066	125060	いねへかの 観 何さ根柢はくだして来やしたいさくきのねへのさ	ほんに	此ごろ) だとりとんなすつたじやあねへか おや判かへ よび出し判はだれをいれなすつた 伝	本	副詞			ホンニ	
		此前まうこう人もまし かわりい足でもつ ちやア ねへか 観 何さ根柢をよくだして来やしたいさくきのねへのさ	ホンニ	此ごろひとりとんなすつたしやアねへか おや判かへ よび出し判はだれを入しなすつた 伝						

キーに表示された「ほんに」が、原文文字列では片仮名表記であることが示される。また、前後文脈の原文 KWIC 表示を見ると、「ただして」が「たどして」、「此ごろ」が「此ころ」であったことがわかる。ただし、これらはいくまで補助的な表示であるため、底本の詳しい状況を確認する場合は、ページ数を元に底本を直接ご参照いただきたい。

〔8〕振り仮名

本行の文字列の右側に小書きされた文字については、「振り仮名」列に表示される。振り仮名については、片仮名や踊り字等への校訂を行っていない。

なお洒落本では、たとえば本行の漢語や外国語に対応する説明、方言形に対応する語など、注釈的な語句が本行の左側に付されることがある。ただし現状のコーパスでは本文をひとつに定める必要があるため、表示対象となっていない。元の XML データには、<IRuby> (左ルビ) というタグで情報が保持されている。

3.2 本文情報

次に、本文の会話・ト書きの別、話者等の表示方法について述べる（表2）。

〔1〕 本文種別

性質の異なる本文を区別し、「本文種別」として表示している。具体的な分類は下記のとおりである。また、それぞれ和歌や俳句、歌等であることが明らかな箇所には、「会話-韻文」のように、「-韻文」が表示される。なお、割書き内の会話文風のもの、認定の難しさや性質の違いを考慮し、「会話」と認定していない。

表2 本文種別の対応

本文種別	説明	
会話	本文中の会話部分	
割書き	割書きされた本文	
地の文（明示なし）	地の文相当のものから割書きをのぞいた本文	
その他	注釈	文頭に記された注記
	引用	台詞以外の引用箇所。 「-典拠」は、文献等の引用を表す。また「引用-典拠・和歌」のように、引用の内容が判明する箇所には「・」で区切り、表示した
	内題	各作品の内題
	尾題	各作品の尾題
	見出し	序・跋等作品内の section を表す
	話者	著者による話者表示
	会話注記	会話文頭に示される「うた」や「フシ」などの表示

〔2〕 話者

本コーパスにおける会話文について話者表示の統一をはかり、「話者」欄に表示した。

【例】

〈原文〉客 → 〈「話者」列の表示〉客志幸（同一作品内で統一）

そのほか、「性別」「身分」「地域」「年齢」についての情報を判明する限りで付与している（表3）。このような話者の属性の詳細は、これまで公開されたCHJ収録作品には付与されていないものである。そのため中納言の検索結果では、「話者」列内に、すべての属性が表示されるように対処した（表4）。なお、各情報はハイフンでつないで表示させており、検索結果のダウンロード後に、個々人で加工・利用することも可能である。

表 3 話者名と同列内に表示される属性情報

種別	説明	記述内容
性別	話者の性別を記述。 「その他」は、複数人の会話文や、動物以外等による性別不明の生命体（例えば『鳩翁道話』における「松茸」）などを想定している。	男，女，その他
身分	話者の身分・職業等を記述。娼妓／芸妓，店員，客（【】にて詳細），その他（【】にて詳細）等を大別。現状は洒落本に特化した記述であり，汎用性は薄い。今後も妥当性等を検討し，改良する予定である。	娼妓・芸妓，店の者，禿，太鼓持，使用人，客，客【上層・むすこ・通人】，客【武士】，客【医者・学者等】，客【町人】，その他，その他【町人】，その他【上層町人】，その他【武士】，その他【医者・学者等】，その他【神・仙人・僧侶等】，その他【俠者】
地域	話者の使用言語（出身地）について，本文記述や会話の言葉遣い等から，判明する限り記述。「上方板洒落本における江戸話者」などの抽出を想定している。	江戸，上方，田舎
年齢	子供，老人など，判明する限りで，比較的有標な事項のみを記述。今後も妥当性等を検討し，改良する予定である。この情報は付与されないこともある。	子供，老人，年少【こめろ・若者・でっち】，年長【母親など】

表 4 出現書字形・本文種別・話者の表示例

作品名	風流裸人形	
出現書字形	本文種別	話者
客	話者	
おき	会話	客志幸-男-客-上方-
やい	会話	客志幸-男-客-上方-
の	会話	客志幸-男-客-上方-
ゆき	話者	
それ	会話	【遊女】ゆき-女-娼妓・芸妓-上方-
でも	会話	【遊女】ゆき-女-娼妓・芸妓-上方-
ほしい	会話	【遊女】ゆき-女-娼妓・芸妓-上方-
もの	会話	【遊女】ゆき-女-娼妓・芸妓-上方-

3.3 作品情報

〔1〕 ジャンル

「ジャンル」には一律に「洒落本」が表示される。これは、今後『江戸時代編』として、別ジャンルの作品が追加されることを想定したものである。

〔2〕 作品名

「作品名」には、個別の作品名が表示される（旧バージョンでは「洒落本大成」が表示されていた）。原則として『洒落本大成』の表示に従ったが、一部省略したものもある。適宜、CHJ の『江戸時代編 概要』ページよりダウンロード可能な作品リスト〈http://pj.ninjal.ac.jp/corpus_center/chj/edo.html〉、底本の『洒落本大成』をご参照いただきたい。

〔3〕 成立年

「成立年」には個別の作品の成立年が表示される。

【訂正】

サンプル ID : 52-洒落 1820_01048 『(河東方言) 箱まくら』について、成立年を 1820 年 → 1822 年に訂正。

〔4〕 巻名等

「巻名等」は空欄表示となる（旧バージョンでは個別の作品名が表示されていたが、前述のとおり「作品名」に表示されるよう変更した）。

〔5〕 部

「部」には、作品の舞台 {大坂／京都／江戸} の別が表示される。これまで CHJ では、地域性についての情報が付与された作品が無かったため、暫定的に「部」列を利用した。検索対象を選択する際、はじめから地域を選択することも可能である（図 5）ほか、特に指定しない場合、検索結果のダウンロード後に、地域別の集計をすることもできる。

江戸 京 非京

洒落本大成
 江戸 京都 大坂

本文種別
 会話 割書き 引用 地の文ほか その他

図 5 『中納言』における検索対象選択画面

3.4 底本情報

〔1〕 底本

「底本」には、「洒落本大成 <2>」のように対応する底本と巻名とを表示した。

〔2〕 ページ番号

【Ver.0.9】「ページ番号」には、対応する底本のページ番号、ならびに丁数が表示される。

【Ver.1.0】「ページ番号」には、**対応する底本のページ番号のみが表示される。**

3.5 外部リンク

30 作品中、26 作品（『原柳巷花語』『誰が面影』『異本郭中奇譚』『新月花余情』をのぞく）については、参照可能な版本画像が公開されている。この項目には、検索結果に応じて、その参照先が表示される。参照先の表示内容は以下のとおり。

表 5 外部リンク先一覧

表示名	参照先
Niji	国文学研究資料館 新日本古典籍総合データベース https://kotenseki.nijl.ac.jp/
Ninjal	国立国語研究所 日本語史研究資料 http://dglb01.ninjal.ac.jp/ninjaldl/
Ninjoji	大阪大学附属図書館 忍頂寺文庫 http://www.museum.osaka-u.ac.jp/db/ninjoji/
UT-Katei	東京大学総合図書館 霞亭文庫 http://kateibunko.dl.itc.u-tokyo.ac.jp/katei/index_srch.html
UT-Kokugo	東京大学文学部国語研究室 資料画像 http://kokugo.l.u-tokyo.ac.jp/
Waseda	早稲田大学図書館 古典籍総合データベース http://www.wul.waseda.ac.jp/kotenseki/

なお、参照の際には以下の点に注意されたい。

- ・『洒落本大成』の底本と、参照画像の版本は必ずしも一致するものではない
- ・参照用画像が存する作品であっても、落丁により対応する丁が表示されないことがある

附録1 本コーパスのタグセット

要素 (タグ) 名	説明
<text>	1 作品全体
<front>	前付相当の箇所 (序文等)
<body>	主本文相当の箇所
<back>	後付相当の箇所 (跋文, 刊記等)
<article>	1 記事の範囲 (「回」相当)
<titleBlock>	記事とは認められない, <text>直下レベルでの表題周り
<p>	段落を表す。タイトルや署名等を除く主本文
<block>	記事中のタイトルなど, 主本文とは切り分けたい段落要素
<speech >	ひとまとまりの会話文。本タグに話者情報を付与。
<quotation>	文献等からの引用や手紙など。
<warigaki>	割書き箇所。
<s>	文
<verse>	謡などの節付け箇所や和歌など韻文であることが明確な箇所
<delivery>	会話文の様式等を指定する記述
<speaker>	話者の表示
<corrSpan>	振り仮名等により文字列の置き換えを行った短単位以上の箇所
<hi>	小書き・傍線・囲みなどの文字列に対する装飾
<SUW>	語 (短単位)
<IRuby>	本行の左側に振られた振り仮名等の文字列
<ruby>	本行の右側に振られた振り仮名文字列。
<add>	本文の補入箇所
<kanbun>	訓み下す際文字位置を置き換えた漢文等の箇所
<vMark>	底本原文が濁点無表記であった箇所
<odoriji>	底本原文が1字分の踊り字であった箇所
<corr>	誤字・脱字・衍字等の本文の修正
<g>	外字・絵文字等準拠する文字セットでは表示できない文字
<char>	1字を表す単位, @script="カタカナ"で, カタカナ表記箇所に使用
<info>	本文テキストに割って入れられなかった記号, 丁付情報等
<pb><lb>	底本の改ページ位置・改行位置
<opb>	原本画像の丁や画像リンクとの対応

参考文献

- 市村太郎(2014)「近世口語資料のコーパス化—狂言・洒落本のコーパス化の過程と課題—」『日本語学 11 月臨時増刊号 日本語史研究と歴史コーパス』33-14 明治書院
- 市村太郎 (2015)「ひまわり版「洒落本コーパス 0.5」利用案内」
http://pj.ninjal.ac.jp/corpus_center/chj/doc/sharebon0.5-doc.pdf (2018年3月28日確認)
- 市村太郎 (2016)「『江戸時代編』の構築と課題」『日本語学会 2016 年度春季大会予稿集』日本語学会
- 市村太郎・小木曾智信 (2016)「文書構造を利用した近世期洒落本の形態素解析」『言語処理学会第 22 回 年次大会発表論文集』言語処理学会
- 市村太郎・河瀬彰宏・小木曾智信(2012)「近世口語テキストの構造化とその課題」『情報処理学会研究報告 人文科学とコンピュータ研究会報告』 2012-1
- 市村太郎・村山実和子 (2017)「洒落本コーパス構築の試行」『国立国語研究所論集』12
- 小木曾智信 (2016)「『日本語歴史コーパス』の現状と展望」『国語と国文学』93-5
- 小木曾智信 (2017)「多重の読みを持つテキストのコーパス化」『言語資源活用ワークショップ 2017 発表論文集』国立国語研究所
- 小木曾智信・市村太郎・鴻野知暁 (2013)「近世口語資料の形態素解析の試み」『第 4 回コーパス日本語学ワークショップ予稿集』国立国語研究所
- 小椋秀樹・小磯花絵・富士池優美・宮内佐夜香・小西光・原裕 (2011)『『現代日本語書き言葉均衡コーパス』形態論情報規程集 第 4 版 (上) (下)』国立国語研究所
- 国立国語研究所コーパス開発センター(市村太郎ほか)編(2015)『ひまわり版「洒落本コーパス」(日本語歴史コーパス江戸時代編)』http://pj.ninjal.ac.jp/corpus_center/chj/edo.html#share (Ver. 0.5)(2018年3月23日確認)
- 村山実和子・小木曾智信・中村壮範 (2017)「形態論情報の多重化による洒落本コーパスの質的拡張」『第 114 回 CH 研究発表会 論文集』言語処理学会
- 国立国語研究所コーパス開発センター (市村太郎ほか) 編 (2015)『ひまわり版「洒落本コーパス」(日本語歴史コーパス江戸時代編)』
http://pj.ninjal.ac.jp/corpus_center/chj/edo.html#share (Ver. 0.5) (2018年3月20日確認)