

『日本語歴史コーパス 室町時代編Ⅱ キリシタン資料』 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 片山久留美

1. はじめに

『日本語歴史コーパス 室町時代編Ⅱ キリシタン資料』は、大英図書館に所蔵されている以下の資料を底本として作成したコーパスである。

[Nifon no cotoba to historia uo narai xiran to fossuru fito no tame ni xeua ni yaua ragueta ru Feige no monogatari.](#)

[Esopo no fabulas : Latinuo uaxite Nippon no cuchia nasu mono nari.](#)

(大英図書館蔵 請求記号 Or.59.aa.1)

以下ではそれぞれ『天草版平家物語』『天草版伊曾保物語』と呼ぶ。これらは口語体・ポルトガル式ローマ字によって書かれており、キリシタン資料の中でも当時の口語・音韻等を知ることができる点で重要な資料として位置付けられている。

本コーパスの特徴として、ポルトガル式ローマ字によって表記されている原本の本文を、独自の漢字仮名交じりテキストに変換して形態論情報を付与していることが挙げられる。これはローマ字表記のままでは UniDic による形態素解析が困難であること、また『日本語歴史コーパス』の他のサブコーパスの検索結果と並べて表示した際に視認性が下がることなどの理由から漢字仮名交じりテキストが必要となるためである。『日本語歴史コーパス』のサブコーパスの一つとして他のサブコーパスとの整合性を保ちつつ、極力原本のローマ字表記を生かした漢字仮名交じりテキストの作成を目指した。

一方で原本のローマ字表記も日本語研究上欠かすことのできない情報であるため、本コーパスでは漢字仮名交じりテキストと同時に原本のローマ字情報も参照できるようにした。

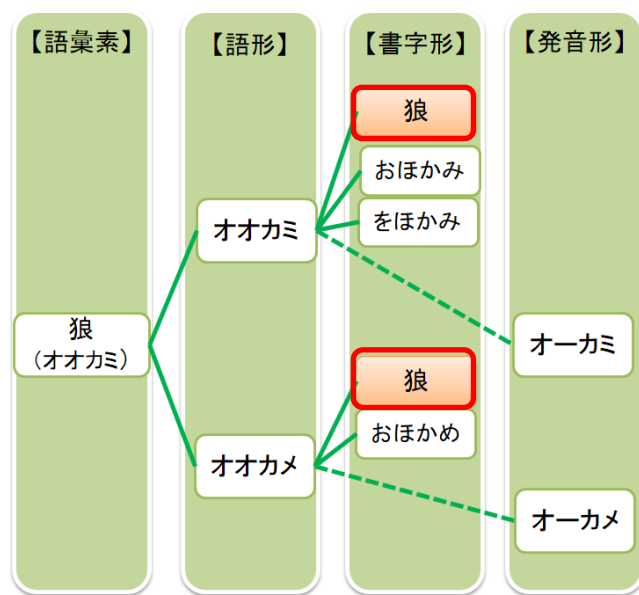
この文書では本コーパスの漢字仮名交じりテキストの性質と、原文文字列の概要、検索アプリケーション「中納言」における表示項目について例を挙げながら述べる。

2. テキストの凡例

2. 1 漢字仮名交じりテキスト

本コーパスではローマ字表記の原本を漢字仮名交じり表記へと変換したものを本文テキストとして使用している。漢字仮名交じりテキストに変換する際には UniDic の語形代表表記の情報を使用した。

UniDic の階層構造には、辞書の見出し語に相当する「語彙素」、異語形を区別する「語形」、異表記を区別する「書字形」、発音を区別する「発音形」の各レベルがある。「語形代表表記」とは、各語形の下に登録されている書字形のうち、最も代表的な表記と考えられるもののことを指す。主に『日本国語大辞典第二版』の表記を参考にして決定しており、地名や人名などの固有名詞を除くすべての語形に対して登録されている。



【図】 UniDic の階層構造（赤枠が語形代表表記）

では「ジョウ」の語形代表表記を「判官」としているため本コーパスでは「右馬の判官」という表記になっている。

UniDic はもともと BCCWJ など現代語を対象としたコーパスのための辞書として作成されたという経緯があるため、語形代表表記もやや現代語に寄った表記に設定されていることがある。そのため室町時代編にはそぐわない表記に見える場合もあるが、そのような場合も解釈を加えずに、修正等はしないという方針を取った。たとえば『天草版平家物語』の語り手の一人である「ウマノジョウ」は当時の官司の通例からは「右馬の允」という表記が妥当であると考えられるが、UniDic

このように語形代表表記を使用してテキストを組み上げることで、同じ語彙素・語形であれば常に同じ表記となり、揺れの少ない斉一なテキストを自動的に作成することが可能となる。語形代表表記によって作成したテキストがどのようなものか、以下に具体例を示しつつ概観する。

● 仮名遣い

語形代表表記を使用することにより、原本のローマ字表記を反映した表音的な性質の強い仮名遣いとなっている。ただし意志推量形の表記などに一部歴史的仮名遣いに近い部分もある（後述）。

【例】 namida uo vosayete → 涙を押さえて

語彙素「押さえる」の語形「オサウ」（文語下二段 - ハ行）の語形代表表記「押さう」

【例】 saru monoguruuaxij cotomoya arurō → 然る物狂わしい事もや有るらう

語彙素「物狂わしい」の語形「モノグルワシイ」の語形代表表記「物狂わしい」

【例】 auoredomo auoredomo → あおれどもあおれども

語彙素「あおる」の語形「アオル」の語形代表表記「あおる」

【例】 buxi no rōdō no narai nareba → 武士の郎等の習いなれば

語彙素「習い」の語形「ナライ」の語形代表表記「習い」

● 長音符の処理

長音を表す「à」「â」は以下のように「あ」と翻字した。

【例】yà → やあ gozàru → 御座ある Hà → はあ

● 語形代表表記の例外

・開合・四つ仮名の違例

語形代表表記によって表記を定めると、原本で書き分けられている四つ仮名や開合の別が明示できない場合には語形代表表記に修正を加えた。

【例】 **Mazzu** Feiqemonogatari no caqifajime niua → 先づ平家物語の書き始めには語彙素「先ず」の語形「マズ」の語形代表表記は「先ず」だが、原本の「zzu /zu」の書き分けを反映して上記のように修正

【例】 Feiqe ua **mučõ** no yama ni gin uo totte → 平家は向かうの山に陣を取って語彙素「向こう」の語形「ムコウ」の語形代表表記は「向こう」だが、原本の「õ」（開長音）の表記を反映して上記のように修正

※本コーパスでは基本的にオ段開長音（õ）は「ア段の仮名＋う」、合長音（ô）は「オ段の仮名＋う」と翻字している。

・代名詞・連体詞の表記

代名詞や連体詞など、語形代表表記のままでは複数の読みが想定され語形の特定が困難になってしまうものについては一部表記を改め仮名に開いた。

【例】「彼の」：「アノ」・「カノ」の二つの語形代表表記となっており、語形の判別が困難になるためそれぞれ「あの」「かの」と仮名に開いた。

● 意志推量形の表記

動詞に意志・推量を表す助動詞「う」「うず」が後続するもののうち、上一段・上二段活用動詞等「イ段＋う・うず」の融合形は拗音形で表記する。四段活用語で原本において融合部分がオ段開長音表記されているものは「ア段の仮名＋う・うず」、下二段活用語で合長音表記されているものは「エ段の仮名＋う・うず」という表記とする。ただし「tçucayôzuru」などハ行（ヤ行）下二段動詞と「う」「うず」が融合して「ヨー（ズ）」という発音形が想定されるものについては、「仕ようずる」のように「よう（ず）」という表記を採用した。

【例】 qitanocatauo imaychido **miôto** vomôua icani → 北の方を今一度見ようと思うは如何に

【例】 govn no **tçuqiôzuru** coto mo catai coto de ua nai. → 御運の尽きようずる事も難い事では無い。

【例】 fSanemori ua condo no icufa ni inochi iqite futatabi Miyaco ye mairō toua
zonjenu →実盛は今度の戦に命生きて再び都へ参らうとは存ぜぬ

【例】 vmano fanauo narabe,caqeôto xeraretareba, → 馬の鼻を並べ、駆けうとせら
れたれば

【例】 xibaxi fafayôzuru guide gozaredomo, → 暫し支ようずる儀で御座れども,

● 外国語の表記

本文中に外国語の普通名詞・地名・人名等が現れる場合は、漢字仮名交じりテキストにおいても原本のアルファベット表記のままとした。

【例】 Nipponno Feiqetoyū Hiftoriato,Morales Sentençasto,Europa no Efopono Fabulasuo vofu mono nari.

→ 日本の平家と言うH i s t o r i aと、M o r a l e s S e n t e n ç a sと、E u r o p aのE s o p oのF a b u l a sを押す物なり。

※ただし漢字仮名交じりテキストでは半角文字を使用できないため、アルファベットは全角となっている。また「l」のように全角での出力が不能な文字は、漢字仮名交じりテキストでは「s」に置き換えている（その場合も原文文字列では原文どおり「l」としているの
で適宜参照されたい）。

● 原本の誤りと見られる箇所の表記

固有名詞以外で明らかに原本の書き誤りであると判断できる箇所は、漢字仮名交じりテキストを正しいと思われる形に修正した。

固有名詞で原本ローマ字どおりの表記では指示対象が不明または間違った指示対象を指してしまう場合には、漢字仮名交じりテキストの該当箇所をカタカナ表記とした。いずれの場合も原文文字列の方は原本どおりのローマ字とし、修正は加えていない。

【例】 Qiyomori cô cofo quabun no coco uoba vôxerature, →清盛公こそ過分の事をば
仰せらるれ。

【例】 Xiguemori icani icani to aqireratureba, → シゲモリ如何に如何にと呆れらる
れば、 ※文脈上は「清盛」とあるべきところを原本が書き誤っている。

● コーパス化の対象外とした箇所

- ・『天草版平家物語』序の末尾にある略号の説明（「f Fito. Q Quan.…」）
- ・『天草版平家物語』巻末の「FEIQE NO CAQIAYAMARI」
- ・『天草版伊曾保物語』巻末の「Efofo no Fabulas no MOCVROCV」

2. 2 原文文字列

本コーパスでは漢字仮名交じりテキストと共に、原本のローマ字テキストを翻字した「原文文字列」を公開している。検索アプリケーション「中納言」上では原文 KWIC 機能を使って、キーだけでなく前後文脈の原文文字列情報を同時に参照することができる（後述）。以下ではこの原文文字列の概要を述べる。

原文文字列は原本のローマ字テキストに忠実に翻字することを第一の目標とした。したがって原本の書き誤りと見られる箇所についても修正を加えずに原本の表記のまま翻字している（2. 1 「原本の誤りと見られる箇所の表記」参照）。開合や四つ仮名の違例と見られる箇所も修正は加えず原本のとおりとした。

大英図書館で原本を閲覧し極力正確な判読に努めたが、文字がつぶれるなどして判読不能な箇所は、文脈上文字が推定できたとしても無理に翻字せず「■」で表した。

● 使用するアルファベット

ポルトガル式ローマ字に固有のアルファベットも原本のまま表示している。たとえば「s」と「ſ」の使い分けも原本どおりになっている。本コーパスで使用しているアルファベットの一覧は以下の通りである。

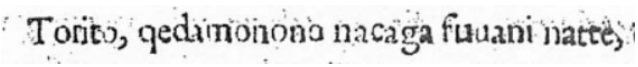
A	a	à	â	ã	B	b	C	c	ç
D	d	E	e	ẽ	ε	F	f	G	g
H	h	I	i	ĩ	j	L	l	M	m
N	n	O	o	ô	õ	ō	P	p	Q
q	R	r	S	s	ſ	T	t	u	ũ
ū	V	v	X	x	Y	y	Z	z	

また「,」「.」「;」「:」「!」「?」等の句読点相当の記号類についても原本で使用されているとおりに翻字した。

● 分かち書きの表示

『天草版平家物語』『天草版伊曾保物語』ともに原本ローマ字では分かち書きをしているが、本コーパスでは原本でアルファベットとアルファベットの間にスペースがあると見られる箇所に「□」を挿入することで分かち書きを再現している。

但し「,」や「.」などの記号の前後についてはスペースの有無の判断が困難である場合が多いため、明らかに記号の前後に空白がある場合でもスペースは表示しないこととした。

【例】 

→ Torito,qedamonono□nacaga□fuuani□natte,

● 『天草版平家物語』に見られる注記記号

『天草版平家物語』では人名や地名などの固有名詞の前に、「f」「t」「c」などの記号が付されていることがある。原文文字列ではこれも原本どおりに翻字し、該当する固有名詞の前に表示している。

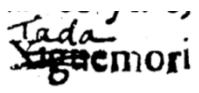
【例】  → qSanuqi□no□cami□fMafamori

※漢字仮名交じりテキストにはこれらの記号を反映させていない。【例】讃岐の守正盛

● 原本への書き入れ等の対応

『天草版平家物語』には巻末に「FEIQE NO CAQIAYAMARI」という正誤表のようなものがある。また原本に手書きで誤字修正の書き入れがされている箇所がある。このような誤字修正の指摘がある箇所については、以下のような記号を付すことで元の状態と修正後の状態を表示している。

※修正前の文字列を{ }で括り、修正後の文字列を<>で括る。

【例】  → {Xigue}<Tada>mori

● 原文文字列の翻字対象外としたもの

- ・単語途中で改行されるとき「-」
- ・本文の欄外に記された柱の語・つなぎ言葉等

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。以下では「中納言」上の主な表示項目とその内容について説明する。

44 件の検索結果が見つかりました。
検索対象語数: 136,123 記号・補助記号・空白を除いた検索対象語数: 123,332

サンプル ID	開始位置	連番	前文脈	キー	後文脈	語彙基読み	語彙基	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	外部リンク
40-天伊 1593_00014	40	40	鳥と、 Caraluto,	狐	ゆ[事]・# 或る[時]鳥[食]を[求]の[得]て、 木 ゆ[上]に[休]み[居]る[に]こ、 狐 noccoto # Arutoqic:caralucxocuuoc motomeyete,qinoc:vyeniyalunimic yruni,qitpunemocxocuuocmotomur	キツネ	狐	キツネ	名詞 -普通 名詞 -一般			qitpune	その他 -タイトル			キリシタ ン資料	天草版 伊曾保 物語	1593	鳥と、 狐の 事。			大英図 書館蔵 天草版 伊曾保 物語	450	
40-天伊 1593_00014	290	240	狐の[事]・# 或る[時]鳥[食]を[求]の[得]て、 木の[上]に[休]み[居]る[に]こ、 Caraluto,qitpunenoccoto, # Arutoqic: caralucxocuuocmotomeyete,qinoc: vyeniyalunimic:yruri,	狐	れ[食]を[求]むれども、 晴いで[馳]せ[帰]る とて、 鳥の[い]きんで[居]る[肉]を[見]て mocxocuuocmotomuredomo,yeides faxecayerutote,caralunoc:fucunde:yruc xiximurauocmle	キツネ	狐	キツネ	名詞 -普通 名詞 -一般			qitpune				キリシタ ン資料	天草版 伊曾保 物語	1593	鳥と、 狐の 事。			大英図 書館蔵 天草版 伊曾保 物語	450	

【中納言検索結果イメージ】

3. 1 形態論情報

「中納言」に表示される形態論情報は UniDic の見出しと対応するものである。基本的には BCCWJ や『日本語歴史コーパス』の他のサブコーパスと同様の情報であるが、以下に利用に際しての注意点を挙げる。

● 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を一括して取得することができる。語彙素の表記がわからない場合は語彙素読みで検索することも有用である。

● 語形

「語形」は、異語形を区別するレベルである。例えば語彙素「狼」に対して語形「オオカミ」と「オオカメ」があるというような非活用語の語形の違いや、活用語の場合には、語彙素「遅れる」の語形に「オクレル」（下一段 - ラ行）、「オクル」（文語下二段 - ラ行）があるといった口語活用と文語活用の区別などがある。語形はカタカナによって表記される。

● 品詞

UniDic の設計に基づいた品詞が付与される。学校文法における「形容動詞」は、語幹に当たる部分が「形状詞」、活用語尾は「助動詞」に分割されている点に注意が必要である。

● 活用型

文語活用として処理されているものには「文語上一段」のように「文語」が表示されるが、口語活用には「上一段」と表示され「口語」との表示はない。形容詞も同様に、文語活用のものは「文語形容詞」、口語のものは「形容詞」と表示される。

● 活用形

活用語の場合その活用形が表示される。活用形の小分類には、「一般」、「○音便」、「補助」（形容詞等の補助活用）のほかに「省略」（「行て」（yte）のように活用語尾が省略されたもの）や「融合」がある。本コーパスに見られる「融合」は、「でござる」などの「で」のみである。これは狂言での処理に倣い、断定の助動詞「なり」の連用形「に」に助詞「て」の付いた「にて」の変化したものと見て語彙素「なり」（助動詞・文語助動詞・ナリ・断定・連用形・融合）としている。

● 原文文字列

本コーパスでは2章で詳述した原文情報が表示される。また「原文 KWIC」機能を用いることで、キーに対応する原文文字だけでなく前後文脈の原文も合わせて参照できる。ただし「原文文字列」に表示される原文の内容は、各サブコーパスによって大きく性質を異にしているので、その点に注意が必要である（詳細は小木曽ほか（2017）を参照）。

また「中納言」の「文字列検索」で対象文字列を「原文」とすることで、原文ローマ字からの検索も可能となっている（下図）。

The screenshot shows the 'Text Search' (文字列検索) interface. The 'Target Text' (対象文字列) dropdown is set to 'Original Text' (原文). Below the search bar, there are several search filters and a list of search results. The search filters include 'Search Text' (検索文字列), 'Target Text' (対象文字列), 'Result Display Unit' (結果表示単位), and 'Key Display Form' (キー表示形式). The search results list includes examples of search terms and their corresponding original text.

検索文字列	対象文字列	結果表示単位	キー表示形式
任意の文字列	例: 国立 研究所	短単位	語
任意の1文字	例: 日 友好	長単位	文字列全体
文字クラス	例: 日本 人国語		
文字クラスの否定	例: 日本 人国語		

● 振り仮名

本コーパスはローマ字表記の資料であるという性質上「振り仮名」の情報は存在しない。

3. 2 本文情報

● 本文種別

本文種別情報を以下のように示している。なお、『天草版平家物語』は全編が喜一検校と右馬允の会話から成るが、本コーパスでは二人の発話については本文種別を会話としていない。二人の発話の中で引用される登場人物たちの発話に対してのみ「会話」という本文種別を付与している。それぞれ和歌や漢詩等の韻文であることが明らかである場合には「韻文」を付与する。

会話：本文中の発話箇所。

引用：発話以外の引用箇所。「典拠」は、何らかの文献等からの引用であることを表す。例えば「引用 - 典拠・和歌 - 韻文」は発話ではなく何らかの典拠を持つ和歌を引用している箇所であることを示す。そのほか手紙、院宣等の命令についても「引用 - 手紙」「引用 - 院宣・宣旨等」などの情報を付与している。

その他 - タイトル：各章・各話のタイトル部分。

● 話者（『天草版平家物語』のみ）

前後文脈等からその発話部分の話者が明らかな場合に話者情報を付与した。文脈等から誰が発話したか特定できない場合は、本文種別が「会話」となっているにもかかわらず話者が空欄となっていることがある。

話者名は基本的に当該人物の発話箇所直前での呼称を用いて入力しているため、同一人

なお『天草版伊曾保物語』には話者情報の付与は行っていない（本文種別のみ）。

● 卷名等

『天草版伊曾保物語』は原本巻末の「Efopo no Fabulas no MOCVROCV」に記載されている話名を UniDic の語形代表表記によって漢字仮名交じり表記にしたものを表示している。そのため各話本文冒頭に出てくる話名とは異なる場合がある。

● 底本・ページ番号

3. 5 原文 KWIC 表示

原文 KWIC は下図の以下の箇所に☑を入れることで表示／非表示を選択することができる。

① 設定を戻す

☐ コーパス情報
☐ サブコーパス名 ☒ サンプル ID ☒ 禁台位置 ☒ 通番 ☒ コア

形態論情報

☒ 前文脈 ☒ キー ☐ 後文脈 ☒ 原文 KWIC ☐ 読要素 ID ☐ 読要素読み ☐ 読要素 ☐ 読要素読分類 ☐ 語形 ☐ 語形代表表記 ☐ 品詞 ☐ 活用型 ☐ 活用法 ☐ 書字形 ☐ 原形名出現形 ☐ 発音名出現形 ☐ 語種 ☒ 原文文字列 ☐ 振リ仮名

☐ 本文情報 ☐ 本文種類 ☐ 話者 ☐ 文体 ☐ 歌番号	☐ 作品情報 ☐ ジャンル ☐ 作品名 ☐ 成立年 ☐ 巻名等 ☐ 部	☐ 作者情報 ☐ 作者 ☐ 生年 ☐ 性別	☐ 紙本情報 ☐ 紙本 ☐ ページ番号 ☐ 校注者 ☐ 出版社	☐ その他 ☐ 外部リンク ☐ permalink
---	--	--	---	---

113 件の検索結果が見つかりました。
 検索対象語数: 138, 123 記号+補助記号・空白を除いた検索対象語数: 123, 332

参考文献

- 池上尚（2016）『日本語歴史コーパス 平安時代編』形態論情報規定集」
http://pj.ninjal.ac.jp/corpus_center/chj/doc/morph-heian-2016.pdf (2018年3月22日閲覧)
- 市村太郎・渡辺由貴（2015・2016）『日本語歴史コーパス 室町時代編Ⅰ狂言』形態論情報
報の概要」 http://pj.ninjal.ac.jp/corpus_center/chj/morph-kyogen-2016.pdf (2018年3月
22日閲覧)
- 小木曾智信・岡照晃・中村壮範・八木豊（2017）『日本語歴史コーパス』における原文 KWIC
表示機能の実装」言語資源活用ワークショップ 2017 発表論文集 pp.252－257
- 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』形態論情報規定集第4版（下）特定領域研究「日本語コーパス」平
成22年度研究成果報告書
- 小椋秀樹・須永哲矢（2012）「中古和文 UniDic 短単位規程集」基盤研究(C)「和文系資料
を対象とした形態素解析辞書の開発」研究成果報告書 2

●漢字仮名交じりテキスト作成・原文ローマ字データの電子化にあたっては以下の資料を
参考にした。

- 江口正弘（1986）『天草版平家物語対照本文及び総索引 本文篇』明治書院
- 江口正弘・溝口博幸編（2005）『天草本平家物語資料大成』尚文出版
- 江口正弘注釈（2009）『天草版平家物語全注釈』新典社
- 亀井高孝・阪田雪子翻字（1966,1980）『平家物語：ハビヤン抄 キリシタン版』吉川
弘文館
- 近藤政美・池村奈代美・浜千代いづみ編（1999）『天草版平家物語語彙用例総索引（1）』
勉誠出版
- 江口正弘編（2011）『天草版伊曾保物語影印及び全注釈 言葉の和らげ影印及び翻刻翻
訳』新典社
- 大塚光信・来田隆編（1999）『エソポのハブラス本文と総索引』清文堂出版

また大英図書館において原本を閲覧し、判読が困難な箇所を確認を行った。貴重な資料の
閲覧をご許可くださいました大英図書館東アジアコレクション日本部に、記して深く感謝
申し上げます。