

『日本語歴史コーパス 室町時代編 I 狂言』テキスト・表示項目について

2015年3月30日 市村太郎

2016年3月29日 渡辺由貴 更新

1. はじめに

『日本語歴史コーパス 室町時代編 I 狂言』は、大塚光信編（2006）『大蔵虎明能狂言集 翻刻註解』を底本としている。この底本をコーパス化するにあたっては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（詳しくは小林・市村 2013、市村 2014 参照）。その様々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

[1] 外字の処理

本文テキストの文字入力は JISX0213 に準拠している。

- 外字となる 11 か所は、下記のように読みが同じで字形・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。

①「まだ利厩をぬかすか、(上 p.93 018_脇_目近籠骨) 左「口」+右「堀」

②やわかうしとは候 (上 p.126 024_脇_牛馬) 「候」に濁点

③平家は摂津国一谷に城郭をかまへ、(上 p.258 047_大名_青海苔) 左「土」+右「郭」

④「きやう人はしれば (上 p.358 067_髯・山伏_鶏髯) 「人」に濁点

⑤ふきやう人もはしる共云 (上 p.358 067_髯・山伏_鶏髯) 「人」に濁点

⑥三千人の客の中に、馮三と云者有 (上 p.572 119_鬼・小名_けいりう) 左「足」+右「藿」

unicode: U+4830

⑦其より馮三をば、鶏鳴とも申ける (上 p.572 119_鬼・小名_けいりう) 左「足」+右「藿」

unicode: U+4830

⑧今など爰元へのぼらふ人ではなひが (下 p.302 200_集_じしやく)

⑨是より鬪刀をも云也、此類多 (下 p.416 227_集_すずきばうちやう) 上「亦」+下「肉」

unicode:U+8114

⑩蝙蝠也底、貉也、壁、狻也 (下 p.459 235_萬集_織女) 左「豸」+右「兪」 unicode:U+3E84

⑪せつしやうすれば、地ごくに入事矢のごとし (下 p.492 273_萬集_ゑさしほうもん外字) 「地」に濁点

- 底本では、台詞の先頭や謡の箇所^ニに上下回^レ転させたものも含めて三種の庵^ニ点^ニが用^レいられ、原著者によるものと節^ニ付^レけ^レ箇所^ニを表^スすために校^ニ注^レ者^ニが付^レした物^ニで使^レい分け^レられている。しかし、JIS JISX0213 の範^ニ圍^ニで用^レいること^ニがで^レきるのは「^ニゝ^ニ」のみである。そのため会話・ト^ニ書^レきの頭^ニに使用^レされる物^ニは「^ニ」に置^レき換^レえ、節^ニ付^レけ^レ箇所^ニに使用^レされるものは「^ニ」に置^レき換^レえて入^レ力^レした。

【例】『毗沙門の福ありのみと聞からに、くらまぎれにて、むかでくひけり』(連歌毗沙門 底本上 p.10)

[2] テキストの校訂

- 本文中 3 か所、処理単位上の問題により振り仮名（または傍記）の文字列をタグ付で本文と置^レき換^レえた。
 - ① 〈原文〉^{デカイタ}手^ニ栖^レ → 〈入力〉^ニで^レか^レいた^ニ (鴈盗人 底本上 p.181 頭書)
 - ② 〈原文〉^{ニツノトワラホウ}莞^ニ爾^ニ笑^レ → 〈入力〉^ニにつ^レこと^ニわ^レら^レほう^ニ (じせんせき 底本上 p.246)
 - ③ 〈原文〉始^ニ春^ニ乃^ニ波^ニ都^ニ禰^ニ乃^ニ家^ニ布^ニ能^ニ多^ニ麻^ニ婆^ニ波^ニ伎^ニ手^ニ爾^ニ等^ニ流^ニ可^ニ良^ニ爾^ニ由^ニ良^ニ久^ニ多^ニ麻^ニ能^ニ乎^ニ → 〈入力〉始^ニ春^ニの^ニは^ニつ^ニね^ニの^ニけ^ニふ^ニの^ニたま^ニば^ニは^ニき^ニて^ニにと^レる^ニから^ニに^ニゆ^ニらく^ニたま^ニの^ニを^ニ (枕物狂 下 p.78)
- 濁音が期待される箇所^ニに濁^レ点^ニが付^レされてい^レない場合^ニは、諸^ニ資^レ料^ニを参^レ考^ニに検^レ討^ニの上^ニ、必要^ニな箇所^ニは濁^レ点^ニを補^レった。た^レだ^レし、清^ニ濁^ニ両^ニ形^ニあり^ニ判^レ断^ニに迷^レう場合^ニには極^ニ力^ニ濁^レ点^ニを付^レ与^レし^レない方針^ニをと^レった。濁^レ点^ニを補^レう前^ニの文字^ニ列^ニは「^ニ原文^ニ文字^ニ列^ニ」に表^レ示^レされる。

【例】〈原文〉ゑびすびしやものことくとまる → 〈入力〉ゑびすびしやものことくとまる (ゑびす大黒 底本上 p.6)
- 仮名 1 字分の踊^レり^レ字^ニは、想^レ定^レされる仮^ニ名^ニに置^レき換^レえた。変^レ換^レ前^ニの文字^ニ列^ニは「^ニ原文^ニ文字^ニ列^ニ」に表^レ示^レされる。な^レお、漢^ニ字^ニの繰^レり^レ返^レし^レや 2 字^ニ分^ニ以上^ニに相^レ当^レする^レく^ニの字^ニ点^ニ等^ニは置^レ換^レし^レない。

【例】〈原文〉よろこびをかさねつ^ニ → 〈入力〉よろこびをかさねつ^ニ (かくすい 底本上 p.33)
- 底^ニ本^ニ本^ニ文^ニ中^ニカ^ニタ^ニカ^ニナ^ニで表^レ記^レされた箇所^ニは、平^ニ仮^ニ名^ニで表^レ示^レした。変^レ換^レ前^ニの文字^ニ列^ニは「^ニ原文^ニ文字^ニ列^ニ」に表^レ示^レされる。

【例】〈原文〉ふくありの^ニ → 〈入力〉ふくありの^ニ (連歌毗沙門 底本上 p.9)
- 見^レせ^レけ^レち^ニなど、原^ニ本^ニ筆^ニ者^ニによる文字^ニの挿^レ入^レ・削^レ除^レ・置^レき換^レえ^ニの指^レ示^ニについては、原則^ニこれ^ニを反^レ映^レさせ^レたもの^ニを本^ニ文^ニテ^ニキ^ニス^ニト・「^ニ原文^ニ文字^ニ列^ニ」^ニとした。

【例】〈原文〉富士をづき^ニの → 〈入力〉富士をづき^ニの (餅酒 底本上 p.27)

- 校注者による本文の誤りの指摘については、頭注等により修正すべき内容が判明する物のみ、本文テキストに反映させた。また、修正される前の文字列は「原文文字列」に表示される。

【例】〈原文〉国をへたでゝあるに → 〈入力〉国をへだてであるに（昆布柿 底本上 p.35）

- 返り点などが付されており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。返り点等のない漢文箇所は「未知語」として扱い、品詞「漢文」とした（「形態論情報の概要」を参照）。

【例】〈原文〉天竺国ニ有_レリ此草葉 → 〈入力〉天竺国に此草葉有り（人馬 底本上 p.240）

- 下記の補読が必要な漢文の箇所は、補読した。

① 〈原文〉是於 → 〈入力〉是に於（牛馬 底本上 p.127）

② 〈原文〉人食_レ之則化シテ馬ト成ル → 〈入力〉人之を食ば化して馬と成る（人馬底本上 p.240）

- 下記の置き字相当とみられる箇所は、置き字相当の文字を表示させていない。

① 〈原文〉^{（チカシ）}幾_レ於道ニ → 〈入力〉道に幾（おひやし 底本上 p.549）

② 〈原文〉自ニヨリ太元 → 〈入力〉太元より（かうやくねり 底本下 p.374）

- 捨て仮名は本文から除いた。

① 〈原文〉刑イ鞭 → 〈入力〉刑鞭（なべやつばち 底本上 p.134）

② 〈原文〉力エ強シ → 〈入力〉力強し（楽阿弥 底本下 p.247）

- 頭注・傍記等の書き入れは、検討の上、本文と同時並行的、または挿入後単位に不均衡を生じてしまう場合は本文とせず、その他の場合は極力台詞・ト書き等に挿入し、本文文字列とした。

3. 中納言における表示項目と内容

『日本語歴史コーパス 室町時代編 I 狂言』の本文には様々なタグ（本稿末参考表）や単語情報（後述）が付されており、その情報は、WEB上のコーパス検索ツール「中納言」上に、検索として表示される（図1・図2）。

以下では、中納言上の主な表示項目とその内容に関して概説する。

図 1 検索画面



図 2 検索結果の表示イメージ

サンプルID	前文脈	キ	後文脈	語彙素読み	語彙素読み	品詞	活用型	活用形	本文種別	話者	作品名	成立年	作者	生年	底本	ページ番号
虎明本狂言 _063_ 大名_ぶあぐ	て、色 / \ あんじて相に、神別して満足したよ「此やうなざるお太刀では、何かしそごなむ	まらせ	うそ、かやうなる物置れはござるまひ「さでであらふ、ゆべれもなむのきされて、代へてしました次刀じやまだこ「私	マイラセル	参らせ	動詞・非自立可能	次語サ行変格	未然形一般	会話	次郎冠者	虎明本狂言	1642	大筋弥次郎虎明	1597	大筋虎明能狂言集 翻訳註解	331
虎明本狂言 _063_ 大名_ぶあぐ	なんだ「たぶんぶあぐじやと思ふが、わごしはせいばいもせいで、成順したと云物じやあらふ理に、ただあき	まらず	まひ「思まいかな事、それはわかれぬ事でござる物が、しつわが申されうか、誰ぞ成順はば	マイラセル	参らせ	動詞・非自立可能	次語サ行変格	終止形一般	会話	主	虎明本狂言	1642	大筋弥次郎虎明	1597	大筋虎明能狂言集 翻訳註解	332
虎明本狂言 _063_ 大名_ぶあぐ	ほどに、はやらこしらへて出さしめ「心得た「あれへ参見て御ざるが、御せぬのごとく、あのやぶはなより見	まらず	ば、かあるのとぶまでも一目に見えまするが、人かかずもござらぬ「誰ぞ放がいごとくじやが、なにも見え	マイラセル	参らせ	動詞・非自立可能	次語サ行変格	已然形一般	引用・会話指示	次郎冠者	虎明本狂言	1642	大筋弥次郎虎明	1597	大筋虎明能狂言集 翻訳註解	334
虎明本狂言 _063_ 大名_ぶあぐ	「あれへ参見て御ざるが、御せぬのごとく、あのやぶはなより見まらずいば、かあるのとぶまでも一目に見え	まらず	が、人かかずもござらぬ「誰ぞ放がいごとくじやが、なにも見えぬ「わたくしのぞんずるは、せいばい「悪いことだ	マイラセル	参らせ	動詞・非自立可能	次語サ行変格	連体形一般	引用・会話指示	次郎冠者	虎明本狂言	1642	大筋弥次郎虎明	1597	大筋虎明能狂言集 翻訳註解	334
虎明本狂言 _063_ 大名_ぶあぐ	ぶあぐでござる「成順したに、何としてそこまはまざるぞ「其事でござる、おしりの御意にぞもぎ	まらず	て、御成順にあつてござる所で、御業へも参りませ、又「其家へも参らば、次道にまよふて	マイラセル	参らせ	動詞・非自立可能	次語サ行変格	連用形一般	会話	武悪	虎明本狂言	1642	大筋弥次郎虎明	1597	大筋虎明能狂言集 翻訳註解	334
虎明本狂言 _064_ 大名_すわらぶあぐ	が有程に、此圖は参る事ならぬ、参宮めさるるならば、やがて下向めされしこいへ「御はこななま	参らせ	られまといへと仰らるるか「あふみどもまいる事がならん「かしたつた「其由申さじ、いや汝はさだめてども	マイラセル	参らせ	動詞・非自立可能	次語下二様・サ行	未然形一般	会話	次郎冠者	虎明本狂言	1642	大筋弥次郎虎明	1597	大筋虎明能狂言集 翻訳註解	338

[1] 形態論情報

「原文文字列」を除いては基本的に BCCWJ や中古和文と同様であり、小椋他(2011)などを参照されたい。中納言において表示される形態論情報（短単位）は、Unidic の見出しに対応している。以下には利用に際して注意すべき点を幾つか挙げる。

● 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一括して取得することができる(図 2 は語彙素読み「マイラセル」による検索結果である)。

● 語形

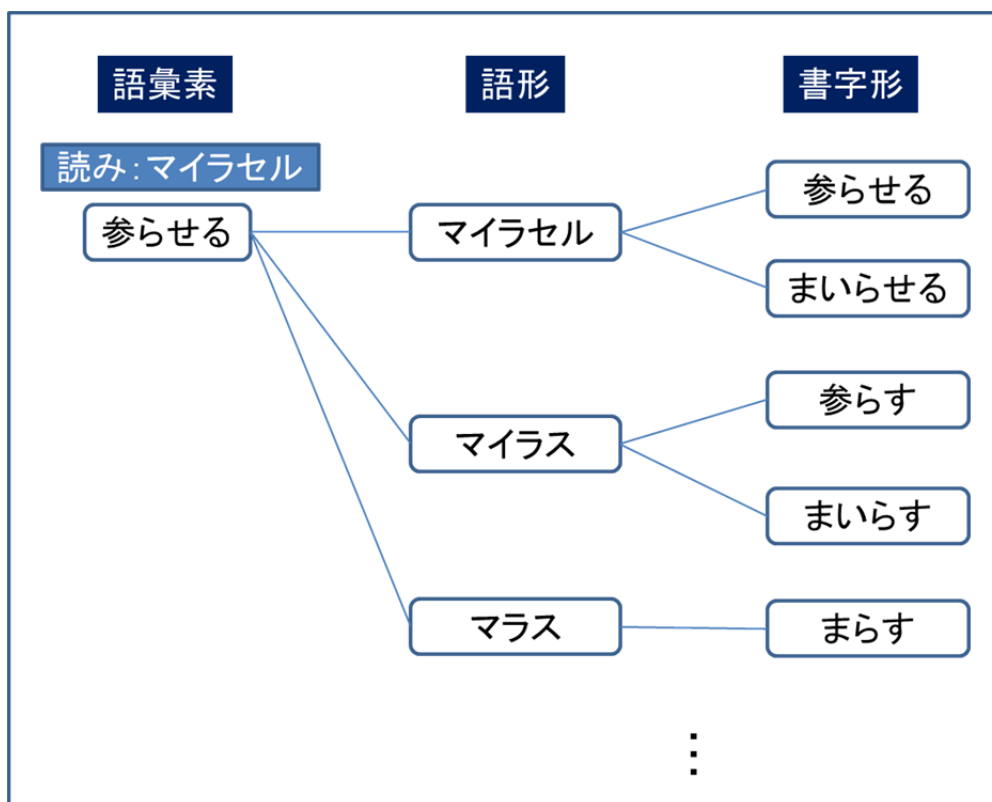
「語形」は、異語形を区別するレベルであり、例えば短縮形の「マラス」は語彙素「参

らせる」の語形として認定される。ただし、2015年3月時点では四段活用と上下二段活用は別語彙素として認定しているため、注意が必要である。

●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 語彙素「参らせる」の語彙素・語形・書字形



●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。

●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。

●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。例えば打消の助動詞「ず」に助詞「は」が付

いた「ずは」の変化した「ざ」は、下記のように処理されている。

【例】さびたらばとがざなるまひ（連歌毗沙門 底本上 p.10）助動詞・文語助動詞-ズ・連用形-融合

また、「でござる」などの「で」は、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、語彙素「なり」（助動詞・文語助動詞-ナリ-断定・連用形-融合）とした。

●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」に表示される（表1）。

また、校注者による修正指示箇所については修正前の本文、筆者によるミセケチや挿入指示等については、それを反映した後の本文を表示している。なお、漢文の箇所は訓読後の形が表示される。

表 1 本文テキスト（出現書字形）と原文文字列

ゑびす大黒		かくすい	
出現書字形	原文文字列	出現書字形	原文文字列
「	「	「	「
あど	アド	さら	さら
しめ	シメ	ば	ハ
を	ヲ	お	お
ひか	ヒカ	いとま	いとま
ば	バ	を	を
やど	ヤド	くださるる	くださるゝ
へ	ヘ	ぞ	ぞ
かへら	カヘラ	わか	わか
ふ	フ	『	『
と	ト	よろこび	よろこび
云	云	に	に
て	テ	、	、

あくまで特定の段階を反映した補助的な表示であるため、踊り字・濁点・片仮名以外の校訂(例えば本文挿入の状況など)、あるいは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。

[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）。

●本文種別

本文種別の対応は下記のとおりである。なお、それぞれ節付け箇所・和歌や謡であることが明らかな箇所には、「ト書き・韻文」のように、「**韻文**」が表示される。

- 会話 : 本文中の台詞箇所。
- ト書き : 割書き、および会話ではない本文。
- 注釈 : 台本の前後に記された注記の段落。内容はト書きと重なることがある。
- 引用 : 台詞以外の引用箇所。
 「**典拠**」は、文献等の引用を表す。
 「**会話指示**」は、ト書き・注釈内の台詞の記述を表す。
 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。

タイトル : 各曲の表題回り。

話者 : 著者による話者表示。校注者による「()」付の話者は本文としていない。

会話注記 : 会話文頭に示される「舞」や「イロ」などの表示。なお、文頭以外の物は、文や単語に対する並行的なテキストとなるため、原則本文としていない。

●話者

対象となる語が含まれる「会話」または「会話指示」の話者を表示した。著者や校注者によって話者が付されている場合はそれを表示し、「会話指示」などで話者が記されていない場合は、判明する限り、タグ付け者がその会話文の話者を、周囲の文内での表示やその作品で用いられている校注者の表示などにより記した。

ただしこれらの話者表示は、ある箇所では「しう」、別の箇所では「主」となっているなど、現時点では同一登場人物（同一の役割の人物）の表記を統一できておらず、そのまま（例えば「しう」と「主」が別物のまま）統計的な材料等に使用するのとは適当ではない。

表 2 出現書字形・本文種別・話者の表示

ゑびす大黒			かくすい		
出現書字形	本文種別	話者	出現書字形	本文種別	話者
「	引用-会話指示		「	会話	奏者
や	引用-会話指示		さら	会話	奏者
、	引用-会話指示		ば	会話	奏者
ゑひや	引用-会話指示		お	会話	奏者
、	引用-会話指示		いとま	会話	奏者

とと	引用-会話指示		を	会話	奏者
、	引用-会話指示		くださる	会話	奏者
やあ	引用-会話指示		ぞ	会話	奏者
、	引用-会話指示		わか	会話注記	百姓
ゑびす	ト書き		『	会話-韻文	百姓
びしやもん	ト書き		よろこび	会話-韻文	百姓
の	ト書き		に	会話-韻文	百姓
ごとく	ト書き		、	会話-韻文	百姓
とまる	ト書き		また	会話-韻文	百姓
「	注釈		よろこび	会話-韻文	百姓
あど	注釈		を	会話-韻文	百姓
しめ	引用-会話指示	アド	かさね	会話-韻文	百姓
を	引用-会話指示	アド	つつ	会話-韻文	百姓
ひか	引用-会話指示	アド	』	会話-韻文	百姓
ば	引用-会話指示	アド	「	ト書き	
やど	引用-会話指示	アド	舞	ト書き	
へ	引用-会話指示	アド	あり	ト書き	
かへら	引用-会話指示	アド	『	会話-韻文	百姓
ふ	引用-会話指示	アド	やら	会話-韻文	百姓
と	注釈		/ \	会話-韻文	百姓
云	注釈		めでた	会話-韻文	百姓
て	注釈		や	会話-韻文	百姓

[3] 作品情報

●ジャンル

「ジャンル」には、「狂言/脇狂言之類」のように、「狂言/その曲の類名」を表示した。

●巻名等

「巻名等」には各曲名を表示した。またその表示方法は、原則底本である大塚(2006)の巻末「曲名索引」に従い、「ほうちょうむこ (庖丁聲)」のように、「一般的な読み (一般的な漢字表記)」とした。

[4] 底本情報

●底本・ページ番号

「底本」項目には、「翻刻註解<上>」のように底本と巻名を表示した。また、底本のページ番号を「ページ番号」項目に表示した。

参考 本コーパスのタグセット

要素(タグ)名	説明
<text>	1作品(曲目)全体
<front>	前付相当の箇所
<body>	主本文相当の箇所
<article>	1記事の範囲
<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。狂言ではタイトルや注釈を除く主本文
<block>	記事中のタイトルや注釈など、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
<stage>	ト書き
<quotation>	他の文献からの引用、ト書き・注釈内の台詞指示
<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>	話者表示
	短単位以上の長さをもった単位、狂言ではカタカナ表記箇所に使用
<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(注記・校訂情報除く)
<ruby>	本行の右側に振られた振り仮名等の文字列(注記・校訂情報除く)
<add>	著者による本文の追加箇所
<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、狂言では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や傍注記等、丁付情報
<pb><lb>	底本の改ページ位置・改行位置

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・渡辺由貴・鴻野知暁・河瀬彰宏・小林正行・山田里奈・堀川千晶・村山実和子・
小木曾智信・田中牧郎 (2015) 印刷中「『虎明本狂言集』コーパスの公開」『日本語学会
2015 年度春季大会予稿集』
- 大塚光信編 (2006)『大蔵虎明能狂言集 翻刻註解』上下巻 清文堂出版
- 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 国立国語研究所
- 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 国立国語研究所
- 小林正行・市村太郎(2013)「『虎明本狂言集』コーパスの構造化—仕様と事例の検討—」第
3 回コーパス日本語学ワークショップ予稿集 pp.323-332