

洒落本を対象とした東西対照コーパスの設計と構築

北崎勇帆^{†1}

洒落本は、近世期に刊行された小説の一形態であり、近世日本語の口語資料としての有用性が高い。この洒落本には、書名や話の粗筋を同一にしなが、江戸板・上方板で内容や語彙に異同のある作品が存在する。上方で刊行されたものが後に江戸で改作された『月花余情』組と、江戸で刊行されたものが後に上方で刊行された『郭中奇譚』組である。本稿ではそのような江戸・上方間で改作が行われた洒落本のテキストを TEI P5 に準拠してマークアップすることにより、当時の東西言語の比較資料として用いることができる対照コーパスを構築した。

Design and Construction of East-West Parallel Corpus for Sharebon

YUHO KITAZAKI^{†1}

Sharebon is a genre of literature from the early modern period of Japan. Some Sharebon were published in Edo/Osaka at first and after that republished in Osaka/Edo. In this report, I make an East-West parallel corpus by marking-up interpolations in texts and creating an XML version of "Gekka Yojo" and "Kakuchu Kitan" based on TEI P5 Guidelines.

1. はじめに

洒落本は、江戸中期から末期にかけて、初めは上方に発生し、後には江戸を中心として刊行された、遊里小説の一形態である。岩波書店『日本古典文学大系』や小学館『日本古典文学全集』に数作品が取り上げられる他、叢書としては林平書店『洒落本大系』(全12巻)、中央公論社『洒落本大成』(全30巻)が発行されている。洒落本はその口語性の高さから、日本語史研究においては、主に近世後期日本語の資料としての有用性を持つ。現在、国立国語研究所によって洒落本コーパスの設計が進められており[1]、その活用が期待される。

さて、この洒落本には、書名を同一にしなが、江戸板・上方板で語彙や内容を異にするものが存在する。当初は上方で刊行されたものが後に他者の手によって江戸で刊行された『月花余情』、これとは逆に、江戸で刊行されたものが改作され、上方で発刊された『郭中奇譚』の二組の作品群である。いずれの改作にも語彙の意図的な改変が認められるため、これらの作品を比較対照することによって当時の上方・江戸語の様相を窺い知ることができることが、矢野準[2]、増井典夫[3]らにより論じられている。本稿では『月花余情』『郭中奇譚』を対照コーパス化し、改作部分に対応付けることによって対照比較のための基礎資料の構築を試みると共に、日本語史研究における対照コーパスの利用可能性を検討する。

2. 対象とする資料

以下、中野三敏氏の解題[4,5]に依り、『月花余情』『郭中奇譚』の二組の刊行・改作の経緯を述べる。

2.1 『月花余情』組

『月花余情』(『洒落本大成』第3巻所収)は延享3年(1746)に上方で刊行されたものとされており、その後編として、「陽台遺編」と「妣閣秘言」の二作品を合刻した『陽台遺編・妣閣秘言』(『洒落本大成』第3巻所収)が寛延3年(1750)以前に発刊されている。これら「月花余情」「陽台遺編・妣閣秘言」の三作品の題をそれぞれ「燕喜篇」「秘戯篇」「自楽篇」と変え、内容を改めて明和年間(1764-1771)に江戸で発刊されたものが異本『月花余情』(『洒落本大成』第3巻所収)として知られる。本稿では便宜的に、上方で発刊された『月花余情』、『陽台遺編・妣閣秘言』を併せて『上方板月花余情』、江戸で発刊された『月花余情』を『江戸板月花余情』と呼ぶ。以上の改作の経緯をまとめると、表1のようになる。

改作前(上方)		改作後(江戸)	
書名	部	書名	部
月花余情	燕喜篇	月花余情 (異本)	燕喜篇
陽台遺篇・ 妣閣秘言	陽台遺篇		秘戯篇
	妣閣秘言		自楽篇

表1 『月花余情』組

また、『上方板月花余情』『江戸板月花余情』中の「燕喜篇」の冒頭部を抜き出し、改作前と改作後が対応している箇所それぞれアルファベットと下線を付して以下に示す。

【上方板】

[客花情] a来ル。bどふじや。cいかふさむいの。

[中居とよ] d花情さん。よふおいなはつた。eサア。おあがりなはれ。fこれお久米どん。g御茶あげませんせ。サアマ。御上りなはれ。

^{†1} 東京大学大学院人文社会系研究科
Graduate School of Humanities and Sociology, University of Tokyo
yuhokitazaki@gmail.com

[廻シ佐助来ル] わたしが所の。ちよと。お尋なさつて。下さりませ。

[とよ] はやいぞや。

[客]_hいかふにぎやかなの。_i神棚のもてなしがよひとみへた。

[とよ]_jナアニおつしやるやら。_k是おなつどん。_i花情さんのおいなはつたぞや。

【江戸板月花余情】

[客花丈]_a来ル _bとうしゃ _cいかふ寒いの

[中居とよ]_d花丈さんよふ御出なさつた _eサア御上りなされ

_fコレおくめどん _g御ちや上さんせ

[花丈]_hいかふにぎやかな _i神棚のもてなしがよひと見へた

[とよ]_jナアニおつしやるやら _kコレおなつどん _i花丈様の御出なさつたぞへ

『陽台遺篇・舩閣秘言』中の「陽台遺篇」と、「江戸板月花余情」中の「秘戯篇」についても、同様に例を示す。

【上方板】

[客]_aウゝゝゝアゝ。 _bついねたそふな。 _cもうなん時じや

[女郎]_dしらんわいな

[客]_eしらんほどなら。 _fもふよいいにちぶんじや

[女郎]_f客のかいなへ喰つく。

[客]_gあいたゝ。 _hこりやめつたにかぶるまい。 _iほうれんそうの汁がついて有ぞ

[女郎]_jだんないわいな。 _kいつそしんだがましじやわいな

【江戸板】

[花丈]_aウゝアゝ。 _bついねたそうな。 _c何時しや

[かしく]_dしらんわいな

[花丈]_eしらんほどならもふよいいに時分じやが

[かしく]_fきやくのかいなへくい付

[花丈]_gあいた _hこりやめつたにかぶるまい _iほうれん草の汁が付て有ぞ

[かしく]_jだんないわいな _kいつそしんだがましじや

2.2 『郭中奇譚』組

『郭中奇譚』(『洒落本大成』第4巻所収)は明和六年(1770)に江戸で発刊された。「船窓笑語」「弄花唇言」「掃臭夜話」の三部に分かれるこの作品から、「船窓笑語」を全て削り、「掃臭夜話」を「掃臭夜帖」と改題して、内容を改めて明和末頃に上方で発刊されたものが『郭中奇譚(異本)』(『洒落本大成』第4巻所収)として知られる。こちらが『江戸板郭中奇譚』『上方板郭中奇譚』と呼ぶ。以上の改作の経緯をまとめると、表2のようになる。

改作前(江戸)		改作後(上方)	
書名	部	書名	部
郭中奇譚	船窓笑語	郭中奇譚 (異本)	×
	弄花唇言		弄花唇言
	掃臭夜話		掃臭夜帖

表2 『郭中奇譚』組

前項同様、『江戸板郭中奇譚』『上方板郭中奇譚』より「弄花唇言」の一部分を抜き出し、対応部に下線を付して以下に示す。『郭中奇譚』組の改作は『月花余情』組に比してその改作が改作元に忠実でない。

【江戸板】

[かもと小菊]_a次の間よりはしり出る

[太夫]_bなんじやいの。 _cさはがしい

[小菊]それでもあの勝弥がわしをこかしてこそぐりやるもの

[太夫]_dこれ勝弥 _eせびらかしやんな

[勝弥]_fいゝへ _gあの子がくらかりけうしろからひつくりさしやるによつて

[小菊]_hゑゝ _iうそ

[亭]_jその様にあがくと内へいふてやるぞ

[客]_j太夫に成るものがそのやうにあがくものか

【上方板】

[かむろ若ば]_aろうかよりはしりいる

[女郎]_bナンタ。 _cさわがしい

[若ば]柴木どのこそぐらツしやるナ

[女郎]_dコレ柴木ヤ。 _eたしなみや

[しばぎ]_fイエ _g若葉どのがくらかりけうおどさツしやるから

[若ば]_hエゝ _iうそいはツしやれ

[茶]_jソノやうにさわくとおさつどのにつげるぞ

[客]_k三歩になるものがそのやうにでうだんするものか

同様に、「掃臭夜話」「掃臭夜帖」の冒頭部を抜き出し、以下に示す。

【江戸板】

[夜たか]申へ

[客]_aでばの熊 _b手ぬぐいほうかむりひより下駄させるこしにさして _cなアまアいだアんぶ _d引へ

[よ]_e熊さんか

[客]_fアゝ _gはやく出たナアゝ _hこの犬めはいまへしいと下太でくはん

[よ]_eゆふべはどふしなさつた _f外にゑいのができたか

【上方板】

[客]_aてんぼの太兵衛 _b手拭ほうかふりして草履下駄にて来り

[惣嫁] 桐の木の間。太兵衛さんじゃないかいな
[客] おれじや。はやう出やつたの
[惣か]。夕部は見へなんたなあ。北の方や南の方みてあたけ
れと。外によいのかてきたかへ

p	段落を示す
said	発話を示す
q	割書等を示す

表 3 使用する XML タグの一覧

3. 文書構造と構築の方針

2節で示した『月花余情』『郭中奇譚』の二組のテキストを対象として、TEI(Text Encoding Initiative) P5[6]に準拠したコーパスの構築を試みる。今回は口語の対照と比較を目的とするため、物語本文のみを対象とし、序文などの前付け部分、跋文・出版情報などの後付け部分は除外する。マークアップの際には同様の TEI P5 準拠の要素を用いた先行事例[7]を参照するが、本稿は東西の比較を目的とし、XML による文書の記述そのものを目指すものではないため、ルビを示す<w>や外字を示す<g>といった文末満の情報、改ページ位置を示す<pb/>、改行位置を示す<lb/>といった位置情報は付与しない。

洒落本テキストの構造は、図1のような階層構造を取る。

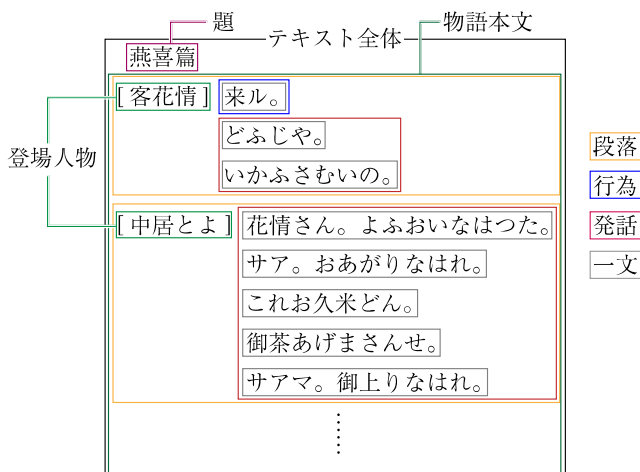


図 1 洒落本テキストの構造

こうした構造に TEI P5 準拠のタグを対応させる場合、例えば一つの段落は<p>に相当し、一つの発話は<said>に相当し、一つの文は<s>に相当する[a]。物語本文のマークアップに用いる TEI P5 準拠の XML タグの一覧を表 3 に示す。

タグ名	タグの説明
text	全体を示す
body	物語本文を示す
head	題を示す
name	固有名詞（話者）を示す

a 尚、河瀬他(2013)[7]では割書き内のテキストに<v>を割り当てているが、<v>タグは"verse line"に用いることが想定されており、韻文ではない文には適さないものと思われるため、本稿では会話中の一文をマークアップする際と同様に<s>を用いた。

この構造に従い「上方板月花余情」のテキストにタグを付与する（図2）。

```
<p>
  <name> 客花情 </name>
  <q>
    <s> 来ル。 </s>
  </q>
  <said>
    <s> どふじや。 </s>
    <s> いかふさむいの。 </s>
  </said>
</p>
<p>
  <name> 中居とよ </name>
  <said>
    <s> 花情さん。よふおいなはつた。 </s>
    <s> サア。おあがりなはれ。 </s>
    <s> これお久米どん。 </s>
    <s> 御茶あげまさんせ。 </s>
    <s> サアマ。御上りなはれ。 </s>
  </said>
</p>
```

図 2 タグ付与後の XML

以上の行程により、一つの作品につきそれぞれ一つのコーパスが構築される。次に、改作前後の対象関係を示すための属性を与えるために、一文に相当する<s>に属性 @xml:id を付与する。図3にその例を示す。

```
<p>
  <name> 客花情 </name>
  <q>
    <s xml:id="K1.0"> 来ル。 </s>
  </q>
  <said>
    <s xml:id="K1.10"> どふじや。 </s>
    <s xml:id="K1.20"> いかふさむいの。 </s>
  </said>
</p>
<p>
  <name> 中居とよ </name>
  <said>
    <s xml:id="K1.30"> 花情さん。よふおいなはつた。 </s>
    <s xml:id="K1.40"> サア。おあがりなはれ。 </s>
    <s xml:id="K1.50"> これお久米どん。 </s>
    <s xml:id="K1.60"> 御茶あげまさんせ。 </s>
    <s xml:id="K1.70"> サアマ。御上りなはれ。 </s>
  </said>
</p>
```

図 3 ID 付与後の XML

更に、ID を付与された<s>に対し、東西で改変が見られる文について、対応関係を与える。『月花余情』組を例に取るならば、図4のような対応関係を示す属性を与えることが出来ればよい。

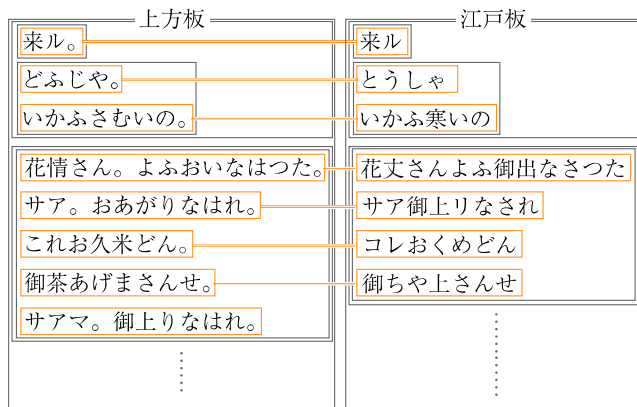


図4 東西対照の構造

本稿ではこの対応関係を示すために、<s>に属性 @corresp[b]を与えることとする。

上方板テキスト中の<s>に対し、対応する江戸板のテキスト<s>のIDを@correspで与え、同様に江戸板テキスト中の<s>に対し、対応する上方板テキストの<s>のIDを@correspで与える、ということになる。以上の行程により、上方板のテキストについては江戸板の対応する箇所が、江戸板のテキストについては上方板の対応する箇所が示される。完成した上方板・江戸板それぞれのXMLの一部を図5、図6に示す。

```
<p>
<name> 客花情 </name>
<q>
<s xml:id="K1.0" corresp="E1.0"> 来ル。 </s>
</q>
<said>
<s xml:id="K1.10" corresp="E1.10"> どふじや。 </s>
<s xml:id="K1.20" corresp="E1.20"> いかふさむいの。 </s>
</said>
</p>
<p>
<name> 中居とよ </name>
<said>
<s xml:id="K1.30" corresp="E1.30"> 花情さん。よふおいなはつた。 </s>
<s xml:id="K1.40" corresp="E1.40"> サア。おあがりなはれ。 </s>
<s xml:id="K1.50" corresp="E1.50"> これお久米どん。 </s>
<s xml:id="K1.60" corresp="E1.60"> 御茶あげませんせ。 </s>
<s xml:id="K1.70"> サアマ。御上りなはれ。 </s>
</said>
</p>
```

図5 @corresp付与後のXML (上方板)

```
<p>
<name> 客花丈 </name>
<q>
<s xml:id="E1.0" corresp="K1.0"> 来ル </s>
</q>
<said>
<s xml:id="E1.10" corresp="K1.10"> とうしゃ </s>
<s xml:id="E1.20" corresp="K1.20"> いかふ寒いの </s>
</said>
</p>
<p>
<name> 中居とよ </name>
<said>
<s xml:id="E1.30" corresp="K1.30"> 花丈さんよふ御出なさつた </s>
<s xml:id="E1.40" corresp="K1.40"> サア御上りなされ </s>
<s xml:id="E1.50" corresp="K1.50"> コレおくめどん </s>
<s xml:id="E1.60" corresp="K1.60"> 御ちや上さんせ </s>
</said>
</p>
```

図6 @corresp付与後のXML (江戸板)

4. 改変の検証

以上の作業によってそれぞれの作品のコーパスを作成し、各テキストに文単位のIDと、各文に対応する改作後・改作前の文のIDを付与した。XML内に対応するIDを記述することにより、一方のテキストからもう一方のテキストを参照することが可能になる。本節ではこのデータを用いて改作前後の比較を行うことにより、対照コーパスを用いることでどのような改変を観察することができるかを検証する。

一言に「改変」と言っても、改作前にはあった場面が改作後では削除されている（若しくは新しい場面が追加されている）か、変更されているといった粗筋ごとの改変や、改作前にあった発話が、改作後では削除されている（若しくは新しい発話が追加されている）か、変更されているという発話単位での改変も見られるが、形態素単位に注目すると、名詞・副詞・動詞（本動詞・補助動詞）・形容詞・助動詞・格助詞・接続助詞といった形態素に追加・削除・変更が見られる。こうした改変は、語彙素単位の書き換え（ばか→あほ／みそ→自慢）の書き換え（言や→言え／やらん→やらぬ）、表記単位の書き換え（さむい→寒い／御休み被成→御やすみなされ）に大別できる。以下、『月花余情』『郭中奇譚』の各組に見られる改変の実例を数例挙げる。

4.1 『月花余情』組

『月花余情』組中で最も顕著なのは、補助動詞の改変である。上方から江戸への改変の過程で、尊敬の補助動詞「なはる」が「なさる」へと変更されている例が、16例見られる。

（上方板）マアせんどはきつうのましなはつたぞへ
 →（江戸板）マウせんどはきつふのませなさつたぞへ
 （上方板）花情さんよふおいなはつた

b ガイドライン[4]では "(corresponds) points to elements that correspond to the current element in some way." と定義されている。

→ (江戸板) 花丈さんよふ御出なさつた

「今はもう何時になってしまったのか」と尋ねる際に、上方板では「もう何時だ」と聴くが、江戸板では副詞「もう」が削除されている。

(上方板) もうなん時じや

→ (江戸板) 何時しや

(上方板) もうなん時じや

→ (江戸板) なん時じや

4.2 『郭中奇譚』組

『郭中奇譚』組については積極的な語彙の改変が見られる。「ばか」を「あほ」に書き換える例が分かりやすい。

(江戸板) ナニ**ばか**ツつらめ

→ (上方板) くそこゝな**あほう**づらめが

(江戸板) **ばか**やめて

→ (上方板) **あほう**らしい事いわずと

敬意を示す補助動詞「しやる」が忌避される傾向が見られる。後者は忌避された表現を改変せず、「しやる」を含む本動詞まで削除した例である。

(江戸板) しは木どのつが**ツしやんな**

→ (上方板) これさよつぎ**やんな**

(江戸板) エ、うそいは**ツしやれ**

→ (上方板) ぬゝうそ

こうした語彙的・文法的側面以外に、音韻の特徴を取り出すこともできる。例えば、次に示すのは形容詞連用形の「〜ク」が、上方板で「〜ウ」のウ音便形に書き換えられている例である。

(江戸板) 今夜はお月様が**よく**さへさしやツた

→ (上方板) こんやはお月さんが**よふ**さへさしやツた

(江戸板) おめエのけふの髪は**高**くゆひなさツたノ

→ (上方板) おまへけふの髪は**高**ふいゝなんしたの

5. 日本語史研究と対照コーパス設計の課題

本稿で対象とした資料については、その改変が本当に「言語変化の一例として観察できるのか」が問題となる。例えば、上方板から江戸板への改変が認められる場合であっても、必ずしも全ての改変に「上方のことばが江戸のことばへ改められた」という説明が与えられるわけではない[c].

c 例えば、『月花余情』組には「今夜」を「今宵」に書き換えた箇所が二箇所見られるが、他の江戸語資料に「今夜」が忌避され「今宵」が好まれる

改作者が明確な意図を持って改変を行ったかどうかを窺い知ることはできないため、明らかに改変が行われている箇所であっても、それが東西の話者としての言語感覚の差異によるものなのか、作者の個人的な言語感覚によるものなのかといった判断がつかない場合が多々ある。「どこからどこまでが意図的な改変なのか」を見極めるためには他の資料の参照が不可欠なものとなるが、今回のような改作前後の単純比較からも一定の成果が得られることが確認できる。

また、今回は TEI/XML に従ってデータを作成したため、永崎研宣氏によって構築された Javascript (jQuery) による、TEI/XML 対応の簡易的なパラレルコーパス表示システムを利用することにより、Web ブラウザ上に東西テキストを並置させ、相互参照を行わせることもできた (図 7)。これはこのまま Web 上に公開することも可能となっている。人文学資料のマークアップにおける国際的なデファクト標準となっている TEI/XML に対応することは、データの再利用可能性を高めるだけではなく、ツールの開発を行うにあたって、その適用可能性を広げられるというメリットがある。

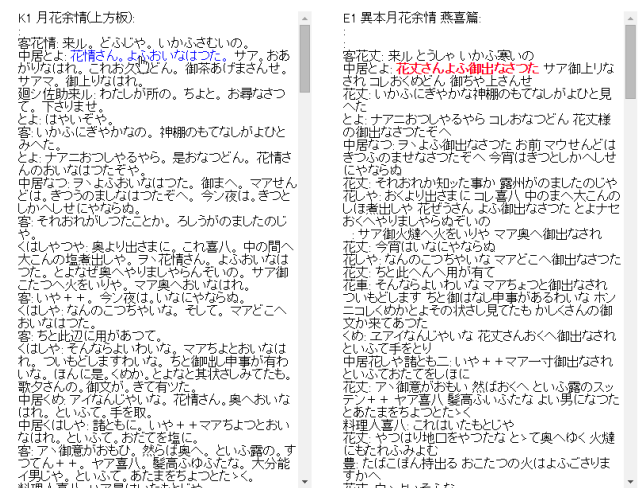


図 7 パラレルコーパス表示システム

なお、本稿では文単位の対応付けのみを行ったが、段落単位、文単位、文節単位、形態素単位での対応付けを行うことが理想的であると思われる。

日本語史研究の手法として対照コーパスを設計・利用したのものには他に、『今昔物語集』『宇治拾遺物語』『日本霊異記』中に見られる同一説話を対象にしたものがある[8,9].

また、2015年3月末には国立国語研究所による共同研究プロジェクト「通時コーパスの設計」の一環として、寛永19年(1642)書写の狂言台本『虎明本』をタグ付きコーパス化した『日本語歴史コーパス室町時代篇 I 狂言篇』が公開された[10,11]. この『虎明本』と、150年下った寛永4

という傾向は別段見られない。こうした改変については、改作者個人の感覚によるものと見る方が穏当であるように思われる。

年（1792）に書かれた『虎寛本』との相違を調査することによって中世から近世の言語変遷を観察する試みはこれまで数多く行われてきている[12,13等].

同一の粗筋を持ちながら、何らかの理由によって内容・語彙に異なりのあるテキストが複数存在する資料における相違を比較することによって、一定の成果が得られることは言うまでもない。そうした資料群に対し、本稿で示した手法を用いてマークアップを行うことで、より客観的且つ再利用の容易な対照コーパスの構築が可能になるものと思われる。

参考文献

- 1) 市村太郎, 河瀬彰宏, 小木曾智信: 洒落本コーパスの構造化, 第3回日本語学コーパスワークショップ予稿集, pp.249-258 (2013).
- 2) 矢野準: 近世後期京坂語に関する一考察 —洒落本用語の写実性—, 国語学, Vol.107, pp.16-33 (1976).
- 3) 増井典夫: 近世後期上方語研究の課題 —近世後期名古屋方言を視野において—, 淑徳国文, Vol.35, pp.47-64 (1994).
- 4) 洒落本大成編集委員会: 『洒落本大成 第3巻』, 中央公論社, (1979).
- 5) 洒落本大成編集委員会: 『洒落本大成 第7巻』, 中央公論社, (1979).
- 6) Text Encoding Initiative, TEI: P5 Guidelines, available from <http://www.tei-c.org/Guidelines/P5/> (accessed 2015-04-01).
- 7) 河瀬彰宏, 市村太郎, 小木曾智信: TEI: P5に基づく近世口語資料の構造化とその問題点, じんもんこん 2013 論文集, pp.7-12 (2013).
- 8) 田中牧郎: 説話のパラレルコーパスの設計 —平安・鎌倉時代の文体変異の研究に向けて—, 第3回コーパス日本語学ワークショップ予稿集, pp.259-268 (2013).
- 9) 田中牧郎, 山元啓史: 『今昔物語集』と『宇治拾遺物語』の同文説話における語の対応 —語の文体的価値の記述—, 日本語の研究, Vol.10, No.1, pp.16-30 (2014).
- 10) 国立国語研究所: 日本語歴史コーパス 室町時代編 I 狂言, 入手先 (<https://maro.ninjal.ac.jp>) (参照 2015-04-01).
- 11) 小林正行, 市村太郎: 『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 (2013).
- 12) 柳田征司: 虎明本狂言と虎寛本狂言との語彙の比較 —困惑の気持ちを表わす感情語彙に就いて—, 安田女子大学紀要 1, pp.31-44 (1967).
- 13) 蜂谷清人: 虎明本から虎寛本へ —語形・用法の変遷とその史的位罫についての試論—, 狂言台本の国語学的研究, 笠間書院 (1977).