

# 『日本語歴史コーパス 室町時代編 I 狂言』形態論情報の概要

2015年3月17日 市村太郎

2016年3月29日 渡辺由貴 更新

## はじめに

本コーパスにおける形態論情報は、検索の便宜・処理上の便宜等を考慮して付与されているため、学術上の通説、あるいは既存の索引類等と異なる尺度で付されたものも存在し、必ずしも「学術的な正しさ」を企図して付与されたものではない。

そのため、場合によっては目的の語がヒットしなかったり、利用者各位の研究目的とは合致しない分類がなされていたりするおそれがある。また「で」(助詞/助動詞)「又」(副詞/接続詞)など、時に品詞分類等が困難なケースも存在する。

研究利用に当たっては、この点を留意の上、目的のものはすべて表示されているかどうか、また付与された情報が研究目的に適うものかどうか、文字列検索・語彙素検索の結果と照合したり、各位において再分類したりする等、多角的に確認・検討することを推奨する。

## 1. 言語単位

『日本語歴史コーパス 室町時代編 I 狂言』では、用例収集を目的とした「短単位」・言語的特徴の解明を目的とした「長単位」の2種類の言語単位を採用している。これは、『現代日本語書き言葉均衡コーパス (BCCWJ)』で採用した単位を基に設計したものである。基となっている『BCCWJ』の言語単位は『日本語話し言葉コーパス (CSJ)』との互換性の保持を図り、国立国語研究所が行った語彙調査の単位を基に設計された。これまでに国立国語研究所が実施してきた語彙調査における言語単位のうち、短い単位の系列に属するものが「短単位」、長い単位の系列に属するものが「長単位」である。

本コーパスの言語単位は、通時的な日本語研究で利用するために、現代語や別の時代のコーパスとの互換性の保持を図っている。その一方で、BCCWJ・平安時代編等の規程をそのまま用いるのではなく、本コーパス用に単位認定規程の修正・拡張を行った。

短単位・長単位とも、代表形(語彙素読み)・代表表記(語彙素)・品詞・活用型・活用形を与える。代表形は国語辞典の見出しに、代表表記はその見出しに与えられた漢字等の表記に相当するものである。

## 2. 短単位の概要

短単位は、言語の形態的側面に着目して規定した言語単位である。短単位の認定にあたっては、まず意味を持つ最小の単位(最小単位)を規定し、その最小単位を文節の範囲内で短単位認定規程に基づいて結合させる(もしくは結合させない)ことで認定する。

### (1) 最小単位

- 最小単位は現代語において意味を持つ最小の単位である。本コーパスにおける最小単位については、現代語との関連を重視して、原則として現代語を対象とした最小単位認定を行うが、必要に応じて、使用実態や平安時代編・近代語コーパスの状況に基づき個別の判断をすることがある。語種等により、次のように認定する。

※「/」は最小単位の分割位置を表す。

和語： /この/あさ/なべ/と/いふ/もの/にて/

漢語： /罪/人/ /ざい/人/ /案/内/者/ /あん/ない/しや/

外来語： /ぼさつ/ /娑婆/ /遮羅婆羅/草/

記号： /。/ / / \ /

人名： /佐奈田/の/与市/義貞/ /白/楽天/

地名： /津/の/国/ /わかさ/の/おばま/

- 上記のように認定した最小単位を、短単位認定のために下表のとおり分類する。

表1 最小単位の分類

| 分類   | 例                                 |
|------|-----------------------------------|
| 一般   | 和語：花ほどしたたかかたじけない笑う…               |
|      | 漢語：連歌来臨…                          |
|      | 外来語：菩薩 娑婆 ぼろおん…                   |
| 付属要素 | 接頭的要素：相(あい) 御(お、ご、み) 大(おお) 不(ぶ)…  |
|      | 接尾的要素：殿(どの) 兼ねる がましい 気(げ) 立て(だて)… |
| その他  | 記号：、。『』…                          |
|      | 数：一二十百千… 幾数何…                     |
|      | 固有名                               |
|      | 助詞・助動詞                            |

(2) 短単位

- 短単位の認定規定は、上表の分類ごとに適用すべき規定が定められる。その規定に基づき、最小単位を結合させる(又は結合させない)ことによって、短単位を認定する。以下、「一般」・「数」・「その他」に分けて、短単位認定規定の概要を示す。  
「|」は短単位の分割位置を、「=」は短単位を切らないことを示す。

[1] 一般

《和語・漢語》

最小単位2つの結合までを1短単位とする。

【例】 |ちち| |はは| |ちち=はは| |わび=こと| |よ=ぶか(夜深)|  
|冠=者| |言=語| |道=断| |理| |不=尽| |案=内| |者|

例外：切る位置が明確でないもの、あるいは切った場合と一まとめにした場合とで意味にずれがあるものは、3最小単位以上の結合であっても1短単位とする。

【例】 |新発意| |殊の外|

例外：最小単位が3つ以上並列した場合、それぞれの最小単位を1短単位とする。

【例】 |雪| |月| |花|

《外来語》

1 最小単位を1短単位とする。

【例】 |遮羅婆羅| |草| |金剛| |夜叉|

[2] 数

「数」以外の最小単位と結合させない。「数」どうしの結合は、一・十・百・千の桁ごとに1短単位とする。「万」「億」等は、単独で1短単位とする。

【例】 |四十| |八| |手| |五百| |八十| |年| |二| |万| |億|

[3] その他

1 最小単位を1短単位とする。

付属要素 |お| |前| (二人称代名詞でない場合) |われ| |ら| |住み| |がたし|

助詞・助動詞 |各| |を| |申し| |れ| |させ| |られ| |たら| |ば| |よう| |ござら| |う|

人名 |源| |実朝| |呂| |洞賓| |くわんむ| |天王|

地名 |伊豆| |の| |国| |ひるがこ嶋| |ひえい| |ざん|

- 短単位データの作成は自動形態素解析と人手修正によって行われている。形態素解析処

理は形態素解析器に「MeCab」、解析用辞書に「近世口語 UniDic」を使用している。

### 3. 長単位の概要

長単位は、言語の構文的な機能に着目して規定した言語単位である。長単位の認定は、文節の認定を行った上で、各文節の内部を規定に従って自立語部分と付属語部分とに分割していくという手順で行う。

#### (1) 文節

- 長単位の認定にあたっては、まず文節の認定を行う。現代語の文節は、一般に付属語又は付属語連続の後ろで切れる。このほかに、本コーパスでは、付属語を伴わない自立語であっても、主語・主題、連用修飾、連体修飾の各成分の後ろで切るといった規定を設けた。
- 文節を認定する上で問題となることの一つに、固有名、「一が～」「一つ～」「一の～」で1短単位と認める体言句、複合辞がある。これらについては、内部にある付属語の後ろでは切らないこととする。
- 複合辞は、本コーパス内での用例数、BCCWJにおける類例の認定状況等を基準に一部認定し、付属語として認めた。

※「|」は文節の分割位置を、「=」は文節を切らないことを表す。

| 木村の源三成綱 | | 雁が音 | | うへつかた | | 竹の子 |  
| に=よつ=て | | で=御ざる |

#### (2) 長単位

- 長単位は、上記の文節を規定に基づいて分割する（又は分割しない）ことによって認定する。文節を超えることはない。以下、長単位認定規定の概要を示す。

※「|」は長単位の分割位置を、「=」は長単位を切らないことを示す。

[1] 記号は1長単位とする。

【例】 | 〔 | 罷出 | たる | 者 | は | 〕 | 東国 | に | かくれ | も | なひ | 大名 | です | 〕 |

[2] 付属語は1長単位とする。

【例】 | 「 | 罷出 | たる | 者 | は | 」、 | 東国 | に | かくれ | も | なひ | 大名 | です | 」、 |

[3] 主語・主題、連用修飾成分、連体修飾成分の後ろで切る。

【例】 | 御年貢 | を | 上る | 事 | めでたう | おぼしめす | 」、 |

[4] 体言に形式的な意味の「為る」「遊ぶ」「致す」「仕る」「為される」「参る」「召さる」「召す」「申す」が直接続く場合、切り離さない。

【例】 | 成就=し | た | | 同道=いたす |

[5] 「御～有る・為る・候う・仕る・為される・召す・申す」「～遊ぶ・有る・致す・おわします・候う・奉る・給う・仕る・為される・参らせる・参る・まします・申す・遣る」のような形式の敬語表現は、全体を1長単位とする。

【例】 | お尋ね=申さ | う | | 申=奉る |

上記形式中に付属語が含まれる場合、切り離さない。

【例】 | まぼら=せ=給ふ |

[6] 同格の関係にある体言連続は切り離さない。

【例】 | さつまのかみ=ただのり |

[7] 並列された語は切り離さない。

【例】 | 西山=東山 | | 金襴=どんす=どんきん |

[8] 係り受けを重視し、付属語を切り出すのは不適切なものを連語として認める。

【例】 | かまひ=て | (※副詞) | さり=ながら | (※接続詞)

- 長単位データの作成は、人手修正済み短単位データを基に、長単位解析器 Comainu によって長単位の自動構成を行っている。

#### 4. 他のコーパスと異なる処理・特殊な処理

本コーパスのデータの作成にあたっては、原則現代語のコーパスと平安時代編における処理を踏襲して行った。ただ、古代語から現代語への過渡的様相を示す虎明本狂言集においては、現代・中古いずれの規程に拠っても一律に処理できないケースがしばしば現れる。そのため、可能な限り既存の枠組みを尊重しつつ、独自の処理、あるいは特殊な処理を行った箇所がある。以下にはそのうち、全体に関わる特に注意すべきものを挙げる。

##### [1] 文語活用と口語活用

現代語のコーパスおよび『日本語歴史コーパス』では、活用語について「文語」「口語(明示なし)」の二大別を行っている。ところが虎明本狂言集は、前述のとおり古代語から現代語への過渡的様相を示す資料であり、いずれに拠っても処理が困難な事例が現れる。そのため、品詞・語により方針を立てて処理を行った。

- 動詞は原則「文語」活用と見、口語活用でなければ対応できないものを「口語」とした。これは「文語」に分類される上・下二段活用動詞の一段化が進んでいないことによる。

【例】 <<文語>>

| 此 | あたり | の | もの | で | 御ざる |

→動詞-非自立可能・文語四段-ラ行・終止形-一般

| 申あぐる | 事 | は | 迷惑 | で | 御ざる |

→動詞-一般・文語下二段-ガ行・連体形-一般

<<口語>>

| 扇 | を | かほ | に | あて | 、 | ねる |

→動詞-一般・下一段-ナ行・終止形-一般 (文語下二段では処理不可)

- 形容詞型活用は原則口語活用と見、文語活用でなければ対応できないものを「文語」とした。ただし「一けれ」の形は文語・已然形とする。詳細は渡辺他(2015)参照。

【例】 <<口語>>

| 一段 | めでたい |

→形容詞-一般・形容詞・終止形-一般

<<文語>>

| 誠 | に | めでたき | 御 | 事 |

→形容詞-一般・文語形容詞-ク・連体形-一般

| 鷹くひ | に | なる | こそ | 、 | めでたけれ |

→形容詞-一般・文語形容詞-ク・已然形-一般

##### [2] 終止形・連体形の別

- 文語サ行変格活用等には、連体形に相当する形態で文末終止を行う場合がある。このような場合は、文末であっても終止形ではなく連体形とした。

【例】 | いそい | で | 罷のぼら | ふ | と | 存る |

→動詞-一般・文語サ行変格・連体形-一般

| いつも | うへとう | へ | 御 | ねんぐう | を | ささぐる |

→動詞-一般・文語下二段-ガ行・連体形-一般

- 終助詞や助動詞に前接する場合、終止・連体形の区別が困難なケースが多い。そこで、形態的に明らかなものはその活用形とし、終止・連体同形の物は、極力平安時代編や小椋他(2011)に合わせ、終止形・連体形いずれかに統一した。

### [3] 助動詞「う」と意志推量形

- 本コーパスでは助動詞「う」・助動詞「むず」の語形「うず」を立て、未然形+助動詞「う」・「むず」と、用言と助動詞を分割する。  
「一人づつゆかう」などの「行こう」は、現代語のコーパスでは口語活用の「意志推量形」とされているが、この「意志推量形」は、現在の規程上、本コーパスの多くの動詞が該当する文語活用としては用いることができない。また、文語認定した動詞について、未然形のみを口語と認定することも考え得るが、「うずる」のような「うず」型の形態に対しては対応が困難である。  
【例】 |くらま|へ|同道|いたい|て|参ら|う|  
→動詞-非自立可能・文語四段-ラ行・未然形-一般+助動詞・無変化型・終止形-一般  
|おれ|こそ|子共|や|孫|を|つかは|ふずれ|  
→動詞-一般・文語四段-ハ行・未然形-一般+助動詞・文語助動詞-ムズ・已然形-一般
- 文語四段動詞の場合、実際の発音は多くの場合オ段であり、本来（口語）五段活用とすべきものであるが、多くの場合、「行かう」のように活用語尾がア段の仮名で表記されており、表記上四段活用として処理することが可能である。そのため、これらのものは便宜的に文語四段活用未然形とした。

- ただし、未然形がオ段の仮名で表記されている場合は、（口語）五段活用の意志推量形と認定した。表記ベースでも「四」段と認定できないためである。

【例】 |まだ|夜ぶか|さう|な|程|に|、|まどろもふ|  
→動詞-一般・五段-マ行・意志推量形

### [4] 語尾が「い」となる命令表現

- 活用型・活用形によって以下のように対応した。  
【例】 <<四段動詞未然形>>  
|めでたい|程|に|、|うたわ|ひ|  
→文語四段・未然形+助動詞「い」命令形  
<<四段動詞命令形>>  
|何|ぞ|ある|か|だせ|ひ|やひ|  
→文語四段・命令形+助詞-終助詞「い」  
<<下一段・下二段型>>  
|つま|と|さだめひ|  
→動詞-一般・下一段-マ行・命令形  
|矢|の|根|を|けづら|れひ|  
→助動詞・文語下二段-ラ行・命令形

### [5] 「お～やる」における助動詞「やる」（短単位）

- 「御」+動詞連用形に後続する「やる」は、原則助動詞「やる」とする。  
【例】 |お|も(持)|ち|やら|う|か|お|もち|やる|まひ|か
- 「おもやる」（「お思いある」の転）と「おりやる」（「お入りある」の転）については融合が進み、部分を切り出すのが困難なため、一短単位と認定した。  
【例】 |な|をし|たひ|と|おもやら|ば|

→動詞一般・文語四段・ラ行・未然形一般  
 |惣名|で|おりやる|よ|なふ|  
 →動詞一般・文語四段・ラ行・連体形一般

[6] 連語での発音の融合が想定される箇所

- 狂言台本には、[出う／イジョー] [今日は／コンニッタ]のように、台本の表記から通常想定される発音と、舞台上で通用される発音が異なるものが存在する。このようなものについては、明らかにそう発音させる意図があることが推測され、かつ言語単位上不都合が生じない表記がなされていない限りは、原則舞台上の通用の読みは反映させない。

[7] 活用語尾が表記されていない語の音便化の判定

漢字表記された動詞には、活用語尾が表記されていないものが見られる。これらのうち、「致た」「云た」「聞て」「持て」など、現代語あるいは舞台上では通常連用形が音便化するものが多く存在するものの、虎明本狂言集の台本で音便化しているのかどうかの判断は難しい。そこで以下の方針で処理を行った。詳しくは渡辺他(2015)参照。

- 「た」が後接する場合、あるいは仮名表記に音便化した形しかないものについては原則音便化した形として認定する。

【例】《助動詞「た」が後接》

|うりて|の|云|た|ごとく|云|  
 →動詞一般・文語四段・ハ行・連用形・ウ音便

|や、|思ひ出|た|、|  
 →語彙素：思い出す・動詞一般・文語四段・ハ行・連用形・イ音便

【例】《仮名表記に音便化した形のみ》

|よそみ|し|て|お|み|やら|ぬ|に|依|て|さやう|に|仰|らるる|  
 →語彙素：因る・動詞一般・文語四段・ラ行・連用形・促音便

- 他のものは積極的に音便化させない。

【例】|当年|も|年|を|とり|に|ゆか|ふ|と|云|て|  
 →動詞一般・文語四段・ハ行・連用形一般

【例】|慥|に|成敗|は|致|て|御ざる|が|  
 →動詞・非自立可能・文語四段・サ行・連用形一般

[8] 表記等による語彙素の統合・分割の判断

小椋他(2011)の「同語異語判別規程」に拠るなどし、意味・機能の別と一致しない場合であっても、表記等によって語彙素の統合・あるいは語彙素の分割の判断を行った場合がある。

【例】

| 語彙素 | 語形  | 書字形 | 備考  |
|-----|-----|-----|---|
| 表   | オモテ | おもて | 「表面」を意味しながら「面」と表記される場合、その漢字表記を優先し語彙素「面」とする。「顔」を意味する「表」表記も同様に「表」とする。仮名書き例は、「顔」の意が明確な場合に限り「面」とする。 |
|     |     | 表   |   |
| 面   | オモテ | おもて |   |
|     |     | 面   |   |
| 物   | モノ  | もの  | 仮名表記の場合、「人」を意味する場合は「者」、それ以外は「物」とした。漢字表記の場合、たとえば   |
|     |     | 物   |   |
| 者   | モノ  | もの  | 「物」表記で「人」を意味する場合であっても語彙素は「物」とする。  |
|     |     | 者   |   |
| 尋ねる | タズヌ | たづぬ | 仮名表記の場合、質問相手の存在が明確な場合は「尋ねる」、それ以外は「訪ねる」とした。漢字表記の場合、たとえば「尋ぬ」表記で「訪問」を意味する場合であっても語彙素は「尋ねる」とする。      |
|     |     | 尋ぬ  |   |
| 訪ねる | タズヌ | たづぬ |   |
|     |     | 訪ぬ  |   |

[9] その他

- 存在動詞「ゴザアル」は1短単位と見、語彙素「御座る」の語形とした。また意味・機能が対応する「ゴザナイ」「ゴザナシ」も1短単位と認め、形容詞「御座無い」とした。
- [5] のとおり「おりやる」を1短単位としたため、意味・機能が対応する「オリナイ」「オリナシ」についても1短単位と認め、形容詞「おりない」とした。
- 解釈が不明な箇所や単語情報付与が困難な箇所等は「未知語」として扱い、その種別を「品詞」欄に表示する。なお文字列検索は可能である。

表2 未知語の種別 (品詞)

| 種別 (品詞) | 内容   | 例   |
|---------|--|---|
| 歌・呪文ほか  | 歌や呪文、笛の音など、語としての切れ目を付け難い、あるいは切れ目の分らないもの。   | どうらろうろらりり   (楽阿弥)<br>  ちりぬるをわか   (いろは)<br>  つれてん / \ てん / \ てん   (こぶうり) |
| 解釈不明    | 解釈不明の箇所。他の参考資料で解釈がなされていても、底本の注記に「不詳」とあれば原則それを優先する。また意味は不詳でも確実に単語や読みの認定可能なものは短単位認定する。 | くわつしぼぼ   (せつぶん)<br>  むたうじやう   (なきあま)<br>  びんななんほうす   (ひいたる舞)            |
| 漢文      | 訓点等のない漢文箇所。訓点のある箇所については原文に従い極力読み下している。   | 不肯永落生死   (ふせなひきやう)  |
| 経文      | 経典の読誦等   | ぐわんにしくどくふぎうお一切   (なきあま)   |
| 言いよどみ   | 言いよどみ  | う   うし   や   (せつぶん)   |
| 言い間違い   | 言い間違い・記憶違い等による臨時的な形態   | 「   名   は   ぐひす   と   やら   申   (さつくわ)                                   |
| 長大な人名   | 長大な人名  | ぎんばばいにばいやれ   (昆布柿)  |
| 洒落      | 洒落や掛詞等による臨時的な形態  | さはぐんなり   (どひつ)  |

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  
[http://www.ninjal.ac.jp/corpus\\_center/bccwj/doc/report/JC-D-10-05-02.pdf](http://www.ninjal.ac.jp/corpus_center/bccwj/doc/report/JC-D-10-05-02.pdf)  
(2015 年 3 月 4 日閲覧)

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  
[https://dl.dropboxusercontent.com/u/73297026/report/unidic-EMJ\\_rulebook2012.pdf](https://dl.dropboxusercontent.com/u/73297026/report/unidic-EMJ_rulebook2012.pdf)  
(2015 年 3 月 4 日閲覧)

渡辺由貴・市村太郎・鴻野知暁（2015）『『虎明本狂言集』のコーパスデータにおける短単位認定の諸問題』第7回コーパス日本語学ワークショップ予稿集 pp.233-240