『日本語歴史コーパス　室町時代編Ⅲ抄物』Ver.0.8
形態論情報の概要

2026 年 3 月 31 日 村山実和子・渡辺由貴・北﨑勇帆

はじめに

本コーパスにおける形態論情報は、検索・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は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）等、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（CHJ）』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。

『日本語歴史コーパス室町時代編Ⅲ抄物』の言語単位は、原則として「『日本語歴史コーパス室町時代編』形態論情報規程集　Ver.1.0」にしたがっている（ただし、抄物については 2026 年 3 月時点では短単位データのみ公開）。短単位には、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、使用実態や室町時代編所収のその他作品・明治・大正編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:　／いつ／の／ま／に／やら／秋風／吹／て／すずしく／なる／ほど／に／

漢 語:　／念／願／　／学／問／　／今／夜／

外来語:　／菩薩／　／菩提／

記 号:　／、／　／。／　／／＼／

人 名:　／陶／淵明／　／項羽／　／諸葛／孔明／

地 名:　／廬山／　／赤壁／　／斉／の／国／

- 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に分類する。

表1 最小単位の分類

分類		例
一般		和語:花 ほど 穏やか 面白い 笑う …
		漢語:艱 難 辛 苦 …
		外来語:菩薩 菩提 …
付属要素		接頭的要素:御(お、ご、み) 不(ぶ) 相(あい)…
		接尾的要素:さ 共(ども) 難い(がたい) 兼ねる …
その他	記号	、 。「 」 ／ ＼…
	数	一 二 十 百 千…数 何…
	固有名	人名:李白 白 楽天…
		地名:磻渓 呉 咸陽…
	助詞・助動詞	の を ぞ こそ まで る・らる ず ごとし やる なり う…

(2) 短単位

- 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「中世口語 UniDic」を使用している。
- 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】｜秋＝風｜　｜聞き＝及ぶ｜　｜作＝者｜　｜言＝語｜道＝断｜　｜白｜牡＝丹｜　｜当世｜かたぎ｜

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】｜身の上｜　｜ほしいまま｜

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】｜眼｜耳｜口｜鼻｜

《外来語》

1 最小単位を 1 短単位とする。

【例】｜阿耨｜菩提｜　｜金剛｜蔵王｜菩薩｜

［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】｜一｜度｜　｜八十｜万｜騎｜　｜六百｜二十｜九｜年

［3］その他

1 最小単位を 1 短単位とする。

付属要素　　｜御｜恩｜　｜我｜等｜　｜たへ｜がたき｜

助詞・助動詞　｜其｜中｜に｜て｜一｜のみ｜にくから｜ん｜を｜胡国｜へ｜つかわさ｜る｜べき｜也｜と｜ある｜処｜で｜

人名　　｜諸葛｜孔明｜　｜白｜楽天｜

地名　　｜秦｜の｜国｜　｜浣花｜渓｜

3. 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、原則として「『日本語歴史コーパス 室町時代編』形態論情報規程集 Ver.1.0」にしたがっている。以下、『日本語歴史コーパス』全体に関わるもののうち、『室町時代編』共通の処理(3.1)、『室町時代編Ⅲ抄物』独自の処理として、特に注意すべきものを挙げる(説明に際し、一部に『室町時代編』の別コーパスの用例を含む)。

3.1. 『室町時代編』以外のシリーズと異なる処理・特殊な処理

[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。ところが抄物資料は、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う。

- 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」とする。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≫

｜徒然｜の｜あまり｜咸陽｜の｜市｜へ｜ゆき｜て｜、｜酒｜を｜のむ｜ぞ｜。｜

→動詞-非自立可能・文語四段-マ行・終止形-一般

｜春雨｜の｜中｜に｜緑｜が｜のぶる｜ほど｜に｜

→動詞-一般・文語上二段-バ行・連体形-一般

≪口語≫

｜人｜の｜帯びる｜こと｜も｜なく｜し｜て｜無事｜に｜し｜て｜ある｜ぞ｜。｜

→動詞-一般・上一段-バ行・連体形-一般

- 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「―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≫

｜此花｜面白い｜と｜云ふ｜心｜は｜ない｜ぞ｜。｜

→ 形容詞-一般・形容詞・終止形-一般

≪文語≫

｜中｜に｜就い｜て｜も｜海棠｜は｜面白き｜花｜なれ｜ども｜

→形容詞-一般・文語形容詞-ク・連体形-一般

｜鳥｜も｜一｜つ｜二｜つ｜こそ｜面白けれ｜

→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

- 文語上下二段・カ変・サ変・ナ変・ラ変動詞には、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】　｜これ｜に｜つい｜て｜、｜三笑｜は｜画｜そらごと｜に｜て｜ある｜。｜
→動詞-一般・文語ラ行変格・連体形-一般
｜それ｜を｜哭する｜と｜云｜也｜。｜
→動詞-一般・文語サ行変格・連体形-一般
｜水｜が｜そろ／＼｜と｜流るる｜也｜。｜
→動詞-一般・文語下二段-ラ行・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・連体同形のものは、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。
「はや酒をのまうとするぞ」等の「飲もう」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】　｜晩景｜に｜はや｜酒｜を｜のま｜う｜と｜する｜ぞ｜。｜
→動詞-非自立可能・文語四段-マ行・未然形-一般＋助動詞・無変化型・終止形-一般
｜我｜も｜帰ら｜うずる｜こと｜なれ｜ども｜、｜帰ら｜ぬ｜ぞ｜。｜
→動詞-一般・文語四段-ラ行・未然形-一般＋助動詞・文語助動詞-ムズ・連体形-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する。
- 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する。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｜(虎明本狂言・なべやつばち)
→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｜（虎明本狂言・びくさだ）
→文語四段・未然形＋助動詞「い」命令形
≪四段動詞命令形≫
｜学問｜を｜めされ｜い｜と｜
→文語四段・命令形＋助詞-終助詞「い」
≪下一段・下二段型≫
｜其｜礼｜に｜所領｜を｜くれい｜と｜云ふ｜。｜
→動詞-一般・下一段-ラ行・命令形
｜矢｜の｜根｜を｜けづら｜れひ｜（虎明本狂言・はなとりずもう）
→助動詞・文語下二段-ラ行・命令形

［5］その他

- 『鎌倉時代編』までは複合動詞を認めていないが、『室町時代編』以降のコーパスでは、「最小単位２つの結合までを１短単位とする」というルールにのっとり、なるべく結合させる方針をとっている。
- 存在動詞「ゴザアル」は1短単位と見、語彙素「御座る」の語形とする。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」とする。

3.2.『室町時代編Ⅲ抄物』独自の処理

※以下の処理は、Ver.0.8 における暫定的な処理であり、今後のアップデートに際して、方針は変更される可能性がある。

［1］語の読みの判定

- 漢字表記された語について、原則として、本コーパス構築のために作成された校訂本文テキストの「読み」に関する情報を参照し、語形を確定している。
- また、活用語尾が明示されない活用語に関しては、コーパス用に本文を変換する際、テキストに送り仮名等を補入した場合がある。それらについては原文文字列（原文KWIC）列にて、補う前の文字列が復元される（詳しくは別紙「テキストの凡例と中納言表示項目について」を参照）。

［2］漢文・抄文中の引用箇所の取り扱い

- 『中華若木詩抄』では、漢詩（詩題・作者も含む）とそれに対する抄文が交互に掲載される構成となっている。まず、漢詩（詩題・作者）については、本文種別を「漢文」と大別したうえで、一律に「未知語」として扱い、品詞を「漢文」とした（表２）。
- 抄文中には、直前の漢詩からの引用である漢字文字列や、別の漢籍等から引用した漢文・

漢詩の一部が含まれることがある。それらに対して、以下のような方針で処理を行った。

① 三字以上の漢字文字列は、一律に「未知語」として扱い、品詞を「漢文」とした（表 2）。

② 二字以下の漢字文字列は、個別の対応を行った。基本的には、漢語の名詞として処理し、文脈上、読み下すことが適当だと思われる箇所については、読み下したうえで形態論情報を付与した。しかしながら、漢詩・漢文の引用で、読みが確定できない、あるいは「語」としての認定が困難なものについては、「未知語」として扱い、品詞を「漢文」とした。

- 品詞を「漢文」として処理したものは、短単位検索においては検索対象外となる（文字列検索や、検索条件式を利用した検索は可能である）。

表 2 漢文相当箇所に対する本文種別と品詞付与の例

キー	品詞	本文種別
聞子規	漢文	漢文-詩題
荀鶴	漢文	漢文-作者
楚天空闊月成輪	漢文	漢文-漢詩
蜀魄声々似告人	漢文	漢文-漢詩
啼得血流無用処	漢文	漢文-漢詩
不如緘口過残春	漢文	漢文-漢詩
一二	名詞-数詞	（空欄）
の	助詞-格助詞	（空欄）
句	名詞-普通名詞-一般	（空欄）
は	助詞-係助詞	（空欄）
、	補助記号-読点	（空欄）
楚天空闊	漢文	（空欄）
は	助詞-係助詞	（空欄）

参考文献

小椋秀樹・小磯花絵・冨士池優美・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 言語変化研究領域（片山久留美）編（2019）「『日本語歴史コーパス 室町時代編』形態論情報規程集 Ver.1.0」https://clrd.ninjal.ac.jp/chj/morph_muromachi_v1_0.pdf.pdf（2026年3月10日最終閲覧）

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