

『日本語歴史コーパス鎌倉代編Ⅲ軍記』 形態論情報の概要

2022年4月30日 片山久留美

1. はじめに

『日本語歴史コーパス鎌倉時代編Ⅲ軍記』（以下「本コーパス」）は、短単位・長単位の2種類の言語単位のデータを公開している。先行して公開された『鎌倉時代編Ⅰ説話・随筆』『鎌倉時代編Ⅱ日記・紀行』（以下「Ⅰ説話・随筆」「Ⅱ日記・紀行」）と同じく、短単位については『『日本語歴史コーパス鎌倉時代編』短単位規程集 Ver1.0』（以下「鎌倉時代編規定集」）、長単位については『『日本語歴史コーパス平安時代編』形態論情報規程集』に基づいて形態論情報を付与している。しかし、本コーパスには「Ⅰ説話・随筆」「Ⅱ日記・紀行」に見られない語彙や語法が見られる場合があり、特に短単位において上記の規程と異なる処理をした箇所がある。以下では、本コーパスを使用する際に注意が必要な点について述べる。

2. 短単位の形態論情報

2.1 意志・推量を表す形

本コーパスでは、意志・推量を表す形式として「あづからう」「参らせう」のような「～う」という形が現れる。この形式は「Ⅰ説話・随筆」「Ⅱ日記・紀行」には見られないが、「室町時代編Ⅰ狂言」「室町時代編Ⅱキリシタン資料」においては頻出する。「室町時代編」ではこの形式を助動詞「う」という語彙素を使用して処理しているが、本コーパスではこれを助動詞「む」からの変化の過渡期的な例と見て、語彙素「む」の発音形出現形「う」として形態論情報を付与している。

【例】

まづ門出のあしさよ。逆櫓をたてうとも、かへさま櫓をたてうとも、殿原の舟には百挺千挺もたて給へ。

語彙素「立てる」動詞-一般, 文語下二段-タ行, 未然形-一般 (30-平家 1250_11001, 13440)

語彙素「む」助動詞, 文語助動詞-ム, 終止形-一般 (30-平家 1250_11001, 13460)

(参考・「室町時代編Ⅰ狂言」での処理)

すこし成ともにとやうにちいさうたてうとおもふほどになんじもよくみておけ

語彙素「立てる」動詞-一般, 文語下二段-タ行, 未然形-一般 (40-虎明 1642_02005, 9280)

語彙素「う」助動詞, 無変化型, 終止形-一般 (40-虎明 1642_02005, 9300)

2.2 姓と地名

本コーパスには、作品の性質上多くの武士・貴族の名前が登場する。武士の名前の中には、本拠とする地名をそのまま家名または名字として名乗るものが見られる。地名に由来する名字については品詞を「名詞-固有名詞-地名-一般」とすることも考えられるが、本コーパスでは文脈上特定の個人を指すことが明らかな場合、「名詞-固有名詞-人名-姓」とした。

【例】

稲毛三郎重成 熊谷次郎直実 田代冠者信綱 土肥次郎実平 (名詞-固有名詞-人名-姓)

2.3 語の読みの認定

鎌倉時代編規定集では、複数の読みの可能性がある語について「基本読み」を定めることでコーパス内での読みの揺れを抑えてきた。しかし、「基本読み」は「平安時代編」の和文資料をもとに設定されており、本コーパスのような和漢混淆文のテキストにはそぐわない場合も多い。そこで、本コーパスでは「基本読み」を採用することはせず、テキストに即して読みを決定することとした。おおむね以下の方針に従っている。

- ・ルビのある例についてはルビを最優先して読みを決定する
- ・ルビがない場合、同作品・同章段内にルビのある同形の例があればそのルビに従う
- ・ルビのある例が本コーパス中に1例もない場合、注釈書・『日本国語大辞典第二版』などの辞典類を参考に適切と思われる読みを決定する

【例】

香染の御 (ルビ：おん) 衣の袖、しぼりもあへ給はねば、大衆もみな涙をぞながしける。

語彙素「御」 語彙素読み「オン」 発音形出現形「オン」 接頭辞

(30-平家 1250_02002, 9800)

この時の内裏 (ルビ：だいら) の有様を見聞きては、耳をも目をも洗ひぬべくぞ思ゆる

語彙素「内裏」 語彙素読み「ダイリ」 発音形出現形「ダイリ」 名詞-普通名詞-一般

(30-平治 1246_01009, 21820)

(参考・「I 説話随筆」「II 日記・紀行」における「基本読み」の例)

大宮院、顕紋紗の薄墨の御 (ルビ：おん) 衣、鈍色の御衣引き掛けさせたまひて、同じ色の小几帳立てられたり。

語彙素「御」 語彙素読み「オオン」 発音形出現形「オオン」 接頭辞

(30-とは 1306_01013, 16420)

蘇我の大臣勅を奏奉して、僧を召して、初めて内裏 (ルビ：だいら) に入れつ。

語彙素「内」 語彙素読み「ウチ」 発音形出現形「ウチ」 名詞-普通名詞-副詞可能

(30-今昔 1100_11001, 15040)

2.4 「一の～」の追加

「一が～」「一つ～」「一の～」で一短単位とする語について、鎌倉時代編規定集「要注意語」で認定されているものに加え、本コーパス独自に下記の一短単位とした。

家の子（イエノコ） 犬の子（エノコ） 剛の者（ゴウノモノ） 関の声（トキノコエ）

3. 長単位の形態論情報

長単位については、基本的に『日本語歴史コーパス平安時代編』形態論情報規程集に基づいており、大きく方針を変更した箇所はない。

なお、『平家物語』には以下の【例】(a)のように、校訂本文にはない読み添えの「の」がルビに示されている例が見られる。本コーパスではこうした場合でも校訂本文に「の」を補うことはしていない。また、短単位の規定において「名詞-固有名詞-人名-姓」および「名詞-固有名詞-地名」以外の語については読み添えの「の」を含む語形の使用を認めていないため、短単位の形態論情報には読み添えの「の」の情報が含まれないことがある。こうした場合、校訂本文および形態論情報に「の」が含まれていなくても、文節・長単位においては校訂本文に「の」がある場合と同様の規定に従って分割または結合することとした。【例】(a)(b)のように、校訂本文は同じでも文節・長単位の処理が異なる場合があるので注意が必要である。

【例】

(a) 平朝臣清盛公、法名浄海、ほしいままに国威をひそかにし、朝政（ルビ：てうのまつりごと）を乱り、内につけ外につけ、恨をなし歎をなす間

(30-平家 1250_04008, 4810)

●短単位

語彙素「朝」 語彙素読み「チョウ」 発音形出現形「チャー」 名詞-普通名詞-一般

語彙素「政」 語彙素読み「マツリゴト」 発音形出現形「マツリゴト」 名詞-普通名詞-一般

●文節（「|」で文節境界を示す）

| 平朝臣清盛公、 | 法名浄海、 | ほしいままに | 国威を | ひそかに | し、 | 朝 | 政を
| 乱り、 | 内に | つけ | 外に | つけ、 | 恨を | なし | 歎を | なす | 間、 |

※文節認定にあたっては、ルビ「てうの」により「の」で文節が切れるとみなし、「朝」と「政を」をそれぞれ一文節とする。

●長単位

語彙素「朝」 語彙素読み「チョウ」 出現発音形「チャー」 名詞-普通名詞-一般

語彙素「政」 語彙素読み「マツリゴト」 出現発音形「マツリゴト」 名詞-普通名詞-一般

(b) ひたすら朝政（ルビ：あさまつりごと）をすすめ申させ給ふ御有様なり。

(30-平家 1250_01007, 14960)

●短単位

語彙素「朝」 語彙素読み「アサ」 出現発音形「アサ」 名詞-普通名詞-副詞可能

語彙素「政」 語彙素読み「マツリゴト」 出現発音形「マツリゴト」 名詞-普通名詞-一般

●文節

■ ひたすら ■ 朝政を ■ すすめ申させ給ふ ■ 御有様なり。 ■

※ルビ「あさまつりごと」とあることから、複合語と見て「朝」と「政」で文節を切ることはしない。

●長単位

語彙素「朝政」 語彙素読み「アサマツリゴト」 名詞-普通名詞-一般

【参考文献】

国立国語研究所コーパス開発センター（池上尚）編（2016）『『日本語歴史コーパス平安時代編』形態論情報規程集』 <https://clrd.ninjal.ac.jp/chj/doc/morph-heian-2016.pdf>

国立国語研究所コーパス開発センター（鴻野知暁）編（2017）『『日本語歴史コーパス鎌倉時代編』短単位規程集 Ver.1.0』

https://clrd.ninjal.ac.jp/chj/doc/morph_kamakura_v1_0.pdf