

『日本語歴史コーパス』への形容詞の語義・用法情報の追加

山崎 誠¹ 村田 菜穂子² 前川 武³ 村山 実和子⁴¹国立国語研究所 研究系 ²大阪国際大学 基幹教育機構³大阪国際大学短期大学部 ライフデザイン学科 ⁴日本女子大学 文学部

yamazaki@ninjal.ac.jp nmurata@oiu.jp tmaekawa@oiu.jp murayamam@fc.jwu.ac.jp

概要

これまで公開されてきた『日本語歴史コーパス』[1] (以下「CHJ」という)には、時代名などのコーパス情報、品詞などの形態論情報、本文種別などの本文情報、ジャンルなどの作品情報、作者などの作者情報、ページ番号などの底本情報、底本の該当箇所へのリンクなどの外部リンクが収録されているが、語義や用法などの情報は付与されていない。

そのため、語義や用法についての分析を行うには、個々の研究者が独自にデータを作成するよりほかなかった。しかし、それでは非効率的であるため、本研究では、CHJに形容詞の語義情報と用法情報を追加するためのデータを作成し、その応用例を示すこととした。

1 はじめに

CHJに対する語義情報の付与については、池上(2017)[2]がCHJ(平安時代編)に出現する形容詞22語を対象に外部リンクにあるジャパンナレッジ新編全集の当該ページの現代語訳を参考に『日本古典対照分類語彙表』[3]の分類語彙表番号を付与する試みを行っている。その後、浅原ら(2017)[4]は同様の手法で平安時代の4作品について古典対照分類語彙表番号のアノテーションを行い、その過程で構築された「古典対照分類語彙表番号-UniDic 語彙素番号対応表」[5]および整備済の「分類語彙表番号-UniDic 語彙素番号対応表」[6]を用いて浅原ら(2022)[7]は、分類語彙表番号を自動で付与する工夫を行いCHJの全データに対して網羅的・体系的に『日本古典対照分類語彙表』の分類語彙表番号および『分類語彙表増補改訂版』[8]の分類語彙表番号を付与している。

形容詞の用法については、土岐(2017)[9]が平安文学作品の会話文における動詞、形容詞の用法(連

体法、準体法、終止法)について量的な分析を行っている。

ここで紹介した「分類語彙表番号」は統語分類である類を示す1桁の数字と意味分類である分類項目を示す4桁の数字をピリオドでつないだものである。ただし、『日本古典対照分類語彙表』の分類語彙表番号はピリオドのない5桁の数字となっている。

例えば、『日本古典対照分類語彙表』では、見出し語「あかし(明・赤)」に対する意味分類として、「33420(人柄)/35010(光)/35020(色)」が割り当てられている。先頭の「3」は形容詞・副詞・連体詞などの「相」を表し、次の1桁は部門と呼ばれ、「5」は「自然」を表すというように、階層的な構造になっている。

この分類語彙表番号は大まかな意味分類を捉えるという目的では重宝されるものだが、元となる分類語彙表の粒度に左右され、語義に識別にも限度があるため、今回は、分類語彙表番号を用いずに、独自の手法で語義分類、用法分類を試みたので、その分類手順と分類した語義・用法の応用例としていくつかの分析結果を報告する。

なお、本原稿は、じんもんこん2021およびじんもんこん2022で発表した内容[10]を基に新たな要素を追加したものである。

2 語義・用法情報の追加方法

CHJの語義・用法情報の追加は、池上、浅原らと同様、CHJのコーパス情報の「サンプルID」と「開始位置」をユニークなキーとしたアノテーションデータの形で行う。このアノテーションデータのイメージは表1のようなものになる。

表1 語義・用法アノテーションデータ

サンプルID	位置情報	見出し語	意味1	意味2	用法
20W拾遺1005_09009	28920	高い	(一)(2)盛り上がる・積もる	(四)(3)数値・程度	連体用法
30-建礼1232_00015	4590	憂い	(一)(2)つらい		終止法(係り結び)
20-源氏1010_00050	188780	安い	(1)安心・平穩		語幹-形容動詞生成

3 データの範囲

コーパス検索ツール「中納言」[11]を利用してCHJから品詞が形容詞のものを抜き出し（短単位および長単位）、その中から「奈良」「平安」「鎌倉」「室町」「江戸」「明治」「大正」のいずれの時代にも1例以上出現し、かつ、その形容詞が構成する複合語の数が異なりで5以上あるもの16語の中から「深い」「憂い」「恥ずかしい」「高い」「安い」「難しい」「痛い」「悲しい」「良い」「悪い」の10語を分類の対象とし（ただし、用法分類については前半の5語のみ）、CHJの用例としては、平安時代および鎌倉時代の用例を対象に、1語あたりの用例数は100とし、用例が100未満の場合は全数とした。

4 語義分類の手順

当初は『デジタル大辞泉』の分類に従って語義分類を行ったが、その後評価した結果、古典語の語義のカバー範囲に不安があることから、古典語の語義のカバー範囲が広い『日本国語大辞典』（ジャパンナレッジ、2022年7月～8月アクセス）の分類に従うこととした。『日本国語大辞典』にないものは、作業者が語義を付与した。また、小学館『日本古典文学全集』（ジャパンナレッジ）記載の現代語訳を参考にした。

この語義分類については、すべて手作業にて行った。形容詞全体を対象とするには膨大な数の用例を処理しなければならず、当然、自然言語処理による自動処理が望ましいが、現時点でのプロジェクトメンバー内に専門家がいなかったことや、自動化のための学習用データがまだ不足していることから、まずは試行的に対象を限定して手作業にて行うこととした。

5 用法分類の手順

用法については、小田[11]、土岐を参考にして詳細な分類基準を策定した。その策定にあたっては、当該形容詞の前後の語の品詞・活用形の情報参照することで、基準を明確にすることを試みた。なお、前後の語の形態論情報については、「中納言」の「インラインタグを使用」という機能を利用し、当該形容詞の前後に現れる語の品詞（中分類）・活用形（小分類）の情報を同時に抽出している。

活用形ごとの分類基準を付録Aに示す。

6 応用例

次に、CHJに追加された語義・用法情報の応用例を示す。まず、「中納言」を用いてCHJから形容詞のデータをダウンロードする。次に、「サンプルID」+「開始位置」をキーに語義・用法アノテーションデータと照合し、語義・用法情報を付与したコーパスデータを作成する。その後は、Excelの機能を用いて集計・分析を行う。以下に応用例を4つ挙げる。1つ目は語義分類の結果を平安時代と鎌倉時代と比較したもの、2つ目は、用法分類の結果を平安時代と鎌倉時代と比較したもの、3つ目は語義と用法の関連、4つ目はジャンルと用法の関連である。

6.1 語義分類の時代間比較

10語のうち、(1)「深い」「憂い」「高い」「難しい」は時代間の差がほとんどない、(2)「恥ずかしい」は平安時代には[気詰まり]や[気おくれ]が多用されるのに対して、鎌倉時代になると大幅に減少し[決まりが悪い]が主流となる、(3)「痛い」は平安時代では[精神的苦痛]が多いのに対して、鎌倉時代は[肉体的苦痛]がほとんどである、(4)「悲しい」は平安時代、鎌倉時代とも[嘆かわしい]が主だが、鎌倉時代では[興味深い]の意が新たに生じている、(5)「良い」は平安時代では[正当・善][適当・相応]が主であるのに対して、鎌倉時代には[優れる]が圧倒的に多くなる、(6)「悪い」は平安時代では[邪悪・不正][けしからぬ][機嫌が悪い]がほぼ均等に使われているのに対し、鎌倉時代では[邪悪・不正][けしからぬ]が中心となっていくことなどがわかった。

特徴的な例として「恥ずかしい」「痛い」「悲しい」のデータをそれぞれ表2～表4に示す。

表2 「恥ずかしい」の語義分類

意味	時代					平安					鎌倉				
	地の文	歌	会話	その他	計	地の文	歌	会話	その他	計	地の文	歌	会話	その他	計
(1) 決まりが悪い	16	0	4	0	20	21	7	0	0	28					
(2) 気詰まり	43	0	10	0	53	9	6	0	0	15					
(3) 気おくれ	22	0	4	0	26	6	0	0	0	6					
Xその他	0	1	0	0	1	1	0	0	0	1					
計	81	1	18	0	100	37	13	0	0	50					

表3 「痛い」の語義分類

意味	時代					平安					鎌倉				
	地の文	歌	会話	その他	計	地の文	歌	会話	その他	計	地の文	歌	会話	その他	計
(-) (1) 肉体的苦痛	7	0	8	0	15	12	1	15	0	28					
(-) (2) 精神的苦痛	20	0	20	0	40	0	0	2	0	2					
(-) (3) 甚だしい(連用形多)	1	7	0	2	10	0	2	0	0	2					
(-) (4) 優れている	9	0	2	0	11	0	0	0	0	0					
計	37	7	30	2	76	12	3	17	0	32					

表 4 「悲しい」の語義分

意味	時代					平安					鎌倉				
	地の文	歌	会話	その他	計	地の文	歌	会話	その他	計	地の文	歌	会話	その他	計
(1) 嘆かわしい	39	34	16	0	89	45	28	10	0	83					
(2) 愛おしい	3	0	1	0	4	2	0	1	0	3					
(3) 興味深い	0	0	0	0	0	11	0	1	0	12					
(4) 見事 (連用形多)	0	0	0	0	0	1	0	0	0	1					
Xその他	3	1	3	0	7	0	0	1	0	1					
計	45	35	20	0	100	59	28	13	0	100					

6.2 用法分類の時代間比較

用法分類については、「深い」「憂い」「恥ずかしい」「高い」「安い」の5語のみの分類だが、「深い」「憂い」「高い」については時代間の差はほとんどなく、「恥ずかしい」では、終止法と副詞法の比率が鎌倉時代に増えること、「安い」では、平安時代ではほとんどが「安からず」の形で用いられていたのに対し、鎌倉時代になると、連体法と副詞法が合わせて50%まで増えることがわかった。

特徴的な例として、「恥ずかしい」と「安い」のデータをそれぞれ表5と表6に示す。

なお、「5用法分類」で詳細な分類基準を策定したと書いたが、その後一部基準の見直しが必要となったため、これ以降の用法に関しては、代表的な用法の1.連体法、2.終止法、3.副詞法と4.その他に集約した形での分類結果を示すこととする。

表 5 「恥ずかしい」の用法分類

用法	平安	鎌倉
1.連体法	8	2
2.終止法	16	13
3.副詞法	18	16
その他	58	19
合計	100	50

表 6 「安い」の用法分類

用法	平安	鎌倉
1.連体法	3	25
2.終止法	1	1
3.副詞法	1	15
その他	95	39
合計	100	80

6.3 語義と用法の関連

「深い」「憂い」「恥ずかしい」「高い」「安い」の5語の語義と用法の関連については、(1)「深い」は、全体としては平安時代、鎌倉時代とも連体法と

副詞法で70%以上を占め、終止法がほとんどないが、個別の語義としては、平安時代には「空間的距離」や「関わり方」の意ではほとんどが連体法である、(2)「憂い」は平安時代、鎌倉時代とも60%以上が連体法、終止法が10%程度、副詞法がほとんどない、(3)「恥ずかしい」は、平安時代には連体法が8%、終止法が16%、副詞法が18%と3用法合わせて40%強であるのに対して、鎌倉時代にはそれぞれ4%、26%、32%と連体法が少なく、終止法と副詞法が増える、(4)「高い」は、全体としては平安時代、鎌倉時代とも連体法と副詞法で60%以上を占め、終止法がほとんどないが、個別の語義としては、平安時代には「身分・地位」の意ではほとんどが連体法であり、「声・音が大きい」の意ではほとんどが副詞法である、(5)「安い」は、連体法、終止法、副詞法の3用法を合わせても5%であったのに対し、鎌倉時代になると、終止法は変わらず、連体法と副詞法を合わせて50%を占めるようになる、などがわかった。

特徴的な例として、「深い」の平安時代および鎌倉時代の語義と用法の関連をそれぞれ表7と表8に、「恥ずかしい」の平安時代および鎌倉時代の語義と用法の関連をそれぞれ表9と表10に、「安い」の平安時代および鎌倉時代の語義と用法の関連をそれぞれ表11と表12に示す。

表 7 「深い」の平安時代の語義と用法の関連

平安	1.連体法	2.終止法	3.副詞法	4.その他	合計
(1)空間的距離	12			3	15
(2)感情・趣き	15	1	24	17	57
(3)関わり方	9		2	3	14
(4)色・香	4	1	1		6
(5)時間経過	3		1		4
Xその他	1		1	2	4
合計	44	2	29	25	100

表 8 「深い」の鎌倉時代の語義と用法の関連

鎌倉	1.連体法	2.終止法	3.副詞法	4.その他	合計
(1)空間的距離	18		9	2	29
(2)感情・趣き	10	2	18	11	41
(3)関わり方	4		8	5	17
(4)色・香	3	2	2	1	8
(5)時間経過	3		2		5
合計	38	4	39	19	100

表 9 「恥ずかしい」の平安時代の語義と用法の関連

平安	1.連体法	2.終止法	3.副詞法	4.その他	合計
(1)決まりが悪い	1	7	4	8	20
(2)気詰まり	2	9	10	32	53
(3)気おくれ	5		4	17	26
Xその他				1	1
合計	8	16	18	58	100

表 10 「恥ずかしい」の鎌倉時代の語義と用法の関連

鎌倉	1.連体法	2.終止法	3.副詞法	4.その他	合計
(1)決まりが悪い	1	7	11	9	28
(2)気詰まり		4	4	7	15
(3)気おくれ		2	1	3	6
Xその他	1				1
合計	2	13	16	19	50

表 11 「安い」の平安時代の語義と用法の関連

平安	1.連体法	2.終止法	3.副詞法	4.その他	合計
(1)安心・平穩	3	1		89	93
(3)自由・気楽				2	2
Xその他			1	4	5
合計	3	1	1	95	100

表 12 「安い」の鎌倉時代の語義と用法の関連

鎌倉	1.連体法	2.終止法	3.副詞法	4.その他	合計
(1)安心・平穩	6		12	30	48
(2)容易	19	1	3	7	30
(3)自由・気楽					
Xその他				2	2
合計	25	1	15	39	80

6.4 ジャンルと用法の関連

「深い」「憂い」「恥ずかしい」「高い」「安い」の5語のジャンルと用法の関連については、いずれの語も1つか2つのジャンルに多用されていて、それが全体の傾向に表れていることがわかった。

そこで、全語を合わせて全体としての傾向を見ると、平安時代は、歌集や歌物語では連体形が多く、作り物語や日記では連体法と副詞法が多いこと、鎌倉時代は歌集では連体形が多く、それ以外は連体法と副詞法が多いこと、また、時代・ジャンルを問わず終止法の比率が低いことなどがわかった。

以下、5語を合わせた全体での平安時代および鎌倉時代のジャンルと用法の関連をそれぞれ表13と表14に示す。

表 13 平安時代のジャンルと用法の関連

平安	1.連体法	2.終止法	3.副詞法	4.その他	合計
歌集	37	4	6	34	81
歌物語	10	3	1	5	19
作り物語	82	18	62	152	314
随筆	1	4	10	10	25
日記	11	3	8	18	40
歴史物語	2		4	15	21
合計	143	32	91	234	500

表 14 平安時代のジャンルと用法の関連

鎌倉	1.連体法	2.終止法	3.副詞法	4.その他	合計
歌集					
説話	1	13	11	15	40
随筆	1		3	3	7
日記			2	1	3
紀行					
合計	2	13	16	19	50

7 おわりに

これまで、形容詞10語の平安時代と鎌倉時代の用例に対して語義の分類を、形容詞5語の平安時代と鎌倉時代の用例に対して用法の分類を行ってきたが、本研究の目的としては、最終的に古代から現代に至る形容詞の通時的な語義の変遷や用法の変遷のプロセスを明らかにするところにあり、全体の構想からするとまだまだデータが不十分である。

今後は、自然言語処理による自動化も視野に入れて、さらに対象となる語や時代範囲を増やしていきたい。

また、用法分類については、助動詞や助詞が後続する場合を1つの用法としていたが、これについては、再度機能面からの見直しを行いたい。

謝辞

本研究は JSPS 科研費 JP21K00279, 20K13059 の助成を受けたものである。

参考文献

1. 国立国語研究所 (2021) 『日本語歴史コーパス』バージョン 2021.3 (引用日: 2023 年 1 月 10 日)
<https://clrd.ninjal.ac.jp/chj/>
2. 『日本語歴史コーパス 平安時代編』出現形容詞に対する古典分類語彙表番号アノテーション. 池上尚. : 言語処理学会 第 23 回年次大会 発表論文集, 2017.
3. 宮島達夫/鈴木泰/石井 久雄/安部清哉編. 日本古典対照分類語彙表. : 笠間書院, 2014.
4. 『日本語歴史コーパス』4 作品に対する分類語彙表番号付与とその分析. 浅原正幸, 加藤祥, 鈴木泰, 池上尚. : 日本語学会 2018 年度秋季大会プログラム予稿集, 2018.
5. 国立国語研究所(2020) 『古典対照分類語彙表分類番号—UniDic 語彙素番号対応表』(ver.0.8.0). (引用日: 2023 年 1 月 10 日.)
https://github.com/masayu-a/WLSP2UniDic_historical
6. 国立国語研究所(2020) 『分類語彙表番号—UniDic 語彙素番号対応表』(ver.1.0.2). (引用日: 2023 年 1 月 10 日.) <https://github.com/masayu-a/wlsp2unidic>
7. 分類語彙表番号を付与した『日本語歴史コーパス』データ. 浅原正幸, 池上尚, 鈴木泰, 市村太郎, 近藤明日子, 加藤祥, 山崎誠. : 日本語学会 2022 年度春季大会プログラム予稿集, 2022.
8. 国立国語研究所編. 国立国語研究所資料集 14 分類語彙表 増補改訂版. : 大日本図書, 2004
9. 平安和文会話文における連体法、準体法、終止法の比較分析. 土岐留美江. : 愛知教育大学研究報告. 人文・社会科学編 66, 2017.
10. 形容詞の通時的語義・用法データベースの作成, 山崎誠, 村田菜穂子, 前川武, 村山実和子. : じんもんこん 2021 論文集, 2021. 形容詞の通時的語義・用法データベースの構想と進捗状況. 山崎誠, 村田菜穂子, 前川武, 村山実和子. : じんもんこん 2022 論文集, 2022.
11. 国立国語研究所 コーパス検索アプリケーション「中納言」(バージョン 2.7.0, データバージョン 2022.10). (引用日: 2023 年 1 月 10 日.)
<https://chunagon.ninjal.ac.jp/>

A 付録 活用形ごとの分類基準

	用法	分類基準 (機能・後続語)
A. 連体形	a.連体用法	連体修飾語を形成する
	b.連体形終止法	そこで文を終止する
	c.終止法(係り結び)	係り助詞「ぞ」「なむ」「や」「か」と呼応して文を終止する
	d.準体用法	連体形をそのまま名詞句として用いる
	e.助動詞に接続	なり・たり (断定), ごとし (比況)
	f.助詞に接続	が (逆接確定条件<~けれど・~のに>, 単純な接続<~と・~ところ>) に・を (順接確定条件<~から・~ので>, 逆接確定条件<~けれど・~のに>, 単純な接続<~と・~ところ>) ものの・ものを・ものから・ものゆゑ (逆接確定条件<~けれど・~のに>)
B. 終止形	a.終止形終止法	そこで文を終止する
	b.終止形名詞法	終止形自体がそのような状態のものの意を表す名詞として用いられる
	c.連体法	終止形が直接名詞に続く
	d.助動詞に接続	らむ・べし・らし・めり (推量), なり(伝聞), まじ (打消推量)
	e. 助詞に接続	とも (逆接仮定条件<たとえ~ても>)
C. 連用形	a.連用修飾(副詞法)	後続の用言を修飾する
	b.中止法	そこで文を終止せずに, いったん止めて, 以下に文を続ける
	c.連用形名詞法	連用形自体が名詞として用いられる
	d.助動詞に接続	き・けり (過去), つ・ぬ・たり(完了), けむ (推量), たし (希望)
	e. 助詞に接続	て・して (単純な接続<~て>) つつ (動作の並行<~ながら>, 反復・継続<~しては・~し続けて>) ながら (動作の並行<~ながら>, 逆接確定条件<~けれど・~のに>)
D. 未然形	a.助動詞に接続	る・らる (受身・自発・可能・尊敬), す・さす・しむ (使役・尊敬), む・むず・まし (推量), ず (打消), じ (打消推量), まほし (希望)
	b.助詞に接続	は (ば) (順接仮定条件<もし~ならば>, (順接確定条件 (原因・理由<~から・~ので>, 偶然条件<~と・~ところ>, 恒常条件<~といつも>))
E. 已然形	a.終止法(係り結び)	係り助詞「こそ」と呼応して文を終止する
	b.助詞に接続	は (ば) (順接仮定条件<もし~ならば>, (順接確定条件 (原因・理由<~から・~ので>, 偶然条件<~と・~ところ>, 恒常条件<~といつも>))
F. 語幹	a.名詞修飾	直接名詞を修飾する
	b.動詞修飾	直接動詞を修飾する
	c.感動	「あな+語幹」, 「あな+語幹+や」, 単独で感動の意を表す
	d.名詞	語幹自体が名詞として用いられる
	e.名詞を作る	接尾辞「さ」を伴って, その状態・程度を表す名詞を作る 接尾辞「み」を伴って, 「その状態の所」の意を表す名詞を作る
	f.連体法	語幹がそのまま連体法に用いられる
	g.形容動詞を作る	接尾辞「~げ」を伴って, 「様子や気配」という意味を付加した形容動詞を作る
	h.助詞に接続	に (逆接確定条件<~けれど・~のに>)