

形容詞の通時的語義・用法データベースの作成

山崎 誠 (国立国語研究所)

村田 菜穂子 (大阪国際大学)

前川 武 (大阪国際大学短期大学部)

村山 実和子 (福岡女子短期大学)

概要：本研究では、『日本語歴史コーパス』(CHJ)を利用して、形容詞の通時的語義・用法データベースシステムの構築について概要を紹介するとともに、データベースを作成するにあたって、どのような問題点があるか、試行的に行った語義分類・用法分類の結果を報告する。試行では、「深い」「憂い」「恥ずかしい」「高い」「安い」の5語について、平安時代の用例を『大辞林』の語義分類に基づいて語義を分類し、出現する文脈によって語義に偏りがあることや特定の活用形に偏る語があることなどを示した。データベース構築上の課題としては、掛詞の取り扱いや語義分類の見直しなどについて指摘した。

キーワード：形容詞、語義、用法、日本語歴史コーパス、データベース、言語変化

Creating a diachronic database of adjective senses and usages

Makoto Yamazaki (National Institute for Japanese Language and Linguistics)

Nahoko Murata (Osaka International University)

Takeshi Maekawa (Osaka International College)

Miwako Murayama (Fukuoka Women's Junior College)

Abstract: This paper outlines the construction of a diachronic database system for adjective senses and usages using the Corpus of Historical Japanese (CHJ), and reports on the results of trial word sense classification and usage classification to see what problems exist in creating the database. In the trial, we classified the meanings of five words, “fukai”(deep), “ui”(sorrow), “hazukashii” (embarrassing), “takai” (high, expensive), and “yasui” (cheap) based on the word sense classification of “Daijirin” in the Heian period, and showed that there were biases in the meanings of words depending on the context in which they appeared, and that some words were biased toward specific conjugations. As issues in the construction of the database, we pointed out the handling of *kakekotoba* (pun in poetry) and the review of word sense classification.

Keywords: Adjective, Meaning, Usage, Corpus of Historical Japanese, Database, Language Change

1. はじめに

筆者らはこれまで主に語彙索引を利用して、中古・中世を中心に語彙史の観点から形容詞・形容動詞の量的な研究を行ってきた([1][2][3])。索引というデータの制約上、語義や用法など文脈に基づく分析は多大な時間がかかるため実現できていなかった。しかし、『日本語歴史コーパス』(Corpus of Historical Japanese, 以下CHJと略す) [4]の利用により、語義や用法による分析も比較的容易に行えるようになってきた。そこで、筆者らは、最終的にはCHJをもとにした、形容詞の語義・用法データベースシステムを構築・公開することによって日本語学研究、特に、語彙研究および語彙史研究の分野に有益な資料を提供することを計画した。

本研究では、通時的語義・用法データベースを作成するにあたって、どのような問題点があるか、

試行的に行った語義分類・用法分類の結果を報告する。

2. 先行研究

形容詞の語義に関する量的な研究として、池上[5]がある。これは、嗅覚を表す形容詞「カグハシ」「カウバシ」「カンバシ」の意味的変遷と交替を量的に分析したものである。安本[6][7]は、古典語の感情形容詞の意味分類を見直したものである。形容詞の用法については、土岐[8]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。現代語では、橋本・青山[9]、宮島[10]がある。これらは、現代語の形容詞を対象に3つの用法(終止、連体、連用)を量的に分析したものである。また、語義と用法の関係では、山崎[11]がある。これは、現代語のいくつかの動詞および形容詞の語義と活用形の関係进行分析し、特徴的な関係を持つ例を紹介したものである。語義を量的

に扱った文献では、比較的少数の語を扱っているものが多い。なお、宮島他[12]は古典作品を対象としたシソーラスであり、古典の意味的な分析に有用であるが、見出し語を単位にして代表的な語義を付したもので、多義語については対応しておらず、今回の目的には利用できない。

3. データ

コーパス検索ツール「中納言」[13]を利用してCHJ から品詞が形容詞であるものを抜き出した（短単位および長単位）。CHJ 全体の形容詞の量的情報を表 1 に示す。

表 1 CHJ の形容詞

	延べ語数	異なり語数
短単位	316814	1330
長単位	72470	1451

長単位の語数が短単位に比べてそれほど多くないのは、短単位のデータが「奈良」から「昭和」までの全時代に渡って整備されているのに対して、長単位のデータは「江戸」「明治」「大正」「昭和」のデータがまだ中納言に実装されていないためである。

今回の試行的作業にあたって、以下の 2 つの条件を満たす語を抜き出した。

- (1) 現代まで使用されている語
- (2) 複合語を作る力の強い一次形容詞

具体的には、条件 (1) については、「奈良」「平安」「鎌倉」「室町」「江戸」「明治」「大正」のいずれの時代においても 1 例以上出現しているもの、条件 (2) については、その形容詞が構成する複合語の数が異なりで 5 以上あるものとした。この条件を満たす語は 16 語であった。表 1 にその全体を示す。表 1 の「複合語数」は、短単位を後部要素に持つ長単位の異なり数である。

表 1 抽出した形容詞とその時代別用例数（短単位）

語彙素読み(語彙素)	1 奈良	2 平安	3 鎌倉	4 室町	5 江戸	6 明治	7 大正	8 昭和	合計	複合語数
カタイ (難い)	10	218	62	7	12	827	85		1221	238
ナイ (無い)	490	5124	4128	1590	5188	42762	25441	1187	85910	101
ヤスイ (易い)	4	45	38	76	21	151	16		351	34
ヨイ (良い)	90	740	429	904	2485	7036	6189	762	18635	28
フカイ (深い)	15	682	511	105	270	2530	1408	147	5668	17
タカイ (高い)	61	264	329	45	162	2483	1341	357	5042	12
クルシイ (苦しい)	59	550	101	118	57	369	298	22	1574	11
チカイ (近い)	41	589	340	55	121	1234	662	77	3119	11
トオイ (遠い)	138	246	186	30	67	1304	494	178	2643	9
アシイ (悪い)	26	294	255	47	63	359	78	2	1124	8
ウイ (憂い)	8	621	195	12	99	91	8	1	1035	8
イタイ (痛い)	37	76	32	41	72	378	137	39	812	6
カナシイ (悲しい)	93	792	452	81	288	511	290	34	2541	6
ヤスイ (安い)	33	127	80	16	55	336	239	12	898	6
ナガイ (長い)	109	270	279	49	177	2269	1550	281	4984	5
ハズカシイ (恥ずかしい)	6	415	50	52	174	260	144	9	1110	5
クサイ (臭い)		5	21	2	8	70	62		168	14
コワイ (強い)		12	5	10	6	2	2		37	5
サワガシイ (騒がしい)		93	15	2	17	51	28	10	216	5
ムズカシイ (難しい)		132	65	26	86	438	255	34	1036	5

※語彙素読み(語彙素)がゴチックになっているのは、今回の試行で対象とした語

今回の調査では、この中から「憂い」「高い」「深い」「恥ずかしい」「安い」の 5 語を選んで平安時代の用例（サブコーパス名で「平安-仮名文学」と「和歌集」に相当するもの）について語義分類を行った。これらの形容詞は、属性形容詞(高

い、深い、安い)、感情形容詞(憂い、恥ずかしい)に分けられ、ある程度の意味的な均衡を考慮したものである。なお、作業上の問題点を洗い出す試行であるため、作業対象とする用例数をそれぞれランダムに抽出した 100 例とした。

4. 語義分類と用法分類の手順

語義分類は、原則として『デジタル大辞泉』(ジャパンナレッジ, 2021年7月~8月アクセス, 以下大辞泉と略す) [14]の分類に従った。例えば、「深い」の語義は大ブランチで6個あり、そのうち1つのブランチは2つの小ブランチを持ち、合計7つのブランチとなる。この小ブランチまでを分類する。また、『大辞泉』に語義がないものは、「X その他」として、作業者が語義を付与した。

今回対象になったデータの多くは、小学館『日本古典文学全集』[15]に収録されており、ジャパンナレッジで閲覧することができるため、そこに記載されている現代語訳を参考にしている。

用法分類は、コーパスの検索結果である「活用形」を利用して行った。まず、活用形による傾向を概観したのち、「終止」「連体」「連用」「準体」の4つの用法について観察した。

5. 結果1・語義の分布

語義分類の結果を表3, 5, 6, 7, 8に示す。CHJには「歌」「詞書」「会話」などの「本文種別」の情報が付いているため、その情報も併せて示した。「本文種別」の情報と合わせて見ることで語義の分布の特徴を把握することができると考えたからである。なお、本文種別の空欄のものは「地の文」として扱った。なお、計が100を超えているのは、掛詞等で2つの意味をそれぞれカウントしたためである。以下、語別に傾向を見る。

なお、掛詞などにより複数の意味が生じた場合は、それぞれをカウントしている。「中納言」の検索結果には、当該の語が掛詞で使われているかどうかを示されるが、以下の例のような場合は、通常の掛詞とは異なり、語としては同じ語であるが異なる語義で使われている場合があり、そのような場合も別々の語義として数えた。

「棹させど底ひも知らぬわたつみの深き心を君に見るかな」(20-土佐 0934_00001,12030)

この例では、「わたつみの深き」で「1 距離」の意味が使われ、「深き心」で、「2 程度・分量」の意味が使われている。

(1) 「深い」

「深い」は、約7割弱が「2 程度・分量」として使われている。これは『大辞林』の分類の粒度がやや粗いことも関係していると考えられる。

『大辞林』のこの語義は『日本語国語大辞典』[16]においては、2つのブランチに対応し、思想、感情、趣き、関わり、縁、罪、学問、経験などさまざまな対象について用いられる。ただ、地の文と歌では語義の分布が違っている。地の文では、約77%が「2 程度・分量」なのに対して、歌では、約33%である。その代わり「1 距離」が約37%

と多くなっている。また、会話では地の文よりも「2 程度・分量」への偏りが大きい。

表3 「深い」の語義

語義	地の文	歌	会話*	その他	計
1 距離	3	10	4	0	17
2 程度・分量	37	9	26	0	72
3 色合い	2	5	0	0	7
4 密度	2	1	1	0	4
5 時間	2	2	0	0	4
6(ア)接辞: 距離	0	0	0	0	0
6(イ)接辞: 程度	0	0	0	0	0
X その他	2	0	0	0	2
計	48	27	31	0	106

*「会話」には「会話-発話引用」を含む

掛詞における2つの語義の関係は、表4のとおりである。「2 程度・分量」とそれ以外の意味を結び受けるパターンが多いことが分かる。

表4 「深い」の掛詞における語義

語義1	語義2
3 色合い	2 程度・分量
1 距離	2 程度・分量
2 程度・分量	4 密度
2 程度・分量	1 距離
2 程度・分量	3 色合い
3 色合い	2 程度・分量

(2) 「憂い」

「憂い」は、約77%が「1 つらい」の意味で用いられている。この傾向は地の文、歌、会話の順に大きくなる。また、全体の約7割が歌で用いられていることも「憂い」の特徴である。

「X その他」の割合が約15%を占める。「X その他」となっている語義は、「情けない(4)、恨めしい(3)、憂鬱(2)、憎い(1)、いやな(1)、不安(1)、うとましい(1)、浮いている(1)、牛(1)」である(括弧内はそれぞれの延べ例数)。語義は『日本古典文学全集』などを参考にしたが、そこで意識されているようなものがここに出ていると推測される。

表5 「憂い」の語義

語義	地の文	歌	会話	その他	計
1 つらい	12	55	12	0	79
2 煩わしい	1	4	0	0	5
3 つれない	0	0	0	0	0
4 切ない	0	2	0	0	2
5 接尾辞	0	1	0	0	1
X その他	5	9	2	0	16
計	18	71	14	0	103

(3) 「恥ずかしい」

「恥ずかしい」は半分強が「2 気詰まり」であり、残りを「1 欠点」「3 気おくれ」が占める。3つの語義がある程度バランスを取って用いられている。また、歌の例がほとんどなく、会話の例が2割を占める。

表6 「恥ずかしい」の語義

語義	地の文	歌	会話	その他	計
1 欠点	16	0	4	0	20
2 気詰まり	43	0	10	0	53
3 気おくれ	22	0	4	0	26
X その他	0	1	0		1
計	81	1	18	0	100

(4) 「高い」

「高い」は半分強が「1(ア)位置」と「1(イ)垂直方向」の物理的な意味で使われている。抽象的な意味の中では「3(ア)程度」が多い。

現在では、「4(イ)金額」の意味も使われるが、今回の調査範囲には出現しなかった。これは、そのような状況を表す場面がなかったからではなく、『日本国語大辞典』の「高い」の語誌欄に「高貴、高価を表わすのは比喩的用法で、前者は平安時代以降、後者は室町時代以降に成立したと考えられる。」とあるように、平安時代にはこの意味がなかったからと思われる。

文体種別では、「会話」では約半数が「3(ア)程度」であることが特徴的である。

表7 「高い」の語義

語義	地の文	歌	会話*	その他	計
1(ア)位置	13	3	2	0	18
1(イ)垂直方向	20	15	3	1	39
2(ア)音・声	7	3	3	0	13
2(イ)評判	2	2	0	0	4
3(ア)程度	15	2	9	1	27
3(イ)高邁	0	0	0	0	0
4(ア)数値	0	2	0	0	2
4(イ)金額	0	0	0	0	0
5 高慢	0	0	0	0	0
X その他	0	0	1	0	1
計	57	27	18	2	104

*「会話」には「会話-発話引用」を含む。「その他」は詞書。

(5) 「安い」

「安い」は、意味のほとんどが「3 安心」に集中している。また、会話での例が約4割と多く、歌の例が少ないことが特徴的である。「1 値段」の意味が現れないのは、「高い」と同じように、その意味が平安時代には成立していなかった、と

いう訳ではない。『日本国語大辞典』によると、廉価の意味が10世紀初頭や12世紀初めの資料に見られるからである(ただし、2つとも訓点資料)。

表8 「安い」の語義

意味	地の文	歌	会話	その他	計
1 値段	0	0	0	0	0
2 価値	0	0	0	0	0
3 安心	51	5	36	1	93
4 親密	0	0	0	0	0
5 気楽	1	0	1	0	2
X その他	3	0	2	0	5
計	55	5	39	1	100

*その他は詞書。

6. 結果2・活用形の分布

表9~12に各形容詞の活用形の分布を示す。活用形の集計は、語義の場合の集計とは違って、CHJにおいて時代が「平安」に該当する全例による集計である。また、活用形の名称は、中納言の検索により得られたものをそのまま使っている。

表9 「深い」の活用形

活用形	頻度	割合
語幹-一般	49	7.18
終止形-一般	12	1.76
未然形-補助	22	3.23
連体形-一般	258	37.83
連体形-補助	2	0.29
連体形-撥音便	4	0.59
連用形-ウ音便	67	9.82
連用形-一般	240	35.19
連用形-補助	17	2.49
已然形-一般	11	1.61
計	682	100.00

表10 「憂い」の活用形

活用形	頻度	割合
ク語法	1	0.16
語幹-一般	14	2.25
終止形-一般	78	12.56
未然形-補助	5	0.81
連体形-一般	447	71.98
連体形-補助	6	0.97
連用形-一般	30	4.83
連用形-補助	25	4.03
已然形-一般	15	2.42
計	621	100.00

表11 「恥ずかしい」の活用形

活用形	頻度	割合
語幹-一般	108	26.02
終止形-一般	53	12.77
未然形-補助	2	0.48
連体形-一般	49	11.81
連体形-補助	3	0.72
連用形-ウ音便	70	16.87
連用形-一般	86	20.72
連用形-補助	4	0.96
已然形-一般	40	9.64
計	415	100.00

表12 「高い」の活用形

活用形	頻度	割合
語幹-一般	27	10.23
終止形-一般	8	3.03
未然形-補助	2	0.76
連体形-一般	91	34.47
連体形-補助	1	0.38
連体形-撥音便	1	0.38
連用形-ウ音便	32	12.12
連用形-一般	93	35.23
連用形-補助	1	0.38
已然形-一般	8	3.03
計	264	100.00

表13 「安い」の活用形

活用形	頻度	割合
語幹-一般	12	9.45
終止形-一般	1	0.79
未然形-補助	107	84.25
連体形-一般	3	2.36
連体形-補助	1	0.79
連用形-一般	2	1.57
已然形-一般	1	0.79
計	127	100.00

これらの表から、以下の傾向が見て取れる

- (1) 「深い」は、「連体形-一般」と「連用形-一般」がほぼ同じで約35%ずつを占め、「終止形-一般」が少ない。
- (2) 「憂い」は「連体形-一般」に集中している(約70%)。
- (3) 「恥ずかしい」は、「語幹-一般」「連用形-一般」が2割を超え、「終止形-一般」「連用形-ウ音便」「連体形-一般」も10%台で出現しており、大きな集中が見られない。

(4) 「高い」は、「深い」と同じ傾向で、「連体形-一般」と「連用形-一般」がほぼ同じで約35%ずつを占め、「終止形-一般」が少ない。

(5) 「安い」は、「未然形-補助」に約85%が集中しているのが特徴的である。具体的には「安からず」の形で副詞的な使われ方である。

7. 結果3・用法の分布

用法は、「連体」「連用」「終止」「準体」「その他」に分類した。「連体」は名詞を修飾する用法、「連用」は述語に係る用法である。「終止」は文を終止させる用法で、ここでは、係り結びの場合を含めている。「準体」は当該の語の後に「こと」「もの」「ひと」などの名詞が省略されると解釈されるような用法であり、文語文で用いられるものである。

今回調査した5つの形容詞において、用法の分布を図1に示した。対象は、活用形の場合と同じで、CHJにおいて時代が「平安」に該当する全例である。「その他」は、上に挙げたいずれの用法にも含まれないものである。活用形でいうと、「語幹」「未然形」「已然形」であるが、形容詞のかり活用の場合は、活用形にかかわらず「その他」とした。なお、図中の数字は用例数である。

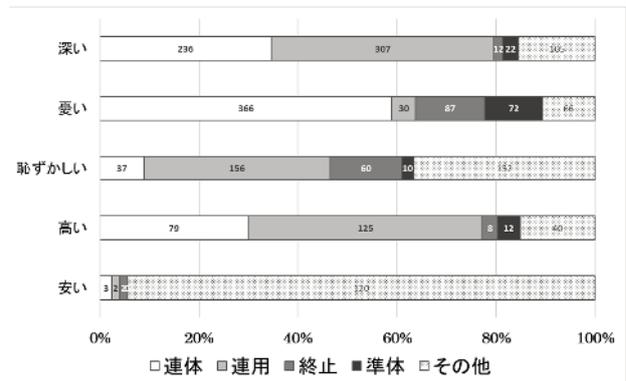


図1 各用法の分布

図1からは、「深い」と「高い」が連用法と連体法とを合わせたものが約8割であり、それぞれの割合も似た分布を示していること、「憂い」は約6割が連体法であり、連用法が少ないこと、「恥ずかしい」と「安い」は連体法が少なく、その他が多いことなどが分かる。

また、語義と用法を組み合わせる場合に傾向が見られる場合がある。図2は、「高い」における語義と用法(「その他」を除く)との関係を示したものである。全体が100例余りであり、それほど多くないため、参考程度であるが、連体法は約半数が「3(ア)程度」の意味であること、連用法では5つの語義がすべて現れていることが分かる。

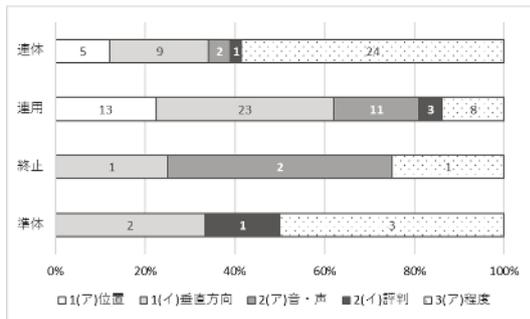


図2 「高い」の語義と用法

8. 試行から分かったこと

今回の試行の結果、形容詞の語義・用法データベースの作成に当たって、以下のような点が明らかになった。これらは今後の構築作業に活かす予定である。

(1) 語義は文体やジャンルによって語義が偏る可能性があるため、それらの情報も情報として取り込む必要があること。とくに、CHJの「本文種別」の情報が重要である。

(2) 掛詞のように意味が複数ある場合があるため、複数の意味を登録できる仕様にすること。

(3) 語義分類の見直し。今回は『大辞林』を用いたが、意味分類の粒度が細かいところと粗いところが見受けられた。適宜、語義の分割や合併を検討する。また、「X その他」として分類された語義については、文脈に添った意識的な意味をどこまで認めるかが課題である。

(4) 全体的な用法の見直し。今回は係り結びを終止用法に含めたが、これは独立させたほうがよいかもしれない。また、カリ活用の扱いも再考の余地がある、など時代的な変遷を捉えようとする場合に特徴的なものを取り出せるような用法分類が望ましいのではないかと。

9. 今後の予定

今回の調査は、平安時代の例だけであり、しかも、5つの形容詞でそれぞれ100例ずつというごく少数の観察であった。

今後、形容詞の数を増やす、作業が終わった5語について情報を付与する用例数を増やす、別の時代の用例を順次対象としていく、などいくつかの方向で増補の展開を考えている。

構築するデータベースは、古典語学習者や研究者が目的に応じて必要な統計データが取り出せるようなシステムを念頭に置いている。今後、今回の試行作業で明らかになった点を検討した上でデータベースの設計に取りかかりたい。なお、このデータベースは、完成後WEB上で公開する。

謝辞

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

参考文献

[1] 村田菜穂子:形容詞・形容動詞の語彙論的研究, 和泉書院(2005).

[2] 村田菜穂子, 前川武:語構成から見た形容詞—中古から中世への変遷—,国語語彙史の研究,Vol.38,pp.139-162,和泉書院(2019).

[3] 村山実和子:「ワル(悪)+形容詞」の消長—形容詞語形成の観点から—,筑紫語学論叢Ⅲ—日本語の構造と変化—,pp.179-198,風間書房(2021).

[4] 国立国語研究所(2021)『日本語歴史コーパス』データバージョン2021.3 <https://ccd.ninjal.ac.jp/chj/>

[5] 池上尚:嗅覚表現形容詞「カグハシ」「カウバシ」「カンバシ」—近世以降における意味・用法の分担過程,国文学研究,162,pp.64-53(2010).

[6] 安本真弓:感情形容詞の意味分類—『日本古典対照分類語彙表』を基盤として—,国語学研究,57,pp.147-133(2018).

[7] 安本真弓:奈良・平安・鎌倉時代の和文作品における感情形容詞の意味分類とシソーラス,国語学研究,59,pp.525-509(2020).

[8] 土岐留美江: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,愛知教育大学研究報告.人文・社会科学編,66,pp.21-29(2017).

[9] 橋本美奈子,青山文啓:形容詞の三つの用法:終止,連体,連用:計量言語学,18巻,5号,pp.201-214(1992).

[10] 宮島達夫:形容詞の語形と用法,計量国語学,Vol.19,No.2,pp.94-104(1993).

[11] 山崎誠:多義語を構成する意味の使用傾向—一品詞と活用形による違い—,言語処理学会第9回年次大会発表論文集,pp.(2011).

[12] 宮島達夫,鈴木泰,石井久雄,安部清哉(共編):日本古典対照分類語彙表,笠間書院(2014)

[13] 国立国語研究所 <https://chunagon.ninjal.ac.jp/> (検索日2021年7月24日)

[14] 小学館:デジタル大辞泉, ジャパンナレッジ <https://japanknowledge.com/> にて利用.

[15] 小学館:日本古典文学全集, ジャパンナレッジ <https://japanknowledge.com/> にて利用.

[16] 小学館:日本国語大辞典, ジャパンナレッジ <https://japanknowledge.com/> にて利用.