

## 参考文献

- 石黒圭 (2008) 『文章は接続詞で決まる』 光文社新書  
グループ・ジャマシイ編 (1998) 『日本語文型辞典』 くろしお出版  
仁田義雄 (1991) 『日本語のモダリティと人称』 ひつじ書房  
益岡隆志 (2008) 「叙述類型論に向けて」 『叙述類型論』 くろしお出版

## 辞書

- 『広辞苑 (第六版)』 (2008) 岩波書店  
『大辞林 (第三版)』 (2006) 三省堂  
『日本国語大辞典 (第二版、第八巻)』 (2001) 小学館  
『明鏡国語辞典 (初版)』 (2003) 大修館書店

## コーパス調査による形容詞の連体形と連用形

李在鎬(情報通信研究機構)  
横森大輔(京都大学大学院)  
土屋智行(京都大学大学院)

### 1. はじめに(背景と目的)

形容詞の活用は、動詞の活用とともに古くから議論されており、数多くの研究が存在する(例えば、橋本(1934)、三上(1955)、川端(1978,1979)など)。しかし、先行研究の多くは活用の体系を記述的に捉えることを中心としており、実際の用法から網羅的に形容詞の分布を明らかにするというものではなかった。加えて、活用形による意味の違いに関してはほとんど議論されていない。

こうした現状を踏まえ、本研究では、コーパス調査の観点から活用における形容詞の活用に注目し、その分布の実態を明らかにする。調査の焦点は連体形と連用形である。これは終止形に比べて、いずれも何らかの後続要素を伴う点で共通しているものの、構文的・意味的振舞いはかなり異なっており、言語記述的に興味深い。例えば以下に挙げる(1)(2)に注目してほしい。

- (1) a. 痛い注射  
b. 痛く感動する
- (2) a. 恐ろしい体験  
b. 恐ろしくまずい

形容詞「痛い」や「恐ろしい」に関して、bの連用形の場合は程度の高さを表しており、形容詞の表す意味がaの連体形とは大きく異なっている。このような相違が見られることは一部の研究で指摘されてきたが、本格的に考察されているとは言い難い。例えば、三上(1955)はこのような現象を指摘することにより、伝統文法において前提とされている形容詞の連体形と連用形を共通の語彙素からの活用形とみなすことに疑問を呈している<sup>1</sup>。以上を踏まえ、本研究では実際の使用例に対する大規模な調査を行うことで、日本語学および日本語教育分野における基礎資料を提供することを目的とする。

調査においては、新聞記事データとして「読売新聞の記事」を、小説データとして「新潮文庫」を、新聞と小説の中間に位置するデータとして「新潮新書」を使用した。これらのコーパスを「茶筌」で形態素解析し、延べ28,700例のサンプルを収集した。さらにこれ

<sup>1</sup> 関連研究として李(他)(2007)および橋本・青山(1992)が挙げられる。李(他)(2007)では動詞の語形によって異なった意味グループの名詞が共起している事実を明らかにしている。また、橋本・青山(1992)は実例に基づき、形容詞の終止用法、連体用法、連用用法の三つの分布関係を調査している。

らを連体形と連用形で集計し、出現頻度の差異係数を求め、カイ二乗検定で統計的有意を確認した。結果として、連体形では用いられるが、連用形では用いられない形容詞(好ましい、苦い、望ましい)の存在が確認される一方、連用形では用いられるが、連体形では用いられない形容詞(おかしい、手早い)が存在することを確認した。

## 2. 使用データと調査の概要

本研究では、実際の使用例から形容詞の活用形を網羅的に抽出し、分布実態を明らかにする。調査対象として以下の三つのテキストコーパスを使用した。

1. 新聞記事として Utiyama&Isahara (2003) が構築した「日英新聞記事対応付けデータ<sup>2</sup>」の日本語の部分(以下、読売新聞)
2. 小説として「新潮文庫の絶版 100 冊」をテキスト化したもの(以下、新潮文庫)
3. 新聞と小説の中間に位置する書籍データとして NICT 所有の「新潮新書 100 冊」(以下、新潮新書)

本調査の実行に指し当たって、以下の2点が問題になる。

- I. データ収集に関する問題：当該テキストに出現する形容詞の一覧があらかじめ存在するわけではないため、文字列をベースに収集することは難しい。
- II. 結果の信頼性に関する問題：分布の実態を明らかにするとは言っても、コーパスの規模が異なるため、出現頻度に対する意味づけが難しい。

以上の問題を解決すべく、I に対しては形態素解析とパターン検索によるデータ収集を行った。II に対しては相関分析を行うことで、コーパス間の分布の偏り具合を検証した上で、差異係数やカイ二乗値を用いて最終的な判断を行った。なお、I の処理は KH Coder を利用し、行った。II の処理は Excel を用いて行った。

## 3. 結果と考察

### 3.1. 全体の分布

まず、調査の対象のコーパスを形態素解析した。その結果を表1に示す。

<sup>2</sup> 本データは、読売新聞と The Daily Yomiuri から自動作成された日英対応付けコーパスで、誰でも利用可能である。なお、本研究では正規表現で日本語の部分のみを取り出し、データとして利用した。

表.1 コーパス規模

コーパス	文字数	延べ語数	異なり語数
読売新聞	7,497,353 字	4,606,346 語	52,557 語
新潮文庫	12,059,478 字	4,621,261 語	61,462 語
新潮新書	4,518,753 字	1,847,806 語	48,908 語
合計	24,075,584 字	11,075,413 語	

表1は、上述の三つのコーパスを形態素解析した結果である。これらのコーパスに含まれる全形容詞の連体形と連用形を収集した。収集に際しては、連体形は形容詞の基本形に名詞が後続する全用例を抽出した。一方の連用形は形容詞に動詞および助動詞が後続する全用例を抽出した。データ抽出後に、目視による確認作業を行った。最終的には、異なり語数で548語の形容詞を収集した。延べ頻度を表2に示す。

表.2 収集データの延べ頻度

	異なり語数	延べ語数		総計
		連体形	連用形	
読売新聞	223	4133(45.7%)	4909(54.3%)	9042
新潮文庫	509	5305(36.1%)	9376(63.9%)	14681
新潮新書	293	2128(42.8%)	2849(57.2%)	4977
総計		11566(40.3%)	17134(59.7%)	28700

表1と表2を突合せながら、全体の分布を検討した。

1. 読売新聞の場合、規模こそ大きいものの、形容詞の使用頻度は相対的に低い。
2. 新潮新書の場合、延べ語数では読売新聞の半分程度であるが、異なり語数では読売新聞より多くの形容詞が用いられ、多様な形容詞が用いられている。
3. 新潮文庫の場合、コーパスの規模は読売新聞とほぼ同サイズであるが、異なり語数と延べ語数のいずれにおいても圧倒的に多く形容詞が用いられている。
4. コーパスの相違に関係なく、連体形より連用形のほうが多く用いられている。

また、全てのコーパスと活用形における、語彙ごとの度数分布は以下の通りであった。

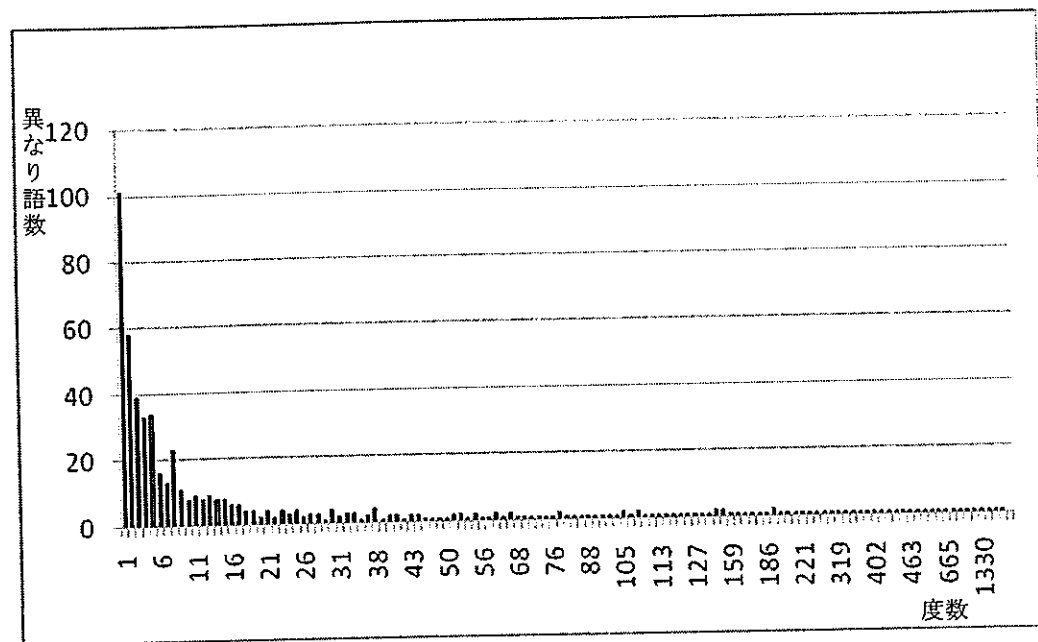


図1 全形容詞の度数分布図

図1より、語彙によって頻度に極端な差があることがわかる。これを踏まえ、分析の際には、より一般性のある結果を出すため、全コーパスでの絶対頻度が「5」以上のものを対象に行った。その結果、延べ頻度は「28,700例」から「28,234例」になり、異なり頻度は「548語」から「316語」になった。

次に、この316語の形容詞に各コーパス内での頻度と全コーパスでの頻度を計算した。これを元に、Pearson 相関係数を求め、コーパス間の分布の類似を調べた。

表3 相関分析

	新潮 新書 連体	新潮 新書 連用	新潮 新書 全体	新潮 文庫 連体	新潮 文庫 連用	新潮 文庫 全体	読売 新聞 連体	読売 新聞 連用
新潮新書連用	0.339							
新潮新書全体	0.832	0.804						
新潮文庫連体	0.684	0.459	0.703					
新潮文庫連用	0.305	0.742	0.63	0.553				
新潮文庫全体	0.499	0.717	0.738	0.806	0.939			
読売新聞連体	0.86	0.425	0.794	0.548	0.356	0.479		
読売新聞連用	0.235	0.864	0.657	0.288	0.672	0.597	0.476	
読売新聞全体	0.605	0.771	0.837	0.473	0.614	0.631	0.832	0.884

表3の結果から、以下の3点が言える。

1. 類似点に注目した場合、読売新聞と新潮新書がもっとも良く似ており、新潮新書と新潮文庫もある程度は似たような分布をしている。一方の新潮文庫と読売新聞は相対的に類似度が低い。
2. 非類似度の点に注目した場合、あるコーパスで連体形としてよく用いられる形容詞が別のコーパスで連用形として使用されるようなずれは見られない。例えば、新潮文庫で連用形として使用される形容詞が読売新聞で連体形として使用されることはあまりない。
3. コーパスの相違に関係なく、連用形と連体形をめぐる形容詞の分布はある程度一貫している。

以上は、本研究の結果が別のコーパスを使ったとしても大きくぶれるものではないことを示唆している。本研究が使用したコーパスは比較的小規模ではある点で十分ではないが、ある程度データの信頼性が保証されているものと考えられる。

### 3.2. 個別の分布

書きことばコーパスにおいて、もっともよく用いられた最頻50語を表4に示す。

表4 形容詞の最頻50語

基本形	新書 連体	新書 連用	文庫 連体	文庫 連用	新聞 連体	新聞 連用	連用形 の合計	連体形 の合計	総計	差異 係数	CHI TEST	有意 水準	グル ープ
大きい	20	293	86	401	43	791	1485	149	1634	-0.818	1.50E-239	p<.001	連用
新しい	465	43	265	93	701	25	161	1431	1592	0.798	2.50E-222	p<.001	連体
強い	81	146	94	268	399	576	990	574	1564	-0.266	7.06E-26	p<.001	連用
早い	9	129	11	755	38	388	1272	58	1330	-0.913	5.70E-243	p<.001	連用
高い	101	103	119	306	366	212	621	586	1207	-0.029	0.313729		両方
長い	90	111	204	249	77	95	455	371	826	-0.102	0.00347	p<.01	連用
深い	52	125	135	292	51	137	554	238	792	-0.399	2.95E-29	p<.001	連用
厳しい	27	23	11	25	339	303	351	377	728	0.036	0.335234		両方
多い	26	183	40	156	78	182	521	144	665	-0.567	2.11E-48	p<.001	連用
若い	129	8	300	46	121	5	59	550	609	0.806	4.38E-88	p<.001	連体
白い	25	23	287	228	33	3	254	345	599	0.152	0.000201	p<.001	連体
広い	47	118	106	47	67	164	329	220	549	-0.199	3.29E-06	p<.001	連用
近い	54	13	88	83	129	123	219	271	490	0.106	0.018818	p<.05	連体
悪い	31	37	85	227	39	44	308	155	463	-0.33	1.16E-12	p<.001	連用
小さい	21	47	132	182	16	50	279	169	448	-0.246	2.03E-07	p<.001	連用

遠い	16	56	84	247	19	18	321	119	440	-0.459	5.97E-22	p<.001	連用
低い	26	42	55	157	86	64	263	167	430	-0.223	3.66E-06	p<.001	連用
激しい	30	51	44	115	75	112	278	149	427	-0.302	4.30E-10	p<.001	連用
古い	78	61	142	26	67	28	115	287	402	0.428	9.61E-18	p<.001	連体
暗い	19	19	164	155	19	9	183	202	385	0.049	0.33288		両方
赤い	27	17	136	148	14	1	166	177	343	0.032	0.552549		両方
重い	21	23	49	78	74	92	193	144	337	-0.145	0.007603	p<.01	連用
黒い	17	12	144	109	36	5	126	197	323	0.22	7.80E-05	p<.001	連体
軽い	10	28	39	180	14	48	256	63	319	-0.605	3.23E-27	p<.001	連用
美しい	34	14	111	125	13	7	146	158	304	0.039	0.491297		両方
明るい	14	17	56	115	54	23	155	124	279	-0.111	0.063465		両方
安い	15	48	12	23	65	105	176	92	268	-0.313	2.88E-07	p<.001	連用
少ない	11	81	7	37	54	76	194	72	266	-0.459	7.42E-14	p<.001	連用
詳しい	6	74	6	44	42	49	167	54	221	-0.511	2.93E-14	p<.001	連用
良い	22	21	24	25	53	54	100	99	199	-0.005	0.943487		両方
短い	21	26	52	59	18	22	107	91	198	-0.081	0.255509		両方
幅広い	2	2	0	1	113	79	82	115	197	0.168	0.018715	p<.05	連体
固い	2	15	25	136	1	8	159	28	187	-0.701	9.74E-22	p<.001	連用
狭い	27	11	62	29	43	15	55	132	187	0.412	1.79E-08	p<.001	連体
正しい	26	37	22	29	27	45	111	75	186	-0.194	0.008299	p<.01	連用
難しい	9	19	3	1	65	84	104	77	181	-0.149	0.044761	p<.05	連用
薄い	10	20	45	62	14	26	108	69	177	-0.22	0.003374	p<.01	連用
遅い	4	32	6	105	2	23	160	12	172	-0.86	1.56E-29	p<.001	連用
無い	7	2	34	120	5	3	125	46	171	-0.462	1.53E-09	p<.001	連用
弱い	12	10	24	37	55	21	68	91	159	0.145	0.068149		両方
細い	9	7	49	70	8	5	82	66	148	-0.108	0.188446		両方
青い	4	6	61	70	4	3	79	69	148	-0.068	0.41108		両方
著しい	4	33	4	9	17	80	122	25	147	-0.66	1.24E-15	p<.001	連用
冷たい	10	14	52	58	4	9	81	66	147	-0.102	0.216021		両方
濃い	11	6	31	67	11	7	80	53	133	-0.203	0.019222	p<.05	連用
苦しい	3	6	33	47	27	13	66	63	129	-0.023	0.791676		両方
鋭い	12	11	40	41	5	18	70	57	127	-0.102	0.248679		両方
甘い	6	6	23	49	17	23	78	46	124	-0.258	0.004057	p<.01	連用
淋しい	5	0	39	71	0	0	71	44	115	-0.235	0.01181	p<.05	連用

差異係数とは、出現の差異度を表す統計的に指標である。本研究で言えば、連用形と連体形としての出現の差異度を表す。計算方法は非常に単純であり、「連体形の出現頻度－連用形の出現頻度／連体形の出現頻度＋連用形の出現頻度」である。負の値である-1に近ければ

ば近いほど、連用形として使用されることが多いことを表し、正の値である1に近ければ近いほど、連用形として使用されることが多いことを表す。「大きい」の場合、連用形としての出現頻度が1485回、連体形としての出現頻度が149回であり、その差異係数は-0.818となる。次にエクセルの「CHITEST 関数」でカイ2乗検定を行うと同時に、有意水準を計算した。こうすることで、単に差異係数だけで、(連用形として使用されることが多いか、連体形として使用されることが多いに関する)判断を行わず、カイ2乗検定の結果を踏まえて結果を出した。これに従った場合、「大きい」は0.001水準で連用形として用いられることになる。同じ方法で見た場合、「大きい、強い、早い、深い」はかなりの確率で、連用形として用いられていることが分かる。一方の「新しい、若い、白い、古い」はかなりの確率で、連体形として用いられていることが分かる。なお、「高い、厳しい、暗い、赤い、美しい」といった形容詞に関して、差異係数としても0.03前後であり、連用形と連体形のいずれにも偏ることなく、使用されていることが分かる。こうした観点で、個々の形容詞を分類していった場合、以下の二極化された形容詞の分布が確認される。

- 連体形としての用法が主な形容詞 (34例): 新しい, 若い, 白い, 近い, 古い, 黒い, 幅広い, 狭い, 貧しい, 乏しい, 幼い, 柔らかい, 好ましい, 素晴らしい, 苦い, 望ましい, 輝かしい, 名高い, 汚い, ほど遠い, 真新しい, 可愛らしい, 古めかしい, 尊い, 凄まじい, 数少ない, やむを得ない, 四角い, 女らしい, 白々しい, 極まりない, 湿っぽい, 血なまぐさい, 手ごわい
- 連用形としての用法が主な形容詞 (132例): 大きい, 強い, 早い, 長い, 深い, 多い, 広い, 悪い, 小さい, 遠い, 低い, 激しい, 重い, 軽い, 安い, 少ない, 詳しい, 固い, 正しい, 難しい, 薄い, 遅い, 無い, 著しい, 濃い, 甘い, 淋しい, 余儀ない, 親しい, 面白い, 細かい, 熱い, 珍しい, 烈しい, 寒い, 数多い, 堅い, 痛い, 忙しい, 楽しい, 速い, 危うい, 素早い, 丸い, 根強い, 暑い, 何気ない, 久しい, 等しい, 注意深い, いち早い, 嬉しい, 快い, 紅い, 手早い, 慌しい, 勇ましい, 空しい, 永い, 遠慮ない, 憎い, 粘り強い, 奥深い, 色濃い, 青白い, 涼しい, 熱っぽい, 手厚い, 蒼い, 恥ずかしい, 欲しい, 可笑しい, 円い, 恋しい, 甚だしい, 心細い, 重々しい, 怖い, 酷い, 目まぐるしい, 眠い, 口惜しい, 頼もしい, 哀しい, 息苦しい, 果てしない, 脆い, 羨ましい, 用心深い, 善い, 委しい, 誇らしい, 情ない, 大人しい, 辛抱強い, 明い, 愉しい, 恭しい, 精しい, 堆い, 腹立たしい, 宜しい, 軽々しい, 手っとり早い, 目ざとしい, 華々しい, 赤黒い, 悪い, 気安い, 気味悪い, 喧しい, 執念深い, 凶々しい, 憎らしい, 容易い, ねばり強い, 歯がゆい, 有難い, 許可ない, 慌ただしい, 心苦しい, 麗々しい, 繁い, 恐い, 恨めしい, 酸い, 小気味よい, 人臭い, 馬鹿らしい, 煩わしい, 痒い

#### 4. 最後に(まとめと課題)

以上の調査結果から、形容詞における連体形と連用形は別の形容詞を要請することがデータから示された。こうした事実を辞書記述に反映させるなら、自動詞と他動詞のアナロジーとして連用形容詞と連体形容詞というカテゴリーを設けることも有効といえるのではないだろうか。

\* 謝辞: 本稿は、小川(他)(2008)をもとに作成された。また、一連の研究を遂行する上で、第一著者は科学研究費補助金若手研究 (B)(課題番号: 19720111)および特定領域研究日本語コーパス (課題番号: 19011003)の援助を受けた。

#### 参考文献

- [1] Utiyama, Masao and Hitoshi Isahara (2003). "Reliable Measures for Aligning Japanese-English News Articles and Sentences." ACL-2003, pp. 72--79.
- [2] 小川典子・李在鎬・横森大輔・土屋智行 (2008). 「コーパス調査による形容詞の連体形と連用形の頻度」, 2008 日本語教育学会国際研究大会(韓国, プサン) (『予稿集 2』 pp.154-157)
- [3] 川端善明 (1978, 1979). 『活用の研究 I・II』, 東京: 大修館書店.
- [4] 丹保健一 (1997). 「形容詞の連体、連用、終止用法の出現頻度と意味の関連性をめぐって: 「高い」「広い」「寂しい」を例として」, 『三重大学教育学部紀要人文・社会科学』, Vol.48, pp.9-18, 三重大学.
- [5] 橋本進吉 (1934). 『國語法要説』, 東京: 明示書院.
- [6] 橋本三奈子・青山文啓 (1992). 「形容詞の三つの用法: 終止、連体、連用」, 『計量国語学』第 18 卷 5 号, pp.201-214
- [7] 三上章 (1955). 『現代語法新説』, 東京: 刀江書院.
- [8] 宮島達夫 (1993). 「形容詞の語法と用法」, 『計量国語学』第 19 卷 2 号, pp.94-104
- [9] 李在鎬・鈴木幸平・永田由香・黒田航・井佐原均(2007). 「動詞「流れる」の語形と意味の問題をめぐって」『計量国語学』第 26 卷 2 号, pp.64-74.

#### 関連 URL

- [1] KH Coder : <http://khc.sourceforge.net/>
- [2] 日英新聞記事対応付けデータ : <http://www2.nict.go.jp/x/x161/members/mutiyama/jca/index-ja.html>

## 「全く」と「全然」の(共起述語類別)頻度比の変遷 —国会会議録のデータより—

服部匡(同志社女子大学)

### 1 目的

1947年以降の国会会議録をデータとして用い、類義性のある副詞「全然」と「全く」のそれぞれ共起する(かつ、「係る」)述語等の傾向を分析する。同時に、未分析の(プレインテキストの)コーパスを用いた共起傾向分析方法の試論ともなる。

### 2 次要素の調査

まず、ごく大雑把に二つの副詞の共起傾向を探るため、IPL+chasen を用い、副詞の次の要素を切り出したのが別紙の表 1・表 2 である。その要素を先頭として 5 要素以内に「ない」か「ません(ます+ん)」がある比率も示している。(★は、その率が 0.5%以上のもの)。<sup>1</sup>

いくつかの傾向が見て取れる。

「全く」に関して

★の増加

「ない」の順位上昇

「同感/同じ」の順位下降

「違う」の順位上昇

「全然」に関して

「違う」の順位上昇

「ない」の順位下降

「別」の順位上昇

「同感」の順位下降 (早期に)

特に、「全く」は、戦後、否定形の述語とよく共起する方向に用法変化した可能性がある。

<sup>1</sup> 1947~の区間については、旧字体の使用や仮名遣いの相違によりやや信頼性に欠ける。