

『計量国語学』アーカイブ

ID	KK290301
種別	調査報告
タイトル	形状詞の「ナ」共起と「ノ」共起のコーパス基盤調査
Title	Corpus-Based Study on the Co-Occurrence of Nominal Adjectives “ <i>Na</i> ” and “ <i>No</i> ”
著者	李 在鎬
Author	LEE Jae-Ho
掲載号	29巻3号
発行日	2013年12月20日
開始ページ	77
終了ページ	95
著作権者	計量国語学会

調査報告

形状詞の「ナ」共起と「ノ」共起のコーパス基盤調査

李 在鎬 (筑波大学)

要旨

形状詞の「ナ」(e.g. 高額なプレゼント)と「ノ」(e.g. 高額のプレゼント)による名詞修飾節の使用実態を均衡コーパスで調査した。調査データとして『現代日本語書き言葉均衡コーパス』を使用し、「名詞-普通名詞-形状詞可能+ノ・ナ+名詞」のパターンを検索し、合計頻度 10 以上の形状詞可能語(トークン頻度 21,734, タイプ頻度 277)に対して詳細な分析を行った。データ分析の方法としてクラスター分析と判別分析を使い、グループ分けを行った上で、グループ間の語義数や語彙親密度に差が見られるかを分散分析で検討した。調査の結果、3つのグループの存在が明らかになった。グループ1(e.g., 最高、大量、大型)はノとの共起が顕著で「量の概念」を表すものが多く、グループ2(e.g., 不明、平等、異質)はナとノのいずれとも共起するグループで「関係概念」を表すものが多い。グループ3(e.g., 複雑、厄介、親密)はナとの共起が顕著で、「出来事的概念」を表すものが多い。一元配置の分散分析を行ったところ、語義数において有意な差が確認された($F(2,274)=3.873, p=.002$)。

キーワード: 形状詞 (形容動詞・ナ形容詞), 共起語, 多変量解析, 多義性, 語彙親密度

1. 背景と目的

日本語の形容動詞による名詞修飾は、1)のようにナとノのいずれとも共起するタイプがある一方で、2)や3)のように一方とのみ自然な共起を許すタイプが存在する。

- | | | |
|----|----------------------|-------------------|
| 1) | a. 高額 <u>の</u> プレゼント | 高額 <u>な</u> プレゼント |
| | b. 健康 <u>な</u> 条件 | 健康 <u>の</u> 条件 |
| 2) | a. 個別 <u>な</u> 症状 | ? 個別 <u>な</u> 症状 |
| | b. 大量 <u>の</u> 雨 | ? 大量 <u>な</u> 雨 |
| 3) | a. 不思議 <u>な</u> 現象 | * 不思議 <u>の</u> 現象 |
| | b. 勝手 <u>な</u> 判断 | * 勝手 <u>の</u> 判断 |

これらの現象は、日本語の品詞論を考える上で、重要な問題であると認識され、日本語学からの分析に限らず、計量言語学や社会言語学の立場からの分析も存在し、様々なアプローチがなされてきた。例えば、日本語学の立場からの研究としては三尾(2003; 原著は

1942)はナ形容詞とノ形容詞という品詞分けを提案し、ノ形容詞が持つ機能に注目している。そして、寺村(1982)や上原・熊代(2007)では、1)の事実は、名詞と形容動詞(寺村(1982)の用語では名詞的形容詞)の連続性に関する論証として議論されるなど、様々な記述的研究がなされてきた。そして、計量言語学や社会言語学の立場からの分析としては次のものがある。水谷・星野(1994)では、昭和以降の小説や新聞記事を対象に名詞から副詞までの品詞を対象に品詞枠の再検討を行っている。とりわけ、本研究と関連が深い形容動詞語幹についても名詞と副詞の間であると位置づけ、「ノ」や「ナ」との共起を取り上げながら実例に対するコード付けを行っている。そして、田野村(2002)および田野村(2008)では新聞記事データや国会会議録データを使用し、特定の形容動詞類の「ナ」と「ノ」の選択傾向の変化を時系列にそって定量的に分析することで、文法変化の実態を記述している。さらに、羅(2009)では母語話者の使用意識をアンケート調査の方法で調査しており、ノ形容詞は中立的で客観的なコンテキストで使用されるのに対して、ナ形容詞は何らかの極性(ネガティブまたはポジティブな評価)をもったコンテキストで使用されやすいことを指摘している。

本研究は計量言語学の立場から、形容動詞類の「ナ」および「ノ」との共起現象を捉えるものであり、均衡コーパスという新しい研究資源とテキストマイニングの方法を使った調査分析を試みる。具体的な研究課題として、1)「ナ」および「ノ」との共起例を語レベルで捉え、Sunakawa et al (2012)が推し進めている日本語学習者向けの語彙データベース作成に役立つ資料を作ること、2)容認度に影響する意味的要因を明らかにし、日本語教育に資すること。1)の課題に対して、現代日本語の縮図と言える『現代日本語書き言葉均衡コーパス』(Balanced Corpus of Contemporary Written Japanese: 以下、BCCWJ)を利用し、「ナ」および「ノ」による修飾節のKWICデータを収集し、多変量解析の方法でグループ分けを行う。2)の課題に対して、単語の基本語性や語義の曖昧性などの要素が使用実態にどのようなバイアスを与えているかを分散分析で検討する。

2. データ

調査データとしてBCCWJを使用した¹。調査においては、品詞によるキーワード検索が必要であると判断し、上記の1)から3)を「茶まめ」(形態素解析辞書は「UniDic」、形態素解析エンジンは「MeCab」)を使用し、解析を行った。その結果、前節の1)から3)で取り上げた「高額、健康、個別、大量、不思議、勝手」はいずれも「名詞-普通名詞-形状詞可能²」という品詞であることが明らかになった。これを踏まえ、検索システムとして「中納言 Ver1.0.5」を使い、「名詞-普通名詞-形状詞可能+ノ・ナ+名詞」のパターンで検索を行い、表1の初期データを得た³。

1 BCCWJの使用を決めた理由は次のとおりである。近年ウェブを利用した大規模なコーパス作成が行われており、TWC (Tsukuba Web Corpus: <http://corpus.tsukuba.ac.jp/>)などは10億語を収録したコーパスである。網羅性を優先するならば、TWCなどを利用すべきであるが、前節の研究課題1)に示した日本語学習者のための語彙データベースの整備という目的においては、教育現場での実装が前提になるため、均衡性を優先するべきと判断した。つまり、学習者に提示すべき学習項目としての語彙を収集するという目的においては、現代日本語の縮図としてデザインされた均衡コーパスを使ったほうが良いと判断したのである。

表1 BCCWJ から抽出した「名詞 - 普通名詞 - 形状詞可能 + ノ・ナ + 名詞」の頻度

BCCWJ のサブコーパス	名詞 - 普通名詞 - 形状詞可能 + ナ + 名詞 (以下, 「ナ共起」)	名詞 - 普通名詞 - 形状詞可能 + ノ + 名詞 (以下, 「ノ共起」)	合計
出版・雑誌	870	468	1338
出版・書籍	3684	1964	5648
出版・新聞	285	229	514
図書館・書籍	7453	4078	11531
特定目的・ブログ	1095	568	1663
特定目的・ベストセラ	580	331	911
特定目的・韻文	21	16	37
特定目的・教科書	78	49	127
特定目的・広報誌	228	88	316
特定目的・国会会議録	643	250	893
特定目的・知恵袋	1121	662	1783
特定目的・白書	587	347	934
特定目的・法律	151	15	166
合計	16796	9065	25861

表1のデータを「名詞 - 普通名詞 - 形状詞可能」の語単位 (以下, 形状詞可能語) で集計し, 出現頻度 10 以上のものをリストアップした. その結果, トークン頻度 21,734, タイプ頻度 277 の分析データを得ることができた (具体例は【巻末資料】参照). この分析データに対して, 詳細な分析を行った⁴.

分析にあたっては, まず, 4)の計算式でもって「ナ共起」と「ノ共起」の差異係数を計算した.

2 形状詞とは UniDic によって導入された品詞であり, マニュアルによれば「[「静か」「健やか」など, いわゆる形容動詞の語幹部分]であるとされている. また, 「名詞 - 普通名詞 - 形状詞可能」とは「名詞 - 普通名詞」の下位分類の一つである. マニュアルによれば, 「名詞 - 普通名詞 - { サ変可能, 形状詞可能, サ変形状詞可能 } : 普通名詞のうち, 「運動 (する)」のように形式的な意味の「する」「できる」などが直接続き, 動詞として用いられることがあるもの, 「安全 (な)」のように「な」(助動詞「だ」の連体形) が直接続き, 形容動詞として用いられることがあるもの, (サ変形状詞可能は)「心配 (する・な)」のように両者が可能なものをそれぞれ, 「名詞 - 普通名詞 - サ変可能」「名詞 - 普通名詞 - 形状詞可能」「名詞 - 普通名詞 - サ変形状詞可能」に分類する」とされている.

3 「中納言」を使用に関しては, 形態素解析などの自動処理による誤解析の問題やアノテーションの漏れなどが懸念される. しかし, 「中納言」のアノテーションに依存した調査を計画したのには, 次の理由からである. というのは, 本研究の調査対象が語単位で予め明らかになっていれば, 語単位で文字列検索などを行い, 実例を収集することも可能であるが, そのような方法を実現させてくれる十分な資料は存在しない. そのため, 形態素解析による品詞情報を手掛かりに分析対象を集めるしかないと判断した.

4 分析データの決定においては, BCCWJ は本来サブコーパス単位で均衡性をとっているため, 理想的にはサブコーパス単位で分析を行うべきであるが, 今回は, 1. の研究背景で示した通り, 辞書記述で利用可能な網羅的なリストの作成を目指すことと定量的な分析のため, ある程度の規模の量のデータが必要であることから, 全体を一つの母集団にして分析を行った.

4)

$$\frac{\text{「ナ共起」の出現頻度} - \text{「ノ共起」の出現頻度}}{\text{「ナ共起」の出現頻度} + \text{「ノ共起」の出現頻度}}$$

差異係数順にソートした高頻度語 30 語の分布を確認したところ、表 2 が明らかになった。

表 2 差異係数順にソートした高頻度語 30 語

No.	項目	ナ共起	ノ共起	合計	差異係数
1	便利	158	0	158	1.00
2	複雑	358	0	358	1.00
3	妙	188	0	188	1.00
4	正確	187	0	187	1.00
5	駄目	236	5	241	0.96
6	変	440	11	451	0.95
7	必要	1891	67	1958	0.93
8	不思議	520	23	543	0.92
9	困難	249	15	264	0.89
10	苦手	150	11	161	0.86
11	危険	291	49	340	0.71
12	嫌い	159	31	190	0.67
13	無理	146	29	175	0.67
14	異常	156	32	188	0.66
15	幸せ	174	37	211	0.65
16	不幸	111	31	142	0.56
17	元気	110	37	147	0.50
18	自由	310	129	439	0.41
19	安全	222	111	333	0.33
20	得意	104	54	158	0.32
21	不安	96	71	167	0.15
22	健康	117	108	225	0.04
23	平和	73	86	159	-0.08
24	不明	55	116	171	-0.36
25	一杯	27	149	176	-0.69
26	固有	13	135	148	-0.82
27	別	90	1596	1686	-0.89
28	普通	27	950	977	-0.95
29	最高	7	338	345	-0.96
30	大量	4	263	267	-0.97

表 2 では、高頻度の 30 語の形状詞可能語における「ナ共起」と「ノ共起」の共起頻度を示している。「便利」から「健康」までは「ナ共起」が多いのに対して、「平和」から「大量」までは「ノ共起」が多い。表 2 の語彙は、均衡コーパスから抽出したものであるため、日本語教育においても優先的に導入すべき形状詞可能語と言える⁵。

3. 調査方法

「ノ共起」が顕著な語と「ナ共起」が顕著な語のグループ分けを行うため、共起頻度をもとにクラスター分析を行った。なお、クラスター分析の妥当性を評価する方法として、李・井佐原(2006)の提案手法にそって判別分析を利用した。そして、グループ分けを行っ

5 表 2 の日本語教育における指導の現状を確認するため、日本語教育におけるもっとも基礎的語彙資料である国際交流基金・日本国際教育支援協会(2002)『日本語能力試験出題基準【改訂版】』における収録状況を確認した。その結果、「大量」を除く 29 語はすべて収録されていた。1 級(上級)語彙として収録されている語彙が 2 語、2 級(中上級)語彙が 13 語、3 級(中級)語彙が 11 語、4 級(初級)語彙が 3 語であり、日本語教育における重要性が確認された。

た上で、それぞれの語が属するクラスターを因子にし、2つの調査を行った。1)形状詞可能語の語義の多様性と所属クラスターの関係性を調べた。2)形状詞可能語の語彙的基本語性と所属クラスターの関係性を調べた。1つ目の調査を行ったのには、多様な語義を持つ語は形式的にも多様なパターンにおいて出現するという予測の妥当性を検証するためである。2つ目の調査を行ったのは、1節の研究課題2)のためである。すなわち、クラスター間で基本語性の差が存在するなら、言語教育上の提示の優先度として、より基本的なものを先に導入し、非基本的なものを後で導入するなどのことができると考えた。1つ目の調査のため、『分類語彙表』を使い、各形状詞可能語の語義数を調べた。2つ目の調査のため『NTTデータベースシリーズ 日本語の語彙特性』を使い、各形状詞可能語の語彙親密度を調べた。そして、一元配置分散分析を使い、統計的な差があるか調べた。

4. 結果

4.1 クラスター分析の結果

「ナ共起」と「ノ共起」の各共起頻度と合計頻度を対数変換した値を独立変数にし、階層的クラスター分析を行った。クラスター化の方法はグループ間平均連結法を、測定方法は、度数のカイ2乗測度を使用した。グループ化における最適な解を見つける方法として李・井佐原(2006)が行った方法を利用した。具体的にはクラスター分析により得られた所属クラスターを従属変数に、クラスター分析時に使用した変数を独立変数にして、正準判別分析を行った。以下では、3つから6つのクラスターに分けた場合の分類結果を示す。

表3 3つのクラスターとして分けた場合の分類結果

		予測グループ			合計
		クラスター1	クラスター2	クラスター3	
観測グループ	クラスター1	68	1	0	69
	クラスター2	1	65	12	78
	クラスター3	0	4	126	130

交差確認済み判別率: 93.9%

表4 4つのクラスターとして分けた場合の分類結果

		予測グループ				合計
		クラスター1	クラスター2	クラスター3	クラスター4	
観測グループ	クラスター1	68	1	0	0	69
	クラスター2	0	51	3	7	61
	クラスター3	0	1	14	2	17
	クラスター4	0	0	8	122	130

交差確認済み判別率: 93.1%

表5 5つのクラスターとして分けた場合の分類結果

		予測グループ					合計
		クラスター1	クラスター2	クラスター3	クラスター4	クラスター5	
観測 グループ	クラスター1	66	3	0	0	0	69
	クラスター2	0	23	0	0	0	23
	クラスター3	0	3	33	0	2	38
	クラスター4	0	0	2	13	2	17
	クラスター5	0	0	1	8	121	130

交差確認済み判別率: 92.8%

表6 6つのクラスターとして分けた場合の分類結果

		予測グループ						合計
		クラスター1	クラスター2	クラスター3	クラスター4	クラスター5	クラスター6	
観測 グループ	クラスター1	66	3	0	0	0	0	69
	クラスター2	0	23	0	0	0	0	23
	クラスター3	0	3	33	0	0	2	38
	クラスター4	0	0	1	13	3	0	17
	クラスター5	0	0	1	6	46	6	59
	クラスター6	0	0	0	0	9	62	71

交差確認済み判別率: 88.4%

表3から表6の結果から誤判別のリスクがもっとも少ない、3つのグループで分けることが適切と判断した。

さて、各クラスターの「ナ共起」と「ノ共起」の詳細を確認すべく、平均頻度を求めた。まず、合計頻度としては、クラスター1が92回、クラスター2が111回、クラスター3が51回使用されており、クラスター2がもっとも生産的であることがわかった。次に、「ナ共起」と「ノ共起」の平均値は、図1のとおりである。

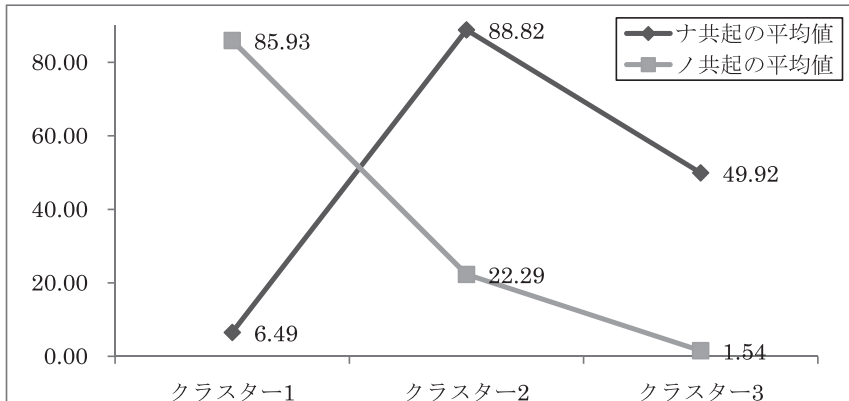


図1 クラスター間のナ共起とノ共起の平均値

図1では、各クラスターにおけるナ共起およびノ共起の平均値を示している。クラスター1はノとの共起が顕著で、ナとの共起はあまり顕著ではないグループで、クラスター3はナとの共起は中程度であり、ノとはほとんど共起しないグループである。そして、クラスター2はナとの共起が顕著であるが、ノともある程度は共起するグループである。以下、具体例を示す。

- 5) 個別, 未知, 架空, 大型, 大量, 生, 最高, 小型, 普通, ブルー, 無限, 別, 秘密, 固有, 匂, 公式, 逆, 最悪, 一杯, ソフト, オリジナル, 不満, 緊急, 偶然
- 6) 不明, 無用, 平和, 高額, 異質, 健康, 幸福, 平等, 不安, 不正, 得意, 安全, 孤独, 自由, 元気, 不幸, 高級, 公平, 幸せ, 異常, 無理, 嫌い, 困難, 不思議, 必要, 駄目, 風
- 7) 過剰, 無駄, 勝手, 違法, 公正, 楽, 余分, 正直, 苦手, 不快, 親切, 皮肉, 正常, 急, 不利, 詳細, 不吉, 慎重, 真剣, 極端, 暇, 高価, 奇怪, 正当, 残酷, 親密, 有能, 完璧

5)はクラスター1, 6)はクラスター2, 7)はクラスター3の具体例である。次節では、これらの語が持つ意味的特性を調べるべく、語義の数に基づく多義性の問題と親密度に基づく基本語性の問題について調査した。

4.2 クラスターの特徴

所属クラスターを因子にし、意味の幅を表す語義の数と基本語性を表す親密度に差があるかを調べるため、一元配置分散分析を行った。その結果、語義の数においては有意な差が確認された ($F(2,274)=3.873, p=.002$)。しかし、親密度においては有意な差は確認されなかった ($F(2,274)=0.289, p=.749$)。

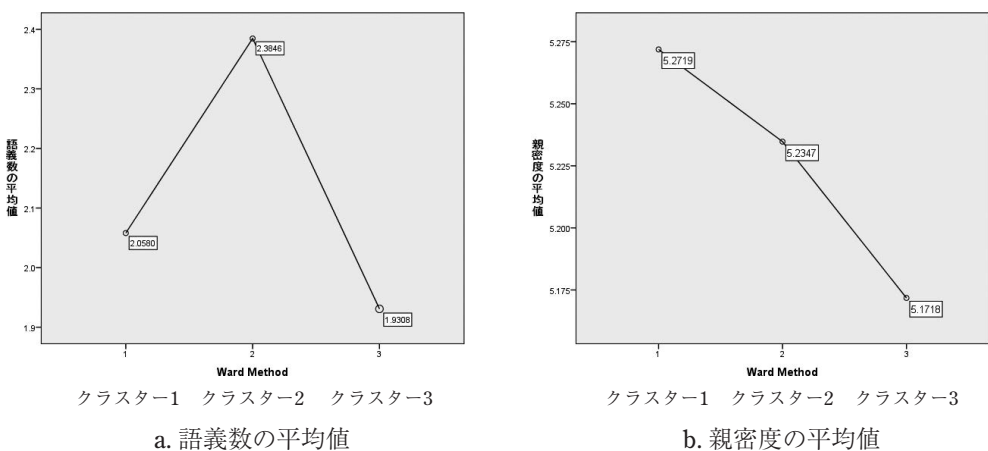


図2 クラスターによる語義数と語彙親密度の平均値

図 2a の語義数の平均値を見ると、クラスター1 の場合、2.0 個、クラスター3 の場合、1.9 個の語義を持っているのに対して、クラスター2 の場合、2.4 個の語義を持っている。なお、Tukey 法による多重比較をしたところ、クラスター2 とクラスター3 では 0.05 水準で有意な差が確認された。次に図 2b の語彙親密度に関しては、クラスター1 は「5.27」、クラスター2 は「5.23」、クラスター3 は「5.17」となっているが、統計的に有意な差は確認されなかった。ただし、クラスター1 からクラスター3 に行くにつれ、下がっていくこと、すなわち親密度は低くなっていく可能性が示唆される。

5. 総合考察

総合考察として、以下の 2 点を検討する。

- 1) 各クラスターにおける意味的特徴（意味分類）は存在するか。
- 2) クラスター2 の語義数が多いのはなぜか。

1)については『分類語彙表』による意味分類の観点から検討する。クラスター1における特徴として、「関係-量」の意味分類が付与される項目が多い。具体的には、次の15項目であり、クラスター1の21%を占めている。「最高、大量、一杯、個別、無限、多量、最低、大、ライト、ミクロ、極度、高温、高速、微量、低温」がある。クラスター2における特徴として、「活動-行為」や「活動-生活」に関わる語彙が多い。具体的には、次の26項目であり、クラスター2の33%を占めている。「困難、幸せ、不明、平和、得意、元氣、不幸、孤独、幸福、幸運、多忙、不能、神秘、無知、無垢、クラシック、従順、ストレート、無能、強気、不運、弱気、軽薄、ラッキー、俗、風流」がある。クラスター3における特徴として「活動-心」の意味分類が付与される項目が多い。具体的には、次の36項目であり、クラスター3の27%を占めている。「正確、勝手、面倒、厄介、詳細、快適、真剣、楽、慎重、悲惨、公正、親切、哀れ、愉快、親密、冷静、不快、奇怪、平気、悲痛、敏感、難解、精密、有望、不可思議、憂鬱、けち、綿密、臆病、不慣れ、鈍感、ラフ、場違い、ランダム、爽快、怪奇」がある。

さて、第二階層の情報をもとに、クラスター間で集計を行った結果、図3のとおりの結果になった。

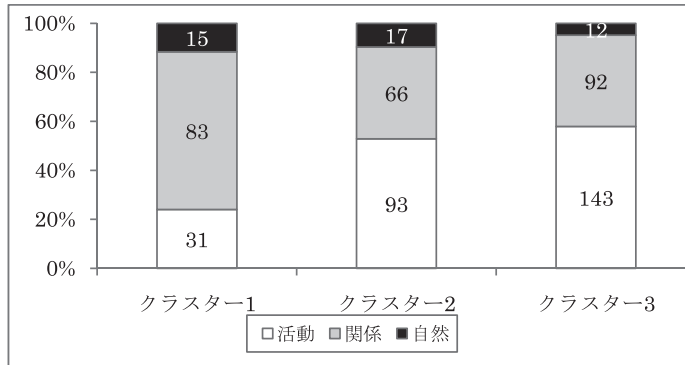


図3 クラスター×意味分類（第二階層）の集計

図3の結果から、クラスター1は関係概念を表す語が多く、クラスター3は活動概念を表す語が多い。クラスター2は、中間的な特徴を持っているが、全体的な特徴としては、クラスター3に類似しており、活動を表す語が多いと結論づけられる。ただし、これはあくまで傾向であって必要十分条件ではない。

2)の考察として、なぜクラスター2は多義的であるのかについて考えてみたい。このことを考察するにあたり、各クラスターにおける合計としての平均出現値を調べた。図4の通りである。

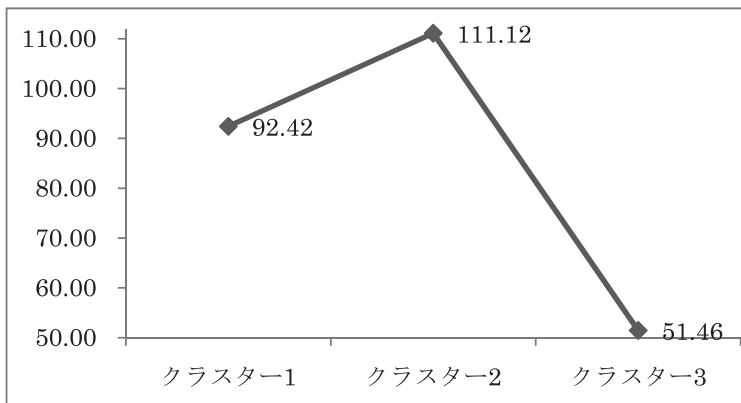


図4 クラスター別の一語の平均出現率

図4では各クラスターが合計で何回使われているかを示している。図4の出現率の変化に関して注目すべきは、図2aで示した語義数と同じ分布を示している点である。つまり、いずれの分布もクラスター2で分布の頂点が存在し、クラスター3<クラスター1<クラスター2の分布になっており、語彙の量的性質に従った振る舞いであると言える。早急な結論は難しいが、一つの見方として、多様な語義を有する語のグループは、ナとノのどちらともよく共起し、形式的にも多様であることが示唆される。

6. まとめと課題

本研究では、形状詞可能語に対するコーパス基盤調査を行った。調査の結果、次の5点が明らかになった。1)形状詞可能語の共起パターンは3つ存在すること、2)形状詞可能語の共起パターンには語義数において有意な差が存在すること、3)「ノ共起」の形状詞可能語は「関係概念」を表すものが多く、「ナ共起」の形状詞可能語は「活動(出来事)概念」を表すものが多いこと、4)多様な語義を有する語は、ナとノのどちらともよく共起することが明らかになった。5)語の基本語性を示す親密度とナ・ノの共起は無関係であること。

最後に本調査の課題として次の3点をあげる。1点目は実際の用例がもつ語義の問題を考慮しなかった点、2点目は短単位による過度な分割問題、3点目は利用した資料と手法の限界である。

1点目の問題について具体例を示す。

- 8) a. ハードな {毎日 / 課題 / 人生}
- b. ハードの {チューニング / 計算能力 / 売り上げ}
- 9) a. クラシックな {デートのマナー / 女優スタイル / エレベーター}
- b. クラシックの {名曲 / オーケストラ / 作曲}

8)のハードは、ナ共起においては、「容易ではない、労力を要する」などの意味で使われているが、ノ共起の場合はソフトウェアの反対語としてのハードウェアの意味でしか使われておらず、aの用法とbの用法を一語として扱うことに問題があると考えられる。また、9)の場合、ナ共起の例は「古典的」という意味での使い方であるが、ノ共起の場合もっぱら音楽ジャンルとしての「古典音楽」の意味でしか使われず、語義の固定化傾向が観察される。今回の調査では、これらの問題を捨象し、すべてを一語として扱った点は問題である。

2点目の問題として、本調査ではBCCWJの「中納言」が採用している短単位をもとに調査を行ったが、そのため、「非」などの接辞の問題を考慮しなかった。例えば、「非公式、非合法、非対称、非均質、非正式」などは、短単位では2形態素として解析されるため、本調査では「公式、合法、対称、均質、正式」の用例としてカウントしているが、「公式」と「非公式」の同一性は保証されないことなどを考えると調査の正確性という意味では問題があったと言えよう。

3点目の問題として、本研究では日本語教育への実装を意図しているため、均衡性を重視し、BCCWJを利用したが、BCCWJはコロケーション抽出のためには、必ずしも十分とは言えないサイズである。実際問題として、トークン頻度10以上のものということで、分析対象のデータを限定したが、このことの理論的な根拠はない。また、分析手法としてMeCabによる事前調査が十分でなかったことと「中納言」のアノテーションに全面的な頼っており、目視によるデータのチェックを行ったとは言え、次のような例が分析データに含まれた点は再考を要する。それは「こんなふうな計算」などの例も「中納言」では語彙素「風」と品詞「名詞-普通名詞-形状詞可能」としてアノテーションされているが、自立語ではないことを考えるなら、水谷・星野(1994)などを参考に何らかの再分類をする

必要があると言える。

以上の問題を考慮した場合、本研究は日本語の辞書記述全体に対する用例分析というよりは、BCCWJと「中納言」を利用した狭義の共起関係の分析と位置づけるべきである。上述の3点の問題に対しては、実データに対する複数人による精査を行うとともに、より大きなデータを使った検証が必要と考えられる。

【参考文献】

- 上原聡・熊代文子(2007)『講座 認知言語学のフロンティア 音韻・形態のメカニズム』研究社。
- 国際交流基金・日本国際教育支援協会(2002)『日本語能力試験出題基準【改訂版】』凡人社。
- 田野村忠温(2002)「形容動詞連体形における「な / の」選択の一要因 - 「有名な」と「無名の」」『計量国語学』23(4), pp.207-213.
- 田野村忠温(2008)「大規模な電子資料に見る現代日本語の動態」『待兼山論叢 文化動態論篇』42, pp.55-76.
- 寺村秀夫(1982)『日本語のシンタクスと意味 I』くろしお出版。
- 三尾砂(2003)『三尾砂著作集 II』ひつじ書房。
- 水谷静夫・星野和子(1994)「名詞から副詞まで - 語類の新しい枠づけ」『計量国語学』19(7), pp.331-340.
- 李在鎬・井佐原 均(2006)「第二言語獲得における助詞「に」の習得過程の定量的分析」、『計量国語学』25(4), pp.163-180.
- Sunakawa, Yuriko, Lee, Jae-ho, and Takahara, Mari (2012) "The Construction of a Database to Support the Compilation of Japanese Learners' Dictionaries", *Acta Linguistica Asiatica* 2(2), pp.97-115 (閲覧用リンク : <http://revije.ff.uni-lj.si/ala/article/view/174/149>)

【言語資源】

- 『現代日本語書き言葉均衡コーパス』(中納言) (<https://chunagon.ninjal.ac.jp/>)
- 『分類語彙表』 (<http://www.ninjal.ac.jp/archives/goihyo/>)
- 『NTT データベースシリーズ 日本語の語彙特性』 (http://www.sanseido-publ.co.jp/publ/ep/ntt_database.html)
- 『UniDic』 (<http://sourceforge.jp/projects/unidic/>)
- 『MeCab』 (<https://code.google.com/p/mecab/>)

(2013年7月22日受付, 同8月19日再受付)

【巻末資料】

クラスター1 (n=69, 合計頻度順)

No.	語	ナ共起頻度	ノ共起頻度	合計頻度	語義数	親密度	差異係数
1	別	90	1596	1686	2	5.25	-0.893
2	普通	27	950	977	2	6.15	-0.945
3	最高	7	338	345	3	6.40	-0.959
4	大量	4	263	267	1	6.35	-0.97
5	一杯	27	149	176	2	4.85	-0.693
6	固有	13	135	148	4	4.05	-0.824
7	最悪	19	121	140	2	6.32	-0.729
8	大型	0	131	131	1	4.98	-1
9	逆	17	112	129	2	5.85	-0.736
10	ソフト	24	105	129	5	5.65	-0.628
11	秘密	10	114	124	2	5.50	-0.839
12	生	0	125	125	4	4.35	-1
13	個別	0	109	109	1	5.28	-1
14	未知	0	91	91	2	5.50	-1
15	ブルー	4	83	87	1	6.25	-0.908
16	無限	4	77	81	2	6.05	-0.901
17	小型	2	74	76	1	5.10	-0.947
18	オリジナル	14	58	72	3	5.88	-0.611
19	偶然	18	54	72	2	6.02	-0.5
20	旬	7	62	69	2	5.98	-0.797
21	公式	7	60	67	4	4.98	-0.791
22	架空	0	66	66	1	1.48	-1
23	緊急	14	48	62	2	5.75	-0.548
24	不満	11	41	52	2	6.02	-0.577
25	グレー	2	47	49	1	5.85	-0.918
26	多量	2	47	49	2	5.85	-0.918
27	最低	9	37	46	3	5.70	-0.609
28	フリー	6	39	45	1	5.80	-0.733
29	大	1	43	44	3	5.15	-0.955
30	コア	5	36	41	1	4.72	-0.756
31	ライト	0	40	40	6	5.12	-1
32	年長	0	40	40	2	4.98	-1
33	万全	4	33	37	2	5.50	-0.784
34	ショック	7	29	36	3	5.88	-0.611
35	ミクロ	4	29	33	3	5.05	-0.758
36	至難	2	29	31	1	4.90	-0.871
37	苦痛	6	25	31	1	5.78	-0.613
38	必須	4	26	30	2	5.42	-0.733
39	やくざ	6	24	30	3	2.42	-0.6
40	極度	0	29	29	2	5.00	-1
41	空白	0	27	27	1	5.40	-1
42	同然	2	25	27	1	5.35	-0.852
43	メジャー	8	17	25	2	2.40	-0.36
44	高温	0	24	24	1	6.02	-1

45	不良	6	18	24	4	4.88	-0.5
46	無念	4	19	23	2	3.05	-0.652
47	高速	3	19	22	3	5.58	-0.727
48	純白	3	19	22	1	5.88	-0.727
49	新鋭	0	19	19	3	4.60	-1
50	好評	3	16	19	2	5.75	-0.684
51	優位	5	14	19	1	5.60	-0.474
52	静寂	5	13	18	2	5.08	-0.444
53	マニア	2	14	16	1	5.92	-0.75
54	陽性	1	14	15	1	3.05	-0.867
55	三角	0	14	14	1	5.45	-1
56	色白	0	14	14	2	6.00	-1
57	無敵	1	13	14	2	5.78	-0.857
58	旧式	3	11	14	3	4.62	-0.571
59	貧困	3	11	14	1	5.42	-0.571
60	長寿	0	13	13	1	6.05	-1
61	純正	1	12	13	4	4.20	-0.846
62	不調	5	8	13	2	6.10	-0.231
63	微量	4	8	12	1	5.48	-0.333
64	不意	0	11	11	2	5.38	-1
65	空っぽ	1	10	11	2	5.85	-0.818
66	不減	1	9	10	2	5.35	-0.8
67	低温	2	8	10	1	5.95	-0.6
68	無効	2	8	10	2	5.68	-0.6
69	対称	3	7	10	2	4.78	-0.4

クラスター2(n=78, 合計頻度順)

No.	語	ナ共起頻度	ノ共起頻度	合計頻度	語義数	親密度	差異係数
1	必要	1891	67	1958	2	6.35	0.932
2	不思議	520	23	543	2	5.98	0.915
3	変	440	11	451	3	3.30	0.951
4	自由	310	129	439	6	6.15	0.412
5	風	335	7	342	2	5.75	0.959
6	危険	291	49	340	2	5.72	0.712
7	安全	222	111	333	2	6.25	0.333
8	困難	249	15	264	4	5.48	0.886
9	駄目	236	5	241	6	5.30	0.959
10	健康	117	108	225	2	5.92	0.04
11	幸せ	174	37	211	2	6.10	0.649
12	嫌い	159	31	190	3	6.05	0.674
13	異常	156	32	188	2	5.62	0.66
14	無理	146	29	175	5	4.32	0.669
15	不明	55	116	171	3	2.65	-0.357
16	不安	96	71	167	2	6.15	0.15
17	平和	73	86	159	3	6.22	-0.082
18	得意	104	54	158	5	4.40	0.316
19	元気	110	37	147	4	5.30	0.497

20	不幸	111	31	142	3	5.92	0.563
21	孤独	68	34	102	3	3.60	0.333
22	幸福	50	44	94	2	6.32	0.064
23	異質	38	39	77	2	4.95	-0.013
24	高額	30	34	64	2	6.32	-0.062
25	高級	47	13	60	2	4.80	0.567
26	平等	32	26	58	2	6.05	0.103
27	無用	22	36	58	2	4.88	-0.241
28	公平	46	11	57	2	5.75	0.614
29	不正	32	20	52	4	5.78	0.231
30	幸運	21	28	49	2	6.45	-0.143
31	ハード	35	11	46	3	4.85	0.522
32	不審	37	8	45	2	5.15	0.644
33	名誉	17	28	45	2	5.38	-0.244
34	多忙	33	11	44	3	5.80	0.5
35	不能	27	17	44	4	4.52	0.227
36	神秘	14	23	37	2	5.42	-0.243
37	好調	19	17	36	2	5.95	0.056
38	上等	16	20	36	2	5.05	-0.111
39	ベスト	15	21	36	3	6.00	-0.167
40	プライベート	29	6	35	1	6.05	0.657
41	無知	23	12	35	2	5.78	0.314
42	大柄	23	11	34	3	5.58	0.353
43	無垢	14	20	34	6	3.10	-0.176
44	悪	24	6	30	1	5.70	0.6
45	リアル	23	6	29	1	6.00	0.586
46	クラシック	13	16	29	2	4.30	-0.103
47	残虐	23	3	26	1	5.38	0.769
48	シャープ	19	6	25	3	4.72	0.52
49	従順	21	3	24	1	5.45	0.75
50	ストレート	11	13	24	1	5.75	-0.083
51	有限	10	14	24	2	4.68	-0.167
52	垂直	13	10	23	2	5.45	0.13
53	四角	11	12	23	2	5.50	-0.043
54	真っ青	18	4	22	2	5.30	0.636
55	ローカル	18	4	22	2	5.02	0.636
56	無能	18	3	21	2	5.68	0.714
57	平穩	18	3	21	2	5.82	0.714
58	強気	8	12	20	4	5.80	-0.2
59	均等	15	4	19	2	5.50	0.579
60	不運	14	5	19	2	5.92	0.474
61	本気	9	10	19	1	6.38	-0.053
62	未練	8	11	19	1	2.18	-0.158
63	フラット	15	3	18	3	4.58	0.667
64	無力	12	4	16	2	5.38	0.5
65	レア	7	8	15	1	5.75	-0.067
66	寛容	12	2	14	1	5.20	0.714

67	弱気	11	3	14	4	5.60	0.571
68	軽薄	10	4	14	1	5.02	0.429
69	スタンダード	10	4	14	2	6.00	0.429
70	ラッキー	11	2	13	1	6.68	0.692
71	俗	9	4	13	4	4.60	0.385
72	風流	9	2	11	2	4.62	0.636
73	病弱	8	3	11	2	5.62	0.455
74	正統	8	3	11	1	4.88	0.455
75	マイナー	8	3	11	1	2.02	0.455
76	合法	6	5	11	3	5.20	0.091
77	勤勉	8	2	10	1	5.52	0.6
78	清浄	7	3	10	2	4.65	0.4

クラスター3(n=130, 合計頻度順)

No.	語	ナ共起頻度	ノ共起頻度	合計頻度	語義数	親密度	差異係数
1	複雑	358	0	358	2	5.95	1
2	妙	188	0	188	2	2.82	1
3	正確	187	0	187	1	6.15	1
4	苦手	150	11	161	1	6.02	0.863
5	便利	158	0	158	2	6.25	1
6	無駄	122	15	137	4	6.08	0.781
7	勝手	119	14	133	3	3.42	0.789
8	極端	122	4	126	3	2.70	0.937
9	面倒	120	1	121	3	5.82	0.983
10	正当	119	2	121	1	5.42	0.967
11	正直	104	8	112	1	6.25	0.857
12	厄介	108	0	108	3	5.18	1
13	詳細	100	4	104	2	5.62	0.923
14	快適	103	0	103	1	6.10	1
15	急	98	5	103	6	4.90	0.903
16	素朴	101	1	102	2	5.40	0.98
17	莫大	99	1	100	1	5.82	0.98
18	豪華	98	0	98	2	5.90	1
19	些細	96	1	97	2	5.58	0.979
20	高価	95	2	97	2	6.20	0.959
21	皮肉	91	5	96	2	1.75	0.896
22	真剣	91	3	94	2	4.45	0.936
23	特異	90	1	91	1	4.42	0.978
24	楽	82	8	90	4	5.15	0.822
25	慎重	86	3	89	2	5.62	0.933
26	悲惨	88	0	88	1	5.58	1
27	小柄	86	1	87	3	5.62	0.977
28	公正	76	8	84	4	5.28	0.81
29	繊細	83	0	83	2	5.25	1
30	完璧	82	1	83	2	6.25	0.976
31	暇	80	2	82	3	3.62	0.951
32	正常	76	4	80	1	5.78	0.9

33	親切	74	5	79	4	5.60	0.873
34	余分	72	7	79	2	5.72	0.823
35	哀れ	77	0	77	3	5.40	1
36	有能	76	1	77	1	5.28	0.974
37	平凡	72	0	72	1	5.92	1
38	半端	72	0	72	5	5.08	1
39	不利	68	3	71	2	5.15	0.915
40	優雅	68	0	68	3	5.50	1
41	愉快	68	0	68	2	5.98	1
42	親密	64	1	65	2	5.70	0.969
43	残酷	64	1	65	1	5.58	0.969
44	冷静	63	0	63	2	6.00	1
45	不快	58	4	62	2	5.88	0.871
46	過剰	53	7	60	1	5.30	0.767
47	奇怪	57	1	58	1	4.25	0.966
48	不吉	51	2	53	2	5.65	0.925
49	違法	46	5	51	1	5.78	0.804
50	神聖	45	1	46	1	4.78	0.957
51	誠実	42	4	46	1	5.68	0.826
52	強大	45	0	45	1	5.18	1
53	善良	42	2	44	1	5.30	0.909
54	平気	41	3	44	2	5.70	0.864
55	清潔	42	1	43	6	5.65	0.953
56	高貴	38	4	42	2	4.55	0.81
57	悲痛	39	1	40	1	4.95	0.95
58	敏感	39	0	39	1	5.92	1
59	重厚	39	0	39	1	4.68	1
60	カジュアル	35	4	39	1	5.45	0.795
61	粹	33	5	38	1	4.05	0.737
62	平坦	37	0	37	1	4.95	1
63	不便	35	1	36	2	5.85	0.944
64	難解	33	1	34	2	5.45	0.941
65	単調	32	0	32	2	5.32	1
66	華麗	32	0	32	1	5.58	1
67	間抜け	29	2	31	1	5.62	0.871
68	無邪気	30	0	30	1	5.32	1
69	古風	29	1	30	2	4.02	0.933
70	ポピュラー	29	0	29	2	5.72	1
71	野蛮	28	0	28	4	3.40	1
72	邪悪	28	0	28	1	4.68	1
73	有益	27	0	27	2	5.48	1
74	精密	26	1	27	2	5.45	0.926
75	残忍	25	1	26	1	5.10	0.923
76	有望	24	0	24	2	5.75	1
77	空虚	24	0	24	2	4.12	1
78	不可思議	23	1	24	1	4.58	0.917
79	賢明	23	1	24	2	5.02	0.917

80	過大	23	1	24	2	4.90	0.917
81	憂鬱	22	2	24	3	5.92	0.833
82	温厚	21	1	22	1	5.62	0.909
83	けち	21	1	22	5	5.80	0.909
84	綿密	21	0	21	1	5.18	1
85	コンパクト	21	0	21	2	4.45	1
86	無礼	19	2	21	2	5.08	0.81
87	質素	20	0	20	2	5.18	1
88	臆病	19	1	20	4	5.98	0.9
89	貧弱	19	0	19	3	4.38	1
90	オーソドックス	19	0	19	1	5.35	1
91	手近	17	2	19	2	4.78	0.789
92	精巧	18	0	18	2	5.32	1
93	冷酷	17	0	17	1	5.50	1
94	無謀	17	0	17	2	4.95	1
95	不慣れ	17	0	17	3	5.68	1
96	卑劣	17	0	17	1	4.48	1
97	鈍感	17	0	17	1	5.90	1
98	ラフ	15	1	16	5	3.18	0.875
99	貧相	15	0	15	3	4.95	1
100	卑猥	15	0	15	1	5.15	1
101	卑怯	15	0	15	1	5.10	1
102	聡明	15	0	15	1	5.00	1
103	殺風景	15	0	15	2	5.35	1
104	奇抜	15	0	15	1	5.20	1
105	豊満	14	0	14	4	4.45	1
106	壮絶	14	0	14	1	3.82	1
107	険悪	13	1	14	1	2.25	0.857
108	軽率	13	0	13	1	5.65	1
109	気弱	13	0	13	2	5.40	1
110	簡潔	13	0	13	2	5.20	1
111	エレガント	13	0	13	1	5.30	1
112	無口	12	1	13	1	5.75	0.846
113	不潔	12	0	12	6	5.92	1
114	パワフル	11	1	12	1	5.85	0.833
115	場違い	11	0	11	2	5.35	1
116	ランダム	11	0	11	2	4.58	1
117	タイト	11	0	11	1	4.88	1
118	優勢	10	1	11	2	5.12	0.818
119	無益	10	1	11	2	4.98	0.818
120	パーフェクト	10	1	11	1	6.15	0.818
121	冷淡	10	0	10	1	3.10	1
122	不向き	10	0	10	2	5.30	1
123	爽快	10	0	10	1	5.92	1
124	堅実	10	0	10	1	4.80	1
125	怪奇	10	0	10	2	4.62	1
126	スリム	10	0	10	2	5.82	1

127	適格	9	1	10	1	4.38	0.8
128	蒼白	9	1	10	1	4.98	0.8
129	高慢	9	1	10	1	4.98	0.8
130	ナチュラル	9	1	10	1	5.98	0.8

Report

Corpus-Based Study on the Co-Occurrence of Nominal Adjectives “*Na*” and “*No*”

LEE Jae-Ho (University of Tsukuba)

Abstract:

The actual status of the usage of a noun-modifying clause by nominal adjectives “*na*” (e.g., *kougakuna purezento*/expensive present) and “*no*” (e.g., *kougakuno purezento*/expensive present) was investigated in a balanced corpus. BCCWJ was used as research data, a search for the pattern of “noun (common-adjectival+*na/no*+noun)” was performed, and an elaborate analysis of potential nominal adjectives with a frequency of appearance of more than 10 was conducted (token frequency 21,734, type frequency 277). Cluster analysis and discriminant analysis were used as methods of data analysis and after dividing into groups we examined the difference in intergroup cluster number or word familiarity with an analysis of variance. The presence of three groups was revealed after the investigation. Group 1 (e.g., *saikou*/excellent, *tairyō*/massive, *oogata*/large-scale) has a pronounced co-occurrence with “*no*” and often represents a “quantity concept” and Group 2 (e.g., *fumei*/unclear, *byōdo*/equal, *ishitsu*/alien) is a co-occurring group for both “*na*” and “*no*” and often represents a “relational concept”. Group 3 (e.g., *fukuzatsu*/complex, *yakkai*/bothersome, *shinmitsu*/close) has a pronounced co-occurrence with “*na*” and often represents a “concept of event”. One-way analysis of variance confirmed a statistically significant difference in cluster number ($F(2,274)=3.873, p=.002$).

Keywords: nominal adjectives, collocation, multivariate statistics, polysemy, word familiarity