

『現代日本語書き言葉均衡コーパス』に基づくオノマトペの分析 —品詞性の検討を中心に—

宮内佐夜香 小木曾智信 小磯花絵 小椋秀樹

人間文化研究機構 国立国語研究所

1 はじめに

日本語のオノマトペは、品詞論的に多様なふるまいをすることがこれまで多数の研究において指摘されている。本研究では大規模コーパスからオノマトペの用例を広く収集し、統計学的手法によってそれらの後接要素に着目した分析を行い、オノマトペの品詞性に基づく分類について検討する。

2 先行研究

加藤・坂口(1996)は、日本語のオノマトペが「副詞としてふるまうことが多い」が、「動詞述部になることもあり」「ダ」「ナ」「ノ」を伴って形容詞としてふるまうこともある」と指摘して、小説、エッセイ等数作品を対象に調査を行い「後接成分」の表れ方を手がかりにオノマトペの分類を行っている。結果オノマトペが大きく「ダ(・ナ・ノ)」を取り得るか否かの二つに分かれ、「ダ」を取らないものは「ニ」も取らないという大きな傾向が指摘されている。ここではオノマトペを「ゼロ格」「ト格」「ニ格」に分けた上で、さらにそれが「ダ」「スル」「用言」に接続するかどうか、という階層的な分類を行っているが「ト格」「ニ格」の場合は「ダ」を取ることがないのは自明であり、後接要素の扱いに不整合が見られる。特に「ニ」を格として捉えてしまうことは「ダ」との関連性を論じる上で問題が大きい。

星野(2005)は「擬態語は短絡的に副詞とされることが多いが、文法的に考察すると副詞でないものも散見される」と述べ、特に9語をあげて統語論的な考察を行なっているが、個別の語の用法記述が主であり、品詞性の記述と意味記述を同時に行っているために、その用法の統一的な判断基準が見えにくい感がある。例えば「ダ」「デ」「ノ」「ニ」の後接する場合を体言用法としているが、これには形容動詞的な働きが認められる場合も考えられる。また、「ゆっくりする」「ゆっくりとする」は同じ意味の動詞述語であるとする等「ト」の有無を区別しない立場であるが、こうした「ト」の扱いの妥当性についても議論の余地があると思われる。

オノマトペは統語的にも意味的にも多様であり、

また手がかりとなる後接要素も、先行研究から分かるように、その統語上の位置付けについて議論すべき問題が大きい。また個々の特性に起因するところが大きく、全体的な傾向を把握するのが困難な面があると思われる。本研究では個別の形態の詳細の検討とは異なるアプローチとして、大規模コーパスを用いた用例収集を行い、統計学的手法を用いた分析を行う。議論の多い後接要素の扱いを単純化して形態のみを手がかりとし、品詞性について全体的な傾向を把握することを試みる。

3 方法

3.1 分析データ

分析データには国立国語研究所を中心に現在構築中の『現代日本語書き言葉均衡コーパス』(Balanced Corpus of Contemporary Written Japanese, 以下BCCWJ)内のサンプルのうち、白書、新聞、雑誌、書籍、Webデータ(Yahoo!知恵袋・Yahoo!ブログ)を使用した¹。BCCWJには短単位・長単位と呼ばれる2種の言語単位に基づき形態論情報が付与される(小椋ほか(2011)参照)。今回はこのうち短単位を言語単位として用例を収集した。今回利用したデータは、短単位を採用した形態素解析用電子化辞書UniDicを用いてMeCabによって自動解析されたものを元に、一部に人手修正を加えたものである。データのサンプル数と延べ語数は表1のとおりである。

表1: 分析データの内訳

コーパスの種類	サンプル数	延べ語数
白書	1,529	5,000,185
新聞	1,489	1,391,029
雑誌	2,439	5,525,606
書籍	22,433	63,961,787
Yahoo!知恵袋	45,725	5,190,722
Yahoo!ブログ	52,680	10,453,668
計	126,295	91,522,997

3.2 用例の収集

分析対象は一般に擬音語・擬態語とされるものを広く対象とし、「かくかく」「がっしり」等の様態を

¹ オノマトペを広く収集するため多ジャンルを選択したもので、本研究においてはこれらのジャンルを区別した分析は行わない。

表3:オノマトペ 後接要素別頻度表(上位 10 語)と総計

	ト	用言	スル	ダ	ナ	デ	ノ	ニ	格助詞	計
はつきり	2,747	2,560	2,979	0	1	6	0	0	0	8,293
ゆっくり	3,745	1,876	404	49	19	52	11	24	1	6,181
たっぶり	583	1,165	39	48	49	78	617	160	2	2,741
どンドン	114	2,329	12	0	0	3	2	1	2	2,463
すっかり	23	2,312	2	0	0	1	0	0	1	2,339
さっ	1,652	11	4	2	1	0	0	0	1	1,671
ぴったり	339	389	88	261	125	67	366	24	0	1,659
ぼんやり	861	273	406	1	2	2	2	2	2	1,551
すつきり	357	225	881	12	3	2	7	1	2	1,490
さっぱり	234	633	371	50	11	13	1	1	1	1,315
...
総計	63,670	30,723	21,028	1,938	640	1,339	3,749	3,702	669	127,458

表すもの、「とんとん」「からん」等の音を表すもの
 の他、和語の実質的な意味との関連も考えられる「つ
 やつや」「ねじねじ」「ひんやり」などの語も対象と
 した。漢語系の「凛々」「燦々」等は対象外とした。
 短単位においては、オノマトペは単体で1語とし「と」
 「に」などの後接要素は別語として切り離す。用例
 収集においてはオノマトペとそれに続く後接要素を
 分析対象として抽出した²。ただし特殊な語形を除く
 ために、1サンプルにしか現れない語は除外した。

次に後接要素についてである。本研究ではオノマ
 トペ直後の語にのみ着目して表2のような要素を指
 標とし、これに当たらない用例は分析対象外とした。
 対象外のは文末での使用(例:よく寝てすつきり。),
 複合語の構成要素(例:しっとり感),間に他
 の成分を挟んで用言を修飾する副詞用法等である。

表2:オノマトペ後接要素一覧

後接要素 ラベル	内訳
ト	出現形が「と」のもの(格助詞・引用)
ニ	出現形が「に」のもの(格助詞・連用修飾)
ナ	出現形が「な」のもの
ダ	出現形が「だ」「だっ(た)」のもの・「です」
デ	出現形「で」(格助詞・「である」等)
ノ	出現形が「の」のもの
スル	動詞「する」
用言	動詞・形容詞・形容動詞
格助詞	「と」「に」「で」「の」を除く格助詞

以上のような基準により抽出された用例は異なり
 1,719 語, 延べ 127,458 語であった。総計上位 10
 語の後接要素別の頻度と用例総数を表3に示す。

3.3 分析

後接要素を指標として、類似する性質を持つオノ

マトペをまとめ上げるために、オノマトペの後接要
 素別頻度表に対して対応分析を行なった。分析には
 RのMASS パッケージの中のcorresp関数を用いた。
 さらに距離の近いオノマトペをグループ化するた
 めに、対応分析で得られた各オノマトペのスコア(2
 次元)を対象にクラスター分析を行なった。これに
 はhclust関数を用い、ユークリッド距離+群平均法
 によって分析した。

4 結果と考察

クラスター分析を行った結果から、オノマトペに
 ついて6つのグループを抽出した。所属語数と後接
 要素の総数を表4, 所属する語の頻度上位 10 語を
 表5に示す。対応分析の結果から、代表語として表
 5に示した上位 10 語と後接要素の相対的位置をプ
 ロットすると、図1のようになる(寄与率は1軸:
 54.8%, 2軸:45.2%)。

後接要素の位置関係をみると、x・y軸方向とも
 に正の方向に広がりは見られるが概ね直線的に配置
 されているように見て取れる。それぞれの位置関係
 を見ると〈ダ〉〈ナ〉〈デ〉〈ノ〉が近接しており、こ
 れらは関連性の強い後接要素であると言えるだろう。
 また〈格助詞〉〈スル〉〈用言〉〈ト〉などと〈ニ〉が
 非常に離れているのが特徴的である。「ト」が付加さ
 れる副詞と「ニ」が付加される副詞は意味的な比較
 において差異が指摘されているが³, 今回の対応分析
 でも最も遠い位置関係になっている。〈ダ〉〈ナ〉〈デ〉
 〈ノ〉はその中間に位置している。

以下、クラスター分析で分けられたオノマトペの
 各グループをこの後接要素の位置に対応させて、表
 4に示した後接要素の頻度を参照しながら見ていく。

[グループ1]は〈ト〉を中心に位置しており、

² 短単位では「と」を含めて1語の副詞として情報を付与しているものもある(「じっと」「どっと」「ぞっと」「ぼうっと」等)。こうした副詞は分析対象外とした。

³ 佐々木(1986)等。

表4:オノマトペ6分類の所属語数と後接要素総計

	グループ1	グループ2	グループ3	グループ4	グループ5	グループ6	計
所属語数・異なり	1,015	355	95	155	90	9	1,719
所属語数・延べ	54,004	43,330	12,216	11,860	5,935	113	127,458
ト	45059	13157	1786	3021	647	0	63,670
用言	6194	19577	1166	3410	375	1	30,723
スル	1687	9340	8691	1048	262	0	21,028
格助詞	212	143	72	203	39	0	669
ダ	189	302	258	699	483	7	1,938
ナ	50	86	50	299	154	1	640
デ	207	201	56	468	402	5	1,339
ノ	249	294	101	1814	1282	9	3,749
ニ	157	230	36	898	2291	90	3,702

表5:オノマトペ・6グループ頻度上位 10 語

グループ1	グループ2	グループ3	グループ4	グループ5	グループ6						
ゆっくり	6,181	はつきり	8,293	すつきり	1,490	たつぷり	2,741	ばらばら	1,044	びしょびしょ	62
さっ	1,671	どんどん	2,463	どきどき	872	びったり	1,659	ぎりぎり	800	こてんぱん	22
ゆったり	1,102	すっきり	2,339	がっかり	777	さらさら	463	ぼろぼろ	593	べろんべろん	10
ぱっ	1,092	ぼんやり	1,551	にこにこ	674	ふわふわ	366	びかびか	339	けちよんけちよん	7
ぐっ	1,018	さっぱり	1,315	うんざり	628	がたがた	326	どろどろ	260	べこべこ	5
ちら	894	のんびり	1,265	わくわく	559	ぼりぼり	291	からから	210	こちんこちん	3
ふっ	699	じゅっ	1,119	ごろごろ	504	ぼちり	277	くたくた	181	ぐちよぐちよ	2
ちらり	691	にっこり	824	うろうろ	491	ごちゃごちゃ	251	ぐちゃぐちゃ	178	ぼっかぼっか	1
すっ	675	そろそろ	803	にやにや	386	ちょい	249	くしゃくしゃ	177	ねじねじ	1
びん	675	きらきら	698	ぐずぐず	319	つつる	235	とろとろ	146		
	14,698		20,670		6,700		6,858		3,928		113

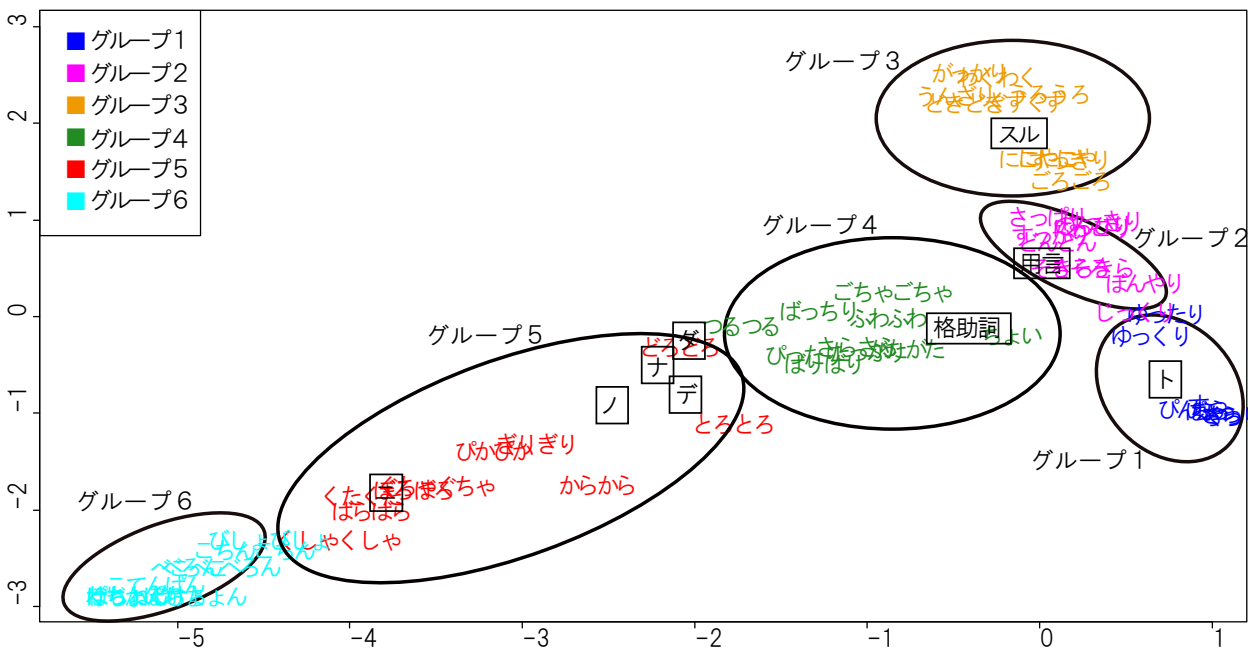


図1:各グループ頻度上位 10 語・後接要素の相対的位置

頻度を見ても〈ト〉の数値が大きい。所属語は1音の「ば(っと)」2音の「ちら(っと)」など通常「ト」を付加するタイプの語が 337 語所属しているが、その他「ゆったり」「ちらちら」等用言に直接付くこと

も可能な語が多い。こうした〈ト〉の有無どちらも可能なものであっても、頻度上〈ト〉付加に偏る語群がここに所属していると言えるだろう。

[グループ2]は主に〈用言〉周辺に位置してお

り、単独で連用修飾成分となる性質の強い語群と言えるが、〈ト〉や〈スル〉の頻度も高い。連用成分として多様な後接要素を取り得る一群と考えられる。

[グループ3]は〈スル〉周辺に位置しており、サ変動詞として使用される頻度が高いものが所属している。

[グループ4]は〈格助詞〉に近い位置にある。〈格助詞〉は全用例で0.5%程度(640/127,458例)だが、このグループでは1.7%(203/11860例)と相対的に高い頻度となっており、特徴的である。後接要素の全体の傾向は〈ト〉や〈用言〉の頻度が特に高いが、〈ダ〉〈ナ〉〈デ〉〈ノ〉の後接する頻度もグループ1, 2などより高くなっていることが指摘される。この一群は副詞的にも形容動詞的にも、名詞的にも働く多機能な語群であると言える。他に〈格助詞〉が現れるのは主に[グループ1, 2]で、概ね副詞的なグループに所属していることが分かる。

[グループ5]は〈ダ〉〈ナ〉〈デ〉〈ノ〉から〈ニ〉周辺にかけて配置される。[グループ6]とは異なり、〈用言〉や〈ト〉の頻度が比較的少ない。この一群は[グループ4]同様に多機能とも言えるが、より形容動詞的な性質の強い語が所属していると言える。

[分類6]は全体に低頻度語であり、他の分類との用例数の差が大きい。これはほとんどの場合〈ニ〉を後接して用いられる語が所属している。

以上から、今回収集したオノマトペの用例群の品詞論性の傾向は表6のようにまとめられる。

表6:オノマトペの品詞性分類

グループ	品詞性の分類	異なり語数
1	ト付加型副詞・名詞	1015語
2	副詞(単独)・ト付加型副詞・スル動詞・名詞	355語
3	スル動詞	95語
4	多機能型(副詞的・名詞)	155語
5	多機能型(形容動詞的)	90語
6	ニ付加型副詞	9語

それぞれの分類の頻度1位の語の後接要素を例として表7にあげる。

表7:各分類頻度1位語の後接要素

	ゆっくり	はっきり	すっきり	たっぷり	ばらばら	びしょびしょ
ト	3745	2747	357	583	88	0
用言	1876	2560	225	1165	13	1
スル	404	2979	881	39	2	0
格助詞	1	0	2	2	0	0
ダ	49	0	12	48	102	6
ナ	19	1	3	49	70	1
デ	52	6	2	78	98	4
ノ	11	0	7	617	95	6
ニ	24	0	1	160	576	44
総計	6181	8293	1490	2741	1044	62

今回の分析で、〈格助詞〉〈ダ〉〈デ〉〈ノ〉〈ニ〉が付加されるものはその頻度の傾向が異なっており、グループが分かるといえる結果が得られた。これらの後接要素の品詞性の判断は研究によってゆれが見られる点だが、「体言的」のようにまとめられるものではなく、名詞的、形容動詞的、副詞的のように分けて考えるべきものと思われる。また統語上同様に副詞として働くために区別せずに扱われることもある副詞(単独)とト付加型副詞だが、ト付加型に偏る語群が認められ、性質の異なるものとして扱うべき側面が確認された。

所属語数ではグループ1のト付加型副詞が最も多く、続いていずれも後接要素が〈用言〉であるグループ2, 3が続く。オノマトペは典型的に副詞として用いられるという従来の指摘通りの分布である。日本語のオノマトペはその多様性が着目されるが、その広がりを中心はト付加型副詞にあると言える。それに対し、意味的分析で比較対象として取り上げられるニ付加型は、それに偏って用いられる語が比較的少ないようである。ニ型は限定的な語群であり、品詞性の分布上はオノマトペとして特異な性質を持つと考えられる。

5 おわりに

以上多量の用例を対象にした統計学的な分析に基づいて、オノマトペの品詞性の分類を試みた結果、品詞性の違いを論じるにおいて、個別の語の検討では見えにくい基準が明確になったものとする。今回考慮しなかった所属語彙の意味的な性質や、後接要素の形態のみでは判別できない統語機能と関連付けた分析は今後の課題とする。

参考文献

- 小椋秀樹ほか(2011)国立国語研究所内部報告書『『現代日本語書き言葉均衡コーパス』形態論情報規程集第4版』
- 加藤久雄・坂口昌子(1996)「後接成分とオノマトペの性質について」『奈良教育大学紀要 人文・社会科学』45:1
- 佐々木文彦(1986)「擬態語類の語尾について」『松村明教授古希記念 国語研究論集』明治書院
- 星野和子(1991)「擬態語の用法—構文論の立場から—」講座日本語教育 26 (早稲田大学日本語研究教育センター)
- 星野和子(2005)「擬態語の文法」『駒沢女子大学研究紀要』12

付記

本研究は、文部科学省科学研究費特定領域研究「日本語コーパス」による補助を得たものである。