

表層格に着目したオノマトペ共起語の抽出と分析

乙武北斗¹ 内田ゆず² 高丸圭一³ 木村泰知⁴

^{*1}福岡大学 ^{*2}北海学園大学 ^{*3}宇都宮共和大学 ^{*4}小樽商科大学

^{*1}ototake@fukuoka-u.ac.jp, ^{*2}yuzu@eli.hokkai-s-u.ac.jp, ^{*3}takamaru@kyowa-u.ac.jp, ^{*4}kimura@res.otaru-uc.ac.jp

1. はじめに

日本語のオノマトペには複数の語義を持つものが多い。例として、日本語オノマトペ辞典[1]には「ばらばら」の語義が4種類記載されている。また、基本となる語義から派生した新たな語義や用法が観察されることもある。

オノマトペの語義は文脈によって決定することが可能であると考えられる。具体的にはオノマトペの共起語（係り先動詞や文の表層格）を手がかりにできると考えられる。そこで本稿では実際の用例を対象に、オノマトペの係り先と表層格の抽出を試みる。抽出結果について観察し、オノマトペ共起語とオノマトペの語義の関係について具体例を挙げて述べる。

2. 係り先と表層格の抽出方法

2.1. 対象データ

本研究では『現代日本語書き言葉均衡コーパス』（以降BCCWJ）を対象に、オノマトペの係り先と表層格を抽出する。抽出対象とするオノマトペは、日本語オノマトペ辞典の「五十音順さくいん」に掲載されている全4,506語のうち、4文字以上¹で構成される3,165語とした。

2.2. 抽出手順

オノマトペ、および、その係り先と表層格の抽出方法を図 1に示す。はじめに、BCCWJから2.1で述べた対象オノマトペを含む文を抽出する(図 1(1))。オノマトペの判定は字面の一致で行う。

次に、抽出した文を対象に形態素解析を行う(図 1

(2))。形態素解析はMeCab[2]+IPA辞書を用いた。対象オノマトペが1形態素として抽出された場合に限り、係り先抽出と表層格抽出(図 1 (3) (4))を行う。

係り先と表層格の抽出にはCaboCha[3]による係り受け解析結果を用いる。図 2, 図 3にそれぞれ例文(i), (ii)の係り受け解析結果を示す。

- (i) 釣りで捻挫のしょんぼり気分を酒でパーッと発散して帰った。
- (ii) 子供は急にしょんぼりした顔であたしを見つめる。

例文(i)に含まれる対象オノマトペ「しょんぼり」は1形態素として抽出されるため、係り先抽出の対象となる。図 2で示すように、例文(i)の「しょんぼり」の係り先が文節「発散して」となっているため、係り先として文節の先頭形態素である「発散」を出力する。例外として、例文(ii)のように、オノマトペの直後に動詞「する」があって、かつオノマトペと同一文節に属する場合は、「する」を係り先として出力する(図 3)。

オノマトペの係り先を決定した後、係り先の係り元文節を対象に表層格の抽出を行う。図 2および図 3の点線で囲まれた部分が表層格抽出対象となる係り元文節である。係り元文節の助詞を手がかりとして、ガ格、ヲ格、ニ格、デ格、カラ格、ヘ格、ト格、ヨリ格、マデ格を抽出する。図 2の例の場合、ヲ格としてオノマトペ自身を含む「しょんぼり気分を」、デ格として「酒で」、ト格として「パーッと」が抽出される。

¹ これまでの研究から3文字以下のオノマトペは形態素解析誤りの割合が高いことがわかっている。形態素解析の時点で誤りが生じると表層格の抽出を適切に行えないため、4文字以上のオノマトペを対象とする。

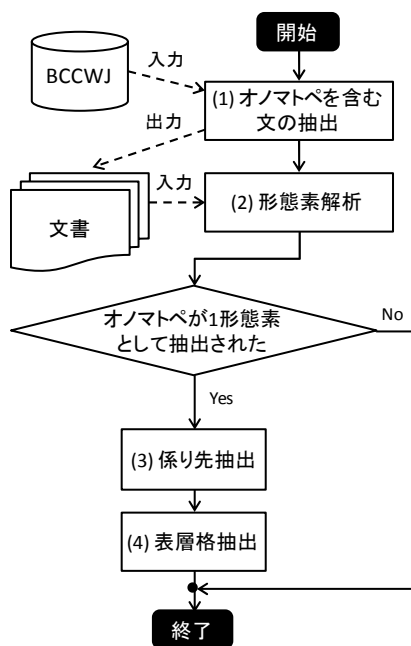


図 1 オノマトペ抽出の流れ

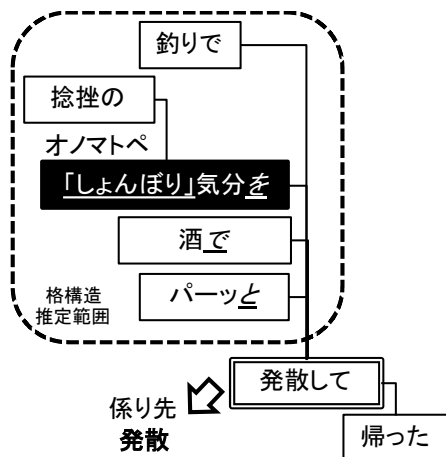


図 2 表層格抽出の例 (基本的な処理)

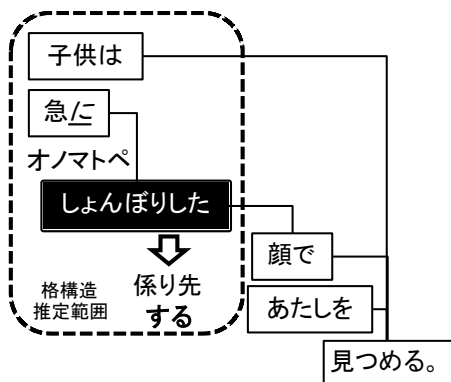


図 3 表層格抽出の例 (オノマトペ+する)

図 1 の (2) 形態素解析においてオノマトペが1形態素として抽出されなかった場合は、続く処理である係り先抽出や表層格抽出は行わない。ただし、その際は係り先や表層格は「なし」として出力する。

3. 係り先と表層格の抽出結果

2.1で述べた対象データについて、2.2で述べた方法で係り先と表層格の抽出を行った。その結果、222,197文から用例が抽出された。

表 1～表 7に、頻度が10以上のオノマトペ+係り先について各表層格を伴う割合が高い順に並べたものを一部示す。例えば、表 1の「ブツブツ―途切れる」は13回出現し、そのうちすべてがガ格を伴っていることを表している。へ格、ヨリ格については、これらを伴う割合が50%以上となるような「オノマトペ+係り先」が見つからなかったため、本稿では割愛した。

3.1. 特徴的な表層格を伴うオノマトペと係り先

本節ではオノマトペが「どのような語と共起しているのか」という点に着目して、表 1～表 7の事例から特徴的なものを取り上げる。

共起語が一意に決まる定型文のような例として、表 2で示すヲ格の例「がっくり―〇〇を―落とす」があり、すべての例でヲ格を伴っている。ヲ格を具体的にみると、30例のうち26例が「肩を」であり、「膝を」が2例、「首を」、「力を」が1例ずつとなっている。肩に代表される身体の部位が多数を占めており、語義も一意に決定づけられる。

一意ではないが類似する共起語を伴って語義が一定になりそうな例として、表 1で示すガ格の例「ブツブツ―〇〇が―途切れる」がある。「ブツブツ―途切れる」はすべての例でガ格を伴っている。ガ格を具体的にみると、13例すべて「動画が」となっており、語義が「ものを細かく切るさま」に一意に決まる。「ブツブツ―できる」もガ格を伴う割合が高いが、この場合はガ格にブツブツ自身が入っており、このときの語義は湿疹などを表す。

3.2. 表層格がオノマトペ自身である場合

表 1～表 7に示した各表層格を伴うオノマトペと係り先の種類として、オノマトペ自身を含む文節が表層格に入る場合とそうでない場合とに分けられる。オノマトペ自身が表層格として入る例に着目すると、表 4に示すニ格の例はほぼすべてオノマトペ自身がニ格に入っていることが確認された。具体的には、「ドロドロ→なる」は90%の例でオノマトペ自身がニ格に入っていた。表 4から、係り先「なる」でニ格を伴うオノマトペは、「(オノマトペ) になる」という表現が多い傾向があることがわかる。

表 1 ガ格を伴うオノマトペと係り先

オノマトペ	係り先	頻度	ガ格割合
ブツブツ	途切れる	13	100%
こんがり	つく	12	100%
だんだん	立つ	10	100%
ブツブツ	できる	14	93%
どンドン	なくなる	14	93%
つつい	出る	12	92%
ギッシリ	(なし)	12	92%
ゴタゴタ	ある	12	92%
ゴロゴロ	鳴る	12	92%
さっぱり	つく	23	91%

表 2 ヲ格を伴うオノマトペと係り先

オノマトペ	係り先	頻度	ヲ格割合
じっくり	かける	36	100%
がっくり	落とす	30	100%
ばちばち	する	29	100%
すっかり	取り戻す	19	100%
たっぷり	吸う	18	100%
びっしょり	かく	18	100%
ガックリ	落とす	15	100%
ゴロゴロ	鳴らす	14	100%
しばしば	伴う	13	100%
しばしば	起こす	12	100%

表 3 カラ格を伴うオノマトペと係り先

オノマトペ	係り先	頻度	カラ格割合
どンドン	離れる	17	59%
こっそり	出る	11	55%

表 4 ニ格を伴うオノマトペと係り先

オノマトペ	係り先	頻度	ニ格割合
くしゃくしゃ	なる	24	100%
くたくた	疲れる	22	100%
ドロドロ	なる	20	100%
バサバサ	なる	20	100%
ひたひた	なる	19	100%
ピカピカ	磨く	19	100%

表 5 デ格を伴うオノマトペと係り先

オノマトペ	係り先	頻度	デ格割合
ゆっくり	煮る	14	64%
ワイワイ	する	11	64%
ばったり	会う	43	58%
ゴシゴシ	する	14	57%
ムンムン	する	11	55%
コツコツ	叩く	13	54%
プチプチ	包む	28	54%
はっきり	確かめる	10	50%
じっくり	焼く	22	50%
ゴシゴシ	洗う	10	50%

表 6 ト格を伴うオノマトペと係り先

オノマトペ	係り先	頻度	ト格割合
きよとん	する	126	100%
キョトン	する	60	100%
もうろう	する	44	100%
シャキッ	する	43	100%
びしゃり	言う	38	100%
ひしひし	伝わる	31	100%
ありあり	浮かぶ	21	100%

表 7 マデ格を伴うオノマトペと係り先

オノマトペ	係り先	頻度	マデ格割合
ギリギリ	寝る	126	100%
ぎりぎり	する	60	100%

4. オノマトペの語義と表層格の関係

本研究では、表層格の情報をオノマトペの語義分類に利用することを目的として、語義と表層格の関係を分析している。ここでは、「ばらばら」の語義とヲ格の関係を考察する。「ばらばら」の5つの語義を表 8に示す。表 8の語義I～IVはオノマトペ辞典に掲載されており、語義Vは我々が検討して独自に付け加えた項目である。2.2で述べた方法により得られた

「ばらばら」の係り先と表層格からヲ格に関する用例のみを取り出し、それらのコロケーションを表 8 に示す5つの語義に人手で分類した。

分類結果について語義ごとに述べる。図 4に語義ごとのコロケーションを示す。ヲ格を伴う語義Iの「ばらばら」は、対象データ中に用例が1つしかなかったため、「ひげを一生やす」のみの結果となっている。

語義IIは語義Iの場合と異なり、複数の用例が存在した。ヲ格には調味料や薬味などの料理の材料、もしくは粒状の小さな物が該当する。係り先は図 4に示すもののほかに、「注ぐ」なども見られた。ヲ格に該当する名詞の種類によって、係り先に該当する動詞に若干の違いが見られた。

語義IIIのヲ格には料理の材料か、料理名が該当する。係り先動詞は「なる」、「作る」などが該当する。「ばらばら」が二格に入る場合は語義IIIが該当することが確認された。

語義IV, Vのヲ格には、両者ともに本や紙に関する名詞が該当するため、ヲ格からのみでは語義を決定することができない。例外として、語義Vの場合は「目を通す」という表現が存在する。係り先まで確認すると語義IV, Vで違いが見られた。語義IVの場合は「めくる」、「繰る」、「弾く」などがあり、語義Vの場合は「見る」、「読む」および係り先なしといった用例が存在した。

5. おわりに

本稿では、表層格に着目したオノマトペ共起語の抽出と分析結果について述べた。分析はまだ十分な量の用例に対して行っていないため、今後も引き続き様々な用例に対して分析を行っていきたい。

謝辞

本研究の一部は科研費基盤(C) No. 26370498による。

参考文献

- [1] 小野正弘編: 日本語オノマトペ辞典, 小学館, 2007.
- [2] Taku Kudo, Kaoru Yamamoto, Yuji Matsumoto: Applying Conditional Random Fields to Japanese Morphological

Analysis, Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP-2004), pp.230-237, 2004.

- [3] 工藤拓, 松本裕治: チャンキングの段階適用による日本語係り受け解析, 情報処理学会論文誌, 43(6), pp1834-1842, 2002.

表 8 「ばらばら」の語義

語義I	数量が少ない物が点在しているさま
語義II	雨や霰や木の葉など軽いものがまばらに落ちてくる音、さま
語義III	一つのまとまりをもったものが散らばって広がるさま
語義IV	本などのページが軽く次々にめくれる音、さま
語義V	(紙をめくるさまから転じて) 簡単に見るさま

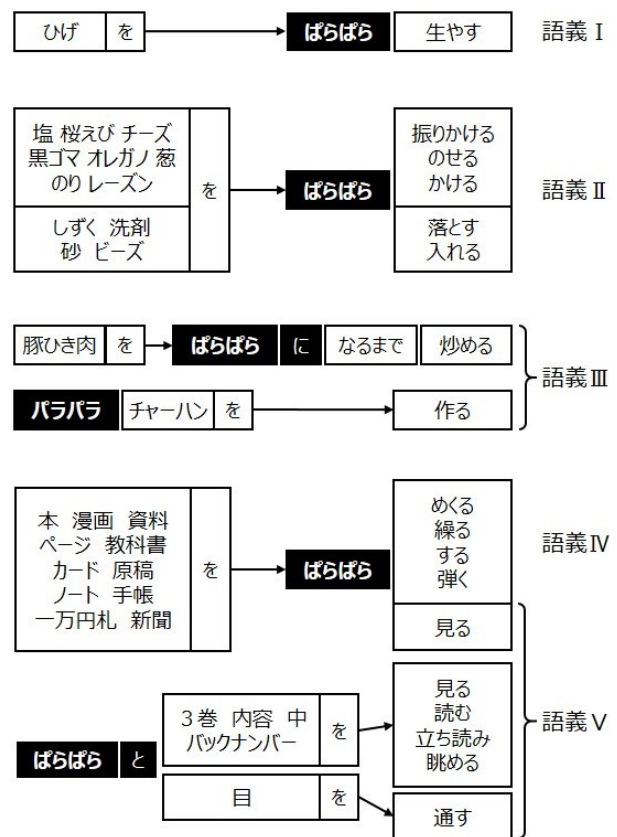


図 4 「ばらばら」のコロケーション