

# BCCWJにおけるオノマトペの語義とコロケーションの分析

## Analysis of Onomatopoeias' Senses and Collocations in BCCWJ

内田ゆず<sup>\*1</sup>  
Yuzu Uchida

高丸圭一<sup>\*2</sup>  
Keiichi Takamaru

<sup>\*1</sup> 北海学園大学  
Hokkai-Gakuen University

<sup>\*2</sup> 宇都宮共和大学  
Utsunomiya Kyowa University

乙武北斗<sup>\*3</sup>  
Hokuto Ototake

木村泰知<sup>\*4</sup>  
Yasutomo Kimura

<sup>\*3</sup> 福岡大学  
Fukuoka University

<sup>\*4</sup> 小樽商科大学  
Otaru University of Commerce

The goal of this study is to develop an onomatopoeia dictionary with a large amount of example sentences. Since most onomatopoeia words have multiple senses, it is necessary to add information concerning senses to each of the sentences. Knowledge of collocation is important for word-sense disambiguation because the sense is influenced immediately by the linguistic context. We analyzed onomatopoeias' senses and collocations in BCCWJ. This analysis allows us to reveal a determining factor of an onomatopoeia's sense.

### 1. はじめに

オノマトペ(擬音語および擬態語)は音や鳴き声、状態、程度を効果的に伝える語である。日本語のオノマトペは複数の語義を持つものが多い。例えば、『日本語オノマトペ辞典』[小野 07]には、オノマトペ「ふわふわ」の語義として、①「やわかくふくらんでいるさま」、②「あてもなく浮かただようさま」、③「心や動作が軽々しく落ち着かないさま」の3つが掲載されている。[小野 15]によると、「ふわふわ」は、中世には主に③の語義で使われており、近世以降に①および②の語義でも使われるようになったとされる。このようにオノマトペがもつ複数の語義は、オノマトペが修飾する対象の広がりなどに伴って、派生的に生成されたものであることが多い。日本語母語話者は、日常会話の中で用いられるオノマトペの意味を直感的に理解することができる。一方、日本語非母語話者にとって、文脈からオノマトペの語義を類推することは困難である。日本語非母語話者の学習支援には、オノマトペの語義と使用される文脈の関係を実例に基づいて整理する必要があると考えられる。さらに、自然言語処理などの工学分野でオノマトペを利活用する場合も、同様のことがいえる。

そこで、筆者らは現代の日本語文における「オノマトペの語義」と「オノマトペと共起する語(コロケーション)」との関係に着目した研究を進めている。[内田 15]では、日本語コーパスにおけるオノマトペの出現実態を表層形態の変異パターンに着目して分析した。[高丸 15]では、オノマトペの語義とオノマトペの係り先動詞の関係について分析し、「オノマトペー係り先動詞」の組のうち約85%は単一語義で出現することを指摘した。また、[乙武 16]では、オノマトペとその共起語を抽出する手法について検討を行った。

係り先動詞について、上述の「ふわふわ」の例について考えると、「ふわふわ」の係り先が「膨らむ」であれば①の語義、「飛ぶ」や「漂う」であれば②の語義、「暮らす」であれば③の語義である可能性が高い。これらの動詞はオノマトペ「ふわふわ」にとって、

語義の区別に寄与するものであると考えられる。しかし、動詞「する」はオノマトペの係り先として頻出するものの語義の区別に寄与しづらいと考えられる。「ふわふわ」の例でも、「ふわふわ」+「する」は①から③のいずれの語義でも使用することができる。オノマトペの語義が係り先動詞によって区別できない場合、次の手がかりとして、文中の(係り先動詞以外の)共起語を用いる必要がある。例えば、「ふとんが」+「ふわふわ」+「する」(語義①)、「雲が」+「ふわふわ」+「する」(語義②)、「気持ち」+「ふわふわ」+「する」(語義③)のように、文のガ格に共起する語は語義の手がかりにすることができると予想される。オノマトペと係り先動詞を中心とした単文の範囲で、語義判断の手がかりとなる共起語が属しやすい格はオノマトペごとに異なる可能性がある。そこで、本研究は共起語が属する格とオノマトペの語義との関係についての傾向を明らかにすることを目的とする。

本稿の構成は以下の通りである。2章では、現代日本語書き言葉均衡コーパス(BCCWJ)における、オノマトペが動詞「する」に係る用例から、オノマトペとその共起語を抽出する。3章では、各オノマトペの語義を手手で分類する。さらに、4章では、オノマトペの語義と共起語が属する格についての分析結果について述べる。

### 2. コロケーションの抽出

この章ではコーパスからオノマトペのコロケーションを抽出する方法とその結果について述べる。ここで、コロケーションとは、オノマトペの係り先動詞と、係り先動詞の係り元のうち格助詞(が、を、に、で、から、へ、と、より、まで)を伴う文節を指す。

#### 2.1 対象データ

コーパスは BCCWJ の全データを用いる。データ数は 5,932,696 文である。BCCWJ には、書籍、新聞、広報紙、ブログ、国会会議録などのジャンルにまたがったデータが収録されているが、本研究ではジャンルの区別は行わない。

対象とするオノマトペは、[小野 07]に掲載されている 4,506 語のうち 4 文字以上のオノマトペ 3,165 語である。これまでの研究

連絡先: 内田ゆず, 北海学園大学工学部電子情報工学科,  
札幌市中央区南 26 条西 11 丁目 1 番 1 号, yuzu@hgu.jp

から、3文字以下のオノマトペは形態素解析誤りの原因になりやすいことがわかっているため、コロケーション抽出の対象としない。

## 2.2 抽出方法と抽出結果

先行研究[乙武 16]と同様の方法でオノマトペを含む文を抽出し、さらにオノマトペのコロケーションを抽出する。以下に手法を簡潔に説明する。

- (1) 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。オノマトペの判定は字面の一致で行う。
- (2) 抽出した文を対象に形態素解析を行う。形態素解析器は MeCab[Kudo 04] + IPA 辞書を用いる。
- (3) (2)でオノマトペが 1 形態素として抽出された文を対象に構文解析を行う。構文解析器は CaboCha[工藤 02] を用いる。
- (4) (3)で得られた係り受け解析結果から、オノマトペの係り先動詞を抽出する。
- (5) (4)の係り先動詞の係り元文節を対象に、格助詞(が, を, に, で, から, へ, と, より, まで)を伴う文節を抽出する。ただし、当該文節にオノマトペ自身が含まれている場合は抽出の対象としない。(例: パーツを**ばらばら**にする)

これら一連の処理は、オノマトペを含む単文から表層格を考慮した文脈情報を抽出することを意図している。たとえば、「平日にゆっくりと家族で食事をした経験がない、ということも考えられるのだ。」という用例からは、オノマトペ「ゆっくり」に対するコロケーションとして、係り先動詞「する」、文節「平日に」、「家族で」、「食事を」が抽出される。

BCCWJ からオノマトペを含む文は 222,197 文抽出された。さらに、上記(2)~(5)の処理によって、104,358 文からコロケーション情報(係り先動詞と1つ以上の文節)が抽出された。ただし、形態素解析誤り、構文解析誤り、助詞「は」によるガ格の表現などの原因で、コロケーションが適切に抽出できないことも多い。コロケーションの抽出精度の向上は今後の課題である。

## 3. 語義分類

多義オノマトペの語義を判断する上で、係り先動詞は重要な役割を果たす[高丸 15]。一方で、オノマトペが動詞「する」に係る場合は語義が一意に定まらないことが多く、語義を決めるには周辺文脈の情報が必要である。そこで、「する」を係り先にもち、出現回数が 100 以上であり、何らかのコロケーションが抽出された、という3つの条件を満たすオノマトペを対象として、語義分類を行う。該当するオノマトペは 26 語、用例は 7,273 文存在する。

語義の定義は原則的には[小野 07]に従うが、実際の用例を辞典の語義に照らし合わせると、分類が難しいことがしばしばある。オノマトペの語義は定義が曖昧で、特に多義オノマトペでは、非常に近い語義をもつ場合がある。たとえば、複数の辞典[小野 07, 山口 03, 飛田 02, 阿刀田 95]で見出し語「はつきり」を参照すると、語義が 3~5 種類に分けて定義されている。このように、多義オノマトペの語義の境界は明確ではない。したがって、語義分類が困難なオノマトペについては、共著者間で協議の上、既存の辞典を参考にしながら用例の実態に即した語義の定義を再構築した。

表 1 に、オノマトペの語義、各語義の用例数、オノマトペが当該語義で用いられる際の共起語が属する格の出現割合を示す。たとえば、「あっさり+する」が語義①で使われている用例は 7 例存在し、そのうち 14.3%の用例がガ格に属する共起語をもつことを示している。これらの値から、オノマトペがある語義で用いられる際に、どのような情報が補足されるのかが示唆される。語義とコ

ロケーションに一定の傾向があれば、語義分類などに有益な情報となる。

なお、現段階では、「頭がすっきりする」のようにオノマトペがサ変名詞としての用法で用いられる場合と、「すっきりした後味」のように連体修飾部としての用法で用いられる場合の区別をしていない。語義分類に有用な知見を得るためには、これらを切り分けて分析する必要があるだろう。

## 4. 語義とコロケーションの関係

語義分類の結果、「オノマトペ」+「する」が単一語義になったものと複数語義になったものに分けて考察する。なお、用例数が十分ではないため、オノマトペの語義とコロケーションの関係について、一般化された知見を示すことは難しい。したがって、コロケーションに特徴のあるオノマトペについて、個別具体的な例を挙げて考察する。

### 4.1 単一語義で用いられたオノマトペ

「うんざり」、「きよとん」、「しばしば」、「どきどき」の 4 語は[小野 07]で語義の定義が 1 種類であった。オノマトペの語義・用法は変化しやすく、辞典に掲載されていない語義で使用される場合がある。したがって、辞典に単一語義として定義されているオノマトペに対しても人手で語義分類を行った。その結果、4 語とも全ての用例が辞典に掲載された語義で用いられていることが確認された。以下に、実際の用例を示す。

#### 【元来単一語義のオノマトペの用例】

- ダラダラ続くリストラに組織全体がうんざりしており…
- 一瞬、誰もがきよとんとしていた。
- ぼくは目をなんどもなんどもしばしばさせて、それを見た。
- 胸がどきどきしていまにも気絶しそうだ。

コロケーション上の特徴をいくつか述べる。「うんざり+する」は、助詞「に」を伴う文節と多く共起する。「うんざり」は「物事がいやになる」という語義をもち、いやになる対象をニ格で表す形で用いられる。「しばしば+する(させる)」は、用例数は少ないものの文節「目を」と高い確率で共起する。また、「どきどき+する」は、157 例で助詞「が」を伴う文節と共起している。そのうち 126 例は{心臓, 心, 胸}+「が」であった。

「いらいら」、「がっかり」、「びっくり」、「わくわく」の 4 語は、多義オノマトペであるものの、ほぼ全ての用例が同一語義で用いられていた。[小野 07]では、「びっくり」の語義として②「鋭くほんの少し動くさま」が定義されているが、この語義は現代日本語では一般的ではない。このように、やや特殊な用法や、方言に基づく語義の用例は出現しなかった。以下に、実際の用例を示す。

#### 【単一語義のみが出現したオノマトペの用例】

- その要点を回避した答え方が中隊長をいらいらさせる。
- 誕生日に娘をがっかりさせる父親なんているかしら？
- 僕はMの告白にびっくりして言葉がでなかった。
- いつも斬新な発想でボクたちをわくわくさせてくれる。

これらのオノマトペの用例では、助詞{が, を, に, で}を伴う文節がやや多く共起しているが、他に目立った特徴はない。4 語とも感情を表すオノマトペであり、能動態・受動態のどちらの用法も一般的である。態によって使用される格が変化する(例:[人]が[原因]でいらいらする/[人]が[原因]にいらいらする/[原因]が[人]をいらいらさせる)ため、{が, を, に, で}が同程度用いられると考えられる。

## 4.2 複数語義で用いられたオノマトペ

語義とコロケーションに特徴が見られたオノマトペについて、以下に用例を挙げる。なお、用例の前に記載されている番号は、表1の語義番号に一致している。

### 【しっとり の用例】

- ① のびがよく、肌が本当にしっとりします。
- ② 歩くだけでしっとりとした中国情緒を満喫できる。
- ③ 大島紬にしっとりとした女らしさを演出してください。

### 【ぶらぶらの用例】

- ① 棧橋にあがって突端に腰をおろし、脚をぶらぶらさせた。
- ② 旅行最後の日は町の中心部をぶらぶらしました。
- ③ 仕事もたず家でぶらぶらしていた。

### 【ばたばた の用例】

- ③ 団扇をばたばたさせる位では、暑さしのぎにならぬ。
- ⑤ 法事の用意で家の中はばたばたしている。
- ⑥ みんな儲からなくなり、ばたばたと閉店をする。
- ⑦ 手足をばたばたとさせるような発作が見られた。

### 【ばらばら の用例】

- ③ 死体をばらばらにするのは余計な仕業ではないか。
- ④ 職場がばらばらで誰が何をしているのか、分からない。

「しっとり」+「する」は3種類の語義に分類されたが、多くは語義①「湿気が軽く全体にいきわたっているさま」での用例であった。語義①に該当する用例は、[具体物(肌, 食べ物)]+「が」が多く含まれる。語義②「静かに落ち着いたさま」、③「女性の、静かな中にも色っぽさが漂っているさま」は、「しっとりとした」+「名詞」の形で用いられることが多く、語義の判別には「しっとりとした」の係り先が有用である。しかし、2.2で述べた手法では「しっとりとした」の係り先は抽出されない。今後、コロケーションの抽出範囲を再考する必要がある。同様のことが、「さっぱり」、「すっきり」などにもいえる。

「ぶらぶら」+「する」は3種類の語義に分類された。語義①「宙に釣り下がって揺れ動くさま」で用いられる場合は、ヲ格に共起語が多く、「[体の一部]をぶらぶらさせる」が典型的な用法である。語義②「あてもなく時間をかけて歩くさま」、③「何の目的もなく怠惰に暮らしているさま」では、[場所]+「を/で」との共起が多いが、[場所]にあてはまる単語が語義②では屋内、語義③では屋内という違いがある。

「ばたばた」は[小野 07]では7種類の語義が定義されている。その中で、「ばたばた」+「する」の用例は4種類の語義に分類された。語義⑤「あわただしく事を行うさま」は、デ格に慌しさの原因をもち、ヲ格の共起語はほとんど見られない。一方、語義③「羽音や団扇などを大きくあおぐ音、さま」、⑦「手足を強く振ったり床に打ちつけたりするさま、音を立てて勢いよく歩くさま」では、ヲ格の共起語が多い。ヲ格に属する共起語の内容は、③では{団扇, 羽}、⑦では{手足, 両手, 腕}などが多く、共起語の意味カテゴリが語義分類に寄与すると考えられる。

「ばらばら」は[小野 07]では5種類の語義が定義されている。その中で、語義①「粒状のものが、散らばりながら広がって、物にぶつかる音、さま」、②「多数が、いっせいに、または次々に立ったり動いたりするさま」、⑤「本などのページが乱暴に次々とめくれる音、さま」は「ばらばら」+「する」の形式では用いられにくいことが示唆された。語義③「一つのまとまりをもったものが、こまかくわかれるさま」では56例中55例が「ばらばらにする」の形で用いられており、原則的に目的語を必要とする。したがって、ヲ格の共起語が多い。語義④「一つにかたまらないであちこちにあ

るさま、それぞれが独立していてまとまりのないさま」の用例は、意図したコロケーションを抽出できていないものが多かった。上記の用例「職場がばらばらで…」からは文節「誰が」、「何を」が抽出されるが、これは適切ではない。オノマトペが形容動詞のように用いられる際の抽出手法の改善が必要である。

3章でも触れたように、多義オノマトペの語義の境界は曖昧で、日本語母語話者でも判別が難しい場合がある。本稿で分析対象とした26語の中では、「あっさり」の語義①②、「うろうろ」の①②、「さっぱり」の①③、「しっかり」の③④⑤、「はっきり」の②③、「ゆっくり」の①③などの区別が特に難しい。これらの語義は、抽出されたコロケーションの特徴も類似している。また、語義が互いに重なり合う部分もあり、実際の運用上は区別する必要がないと考えられるものもある。今後、オノマトペ使用の実態に即した語義分類を検討する予定である。

## 5. おわりに

共起語が属する格とオノマトペの語義との関係についての傾向を明らかにすることを目的とし、BCCWJにおける「オノマトペ」+「する」の用例の語義とコロケーションを分析した。その結果、一部のオノマトペでは、語義によって共起語の属する格に一定の傾向があること、共起語に意味的な類似性があることが明らかになった。代表的な例として、「ぶらぶら」は「揺れ動くさま」を表す場合は[体の一部]+「を」と共起するという特徴が挙げられる。

今後は、オノマトペと共起する語を意味の観点からクラスタリングし、そのオノマトペを使用できる文脈、複数の語義の違いなどを提示するシステムの構築を目指す。

## 謝辞

本研究は科研費(No.26370498)の助成を受けたものである。

## 参考文献

- [Kudo 04] 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.
- [阿刀田 95] 阿刀田稔子, 星野和子: 擬音語擬態語使い方辞典, 創拓社, 1995.
- [内田 15] 内田ゆず, 高丸圭一, 乙武北斗, 木村泰知: BCCWJ コアデータにおけるオノマトペ出現実態の分析—現代オノマトペ事例辞書アプリの構築に向けて—, 人工知能学会全国大会論文集 29, 3G4-OS-05b-5, 2015.
- [乙武 16] 乙武北斗, 内田ゆず, 高丸圭一, 木村泰知: 表層格に着目したオノマトペ共起語の抽出と分析, 言語処理学会第22回年次大会予稿集, P7-7, 2016.
- [小野 07] 小野正弘編: 日本語オノマトペ辞典, 小学館, 2007.
- [小野 15] 小野正弘: 感じる言葉オノマトペ, 角川選書, 2015.
- [工藤 02] 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.
- [高丸 15] 高丸圭一, 内田ゆず, 乙武北斗, 木村泰知: 現代オノマトペ・コロケーション・データベース構築のための係り先動詞の分析, ファジィシステムシンポジウム講演論文集 31, 415-420, 2015.
- [飛田 02] 飛田良文, 浅田秀子: 現代擬音語擬態語用法辞典, 東京堂出版, 2002.
- [山口 03] 山口仲美: 暮らしのことば 擬音・擬態語辞典, 講談社, 2003.

