

## 用例に基づくオノマトペ用法辞書の編纂

古 武 泰 樹 佐 藤 理 史

名古屋大学大学院工学研究科

furutake@sslslab.nuee.nagoya-u.ac.jp, ssato@nuee.nagoya-u.ac.jp

### 1. はじめに

オノマトペとは、「ばんばん」、「さらし」、「しーん」などの擬音語・擬態語の総称である。日本語は、他の言語に比べてオノマトペが非常に豊富な言語であり、動作のニュアンスや感情などを伝える際に、頻繁にオノマトペが用いられる。

自然言語処理において、一般に、オノマトペは形態素解析誤りを引き起こしやすい。これは、既存の形態素解析用辞書がオノマトペに対して十分な情報を有していないことに原因がある。形態素解析用辞書に、未定義のオノマトペを追加すれば、この問題は解決できると考えられる。しかしながら、それぞれのオノマトペを、どのような語形で、どのような品詞として登録すればよいかは、それほど自明ではない。

このような背景から、我々は、オノマトペの語形と用法を整理した辞書(オノマトペ用法辞書)の編纂を行っている。本稿では、これまでに編纂した辞書の概要と、その編纂手法、および、評価について述べる。

### 2. オノマトペ用法辞書の概要

図 1 に、本研究で編纂したオノマトペ用法辞書のエントリの例を示す。辞書のエントリは、見出し語に対して定義される表記リストと音韻形態パターンの 2 つの情報と、複数の用法別サブエントリから構成される。ここで、音韻形態パターンとは、1 モーラないし 2 モーラの語基から、オノマトペがどのように構成されているかを示すものである<sup>1)</sup>。

用法別サブエントリは、そのオノマトペが使用される用法を示すものである。本辞書では、オノマトペの用法を、連用修飾用法、連体修飾用法、用言用法の 3 種類に分けて記述する<sup>\*</sup>。各用法の記述では、とりうる語形に対して、共起単語と用例が記述されている。

#### 2.1 連用修飾用法

連用修飾用法は、「きらきらとかがやく」のように、副詞的にオノマトペが使われる用法をいう。この用法は、次の 3 種類の語形をとる。

- (1) 「～(と)」

<sup>\*</sup> これらの用法以外に、体言と文外独立の二つの用法があるが、現段階では記述していない。

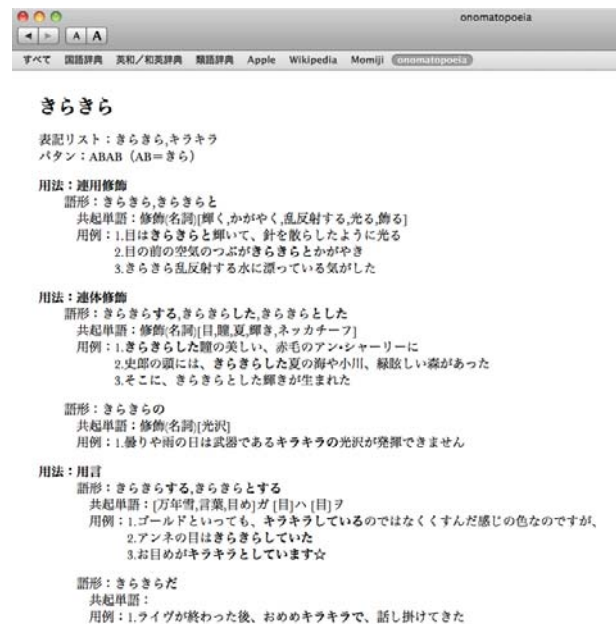


図 1 辞書のエントリの例

- (2) 「～に」

- (3) 「～(と)して」

共起単語には、被修飾語である用言(現時点では動詞に限定)が記述されている。

#### 2.2 連体修飾用法

連体修飾用法は、「きらきらした瞳」のように、体言を修飾するオノマトペの用法である。この用法は、次の 4 種類の語形をとる。

- (1) 「～(と)した」「～(と)する」

- (2) 「～な」

- (3) 「～の」

- (4) 「～という」

共起単語には、被修飾語である体言(現時点では名詞に限定)が記述されている。

#### 2.3 用言用法

用言用法は、「目がきらきらしている」のように、用言として使われる用法である。おもに、次の 2 種類の語形をとる。

- (1) 「～(と)する」

- (2) 「～だ」

共起単語には、用言の直前に現れる体言を助詞ごとに整

理した形で記述されている。

### 3. 見出し語の作成

辞書の編纂は以下の3ステップで行った。

- ステップ1. 見出し語リストの作成
- ステップ2. 見出し語エントリの作成
- ステップ3. 用法別サブエントリの作成

ここでは、ステップ1と2の内容について述べ、ステップ3は次節で述べる。

#### 3.1 見出し語リストの作成

今回は、辞書の収録対象を比較的良好に用いられていると考えられるオノマトペに限定した。具体的には、『日本語オノマトペ辞典』<sup>2)</sup>(見出し語 4,060 語)と『現代擬音語擬態語用法辞典』<sup>3)</sup>(見出し語 1,072 語)の両方に収録されているオノマトペ 1,035 語を見出し語として採用した。

#### 3.2 見出し語エントリの作成

各見出し語に対して表記リストと音韻形態パターンを作成した。

オノマトペは、一般に、複数の表記を持つ。各オノマトペに対して、Web や辞典を参考にして表記リストを作成した。今回は、以下の3種類を異表記として認定した。

- (1) 平仮名・片仮名の表記揺れ  
例:「ぼろぼろ」「ポロポロ」
- (2) 長音の表記の違い  
例:「ぜーぜー」「ぜえぜえ」「ぜいぜい」
- (3) 末尾の促音の有無  
例:「びくびく」「びくびくっ」

音韻形態パターンの記述は、田守<sup>1)</sup>に従った。1モーラの語基をA、2モーラの語基をAB(2種類ある場合はABとCB)と記述した。このような抽象化により、音韻形態パターンは全部で23種類となった。なお、これ以外に、典型的な音韻形態パターンを持たないオノマトペが存在する。これは「特殊」と記述した。

### 4. 用法別サブエントリの作成

用法別サブエントリは、コーパスを利用し、自動的に作成した。この作成手順は、(1)用例の収集と、(2)サブエントリの決定の2つのステップからなる。図2に、その作成過程を示す。

#### 4.1 用例の収集

本辞書では、コーパスに用例が存在する場合に、その用法が存在するとみなし、サブエントリを作成するという方針をとる。用法別サブエントリ作成の中心は、各オノマトペに対して、連用修飾用法、連体修飾用法、用言用法の各用法の用例が存在するかどうかを判定する処理である。

たとえば、オノマトペ「きらきら」は、連用修飾用法の場合に、「きらきら(と)」という語形をとる。これに対して、「びしょびしょ」は、「びしょびしょに」という語形をとる。このように、各オノマトペによって、同じ用法

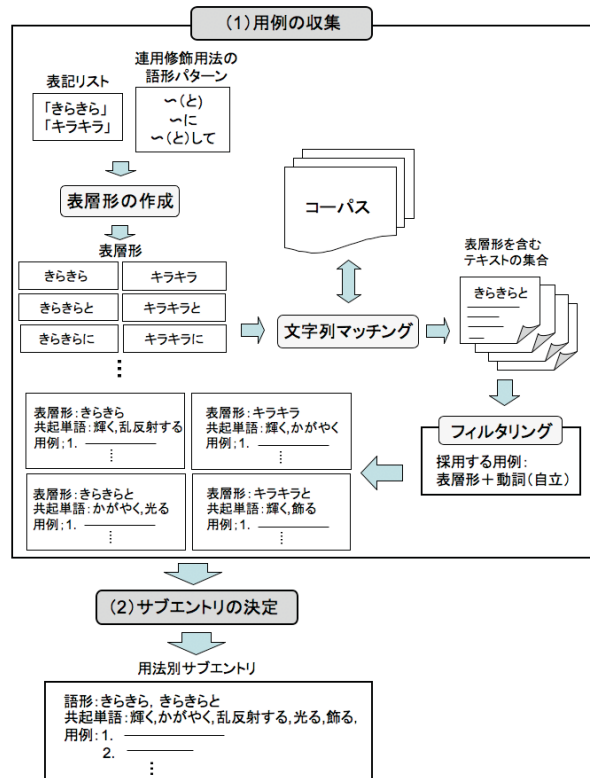


図2 用法別サブエントリ作成の流れ

であっても用いられる語形が異なる。どのような用法にどのような語形が用いられるかについては、既存の辞書にも記述はあるが、その記述は網羅的ではない。

我々は、まず、見出し語にどのような付属語が付与されるかを整理し、これを語形パターンとして定義した。定義した語形パターンを表1に示す。活用する用言用法に対しては、活用済語形パターンをあわせて定義した。これらの語形パターンに、見出し語の表記を代入すれば、実際の文に現れる表層形が得られる。但し、実際には存在しない表層形も過剰に生成される。

原則として、各オノマトペの全ての表記に対し、定義した全ての語形パターンを用いて表層形を生成する。但し、2モーラの表記(「どん」など)、および、末尾が促音「っ」の表記(「たつたつ」など)に対しては、「と」がオノマトペの直後に接続する語形パターンのみ限定する。これは、表層形の過剰生成を抑えるためのヒューリスティックである。

こうして得られた表層形のそれぞれに対して、その表層形を含む文を文字列マッチングによって収集する。この時点で、その表層形を含む文がひとつも見つからなかった場合は、その表層形は存在しないもの(過剰に生成されたもの)と判断する。

文字列マッチングによって収集された文は、かならずしもその表層形の用例として適切であるとは限らない。次のような場合は、用例として明らかに不適切である。

- (1) オノマトペが他の語の一部となっている

表 1 各用法の語形パターンと活用済語形パターン

用法	語形パターン	活用展開済み語形パターン
連用修飾用法	～(と), ～に, ～(と)して	
連体修飾用法	～(と)した・する, ～な, ～の, ～という	
用言用法	～(と)する	～(と)する, ～(と)して,～(と)した, ～(と)しろ,～(と)させる,～(と)させて, ～(と)させた,～(と)させろ
	～だ	～だ,～で,～なのだ,～なので,～なのに

表 2 採用する用例

用法	採用する用例
連用修飾用法	…表層形+動詞(自立)…
連体修飾用法	…表層形+名詞(自立)…
用言用法	全て <sup>☆</sup>

☆ 「～する」「～した」の表層形は表層形+名詞となる用例を、「～して」の表層形は、表層形+ (動詞(自立語)・形容詞・形容動詞)となる用例をそれぞれ除いた。

表 3 収集する共起単語

用法	共起単語
連用修飾用法	表層形の直後の動詞
連体修飾用法	表層形の直後の名詞
用言用法	表層形の直前の名詞+助詞

表 4 語形パターンの分布

用法	語形パターン	作成数
連用修飾用法	～(と)	803
	～に	120
	～(と)して	100
連体修飾用法	～(と)した・する	324
	～な	45
	～の	191
	～という	109
連用修飾用法	～(と)する	435
	～だ	190

表 5 各用法の正誤

用法	正	誤	合計	精度
連用修飾用法	817	17	834	98.0%
連体修飾用法	451	51	502	89.8%
用言用法	483	35	518	93.2%

(2) 用法としての制約を満たしていない

オノマトベが他の語の一部となっている文とは、以下のような文である。

- 表層形「どん-と」
  - 文「壁をどん どんと 叩く」
- 表層形「まったり-する」
  - 文「とど まったりする」

このような文を排除するために、既知語との照合を行う。具体的には、表層形の直前の任意の長さのひらがな列とオノマトベの表記(表層形を作る際に語形パターンに代入した表記)との連結が、既知語あるいは既知語の活用形等である場合は、収集した文を破棄する。上記の例では、「どんどん」、「とどまつたり」が既知語であるため、これらの文は破棄される。

用法としての制約を満たしていない文とは、たとえば以下のように、連体修飾用法の用例を求めているのに、直後に体言が存在しない文である。

- 表層形「どきどき-した」
  - 文「もっと どきどきしたい」

このような文を除去するために、各用法に対し、用例の採否を決定するための条件を表2のように設定した。この条件を満たしているかどうかを、表層形より後ろの文字列を切り出し、それを形態素解析することによってチェックする。条件を満たさなかったものは、破棄する。

以上の不適切な文の除去をパスしたものを用例として採用する。この後、採用した用例に含まれる表層形の前方、もしくは、後方の文字列を切り出して形態素解析し、共起単語を収集する。どのような語を共起単語として抽出するかは、用法ごとに異なる(表3)。

なお、今回は、用例を収集するためのコーパスとして、現代日本語書き言葉均衡コーパス 2009 年度モニター公開版<sup>☆</sup>を使用した。用例の形態素解析には Mecab<sup>☆☆</sup>を、形

態素解析用辞書には IPAdic<sup>☆☆☆</sup>を使用した。

4.2 サブエントリの決定

収集した用例を語形パターンごとに整理し、それらを更に用法ごとに整理して、用法別サブエントリを作成する。この結果、最終的に、927 語のオノマトベに対して、総計で 1,854 件の用法別サブエントリが作成された。各用法の語形パターンの分布を表4に示す。

5. 評価

編纂したオノマトベ用法辞書の各用法別サブエントリに対し、実際にそのオノマトベがその用法を持つかどうかを、既存のオノマトベ辞典等を参考に、人手で調査した<sup>☆4</sup>。その結果を表5に示す。この表からわかるように、連用修飾用法のサブエントリの精度は 98%と非常に高い。これに対して、連体修飾用法および用言用法は、90%程度である。

☆☆☆ <http://sourceforge.jp/projects/ipadic/>

☆4 この調査は、用法別に行った。すなわち、用法別サブエントリに記述されている、いずれかの語形で、その用法が存在すれば、正解とした。

☆ <http://www.kokken.go.jp/kotonoha/>

☆☆ <http://mecab.sourceforge.net/>

## 5.1 用法の判定誤りの分析

各用法に共通して見られる誤りで最も多かったものは、オノマトペ以外の同音異義語との区別がつかないことに起因する誤りである。たとえば、オノマトペ「パリ」に対して、「パリに行く」という用例が採用され、連用修飾用法のサブエントリが作成された。

次に多かったのが、オノマトペが体言用法を持つことに起因する誤りである。たとえば、オノマトペ「いざこざ」に対して、以下のような用例が採用され、連用修飾用法、および、連体修飾用法のサブエントリが作成された。

- 「いざこざに巻き込まれる」
- 「いざこざの原因」

用法別に見ると、連体修飾用法では、「～(と)した」と「～という」の2つの語形パターンに誤りが多く見られた。これは、以下の例のように、オノマトペが連体修飾節の用言として用いられている場合を、連体修飾用法と区別できないことに原因がある。

- 「岩がごろごろとした道」
- 「ごはんはちよっぴりという状態」

## 5.2 辞書の利用

今回編纂した辞書を用いると、用言や体言に対して、それらをオノマトペで修飾した表現を自動的に作成することができる。これを確認するプログラムを作成した。たとえば、名詞「光」というオノマトペを入力すると、プログラムは以下のような表現を出力する。

- 「ぼんやりとした光」
- 「ギラギラした光」
- 「しんとした光」

また、動詞「眺める」というオノマトペを入力すると、以下のような表現を出力する。

- 「うっとり眺める」
- 「のんびり眺める」
- 「にやにやして眺める」

現時点で辞書に収録されている共起単語は、その数、および、信頼性ともに十分ではない。これらの点を改善することができれば、このようなプログラムは、作文支援などの応用に利用できると考えられる。

## 6. 関連研究

オノマトペ辞書の編纂に関する研究として、Web上のコーパスからオノマトペの辞書を自動編纂する奥村ら<sup>4)</sup>の研究がある。この研究は、既存の知識源に掲載されていないオノマトペを発見し、その語義や用法を得ることを想定して、典型的な音韻形態パターンからオノマトペ候補語を生成し、その用例を収集するというアプローチをとっている。この研究は、コーパス中の用例に基づいて辞書を編纂するという点で、我々の研究との共通点がある。しかしながら、対象としているオノマトペは異なる。彼らの研究は、主に、未知のオノマトペを発見することに主眼が置かれているのに対し、我々は、既知のオ

ノマトペに対して、用法と語形を網羅的に記述した辞書を編纂することを目標としている。

## 7. まとめと今後の方針

本稿では、オノマトペの語形と用法を整理した辞書を、コーパスを用いて編纂する方法について述べた。この方法を、比較的よく用いられていると考えられるオノマトペ1,035語に対して適用した結果、927語の見出し語に対して、総計で1,854件の用法別サブエントリを作成することができた。作成された用法別サブエントリに対して、実際にその用法が存在するかどうかの調査を行ったところ、比較的良好な結果を得た。

上記の調査は、用法別の調査であり、語形ごとの調査は行っていない。今後は、この点を調査し、用例の収集法を改良していく予定である。

## 参考文献

- 1) 田守育啓. 2003. 日本語オノマトペの音韻形態オノマトピア 擬音語・擬態語の楽園, pp.1-15, 勁草書房
- 2) 小野正弘(編). 2007. 日本語オノマトペ辞典, 小学館
- 3) 飛田良文・浅田秀子. 2007. 現代擬音語擬態語用法辞典, 東京堂出版
- 4) 奥村敦史・齋藤豪・奥村学. 2003. Web上のテキストコーパスを利用したオノマトペ概念辞書の自動構築 情報処理学会自然言語処理研究会 2003-NL-154-10, pp.63-70