

オノマトペの多義性解消における名詞の有効性

†福島 弘識, ‡内田ゆず, †荒木健治

†北海道大学大学院情報科学研究科, ‡北海学園大学

†{Fukushima.Hironori,araki}@ist.hokudai.ac.jp, ‡yuzu@hgu.jp

概要 ブログや SNS には話し言葉による投稿が多く、オノマトペが多数存在する。それらオノマトペは、日本語に多く存在し、自然言語処理のタスクにおいても無視できない。特に複数の意味を持つオノマトペは、文中で使われている意味を判別する必要があるが、機械翻訳などのタスクで問題となる。オノマトペの多義性解消にはオノマトペの係り先動詞が有効であることが確認されているが、ブログや SNS には係り先動詞がない文が存在したり、係り先動詞だけでは判別できない文も存在する。そこで本稿では、一語で擬音語と擬態語の 2 つの意味を持つオノマトペが文中においてどちらの意味で使われているか自動的に判別する手がかりとして、名詞の有効性を検証した。オノマトペの意味を決定している名詞を自動で文から抽出する手法を構築し、抽出した名詞を使用して、9 種類のオノマトペの多義性の解消を行った。オノマトペの意味を決定する名詞の自動抽出手法の精度および、文中でのオノマトペが擬音語か擬態語かの判別精度とともに、8割を大きく超える精度を達成し、オノマトペの多義性解消における名詞の有効性が確認された。

キーワード 自然言語処理, 意味解析, 機械翻訳

1 はじめに

近年、外国人観光客数や外国人労働者数の増加のニュースを耳にすることが多い。さらに 2020 年東京五輪の開催も決定し、日本における異文化間コミュニケーションの増加は不可避である。また、Twitter¹や Facebook²などの SNS の普及に伴い、異なった母国語を持つもの同士の Web 上での交流や、情報交換を行う機会も増大している。それらを補助するために、正確な機械翻訳ツールは大きな役割を果たすと考えられる。しかし、現状では正確な機械翻訳が常に可能であるとは言い難い。特に日本語のオノマトペは 1 つの単語で擬音語としてだけでなく擬態語としても用いられ、擬態語の中でも複数の意味を持つものがあり、それらの意味を自動的に判別し、翻訳することは困難である。

先行研究[1]では、オノマトペを含む文から名詞、形容詞、動詞を抽出して素性とし、機械学習を使用してオノマトペの多義性を解消する手法を提案し、係り先動詞と名詞を素性とすることで、一定の成果を得ている。機械学習はオノマトペの多義性解消に有効な手法の 1 つといえるが、精度が訓練データの規模に依存してしまうという問題がある。オノマトペが使用される場面は小説から SNS、ブログまで多岐にわたるうえ[2]、話し言葉にも多く出現する。それらのオノマトペと共起する名詞を、オノマトペの種類ごとに網羅し、素性とするのは困難である。

そこで本研究では、日本語母語話者が同一単語で擬音語と擬態語の 2 つの意味を持つオノマトペの意味が、文中でどちらの意味で使用されているかを判別する際に、オノマトペが含まれる文中のどの単語に着目するか調査を行った。その調査を基にルールを作成し、9 種類のオノマトペについて多義性解消に必要な名詞を抽出した。抽出された名詞を解析し、名詞の持つどのような性質がオノマトペの多義性解消に有効であるかの検討を行った。

2 予備調査

2.1 多義性解消に有効な単語の調査

オノマトペの多義性解消を行うルールを作成するため、日本語母語話者が、オノマトペを含む文のどの単語に着目し、意味を決定しているのか調査を行った。本実験では、「外国人のための基本語用例辞典[3]」の擬音語、擬態語の項目に含まれ、その中で擬音語としても擬態語としても使用されるオノマトペを対象とした。用法で「～の音」と説明されているものを擬音語としての用法、「～の音」以外で説明されているものを擬態語としての用法とし、その両方の用法をもつ全 9 種類をピックアップした。

その 9 種類のうち、Google³ 検索結果数が最も多かった、「ガタン」についてアメーバブログコーパスを基に作成された YACIS コーパス[4]および Yahoo!リアルタイム検索⁴から抽出した文 192 文について、調査を行った。

Copyright is held by the author(s).

The article has been published without reviewing.

1) <https://twitter.com/> 2) <http://www.facebook.com>

3) www.google.com 4) <http://www.yahoo.co.jp/>

調査は20代理系男子大学院生2名、20代社会人女性1名を被験者とした。まず、オノマトペが含まれる1文を MeCab⁵ を使用して形態素解析を行い、単語単位に分割した。その後、それぞれの文において“ガタン”が擬音語として使用されているか擬態語として使用されているかを判断する単語がどの部分かマークしていただいた。3人中2人以上が同じ単語にマークしていれば、その単語をオノマトペの意味決定単語とした。調査例を図1に示す。

例文：運動/量/が/ガタン/と/落ちる/後半
結果：2人以上が量をマークした→“量”が必要な単語

図1 調査例(“ガタン”が擬態語として使用される例)

2.2 調査結果の分析

日本語母語話者が、“ガタン”が文中で擬音語、擬態語のどちらで使用されているか判断するために、感覚的に抽出する単語にどのような共通点があるか分析を行った。先行研究によりオノマトペの多義性判別には、オノマトペの係り先動詞が有効であることがわかっている[5][6]。したがって、オノマトペ意味決定単語とオノマトペの係り先が一致しているかどうかについての分析を行った。結果を表1に示す。

結果から考察を行うと、オノマトペの係り先が名詞であった文については、オノマトペの係り先が9割近く意味決定単語となっているが、オノマトペの係り先が動詞であった文については、3割程度にとどまっている。

オノマトペの係り先が動詞だが、意味決定単語とならなかった文には以下のような2つのパターンが存在した。

I：オノマトペの係り先動詞が、オノマトペの使用意味に関わらず、どちらの意味でも出現する場合。

- 例 (1)イスがガタンと落ちる仕掛けのアトラクション
(2)昨日はアクセスがガタンと落ちましたね

II：オノマトペを含む文に“音”という単語が存在する

- 例 (1)洗濯機と乾燥機のたてるガタンという音
(2)あの日の朝、ガタンという音で目が覚めた

また擬音語か擬態語か判別できなかった5文を除く187文のうち、オノマトペの係り先以外が意味決定単語にな

表1 オノマトペ“ガタン”の係り先品詞と抽出単語の関係

	オノマトペの係り先		
	動詞	名詞	その他
対象文数	164	18	5
抽出単語	52	16	0

注)擬音語か擬態語か判別不明な5文を除く(単位:文)

ったものは、119文あり、その全ての単語が名詞であった。これらの結果から、日本語母語話者は、オノマトペの意味を判別する際にオノマトペの係り先動詞だけでなく、オノマトペの係り先名詞、さらには“音”といった特定の名詞により総合的にオノマトペの使用意味を判断していると断定できる。

3 実験準備

3.1 対象オノマトペの決定

本研究で対象とするオノマトペは、2.1で述べた「外国人のための基本語用例辞典」の擬音語、擬態語に含まれる語の中で「～の音」として意味を定義され、かつ、それ以外に「～の音」以外の擬態語の意味も定義されている9種類を調査対象とした。9種類は、“ガサガサ”、“ガタガタ”、“ガタン”、“ガンガン”、“ポンポン”、“バタバタ”、“ピイピイ”、“ビリビリ”、“コツコツ”である。

3.2 動詞のみで意味が判別できる場合

オノマトペの多義性の解消が動詞のみで行われる場合があることは、本稿2.1及び2.2の予備調査で裏付けられている。予備調査では“ガタン”を例に調査を行っているが、オノマトペが擬音語か擬態語か判別する際に、オノマトペの係り先動詞を感覚的に抽出している場合が、164文中、52文存在している。したがって、オノマトペと動詞の組み合わせだけで、オノマトペの意味を判別できるような動詞が存在する。そのような動詞が文中にオノマトペの係り先動詞として存在していれば、その他の単語を考慮せずに意味を判別できるため、判別は容易である^{注1}。したがって、本実験では、オノマトペと動詞だけによって意味が決定されるような動詞が、オノマトペの係り先となっている場合は実験対象としないこととした。

そのために、9種類のオノマトペそれぞれについて、対象としない動詞をリスト化する。

まず、京都大学格フレーム[7]を用いて、オノマトペと関連のある用言を検索し、その中で出現頻度が2以上の動詞をピックアップした。

つぎにピックアップした動詞とオノマトペをペアとし、そのペアのみでオノマトペが擬音語か擬態語か判別できるか否か、20代理系男子大学院生1名、20代社会人女性1名、40代社会人男性1名に調査を行った。調査方法は5段階のリッカート尺度を用いて行う。例えば、(ガタン,鳴る)が擬音語で用いられると考える場合は1、どちらかといえば擬音語で用いられると考える場合は2、といった具合である。5に近づくほど、オノマトペが擬態語として用いられ、1に近づくほど擬音語として用いられるイメージが強くなる。3人の調査結果の平均をscoreとして、以下のように定義した。

5)MeCab:
Yet Another Part-of-Speech and Morphological Analyzer

注1)したがって、動詞のみで決定されるオノマトペは、容易に自動で判別することが可能である。

- ・動詞だけで擬音語と判断する動詞($1 < score \leq 2$)
- ・動詞だけで擬態語と判断する動詞($4 < score \leq 5$)

3.3 オノマトペの意味決定名詞抽出法の構築

オノマトペの多義性を解消するにあたり、必要な名詞を抽出する手法を構築する。

本手法では、予備調査の結果をもとに、5つのパターンを作成し、図2に示す。ここで、抽出を試みる名詞をN、それ以外の全ての名詞をn、オノマトペをO、動詞をVとする。また“→”は係り先関係を示す。ただし、名詞Nは、MeCabに使用されているIPA品詞体系⁶の名詞の中で、予備調査のオノマトペ意味決定単語の名詞2語以上で出現した、“一般”、“サ変接続”、“形容動詞語幹”、に加え、それらが接尾辞として使用される“接尾-一般”、“接尾-サ変接続”、“接尾-形容動詞語幹”とした。

図2に示す5つのパターンを用いて、名詞を抽出す

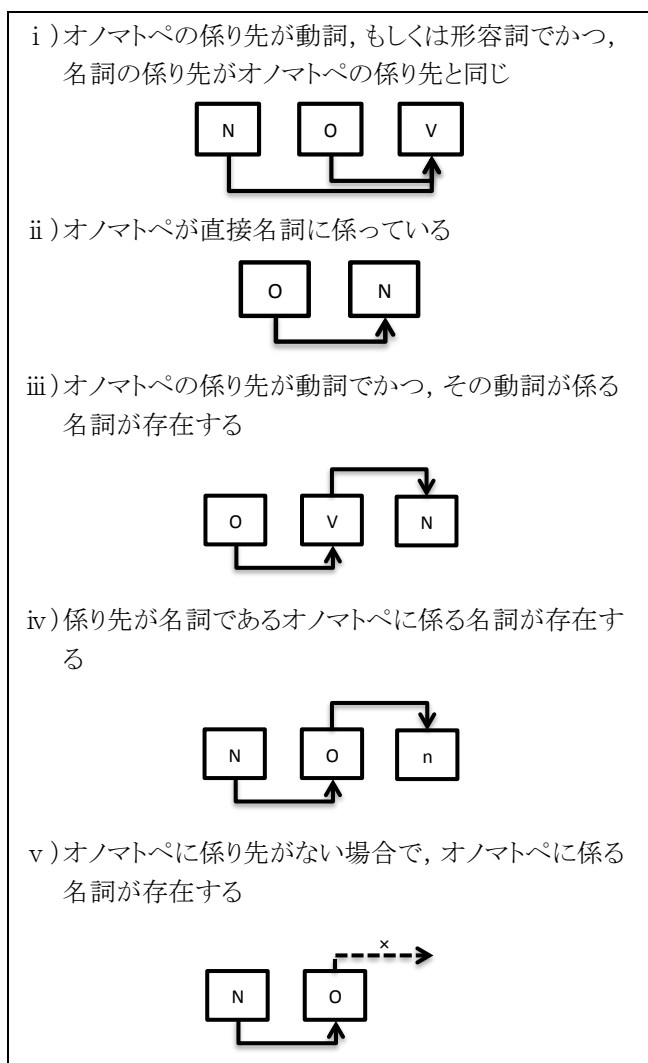


図2 多義性解消に必要な名詞を抽出するパターン

6) 参照 ipadic version 2.7.0 ユーザーズマニュアル(浅原, 松本)

<http://chasen.naist.jp/snapshot/ipadic/ipadic/doc/ipadic-ja.pdf>

7) CaboCha: Yet Another Japanese Dependency Structure Analyzer

<http://chasen.org/taku/software/cabochoa>

る。係り受け解析には CaboCha⁷を用いた。

しかし、1文あたりに存在する名詞の個数は1つとは限らないため、 $i \sim v$ に該当する名詞が複数存在することも考えられる。そこで抽出した名詞が複数だった場合、優先順位をつけることとした。まず、予備調査の結果から、名詞を意味決定単語としたもののうち、約75%がパターン*i*に当てはまることがわかった。また、オノマトペの係り先が名詞である18文のうち16文でその名詞が抽出単語となっていることから、パターン*ii*も同様に重要である。また、オノマトペに係り先がないパターン*v*の場合は、パターン*i \sim iv*と同時に起こり得ないので、重要度としては*i*および*ii*と同じとした。パターン*iii*とパターン*iv*は、パターン*i*とパターン*ii*の合計に比べ、予備調査における出現頻度が1割程度と低かったため、重要度を下げた。

したがって、複数の名詞が抽出された場合は、以下のように重要度を定め、重要度が低い名詞は抽出しない。

$$\text{重要度} : i = ii = v > iii = iv$$

続いて、重要度が同一のパターンの名詞が複数存在する場合の優先度を決定する。その際、以下の式(1)、式(2)で計算される2つのスコアを用いて優先度を決定する。文を文節ごとに区切り、最初の文節から順に割り当てた整数をチャンク番号とすると、 N_{pos} は対象名詞のチャンク番号で、 K_{pos} はオノマトペの係り先のチャンク番号、 O_{pos} は、オノマトペのチャンク番号を示す。

$$D_{on} = |N_{pos} - O_{pos}| \quad (1)$$

$$D_{kn} = |N_{pos} - K_{pos}| \quad (2)$$

式(1)は、対象名詞とオノマトペそれぞれが含まれるチャンク番号の差を、式(2)は対象名詞と、オノマトペの係り先のチャンク番号の差をそれぞれ絶対値で示している。予備調査を分析した結果、オノマトペの意味決定名詞と、オノマトペそのものとの距離 D_{on} は、平均1.4となり、極めて距離が近いことが明らかとなった。また、どのパターンにも対象名詞とオノマトペは存在することから、同一パターンの名詞が複数存在した場合は、 D_{on} の距離が短い方を優先することとした。

パターンが同一となり、 D_{on} も等しくなった場合は、 D_{kn} についても算出し、同様に距離が短い名詞を優先して抽出する。 D_{kn} のスコアも等しくなった場合は、判定する要素が極めて乏しくなるため、抽出不可とした。

3.4 オノマトペを含む文の収集

オノマトペを含む文の収集を行った。本実験では、収集元として、2.1で用いた YACIS コーパスと Twitter を用いた。ブログは、性質上、口語表現が多くオノマトペが出現しやすいと考えられる。また、Twitterはブログよりさらにカジュアルな表現が用いられたり、短文での投稿

が多く、動詞の省略されたツイートが頻繁に出現する。このため、多義性解消における、名詞の有効性の確認に適していると判断し、用いた。

3.1 で述べた 9 種類のおノマトペについてそれぞれブログコーパスから 25 文、Twitter から 25 文の計 50 文を収集し、全 450 文のおノマトペデータベースを構築した。ただし、おノマトペ“パイパイ”については、ブログコーパスに出現した文が 18 文のみだったため、Twitter から 32 文を収集した。

また、文に含まれるおノマトペの係り先動詞が、3.2 で構築した動詞のリストに含まれ、おノマトペと動詞のみで判別できると想定される文や、係り受け解析に明らかに失敗し、3.3 で構築した手法でどのパターンにも該当せず名詞が抽出されない文についてはあらかじめ、取り除いて収集した。

4 おノマトペの意味決定名詞の抽出

4.1 抽出結果および精度

3.2 および 3.3 で述べた手法により、450 文全てからおノマトペの意味決定名詞の抽出を行った。日本語母語話者が、おノマトペの意味を判断するために抽出する名詞と、構築したルールにより、システムが自動抽出した名詞がどれくらい一致しているかの確認を行った。一致しているか否かは、日本語母語話者である 20 代社会人女性 1 名に回答していただいた。結果として、9 種類のおノマトペの意味決定名詞の抽出の平均精度は、86.2%という高い正解率が得られた。特に、“ガタン”、“ガタガタ”、“ガンガン”、“ポンポン”、“コツコツ”については 9 割を超えており、高確率でおノマトペの意味決定名詞を抽出できている。表 2 に意味決定名詞の抽出精度をおノマトペ別に示す。

4.2 抽出誤りの分析

抽出する名詞を誤った場合について例を挙げて説明する。誤ったのは大きく分けて次の 2 つのパターンである。

表 2 おノマトペ別意味決定名詞の抽出精度

	正解(文)	不正解(文)	精度(%)
ガタン	48	2	96.0
ガタガタ	46	4	92.0
ガンガン	45	5	90.0
ポンポン	45	5	90.0
コツコツ	45	5	90.0
バタバタ	43	7	86.0
ビリビリ	42	8	84.0
ガサガサ	41	9	82.0
パイパイ	33	17	66.0
全体	388	62	86.2

- ① おノマトペの係り先動詞のみでおノマトペの意味が決定されるが、係り先動詞が動詞リストに含まれておらず名詞を抽出してしまう誤り。

以下の例ではおノマトペの意味は下線部の動詞によって決定されるが、名詞を抽出している。

例) 午後もバタバタと空を飛べるほど動き回るのであります。(抽出された名詞「空」)

- ② 抽出される名詞を誤った場合

例) 背中をポンポンしていつてくれた言葉
(抽出された名詞「言葉」)

これらの誤りの原因について考察すると、①の例に関しては、係り先動詞のみでおノマトペの意味を決定する際の動詞リストの種類が少なかったためである。全ての動詞を網羅することは非常に困難だが、“パイパイ”には動詞リストに 1 種類しかないため、おノマトペの意味決定名詞抽出の際の精度もかなり低くなってしまったと考えられる。続いて②の例に関しては、係り受け解析の誤りである。Twitter やブログには、助詞の欠落やひらがなの多用など係り受け解析を困難とする様々な要因がある。しかし、それが特色でもあるため、本実験では、係り受け解析に失敗する可能性が高い、文中への顔文字挿入や、句読点がなく文が続くもの、3.3 で構築したルールを使用しても名詞が全く抽出できないもの以外は、文に対して処理を施していない。したがって例②のような誤りも発生してしまった。

5 意味決定名詞を用いた多義性判別法の提案

5.1 手法の提案

本手法では、JUMAN⁸ の意味情報を用いる。JUMAN の意味情報では、カテゴリ情報 (22 種類) とドメイン情報 (12 種類) が与えられている。本実験では、このうちカテゴリ情報を使用して分類を行う。抽出された名詞に付与されたカテゴリ情報をもとに、おノマトペを含む文を仮説によって分類し、擬音語と擬態語に分類できているか精度を測定する。

分類の精度は、日本語を母語とする 20 代理系男子大学院生 1 名、20 代社会人女性 1 名、40 代社会人男性 1 名に判定していただき、2 人以上が正解と判断した場合は正解とする。

5.2 仮説の提唱と検証

予備実験で扱った“ガタン”について抽出された名詞

8) 日本語形態素解析システム JUMAN
<http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN>

の分析を行った。“ガタン”を含む文 192 文については、先行研究[1]で予め“重いものが落ちたりぶつかったりする音やその様子”と“成績、値段などの物事が急激に落ちる様子”という2つの意味に分類されており、それぞれラベルが付与されている。このうち、前者を擬音語として定義し、後者を擬態語とする。

2.1 に記述した予備調査で、被験者が多義性解消に必要であると判断し、抽出した名詞に JUMAN の意味情報を付与し分析した結果を表 3 に示す。表 3 において、抽象物のカテゴリに含まれる「音」とは、抽出単語に「音」という文字が含まれる場合のことを示す。

分析結果として、擬音語は金銭を除く、「人工物」や「場所」といったカテゴリに含まれる名詞が抽出され、擬態語は、「抽象物」や「数量」といったカテゴリに含まれる名詞が抽出されていた。しかし、「抽象物」の中でも、抽出名詞に“音”が含まれるものについては、全て擬音語であった。これらから、以下の2つの仮説を提唱する。

- I： 擬音語と擬態語の分類は基本的に、オノマトペの意味を決定する名詞が具体物か、抽象物かに依存する。
- II： 抽出した名詞に“音(カテゴリは抽象物)”が含まれる場合については、擬音語となる。

以上の2つの仮説から、まず、JUMAN に与えられている22種類のカテゴリ情報を具体物と抽象物に分類した。抽象物として分類したのは、「抽象物」「色」「数量」「時間」「形・模様」「人工物-金銭」⁸である。それ以外は具体物とした。

また、抽出した単語そのものにカテゴリが付与されていなくても、その語に上位語があれば、上位語のカテゴリを付与した。

表 3 “ガタン”を含む文の抽出単語のカテゴリ情報

		擬音語	擬態語
抽象物	音	36	0
	音以外	0	45
人工物	乗り物	2	0
	金銭	0	3
	その他	5	3
	場所	4	0
数量		0	21
時間		0	1
自然物		0	1
動物		0	1
カテゴリなし		0	7

(単位:文)

例えば、「ストッキング」はカテゴリが付与されていないが、上位語として、「長靴下」が出力された。その場合は、「長靴下」を JUMAN によりもう一度解析し、カテゴリ「人工物-衣類」を得た。

これら2つの仮説を検証するため、9種類のオノマトペについて、4.1 で意味決定名詞が正しく抽出された388文を対象として実験を行った。388文のうち、カテゴリが付与されなかった12文を除く計376文について、自動で擬音語か擬態語かを判別した。その後、それらが正解か不正解かを5.1 で述べた被験者に判定していただいた。

5.3 意味判別の結果と考察

判別結果を以下の表 4 に示す。オノマトペの意味判別結果は、全450文の平均値をとると、71.5%の高い精度が得られた。特に、“コツコツ”や“ガタン”は9割を超えており非常に高い精度であった。一方で、“ビリビリ”は5割に満たないという結果になり精度が低い。そこで、最も精度が低くなってしまったオノマトペである、“ビリビリ”において判別結果が不正解となった名詞に付与されたカテゴリの割合から、原因の分析を行った。図 3 に、不正解となった文から抽出されている意味決定名詞に付与されているカテゴリの種類の存在割合を示す。

図 3 を分析すると、誤っている文から抽出された名詞に多いカテゴリは、「動物-部位」である。この誤りがあった文には次のようなものがある。

例)寝ようとすると皮膚がビリビリする

ここでの“ビリビリ”は被験者は擬態語と判断しており、痛みを表す擬態語だと推測できる。皮膚や手足など動物の部位は具体物であるが、痛みを表す際は擬態語となるため誤りが生じた。

表 4 オノマトペ別判別精度

	正解(文)	不正解(文)	正解率(%)
コツコツ	43	1	97.7
ガタン	43	4	91.5
バタバタ	33	7	82.5
ガンガン	33	11	75.0
ガタガタ	32	13	71.1
ポンポン	26	18	59.1
ガサガサ	23	17	57.5
ピイピイ	17	14	54.8
ビリビリ	19	22	46.3
全450文	269	107	71.5

⁸JUMAN に与えられているカテゴリについては、JUMAN のチュートリアルを参照した。(http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN)

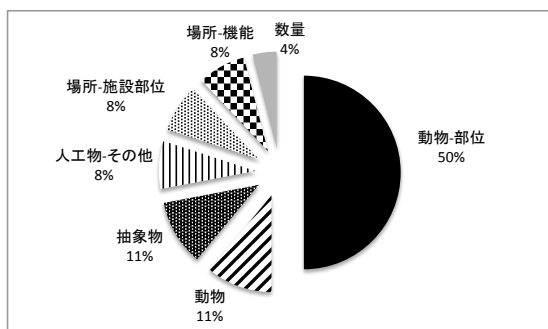


図3 誤って付与したカテゴリの割合<“ビリビリ”>

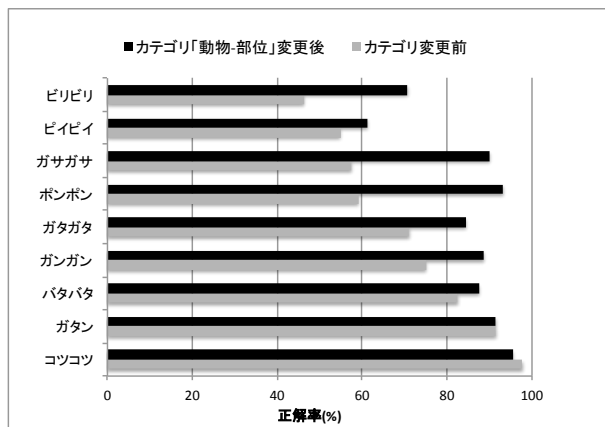


図4 「動物-部位」カテゴリを変更後の正解率

また「動物-部位」のカテゴリを付与された名詞が抽出されている“ビリビリ”を含む文は全て擬態語と判断されていた。その他のオノマトペにおいて、「動物-部位」のカテゴリが付与された意味決定名詞を抽出した場合は、全体の91.0%が擬態語で使用されていた。したがって、カテゴリ「動物-部位」が付与された名詞が抽出された時は、擬態語として、再度正解率を測定した。結果を図4に示す。

カテゴリのグループを変更したところ、9種類の平均正解率は、85.6%となり、14.1ポイント上昇した。また、“ガタン”を除く全ての種類のオノマトペで正解率が上昇した。

この結果から、オノマトペの擬音語、擬態語の分類は基本的に抽象物と具体物に依存するが、動物の部位がオノマトペの意味判別決定名詞となっているときは、擬態語に分類する方が精度が高くなることが確認された。

また、2つめの仮説は、「音」を含む全36文のうち、35文が擬音語であったことから、妥当であるといえる。

6 まとめ

本研究では、オノマトペの多義性解消のためにどのような名詞が有効であるか、そしてどのくらい有効であるのかについての調査を行った。日本語母語話者は、オノマトペの意味を周辺の単語や文脈によって判断していると考え、アンケート調査によりオノマトペを含む文のどの単語によって、多義を持つオノマトペの意味を判別し

ているのかを分析した。その分析に基づき、オノマトペの多義性解消に有効な名詞である、オノマトペの意味を決定している名詞を自動で抽出する手法を提案した。オノマトペの係り先やオノマトペに係る名詞に注目した、単純な手法ではあるが意味を決定する名詞の自動抽出は85%を超える精度を達成した。

その手法により、9種類のオノマトペについて、分類を行った。その結果、オノマトペの多義性の解消には、意味を決定している名詞が、具体物か抽象物のどちらであるかということや、「音」という単語を含むか否かが非常に関係していることが明らかとなった。さらには、具体物の中でもオノマトペの意味を決定する名詞が、動物の部位を表す場合は、擬態語であることが多いと明らかになった。これらの知見を利用して意味判別を行ったところ、85%を超える正解率を達成した。

本研究で扱った、擬音語と擬態語の2種類の意味ではなく、擬態語だけで複数の意味を持つオノマトペも存在することから、今後はそれらの多義性の解消へのアプローチを研究する必要がある。また、オノマトペはそれぞれの語によって、意味判別の精度にばらつきがあることからわかるように、1つ1つが異なる特徴を持っている。それらの特徴を分析し、さらに意味判別の精度を向上し、高精度で自動的に、全てのオノマトペの多義性の解消を行うシステムの構築につなげたいと考えている。

参考文献

- [1] Hironori Fukushima, Kenji Araki and Yuzu Uchida: Disambiguation of Japanese Onomatopoeias Using Nouns and Verbs, P.Sojka et al. (Eds.), TSD2014, LNAI 8655, pp.141-149, 2014.
- [2] 内田ゆず, 荒木健治, 米山淳: ブログ記事から抽出したオノマトペの多義性について, 第27回ファジィシステムシンポジウム講演論文集, pp.853-856, 2011.
- [3] 文化庁, 外国人のための基本語用例辞典(第三版), 大蔵省印刷局, 1990.
- [4] Michal Ptaszynski, Pawel Dybala, Rafal Rzepka, Kenji Araki and Yoshio Momouchi: YACIS: A Five-Billion-Word Corpus of Japanese Blogs Fully Annotated with Syntactic and Affective Information, in Rzepka R., Ptaszynski, M. and Dybala, P. (Eds.), Linguistic and Cognitive Approaches To Dialogue Agents (AISB/IACAP Symposium), 40-49, 2012.
- [5] 浅賀千里, ユスフムカルラマー, 渡辺知恵美: オノマトペ用例辞典における用例を意味により分類するためのクラスタリング手法の諸検討, 日本データベース学会 Letters Vol.6, No.2, pp.45-48, 2007.
- [6] Chisato Asaga, Mukarramah Yusuf, Chiemi Watanabe: Onomatopedia: Onomatopoeia Online Example Dictionary System from Data on the Web, The 10th Asia Pacific Web Conference(APWeb), 2008.
- [7] 河原大輔, 黒橋禎夫: 自動構築した大規模格フレームに基づく構文・格解析の統合的確率モデル, 自然言語処理, Vol.14, No.4, pp.67-81, 2007.