2つの意味をもつオノマトペの意味判別における素性の検討

[1]福島　弘識，[2]内田　ゆず，[3]荒木　健治，
[1]北海道大学工学部情報エレクトロニクス系　[2]青山学院大学理工学部
[3]北海道大学大学院情報科学研究科

1　はじめに

日本政府観光局が発表している，年別訪日外国人客数は，年ごとに多少の増減はあるものの，長期的にみると増加傾向にある．グローバル化は益々進むと考えられ，日本語を第一言語とする人々とそれ以外の人々がコミュニケーションを取り合う機会の増加も予測される．日本語でコミュニケーションをとる上で，“オノマトペ”は必要不可欠である．感覚に依存する部分も多く，日本語を第一言語としない人に，“オノマトペ”がもつ意味を明確に伝えることは非常に困難なことである．

その中でも多義をもつオノマトペの意味の伝達は，非常に困難である．しかし，複数の意味をもつオノマトペが文中でどの意味で使用されるか判別ができれば，人間同士の会話はもちろん，機械の言語理解や，翻訳など多様な場面で有益である．

古武ら[1]の研究では，オノマトペの言い換え表現を自動収集する方法を提案している．オノマトペには，典型的に修飾する動詞が存在することから，オノマトペが修飾する動詞のみに着目して研究を行っている．最終的に，適切な言い換えができた割合が80.6%となりオノマトペが修飾する動詞は，オノマトペの意味判別に重要な品詞である，と考えられる．

また，内田ら[2]の研究では，ブログ記事から抽出したオノマトペの多義性についての分析をオノマトペの係り先に注目して行っている．分析の結果として，オノマトペが係る単語の品詞は動詞が最も多く，続いて，名詞，形容詞となっていることが言及されている．

そこで，本研究では，多義をもつオノマトペを含む文から名詞，形容詞，動詞を抽出し，それらの組み合わせによって，意味判別の精度がどの程度変化するのかを考察する．多義をもつオノマトペには意味が2つのものや3つのもの，あるいはそれ以上のものが存在する．オノマトペのもつ意味が増えるほど，意味の選択肢が増えるため判別が困難となる．そこで本稿では，3つ以上の意味をもつオノマトペは今後の課題とし，2つの意味をもつオノマトペを調査対象とする．

2　提案手法の概要

調査対象のオノマトペを決定し，そのオノマトペを含む文をアメーバブログ[3]から収集する．

調査対象オノマトペそれぞれについて，10人の被験者に意味を判別してもらい，8人が一致した意味をその文におけるオノマトペの意味とする．

続いて，1文ごとにMeCab[4]を使用して形態素解析を行い，名詞，形容詞，動詞を抽出した．

抽出した品詞を組み合わせて素性とし，SVM-Light[5]を用いて機械学習を行い，どの程度の意味判別が可能か精度を測定し，考察を行う．

3　実験準備

3.1　調査対象オノマトペ

まず「外国人のための基本語用例辞典[6]」の中の擬音語，擬態語の中から，複数の意味をもつものをピックアップした．それらは，使用される意味によって品詞が異なるか否かによって，2種類に分類される．使用される意味が品詞で区別されるものに関しては，オノマトペ自体の品詞を判別すればよく，意味の判別が難しいオノマトペであるとは言い難い．よって本研究では，意味が判別しにくいオノマトペは，“文中でオノマトペ自体が同じ品詞で使用され，かつ，その意味が異なるもの”と仮定する．意味が判別しにくいと仮定したオノマトペの例として“ばりばり”を表1に示す．“ばりばり”を副詞として使用する場合，以下の2つの意味がある．

意味①いきおいよく紙などをやぶるようす。
-彼は書きかけの手紙をばりばりやぶった。
意味②いきおいよく仕事をして能率をあげるようす。
-休みが終わったら，ばりばり働くぞ。

「外国人のための日本語用例辞典」の中で，この仮定にあてはまり，かつ意味が2つのものは21語存在した．

本研究では，日常的に広く使用されているオノマトペを調査すべき，という考えからGoogle検索[7]の検索ヒット件数の多い上位2語を調査対象とした．検索結果より，意味が2つのオノマトペからは，“がたんと”と“さらりと”を，調査対象とすることとした．“がたんと”および“さらりと”は

“オノマトペ+助詞(と)”の形であるが，どちらも“多く「と」を伴って”と記述されている(大辞林第三版[8]より)ことや，形態素解析での誤りを減少させるため，“がたんと”および“さらりと”のまま調査を行う．

3.2 オノマトペがもつ意味の定義

調査の対象とした多義をもつオノマトペ“がたんと”，“さらりと”がそれぞれどのような意味をもつのかをあらかじめ定義しておく必要がある．「外国人のための基本語用例辞典」の意味，用法の分類だけではどこに分類すればいいのか分からない文が存在したため，“さらりと”，“がたんと”，ともに独自に意味表現を定義した．

定義した“さらりと”の意味は以下の 2 つである．

意味①感触（手で触ってみたり，肌で感じた様子），味わい．

-さらりとした風が吹く．

意味②惜しいと思わない気持ち，こだわりがなく思い切りのよい様子．

-さらりと言い放った．

定義した“がたんと”の意味は以下の 2 つである．

意味①重いものが落ちたりぶつかったりする音やその様子．

-本ががたんと棚から落ちた．

意味②成績，値段などの物事が急激に落ちる様子．

-売り上げががたんと落ちた．

3.3 オノマトペを含む文の収集

日常的かつ口語的な表現が多く意味の判別が難しいオノマトペが多数存在すると考えたため，アメーバブログからオノマトペを含む文を収集した．それぞれ調査対象とするオノマトペを含む文を取り出し，1 文ごとに第一著者が手作業で意味を判定していく．

ブログにはノイズとなる文が存在するため，著者の手作業で判定し，取り除く必要がある．その例を以下の表 1 に示す．例 1 のように，そのオノマトペ単体で文として成り立っているようなものは，解析不能として取り除く．続いて，例 2 に示すように，著者が読んで一般的に意味が分からないであろう，もしくは，オノマトペと定義できない，と推測する文は取り除く．次に，例 3 に示すように，オノマトペそのものが，固有名詞として使用されている場合は取り除く．固有名詞であるかどうかの判定は著者が判断した．最後に，例 4 に示すように，オノマトペを構成する文字列の要素がそのオノマトペの直前，あるいは直後で繰り返されているものは取り除く．この規則に従って，意味 1 つにつき 100 文を目安に収集を行う．データの量が不十分な場合は，Yahoo!リアルタイム検索[8]において文を収集した．

表1　調査しないと判定した文の例

例1	がたんと。
	さらり、さらりと。
例2	生で見れるなんて、さすがたんとんさん。
例3	さらりとした梅酒
例4	おとなりの空き室からがたんがたんと戸が開く音

3.4 正解データの作成

機械学習装置(SVM-Light)に与えるための正解データの収集を行った．本研究では著者が手作業で収集した文中のオノマトペがどの意味で扱われているのかを日本語を母国語とする 10 人に回答していただき，8 人以上が同じ答えを示せば，その意味を正解データとした．

“さらりと”については，10 代女性 1 人，20 代女性 3 人，30 代女性 1 人，20 代男性 4 人，30 代男性 1 人の計 10 人に回答していただいた．“がたんと”については，20 代女性 4 人，30 代女性 1 人，40 代女性 1 人，20 代男性 3 人，30 代男性 1 人の別の計 10 人に回答していただいた．

データを収集した結果を以下の表 2，表 3 に示す．なお，単位は文である．この表において，収集件数は，著者がオノマトペの意味を人手で判別する前の文の数であり，著者判別件数とは，著者が人手で判別を行い，調査に使用することに決めた文の数，使用データ件数は，10 人の調査のうち 8 人以上が同じ意味を回答した文の数である．

表2　“さらりと”に関する調査結果

	収集件数	著者判別件数		正解データ件数
		アメーバコーパス	Yahoo!リアルタイム検索	
意味1	431	78	0	70
意味2		124	0	119

表3　“がたんと”に関する調査結果

	収集件数	著者判別件数		正解データ件数
		アメーバコーパス	Yahoo!リアルタイム検索	
意味1	379	96	4	98
意味2		66	34	94

調査を行った 10 人のうち 3 人以上の意味解釈が異なった文は，“さらりと”に関しては 202 文中 13 文，およそ 6.4%，“がたんと”に関しては 200 文中 8 文の 4%であった．このことから，多義をもつオノマトペの意味判別は日本語を母国語とする人間が行っても解釈にばらつきが出ていることが分かる．また，意味の判別において，ばらつきが出た文は，下に示す表 4 のように，3 つの例に分類することができた．例 1 は「どちらの意味でも判別できる場合」，例 2 は日常ではあまり表現しない場合例 3 は，文中の情報量が不足している場合

である．

表4　判別不可能文章例　（　調査対象オノマトペ別）

例1	さらりと	寒いけどパウダーシュガーをさらりとかけた様な
		淡く雪化粧した冬の景色が大好きです。
	がたんと	ちょっと先の方でガタンと倒れた方がいらっしゃいました。
例2	さらりと	彼の声も、さらりとしていて大好きです。
	がたんと	ゆっくりガタンと上昇する。
例3	さらりと	うーん、さらりと言えばさらりとかも。
	がたんと	ガタンと落ちてきたときかな。

4　意味判別実験

4.1　オノマトペの意味判別法

まず，使用する文に MeCab を用いて形態素解析を行い，名詞，形容詞，動詞を抽出し，辞書を作成した．

1 文ごとに，その文中の，辞書中の単語の出現回数を調べ，単語と出現回数を組にしたテキストデータを作成した．そのテキストデータを用いて，オノマトペがどの意味で使用されているのかを分類した．分類の方法は機械学習装置(SVM-Light)を用いた．また，データ量が少ないため，今回の実験では 10 分割交差検定の正解率を比較する．正解率とは，予測結果全体と答えがどの程度一致しているのかを示す値である．2 つの意味をもつオノマトペについて，名詞，動詞，形容詞の全ての組み合わせについてデータを作成し，検定を行った．

この結果を分析し，SVM における素性の選択法を改善し，再度検定を行った．

4.2　SVM を用いた意味判別実験の実施

10 分割交差検定を SVM-Light を用いて行った．今回はそれぞれのオノマトペについて，10 分割交差検定を 12 回行い，それぞれの項目について，値が一番高いものと，一番低いものを除く，10 回の平均値を求めた．求めた値は，正解率，適合率，再現率，F 値である．実験結果を表 5，表 6 に示す．

表5　”さらりと”の検定結果

	正解率	適合率	再現率	F値
名詞のみ	75.66	73.59	96.70	83.52
形容詞のみ	67.06	65.99	95.43	78.07
動詞のみ	79.81	87.20	79.56	83.21
名詞&形容詞	76.07	73.48	97.75	83.86
名詞&動詞	**83.50**	88.74	84.37	86.53
動詞&形容詞	78.09	84.28	80.78	82.30
名詞&形容詞&動詞	**83.66**	88.78	85.16	86.96

表6　”がたんと”の検定結果

	正解率	適合率	再現率	F値
名詞のみ	75.70	73.70	85.41	79.40
形容詞のみ	52.60	51.97	96.03	67.40
動詞のみ	72.59	73.70	85.41	72.96
名詞&形容詞	74.10	72.29	85.11	78.04
名詞&動詞	**80.50**	77.25	88.99	82.69
動詞&形容詞	73.32	73.83	73.94	74.01
名詞&形容詞&動詞	**80.77**	76.82	89.16	82.53

4.3 実験結果の考察

本実験では多義オノマトペの意味の判別の精度が，正解率によって示されている．“さらりと”，“がたんと”ともに意味判別において，正解率が最も高かったのは，表の太字傍線部の名詞と動詞と形容詞を組み合わせたものであり，2 番目に正解率が高かったものは，表の太字のみの名詞と動詞を扱ったときであった．

これら 2 つの結果から，考察すると，名詞と形容詞と動詞の組み合わせと，名詞と動詞を扱ったものの差が非常に小さく，形容詞は正解率の向上には微力であると考えられた．ただ，“さらりと”は，名詞のみの場合と，名詞と形容詞を組み合わせた場合を比較したとき，形容詞を組み合わせた方が正解率が上昇していることから，形容詞は多義オノマトペの意味判別に関して完全に雑音となるのではなく，場合によっては正解率の向上に貢献することもあることが分かる．

これは，形容詞が，調査対象としたオノマトペおよび調査対象オノマトペと共起している動詞と，どのくらい共起しているのかということに関係していると考えられる．

また，名詞のみ，形容詞のみ，動詞のみの正解率を比較したところ，“さらりと”に関しては動詞のみの場合が，“がたんと”に関しては，名詞のみの場合がそれぞれ最も高かった．このことから，オノマトペは，その種類によって，意味判別に使用するべき品詞の重要度が異なると考えられる．

5　意味判別実験法の改善

5.1 改善手法

まず，動詞について考える．関連研究[2]について記述した中にもあるように，オノマトペの意味判別には，調査対象オノマトペが係る動詞が重要であることがわかっている．したがって，オノマトペが係る動詞のみを文中から抽出して，その動詞のみをデータとして使用することとした．ここで係り先の判定には CaboCha[9]を用いた．また，動詞に分類される中でも接尾辞等は除き, Me Cab の出力における品詞分類のうち動詞の中でも自立語のみを抽出する．接尾辞は，「～おります」の，「おり」や「～います」の「い」を含んでしまうからである．

続いて名詞について考える．“さらりと”および“がたんと”を含む文に出現した名詞には「小川」，「ダルビッシュ」といった固有名詞や，数詞や，

「私」といった代名詞も出現した．これらは名詞のカテゴリーに含まれているが，判別に影響するとは考えにくく，除くこととした．また，MeCabの出力における品詞分類のうち，名詞の中でも非自立語に分類されるもの，たとえば「～することになった」の「こと」といったものも，判別に関係なく，除くこととした．したがって，形態素解析をする段階で，今回は，MeCabにおいて品詞IDの36番～39番に限定して抽出する．36番～39番とは，名詞の中で，サ変接続のもの，ナイ形容詞語幹のもの，一般のもの，引用文字列のものである．

このように，素性の選択法を改善し，再度，多義オノマトペ“さらりと”と“がたんと”について10分割交差検定を行った．今回のデータは，係り先動詞のみの場合と，固有名詞や代名詞等を除いた名詞のみについて判定した．

5.2 素性選択法改善後の検定結果

“さらりと”と“がたんと”に関して，固有名詞や代名詞等を除いた名詞のみを素性とした場合と，係り先動詞のみを素性とした場合の２つに関しての調査を行い，名詞の抽出範囲を変える前と後，動詞のみの場合と係り先動詞のみの場合でそれぞれ結果を比較する．比較結果を表7に示す．

表7　オノマトペ別の名詞と動詞の抽出法における正解率の比較

<結果比較:名詞の抽出法>

	“さらりと”	“がたんと”
名詞のみ	75.66	75.70
名詞(固有名詞等を除く)のみ	72.70	80.73
変化	**-2.96**	**5.03**

<結果比較:動詞の抽出法>

	“さらりと”	“がたんと”
動詞のみ	79.81	72.59
係り先動詞のみ	85.23	81.67
変化	**5.42**	**9.08**

5.3　素性選択法改善後の検定結果の考察

まず，名詞の抽出法の変化に注目すると“がたんと”の正解率は，素性選択法改善後の方が上昇しているが，“さらりと”に関しては減少している．

次に動詞の抽出法の変化に注目すると，“がたんと”の方が上昇率は大きいが，どちらも大幅に正解率が向上している．このことから，係り先動詞はオノマトペの意味判別においてオノマトペの種類に関係なく正解率の向上に貢献すると考えられる．

名詞の抽出においては，素性選択法改善前では2つのオノマトペ間で正解率の差は0.04ポイントしかない．しかし素性改善後は，約8ポイントまで広がった．これは，固有名詞等を除くことが，“がたんと”に関しては正解率向上に貢献したが，“さらりと”に関しては悪影響を及ぼしたということである．したがって，名詞の抽出は，オノマトペごとに，抽出すべきものを選択することが，正解率改善への重要な課題である．

6　まとめと今後の課題

本稿では，オノマトペ2語を対象として，複数の意味をもつオノマトペの意味を判別するために，どの品詞が意味判別に重要か調査を行った．まず2つの意味をもつオノマトペを対象として，名詞，形容詞，動詞を抽出する組み合わせを変えながら，意味判別の正解率を機械学習によって測定した．その結果，オノマトペの係り先動詞の抽出は，安定して，正解率が8割に達し，オノマトペの種類に関係なく正解率を向上できることが明らかとなった．また，名詞，形容詞，動詞を全て組み合わせ，抽出品詞とした場合も安定して正解率が8割に達した．

しかし，オノマトペは特異的なものが多いため，多義を判別する際，オノマトペの種類によって，素性を変える必要性があることが明らかとなった．

今後の課題として，オノマトペの種類によって，抽出する品詞の種類を変更したり，オノマトペを含む文に出現する名詞と，そのオノマトペの共起頻度や，出現する名詞と，オノマトペの係り先動詞との共起頻度の高い名詞のみを素性とし，係り先動詞と組み合わせることで，正解率が向上を目指していきたいと考えている．

参考文献

[1] 古武泰樹, 佐藤理史, 駒谷和範, “オノマトペを言い換える表現の自動収集”, 言語処理学会第17回年次大会発表論文集, pp.904-907，2011.

[2] 内田ゆず, 荒木健司, 米山淳, “ブログ記事から抽出したオノマトペ多義性について”,第27回ファジィシステムシンポジウム, pp. 853-856，2011.

[3] http://www.ameba.jp

[4] MeCab: Yet Another Part-of-Speech and Morphological Analyzer :http://MeCab.sourceforge.net/

[5] SVM-Light: Support Vector Machine: http://svmlight.joachims.org

[6] 外国人のための基本語用例辞典第三版著者:文化庁 出版社:大蔵省印刷局，1990.

[7] https://www.google.co.jp/

[8] 大辞林第三版:http://kotobank.jp より検索

[9] http://search.yahoo.co.jp/realtime

[10] CaboCha:YetAnotherJapaneseDependencyStructure Analyzer :http://chasen.org/ taku/software/cabocha