

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 10 日現在

機関番号：32601

研究種目：若手研究(B)

研究期間：2011～2013

課題番号：23700256

研究課題名(和文)感情情報を含んだオノマトペデータベースの構築に関する研究

研究課題名(英文)Construction of Onomatopoeia Database Including Emotional Information

研究代表者

内田 ゆず (Uchida, Yuzu)

青山学院大学・理工学部・助教

研究者番号：80583575

交付決定額(研究期間全体)：(直接経費) 1,300,000円、(間接経費) 390,000円

研究成果の概要(和文)：本研究の成果は、以下の3点である。

1つ目は、web上の文書からオノマトペを含む大量のデータを高い精度で自動抽出する手法を確立した点である。2つ目は、324語のオノマトペをPlutchikの基本感情に基づいて8種の感情カテゴリへの分類した点である。3つ目は、ドメインとオノマトペの関連の強さを、階層的クラスタリング手法を用いた実験で明らかにした点である。これらの成果により、オノマトペ用例文を大量に取得し、評判分析や日本語学習者への支援などに有効なデータベースが構築された。

研究成果の概要(英文)：We have three major results.

First, we have developed a method to extract large quantities of sentences including onomatopoeia words with a high degree of accuracy. Second, we have classified 324 onomatopoeia words into Plutchik's basic 8 emotions. Third, we have evaluated the association between a domain and an onomatopoeia with hierarchical cluster analysis.

These results will be helpful to sentiment analysis and support for Japanese-learners.

研究分野：感性情報学・ソフトコンピューティング

科研費の分科・細目：若手研究(B)

キーワード：オノマトペ 感情分析 クラスタ分析 評判抽出

### 1. 研究開始当初の背景

オノマトペとは、擬音語・擬態語を包括的に指す語であり、日本語に豊富に存在し、日常生活において多用される。オノマトペは豊かな表現力を持ち、話し手の細やかな心情を表したり、様々な事物の様態を生き生きと描写することができる。日本語を母語とする人は、ごく自然にオノマトペの用法を身につけ、「感覚的に」使用する。ほかの言葉に置き換えたり、その意味を明確に説明することはできない。したがって、日本語を学ぶ人がオノマトペの意味・用法を習得することは難しいと言われている。

これまで、オノマトペは言語学や認知科学の分野で盛んに研究されてきたが、近年、情報学、特に自然言語処理の分野でも注目されている。日本語学習者がオノマトペの用法を習得するためには、実際にオノマトペが使用されている場面に多く触れることが有効であるという観点から、オノマトペの用例を提示するシステムの開発を目指した研究について以下で説明する。

香林らは、オノマトペの用例を日本語、英語、中国語、韓国語で表示するオンライン多言語辞書を開発した。この辞書は、小説から手動で用例を抽出する手法で構築されているため、開発に多大なコストがかかる。この問題を解決するために、Asagaらはオノマトペが用いられている文章を Web コーパスから自動抽出し、提示する辞典を開発した。しかし、研究成果として一般に公開されているデータは 80 語余りのオノマトペを対象にしたもので、4,500 語にも達するとされる日本語オノマトペを十分に網羅しているとは言えない。

このように、オノマトペのデータベースに関する研究が興隆しつつあるものの、どれも大規模なデータベースの公開には至っていない。さらに、先行研究では、オノマトペが含まれる文を提示することに力点が置かれ、そのオノマトペが使用された状況や話し手(あるいは書き手)の心理が軽視されてきた。

図1にオノマトペを含むブログ記事の抜粋を例として挙げる。例ではオノマトペを含む文に感情情報も記述されているため「きよときよと」の意味が伝わりやすい。しかし、例ではオノマトペを含む文のみを提示してもオノマトペの意味を伝えることはできない。このような場合は、周囲の文脈も考慮する必要がある。

雪でした。ベルは、きよときよとして、びっくりした様子。  
みんなきゃーきゃー言っていましたね。景品、でかいと素直に嬉しいよね。

図1 オノマトペの使用例

本研究は「オノマトペに含まれる話し手の感情」というこれまでにない新しい要素を考慮した大規模なオノマトペデータベース構築に、自然言語処理の技術を用いて挑戦するものである。

### 2. 研究の目的

本研究は、感情情報を付与したオノマトペデータベースの公開を目的としている。この目的の達成のために、(1)Web上に存在するオノマトペを含むデータを大量に、かつ高精度で自動抽出し、(2)その周辺の感情情報を関連付ける手法を開発しなければならない。したがって、研究期間内に次の二つの研究を段階的に進める。

第一に、Web上の文章を対象としてオノマトペを含む文を抽出する場合、どのような雑音が含まれるのか、その雑音を除去するにはどのような手法を用いるべきなのかを明らかにする。ここでは、主に自然言語処理の技術を応用する。また、本研究で構築するオノマトペデータベースは、一般への公開を前提としているため、データの質の高さが要求される。したがって、再現率よりも適合率を優先する方針をとる。

第二に、オノマトペに含まれる書き手の感情を周辺文脈から判定する手法を明らかにする。Ptaszynskiらによって考案された日本語の感情分析に用いられた感情表現要素に基づき、感情の判定を行う予定である。この感情表現要素は、中村5が日本語の感情を分類した{喜、怒、哀、怖、恥、好、厭、昂、安、驚}の10カテゴリに対応している。

### 3. 研究の方法

まず、形態素解析や構文解析などの自然言語処理の技術を用いてWeb上の文書からオノマトペを含む大量のデータを高い精度で自動抽出する手法を提案する。予備実験により、Web上にはデータベース構築に十分な量のオノマトペの用例が存在することがわかっている。

オノマトペは複数の意味を持つことが多い。例えば、「ぺらぺら」は「外国語などをよどみなく話すさま」や、「軽薄な調子でよくしゃべるさま」、「紙や布などの薄くて弱いさま」、「薄いものがこきざみにゆれたり、ひるがえったりするさま」などの意味を持つ。オノマトペが表す意味を正しく解釈するた

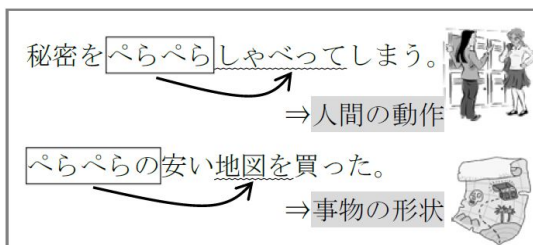


図2 オノマトペの多義性  
めには、複数の意味の使い分けも重要となる。

つまり、オノマトペの用例を「人間の動作」、「事物の動き」、「事物の形状」などのカテゴリに分類してデータベースを構築するべきである。これを実現するためには、構文解析を用いて格関係やオノマトペの係り先を考慮する手法が考えられる。

次に、既存の日本語感情表現辞典と、それを拡張した新たな感情表現を用いて、オノマトペに含まれる書き手の感情を周辺文脈から判定する手法を提案する。上述の方法で収集したデータに感情情報を付与し、オノマトペデータベースを完成させる。日本語を母語とする被験者によってオノマトペと感情の関連度を評価し、日本語を母語としない被験者によって既存の国語辞典との比較実験を行う。

完成したデータベースは、(1)オノマトペと感情情報を適切に関連付けているか、(2)このデータベースによって日本語オノマトペの理解が深まるか、の2つの観点から評価する。

(1)の観点に対する評価方法としては、日本語を母語とする人たちによる主観評価が最も適しているだろう。なぜなら、オノマトペと感情の関連度を、何らかの基準を定めて数値で評価することは難しいためである。具体的には、データベースに含まれるオノマトペの用例を提示し、書き手がどのような感情を抱いているかを自由記述形式、あるいは選択形式で回答するというアンケートを実施する予定である。アンケート結果とデータベース内の感情情報との一致率によって本研究の妥当性を検証する。

(2)の観点に対する評価方法としては、日本語を母語としない被験者による評価が適しているだろう。本研究で得られた感情情報付きのオノマトペ用例と、既存の国語辞書の記述を比較して理解に差があるかどうかを調査する。

#### 4. 研究成果

本研究の成果は、以下の3点である。

(1) web上の文書からオノマトペを含む大量のデータを高い精度で自動抽出する手法を確立した。この手法は、オノマトペの後続要素と係り先の分析から明らかになった、オノマトペを含む文の特徴を利用したものである。41,315件分のブログ記事から15,437文のオノマトペ用例文を抽出し、抽出精度は96.2%であった。大量の文書集合や語彙資源を必要としないシンプルな手法であるが、良好な結果を得た。

(2) 324語のオノマトペをPlutchikの基本感情8種に基づく感情カテゴリに分類する実験を行った。10名の評価者にオノマトペを提示し、分類させ、評価者間の意見がどの程度一致するかを定量的に調査した。その結果、290語のオノマトペが8種の感情カテゴリに分類された。

(3) オノマトペは使用されるドメインによ

って意味が異なることがあるため、ドメイン-オノマトペ間の関係を明らかにする必要がある。ドメインとオノマトペの関連の強さを、階層的クラスタリング手法を用いて評価する実験を行った。この実験では、オノマトペの出現頻度によって商品カテゴリを類型化する。結果の一例として、図3に“やわらかい”ことを表すオノマトペを変数にした場合のデンドログラムを示す。このような手法を用いて、614語のオノマトペを対象として729,865件の商品レビューを分析したところ、658組の商品カテゴリ-オノマトペのペアが抽出された。

これらの成果により、これまで研究されてこなかった実生活に根ざしたオノマトペの用法について、利用可能なデータが得られた。このデータは日本語学習者への支援(日本語教育分野)、発話分析や人とロボットのコミュニケーション支援、評判分析、機械翻訳システムへの応用(自然言語処理分野)、言語学や認知科学研究へのデータ提供など、幅広い分野で活用されることが期待できる。

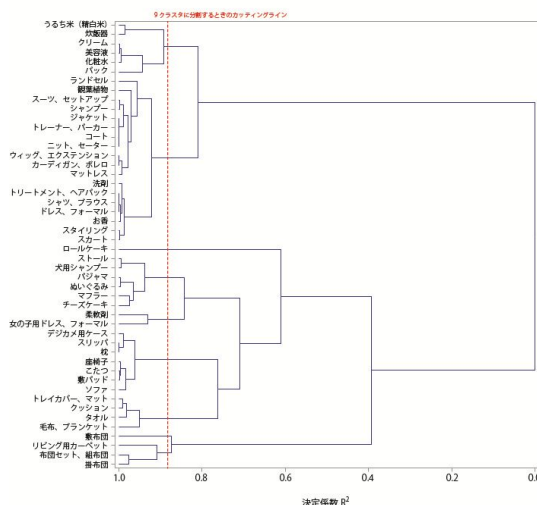


図3 デンドログラム

#### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計2件)

内田ゆず, 荒木健治, 米山淳, “ブログ記事からのオノマトペ用例文の自動抽出手法”, 知能と情報(日本知能情報ファジィ学会誌), Vol.24, No.3, pp.811-820, 2012.

Yuzu Uchida, Kenji Araki and Jun Yoneyama, “Classification of Emotional Onomatopoeias Based on Questionnaire Surveys”, Asian Language Processing (IALP), pp.1-4, Hanoi, Vietnam, Nov. 2012.

〔学会発表〕(計11件)

高丸圭一, 内田ゆず, 乙武北斗, 木村泰知, “地方議会会議録におけるオノマトペの

出現傾向に関する基礎的検討 少数の自治体に高頻度で出現するオノマトペについて ”, 言語処理学会第20回年次大会発表論文集, pp.566-569, 2014.

福島弘識, 内田ゆず, 荒木健治, “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”, 言語処理学会第20回年次大会発表論文集, pp.181-184, 2014.

内田ゆず, 荒木健治, 米山淳, “商品レビュー文に含まれるオノマトペに基づく商品カテゴリの分類”, 第29回ファジィシステムシンポジウム講演論文集, 2013.

佐藤宏亮, 内田ゆず, 米山淳, “否定表現を考慮したカスタマーレビューに出現するオノマトペとドメインの関係について”, 第29回ファジィシステムシンポジウム講演論文集, 2013.

内田ゆず, 長谷川大, 荒木健治, 米山淳, “商品レビュー文におけるオノマトペと商品カテゴリの関係についての予備調査”, 言語処理学会第19回年次大会発表論文集, pp.810-813, 2013.

上野駿太, 内田ゆず, 長谷川大, 佐久田博司, “エージェントによるオノマトペ教育システムの開発”, 情報処理学会第75回全国大会講演論文集, pp.4661-4662, 2013.

佐藤宏亮, 内田ゆず, 長谷川大, 米山淳, “カスタマーレビューにおけるレビュー評価点とオノマトペの関係の分析”, 情報処理学会第75回全国大会講演論文集, pp.2309-2310, 2013.

内田ゆず, 荒木健治, 米山淳, “感情を表すオノマトペのアンケート調査による分類”, 第28回ファジィシステムシンポジウム講演論文集, pp.175-180, 2012.

内田ゆず, 荒木健治, 米山淳, “複数評価者による感情を表す日本語オノマトペの分類”, 言語処理学会第18回年次大会発表論文集, pp.227-230, 2012.

内田ゆず, 荒木健治, “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.

内田ゆず, 荒木健治, 米山淳, “ブログ記事から抽出した用例文に基づくオノマトペの感情分析”, 第10回情報科学技術フォーラム (FIT2011) 講演論文集, 第2分冊, pp.273-279, 2011.

## 6. 研究組織

### (1) 研究代表者

内田 ゆず (UCHIDA, Yuzu)  
青山学院大学・理工学部電気電子工学科・助教

研究者番号 : 80583575