

令和 7 年 4 月 17 日現在

機関番号：30107

研究種目：基盤研究(C)（一般）

研究期間：2021～2024

課題番号：21K12584

研究課題名（和文）現代日本語オノマトペの語義とコロケーションに基づくオノマトペ実用辞書の構築

研究課題名（英文）Development of a Practical Dictionary of Contemporary Japanese Onomatopoeia Based on Semantic Meanings and Collocational Patterns

研究代表者

内田 ゆず（Uchida, Yuzu）

北海学園大学・工学部・教授

研究者番号：80583575

交付決定額（研究期間全体）：（直接経費） 2,100,000円

研究成果の概要（和文）：オノマトペの語義分類を低コストで実現するために、オノマトペの辞書にある用法から生成したルールを用いて語義分類を行う手法を提案した。

妊娠・出産・育児に関するオンラインコミュニティにおける検索履歴を対象として、オノマトペの出現傾向を分析し、ネガティブな感情や心身の不調を表すオノマトペが頻出することを明らかにした。

コミックのコマ画像中出现するオノマトペの特徴の調査を行い、コミックの画像中のオノマトペは多種多様であり、辞書に定義されていない未知語や、辞書で定義されていても語の一部が変化している語も存在することを確認した。

研究成果の学術的意義や社会的意義

本研究では、辞書情報と格解析情報を用いた、オノマトペの低コストな語義分類手法を提案した。従来、困難だったオノマトペの語義の体系化を可能にし、言語学や自然言語処理に新たな知見をもたらすものである。大規模な学習データを必要とせず、辞書の用法に基づくアプローチをとった点に独自性がある。

妊娠・出産・育児に関する検索履歴を分析した結果、心身の不調を示すオノマトペが多く使われていることが判明した。これは、産後うつなどの早期検出に活用できる可能性を示唆するものである。加えて、マンガ画像中の多様なオノマトペの調査により、視覚メディアにおけるオノマトペの特徴を明らかにした。

研究成果の概要（英文）：We proposed a low-cost method for onomatopoeia sense classification using rules derived from usage examples in existing dictionaries. Additionally, we introduced an algorithm that classifies expressions based on a "sense score," which favors senses that match more case analysis features.

We analyzed search histories from online communities related to pregnancy, childbirth, and parenting, and found that onomatopoeia expressing negative emotions or physical/mental distress frequently appeared. This suggests that such expressions may serve as useful cues for the early detection of postpartum depression and related conditions. Clustering co-occurring words further enabled us to visualize the usage patterns of onomatopoeia in this domain.

We also examined onomatopoeia in comic panel images and found a wide variety, including modified forms of known words and previously unlisted expressions.

研究分野：自然言語処理

キーワード：オノマトペ

1. 研究開始当初の背景

オノマトペは、日本語の語彙に豊富に存在している。言語学の分野において、オノマトペは幼稚な表現だとされていた時代もあるが、近年はその重要性が論じられるようになり、特に日本語学領域や日本語教育領域での研究が盛んになりつつある。

応募者は、オノマトペを工学的に応用することを目指し、オノマトペに関する研究を継続的に行っている。具体的には、オノマトペを高い精度で自動抽出する手法の確立、オノマトペを利用してテキスト中の感情を分析する手法の確立、日本語学習者のためのオノマトペ学習アプリケーションの開発を達成している。

このような動向の中でいまだに十分な知見が得られていないのが、本研究で取り組む「現代日本語オノマトペの語義の問題」である。オノマトペは多義性が高いことが知られている。言語学分野において、オノマトペの語義変化を分析する取り組み^[c]も見られるが、奈良時代～近代での語義の展開に焦点をあてており、現代日本語での使用実態を十分に反映しているとはいえない。また、個別具体的な少数のオノマトペを取り扱うにとどまり、大規模な分析は行われていない。つまり、現状では現代日本語で使用されているオノマトペの語義を体系的に捉える研究がなされていない。

2. 研究の目的

本研究は、多様なコーパスを分析して、オノマトペの出現傾向および語義の傾向、新出の語義・用法、オノマトペの語義と文脈情報の関係などについての基礎研究を完成させ、『現代日本語オノマトペ実用辞書』を構築することを目的とする。

オノマトペの語義についての従来研究は、少数の事例や近代までの事例を対象とした報告にとどまっている。そこで本研究では、応募者が大量のオノマトペを用法ごとに詳細に分析してきた経験を活かし、多様なコーパスから抽出したオノマトペと文脈情報から、現代日本語オノマトペの実態に即した言語的特徴・語義的特徴を俯瞰する。本研究で得られる知見は、日本語教育分野、言語学分野などの研究を進展させると同時に、機械翻訳・対話システムのような人工知能技術の向上にも貢献する。

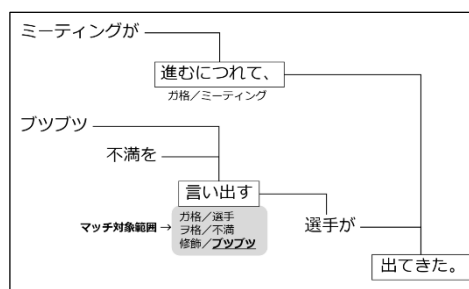
3. 研究の方法

本研究では、ジャンルや文体の異なる種々のコーパスに含まれるオノマトペおよび文脈情報を分析し、『現代日本語オノマトペ実用辞書』の構築を目指す。具体的な手法を以下に示す。

オノマトペと文脈情報の抽出

コーパスからオノマトペとその周辺の語を抽出する。対象となるコーパスは、Web 文書、検索履歴、議会会議録、コミックといった多ジャンルの言語資源から成る。コーパスには研究期間中に最新のデータを追加し、随時更新する。

大量のオノマトペ用例を短時間で得るための方法として、形態素解析による自動抽出が考えられる。しかし、オノマトペはひらがな・カタカナで表記される短い単語である上、品詞が一つに定まらないため、既存の形態素解析器では抽出精度が不十分であることを確認している。そこで、形態素解析器で抽出されたオノマトペ用例を手で確認し、誤抽出されたデータを排除する。次に、構文解析器によって構文木を生成し、文脈情報(係り先や修飾語)を自動抽出する。ここでも手確認を行う。このステップで抽出されるデータは本研究の基盤となるため、人的コストをかけて質の高さを担保する。



オノマトペの語義と文脈情報の分析

日本語オノマトペの多くは複数の語義を持ち、それらの語義は周辺の文脈によって判別される。オノマトペ実用辞書を構築するためには、収集した用例を語義ごとに分類する必要がある。大量の用例を語義分類するために、BERT から得られるオノマトペのベクトル表現を用いることで、小規模な訓練データによる語義分類の自動化を行う。人手で分類する場合は既存の日本語オノマトペ辞典の語義を基準とするが、辞典に掲載されていない語義で使用されている例が発見された場合は、新たな語義を定義する。

語義の分類が完了した後、語義と文脈情報の関係について調査する。オノマトペの語義の決定には係り先が重要な役割を果たすが、係り先よりも修飾語が有用な場合も確認されている。以上

表1 「ふらふら」の語義

語義番号	語義
①	不安定な状態で、力なく連続して揺れ動くようす。
②	疲れきって体が正常に動かないようす。
③	意志力に欠け、考えや気持ちが不安定であるようす。

を踏まえて各オノマトペの語義を決定する要素を、大量の語義分類済みデータから統計的に判断する。

4. 研究成果

(1) オノマトペの語義分類手法の構築

BERT から得られるオノマトペのベクトル表現を用いることで、小規模な訓練データにおける語義分類の自動化を試みたが、訓練データが少ない場合や文脈に特徴がない場合では分類精度が悪く、精度が不十分であることが明らかになった。また、学習データを作成するためのアノテーションコストが高く、全てのオノマトペに対して訓練データを豊富に用意することは難しいという問題があった。したがって、オノマトペの辞書にある用法から生成したルールによる語義分類を行う手法を提案した。多義のオノマトペを対象として分類実験を行った結果、いくつかの語義において高い精度が得られた。さらに、より特徴的で多くの格解析情報にマッチした語義に分類するような「語義スコア」を導入したアルゴリズムを提案した。

例：私が近付くと **ごろごろ** と **のど** を **囁かせ** ます。

【正解語義 ①】

語義	用法
①	<雷・物・腹>が……(と) 鳴る <猫>が<のど>を……(と) 囁らす
②	<物・肉体>が……(と) ころがる, 落ちる, する
③	<人>が……(と) 暮らす, 過ごす, する
④	<物・人>が……している, ある
⑤	<目・寝具・袋>が……する <突起>が……(と) 当たる

(2) 種々のドメインにおけるオノマトペの用法分析

・ コミック

小説やコミックといった作品においては、繊細な表現を読者に伝えるためにオノマトペが多用される。また近年は日本語学習におけるオノマトペの活用に関しても多く検討されている。特にコミックに出現するオノマトペは“画像”と“オノマトペ”がセットになって出現することから、オノマトペの繊細な表現を日本語学習者や人工知能が理解するために非常に有効と考えられる。

そこで、12種類のジャンルのコミックを対象とし、計360のコマ画像で使用されるオノマトペの特徴を調査した。360コマの画像に662種のオノマトペが計992回出現していた。そのうち、オノマトペ辞典の見出し語に掲載されている語は40.3%に過ぎないことが明らかになった。さらに、そのうち26.4%はオノマトペ辞典で定義されていない語義で使用されていた。これらのことから、コミックの画像中のオノマトペは非常に多種多様であり、辞書で定義されていても語の一部が変化している語や、辞書に定義されていない未知語の存在が明らかになった。

コミック画像中のオノマトペ分析を通じて、視覚メディアにおける言語表現の多様性が明らかになり、コミック工学分野への応用につながる。

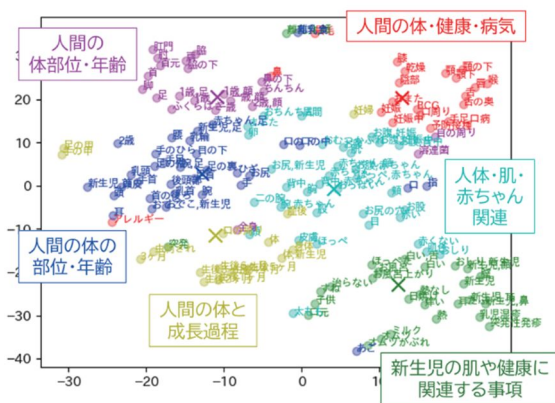
・ 出産・育児に関するテキスト

母親向けのオンラインコミュニティ上の検索履歴を対象として、オノマトペの出現傾向を分析した。約1億件の検索クエリにおいて、10回以上出現したオノマトペは578種存在することが確認された。頻度上位5語を挙げると、「いらいら」「ちくちく」「ぐりぐり」「ぶつぶつ」「かさかさ」となる。他のドメインと比較して、ネガティブな感情や心身の不調を表すオノマトペが頻出することが明らかになった。

検索クエリは「いらいら ワンオペ」

「かさかさ ほっぺ」「胃 ムカムカ 眠れない」のように、複数のキーワードを組み合わせられて構成されることが多い。そこで、オノマトペの共起語をクラスタリングすることで、出産・育児ドメインにおける各オノマトペの用法を可視化した。その結果、母親の悩みや母子の健康状態を示す表現を抽出することが可能になった。たとえば、「ぶつぶつ」の共起語は6つのクラスタに分類され、そのうち「人間の身体・健康・病気」クラスタには「手足口病」「アレルギー」「BCG」「予防接種」等のキーワードが属しており、母親の関心の対象を具体的に読み取ることができる。また、同じオンラインコミュニティでの質問文・回答文を分析したところ、母親の感情や心身の不調、子どもの疾病を表すオノマトペが頻出することが明らかになった。感情を表すオノマトペでは「イライラ」が圧倒的に高頻度である。「イライラ」と共起する語には「旦那」「義母」「産後」などが挙げられる。

これらの分析を通じて、母親の Web 上での行動からネガティブな感情や心身の不調に関連する表現の早期検出が可能になる可能性が示される。これは、産後うつなどの精神的問題の早期発見・支援に寄与するものであり、母親のメンタルヘルスを支える仕組みづくりに資する重要な知見である。



5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2件/うち国際共著 0件/うちオープンアクセス 2件）

1. 著者名 UCHIDA Yuzu, TAKAMARU Keiichi, OTOTAKE Hokuto, KIMURA Yasutomo	4. 巻 37
2. 論文標題 Clustering Analysis of Onomatopoeia Co-Occurrence Words in Search Queries on Mother-Targeted Websites	5. 発行年 2025年
3. 雑誌名 Journal of Japan Society for Fuzzy Theory and Intelligent Informatics	6. 最初と最後の頁 515 ~ 519
掲載論文のDOI (デジタルオブジェクト識別子) 10.3156/jsoft.37.1_515	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 FUJITA Michito, UCHIDA Yuzu, ARAKI Kenji	4. 巻 35
2. 論文標題 Onomatopoeia Sense Classification Algorithm Based on Dictionary Definitions	5. 発行年 2023年
3. 雑誌名 Journal of Japan Society for Fuzzy Theory and Intelligent Informatics	6. 最初と最後の頁 501 ~ 505
掲載論文のDOI (デジタルオブジェクト識別子) 10.3156/jsoft.35.1_501	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

〔学会発表〕 計7件（うち招待講演 0件/うち国際学会 0件）

1. 発表者名 内田ゆず, 高丸圭一, 乙武北斗, 木村泰知
2. 発表標題 母親向け情報サイトの検索クエリにおけるオノマトベ共起語のクラスタリング分析
3. 学会等名 第40回ファジィシステムシンポジウム
4. 発表年 2024年

1. 発表者名 内田ゆず, 高丸圭一, 乙武北斗, 木村泰知
2. 発表標題 母親向け情報サイトにおける検索語と投稿文を対象とした感情分析
3. 学会等名 第39回ファジィシステムシンポジウム
4. 発表年 2023年

1. 発表者名 内田ゆず, 高丸圭一, 乙武北斗, 木村泰知
2. 発表標題 妊娠・出産・育児に関する情報サイトにおける母親の検索行動の予備的分析
3. 学会等名 言語処理学会第29回年次大会
4. 発表年 2023年

1. 発表者名 藤田実智斗, 内田ゆず, 荒木健治
2. 発表標題 オノマトベの語義決定に寄与するコロケーションの分析
3. 学会等名 言語処理学会第29回年次大会
4. 発表年 2023年

1. 発表者名 藤田実智斗, 内田ゆず, 荒木健治
2. 発表標題 オノマトベ辞書に基づいたルールによるオノマトベ語義分類手法の提案
3. 学会等名 第38回ファジィシステムシンポジウム
4. 発表年 2022年

1. 発表者名 福島弘識, 内田ゆず, 荒木健治
2. 発表標題 コミックのコマ画像に出現するオノマトベの特徴分析
3. 学会等名 第38回ファジィシステムシンポジウム
4. 発表年 2022年

1. 発表者名 藤田実智斗, 内田ゆず, 荒木健治
2. 発表標題 辞書中の用法から生成したルールによるオノマトペ語義分類に関する一考察
3. 学会等名 令和3年度 電気・情報関係学会北海道支部連合大会
4. 発表年 2021年

〔図書〕 計1件

1. 著者名 木村泰知、高丸圭一、内田ゆず、本田 正美、河村 和徳、乙武 北斗、吐合 大祐、遠藤 勇哉	4. 発行年 2022年
2. 出版社 日本経済評論社	5. 総ページ数 288
3. 書名 自治体DX推進とオープンデータの活用	

〔産業財産権〕

〔その他〕

ONO-COLLO http://ono-collo.com/
--

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分 担 者	荒木 健治 (Araki Kenji) (50202742)	北海道大学・情報科学研究院・特任教授 (10101)	

6. 研究組織（つづき）

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	木村 泰知 (Kimura Yasutomo) (50400073)	小樽商科大学・商学部・教授 (10104)	
研究分担者	高丸 圭一 (Takamaru Keiichi) (60383121)	宇都宮共和大学・シティライフ学部・教授 (32207)	
研究分担者	乙武 北斗 (Ototake Hokuto) (20580179)	福岡大学・工学部・助教 (37111)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関