

令和 5 年 5 月 30 日現在

機関番号：30107

研究種目：若手研究(B)

研究期間：2017～2022

課題番号：17K12791

研究課題名(和文)オノマトペ語義の体系化を目指した文脈情報付きオノマトペデータベースの構築

研究課題名(英文) Construction of a Contextualized Onomatopoeia Database for Systematizing Onomatopoeic Meanings

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) 1,600,000円

研究成果の概要(和文)：本研究では、多様なコーパス(地方議会会議録、Web文書、ユーモア文、オンラインコミュニティ上の検索語、雑誌等)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行った。
また、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ことから、コンピュータでオノマトペの語義を自動的に判別するためのアルゴリズムの構築を行った。

研究成果の学術的意義や社会的意義

本研究の成果により、それぞれのオノマトペがどのような語と強く結びつくのか、どのような文脈で使用可能なのか、文脈によってどのように語義が変化するかを示すことができる。これは、従来の辞典のような語義といくつかの用例を提示するものとは本質的に異なるものである。日本語非母語話者にとってオノマトペの使い分けは難しいとされているが、本研究で得られた知見を日本語学習に活用することも期待できる。

研究成果の概要(英文)：In this study, we conducted a basic analysis of the occurrence tendencies of Japanese onomatopoeic expressions, their meanings, newly emerging meanings/usages, and the relationship between the meanings of onomatopoeia and contextual information by analyzing various corpora, including local assembly meeting minutes, web documents, humorous texts, search terms on online communities, and several types of magazines.
Moreover, since many Japanese onomatopoeic expressions have multiple meanings, and these meanings are discerned by the surrounding context, we developed an algorithm to automatically determine the meaning of onomatopoeic expressions using computers.

研究分野：コーパス言語学

キーワード：オノマトペ 語義 コーパス コロケーション

1. 研究開始当初の背景

オノマトペ(擬態語、擬音語の総称)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[a]、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[b]。しかし、「現代日本語オノマトペの語義」の問題については、いまだに十分な知見が得られていない。オノマトペは多義性が高いことが知られている。これは、オノマトペを機械的に処理する際や、非日本語母語話者がオノマトペを運用する際の支障となる。したがって、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求められている。

[a] 苧阪直行: 感性のことばを研究する 擬音語・擬態語に読む心のありか, 新曜社, 1999.

[b] 橋本喜代太 他: 外国人日本語学習者のオノマトペ習得支援システムのプロトタイプ開発, 日本教育工学会論文誌, Vol. 34, 2010.

2. 研究の目的

応募者は、オノマトペを工学分野で真に利活用するためにはオノマトペの語義についての知見を得ることが最優先の課題であると考えている。したがって、本研究計画は、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、『文脈情報付き現代日本語オノマトペデータベース』を構築することを目的とする。

3. 研究の方法

(1) コーパスの準備

オノマトペの語義や用法は、場面・対象・話し手の感情や意思などの複合的な要素によって決まる。本研究の目的達成のためには、様々なジャンルのコーパスを分析する必要がある。本研究では新聞、小説、地方議会会議録、商品レビュー、ブログ、雑誌からコーパスを整備する。雑誌のコーパスについては、ファッション、料理、育児などの各ジャンルから選定した雑誌を電子化し、オノマトペの抽出を行う。

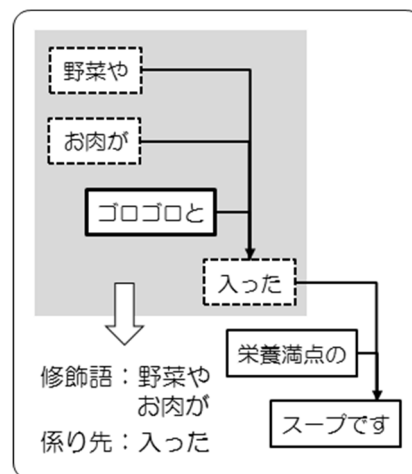
(2) オノマトペと文脈情報の抽出

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。したがって、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右図に示すような構文木を生成し、文脈情報(係り先や修飾語)を自動抽出する。ここでも、手で確認を行い、誤抽出を排除する。

(3) オノマトペの語義と文脈情報の分析

収集したデータを詳細に分析し、それぞれのオノマトペの語義を分類する。分類に際し、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については研究協力者と協議の上、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。調査から得られた知見に基づき、各オノマトペの語義を決定する要素を自動抽出する手法を考案する。



オノマトペの文脈情報抽出例

4. 研究成果

(1) 10,848,883 文の地方議会会議録コーパスから 982 語のオノマトペ(異なり数)を含む文を 186,416 例抽出した。それらを分析したところ、フォーマルな場でのオノマトペの役割、オノマトペ使用の地域差、議会特有の用法が明らかになった。具体的には、議会の場では強調を表すオノマトペが副詞として用いられることが多く、特に「しっかり」「どンドン」「はっきり」の頻度

が高いといった傾向がみられた。

また、Web データと地方議会会議録を対象としてオノマトペの出現傾向やオノマトペの係り先動詞の特徴を分析した。その結果、Web データは会議録よりもエントロピーが高い傾向にあることが明らかになった。また、ジャンルや文体が異なるコーパスでの動詞「する」に接続するオノマトペの用法、係り先動詞の偏りを比較し、新たな知見を得た。

(2) ユーモアの文脈でのオノマトペの用法を明らかにするために、駄洒落に出現するオノマトペの出現傾向を定量的に示し、オノマトペの出現位置や音の変化について考察を行った。対象としたコーパスにおいて、553 種のオノマトペが 1,715 件の駄洒落に合計 1,737 回出現した。それらを分析したところ、2~3 文字の短いオノマトペの出現頻度が高いこと、大部分のオノマトペが変形表現(駄洒落の前半部分)に出現することなどが明らかになった。オノマトペは他の語群にはない独特な音韻の特徴をもつため、駄洒落を作るために種表現の音に近い言葉を探するときオノマトペが多用され、変形表現に出現すると考えられる。また、駄洒落を成立させるためにオノマトペの音を様々な形式で変化させていることが明らかになった。

(3) 雑誌からオノマトペを抽出すると同時に、各オノマトペの係り先や修飾対象を人手で抽出し、文脈情報としてまとめた。さらに、オノマトペの表記揺れを人手で正規化し、実態に即した出現傾向を明らかにした。ここで得たデータにより、特定のドメインでのオノマトペの用法を明らかにできる。

幼児向け雑誌：11 冊から 737 語のオノマトペ(延べ数)を抽出した。他のドメインと比較すると、係り先の動詞が省略される例やオノマトペを単独で用いる例、辞書にはないオノマトペを使用する例、同じオノマトペを繰り返し用いる例が特徴的であった。幼児向け雑誌を対象としたオノマトペの文脈分析はこれまでに行われておらず、あらたな知見が得られたといえる。

女性向け情報誌・料理雑誌・男性向けファッション誌：各雑誌から人手で 3,018 語のオノマトペを抽出した。次に、オノマトペの周辺語彙のうち、語義の決定に寄与する単語を人手で抽出した。そして、オノマトペの語義を人手で分類し、ドメインによるオノマトペの使用傾向の違いを明らかにした。例えば、「すっきり」は女性向け情報誌では「余計なものがなく、はっきりしているさま」の語義、男性向けファッション誌では「身なりなどが垢抜けて洗練されているさま」の語義で使用される傾向がある。これらの結果から、オノマトペの語義曖昧性解消にドメイン情報が有効であることが示唆される。また、料理雑誌においては、オノマトペの出現頻度だけでなく、記事を通年で調査することによりオノマトペの季節感や特定の食材との結びつきなどを明らかにした。

(4) 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た。特筆すべき傾向として、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。このことから、母親の Web 上での行動から産後うつなどの精神的な不調を早期に検出する際に、オノマトペが手がかりとして有効であることが示唆された。

(5) オノマトペデータベース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。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。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」を導入することにより、提案手法を改良した。

5. 主な発表論文等

〔雑誌論文〕 計5件（うち査読付論文 3件／うち国際共著 0件／うちオープンアクセス 3件）

1. 著者名 藤田 実智斗, 内田 ゆず, 荒木 健治	4. 巻 35
2. 論文標題 用法情報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	5. 発行年 2023年
3. 雑誌名 知能と情報（日本知能情報ファジィ学会誌）	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 -

1. 著者名 内田ゆず, 荒木健治	4. 巻 32
2. 論文標題 駄洒落に使用されるオノマトベの特徴分析	5. 発行年 2020年
3. 雑誌名 知能と情報（日本知能情報ファジィ学会誌）	6. 最初と最後の頁 507-511
掲載論文のDOI（デジタルオブジェクト識別子） 10.3156/jsoft.32.1_507	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 乙武北斗, 内田ゆず, 高丸圭一, 木村泰知	4. 巻 32
2. 論文標題 BERTによる周辺文脈を考慮したオノマトベの語義分類手法の提案	5. 発行年 2020年
3. 雑誌名 知能と情報（日本知能情報ファジィ学会誌）	6. 最初と最後の頁 518-522
掲載論文のDOI（デジタルオブジェクト識別子） 10.3156/jsoft.32.1_518	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 内田ゆず	4. 巻 30
2. 論文標題 計算機で言葉の意味を扱う難しさ オノマトベのあいまいさを例に	5. 発行年 2018年
3. 雑誌名 知能と情報（日本知能情報ファジィ学会）	6. 最初と最後の頁 196-198
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 内田ゆず	4. 巻 28
2. 論文標題 日本語オノマトペを利用した感性情報処理に向けて	5. 発行年 2019年
3. 雑誌名 経営システム	6. 最初と最後の頁 91-96
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計11件 (うち招待講演 1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 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 オノマトペ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 オノマトペ辞書に基づいたルールによるオノマトペ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ペの特徴分析
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

1. 発表者名 藤田実智斗, 内田ゆず
2. 発表標題 多ジャンルの雑誌に出現するオノマトペのコロケーション分析
3. 学会等名 情報処理学会第82回全国大会
4. 発表年 2020年

1. 発表者名 内田ゆず, 荒木健治
2. 発表標題 オノマトペに着目した駄洒落の面白さの分析 駄洒落の自動生成に向けて
3. 学会等名 第35回ファジィシステムシンポジウム
4. 発表年 2019年

1. 発表者名 乙武北斗, 内田ゆず, 高丸圭一, 木村泰知
2. 発表標題 オノマトベの語義分類を目的としたBERTの適用方法に関する検討
3. 学会等名 第35回ファジィシステムシンポジウム
4. 発表年 2019年

1. 発表者名 内田ゆず
2. 発表標題 オノマトベから見る言葉のあいまいさ
3. 学会等名 第237回自然言語処理研究会（招待講演）
4. 発表年 2018年

1. 発表者名 内田ゆず
2. 発表標題 2種のコーパスを対象とした係り先動詞のエントロピーに基づくオノマトベの特徴分析
3. 学会等名 第33回ファジィシステムシンポジウム
4. 発表年 2017年

1. 発表者名 内田ゆず, 荒木健治
2. 発表標題 駄洒落に含まれるオノマトベの特徴分析
3. 学会等名 言語処理学会第24回年次大会
4. 発表年 2018年

〔図書〕 計2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	

1. 著者名 早稲田文学会（編）	4. 発行年 2021年
2. 出版社 筑摩書房	5. 総ページ数 276
3. 書名 早稲田文学 2021年春号	

〔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
---------------------------	-----------------------	----

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
---------	---------