

ブログ記事から収集した擬態語および擬音語用例の感情表現分析

内田ゆず[†] 渡部純平[†] 米山淳[†] 荒木健治[‡]
 Yuzu Uchida Junpei Watanabe Jun Yoneyama Kenji Araki
 yuzu@ee.aoyama.ac.jp

[†]青山学院大学
 Aoyama Gakuin University

[‡]北海道大学大学院 情報科学研究科
 Graduate School of Information Science and Technology, Hokkaido University

1. はじめに

日本語の語彙には、動作や事物の姿形、状態を感覚的に表す擬態語、事物の音や声を表す擬音語が豊富に存在している。(以降、擬態語・擬音語をまとめて「オノマトペ」と呼ぶ。)これらの語は生き生きとした表現力をもち、日本語でのコミュニケーションには欠かせないものとなっている。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。

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が行われている。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[1]。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[2]。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。

我々は、実際にオノマトペが使用されている用例を大量に収集することで、様々な分野に応用しようと考え、ブログ記事を対象としてオノマトペの用例の収集を行ってきた。さらに、各オノマトペが使用される場面で話し手が抱いている感情情報を提示できるようなデータベースの構築を目指

表1 収集対象としたオノマトペの例

動作 カテゴリ	オノマトペ	個数
騒ぐ	がやがや、どたばた...	32
疲れる	うんざり、くたくた...	13
働かない	ごろごろ、のんびり...	15
吐く	がらがら、げろげろ...	11
起きる	がばっ、むっくり...	21
飲む	がぶがぶ、ごくごく...	67
食べる	かりかり、ぱくり...	71
見る	きょろきょろ、まじまじ...	53
咳をする	けほけほ、こほん...	16

している。本稿では、収集した用例を感情表現の観点から分析した結果について述べる。

2. オノマトペを含む用例の収集

2.1 収集方法

Yahoo! ブログ検索 Web API を用いて、オノマトペ1語を検索クエリとしたときの検索結果上位20件分のスニペットを取得した(2010年6月20日時点)。検索クエリの対象としたオノマトペは、「日本語オノマトペ辞典」[3]に、動作を表すオノマトペとして掲載されている299語である。表1にその一部を示す。

この方法で、5,680件のブログ記事に対するスニペットが得られた。本研究では、オノマトペを含む文とその近傍を分析対象とするため、ブログ記事のスニペットからTitle要素とDescription要素を抽出している。

2.2 タグ付け

2.1で述べたスニペットには正しくオノマトペ

表 2 取得したスニペット中の正抽出率

オノマトペ	正抽出率
えへん	5.0% (1/20)
おほん	55.0% (11/20)
かーっ	10.0% (2/20)
かつ	5.0% (1/20)
ぐぐっ	25.0% (5/20)
けほけほ	90.0% (18/20)
げほげほ	70.0% (14/20)
ごほごほ	85.0% (17/20)
こほこほ	75.0% (15/20)
こん	0.0% (0/20)
こんこん	0.0% (0/20)
ごほん	65.0% (13/20)
こほんこほん	100.0% (20/20)
ごほんごほん	100.0% (20/20)
ぜーぜー	100.0% (20/20)
ぜりぜり	100.0% (1/1)

が含まれていないものも多く存在する。例えば、「えへん（せきばらいの音を表すオノマトペ）」をクエリとして検索を行った結果、「今日はもう外に出えへんの。」という文が誤って取得されていた。

そこで、取得したスニペット中で正しく用いられているオノマトペに人手でタグを付与した。この作業は、第二著者が行った。その結果、全体の47.7%のスニペットには正しくオノマトペが含まれていることが明らかになった。結果の一部として、表 2 に「咳をする」カテゴリに分類されたオノマトペの正抽出率を示す。表中の正抽出率は、取得されたスニペット（最大 20 件）のうち、正しくオノマトペが含まれている割合を表している。

3. 感情表現の抽出

3.1 使用するデータ

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ため、2.2 で正しくオノマトペが含まれていると判断された（つまり、タグを付与された）スニペットから感情表現を抽出する。ここでは、Ptaszynski ら[4]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。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。以下に感情表現要素の一部を示す。

- ・ 喜：めでたい, 幸福感, にんまり...

- ・ 怒：殺意, 逆鱗に触れる...
- ・ 哀：悲しみ, 嗚咽, 寂しい...
- ・ 怖：不気味, 寒気, 頼りない...
- ・ 恥：恥ずかしい, 赤らめる...
- ・ 好：友情, 慈悲, 蕩ける...
- ・ 厭：いやらしい, 忌々しい, 絶望...
- ・ 昂：焦らす, どよめく, 感嘆...
- ・ 安：落ち着く, 平然, びくともしない...
- ・ 驚：ショック, 度肝を抜かれる...

3.2 感情スコア

3.1 で述べた感情表現要素がスニペット中出现する頻度を調査し、それぞれの感情に対するスコアを算出する。例えば、「げーげー」をクエリとして検索して取得されたスニペット中に「悲しみ」が 4 度、「嗚咽」が 1 度、「赤らめる」が 3 度出現したとする。この場合、「哀」カテゴリのスコアは 5（「悲しみ」, 「嗚咽」の頻度の合計）, 「恥」カテゴリのスコアは 3（「赤らめる」の頻度）となる。このスコアを感情スコアと呼ぶ。

なお、頻度を算出する前にスニペットには形態素解析を施し、全ての単語を見出し語に変換して連結する処理を行っている。これは、感情表現要素が原形で収録されているためである。形態素解

表 3 「くしゃくしゃ」の感情スコア

感情カテゴリ	スコア
喜	11
怒	20
哀	6
怖	1
恥	20
好	1
厭	21
昂	2
安	2
驚	1

表 4 「くしゃくしゃ」のアンケート結果

感情カテゴリ	平均値
喜	1.60
怒	2.10
哀	2.25
怖	1.60
恥	2.10
好	1.45
厭	2.00
昂	2.05
安	1.50
驚	1.20

析器には日本語形態素解析システム JUMAN6.0[6]を用いた。形態素辞書は JUMAN の附属辞書である。結果の一部として、表 3 に「くしゃくしゃ」の感情スコアを示す。

4. アンケート調査

オノマトペに込められた話し手の感情を主観的に評価するため、アンケート調査を行う。このアンケートは、オノマトペ1語を単独で提示し、その単語に 10 カテゴリーの感情が含まれているかど

うかを {全く感じない, 感じない, 感じる, とても感じる} の 4 段階で評価するものである。提示するオノマトペは、3.2 で算出された感情スコアの合計が 15 を超えたオノマトペの中から、ランダムに 10 語 {むかむか, のんびり, むくむく, こほんこほん, ぐびっ, うんざり, わいわい, ぎろり, くしゃくしゃ, ほろり} を選択した。アンケートの回答者は日本語を母語とする 20 名である。

アンケートの結果は、「全く感じない」を 1,

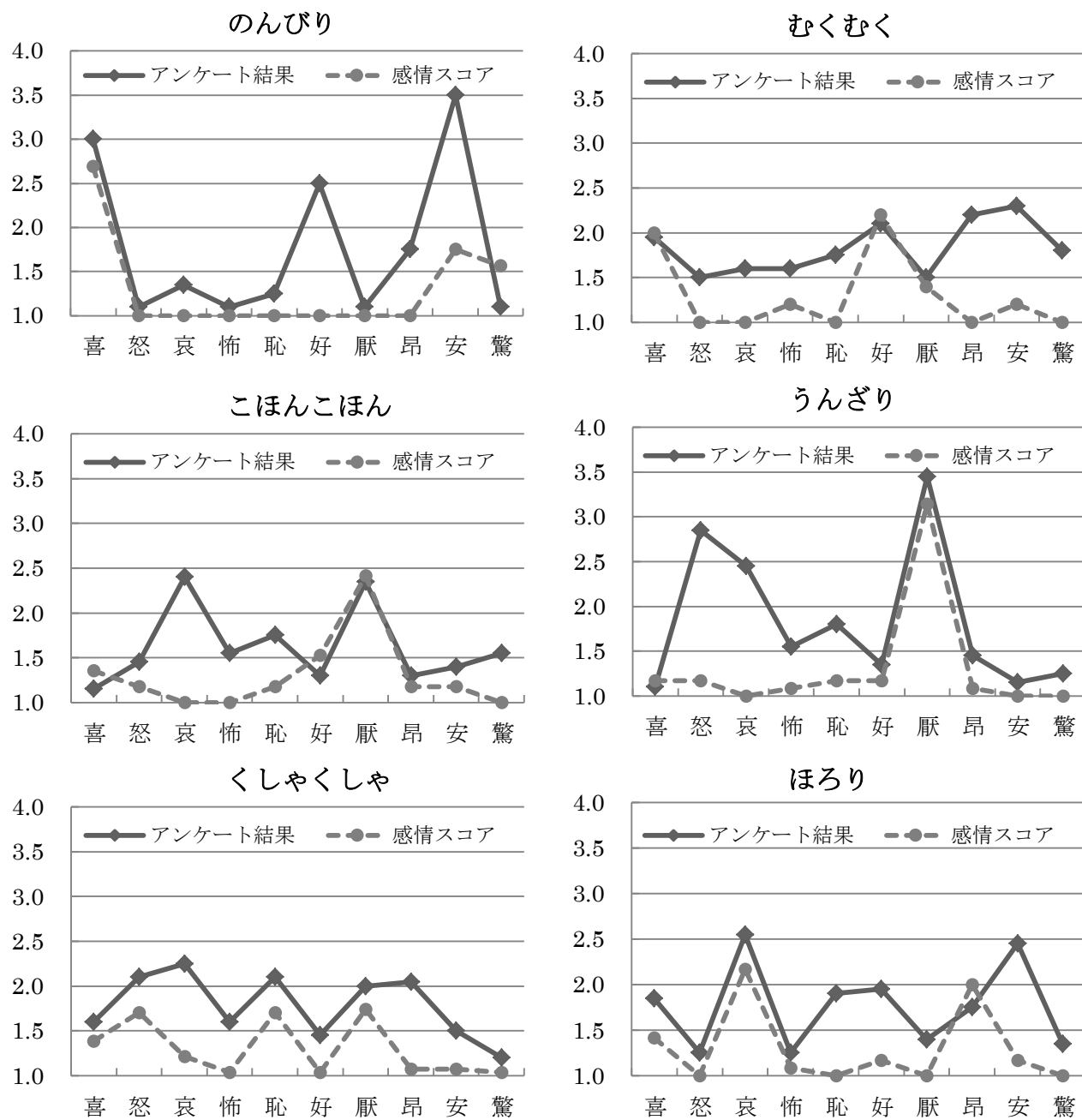


図 1 感情スコアとアンケート結果の比較

「感じない」を2, 「感じる」を3, 「とても感じる」を4と数値化し, 回答者全員の結果の平均値として求めた. 表4に「くしゃくしゃ」のアンケート結果を示す.

5. 比較実験

5.1 実験方法

3.2 で算出した感情スコアと 4. で得られたアンケート結果を比較し, 本研究における感情スコアの妥当性を評価する.

比較にあたり, 感情スコアを 1.00~4.00 の値に正規化している.

5.2 実験結果と考察

図1にアンケート結果と感情スコアを比較したグラフを示す. 紙面の都合上, 10語についての結果を載せるのは難しいため, 6語のオノマトペの結果を抜粋して載せる.

「喜」, 「厭」カテゴリの感情スコアは平均 1.73, 1.63 と高い傾向にある. (他のカテゴリは 1.07~1.37) 「喜」カテゴリに関しては, アンケート回答者にとって判断が易しいことやブログ記事になりやすい感情であることなどが要因だと考えられる. 「厭」カテゴリは感情表現要素が 533 種収録されており, 10 のカテゴリの中で最多であった. (他のカテゴリは 106~269 種) 感情表現要素が多いため, 感情スコアが上昇しやすいと考えられる.

一方で, 「昂」カテゴリの感情スコアは「ほろり」を除くオノマトペで 1.00~1.78 と低い値になっている. これは, 「昂」カテゴリ内の感情表現要素にブログ記事では使用頻度が低いと思われる表現が含まれていたからであると考えられる. 例えば, 「熟としていられない」, 「気が置かれる」, 「硬い気持ち」, 「押された気分」, 「激情」, 「激賞」などが挙げられる. 「ほろり」で 2.00 という高い値を取っているのは, 「昂」カテゴリの感情表現要素に「ほろり」が含まれているためである.

今回の実験では, 中村の感情表現辞典から抽出した感情表現要素をそのまま用いた. したがって, 表記の揺れの影響が見られた. 例えば, 「イライラする」という表現はブログ記事にも高い頻度で現れるが, 感情表現要素での表記が「苛苛する」となっており, 感情スコアに反映されていなかった. 今後の実験では表記の違いを考慮したデータを用いる必要がある.

6. まとめ

299 語のオノマトペが使用されているブログ記事のスニペットを取得した. 取得した 5,680 件分のスニペット全てを人手で調査したところ, 47.7%は実際にオノマトペが用いられているブログ記事であった.

さらに, あるオノマトペが使用される場面で話し手が抱えている感情情報を提示できるようなデータベースの構築を目指し, 収集した用例に感情スコアを付与した. アンケートによる主観評価との比較実験の結果, 「喜」, 「厭」は感情スコアとアンケート結果の平均誤差が 0.50 以下となった. したがって, 感情の種類によってはシステムによる評価が主観評価と一致することが明らかになった.

今後は, より高い精度でオノマトペ用例を取得する手法を開発し, 大規模なデータベースの構築を行う. さらに, そのデータベースを用いてオノマトペの感情評価を行う予定である.

参考文献

- [1] 香林隆子, 増永良文: "オノマトペのオンライン多言語辞書の構築", DEWS2002 論文集, A4-4, 2002.
- [2] Chisato Asaga, Mukarramah Yusuf, Chiemi Watanabe: "Onomatopedia: Onomatopoeia Online Example Dictionary System Extracted from Data on the Web", The 10th Asia Pacific Web Conference(APWeb), 2008.
- [3] 小野正弘: "擬音語・擬態語 4500 日本語オノマトペ辞典", 小学館, 2007.
- [4] Michal Ptaszynski, Pawel Dybala, Wenhan Shi, Rafal Rzepka and Kenji Araki: "A System for Affect Analysis of Utterances in Japanese Supported with Web Mining", Journal of Japan Society for Fuzzy Theory and Intelligent Informatics, Vol. 21, No. 2 (April), pp. 30-49 (194-213), 2009.
- [5] 中村明, "感情表現辞典", 東京堂出版, 1993.
- [6] 日本語形態素解析システム JUMAN 6.0, <http://www-lab25.kuee.kyoto-u.ac.jp/nl-resource/juman.html>.