E-031

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
Affect Analysis of Onomatopoeia Sentences Extracted from Blog Entries

内田 ゆず† 荒木 健治‡ 米山 淳†
Yuzu Uchida Kenji Araki Jun Yoneyama

1. はじめに

日本語の語彙には，動作や状態，事物の姿形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．そのため，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困難である．

オノマトペは日常生活の中で多用されるが，日本語教育の現場ではその使用頻度や重要性に見合うだけの十分な学習や指導がなされているとは言えない[1]．日本語学習者がオノマトペの意味や用法を習得することは難しく，有効な指導法・学習法の確立が求められている．

オノマトペを学ぶ方法の一つとして，オノマトペを用いた様々な表現に触れることが挙げられる．これを実践するためには，オノマトペが含まれる大量の用例が必要となる．

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[2]．この研究は，小説から手動で用例を抽出する手法を用いているため，質の高い辞書を実現している．しかし，大量の用例を人手で抽出（あるいは作成）するとコストが高い上，オノマトペの新しい用法を網羅できない．

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[3]．しかし，用例抽出手法が非常に単純で，Web 上に存在する現代のオノマトペの用法を網羅的に抽出しているとは言えない．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と比較するとデータ量が不十分である．

一方，多くの用例に触れるという学習方法には日本語母語話者が抱いている細かな感覚が伝わりにくいという問題点がある．先行研究では，オノマトペから感じる印象を，音象徴などを用いて評価したものがある[4-6]．

我々は，オノマトペが使用されたときの話し手，あるいは書き手の感情という側面からこの問題を解決しようとしている．したがって，大規模なオノマトペ用例文データベースを構築し，データベース中の各オノマトペに感情情報を付加することを目指している．

本論文では，次の 2 点について述べる．まず，オノマトペ用例文を自動的に抽出する手法を提案する．本手法はオノマトペの後続要素を詳しく分析した結果に基づくものである．次に，オノマトペの周辺文脈から感情情報を抽出し，各オノマトペが使用されるときの話し手，書き手の感情を判定し，その妥当性を評価する．

本研究では，Web 上のブログ記事をコーパスとして用例を収集する．大量のデータを低コストで入手できる点，現代日本語の用法を反映している点，新聞等のコーパスよりも感情情報が含まれやすい点でブログ記事は有用である．

2.ではオノマトペ用例文の特徴分析について述べる．**3.**では，**2.**の結果を踏まえてオノマトペ用例文の自動抽出手法を提案し，実際に抽出した用例文について説明する．**4.**ではオノマトペ用例文に含まれる感情情報を数値化した感情スコアについて述べる．**5.**では **4.**の感情スコアの妥当性を主観評価との比較によって検証する．最後に，**6.**で本論文の結論と今後の課題を示す．

2. オノマトペ用例文の特徴分析

ブログ記事からオノマトペ用例文を大量に抽出する手法を構築するために，オノマトペが使用されている文の特徴を分析する．本章では，分析対象となるオノマトペ用例文データの生成と，分析結果について述べる．

2.1 オノマトペ用例文データの生成

2.1.1 ブログ記事収集

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[7]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，5,802 件分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，各スニペットの Title 要素（ブログ記事のタイトル）と Description 要素（ブログ記事の本文）を使用する．したがって，合計 11,604 の要素を扱うことになる．

2.1.2 タグ付け

2.1.1 で収集したスニペットには正しくオノマトペが含まれていないものが多く存在する．例えば，「えへん（せきばらいの声を表すオノマトペ）」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．」という表現が取得される．オノマトペ用例文の自動抽出手法を考案するにあたり，正しいオノマトペ用例文と誤ったオノマトペ用例文を区別して分析しなければならない．

そこで，取得したスニペット中の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．作業者は日本語を母語とする 20 代男性 1 名である．

その結果，99 の Title 要素と 3,434 の Description 要素にタグが付与された．これにより，全体の 30.5%の要素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．

2.1.3 文分割

Description 要素は複数の文から成り立っていることが

† 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

表 1　収集対象としたオノマトペの語数と抜粋

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

多いため，1 文単位に分割し，オノマトペが含まれる文のみを抽出する．文の区切り位置には，句点，疑問符，感嘆符，改行に加えて，ブログ記事の文末に多く見られる表現を用いる．具体的には，以下の表現である．

- ★，☆
- ♪
- （笑），(笑)
- （汗），(汗)
- 連続した“w”あるいは“w”（2 回以上）
- ＾＾，^^
- …，・・・，．．．
- 鍵括弧

Title 要素と文分割された Description 要素のうち，オノマトペ用例文として不適切なものを除外する．ここでは，3 つのヒューリスティクスを用いる．

A) URL やメールアドレスが含まれる文を除外するため，半角英数字の連続（6 字以上）をストップワードとする．

B) 固有名詞として使用されているオノマトペを除外するため，括弧で括られているオノマトペ，直後に敬称が付与されたオノマトペをストップワードとする．使用した敬称は，｛さん，ちゃん，君，くん，様，さま，殿｝である．

C) 名詞が含まれていない文，短い文（オノマトペ+4 文字以下）は 1 文では意味を成さないと判断し，除外する．

2.1.2 でタグを付与された Title 要素と Description 要素に対して上記の処理を施した結果，2,861 のオノマトペ用例文が得られた．これらを正しいオノマトペ用例文として扱い，正解データと呼ぶ[1]．

タグを付与されなかった――人手で誤りだと判断された――8,071 の要素のうち，上記の処理を施した結果得られた 1,194 のオノマトペ用例文を不正解データと呼ぶ[2]．単純なヒューリスティクスを用いたフィルタリングだけでも，誤った用例文を 7 分の 1 程度まで削減できるきることが明らかになった．

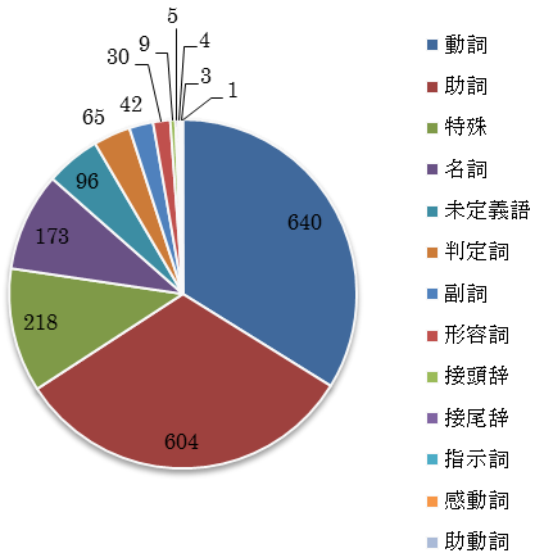

図 1　正解データにおけるオノマトペの後続単語の品詞
（形態素解析でオノマトペが正しく分割された場合）

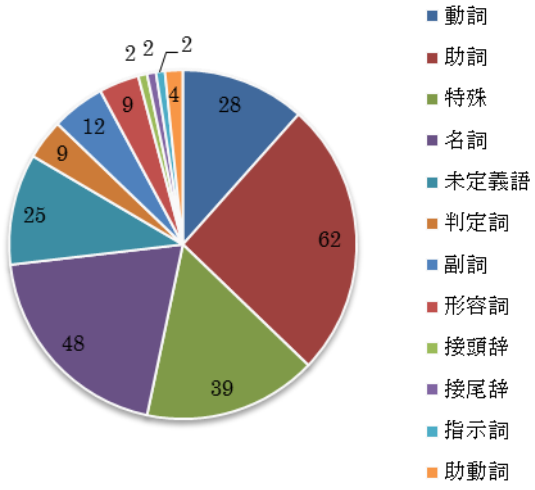

図 2　不正解データにおけるオノマトペの後続単語の品詞
（形態素解析でオノマトペが正しく分割された場合）

2.2 用例文データの分析

2.2.1 形態素解析による後続単語の分析

オノマトペに後続する単語の特徴を分析するため，正解データ，不正解データに形態素解析を行う．形態素解析器は日本語形態素解析システム JUMAN6.0[8]を用いた．形態素辞書は JUMAN の附属辞書に本研究で使用する 299 語のオノマトペを副詞，普通名詞として加えたものである．

形態素解析の結果，オノマトペ，あるいはオノマトペに“と”が接続された語[3]が 1 形態素として分割されたもの

[1] 1 文中にオノマトペが複数回出現する場合があることを考慮すると，正解データ中で分析対象となるオノマトペの延べ数は，3,156 である．

[2] 同様に，不正解データ中で分析対象となるオノマトペの延べ数は 1,286 である．

[3] “きっと”，“ぐっと”のように，オノマトペ＋“と”が形態素辞書に 1 語として登録されている場合があるため．

は，正解データ中に 1,765 例，不正解データ中に 223 例存在した．それぞれについて，後続単語の品詞の内訳を図 1，図 2 に示す．

動詞は正解データ中のオノマトペに後続する頻度が最も高く，全体の 33.9%である．不正解データにおける割合（11.6%）の約 3 倍である．

次に高い頻度で正解データ中のオノマトペに後続する品詞は助詞で，32.0%である．また，オノマトペ＋助詞に後続する単語は，名詞，動詞が 79.3%を占めている．不正解データの場合，助詞が後続する割合は全体の 25.6%であり，そのうち 50.8%が名詞，動詞に接続する．

3 番目に高い頻度で正解データ中のオノマトペに後続するのは，句読点や括弧などの記号を表す“特殊”である．動詞，助詞と比較すると大きな割合ではないが，その内容に特徴がある．正解データでは句読点が 148 例存在し，67.9%を占めている一方，不正解データでは 16 例，41.0%である．

判定詞は正解データ，不正解データともに低頻度であるが，具体的な単語に相違点がある．正解データには“な”（判定詞“だ”の連体形）が 20 例存在したが，不正解データには 1 例も存在しなかった．

2.2.2 形態素解析で正しく分割されないオノマトペの後続単語の分析

2.2.1 で述べた形態素解析の結果，オノマトペ，あるいはオノマトペに“と”が接続された語が 1 形態素として分割されないものも多数存在した．具体的には，AB ン型[4]（“えへん”，“ごほん”，“とろん”），AB ッ型（“きょろっ”，“じろっ”），AB リ型（“じろり”，“こくり”）等のオノマトペが正しく形態素解析されず，正解データ中の 1,096 例，不正解データ中の 971 例が該当した．

特定のオノマトペに関する用例文が分析対象にならないことは望ましくない．そこで，各文をオノマトペで区切り，オノマトペの直後の部分のみを形態素解析することで後続単語の分析を行う．正解データ，不正解データそれぞれについて，後続単語の品詞の内訳を図 3，図 4 に示す．

正解データ中のオノマトペに後続する単語は，助詞が最も多い．後続する助詞には“と”，“が”，“っと”の 3 語があり，“と”が 96.8%（301 例）を占めている．不正解データでは，助詞が後続する割合は 5.6%であり，低頻度である．

未定義語には疑問符や句読点，長音記号などが含まれ，正解データと不正解データで特に大きな差は見られなかった．

動詞は正解データ，不正解データで同程度の頻度だが，具体的な単語に相違点がある．

正解データでは，“する”（53 例，24.7%），“にる”（40 例，18.6%）の 2 語が突出して多く，“食べる”，“騒ぐ”，“飲む”などが続く．ただし，“にる”という動詞が高頻度になった背景には，形態素解析誤りがある．オノマトペには助詞“に”が接続することが多くあるが，JUMAN による形態素解析結果では助詞“に”が動詞“にる”の連用形となっていた．これは，用例文中のオノマトペの後ろに続く部分のみを切り出して形態素解析を行った結果，オノマトペの直後の助詞が文頭に置かれたことが原因である．

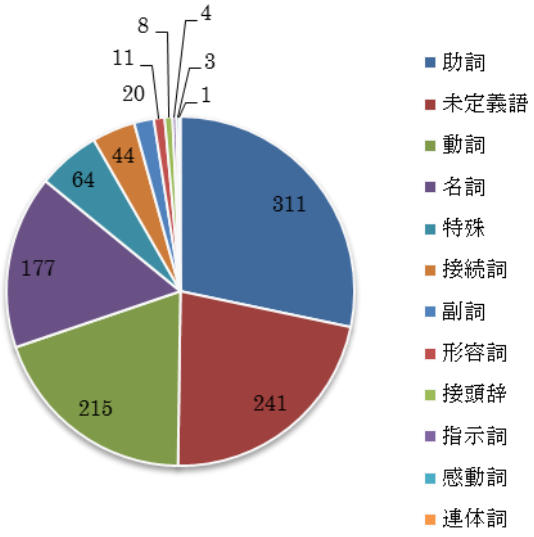

図 3　正解データにおけるオノマトペの後続単語の品詞（形態素解析でオノマトペが分割されなかった場合）

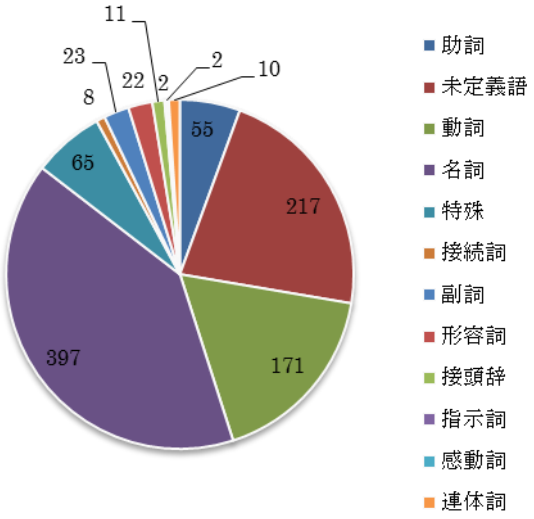

図 4　不正解データにおけるオノマトペの後続単語の品詞（形態素解析でオノマトペが分割されなかった場合）

不正解データでは，“くる”（18 例，10.5%）が最も多く，“ぬく”，“ひる”，“ちる”などが続く．動詞の種類，頻度に明確な特徴は見られない．

3. オノマトペ用例文の自動抽出

3.1 提案手法

2.2 で述べた用例文データの分析結果に基づき，オノマトペが正しく含まれる用例文を抽出する手法を提案する．再現率よりも適合率を優先させることを基本方針とする．

具体的なアルゴリズムを図 5 に示す．viii.にある“カテゴリ代表動詞”とは，オノマトペの動作カテゴリを明示する動詞を指し，｛見る，見つめる，見える，睨む，騒ぐ，

[4] 日向[9]の方法に従い，オノマトペの語基をアルファベットで表記する．1 つのアルファベットは 1 拍を表す．

a.収集したスニペットを **2.1.3** で述べた方法で文分割，フィルタリングを行う．

b.各文に形態素解析を行い，オノマトペが 1 形態素として分割された場合，以下の条件を満たす文を適切なオノマトペ用例文と判断する．
- i. オノマトペ＋助詞＋動詞
- ii. オノマトペ＋助詞＋名詞
- iii. オノマトペ＋動詞
- iv. オノマトペ＋判定詞
- v. オノマトペ＋句読点

c.b.でオノマトペが 1 形態素として分割されなかった場合，以下の条件を満たす文を適切なオノマトペ用例文と判断する．
- vi. オノマトペ＋助詞”と”
- vii. オノマトペ＋動詞”する”
- viii. オノマトペ＋カテゴリ代表動詞
- ix. オノマトペ＋“に”

図 5　オノマトペ用例文抽出アルゴリズム

(ア) すごくおもしろくて素敵な曲ばかりで、また聴きたいぞーとむくむく思ってきました。
(イ) もう上辺だけその場だけの取り繕いはバカ政権だけでうんざりです。
(ウ) かすみちゃん、ごほんにのるのいけないよ…
(エ) ケーキバイキングもあったり熊本ラーメンのおっぺしゃんとか一風堂なんかもあって…

図 6　抽出された用例文の一部

働く，怠ける，起きる，立つ，疲れる，吐く，もどす，食べる，飲む，呑む，咳き込む，むせる〕の 17 語とした．

2.1.1 で収集したスニペットをテストデータとして提案手法に適用すると，適合率 0.925，再現率 0.587 となる．

3.2 自動抽出したオノマトペ用例の評価

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 500 件分のスニペットを取得した（2011 年 4 月 11 日時点）．検索クエリの対象としたオノマトペは，**2.1.1** で述べた 299 語と同じものとした．

この方法で，125,730 件分のブログ記事に対応するスニペットが得られた．これらのスニペットの Title 要素と Description 要素，合計 251,460 の要素に図 5 のアルゴリズムを適用する．なお，これらの要素のうち，1,122 要素が **2.1.1** で述べた 11,604 要素と重複している．

まず，a.の処理によって，92,185 文が生成された．次に，JUMAN を用いて形態素解析を行い，b.の処理で 31,822 のオノマトペ用例文が抽出された．最後に c.の処理で 10,101 のオノマトペ用例文が抽出された．

抽出されたオノマトペ用例文から，9 つの動作カテゴリごとに 20 文ずつランダムに選択し，人手で正否を判断した．180 文中 168 文が正しいオノマトペ用例文であった．この結果から，適合率 0.933 が期待され，テストデータに本手法を適用した場合と同程度の性能が確認された．

図 6 に実際に抽出されたオノマトペ用例文の一部を示す．(ア)，(イ)は正抽出例で，(ウ)，(エ)は誤抽出例である．(ウ)は，「ごほん（重い咳を 1 回する音を表すオノマトペ）」の用例文として抽出されたものであるが，「本」の丁寧語である「ごほん」という同音異義語が原因で誤抽出となっている．(エ)は「しゃん（体を起こして姿勢正しいさまを表すオノマトペ）」の用例文として抽出されたものであるが，「おっぺしゃん」という固有名詞の一部を誤って抽出している．

4. 感情表現の抽出

4.1 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，**3.2** で抽出したオノマトペ用例文から感情表現を抽出する．ここでは，Ptaszynski ら[10]によって考案された日本語感情分析手法に用いられた感情表現要素を用いる．この感情表現要素は，中村[11]が日本語の感情を分類した〔哀，恥，怒，嫌，怖，驚，好，昂，安，喜〕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．

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

4.2 感情スコア

4.1 で述べた感情表現要素がオノマトペ用例文に出現する頻度を調査し，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”に対応するオノマトペ用例文に“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，“げーげー”の“哀”カテゴリのスコアは 5 ポイント（“悲しみ”，“嗚咽”の頻度の合計），“恥”カテゴリのスコアは 3 ポイント（“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オノマトペ用例文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．形態素解析器は日本語形態素解析システム JUMAN6.0 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 2 に“ぐびっ”と“くしゃくしゃ”の感情スコアを示す．

5. 感情スコアの評価

5.1 実験方法

4.2 で算出した感情スコアが日本語母語話者の感覚を正確に反映しているか否かを評価する必要がある．比較対象として，人間がオノマトペから受ける印象を，アンケートによって調査した．アンケートの詳細は **5.2** で説明する．

表 2　感情スコア（ぐびっ，くしゃくしゃ）

感情	ぐびっ	くしゃくしゃ
哀	14	31
恥	1	4
怒	0	244
嫌	15	252
怖	1	7
驚	4	2
好	20	12
昂	6	27
安	12	4
喜	21	56

表 3　アンケート結果（ぐびっ，くしゃくしゃ）

感情	ぐびっ	くしゃくしゃ
哀	28	45
恥	29	42
怒	28	42
嫌	30	40
怖	36	32
驚	40	24
好	38	29
昂	51	41
安	37	30
喜	46	32

評価対象のオノマトペは，299 語の中からランダムに選択した 10 語｛むかむか，ぐびっ，のんびり，ほろり，ぎろり，くしゃくしゃ，むくむく，わいわい，こほんこほん，うんざり｝とした．

感情スコアと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．

5.2 アンケートの詳細

アンケートは，10 語のオノマトペを 1 語ずつ単独で提示し，10 カテゴリ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．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，「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合計とした．表 3 に“ぐびっ”と“くしゃくしゃ”のアンケート結果を示す．

5.3 実験結果と考察

図 7 に 10 語のオノマトペに関するアンケート結果と感情スコアを比較したグラフを示す．両者の結果は平均値や標準偏差が異なるため，標準化したデータとなっている．

各々のオノマトペについて，結果を考察する．感情スコアとアンケートで大きく差がある結果，特に符号が異なるものについては原因を詳しく述べる．感情スコアの算出方法の性質上，感情スコアの方が大きい――つまり，ノイズが多く含まれた――項目の分析を行う．

【むかむか】

感情スコアとアンケート結果が概ね同様の傾向を示している．しかし，“好”の感情スコアは 0.79，アンケート結果は-1.09 であった．“むかむか”は怒りがこみ上げるさまや，吐き気をもよおすさまを表すオノマトペであり，“好”という感情は相応しくない．それにも関わらず，感情スコアが上昇したのは，“好”の感情表現要素に含まれている“酔う”，“熱”，“ふらふら”が原因である．例えば，「その直後から胃がむかむかし，芝居が終わる頃には胃がしくしく痛み始め，帰ってからは熱も３８度出て，ぶっ倒れてた」という用例文が該当する．

【ぐびっ】

図 7 のグラフから，突出して強く表れる感情は存在しないことがわかる．“哀”の感情スコアが高くなっているのは，「お風呂あがりに冷えた麦茶をぐびっと」のような例に含まれる“冷える”が原因であった．

【のんびり】

“驚”の感情スコアがアンケート結果より非常に高くなっている．“のんびり”はくつろいでいるさまを表すオノマトペであるので，“驚”という感情は明らかに不適切である．これは「普段ガラガラの映画館なのでのんびり行ったら結構人が入ってて びっくり」のような，オノマトペと感情表現要素の修飾先が異なる例によって起きたノイズであった．さらに，10 語のオノマトペを通して，“驚”の感情スコアは低い傾向があり，“のんびり”に付与された 6 ポイントが最高であった．このことから，数少ないノイズが大きな影響を及ぼしたと考えられる．

【ほろり】

“昂”の感情スコアがアンケート結果に比べて 2.67 高い．これは，“昂”の感情表現要素に「ほろり」が含まれているためである．感情表現要素に含まれるオノマトペが感情コアに反映されないような算出方法を導入することで改善が可能である．

【ぎろり】

アンケート結果では“怖”が 2.56 と高くなっているが，感情スコアは 0.17 である．これは，感情表現要素の種類に問題があると考えられる．Web 上で使用されるような現代日本語に即した感情表現要素を追加することが必要である．

【くしゃくしゃ】

“昂”，“喜”の感情スコアが高くなっている．これは，「顔をくしゃくしゃにして」という表現に続く｛泣く，涙を流す，号泣する，喜ぶ，笑う｝によるものである．このような例には，慣用表現を 1 つの単位とした分析を行うのが適当である．また，今回のアンケートでは，1 語ずつ提示されたオノマトペを評価する設問を使用したが，文脈を考慮した設問を用いることでより詳細な感情分析が可能になると考えられる．

【むくむく】

アンケート結果がすべての感情で-1.0～1.0 の間に収まっている．“むくむく”は感情や考えが急激に膨れ上がるさまなどを表すオノマトペであり，特定の感情との結びつきは低いためだと考えらえる． 感情スコアも“恥”が 1.66 とやや高くなっている以外は似た傾向を示している．“恥”の感情スコアは“のんびり”と同様，オノマトペと感情表現要素の修飾先が異なる用例文によって起きたノイズで上昇していた．

【わいわい】

感情スコアとアンケート結果が概ね同様の傾向を示して

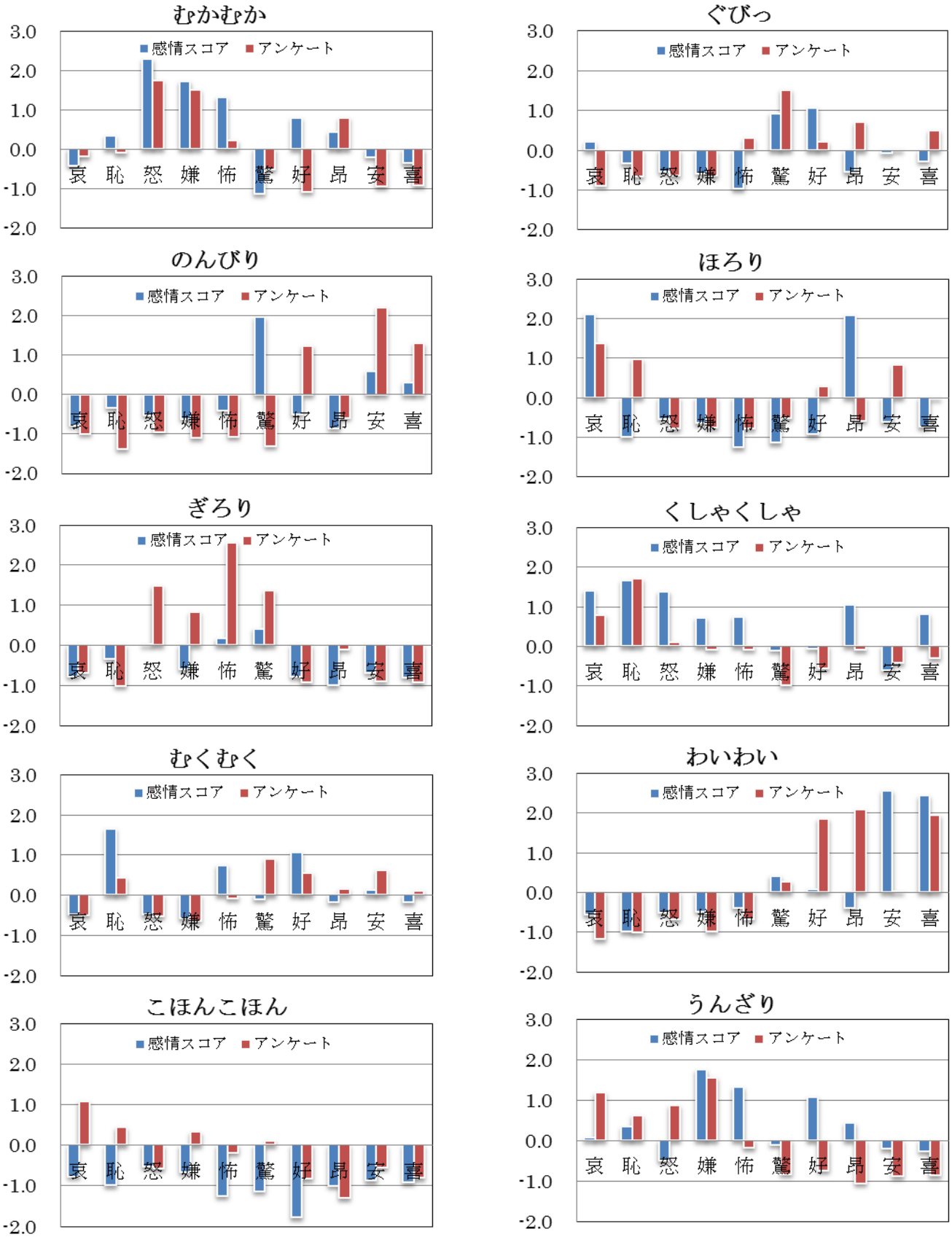

図 7　各オノマトペの感情スコアとアンケート結果の比較

いる．しかし，“安”に対応する感情スコアとアンケート結果の間に 2.52 の差があった．これは，“安”の感情表現要素に“楽（らく）”が含まれていたことによるものである．「初心者ばかりなのでとにかく楽しくわいわいやりたいと思っています」の例にみられるように，「たのしい」の意味で使われている“楽”が 47 例存在し，結果的に“安”の感情スコアを増加させていた．感情表現要素の読み仮名を考慮するなどの対策が必要である．

【こほんこほん】

標準化する前の感情スコアは，全ての感情で 0 ポイントであった．自動抽出されたオノマトペ用例文が 5 文のみであったことが原因である．

【うんざり】

“怖”，“昂”の感情スコアがアンケート結果よりも 1.5 ほど高い．これには時事的な要因が少なからず関わっている．この実験で用いたブログ記事は 2011 年 4 月 11 日に収集したもので，東日本大震災に関する話題が多く含まれている．特に，大地震発生から 1 か月が経過したにもかかわらず，余震が続く状況に「うんざりした」というような記述が多く存在した．例えば，「昨日の夜の地震も怖かったですがいくら天災とはいえ，もうんざりしてしまいます」などの文が抽出されている．そのため，“怖”の感情スコアが上昇したものと考えられる．また，“昂”の感情表現要素に含まれる“揺れる”が，「日本の半分以上が揺れていたようで，また，うんざり」などの例によって“昂”のスコアを上昇させていた．

ここから明らかなように，本研究は日々更新されるブログ記事をコーパスとしているため，最新の用例文を結果に反映させることが可能である．しかし，大地震のような大きな出来事に影響を受けやすいというデメリットもある．この問題は，長期間にわたって取得したブログ記事をコーパスとすることで解決できると考えられる．

6. おわりに

本論文では，大規模なオノマトペ用例文データベースの自動構築を目指し，ブログ記事からのオノマトペ用例文自動抽出手法について述べた．

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，オノマトペの後続単語について詳しく分析した．オノマトペの後ろには助詞＋動詞や助詞“と”が接続しやすいなどの特徴が明らかになった．分析結果に基づいたオノマトペ用例文の自動抽出手法を用いて，125,730 件のスニペットから 41,923 文のオノマトペ用例文を抽出した．抽出の精度は 93.3%であった．

次に，オノマトペと感情の関係を明らかにするために，オノマトペの周辺文脈に現れる感情情報を抽出し，書き手の感情を判定する実験を行った．

収集した用例文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，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．人間の評価により近い感情スコアを算出するためには，ブログ記事に含まれる感情表現に適した感情表現要素を検討する必要性がある．絵文字や顔文字の分析なども有効であろう．

今後は，長期的にブログ記事を収集し，一時的な話題による感情表現の偏りを抑制したデータベースを構築する予定である．

謝辞

本研究を進めるにあたり，タグ付けなどの作業に協力いただいた青山学院大学理工学部（当時）の渡部純平氏に心から感謝致します．

本研究は科研費（課題番号：23700256）の助成を受けたものである．

参考文献

[1]三上 京子, “日本語教育のための基本オノマトペの選定とその教材化”, ICU 日本語教育研究 3, pp.49-63 (2006).

[2]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).

[3]Chisato Asaga, Mukarramah Yusuf and Chiemi Watanabe, “Onomatopedia: Onomatopoeia Online Example Dictionary System Extracted from Data on the Web”, The 10th Asia Pacific Web Conference (APWeb) (2008).

[4]市岡 健一, 福本 文代, “Web 上から取得した共起頻度と音象徴によるオノマトペの自動分類”, 電子情報通信学会論文誌. D, 情報・システム J92-D(3), pp.428-438 (2009).

[5]小松 孝徳, 清河 幸子, 秋山 広美, “オノマトペか感じる印象を表現する属性とその客観的数値化”, HAI シンポジウム 2009, 2B-4 (2009).

[6]中部文子, 浅賀千里, 渡辺知恵美, “感性情報を利用したオノマトペ学習システムの開発”, 第 1 回データ工学と情報マネージメントに関するフォーラム(DEIM2009), E5-1 (2009).

[7]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).

[8]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.

[9]日向 茂男, “語形からみた擬音語・擬態語”, 東京学芸大学紀要. 第 2 部門, 人文科学 42, pp.59-70 (1991).

[10] 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).

[11] 中村 明, “感情表現辞典”, 東京堂出版 (1993).