

顔文字と日本語オノマトペの関連性分析

The analysis of the relation of emoticons and Japanese onomatopoeia

石井 直人 梶井 文人 プタシンスキ ミハウ
Naoto ISHII Fumito MASUI Michal PTASZYNSKI

北見工業大学

Kitami Institute of Technology

Along with diversification of communication methods with the use of modern technology, emoticons have become frequently used as expressions user feelings, attitudes and emotions. However, using the emoticons alone the nuance of the message is not transmitted correctly. Therefore it is necessary to supplement the meaning conveyed by emoticons with other information. In this paper, we focus on studying the relation between emoticons and onomatopoeia, as one of such supplementary means often corresponding to and co-occurring with emoticons, and report about initial findings regarding this relation.

1. はじめに

近年、メールやLINE・掲示板など、数多くのコミュニケーション手段が日常的に使われるようになってきている。これらは文字による情報伝達を主とするため、文章だけでは相手の表情や気持ちなどがわからず、誤解を与える場合がある。そのため、文字による情報伝達を支援する手段が数多く提案されてきた。例えば、単語を入力したらそれを簡単な絵に替える絵文字や、LINE・Facebookなどで普及した“スタンプ”(高解像度の絵文字)などが存在する。そのうち、高解像度の画像をユーザ同士で送りあうことがまだできなかった時代から使用されている手段として、顔文字が挙げられる。

顔文字とは、特殊文字を含めてあらゆる記号で作られ、顔の表情や姿勢を真似した表現であり、文字による情報伝達をベースにしたオンラインのコミュニケーション環境では、人の表情で現れる気持ちや感情などのような非言語的情報を伝えるときに使用される。その例として、「\(^ \circ ^)ノ」や「(;´ `)」などが挙げられる。

また顔文字はオノマトペや感嘆詞などと同時に使用されることが多い。それらの表現は顔文字と同時に文章中出现することもあるだけでなく、顔文字の一部として使用されることもあり、顔文字だけでは気持ちや感情を表現できない場合、顔文字へ不可欠な追加の言語情報として機能する。

オノマトペとは、擬音語や擬態語などの音や声を表した言葉である。擬音語の例には、物が発する音を字句で模倣した「ドカーン」や「パチパチ」といった例がある。擬態語の例には、精神状態、身体状態、感情などの音を発しないものを字句で模倣した「ピカピカ」や「ふわふわ」といった例がある。

また、顔文字とオノマトペが合わせて用いられる例として、例えば「しょぼーん」に対し「(´・`)」、「シャキーン」に対し「(´・`)」などの組み合わせが挙げられる。

しかしながら、顔文字とオノマトペの関連性はまだ解明されていない。そこで、本研究では文字による情報伝達における誤解を防ぎ、文字による情報伝達をより円滑に行うことを目的として、顔文字と合わせて使用されることの多いオノマトペに焦点を当て、この両者の関連性を調査する。

以下、2章では関連研究について述べ、3章では予備調査の結果と考察について述べる。また、4章では顔文字とオノマト

ペの理解度における調査の結果と考察について述べる。最後に5章では結論及び今後の課題について述べる。

2. 関連研究

2.1 顔文字に関する研究

顔文字を用いた研究としては、川上 [1] が、顔文字が表す感情と強度を数値化し、顔文字データベースの作成を行っている。川上の調査は、1つの顔文字が複数の感情を表す場合があることを示した。

人工知能の分野で顔文字を用いた研究として、Ptaszynski [2] が構築した顔文字解析システム CAO がある。CAO システムとは入力文から顔文字を抽出し、その感情の種類を推定するシステムである。CAO システムが網羅する顔文字数はおよそ 300 万個にも及び、2010 年の段階では 99.5 % の顔文字を正しく抽出することが可能で、顔文字の感情推定精度は 93.5 % と高い精度である。

2.2 オノマトペに関する研究

オノマトペを用いた研究としては、橋本ら [3] はオンライン学習システム ONOMATOPENARAI を開発した。このシステムは、日本語学習者のうち、特に職業研修・就業を目的として訪日する者を対象としている。

また、内田ら [4] はオノマトペを 10 個の感情ごとに分類したオノマトペ DB を作成した。内田らの調査では、人間であってもオノマトペに含まれる感情の判断は困難であることを示した。

また、顔文字とオノマトペの両者を対象とした研究として、瀧下ら [5] は顔文字とオノマトペの複合要素から抽出される感情成分に着目し、自動的に感情の抽出を行う研究を行っている。しかし、この両者を対象とした研究はあまり行われていない。

3. 予備調査

3.1 アンケートの設計・実施

顔文字とオノマトペの関連性を調べるための方法として、アンケート調査を行った。以下ではアンケートの設計方法について記述する。

まず、アンケートの項目を作るにあたって資源となる顔文字とオノマトペのリストが必要になった。インターネット上には

顔文字の辞書は数多く存在する。しかし、多くの顔文字辞書では顔文字をそれが伝える意味(例“うれしい”、“悲しい”、“こんにちは”、など)で検索しなければならない。それと違い、顔文字をそれが一緒に頻出する表現(例：ガーン、ガクガクブルブル)を使って検索する「2ちゃんねる定番顔文字」[6]も存在する。本辞書の項目数は、477個にも及び、その検索キーワードの表現には、オノマトベも含まれる。そのため、今回の実験では、アンケート作成の資源としてこの辞書を使用した。

本アンケートでは、被験者はまず顔文字のリストを紹介され、それぞれの顔文字に関連したオノマトベを書く。また、次にオノマトベのリストを紹介され、それに合った顔文字または、顔文字が浮かばないときに似顔絵などの図を書く。回答が浮かばない場合には「わかりません」という回答も認めた。こうすることにより、オノマトベの解釈、または顔文字の解釈にどのような違いが生じるかを確かめる。実際に設計したアンケートの設問例を図1に示す。

設問1
次の顔文字を見て、どのようなオノマトベが浮かびますか
浮かばない場合は「わかりません」と答えてください
例) (°ω°) : しよぼーん
(1) (つゝ)つ
(2) (・▽・)
(3) (= 3 =)
(4) \ (´-´)ノ
(5) (´▽´)
(6) \ (´^0^´)ノ
(7) f(^^)
(8) (´ρ´)
(9) (´Д´)
(10) Σ (´皿´)

設問2
次のオノマトベを見て、どのような顔文字が浮かびますか。
浮かばない場合は「わかりません」と答えてください。
また、顔文字がわからない場合には、絵などで表しても構いません。
例) しよぼーん : (°ω°)
(1) ガーン
(2) どよーん
(3) シヤキーン
(4) きよるきよる
(5) こくり
(6) シダバタ
(7) イライラ
(8) グラグラ
(9) ガクガクブルブル
(10) くわっ

図 1: アンケートの設問例

次に、設計した設問と解答欄を用いてアンケートを実施した。アンケートの対象者は20代の男性8名、女性4名の合計12名であった。

3.2 結果と考察

実施したアンケートの回答の詳細分析を行った。例えば、設問1の(9) (´▽´) (10) (´皿´)には項目「ガーン」が1票ずつ入っており、設問2の(1) ガーンには回答項目9票のうち7票が「°ω°」という目と口を表現した顔文字が用いられていた。このことから、「°ω°」という顔文字は「ガーン」というオノマトベと対応付けられていると考えられる。

また、設問1の(3) (= 3 =)の投票を集計した結果、怒りの感情を表す「むー」や「ブーブー」と、愛好の感情を表す「ちゅー」が混在していることがわかった。これは、被験者の回答の多くが一致する関係とあまり一致しない関係があるということを示している。このことから、顔文字とオノマトベは1対1に対応しないことが考えられる。

更に、各設問(1)～(10)までの「わかりません」の割合について調べた。その結果を表1、表2に示す。なお、「わかりません」の割合の平均を表1に、「わかりません」の割合における最小値、最大値を表2に示す。また、男女間での顔文字・

オノマトベの理解度の違いを確かめるため、男女別の回答率を求めた。その平均値を男女別に分けて表3に示す。

表 1: 「わかりません」の割合における平均値

		割合の平均
顔文字	オノマトベ	0.36
オノマトベ	顔文字	0.46

表 2: 「わかりません」の割合における最小値と最大値

		最小値	最大値
顔文字	オノマトベ	0.00	0.57
オノマトベ	顔文字	0.21	0.78

表 3: 男女別平均回答率

		男性	女性
顔文字	オノマトベ	0.50	0.90
オノマトベ	顔文字	0.50	0.75

表1、表2より、「わかりません」の割合における平均「わかりません」の割合における最大値ともに顔文字をオノマトベに変換するよりも、オノマトベを顔文字に変換する方が高い数値が得られた。このことは、顔文字からオノマトベへの変換においては回答が一意に決まるのに対し、オノマトベから顔文字への変換においては、回答者が多数の変換候補を想起してしまい、当てはまるオノマトベがわからないという状況に直面することが理由であると考えられる。これは顔文字が多義であるケースが多いことを示唆していると考えられる。よって、顔文字からオノマトベへの変換よりも、オノマトベから顔文字への変換の方がより難しいということがわかる。また、顔文字をオノマトベに変換する際、「まじ!」や「しまった」といったオノマトベではない単語も幾つか混在しており、顔文字の持つニュアンスの曖昧さを示唆している。

表3では、男性よりも女性の方が平均回答率が高いことが考察できる。その理由として、女性の方が男性よりも絵文字や顔文字を用いる機会が多いと言われており[7]、それらに対する知識の多さや慣れが影響していると考えられる。

3.3 予備調査の問題点

3.2で述べたように、予備調査に用いたアンケートでは選択肢がないため顔文字・オノマトベが想起しにくく、「わかりません」と答えた被験者が多く見られた。また、人によって想起される顔文字・オノマトベが違う、または似ていても同じではないため、回答の曖昧性が高くなってしまった。これらの課題を解決し、より正確な調査を行うため、アンケートの再設計を行った。4章では再設計したアンケートについて述べる。

4. 顔文字とオノマトベの理解度における調査

4.1 アンケートの設計・実施

顔文字・オノマトベはともに感情を表す表現であり、何らかの感情とともに使用していると考えられる。そこで、顔文字とオノマトベの関連性を調べるために、オノマトベまたは顔文字を対象とし、それらを感情と結びつく研究やデータが必要になった。まずは、感情を媒体にして調査を進めるのには、

日本の現状に合った感情の種類における既存研究が必要になった。そのため、中村の感情表現辞典 [8] を利用した。この感情表現辞典では、見出し語が喜、怒、哀、怖、恥、好、厭、昂、安、驚の 10 種類に分類されている。この 10 種類の感情を表現する顔文字・オノマトペをそれぞれ 20 個無作為に用意し、アンケートを設計した。このとき、顔文字は CAO システム [2] の DB を利用した。また、オノマトペは内田らの考案したオノマトペ DB [4] を利用した。

予備調査 (第 3 章参照) を実施した際には、選択肢がないため顔文字・オノマトペが想起しにくく、回答が困難であるという意見が見られた。

そこで、今回のアンケートでは、20 個の顔文字を無作為に並べ、それらを最も適切に表しているオノマトペを語群から選ぶ形式を採用した。また、当てはまるオノマトペがないと感じた場合、無回答で良いという条件を加えた。このように、「無回答・わかりません」という選択の自由度を維持したうえで、回答が浮かんだ場合は、それに一番合った項目をリストから選び、自由記述において起きる表記の揺れ問題を解決した。また、回答は自分で考える必要はなく、リストから選ぶことは、回答者の労力低下につながり、無回答が少なくなかつ、全般的により正確な結果が得られることを仮定した。

設計したアンケートの設問例を図 2 に示す。



図 2: アンケートの設問例

次に、設計した設問と解答欄を用いてアンケートを実施した。アンケートの対象者は 20 代の男性 29 名、女性 6 名の合計 35 名であった。

4.2 一致度の検証

顔文字とオノマトペの理解度や関連性を客観的に調査するため、顔文字とオノマトペに関して回答者の一致度を数値化する必要がある。一致度を求めるための方法として、Cohen の

係数 [9] があり、この方法は被験者間の回答の一致度を評価するものである。しかし、係数は、2 人の被験者が 2 つの項目についてどのくらい一致しているかを数値化する方法である。複数の被験者あるいは複数の項目を扱う も存在するが、両方が複数かつ事前から未定であるという高自由度が求められる一致度の係数はなかったため、 を使って正確に一致度を測ることはできなかった。

そこで、今回以下に示す方法で一致度を求めた。まず、各顔文字についてのオノマトペの投票率を求めた。次に、それらを順位で重み付けした。最後に全項目の平均を取った。投票率を a 、順位を x 、項目数を y とし、以下の式 (1) で求められる。

$$\sum \frac{a * (y - x - 1)}{y} \quad (1)$$

次に、各顔文字の一致度を求めた結果を一致度の高い順に表 4 に示す。

表 4: 各顔文字の一致度

順位	顔文字	一致度
1	(; _ ;)	0.8667
2	\ (> o <) /	0.8615
3	(° ° ;)	0.8378
4	(` `)	0.7431
5	(J) J	0.6781
6	-(; - ` A)	0.6585
7	(` -) J	0.6296
8	((o(o・・o)o))	0.6250
8	f(^ ^ ;)	0.6250
10	(^ 2 ^)	0.5789
11	(- o -)	0.5754
12	(= =)	0.5397
13	(° □ ° 中) 中	0.5104
14	(* J J)	0.5099
15	(J ・ o)	0.5000
16	(° v °)	0.4986
17	(^ ^ *)	0.4889
18	(` *)	0.4662
19	(° °)	0.4587
20	(` ・ `)	0.4571

一致度の高い顔文字は仕草や表情などのアクションが比較的明確なものが多く意味がわかりやすいため、一致度が高くなったと考えられる。一方、一致度の低い顔文字を見ると、投票されたオノマトペの項目数が 10 件以上と多い顔文字が多数存在していた。このことから、一致度の低い顔文字は語義が複数あるということが考えられる。

4.3 顔文字の詳細分析

顔文字「(` -) J」と「(= =)」について、これらはともに「ちゅー」と「ぶーぶー」の投票率が高かった。このことから、この 2 つの顔文字は“ ”という記号から意味を想起したと考えられる。

また、顔文字「(` ・ `)」は「げんなり」や「くよくよ」などの投票率が高く、顔文字「(` `)」は「むすっ」と「ぶーぶー」が多数を占めた。さらに、「(J ・ o)」は「しくしく」と「くよくよ」の投票率が高かった。従って、“ ”という記号は怒りや悲しみなどのネガティブな感情を示すことが多いと考えられる。これらのことから、記号により大まかに分類できる関係があることが考えられる。

顔文字「(- o -)」と「(^ 2 ^)」はともに無回答が多く見られた。このうち、「(- o -)」はアクションの大きさが抑えられており、無表情に見えるため感情が想起しにくく、結果として無回答が多く見られたと考えられる。また、「(^ 2 ^)」について、“ 2 ”という記号の指すニュアンスがわかりにくいため、無回答が多く見られたと考えられる。

4.4 性別による差異についての考察

顔文字とオノマトベの関連性をより正確に把握するためには、ユーザーの属性の違いから顔文字・オノマトベの解釈にどのような差異が生じるかを確かめる必要があった。そのため、被験者の性別に着目し、その差異を考察した。

例えば、顔文字「(ー〇ー)」を見ると男性は「るんるん」11票、「わいわい」2票、「うきうき」「わくわく」がそれぞれ1票ずつであった。それに対して、女性は「ほっ」2票、「うっとり」1票であった。これは、男性は顔文字「(ー〇ー)」を活動的な様子を示していると感じているのに対し、女性は落ち着いた様子を示していると感じているように見られる。このことは、男性と女性では顔文字の捉え方が異なるということを示唆していると考えられる。

また、顔文字「(= =)」の投票を集計した結果、男性は「むすっ」10票「ぶーぶー」10票「ちゅー」13票であったのに対し、女性は「むすっ」2票と「ぶーぶー」5票のみであった。このことは、男性は怒りの感情を表す「むすっ」「ぶーぶー」と、愛好の感情を表す「ちゅー」とで回答が対立しているのに対して、女性は怒りの感情で一致していることを示している。このことから、男女によって顔文字の捉え方に差が確認できるうえで、同じ性別においても顔文字の理解に差異があると考えられる。

4.5 オノマトベについての考察

被験者のオノマトベに対する理解度を調査するため、語群に使用したオノマトベの投票数を求めた。また、一般人のオノマトベの使用頻度を求めるため、YACIS ブログコーパス [10] を用いてそれらの使用件数を求めた。その結果を投票数の高い順に表5に示す。

表5: 各オノマトベの投票数と使用件数

順位	オノマトベ	投票数	使用件数
1	やれやれ	52	14530
2	ぎゃー	51	15054
3	にやにや	48	5772
3	しくしく	48	6541
5	るんるん	43	10686
6	わいわい	42	14203
7	どきっ	40	12152
8	うきうき	36	12869
8	ほっ	36	255567
8	ぶーぶー	36	2518
11	てれっ	33	1917
12	むすっ	29	562
12	ちゅー	29	49197
14	わくわく	26	80032
15	げんなり	24	3577
16	くよくよ	21	4288
17	おろおろ	18	2843
17	うっとり	18	58392
19	うじうじ	8	1995
20	もじもじ	7	2613

投票数と使用件数の Pearson 相関係数を調べたが、相関は認められなかった。しかし、網掛けされている部分を見ると、一部例外があるものの、投票数の高いものは使用件数が1万数千件以上と高く、投票数の低いものは使用件数が数千件程度と低くなった。このことから、日常的なやりとりで頻繁に使わ

れるオノマトベは被験者の回答として使用されやすいと考えられる。

5. おわりに

本研究では、アンケート調査を行うことにより、両者の関連性を調べた。その結果、顔文字は多義であり、曖昧性の高い表現であることや、オノマトベと顔文字は1対1対応ではないことがわかった。また、顔文字毎にその理解度は異なるということ、さらに、人によって顔文字に対する理解に差異が生じることなどがわかった。今後はより多人数を対象にアンケートを実施する予定である。またアンケートの結果を実用し、顔文字オノマトベ翻訳システムの開発に取り組む予定である。

参考文献

- [1] 川上正造:顔文字が表す感情と強調に関するデータベース、大阪樟蔭女子大学人間科学研究紀要 7, pp.67-82, 2008
- [2] Michal Ptaszynski, Jacek Maciejewski, Pawel Dybala, Rafal Rzepka and Kenji Araki: CAO: A Fully Automatic Emoticon Analysis System Based on Theory of Kinesics, IEEE Transactions on Affective Computing, vol. 1, no. 1, pp. 46-59, Jan.-June 2010.
- [3] 橋本喜代太, 竹内和広: 外国人日本語学習者のオノマトベ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 34(Suppl.), pp.69-72, 2010-12-20.
- [4] 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ベの分類, 言語処理学会 第18回年次大会 発表論文集, 2012年3月
- [5] 瀧下祥, 奥村紀之: 顔文字とオノマトベに基づく文章からの感情抽出, 情報処理学会第77回全国大会, IR05, 2015年3月
- [6] 顔文字ちゃんねる: <http://kaomoji.uunyan.com>
- [7] 西川健, 井手口哲夫, 奥田隆史, 村田嘉利: 携帯電話における絵文字の利用分析, 経営情報学会 全国研究発表大会要旨集 2004
- [8] 中村明: 感情表現辞典, 東京堂出版, 1993.
- [9] Cohen, J.: A coefficient of agreement for nominal scales, Educational and Psychological Measurement, 20, pp.37-46, 1960.
- [10] Michal Ptaszynski, Pawel Dybala, Rafal Rzepka, Kenji Araki and Yoshio Momouchi: YACIS: A Five-Billion-Word Corpus of Japanese Blogs Fully Annotated with Syntactic and Affective Information, In Proceedings of The AISB/IACAP World Congress 2012 in Honour of Alan Turing, 2nd Symposium on Linguistic and Cognitive Approaches To Dialog Agents (LaCATODA 2012), pp. 40-49, 2-6 July 2012,