Classification of Emotional Onomatopoeias Based on Questionnaire Surveys

Yuzu Uchida
Department of Electrical Engineering and Electronics
Aoyama Gakuin University
Sagamihara, Japan
yuzu@ee.aoyama.ac.jp

Kenji Araki
Graduate School of Information Science and Technology
Hokkaido University
Sapporo, Japan
araki@media.eng.hokudai.ac.jp

Jun Yoneyama
Department of Electrical Engineering and Electronics
Aoyama Gakuin University
Sagamihara, Japan
yoneyama@ee.aoyama.ac.jp

Abstract—Emotion estimation from sentences is helpful in semantic analysis and human-computer interaction. People sometimes express their emotions directly and sometimes describe experiences that produce various emotions. Emotion estimation requires both verbal expressions. We believe Japanese onomatopoeias could serve as a landmark of emotions. In this study, multiple raters judged emotions found in 324 onomatopoeias. We quantitatively evaluated the degree of agreement among the raters and attempted to improve it.

Keywords—Japanese onomatopoeia, emotion, agreement rate, Kappa coefficient

I. INTRODUCTION

A method is needed to estimate emotions from verbal expressions in the fields of natural language processing and artificial intelligence. This method would play an important role in the development of semantic analysis and human-computer interaction.

People sometimes express their emotions directly and sometimes describe experiences that produce various emotions. We need to exhaustively collect both expressions for accurate emotion estimation.

Several studies have collected emotional expressions in the Japanese language. Nakamura published a dictionary of emotional words and phrases[1]. Tokuhisa et al. annotated emotion tags to text dialogs in Japanese comics with focusing facial expressions and constructed a reliable dialog corpus[2].

We aim to collect emotion-provoking events using onomatopoeias. There are numerous onomatopoeias in Japanese that can be used to express emotions.

First, we must understand the relationship between onomatopoeia words and emotions. Some dictionaries that contain many onomatopoeias have already been published[4][5]. However, these dictionaries are not sufficient to satisfy our purpose. As onomatopoeias are not clearly defined, readers (or listeners) have different perceptions of these words.

In this study, multiple raters classified 324 onomatopoeia words into emotion categories, and then, we quantitatively analyzed the degree of agreement among the raters. These onomatopoeias are perceived as words that refer to people’s senses and emotions. Moreover, we attempted to improve the degree of agreement by investigating the cause of disagreement and refining the experimental procedure.

Section II describes various types of emotions. Section III presents a preliminary experiment and formulates a hypothesis about the cause of disagreement. Section IV describes an experiment that was conducted to test the hypothesis. Finally, Section V concludes the paper with a discussion of future work.

II. TYPES OF EMOTIONS

It is difficult to determine the exact number of human emotions. Some researchers have attempted to subdivide emotions into categories. Ekman defined the basic emotions as anger, disgust, fear, happiness, sadness, and surprise[6]. In addition, he mentioned these emotions are universal to all humans. Nakamura developed an emotive expressions dictionary after a long-time study on words that describe emotional states in Japanese[1]. In Nakamura’s dictionary, direction words are classified into 10 emotions: joy, anger, sadness, fear, shame, fondness, dislike, excitement, relief, and surprise. Wundt claimed that emotions consist of three basic dimensions, each of which is a pair of opposite states: pleasantness/unpleasantness, excitement/relaxation, and tension/release[7]. Plutchik suggested that there are eight basic emotions grouped into four pairs of opposites: joy/sadness, anger/fear, trust/disgust, and surprise/anticipation[8]. In Plutchik’s view, all emotions are a combination of these basic emotions.

We used Plutchik’s classification of eight emotions in this study. The goal of this research is to capture emotions from
onomatopoeias, which are supposed to contain complex emotions. Since Plutchik’s classification can represent many emotions by combining the basic emotions, it is favorable to our goal.

III. PRELIMINARY EXPERIMENT

A method to classify onomatopoeias into emotional categories has not been established yet. We started with a hypothesis.

Hypothesis 1: If a rater is a native Japanese speaker, he/she can intuitively detect emotions contained in an onomatopoeia, and it will be mostly consistent with other raters.

To test this hypothesis, we conducted a preliminary experiment.

A. Experimental Procedure

We had raters decide what emotions are contained in each onomatopoeia. The raters were asked to check all appropriate emotions on the questionnaire form printed with onomatopoeias and emotional categories. They worked separately. As each onomatopoeia was presented in the form of a single word, context did not affect the judgment of the raters. In this preliminary experiment, the raters were 10 Japanese students (eight male and two female) in their 20s.

We used 324 onomatopoeias that are listed in the “Dictionary of Japanese Onomatopoeias[5]” as words referring to people’s senses and emotions. Table I shows some of these onomatopoeias.

<table>
<thead>
<tr>
<th>category in dictionary</th>
<th>number</th>
<th>examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>smile</td>
<td>75</td>
<td>abhahha, nikori</td>
</tr>
<tr>
<td>cry</td>
<td>47</td>
<td>uru-uru, gusun</td>
</tr>
<tr>
<td>anger</td>
<td>39</td>
<td>kachin, busuri</td>
</tr>
<tr>
<td>depressed</td>
<td>30</td>
<td>shun, tobo-tobo</td>
</tr>
<tr>
<td>rejoice</td>
<td>11</td>
<td>uki-uki, run-run</td>
</tr>
<tr>
<td>marvel</td>
<td>38</td>
<td>gikuri, doki-doki</td>
</tr>
<tr>
<td>fluster, struggle</td>
<td>39</td>
<td>jitarata, hara-hara</td>
</tr>
<tr>
<td>hesitate</td>
<td>10</td>
<td>guda-guda, moji-moji</td>
</tr>
<tr>
<td>feel</td>
<td>20</td>
<td>kyun, moya-moya</td>
</tr>
<tr>
<td>hurt</td>
<td>44</td>
<td>hiri-hiri, chiri-chiri</td>
</tr>
</tbody>
</table>

The emotional categories were Plutchik’s eight basic emotions, [none], and [unclear]. Category [none] means that there is no emotion in the target onomatopoeia, and category [unclear] means that a rater does not know the target onomatopoeia because it has become obsolete.

B. Experimental Results

1) Features of Classified Onomatopoeias: The raters chose 1.29 categories per onomatopoeia on average. The most chosen category was [joy], and the raters answered that on average, 55.1 onomatopoeias contained [joy]. The least chosen category was [uncertainty], and they answered that on average, 10.8 onomatopoeias contained [uncertainty].

Although the onomatopoeias used in this experiment refer to people’s senses and emotions, more than one rater chose [none] for 124 onomatopoeias. The onomatopoeias for which more than half of the raters chose [none] were “gasa-gasa,” “koro-koro,” “chika-chika,” “hiri-hiri,” among others. These words are usually used when people express the movement or condition of objects. Therefore, this result is reasonable.

The raters chose [unclear] for 39.5 onomatopoeias on average. We attempted to search those words on the Internet and received a limited number of hits. On the basis of this result, it may be said that the onomatopoeias judged to be [unclear] are words used less frequently in modern Japanese.

2) Agreement Rate: We used agreement rate to measure inter-rater agreement.

Agreement rate is defined as follows:

\[
AR(A, B) = \frac{N(R_A \cap R_B) \times 2}{N(R_A) + N(R_B)} \times 100
\]

where \(N(R_A)\) is the number of emotional categories that rater A chose, \(N(R_B)\) is the number of emotional categories that rater B chose, and \(N(R_A \cap R_B)\) is the number of emotional categories that both raters A and B chose.

Table II shows the agreement rate we calculated using the eight categories and eliminating [none] and [unclear]. The agreement rate among the 10 raters was 41.4% on average and ranged from 28.3% to 54.0%.

3) Kappa Coefficient: In addition, we used Cohen’s Kappa to measure inter-rater agreement[9]. The Kappa coefficient is a better measure of inter-rater agreement because it considers the effect of chance agreement.

The Kappa coefficient \(\kappa\) is defined as follows:

\[
\kappa = \frac{P_o - P_c}{1 - P_c}
\]

where \(P_o\) is the relative observed agreement among raters and \(P_c\) is the hypothetical probability of chance agreement. They are defined as:

\[
P_o = \sum_{i=1}^{k} p_i \cdot p_i, \quad P_c = \sum_{i=1}^{k} p_i \cdot p_{i,}
\]

where \(p_i\) is the probability that both raters choose category \(i\), \(\cdot\) is an arbitrary category, and \(k\) is the number of categories.

Table III shows the Kappa coefficient of all pairs of the 10 raters. The Kappa coefficient among the 10 raters was 0.600 on average and ranged from 0.338 to 0.835.

C. Consideration

The agreement rates between rater D and the other raters were relatively low. Rater D chose 1.07 categories per onomatopoeia, whereas the others chose 1.19 categories. Rater D classified 12 onomatopoeias into [joy], whereas...
the others classified an average of 55.1. Likewise, rater D classified 18 onomatopoeias into [sadness], whereas the others classified an average of 49.1. Rater D tended to choose multiple emotions less than the others did; therefore, his/her criteria for judging was distinct from the others. These factors are the likely cause of the low agreement rate.

Table IV shows a commonly used scale of the Kappa coefficient. According to the interpretation, the pairs of raters D–E, D–F, D–G, F–H, and D–J are in “fair agreement,” 17 pairs are in “moderate agreement,” 22 pairs are in “substantial agreement,” and the other pair is in “almost perfect agreement.”

More than half of the pairs are in “substantial agreement” or “almost perfect agreement,” and there is a certain level of agreement among the raters. However, the agreement rate and Kappa coefficient of some pairs were around 30% and less than 0.4, respectively. These results mean that it can be difficult to decide the emotion contained in onomatopoeias even for native Japanese speakers. Thus, hypothesis 1 is not completely satisfied.

An onomatopoeia often has multiple meanings. For example, in the Dictionary of Japanese Onomatopoeias, the word “jiri-jiri” is defined as (1) the sound of an alarm bell, (2) something that has been burned, (3) the scorching sun, (4) something that draws closer little by little, and (5) someone running out of patience and fretting. Here, the raters were expected to judge on the basis of meaning (5). However, some raters chose [disgust], [fear], or [none], which means they judged on the basis of meanings (1), (2), or (3).

The raters should have judged based on the same criteria. In order to unify criteria for judgment, we decided to use definition sentences of onomatopoeias.

We hypothesized that providing definitions of onomatopoeias would improve agreement among the raters.

Hypothesis 2: If given definitions of onomatopoeias, the degree of agreement among raters will be improved.

IV. EXPERIMENT WITH DEFINITION SENTENCES

A. Experimental Procedure

While the general procedure was the same as mentioned in Section III-A, we used the 122 onomatopoeias that fewer than half of the raters agreed on in the preliminary experiment. We reduced the target onomatopoeias for two reasons: first, it would not be difficult to judge the onomatopoeias with which more than half of the raters agreed; second, it is necessary to ease the burden on the raters. Incidentally, the average agreement rate and Kappa coefficient that were calculated using the 122 onomatopoeias in the preliminary experiment were 20.6% and 0.293, respectively.

In the preliminary experiment, the raters chose some emotion categories judging only by a single onomatopoeia. In this experiment, we used a questionnaire form printed with onomatopoeias and their definition sentences. The definition sentences were from the Dictionary of Japanese Onomatopoeias and were related to emotions.
The raters were five male Japanese students in their 20s and were different from those in the preliminary experiment.

B. Experimental Results

The average agreement rate among the five raters was 31.5% ranging from 16.8% to 48.2%. Moreover, the average Kappa coefficient among the five raters was 0.359, ranging from 0.275 to 0.447. Figure 1 shows the agreement rate and Kappa coefficient.

C. Consideration

Compared with the results of the preliminary experiment, the agreement rate increased from 20.6% to 31.5% and the Kappa coefficient increased from 0.293 to 0.359. However, considering that the average agreement rate and Kappa coefficient were calculated to be 57.4% and 0.671, respectively, using the onomatopoeias that more than half of the raters agreed on in the preliminary experiment, this is not sufficiently high. These results show that providing definition sentences of onomatopoeias works to some extent, but it is not enough. Therefore, it can be concluded that hypothesis 2 is confirmed, but is not perfect.

Fewer than half of the raters agreed on 34 onomatopoeias when given the definition sentences. Specific examples include “atafuta,” “shidomodo,” and “dogimagi.” These onomatopoeias mean “to fall into a flutter,” and they contain very complex emotions. The raters were divided according to the emotion categories they chose, including [disgust], [fear], [surprise]. Other examples are “kyapi-kyapi,” which means “to brim with youthful enthusiasm,” and “tsukuzuku,” which means “to be deeply implanted in someone’s mind.” Although the raters should have chosen [none], some raters chose [trust], [anticipation], [sadness] and [anger] judging from impressions of these words. This led to inconsistent results.

These 34 onomatopoeias would be unfit for Plutchik’s basic emotions or contain no emotion. Therefore, it is reasonable to say that these onomatopoeias should be eliminated from the list of words to classify.

V. CONCLUSION AND FUTURE WORK

In this paper, we presented emotion classification of 324 Japanese onomatopoeias. Ten raters chose emotions contained in each onomatopoeia from Plutchik’s eight basic emotions. As a result, the agreement rate was 41.4% on average and the Kappa coefficient was 0.600 on average. We achieved a certain level of inter-rater agreement; however, it was suggested that native Japanese speakers cannot always decide the emotion contained in an onomatopoeia. When the raters were given definition sentences of each onomatopoeia, the degree of agreement improved. Nevertheless, there were 34 onomatopoeia words that fewer than half of the raters agreed on, and we concluded that these onomatopoeias should not be classified according to Plutchik’s basic emotions.

Future work will aim to develop a method of extracting emotion-provoking events with the help of onomatopoeias.

ACKNOWLEDGMENT

This work was supported by MEXT/JSPS KAKENHI Grant Number (23700256).

REFERENCES