Use of Sound Symbolism in Sentiment Classification

Takuma Igarashi[†], Ryohei Sasano[†], Hiroya Takamura[†] and Manabu Okumura[†]

In linguistics, sound symbolism is an idea that the vocal sounds of certain words carry meaning in themselves. This paper focuses on the sound symbolism of onomatopoeic words and demonstrates the close relationship between sound symbolism and sentiment polarity. Because onomatopoeic words imitate the sounds they represent, they can help us better understand the sentiment of a sentence when utilizing sound symbolism. Therefore, we modeled sound symbolism with N-gram-based features and applied the model to a series of sentiment classification tasks. The experimental results show that this method with sound symbolism significantly outperformed the baseline method without sound symbolism, which effectively demonstrates that a close relationship exists between sound symbolism and sentiment polarity.

Key Words: *Sound Symbolism, Sentiment Classification, Onomatopoeia*

1 Introduction

According to Köhler (Köhler 1929), a close relationship exists between vocal sounds and visual impressions of shape. After experimentation, Ramachandran and Hubberd (Ramachandran and Hubbard 2001) named this phenomenon the **Bouba/Kiki effect** (Figure 1). A similar phenomenon is referred to as **sound symbolism** (David 2008), which is the idea that the vocal sounds of certain words carry impressions of the thing indicated by the words. Because an analysis of these impressions is beneficial to several natural language processing (NLP) applications, this paper examines sentiment classification to show the overall benefits of sound symbolism.

Fig. 1 Test subjects were asked the following question: “In the Martian language, one of these two figures is a ‘bouba’ and the other is a ‘kiki’. Try to guess which is which.” The findings show that 95% of the test subjects selected the left figure as “kiki” and the right figure as “bouba.”

[†] Tokyo Institute of Technology

Sentiment classification is the task of classifying words, sentences, and documents according to sentiment polarity. For example, the polarity is “positive” when the text suggests good sentiment, “negative” when the text suggests bad sentiment, and “neutral” when the text suggests neither good nor bad sentiment. The most common approach to sentiment classification is to apply adjectives (such as “delicious”) in context (Hu and Liu 2004; Kamps, Marx, Mokken, and de Rijke 2004). In these studies, the researchers prepared a lexicon that contained adjectives and their polarities in advance, and then utilized the lexicon to estimate the polarity of the target text. Consider the following sentences:

- (1) This biscuit is delicious.
- (2) I lost the key for my house.

Because “delicious” usually includes positive polarity in common lexicons, Sentence (1) is classified as “positive.” This adjective-based approach, however, cannot estimate the polarity of sentences that do not include adjectives, such as Sentence (2). To resolve this problem, some researchers have expanded the adjective polarity lexicon to other parts of speech by using co-occurrence information in corpus (Turney and Littman 2003; Takamura, Inui, and Okumura 2005). This can be considered as a type of bootstrap approach in which the polarities of various words are estimated from several seeds, the polarities of which are known from the lexicon.

However, such approach cannot adequately deal with words that are not included in training data for learning models. In particular, Japanese onomatopoeia is used at a relatively high frequency. We searched for onomatopoeic words in a variety of restaurant reviews by using part-of-speech (POS) taggers (JUMAN)¹ in advance and found that approximately 6.6% of sentences in the restaurant reviews contained onomatopoeic words, some of which were not included in common dictionaries. This finding shows that there are a number of unknown onomatopoeic words in Japanese texts. Because some sentences owe their sentiment polarity to onomatopoeic words, the effort to estimate the polarity of onomatopoeic words is extremely important.

Consider the following example sentences:

- (3) *Te-ga **bettori** suru.*
hands sticky are
My hands are sticky.
- (4) *Te-ga **becho-becho** suru.*
hands humid are

¹ <http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN>

My hands are humid.

Bettori and *becho-becho* are onomatopoeic words in Japanese and both sentences above have negative polarity. *Bettori* is commonly used and included in a typical lexicon, whereas *becho-becho* is not normally used because lexicon-based approaches cannot adequately deal with such unknown onomatopoeic words.

On the other hand, onomatopoeia reflects sound symbolism to a significant extent (Hamano 1998). We can understand the nuance of onomatopoeic words from the impression of sound symbolism even if we do not understand the meaning of the words themselves. Because modeling sound symbolism can help estimate the polarity of unknown onomatopoeic words, we present a modeling method in order to demonstrate its effectiveness in sentiment classification. The reason for selecting this task is that onomatopoeic words often reflect sound symbolism and are used for representation of sentiment.

2 Background

Onomatopoeic words either imitate the sounds they represent or suggest a relevant situation or emotion. There are literally thousands of onomatopoeic words in the Japanese language. Consider the following examples:

- (5) *Kare-ha **niko-niko** waratta.*

He smiled.

- (6) *Kare-ha **niya-niya** waratta.*

He simpered.

Niko-niko is an onomatopoeic word that indicates a bland smile, and *niya-niya* is an onomatopoeic word that represent a simper. *Waratta* is a generic verb that indicates a smile, laugh, chuckle, simper, or guffaw. In Japanese, such situational or emotional information is represented by onomatopoeic words that are separate from verbs; whereas in English, verbs tend to include such information.

Onomatopoeic words have phonological and morphological systematicity. For example, *niko-niko*, *nikotto*, and *nikkori* suggest similar situations. *Niko* is the **base** form of these words and morphemes are created from base forms through a type of conjugation. According to Waida (Waida 1984), the patterns of conjugation are known as **onomatopoeia markers**. Kadooka (Kadooka 2002) listed these markers in the manner shown in Table 1.

The meaning of onomatopoeic words can often be guessed from sound symbolism.

Table 1 Onomatopoeia Markers in Japanese. Reduplication pattern is the most frequent among them.

Markers	Examples	Meaning
bare stem	<i>hishi</i>	to hug firmly
altered reduplication	<i>gasa-goso</i>	to look for something
doubled base	<i>butsu-kusa</i>	grumbling voice
reduplication	<i>koro-koro</i>	something rolling
others	<i>gui</i>	jerking action

(7) *Kare-ha **gehi-gehi** waratta.*

He vulgarly laughed.

As observed in Sentence (7), people can guess that the onomatopoeic word *gehi-gehi* includes negative nuances even if the word is unknown. Hamano (Hamano 1998) studied the sound symbolism of Japanese onomatopoeic words and found that voiced consonants often created a negative impression. In addition, she qualitatively manually analyzed nuances of phonemes in Japanese. For example, /p/ and /b/ have nuances of explosion, breaking, and decisiveness, and /r/ has nuances of rolling and fluid movement. On the other hand, our method focuses on sentiment polarity (positive, negative, and neutral) and obtains features of sound symbolism automatically. (Sokolova and Bobicev 2009) studied the classification of emotion words in the Russian and Romanian languages, and they also utilized word sounds for the classification of words. However, they used alphabetical letters to approximately represent the sound, whereas we apply the International Phonetic Alphabet (IPA)², which is a more accurate approach to representing sounds (Section 3.2). In addition, their target was the classification of words without context, whereas we focus on how words behave in terms of sentiment polarity.

Recently, research on onomatopoeia has become popular in the computer science field, including the area of NLP. For example, (Komiya and Kotani 2011) proposed a method for clustering Japanese onomatopoeic words by using single-link hierarchical clustering based on context. (Ichioka and Fukumoto 2008) proposed a method for clustering Japanese onomatopoeic words on the basis of the co-occurrence of onomatopoeic words on the Web and the vocal sound similarity of the words. However, the vocal sounds they considered were insufficient to represent pronunciation. The purpose of these aforementioned studies was to help people understand the meaning of onomatopoeic words, especially for non-native speakers. In contrast, the goal of this paper is to demonstrate the effectiveness of sound symbolism in sentiment classification.

² <http://www.langsci.ucl.ac.uk/ipa/>

In regard to sentiment classification tasks, recent studies have used various features and advanced models. For example, (Kudo and Matsumoto 2004) used text structures, whereas (Täckström and McDonald 2011) utilized star ratings and hidden conditional random fields (CRFs). However, we simply use vocal sound features and bag-of-words features because our primary goal is to demonstrate the effectiveness of sound symbolism for interpreting the nuances of onomatopoeic words.

3 Proposed Method

Our method estimates the polarity label of a sentence that includes onomatopoeic words. As shown in Figure 2, the system receives a sentence that includes onomatopoeic words, classifies it into one of the three classes (positive, negative, and neutral) by using contextual and sound symbolism features, and outputs the polarity label. The proposed system is based on the support vector machine (SVM) introduced by (Cortes and Vapnik 1995). The bag-of-words features and the context polarity feature are baseline features that are not related to sound symbolism (these are described in more detail in Section 3.1). The vocal sound features and onomatopoeia dictionary feature are sound symbolism features, which are described in Sections 3.2 and 3.3. The sound symbolism model, which is based on an onomatopoeia dictionary using vocal sound features, takes an onomatopoeic word as input and outputs a binary value of 1 if the inputted

Fig. 2 Overview of our system.

word is estimated to be positive and 0 otherwise. The main SVM classifier then uses this value as an onomatopoeia dictionary feature.

3.1 Baseline Features

Because context information is often used for sentiment classification, this study first models context information with the bag-of-words model and then selects content words in the input sentence and lemmatizes them. Finally, the number of occurrences of each word for feature values are accounted for and normalized.

In addition, we use a sentiment lexicon created by (Takamura et al. 2005), which they assessed the polarity probability of words in WordNet (Fellbaum 1998) with a **spin model**, assuming that adjacent words in context tend to have the same polarity. This lexicon includes words and their polarity values and the range of value is $[-1, 1]$. Polarity values of less than 0 indicate negative polarity and values of more than 0 indicate positive polarity. Furthermore, the absolute values denote the confidence value. For example, the value of the word “beautiful” is $+0.371463$ and that of “noisy” is -0.218053 . This shows that “beautiful” is positive and its confidence value is 0.371463 , and “noisy” is negative and its confidence value is 0.218053 . We selected 2,000 words for each polarity in order of decreasing confidence value³ (4,000 words in total) and used them as a clue to interpret contextual polarity. Contextual polarity is a binary feature that is based on the sum of the words’ values in context, in which the feature value is 1 if the sum is more than 0, and 0 otherwise. For example, in the sentence “the smell is nice,” the polarity value of “smell” is -0.0174731 and “nice” is $+1$. Therefore, the sum is 0.9825269 and the context polarity feature is 1. The baseline features consist of the context polarity feature and bag-of-words features.

3.2 Vocal Sound Features

This study considers three overall forms in the representation of Japanese vocal sounds:

(1) **Kana characters**

“Kana” is a type of Japanese character. Each kana consists of a consonant onset and vowel nucleus. For example, “は (ha)” or “い (i).”

(2) **Phoneme symbols**

A phoneme is a basic unit of a language’s phonology that describes vocal sounds recognized by native speakers in a language. The phoneme /k/ in the words **kit** or **skill** is recognized to be the same sound even though they are different sounds, as seen in $[k^h]$ (aspirated) for

³ Strictly speaking, we use Japanese sentiment lexicon (not English) that is published by Takamura. <http://www.lr.pi.titech.ac.jp/takamura/pubs/pn-ja.dic>.

kit and [k] (unaspirated) for skill.

(3) Phonetic symbols

Phonetics is an objective and a strict representation of all vocal sounds. Unlike phonemes, vocal sounds such as [k^h] and [k] are distinguished in phonetic symbols. This study utilizes the International Phonetic Alphabet (IPA) for this representation.

Table 2 shows the examples of Japanese kana with phonemes and phonetic representations. In general, phoneme symbols are written with slant brackets (/ /), whereas phonetic symbols are written with square brackets ([]). Kana characters are conventionally classified by consonant and vowel sounds (Miller 1967). Phoneme symbols are assigned to each kana character on the basis of this interpretation, and thus, different sounds tend to be regarded as the same phoneme symbol, as seen in the difference between “は (/ha/, [ha])” and “ひ (/hi/, [çi]).” In general, phonetic symbols are defined in detail and they do not depend on languages.

These representations are used to create sound symbolic features and we convert onomatopoeic words to each of the three forms and create N-gram features from them. For example, an onomatopoeic word “しとしと (*shito-shito*, drizzingly)” is converted to [ʃitofitof] with phonetic symbols. The N-gram features are derived from phonetic representation: [ʃ], [i], [t], [o], [ʃi], [it], [to], and [of] (unigram and bigram), whereas the feature values are calculated from the feature frequency in the target word and then normalized with total occurrences for each unigram and

Table 2 Kana, phonemes and phonetic transcriptions in Japanese.

	φ	/k/	/s/	/t/	/n/	/h/	/m/	/y/	/r/	/w/	
	あ	か	さ	た	な	は	ま	や	ら	わ	ん
/a/	/a/	/ka/	/sa/	/ta/	/na/	/ha/	/ma/	/ya/	/ra/	/wa/	/n/
	[a]	[ka]	[sa]	[ta]	[na]	[ha]	[ma]	[ja]	[ra]	[wa]	[N]
	い	き	し	ち	に	ひ	み		り	ゐ	
/i/	/i/	/ki/	/shi/	/chi/	/ni/	/hi/	/mi/		/ri/	/wi/	
	[i]	[k ⁱ i]	[ʃi]	[tçi]	[n ⁱ i]	[çi]	[m ⁱ i]		[r ⁱ i]	[i]	
	う	く	す	つ	ぬ	ふ	む	ゆ	る		
/u/	/u/	/ku/	/su/	/tsu/	/nu/	/fu/	/mu/	/yu/	/ru/		
	[u]	[k _ɥ u]	[s _ɥ u]	[ts _ɥ u]	[n _ɥ u]	[f _ɥ u]	[m _ɥ u]	[j _ɥ u]	[r _ɥ u]		
	え	け	せ	て	ね	へ	め		れ	ゑ	
/e/	/e/	/ke/	/se/	/te/	/ne/	/he/	/me/		/re/	/we/	
	[e]	[ke]	[se]	[te]	[ne]	[he]	[me]		[re]	[e]	
	お	こ	そ	と	の	ほ	も	よ	ろ	を	
/o/	/o/	/ko/	/so/	/to/	/no/	/ho/	/mo/	/yo/	/ro/	/wo/	
	[o]	[ko]	[so]	[to]	[no]	[ho]	[mo]	[jo]	[ro]	[o]	

Table 3 Consonant categories in the IPA. This study considered only sounds that appear in Japanese.

1. plosive (e.g., [p], [k])	2. fricative (e.g., [f], [s])	3. affricate (e.g., [tʃ], [dʒ])
4. nasal (e.g., [m], [n])	5. flap or tap (e.g., [ɾ], [ɽ])	6. approximant (e.g., [j], [w])
7. bilabial (e.g., [p], [b])	8. alveolar (e.g., [t], [n])	9. palatal (e.g., [ç], [j])
10. velar (e.g., [k], [g])	11. coronal (e.g., [z], [ʃ])	12. dorsal (e.g., [ɖ], [ɳ])
13. voiced (e.g., [g], [z])		

bigram. These features denote what sounds exist and what connections of the sounds exist in a word, but they do not consider relationships among them. For example, [m] and [p] are simply regarded as different sounds, although both are bilabial consonants. Therefore, this study utilizes consonant categories in the IPA to consider such relationships. As shown in Table 3, items 1–6 are categories on **manner of articulation**, 7–10 are categories on **place of articulation**, 11–12 are superior categories of **place of articulation**, and 13 indicates whether a sound is voiced or voiceless. We use these categories as binary features for the first and second consonants of an onomatopoeic word on the basis of (Hamano 1998), which states that the **base** of an onomatopoeic word generally includes the first and second consonants, and that consonants in the base play a key role in sound symbolism. In this case, the first consonant in an onomatopoeic word is considered more important than the second consonant.

3.3 Which Representation is the Best?

Our hypothesis is that sound symbolic impressions of onomatopoeic words are composed of the vocal sounds of words. In addition, we believe that a more detailed representation is better for modeling sound symbolism. In other words, phoneme symbols are better than kana, and phonetic symbols are better than phoneme symbols. Therefore, we conducted experiments to test this hypothesis.

When the classifier takes an onomatopoeic word, it classifies the word as either positive or negative. For this experiment, we began by first training an SVM classifier⁴. Then, we used a Japanese onomatopoeic dictionary (Hida and Asada 2002) that consisted of 1,064 onomatopoeic words along with the impressions of each categorized into five values: “positive,” “a little positive,” “neither positive nor negative,” “a little negative,” and “negative.” Out of the total number of onomatopoeic words, we selected 845 that did not contain contradictory impressions (such as “positive” and “a little negative” in the same word), and then classified 225 onomatopoeic words

⁴ We used SVM^{light} <http://svmlight.joachims.org/>.

that were “positive” or “a little positive” into the positive class. Finally, we selected 620 onomatopoeic words that were “negative” or “a little negative” and classified them into the negative class. We evaluated the performance for accuracy by performing 10-fold cross-validation.

As shown in Table 4, *Baseline* is the model that classifies test instances into the most frequent class in training data, i.e., it always classifies test data as “negative.” *Kana*, *Phoneme*, and *IPA* use only the N-gram features with each vocal sound representation, as detailed in Section 3.2. *IPACat* is a model with consonant category features in addition to *IPA*, and *IPACat2* is the model with consonant bigram features in addition to *IPACat*. The consonant bigram features are created from sequences of only those consonants in the IPA that are obtained to eliminate vowels from the phonetic representation of onomatopoeic words. For example, the onomatopoeic word “しとしと (*shito-shito*)” is converted into a phonetic representation [ʃitofʃito] and then further converted to [ʃtʃt] by eliminating the vowels. In this case, the consonant bigram features are [ʃt] and [tʃ]. We considered the consonant bigram features in the experiment because we believe that consonant connections can affect the overall impression of sound symbolism.

The experimental result is shown in Figure 3. The horizontal axis indicates the classification accuracy⁵. Overall, *Phoneme* achieved a higher accuracy compared with *Kana*, and *IPA* achieved a higher accuracy than *Phoneme*. *IPA*, *IPACat*, and *IPACat2* use phonetic symbols. In addition, *IPACat* was better than *IPA* with simple N-grams, and *IPACat2* proved to be the best overall. *IPACat* significantly outperformed *Kana* using McNemar’s test ($P < 0.01$), and *IPACat2* also significantly outperformed *Phoneme* in the same test ($P < 0.05$). These findings suggest that the hypothesis in Section 3.2 is valid and IPA phonetic representation is the best representation for modeling sound symbolism among the kana, phoneme, and phonetic representations. Thus, we used the *IPACat2* model as the sound symbolism model in the proposed system (Figure 2).

Table 4 Models for estimating the polarity of onomatopoeic words.

Model	Features
<i>Baseline</i>	(Classified as the most frequent class in training data.)
<i>Kana</i>	Kana N-gram (uni-, bi-, and trigram)
<i>Phoneme</i>	Phoneme N-gram (uni-, bi-, and trigram)
<i>IPA</i>	Phonetic N-gram (uni-, bi-, and trigram)
<i>IPACat</i>	Phonetic N-gram (uni-, bi-, and trigram), IPA category
<i>IPACat2</i>	Phonetic N-gram (uni-, bi-, and trigram), IPA category, consonant bigram

⁵ We empirically choose the best C parameter for SVM from {0.01, 0.1, 1, 10, 100, 1000} for each model.

Fig. 3 Experimental results of estimating the polarity of onomatopoeic words. Phonetic representation based on the IPA is the best form of representation among the kana, phoneme, and phonetic representations.

4 Experiments

The proposed system estimates the polarity of a sentence that includes onomatopoeic words. In this study, we present experiments to demonstrate that sound symbolism features are useful in sentiment classification. When the system takes as input a sentence that includes an onomatopoeic word, it classifies the sentence into one of the three classes (positive, negative, or neutral) and outputs the class label of the sentence.

4.1 Experimental Setup

In the experiment, we applied documents collected from tabelog.com⁶, which is a restaurant review web site in Japan. The variety of reviews includes more onomatopoeic words compared with other domains because onomatopoeic words are subjective and convenient for describing the tastes of food and drink. We extracted sentences that included onomatopoeic words with a Japanese POS tagger (JUMAN 7.0) that could automatically recognize onomatopoeic words even if they were not contained in the JUMAN dictionary by using morphologic patterns. Then, we manually selected 1,000 sentences for 50 onomatopoeic words⁷ (20 sentences per onomatopoeic word). Because our goal was to focus on the sound symbolism of onomatopoeia, we excluded the sentences that included negation words that would affect the overall sentence polarity.

The gold standard labels were tagged by two annotators, and any discrepancies were discussed by the annotators in order to decide on a label. The labels, which are shown in Table 5, are

⁶ <http://tabelog.com/>

⁷ We listed them in appendix.

Table 5 Labels annotated for each instance in a dataset.

Label	No. of Instances	Details	Examples
Label 1	276	Target sentence is positive. The polarity depends on the onomatopoeic word included in the sentence.	<i>Sakkuri-shita shokkan.</i> crisp food texture Crisp food texture.
Label 2	76	Target sentence is positive. The polarity does not depend on the onomatopoeic word included in the sentence.	<i>Piri-piri-shita sansho-ga oishi.</i> hot pepper nice The hot pepper is nice.
Label 3	170	Target sentence is negative. The polarity depends on the onomatopoeic word included in the sentence.	<i>Teburu-ga beto-beto shiteru.</i> table sticky The table is sticky.
Label 4	29	Target sentence is negative. The polarity does not depend on the onomatopoeic word included in the sentence.	<i>Gun-to kakaku-ga agatta.</i> rapidly price rose The price rose rapidly.
Label 5	272	Target sentence is neutral: neither positive nor negative.	<i>Akari-ha bou-tto-shita orenji-iro.</i> lighting dim orange The lighting is dim orange.
Label 6	177	The polarity cannot be identified from target sentence.	<i>Kuruma-ga zurari-to narandeta.</i> cars a line of There was a line of cars.

grouped into positive (Labels 1 and 2), negative (Labels 3 and 4), neutral (Label 5), and other (Label 6). We removed the data of Label 6 from the dataset because they could not be identified by human annotators, and therefore, they were considered as “noise.” To investigate whether sound symbolism is more useful in a limited dataset (Labels 1, 3, and 5), positive and negative classes were divided into two on the basis of polarity dependence of an onomatopoeic word. We designated the complete dataset as Dataset1 and the limited dataset as Dataset2. Dataset1 consisted of 823 instances⁸ and Dataset2 consisted of 718 instances⁹, and 49 unique onomatopoeic words were included in each dataset¹⁰.

Table 6 lists the label distribution, and shows that the concordance rate was not high: the kappa coefficient was $\kappa = 0.57$ in Dataset1 when Labels 1 and 2 were designated as one set, and Labels 3 and 4 were regarded as another. However, the annotators did not tag opposite labels (e.g., Label 1 versus Label 3). For example, the sentence *Okyaku-san ga **choko-choko** kiteta* was

⁸ The number is the sum of 276, 76, 170, 29 and 272.

⁹ The number is the sum of 276, 170 and 272.

¹⁰ All of the 20 sentences for a certain onomatopoeic word were labeled as Label 6, and thus, the number of unique onomatopoeic words was reduced to 49.

Table 6 Distribution of annotation results.

	Annotator A						
		Label 1	Label 2	Label 3	Label 4	Label 5	Label 6
Annotator B	Label 1	156	4	0	0	17	16
	Label 2	75	14	0	0	14	6
	Label 3	0	0	101	1	7	17
	Label 4	0	0	25	6	2	8
	Label 5	44	6	34	9	251	72
	Label 6	11	1	12	2	43	46

labeled as Label 2 because the annotator thought *Okkyaku-san ga kiteta* itself was positive. On the other hand, the same sentence was labeled as Label 1 because the annotator believed that *Okkyaku-san ga kiteta* was neutral. Such differences of sensitivity caused certain discrepancies in annotation.

In this experiment, we evaluated the performance of classifying sentences that included unknown onomatopoeic words, i.e., onomatopoeic words that appeared in the test and not the training set. We divided the dataset into 10 parts and used one part for testing, another for development, and the remainder for training. In addition, we used a linear SVM with a one-versus-rest method and evaluated the performance for accuracy by performing 10-fold cross-validation¹¹.

We tested the following five settings:

M_B	BOW, <i>Context Polarity</i> (Baseline)
M_{BP}	BOW, <i>Context Polarity</i> , <i>Phonetic</i>
M_{BPD}	BOW, <i>Context Polarity</i> , <i>Phonetic</i> , <i>Onomatopoeia Dictionary</i>
M_{BD}	BOW, <i>Context Polarity</i> , <i>Onomatopoeia Dictionary</i>
M_D	<i>Onomatopoeia Dictionary</i>

The baseline method M_B utilizes a bag-of-words model (BOW) and context polarity (as described in Section 3.1). This model does not consider sound symbolism, unlike the other models. The *Phonetic* features are the same as *IPACat2* in Section 3.3, which are the N-gram features of phonetic symbols (uni-gram, bi-gram, and tri-gram), consonant categories in the IPA, and consonant bi-gram. The *Onomatopoeia Dictionary* feature is a binary feature that is the output of the sound symbolism model (see Figure 2). This model is the same as *IPACat2*, and

¹¹ The C parameter of SVM was automatically chosen as the best value in {0.1, 1, 10, 100} after testing with the development set.

it was trained on a dataset that contained 820 onomatopoeic words¹². Its feature value was designated as 1 if the output was 0 or more and 0 otherwise.

4.2 Results and Discussion

According to the experimental results shown in Figure 4, Dataset1 (baseline) had 47.27% accuracy, whereas M_{BD} , at 62.94%, had the strongest performance. There was a significant difference between M_{BD} and the baseline using McNemar’s test ($P < 0.01$). The other methods (M_{BP} , M_{BPD} , and M_D) outperformed the baseline by a significant margin as well ($P < 0.01$). The *Onomatopoeia Dictionary* feature was effective because the sound symbolism model acquired sound symbolic impressions from numerous onomatopoeic words. M_D achieved an accuracy of 59.78%, even though it used only the *Onomatopoeia Dictionary* feature. The accuracy of M_{BPD} using all of the features was lower compared with that of M_{BD} because of the overfitting of *Phonetic* features, which were acquired from approximately 40 onomatopoeic words in the training data.

In Dataset2, the accuracy of the baseline was just 44.85%, which was lower than Dataset1. This decline is understandable because Dataset2 contained only limited data in which the polarity of a sentence was dependent on an onomatopoeic word. M_{BD} again achieved the best performance, with an accuracy of 62.40% and M_D also had a high accuracy because Dataset2

Fig. 4 Result of experiment to estimate the polarity of sentences that include onomatopoeic words. Proposed methods significantly outperform the baseline method M_B .

¹² No onomatopoeic words in this dataset appeared in the tabelog.com data.

contained only limited data where the polarity of a sentence was dependent on an onomatopoeic word. Models that used *Phonetic* features had poor results because of overfitting, but they showed improvement if they were fed additional training data.

5 Conclusion and Future Work

This study modeled sound symbolism with vocal sound representations of onomatopoeic words and showed that the polarity of unknown onomatopoeic words can be estimated by using the N-gram features of phonetic representation and consonant category features of phonetic symbols. The experimental results clearly showed that the methods using sound symbolism significantly outperformed the baseline model, which indicates that sound symbolism can be useful in sentiment classification tasks, and a close relationship exists between sound symbolism and sentiment polarity. However, the accuracy of the proposed methods was not sufficiently high. When an individual infers the polarity of onomatopoeic words, there can be many different interpretations depending on the person. Needless to say, further improvement on the accuracy of this method is a challenging task. However, even though weighting the features and feature selection is one approach of increasing accuracy, we did not include these in the present method because our primary goal was to demonstrate the effectiveness of sound symbolism without overcomplicating the process.

In future, we have three overall goals regarding research on this topic. The first is to model sound symbolism in more detail. To date, we have primarily considered the phonological features of onomatopoeia; however, morphological features can be related to sentiment. Considering that **onomatopoeia markers** is a key factor in this task, it will be interesting to see which markers indicate which impressions.

The second is to create higher quality training data by collecting more opinions in order to correctly model the sound symbolism of onomatopoeic words, as in the semantic differential method (Osgood, Suci, and Tennenbaum 1957). In the present study, the data was only tagged by two annotators.

The third is to expand the domain adaptation because we applied our method to only the food and drink domain. Because only context information and not auditory impressions are related to this domain, we believe that sound symbolism features can be expanded and applied to sentiment classification in various domains.

Acknowledgement

A portion of this paper has been presented at the 12th Pacific Rim International Conference on Artificial Intelligence (PRICAI 2012) (Igarashi, Sasano, Takamura, and Okumura 2012).

Reference

- Cortes, C. and Vapnik, V. (1995). “Support-vector networks.” *Machine Learning*, **20** (3), pp. 273–297.
- David, C. (2008). *A Dictionary of Linguistics and Phonetics*. Wiley-Blackwell.
- Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*. MIT Press.
- Hamano, S. (1998). *The Sound-Symbolic System of Japanese*. CSLI Publications.
- Hida, Y. and Asada, H. (2002). *Gendai Giongo Gitaigo Yoho Jiten (In Japanese, ISBN: 9784490106107)*. Tokyo-do.
- Hu, M. and Liu, B. (2004). “Mining and Summarizing Customer Reviews.” In *Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining (KDD2004)*, pp. 168–177.
- Ichioaka, K. and Fukumoto, F. (2008). “Graph-Based Clustering for Semantic Classification of Onomatopoeic Words.” In *Proceedings of the 3rd Textgraphs workshop on Graph-based Algorithms for Natural Language Processing in the 22th International Conference on Computational Linguistics (Coling2008)*, pp. 33–40.
- Igarashi, T., Sasano, R., Takamura, H., and Okumura, M. (2012). “The Use of Sound Symbolism in Sentiment Classification.” In *Proceedings of the 12th Pacific Rim International Conference on Artificial Intelligence (PRICAI2012)*, pp. 746–752.
- Kadooka, K. (2002). “Onomatopoeia Markers in Japanese.” *LACUS FORUM*, **28**, pp. 267–275.
- Kamps, J., Marx, M., Mokken, R. J., and de Rijke, M. (2004). “Using WordNet to Measure Semantic Orientations of Adjectives.” In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC2004)*, pp. 1115–1118.
- Köhler, W. (1929). *Gestalt Psychology*. New York: Liveright.
- Komiya, K. and Kotani, Y. (2011). “Classification of Japanese onomatopoeias using hierarchical clustering depending on contexts.” In *Proceedings of the 8th English International Joint Conference on Computer Science and Software Engineering (JCSSE2011)*, pp. 108–113.
- Kudo, T. and Matsumoto, Y. (2004). “A Boosting Algorithm for Classification of Semi-Structured Text.” In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language*

Processing (EMNLP2004), pp. 301–308.

Miller, R. A. (1967). *The Japanese language*. Chicago: University of Chicago Press.

Osgood, C. E., Suci, G. J., and Tennenbaum, P. (1957). *The Measurement of Meaning*. University of Illinois Press.

Ramachandran, V. S. and Hubbard, E. M. (2001). “Synaesthesia: A window into perception, thought and language.” *Journal of Consciousness Studies*, **8** (12), pp. 3–34.

Sokolova, M. and Bobicev, V. (2009). “Classification of Emotion Words in Russian and Romanian Languages.” In *Proceedings of Recent Advances in Natural Language Processing (RANLP2009)*, pp. 416–420.

Täckström, O. and McDonald, R. (2011). “Discovering fine-grained sentiment with latent variable structured prediction models.” In *Proceedings of the 33rd European conference on Advances in information retrieval (ECIR2011)*, pp. 368–374.

Takamura, H., Inui, T., and Okumura, M. (2005). “Extracting semantic orientations of words using spin model.” In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics (ACL2005)*, pp. 133–140.

Turney, P. D. and Littman, M. L. (2003). “Measuring Praise and Criticism: Inference of Semantic Orientation from Association.” *ACM Transactions on Information Systems (TOIS)*, **21** (4), pp. 315–346.

Waida, T. (1984). “English and Japanese Onomatopoeic Structures.” *Bulletin of Osaka Women’s University, Studies in English*, **36**, pp. 55–79.

Appendix

Table 7 shows the onomatopoeic words used in Section 4.1. Apostrophes at the end of the words represent double consonants called *sokuon*.

Table 7 Onomatopoeic words in the dataset.

<i>choko-choko</i> (ちょちょこ)	<i>gusshori</i> (ぐっしり)	<i>pittari</i> (びったり)	<i>hara-hara</i> (はらはら)
<i>fura-fura</i> (ふらふら)	<i>kongari</i> (こんがり)	<i>gakkari</i> (がっかり)	<i>sakkuri</i> (さっくり)
<i>beto-beto</i> (べとべと)	<i>jiri-jiri</i> (じりじり)	<i>zoro-zoro</i> (ぞろぞろ)	<i>kiri'</i> (きりっ)
<i>mekkiri</i> (めっきり)	<i>sukkiri</i> (すっきり)	<i>sara-sara</i> (さらさら)	<i>hoku-hoku</i> (ほくほく)
<i>piri-piri</i> (ぴりぴり)	<i>koro-koro</i> (ころころ)	<i>mera-mera</i> (めらめら)	<i>bara-bara</i> (ばらばら)
<i>bassari</i> (ばっさり)	<i>gisshiri</i> (ぎっしり)	<i>hishi-hishi</i> (ひしひし)	<i>horori</i> (ほろり)
<i>sowa-sowa</i> (そわそわ)	<i>kusu-kusu</i> (くすくす)	<i>gorori</i> (ごろり)	<i>kichi'</i> (きちっ)
<i>hera-hera</i> (へらへら)	<i>zushi'</i> (ずしっ)	<i>mou-mou</i> (もうもう)	<i>surari</i> (すらり)
<i>shan</i> (しゃん)	<i>sâ'</i> (さーっ)	<i>gûn</i> (ぐーん)	<i>uki-uki</i> (うきうき)
<i>bou'</i> (ぼうっ)	<i>huu-huu</i> (ふうふう)	<i>gururi</i> (ぐるり)	<i>sû'</i> (すーっ)
<i>usu-usu</i> (うすうす)	<i>potsun</i> (ぽつん)	<i>zushiri</i> (ずしり)	<i>sa'</i> (ざっ)
<i>choko'</i> (ちょこっ)			

Takuma Igarashi: received his B.E. in 2008 from Aoyama Gakuin University, Japan, and his M.E. in 2012 from Tokyo Institute of Technology, Japan. He is currently employed as an engineer at Dwango Co., Ltd.

Ryohei Sasano: received his B.E., M.S., and Ph.D. in Information Science and Technology from the University of Tokyo in 2004, 2006, and 2009, respectively. He was a researcher at the Graduate School of Informatics at Kyoto University, and he is currently an assistant professor in the Precision and Intelligence Laboratory at the Tokyo Institute of Technology. His research interests include natural language processing, especially predicate-argument structure analysis and anaphora resolution.

Hiroya Takamura: was born in 1974. He received his B.E. and M.E. from the University of Tokyo in 1997 and 2000, respectively. In 1999, he was a research student at the Technische Universitaet von Wien in Austria. He received his Dr.Eng. from the Nara Institute of Science and Technology in 2003. He was an assistant professor at the Tokyo Institute of Technology from 2003 to 2010.

He is currently an associate professor at Tokyo Institute of Technology. His current research interests include computational linguistics. He is a member of the Information Processing Society of Japan and the Association for Computational Linguistics.

Manabu Okumura: was born in 1962. He received his B.E., M.E. and Dr.Eng. from the Tokyo Institute of Technology in 1984, 1986, and 1989, respectively. He was an assistant professor in the Department of Computer Science at the Tokyo Institute of Technology from 1989 to 1992, and an associate professor in the School of Information Science at the Japan Advanced Institute of Science and Technology from 1992 to 2000. He was also a visiting associate professor in the Department of Computer Science at the University of Toronto from 1997 to 1998. He is currently a professor in the Precision and Intelligence Laboratory at the Tokyo Institute of Technology. His current research interests include natural language processing, especially automatic text summarization, computer-assisted language learning, sentiment analysis, and text data mining. He is a member of the Information Processing Society of Japan, the Japanese Society for Artificial Intelligence, the American Association for Artificial Intelligence, the Association for Computational Linguistics, and the Japanese Cognitive Science Society.

(Received October 24, 2012)

(Revised February 15, 2013)

(Accepted March 28, 2013)