# Japanese Emotion Corpus Analysis and its Use for Automatic Emotion Word Identification

Junko Minato, David B. Bracewell, Fuji Ren, and Shingo Kuroiwa *

**Abstract**

In this paper, the creation of a Japanese emotion corpus and its use in automatic emotion word identification are examined. The corpus was created by manually tagging words in just under 1,200 dialog sentences with emotion. Using the tagged corpus, statistical analysis was performed to determine the characteristics of emotional expression in Japanese dialog. This type of analysis should prove beneficial for understanding how emotion is expressed and how to identify, classify, etc. emotion in Japanese. To test this theory an automatic emotion word identification system was built using machine learning based classifiers with features taken from the statistical analysis. In total, four different classifiers were trained and compared to a baseline dictionary approach. It was found that classifier based identification was able to significantly increase recall.

*Keywords: Natural Language Processing, Statistical Analysis, Affective Computing, Corpus Creation*

## 1 Introduction

Development of information technology has increased the chances of interaction between humans and computers. As computers become more prevalent in daily life, the focus is transitioning from merely transmitting factual information in order to allow for a more enjoyable and comfortable experience for humans. The new trend is to include the communication of information such as how the speaker/writer feels about the fact or how they want the listener/reader to feel. The first step in realizing this new advanced communication is to recognize emotion in text from document to word level.

In order to obtain knowledge and information from emotional text it is necessary to have reliable linguistic resources, such as tagged emotion corpora and emotion dictionaries. As the study of emotion recognition combined with natural language processing is rather new, it is still difficult to obtain such linguistic resources. For that reason many researchers have started creating such corpora, such as [16], [5] and [9].

Additionally, research has been done on emotion dictionary construction. Takamura et.al [14] regarded semantic orientations as semantic electrons and tried to make a list of desirable and undesirable words. Some of the earliest work done on emotions in Japanese was by Nakamura who compiled and emotion expression dictionary made up of text written by famous Japanese writers categorized into 10 different kinds of emotions [11].

In this paper we build a Japanese emotion corpus. The corpus is made up of about 1,200 sentences taken from [3]. Then we manually tagged the words for emotion using 8 categories of emotion. From the tagged corpus we collected statistics to describe emotion in Japanese dialog and to help deal with emotion in natural language processing (NLP). Finally, we use the corpus to train various classifiers to identify emotion carrying words in text. This should help in future research through partially automating the annotation process by automatically finding emotion carrying words.

This paper will proceed as follows, in section 2 a background on affective computing and its joining with natural language processing is examined. In section 3, the annotation of the corpus is described. In section 4 , the collected statistics are examined in detail. Section 5 looks at automatic emotion word identification and the obtained results. Finally, in section 6, we discuss future work and make concluding remarks.

## 2 Background

Affective computing is an area of artificial intelligence that focuses on emotion understanding and creation. It hopes to improve such interactions, opening the door to many new possibilities. Text based emotion understanding relies heavily on natural language processing, which is mostly focused on understanding the semantics of text.

By analyzing the texts and obtaining not only semantic, but also emotional information the computer can deal with more interpersonal matters such as understanding the relationships between people, which creates a better user experience when dealing with computers. Both af-

*Department of Information Science and Intelligent Systems The University of Tokushima, JAPAN Email: {j_minato,davidb,ren,kuroiwa}@is.tokushima-u.ac.jp Fuji Ren is also with the School of Information Engineering, Beijing University of Posts and Telecommunications Beijing 100876, China

**(Advance online publication: 19 February 2008)**

Table 1: Corpus Statistics

| Description | Frequency/Percentage |
|---|---|
| Total # of Sentences | 1,191 |
| Total # of Words | 14,195 |
| Average Words per Sentence | 11.9 |
| Total # of Unique Words | 2,338 |
| Total # of Emotion Words | 1,160 |
| Total # of Unique Emotion Words | 653 |
| Total # of Emotion Idioms | 275 |
| Average Emotion Words/Idioms per Sentence | 1.2 |
| Average # of times an emotion word did not carry emotion | 0.22 |

fective computing and natural language processing are needed to reach this goal. Natural language processing algorithms are needed to understand the semantics or explicit message of text, while affective computing is needed to understand the implicit message in text manifested through emotion.

When the two are brought together the potential applications are endless. Saeki et.al, suggested the application to machine translation by focusing on emotional adverbs and their sentence structures in Japanese and English to find the translation tendency among both languages [13]. Ma et.al, created an internet chat system that assed the emotion in the textual messages and presented the emotion through avatars in order to realize natural and social communication between distant communication partners [6]. Holzman and Pottenger, also attempted to detect the emotional attributes in chat data for homeland security purposes [4]. Matsumoto et al. looked at improving human-robot interaction by giving the robot the ability to understand and mimic human emotion [8].

## 3 Corpus Annotation

Japanese-English bilingual sentence pairs from [3] were manually inputted into plain text format to create the initial untagged corpus. In this paper, we only focus on the Japanese portion of corpus. Then, using chasen [10] morphological analysis in the form of word segmentation and part-of-speech tagging was performed on the inputted sentences. After morphological analysis a manual correction was done to correct any errors by chasen.

Next, the emotion words and idioms were manually tagged by ourselves. Eight basic emotions were decided on; Joy, Hate, Love, Sorrow, Anxiety, Surprise, Anger, and Respect. Single word emotions were tagged with an 'S' meaning "single." Words in an idiom were tagged with 'B,' 'I,' and 'E' for beginning, intermediate, and end. This tagging scheme was taken from [15].

In addition to annotating emotion, emotion modifiers were annotated. An emotion modifier is a word or word phrase that changes the degree of or the type of emotion a word has within the context of the sentence. A simple example would be "not," which would negate the emotion within in the sentence. Four types of modifiers were chosen (plus, minus, neutral and negative) and they were tagged in the same manner as emotion words and idioms.

## 4 Statistical Analysis

After the annotation of the corpus was completed we performed statistical analysis on it to try and find any characteristics of emotional expression in Japanese. First, we gathered basic statistics on the corpus that included the number of words, sentences, etc. These results are shown in table 1.

The corpus was made up of 1,191 sentences, 14,195 words of which 2,338 were unique. There were 1,160 single word emotions and 275 emotion idioms. 653 unique words were used to describe emotion in the corpus. On average, in single word emotion words, a word that was marked as an emotion was used as a non-emotion only 0.22 times.

Figure 1, shows the distribution of emotion categories in the corpus. Interestingly, the four most frequent emotions are two sets of converse emotions; love and hate and joy and sorrow. Except for surprise and respect, each of the categories were well represented.

Next, the percentage breakdown for emotional part-of-speech was examined. The results are shown in figure 2. What can be seen is that nouns and verbs tend to carry emotion the most in Japanese. Combined they account for 75% of all the emotion words. Some parts-of-speech, like particles, show up in the graph due to their inclusion in emotion idioms.

Next, we combined the previous forms of information to look at how the parts-of-speech broke down for each of the emotion categories. The results are shown in figure 3. What can be seen, is that for the most part the parts-of-speech are evenly distributed among the different emotion categories. One notable exception is joy and sorrow's lack of auxiliary verbs.

The final statistic we looked at was if the position of the

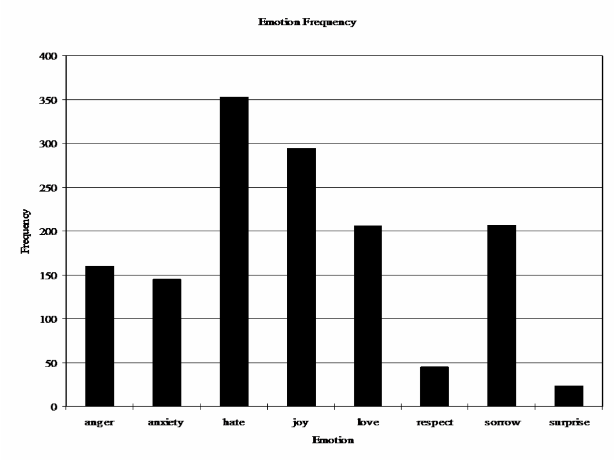


Figure 1: Emotion Frequency

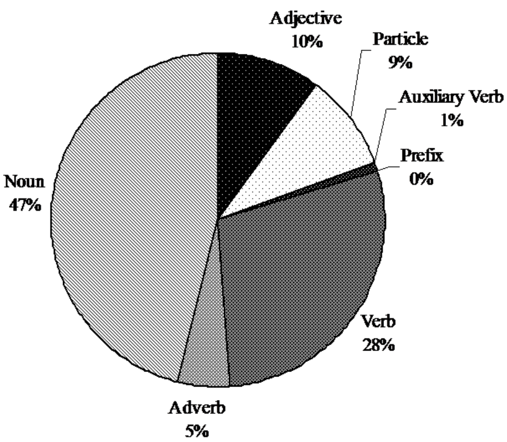


Figure 2: Part-of-Speech Frequency for Emotion Words

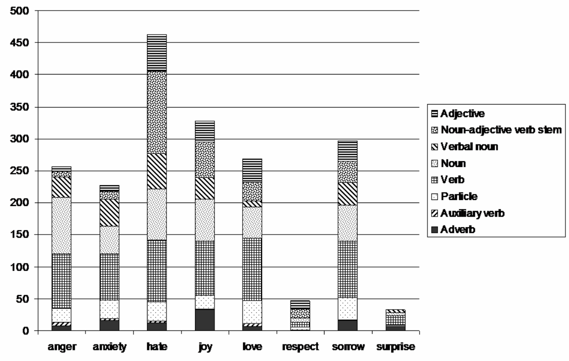


Figure 3: Part-of-Speech per Emotion Frequency

word in the sentence had played a role in emotional expression. The results are shown in figure 4. The sentence lengths were normalized so they could be combined together to determine the distribution. As we can see from the chart, emotion often takes place right after the middle of the sentence in Japanese.

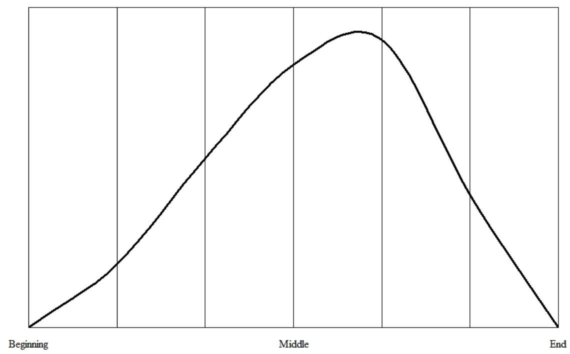


Figure 4: Distribution of Emotion in a Sentence

The statistics show that emotion in Japanese is predominately described using nouns and verbs. Emotion also tends to occur near the middle of the sentence. Also, no emotion category seems to overwhelmingly favor one part-of-speech over another. This information should prove useful when dealing with emotion in Japanese sentences.

## 5 Automatic Emotion Word Identification

One application that comes about through creating and analyzing a corpus, like the one in this paper, is the automatic identification of emotion words. Emotion word identification can be defined in many ways, but in this paper it is defined as determining if a word carries or does not carry emotion. The most straight forward approach is to construct an emotion dictionary and then identify words as emotions if they are present in the dictionary. This approach should yield a high precision, but it is impossible to identify words not in the dictionary.

Another approach is to treat the identification as a binary classification problem and use machine learning. The advantage of this approach is that it should be possible to classify words that have not been seen before. The disadvantage is that there will probably be a lose in precision and that non-emotion carrying words could be classified as carrying emotion. However, since even humans sometimes have a hard time determining emotion, this noise may be beneficial.

In this paper we examined four different classifiers; naive bayesian [7], decision tree [12], naive possibilistic [2] and maximum entropy [1]. A combination of seven different features, listed below, were used with the classifiers.

These features make up information about the current word being examined and the words immediately around it.

- Previous word
- Previous word's part-of-speech
- Current word
- Current word's part-of-speech
- Next word
- Next word's part-of-speech
- Position in sentence

For evaluation, the tagged corpus was split into 80% training and 20% testing data. The classifiers were then trained using various combinations of features. As a baseline, the training set was mined to create an emotion dictionary and the words in the test set that were in the dictionary were classified as carrying emotion and words not in the dictionary were classified as not carrying emotion. Table 2 shows its results.

Table 2: Baseline Results

| | Recall | Precision | F-Measure |
|---|---|---|---|
| Non-Emotion | 99.8% | 91.2% | 95.3% |
| Emotion | 24.9% | 95.3% | 39.5% |
| Averaged | 62.3% | 93.3% | 67.4% |

The first classification test performed using only the current word and current word's part-of-speech features. The results are shown in table 3. What can be seen from the table is that by using a classification based method, with even just the current word and its part-of-speech, recall can be increased if a decrease in precision is acceptable. The maximum entropy classifier had an average recall of about 13% higher than the baseline with a decrease in precision of a little less than 10%.

The next test added context information in the form of the previous and next word and their parts-of-speech. Table 4 shows the results. The added features helped all but the maximum entropy classifier. With these set of features the best classifier was the nave bayesian classifier with an average recall of 66.5% and average precision of 83.8%. This shows only a slight improvement over the baseline method, but resulted in a higher recall and f-measure.

Next, the position of the word in the sentence was added as a feature. The results are shown in table 5. There were no real significant changes when adding this feature.

As was to be expected, the machine learning based results were able to achieve a higher recall at the cost of

Table 3: Classification with Current Word and Current Word's Part-of-Speech

| Naive Bayesian | | | |
|---|---|---|---|
| | Recall | Precision | F-Measure |
| Emotion | 25.5% | 85.4% | 39.3% |
| Non-Emotion | 99.3% | 88.9% | 93.8% |
| Averaged | 62.4% | 87.2% | 66.5% |
| Decision Tree | | | |
| | Recall | Precision | F-Measure |
| Emotion | 26.2% | 83.1% | 39.8% |
| Non-Emotion | 99.1% | 88.9% | 93.7% |
| Averaged | 62.6% | 86.0% | 66.8% |
| Naive Possibilistic | | | |
| | Recall | Precision | F-Measure |
| Emotion | 25.5% | 85.4% | 39.3% |
| Non-Emotion | 99.3% | 88.9% | 93.8% |
| Averaged | 62.4% | 87.2% | 66.5% |
| Maximum Entropy | | | |
| | Recall | Precision | F-Measure |
| Emotion | 54.4% | 73.2% | 62.4% |
| Non-Emotion | 96.7% | 92.7% | 94.7% |
| Averaged | 75.5% | 83.0% | 78.5% |

Table 4: Classification with Word and Part-Of-Speech Information for Current, Previous and Next Words

| Naive Bayesian | | | |
|---|---|---|---|
| | Recall | Precision | F-Measure |
| Emotion | 34.7% | 77.7% | 48.0% |
| Non-Emotion | 98.3% | 90.0% | 94.0% |
| Averaged | 66.5% | 83.8% | 71.0% |
| Decision Tree | | | |
| | Recall | Precision | F-Measure |
| Emotion | 28.4% | 83.6% | 42.4% |
| Non-Emotion | 99.1% | 89.2% | 93.9% |
| Averaged | 63.8% | 86.4% | 68.2% |
| Naive Possibilistic | | | |
| | Recall | Precision | F-Measure |
| Emotion | 29.9% | 86.0% | 44.4% |
| Non-Emotion | 99.2% | 89.4% | 94.0% |
| Averaged | 64.5% | 87.7% | 69.2% |
| Maximum Entropy | | | |
| | Recall | Precision | F-Measure |
| Emotion | 20.6% | 52.1% | 29.5% |
| Non-Emotion | 96.8% | 88.0% | 92.2% |
| Averaged | 58.7% | 70.0% | 60.9% |

Table 5: Classification with All Features

| Naive Bayesian | | | |
|---|---|---|---|
| | Recall | Precision | F-Measure |
| Emotion | 34.7% | 77.7% | 48.0% |
| Non-Emotion | 98.3% | 90.0% | 94.0% |
| Averaged | 66.5% | 83.8% | 71.0% |
| **Decision Tree** | | | |
| | Recall | Precision | F-Measure |
| Emotion | 27.4% | 82.5% | 41.1% |
| Non-Emotion | 99.0% | 89.1% | 93.8% |
| Averaged | 63.2% | 85.8% | 67.5% |
| **Naive Possibilistic** | | | |
| | Recall | Precision | F-Measure |
| Emotion | 29.9% | 86.0% | 44.4% |
| Non-Emotion | 99.2% | 89.4% | 94.0% |
| Averaged | 64.5% | 87.7% | 69.2% |
| **Maximum Entropy** | | | |
| | Recall | Precision | F-Measure |
| Emotion | 21.1% | 56.1% | 30.7% |
| Non-Emotion | 97.2% | 88.1% | 92.4% |
| Averaged | 59.2% | 72.1% | 61.6% |

precision. In some cases the added recall may be so desirable that a little extra noise is acceptable. Interestingly, extra feature information, such as word position, was not as useful as we thought it might be. During the statistical analysis, whether a word carried emotion or not seem to be correlated to its position in the sentence, but this did not seem to help the classifiers.

## 6 Conclusion and Future Work

In this paper almost 1,200 Japanese sentences had emotion assigned to their words creating an emotion corpus. From the corpus a thorough statistical analysis was done. This analysis, we believe, can benefit NLP researchers in dealing with emotion in text. Then using the corpus various classifiers were trained to identify Japanese emotion words based on features inspired by the statistical analysis.

Through the statistical analysis, we found that Japanese tends to use nouns and verbs to describe emotion. Typically emotion is described in only one word and in a sentence emotion tends to occur near the middle. Finally, we found that for single emotion words, the word usually always carries emotion. These features suggested how to identify emotion words in sentences.

Using the statistical information, naive bayesian, naive possibilistic, decision tree and maximum entropy classifiers were trained using varying features. Their results were compared to using a simple emotion dictionary created from the training data. It was found that the classifiers were able to achieve a better recall for identification over the dictionary, but at the expense of precision.

In the future, we hope to expand the identification process and look at emotion classification for words an sentences. We also are currently tagging the English sentences and their words with emotion. After the tagging is completed statistical analysis will be done for the English corpus and comparisons can be made between Japanese and English.

## Acknowledgment

This research has been partially supported by the Ministry of Education, Culture, Sports, Science and Technology of Japan under Grant-in-Aid for Scientific Research (B), 14380166, 17300065, Exploratory Research, 17656128.

## References

[1] Adam Berger, Stephen Della Pietra, and Vincent Della Pietra. A maximum entropy approach to natural language processing. *Computational Linguistics*, 22, 1996.

[2] Christian Borgelt and Jrg Gebhardt. A naive bayes style possibilistic classifier. In *Proc. 7th European Congress on Intelligent Techniques and Soft Computing*, Aachen, Germany, 1999.

[3] I. Hiejima. *Japanese-English Emotion Dictionary.* Tokyodo Shuppan, 1995.

[4] L. E. Holzman and W. M. Pottenger. Classification of emotions in internet chat: An application of machine learning using speech phonemes. Technical report, Lehigh Tech Report 03-002, 2003.

[5] T. Koshino, M. Tokuhisa, J. Murakami, and S. Ikehara. Error analysis of emotion annotated dialogue corpus. In *Proceedings of the 18th Annual Conference of the Japanese Society for Artificial Intelligence*, 2004.

[6] C. Ma, A. Osherenko, H. Prendinger, and M. Ishizuka. A chat system based on emotion estimation from text and embodied conversational messengers. In *proceedings of the International Conference on Entertainment Computing*, pages 535–538, 2005.

[7] Christopher D. Manning and Hinrich Schutze. *Foundations of Statistical Natural Language Processing.* The MIT Press, Cambridge, Massachusetts, May 1999.

[8] K. Matsumoto, J. Minato, F. Ren, and S. Kuroiwa. Estimating human emotions using wording and sentence patterns. In *Proceedigns of the IEEE International Conference on Information Acquisition*, pages 421–426, 2005.

[9] Kazuyuki Matsumoto, David B. Bracewell, Fuji Ren, and Shingo Kuroiwa. Development of an emotion corpus creation system. Technical report, IPSJ SIG Technical Report, 2005.

[10] Yuji Matsumoto, Akira Kitauchi, Tatsuo Yamashita, Yoshitaka Hirano, Hiroshi Matsuda, Kazuma Takaoka, and Masayuki Asahara. Morphological analysis system chasen version 2.2.9 manual. Technical report, Nara Institute of Science and Technology, 2002.

[11] Akira Nakamura. *Emotion expression dictionary*. Tokyo-do Shuppan, 1993.

[12] J.R. Quinlan. Induction of decision trees. *Machine Learning*, 1:81–106, 1986.

[13] M. Saeki, M. Tokuhisa, J. Murakami, and S. Ikehara. Comparative analysis of japanese-english adverb focused on emotion expression (in japanese). In *Proceeding of the 11th Language Processing Annual Conference*, pages 33–36, 2005.

[14] Hiroya Takamura, Takashi Inui, and Manabu Okumura. Extracting semantic orientations of words using spin model. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, pages 133–140, 2005.

[15] Y. Tanaka, H. Takamura, and M. Okumura. Research concerning facemarks in text based communication (in japanese). In *Proceedings of the 10th Annual Natural Language Processing Conference*, 2004.

[16] R. Tokuhisa, K. Inui, M. Tokuhisa, and N. Okada. Two complementary case studies for emotion tagging in text corpora. In *Proceedings of the Japanese Society for Artificial Intelligence SIG-SLUD*, 2001.