HOKKAIDO UNIVERSITY

Title	The Limits of Words : Expanding a Word-Based Emotion Analysis System with Multiple Emotion Dictionaries and the Automatic Extraction of Emotive Expressions
Author(s)	Wang, Lu; Isomura, Sho; Ptaszynski, Michal et al.
Citation	Applied Sciences, 14(11), 4439 https://doi.org/10.3390/app14114439
Issue Date	2024-05-23
Doc URL	https://hdl.handle.net/2115/95896
Rights	© 2024 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/).
Rights(URL)	https://creativecommons.org/licenses/by/4.0/
Type	journal article
File Information	Appl. Sci.14-11_4439.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

applied sciences

MDPI

Article

The Limits of Words: Expanding a Word-Based Emotion Analysis System with Multiple Emotion Dictionaries and the Automatic Extraction of Emotive Expressions

Lu Wang [1,*], Sho Isomura [1,†], Michal Ptaszynski [1,*], Pawel Dybala [2], Yuki Urabe [3], Rafal Rzepka [4] and Fumito Masui [1]

[1] Text Information Processing Laboratory, Kitami Institute of Technology, Kitami 090-8507, Japan
[2] Institute of Middle and Far Eastern Studies, Faculty of International and Political Studies, Jagiellonian University, 30-387 Krakow, Poland
[3] Independent Researcher, Sapporo 060-0808, Japan
[4] Langauge Media Laboratory, Hokkaido University, Sapporo 060-0808, Japan
* Correspondence: luwang2753@gmail.com (L.W.); michal@mail.kitami-it.ac.jp (M.P.)
† Graduated in March 2022.

Abstract: Wide adoption of social media has caused an explosion of information stored online, with the majority of that information containing subjective, opinionated, and emotional content produced daily by users. The field of emotion analysis has helped effectively process such human emotional expressions expressed in daily social media posts. Unfortunately, one of the greatest limitations of popular word-based emotion analysis systems has been the limited emotion vocabulary. This paper presents an attempt to extensively expand one such word-based emotion analysis system by integrating multiple emotion dictionaries and implementing an automatic extraction mechanism for emotive expressions. We first leverage diverse emotive expression dictionaries to expand the emotion lexicon of the system. To do that, we solve numerous problems with the integration of various dictionaries collected using different standards. We demonstrate the performance improvement of the system with improved accuracy and granularity of emotion classification. Furthermore, our automatic extraction mechanism facilitates the identification of novel emotive expressions in an emotion dataset, thereby enriching the depth and breadth of emotion analysis capabilities. In particular, the automatic extraction method shows promising results for applicability in further expansion of the dictionary base in the future, thus advancing the field of emotion analysis and offering new avenues for research in sentiment analysis, affective computing, and human–computer interaction.

Keywords: affect analysis; emotive expressions; affect lexicon; emotion lexicon; emotive expression extraction; part of speech; natural language processing; emotion category; emotion analysis; emotion corpus

check for updates

Citation: Wang, L.; Isomura, S.; Ptaszynski, M.; Dybala, P.; Urabe, Y.; Rzepka, R.; Masui, F. The Limits of Words: Expanding a Word-Based Emotion Analysis System with Multiple Emotion Dictionaries and the Automatic Extraction of Emotive Expressions. *Appl. Sci.* **2024**, *14*, 4439. https://doi.org/10.3390/app14114439

Academic Editor: Douglas O'Shaughnessy

Received: 5 April 2024
Revised: 16 May 2024
Accepted: 20 May 2024
Published: 23 May 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

As social networking services (SNSs) have become popular in recent years, Internet users have been posting opinions, feelings, reviews, and criticisms about things, services, products, events, and disasters on the Internet. By accurately detecting, efficiently collecting, and further analyzing user emotions and other information contained in these Internet postings, it will be possible to automatically collect public opinions and timely information about these things, services, products, events, and disasters and to make efficient use of the information to improve those products and services and support decision-making.

For this reason, along with the spread of social media, research on the automatic extraction, collection, and analysis of emotional information from text data contained in these media has gained in popularity [1–5]. An example is ML-Ask (https://github.com/ptaszynski/mlask (accessed on 19 May 2024)) [6], an affect analysis system for textual input in Japanese developed by Ptaszynski et al.

The original version of ML-Ask was developed a few years ago, and its source code was all uploaded to GitHub since it was the first open-sourced affect system for Japanese [6]. The fact is that the affect lexicon contained in the database used in ML-Ask was constructed based on the Dictionary of Emotive Expressions [7], which was created in the 1990s, and thus many contemporary emotive expressions have not yet been included. Furthermore, the database has not been updated for many years, and it is still unable to respond to new emotive expressions that have started to be used since the spread of the Internet.

To solve this problem and improve the system performance, it is necessary to collect new emotive expressions, and to cope with the ever-changing Internet languages, it is necessary to have a method that enables the periodic automatic or semi-automatic updating of the emotive expression database. Therefore, we surveyed studies related to the expansion of the affect lexicon and emotion dictionaries. According to our investigation, although, indeed, the lexicon databases of Japanese emotion analysis systems developed by other researchers use different emotion dictionaries, unfortunately, these authors do not explain anything about how to unify lexicons that have different emotion categories with different definitions of the emotion vocabulary. We will discuss the current state of emotion dictionaries used in emotion analysis systems for Japanese and related research in the next section. Therefore, we decided to make the first attempt in this field to find a scientific and heuristic method to unify the Japanese emotion vocabulary in different emotion dictionaries which have different categories of emotions manually. Hopefully, with this method we presented, other researchers can easily go ahead and apply this method to keep manually expanding the vocabulary of ML-Ask or other databases of emotion lexicons if they have other available dictionaries of emotive expressions. Besides the manual expansion method, we also decided to propose an automatic extraction method of emotive expressions to achieve the semi-automatic expansion of emotion lexicons to reduce the manual burden of data processing while improving system performance.

The goal of this study is to expand the lexicon database of the ML-Ask affect analysis system by achieving the following key objectives:

1. Investigate how to update and extend the affect lexicon of ML-Ask in a way of manual expansion, proposing a scientific and heuristic method for manually adding Japanese emotive expressions through some available Japanese dictionaries of emotive expressions which have different classifications of emotion categories from that of the affect lexicon in ML-Ask.
2. Propose a method for automatically extracting new emotive expressions to update the dictionary periodically.
3. Improve the performance of ML-Ask after the expansion of the lexicon database.

The subsequent sections of this paper are organized as follows: In Section 2, we present a review of related works on emotion dictionaries and emotion analysis systems. Section 3 elaborates on the methods for the manual expansion of the lexicon by adding two different emotion dictionaries. Section 4 describes the method for the automatic expansion of the lexicon, especially the extraction of emotive expressions from the corpus. In Section 5, we present the results of experiments and performance evaluation of our approach to the integrated system. In Section 6, we discuss this study consisting of several aspects. Finally, Section 7 outlines the conclusion made and explores potential avenues for furthering this research in the future.

2. Literature Review

In this section, we present a review of the related works on types of emotion analysis systems and discuss the problem of the emotion analysis system of the Japanese language using the Japanese emotion dictionaries.

ML-Ask uses a lexicon-based method using a pre-built database, which contains a lexicon of words, phrases, and sentence patterns that express positive and negative emotions such as “happy”, “sad”, etc. When analyzing a new input sentence, this lexicon is used as a reference. If an input sentence contains one of these words, the sentiment

score is set to either negative or positive, and the sentence is annotated with the type of emotion expressed by the word (e.g., "happy" to "joy", and "afraid" to "fear"). Besides the lexicon-based method, there are two other natural language processing methods for affect analysis. One is rule-based methods that use a list of rules for emotive expressions extracted from multiple emotive sentences (such as n-grams, which are chains of words), and another is machine learning-based methods that use machine learning algorithms to learn rules automatically.

2.1. Lexicon-Based Emotion Analysis Systems

For example, Sharma et al. [8] proposed an emotion analysis system applying an emotion lexicon. This system can classify text into seven different categories, including anger, disgust, sadness, surprise, fear, joy, and neutral. Its emotion lexicon consists of 3000 words which are generated using movie reviews as the baseline.

Toçoğlu et al. [9] presented an emotion lexicon for emotion analysis in Turkish. This lexicon contains six emotion categories which are happiness, fear, anger, sadness, disgust, and surprise. They generated the lexicon based on a Turkish dataset. They compared the performance of the lexicon-based approach with the machine learning-based approach by using their proposed lexicon. According to their conclusion, using the proposed lexicon produces comparable results efficiently in the emotion analysis task of Turkish text.

Kamal et al. [10] proposed a framework to analyze Twitter data for crowdsource sensing and decision-making with a lexicon-based emotion analysis approach. This framework can obtain real-time Twitter data and classify the data into eight different emotion categories, including anger, anticipation, disgust, fear, joy, sadness, surprise, and trust.

2.2. Rule-Based Emotion Analysis Systems

Asghar et al. [11] proposed a rule-based emotion analysis system using five rule-based modules which are lexical resource generation, emotion word classifier, emoticon classifier, slang classifier, and mixed-mode classifier. The system uses different lexicons with a total of eight different emotion categories, which are fear, anger, happiness, disgust, surprise, sadness, embarrassment, and reactive.

Gao et al. [12] presented a rule-based emotion analysis system based on an emotional model to detect emotion causes for Chinese micro-blogs. The emotion lexicon used in this system was constructed with 22 types of emotions.

2.3. Machine Learning-Based Emotion Analysis Systems

Nasir et al. [13] proposed an emotion prediction and classification model applying a machine learning algorithm. The text dataset used to train the model was labeled into six emotion categories, which are anger, disgust, fear, sadness, joy, and guilt.

Xu et al. [14] proposed a machine learning-based emotion analysis model for emotion classification in Chinese microblog texts. This model only classifies texts into two categories, positive and negative. It has the highest accuracy in emotion classification compared with the other related seven models.

From a technical point of view, methods using lexicons and predefined rules, unlike machine learning-based methods, do not use complex computational algorithms and perform simple pattern matching; thus, they have the advantage of being able to analyze large numbers of documents in a short time. However, they also have some disadvantages. One is not being able to analyze new terms that are not in the lexicon because the analysis is performed using only the limited data in the lexicon. Another is not being able to properly process idiomatic phrases and emotive expressions consisting of multiple words distributed in a sentence, whose nuances are ambiguous depending on the wider context.

On the other hand, machine learning-based methods learn from a large amount of text data and calculate sentiment scores, and have the advantage of being more accurate than rule-based and lexicon-based methods if a large amount of data can be prepared. However, it is an expensive method, as it requires collecting such large datasets, performing data

cleaning, accurate expert label annotation, and post-processing. In addition, when using machine learning, it has been pointed out that the more classification classes there are, the less accurate the classifier (classification algorithm) will become [15]. Furthermore, people often express multiple emotions in natural speech, even in one sentence or a single utterance, or in writing, etc. Therefore, even a classifier that supports multiple classes can output only one final class, making it difficult to analyze crowded and complex emotional states.

From the above, it can be seen that lexicon-based and rule-based affect analysis methods are still highly applicable. However, the performance of such methods is highly dependent on the lists of words and rules contained in lexicons, and new words and phrases must be regularly added to the affect lexicon from which the rules are derived to ensure and improve the accuracy.

Since ML-Ask was originally developed using Nakamura's dictionary as the system's database, we used it as a benchmark to investigate other emotion analysis systems for the Japanese language that also use the contents of this dictionary as their database, and whether the system adds words from other dictionaries to this dictionary. Since the compilation of dictionaries is limited to the authors' dictums and biases, and each author's views and opinions are different, almost all dictionaries, especially emotion dictionaries, use a different categorization of the kinds of emotions. For example, the two available dictionaries of Japanese emotive expressions that we decided to add to ML-Ask categorize emotions into eight major categories [16], two major categories, plus many subcategories [17], respectively. Therefore, if other researchers added the vocabulary of other dictionaries to Nakamura's dictionary in their emotion analysis or classification system based on Nakamura's dictionaries, we further investigated what approaches other researchers have used to unify the vocabulary of various dictionaries with different emotion categories.

Kobayashi et al. [18] developed a high-speed emotion analysis system by creating an emotion expression dictionary containing 6660 headwords associated with binary values representing emotions in 10 categories referring to Nakamura's dictionary. This system enables the rapid analysis of 10,000 free-text sentences in approximately one second, catering to the growing need for automated emotion analysis in social networking services, online word-of-mouth, and reviews. In this study, they added vocabulary from two other Japanese dictionaries of emotive expression in addition to Nakamura's dictionary; however, they did not provide any detailed description of the maneuvers used and how the integration of the dictionary of emotive expression was carried out.

Takeuchi et at. [19] proposed an emotion estimation method for words to construct a dictionary of emotion words. However, this kind of task is to generate lexicons from zero, which is not the same as the unification of multiple lexicons that we are going to study.

According to our survey, very few studies are even remotely related to the integration of multiple dictionaries of emotive expressions, and instructions on how to integrate dictionaries are never mentioned. Therefore, we have to assume that these researchers have integrated the dictionaries rather arbitrarily and that we need to come up with a scientific and heuristic method to realize the integration of the dictionaries of emotive expressions.

Therefore, we decided to propose a methodology of the manual quantitative analysis of the differences between emotion categories in different dictionaries of emotive expressions and the integration of the lexicon of ML-Ask with the other two available Japanese dictionaries of emotive expressions and a methodology of manual quantitative analysis of the differences between the part-of-speech in different dictionaries and automatic extraction of emotive expressions.

The two dictionaries of emotive expressions are "A Short Dictionary of Feelings and Emotions in English and Japanese" made by Hiejima and "Love, Hate and Everything in Between: Expressing Emotions in Japanese" made by Murakami. In the following, we will refer to them as Hiejima's dictionary and Murakami's dictionary, respectively.

3. Materials and Methods for Manual Expansion

In this section, we describe the methods for the manual addition of Hiejima's dictionary and Murakami's dictionary used in this research, respectively.

3.1. Differences in Emotion Categories in Hiejima's Dictionary and Nakamura's Dictionary

To update and expand the database of emotive expressions in the ML-Ask affect analysis system, we decided to add and integrate a new dictionary of emotive expressions, namely, *A Short Dictionary of Feelings and Emotions in English and Japanese* created by Hiejima [16]. This dictionary has also been used as a dictionary of emotive expressions in other research [20].

In this dictionary, Japanese emotive expressions are classified into eight types: expressions of joy, expressions of love, expressions of anger, expressions of suffering, expressions of sadness, expressions of blame, expressions of enjoyment, and expressions of surprise.

On the other hand, the dictionary of emotive expressions used in the affect lexicon database of ML-Ask is Nakamura's dictionary of Emotive Expressions [7], which contains ten emotion categories: joy, fondness, relief, gloom, dislike, anger, fear, shame, excitement, and surprise. At first glance, there are similar types of emotion: joy and expressions of joy, fondness and expressions of love, anger and expressions of anger, gloom and expressions of sadness, and surprise and expressions of surprise.

In Nakamura's dictionary, however, the type of joy includes not only expressions of joy but also expressions of enjoyment. For a simple example, "enjoyable ("tanoshii" in Japanese)" and "enjoy ("tanoshimu" in Japanese)" can be found in both "joy" and "expressions of enjoyment". Therefore, to more precisely classify the emotive expressions, we had to verify how many existing emotive expressions there are in each emotion type in Hiejima's dictionary, and how those two dictionaries align together regarding the understanding of emotion categories.

3.2. Checking for Existing Emotive Expressions in Hiejima's Dictionary

The number and percentage of existing emotive expressions within each emotion type in Hiejima's dictionary are fully shown in Table 1. Emotive expressions that were not found in the original dictionary are classified as out-of-vocabulary (OOV) expressions. Additionally, to more clearly illustrate the emotion categories in Hiejima's dictionary, we visualized the data using bar charts for each emotion category separately. Figures 1–8 show the data for each emotion category, respectively.

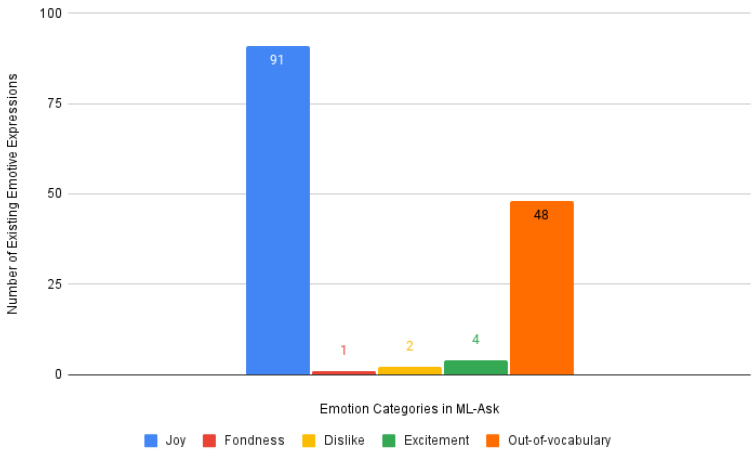

Figure 1. Existing emotive expressions in expressions of joy.

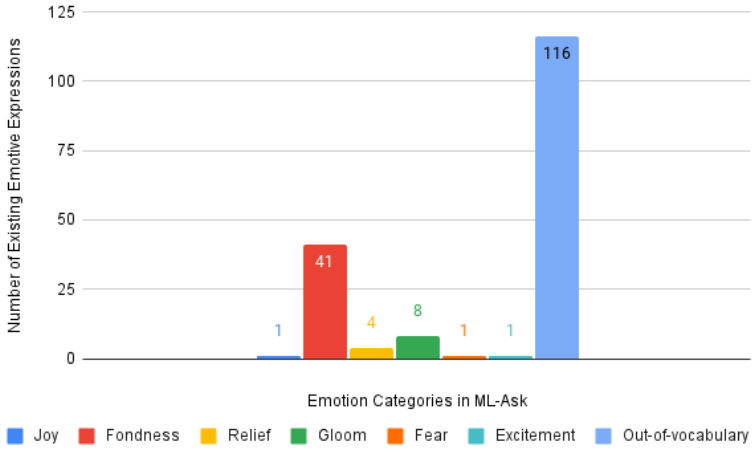

Figure 2. Existing emotive expressions in expressions of love.

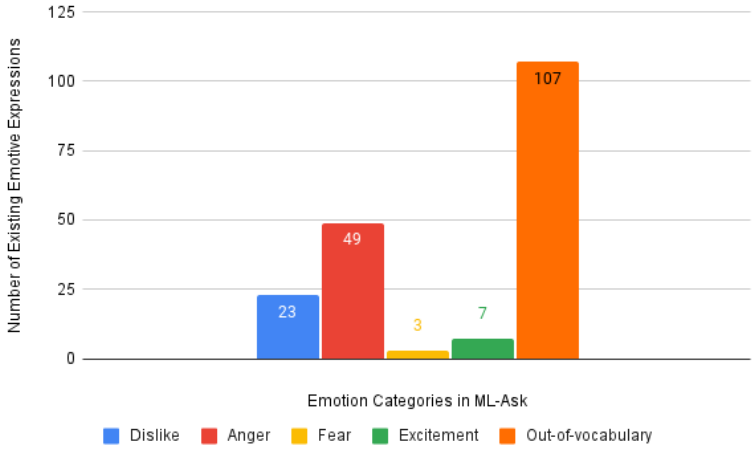

Figure 3. Existing emotive expressions in expressions of anger.

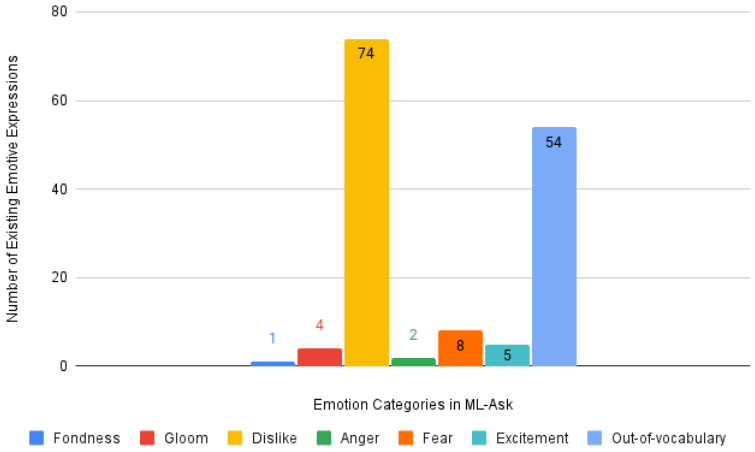

Figure 4. Existing emotive expressions in expressions of suffering.

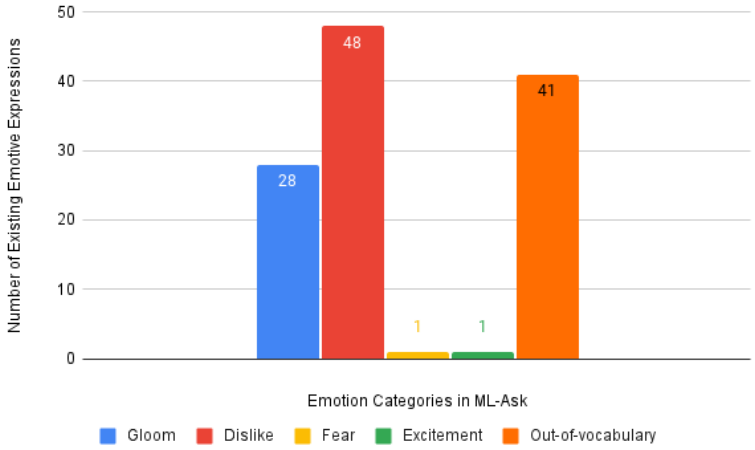

Figure 5. Existing emotive expressions in expressions of sadness.

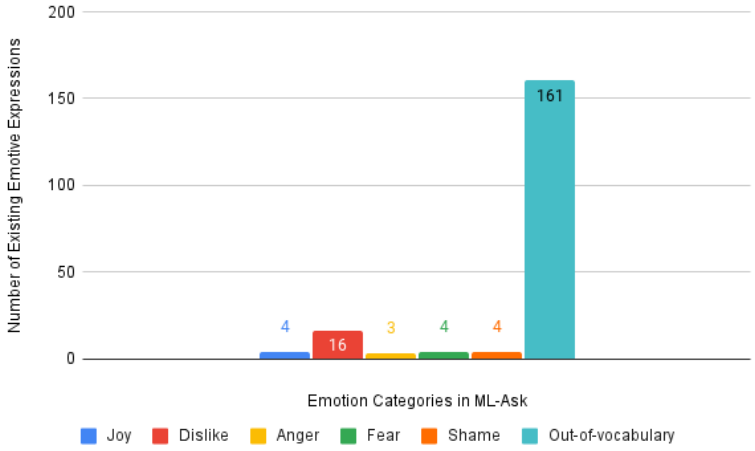

Figure 6. Existing emotive expressions in expressions of blame.

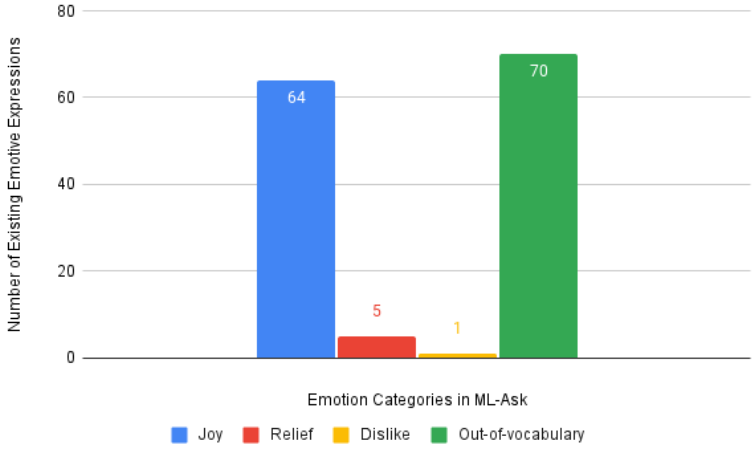

Figure 7. Existing emotive expressions in expressions of enjoyment.

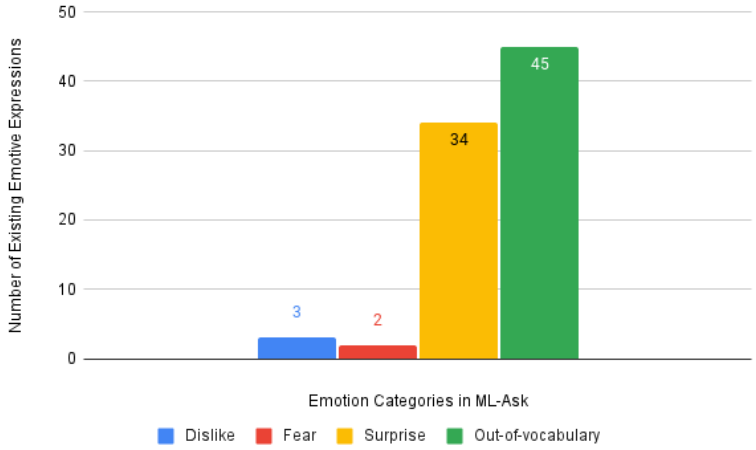

Figure 8. Existing emotive expressions in expressions of surprise.

As shown in Table 1 and the figures, the numbers and percentages of overlapping emotive expressions per category/type are the highest for expressions of joy, expressions of suffering, and expressions of sadness, and are higher than those for OOV expressions. Originally, we would have classified them into the type with the highest percentage, but after looking at these results, we decided that this was inappropriate because we found that almost all of the expressions were OOV in other emotion types. All OOV expressions had to be processed in some other way.

Table 1. Number and percentages of existing emotive expressions.

	Emotion Categories	**Nakamura's Dictionary**										
		Joy	**Fondness**	**Relief**	**Gloom**	**Dislike**	**Anger**	**Fear**	**Shame**	**Excitement**	**Surprise**	**OOV**
Hiejima's dictionary	Joy	91 (62.33%)	1 (0.68%)			2 (1.37%)				4 (2.74%)		48 (32.88%)
	Love	1 (0.58%)	41 (23.84%)	4 (2.33%)	8 (4.65%)			1 (0.58%)		1 (0.58%)		116 (67.44%)
	Anger					23 (12.17%)	49 (25.93%)	3 (1.59%)		7 (3.70%)		107 (56.61%)
	Suffering		1 (0.68%)		4 (2.70%)	74 (50.00%)	2 (1.35%)	8 (5.41%)		5 (3.38%)		54 (36.49%)
	Sadness				28 (23.53%)	48 (40.34%)		1 (0.84%)		1 (0.84%)		41 (34.45%)
	Blame	4 (2.08%)				16 (8.33%)	3 (1.56%)	4 (2.08%)	4 (2.08%)			161 (83.85%)
	Enjoyment	64 (45.71%)		5 (3.57%)		1 (0.71%)						70 (50.00%)
	Surprise					3 (3.57%)		2 (2.38%)			34 (40.48%)	45 (53.57%)

3.3. Processing OOV Expressions in Hiejima's Dictionary

Sakai et al. designed a questionnaire survey on the type of emotion expressed by each emoticon generated by their automatic emoticon generation algorithm [21]. In our research, to determine the most appropriate type of emotion expressed by the emotive expressions of uncertain emotion types from Hiejima's dictionary, we decided that it would also be appropriate to prepare a similar questionnaire survey, containing not emoticons but rather example sentences containing those OOV emotive expressions and conduct the questionnaire among native speakers of Japanese.

First, for each OOV word, 10 example sentences containing it were prepared. In the questionnaire, respondents participating in the survey were asked to choose the type of emotion expressed in the example sentences.

However, as can be seen in Table 1, there were 642 OOV expressions overall. If all of these OOV expressions were to be processed using the above method, 6420 example sentences would need to have been first prepared. This would not only be burdensome in terms of preparing the dataset of example sentences but would also require a considerable amount of time to complete the questionnaire survey by the participants, resulting in fatigue and errors. Since the less time it takes to process the OOV expressions, the better, we needed to reduce the number of example sentences to 5 (since the maximum number of example sentences for a word in Hiejima's dictionary is 5) and reduce the number of OOV expressions in some other way to make the study realizable in the given time constraints.

3.3.1. Reducing Number of OOV Expressions in Hiejima's Dictionary

We again studied Hiejima's dictionary and found that most of the Japanese emotive expressions in Hiejima's dictionary are grouped into smaller groups of 4 synonymous emotive expressions. Specifically, in this dictionary, the main expression is highlighted in bold. It is followed by a single parenthesis, inside of which there are 3 (very rarely 2 or 4) emotive expressions that have a similar meaning to the main emotive expression. In other words, there are typically 4 synonyms in each group.

Knowing that there are synonyms, we first had to process the synonyms and reconfirm the existing emotive expressions before reducing the number of OOV expressions. Also, some expressions were duplicated, so they had to be dealt with as well.

The number and percentages of existing emotive expressions in each emotion category in Hiejima's dictionary after processing the synonyms and duplicate expressions are shown in Table 2. Also, to clearly illustrate the number of existing emotive expressions in each emotion category after the process of synonyms and duplicates, we again visualized the data using bar charts for each emotion category separately as Figures 9–16.

Table 2. Number and percentages of existing emotive expressions after processing synonyms and duplicate expressions.

	Emotion	**Nakamura's Dictionary**										
	Categories	**Joy**	**Fondness**	**Relief**	**Gloom**	**Dislike**	**Anger**	**Fear**	**Shame**	**Excitement**	**Surprise**	**OOV**
Hiejima's dictionary	Joy	115 (82.73%)	4 (2.88%)			4 (2.88%)				4 (2.88%)		12 (8.63%)
	Love	4 (2.45%)	59 (36.20%)	11 (6.75%)	8 (4.91%)			4 (2.45%)		1 (0.61%)		76 (46.63%)
	Anger					39 (22.16%)	61 (34.66%)	3 (1.70%)		13 (7.39%)		60 (34.09%)
	Suffering		1 (0.68%)		4 (2.72%)	101 (68.71%)	2 (1.36%)	9 (6.12%)		10 (6.80%)		20 (13.61%)
	Sadness				36 (31.03%)	59 (50.86%)		1 (0.86%)		4 (3.45%)		16 (13.79%)
	Blame	4 (2.26%)				27 (15.25%)	8 (4.52%)	4 (2.26%)	4 (2.26%)			130 (73.45%)
	Enjoyment	83 (64.34%)		6 (4.65%)		4 (3.10%)						36 (27.91%)
	Surprise					4 (5.00%)		4 (5.00%)			40 (50.00%)	32 (40.00%)

As shown in Table 2, there were a total of 382 OOV expressions. The number of OOV expressions was reduced by a little less than half compared to the previous number.

Furthermore, since there were synonyms among the OOV expressions, if the emotion type of only the main emotive expression was determined for a group of emotive expressions, the emotion types of the other synonyms could also be assumed to belong

to the same emotion type as the main emotive expression. In this way, after eliminating the synonyms and duplicates, the number of OOV expressions that had to be processed decreased to 93. Thus, the number of OOV expressions was reduced to almost a quarter, which considerably shortened the process of preparing the dataset of example sentences and the questionnaire survey.

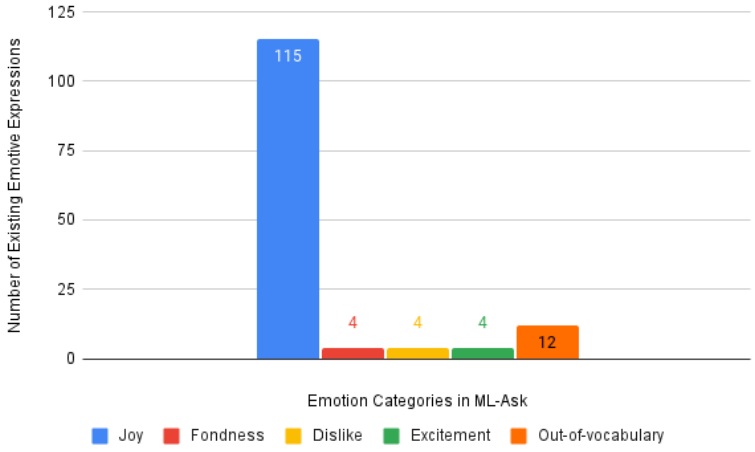

Figure 9. Existing emotive expressions in expressions of joy (new).

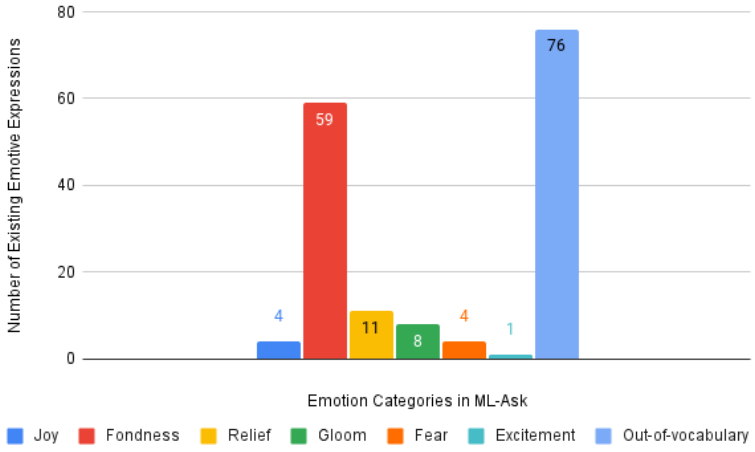

Figure 10. Existing emotive expressions in expressions of love (new).

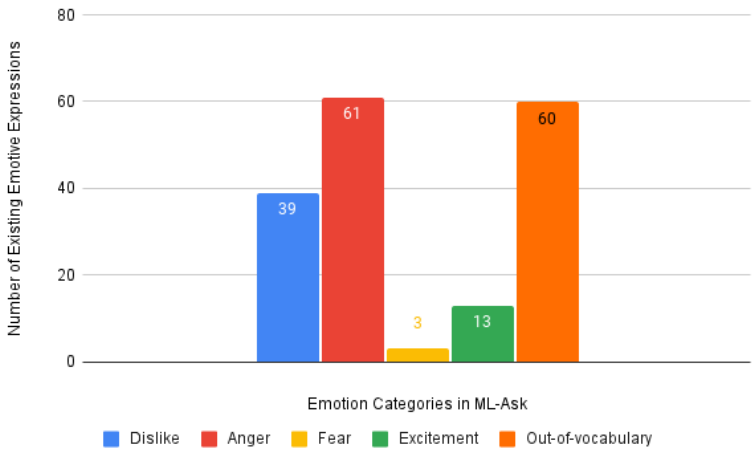

Figure 11. Existing emotive expressions in expressions of anger (new).

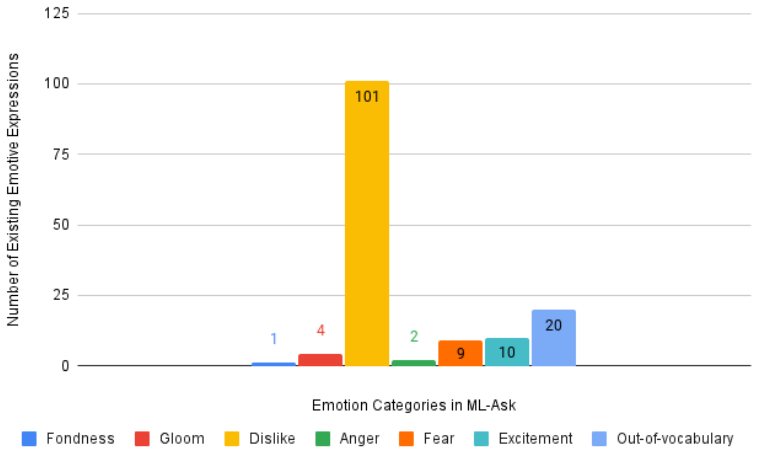

Figure 12. Existing emotive expressions in expressions of suffering (new).

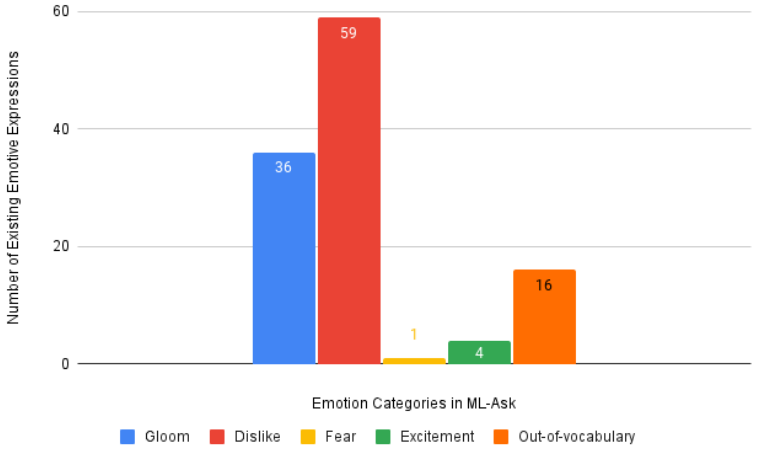

Figure 13. Existing emotive expressions in expressions of sadness (new).

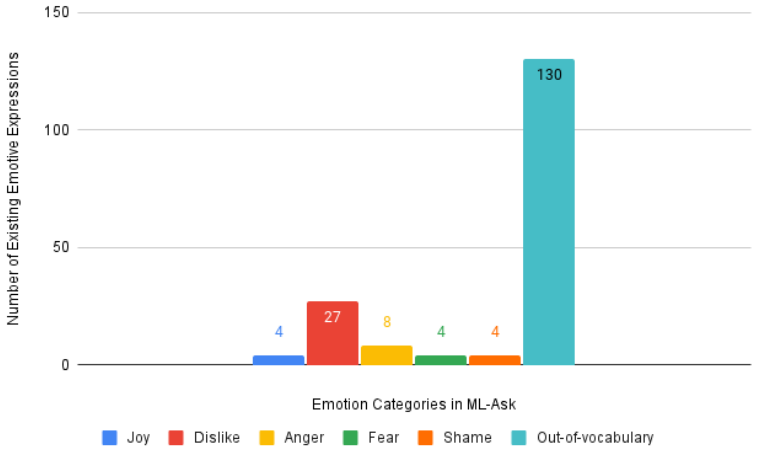

Figure 14. Existing Emotive Expressions in Expressions of Blame (new).

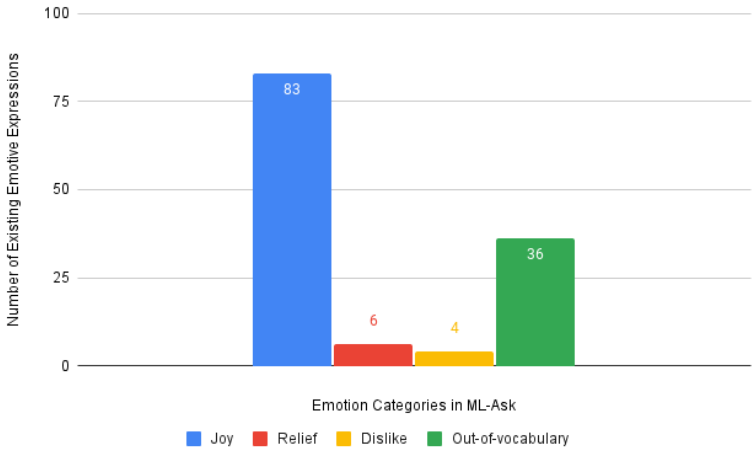

Figure 15. Existing emotive expressions in expressions of enjoyment (new).

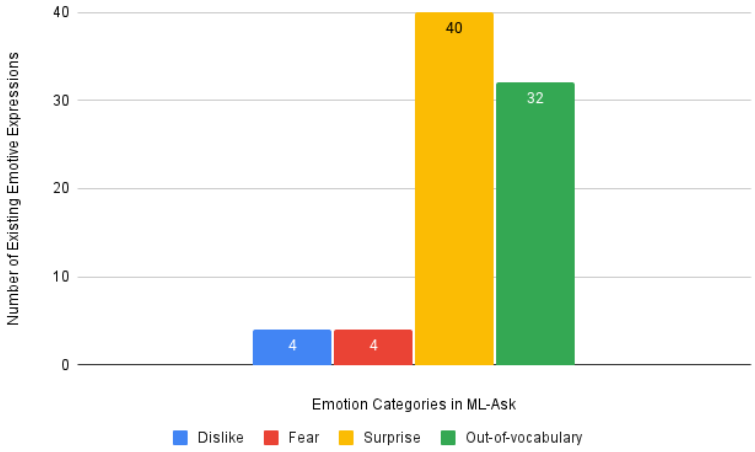

Figure 16. Existing emotive expressions in expressions of surprise (new).

3.3.2. Questionnaire Survey 1

To process the OOV emotive expressions in Hiejima's dictionary in the questionnaire survey, we first had to determine the source text data for the example sentences. Obviously, it is best to use raw Japanese sentences to process Japanese expressions. In this research, we selected a Japanese corpus for the extraction of the Japanese texts used for the questionnaire survey. The Japanese corpus will be further explained in the next subsection.

- YACIS Large-scale Japanese Blog Corpus
  The YACIS blog corpus is currently the largest Japanese blog corpus with 5.6 billion words [22]. The YACIS corpus, made in 2010, was collected from the Ameba blogs, and when it was created, it included roughly one-third of the Ameba blog. Furthermore, the YACIS corpus has been applied to several studies about affect analysis [23–25]. Therefore, we considered it to be suitable for application to this study.

3.3.3. Preparation of Example Sentence Dataset for the Questionnaire

In preparation for the questionnaire, we needed to collect proper examples. Firstly, the example sentences were collected from Hiejima's dictionary, which already contains some examples for each emotive expression. If this was not enough, we used example sentences from the YACIS corpus, and if this was still not enough, we searched for sentences on Twitter containing the specific expression.

However, if the sentences extracted from the YACIS corpus or Twitter are too long or too short, they could cause additional bias in the responses. To extract example sentences

from the YACIS corpus, we first needed to determine how long the sentences should be. We aimed at a length similar to the examples from the dictionary.

To specify the optimal sentence length, the example sentences from Hiejima's dictionary were segmented into words using Mecab [26], and the number of words in each sentence was calculated and used as the length of the sentence. Then, the mean and standard deviation of the example sentence lengths in the dictionary were calculated to be 10.96 and 2.80, respectively. The mean plus or minus standard deviation was used to determine the final optimal range of sentence lengths to be extracted from the YACIS corpus. Based on the calculation, the range of sentence lengths should be from 8 to 13 words. Finally, we extracted the example sentences from the YACIS corpus of the specific lengths using regular expressions.

3.3.4. Questionnaire Setup

The setup of the questionnaire was designed as follows. Firstly, 465 example sentences were prepared for 93 OOV emotion words (five sentences per word). All the example sentences were distributed into four sets on average, and four sets of questionnaires were created. Respondents were asked to respond to each sentence in terms of the type of emotion expressed in the sentence.

The questionnaire was administered in a multiple-choice format. However, if all the emotion types used in the database of ML-Ask were used as options, the respondents would be overloaded, and this would affect the quality of their responses, so only four options were given. Specifically, the questionnaire responses were given four options: (1) the original emotion type (treated as the potential correct answer), (2) a similar emotion type, (3) the opposite emotion type, and (4) other.

The similar and opposite emotion types were selected by referring to ML-Ask emotion types mapped onto the two-dimensional emotion model [27] as in [6]. Since the labels of the emotion types in the original database include Japanese characters that are not usually used (e.g., "iya", or "takaburi" in Japanese), instead of using just those labels, we referred to the database of ML-Ask and used easy-to-understand expressions as the options. Thus, for example, if the correct answer was "happy", the options were (1) happy, (2) like, (3) dislike, and (4) other.

Moreover, after an initial trial run of the questionnaire, we found out that the respondents had trouble recognizing the correct emotion type based only on one-word descriptions of the labels. Therefore, we added several descriptions per each label to make the meaning of the label more understandable. For example, for the option "happy", we eventually included words like "happy", "joyful", "cheerful", etc.

In addition, we randomized the options from (1) to (3) for each example sentence so that the correct option would not be easily guessed by the respondents after answering a few questions. In case no emotion type matched the respondent's idea, or in case the respondent did not know the emotion type, the option "other" was given, and the respondent was asked to describe the emotion type that matched his/her idea. Thus, four sets of questionnaires were created and surveyed.

3.3.5. Initial Results of Questionnaire

The following is a description of the responses submitted by the respondents participating in the survey. Before responding to questions regarding the emotion types, we first asked them to indicate their nationality, gender, and age. To investigate the emotion types of Japanese texts, we will focus on responses from Japanese respondents only. If a respondent responded in a jokey manner or ridiculed the questionnaire, we will also delete the response. In the future, we aim to analyze the responses of non-Japanese respondents as well to see the differences in how Japanese language learners of various nationalities perceive Japanese emotive expressions.

As a preliminary survey, we aimed for a minimum of 10 respondents in each set, but the number of respondents dropped to 8 in the first set, and then to 3 in the fourth set.

The percentage of male respondents was higher than that of female respondents in each set; thus, we would like to adjust the percentage of female respondents to be more equal in the future to increase the statistical reliability and investigate the differences in the perception of emotive expression between male and female respondents. The statistical information of the respondents for each set is shown in Table 3.

Table 3. Statistics on respondents per set in the 1st survey.

Set	Number of Respondents	Age (Generations/%)					Percentage by Gender	
		10s	20s	30s	40s	50s	Male	Female
1	8	12.5%	75%	0%	0%	12.5%	87.5%	12.5%
2	6	16.7%	66.7%	0%	0%	16.7%	83.3%	16.7%
3	4	0%	100%	0%	0%	0%	100%	0%
4	3	0%	100%	0%	0%	0%	100%	0%

Furthermore, in the initial questionnaire run, the respondents commented that there were few options for the emotion labels and that the objective descriptions made it difficult to determine the type of emotion. Therefore, we had to improve the emotion labels, provide additional explanations about the questionnaire, change a few difficult-to-understand example sentences, and conduct the questionnaire once more.

3.3.6. Improvement of the Questionnaire

To improve the labels of the emotion types, three easy-to-understand expressions were selected by referring to the emotive expression database of ML-Ask, and the emotion type of ML-Ask to which they belonged was added in front of them. Thus, for example, if the potential correct emotion type was joy, the choices were (1) [joy] (enjoyment, happiness, and fun), (2) [like] (love, dear, and addicted to), (3) [dislike] (aversion, disgust, and unpleasant), and (4) others.

The example sentences used in the questionnaire were extracted from dictionaries and blogs (YACIS corpus); therefore, many sentences directly express the speaker's emotions. However, other sentences are descriptions of a person's behavior or short sentences with a narrow range of context. Since it was difficult to directly judge the emotion type of these sentences, we asked the participants to simply answer which emotion they felt was being referred to when they read a sentence, rather than considering the standpoint of the speaker expressing it. Then, several difficult-to-understand example sentences were changed.

In this way, four sets of questionnaires were improved and re-surveyed.

3.3.7. Final Results and Summary of Both Surveys

All four sets of questionnaires in the 2nd survey were answered by university students in their 20s. The information for each set of respondents is shown in Table 4.

Table 4. Statistics of respondents per set in the 2nd survey.

Set	Number of Respondents	Age (Generations/%)					Percentage by Gender	
		10s	20s	30s	40s	50s	Male	Female
1	7	0%	100%	0%	0%	0%	100%	0%
2	10	0%	100%	0%	0%	0%	90%	10%
3	12	0%	100%	0%	0%	0%	83.3%	16.7%
4	15	0%	100%	0%	0%	0%	100%	0%

We had to integrate the results obtained from the new survey with those from the previous survey. Specifically, first, the emotion labels from the previous survey were changed to the current labels. Then, all of the "other" responses were initially processed by the ML-Ask affect analysis system to fish out those emotion labels that could be unified automatically. Responses that could not be processed by the system were considered outliers. In this way, the percentage of each emotion type for each emotive expression could be calculated.

The survey results show that most of the emotive expressions tend to have a high percentage of association with two emotion types. However, there remained expressions with more than one or two responses, for which the emotion label was ambiguous. To disambiguate those emotion labels, we applied a method similar to the one proposed by Ptaszynski et al. [28] for the automatic estimation of ambiguity in the meaning of emoticons, with the assumption that it may be possible to estimate ambiguity in the meaning of whole texts as well.

However, since there was no emotive expression with the highest percentage of the label "other" in the current survey, we determined that there was no need to estimate the ambiguity of the emotive expressions in this case.

To determine the emotion type to which an emotive expression belongs, we had to determine a threshold for the percentage of emotion types. This time, to cover all emotive expressions as much as possible, each emotive expression was classified into the emotion type with the highest percentage if the highest percentage of the emotion type was 67.5% or higher. If the highest percentage of the emotion type is less than 67.5%, the emotive expression is classified into the emotion type with the first and the second highest percentages. Thus, we were able to classify all of the emotive expressions except one emotive expression whose sum of the first and second highest percentages was less than 67.5%.

In addition, the majority of the survey respondents are males in their 20s. Thus, we would like to widen the diversity of respondents to make the results more relevant and investigate the differences in the perception of emotive expression between male and female respondents.

3.4. Differences in Emotion Categories in Murakami's Dictionary and Nakamura's Dictionary

The second available dictionary of emotive expressions we decided to add and integrate with the ML-Ask affect lexicon database was made by Mamiko Murakami, namely, *Love, Hate and Everything in Between: Expressing Emotions in Japanese* [17].

In this dictionary, expressions of affection are divided into two main categories, which are "From uncertainty to love" and "From uncertainty to hate". In other words, the author of this dictionary has collected the expressions and words based on the two categories of love and hate. Within each category, there are many subcategories.

First of all, there are 14 subcategories in the category "from uncertainty to love", which are Ambivalence, Flattery, Sympathy, Friends and neighbors, Higher love, Thankfulness, Beloved things, Parents and children, Tough love, Love and liking, Coming on strong, Walking on air, Loving to excess, and Off the beaten path. On the other hand, there are 24 subcategories in the category "From uncertainty to hate", which are Ambivalence, Frosty silence, Derision, Arrogance and pride, Knowing no shame, Ingratitude, The cold shoulder, Scowling and frowning, Sarcasm and bullying, Finding fault, Getting mad, Fed up, Cannot stand, Unlikable things, Jealousy, Fighting, defiance, Hate, Holding a grudge, Revenge, Betrayal, Horrible things, and "We've been through a lot together. Now let's break up!".

Since there are no specific descriptions for these subcategories, it is difficult for the reader to understand their meaning at once. Therefore, we considered that it would be inappropriate to put them directly into the categories of love and hate in the database of ML-Ask and that it would be necessary to process the existing emotive expressions in the same way as in the first dictionary.

3.5. Checking for Existing Emotive Expressions in Murakami's Dictionary

Firstly, to check for the existing expressions, we processed all the expressions in Murakami's dictionary by the ML-Ask system. Since the results of the ML-Ask processing are not always completely correct, the output needed to be manually reconfirmed. So, after reconfirmation, we obtained the result of the number of existing emotive expressions.

This time, we also visualized the data of existing emotive expressions using bar charts for the two big emotion categories, which are "From uncertainty to love" and "From uncertainty to hate", separately. Figures 17 and 18 show the data for the two big emotion categories, respectively.

The number and percentage of existing emotive expressions in Murakami's dictionary are as follows. Emotive expressions that were not found in the ML-Ask database are still classified as out-of-vocabulary (OOV) expressions. There are 153 expressions from ML-Ask that exist in this dictionary and 212 out-of-vocabulary expressions. Meanwhile, the percentages of existing expressions and OOV expressions are 41.92% and 58.08%, respectively. In other words, about half of the emotive expressions are out-of-vocabulary, so they need to be processed separately.

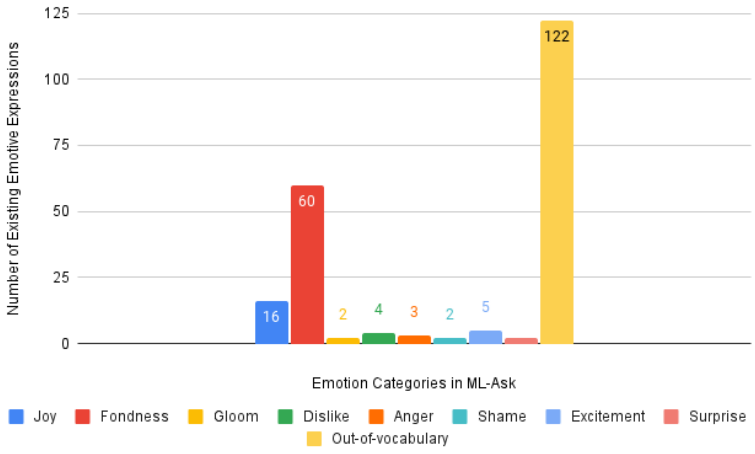

Figure 17. Existing emotive expressions in "From uncertainty to love".

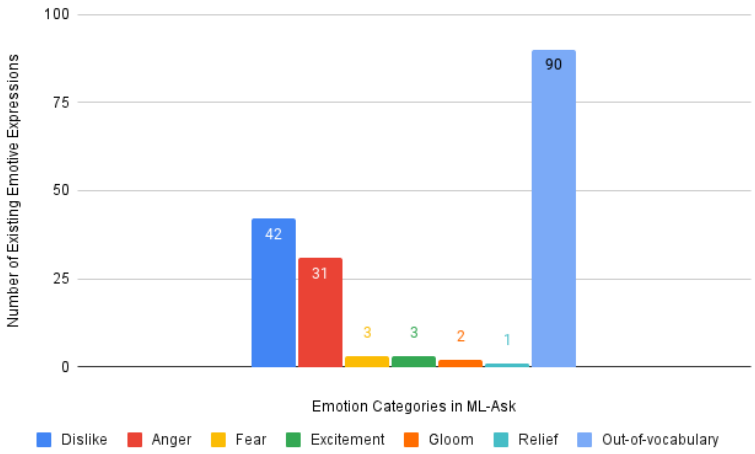

Figure 18. Existing emotive expressions in "From uncertainty to hate".

3.6. Processing OOV Expressions in Murakami's Dictionary

3.6.1. Basic Concept

When dealing with this dictionary, we considered that although the categories of emotions in OOV expressions can be effectively dealt with through questionnaires by

native speakers, the more words are dealt with using questionnaires, not only would we need to collect more examples, which would make the time spent on preparing the questionnaires and waiting for the results longer, but the more it will increase the burden on the respondents, and thus affect the quality of the responses.

For example, this time, if all the remaining OOV words were to be processed through the questionnaire like last time, we would need to prepare about 1000 example sentences, which would take a lot of time. However, our research needed to be sped up without a lot of waiting time. So this time, we wanted to first try other ways to minimize the number of emotive expressions that need to be solved by a native speaker's questionnaire survey, that is, to reduce the number of the remaining OOV emotive expressions by processing them in other ways.

From the integration of Hiejima's dictionary, it is known that if there is a synonym for each word, then it is sufficient to deal with one of the words of a group of words rather than with all the words in the group. So this time, we decided to process the OOV words by their synonyms. Then, we selected three methods to get synonyms of OOV words. These three methods are the pre-trained Word2vec model, the pre-trained fastText model, and the Weblio thesaurus, which will be further described in the next section.

In addition, to improve the query matching of OOV words with existing words, to increase the number of words that can be processed as OOV words, and to reduce the number of words that ultimately require a questionnaire to decide on their emotion categories, we decided to use the same three methods on all words in the original ML-Ask affect lexicon database to obtain the corresponding synonyms as well.

3.6.2. Word2vec and fastText Models

Word2vec is a technique in natural language processing (NLP) for obtaining the vector representations of words, created by a team of researchers at Google [29,30]. The word2vec algorithm estimates information about word meanings and their usage in context in the form of vectors by modeling text from large corpora. The word2vec model, if trained, can detect synonyms or suggest more words for part of a sentence.

Moreover, to obtain appropriate and reliable synonyms as much as possible, the words inside the word vectors should have a relatively high frequency of occurrence and relevance in the corpus. Therefore, the number of words inside the vectors to be selected will be correspondingly small. So this time, we selected chiVe (Sudachi Vector) [31], which is based on the Skip-gram algorithm and uses Word2vec (gensim Python library) to construct a Japanese word distribution representation. And the selected version was 1.2 mc90. Additionally, the National Institute for Japanese Language and Linguistics' Japanese Web Corpus (NWJC), which contains about 100 million web page sentences, was used for training, and Sudachi, a morphological analyzer from Works Applications, was used for segmentation.

Another method used in obtaining synonyms is fastText. It is a library for the efficient learning of word representations and sentence classification created by Facebook's AI Research lab [32–34]. Facebook also has already made available pre-trained models for 294 languages. So this time, we selected the Japanese pre-trained model.

3.6.3. Weblio Thesaurus

Considering that we might not be able to obtain enough proper synonyms just by using word2vec and fastText models, we also decided to use the Weblio Thesaurus web service (https://thesaurus.weblio.jp/ (accessed on 19 May 2024)) to collect as many synonyms as possible.

Weblio Thesaurus is a web service, where everyone can use the synonym search function for keywords in Japanese online through its website. Weblio Thesaurus has the following features. First, it contains approximately 400,000 Japanese expressions of various synonyms. The thesaurus is categorized and organized by usage situation and nuance. It introduces thesaurus words regardless of their parts of speech, such as nouns,

adjectives, and exclamation marks. It also includes many common expressions and sayings. This time, we used it as a complement to the insufficient synonyms from the word2vec and fastText models.

3.6.4. Synonym Query Matching

Before the synonym query matching, we used word2vec and fastText models to generate the top five words with the highest score most related to the out-of-vocabulary expressions and the same with expressions in the ML-Ask database. As a result, the words output from the word2vec model are more appropriate as synonyms, while those output from the fastText model are more relevant to the main expressions. So, we mainly used the synonymous expressions from the word2vec model and Weblio Thesaurus for synonym query matching.

After the synonym query matching, we obtained the result that 159 out-of-vocabulary expressions were processed successfully, with the remaining 46 expressions to be processed later. We confirmed them manually and found that 3 expressions were ambiguous because they are usually used in the situation of a third-party declarative sentence. Also, there is an expression "まあまあ maa-maa" in Japanese which is hard to consider as an emotive expression, so we decided to remove these 4 expressions. Eventually, 42 out-of-vocabulary expressions remained that needed to be processed by questionnaire survey.

3.7. Questionnaire Survey 2

We conducted a questionnaire survey that had similar basic concepts as when dealing with the out-of-vocabulary expressions in Hiejima's dictionary on Japanese university students in their 20s. This time, as the maximum number of example sentences for an expression in Murakami's dictionary is 3, we collected 3 example sentences per out-of-vocabulary expression.

Regarding the results of this questionnaire survey, since this time, unlike the previous one, there were no diffuse and ambiguous results of expressions that needed special treatment, we decided to categorize the survey results according to the emotion category with the highest percentage of votes and the second highest. If the percentage of the emotion category with the highest number of votes was greater than its average, we classified the emotion category of the OOV expressions in that category. If the highest percentage of the emotion category was less than the average, we classified it into the emotion category with the highest and second highest percentages. Thus, we successfully classified all the OOV expressions.

4. Materials and Methods for Automatic Expansion

In this section, we described the method for the automatic expansion of the affect lexicon, especially the extraction of emotive expressions from the corpus.

4.1. On Preparing Datasets for Emotive Expression Extraction

To automatically update and extend the database of emotive expressions in the affect analysis system ML-Ask, we first had to determine the textual data from which to extract them. In this case, we used the YACIS blog corpus, which is textually diverse and already annotated with emotional information [22,35]. The YACIS corpus has been described in the previous section. Moreover, the YACIS corpus is multidimensionally annotated and has been applied and referenced in numerous studies [25,36–39]. Therefore, it was judged still suitable for application to this part of the study.

4.1.1. Emotional Information in the YACIS Corpus

YACIS is already annotated with emotional information, making it highly applicable to new emotive expressions. Specifically, the ratio of emotive to non-emotive sentences in YACIS is 2:1 [22], which means that YACIS is expected to contain a wealth of emotion-filled daily expressions. In addition, since the data are collected based on a language resource

(Ameblo) that does not have the character limit found in microblog-style SNSs such as Twitter, there is no tendency for messages to be unnatural, as users are forced to conform to the character limit, and since users can express themselves as they wish, it is thought to contain more natural daily expressions than microblogs. This is considered to be more natural than microblogging.

The specific linguistic resource used in this study is not the entire YACIS but a part of YACIS used to construct an ontology of emotion objects created based on the annotations of emotional information in YACIS by Ptaszynski et al. [40].

This linguistic resource contains emotional information for each sentence from among ten emotion categories according to Nakamura's classification of emotions appropriate for Japanese and was used as text data to investigate emotive expressions not yet published in ML-Ask and to extract new emotive expressions.

4.1.2. Preparation of Dataset for Emotive Expression Extraction

The specific dataset used in the emotive expression extraction study was created as follows. First, from the above dataset with the automatic annotation of emotional information, we randomly selected 200 sentences for each emotion category, for a total of 2460 sentences, and all sentences were further manually annotated by two human annotators.

In the annotation process, 123 sets of 20 sentences were created for every sentence, and 105 annotators were asked to respond in the form of an Internet questionnaire. Of these, most respondents responded once, and 13 responded multiple times, each time to a different set. The average age of the respondents was 26.6 years old, 54.3% male and 45.7% female, and most of them were university students. The average age of the respondents was 26.6 years old. Of these, 71.4% were students in the sciences and 25.7% in the humanities.

Furthermore, in 546 (21.52% of the total) of the sentences, two respondents chose an auxiliary "free description" for the same sentence rather than a sentiment category, of which 466 sentences were matched with answers such as "nothing in particular" or "unknown".

Next, the percentage of consistency among respondents was determined. The consistency rate was calculated by computing Krippendorff's alpha and Cohen's Kappa. Then, the consistency rate between the two subjects who responded to one set of questionnaires was calculated and averaged for the entire set (123 sets).

The results showed that Krippendorff's alpha = 0.372 (MAX = 0.781, MIN = −0.165, MEDIAN = 0.4), Cohen's Kappa = 0.373 (MAX = 0.777, MIN = −0.05, MEDIAN = 0.394).

Furthermore, Cohen's Kappa resulted in Kappa equal to 0.21 to 0.40, which is a fair consistency between the annotators, i.e., not a high and weak consistency.

Finally, only the sentences in which there was complete agreement between the two annotators were kept as a dataset to be used for the extraction of emotional expressions. The dataset thus created contains 721 sentences, each of which is assigned a single emotion category. The distribution of sentences per emotion category is shown in Table 5.

Table 5. Number of sentences per emotion category in the emotive sentence dataset used for emotive expression extraction.

Emotion Category	Number of Sentences
Surprise	83
Anger	75
Dislike	72
Excitement	63
Fear	70
Joy	84
Fondness	83
Relief	67
Gloom	69
Shame	55
Total	721

4.2. Emotive Expression Extraction Method

To propose a method for extracting emotive expressions from the above dataset, we conducted a detailed analysis of the dataset.

First, the dataset was subjected to word segmentation and morphological analysis using Mecab (https://taku910.github.io/mecab/ (accessed on 19 May 2024)) , a morphological analyzer for Japanese, to extract the part-of-speech information for each word.

The frequency of occurrence of each word in the dataset was determined, and a list was created in descending order of frequency. In addition, we investigated how many emotive expressions were included within the top 50 words. When checking whether a word or phrase was an emotive expression or not, we consulted various dictionaries (Japanese dictionary, Kojien, etc.) to confirm the meaning of each word or phrase, and treated it as an emotive expression if it was described as being used to express emotion.

However, the process of checking the expressions by looking through the dictionary must be performed manually. In this study, we developed a method for the automatic identification and removal of words that are not emotive expressions to automate the affect lexicon expansion.

To investigate what parts of speech are usually used for emotive expressions, we performed morphological analysis on all the expressions in the ML-Ask affect lexicon database and confirmed the distribution of morphemes and parts of speech.

4.3. Investigation of Parts of Speech Appearing in Emotive Expressions

In this investigation, we examined the parts of speech that usually appear in emotive expressions and the parts of speech that do not appear in emotive expressions, and confirmed their distribution. The purpose of the investigation was to identify the parts of speech that do not appear in emotive expressions so that such expressions can be automatically removed in the future when extracting candidate emotive expressions, thereby reducing the effort required to expand the affect lexicon and bringing it closer to fully automated expansion.

First, morphological analysis was performed using MeCab on the top 50 most frequently occurring expressions created in Section 4.2 to confirm the types of parts of speech that occur in the word groups and to create a distribution. Tables 6–15 show the distribution.

In addition, the same survey was conducted on the part-of-speech distribution of emotive expressions in the ML-Ask affect lexicon, and the results are also shown in Tables 6–15 for comparison.

Because the data used in this comparative study contained only a small number of emotive expressions, and the part-of-speech distribution in emotive expressions had to be confirmed with a small number of data, a thorough comparative study of the part-of-speech distribution based only on the automatically extracted data was not possible. For this reason, part-of-speech distributions other than emotive expressions and the part-of-speech distribution in ML-Ask's emotive expression dictionary were also used as supplements to the survey. As a result, it was found that there is no part-of-speech (POS) for any particles, auxiliaries, and participles in the emotive expressions.

This result suggests that it is effective to remove candidate words containing particles, auxiliary verbs, and coordinating verbs as a preprocessing step when extracting new emotive expressions.

Table 6. Part -of-speech distribution of surprise.

Part-of-Speech	Non-Emotive Expression	Emotive Expression Candidate Words	Emotive Expression (ML-Ask)
Nouns	11	6	36
Verbs	16	1	11
Adverbs	1	0	10
Auxiliaries	7	0	0
Auxiliary verbs	4	0	0
Pre-noun adjectival	3	0	0
Adjectives	1	0	0

Table 7. Part-of-speech distribution of anger.

Part-of-Speech	Non-Emotive Expression	Emotive Expression Candidate Words	Emotive Expression (ML-Ask)
Nouns	11	6	60
Verbs	16	1	39
Adverbs	11	0	14
Adjectives	3	0	2
Auxiliaries	11	0	0
Pre-noun adjectival	4	0	0
Auxiliary verbs	2	0	0

Table 8. Part-of-speech distribution of shame.

Part-of-Speech	Non-Emotive Expression	Emotive Expression Candidate Words	Emotive Expression (ML-Ask)
Nouns	9	1	20
Verbs	14	1	10
Adjectives	1	3	5
Adverbs	5	0	2
Auxiliaries	11	0	0
Auxiliary verbs	3	0	0
Pre-noun adjectival	2	0	0
Interjection	1	0	0

Table 9. Part-of-speech distribution of dislike.

Part-of-Speech	Non-Emotive Expression	Emotive Expression Candidate Words	Emotive Expression (ML-Ask)
Nouns	11	5	185
Verbs	10	1	92
Adjectives	2	1	54
Adverbs	5	0	20
Auxiliaries	14	0	0
Auxiliary verbs	3	0	0
Pre-noun adjectival	2	0	0
Interjection	0	0	3

Table 10. Part-of-speech distribution of fondness.

Part-of-Speech	Non-Emotive Expression	Emotive Expression Candidate Words	Emotive Expression (ML-Ask)
Nouns	9	6	93
Verbs	13	3	46
Adjectives	2	1	8
Adverbs	3	0	4
Auxiliaries	9	0	0
Auxiliary verbs	3	0	0
Pre-noun adjectival	1	0	0
Conjunction	1	0	0

Table 11. Part-of-speech distribution of fear.

Part-of-Speech	Non-Emotive Expression	Emotive Expression Candidate Words	Emotive Expression (ML-Ask)
Nouns	6	5	92
Verbs	15	3	36
Adverbs	2	0	13
Auxiliaries	11	0	0
Auxiliary verbs	1	0	0
Pre-noun adjectival	2	0	0
Adjectives	3	1	19

Table 12. Part-of-speech distribution of excitement.

Part-of-Speech	Non-Emotive Expression	Emotive Expression Candidate Words	Emotive Expression (ML-Ask)
Nouns	10	6	86
Verbs	16	2	67
Adverbs	1	0	16
Adjectives	0	1	8
Auxiliaries	8	0	0
Pre-noun adjectival	1	0	0
Auxiliary verbs	4	0	0

Table 13. Part-of-speech distribution of joy.

Part-of-Speech	Non-Emotive Expression	Emotive Expression Candidate Words	Emotive Expression (ML-Ask)
Nouns	14	1	103
Verbs	17	0	33
Adjectives	3	2	16
Adverbs	2	0	15
Auxiliaries	5	0	0
Auxiliary verbs	3	0	0
Pre-noun adjectival	2	0	0

Table 14. Part-of-speech distribution of relief.

Part-of-Speech	Non-Emotive Expression	Emotive Expression Candidate Words	Emotive Expression (ML-Ask)
Nouns	10	3	39
Verbs	15	1	18
Adverbs	5	0	8
Adjectives	3	0	4
Auxiliaries	10	0	0
Pre-noun adjectival	2	0	0
Auxiliary verbs	2	0	0

Table 15. Part-of-speech distribution of gloom.

Part-of-Speech	Non-Emotive Expression	Emotive Expression Candidate Words	Emotive Expression (ML-Ask)
Nouns	15	1	96
Verbs	15	2	37
Adjectives	1	2	19
Adverbs	2	1	13
Auxiliaries	7	0	0
Pre-noun adjectival	2	0	0
Auxiliary verbs	3	0	0

4.4. On the Segmentation Process and Word Weight Coefficients

Next, to propose a more accurate emotive expression extraction method, we examined several pre-processing methods for the dataset introduced in Section 4.1.2, and compared several weight coefficients to more accurately calculate the importance of words in a sentence.

The top 50 words in the list created with each segmentation method and weight coefficient were examined to see how many emotive expressions were included and how many emotive expressions that were not included in the ML-Ask emotive expression dictionary were included, and the combination of the segmentation process and weight coefficient that best-extracted emotion was investigated.

Specifically, we compared three methods of sharpening and two types of weight coefficients. These methods are described below.

- Segmentation Process
  The segmentation process is a method of separating and segmenting individual words in a sentence. The following three types of processing were used in this study.
  1. Word Segmentation
     Word segmentation is a method of splitting a given sentence into the smallest units (morphemes) of meaningful words. Word segmentation is the simplest and most representative segmentation method that does not use any other post-processing. In this study, MeCab was used for word segmentation.
  2. Prototype Processing
     Prototype processing is a method of dividing an input sentence into morphemes, which are the smallest units of the final meaningful word group, and returning the words to their prototype (or dictionary type). The prototype processing is used in this experiment because it unifies grammaticalized words (e.g., した "Shi Ta", している "Shi Te I Ru") into the lexical form (e.g., する "Su Ru"), which reduces the number of word types but makes important words more noticeable. This time, MeCab was used for prototype processing.
  3. Chunk Processing
     In chunk processing, the input sentence is not divided into morphemes but into phrases. This enables the extraction of not only words but also long phrases consisting of multiple words, and was used in this study. This time, CaboCha (https://taku910.github.io/cabocha/ (accessed on 19 May 2024), [41]) was used for chunk processing.
- Weighting Coefficients
  The weighting coefficients used in this study are coefficients that indicate the importance of words in a sentence. In this study, two types of weighting coefficients are used:
  1. tf*idf
     tf*idf (also tf-idf) is the value obtained by multiplying tf (word frequency) [42] by idf (inverse document frequency) [43]. Specifically, tf is the frequency of occurrence of each word per sentence, df is the total number of times the word appears in a document (e.g., dataset), and idf is the inverse of df.
  2. OkapiBM25
     OkapiBM25 [44] was used as the weighting coefficient for the other type of word importance. OkapiBM25 is a method that introduces the total number of words in a document into the tf*idf approach to calculating word importance. The weighting coefficients are calculated by taking into account the values of the number of words in the document and the average number of words, which are not taken into account in tf*idf, thus limiting the weight of unimportant words in the document more than in tf*idf. For this reason, we used it in this case for comparison with tf*idf.

4.5. Extraction Results

First, word segmentation, prototype processing, and chunk processing were performed on the prepared datasets. For each of the three preprocessed datasets, two weight calculation methods were used to extract words and chunks and to create a list of words and chunks sorted by importance in descending order.

The top 50 words/chunks from each list were then checked to see how many emotive expressions they contained. The reason for selecting the top 50 chunks was that most of the words below the top 50 were considered to be noise when we checked the top 50 chunks in the preliminary study to see how many new emotive expressions could be found.

In addition, since the dictionary cannot be used as an extension if only existing expressions can be extracted, we also checked how many of the emotive expressions that

appeared in the top 50 contained new emotive expressions that had not yet been published in ML-Ask.

The results are shown in Tables 16 and 17.

When calculating the extraction rate of emotive expressions in Table 16, the formula presented in Equation (1) was used:

$$E_r(\%) = \frac{E_{avg}}{50} \times 100 \quad (1)$$

where,

E_r is the emotive expression extraction rate,

E_{avg} is the average number of emotive expressions extracted.

In addition, the extraction rate of the unpublished emotive expressions in Table 17 was calculated using the formula presented in Equation (2):

$$E_n(\%) = \frac{E_{avgn}}{E_r} \times 100 \quad (2)$$

where

E_n is the new emotive expression extraction rate,

E_{avgn} is the average number of new emotive expressions, and

E_r is the emotive expression extraction rate.

It was found that the method using chunking as preprocessing has a higher extraction rate of emotive expressions than word segmentation and prototype processing. In the case of chunk processing, when the weighting coefficient is set to OkapiBM25, the extraction rate is 3% higher than that of tf*idf.

The extraction rate of emotive expressions not yet listed in the emotive expression dictionary (new emotive expressions) was also found to be higher when chunk processing was used for preprocessing, and when OkapiBM25 was used for the weighting coefficient than tf*idf.

As a preprocessing method, the results show that the chunking method is superior for extracting emotional expressions and unpublished emotional expressions.

As a result, the best weighting coefficient is OkapiBM25.

Table 16. Emotive expression extraction rate (average of 10 emotion categories).

	Word Segmentation	Prototype Processing	Chunk Processing
tf*idf	23%	21%	25%
OkapiBM25	19%	20%	28%

Table 17. Percentage of new emotive expressions among all extracted emotive expressions (average of 10 emotion categories).

	Word Segmentation	Prototype Processing	Chunk Processing
tf*idf	5.0%	2.4%	28%
OkapiBM25	7.5%	13%	51%

5. Experiments and Results

In this section, we describe the experiments conducted to evaluate the performance of the improved ML-Ask system by manual affect lexicon expansion and automatic expansion, respectively.

To confirm the coverage of Hiejima's dictionary, the coverage of Murakami's dictionary, and the overall performance of the integrated system, we conducted two evaluation experiments. Also, to select the more accurate emotion extraction method from the combination of the three methods of segmentation and the two methods of calculating word importance weights, we conducted an evaluation experiment. As a result, we first conducted four similar experiments.

To conduct evaluation experiments, it was first necessary to create test data. In these evaluation experiments, two sets of test data created in previous studies were combined and used. Specifically, we used the dataset used by Sakai et al. [21] for the affect analysis of emoticons and the dataset created by Ptaszynski et al. [6] for the development of ML-Ask. The test data contained 280 emotive and non-emotive sentences labeled with 10 emotion categories, which were treated as correct data for evaluation.

5.1. Evaluation Experiment 1: Coverage of Hiejima's Dictionary and Murakami's Dictionary

5.1.1. Experimental Setup

The setup for the evaluation experiment was as follows. First, we created a list of candidate emotive expressions using only the emotive expressions from Hiejima's dictionary and that from Murakami's dictionary. Next, we added each of the prepared lists to the affect lexicon of ML-Ask, evaluated them on the test data using the ML-Ask baseline model, and checked for changes in the results.

Furthermore, when checking the accuracy of the affect analysis, we first checked whether the output matched the correct data completely (exact match rate). However, since it is impractical to obtain the exact match rate when multiple emotion types are expressed in a single sentence from the correct data, we also checked whether at least one emotion type per sentence was detected (partial match rate) in addition to the exact match rate. In addition to the specific emotion type, we also used Russell's two-dimensional model of emotion [27], which is incorporated in ML-Ask, to determine whether the emotion expressed in the sentence was positive or negative, that is, the probability that the polarity of the emotion matched, as well as the probability that the activation dimension of the emotion was consistent. In addition, we also checked the probability that the system incorrectly extracted emotions from sentences that did not express emotions. The results are shown in Table 18.

Table 18. Results of Experiments 1 and 2.

	ML-Ask Baseline	Hiejima's Dict.	Murakami's Dict.	Integration
partial match rate	66%	19%	1%	68%
exact match rate	40%	11%	0%	40%
polarity dimension match rate	65%	21%	2%	64%
activity dimension match rate	58%	18%	2%	59%
incorrect extraction rate	3%	2%	0%	3%

5.1.2. Results and Discussion of Experiment 1

The results in Table 18 show that all indicators in this experiment were significantly reduced compared to the ML-Ask baseline. This suggests that the coverage of Hiejima's dictionary and the coverage of Murakami's dictionary are both low since it is also consistent with the fact that there are about 1100 emotive expressions in Hiejima's dictionary and 365 emotive expressions in Murakami's dictionary, while there are about 2620 emotive expressions in Nakamura's dictionary. In other words, there is a significant difference between the emotive expressions in Hiejima's dictionary and those in the database of ML-Ask, as well as that in Murakami's dictionary. Therefore, we expected that integrating Hiejima's dictionary and Murakami's dictionary into ML-Ask will improve the accuracy rate. Consequently, we integrated Hiejima's dictionary and Murakami's dictionary into ML-Ask and conducted an additional experiment.

5.2. Evaluation Experiment 2: Expansion of the Affect Lexicon

In the additional experiment, we added emotive expressions from Hiejima's dictionary and Murakami's dictionary to the database of ML-Ask by subtracting duplicated emotive

expressions. The results of the re-evaluation after the lexicon was integrated are shown in Table 18.

Results and Discussion of Experiment 2

In the results of Experiment 2, the partial match rate and the polarity match rate were improved by 2 pt. (percentage points), and 1 pt., respectively. However, the exact match rate decreased by 4 pt., while the polarity match rate also decreased but only by about 1 pt. The incorrect emotion extraction rate was unchanged from that of the ML-Ask baseline model, suggesting that the improvement was successful.

An investigation of the reasons for the decrease in the exact match rate revealed that emotive expressions were added to some of the analysis results, and the number of emotion types analyzed was also increased. This suggests that the more emotive expressions there are in the database, the lower the percentage of exact matches.

The experiment also suggests that other available dictionaries should also be added to improve the performance of the ML-Ask system. Moreover, the results of the experiment show the limitations of such manually collected dictionaries. Therefore, the experiment of the automatic expansion of the affect lexicon was expected to attain a good result. In addition, the test data for the evaluation experiment need to be increased to show a more diverse spectrum of how emotions are expressed.

5.3. Evaluation Experiment 3: Automatic Expansion of Affect Lexicon

5.3.1. Experimental Setup

The setup of this evaluation experiment was as follows:

1. Six combinations of each preprocessing and weighting coefficient (three types of segmentation times two types of weighting coefficient) were used to create a list of candidate emotive expressions.
2. Evaluated with test data using ML-Ask baseline model.
3. Added each of the lists prepared in (1), above, to the ML-Ask affect lexicon, evaluated each combination with test data, and confirmed changes in the results.

As described in (1) above, the combinations of preprocessing and weighting coefficients used in this experiment were as follows:

- Word Segmentation + tf*idf;
- Word Segmentation + BM25;
- Prototype Processing + tf*idf;
- Prototype Processing + BM25;
- Chunk Processing + tf*idf;
- Chunk Processing + BM25.

To confirm whether and to what extent the expansion of the affect lexicon can be fully automated, we first added the top 50 words from the above list of candidate emotive expression words to the ML-Ask affect lexicon database as they are, and conducted an evaluation experiment using the above procedure.

In addition, to check the accuracy of the affect analysis, the output was compared to the correct data, as in the previous experiment, to check the percentage of exact and partial match rates.

The results of the exact and partial match analysis are shown in Table 19.

Also, we checked the polarity dimension match rate, the activity dimension match rate, and the incorrect extraction rate. Their results are shown in Table 20.

Table 19. Result of Experiment 3-1.

	Partial Match Rate	Exact Match Rate
ML-Ask baseline	66%	41%
Word Segmentation +tf*idf	76% (+10%) *	19% (−22%)
Word Segmentation +BM25	72% (+6%)	23% (−18%)
Prototype Processing +tf*idf	74% (+8%)	15% (−26%)
Prototype Processing +BM25	72% (+6%)	23% (−18%)
Chunk Processing +tf*idf	75% (+9%)	26% (−15%)
Chunk Processing +BM25	66% (+0%)	35% (−6%)

* The improvement is shown in red color font and the deterioration is shown in blue color font.

Table 20. Results of Experiment 3-2.

	Polarity Dimension Match Rate	Activity Dimension Match Rate	Incorrect Extraction Rate
ML-Ask baseline	65%	58%	2%
Word Segmentation +tf*idf	72% (+7%) *	62% (+4%)	7% (+5%)
Word Segmentation +BM25	70% (+5%)	64% (+6%)	6% (+4%)
Prototype Processing +tf*idf	71% (+6%)	59% (+1%)	8% (+6%)
Prototype Processing +BM25	70% (+5%)	64% (+6%)	6% (+4%)
Chunk Processing +tf*idf	74% (+9%)	70% (+12%)	3% (+1%)
Chunk Processing +BM25	73% (+8%)	66% (+8%)	3% (+1%)

* The improvement is shown in red color font and the deterioration is shown in blue color font.

5.3.2. Results and Discussion of Experiment 3

The results in Tables 19 and 20 show that, compared to the ML-Ask baseline, the current experiment improved the partial match rate by an average of 6.5%, the polarity match rate by an average of 6.7%, and the activity match rate by an average of 6.16%. However, the exact match rate is not improved and is 17.5% lower on average than the baseline, and the incorrect extraction rate is also 3.5% higher on average than the baseline.

From the experimental results, the partial match rate increased up to 10%, depending on the method, indicating that many emotive expressions can be automatically extracted in the evaluation using the present dataset. However, the fact that the incorrect emotion extraction rate also increased at the same time suggests that the full automation of the extraction process may result in the extraction of expressions that are not emotive expressions, which may have reduced the evaluation of the system. Specifically, in this experiment, the top 50 expressions for each emotion category in the list of candidate emotive expressions from the emotive sentence dataset (Section 4.1.2) were added to the dictionary as candidate emotive expressions, so the list includes expressions that have nothing to do with emotive expressions, which was the source of the lower results.

Since the partial match rate, emotional polarity match rate, and emotional activity match rate have increased, the full automation of emotive expression expansion is basically possible. However, since the exact match rate and the incorrect emotion extraction rate

have also increased, there remains the issue of removing expressions other than emotive expressions after the fully automated extraction.

Therefore, we conducted an additional evaluation experiment this time by manually removing non-emotive expressions.

5.4. Evaluation Experiment 4: Improvement of the Auto-Expanded Affect Lexicon

In this additional evaluation experiment, we manually checked the expressions added by automatic expansion to the ML-Ask affect lexicon database, checked the meaning of each expression, whether it was an emotive expression and in what context it was used, by consulting a general Japanese dictionary, and manually removed expressions other than emotive expressions.

Table 21 shows the partial and exact match rates for the re-evaluation after manually removing non-emotive expressions, and Table 22 shows the polarity match rate, activity match rate, and incorrect extraction rate.

Table 21. Results of Experiment 4-1.

	Partial Match Rate	**Exact Match Rate**
ML-Ask baseline	66%	41%
Word Segmentation +tf*idf	67% (+1%) *	40% (−1%)
Word Segmentation +BM25	67% (+1%)	40% (−1%)
Prototype Processing +tf*idf	65% (−2%)	39% (−2%)
Prototype Processing +BM25	67% (+1%)	39% (−2%)
Chunk Processing +tf*idf	67% (+1%)	38% (−3%)
Chunk Processing +BM25	67% (+1%)	40% (−1%)

* The improvement is shown in red color font and the deterioration is shown in blue color font.

Table 22. Results of Experiment 4-2.

	Polarity Dimension Match Rate	**Activity Dimension Match Rate**	**Incorrect Extraction Rate**
ML-Ask baseline	65%	58%	2%
Word Segmentation +tf*idf	72% (+7%) *	65% (+7%)	2% (+0%)
Word Segmentation +BM25	72% (+7%)	66% (+8%)	2% (+0%)
Prototype Processing +tf*idf	72% (+7%)	65% (+7%)	2% (+0%)
Prototype Processing +BM25	72% (+7%)	66% (+8%)	2% (+0%)
Chunk Processing +tf*idf	72% (+7%)	66% (+8%)	2% (+0%)
Chunk Processing +BM25	72% (+7%)	66% (+8%)	2% (+0%)

* The improvement is shown in red color font.

Results and Discussion of Experiment 4

The results of the current evaluation Experiment 4 show a similar decrease in the exact match rate but only by about 1.7% on average, while the other indices are improved.

The incorrect extraction rate is also unchanged from that of the ML-Ask baseline model, suggesting that the improvement was successful.

Additionally, the segmentation method that led to the greatest improvement in accuracy is chunking, and the weighting coefficient is OkapiBM25.

Compared to the baseline system, only the exact match rate decreased slightly, but all other measures improved by an average of 3.3 percentage points.

We investigated the reason for the decrease in the percentage of exact match rate and found that all of the analyses that did not result in exact matches included emotive expressions (e.g., 情 "Jyou") consisting of a single character with ambiguous meanings.

This suggests that emotive expressions consisting of a single character should be removed from the dictionary.

In addition, to improve the system in this experiment, the emotive expressions had to be removed manually. To solve this problem in the future, it will be necessary to develop a method to automatically remove expressions other than emotive expressions.

The improvement in results was small, but this was due to the fact that the number of words that could be added to the affect lexicon was very small because only the top 50 candidate emotive expressions were used in this experiment.

To solve this problem, it is necessary to expand the range of candidate words for emotive expressions (e.g., from the top 50 to 100), increase the data from which the candidate words are extracted, and recheck the results on a larger dataset.

5.5. Additional Experiment

To increase the data for evaluation Experiments 1 and 2, we decided to use questionnaire data from the study of automatic expansion of the affect lexicon. The additional data included 1012 emotive sentences with 10 different emotion labels, which is roughly 100 sentences for each emotion category, which were treated as the correct data for the additional experiment.

Results and Discussion of Additional Experiment

The results of the additional experiment are shown in Table 23. This time, since the test data were annotated with only one emotion label, the "partial match rate" metric in the experiment meant that one of the detected emotion categories was the correct emotion label, and therefore, no other metrics were used.

Table 23. Results of additional experiment.

	ML-Ask Baseline	**Integration**
partial match rate	93.28%	93.38%
number of sentences correctly analyzed	944	945

Since the additional data are created based on the original ML-Ask system, it should be expected that a good result, i.e., a relatively high "partial match rate", will be obtained. Also, the expected increase in the "partial match rate" between unintegrated and integrated systems would not be high.

As a result of additional experiments, which were not unexpected, the "partial match rate" remained unchanged at 93%. Specifically, the unintegrated system could analyze 944 sentences correctly, while the integrated system could correctly analyze 945 sentences. The difference is the word "汚い Kitanai", which means dirty in English.

Therefore, depending on the test data used in the evaluation experiments, the degree of improvement in the performance of the system is different. In general, the improvement in system performance is not that significant.

6. Discussion

In this section, we discussed this study, including the main findings, theoretical and practical implications, originality of the study, and study limitations.

6.1. Main Findings

Our research demonstrated that integrating additional emotion dictionaries into the ML-Ask system improved its performance, albeit modestly, by approximately 2 percentage points. The manual expansion process highlighted significant differences in the interpretation of emotive expressions across dictionaries, underscoring the complexity of emotion lexicon development. Additionally, our automatic extraction method, while promising, introduced non-emotive expressions, necessitating manual curation to maintain accuracy. Specifically, chunk processing and the OkapiBM25 weighting coefficient proved most effective for segmentation and weighting, respectively.

6.2. Theoretical and Practical Implications

This study contributes to the field of emotion analysis by demonstrating a method for enhancing lexicon-based systems through dictionary integration and automatic expression extraction. The findings suggest that while manual expansion remains labor-intensive, it provides a significant improvement in the system's accuracy and granularity. The automatic extraction method shows potential for scalability, although it requires further refinement to reduce the inclusion of non-emotive expressions. These advancements have practical implications for improving sentiment analysis, affective computing, and human–computer interaction applications by providing more nuanced emotion detection capabilities.

6.3. Originality of the Study

The originality of this study lies in its dual approach to expanding the affect lexicon of the ML-Ask system: combining manual integration of multiple dictionaries with an automated extraction mechanism. This hybrid method not only enhances the system's vocabulary but also addresses the challenges associated with integrating diverse emotion expression standards.

6.4. Study Limitations

A limitation of our study is the reliance on manual curation to remove non-emotive expressions from the automatically expanded lexicon, which is time-consuming and reduces scalability. Additionally, in the evaluation experiment, the exact match rate decreased slightly, but the other indices were improved. Additionally, the decreased exact match rate due to single-character expressions (e.g., 喜 "Ki" and 情 "Jyou") highlights the need to exclude emotive expressions consisting of a single character from the dictionary to further improve the system's hit probability. The study's focus on Japanese text, which requires word segmentation before processing, may limit the generalizability of our findings to languages with different linguistic structures.

7. Conclusions and Future Work

This study aimed to enhance the emotion analysis capabilities of the ML-Ask system by expanding its affect lexicon through the integration of multiple emotion dictionaries and implementing an automatic extraction mechanism for emotive expressions. Our approach involved the manual addition of emotive expressions from two external dictionaries and the automatic extraction of candidate expressions from a large-scale blog corpus.

According to the results of the evaluation experiments, it can be concluded that the two available dictionaries of emotive expressions and the dictionary used in the ML-Ask database have a significant difference in the understanding of emotive expressions, and this study succeeded in improving the performance of the system, although by a small amount (2 percentage points).

On the other hand, we extracted emotive sentences from a large-scale blog corpus and automatically extracted candidate emotive expressions from them to expand the affect lexicon of a lexicon-based affect analysis system.

Specifically, to expand the affect lexicon, we first created a dataset with information about user emotions. We then compared the dataset for three types of segmentation and two types of weighting coefficients. The results showed that chunk processing was the most effective segmentation method, and OkapiBM25 was the most effective weighting coefficient.

Although the automatic expansion of the affect lexicon was possible, the fully automatic method included expressions other than emotive expressions, which caused an increase in the incorrect extraction rate of emotions. Therefore, we further manually eliminated non-emotive expressions from the automatically expanded affect lexicon, thereby solving the problem of the incorrect extraction rate.

Future research will explore the expansion of the affect lexicon by converting Kanji emotive expressions into Kana expressions and evaluating their usage in the Japanese YACIS corpus. We also plan to develop a fully automated lexicon expansion method using large language models, aiming to enhance scalability and accuracy. Moreover, integrating transformer-based models with the ML-Ask system will be investigated to leverage advanced natural language processing capabilities for more effective emotion analysis.

By addressing these avenues, we aim to further advance the field of affect analysis, offering robust tools for sentiment analysis and improving human-computer interactions across diverse applications.

Author Contributions: Conceptualization, L.W., S.I. and M.P.; methodology, L.W. and S.I.; software, M.P., P.D., R.R. and F.M.; validation, L.W., S.I., M.P. and F.M.; formal analysis, L.W. and S.I.; investigation, L.W. and S.I.; resources, L.W., S.I., M.P., P.D., Y.U. and R.R.; data curation, L.W., S.I., M.P., P.D., Y.U. and R.R.; writing—original draft preparation, L.W. and S.I.; writing—review and editing, L.W. and M.P.; visualization, L.W. and S.I.; supervision, M.P. and F.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The emotion dataset for the Japanese language used to support the findings of this study is available upon reasonable request to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
CVS	Contextual Valence Shifters
MeCab	Yet Another Part-of-Speech and Morphological Analyzer
ML-Ask	eMotive eLement and Expression Analysis system
NLP	Natural Language Processing
OkapiBM25	Term weighting for Okapi information retrieval system with Best Matching 25
OOV	Out-of-Vocabulary
POS	Part-of-Speech
SNS	Social Networking Services
TF*IDF	Term weighting using Term Frequency with Inverse Document Frequency
TF-IDF	See TF*IDF
YACIS	Yet Another Corpus of Internet Sentences

References

1. Beigi, G.; Hu, X.; Maciejewski, R.; Liu, H. An Overview of Sentiment Analysis in Social Media and Its Applications in Disaster Relief. In *Sentiment Analysis and Ontology Engineering: An Environment of Computational Intelligence*; Springer: Cham, Switzerland, 2016; pp. 313–340.
2. Jain, V.K.; Kumar, S.; Fernandes, S.L. Extraction of emotions from multilingual text using intelligent text processing and computational linguistics. *J. Comput. Sci.* **2017**, *21*, 316–326. [CrossRef]
3. Gaind, B.; Syal, V.; Padgalwar, S. Emotion detection and analysis on social media. *arXiv* **2019**, arXiv:1901.08458.
4. Yue, L.; Chen, W.; Li, X.; Zuo, W.; Yin, M. A survey of sentiment analysis in social media. *Knowl. Inf. Syst.* **2019**, *60*, 617–663. [CrossRef]
5. Avasthi, S.; Chauhan, R.; Acharjya, D.P. Information Extraction and Sentiment Analysis to Gain Insight into the COVID-19 Crisis. In Proceedings of the International Conference on Innovative Computing and Communications: ICICC 2021, Delhi, India, 20–21 February 2021; Springer: Berlin/Heidelberg, Germany, 2022; Volume 1, pp. 343–353.
6. Ptaszynski, M.; Dybala, P.; Rzepka, R.; Araki, K.; Masui, F. ML-Ask: Open source affect analysis software for textual input in Japanese. *J. Open Res. Softw.* **2017**, *5*, 16. [CrossRef]
7. Nakamura, A. *Kanjo Hyogen Jiten [Dictionary of Emotive Expressions]*; Tokyodo Publishing: Tokyo, Japan, 1993. (In Japanese)
8. Sharma, S.; Kumar, P.; Kumar, K. LEXER: Lexicon Based Emotion Analyzer. In Proceedings of the International Conference on Pattern Recognition and Machine Intelligence, Kolkata, India, 5–8 December 2017; Springer: Berlin/Heidelberg, Germany, 2017; pp. 373–379.
9. Toçoğlu, M.A.; Alpkocak, A. Lexicon-based emotion analysis in Turkish. *Turk. J. Electr. Eng. Comput. Sci.* **2019**, *27*, 1213–1227. [CrossRef]
10. Kamal, R.; Shah, M.A.; Maple, C.; Masood, M.; Wahid, A.; Mehmood, A. Emotion classification and crowd source sensing—A lexicon based approach. *IEEE Access* **2019**, *7*, 27124–27134. [CrossRef]
11. Asghar, M.Z.; Khan, A.; Bibi, A.; Kundi, F.M.; Ahmad, H. Sentence-level emotion detection framework using rule-based classification. *Cogn. Comput.* **2017**, *9*, 868–894. [CrossRef]
12. Gao, K.; Xu, H.; Wang, J. A rule-based approach to emotion cause detection for Chinese micro-blogs. *Expert Syst. Appl.* **2015**, *42*, 4517–4528. [CrossRef]
13. Nasir, A.F.A.; Nee, E.S.; Choong, C.S.; Ghani, A.S.A.; Majeed, A.P.A.; Adam, A.; Furqan, M. Text-based emotion prediction system using machine learning approach. *IOP Conf. Ser. Mater. Sci. Eng.* **2020**, *769*, 012022. [CrossRef]
14. Xu, D.; Tian, Z.; Lai, R.; Kong, X.; Tan, Z.; Shi, W. Deep learning based emotion analysis of microblog texts. *Inf. Fusion* **2020**, *64*, 1–11. [CrossRef]
15. Gupta, M.R.; Bengio, S.; Weston, J. Training highly multiclass classifiers. *J. Mach. Learn. Res.* **2014**, *15*, 1461–1492.
16. Hejima, I. *A Short Dictionary of Feelings and Emotions in English and Japanese*; Tokyodo Shuppan: Tokyo, Japan, 1995.
17. Murakami, M. *Love, Hate and Everything in Between: Expressing Emotions in Japanese*; Kodansha International: Tokyo, Japan, 2002.
18. Kobayashi, T.; Ishii, K.; Edani, N.; Kondo, Y.; Adachi, Y. Kanjogo Jisho to Kanjo Bunseki Shisutemu EEAS [Dictionariy of Emotion Expression and Emotion Analysis System EEAS]. In Proceedings of the 83rd National Convention of IPSJ, Online, 18–20 March 2021; Volume 2021, pp. 47–48. (In Japanese)
19. Tatsuya, T.; Masafumi, H. A Proposal of Emotion Estimation Method for Words and Construction of Word-Emotion Dictionary. *Trans. Jpn. Soc. Kansei Eng.* **2019**, *18*, 273–278.
20. Minato, J.; Bracewell, D.B.; Ren, F.; Kuroiwa, S. Statistical Analysis of a Japanese Emotion Corpus for Natural Language Processing. In *Computational Intelligence*; Huang, D.S., Li, K., Irwin, G.W., Eds.; Springer: Berlin/Heidelberg, Germany, 2006; pp. 924–929.
21. Sakai, T.; Ptaszynski, M.; Masui, F. Kao Moji Patsu ni Motozui ta Kao Moji no Jido Seisei no Kano Sei ni Kansuru Chosa [Study on Potential of Automatic Emotion Generation based on Emoticon Parts]. In Proceedings of the 33rd Annual Conference of the Japanese Society for Artificial Intelligence, Tokyo, Japan, 4–7 June 2019; Volume JSAI2019, p. 2E4OS903. (In Japanese) [CrossRef]
22. Ptaszynski, M.; Dybala, P.; Rzepka, R.; Araki, K.; Momouchi, Y. YACIS: A Five-Billion-Word Corpus of Japanese Blogs Fully Annotated with Syntactic and Affective Information. In Proceedings of the AISB/IACAP World Congress, Birmingham, UK, 2–6 July 2012; pp. 40–49.
23. Vo, B.K.H.; Collier, N. Twitter emotion analysis in earthquake situations. *Int. J. Comput. Linguist. Appl.* **2013**, *4*, 159–173.
24. Steinborn, V.; Maronikolakis, A.; Schütze, H. Politeness Stereotypes and Attack Vectors: Gender Stereotypes in Japanese and Korean Language Models. *arXiv* **2023**, arXiv:2306.09752.
25. Ptaszynski, M.; Rzepka, R.; Araki, K.; Momouchi, Y. Automatically annotating a five-billion-word corpus of Japanese blogs for sentiment and affect analysis. *Comput. Speech Lang.* **2014**, *28*, 38–55. [CrossRef]
26. Kudo, T. Mecab: Yet Another Part-of-Speech and Morphological Analyzer. 2005. Available online: http://mecab.sourceforge.net/ (accessed on 19 May 2024).
27. Russell, J.A. A circumplex model of affect. *J. Personal. Soc. Psychol.* **1980**, *39*, 1161. [CrossRef]
28. Ptaszynski, M.; Masui, F.; Ishii, N. A method for automatic estimation of meaning ambiguity of emoticons based on their linguistic expressibility. *Cogn. Syst. Res.* **2020**, *59*, 103–113. [CrossRef]
29. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient estimation of word representations in vector space. *arXiv* **2013**, arXiv:1301.3781.

30. Le, Q.; Mikolov, T. Distributed Representations of Sentences and Documents. In Proceedings of the International Conference on Machine Learning, PMLR, Beijing, China, 22–24 June 2014; pp. 1188–1196.
31. Manabe, H.; Oka, T.; Umikawa, S.; Takaoka, K.; Uchida, Y.; Asahara, M. Fukusu Ryudo no Bunkatsu Kekka ni Motozuku Nihongo Tango Bunsan Hyogen [Japanese Word Distributed Representation Based on Multi-Granular Segmentation Results]. In Proceedings of the 25th Annual Conference of the Association for Natural Language Processing (NLP2019), Nagoya, Japan, 12–15 March 2019; The Association for Natural Language Processing: Tokyo, Japan, 2019; p. NLP2019-P8-5.
32. Bojanowski, P.; Grave, E.; Joulin, A.; Mikolov, T. Enriching Word Vectors with Subword Information. *Trans. Assoc. Comput. Linguist.* **2017**, *5*, 135–146. [CrossRef]
33. Joulin, A.; Grave, E.; Bojanowski, P.; Douze, M.; Jégou, H.; Mikolov, T. FastText.zip: Compressing text classification models. *arXiv* **2016**, arXiv:1612.03651.
34. Joulin, A.; Grave, E.; Bojanowski, P.; Mikolov, T. Bag of Tricks for Efficient Text Classification. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, Valencia, Spain, 3–7 April 2017; Short Papers; Association for Computational Linguistics: Stroudsburg, PA, USA, 2017; Volume 2, pp. 427–431.
35. Ptaszynski, M.; Momouchi, Y.; Maciejewski, J.; Dybala, P.; Rzepka, R.; Araki, K. Annotating Japanese Blogs with Syntactic and Affective Information. In *Mining User Generated Content*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2014; pp. 229–262.
36. Leonova, V. Review of Non-English Corpora Annotated for Emotion Classification in Text. In Proceedings of the Databases and Information Systems: 14th International Baltic Conference, DB&IS 2020, Tallinn, Estonia, 16–19 June 2020; Proceedings 14; Springer: Berlin/Heidelberg, Germany, 2020; pp. 96–108.
37. Ihasz, P.L.; Van, T.H.; Kryssanov, V.V. A Computational Model for Conversational Japanese. In Proceedings of the 2015 International Conference on Culture and Computing (Culture Computing), Kyoto, Japan, 17–19 October 2015; pp. 64–71.
38. Nasser, A. Large-Scale Arabic Sentiment Corpus and Lexicon Building for Concept-Based Sentiment Analysis Systems. Ph.D. Thesis, Graduate School of Science and Engineering of Hacettepe University, Ankara, Turkey, 2018.
39. Barbaresi, A. Efficient Construction of Metadata-Enhanced Web Corpora. In Proceedings of the 10th Web as Corpus Workshop, Berlin, Germany, 12 August 2016; pp. 7–16.
40. Ptaszynski, M.; Rzepka, R.; Araki, K.; Momouchi, Y. A Robust Ontology of Emotion Objects. In Proceedings of the Eighteenth Annual Meeting of The Association for Natural Language Processing (NLP-2012), Silchar, India, 16–19 December 2012; pp. 719–722.
41. Taku Kudo, Y.M. Japanese Dependency Analysis using Cascaded Chunking. In Proceedings of the CoNLL 2002: 6th Conference on Natural Language Learning 2002 (COLING 2002 Post-Conference Workshops), Taipei, Taiwan, 31 August–1 September 2002; pp. 63–69.
42. Luhn, H.P. A statistical approach to mechanized encoding and searching of literary information. *IBM J. Res. Dev.* **1957**, *1*, 309–317. [CrossRef]
43. Jones, K.S. A statistical interpretation of term specificity and its application in retrieval. *J. Doc.* **1972**, *28*, 11–21. [CrossRef]
44. Robertson, S.E.; Walker, S.; Jones, S.; Hancock-Beaulieu, M.M.; Gatford, M. *Okapi at TREC-3*; NIST Special Publications (SPs); NIST: Gaithersburg, MD, USA, 1995; Volume 109, p. 18.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.