Utilizing Figurative Language Examples for Recognizing Japanese Sentences with Metaphors

Mateusz Babieno, Shō Takishita, Rafał Rzepka and Kenji Araki

Graduate School of Information Science and Technology, Hokkaido University,
Sapporo, Kita-ku, Kita 14 Nishi 9, 060-0814, Japan
{babieno,takishita,rzepka,araki}@ist.hokudai.ac.jp,
WWW home page: http://arakilab.media.eng.hokudai.ac.jp/

Abstract. Recently we have proposed highly efficient method for recognizing Japanese sentences containing metaphors by utilizing figurative language examples from a dictionary. Having proven high efficiency of the proposed method when trained on distinctly metaphorical and non-metaphorical data, we proceeded to test it against text data containing a mix of figurative and non-figurative language. In order to do so, we have excerpted test data from pieces of Japanese literature available at *Aozora Bunko* digital library. We introduce the experimental results and discuss some general issues regarding understanding the notion of figurativeness.

1 Introduction

Figurative expressions are ubiquitous in human language and thus their processing from computational perspective should be considered a task of great importance. For example, in her research from 2003, Cameron shows that on average words are being used metaphorically 50 out of 1000 times during everyday conversation [8]. Also, according to Shutova and Teufel's corpus analysis from 2010, statistically in the span of 3 sentences there is one verb used figuratively [13]. However, metaphorical expressions are difficult to be processed by computers because exactly the same phrase, e.g. "this is a disaster", may have literal or figurative meaning depending on the context.

As the ultimate goal of our research is to construct an efficient figurative expressions' generating system, the issue highlighted by Utsumi in his work [15] becomes very interesting from our perspective. He discusses the relation of metaphor to simile, and proposes the category of interpretative diversity which might be a key factor in choosing whether metaphor or simile is more appropriate in a given context. In order to be able to investigate contextual influences on figurative language with statistical approach, we decided to train a classifier to help us to mechanize the process of finding metaphors in large-test corpora.

In the previous work [6] we introduced results of a classification experiment designed to recognize sentences containing figurative expressions in Japanese text [6]. For the experiment we have utilized two datasets: one of figurative expressions and another consisting mostly literal sentences. The former was taken from a collection of manually chosen examples from literature [12]. The "non-figurative" set was created by retrieving data from three different sources: Japanese Wikipedia, local assembly minutes and news articles.

In this paper we describe further experiments aiming at testing how effectively the classifier labels Japanese text. They were conducted in the following three steps. First, we have used 9,152 simile-containing sentences and the same number of literal ones. These data have already been used in our previous experiment [6]. In the second stage we have replaced sentences comprising similes with the ones containing narrowly-defined metaphors[1]. Lastly, we have combined sentences including both types of figurative expressions, in the end achieving 18,304 of those. By adding another 9,152 literal sentences we have also prepared "non-figurative" set of the corresponding size.

As for the test data, we have randomly excerpted sentences from novels' texts stored in *Aozora Bunko* digital library [1]. These were labelled by two annotators: one Japanese language native-speaker and one foreigner (the first author) with considerable level of fluency in Japanese language. As their estimations turned out to be quite different – they used different labels 40 out of 100 times (κ coefficient = 0.297) – we decided to conduct each experiment twice.

As one could have imagined, the results prove that it is indeed difficult to computationally discern figurative usage of language within the text which does not belong to distinct metaphorical or non-metaphorical set of sentences. The results of the tests are shown in Section 4.

2 Related research

Contrary to back then popular view, metaphors have been proposed not to be mere rhetorical figures used almost solely in poetry but rather a profound cognitive mechanism construing large part of any given utterance and widely noticeable in everyday language [11]. Metaphorical expressions, or to put it more broadly, figurative language, is so prevailing in our daily communication, that people can hardly distinguish it from literal language. Probably most explicitly expressed by Lakoff and Johnson in their *Metaphors We Live By*, this discovery has become a turning point in scholarship, mainly linguistics. From that point on, vast number of research papers and whole books have been dedicated solely to the analysis of this phenomenon.

Alongside pragmatics and semantics, natural language processing has become another field in which figurative language's analysis began to be an area of great interest. Bulat et al. [7] has recently proposed the first metaphor identification method using representations constructed from embedding-based property norms (approx. 75% accuracy). Tsvetkov and colleagues [14] have shown that

[1] In the previous study we have utilized examples from Onai's dictionary which contained comparative connectors, e.g. *-no yō na* or *-mitai na*. As the term "metaphor" is sometimes used too loosely, in order to avoid confusion, throughout the paper we use the name "narrowly-defined metaphors" to address expressions such as *love is a travel* in contrast to *love is like a travel* type similes.

their method using subject-verb-object and adjective-noun sets can be used for English, Spanish, Farsi and Russian. Although they claim their model to be effective irrespectively of the target language, it is doubtful when thinking about using raw, "real-world" data. To quote Dobrzyńska: "Problems of metaphor can be most clearly seen and defined when a metaphorical expression is to be translated, that is, when its sense is to be conveyed in another language. Another language also means another cultural background and another value system of other listeners or readers" [9]. Recognizing metaphorical, not straightforward meaning in a sentence may be useful also in machine translation.

As mentioned in Introduction, recently we have proposed highly efficient method for recognizing Japanese sentences containing metaphors by utilizing figurative language examples from a dictionary [6]. The model was trained also on non-figurative language sentences taken from Japanese Wikipedia [5], local assembly minutes [2] and news articles [3]. After testing three basic text classification methods (Naïve Bayes, Support Vector Machines and Neural Network) we confirmed very high precision and recall (94-98% F-score depending on the training-testing data size ratio) achieved by the algorithm.

Nonetheless, it was rather clear that the very high score achieved by our model was gained thanks to the high level of uniformity of the data used. In order to check whether the model works effectively independently of its target data, we have decided to use texts from Japanese-language *Aozora Bunko* digital library as a source for the new test-set. Indeed, the results achieved through several experiments show great decline of the model's efficiency and thus prove that our considerations were correct.

3 Datasets for Classification

Data we have used for training can be largely divided into two groups: one of the figurative expressions and the other of the literal ones. For the first and the second stages of our experiment 9,152 sentences previously used in our former experiment have been adopted as a "non-figurative" train-set. As for the figurative expressions, we have utilized equivalent number of sentences comprising only similes – also the ones used in the previous experiment. In the second stage, we have replaced sentences containing similes with the ones including narrowly-defined metaphors from [12]. In the last step, we have concatenated both "figurative" groups, eventually achieving 18,304 sentences. In order to get an equal amount of non-figurative data, we have added another 9,152 sentences taken from sources generally deemed as literal. Those were: the latest Japanese Wikipedia dump, local assembly minutes and articles from Livedoor News.

As for the test-data, we have randomly excerpted 100 sentences from Japanese novels' texts, using *Aozora Bunko* digital library. Because the evaluation of whether a certain expression is used figuratively or literary often depends on the context [9], we have also provided adjacent sentences (the previous and the next) prior to asking for an annotation. There were two annotators – one Japanese language native speaker and one foreign student equipped with considerable level of fluency (the first author of this paper). In numerous cases, their evaluation was different: we have therefore decided that the experiment should be conducted twice for each evaluation to empirically show the level of variation. As a result, we have achieved two different sets of scores.

4 Experiment and Results

Just like before, we have used Python's scikit-learn library for the classification [4]. As for the features, we used word count vectors of 300 dimensions for each sentence of training set. Each word count vector contains the frequency of 300 words in the training file. We have used standard state-of-the art models, namely Naïve Bayes (multinomial), Support Vector Machines (Linear SVC) and Artificial Neural Nets (multi-layer perceptron with stochastic gradient descent, 10,000 max. iterations). Five-fold cross-validation was used.

The experiment has been conducted in three stages. First, we have utilized datasets from already mentioned previous experiment without adding any changes. The train-sets were constructed using sentences containing similes on the one hand and literal expressions on the other: precisely 9,152 sentences from each group. Irrespectively of the method used, the accuracy score did not reach even 40% using foreigner's annotation. It got significantly better after changing labels to the ones prepared by the native speaker (78% accuracy when SVM and artificial neural network were used).

In the next stage, we have used sentences containing narrowly-defined metaphors instead of similes. The accuracy was almost equally low using foreigner annotator's labels; it has also visibly dropped using native speaker's labelling.

As for the third part of our experiment, we have added sentences comprising metaphors to the ones with similes, gaining 18,304 "figurative" sentences in total. While working on foreigner-annotated data the accuracy got slightly better comparing it with the previous stages, but calling it good would be quite an overstatement. Interestingly enough, it got significantly worse in case of the native speakers' annotation, which was giving better results so far.

All the results can be compared in Tables 1 and 2. The efficiency of our model – at least having it work on a "real-world data" – has to be questioned and further improvements are clearly needed.

5 Result Analysis and Considerations

Within sentences adopted as a test-data there were numerous expressions whose usage's evaluation regarding figurativeness led to disagreement between the annotators. Those were for example[2]:

- *rekishi no shita dewa*: 'historically' [lit. 'under history'];

[2] In this work Japanese expressions are presented in italics. We have adpoted widely used Hepburn's romanization.

Table 1. Results for the native-speaker evaluation.

Training Data	Measure	SVM	Naïve Bayes	ANN
Similes	Precision	1.00	0.59	**1.00**
	Recall	0.21	0.36	**0.60**
	F-score	0.35	0.44	**0.75**
	Accuracy	0.78	0.75	**0.78**
Metaphors	Precision	0.86	0.60	**1.00**
	Recall	0.21	0.21	0.21
	F-score	0.34	0.32	**0.35**
	Accuracy	**0.78**	0.74	**0.78**
Similes & metaphors	Precision	0.24	**0.28**	0.25
	Recall	0.50	**0.57**	**0.57**
	F-score	0.33	**0.38**	0.35
	Accuracy	**0.42**	0.38	0.40

Table 2. Results for the non native-speaker evaluation.

Training Data	Measure	SVM	Naïve Bayes	ANN
Similes	Precision	**0.83**	0.59	**0.83**
	Recall	0.07	0.15	0.08
	F-score	0.14	**0.24**	0.14
	Accuracy	**0.39**	0.38	**0.39**
Metaphors	Precision	**0.86**	0.60	0.83
	Recall	**0.09**	**0.09**	0.07
	F-score	**0.17**	0.16	0.14
	Accuracy	0.4	0.37	**0.39**
Similes & metaphors	Precision	0.64	0.64	**0.66**
	Recall	0.57	**0.65**	**0.65**
	F-score	0.60	0.64	**0.65**
	Accuracy	0.51	0.53	**0.55**

– *jijō to musubitsuita*: 'circumstances-related' [lit. 'tied to circumstances'];
– *ki o momanakute wa naranai*: 'cannot help but to worry' [lit. 'cannot help but to rub one's *ki* (internal energy)'];
– *fuan na kimochi de matte iru uchi ni*: 'while waiting worried' [lit. 'while being inside of anxious mood'];
– *omoi mo tsukanai koto*: '(sth) unexpected' [lit. 'thing that even a thought doesn't stick to'];
– *dokki o nukarete*: 'getting dumbfounded', 'getting disarmed' [lit. 'getting one's malice taken away from'];
– *uwasa wa uwasa o umu*: 'rumor feeds upon rumor' [lit. 'rumor gives birth to rumor'].

Element of subjectivity plays an important role during classification of such expressions; it is therefore not surprising that number of times (40 out of 100) they have received different labels depending on the annotator. As Tsvetkov and colleagues point out: "humans may disagree whether a particular expression is used metaphorically or not" [14].

5.1 Problems with Definitions

Largely thanks to the popularity gained throughout the years by the cognitive linguistics, countless publications regarding metaphor have been published. Nevertheless, one might get an impression that none of the metaphor's definitions coined so far can be called "complete" or comprehensive enough in order to cover all of the metaphor's usage instances in any kind of discourse.

The level of subjectivity involved in understanding of the notion is enormous. Certain expression called by one metaphoric might be as well categorized differently by someone else and often it is by no means a consequence of lacking a scholarship. Among numerous cases in which identifying figurativeness gets difficult, one should point at conventional metaphors.

Some of the most seasoned scholars out there assume that "so-called dead metaphor is not a metaphor at all, but merely an expression that no longer has a pregnant metaphorical use" [15]. On the other hand, others would treat it as the most prototypical, the "best" of its kind. For example, Kövecses [10] thinks that "they are *alive* in the most important sense – they govern our thought: they are *metaphors we live by*". It is thus impossible to annotate given data in a way, that virtually no one could rise an objection to.

Also, the boundaries between metaphor-related notions are typically vague, therefore to distinguish metaphor from idiom, metonymy from synecdoche, dead metaphor from living metaphor is no easy task. Should one consider *to understand*, *to put (sth) straight* or Japanese *ki wo tsukeru* 'to watch out', 'to be careful' figurative or literal? There is probably no easy answer.

5.2 Influence of Annotators' Background

Another interesting problem regarding classification of "dead" and "living" metaphors comes to light, when one tries to analyze foreign language acquisition process. It is well-known that in any given language figurative expressions are pervasive. As their usage is frequent not only in stereotypically "metaphorical" genres as poetry, but also in an every-day conversation, it is often the case that native speakers do not recognize the presence of a "frozen" metaphor in a given utterance. On the other hand, it becomes relatively easily discernible for a foreigner, in whose native language the same or nearly the same meaning would be expressed differently. One might say that to a foreigner the "dead" metaphor becomes once again "living". Because of this impression of novelty and "freshness", it is not rare for foreign students to identify more of actually existing metaphors than in the native speakers' case. This problem is multi-dimensional and fairly complicated, but in our opinion worth further studies.

Classification of compound-verbs like *mezameru* 'to wake up', *tabedasu* 'to start to eat' or *yarinaosu* 'to do (sth) over' also may be confusing. As these compounds' constituent elements mean something else in isolation, should one consider the foundation underlying their bindings as metaphorical in nature? As for the examples from English language, we can point at above-mentioned *to understand*.

How an annotator should treat proverbs, parables, allegories and so on seems like yet another big problem. These rhetorical devices are often – if not always – used metaphorically, although they differ from short "creative" metaphors.

5.3 Context-dependency

It can be easily noticed that an interpretation of whether a certain expression is used figuratively or not, strongly depends on given context. For example, *you're killing me* cannot be classified without knowing, in what kind of situation these words have been uttered.

Often may be this interpretation affected also by a specific cultural background, certain system of values differing between nations, social groups and so on. As Dobrzyńska claims: "The sets of associations fixed in the consciousness of native speakers of a given language, with all their different degrees of generality – varying from merely individual, through fairly common and stereotyped in a social group, to those shared by all speakers of the language – make metaphorical communication always extremely sensitive to the communicational context" [9]. We realize that vector-based classifiers still treat "context" quite shallowly especially in cases like ours, where the dictionary entries consist of sentences taken from literature separately (without neighboring sentences) due to the copyright restrictions. However, we believe that when our classifier is improved, it will help collecting figurative speech examples accompanied by the broader context. Eventually it should lead to creating the corpus and aiding metaphor processing in general.

6 Conclusions and Future Work

Figurative language's analysis from NLP perspective should be definitely considered a vital issue. As Veale and colleagues discuss in their monograph, "metaphor processing would not only help human users to be more systematic, comprehensive, and efficient in obtaining data to satisfy their information needs, but would also help to augment the creative reach and expressiveness of the user" [16]. As for the possible applications for an efficient metaphor-detecting or – even better – metaphor-generating system, one can think of not only typically NLP-related fields like sentiment analysis or dialogue systems but also completely independent domains such as psychotherapy or public relations.

Due to the phenomenon's level of complexity it is most probably impossible to create an ideal metaphor-recognizing model for Japanese language, however, we believe that improving efficiency of metaphor recognition can help with collecting more examples of Japanese figurative speech necessary for utilizing generative models. In this paper work we have presented the results of experiments conducted on texts not distinctively metaphorical and showed how different types of data and evaluation influence the outcome. In some cases the accuracy was similar to or even higher that of the state-of-the-art methods; we therefore plan to extend our experiments by increasing the number of evaluators and testing more recent texts, e.g. blogs in the near future.

7 Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 17K00295.

References

1. Aozora bunko digital library, `https://www.aozora.gr.jp`
2. Japanese local assembly minutes corpus project, `http://local-politics.jp`
3. Livedoor news, `http://news.livedoor.com`
4. Scikit-learn machine learning in python, `https://scikit-learn.org/stable/`
5. Wikpedia dumps list, `https://dumps.wikimedia.org/jawiki/latest`
6. Babieno, M., Rzepka, R., Takishita, S., Araki, K.: Utilizing figurative language examples for recognizing japanese sentences with metaphors. Tech. rep., 2018 Joint Convention Record, the Hokkaido Chapters of the Institutes of Electrical and Information Engineers, paper 134 (2018)
7. Bulat, L., Clark, S., Shutova, E.: Modelling metaphor with attribute-based semantics. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers. vol. 2, pp. 523–528 (2017)
8. Cameron, L.: Metaphor in educational discourse. A&C Black (2003)
9. Dobrzynska, T.: Translating metaphor: Problems of meaning. Journal of pragmatics 24(6), 595–604 (1995)
10. Kovecses, Z.: Metaphor: A practical introduction. Oxford University Press (2010)
11. Lakoff, G., Johnson, M.: Metaphors we live by. University of Chicago press (2008)
12. Onai, H.: Great Dictionary of 33800 Japanese Metaphors and Synonyms (in Japanese). Kodansha (2005)
13. Shutova, E., Teufel, S.: Metaphor corpus annotated for source-target domain mappings. In: LREC. vol. 2, pp. 2–2 (2010)

14. Tsvetkov, Y., Boytsov, L., Gershman, A., Nyberg, E., Dyer, C.: Metaphor detection with cross-lingual model transfer. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). vol. 1, pp. 248–258 (2014)
15. Utsumi, A.: Interpretive diversity explains metaphor–simile distinction. Metaphor and Symbol 22(4), 291–312 (2007)
16. Veale, T., Shutova, E., Klebanov, B.B.: Metaphor: A computational perspective. Synthesis Lectures on Human Language Technologies 9(1), 1–160 (2016)