

Evaluation Metrics for Automatically Generated Metaphorical Expressions

Akira Miyazawa and Yusuke Miyao

The Graduate University for Advanced Studies / National Institute of Informatics

INTRODUCTION

Automatic generation of metaphors helps us write novels, poems, etc. Existing systems have focused on only **similes** such as “*T* like *S*” (Kitada and Hagiwara, 2001; Abe et al., 2006).

We propose **metaphoricity**, **novelty**, **comprehensibility**, and **overall evaluation** to get “good metaphors”.

EXPERIMENTAL SETTINGS

By **crowdsourcing**, we got 10 scores for each expression and metric. The scores were evaluated on a **five-point scale**.

次の表現を比喩だと感じますか。5段階で評価してください。

「愛が溢れる(あふれる)」

5: 比喩だと感じる

4

3

2

1: 比喩だとは感じない

選択数は1〜5で段階的に変化するものとします。

Do you feel that the expression is metaphorical?

ai ga ahureru (love overflows)

5. It seems to be metaphorical.

1. It doesn’t seem to be metaphorical.

TARGETS

We made 1,360 Japanese expressions by combining 40 nouns and 34 verbal phrases following the method of Nabeshima (2011).

Example

ai (love) + *X ga ahureru* (*X* overflows) → *ai ga ahureru* (love overflows)

Nouns

◆ FLUID AND SOLID

mizu (water) 💧, *sun*a (sand) 🏖️, *iwa* (rock) 🏔️, etc.

◆ EMOTION

kimoti (feeling) 😊, *ai* (love) 🍷, *tanosisa* (enjoyment) 😄, *kyouhu* (fear) 😱, *ikari* (anger) 😡, *kanasimi* (sorrow) 😞, etc.

◆ IDEA

rikai (understanding) 👍, *jouhou* (information) 📄, etc.

◆ OTHERS

kinsen (money) 💰, *jikan* (time) 🕒, *neko* (cat) 🐱, etc.

Verbal Phrases

◆ PHYSICAL ACTION RELATED TO WATER

X ga ahureru (*X* overflows), *X wo nomu* (somebody drinks *X*), etc.

METRICS

Metaphoricity

Metaphoricity measures how metaphorical an expression is.

| Rank | Noun (X) | Verbal phrase | Score |
|------|--------------------------|------------------------------------------------------------|-------|
| 1 | <i>kotoba</i> (word) | <i>X ga huttou-suru</i> (<i>X</i> boils) | 3.9 |
| 2 | <i>kanjou</i> (emotion) | <i>X ni oboreru</i> (<i>Y</i> almost drowns in <i>X</i>) | 3.8 |
| 3 | <i>zetubou</i> (despair) | <i>X ga ahureru</i> (<i>X</i> overflows) | 3.7 |
| | | ⋮ | |
| 1356 | <i>mizu</i> (water) | <i>X ga huttou-suru</i> (<i>X</i> boils) | 0.0 |
| 1356 | <i>mizu</i> (water) | <i>X ga nagareru</i> (<i>X</i> flows) | 0.0 |
| 1356 | <i>mizu</i> (water) | <i>X wo nomu</i> (<i>Y</i> drinks <i>X</i>) | 0.0 |

High-ranked expressions often use nouns related to emotion such as *zetubou* (despair).

Novelty

Novelty measures how novel an expression looks or sounds.

| Rank | Noun (X) | Verbal phrase | Score |
|------|---------------------|-------------------------------------------------------|-------|
| 1 | <i>ari</i> (ant) | <i>X ga simiru</i> (<i>X</i> soaks into <i>Y</i>) | 4.0 |
| 1 | <i>neko</i> (cat) | <i>X wo siboridasu</i> (<i>Y</i> squeezes <i>X</i>) | 4.0 |
| 1 | <i>neko</i> (cat) | <i>X ga nagaderu</i> (<i>X</i> flows out) | 4.0 |
| | | ⋮ | |
| 1353 | <i>mizu</i> (water) | <i>X ga nagareru</i> (<i>X</i> flows) | 0.0 |
| 1353 | <i>mizu</i> (water) | <i>X ga waku</i> (<i>X</i> gushes out) | 0.0 |
| 1353 | <i>mizu</i> (water) | <i>X wo nomu</i> (<i>Y</i> drinks <i>X</i>) | 0.0 |

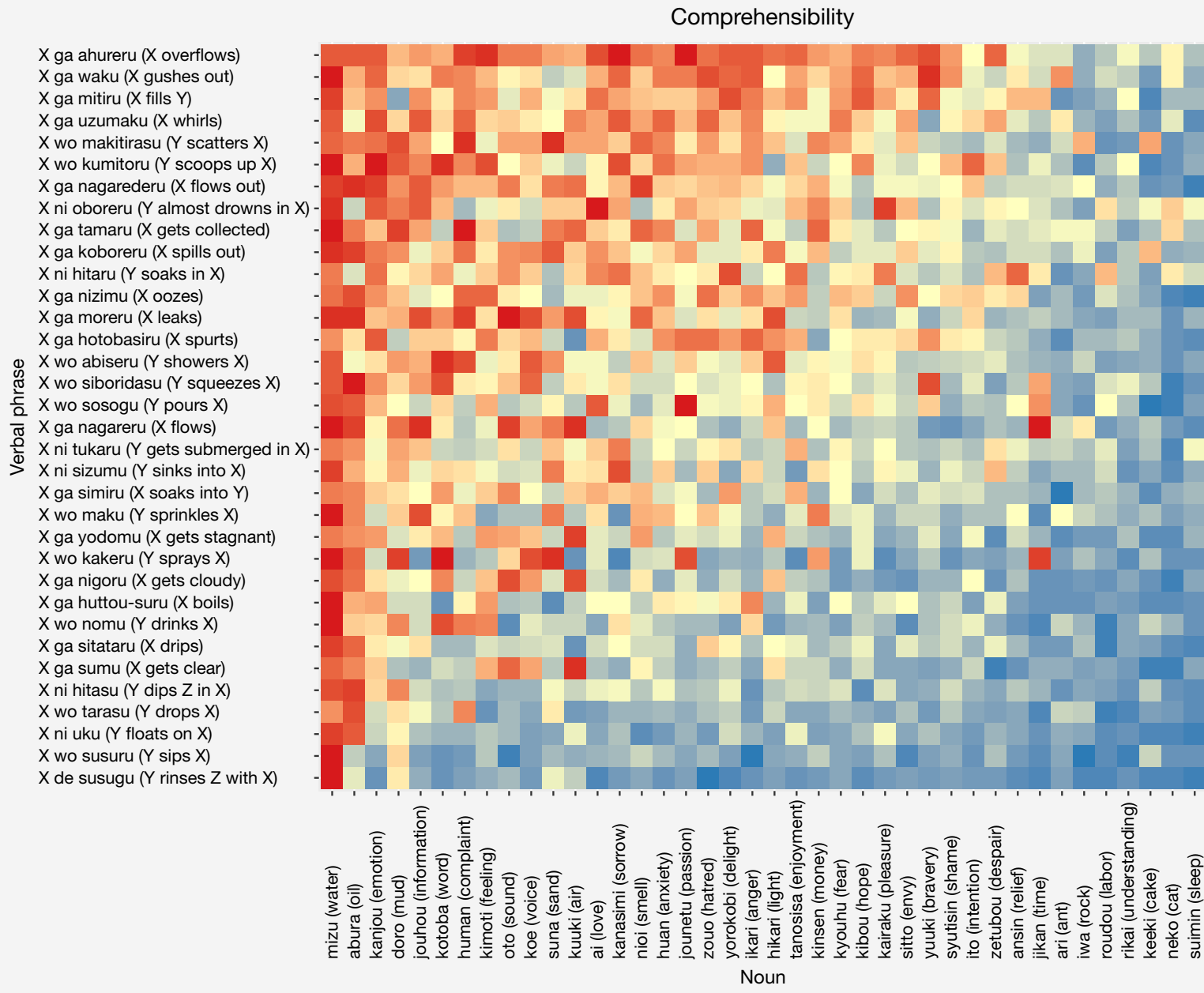
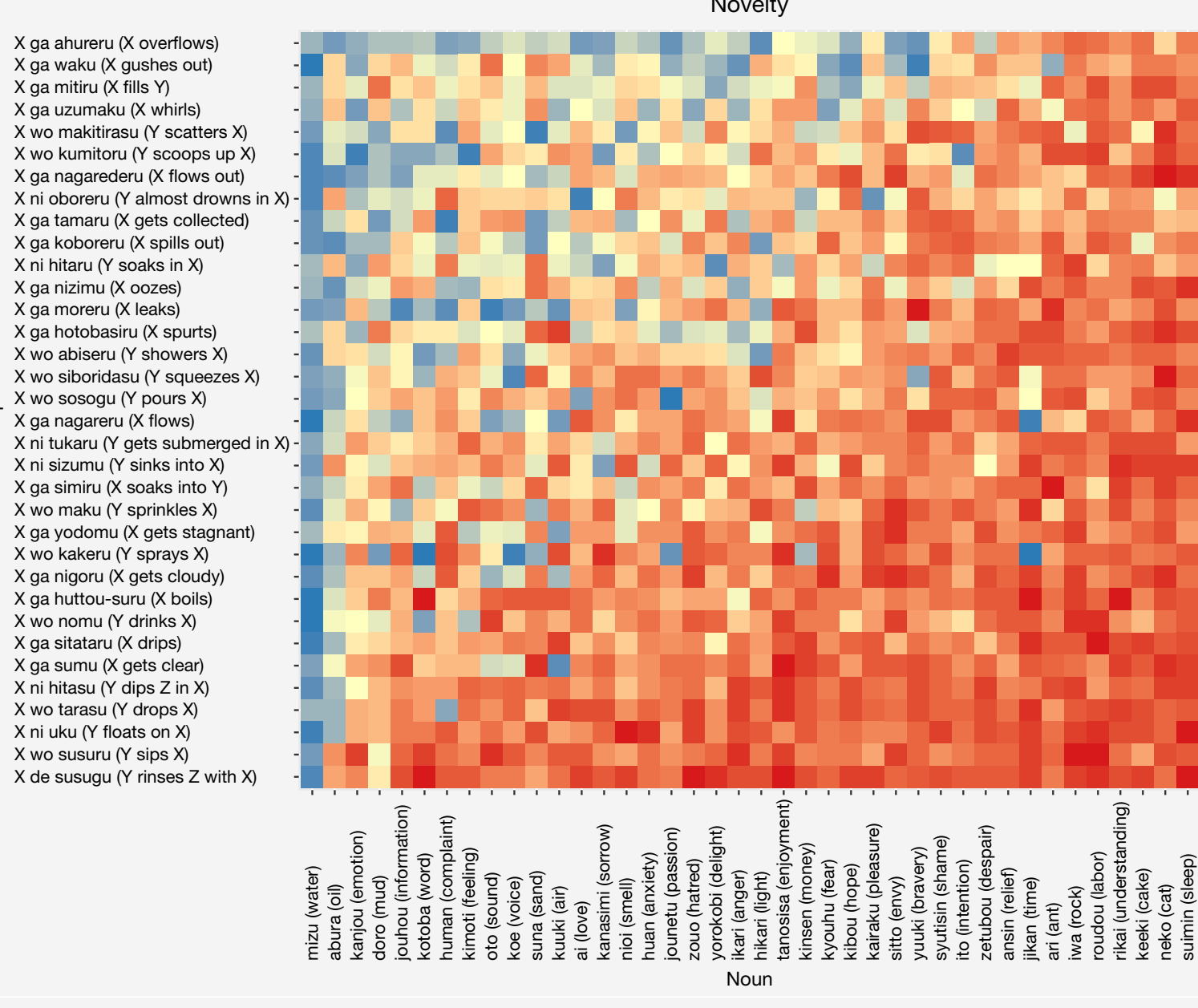
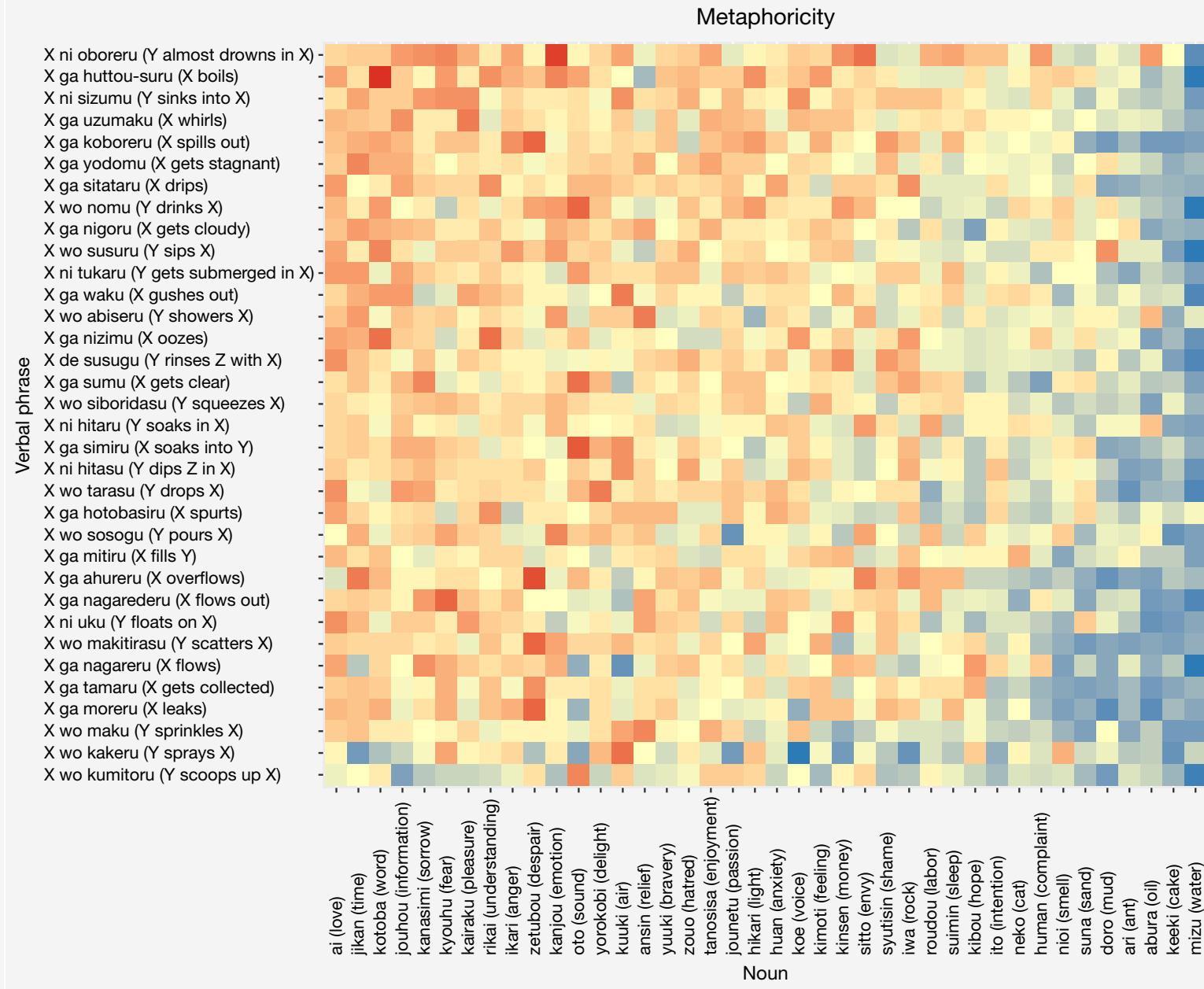
Expressions that use nouns for concrete objects were ranked high.

Comprehensibility

Comprehensibility measures how easy it is to understand the meaning of expressions.

| Rank | Noun (X) | Verbal phrase | Score |
|------|-------------------------|---------------------------------------------------------------|-------|
| 1 | <i>ai</i> (love) | <i>X ni oboreru</i> (<i>Y</i> almost drowns in <i>X</i>) | 4.0 |
| 1 | <i>kanjou</i> (emotion) | <i>X wo kumitoru</i> (<i>Y</i> scoops up <i>X</i>) | 4.0 |
| 1 | <i>mizu</i> (water) | <i>X de susugu</i> (<i>Y</i> rinses <i>Z</i> with <i>X</i>) | 4.0 |
| | | ⋮ | |
| 1356 | <i>ari</i> (ant) | <i>X ga simiru</i> (<i>X</i> soaks into <i>Y</i>) | 0.0 |
| 1356 | <i>iwa</i> (rock) | <i>X wo susuru</i> (<i>Y</i> sips <i>X</i>) | 0.0 |
| 1356 | <i>zouo</i> (hatred) | <i>X de susugu</i> (<i>Y</i> rinses <i>Z</i> with <i>X</i>) | 0.0 |

This metric shows the opposite tendency to the novelty. The corr. coefficient was −0.92.



OVERALL EVALUATION

Overall Evaluation

Here we defined it as the average of the three metrics.

Goodness

A volunteer chose one from each pair of high- and low-ranked expressions that made him more inclined to use (marked red in the table).

| High-ranked expression [rank in overall eval.] | Low-ranked expression [rank in overall eval.] |
|------------------------------------------------------------------------|------------------------------------------------------------------------|
| <i>human wo nomu</i> (<i>Y</i> drinks complaints) [23] | <i>abura wo kumitoru</i> (<i>Y</i> scoops up oil) [1087] |
| <i>ikari ga koboreru</i> (anger spills out) [6] | <i>iwa ni oboreru</i> (<i>Y</i> almost drowns in a rock) [1117] |
| <i>syuutisin ga tamaru</i> (shame gets collected) [44] | <i>syuutisin wo sosogu</i> (<i>Y</i> pours shame) [856] |
| <i>jouhou ga nigoru</i> (information gets cloudy) [106] | <i>kuuki wo makitirasu</i> (<i>Y</i> scatters the air) [212] |
| <i>kanasimi ga simiru</i> (sorrow soaks into <i>Y</i>) [32] | <i>rikai ga nagareru</i> (understanding flows) [721] |
| <i>tanosisa ga uzumaku</i> (enjoyment whirls) [81] | <i>human ni tukaru</i> (<i>Y</i> gets submerged in complaints) [1241] |
| <i>kotoba ga nizimu</i> (words ooze) [14] | <i>kyouhu ga nagareru</i> (fear flows) [307] |
| <i>kanjou wo sosogu</i> (<i>Y</i> pours emotion) [44] | <i>ito X ni tukaru</i> (<i>Y</i> gets submerged in intention) [654] |
| <i>huan ga nagaderu</i> (anxiety flows out) [44] | <i>jounetu wo kumitoru</i> (<i>Y</i> scoops up passion) [165] |
| <i>jouhou ni oboreru</i> (<i>Y</i> almost drowns in information) [23] | <i>abura ga tamaru</i> (oil gets collected) [1241] |

Metaphoricity

The author judged 8 of the 10 expressions as metaphorical by using MIPVU (Steen et al., 2010).

CONCLUSION

- ◆ We proposed metrics to evaluate automatically generated metaphors.
- ◆ We actually evaluated expressions by crowdsourcing.
- ◆ The result shows the validity of the metrics and their relationship.
- ◆ High-ranked expressions in the overall eval. are good metaphors.

The results are available at <https://github.com/pecorarista/metaphor-evaluation-result>.

REFERENCE

- Abe, K., K. Sakamoto, and M. Nakagawa (2006). “A computational model of metaphor generation process”. In: *Proceedings of the 28th Annual Meeting of the Cognitive Science Society*, pp. 937–942.
- Kitada, J. and M. Hagiwara (2001). “Figurative composition support system using electronic dictionaries (in Japanese)”. In: *Transactions of Information Processing Society of Japan* 42.5, pp. 1232–1241.
- Nabeshima, K. (2011). *Nihongo no metafā (in Japanese)*. Kurosio Publishers.
- Steen, G. J. et al. (2010). *A Method for Linguistic Metaphor Identification: From MIP to MIPVU*. John Benjamins Publishing.