

Exploring metaphorical expressions in Japanese newspaper-article corpora

Sachi Kato, Masayuki Asahara
National Institute for Japanese Language and Linguistics
yasuda-s@ninjal.ac.jp

Keywords: metaphorical expression, corpus, newspaper, semantic tagged corpus

We constructed a metaphorical expression database from the Balanced Corpus of Contemporary Written Japanese (hereafter, BCCWJ), comprising newspaper samples, and explored the metaphorical expression usage in the databases. The data include the article categories following Kinokuniya (2000), such as politics, economy, and culture. A subset of the data (117,543 words) consists of annotated semantic labels from the "Word list by semantic principles," hereafter WLSP (Kato et al., 2018). We exhaustively annotated about 3,000 metaphorical expressions in the subset of the data, using human judges (following MIP). The metaphorical expressions are subcategorized into "metaphor" (61.7%), "metonymy (including synecdoche)" (27.0%), "idiomatic" (10.2%) and "others" (0.5%).

Previous work constructing a metaphorical expression example database is limited to literary works (Nakamura, 1977; Nakamura, 1995; etc.). Moreover, such corpora are not exhaustive and instead pick examples at random. The data cannot be used for statistical analysis, as could exhaustive annotation data in a balanced corpus. Our work is the first comprehensive annotated database of metaphorical expression in the balanced Japanese corpus. We performed several analyses of the corpus.

First, we evaluated the distribution by metaphorical expression types. Metaphors were subcategorized with or without cue phrases. We also subcategorized metaphors by the source domain (SD) and target domain (TD) of semantic labels. The distribution shows that the prosopopoeia is 5.5% (e.g., SD: object [inanimate], TD: animal [animate]); and objectification is 34.0% (e.g., SD: concept [abstract], TD: object [concrete] is 33.1%, SD: person [animate], TD: object [inanimate] is 0.9%). In contrast to the results obtained by classifying the data of Nakamura (1977) in the same way, the percentage of prosopopoeia is exceedingly low and the objectification is higher.

Second, we explored the distribution by the article categories. The BCCWJ newspaper samples are also annotated with article categories. The sports (16.7%), international (8.0%), economy (18.2%), politics (20.2%), and culture (17.1%) articles contain the largest proportions of the corpus' total number of metaphorical expressions. Furthermore, metonymy appears frequently in international articles (49.1%).

Third, we explored the distribution of connective patterns. The semantic categories of elements as contextual binding were investigated in recent corpus-based metaphor research. For example, metaphors in the corpus of "fighting against disease" speeches tend to fall in the semantic categories of war and travel (Semino et al., 2015). In this way, contextual heterogeneous semantic elements are useful for identifying metaphors. We analyzed semantic categories such as "architecture" in articles of economy.

Our future work will extend the genres to books and magazines. We expect additional analysis of collected examples to reveal central features of Japanese metaphorical expressions.

References

- Kato, Sachi, Masayuki Asahara, Makoto Yamazaki. (2018). Annotation of "Word List by Semantic Principles" Labels for the Balanced Corpus of Contemporary Written Japanese. PACLIC-32.
- Kinokuniya Shoten. (2000). *CD-HIASK (Asahi Newspaper Article Database) Manual*. (in Japanese)
- Nakamura, Akira. (1977). *Theory and Category of Metaphor Expression*. Shuei Shuppan. (in Japanese)
- Nakamura, Akira. (1995). *Japanese Dictionary of Metaphorical Expressions*. Kadokawa Shoten. (in Japanese)
- Pragglejaz Group. (2007). MIP: A method for identifying metaphorically used words in discourse. *Metaphor and Symbol* 22(1): 1-39.
- Semino, Elena, Zsófia Demjén, Jane Demmen, Veronika Koller, Sheila Payne, Andrew Hardie, Paul Rayson. (2015). The online use of violence and journey metaphors by patients with cancer, as compared with health professionals: a mixed methods study. *BMJ Supportive & Palliative Care* 2015(0): 1-7.