Extraction of Japanese Conceptual Metaphors Using an LLM

Naranbuuvei Ganbat[1] Rowan Hall Maudslay[2] Kanako Komiya[1]
Masayuki Asahara[3] Sachi Kato[4] Ryoko Uno[1]
[1]Tokyo University of Agriculture and Technology [2]Université Paris Sciences et Lettres (PSL)
[3]National Institute for Japanese Language and Linguistics [4]Hokkaido University
s248894ra@st.go.tuat.ac.jp, r.hallmaudslay@gmail.com, kkomiya@go.tuat.ac.jp,
masayu-a@ninjal.ac.jp, katosachi@let.hokudai.ac.jp, ryokouno@cc.tuat.ac.jp

Abstract

This study investigates whether typical Japanese conceptual metaphors can be derived from the BCCWJ-Metaphor corpus using a large language model (LLM). We automatically extract source and target domains from metaphorical constructions and compare two approaches: (i) providing an LLM with a predefined domain inventory and (ii) inferring conceptual metaphors by leveraging higher-level semantic categories. Based on evaluations by native speakers, we find that the former approach achieves higher recall but lower precision, while the latter yields higher precision with lower recall. In particular, the first approach enables the extraction of 72 or 76 conceptual metaphors, demonstrating its effectiveness in broad coverage of Japanese conceptual metaphor usage.

1 Introduction

According to Conceptual Metaphor Theory [1], abstract concepts are understood through more concrete ones, reflecting a fundamental mechanism of human cognition. For example, in Japanese, the conceptual metaphor 感情は水 (EMOTION IS WATER) is commonly used, seen in expressions such as 気分にひたる (to be immersed in a feeling, lit. to be soaked in a feeling) and 愛着が沸く (to develop an attachment, lit. attachment boils up). In this example, EMOTION is the target domain, which is the abstract concept we try to understand, and WATER is the source domain used to describe EMOTION.

Previous studies on conceptual metaphors in Japanese have mainly focused on analyzing metaphors related to a specific target domain [2, 3, 4, 5]. These studies typically examine how a target domain is structured across multiple source domains, often through qualitative analyses of selected examples. However, it remains unclear to what extent these conceptual metaphors are realized in large-scale corpora.

To address this gap, we propose an automatic method for extracting conceptual metaphors in Japanese using a large language model (LLM). Kato et al. [6] annotated metaphorical expressions in the Balanced Corpus of Contemporary Written Japanese (BCCWJ) [7], creating the BCCWJ-Metaphor corpus, which provides explicit information on metaphors in Japanese. Using the BCCWJ-Metaphor corpus, we investigate how many typical Japanese conceptual metaphors can be recovered from metaphorical expressions. Specifically, we focus on the inventory of Japanese conceptual metaphors proposed by Nabeshima [8] and use it as a reference list. This enables us to examine to which extent these theoretically proposed conceptual metaphors appear in corpus data and whether their source and target domains can be systematically identified. In particular, we focus on noun–verb, noun–adjective, and noun–noun constructions, and propose a method for automatically inferring source and target domains from these expressions. Native speakers then evaluate whether the conceptual metaphors we derive from these metaphorical expressions are plausible.

2 Related Work

Several studies have examined specific target domains in Japanese conceptual metaphors. Oishi [2] conducted a comparative study of 感情 (EMOTION) metaphors in Japanese and English, proposing a classification aimed at cross-linguistic comparison. Other work has also focused on a single target domain in Japanese, including 情報 (INFORMATION) [3], 心 (MIND/HEART) [4], and 時間 (TIME) [5], using corpus-based, cultural, or structural

This work is published without peer review and is licensed by the author(s) under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

analyses. From a computational perspective, Zhu et al. [9] proposed a system for detecting metaphorical expressions using the BCCWJ-Metaphor corpus.

Although LLMs have been applied to extracting source and target domains of conceptual metaphors, most prior work has focused on English, with little attention to Japanese. Recent studies have explored LLM-based metaphor processing in various settings. For example, Boisson et al. [10] examined metaphor extraction and conceptual mapping in literary texts, while Wachowiak et al. [11] evaluated GPT-3's ability to detect metaphorical expressions and predict source domains without predefined domain lists. In addition to these studies focusing on metaphor identification, other work has addressed related tasks involving source–target domain modeling, including domain based metaphor generation [12] and word-pair based metaphor detection [13]. More recently, Zheng et al. [14] proposed an emotion-guided LLM approach to metaphor detection that infers source and target domains through emotion-aware domain mapping.

3 Data

3.1 BCCWJ-Metaphor

In this study, we used the BCCWJ-Metaphor corpus released by Kato et al. [6]. In addition to the annotation of metaphors, the corpus also includes metaphorical constructions and metaphor types. Figure 1 shows an example from the BCCWJ-Metaphor corpus.

短単位語彙素 (Lemma)	短単位書字形 (Surface form)	比喩BIO (Metaphor BIO)	結合 (Construction)
想像	想像	O	
為る	する	O	
だけ	だけ	O	
だ	で	O	
、	、	O	
幸福	幸福	O	
だ	な	O	
気分 (feeling)	**気分**	**B**	
に	**に**	**I**	
浸る (to be soaked)	**ひたる**	**I**	**気分にひたる to be immersed in a feeling**

Figure 1 Example of BCCWJ-Metaphor

In the corpus, a BIO label is assigned to every short-unit word, indicating whether it is the beginning of a metaphorical expression (B), inside a metaphorical expression (I), or outside any metaphorical expression (O). The metaphorical expression in this example is 気分に浸る (to be immersed in a feeling). The noun 気分 (feeling) is labeled B, marking the beginning of the metaphorical span. The following particle に and the verb 浸る (to be soaked/immersed) are labeled I, indicating that they belong to the same metaphorical expression. A continuous sequence of B and I labels constitutes a single metaphorical expression span. In this study, we emphasize the construction (結合) corresponding to each span, as it represents the core of the metaphorical expression. Japanese constructions resemble frames in English, but often involve noun–verb or noun–noun word pairs rather than being verb centered. Specifically, we focus on constructions consisting of noun–verb, noun–adjective, and noun–noun combinations. Out of the total 11,364 constructions in the corpus, 8,411 constructions fell within the scope of our target construction types.[1)] Of these target constructions, 3,739 were noun–verb constructions, 415 were noun–adjective constructions, and 4,257 were noun–noun constructions.

3.2 Reference Domain List

We used the conceptual metaphors presented in the book by Nabeshima [8], 日本語のメタファー (Japanese Metaphor), as reference conceptual metaphors. The book describes 255 conceptual metaphors, and this list serves as the basis for identifying the underlying conceptual domains involved in the metaphorical constructions. Based on the conceptual metaphors, we constructed two domain inventories: a list of 111 unique target domains and a list of 155 source domains.

4 Experiments

We conducted experiments with two methods using the OpenAI API with GPT-4o[2)] model to extract conceptual metaphors from Japanese metaphorical expressions. Experiments aim to infer source and target domains of the conceptual metaphors from metaphorical construction in BCCWJ-Metaphor and to compare the results with the reference conceptual metaphors in the predefined list. Figure 2 illustrates the overall workflow of the two methods conducted in this study. The two methods differ in how

1) Constructions that were not included in the experiments consisted of those containing, for example, two verbs or other combinations outside our scope.

2) https://platform.openai.com/docs/models/gpt-4o

This work is published without peer review and is licensed by the author(s) under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

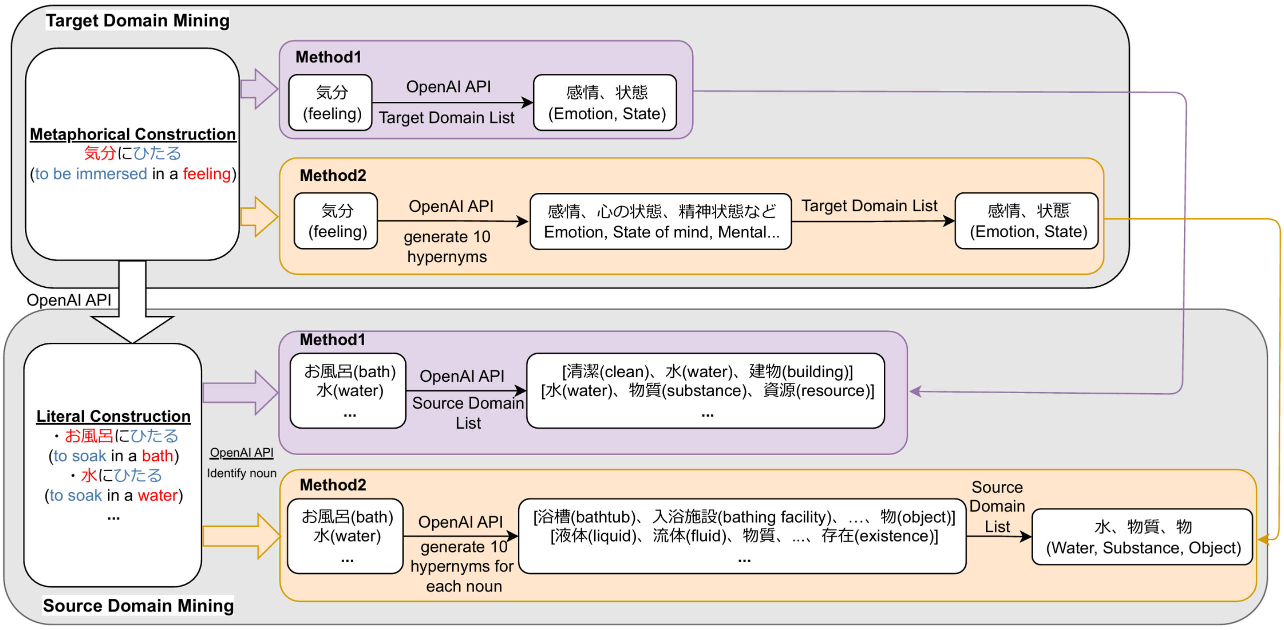

Figure 2 Workflow of Methods (*prompts used in noun-verb construction experiments are shown in Fig 3 and Fig 4)

candidate domains are identified.

- **Target Domain Mining** Target domain mining begins with metaphorical constructions in the corpus, for which part-of-speech information is used to identify the noun (first noun in noun–noun constructions). For example, in 気分に浸る ("to be immersed in a feeling"), the noun 気分 ("feeling") is extracted as input to the LLM. In method 1, an LLM directly selects applicable target domains from the target domain list based on the noun. For example, 気分 (feeling) is mapped to target domains such as Emotion and State. In method 2, an LLM first generates 10 hypernyms for the target noun, such as emotion, state of mind, and mental state. These hypernyms are then compared with the target domain list, and matching domains are selected as target domain candidates.
- **Source Domain Mining** Initially, an LLM is provided with the original sentence, the metaphorical span, the construction, and the part-of-speech of the construction. Based on this information, an LLM generates five literal constructions in which the verb or adjective (second noun in a noun-noun construction) is used in a literal sense, such as お風呂に浸る (to soak in a bath) and 水に浸る (to soak in water). Then, an LLM identifies the nouns in these literal constructions, such as お風呂 (bath) and 水 (water). In method 1, an LLM directly selects applicable source domains from the list of source domains. In method 2, an LLM generates ten hypernyms for each noun in the literal constructions. These hypernyms are then matched with the source domain list to identify source domain candidates.
- **Evaluation** In both approaches, extracted target–source domain pairs were compared with Nabeshima's conceptual metaphor inventory, and matching pairs were evaluated by two native speakers. Annotators assessed the compatibility between each conceptual metaphor and its construction using four labels: yes (clearly appropriate), maybe (possibly appropriate), miss (incorrect and how the error occurred is predictable), and no (clearly inappropriate). The union of matched pairs from both methods, which includes 2,116 unique pairs were annotated. Inter-annotator agreement was measured using Cohen's kappa score (N = 2,116, n = 2, k = 4). The kappa score was 0.36 with the four-label setting and increased to 0.51 when the labels were merged into a binary classification (yes/maybe as positive and miss/no as negative), indicating moderate agreement.

5 Results

Table 1 shows the precision results for the two methods. Precision is calculated as the proportion of conceptual metaphor–construction pairs annotated as yes or maybe.

This work is published without peer review and is licensed by the author(s) under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

A total of 1,758 conceptual metaphor–construction pairs were extracted in method 1, while 579 pairs were extracted in method 2. Overall, method 2 achieved higher precision than method 1, although it extracted fewer conceptual metaphor–construction pairs.

	Method 1	Method 2
Annotator 1	0.251 (442/1,758)	0.345 (200/579)
Annotator 2	0.237 (418/1,758)	0.24 (139/579)
Average	0.244	0.293

Table 1 Precision of Conceptual Metaphor-Construction pair

We also evaluated recall at the level of conceptual metaphor types, using Nabeshima's inventory of 255 conceptual metaphors as the reference set. Method 1 extracted 145 distinct conceptual metaphors, whereas method 2 extracted 45. Recall was defined as the proportion of conceptual metaphor types with at least one associated pair annotated as yes or maybe. Table 2 reports the recall values. The higher recall of method 1 suggests that directly mapping constructions to predefined domains enables broader coverage of conceptual metaphor types. Figure 5 presents the list of conceptual metaphors that were judged as correct (yes or maybe) by both annotators in method 1.

	Method 1	Method 2
Annotator 1	0.282 (72/255)	0.133 (34/255)
Annotator 2	0.298 (76/255)	0.125 (32/255)

Table 2 Recall of Conceptual Metaphor Type

6 Discussion

In this section, we discuss cases in which the method succeeded as well as cases in which it failed. When conceptual metaphors are formulated as "A is B", diverse possible linguistic realizations are reduced to a single expression. As a result, our system may inadvertently exclude other valid alternatives. A representative case in which method 1 succeeded but method 2 failed is 契約が切れる (a contract breaks), which formed the conceptual metaphor RELATIONSHIP IS A LINE. In method 1, literal constructions such as ロープが切れる (a rope breaks) and 糸が切れる (a thread breaks) enabled the model to identify source domain 線 (LINE) from rope and thread. In method 2, however, hypernyms generated for rope and thread did not include LINE. Instead, they consisted of more general categories such as material and object, which did not match the source domain list.

On the contrary, a case in which method 2 succeeded while method 1 failed is 話題をよぶ (to attract attention, lit. to call a topic), which formed the conceptual metaphor IDEA IS PERSON. In method 2, hypernym generation for the target noun 話題 (topic) produced アイデア (idea), enabling alignment with the target domain IDEA. However, in method 1, 話題 (topic) was mapped to domains like COMMUNICATION and LANGUAGE, which did not directly correspond to the predefined target domain list. This example demonstrates that the hypernym based method can facilitate the recovery of higher level conceptual metaphors by bridging lexical variation at the noun level.

Some failures reveal limitations of the literal construction based approach. For example, the construction 研究が一流 (the research is top-class) was incorrectly associated with the conceptual metaphor THEORY IS A BUILDING. In this case, literal constructions such as レストランが一流 (the restaurant is top-class) and 美術館が一流 (the museum is top-class) led the model to identify BUILDING as a source domain. However, the adjective 一流 (top-class) is widely used across diverse entity types, including people, organizations, and abstract activities, and primarily expresses evaluative quality rather than spatial or structural properties. As a result, inferring source domains from nouns in such cases leads to a semantic mismatch. This example shows that inferring source domains from literal constructions can place too much weight on the substituted nouns, resulting in incorrect source domain assignments.

7 Conclusion

This paper examined the automatic extraction of Japanese conceptual metaphors from the BCCWJ-Metaphor corpus using an LLM. By comparing a domain inventory based approach with a hypernym based inference approach, we showed that LLMs can recover a substantial portion of Japanese conceptual metaphors from corpus data, with a trade-off between recall and precision. In particular, domain inventory based approach extracted 72 or 76 conceptual metaphors, demonstrating broad coverage. At the same time, our analysis revealed systematic errors caused by a semantic shift. Generated literal constructions often led to inappropriate source domain assignments, especially for evaluative adjectives and idiomatic expressions, indicating directions for future improvement.

This work is published without peer review and is licensed by the author(s) under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

Acknowledgments

This work was supported by JSPS KAKENHI Grants JP22K12145, 25K15839 and JP25K00459, the JSPS Postdoctoral Fellowship for Research in Japan (for Foreign Researchers), the NINJAL Collaborative Research Project "Empirical Computational Psycholinguistics Using Annotated Data", and the Kayamori Foundation of Informational Science Advancement Research Grant "Extraction of Conceptual Metaphors Using Natural Language Processing". We also would like to thank Masaki Kato for annotating the data used in this study.

References

[1] George Lakoff and Mark Johnson. **Metaphors We Live By**. University of Chicago Press, Chicago, 1980.

[2] 大石亨. 日英対照研究に向けた感情のメタファーの分類. 言語処理学会第 13 回年次大会論文集, 2007.

[3] 李文鑫. <情報>を目的領域とする概念メタファー - メタファーの認定手順の提案とコーパスに基づく分析 -. 国際日本研究, Vol. 9, pp. 129–146, 02 2017.

[4] 松井真人. 日本文化における「心」の概念メタファー. 山形県立米沢女子短期大学紀要, Vol. 52, pp. 11–20, 12 2016.

[5] 益子望. 「時間が動く」・「時間を動く」：現代日本語における時間の概念メタファー構造. 藝文研究, Vol. 116, pp. 146 (119)–159 (106), 06 2019.

[6] 加藤祥, 菊地礼, 浅原正幸. 『現代日本語書き言葉均衡コーパス』に対する mip に基づく比喩表現情報の付与. 言語処理学会第 28 回年次大会発表論文集, 2022.

[7] Maekawa Kikuo, Makoto Yamazaki, Toshinobu Ogiso, Takehiko Maruyama, Hideki Ogura, Wakako Kashino, Hanae Koiso, Masaya Yamaguchi, Makiro Tanaka, and Yasuharu Den. Balanced corpus of contemporary written japanese. **Language Resources and Evaluation**, Vol. 48, pp. 345–371, 2014.

[8] 鍋島弘治朗. 日本語のメタファー. くろしお出版, 2011.

[9] Hang Zhu, Rowan Hall Maudslay, Kanako Komiya, Sachi Kato, and Masayuki Asahara. Large-scale japanese metaphor corpus construction: Expanding bccwj-metaphor with automated annotation. **Proceedings of PACLIC39: The 39th Pacific Asia Conference on Language, Information and Computation**, 2025.

[10] Joanne Boisson, Zara Siddique, Hsuvas Borkakoty, Dimosthenis Antypas, Luis Espinosa Anke, and Jose Camacho-Collados. Automatic extraction of metaphoric analogies from literary texts: Task formulation, dataset construction, and evaluation. In Owen Rambow, Leo Wanner, Marianna Apidianaki, Hend Al-Khalifa, Barbara Di Eugenio, and Steven Schockaert, editors, **Proceedings of the 31st International Conference on Computational Linguistics**, pp. 6692–6704, Abu Dhabi, UAE, January 2025. Association for Computational Linguistics.

[11] Lennart Wachowiak and Dagmar Gromann. Does GPT-3 grasp metaphors? identifying metaphor mappings with generative language models. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki, editors, **Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 1018–1032, Toronto, Canada, July 2023. Association for Computational Linguistics.

[12] Kevin Stowe, Tuhin Chakrabarty, Nanyun Peng, Smaranda Muresan, and Iryna Gurevych. Metaphor generation with conceptual mappings. In Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli, editors, **Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)**, pp. 6724–6736, Online, August 2021. Association for Computational Linguistics.

[13] Yuan Tian, Ruike Zhang, Nan Xu, and Wenji Mao. Bridging word-pair and token-level metaphor detection with explainable domain mining. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar, editors, **Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 13311–13325, Bangkok, Thailand, August 2024. Association for Computational Linguistics.

[14] Li Zheng, Sihang Wang, Hao Fei, Zuquan Peng, Fei Li, Jianming Fu, Chong Teng, and Donghong Ji. Enhancing hyperbole and metaphor detection with their bidirectional dynamic interaction and emotion knowledge. In Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, editors, **Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 489–499, Vienna, Austria, July 2025. Association for Computational Linguistics.

This work is published without peer review and is licensed by the author(s) under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

prompt1

You are given a sentence containing a metaphorical expression, the metaphorical span, the construction that forms the metaphor, and the part-of-speech information for the construction.

Sentence: **{sentence}**
Metaphorical expression: **{span}**
Construction: **{construction}**
Part-of-speech: **{construction_pos}**

List five constructions in which the verb **{verb}** in the construction is used literally, not metaphorically. For each literal construction, change only the noun, and keep particles and other words the same as in the given construction.

Follow the output format below:
1. construction1
2. construction2
3. construction3
4. construction4
5. construction5

prompt2

Please identify the nouns contained in the construction of the above output.

Follow the output format below:

1. noun1
2. noun2
3. noun3
4. noun4
5. noun5

prompt3

For each of the following nouns, answer all applicable items from the "conceptual domain list".

Follow the rules below:
- You must choose only from the candidate domains
- Do not write any explanations or extra text
- Never choose domains that are unrelated to the noun
- If there is no applicable domain, leave the part after the colon empty

Nouns: [**{noun1}**, **{noun2}**, **{noun3}**,**{noun4}**,**{noun5}**]

Conceptual domain list: **{source_domain_list}**

Follow the output format below:
noun: domain1, domain2,...

Method1

For each noun, list 10 hypernyms.

Follow the rules below:
- The level of abstraction of the hypernyms should increase step by step, similar to WordNet (Example: green apple → apple → fruit → food)

Nouns: [**{noun1}**, **{noun2}**, **{noun3}**,**{noun4}**,**{noun5}**]

Follow the output format below:
1. hypernym1, hypernym2, ... , hypernym10
2. hypernym1, hypernym2, ... , hypernym10
3. hypernym1, hypernym2, ... , hypernym10
4. hypernym1, hypernym2, ... , hypernym10
5. hypernym1, hypernym2, ... , hypernym10

Method2

Figure 3 Translated prompts of source domain mining

For the given noun, answer all applicable items from the "conceptual domain list".

Follow the rules below:
- You must choose only from the candidate domains
- Do not write any explanations or extra text
- Never choose domains that are unrelated to the noun
- If there is no applicable domain, leave the part after the colon empty

Noun: **{noun}**

Conceptual domain list: **{target_domain_list}**

Follow the output format below:
noun: domain1, domain2,...

Method1

For the given noun, list 10 hypernyms.

Follow the rules below:
- The level of abstraction of the hypernyms should increase step by step, similar to WordNet (Example: green apple → apple → fruit → food)

Noun: **{noun}**

Follow the output format below:
1. hypernym1
2. hypernym2
...
3. hypernym10

Method2

Figure 4 Translated prompts of target domain mining

アイデアは人 (IDEA IS PERSON)	困難は障害物 (DIFFICULTY IS OBSTACLE)	感情は生物 (EMOTION IS LIVING THING)	組織は生命体 (ORGANIZATION IS LIVING
アイデアは物 (IDEA IS OBJECT)	変化は移動 (CHANGE IS MOVEMENT)	感情は膨らむもの (EMOTION IS SOMETHING THAT SWELLS)	群集は水 (CROWD IS WATER)
コミュニケーションは導管 (COMMUNICATION IS PIPE)	大統領選は戦い (PRESIDENTIAL ELECTION IS BATTLE)	時間は価値ある商品 (TIME IS VALUABLE COMMODITY)	視野は容器 (VISION IS CONTAINER)
コミュニケーションは送ること (COMMUNICATION IS SENDING)	存在は見えること (EXISTENCE IS VISIBILITY)	時間は移動物 (TIME IS MOVING THING)	言葉は水 (WORD IS WATER)
マイナスの感情は下にある水 (NEGATIVE EMOTION IS WATER BELOW)	存続する組織は直立した物理的構造物 (PERSISTING ORGANIZATION IS UPRIGHT PHYSICAL STRUCTURE)	時間は金 (TIME IS MONEY)	議論は戦争 (ARGUMENT IS WAR)
人は植物 (PERSON IS PLANT)	希望は上の光 (HOPE IS LIGHT ABOVE)	時間は限りある資源 (TIME IS LIMITED RESOURCE)	身体は容器 (BODY IS CONTAINER)
人間は動物 (HUMAN IS ANIMAL)	希望は所有物 (HOPE IS POSSESSION)	機械は人間 (MACHINE IS HUMAN)	過去は後ろ (PAST IS BEHIND)
動物は人間 (ANIMAL IS HUMAN)	希望は膨らむもの (HOPE IS SOMETHING THAT SWELLS)	活力は物質 (VITALITY IS SUBSTANCE)	選挙は戦い (ELECTION IS BATTLE)
問題は植物 (PROBLEM IS PLANT)	怒りは火 (ANGER IS FIRE)	活動は移動 (VITALITY IS MOVEMENT)	金銭は水 (MONEY IS WATER)
問題は火 (PROBLEM IS FIRE)	怒りは熱 (ANGER IS HEAT)	無意識は下 (UNCONSCIOUSNESS IS DOWN)	関係は建物 (RELATION IS BUILDING)
問題は障害物 (PROBLEM IS OBSTACLE)	意識は上 (CONSCIOUSNESS IS UP)	物理・感情状態は存在物 (PHYSICAL AND EMOTIONAL STATE IS	関係は線 (RELATION IS LINE)
因果は連結 (CASUALITY IS CONNECTION)	感情は力 (EMOTION IS FORCE)	状態は場所 (STATE/CONDITION IS PLACE)	
困難は重荷 (DIFFICULTY IS BURDEN)	感情は所有物 (EMOTION IS POSSESION)	状況は天候 (SITUATION IS WEATHER)	
困難は障害 (DIFFICULTY IS OBSTACLE)	感情は水 (EMOTION IS WATER)	組織は物理的構造物 (ORGANIZATION IS PHYSICAL STRUCTURE)	

Figure 5 List of conceptual metaphors extracted on which both annotators agreed

This work is published without peer review and is licensed by the author(s) under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).