Large-Scale Japanese Metaphor Corpus Construction: Expanding BCCWJ-Metaphor with Automated Annotation

Hang Zhu[1] **Rowan Hall Maudslay**[1,2] **Kanako Komiya**[1]
Sachi Kato[3] **Masayuki Asahara**[4]
[1]Tokyo University of Agriculture and Technology, Japan
[2]University of Cambridge, United Kingdom
[3]Hokkaido University, Japan
[4]National Institute for Japanese Language and Linguistics, Japan
{s259390y@st.go.tuat.ac.jp, rh635@cam.ac.uk,
kkomiya@go.tuat.ac.jp, katosachi@let.hokudai.ac.jp, masayu-a@ninjal.ac.jp}

Abstract

This paper presents the construction of a large-scale Japanese metaphor corpus through automated annotation of the Balanced Corpus of Contemporary Written Japanese (BCCWJ). Building upon recent advances in Japanese metaphor detection using WLSP-enhanced models, we apply automated metaphor detection to create comprehensive metaphor annotations across the entire BCCWJ, including both the manually annotated BCCWJ-Metaphor subset and portions of BCCWJ beyond this core dataset with automatic WLSP semantic annotations. To validate the quality of our automated annotations, we conduct a systematic evaluation on the existing BCCWJ-Metaphor corpus, revealing that 60.3% of newly predicted metaphor-related words represent genuine metaphorical expressions that demonstrate the reliability of our approach. We provide a comprehensive analysis across four Japanese metaphor types—word-level metaphors, metonymy, synecdoche, and discourse-level metaphors—revealing systematic patterns in automated detection capabilities. The resulting corpus represents the largest Japanese metaphor resource, enabling large-scale studies of metaphor usage patterns across diverse text types and providing essential training data for future metaphor detection research.

1 Introduction

The development of large-scale metaphor corpora is crucial for advancing computational metaphor research, yet most existing resources remain limited in scope due to the time-intensive nature of manual annotation. The BCCWJ-Metaphor corpus (Kato et al., 2022, 2025) provides the first comprehensive Japanese metaphor dataset, featuring systematic annotation of four metaphor types across newspaper, magazine, and book samples from the Balanced Corpus of Contemporary Written Japanese.

The metaphors targeted in this paper are four types of figurative expressions annotated in this corpus: word-level metaphors, metonymy, synecdoche, and discourse-level metaphors. While these categories encompass different rhetorical mechanisms, they are collectively treated as metaphorical phenomena within the Japanese linguistic framework established by the BCCWJ-Metaphor corpus.

While this corpus represents a significant milestone in Japanese metaphor research, the limited scope of manual annotation due to the time-intensive nature and the potential for oversight in comprehensive coverage suggests opportunities for systematic expansion through computational methods to enhance both scale and completeness.

The computational expansion of metaphor resources have become feasible through recent advances in automatic semantic annotation. Asada et al. (2024) created comprehensive WLSP semantic annotations for the entire BCCWJ, achieving 88.05% accuracy through BERT-based all-words word sense disambiguation. This high-quality semantic annotation infrastructure enables prototypical sense-based metaphor detection approaches to be applied across previously unexplored text types and genres.

Building upon this semantic annotation infrastructure, recent transformer-based approaches have demonstrated promising capabilities for identifying metaphorical expressions in Japanese text. However, the application of these models to create large-scale metaphor corpora that extend beyond the limited scope of manually annotated resources remains underexplored. While previous work focused on model development and evaluation against manually annotated data, the potential for applying trained models to construct comprehensive metaphor corpora across diverse text types has not been systematically investigated.

This paper addresses this gap by systematically applying our trained metaphor detection model to the entire BCCWJs, creating a large-scale metaphor resource that encompasses both the existing BCCWJ-Metaphor subset and the broader portions of BCCWJ beyond this core dataset (hereafter referred to as BCCWJ-noncore). To validate the quality of our automated annotations, we conduct comprehensive evaluation on the BCCWJ-Metaphor corpus, revealing that 60.3% of newly predicted metaphor-related words represent authentic metaphorical usage, demonstrating the reliability of our approach for large-scale corpus construction.

We make several contributions: First, we present the first large-scale automated construction of a Japanese metaphor corpus, extending metaphor annotation beyond the limited scope of manually annotated resources. Second, we demonstrate the reliability of automated metaphor detection through systematic validation on BCCWJ-Metaphor. Third, we provide a comprehensive analysis across four metaphor types and diverse text genres, revealing systematic patterns in automated detection capabilities across different linguistic contexts.

2 Related Work

2.1 Semi-Automatic Corpus Construction

Semi-automatic corpus construction has emerged as an effective approach for creating large-scale annotated linguistic resources. Komiya et al. (2018) demonstrated that for Named Entity Recognition corpora, semi-automatic annotation—where automatic tagging is followed by manual correction—proves more efficient than manual annotation which is conducted from scratch. Their comparative study showed that this approach not only reduces annotation time but also maintains high annotation quality across different annotation methods.

Similar approaches have been successfully applied to semantic annotation tasks. Scarlini et al. (2020) automatically assigned word sense information to corpora in five languages (English, French, Italian, German, and Spanish), demonstrating that automatically annotated semantic information proves useful for training machine learning models. For Chinese, Zan et al. (2018) annotated a 1.87 million-word corpus using automatic annotation methods for new domain corpora.

This methodology is particularly valuable for complex linguistic phenomena requiring expert judgment, such as metaphor detection, where manual annotation challenges limit comprehensive coverage.

2.2 Metaphor Corpora and Annotation

Research on metaphor annotation and detection has advanced significantly across multiple languages. In English, MIPVU (Steen et al., 2010; Krennmayr and Steen, 2017) is a widely used annotation method, which is an extendion of the Metaphor Identification Procedure (MIP) (Pragglejaz, 2007). Several other annotation proecdures have been proposed, including the Deliberate Metaphor Identification Procedure (DMIP) (Reijnierse et al., 2018), which is designed to detect potentially deliberate metaphors, and the Procedure for Identifying Metaphorical Scenes (PIMS) (Johansson Falck and Okonski, 2022), which is aimed at capturing metaphorical scenes at the sentence or phrase level.

Similar efforts have been conducted in various other languages, including French (Reijnierse, 2010), German (Herrmann et al., 2019), Dutch (Pasma, 2019), Russian (Badryzlova et al., 2013), Spanish (Sanchez-Bayona and Agerri, 2022), Mexican Spanish (Sánchez-Montero et al., 2024, 2025), and Polish (Hajnicz, 2022). Each has resulted in language-specific metaphor annotation corpora and analyses.

3 Data

3.1 Balanced Corpus of Contemporary Written Japanese (BCCWJ)

The Balanced Corpus of Contemporary Written Japanese (BCCWJ) (Maekawa et al., 2014) serves as the foundation for our corpus construction. BCCWJ contains 104.3 million words across diverse genres, providing comprehensive coverage of contemporary Japanese written language. The corpus includes production-reality samples corresponding to books (PB), magazines (PM), and newspapers (PN), circulation-reality samples corresponding to books (LB), and special-purpose samples including white papers (OW), textbooks (OT), public relations papers (OP), bestsellers (OB), Yahoo! Chiebukuro (OC), Yahoo! blogs (OY), verse (OV), legal documents (OL), and Diet proceedings (OM).

BCCWJ employs systematic sampling methods for each genre and provides morphological analysis with short-unit word segmentation. The core portion of BCCWJ, approximately 1.2 million words from PB, PM, PN, OW, OC, and OY samples, includes manually validated morphological information. This diverse and balanced structure makes BCCWJ an ideal foundation for large-scale corpus annotation projects.

3.2 BCCWJ Automatic Semantic Annotation with WLSP

The Word List by Semantic Principles (WLSP) (for Japanese Language and Linguistics, 2004) is a comprehensive Japanese semantic classification system containing 101,170 entries organized into hierarchical categories. The semantic categories are structured as five-digit numbers, where the left digit represents classes (1. entity, 2. function, 3. relation, 4. other) and the first right digit represents divisions (.1 relation, .2 subject, .3 activity, .4 product, .5 nature).

A critical development enabling large-scale metaphor corpus construction was the automatic semantic annotation of BCCWJ (denoted as BCCWJ-WLSP-auto) by Asada et al. (2024). This work applied BERT-based word sense disambiguation to assign WLSP (Kato et al., 2018) concept IDs to content words throughout the entire BCCWJ.

The automatic annotation achieved 88.05% accuracy through 5-fold cross-validation on manually annotated BCCWJ-WLSP data, demonstrating reliable performance across multiple text registers, including books, magazines, newspapers, legal documents, blogs, and other genres. This high-quality semantic annotation infrastructure provides the essential foundation for prototypical, sense-based approaches to metaphor detection, enabling the systematic retrieval of usage examples that represent prototypical word meanings according to WLSP classifications.

3.3 BCCWJ-Metaphor Corpus

The BCCWJ-Metaphor corpus (Kato et al., 2022, 2025) represents the first comprehensive Japanese metaphor dataset, providing systematic annotation of figurative expressions within BCCWJ samples that have been assigned WLSP concept IDs. The corpus encompasses newspaper (PN), magazine (PM), and book (PB) samples totaling 347,094 words.

The annotation guidelines combine the Metaphor Identification Procedure (MIP; Pragglejaz, 2007) with Japanese-specific linguistic concepts, particularly Nakamura's syntactic construction theory (Nakamura, 1977). The Japanese concept of syntactic construction is somewhat similar to the English concept of a "frame", but is distinct in that, while frames in English are typically verb-centered, Japanese syntactic constructions often involve verb–noun or noun–noun pairs. This theory categorizes metaphor recognition into three types: A-type (indicator-based recognition), B-type (construction-based recognition through deviations from conventional usage), and C-type (context-based recognition). The corpus includes four main categories of figurative expressions:

- **Word-level metaphors** (結合比喩): Representing the majority of metaphorical expressions, where unnatural constructions between words create metaphorical meanings through similarity-based transfers. For example, 心を開く (to open one's heart) uses the concrete action of opening to describe the abstract concept of becoming emotionally receptive. These correspond to Type B recognition in Nakamura's framework, where metaphors are identified through unconventional word constructions that deviate from typical selectional restrictions.

- **Metonymy** (換喩): Involves contiguous relationships where one entity is referred to by mentioning another closely associated entity. For example, using "the crown" to refer to the monarchy.

- **Synecdoche** (提喩): Representing part-whole relationships where a part stands for the whole or vise versa. For example, 言葉で語る (to speak with words), where 言葉 (words), as a component of language, represents the entire linguistic expression system.

- **Discourse-level metaphors** (文脈比喩): Expressions whose figurativeness is determined through broader contextual understanding rather than through individual word meanings. These correspond to Type C recognition as described by Nakamura's framework, where incongruity with the surrounding context signals metaphorical meaning. For example, in a sentence about "climbing a hill" in a business context, the metaphorical nature becomes apparent only through understanding the broader discourse context.

The corpus follows BIO tagging conventions, where B-tags mark the beginning of figurative expressions, I-tags mark continuations, and O-tags mark non-figurative words. This annotation approach captures not only individual metaphorical words but also relevant contextual information that contributes to metaphorical interpretation. The

BCCWJ-Metaphor corpus is planned to be publicly released to support future research in Japanese metaphor detection and analysis.

4 Methodology

4.1 Corpus Construction Pipeline

Our approach constructs a large-scale Japanese metaphor corpus through systematic automated annotation of the entire BCCWJ[1]. The pipeline consists of three main stages.

Target Corpus Preparation We target two complementary datasets within the BCCWJ framework:

- **BCCWJ-Metaphor** (347,094 words): The manually annotated subset covering newspaper (PN), magazine (PM), and book (PB) samples. This serves primarily for quality validation and model reliability assessment.

- **BCCWJ-WLSP-auto** (approximately 100 million words): The broader BCCWJ with automatic WLSP semantic annotations (Asada et al., 2024), including legal documents (OL), textbooks (OT), blogs (OY), white papers (OW), and other genres. This constitutes the main target for large-scale corpus construction.

Target Word Selection and Preprocessing We focus on content words that are suitable for metaphor detection. Content words (nouns, verbs, adjectives, adverbs) are selected because they carry semantic content that can be compared between contextual and prototypical usage according to MIP principles. Function words, pronouns, numerals, and symbols are excluded because they primarily serve grammatical or referential functions rather than conveying metaphorical meanings through semantic transfer [2]. This categorization ensures our metaphor detection focuses on linguistically meaningful metaphorical usage while maintaining computational efficiency. Then we extract sentence-level contexts for each target word, ensuring sufficient contextual information for accurate metaphor detection while maintaining consistency with BCCWJ sentence boundaries.

Model Selection and Application For large-scale corpus construction, we selected the Fold 4 model from our 5-fold cross-validation experiments, which achieved the highest F1-score of 75.65%, to predict metaphorical expressions across the BCCWJ-noncore data. For the BCCWJ-Metaphor portion, we utilized the existing test set results from the original 5-fold cross-validation experiments, where each instance was predicted by the model that did not see it during training, ensuring unbiased evaluation.

Automated Metaphor Prediction Apply our trained WLSP-based model to systematically predict metaphorical expressions across all selected content words. For each word, the model performs a sequential process: it begins by retrieving prototypical usage examples based on the word's WLSP concept ID and incorporates explicit WLSP semantic classification features to enrich the input. Based on the comparison between the contextual and prototypical usage, the model then generates a binary prediction (1 for metaphor, 0 for non-metaphor) for the word, allowing us to automatically label metaphorical words throughout the entire BCCWJ.

4.2 WLSP-Enhanced Metaphor Detection Model

To implement the corpus construction pipeline described in Section 4.1, our approach builds upon a WLSP-enhanced metaphor detection model that adapts transformer-based architectures for Japanese metaphor detection.

The core methodology follows these principles: First, for each target word in context, the model retrieves prototypical usage examples based on the word's WLSP concept ID, ensuring systematic determination of prototypical senses according to established semantic classifications. Second, the model incorporates explicit semantic features from WLSP hierarchical categories to enrich contextual representations. Third, through transformer-based comparison mechanisms, the model evaluates the semantic distance between contextual usage and prototypical examples to generate metaphor predictions.

This approach addresses key limitations in previous metaphor detection methods by providing theoretically principled prototypical sense determination rather than relying on arbitrary non-metaphorical examples or dictionary definitions.

Our model processes metaphor detection through

[1]BCCWJ-Metaphor will be made available in the future on the BCCWJ subscribers' website.

[2]See Appendix A for detailed part-of-speech categories included and excluded in our analysis.

three main stages as detailed below.

Prototypical Sense Determination A fundamental challenge in applying MIP to computational metaphor detection lies in systematically determining what constitutes the prototypical sense of polysemous words. Unlike previous approaches that rely on frequency statistics or arbitrary selection, we leverage the manually curated prototypical sense annotations in WLSP.

Given a target word w with lemma l appearing in context with concept ID c_{context}, we determine its prototypical sense through a confidence-based selection process. We first define the set of highest-scoring concept IDs:

$$C_{\text{top}} = \arg\max_{c \in C_l} \text{confidence}(l, c) \tag{1}$$

where C_l represents all possible concept IDs associated with lemma l in WLSP.

The confidence function reflects the manual annotation scheme used by linguistic experts, with scores ranging from -1 (not prototypical sense) to 4 (confirmed prototypical sense): 4 for confirmed prototypical sense, 3 for high confidence, 2 for medium confidence, 1 for uncertain, 0 for no annotation, and -1 for not prototypical sense.

To select c_{proto} from C_{top}: if $c_{\text{context}} \in C_{\text{top}}$, we select it; otherwise, we randomly select one from C_{top}. If no candidate is found in WLSP, we use $c_{\text{proto}} = c_{\text{context}}$.

Prototypical Usage Example Retrieval Using c_{proto}, we retrieve corresponding prototypical usage examples from BCCWJ-WLSP-auto by finding all instances where the lemma matches l and the concept ID matches c_{proto}. We denote the selected prototypical usage example as u_{proto}, which is chosen as follows: if multiple examples exist, we randomly select one; otherwise, we use the lemma l itself.

Japanese Metaphor Detection Model Given a target sentence $S = \{w_1, \ldots, w_n\}$ containing the target word w_t, we first enrich it with semantic classification information derived from the WLSP database by its concept ID c_{context}. This forms an extended sequence:

$$S' = (w_1, \ldots, w_n, \text{[SEP]}, f_1, \ldots, f_k) \tag{2}$$

where $\{f_i\}$ are features representing semantic classification information from WLSP. If the information cannot be found, [MASK] is used as f_i. These features consist of WLSP's semantic classification information including 類 (classes), 部門 (divisions), 中項目 (sections), and 分類項目 (subcategories).

We employ a multi-level Token Type IDs system to distinguish different types of information in the extended sequence. Each token is assigned a role: Target Word, Local Context (words within punctuation boundaries around the target word), Semantic Features (semantic classification information from WLSP), and Background (all other tokens). This role-based encoding scheme enables BERT to process different types of linguistic information with specialized attention patterns.

After encoding both the target sentence S' and prototypical example $P = u_{\text{proto}}$ containing the same lemma l using Japanese BERT:

$$\mathbf{v}_{S',1}, \ldots, \mathbf{v}_{S',n} = \text{BERT}(S') \tag{3}$$

$$\mathbf{v}_{P,1}, \ldots, \mathbf{v}_{P,m} = \text{BERT}(P) \tag{4}$$

where $\mathbf{v}_{S',t} \in \mathbb{R}^{h \times 1}$ and $\mathbf{v}_{P,t'} \in \mathbb{R}^{h \times 1}$ are the contextualized embedding vectors for the target word at positions t and t' respectively, and h is the dimension of BERT's hidden state.

We then compute a vector $\mathbf{h}_{\text{MIP}} \in \mathbb{R}^{h \times 1}$ that captures the semantic relationship between contextual and prototypical senses:

$$\mathbf{h}_{\text{MIP}} = f([\mathbf{v}_{S',t}; \mathbf{v}_{P,t'}]) \tag{5}$$

where $f(\cdot)$ is a linear transformation that learns the semantic difference between the contextual usage $\mathbf{v}_{S',t}$ and the prototypical sense $\mathbf{v}_{P,t'}$.

The model is trained using cross-entropy loss:

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^{N} [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \tag{6}$$

This model achieved an F1-score of 75.1 on BCCWJ-Metaphor through 5-fold cross-validation, providing the foundation for large-scale automated corpus construction.

5 Evaluation

To assess the quality of our automatically generated metaphor corpus, we conducted a manual validation of the novel annotations produced by our model to demonstrate its ability to augment and enrich existing resources.

5.1 Validation of Novel Metaphor Annotations

While the BCCWJ-Metaphor provides a crucial benchmark, manual annotation processes, despite their high quality, may have occasional oversights due to the complexity and time-intensive nature of comprehensive metaphor identification. A key part of our contribution, therefore, is to demonstrate our model's ability to identify and fill these gaps.

We applied our trained model to the entire BCCWJ to generate comprehensive metaphor predictions. To validate the quality of our automated annotations, we focused on the BCCWJ-Metaphor portion and identified instances that were originally not labeled as metaphorical but were newly predicted as metaphors by our model. From these newly predicted instances, we randomly sampled 500 cases for manual validation by an expert linguist specializing in Japanese metaphor research.

The results of this validation demonstrate the effectiveness of our approach for corpus augmentation. The analysis revealed that 60.3% of the newly identified instances were judged to be genuine metaphorical expressions that had been missed in the original manual annotation. Among these 302 validated genuine metaphors, the distribution across metaphor types was as follows: word-level metaphors accounted for 59.2% (179 instances), metonymy for 11.3% (34 instances), synecdoche for 15.3% (46 instances), and discourse-level metaphors for 0.6% (2 instances). The remaining 13.6% (41 instances) were classified into other categories or required further analysis.

It proves that our automated method is not merely replicating human work, but is acting as a powerful tool to enhance it by discovering valid omissions. This validates the quality of our new, larger corpus as a more complete resource for metaphor research.

6 Error Analysis

Our evaluation quantitatively demonstrated the model's overall effectiveness. A deeper analysis of its performance, however, reveals important patterns in its errors. The following discussion explores the linguistic reasons behind these challenges, drawing on specific examples from our analysis.

6.1 Word-level Metaphor Detection Challenges

Among the compound metaphors annotated in BCCWJ-Metaphor, 21.23% were found to be incorrect. A notable tendency was that the system failed to detect compound metaphors when the dictionary definition of a word included senses explicitly labeled as "figurative" or "by extension." Such cases are so highly conventionalized that human annotators can readily recognize them as metaphors; however, the system often misclassified them. For example, the system judged the expression "誕生" ("birth") in "無党派組織『２１世紀の千葉を創る県民の会』が知事誕生の原動力となった" ("The nonpartisan organization 'Prefectural Citizens' Association to Create 21st Century Chiba' became the driving force behind the governor's birth") as non-metaphorical. In the construction "知事の誕生" ("the governor's birth"), the literal meaning of "誕生" refers to human birth. However, in this context, it is used in the figurative sense of "the emergence or establishment of something" (Shogakukan Inc., 2000–2002). Consequently, the annotator labeled it as a metaphor, but the system failed to do so.

6.2 Metonymy Detection Challenges

Among the metonymies annotated in BCCWJ-Metaphor, 23.30% were incorrect. A frequent error occurred when the system targeted content words within company names. Metonymic expressions tend to be identified in constructions involving selectional restriction violations, such as "a company sells". Since company names are often used in metonymic contexts, this is a reasonable target; however, in this study, parts of proper nouns were sometimes misidentified.

Errors also occurred when interpretation required sentence-level understanding rather than solely relying on lexical meaning or construction patterns. For example, in [赤松円心が願っていたのも、]中央の犠牲となることなく、穏やかに安定してその地方を治めていくことにあったはずで" ("[It must have been Akamatsu Enshin's wish as well] to govern the local region peacefully and stably without becoming a sacrifice to the center [central government,]") the construction "中央の犠牲" ("a sacrifice to the center") can be judged metaphorical. Nevertheless, the system labeled it as non-metaphorical. In this case, recognizing the metonymy requires understanding that "中央" ("the center") corresponds to "地方" ("the local region") in the sentence, a relationship that cannot be easily detected from the lexical meaning or construction alone.

6.3 Synecdoche Detection Challenges

Synecdoche exhibited the highest error rate among metaphor categories in BCCWJ-Metaphor, with 39.50% incorrect. This category often requires interpretation based on context or background knowledge, making it difficult to identify based solely on lexical meaning or deviations from conventional constructions. One contributing factor is the prevalence of euphemistic and exemplary expressions.

For example, in "大和の国の中には、水にめぐまれへん村がぎょうさんあった" ("In Yamato Province, there were many villages not blessed with water"), the system failed to identify "めぐむ" ("bless") as synecdoche. The basic sense of "めぐむ" is "to feel affection for," with extended meanings such as "to show compassion" and "to give alms" (Shogakukan Inc., 2000–2002). In the idiomatic expression "水にめぐまれる" ("to be blessed with water"), the phrase euphemistically refers to the availability of water, which can be interpreted as a type–category relationship, hence synecdoche.[3] However, the system failed to detect this. Such euphemistic expressions are highly conventionalized, making them difficult to classify as selectional restriction violations at the word or construction level, even though they are easily recognized as figurative by humans. Nevertheless, in the present method, certain euphemistic expressions were successfully detected—for example, "世を去る" ("to leave the world," a euphemism for "to die") in "[誰かが]この世を去った時に" ("when [someone] passes away"). Like "水にめぐまれる", this is idiomatic, but its appearance in a hypothetical context may have influenced the system's judgment.

Expressions closer to prototypical exemplification also proved challenging when they required contextual or background knowledge. For instance, in "下戸の彼氏もコーラで付き合う" ("Even my boyfriend, who cannot drink alcohol, joins in with cola"), "コーラ" ("cola") refers to a category of soft drinks. In this context, given "下戸" ("unable to drink alcohol") and a drinking-party scenario, "cola" functions as a prototypical example of a non-alcoholic beverage used to participate in the event. However, the system judged it as non-metaphorical.

Interestingly, there were also cases where the system successfully detected expressions that humans might not easily recognize as figurative. For example, in "京都に店を出す" ("open a shop in Kyoto"), "出す" ("to put out") is interpreted in an extended sense beyond its basic meaning, constituting synecdoche. Although humans may perceive this as a natural semantic extension, the system correctly identified it.

[3] In Japan, with its animistic cultural background, there isn't a single fixed deity that performs blessings. As a result, the idea of "blessed with water" is understood as a metaphorical expression.

6.4 Discourse-level Metaphor Detection Challenges

Among the metaphor categories in BCCWJ-Metaphor, contextual metaphors had the second-highest error rate at 30.20%. The present approach is generally ill-suited to detecting such cases. When an expression is not metaphorical at the sentence level, metaphor recognition cannot be achieved through lexical meaning alone. However, the system did correctly identify certain examples. For instance:

- "[船頭、つまり経営者が]仕かけを工夫して、[釣り人、つまり社員に]釣り方を教える" ("[The boatman, that is, the manager] devises traps and teaches [the fisherman, that is, the employee,] how to fish") occurs in a business-strategy context, interpreted as a figurative expression that, at the sentence level, deviates from the context.
- "種をまいても収穫を急がんことやで" ("Even if you sow seeds, don't rush the harvest") is used in the context of business know-how, also interpreted as a figurative expression.
- "目が画面に釘づけになった" ("My eyes were nailed to the screen") appears in a surveillance-camera playback context, making it a well-established metaphor.
- "「ヤマタノオロチ」が暴れている時に、政府は「草薙の剣」を使うことができなかったようなものです" ("It is as if, when the 'Yamata no Orochi,' a famous giant eight-headed serpent in Japanese mythology, was rampaging, the government could not use the 'Kusanagi sword,' a legendary Japanese sword") occurs in the context of the government's response to a financial crisis. The quotation marks and the inherently metaphorical construction "government uses the 'Kusanagi sword'" likely contributed to the system's detection.

These examples suggest that when unnatural constructions or typographical indicators (e.g., quotation marks) are present, metaphor detection may be aided by features beyond long-range discourse context.

In summary, the system's correct and incorrect judgments did not necessarily align with the ease of human annotation. While there are cases where the system failed on expressions that humans find easy to classify, it could also capture instances that humans may overlook. This suggests that such a system has strong potential as a supplementary tool for human metaphor annotation.

7 Conclusion

This paper presents the first large-scale automated construction of a Japanese metaphor corpus, extending metaphor annotation across the entire BCCWJ through systematic application of WLSP-based detection models. Our approach successfully creates the largest Japanese metaphor resource to date, encompassing diverse text types beyond the scope of manual annotation.

The validation study provides compelling evidence for the effectiveness of automated corpus construction. With 60.3% of newly predicted instances representing genuine metaphorical expressions, our method demonstrates the ability to identify valid metaphors missed in manual annotation, enhancing rather than merely replicating existing resources.

Our analysis reveals systematic patterns in detection capabilities across metaphor types, with particular challenges in synecdoche and discourse-level metaphors due to conventionalization and contextual dependencies. These findings provide valuable insights for future model development and highlight the complexity of automated metaphor detection in Japanese.

The resulting corpus enables large-scale studies of metaphor usage patterns across diverse genres and provides essential training data for advancing computational metaphor research.

Future work should focus on improving the detection of highly conventionalized expressions and incorporating broader contextual information to better handle discourse-level metaphors.

Acknowledgements

This work was supported by JSPS KAKENHI Grants JP22K12145 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 are also grateful to Professor Makoto Yamazaki and Professor Wakako Kashino for providing us with the basic sense data of the WLSP.

References

Soma Asada, Kanako Komiya, and Masayuki Asahara. 2024. Comprehensive assignment of word list by semantic principles concept ids to the balanced corpus of contemporary written japanese. In *Proceedings of the 30th Annual Conference of the Association for Natural Language Processing*, pages 2767–2772.

Yulia Badryzlova, Natalia Shekhtman, Yekaterina Isaeva, and Ruslan Kerimov. 2013. Annotating a Russian corpus of conceptual metaphor: a bottom-up approach. In *Proceedings of the First Workshop on Metaphor in NLP*, pages 77–86, Atlanta, Georgia. Association for Computational Linguistics.

National Institute for Japanese Language and Linguistics, editors. 2004. *Bunrui Goi Hyo: Expanded and Revised Edition*. Dai Nihon Tosho. In Japanese.

Elżbieta Hajnicz. 2022. Annotation of metaphorical expressions in the basic corpus of Polish metaphors. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 5648–5653, Marseille, France. European Language Resources Association.

JBB Herrmann, Karola Woll, and Aletta Dorst. 2019. Linguistic metaphor identification in german. *Metaphor identification in multiple languages. MIPVU around the world*, pages 113–135.

Marlene Johansson Falck and Lacey Okonski. 2022. Procedure for identifying metaphorical scenes (pims): A cognitive linguistics approach to bridge theory and practice. *Cognitive Semantics*, 8(2):294–322.

Sachi Kato, Masayuki Asahara, and Makoto Yamazaki. 2018. Annotation of 'word list by semantic principles' labels for the Balanced Corpus of Contemporary Written Japanese. In *Proceedings of the 32nd Pacific Asia Conference on Language, Information and Computation*, Hong Kong. Association for Computational Linguistics.

Sachi Kato, Rei Kikuchi, and Masayuki Asahara. 2022. Assignment of metaphorical expression information based on mip to the balanced corpus of contemporary written japanese. In *Proceedings of the 28th Annual Conference of the Association for Natural Language Processing, In Japanese*, pages 1427–1431.

Sachi Kato, Rei Kikuchi, and Masayuki Asahara. 2025. Metaphor recognition and information assignment procedures in bccwj-metaphor. In *Proceedings of the 31st Annual Conference of the Association for Natural Language Processing, In Japanese*.

Kanako Komiya, Masaya Suzuki, Tomoya Iwakura, Minoru Sasaki, and Hiroyuki Shinnou. 2018. Comparison of methods to annotate named entity corpora. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, 17(4):34.

Tina Krennmayr and Gerard Steen. 2017. Vu amsterdam metaphor corpus. In *Handbook of linguistic annotation*, pages 1053–1071. Springer.

Kikuo Maekawa, Makoto Yamazaki, Takehiko Ogiso, Masaya Maruyama, Hideki Ogura, Wakako Kashino, Hiroshi Koiso, Makoto Yamaguchi, Masahiro Tanaka, and Yasuharu Den. 2014. Balanced corpus of contemporary written japanese. *Language Resources and Evaluation*, 48(2):345–371.

Akira Nakamura. 1977. *Theory and Classification of Metaphorical Expressions*. Shuei Shuppan, Tokyo.

Tryntje Pasma. 2019. Chapter 5. linguistic metaphor identification in dutch. In *Metaphor Identification in Multiple Languages: MIPVU around the world*, pages 91–112. John Benjamins Publishing Company.

Group Pragglejaz. 2007. Mip: A method for identifying metaphorically used words in discourse. *Metaphor and Symbol*, 22(1):1–39.

G Reijnierse. 2010. Making mip operational for french: practical and theoretical issues concerning the choice of a dictionary. In *8th International Conference on Researching and Applying Metaphor (RaAM 10)*.

W. G. Reijnierse, C. Burgers, T. Krennmayr, and G. J. Steen. 2018. Dmip: A method for identifying potentially deliberate metaphor in language use. *Corpus Pragmatics*, 2(2):129–147.

Elisa Sanchez-Bayona and Rodrigo Agerri. 2022. Leveraging a new Spanish corpus for multilingual and cross-lingual metaphor detection. In *Proceedings of the 26th Conference on Computational Natural Language Learning (CoNLL)*, pages 228–240, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.

Alec Sánchez-Montero, Gemma Bel-Enguix, and Sergio-Luis Ojeda-Trueba. 2024. Evaluating the development of linguistic metaphor annotation in Mexican Spanish popular science tweets. In *Proceedings of the 4th Workshop on Figurative Language Processing (FigLang 2024)*, pages 59–64, Mexico City, Mexico (Hybrid). Association for Computational Linguistics.

Alec Sánchez-Montero, Gemma Bel-Enguix, Sergio-Luis Ojeda-Trueba, and Gerardo Sierra. 2025. Disagreement in metaphor annotation of Mexican Spanish science tweets. In *Proceedings of Context and Meaning: Navigating Disagreements in NLP Annotation*, pages 155–164, Abu Dhabi, UAE. International Committee on Computational Linguistics.

Bianca Scarlini, Tommaso Pasini, and Roberto Navigli. 2020. Sense-annotated corpora for word sense disambiguation in multiple languages and domains. In *International Conference on Language Resources and Evaluation*.

Shogakukan Inc. 2000–2002. , 2nd edition. Shogakukan Inc., Tokyo. Digital version accessed via JapanKnowledge, released July 2, 2007.

Gerard J. Steen, Aletta G. Dorst, J. Berenike Herrmann, Anna A. Kaal, Tina Krennmayr, and Tom Pasma. 2010. *A Method for Linguistic Metaphor Identification: From MIP to MIPVU*. John Benjamins Publishing.

Hongying Zan, Junyi Chen, Xiaoyu Cheng, and Lingling Mu. 2018. Construction of word sense tagging corpus. In *Chinese Lexical Semantics - 19th Workshop, CLSW 2018, Chiayi, Taiwan, May 26-28, 2018, Revised Selected Papers*, volume 11173 of *Lecture Notes in Computer Science*, pages 679–690. Springer.

A Part-of-Speech Categories for Target Word Selection

Our content word filtering process targets word classes that are typically subject to metaphorical usage according to MIP and Japanese linguistic characteristics. The selection is based on UniDic part-of-speech categories used in BCCWJ morphological analysis.

Table 1: Part-of-Speech Categories for Metaphor Detection

Japanese	English
Selected Categories (Content Words)	
名詞-数詞	Noun-Numeral
名詞-助動詞語幹	Noun-Auxiliary Verb Stem
名詞-普通名詞-一般	Noun-Common-General
名詞-普通名詞-サ変可能	Noun-Common-Verbal
名詞-普通名詞-形状詞可能	Noun-Common-Adjectival
名詞-普通名詞-副詞可能	Noun-Common-Adverbial
名詞-普通名詞-助数詞可能	Noun-Common-Counter
名詞-固有名詞-一般	Noun-Proper-General
名詞-固有名詞-地名-一般	Noun-Proper-Place-General
名詞-固有名詞-地名-国	Noun-Proper-Place-Country
名詞-固有名詞-人名-一般	Noun-Proper-Person-General
動詞-一般	Verb-General
動詞-非自立可能	Verb-Bound
形容詞-一般	Adjective-General
形容詞-非自立可能	Adjective-Bound
副詞	Adverb
形状詞-一般	Adjectival Noun-General
連体詞	Attributive
代名詞	Pronoun
Excluded Categories (Function Words)	
助詞-格助詞	Particle-Case
助詞-係助詞	Particle-Binding
助詞-接続助詞	Particle-Conjunctive
助詞-副助詞	Particle-Adverbial
助詞-終助詞	Particle-Final
助動詞	Auxiliary Verb
補助記号-読点	Symbol-Comma
補助記号-句点	Symbol-Period
補助記号-括弧閉	Symbol-Bracket Close
補助記号-括弧開	Symbol-Bracket Open
接頭辞	Prefix
接続詞	Conjunction
接尾辞-名詞的-一般	Suffix-Nominal-General
接尾辞-名詞的-助数詞	Suffix-Nominal-Counter
接尾辞-形状詞的	Suffix-Adjectival Noun
記号-一般	Symbol-Character
感動詞-一般	Interjection-General
空白	Whitespace