Automatic Detection of Metaphorical Expressions in Classical Japanese Using WLSP-Enhanced BERT

Hang Zhu[1], **Mitoki Ohara**[2], **Rei Kikuchi**[3], **Kanako Komiya**[1],
Masayuki Asahara[4], **Sachi Kato**[5]
[1]Tokyo University of Agriculture and Technology, Tokyo, Japan
[2]Chuo University, Tokyo, Japan
[3]National Institute of Technology, Nagano College, Nagano, Japan
[4]National Institute for Japanese Language and Linguistics, Tokyo, Japan
[5]Hokkaido University, Hokkaido, Japan
s259390y@st.go.tuat.ac.jp, 82303991araho@gmail.com, r_kikuchi@nagano-nct.ac.jp,
kkomiya@go.tuat.ac.jp, masayu-a@ninjal.ac.jp, katosachi@let.hokudai.ac.jp

Abstract

Metaphor detection is a fundamental task in natural language processing, yet research on historical languages remains limited. While progress has been made in modern Japanese metaphor detection, classical Japanese texts present unique challenges due to their distinct vocabulary, grammar, and metaphorical patterns. This paper addresses this gap by applying a BERT-based metaphor detection method enhanced with semantic classification information from the Word List by Semantic Principles (WLSP) to classical Japanese texts. We evaluate our approach on CHJ-Metaphor, a newly available corpus featuring metaphor annotations for three medieval Japanese works from the Corpus of Historical Japanese (CHJ). Our method achieves an F1-score of 82.18 through 5-fold cross-validation. Notably, qualitative analysis by domain experts reveals that our model successfully identifies genuine metaphors overlooked during manual annotation, demonstrating its potential as a tool for improving annotation quality in large-scale corpus construction. These results confirm the effectiveness of WLSP-enhanced approaches for metaphor detection in classical Japanese and suggest promising directions for applying similar techniques to other historical languages.

Keywords: Metaphor Detection, Classical Japanese, Historical Language Processing, BERT

1. Introduction

Metaphorical expressions play a crucial role in language understanding, and their automatic detection remains an important challenge in natural language processing (Group Pragglejaz, 2007). In this framework, a word is considered metaphorical when its meaning in context differs from its most basic, concrete, or historically earliest sense (Group Pragglejaz, 2007). Recent years have witnessed significant advances in metaphor detection research using Transformer-based models such as BERT (Choi et al., 2021; Li et al., 2023a), though most studies focus primarily on modern languages.

For Japanese, Kato et al. (2022) constructed BCCWJ-Metaphor based on the Balanced Corpus of Contemporary Written Japanese (Maekawa et al., 2014), providing systematic metaphor annotations following the Metaphor Identification Procedure (MIP) (Group Pragglejaz, 2007). Building upon this foundation, Zhu et al. (2025) leveraged semantic classification information from the Word List by Semantic Principles (National Institute for Japanese Language and Linguistics, 1964) to achieve an F1-score of 75.07 for modern Japanese metaphor detection.

However, automatic metaphor detection in classical Japanese texts remains largely unexplored. Classical Japanese differs substantially from modern Japanese in vocabulary and grammar, and metaphorical usage patterns may have evolved over time. The Corpus of Historical Japanese (National Institute for Japanese Language and Linguistics, 2023) provides a rich resource spanning texts from the Nara period (710-794) to the Meiji era (1868-1912), but metaphor detection for these historical texts presents unique challenges.

To address this research gap, we apply the WLSP-based BERT approach, previously validated on modern Japanese, to classical Japanese metaphor detection. We evaluate our method on CHJ-Metaphor, which provides metaphor annotations for classical Japanese texts from CHJ enhanced with WLSP semantic classifications (Asahara et al., 2022). Specifically, we conduct experiments on three currently annotated works from the medieval period (1185-1573, comprising the Kamakura and Muromachi eras): *Konjaku Monogatari-shu* (Tales of Times Now Past), *Hojoki* (An Account of My Hut), and *Toraakira-bon Kyogen-shu* (Toraakira's Kyogen Collection).

Our contributions are threefold: (1) We demonstrate the first application of automatic metaphor detection to classical Japanese texts, achieving an F1-score of 82.18 through 5-fold cross-validation. (2) We provide a comprehensive analysis across works with varying characteristics, revealing that performance remains stable across metaphor rates

Proceedings of LT4HALA 2026 @ LREC 2026, pages 89–95.

ranging from 4.41% to 25.82%. (3) Through expert validation, we demonstrate that our model identifies genuine metaphors overlooked during manual annotation, suggesting its potential utility for improving annotation quality in large-scale corpus construction.

2. Related Work

2.1. Metaphor Detection Approaches

Metaphor detection approaches can be broadly categorized into three types: dictionary and rule-based methods (Dodge et al., 2015), statistical approaches (Klebanov et al., 2014), and deep learning-based methods (Mao et al., 2019). Recent advances have been dominated by Transformer-based models (Tong et al., 2021), with various architectural innovations proposed to capture metaphorical meaning.

MelBERT (Choi et al., 2021) builds upon RoBERTa (Liu et al., 2019) and compares target word embeddings with contextual embeddings to detect metaphors. FrameBERT (Li et al., 2023b) incorporates semantic frame information from FrameNet (Baker et al., 2024). BasicBERT (Li et al., 2023a) learns basic sense representations from non-metaphorical examples in the training data. Other approaches include MIss RoBERTa WiLDe (Babieno et al., 2022), which leverages Wiktionary definitions, and MisNet (Zhang and Liu, 2022), which enhances linguistic knowledge for metaphor detection.

Our approach differs from these methods by leveraging the systematic semantic classification structure of WLSP, providing theoretically grounded basic sense determination.

Recent work has also explored metaphor detection beyond English, including cross-lingual transfer approaches for low-resource languages (Hülsing and Schulte Im Walde, 2024), highlighting the broader challenge of extending metaphor detection methods across linguistic boundaries.

2.2. Japanese Historical Corpora and Metaphor Research

The Corpus of Historical Japanese (National Institute for Japanese Language and Linguistics, 2023) encompasses diverse textual genres spanning from the Nara period (710-794) to the Meiji era (1868-1912). Asahara et al. (2022) developed CHJ-WLSP by annotating CHJ texts with WLSP (National Institute for Japanese Language and Linguistics, 1964) semantic classification codes. Building upon CHJ-WLSP, Kikuchi et al. are constructing CHJ-Metaphor through manual metaphor annotation. This study uses the currently completed portions of the corpus under construction.[1]

3. Dataset

3.1. CHJ-Metaphor

CHJ-Metaphor is a metaphor expression dataset for classical Japanese, created by adding metaphor annotations to CHJ-WLSP (Asahara et al., 2022). The corpus is designed to include medieval period works across three genres: narratives (*Konjaku Monogatari-shu*, *Uji Shui Monogatari*, *Jikkinsho*), essays (*Hojoki*, *Tsurezuregusa*), and kyogen plays (*Toraakira-bon Kyogen-shu*).

Metaphor identification follows the same procedure as BCCWJ-Metaphor for modern Japanese, employing MIP and its extended version MIPVU (Steen et al., 2010), along with Nakamura's (Nakamura, 1977) metaphor type classification. MIP identifies metaphorical words by comparing their contextual meaning with their most basic (prototypical) sense: a word is metaphorical if these two meanings differ in a systematic way (Group Pragglejaz, 2007). MIPVU extends MIP to handle implicit metaphors, personification, and metaphorically used proper nouns. Nakamura's classification further distinguishes metaphorical expressions by how they are comprehended: Type A (simile, where a linguistic marker signals the comparison), Type B (combinatorial metaphors, arising from selectional restriction violations), and Type C (contextual metaphors, where metaphoricity emerges from broader discourse context).

The annotation covers six types of metaphorical expressions: 指標比喩 "simile" (explicit comparison using indicators such as ように "like" or 如し "as"), 換喩 "metonymy" (referring to something by mentioning an associated concept), 提喩 "synecdoche" (using a part to represent the whole or vice versa), 文脈比喩 "Discourse-level metaphors" (metaphorical meaning emerging from context), 結合比喩 "Word-level metaphors" (metaphorical meanings created by unconventional word combinations that violate selectional restrictions), and その他 "others" (including 枕詞 "makurakotoba," a conventional poetic technique in classical Japanese).

This study uses three currently completed works. As shown in Table 1, *Konjaku Monogatari-shu* contains the largest amount of data with 79,892 tokens, while *Hojoki* (2,668 tokens) and *Toraakira-bon Kyogen-shu* (2,542 tokens) are substantially smaller. The corpus contains 85,102 tokens in total.

Table 2 presents the metaphor rate (proportion of metaphorical tokens) for each work, which

[1] The CHJ-Metaphor corpus will be made publicly available after the annotation is completed.

Work	Tokens
Konjaku Monogatari-shu	79,892
Hojoki	2,668
Toraakira-bon Kyogen-shu	2,542
Total	85,102

Table 1: Dataset Size

Work	Rate (%)
Konjaku Monogatari-shu	4.41
Hojoki	25.82
Toraakira-bon Kyogen-shu	12.79
Overall	5.33

Table 2: Metaphor Rate

varies substantially across the three texts. *Konjaku Monogatari-shu*, despite having the most data, exhibits the lowest metaphor rate at 4.41%. In contrast, *Hojoki* shows the highest metaphor rate at 25.82%, nearly six times higher. *Toraakira-bon Kyogen-shu* falls in between with a metaphor rate of 12.79%. The overall metaphor rate across all works is 5.33%.

The metaphor rate varies across literary works. The influence of genre and style on metaphor rate may be analyzed further as annotation of additional works progresses.

4. Methodology

We apply the WLSP-based BERT approach, previously validated on modern Japanese (Zhu et al., 2025), to classical Japanese metaphor detection. This section describes the core components of our method.

4.1. Prototypical Sense Determination

Determining the prototypical sense of polysemous words is a fundamental challenge in applying MIP. Our approach leverages the prototypical sense information annotated to WLSP by Yamazaki et al. (2015); Yamazaki and Kashino (2017).

Notably, the prototypical sense used in this study is based on modern Japanese rather than classical Japanese, as WLSP was originally developed for contemporary language. This methodological choice reflects the practical consideration that annotators, as contemporary speakers, inevitably interpret classical texts through a modern lens when identifying metaphorical expressions.

Yamazaki et al. assigned confidence scores on a five-level scale to each sense of polysemous words in WLSP, indicating the degree to which each sense represents the prototypical sense. For a target word w (lemma l) appearing in context with classification code c_{context}, we determine the prototypical sense as follows. We first identify the set C_{proto} containing all classification codes with the highest confidence score for lemma l in WLSP. If $c_{\text{context}} \in C_{\text{proto}}$, we select it as the prototypical sense classification code c_{proto}. Otherwise, we randomly select one code from C_{proto}. If no candidates exist in WLSP, we set $c_{\text{proto}} = c_{\text{context}}$.

4.2. Prototypical Example Retrieval

Using the determined prototypical sense classification code c_{proto}, we retrieve prototypical usage examples from BCCWJ-WLSP-auto (Asada et al., 2024). We extract all instances where both the lemma l and classification code c_{proto} match. If multiple examples exist, we randomly select one. If no matching example is found, we use the lemma l itself as the prototypical example.

4.3. BERT-based Metaphor Detection Model

For a sentence $S = \{w_1, \ldots, w_n\}$ containing the target word w_t, we first augment it with WLSP semantic classification information:

$$S' = (w_1, \ldots, w_n, \text{[SEP]}, f_1, \ldots, f_k) \tag{1}$$

where $\{f_i\}$ represents WLSP semantic features appended after `[SEP]`. WLSP organizes Japanese vocabulary into a four-level hierarchical taxonomy: *division* (類), *section* (部門), *medium category* (中項目), and *classification item* (分類項目), covering approximately 96,000 headwords. We use [MASK] tokens when classification information is unavailable.

We employ a multi-layer token type ID system to distinguish different information types within the augmented sequence. Each token is assigned a role: target word, local context (words within punctuation boundaries around the target), semantic features (WLSP classification information), or background (other tokens).

We encode the target sentence S' and the prototypical example $P = u_{\text{proto}}$ using Japanese BERT:

$$\mathbf{v}_{S',1}, \ldots, \mathbf{v}_{S',n} = \text{BERT}(S') \tag{2}$$
$$\mathbf{v}_{P,1}, \ldots, \mathbf{v}_{P,m} = \text{BERT}(P) \tag{3}$$

We then compute a vector $\mathbf{h}_{\text{MIP}}$ that captures the semantic relationship between the contextual meaning and the prototypical sense:

$$\mathbf{h}_{\text{MIP}} = f([\mathbf{v}_{S',t}; \mathbf{v}_{P,t'}]) \tag{4}$$

where $f(\cdot)$ is a linear transformation and $[\cdot;\cdot]$ denotes concatenation. The model is trained using cross-entropy loss:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(p_i) + (1 - y_i)\log(1 - p_i)] \quad (5)$$

To address class imbalance (approximately 5% metaphorical tokens in the dataset), we apply weighted cross-entropy loss by assigning a higher weight to metaphorical instances, ensuring that the model learns to detect rare metaphorical patterns effectively.

4.4. Experimental Settings

We use bert-base-japanese-v3[2] as our base model, with a maximum sequence length of 512, a batch size of 16, a learning rate of 3e-5, and train it for 5 epochs. We evaluate performance using 5-fold cross-validation.

Evaluation metrics include F1-score, precision, and recall, all computed at the token level based on CHJ-WLSP morphological units (short units).

5. Experiments and Results

5.1. Overall Performance

Table 3 presents the overall results of 5-fold cross-validation on CHJ-Metaphor.

Model	F1	Prec.	Rec.
WLSP-enhanced	82.18	82.83	81.58

Table 3: Overall Performance

Our WLSP-enhanced BERT approach achieves an F1-score of 82.18%, precision of 82.83%, and recall of 81.58% on classical Japanese texts, demonstrating the effectiveness of applying methods validated on modern Japanese to classical texts.

5.2. Performance by Work

Tables 4 and 5 show a performance breakdown across the three works.

Konjaku Monogatari-shu achieves the highest accuracy (98.41%) with a precision of 82.55% and a recall of 81.21%. *Hojoki* shows a precision of 85.26% and a recall of 84.76% with an accuracy of 92.28%. *Toraakira-bon Kyogen-shu* demonstrates a precision of 80.00% and a recall of 78.76% with an accuracy of 94.77%.

[2]https://huggingface.co/tohoku-nlp/bert-base-japanese-v3

Work	Acc.
Konjaku Monogatari-shu	98.41
Hojoki	92.28
Toraakira-bon Kyogen-shu	94.77

Table 4: Accuracy by Work

Work	Prec.	Rec.
Konjaku Monogatari-shu	82.55	81.21
Hojoki	85.26	84.76
Toraakira-bon Kyogen-shu	80.00	78.76

Table 5: Precision and Recall by Work

All three works achieve over 78% precision and recall with well-balanced performance. *Konjaku Monogatari-shu* shows the highest accuracy, *Hojoki* achieves the most balanced precision-recall performance, and *Toraakira-bon Kyogen-shu* shows intermediate performance. Detailed analysis is provided in Section 6.

6. Discussion

6.1. Performance Differences Across Works

The performance differences across the three works can be attributed to their distinct characteristics in data size and metaphor rate.

Konjaku Monogatari-shu achieves strong performance with the largest data size (79,892 tokens), which contributes to learning stability. Despite the low metaphor rate (4.41%), the model maintains balanced precision-recall, demonstrating that the weighted loss function and abundant training data enable the model to learn metaphorical patterns effectively while handling class imbalance.

Hojoki achieves the most balanced precision-recall performance (85.26% precision, 84.76% recall). The high metaphor rate (25.82%) allows the model to learn metaphorical patterns more effectively, as metaphorical instances are well represented during training. The relatively lower accuracy (92.28%) reflects the increased proportion of metaphorical tokens, which reduces the relative contribution of non-metaphorical tokens to overall accuracy calculations.

Toraakira-bon Kyogen-shu shows a precision of 80.00% and a recall of 78.76%, reflecting its intermediate characteristics in both data size and metaphor rate.

Overall, the method demonstrates stable performance across metaphor rates ranging from 4.41% to 25.82%, confirming the effectiveness of the proposed approach for classical Japanese texts with varying characteristics.

6.2. Comparison with Existing Approaches

Direct comparison with existing metaphor detection models such as MelBERT and FrameBERT is not feasible in this setting, as these models rely on language resources specific to modern languages (e.g., FrameNet, English training corpora) that have no equivalent for classical Japanese.

Regarding the contribution of WLSP semantic features, ablation experiments on modern Japanese metaphor detection (Zhu et al., 2025) demonstrate that the WLSP-enhanced model significantly outperforms a vanilla BERT baseline, confirming the effectiveness of the semantic augmentation. The present study applies this validated framework to classical Japanese, where constructing a comparable ablation baseline would require additional annotation effort beyond the scope of this work.

6.3. Differences from Modern Japanese

The strong performance on classical Japanese texts may be attributed to several linguistic features.

First, classical Japanese vocabulary is more limited compared to modern Japanese, potentially resulting in lower polysemy. Second, classical literary texts are characterized by formulaic expressions, and metaphorical patterns tend to be more regular. These factors may collectively contribute to the model's strong performance on classical texts.

6.4. Qualitative Evaluation of Detection Results

Following the experiments, the creators of CHJ-Metaphor conducted a detailed qualitative analysis of the detection results, examining instances classified as false positives by the model to determine whether they constituted genuine metaphors overlooked during the original annotation phase. The analysis reveals both strengths and limitations of the proposed method.

6.4.1. Successful Detection of Overlooked Metaphors

Some instances classified as false positives were revealed to be genuine metaphors that had been overlooked during manual annotation. The following examples illustrate cases where the model successfully detected such subtle metaphors.

In *Konjaku Monogatari-shu*, the expression 音 "sound" was identified as a metaphor by the model despite being missing during annotation:

> (1) 聖人は，誓ひを発して，<u>音</u>を<u>挙て</u>法花経を読奉る．(Classical Japanese)
> (*Hijiri wa, chikai wo okoshite, <u>oto</u> wo <u>agete</u>, hokkekyō wo yomi-tatematsuru*)
> 'A saint made a vow and, raising his voice, read the Lotus Sutra in reverence.'

Here, 音 (*oto*, "sound") is used to represent 声 (*koe*, "voice"), where the superordinate category stands for a specific member, constituting 提喩 "synecdoche." Furthermore, the verb 挙て (*agete*, "to raise") is also metaphorical: its prototypical sense involves physically lifting a concrete object, whereas here it is applied to sound/voice (a perceptual rather than manipulable entity), constituting a cross-domain mapping annotated as 換喩 "metonymy" under the CHJ-Metaphor scheme. Such synecdoche and metonymy are particularly difficult to notice during manual annotation due to their subtle nature, yet they occur frequently in classical texts.

Similarly, 入滅 "entering nirvana" was detected as a metaphor in the following passage:

> (2) 専ら<u>入滅</u>の日を思て此の会に参り<u>値は</u>ば，罪業を滅して浄土に生れむ事を疑ひ有じ．(Classical Japanese)
> (*Moppara <u>nyūmetsu</u> no hi wo omoite kono-e ni mairi <u>atawaba</u>, zaigō wo messhite jōdo ni umaren koto wo utagahi araji*)
> 'If one attends this gathering with the sole intention of the day of entering nirvana, one would have no doubt about being reborn in the Pure Land after extinguishing one's karmic sins.'

The expression 入滅 (*nyūmetsu*, "entering nirvana") encodes a virtual event script where *dying* is expressed as *entering nirvana*, constituting 換喩 "metonymy" based on event-level contiguity.

These examples demonstrate that the model is capable of detecting subtle metaphors that are easily overlooked during manual annotation, particularly 提喩 "synecdoche," 換喩 "metonymy," and conventionalized metaphors. From a language resource construction perspective, this suggests that the system can serve as an effective supplement to human annotation efforts.

6.4.2. Limitations in Metaphor Detection

However, the analysis also reveals limitations, particularly in handling 結合比喩 "word-level metaphors." In *Hojoki*, the following sentence contains the word-level metaphor 浮雲の思ひ "thoughts like floating clouds":

> (3) 古京はすでに荒れて，新都はいまだ成らず．ありとしある人は皆<u>浮雲</u>の<u>思ひ</u>をなせり．(Classical Japanese)

(*Kokyō wa sudeni arete, shinto wa imada narazu. Aritoshi aru hito wa mina ukigumo no omoi wo naseri*.)
'The ancient capital has already fallen to ruin, and the new capital is not yet complete; all people have minds like floating clouds.'

Both 浮雲 (*ukigumo*, "floating clouds") and 思ひ (*omoi*, "thoughts") (underlined in romanization) should be identified as metaphorical, forming a 結合比喩 "word-level metaphor." However, the model failed to detect these core metaphorical tokens while incorrectly marking the surrounding tokens 皆 (*mina*, "all") and なせり (*naseri*, "made/have") as metaphorical. This suggests that the model struggles with identifying the precise token boundaries of extended metaphorical expressions within complex 結合比喩 "word-level metaphors."

Hojoki is particularly rich in 文脈比喩 "discourse-level metaphors" and extended 指標比喩 "similes," leading to mismatches between annotators'and the model's decisions regarding which tokens within a metaphorical passage should be marked. Nonetheless, the model does not appear to miss entire metaphorical sentences or clauses.

6.5. Future Work

Several directions remain for future research.

First, dataset expansion is necessary. This study uses three currently completed works from CHJ-Metaphor (*Konjaku Monogatari-shu*, *Hojoki*, *Toraakira-bon Kyogen-shu*), but the corpus includes additional works under annotation (*Uji Shui Monogatari*, *Jikkinsho*, *Tsurezuregusa*). Once annotation is completed, a more comprehensive evaluation will enable detailed verification of the method's generalizability.

Second, cross-period generalization requires investigation. This study focuses on works from the medieval period, but the applicability to other historical periods remains unexplored.

7. Acknowledgements

This work was supported by JSPS KAKENHI Grants JP22K12145 and JP25K00459, the NINJAL Collaborative Research Projects "Empirical Computational Psycholinguistics Using Annotated Data" and "Extending the Diachronic Corpus through an Open Co-construction Environment", the Kayamori Foundation of Informational Science Advancement Research Grant "Extraction of Conceptual Metaphors Using Natural Language Processing," and the Mitsubishi Foundation Research Grant. We are also grateful to Professor Makoto Yamazaki and Professor Wakako Kashino for providing us with the basic sense data of the WLSP.

8. Bibliographical References

Soma Asada, Kanako Komiya, and Masayuki Asahara. 2024. Comprehensive assignment of Word List by Semantic Principles sense numbers 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. In Japanese.

Masayuki Asahara, Nao Ikegami, Tai Suzuki, Taro Ichimura, Asuko Kondo, Sachi Kato, and Makoto Yamazaki. 2022. CHJ-WLSP: Annotation of 'word list by semantic principles' labels for the corpus of historical Japanese. In *Proceedings of the Second Workshop on Language Technologies for Historical and Ancient Languages*, pages 31–37, Marseille, France. European Language Resources Association.

Mateusz Babieno, Masashi Takeshita, Dusan Radisavljevic, Rafal Rzepka, and Kenji Araki. 2022. MIss RoBERTa WiLDe: Metaphor identification using masked language model with Wiktionary lexical definitions. *Applied Sciences*, 12(4):2081.

Minjin Choi, Sunkyung Lee, Eunseong Choi, Heesoo Park, Junhyuk Lee, Dongwon Lee, and Jongwuk Lee. 2021. MelBERT: Metaphor detection via contextualized late interaction using metaphorical identification theories. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1763–1773, Online. Association for Computational Linguistics.

Ellen Dodge, Jisup Hong, and Elise Stickles. 2015. MetaNet: Deep semantic automatic metaphor analysis. In *Proceedings of the Third Workshop on Metaphor in NLP*, pages 40–49, Denver, Colorado. Association for Computational Linguistics.

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

Anna Hülsing and Sabine Schulte Im Walde. 2024. Cross-lingual metaphor detection for low-resource languages. In *Proceedings of the 4th Workshop on Figurative Language Processing (FigLang 2024)*, pages 22–34, Mexico City, Mexico (Hybrid). 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*, pages 1427–1432. In Japanese.

Beata Beigman Klebanov, Ben Leong, Michael Heilman, and Michael Flor. 2014. Different texts, same metaphors: Unigrams and beyond. In *Proceedings of the Second Workshop on Metaphor in NLP*, pages 11–17, Baltimore, MD. Association for Computational Linguistics.

Yucheng Li, Shun Wang, Chenghua Lin, and Frank Guerin. 2023a. Metaphor detection via explicit basic meanings modelling. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 91–100, Toronto, Canada. Association for Computational Linguistics.

Yucheng Li, Shun Wang, Chenghua Lin, Frank Guerin, and Loic Barrault. 2023b. FrameBERT: Conceptual metaphor detection with frame embedding learning. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 1558–1563, Dubrovnik, Croatia. Association for Computational Linguistics.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A robustly optimized BERT pretraining approach. *arXiv preprint arXiv:1907.11692*.

Rui Mao, Chenghua Lin, and Frank Guerin. 2019. End-to-end sequential metaphor identification inspired by linguistic theories. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3888–3898, Florence, Italy. Association for Computational Linguistics.

Akira Nakamura. 1977. *Theory and Classification of Figurative Expressions*. Number 57 in Report of the National Institute for Japanese Language and Linguistics. Shuei Shuppan, Tokyo. In Japanese.

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.

Xiaoyu Tong, Ekaterina Shutova, and Martha Lewis. 2021. Recent advances in neural metaphor processing: A linguistic, cognitive and social perspective. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4673–4686, Online. Association for Computational Linguistics.

Makoto Yamazaki and Wakako Kashino. 2017. Annotation of prototypical sense information for polysemous words in the Word List by Semantic Principles. *Journal of Natural Language Processing*, 24(3):111–126. In Japanese.

Makoto Yamazaki, Wakako Kashino, Kiyoko Uchiyama, Kazuko Sunaoka, Ikuto Tajima, Keiko Yamamoto, Yu-Seok Han, and Yun-Soo Seol. 2015. Annotation of basic sense determination information to the Word List by Semantic Principles, revised and expanded edition. In *Proceedings of the 58th Conference of the Mathematical Linguistics Society*. In Japanese.

Shenglong Zhang and Ying Liu. 2022. Metaphor detection via linguistics enhanced siamese network. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 4149–4159, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

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

9. Language Resource References

Baker, Collin F. and Fillmore, Charles J. and Lowe, John B. 2024. *FrameNet*. International Computer Science Institute.

Maekawa, Kikuo and Yamazaki, Makoto and Ogiso, Takehiko and Maruyama, Masaya and Ogura, Hideki and Kashino, Wakako and Koiso, Hiroshi and Yamaguchi, Makoto and Tanaka, Masahiro and Den, Yasuharu. 2014. *Balanced Corpus of Contemporary Written Japanese*. National Institute for Japanese Language and Linguistics.

National Institute for Japanese Language and Linguistics. 1964. *Word List by Semantic Principles*. Shuei Shuppan.

National Institute for Japanese Language and Linguistics. 2023. *Corpus of Historical Japanese*. National Institute for Japanese Language and Linguistics.