

分類語彙表に基づく日本語古文比喩表現の自動検出

Hang Zhu¹ 小原みと希² 菊地礼³古宮嘉那子¹ 浅原正幸⁴ 加藤祥⁵¹ 東京農工大学 ² 中央大学 ³ 長野工業高等専門学校⁴ 国立国語研究所・総合研究大学院大学 ⁵ 北海道大学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

概要

本研究では、日本語歴史コーパス (CHJ) における古文比喩表現の自動検出手法を提案する。現代日本語を対象とした比喩抽出の研究は存在するが、古文における比喩表現検出は語彙や文体の変化により困難である。我々は、分類語彙表 (WLSP) の意味分類情報と多層 Token Type IDs を活用した BERT ベースの手法を古文データに適用し、5 分割交差検証により F1 スコア 82.18% を達成した。実験結果は、現代語で有効な手法が古文比喩検出にも適用可能であることを示している。

1 はじめに

比喩表現は言語理解において重要な役割を果たし、その自動検出は自然言語処理の重要な課題である [1]。近年、BERT 等の Transformer ベースモデルを用いた比喩検出研究が進展しているが [2, 3]、その多くは現代語を対象としている。

一方、日本語における比喩表現研究では、加藤ら [4] によって現代日本語均衡コーパス (BCCWJ) に基づく BCCWJ-Metaphor が構築され、MIP (Metaphor Identification Procedure) [1] に基づく体系的な比喩アノテーションが行われている。さらに、我々の先行研究 [5] では、分類語彙表 (WLSP) [6] の意味分類情報を活用することで、現代日本語の比喩検出において 75.07% の F1 スコアを達成した。

しかし、日本語古文における比喩表現の自動検出については、十分な研究が行われていない。古文は現代語と比較して語彙や文法が異なり、また比喩表現の使用パターンも時代により変化している可能性がある。本研究では、現代語で有効性が示された WLSP 基盤手法を日本語古文に適用し、その有効性を検証する。

2 関連研究

2.1 比喩検出手法

比喩検出手法は、辞書・ルールベース [7]、統計的手法 [8]、深層学習ベース [9] の 3 つに分類される。近年は Transformer ベースモデルを用いた研究が主流となっており [10]、様々なアプローチが提案されている。

MelBERT [2] は RoBERTa [11] ベースで目標語埋め込みと文脈埋め込みを比較する手法を採用した。FrameBERT [12] は FrameNet [13] の意味フレーム情報を組み込んだ。BasicBERT [3] は訓練データから非比喩の用例を用いて基本義を表現する。その他にも、Wiktionary の定義を利用する Miss RoBERTa WiLDe [14] や、言語学的知識を強化した MisNet [15] などが提案されている。

我々の手法は WLSP の体系的な意味分類を活用し、理論的に根拠のある基本義決定を行う点でこれらと異なる。

2.2 日本語歴史コーパスと比喩研究

日本語歴史コーパス (CHJ) [16] は奈良時代から明治時代までの多様なジャンルのテキストを収録している。浅原ら [17] は、CHJ に対して分類語彙表の分類番号を付与した CHJ-WLSP を構築した。菊地らは、この CHJ-WLSP に対して比喩表現のアノテーションを施した CHJ-Metaphor の構築を進めている¹⁾。本研究では、現時点で比喩アノテーションが完了している作品を用いて、古文における比喩表現の自動検出を行う。

1) 本研究では、構築途中のコーパスのアノテーションが終わった部分を利用している。

3 データセット

3.1 CHJ-Metaphor

CHJ-Metaphor は日本の古文における比喩表現のデータセットであり、『分類語彙表』分類番号を付与した日本語歴史コーパス (CHJ-WLSP) [17] に対して、比喩に関するアノテーションを施したものである。収録予定文章は中世 (鎌倉・室町) の説話 (『今昔物語集 (本朝部) (1120 年以降成立)』『宇治拾遺物語 (1213~1221 年ごろ成立)』『十訓抄 (1252 年成立)』), 随筆 (『方丈記 (1212 年成立)』『徒然草 (1330~1331 年ごろ成立)』), 狂言 (『虎明本狂言集 (1642 年成立)』) である。

比喩の同定は現代日本語の比喩表現のデータセットである BCCWJ-Metaphor と同様, MIP およびその拡張版である MIP-VU[18] に加え, 中村 [19] の比喩種別の判定を採用した。これに基づき, 日本古典文学を専門とする者がアノテーションを実施している。

本研究では, 現時点で比喩アノテーションが完了している以下の 3 作品を用いた。

- 今昔物語集 (本朝部) (1120 年以降成立): 79,892 例
- 方丈記 (1212 年成立): 2,668 例
- 虎明本狂言集 (1642 年成立): 2,542 例

表 1 にデータセットの統計情報を示す。全体で 85,102 例を含み, 比喩率 (比喩表現に利用されるトークンの数/全表現のトークンの数) は作品により大きく異なる。今昔物語集は最も多くのデータを含むが, 比喩率は 4.41% と低く, 方丈記は 25.82% と最も高い比喩率を示す。

表 1 CHJ-Metaphor データセット統計

作品名	例数	比喩数	比喩率 (%)
今昔物語集	79,892	3,524	4.41
方丈記	2,668	689	25.82
虎明本狂言集	2,542	325	12.79
合計	85,102	4,538	5.33

4 提案手法

本研究では, 現代日本語比喩検出で有効性が示された WLSP 基盤の BERT 手法 [5] を日本語古文に適用する。手法の概要を以下に述べる。

4.1 基本義決定

多義語の基本義決定は MIP 適用における根本的な課題である。我々の手法では, 山崎ら [?, ?] により WLSP に付与された基本義情報を活用する。

山崎らは WLSP の各語彙素の各意味に対して, 基本義 (代表義) としての信頼度を 5 段階で付与した。目標語 w (語彙素 l) が分類番号 c_{context} の文脈で出現する場合, この信頼度に基づき基本義を決定する。具体的には, 語彙素 l の WLSP における全ての意味候補の中から, 最高スコア (最も高い信頼度) を持つ分類番号の集合 C_{top} を定義する。 $c_{\text{context}} \in C_{\text{top}}$ であればそれを選択, そうでなければ C_{top} からランダムに選択する。WLSP に候補がない場合は $c_{\text{proto}} = c_{\text{context}}$ とする。

4.2 基本義用例検索

決定した基本義分類番号 c_{proto} を用いて, BCCWJ-WLSP から基本義用例を検索する。語彙素 l と分類番号 c_{proto} が一致する全用例を抽出し, 複数存在する場合はランダムに選択する。用例が存在しない場合は, 語彙素 l 自体を使用する。

4.3 BERT 比喩検出モデル

目標語 w_l を含む文 $S = \{w_1, \dots, w_n\}$ に対し, まず WLSP の意味分類情報を追加する。

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

ここで $\{f_i\}$ は WLSP の意味分類情報 (類・部門・中項目・分類項目) である。情報が見つからない場合は [MASK] を使用する。

多層 Token Type IDs システムを用いて, 拡張系列内の異なる情報タイプを区別する。各トークンには役割が割り当てられる: 目標語, 局所文脈 (目標語周辺の句読点境界内の語), 意味素性 (WLSP の意味分類情報), 背景 (その他のトークン)。

日本語 BERT で目標文 S' と基本義用例 $P = u_{\text{proto}}$ を符号化した後,

$$\mathbf{v}_{S',1}, \dots, \mathbf{v}_{S',n} = \text{BERT}(S') \quad (2)$$

$$\mathbf{v}_{P,1}, \dots, \mathbf{v}_{P,m} = \text{BERT}(P) \quad (3)$$

文脈の意味と基本義の意味関係を捉えるベクトル \mathbf{h}_{MIP} を計算する。

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

ここで $f(\cdot)$ は線形変換である。交差エントロピー損失により訓練を行う。

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

4.4 実験設定

モデルは BERT[20] の日本語版 bert-base-japanese-v3²⁾ を使用し、最大系列長 512、バッチサイズ 16、学習率 $3e-5$ で 5 エポック訓練した。5 分割交差検証により性能を評価した。

評価指標として、F1 スコア、適合率、再現率を使用した。これらの指標は、CHJ-WLSP の形態素単位 (短単位) をトークンとし、すべてトークンベースで計算した。比喩クラスの不均衡 (約 5%) を考慮し、重み付き交差エントロピー損失を適用した。

5 実験結果

5.1 全体性能

表 2 に 5 分割交差検証の全体結果を示す。提案手法は F1 スコア 82.18%、適合率 82.83%、再現率 81.58% を達成した。

手法	F1	適合率	再現率
提案手法	82.18	82.83	81.58

この結果は、現代日本語 BCCWJ-Metaphor で達成した F1 スコア 75.07% を大きく上回っている。古文データの方が高い性能を示した理由として、まずデータサイズが大きく、特に今昔物語集の 79,892 例が安定した学習に寄与したことが挙げられる。また、古文の語彙や表現パターンが現代語より限定的で、モデルが学習しやすかったこと、さらに古文特有の比喩表現パターンの規則性が高かったことも性能向上に貢献したと考えられる。

5.2 作品別性能

表 1 に作品別のデータ統計を、表 3 に性能指標を示す。今昔物語集で最も高い正解率 (98.41%) を達成し、適合率 82.55%、再現率 81.21% と良好な性能

を示した。方丈記は正解率が 92.28% とやや低いものの、適合率 85.26%、再現率 84.76% と最もバランスの取れた性能を示した。虎明本狂言集は適合率 80.00%、再現率 78.76% であった。

表 3 作品別性能

作品名	正解率	適合率	再現率
今昔物語集	98.41	82.55	81.21
方丈記	92.28	85.26	84.76
虎明本狂言集	94.77	80.00	78.76

3 作品とも 80% 以上の適合率と再現率を達成し、Precision-Recall のバランスが良好である。今昔物語集は正解率 98.41% と最も高く、方丈記は適合率 85.26%、再現率 84.76% と最もバランスの取れた性能を示した。作品間の性能差の詳細な分析は考察で述べる。

6 考察

6.1 作品間の性能差

表 3 に示した作品間の性能差について、より詳細に考察する。

今昔物語集は正解率 98.41%、適合率 82.55%、再現率 81.21% と良好な性能を示した。データ量が最も多い (79,892 例) ことが学習の安定性に寄与し、比喩率が低い (4.41%) にもかかわらず、Precision-Recall のバランスが取れている。これは、豊富な学習データと本手法のクラス不均衡への対応が有効に機能したことを示唆している。

方丈記は適合率 85.26%、再現率 84.76% と 3 作品中で最もバランスの取れた性能を達成した。比喩率が 25.82% と高いことで、モデルが比喩パターンを十分に学習でき、クラス不均衡の影響を受けにくかったと考えられる。一方、正解率は 92.28% と相対的にやや低いが、これは比喩トークンの増加により、全体の正解率の計算において非比喩トークンの寄与が相対的に減少するためである。

虎明本狂言集は適合率 80.00%、再現率 78.76% と、データ量と比喩率の両面で中間的な特性を反映した性能を示した。全体として、比喩率 4.41% から 25.82% の範囲において、提案手法が安定した性能を維持することが確認された。

2) <https://huggingface.co/tohoku-nlp/bert-base-japanese-v3>

6.2 現代語との差異

古文データで現代語より高い性能が得られた理由として、複数の言語的特徴が考えられる。まず、古文の語彙は現代語より限定的で、多義性が低い可能性がある。また、古典文学の文体は定型表現が多く、比喩パターンも規則的である。さらに、専門家による高品質なアノテーションが性能に寄与したことも重要な要因として挙げられる。

6.3 検出結果の質的評価

実験後、データセット作成者により実験結果の分析を行った。その結果、誤検出と分類された事例の一部は、元のアノテーションの際に見逃されていた比喩であることが分かった。例えば、今昔物語集における「音」「年頃」「失せる」「入滅」「後」（時間概念メタファー）など、コーパス作成の際にアノテーションしていたが、アノテーション時に一部取りこぼしていた比喩を、提案手法が正しく検出できた事例が複数確認された。

この結果は、本手法が人手アノテーションを補完し、より包括的な比喩コーパス構築を支援できる可能性を示している。大規模コーパスにおける人手アノテーションでは一定の取りこぼしが不可避であることを考慮すると、本手法の実用的価値は高いと言える。

6.4 今後の課題

本研究にはいくつかの課題が残されている。

まず、データセットの拡張が必要である。本研究では現時点で比喩アノテーションが完了している3作品（今昔物語集、方丈記、虎明本狂言集）を用いたが、CHJ-Metaphorには他にも宇治拾遺物語、十訓抄、徒然草が収録予定である。これらの作品のアノテーションが完了次第、より包括的な評価を行い、手法の汎用性を詳細に検証する必要がある。

次に、時代間の汎化性能の検証が求められる。本研究は中世（鎌倉・室町）の作品を対象としたが、他の時代（上代・中古・近世等）への適用可能性は未検証である。時代横断的な評価により、手法の一般性を明らかにすることが重要である。

また、比喩タイプ別の詳細分析も今後の課題である。本研究では比喩全体の性能を評価したが、結合比喩、換喩、提喩、文脈比喩等、比喩タイプ別の検出困難性は明らかになっていない。各タイプの特性を

を分析することで、手法の改善点を特定できる。

さらに、誤検出や未検出の事例を詳細に分析し、モデルの弱点を明らかにすることで、さらなる改善が期待できる。

7 おわりに

本研究では、分類語彙表に基づくBERT手法を日本語古文比喩検出に適用し、現時点で比喩アノテーションが完了しているCHJ-Metaphorの3作品（今昔物語集、方丈記、虎明本狂言集）でF1スコア82.18%を達成した。この結果は、現代日本語で有効な手法が古文にも適用可能であることを示している。

今後は、CHJ-Metaphorの他の作品のアノテーション完了を待ち、より包括的な評価を行う。また、時代横断的な評価、比喩タイプ別の詳細分析、エラー分析を通じて、より精緻な古文比喩検出システムの構築を目指す。検出結果を用いた古典文学研究への応用も検討する。

謝辞

本研究はJSPS 科研費JP22K12145, JP25K00459, 及び国立国語研究所共同研究プロジェクト「アノテーションデータを用いた実証的計算心理言語学」の助成を受けたものです。

参考文献

- [1] Group Pragglejaz. Mip: A method for identifying metaphorically used words in discourse. **Metaphor and Symbol**, Vol. 22, No. 1, pp. 1–39, 2007.
- [2] Minjin Choi, Sunkyung Lee, Eunseong Choi, Heesoo Park, Junhyuk Lee, Dongwon Lee, and Jongwuk Lee. Melbert: Metaphor detection via contextualized late interaction using metaphorical identification theories. In **Proceedings of NAACL-HLT 2021**, pp. 1763–1773, 2021.
- [3] Yucheng Li, Shun Wang, Chenghua Lin, and Frank Guerin. Metaphor detection via explicit basic meanings modelling. In **Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics**, Vol. 2, pp. 91–100, 2023.
- [4] 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, pp. 1427–1432, 2022.
- [5] Zhu Hang and Kanako Komiya. Japanese metaphor detection using representation word meaning from bunrui-goi-hyou. In **Proceedings of the 39th Pacific Asia Conference on Language, Information and Computation**, 2025. to appear.
- [6] 国立国語研究所. 分類語彙表. 秀英出版, 1964.
- [7] Ellen Dodge, Jisup Hong, and Elise Stickles. Metanet: Deep semantic automatic metaphor analysis. In **Proceedings of the Third Workshop on Metaphor in NLP**, pp. 40–49, 2015.
- [8] Beata Beigman Klebanov, Ben Leong, Michael Heilman, and Michael Flor. Different texts, same metaphors: Unigrams and beyond. In **Proceedings of the Second Workshop on Metaphor in NLP**, pp. 11–17, 2014.
- [9] Rui Mao, Chenghua Lin, and Frank Guerin. End-to-end sequential metaphor identification inspired by linguistic theories. In **Proceedings of ACL 2019**, pp. 3888–3898, 2019.
- [10] Xiaoyu Tong, Ekaterina Shutova, and Martha Lewis. Recent advances in neural metaphor processing: A linguistic, cognitive and social perspective. In **Proceedings of NAACL-HLT 2021**, pp. 4673–4686, 2021.
- [11] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. **arXiv preprint arXiv:1907.11692**, 2019.
- [12] Yucheng Li, Shun Wang, Chenghua Lin, Frank Guerin, and Loic Barrault. Framebert: Conceptual metaphor detection with frame embedding learning. In **Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics**, pp. 1558–1563, 2023.
- [13] Collin F. Baker, Charles J. Fillmore, and John B. Lowe. The berkeley framenet project. In **Proceedings of ACL/COLING 1998**, Vol. 1, pp. 86–90, 1998.
- [14] Mateusz Babieno, Masashi Takeshita, Dusan Radisavljevic, Rafal Rzepka, and Kenji Araki. Miss roberta wilde: Metaphor identification using masked language model with wiktionary lexical definitions. Vol. 12, p. 2081, 2022.
- [15] Shenglong Zhang and Ying Liu. Metaphor detection via linguistics enhanced siamese network. In **Proceedings of the 29th International Conference on Computational Linguistics**, pp. 4149–4159, Gyeongju, Republic of Korea, 2022. International Committee on Computational Linguistics.
- [16] 国立国語研究所. 日本語歴史コーパス. <https://clrd.ninjal.ac.jp/chj/>, 2023.
- [17] Masayuki Asahara and Toshinobu Ogiso. Annotating the corpus of historical japanese with the word list by semantic principles. In **Proceedings of the Language Technologies for Historical and Ancient Languages Workshop (LT4HALA 2022)**, pp. 26–35, 2022.
- [18] Gerard J. Steen, Aletta G. Dorst, J. Berenike Herrmann, Anna A. Kaal, Tina Krennmayr, and Tom Pasma. **A Method for Linguistic Metaphor Identification: From MIP to MIPVU**. John Benjamins Publishing, 2010.
- [19] 中村明. 比喩表現の理論と分類. 国立国語研究所報告, No. 57. 秀英出版, 1977.
- [20] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In **Proceedings of NAACL-HLT 2019**, Vol. 1, pp. 4171–4186, 2019.