

# 結合比喩に対する字義・基本義のアノテーションと評価

岡 隆之介<sup>1</sup>

<sup>1</sup>追手門学院大学大学院

r-oka@otemon.ac.jp

## 概要

本研究では、『比喩表現の理論と分類』で報告されている結合比喩 (表現を構成している要素の関係が、文字が持つ慣用的な用法から著しくずれた比喩表現)を対象に、字義表現と基本義表現のそれぞれをアノテーションし、作成したデータセットに対して比喩表現検出モデルの性能評価を行った。著者のアノテーションにより、5,792件の結合比喩に対するアノテーションが完了した。また、5分割交差検定の結果、言語学的な比喩表現の抽出手法に動機づけられたBERTモデル (BasicBERT)が、そうでないBERTモデル (Sentence-BERT)よりも、高性能であることを確認した。

## 1 はじめに

私たちは、「担当者が連絡する」のように、表現を構成している要素の関係が、文字が持つありのままの意味を参照した場合に自然な表現である、字義表現を用いて物事を説明することがある。これに対して、「ごめんなさいねが連絡する」のように、表現を構成している要素の関係が、文字が持つ慣用的な用法から著しくずれた表現である、比喩表現を用いて物事を説明することもある。Cameron[1]のメタ分析によると、比喩表現は日常会話において1000単語あたりに50単語の割合で使用されており、字義表現と比較するとその登場頻度は少ないことが示唆されている。

言語学者はコーパスや発話中から比喩表現を検出する手法を提案している。特に、Pragglejaz Group [2]は、Metaphor Identification Procedure (以下、MIPと呼ぶ)と呼ばれる比喩表現の抽出手法を提案している。MIPは、文章全体を読み、語をすべて抽出し、各語の文脈の意味と基本的意味 (以下、基本義と呼ぶ)を調べ比較し、その文脈の意味が基本義と対照的であれば比喩表現として認定するという、4つの段階を踏んで比喩表現を特定する。MIPは手

続きが明確であるという点と、英語においては比喩表現に対してアノテーションがなされたデータセットが充実している (e.g., VUA18[3], VUA20[4])という点から、MIPの考えを計算機に実装した様々な比喩表現の検出手法が提案されている (e.g., MelBERT[5], FrameBERT[6], BasicBERT[7])。

一方で、日本語の比喩表現を対象にした比喩表現の検出に関する研究は、データセットの観点で不足している。現在のところ本邦で最も大規模な日本語を対象とした比喩表現のデータセットは加藤ら[8]のBCCWJ-Metaphorであるが、このデータは公開されていない。このような状況なため、日本語で利用可能な比喩表現の検出のモデルの構築は難しく、日本語におけるオープンな比喩表現のデータセットの公開が待たれている。

このような状況を受けて、本研究では、中村[9]の『比喩表現の理論と分類』を電子化した加藤・浅原[10]の結合比喩 (表現を構成している要素の関係が、文字が持つ慣用的な用法から著しくずれた比喩表現)を対象に、著者が字義表現と基本義表現のアノテーションを行う。そして、作成したデータセットの予備的評価として、MIPを用いた比喩表現検出モデル[7]と、MIPを用いず比喩表現と字義表現の分類を行うモデル (Sentence-BERT[11])で、モデルの精度を比較する。

本研究の貢献は、字義表現と基本義表現のアノテーション結果を公開する点 ([https://github.com/okaexp/metaphor\\_literal\\_basic\\_dataset](https://github.com/okaexp/metaphor_literal_basic_dataset))と、そのデータセットに対するMIPを用いたモデルの有効性を確認する点にある。

## 2 関連研究

MIPを用いた比喩検出モデルの中でも、MelBERT[5]はターゲット語 (比喩性を検出する単語)の埋め込み (例: 連絡する)、ターゲット語の文脈埋め込み (例: ごめんなさいねが連絡する; 下線部)、そして文脈埋め込み ([CLS]トークン)を結合し

て、線形層で比喩表現と字義表現を区別する手法が提案されている。FrameBERT[6]は、MeiBERTの埋め込みに対して、格フレームの情報を用いてファインチューニングすることで高精度化を図っている。最後に、BasicBERT[7]は、コーパス中の比喩表現ラベルがついていない(字義表現)ターゲット語の埋め込みの平均をターゲット語の基本義の意味とみなし、この基本義の意味とターゲット語の文脈埋め込みを結合して、比喩表現と字義表現を区別する手法が提案されている。これらの手法はいずれもMIPをベースとしているが、基本義を字義表現から抽出する手順が踏まれているという点で、BasicBERTは最もMIPに近い手順を踏んでいると考えられる。

日本語の比喩表現のデータセットである加藤ら[8]のBCCWJ-Metaphorについては、古宮らの一連の研究(Zhu et al[12], Ganbat et al[13])によって、BERTを用いた比喩表現の検出が検討されている。特にZhu et al[12]は、BCCWJ-Metaphorに付与された分類語彙表を用いて基本義を特定した上で、BasicBERTを適用し、5分割交差検証で結合比喩に対して平均0.813のprecisionを報告しており、結合比喩に対してBasicBERTを適用することの有効性を示している。

### 3 結合比喩に対する字義・基本義・ターゲット語のアノテーション

はじめに、本稿で新たに追加のアノテーションを行う。中村[9]および加藤・浅原[10]の結合比喩データセットについて紹介する。中村[9]に掲載された結合比喩の用例を電子ファイルに再編した加藤・浅原[10]は、結合比喩例文6,907文をMicrosoft Excel形式にして、付加情報(自然さ、わかりやすさ、古さ、新しさ、比喩性についてのクラウドワーカーの評価結果;それぞれ6件法)を付与した上でまとめている。それぞれの結合比喩は、「ごめんなさいねがどの道に連絡する」や「空気が緊張する」のように短く、文字数の平均値は7.24( $SD = 2.46$ )となっている。

本稿では、このデータセットに対して新たに、字義表現と基本義表現とターゲット語の追加を行うことを試みた。字義表現については、京都大学格フレーム検索(<https://lotus.kuce.kyoto-u.ac.jp/cf-search/>)を参考にしながら、著者が元の結合比喩例と同じ格を共有する別の字義的な項を選出することとした

(例:「彼が変色する」→「葉が変色する」)。一方で、基本義表現については、cl-tohoku/bert-base-japanese-v3(<https://huggingface.co/tohoku-nlp/bert-base-japanese-v3>)を用いて、項の部分を[MASK]トークンで置き換えた文を作成し(例:「足が動揺する」という比喩表現を元に、「[MASK]が動揺する」という表現を作成)、この[MASK]に入るトークンをfill-mask pipelineで埋めたものを参考にしながら、字義表現と同様に著者が元の結合比喩例と同じ格を共有する別の字義的な項を選出することとした(例:子供)。また、字義表現・基本義表現のどちらも、参考元の情報では適切な表現が見当たらない事例については、市販の辞書などを参考にしながら適切な表現に置き換えた。また、用例の中には適切な字義表現の用意が難しい事例(例:「男性への夢」「人生と私の間に金閣が立ちあらわれる」)も存在したため、これらの事例についてはアノテーション対象から除外した。最後に、ターゲット語は、主に格に相当する語とした。これにより、比喩表現と字義表現(例:「ごめんなさいねがどの道に連絡する」と「担当者がどの道に連絡する」)で判別すべきターゲット語(連絡する)を固定することを試みた。

以上の作業を通じて、最終的に、中村[9]および加藤・浅原[10]で報告された事例のうち、5,792事例について、字義表現と基本義表現の用例を追加した。比喩表現、字義表現、基本義表現、そしてターゲット語の例を表1にまとめた。

表1. 比喩表現, 字義表現, 基本義表現, ターゲット語の具体例

比喩	字義	基本義	格	ターゲット
正義感	感情	水	が	発露する
愛情	新作	東京	が	登場する
頭と頭	意見	両者	が	対立する
欲情	資格	体力	が	消滅する
空気	筋肉	体	が	緊張する
彼	葉	皮膚	が	変色する
毒	住民	人	が	運動する

注: 比喩表現は中村[9]および加藤・浅原 [10]で報告された事例

## 4 実験

本節では、3 節でまとめた結合比喩、字義表現、基本義表現、およびターゲット語のデータセットを用いて、BasicBERT[7]とそのアブレーションモデル (AMIP, SPV, BMIP) による比喩表現と字義表現の分類性能を評価する。また、選択制限違反 (SPV) や基本義からの乖離 (AMIP, BMIP) などの言語学的な知識を用いずに分類するベースラインモデルとして、Sentence-BERT[11]を採用し、モデルの分類性能を評価する。

### 4.1 モデル

**BasicBERT** BasicBERT[7]は BasicMIP, AMIP, そして SPV の 3 つのモデルの組み合わせによって表現される。いずれも cl-tohoku/bert-base-japanese-v2 を事前学習済み BERT モデルとして用いた。

BasicMIP は 3 段階からなる。はじめに、ターゲット語の文脈の意味を BERT モデルから獲得する。BasicMIP は文  $S = (w_1, \dots, w_t, \dots, w_n)$  が与えられ、 $w_t$  がターゲット語 (比喩性を判断する格) であるときに、 $H = BERT(emb_{cls}, \dots, emb_t, \dots, emb_n)$  によって文表象を得る ( $cls$  は文章の冒頭のトークン)。そして、 $H = (h_{cls}, \dots, h_t, \dots, h_n)$  が隠れ層の出力として与えられた時に、ターゲット語  $w_t$  の文脈的意味  $v_{S,t} = h_t$  で与えられるものとする。続いて、ターゲット語の基本義における意味を、字義表現と基本義表現のアノテーションデータセットをもとに獲得する。ターゲット語の基本義の文脈的意味は、 $H$  と同様に基本義表現を入力とした文表象からターゲット語が含まれる  $t$  番目の隠れ層の状態  $h_t$  から獲得する。BasicMIP の元論文では、ターゲット語の基本義における意味は、複数の字義表現ラベルが付与された文の平均から獲得されるが、本研究では表 1 に示すように基本義表現は比喩表現と字義表現に対して 1 つ用意されていたため、上述の  $h_t$  をそのまま基本義の文脈的意味として用いた。最後に、前の 2 段階で得られた 2 つの表象を比較することで比喩性を判定するため、BasicMIP に基づく隠れベクトル  $h_{BasicMIP}$  を、 $v_{S,t}$  と  $v_{B,t}$  を結合することによって求める。

AMIP は、非文脈化されたターゲット語の表象と文脈中のターゲット語の表象を比較することで比喩性を判定するため、 $h_{AMIP}$  は  $v_{F,t} = BERT(emb_t)$  (すなわち、ターゲット語  $w_t$  だけを入力した出力の隠

れ層の状態  $h_t$ ) と  $v_{S,t}$  を結合することによって求める。

SPV は、文全体の意味の表象と文脈中のターゲット語の表象を比較することで比喩性を判定するため、 $h_{SPV}$  は  $v_S = BERT(emb_{cls})$  と  $v_{S,t}$  を結合することによって求める。

最後に、BasicBERT は、 $h_{BasicMIP}$ ,  $h_{AMIP}$ , そして  $h_{SPV}$  の 3 つの隠れ層を結合することによって求められる。

それぞれのモデル (BasicBERT, AMIP, SPV, BasicMIP) は上述の手法で求めた隠れ層に線形層を通して最終的な 2 値分類を行う。出力層には `log_softmax` 関数を適用する。

**Sentence-BERT** Sentence-BERT[11]は cl-nagoya/sup-simcse-ja-base を事前学習済み Sentence-BERT モデルとして用いた。Sentence BERT は、比喩表現 (あるいは字義表現) とターゲット語をそれぞれ Sentence BERT に入力し、比喩表現 (あるいは字義表現) とターゲット語の隠れベクトル、それら 2 つのベクトルの差ベクトル、そしてそれら 2 つのベクトルの内積のそれぞれを結合したベクトルを特徴量として用いた。これによって、比喩表現とターゲット語の距離と、字義表現とターゲット語の距離がそれぞれ遠くなるようにモデルが学習することを意図した。

作成された特徴量は scikit-learn の MLPClassifier 関数を用いて分類モデルを学習した。層の数は 512, ニューロン数は 128, 活性化関数は `relu`, 最適化手法は `adam`, 学習の反復の最大回数は 20 回とした。

### 4.2 モデルの評価

モデルの評価には、3 節で紹介した比喩表現・字義表現・基本義表現のデータセットを 5 分割交差検証で評価した。BasicBERT とそのアブレーションモデルおよび Sentence-BERT のそれぞれで、各 Fold の訓練/テストデータが等しくなるようにデータを分割した。また、訓練データのバッチサイズは 32 で固定した。

また、それぞれの比喩表現、字義表現、基本義表現の格は句からなるものも含まれた (例: どの道に連絡する) が、ターゲット語は動詞 (連絡する) のみとした。また、ターゲット語の中には「消滅する」と「滅する」のように、サブワードの水準で共通するものも見られたため、各 Fold でサブワードが訓

練/テストで分かれなないように、文字列部分一致がある項目をサブワードとして事前にクラスタリングして、この問題を解決した。

最後に、モデルの評価は accuracy, precision, recall, F1 値の4つの指標で評価し、Foldの平均値を代表値としてまとめた。

## 5 結果と考察

モデルごとの5分割交差検証の評価結果 (accuracy, precision, recall, F1)の平均値と標準偏差を表2にまとめた。

表2. モデルごとの5分割交差検証の評価結果の平均値 (括弧内は標準偏差)

モデル	Acc.	Prec.	Rec.	F1.
Basic-BERT	<b>0.911</b> ( <b>0.005</b> )	0.919 (0.015)	<b>0.901</b> ( <b>0.014</b> )	<b>0.910</b> ( <b>0.005</b> )
AMIP	0.906 (0.010)	0.915 (0.010)	0.896 (0.024)	0.905 (0.011)
SPV	0.904 (0.005)	0.917 (0.008)	0.889 (0.015)	0.903 (0.006)
Basic-MIP	<b>0.911</b> ( <b>0.010</b> )	<b>0.931</b> ( <b>0.028</b> )	0.889 (0.031)	0.909 (0.011)
Sentenc e-BERT	0.856 (0.019)	0.882 (0.049)	0.829 (0.089)	0.850 (0.032)

表2の通り、precisionを除く全ての指標において、BasicBERTモデルが他のモデルの性能を上回った。また、precisionにおいてはBasicMIPの性能が他のモデルの指標を上回っていた。また、SPVやAMIPについても、Sentence-BERTよりも概して高い性能であることが確認された。

これらの結果は、結合比喩の比喩性判定において、選択制限違反 (SPV)や基本義からの逸脱 (BasicMIP, AMIP)といった、MIPにおける手続きを学習したモデルが、比喩表現と字義表現の格の使い方を区別するように学習したモデル (Sentence-BERT)よりも優れている可能性を示唆した。

また、本研究において、BasicBERTのprecisionの平均値は0.919であり、Zhu et al[12]で報告された0.813よりも高かった。これは、先行研究で使用されたBCCWJ-Metaphorにおけるデータセット (比喩/字義が一文内で混在しているデータセット)と比べて、本研究で扱っているデータセットは格と項だ

けからなる単純な課題となっているためであると考えた。

## 6 おわりに

本研究では、中村[9]の『比喩表現の理論と分類』を電子化した加藤・浅原[10]の結合比喩を対象に、著者が字義表現と基本義表現のアノテーションを5,792件の結合比喩に対して行い、作成したデータセットに対して、BasicBERTとSentence-BERTを適用した比喩表現と字義表現の分類モデルの評価を行った。実験の結果、MIPに基づくBasicBERTはMIPに基づかないSentence-BERTよりもすべての指標において性能が高かった。

今後の課題として、大きく2つの問題を取り上げる。1つ目は、アノテーターの拡充である。2章で説明した通り、本研究のアノテーションは全て、比喩表現について研究実績のある著者のみが行った。これは、字義表現と比喩表現の弁別は手続き (MIP)は明瞭であるものの、その分類自体は難しく (日本語比喩表現のアノテーションの難しさに対する同様の指摘として Ganbat et al[13]), 一般的なクラウドワーカーなどに作業依頼をすることは困難であると考えたための措置であった。一方で、著者の分類が信用できるものであるかについては改善の余地がある。今回のデータセットの公開を踏まえて、今後は他の言語学者などの追加アノテーションと、その分類一致率などの評価も必要になる。

もう1つの課題は、BasicBERTはどのような問題を解けて、どのような問題が解けなかったかの定量的な分析が必要となる。この問題に対しては、加藤・浅原[10]が付与した付加情報 (自然さ、わかりやすさ、古さ、新しさ、比喩性)が、BasicBERTが正しく分類できた事例とそうでない事例とどのように関係しているかを明らかにすることが有効であると考えられる。

## 謝辞

本研究は JSPS 科研費 JP25K21542 の助成を受けたものです。

## 参考文献

- [1] Lynne Cameron. **Metaphor in educational discourse**. London: Continuum, 2003.
- [2] Pragglejaz Group. MIP: A Method for Identifying Metaphorically Used Words in Discourse. **Metaphor and Symbol**, Vol. 22, pp.1-39, 2007.
- [3] Chee Wee Leong, Beata Beigman Klebanov, and Ekaterina Shutova. A report on the 2018 via metaphor detection shared task. **Proceedings of the Workshop on Figurative Language Processing**, pp. 56–66, 2018.
- [4] Chee Wee Leong, Beata Beigman Klebanov, Chris Hamill, Egon Stemle, Rutuja Ubale, and Xianyang Chen. A report on the 2020 via and toefl metaphor detection shared task. **Proceedings of the Second Workshop on Figurative Language Processing**, pp. 18–29, 2020.
- [5] 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. **Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**, pp. 1763–1773, 2021.
- [6] Yucheng Li, Shunyu Wang, Chenghua Lin, Frank Guerin, and Loïc Barrault. Framebert: Conceptual metaphor detection with frame embedding learning. **Conference of the European Chapter of the Association for Computational Linguistics**, pp. 1558–1563, 2023.
- [7] Yucheng Li, Shun Wang, Chenghua Lin, and Frank Guerin. Metaphor detection via explicit basic meanings modelling. **Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics**, Vol. 2, pp. 91–100, 2023.
- [8] 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, pp. 1427–1431, 2022.
- [9] 中村明. 比喩表現の理論と分類. 秀英出版, 1977.
- [10] 加藤祥, 浅原正幸. 『比喩表現の理論と分類』データの電子化および情報付与. 国立国語研究所論集, Vol. 25, pp. 1–19, 2023.
- [11] Reimers, Nils, Iryna Gurevych. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. **Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)**. pp. 3982–3992, 2019.
- [12] ZHU HANG, 古宮嘉那子, 浅原正幸. 分類語彙表の基本義を利用したメタファー検出. 言語処理学会 第 31 回年次大会 発表論文集, pp. 1590–1595, 2025.
- [13] Ganbat Naranbuuvei, 尾崎太亮, 古宮嘉那子, 浅原正幸. 系列ラベリングを用いた日本語の比喩表現抽出. 言語処理学会 第 31 回年次大会 発表論文集, pp. 1569–1573, 2025.