

日本語形容表現における多義性の検出と分析 Detection and Analysis of Polysemy in Japanese Adjectives

大澤 拓巳¹⁾ 寺岡 丈博¹⁾
Takumi Osawa Takehiro Teraoka

1 はじめに

1.1 背景と目的

日本語の形容詞・形容動詞は、日常の会話や文章の中で、物事の状態や性質、そして感情を表現するのに使用されている。そのような形容表現には、語義が一意ではないものが存在しており、活用の幅は多種多用である。加えて、時代の変化に伴い語義が増減しているものもあり、自動でこれらの語義の違いを高い精度で認識することは困難である。例の一つに形容表現の“適当だ”がある。この語は、“程よく当てはまる”という肯定的な語義として使われるのが普通であった。しかし、近年では“いい加減”という否定的な語義としても使用されるようになった。

- (1)彼の掃除は適当だから部屋が汚い。
- (2)その空欄に適当な語を埋める。

(1)の文章では、“適当だ”が“いい加減”といった語義で使用されており、「部屋が汚い理由は、彼の掃除のいい加減さのせい。」という意味になっている。(2)の文章では、“適当だ”が“程よく当てはまる”といった語義で使用されており、「空欄の中に正しく当てはまるような語を入れてください。」という意味になっているが、両者を区別するのは難しい。

このような語義曖昧性解消のタスクについては、これまで様々な研究がなされてきた。しかし、それらは事前に多義性がある単語のみを対象としているため、多義性の有無がわからない場合や、語義の増減があった場合には対応することができない。そこで本研究では、前提知識のない形容表現の多義性を自動で検出することにより、語義の増減があった場合にも語義曖昧性解消の対象となることを目的とした。また、多義性を持っていると判断した形容表現に対して、いくつかの分析を行うことによって、形容表現の語義の違いはどこから生じるものなのかを調べた。

1.2 ChatGPT について

近年、自然言語処理の分野を騒がせている GPT モデルを使用した ChatGPT は、あたかも人間と対話しているかのような文を生成でき、様々な分野において広く普及してきている。直近では GPT-4omni モデル (GPT-4o) が発表され、従来の GPT4 モデルよりも性能が向上し、ますます注目を浴びている。ChatGPT は文脈を正しく理解しているように見えるが、一方で本研究の対象である形容表現においてに多様な語義を区別・判断することができていない。下記に GPT-4o モデルを使用した ChatGPT が、区別できなかった例として“過酷だ”、“清楚だ”が挙げられる。両単語ともに多義性を持たない形容表現である。[1]しかし、ChatGPT は多義性を持つと判断し、付与した語義が、人間から見たら区別する必要のないものとなっていた。

- 過酷だ (厳しすぎるさま. ひどすぎるさま)
 1. 非常に厳しいさま (過酷な環境)
 2. 過酷な状況 (過酷な労働)
- 清楚だ (飾り気がなく、清らかなさま. また、そのさま.)
 1. 上品な (清楚な女性)
 2. 純粋な (清楚なイメージ)

以上から、大規模言語モデルでもハルシネーションが生じていることがわかる。

2 関連研究

梶澤ら [2] の研究は、話題に基づく分類辞書を作成し、文の話題を考慮したカテゴリを単語に付与することによって語義曖昧性解消を行っている。例えば“ドライバー”という対象単語に対して、“工具”“自動車”“ゴルフ”のように、話題に基づくカテゴリを付与することによりデータセットを構築し、精度が向上しているかどうかを検証した。この研究では、話題の情報が語義曖昧性解消の精度向上に貢献している事がわかった。

田淵ら [3] の研究は、教師なし学習による語義曖昧性解消の精度向上のため、日本語 WordNet で定義されている上位語と下位語の関係をを用いた素性の検討を行っている。この研究では、対象単語に対する上位語の分散表現、下位語の分散表現はいずれも類似すると考え、上位語と下位語を、教師なし学習による語義曖昧性解消の素性として有益であるかを検討するものである。結果としては、上位語および下位語が、素性として適している場合と適していない場合があり、どのような単語に有効であるかを判断できない問題点が生じていた。

橋口ら [4] は、WordNet に存在している上位下位関係を使った、従来の語義サイズの縮約手法に lexicographer カテゴリモデルを追加し、推定した lexicographer カテゴリで語義ラベルを置き換えることによって語義曖昧性解消の精度向上を目指したものである。lexicographer カテゴリとは、語義の種類を表す番号のことであり、WordNet の語義ラベルに付与されている。語義曖昧性解消を目指すタスクの多くは対象単語が名詞のみに絞られている場合が多いが、橋口らの研究では、対象単語が名詞、動詞、形容詞、副詞の4つの品詞についての研究を行なっている。実験では、語義ラベル推定モデルの学習に“古典的な埋め込み”を用いた場合と“BERT”を用いた場合の2種類で行い、それぞれの推定結果を比較し、lexicographer カテゴリ推定手法が有効かどうかを調べた。結果として、古典的な埋め込みを用いた場合は、名詞のスコアが従来の縮約手法と比べて 1.1 ポイント程増加していた。しかし、BERT を用いた場合、全体的にスコアが大きく減少していた。これは、lexicographer カテゴリ推定システムの精度が、システムを追加しない場合の語義推定より低いからであった。そのため、lexicographer カテゴリ推定の精度を上げることが改善点として挙げられた。

1) 拓殖大学大学院工学研究科, Graduate School of Engineering, Takushoku University

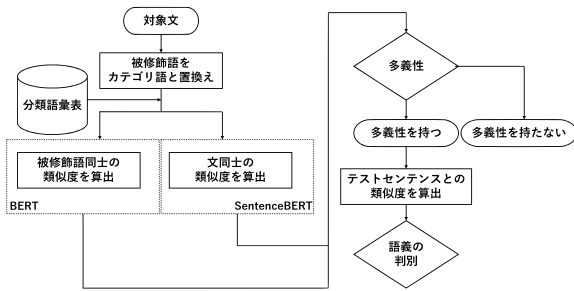


図 1: 手法の概要

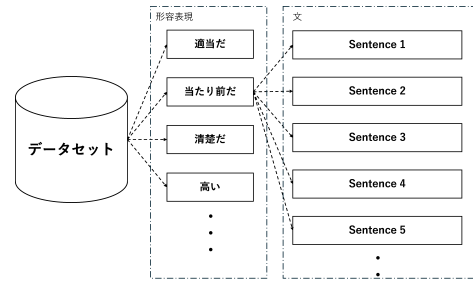


図 2: データセットの構成

3 提案手法

3.1 概要

はじめに、形容表現が多義性を持つかを判断する。その後、多義性を持つ形容表現だった場合、語義毎で類似度のばらつき、違いがあるかどうかを調査する。図 1 は提案手法の処理の流れを示している。

3.2 データセットの構築

本研究では、被修飾語の異なるカテゴリが、特定の形容詞表現に多義性をもたらすという仮定に基づいている。

- (3)あの山は高い。
- (4)あの財布は高い。

“高い”は“地面などの基準点よりも上の位置にある”や“値段がかさむさま”、“音声の振動数が多い”などの語義があるため多義性を持つといえる。上記の文章(3)と(4)の間には、被修飾語以外に違いはなく、非常に似ている文章である。しかし、(3)の“高い”は、“地面などの基準点よりも上の位置にある”の語義である。(4)では、“値段がかさむさま”として使用されている。被修飾語が“財布”の場合、前者の語義を適用できない。これは、“山”の場合でも同様である。そのため、これらの語義の違いは、被修飾語のカテゴリの違いにあると考えられる。

ここでは、形容表現が使われた文章について、被修飾語をカテゴリ語と置き換えることによってデータセットを構築する。形容表現が含まれた文章はブログ投稿サイトである“はてなブログ”¹⁾から抽出を行った。カテゴリ語は国立国語研究所の分類語彙表[5]を使用した。分類語彙表を構成する項目は見出し番号、レコード種別、類、中項目、見出しなどがあるが、本研究では中項目に記載されている単語をカテゴリ語とした。中項目には“言語”、“空間”、“土地”など全 49 種類存在している。

- (5)友人と合流し、適当な店へ。
- (6)友人と合流し、適当な社会へ。

(5)が元の文章であり、“適当だ”の被修飾語である“店”が、分類語彙表の中項目で“社会”に属していたため、(6)のような置換えが起こる。また、分類語彙表に記載されていない単語に対しては、Word2Vec を利用し単語の意味をベクトル化させる。そのようにして得たベクトル表現を使って単語間のコサイン類似度を算出させる。カテゴリ語の中で最もコサイン類似度の数値が高いものをカテゴリ語として扱う。

データセットには、図 2 のように形容表現が 72 個あり、多義性を持つものと多義性を持たないものが 36 語

ずつ存在する。また、形容表現 1 語につき、その語を含む文が複数あるデータセットとなっている。

3.3 多義性の検出

本研究では、特定の形容表現が文中で使用されている場合、被修飾語同士の類似度が低いのは、対象単語が別の語義として使用されていると仮定した。ここでは、データセット内の文に対して、被修飾語同士の類似度、文同士の類似度を総当たりで算出する。その数値を使用することによって、1つの形容表現に対する類似度のばらつきを調べ、多義性を検出する。

4 評価実験

4.1 被修飾語による類似度の変化

本研究では、被修飾語のカテゴリの違いが多義性をもたらしている、と仮定して実験を行ったため、その仮説が正しいかどうかを検証する必要がある。そのため、BERT を使用して、任意の形容表現が使用された文に存在する被修飾語同士の類似度と、SentenceBERT を利用した文同士の類似度をそれぞれ算出し、ヒートマップにまとめることで可視し、それらの変化を明らかにする。

4.2 多義性の検出

3.2 で作成したデータセットに、多義性を持つ場合は 1、多義性を持たない場合は 0 のラベルを付与した。その後、SVM でモデルを構築することで評価をした。SVM の素性には、被修飾語同士または文同士の類似度の平均、最小値を使用した。被修飾語同士の類似度を求める際には、BERT[6] を使用し、文同士の類似度を求める際には、SentenceBERT[7] と BERTScore[8] の両方を使用した。BERT は、東北大学の事前学習済みモデルを利用した。SentenceBERT は、BERT と同様、東北大学の事前学習済みモデルに加えて、Laba ら [9] の研究にて、文脈埋め込みの精度が最も高かった多言語モデルである paraphrase-multilingual-mpnet-base-v2(PMMBV2) を利用した。いずれも最も精度の高い mean-pooling を利用した。また、LLM を使用したベースラインとして GPT-4o を用いたものも用意した。提案手法は、被修飾語に着目した BERT に加えて、文全体を考慮した BERTScore、SentenceBERT を素性としたものである。多義性の有無はデジタル大辞泉 [1]²⁾ を使用し、複数の語義があるものに対して多義性を持つと判断する。

4.3 語義の異なり

多義性を持つ形容表現は、異なる語義で使用される。そのため、3.3 で検出された形容表現について、語義毎で被修飾語同士の類似度にばらつきが生じるかどうかを調べる。任意の形容表現について、異なる語義毎にモデルセンテンスを作成し、モデルセンテンスとデータセッ

1) <https://hatenablog.com/>

2) <https://japanknowledge.com/contents/daijisen/>

表 1: 評価結果

	正解率	適合率	再現率	F 値
SentenceBERT(PMMBV2)	0.57	0.54	0.97	0.69
SentenceBERT(東北)	0.74	0.71	0.81	0.75
BERT	0.76	0.72	0.86	0.79
GPT-4o	0.50	1.00	0.50	0.73
BERTScore + BERT	0.79	0.75	0.92	0.83
SentenceBERT + BERT	0.79	0.80	0.86	0.83
提案手法	0.82	0.77	0.92	0.84

ト内にある文の被修飾語の類似度を総当たりで算出し、その値をストリッププロットで可視化する。

5 結果

5.1 被修飾語による類似度の変化

多義性を持つ“当たり前だ”と多義性を持たない“巨大だ”の、被修飾語をカテゴリ語に置換えた時の BERT を使用した被修飾語同士の類似度を図 3, SentenceBERT の PMMBV2 を使用した文同士の類似度を図 4, 東北大学の事前学習済みモデルを使用した文同士の類似度を図 5 に示す。

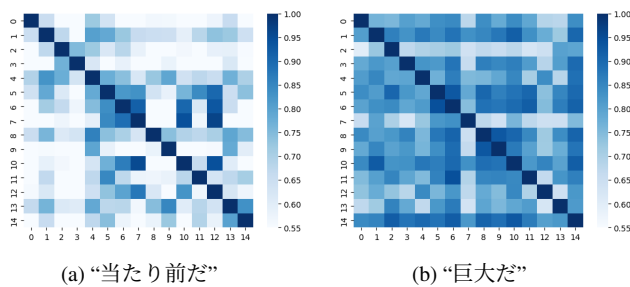


図 3: BERT: 被修飾語同士の類似度

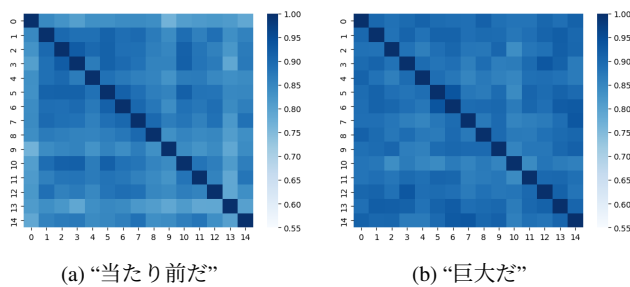


図 4: SentenceBERT(東北): 文同士の類似度

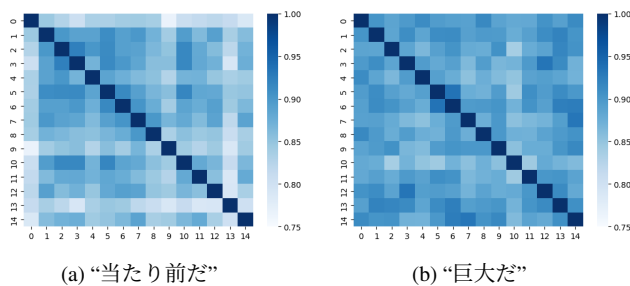


図 5: SentenceBERT(PMMBV2): 文同士の類似度

5.2 多義性の検出

多義性の検出結果を表 1 に示す。SentenceBERT を使用したものよりも、BERT 単体で被修飾語同士の類似度を使用した方が精度が高い。GPT-4o では、多義性を持つものに対しては正しく判断できるが、多義性を持たないものについて検出することができていなかった。

5.3 語義の異なり

形容表現“痛い”の例を図 6 に示す。“痛い”には、4 つの語義が存在する。それぞれの語義に対するモデルセンテンスと、データセット内にて形容表現“痛い”が使用されている文にて、被修飾語の類似度を総当たりで算出した。

- A. 肉体に痛みや苦しみを感ずるさま
- B. 心に苦痛を感じるさま
- C. 弱点を攻撃されたり、打撃や被害を被ったりして、閉口するさま
- D. さも得意そうな言動がひどく場違いで、見るに堪えないさま

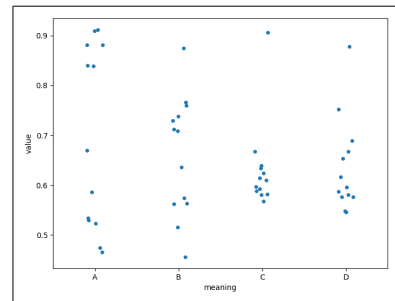


図 6: 形容表現“痛い”に対する語義の異なり

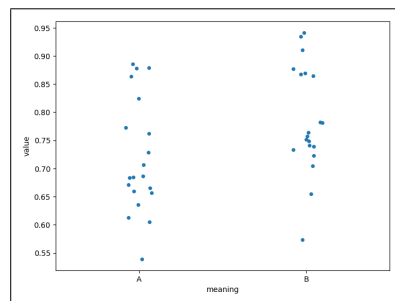


図 7: 形容表現“適当だ”に対する語義の異なり

次に、形容表現“適当だ”の例を図 7 に示す。1.1 で提示した通り、“適当だ”には肯定的な語義と否定的な語義の両方が存在する。図 7 の A は“適当だ”が肯定的な語義として使用されている文をモデルセンテンスとし、

被修飾語同士の類似度をプロットしたものである。B は否定的な語義をモデルセンテンスとしたものを使用している。

6 考察

6.1 被修飾語による類似度の変化

5.1 で示したように、多義性を持つ形容表現においては、被修飾語同士の類似度の散らばりが大きくなっている。本研究の場合は (a) が多義性を持ち、(b) は多義性を持たない形容表現であった。図 3 を見ると、多義性を持つ形容表現と持たない形容表現の間には、大きな差が存在しており、この差が形容表現の多義をもたらすと考えられる。しかしながら、SentenceBERT を使用した図 4, 5 を見ると、両者の間に大きな差が見られず、多義性を持つ (a) の“当たり前だ”でも、類似度が大きく変わる箇所は存在していなかった。東北大学の事前学習済みモデルを使用した SentenceBERT では、文全体の類似度を算出するため、同じ形容表現が使用されていたら、語義に関係なく類似度が高くなる傾向があった。また、PMMBv2 を使用した SentenceBERT では、日本語に特化したモデルではないため、全ての文に対して類似度が低くなり、多義性の有無についての判断が難しいものになっていた。

以上のことから、形容表現の多義性を判断するには、文全体に着目するのではなく、被修飾語の違いに視線を向けるべきであることがわかる。

6.2 多義性の検出

多義性の検出では、Laba らの研究で最も精度の高かった PMMBv2 モデルを使用したものの精度が最も低く、日本語の事前学習済みモデルである東北大学の SentenceBERT も単体で見ると、BERT を使用したモデルより精度が低いことが分かった。そのため、6.1 のヒートマップを見て述べたのと同様に、形容表現の多義性は文全体を見るのではなく、被修飾語の違いを見ることが有効であるといえる。また、以前の研究 [10] の手法である BERT+BERTScore モデルより、本研究で提案した手法の方が僅かに優れていた。

大規模言語モデルである GPT-4o ではハルシネーションが起こっており、多義性を持たない形容表現に関して、全て多義性を持つと判断されていた。本評価実験では多義性を持つ形容表現と持たない形容表現の数を等しくしていたが、実際は多義性を持たない形容表現の方が多い。そのため、全ての形容表現に対して多義性を持つと判断するきらいがあるため、実際は本評価実験よりも GPT-4o の F 値が下がることが考えられる。本研究の手法では、GPT-4o が誤って検出した一部の形容表現について、多義性を持たないと正しく判定することが可能であったため、本手法の方が多義性の検出については優れているといえる。

6.3 語義の異なり

形容表現“痛い”において、図 6 から、それぞれのモデルセンテンスとの類似度が上位のものと同位のものではっきりと分かれているため、被修飾語のカテゴリの違いで語義が見分けられるといえる。しかし、“頭が痛い”や“胸が痛い”などの、痛む対象は体の一部であるが、B の語義として使用されている慣用語について、本研究で

は A の語義として分けられてしまっていた。

形容表現“適当だ”において、図 7 の A を見ると、value が 0.80 以上と高い文については、A のモデルセンテンスで使われている被修飾語のカテゴリと近いものであるから、使用されている語義も同じものであると考えられる。また、B においても同様であり、value が 0.80 以上であった文の形容表現は、B のモデルセンテンスで使用されている語義と同じものである。また、A の語義は肯定的な語義として使用されているため、A のモデルセンテンスに対して上部にプロットされた文の形容表現“適当だ”は、正の極性を自動で付与することが可能である。一方で、B の語義は、否定的な語義で使用されているため、上部にプロットされた文の形容表現は負の極性を自動で付与することが出来る。

7 今後のまとめ

本研究では、日本語形容表現の語義曖昧性解消を行うための手掛かりとして、多義性を検出し、語義の異なりについて調べた。結果として、形容表現の多義は文の異なりから生じるものではなく、被修飾語のカテゴリの違いから生じることがわかった。また、肯定的用法と否定的用法の両方で使用される形容表現に対しても、同様の処理で正しい極性を自動で付与することが可能となった。今後は、本研究で検出された形容表現に対して語義曖昧性解消を行う。また、正しい極性を自動で付与できることから、感情分析やリコメンドに応用することで、精度の向上が期待できる。

8 謝辞

本研究は JSPS 科研費 JP22K00646 の助成を受けたものです。

参考文献

- [1] デジタル大辞泉. 小学館.
- [2] 榎澤優希, 山本和英. 話題に基づく語義曖昧性解消. 言語処理学会第 24 回年次大会 発表論文, pp. 248–251, 2018.
- [3] 田淵知明, 大沢英一. 教師なし学習による語義曖昧性解消の素性検討. 人工知能学会全国大会論文集第 36 回 (2022), pp. 2B5GS601–2B5GS601. 一般社団法人人工知能学会, 2022.
- [4] 橋口卓弥, 佐々木稔. Wordnet lexicographer カテゴリ推定による語義サイズ縮約を用いた語義曖昧性解消. 言語処理学会第 29 回年次大会発表論文集, pp. 1038–1042, 2023.
- [5] NINJAL(2004). 『分類語彙表増補改訂版データベース』(ver1.0), 2004.
- [6] Jacob Devlin Ming-Wei Chang Kenton and Lee Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of naacL-HLT*, Vol. 1, p. 2, 2019.
- [7] Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*, 2019.
- [8] Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. Bertscore: Evaluating text generation with bert. *ICLR*, 2020.
- [9] Yurii Laba, Volodymyr Mudryi, Dmytro Chaplynskyi, Mariana Romanyshyn, and Oles Dobosevych. Contextual embeddings for ukrainian: A large language model approach to word sense disambiguation. In *Proceedings of the Second Ukrainian Natural Language Processing Workshop (UNLP)*, pp. 11–19, 2023.
- [10] 大澤拓巳, 寺岡丈博. 曖昧性を持つ形容表現の極性変化. 情報処理学会第 86 回年次大会 講演論文, pp. 453–454, 2024.