

# 直喩表現を利用した名詞の特徴抽出

佐藤 遼河<sup>†</sup>, 杉本 徹<sup>‡</sup>

芝浦工業大大学院 理工学研究科<sup>†</sup>, 芝浦工業大学 工学部<sup>‡</sup>

{ma18049, sugimoto}@shibaura-it.ac.jp

## 1. はじめに

本研究は名詞と対になる特徴語を抽出することが目的である。特徴語とは名詞が持つ性質を表す動詞、形容詞、形容動詞のことである。

名詞と特徴語のペアを用いることで名詞の印象推定や連想語を取得するタスクに応用できる。例えば「花」という名詞が「美しい」という特徴を持っていた場合、「花」はポジティブな印象であると推定できる。また名詞「林檎」と名詞「血」が共に「赤い」という特徴を持っていた場合、「林檎」と「血」は「赤い」という共通点によって連想できる。

一般的に名詞と対になる特徴語を抽出する場合、名詞と共起する語[1][2]や名詞と係り受け関係になる語[2]によって求める。しかし名詞と共起する語や名詞と係り受け関係になる語は必ずしも名詞の一般的な特徴とは限らない。例えば「赤い服」という句の場合、「赤い」と「服」は共起し、また「赤い」は「服」と係り受けの関係になる。しかし「赤い服」という句は文中に登場した特定の「服」が「赤い」ことを表しているのであり、「服」が一般的に「赤い」ことを表さない。

本研究では直喩表現を用いることで名詞が一般的に持つ特徴を抽出する。直喩表現は喩表現の一種である。喩表現は喩えられる語である被喩辞と喩える語である喩辞、喩辞の特徴を表す特徴語から成り立つ。また直喩は被喩辞と喩辞の間に類似性や関連性を示す標識が言語的に明示される。「のように」や「みたいに」といった表現が標識の例である。本研究では標識を手がかり語として直喩表現における喩辞と特徴語の関係から名詞と名詞の特徴の対を抽出する。また評価実験として直喩表現から抽出した名詞・特徴語対と係り受けから抽出した名詞・特徴語対をアンケート調査によって比較する。

## 2. 提案手法

提案手法の概要を図1に示す。

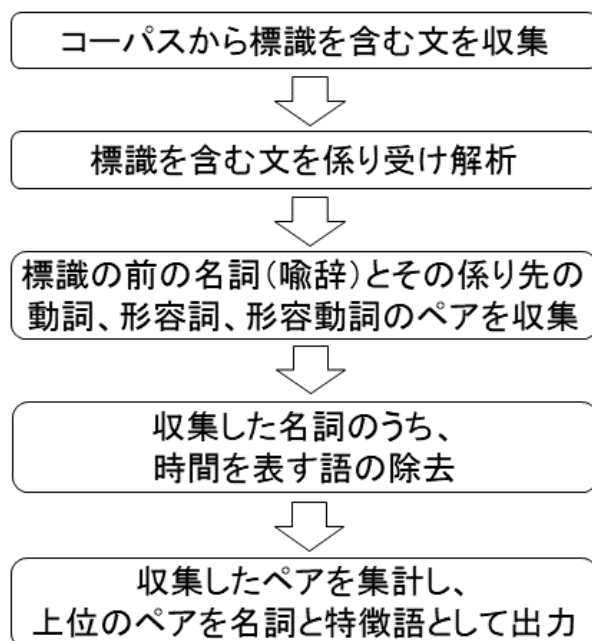


図1 提案手法の概要

まずコーパス中から直喩表現の標識である「のように」、「みたいに」、「のやうに」を手がかり語として直喩表現を含む文を収集する。

次に収集した文を係り受け解析をする。係り受け解析にはCaboChaを用いる。

係り受け解析をした文から喩辞を求める。喩辞は標識の前にある名詞とする。名詞を連体修飾する語がある場合には名詞と共に収集する。これは「兎の眼のように赤い」という表現があった場合「眼」が一般的に「赤い」のではなく「兎の眼」が「赤い」特徴を持つからである。名詞とペアになる語を喩辞と標識を含む文節の係り先から収集する。名詞とペアになる語は動詞、形容詞、形容動詞のみとする。

次に収集したペアのうち名詞が時間を表す語

であるペアを除去する。例えば「毎日のように遊ぶ」という表現がある場合「毎日」自体に「遊ぶ」特徴を持たないので除外する。時間を表す語は分類語彙表[3]の分類項目のうち「毎日・毎度」、「年」、「現在」、「過去」に属する語とする。

最後に収集したペアを集計し、2回以上出現したペアのうち関連度の高いペアを出力する。

名詞Aからペアとなる特徴語Bの関連度は次の式①で定義される。

$$\text{関連度} = \frac{\text{名詞Aと特徴語Bのペアの出現数}}{\text{特徴語Bを含むペアの出現数}} \text{---①}$$

また特徴語Bからペアとなる名詞Aの関連度は次の式②で定義される。

$$\text{関連度} = \frac{\text{名詞Aと特徴語Bのペアの出現数}}{\text{名詞Aを含むペアの出現数}} \text{---②}$$

### 3. 評価実験

#### 3.1 実験の概要

本研究で提案する直喩表現を利用して特徴語を抽出する手法と係り受け関係から特徴語を抽出する手法をアンケート調査で比較した。

#### 3.2 使用したデータ

使用したデータについて表1にまとめる。

表1 評価実験に用いたデータ

	青空文庫	毎日新聞
データ量 (UTF-8)	558MB	715MB

コーパスとして青空文庫と毎日新聞を用いた。青空文庫は古文を取り除くために手がかり語として助動詞「けむ」を含む文章を除外した。また漢字と片仮名のみを除外するため助詞「ヲ」を含み、かつ助詞「を」を含まない文章を除外した。

毎日新聞は5年分の本文データを用いた。

#### 3.3 比較する手法

CaboChaを用いてコーパスを係り受け解析し、名詞とその名詞に係る文節に含まれる動詞、形容詞、形容動詞のペア、および名詞とその係り

先の文節に含まれる動詞、形容詞、形容動詞のペアを抽出する。

10回以上出現したペアを対象に関連度の高いペアを出力する。関連度は2節と同様に式①、②で定義する。

#### 3.4 アンケートの方法

抽出した名詞・特徴語ペアに対してアンケートによる定量的評価を行った。アンケートでは被験者に名詞と特徴語をそれぞれ提示し、特徴語が名詞の性質を適切に表している語であるかを調べた。5個の名詞に対してペアとなる特徴語を各5個、5個の特徴語に対してペアとなる名詞を各5個、計50ペアを直喩表現を利用した方法と係り受けを利用した方法でそれぞれ出力し、それらを10人の被験者に5段階で評価してもらった。

評価に用いた名詞・特徴語の例を表2に示す。

表2 評価に用いた名詞・特徴語の例

直喩表現を用いて抽出		係り受けを用いて抽出	
名詞	特徴語	名詞	特徴語
針	突き刺す	針	刺さる
針	刺す	針	刺す
針	尖る	針	指す
針	とがる	針	戻す
針	鋭い	針	細い
夢	ふしぎ	夢	かなえる
夢	薄れる	夢	ついでる
夢	覚えてる	夢	うなされる
夢	思い浮かべる	夢	かなう
夢	見とれる	夢	醒める
楊白花	美しい	抜群	美しい
珠玉(しゆぎよく)	美しい	非凡	美しい
絵はがき	美しい	秘訣	美しい
女神	美しい	木目	美しい
絵画	美しい	ハーモニー	美しい
大粒の宝石	輝く	金賞	輝く
リング	輝く	MVP	輝く
銀箔	輝く	ベストナイン	輝く
黄金	輝く	最優秀	輝く
ダイヤモンド	輝く	銀色	輝く

5段階の評価基準を以下に示す。

- (5): 「名詞」は「特徴語」の性質を必ず持つ。
- (4): 「名詞」は「特徴語」の性質をたいてい持つ。
- (3): 「名詞」は「特徴語」の性質を場合によっては持つ。
- (2): 「名詞」は「特徴語」の性質を決して持たない。
- (1): 名詞や特徴語の意味が分からない。知らない語が含まれる。

またアンケートの評価基準において注意点を 2 つ設けた。

まず、名詞固有の性質とはいえない特徴語であっても、その性質を持つと判断してよい。例えば名詞「袋」、特徴語「ある」の場合、「ある」は実在するすべての名詞にあてはまる。よって「ある」が「袋」固有の性質ではないが、「袋」は「ある」という性質を満たしているため名詞と特徴語のペアとして適切であると判断する。

次に名詞は特徴語の主体とする。例えば名詞「料理」、特徴語「食べる」の場合、「料理」は「食べられる」性質を持っているが主体的に「食べる」性質を持っていない。よって「料理」と「食べる」は名詞と特徴語のペアとして不適とする。

### 3.4 実験結果

コーパスから取得したペアの数を表 3 に示す。

表 3 コーパスから取得したペアの数

	青空文庫	毎日新聞
直喩から得たペア数	53,278	12,263
係り受けから得たペア数	2,249,880	7,413,015

直喩からペアを得る方法を用いた場合、青空文庫から 53,278 ペアに対し毎日新聞から 12,263 ペアが得られた。青空文庫は小説やエッセイが多いためレトリックを用いた文章が多く含まれ直喩が多く得られた。係り受けからペアを得る方法を用いた場合、青空文庫から 2,249,880 ペア、毎日新聞から 7,413,015 ペアが得られた。このことから直喩表現から名詞・特徴語ペアを抽出する場合は青空文庫コーパスから得られたペアの割合が大きくなり、係り受けから名詞・特徴語ペアを抽出する場合は毎日新聞コーパスから得られたペアの割合が大きくなる。

アンケートの結果を図 2 に示す。

「名詞」は「特徴語」の性質を必ず持つと判断された割合は直喩から得られたペアが 32%なのに対し、係り受けから得られたペアは 22%だった。また「名詞」は「特徴語」の性質を決して持たないと判断された割合は直喩を利用したペアが 22%なのに対し、係り受けを利用したペアは 33%だった。このように、直喩から得られたペアは係り受けから得

られたペアよりも適切と評価されたものの割合が大きかった。

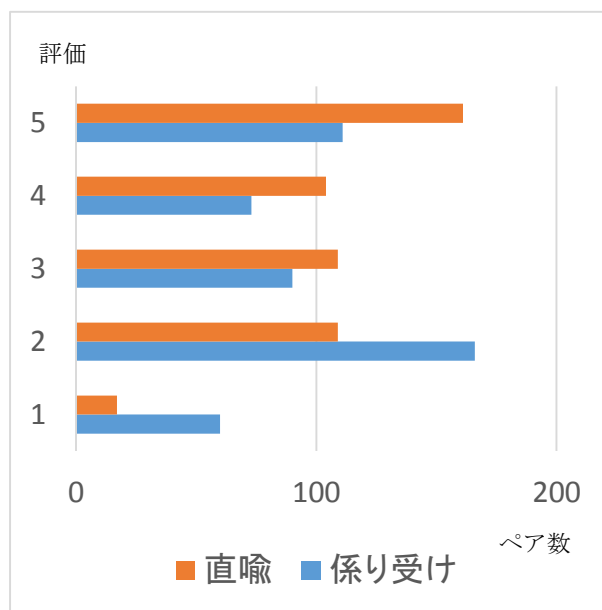


図 2 評価実験の結果

### 3.5 考察

アンケートにおいて 10 人中 8 人以上が「名詞」は「特徴語」の性質を必ず持つと評価した名詞・特徴語のペアを表 4 に示す。

表 4 評価の高かった名詞・特徴語ペア

直喩表現を用いて抽出		係り受けを用いて抽出	
名詞	特徴語	名詞	特徴語
火	熱い	夢	かなう
鳥	飛ぶ	火	燃え移る
針	鋭い	火	燃え広がる
りんご	赤い	鳥	羽ばたく
トマト	赤い	鳥	飛ぶ
血	赤い		
大粒の宝石	輝く		
ダイヤモンド	輝く		

係り受けを用いて抽出した特徴語の全てが動詞だったのに対し、直喩表現を用いて抽出した特徴語は動詞と形容詞の両方を含む。よって直喩表現を用いて名詞・特徴語ペアを抽出することにより、多様なペアが得られると考えられる。

次にアンケートにおいて 10 人中 8 人以上が「名詞」は「特徴語」の性質を決して持たない

と評価した名詞・特徴語のペアを表5に示す。

評価の低いペアにおいて直喩表現を用いて抽出したペアと係り受けを用いて抽出したペアの両方に名詞と特徴語が二格やヲ格をとる関係がある。そこで名詞と特徴語から一般的にとりうる格を統計的に推定することで不適切なペアを除外することができると考えられる。

また「嘘のように」のような特徴語が喩辞の特徴を表していない表現も比喩に含まれることが分かった。

表5 評価の低かった名詞・特徴語ペア

直喩表現を用いて抽出		係り受けを用いて抽出	
名詞	特徴語	名詞	特徴語
夢	覚えてる	夢	かなえる
夢	思い浮かべる	夢	うなされる
夢	見とれる	火	消し止める
鳥	羽ばたかす	針	指す
嘘	死ぬ	針	戻す
		秘訣	美しい

## 4. おわりに

本研究では直喩表現を利用して名詞と名詞の性質を表す特徴語のペアをコーパスから抽出する方法を提案した。今回行った実験により提案手法は係り受けを用いた名詞・特徴語ペアの抽出方法よりも高い割合で適切なペアを抽出できることが示された。

今後の課題として名詞と特徴語が二格やヲ格をとる関係の場合を除外する必要がある。また特徴語が喩辞の特徴を表していない比喩表現を分類し、名詞・特徴語ペアから除外できるようにしたい。

## 参考文献

- [1] 仁科俊晴, 鍛冶伸裕, 吉永直樹, 豊田正史: 対義形容詞対との相互情報量を利用した概念語の順序付け, 情報処理学会研究報告, 2013-NL-214, 8, 2013.
- [2] 後藤慎也, 鈴木良生, 田添丈博: 意味関連辞書構築のための単語間関連度収集手法の検討, 言語処理学会 第19回年次大会, 2013.
- [3] 国立国語研究所: 分類語彙表 増補改訂版, 大日本図書, 2004.