

# ヒューリスティック手法による隠喩判定システムの精度改善

猫本隆哲<sup>†</sup> 澤山大陸<sup>†</sup> 中島陽子<sup>‡</sup> 本間宏利<sup>‡</sup> M. Ptaszynski\* 榎井文人\* 増山 繁\*\*

<sup>†</sup> 釧路高専 情報工学科, <sup>‡</sup> 釧路高専 創造工学科

\* 北見工業大学 地域未来デザイン工学科

\*\* 豊橋技術科学大学 情報・知能工学系

{yoko, honma}@info.kushiro-ct.ac.jp

## 1 はじめに

比喩とは、ある概念を他の概念によって説明または強調する修辞技法の一つである。心理学や言語学、また自然言語処理の分野についても、機械翻訳における比喩表現の誤訳回避などの目的で比喩の抽出・解析の研究が行われている [1, 2]。比喩表現については、様々なアプローチによる研究が行われているが、その多くは直喩を対象としたものであり、隠喩に関する研究は多いとはいえない [3]。直喩に見られる「のような」、「みたいに」などの比喩指標が存在しないことが隠喩の検出を困難にさせる原因の一つになっている。

本研究では、以前より開発を行ってきた隠喩判定システムの精度改善を試み、文章中から隠喩表現を抽出するシステムの開発を目的とする。コンピューターでの隠喩判定は、構文情報のみでは実現がほぼ不可能であるため、意味情報や概念情報を取り扱う必要がある。そこで、形態素解析エンジンと日本語語彙大系を用いた新たな判定モデルを構築し、従来までの隠喩判定システムについて、新たな判定システムを提案する。

## 2 隠喩の判定

本研究では「人生は旅だ」、「鋼の肉体」のような単純な短文から、「彼は長身で鋼の肉体をもつ大男だった」のような助詞や接続詞などを含む文章までの隠喩表現判定を試みる。隠喩表現の判定指標として、最も表現が単純である「A は B」、「B の A」という形をベースとして、A と B が隠喩関係にあるかどうかを判定する。なお、A と B はどちらも名詞に限定し、動詞を含む隠喩表現は判定対象としては扱わない。比喩表現において、例えられる側である名詞 A を被喩辞、例える側である名詞 B を喩辞とよぶ。隠喩の形式として以下の 3 種類を想定する。

### 強調型隠喩

喩辞が被喩辞の特徴を強調する形式。強調型隠喩の例として、「彼女は天使だ」のように「彼女」という人物を「天使」という言葉が持つ「美しい」、「優しい」といった特徴で割り当て、被喩辞を強調する。

### 連想型隠喩

被喩辞と喩辞が共通の特徴を持つ形式。連想型隠喩の例として、「時は金なり」のように「時」と「金」はどちらも「貴重である」、「大切である」といった、被喩辞と喩辞が共通のイメージを有する。

### 擬人型隠喩

被喩辞が人体に係る名称である形式。擬人型隠喩の例として、「鋼の肉体」、「鉄の拳」のように「肉体」、「拳」はどちらも人体に関係のある名称であり、それを名詞で例えている。

事前研究では連想度と類似度を隠喩判定の素性として用いることで、一定数の隠喩判定が可能であったが、これらのみでは強調型と連想型以外の隠喩を判定、抽出することが困難であった。また、非隠喩表現 (以下、リテラル) の判定が考慮されていなかったため、リテラルを隠喩と誤判定してしまう傾向があった。本研究では、前述した 3 種類の隠喩形式以外の隠喩判定の精度を高めることに重点を置き、形態素解析エンジンと日本語語彙大系を利用した新たなアプローチを試みた。

### 2.1 連想度

ある言葉から連想できる別の語のことを連想語と定義する。連想度は、2 つの単語のそれぞれの連想語から、共通する連想語が存在する割合を表す数値である。

2つの単語組を  $a, b$ , 抽出した連想語集合を  $A, B$  とすると, 連想度  $assoc$  は式 (1) で導出できる.

$$assoc(a, b) = \frac{A \cap B}{A + B} \quad (1)$$

連想型隠喩の連想度の値は大きくなり, 非隠喩の連想度は小さくなる, もしくは0になる傾向がある. 判定尺度として利用する連想語は kizasi ラボの連想語抽出ツール **keygram**<sup>1</sup> を使用して抽出する. keygram から抽出される連想語の例を表 1 に示す.

表 1: “りんご” と “みかん” の連想語例

単語	りんご	みかん
連想語	紅玉	柑橘類
	コンポート	デコポン
	シナモン	ミカン
	リンゴ	寒天
	キウイ	こたつ
	ぶどう	りんご
	林檎	柑橘
	バナナ	蜜柑
	レーズン	缶詰
	みかん	皮

keygram では, 連想語検索を行なった語に対し, Web 上でのエントリ数, 掲載数などの情報から最新の連想語と考えられる語が抽出されるが, 稀に連想語とは程遠い語が混在することがある. 以下に, 連想語の精度 (以下, 連想精度) がより高い連想語辞書の構築のための精度改善アプローチ手法を示す.

### 2.1.1 閾値設定による精度改善アプローチ

keygram より抽出される連想語には, 元の語との関連性の大きさを示す関連度というパラメーターが設定されている. この閾値を適切に設定することで, 連想精度が高い語のみを抽出することを可能とした.

### 2.1.2 相互連想を利用した精度改善アプローチ

元となった語と連想語の連想関連は, 相互的であるものの方が連想度は高くなると考える. 例えば, 表 1 では “りんご” の連想語に “みかん” が存在し, “みか

ん” の連想語に “りんご” が存在するため, これらは相互連想の関係にある. 相互関係を満たさない連想語はすべて連想語辞書から削除する.

## 2.2 類似度

類似度はシソーラスにおける 2 単語間の距離を 0 から 1 の数値で表す尺度である.

2つの単語組を  $a, b$  の根からの深さをそれぞれ  $d_a, d_b$  とし, これらの共通する親からの深さを  $d_c$  とすると, 単語組  $a, b$  の類似度  $sim$  は式 (2) で導出できる.

$$sim(a, b) = \frac{d_c \times 2}{d_a + d_b} \quad (2)$$

強調型隠喩の類似度を算出すると値は小さくなり, 隠喩ではない非隠喩の類似度は大きくなる傾向がある. 本研究では日本語の大規模シソーラスである日本語語彙大系<sup>2</sup>を使用する.

## 3 リテラルの判定

リテラルを正しく判定することで, 隠喩判定の精度向上が実現できる. 対応する 2つの名詞  $A, B$  の意味領域が互いに排斥している場合は比喩表現である傾向が高く, 反対に名詞  $A$  の意味領域が名詞  $B$  の意味領域に含まれている場合はその傾向は低くなる. 以下の 3つの性質を有する場合, リテラルと断定できると考える.

### 対比リテラル

被喩辞が喩辞の概念に含まれている形式. 対比リテラルの例として, 「知床の大自然」のように「大自然」は「知床」の下位概念, また保有物である.

### 動的リテラル

被喩辞または喩辞の表現が状態, 状況である形式. 動的リテラルの例として, 「財政の問題」のように「問題」は場合によって意味合いが変動する表現であることから, 例示を示すリテラルである.

### 特定リテラル

喩辞が人称代名詞, 現在過去の時制名詞, 物事の評価を表す名詞である形式. 特定リテラルの例として, 「従来の研究」のように「従来」は過去を表す時制名詞であり, 対比リテラルと同様に比喩としての意味を持たない.

<sup>1</sup><http://kizasi.jp/lab/keygram/keygram.py/>

<sup>2</sup><http://www.kecl.ntt.co.jp/icl/lirg/resources/GoiTaikei/>

隠喩判定対象が上記のいずれにも該当しない場合、隠喩である可能性が高いと判断する。リテラルの判定には、名詞が持つ品詞の種類を形態素解析エンジンを用いて解析し、日本語語彙大系から意味情報を検出することで表現の除外を行なう。

### 3.1 形態素解析エンジン

本研究では形態素解析エンジンとして MeCab<sup>3</sup> を用いる。MeCab は、辞書とシステムが完全分離されており、他の最新の IPA 辞書を参照することができる。本研究では、固有名詞を含めた新語が多く収録されている mecab-ipadic-NEologd<sup>4</sup> を利用し、品詞情報を解析することでリテラルを除外をする。

### 3.2 意味情報検出

日本語語彙大系には、単語や単語属性の他に意味属性番号が割り当てられている。隠喩は抽象名詞を異種概念、またはそれに属する具体物で例えることが多い。動詞リテラルを判定する際、語彙系から意味属性番号を参照することで被喩辞が喩辞に属するか調べることができる。

## 4 隠喩判定実験

精度改善を施した連想語辞書と類似語辞書にリテラル判定システムを組み合わせた新規の段階判定モデルにより、隠喩の判定実験を行なった。自作した隠喩表現 100 組、リテラル 100 組をテストケースとして、従来モデルと本研究モデルを用いて実験を行なった。

### 4.1 従来モデルの判定実験

従来モデルの判定手順概略を図 1 に示す。

従来モデルでは「A は B」、 「B の A」という単純な文体に対して、名詞 A と B の連想度 0 以上、または、類似度 0.6 以下を隠喩と判定する。このモデルの判定結果を表 2 に示す。

隠喩 100 組中 98 組と非常に高い判定精度となったが、リテラルは 100 組中 4 組しか正常に判定されていない。原因として閾値の緩さ、過去に利用したテストケースの品質の悪さなどが挙げられるが、リテラルを考慮した判定モデルではないことが最大の要因である

<sup>3</sup><http://taku910.github.io/mecab/>

<sup>4</sup><https://github.com/neologd/mecab-ipadic-neologd>

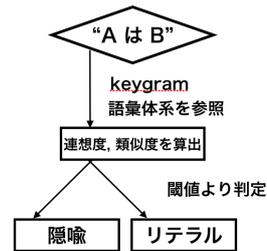


図 1: 従来モデルの判定手順概略

表 2: 従来モデルによる判定結果 [個]

	隠喩	リテラル
判定結果	98/100	4/100

と考えられる。適合率は 0.51、再現率は 0.98 となり、F 値は 0.67 となった。

### 4.2 本研究モデルの判定実験

本研究での提案手法を用いた新規モデルの判定手順概略を図 2 に示す。

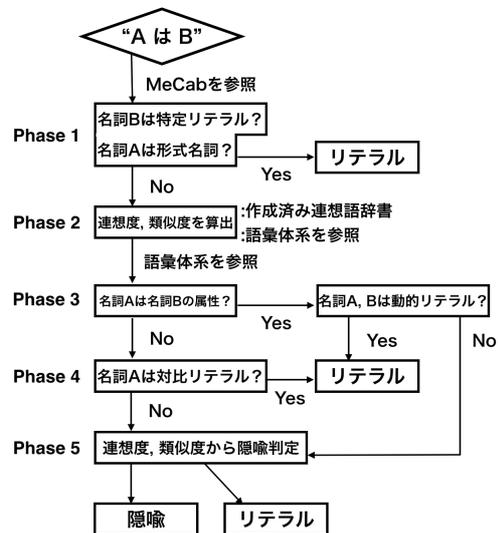


図 2: 本研究モデルの判定手順概略

本研究モデルでは、連想度と類似度の算出値を保持しつつ、初期の判定段階でリテラルと判定された文を隠喩候補から除外する。最後までリテラルと判定されなかった表現に対して、閾値を設定し最終判定を行なう。新規モデルでは、連想度 0.03 以上と類似度 0.25 以下を隠喩と判定する。判定結果を表 3 に示す。

表 3: 本研究モデルによる判定結果 [個]

	隠喩	リテラル
判定結果	75/100	81/100

従来モデルに比べ、隠喩判定精度はやや低下したがリテラルを 100 組中 81 組正常に判定できた。リテラルの除外判定、また連想語辞書の精度改善から隠喩とリテラルが共に均衡した判定結果が得られた。適合率は 0.78、再現率は 0.75 となり、F 値は 0.76 となった。

### 4.3 判定結果検証

図 2 に示した各フェーズ毎に検出されたリテラル、誤判定表現の一部を表 4 に示す。

表 4: 比喩性判定結果

検出箇所	リテラル	
<b>Phase1:</b>	太平洋は広い 明日の天気 嘘は得意だ	手術は成功 火力が弱い
<b>Phase3:</b>	国民の信頼 子供はやんちゃだ	今回の調査 お金は大事だ
<b>Phase4:</b>	日本の国技 東京は都会だ 住宅の基礎	アメリカの情勢 紅茶は飲み物だ
<b>Phase5:</b>	犬は従順だ 夏の季節 車は便利だ	夫は医師だ 焦りは禁物だ 道路の勾配 etc..
誤判定リテラル		
立場は逆だ 機械の修理	財布は空だ 動物の餌	教育の義務 環境の変化
誤判定隠喩表現		
景色の圧力 雪の肌 灰色の日	詩の叫び 紅葉の手 無重力の海	ガラスの心 芝の絨毯 言葉は刃物

100 組のリテラルから、隠喩判定前に除外されたりテラルは 100 組中 25 組、残る 75 組は隠喩判定まで除外されず、最終的には連想度と類似度から多くのリテラルが正常に判定される結果となった。連想語辞書の精度改善から隠喩判定時に多くのリテラルを検出可能

にした一方で、依然として判定しきれていないリテラルと隠喩表現が存在し、隠喩判定前のリテラル除外数も不十分であることから、今後どのようなアプローチを試みるか検討する必要がある。

## 5 おわりに

本研究では「A は B」, 「B の A」という文に対して隠喩であるかリテラルであるかの判定を行なうため、隠喩表現の被喩辞と喩辞の関係性から辞書を参照し、連想度と類似度を算出した。また、形態素解析エンジンと日本語語彙大系を利用し、リテラルの判定も行なうことで連想度と類似度のみでは判定困難な非隠喩表現の判定精度向上を実現した。

現在、不特定の文章から隠喩を検出し、抽出するシステムを本研究で作成した隠喩判定モデルを利用し、開発を続けている。形態素解析エンジンによる品詞情報の出力の他に分かち書きによる出力を加えることで、形態素ごとの品詞情報を取得し、隠喩判定の対象とする文を限定することで、不特定の文から隠喩表現を抽出できると考える。

しかし、辞書を参照する現状の隠喩判定モデルでは、文脈の前後表現が考慮されていないため、文脈に依存した比喩性の判定はできない。隠喩表現の場合、「彼は鬼だ」という単体の文は隠喩と判定されるが、「鬼ごっこで彼は鬼だ」という前表現が存在する場合はリテラルとなる。「鬼」というのは「彼」の風貌について例えているのか、「鬼ごっこ」という枠組みの中での役割としての「鬼」なのか、このような文脈の処理手法を考案し、判定可能としていくことが比喩判定システムの今後の永続的な課題となるだろう。

## 参考文献

- [1] 榎井文人, 他, “確率的尺度を用いた比喩性検出手法,” 自然言語処理, vol.9, no.5, pp.71-92, 2002.
- [2] 田添丈博, 他, ““名詞 A のような名詞 B” 表現の比喩性判定モデル,” 自然言語処理, vol.10, no.2, pp.43-58, 2003.
- [3] 関口育栄, 他, “類似度と連想度を用いた隠喩表現判定に関する研究,” 言語処理学会第 23 回年次大会, P2-1, 2017.