

文末表現“ようだ”に注目した小説テキストからの直喩文抽出手法

宮脇 星名[†] 安藤 一秋[‡]香川大学大学院創発科学研究科[†] 香川大学創造工学部[‡]

1. はじめに

比喩表現は、ある物事を別の事柄に例えることで、文字通りの表現以外の情報を表現・伝達する機能と詩的・審美的効果を喚起する機能の2つを持つ[1]。比喩表現は、主に「直喩」、「隠喩」、「換喩」、「提喩」に分類され、比喩性を感じ取る直接の契機になる特定の言語形式を比喩指標と呼ぶ[2]。特に直喩表現では、「よう」といった特定の比喩指標が使用されやすい。

我々の先行研究[3]では、「名詞Aのような名詞B」タイプの直喩表現を含む文を対象に事前学習済み言語モデルを用いた抽出手法を提案し、実験により、Precisionの最良値が90.91%であることを確認した。また、「動詞Aのような名詞B」タイプの直喩表現を含む文を対象に5素性によるSVMを用いた抽出手法を提案し、Precisionの最良値が80.65%であることを確認した。

本稿では、比喩指標の中でも出現数の高い「よう」に注目し、表1に示す文末に出現する「名詞Aのようだ」タイプと「動詞Aようだ」タイプに属する直喩を含む文を抽出する手法について提案する。

表1 文末表現「ようだ」を含む文の具体例

| 文法パターン | 比喩例 | リテラル例 |
|-----------|--------------------------|-------------------------|
| 「名詞Aのようだ」 | 肌は雪のようだった、 この島は天国のようだ | 父からの連絡のようだ、 好評のようだ |
| 「動詞Aようだ」 | 不安が皮膚を這うようだ、 身が凍るようだ | 車が来たようだ、 彼女は困ったようだった |

2. 文末表現「ようだ」を含む比喩文の分析

菊池[4]は、「ようだ」、「みたいだ」といった助動詞は、命題の真偽や事象の成立に関係し、比喩や推量、婉曲といった複数の用法を持つと述べている。また、「ようだ」が文末で使用された場合、モダリティに相当する働きを見せるが、修飾節に使用されるとモダリティとしての機能は潜在化すると指摘している。

以上より、「よう」を含む文であっても必ずしも比喩であるとはいえず、「名詞Aのような名詞B」よりも文末に出現する「名詞Aのようだ」と「動詞Aようだ」は、モダリティ機能が強くなる。

予備調査として「青空文庫」と「小説家になろう」より取得した文末「ようだ」を含む一文を個人でアノテーションした結果、表2の結果が得られた。動詞を含むパターンの方がリテラルは多くなり、特に文末に「ようだ」が出現するとその傾向は強くなる。助動詞「よう」の出現場所と品詞の差で比喩性の強さが異なるといえる。

3. データセット

本実験で使用する小説テキストは「青空文庫」から100作品、「小説家になろう」の6ジャンル（ドラマ・歴

A Method for Extracting Simile Sentences from Novel Texts Focused on the Sentence-Final Expression “Youda”

[†] Seina Miyawaki, Graduate School of Science for Creative Emergence, Kagawa University

[‡] Kazuaki Ando, Faculty of Engineering and Design, Kagawa University

表2 各文法パターンの比喩出現率

| 文法パターン | 比喩出現率(比喩文数) | アノテーションした文数 |
|--------------|-------------|-------------|
| 「名詞Aのような名詞B」 | 66.90%(285) | 426 |
| 「動詞Aのような名詞B」 | 47.50(304) | 640 |
| 「名詞Aのようだ」 | 35.57%(494) | 1,389 |
| 「動詞Aようだ」 | 11.57%(577) | 4,987 |

史・ホラー・推理・ハイファンタジー・ローファンタジー)から任意抽出した各100作品を用いる。それらの本文から「名詞Aのようだ」を含む文を抽出し、比喩である文を294文、リテラル文を294文、計588文を人手でラベリングしてデータセットを構築した。また、「動詞Aのようだ」を含む文についても、比喩である文を269文、リテラル文を269文、計538文を人手でラベリングしてデータセットを構築した。なお、ラベリングは3人のアノテーターに依頼し、強い比喩、弱い比喩、弱いリテラル、強いリテラル、判断不可能の5段階で評価してもらい、過半数が強い比喩または弱い比喩と評価した文を比喩文、過半数が強いリテラルまたは弱いリテラルと評価した文をリテラル文とした。

4. 「名詞Aのようだ」を含む文の抽出手法

文末に出現する「名詞Aのようだ」表現を含む文を規則に基づいて抽出する手法を提案する。

4.1 提案手法

「名詞Aのようだ」は、出現する主語を名詞Bとして扱うことで、「名詞Aのような名詞B」に言い換えることができる。たとえば、「瞳は真珠のようだ。(名詞Bは名詞Aのようだ)」は「真珠のような瞳だ。(名詞Aのような名詞B)」のように言い換えできる。この法則を利用し、田添らの先行研究[5]で提案された、「名詞Aのような名詞B」に対する日本語語彙体系を利用した比喩性判定モデルを参考に、規則に基づいて抽出する手法を提案する。図1に提案手法を示す。なお本稿では、紙面の都合上、先行研究[5]と異なるステップのみ説明する。

図1に示す提案手法では、まず、係り受け解析器Cabochaを用いて「名詞Aのようだ」を含む文から名詞Aと主語にあたる名詞Bを特定する。ステップ1では、表3に示す特定の副詞が出現するか否かを判定し、出現する場合はリテラルとする。工藤らの研究[6]によると、特定の副詞は、特定のモダリティ表現と共起する関係にあると述べられていることから、推量副詞を人手で指定した。ステップ2では、表4に示す特定の比喩指標を含む文を比喩と判断する。比喩指標は、菊池の研究[4]で示された比喩文中に出現した副詞を参考に、強い比喩指標に該当するものを人手で指定した。主語が存在しない文についても副詞を手掛かりとすることで、一定数が判定可能となる。ステップ4では、名詞Bが形式名詞か代名詞であれば特定の名詞とする。また、名詞Bが出現しなければ判定不可能としてリテラル扱いとする。ステップ5では、特定の名詞Bという条件下において、名詞Aが具体物であれば単に説明している可能性が高いためほぼリテラル、抽象物(意味属性番号1000-2715)であれば比喩とする。

表3 特定の副詞群

やはり、やっぱり、どうやら、どうも、
おそらく、きっと、よほど

表4 特定の比喩指標群

まるで、あたかも、ちょうど、まさに、さながら、
いわば、たとえば、たとえるなら、たとえば

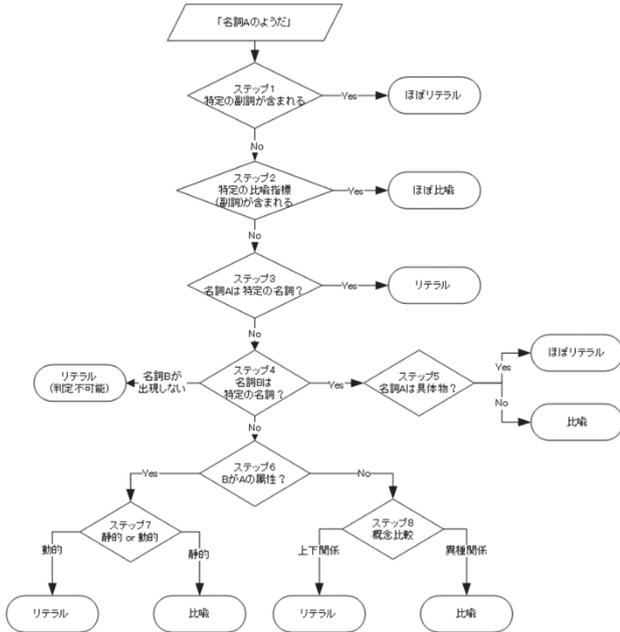


図1 「名詞Aのようだ」を含む比喩文抽出手法

4.2 評価実験

提案手法を588文に適用し, Precision, Recall, F1-scoreで評価した結果, Precisionが0.8052, Recallが0.7313, F1-scoreが0.7665となることを確認した.

5. 「動詞Aようだ」を含む文の抽出手法

文末に出現する「動詞Aようだ」表現を含む文を規則に基づいて抽出する手法を提案する.

5.1.提案モデル

「名詞Aのようだ」と同様, 「動詞Aのような名詞B」に言い換えるため, 対象文から係り受け解析によって動詞Aと名詞Bにあたる主語, 目的語をそれぞれ抽出する. 提案手法を図2に示す. なお本稿では, 紙面の都合上「名詞Aのようだ」と異なるステップのみ説明する.

ステップ3では, 菊池らの研究[4]と予備調査に基づき, 著しく比喩率が低いと判明したサ変動詞と知覚動詞をほぼリテラルとして扱う. ステップ5では, 名詞Bが存在しない文に対して, 日本語語彙体系内の用言が持つ文型パターンに一致しない目的語が使用されている場合, 比喩であると判定する. ステップ6では, 動詞Aの日本語語彙体系における意味情報が抽象的關係(3-8), 精神的關係(9-14), 自然現象(15), 物理的行動(17-27), 精神的行動(28-32), その他(33-36)のいずれかを複数持つかを判定する. ステップ7では, 菊池[7]が提示した程度型直喩を参考に, 動詞Aが身体と感覚の意味情報を持ち, かつ名詞Bが抽象物である場合を比喩と判定する. また, 動詞Aが属性と身体の意味情報を持ち, 名詞Bが抽象物である場合を比喩, 動詞Aが属性と感情の意

味情報を持ち, 名詞Bが事・人間活動(1235-2303)以外である場合を比喩とする. ステップ8では, 用言が持つ文型パターンに, 名詞Bの意味属性番号が一致しなければ比喩とする.

5.2 評価実験

提案手法を538文に適用し, Precision, Recall, F1-scoreで評価した結果, Precisionが0.7219, Recallが0.5019, F1値が0.5921となることを確認した.

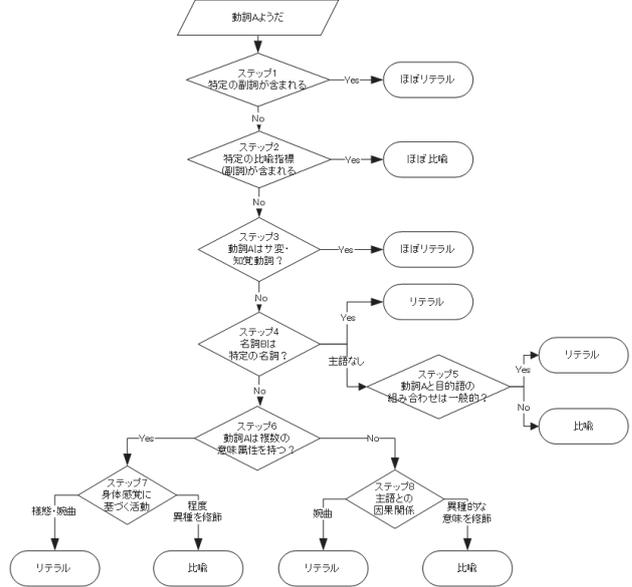


図2 「動詞Aようだ」を含む比喩文抽出手法

6. おわりに

本稿では, 文末に出現する「名詞Aのようだ」と「動詞Aようだ」という直喩表現を含む文を対象に, 規則に基づく抽出手法を提案した. それぞれの手法のPrecisionは, 「名詞Aのようだ」において0.8052, 「動詞Aようだ」において0.7219となった. 先行研究[3]で提案した「名詞Aのような名詞B」を含む文の抽出手法のPrecisionが0.9091であったことから, 文法パターンの違いによる抽出難易度の差と, 規則に基づく抽出手法の限界を確認した. 本稿で注目した文末「ようだ」タイプには, 他の文法パターンにおいてあまり効果の得られなかった特定の比喩指標が効果的であったことから, 文末「ようだ」における人間の直喩判断には呼応の副詞が大きく影響していると考察できる. 今後は, これらの知見を活かして, 素性に基づく抽出手法や事前学習済み言語モデルを用いた手法について検討する.

参考文献

[1]内海, “比喩によってどのような詩的効果が喚起されるか 比喩の鑑賞仮定の認知モデルに向けて”, JSAI2003, pp.1-4, 2003.
 [2]中村, “比喩表現の理論と分類”, 秀英出版, 1977.
 [3]宮脇他, “小説テキストからの「ような」表現に基づく直喩文抽出手法の検討”, IPSJ2024, pp.4-809-4-810, 2024.
 [4]菊池, “現代日本語における直喩の構文論的研究”, 博士学位論文(中央大学大学院), 2022.
 [5]田添他, “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, 10巻, 2号, pp.43-58, 2003.
 [6]工藤他, “日本語の文法3モダリティ”, 岩波書店, 2000.
 [7]菊池, “直喩による程度表現の働き”, 表現研究, 14巻, 115号, pp.14-23, 2022.