

## 単語間の連想関係を考慮した換喩解析

寺岡 丈博†

†東京工科大学

石崎 俊‡

‡慶應義塾大学

### 1 はじめに

換喩（メトニミー）は隣接性に基づいた比喩であり、一般的には表1のように空間的隣接性 [1] と時間的隣接性 [1, 2] に大別することができる。表中の例文のように、リテラルな（字義通りの）解釈と換喩表現を考慮した解釈を比べると、いずれも文法的には正しいにも関わらず、意味的な差異が大きい。そのため、換喩表現かりテラルかの誤解釈によって全く異なった出力になる可能性があることから、自然言語処理の分野では、従来から換喩解析は重要な問題である [3, 4]。

これまで日本語の換喩解析に関しては、新聞コーパスから収集した名詞の用例や共起関係を統計的に利用した研究 [5, 6] が代表的なものとして挙げられる。村田らは、人手で作成した格フレーム辞書を用いて換喩表現の検出も行っていたが、いずれの研究も換喩表現の言い換えについて主に議論が行われている。

一方で、著者らの先行研究 [7] では、換喩が心理的な連想のリンクと密接に関わっていること [1] に着目し、文中における名詞や動詞の連想関係を用いることで精度の高い換喩表現の検出を実現している。これにより換喩表現の検出に関しては、共起情報の代わりに連想情報を用いることの有効性が示されたが、言い換えに関しては議論が不十分であった。そこで本研究では、換喩表現の言い換えに関して単語間の連想関係を考慮した手法を提案し、従来手法と比較することで有効性を検証する。

表 1: 換喩表現の主な関係性

関係性	例文 (換喩表現が意味する内容)
「空間的隣接性」	
容器 - 中身	一升瓶 (→酒) を飲み干す
原料 - 製品	アルコール (→酒) を飲む
手段 - 主体	白バイ (→警官) が逮捕する
作者 - 作品	マーラー (→音楽) を聴く
付属物 - 主体	詰め襟 (→学生) が歩く
その他	理論 (→提唱者) は～を主張する
「時間的隣接性」	
原因 - 結果	杯を傾ける (→お酒を飲む)

Analysis of Metonymic Expressions Using Associative Relation between Words

†Takehiro TERAOKA ‡Shun ISHIZAKI

†Tokyo University of Technology

‡Keio University

表 2: 名詞連想概念辞書のデータ例

連想の課題	「辞書」の連想語 (連想距離)
上位概念	本 (1.17), 書物 (3.33), 書籍 (4.88)
部分・材料概念	紙 (1.23), ページ (3.31)
属性概念	重い (2.00), 難しい (5.54)
環境概念	図書館 (1.66), 本棚 (2.22)

表 3: 動詞連想概念辞書のデータ例

連想の課題	「運ぶ」の連想語 (連想距離)
動作主	私 (3.60), 引越し業者 (4.21)
対象	荷物 (1.36), 人 (4.45), 家具 (7.78)
場所	倉庫 (3.73), 工場 (4.50), 会社 (6.22)
手段	車 (1.62), トラック (2.34), 手 (3.47)

### 2 提案手法

#### 2.1 連想概念辞書

本研究では、名詞連想概念辞書 [8] と動詞連想概念辞書 [7] そして日本語 WordNet [9] を用いて単語間の連想関係を計算し、換喩表現の検出と言い換えを行う。連想概念辞書は、人間が名詞や動詞に対して連想した情報から構築されており、著者らの先行研究では日本語 WordNet の synset と組み合わせて検出に有効な概念間距離を計算している。表2と表3は、全体の一部にあたる連想語のデータ例をそれぞれ表している。尚、動詞連想概念辞書は先行研究から拡張されており、2015年現在は刺激語が動詞 519 語に対して、連想語が延べ約 220,000 語、異なり約 45,000 語となっている。

#### 2.2 手法の概要

以下で、提案手法の概要を3段階に分けて説明する。

- (1) 換喩表現の検出 入力文に対して形態素解析と構文解析を行い、先行研究 [7] で用いた学習データと学習モデルを用いて、文中の名詞に関して換喩表現かりテラルかを判別する。
- (2) 言い換え候補の抽出 換喩表現と判断された名詞とその係り先の用言に関する連想の情報を連想概念辞書からそれぞれ抽出する。その際に、名詞連想概念辞書からは部分・材料概念や環境概念の連想語を抽出する。また、動詞連想概念辞書からは、その名詞の格助詞と対応した連想語を抽出する。

そして、これらの中で重複した連想語があれば言い換え候補とする。重複する連想語が無い場合は、上位語から重複する連想語を抽出する。

- (3) 確信度による出力 言い換え候補の単語に対して、連想概念辞書から抽出した連想距離の値を用いて下記のように定義した確信度  $C$  を計算する。  $L_N$  と  $L_V$  は、名詞連想概念辞書における文中名詞と言い換え候補の連想距離、動詞連想概念辞書における用言と言い換え候補の連想距離をそれぞれ指している。また、  $k$  はこれらの積の逆数を、  $L_H$  は上位語からの連想距離を指している。そして最後に、確信度  $C$  の降順で言い換え候補を出力する。

$$C = \frac{k}{\sum_{i=1}^n k_i} \quad (0 < C \leq 1) \quad (1)$$

$$k = \frac{1}{L_V(L_N + L_H)} \quad (L_V L_N \neq 0) \quad (2)$$

### 3 評価実験

#### 3.1 実験材料

ここでは、用例ベースの換喩解析で代表的な村田らの手法 [5] をベースラインとして比較する。そのため、村田らが実験で用いた全 23 文の中で、換喩であることを判定できた 17 文を入力文とする。尚、これらの入力文は文献 [1] から抽出されており、換喩表現の解釈にあたる単語も明示されている。

#### 3.2 結果と考察

表 4 は、提案手法とベースラインの結果を比較したものである。精度は、ベースライン (0.41) よりも提案手法 (0.77) の方が高く、全 17 文に対して 13 文で換喩表現の言い換え候補を正しく出力していた。そして確信度が最も高い候補のみ、つまり各入力文に対して 1 語のみを出力していたことから、複数の候補から絞り込む際に今回定義した確信度が有効に機能したことが考えられる。

しかし、実験で使用した入力文の数が少ないことや、村田らの用いた格フレーム辞書が人手で作成していたために小規模であったことを踏まえると、今回の実験で提案手法の有効性が示されたとは必ずしもいえない。そのため、換喩表現を含めた文だけでなくリテラルな文も入力文として加え、換喩表現の検出と言い換えの両方を合わせて評価することが今後の課題である。

### 4 おわりに

本研究は、単語間の連想関係に基づいて換喩表現を言い換える手法を提案した。用例ベースの従来手法を

表 4: 提案手法とベースラインの実験結果

	ベースライン	提案手法
精度	0.41 (7/17)	0.77 (13/17)

ベースラインとして評価実験を行ったところ、提案手法における確信度の有効性が示唆された。今後は、検出と言い換えの機能を合わせて大規模な評価実験を行いたい。

### 謝辞

本研究の一部は JSPS 科研費 15K16787 の助成を受けたものである。

### 参考文献

- [1] 山梨正明. 比喩と理解. 東京大学出版会, 1988.
- [2] 瀬戸賢一. 日本語のレトリック. 岩波書店, 2002.
- [3] Shinichiro Kamei and Takahiro Wakao. Metonymy; Reassessment, Survey of Acceptability, and its Treatment in a Machine Translation System. In *Proceedings of the 30th Annual Meeting of the Association for Computational Linguistics*, pp. 309–311, 1992.
- [4] 清田陽司, 黒橋禎夫, 木戸冬子. 自動抽出した換喩表現を用いた係り受け関係のずれの解消. 自然言語処理, Vol. 11, No. 4, pp. 127–145, 2004.
- [5] 村田真樹, 山本専, 黒橋禎夫, 井佐原均, 長尾真. 名詞句「A の B」「AB」の用例を利用した換喩解析. 人工知能学会誌, Vol. 15, No. 3, pp. 503–510, 2000.
- [6] 内山将夫, 村田真樹, 馬青, 内元清貴, 井佐原均. 統計的手法による換喩の解釈. 自然言語処理, Vol. 7, No. 2, pp. 91–116, 2000.
- [7] 寺岡丈博, 東中竜一郎, 岡本潤, 石崎俊. 単語間連想関係を用いた換喩表現の自動検出. 人工知能学会論文誌, Vol. 28, No. 3, pp. 335–346, 2013.
- [8] 岡本潤, 石崎俊. 概念間距離の定式化と既存電子化辞書との比較. 自然言語処理, Vol. 8, No. 4, pp. 37–54, 2001.
- [9] Hiroshi Isahara, Francis Bond, Kiyotaka Uchimoto, Masao Utiyama, and Kyoko Kanzaki. Development of Japanese WordNet. In *Proceedings of the 6th International Conference on Language Resources and Evaluation*, pp. 2420–2422, 2008.