

動詞と名詞の連想による換喩検出法の提案

寺岡 丈博[†] 岡本 潤[‡] 石崎 俊[†]

[†] 慶應義塾大学大学院 政策・メディア研究科

[‡] 慶應義塾大学 SFC 研究所

{teraoka, juno, ishizaki}@sfc.keio.ac.jp

1 はじめに

換喩（メトニミー）は比喻の一種であるが、喩えるもの（喩詞）と喩えられるもの（被喩詞）が類似性の関係で表される直喩や隠喩とは異なり、主にそれらの隣接性や近接性に基づく関係がある。例えば、「一升瓶を飲み干す」という換喩に関しては、リテラルな解釈すなわち字義通りならば「一升瓶」を丸ごと飲み干してしまったことになってしまうが、本来ならば「一升瓶の中に入っている酒を飲み干す」という解釈となる。これは「酒」を空間的な隣接性の関係のある「一升瓶」に置き換えて表されているためであり、このような換喩の機能は、基本的にあるものをそれに関係した別のもので表す点にある。この換喩表現を特徴づける関係は「一升瓶」と「酒」の関係に見られる「容器と中身」の他に「原料と製品」、「主体と手段」、「主体と付属物」、「作者と作品」、「原因と結果」など、空間的な隣接性や近接性、共存性のみならず、時間的な前後関係や因果関係などもある [12]。

本研究では換喩の検出と解釈を含めた換喩解析の内、主に換喩の検出について取り上げる。従来の換喩の研究としては、人手で作成された知識や意味ネットワークを利用する手法 [1] や新聞コーパスから収集した名詞句の用例や名詞の共起関係を統計的なデータとして利用する手法 [5, 11]、そして名詞を中心とした連想の情報から構築された連想概念辞書を利用する手法 [8] などがあるが、換喩の解釈についてのみ深く議論が成されている。村田ら [5] と須賀ら [8] は換喩の検出法にも触れているが、いずれも換喩表現を含んだ文のみを対象に実験を行っており、換喩表現を含まないリテラルな文も実験文と見なした上での検出の評価を行っていない。換喩表現を含む文は前述した「一升瓶を飲み干す」のように文法的には正しいため、本来解釈される意味とは全く違う場合でもコンピュータにとってはリテラルな文として捉えられる可能性が大きい。ゆ

えに自然言語処理の分野において、換喩を考慮した処理を行う場合は、処理する文は換喩表現を含んでいるのか、そして文中のどこが換喩表現に該当するのかを正しく検出することが、換喩表現を正しく解釈する以前に非常に重要であると考えられる。

先に述べたように換喩を特徴づける隣接性や近接性などをはじめとする関係は、言語的・認知的な制約により規制されるものもあるが、心理的な連想のリンクと密接に関わっていることが言える [12]。そのため、本研究では連想の関係という点に着目し、動詞を中心とした連想の情報から成る動詞連想概念辞書を用いて動詞と喩詞の関係性の有無を判別することで換喩の検出を図る。そして最後に、村田らや須賀らの研究で用いられた手法として日本語語彙大系 [2] の構文体系と意味体系を用いたベースラインを設定した上で、換喩表現を含む文とリテラルな文を含んだテスト文を用意し、それぞれの換喩検出精度を比較することで本研究の手法の有効性を確認した。

2 本研究で扱う換喩表現

換喩は1節で記したように、喩詞と被喩詞の関係が空間的な隣接性や時間的な因果関係などで特徴づけられている。また、時間的な前後関係や因果関係を事象の隣接性として捉えることで、換喩表現を特徴づける関係を空間的な隣接性と時間的な隣接性に大別することができる [9]。表1と表2は、前者と後者を換喩の事例とその解釈内容をそれぞれまとめたものである [12, 9]。ここからも分かるように後者の時間的な隣接性に関する換喩は、表に記した解釈内容とリテラルな文として解釈した場合の内容を比べると、それぞれの事象の関係は前後関係や因果関係があるため、文脈内での使われ方によって事象の捉え方が換喩表現なのかリテラルなのかが決まる。そのため、1文だけでは両方の立場での解釈が可能になってしまい判別が難しいことが考え

表 1: 空間的隣接関係の換喩表現と解釈の例

空間的隣接性	換喩表現と解釈
容器—中身	一升瓶を飲み干す 一升瓶に入った酒を飲み干す
原料—製品	アルコールを飲む 酒を飲む
主体—手段	白バイが捕まえる 白バイに乗った警察官が捕まえる
主体—付属物	詰め襟が歩く 詰め襟を着た学生が歩く
作者—作品	マーラーを聴く マーラーの曲を聴く
その他	理論が主張する 理論の提案者が主張する

表 2: 時間的隣接関係の換喩表現と解釈の例

時間的隣接性	換喩表現（解釈例）
原因—結果	杯を傾ける 酒を飲む 手を洗う トイレで用を足す

られる。村田らや須賀らの従来の研究においても、喩詞と被喩詞が時間的隣接性の関係を持つ換喩の文は扱われていないことから、本研究では時間的隣接性の換喩は扱わず、空間的隣接性に特徴づけられた換喩表現のみを対象とする。

3 関連研究と提案手法

従来の換喩解析の村田らと須賀らの研究では、文中に換喩表現が含まれていることを検出した上で、文中にある喩詞に対して被喩詞を補うことで換喩の解釈を行っていた [5, 8]。村田らは、日本語語彙大系から動詞の格フレーム辞書を作成し、各格要素の名詞の意味属性と入力文の名詞の意味的整合性を調べ、これを満たさない場合は換喩と判断している。尚、上位下位関係や意味属性を調べる際にも日本語語彙大系の意味体系と単語体系を利用している [5]。一方、須賀らは日本語語彙大系の構文体系、意味体系、単語体系を用いて作成したデータベースから入力文が文法的に正しいか否かを調べることで換喩を判断している [8]。しかし、いずれの研究も換喩表現が含まれている文のみを対象に実験を行っており、検出手法における定性的な評価まで至っていない。そこで、換喩を特徴づける特徴は心理的な連想のリンクと密接な関係があること [12] に注目し、本研究では動詞や名詞の連想の情報

に基づいて構築された動詞連想概念辞書と名詞連想概念辞書、そして心理学の立場で構築されたものを日本語に置き換えた日本語 WordNet [3] を用いて、従来の網羅的な構文情報や意味情報を含んだシンソーラスを用いた手法とは異なった、言葉の連想の情報を中心とした手法を提案する。本手法は、入力文に対して各文節における名詞とその係り先の動詞との関係性を判別することでその名詞が換喩における喩詞になっているかを判断する。この動詞と名詞の関係性を判断するに当たって、動詞連想概念辞書から得られた情報を元にして日本語 WordNet の synset の情報を用いる。そのため、以下では動詞連想概念辞書について概要を説明した後、詳細な説明を述べていきたい。

3.1 動詞連想概念辞書

動詞連想概念辞書は、動詞を刺激語した連想実験から得た情報を用いて構築しており、252 語の動詞に対して連想語数が約 101,000 語、異なり語数が約 24,000 語となっている [10]。連想実験では、刺激語を小学校の国語の教科書で扱われている動詞 [4] とし、「動作主」、「対象」、「始点」、「終点」、「時点」、「場所」、「手段」、「様態」、「理由」、「目的」の 10 個の深層格の内の 1 つが課題として一緒に提示されるため、実験参加者は刺激語動詞の深層格に入る単語を連想する。刺激語動詞 1 語につき実験参加者は 40 人であり、慶應義塾大学湘南藤沢キャンパスに所属する大学生と大学院生を対象に、延べ約 1200 人となっている。さらに名詞連想概念辞書 [6] と同様に刺激語 x と連想語 y の間の連想距離 $D(x, y)$ を式 (1) のように定量化している。尚、式 (1) の係数は連想距離が最小と最大を取り得る時の条件を境界条件に設定し、実験参加者が 20 人で刺激語動詞が 54 語の時の頻度と順位と連想時間の平均の値をそれぞれ目的関数に用いてシンプレックス法で求めた値である。

$$D(x, y) = \frac{7}{10}F(x, y) + \frac{1}{3}S(x, y) \quad (1)$$

$$F(x, y) = \frac{N}{n + \delta} \quad (2)$$

$$\delta = \frac{N}{10} - 1 (N \geq 10) \quad (3)$$

$$S(x, y) = \frac{1}{n} \sum_{i=1}^n s_i \quad (4)$$

$F(x, y)$ は刺激語 x に対して連想語 y が連想された頻度の逆数、 $S(x, y)$ は連想語 y が連想された順位 s_i の相加平均を表している。また $F(x, y)$ は補正值 δ を分母に加えることで正規化しており、参加者数 N を

表 3: 評価用入力文の例

換喩表現を含んでいる文	リテラルな文
一升瓶 を飲む	祭壇に一升瓶を置く
頭 を刈る	田んぼで稲を刈る
仮説 が原理を説明する	関係者が説明する
白バイ が違反者を逮捕する	警察官が容疑者を逮捕した
指揮者が クラリネット を笑う	バイオリンを弾く

大幅に増加させた時に連想した人数 n が少ない場合に $F(x, y)$ の極端な増加を防ぐことで、それに伴う連想距離 $D(x, y)$ の極端な変動を抑えることが可能となっている。

3.2 連想による換喩検出法

提案手法は入力文に対して、まず始めに形態素解析¹と構文解析²を行い、その結果から各文節内の助詞と係り先の用言の情報を得る。用言が動詞の場合は、動詞連想概念辞書の刺激語動詞と助詞と対応させた深層格 (e.g., 「～が」は「動作主」, 「～を」は「対象」など) から、連想語を連想距離が短い順に数語抽出し、それらの語に関して日本語 WordNet の synset を得る。入力文の文節内にある名詞に対しても synset を得た後に、上位の synset で重複があるかを調べ、重複がなければ換喩表現と判断する。また、用言が動詞でない場合は、名詞連想概念辞書の「属性概念」の関係をを用いる。これは、刺激語名詞「料理」に対して連想語「美味しい」など、刺激語名詞に対して連想語はその特徴を表している [6]。例えば「桜は満開だ」に対して、刺激語名詞「桜」に対して「属性概念」の連想語は「きれい」や「満開」などがある。用言が動詞ではなく形容詞や名詞の時「桜」の連想語が用言であれば、換喩表現と判断されない。そして文節内の名詞と、その述語の関係が「属性概念」で表されない場合は換喩表現と判断する。また文節内の名詞が固有名詞の場合は、構文解析の結果から「大阪」なら「地域」や「シェイクスピア」なら「人名」など、そのプロパティを名詞の意味として反映させ、WordNet の synset の比較を行う。

4 評価実験

提案手法の評価を目的に、以下では評価用入力文やベースラインなどの実験材用と実験結果を順に述べる。

¹MeCab 0.98pre3 を使用。

²CaboCha 0.60pre4 を使用。

4.1 実験材料

評価用入力文

実験に用いた入力文は、従来の研究で用いられてきた換喩表現を含む文 [12, 5, 11, 8] の他に新聞コーパス (CD-毎日新聞 93-95, 03-04) から抽出した換喩表現を含む文と、それを含まない文すなわちリテラルな文を各々45文、合計90文とした。このリテラルな文は、換喩表現の文中で使用されている単語を含んでいるものを優先させている。表3は、入力文の一部を表したものであり、下線部が換喩表現の部分である。本評価では、単語ごとに換喩表現が否かを判別する。合計90文に対して判別部分は全部で113箇所あり、その内換喩表現は45箇所である。

ベースライン

本手法と比較するベースラインは、従来の研究で扱われてきた日本語語彙大系を用いた手法とする。このシソーラスから村田らは格フレーム辞書を、須賀らはデータベースを各々作成していたが、ここでは汎用的な手法として日本語語彙大系のみデータを用いる。入力文に対して形態素解析と構文解析を行う段階までは本研究の提案手法と同じである。構文解析の結果を受けて文中の名詞の係り先の用言に関して、構文体系から構文情報が含まれているパターン対を抽出する。このパターン対に関しては優先順位に基づいて順番が決まっており [7]、入力文と同じ格を持つパターン対がある場合は最も優先順位の高いものを選ぶ。そして構文情報から格に入る名詞の意味属性の情報を得た後に入力文に対応する名詞の意味属性の上位と、全く重複しない場合は換喩表現があると判断する。

4.2 実験結果と考察

表4は、換喩表現の検出に関して本研究の提案手法と従来の手法を元にしたベースラインの実験結果をまとめたものである。表の適合率と再現率は、「90文の入力文に含まれる113箇所の中で換喩と判定したもの、正しく換喩と判定できた割合」と「45文に含まれる45箇所の換喩表現の内、正しい換喩と判定できた割合」をそれぞれ示している。両方の数値とも提案手法がベースラインよりも高く、これらから求められるF値もベースラインを大きく上回っている。

村田らの研究では、換喩表現を含んだ23文の内17文が換喩であると判定していることから、全ての文が換喩表現を含んでいる状況での検出精度は74.9%と捉えることができる [5]。本研究の実験においても、全ての文が換喩表現を含んでいる状況 (i.e., 換喩表現を含

表 4: 提案手法とベースラインの比較結果

	ベースライン	提案手法
換喩検出の適合率 (%)	63.3(31/49)	72.0(36/50)
換喩検出の再現率 (%)	68.9(31/45)	80.0(36/45)
F 値	66.0	75.8

む 45 文のみ)での検出精度を求めると,表 4 の再現率と同義となるため,提案手法は 80.0%となり 5%程高い.そのため入力文の総数の違いも考慮すると,従来の手法よりも提案手法が有効であることが分かる.

最後に,提案手法とベースラインの結果から連想による換喩検出とシソーラスによる換喩検出の違いについて言及したい.ベースラインでは検出できず,提案手法では検出可能だった換喩表現の例としては,表 3 の「指揮者がクラリネットを笑う」や「一升瓶を飲む」というのが挙げられる.従来の手法では,「笑う」や「飲む」のヲ格に入る名詞の意味属性が全ての意味属性が当てはまる「*」と「具体」³とそれぞれなっており,「クラリネット」や「一升瓶」の意味属性が上位を辿ると重複するために換喩表現として判断できなかった.確かに,これらはリテラルな文としても捉えることはできるが,我々の経験上では「指揮者がクラリネットという楽器自体を笑う」ことはまだあるかもしれないが,「一升瓶を瓶ごと飲む」ということは全く考えられないことである.つまり,シソーラスを利用することでリテラルな文として可能な限り見なすことができるようになり,我々の常識を超えた意味も可能にしてみうため,日常的に用いられる換喩表現の検出に沿わない場合がある.このようなことから,提案手法は連想の情報をを用いることで,ベースラインと比べて我々の常識により近い範囲で換喩表現の検出を可能にしていると考えられる.

5 おわりに

本研究では,連想による換喩の検出法を提案し,従来の研究のシソーラスを用いた手法をベースラインに設定して評価実験を行なった.その結果,提案手法の F 値が 75.8 であり,ベースラインの 66.0 よりも高い精度を得ることができ,有効性を示すとともに換喩の検出における連想の長所を確認した.今後は換喩の解釈についても取り組み,検出と合わせて換喩解析に取り組むつもりである.

³ベースラインの手法では優先順位が一番高いパターン対を選ぶ.「具体」を「飲む」(i.e., engulf)のパターン対が「液体」を「飲む」(i.e., drink)のパターン対よりも優先順位が高いため前者が選択されている.

参考文献

- [1] Jacques Bouaud, Bruno Bachimont, and Pierre Zweigenbaum. Processing metonymy: a Domain-Model Heuristic Graph Traversal Approach. In *Proceedings of the 16th International Conference on Computational Linguistics*, Vol. 1, pp. 137–142, 1996.
- [2] 池原悟, 宮崎正弘, 白井諭, 横尾昭男, 中岩浩巳, 小倉健太郎, 大山芳史, 林良彦. 日本語語彙大系 CD-ROM 版. 岩波書店, 1999.
- [3] 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.
- [4] 森田良行. 基礎日本語辞典. 角川学芸出版, 1989.
- [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] 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.
- [7] 白井諭, 横尾昭男, 中岩浩巳, 渡邊いづみ, 高橋直美, 関嘉美, 池原悟, 宮崎正弘. 構文意味辞書と構文体系. 言語処理学会第 4 回年次大会発表論文集, pp. 194–197, 1998.
- [8] 須賀智秋, 石崎俊. 連想概念辞書を用いたメトニミー意味解析システムの構築. 言語処理学会第 12 回年次大会発表論文集, pp. 817–820, 2006.
- [9] 谷口一美. 認知意味論の新展開. 研究社, 2003.
- [10] Takehiro Teraoka, Jun Okamoto, and Shun Ishizaki. An associative concept dictionary for verbs and its application to elliptical word estimation. In *Proceedings of the 7th International Conference on Language Resources and Evaluation*, pp. 3851–3856, 2010.
- [11] 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.
- [12] 山梨正明. 比喩と理解. 東京大学出版会, 1988.