

概要

機械翻訳では、比喩文の翻訳が問題となっている。比喩をそのまま翻訳しても目的言語の表現からは、その意味が解釈できない。そこで翻訳の前段階として、比喩判定を行ない独自の比喩解釈により翻訳を行う必要がある。

比喩とは、あることがらを表現するために、そのことがらとなんらかの関係があることがらに喩えて文を表現することである。比喩には、様々な分類があるが本研究では、中村明の分類に着目する。中村は、比喩を大きく3分類しており、中でも結合比喩は、語と語の異常性という観点から分類されたものである。語と語の正常な関係は、日本語語彙体系でまとめられており、計算機処理が可能であることと比べると、中村分類も計算機処理になじみやすいといえる。

本研究では、まず中村分類の中でも結合比喩に属する比喩の自動判定を目指す。結合比喩の一つに、名詞と動詞の関係に異常が見られるものがある。名詞と動詞の関係が異常であるのは、非文である。つまり、そこに、ある特定の比喩の条件が加われれば、結合比喩が成り立つ。しかし、ある特定の条件を判別することは、不可能に近い。そこで、本研究では類似文検索に注目し、既存の比喩から類似的な文を取得することにより、比喩判定を行う方法を提案する。本手法では、類似文検索の一つであるベクトル空間法を参考にする。今回、用言を見出し語としたベクトルを定義し、その基底に格要素(名詞+格助詞)を使うことにより、既存の比喩から構造的に近い比喩が検出することが期待できる。さらに、名詞は意味属性を使用している。これは単語において字面での類似文検索をするよりも、意味属性なら同義語、類義語などにおいて、字面では不可能なことに対応できる。

既存の比喩として、中村明の結合比喩例文 3,750 文を用いた。これらは、本研究での DB として使用される。判定結果から適合率が 96%、再現率が 27%の結果が得られ類似文的比喩判定の効果が証明された。

目次

1	はじめに	1
2	比喩	2
2.1	比喩定義	2
2.2	比喩分類	2
2.3	中村明の比喩分類	4
2.4	先行研究における比喩判定	4
2.4.1	顕現性落差	4
2.4.2	意外性	5
2.5	本研究で扱う比喩	5
3	ベクトル空間を用いた比喩判定方法	6
3.1	比喩の判定条件	6
3.2	類似文検索	6
3.3	比喩判定用 DB 文	6
3.4	特性ベクトル	7
3.4.1	特性ベクトル定義	7
3.4.2	一般名詞意味属性体系	8
3.4.3	ベクトル次元縮退方法	9
4	実験	10
4.1	閾値調査	10
4.2	比喩判定実験	11
4.2.1	閾値による判定	11
4.2.2	類似文検索に基づく相対的な比喩判定	12
4.2.3	総合的評価	12
5	考察	14
5.1	結合価パターンのベクトル化における問題	14
5.2	慣用句を DB 文とした際の効果	14
5.2.1	閾値調査	14
5.2.2	比喩判定実験	15
5.2.3	総合的評価	16
6	おわりに	18

目 次

1	ベクトル例	7
2	一般名詞意味属性体系	8
3	ベクトル例 (次元数縮退後)	9
4	内積統計 (比喩文)	10
5	内積統計 (慣用句)	14

表目次

1	判定結果 1	11
2	判定結果 2	12
3	判定された比喩文	13
4	判定結果 3	15
5	判定結果 4	16
6	判定された比喩文 (慣用)	17

1 はじめに

機械翻訳において、問題の一つに比喩がある。比喩を翻訳をする上で、翻訳機が意図した通りに訳してくれない場合がある。例えば「彼は鍋を食べる」という例文があるとするとこの例文は、現在の翻訳機においては、“He eats a pot”とし、現状のまま普通に「鍋」そのものを食べるという風に解釈されてしまう。人間は、これを意識的に「鍋料理」と理解して「鍋料理を食べる」と解釈できる。そこでもし比喩かどうかの判定ができるのであれば、先の例文も比喩と判定し、なんらかの計算機における解釈により翻訳ができるのではないかと考えることができる。つまり、先程の例でいうと、“He eats a pot cooking”と翻訳することができる。本研究では、比喩翻訳における初段階として比喩判定を行う。

比喩分類は様々であるが、中村明は、比喩の中で語と語の関係に異常性がみられるものを結合比喩と分類している [3]。比喩の自動判定の先行研究において、名詞と名詞の異常性から比喩を判定することは、梶井らや、内海により行われている [1][2]。

梶井らは、名詞を“概念”とし、それらが持つ特徴を求め、2つの概念同士の意外性や、二つの概念が共有する特徴つまり形や味など様々なイメージがその概念にいかにか斬新であるかによって比喩性の尺度を求めている。そして落差などを計算モデルに表し、計算機上で判定できるようにしている。しかし、これまでに名詞と動詞の異常性からの比喩判定をすることはなされていない。同様の方法でするには、動詞のイメージ的な特徴を網羅するには調査が必要となり、現段階では比喩性の判定を行うことは困難である。

正常文においての名詞と動詞の関係は、語彙体系によって定義づけがなされている。語彙体系でマッチしないのは、非文と考えられる。つまり、非文であるということは、名詞と動詞の関係が異常であるということで、そこである特定の比喩の条件が加われば、結合比喩になる。しかし、特定の条件を見分けることは、不可能に近い。そこで本研究では、名詞と動詞の関係に異常性が見られる結合比喩について、既存の比喩文との類似性を、格要素と用言の共起関係から作成した意味ベクトルにより比較することで、自動的に判定することを目的とする。

その判定方法として、本研究では2つの方法を提案する。一つは、絶対値である判定の閾値を予備実験などで求め、入力文のベクトルが、閾値を超えるかどうかで判定する方法である。もう一つは、相対的に入力文が正常文のベクトル空間と比喩文のベクトル空間のどちらに距離的に近いことによって比喩性を判定する方法である。最終的に、この2つの方法により求めた比喩を総合的に見て結果を出す。その結果、適合率が96%、再現率が27%であった。

本論文の構成は以下のとおりで、第2章で比喩と分類および先行研究について述べ、第3章で本研究での比喩判定方法について述べ、第4章で実験を行う、第5章では、考察として慣用句を考慮した判定方法を試みる、第6章でまとめる。

2 比喩

本章では、比喩について説明し、語学的な比喩分類と本研究で使用した分類について説明する。そして、先行研究における比喩判定と本研究での目標を述べる

2.1 比喩定義

比喩とは、広い意味では、あることがらを表現するために、そのことがらと何らかの関係のあるもう1つのことがらにたとえて表現する働きである、またその働きの結果である表現そのものを指すことができる。

例 彼はオオカミだ。

「彼」というのは、「彼」という字面通りのものを表現する一方、オオカミの獰猛さと関係付けられており、「彼は獰猛である」という意味を表している。

2.2 比喩分類

比喩には、様々な分類がなされている。分類は、言語学において様々な提案がなされている [9]。分類を7つのグループに分けて説明する。

直喩と隠喩

直喩や隠喩とは、喩えるものと喩えられるものにある類似性がみられるもので、直喩では、比喩である単語が含まれている。例えば「バナナのような女の子」ならば「のような」が比喩であることを示す単語である。一方隠喩には、それが示されず暗黙のうちに表現されているものである。

提喩と換喩

換喩や提喩とは、喩えらえるものを喩えるものに関係した他のものによって言い変えて表すものである。換喩は2つの間に隣接関係があるものをさす。例えば「鍋を食べる」では、鍋そのものではなく鍋と隣接関係にある鍋の中身を表している。提喩は、包含関係があるものをさす。例えば「花見に行く」では、花という全体の名詞を桜という部分の名詞で表している。

活喩と物喩

活喩と物喩とは、喩えるものと喩えられるものを擬人化したり戯物化したりして表すものである。活喩は無生物を生物に喩える方法である。例えば「砂漠は生きている」では、砂漠をあたかも生物に喩えて表現している。逆に物喩は、生物を無生物に喩える方法である。例えば「彼は、雲のように生きている」では、彼の生き方を無生物に喩えて表現している。

諷喩と引喩

諷喩と引喩とは、喩えるものだけを言語面に明示したもので、喩えられるものだけをなんとなく感じとらせるものである。諷喩は、主にことわざのように文全体が比喩となり教訓含むものである。例えば「花より団子」などである。引喩は、文の一部に他人の文(ことわざ、格言、などつまり諷喩)をはさみ込むもので、つまり自分の文章を着飾るためだけのものである。例えば「やっぱり花より団子でいくかね」なのである。

張喩と類喩

張喩と類喩とは、事物の特徴となる点をおおげさに表現し、印象を強めるものである。張喩は、表現したいものの特徴となる点を誇張して表現するものである。例えば「ワニのような蟻をみた」というものである。類喩は、表現したい物を関連単語を文章中にまぜて表現したいものである。例えば、炎を表現したいなら文章中に「赤々と、熱い」のような単語をちりばめる。

声喩と字喩

声喩と字喩とは、喩えられるものが、語句本来の概念や意味に関係しないものである。声喩は、喩えられるものを語音つまり音声的に表す方法である。例えば「ハハハと笑う」なのである。字喩は、文字の形などからその文(単語)を表す方法である。例えば「酒」は「水辺の酉」などである。

詞喩

詞喩とは、掛けことば、語呂などである。

2.3 中村明の比喩分類

2.2 節で説明した比喩分類は，細かく分類されているが，分類同士で重複するところがあり研究対象として扱いきれていない．一方，語学的分類の中でも中村明は，比喩を大きく3分類している [3]．3つの分類自体に深い繋がりが無いために，他の分類を考慮にいかず，単体で考えることができ，一つの分類を研究対象にすることができる．

指標比喩

比喩の成立に関わってくる語が含まれるもの．ちなみにその語を指標表現という．

例 狼のような男

この場合『のような』が指標表現である．

結合比喩

語と語の間になんらかの異常性がみられるもの．

例 心にふれる

この場合『心』と『ふれる』に結合的な異常がみられる．

文脈比喩

指標比喩，結合比喩ではなく，文章としては正しいが文の意味しているものが文脈に関係して違う意味を類推してしまうもの．

2.4 先行研究における比喩判定

比喩とは，対象となる物を表現するのにその対象となんらかの関係にあるものに喩えることである．そこで榊井らは，2つの概念の間になんらかの関係があり，「2つの概念関係における特徴の明確さ」および「2つの概念の斬新さ」を判定基準としている．つまり「彼はオオカミだ」という例文があるなら「彼」と「オオカミ」の間の獐猛という特徴がいかに明確で「彼」と「オオカミ」の組合せがどれほど斬新かによって比喩性の判定をしている！その関係の特徴がいかに明確かを“顕現性落差”とし「2つの概念の斬新さ」を“意外性”という尺度としている．

2.4.1 顕現性落差

顕現性落差とは，2つの概念でクローズアップされる特徴を抽出し，それらの特徴がいかに明確であるかを量る尺度である．概念の特徴を属性とよぶ．一般的に概念は属性値の集合を持っている．顕現性落差を計算機上で扱うようにするためには，顕現性落差を定量化する必要性がでてくる．[13]らは，この概念を属性値とその生起確率の集合で表している．概念 W が属性値 w_i をもち，生起確率が p_i であるとき， W は，

$$W = \{p_1\#w_1, p_2\#w_2, \dots, p_i\#w_i, \dots, p_m\#w_m\} \quad (1)$$

但し $\#$ は、その属性値にたいする生起確率を表す。

これから，概念の顕現性は，属性値集合の冗長度から判定できる．(式 1) より，概念 W が， m 種類の属性値からなる属性集合として記述される場合，冗長度 $r(W)$ は，(式 2) より求められる．

$$r(W) = \begin{cases} 1 - \frac{\sum_{i=1}^m p_i \log \frac{1}{p_i}}{\log m} & (m \neq 1) \\ 1 & (m = 1) \end{cases} \quad (2)$$

顕現性落差を求めるには，2つの概念間の共有特徴，つまり，対象となる概念を軸とした特徴を求める．それにより顕現性落差を考えるための共有属性を選びだされる．そこで2つの共起確率を用いた冗長度を計算し，比較することで顕現性落差が求められる．この手法は，顕現性落差が高い程比喩性が高い．

2.4.2 意外性

比喩での意外性とは，概念の組み合わせの斬新さである．そこで，いかに斬新であるかを定めるためには，日常的に用いられているテキストデータからどれだけ頻繁に共起しているかを見る必要がある．よって，テキストデータの単語情報に基づく意味距離を利用すれば，単語2つの単語の共起頻度を用いて意外性を定量化することができる．梶井らは，計算対象となる頻度が小さい場合でも，比較的信頼性できる dice 係数を使用している．dice 係数とは，単語間の意味距離を示す値であるので，結びつきが強いほど値が大きくなる．そこで，意外性という観点から，逆数を求めることにより意外性が得られる．したがって，単語 W_s, W_t の頻度を p_s, p_t とし，両者の共起頻度を $p_s \cdot p_t$ とするならば，意外性 $Nov(W_s, W_t)$ は，(式 3) となる．この手法は，意外性が高いほど比喩性が高い．

$$Nov(p_s, p_t) = \frac{p_s + p_t}{2(p_s \cdot p_t)} \quad (3)$$

2.5 本研究で扱う比喩

本研究で扱う比喩は，格要素と用言の比喩独自の繋がり，つまり語と語の繋がりを対象としている．そこで中村明の分類の一つである結合比喩を研究の対象にする．しかし，格要素と用言の関係からは，概念同士の比較により比喩判定を行うことはできない．そこで本手法として，先行研究においての研究でカバーできない比喩について比喩判定を目指す．

3 ベクトル空間を用いた比喩判定方法

本章では，類似文検索方法に基づいた比喩判定方法について述べる．

3.1 比喩の判定条件

語と語の関係が正常でない文は，非文といわれている．しかし，全ての非文が結合比喩とは限らない．すなわち，[4]における結合価パターンに適合しない文であり，かつ，ある特定の関係のあるものが，結合比喩の文であるといえる．

そこで，本研究では，既存の結合比喩文に類似すること，という条件の下で結合比喩の判定を行う．

3.2 類似文検索

既存の比喩文と類似するかどうかを判定するために，類似文検索の方法を用いる．その検索方法の一つにベクトル空間法がある．ベクトル空間法による類似文検索では，文中に含まれるある要素を基底とする特性ベクトル (式 4) で文を表す．

$$\vec{V} = (w_1, w_2, \dots, w_i, \dots, w_n) \quad (4)$$

データベース文におけるベクトル (\vec{V}_i) と，検索キーとなる文のベクトル (\vec{V}_j) とを，ベクトル間の距離 ($sim(D_i, D_j)$) により類似性を判定する方法である．

$$sim(D_i, D_j) = \vec{V}_i \cdot \vec{V}_j \quad (5)$$

但し \vec{V}_i, \vec{V}_j は，検索キーとなる文 D_i ，データベース文 D_j の特性ベクトルを表す．

3.3 比喩判定用 DB 文

比喩判定のための類似文検索に用いるデータベース (DB) の文を「比喩判定用 DB 文」と呼ぶ (DB 文と略記する) ．

今回，DB 文として，2つの DB 文を用意する．一つは，結合比喩の例文，もう一つは，結合パターンである．結合比喩は，[3]で紹介されている．その中から，3,750 文を DB 文として登録する．結合パターンは，慣用句のパターンを抜いた 11,481 パターンを DB 文として登録する．

なお，DB 文を検索する際，用言を見出し語として検索する．そして，類似性判定には，その検索で得られた文の特性ベクトル (次節参照) を用いる．

3.4 特性ベクトル

3.4.1 特性ベクトル定義

名詞と動詞の異常性をとらえるため，特性ベクトルは，格要素（名詞および格助詞）と用言の共起関係により定義する．

この共起関係から，格助詞が付属している名詞の意味が用言を用いる際，適切であるかがわかる．たとえば「太郎がコップを（壊す）」という文には「が格」と「を格」があり「が格」には動作可能な名詞（主体）が「を格」には具体物がそれぞれ対応している．「壊す」という用言において，これらの組合せは正常である．

さて，結合価パターンを参考にすると，格助詞は 13 個，名詞の意味属性は 2,710 個である．特性ベクトルは，格助詞毎の意味属性の使用可否を 0 / 1 で表すと， $13 \times 2,710$ 次元になる．

ここでベクトル定義を例に基づき説明する．

例 「彼 (23,48) は苗木 (673,689) を植える」

用言として「植える」は見出し語になる．次に格要素は「彼 (23,48) は」と「苗木 (673,689) を」となり「23」と「48」の「は」格「673」と「689」の「を」格が 1 となる．他は，0 となる．

格助詞	他格												
	で格												
	と格												
	の格												
	も格												
	に格												
	まで格												
	へ格												
	より格												
	から格												
	を格								1		1		
	は格				1		1						
	が格												
	1	2	...	23	...	48	...	673	...	689	...	2715	
	意味属性												

図 1: ベクトル例

3.4.2 一般名詞意味属性体系

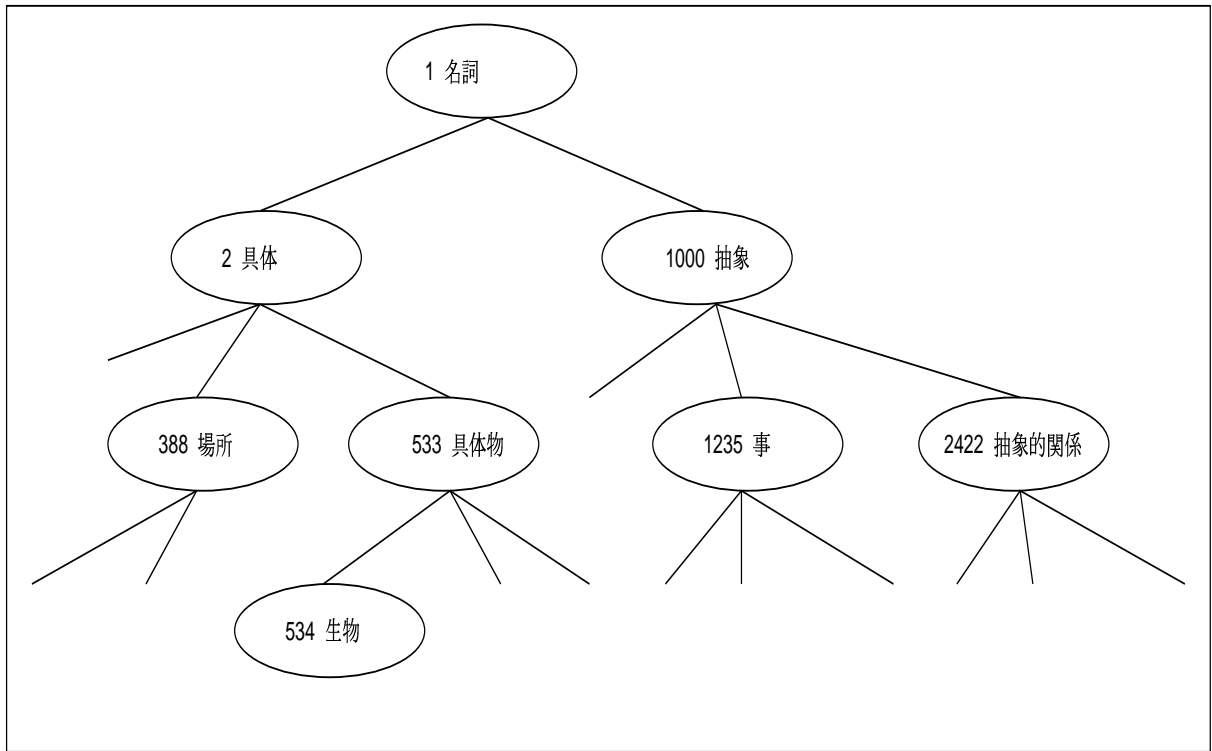


図 2: 一般名詞意味属性体系

特性ベクトルの基底に名詞の意味属性を使用している。意味属性とは、日本語名詞の意味的な用法を約 2,710 種類に分類したものである。意味属性相互間の意味的關係を最大 12 段の木構造で表現している。

また、単語意味属性において、日本語名詞(約 30 万語)に対して、それぞれの語がどのような意味属性を持つか(1 つ以上)が規定されている。

3.4.3 ベクトル次元縮退方法

定義したベクトルでは，スパースなベクトルになるので，次の考えから次元を縮退させる．

まず，格助詞は『が』『は』は主格『を』は対象『から』『より』は起点『へ』『まで』『に』は終点として5次元にする．次に，意味属性は，木構造であり，上位ノードの意味属性を下位ノードが継承していることから，最下位ノードだけをベクトルの次元とする．したがって，名詞が中位ノードを指定している場合には，それを継承している最下位ノードがベクトル化に使われる．こうして， $5 \times 1,921$ 次元のベクトルを用いる．

先程の例に基づいて説明する．

例 「彼 (23,48) 『主体』 苗木 『対象』 植える」

方法は，前章で説明した通りである．

格助詞	その他											
	終点											
	起点											
	対象							1	1		1	
	主体				1		1					
		8	9	...	23	...	48	...	674	675	...	689
意味属性												

図 3: ベクトル例 (次元数縮退後)

4 実験

本章では、まず始めに、比喩判定における閾値の調査を行う。閾値は、比喩判定をする上で必要となり、今回内積値とその頻度により決定する。次に、オープンテストとして比喩判定の実験を行う。判定実験は、2種類行う。

4.1 閾値調査

類似文検索による比喩判定のために閾値の調査を行う。閾値は内積値の頻度により決定する。今回入力として、正常文と比喩文を用意し、類似度の統計をとり正常文が極端にヒットしない点を閾値を設定する値としたい。

(1) 入力

正常文は [7] から 1,141 文そして比喩文は [3] から 965 文について DB 文との類似度 (式 6) を求める。

$$\text{類似度} : \text{sim}(V_i, V_h) = \vec{V}_i \cdot \vec{V}_h \quad (6)$$

但し \vec{V}_i は入力文 \vec{V}_h は結合比喩 DB 文の特性ベクトル

(2) 結果

図 1 に集計結果を示す。表の横軸は類似度、縦軸は頻度である。なお、図中の慣用句については 4.2 節で説明する。

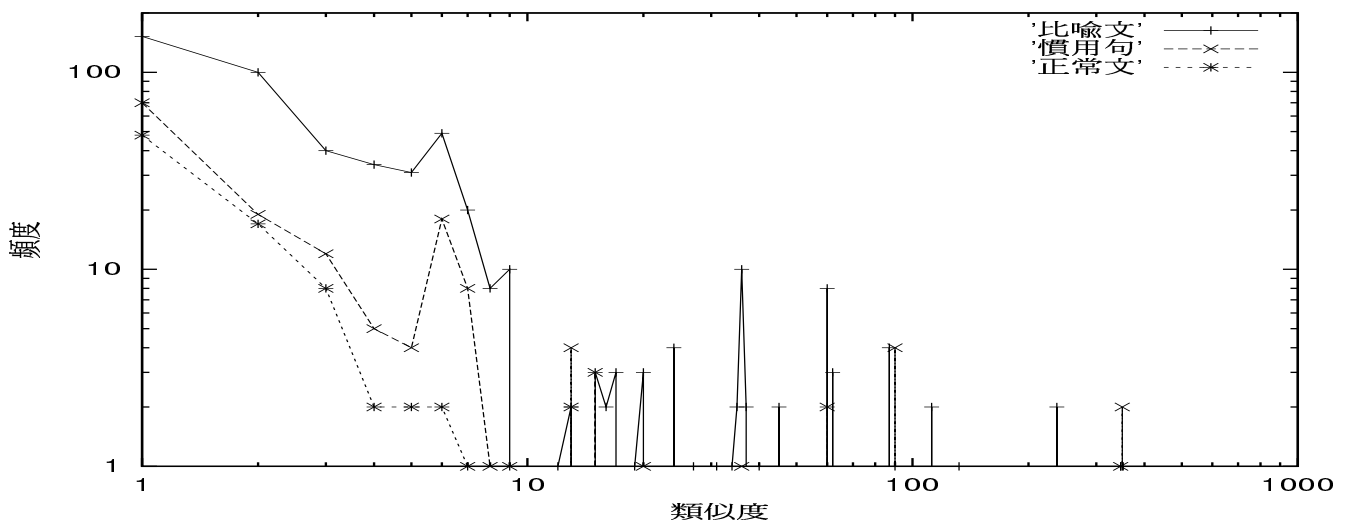


図 4: 内積統計 (比喩文)

類似度 3 の地点において正常文が 10 以下のヒット数である。これは、入力文に対して 0.1% 以下である。本研究では、これ以降をあまり当たらない事として、閾値を 3 に設定する。

4.2 比喩判定実験

比喩判定のオープンテストを行う。オープンテストは、2種類行う。一つは閾値を絶対値として入力文を判定する方法。もう一つは、2つのDB文(結合価パターン, 結合比喩)を用意し、入力文が正常文に近いか比喩文に近いかを、相対的に見る方法である。

4.2.1 閾値による判定

比喩判定をする上で、入力文とDB文の類似度を見るのだが、距離から比喩性を判定するのは困難である。そこで先の調査で求めた閾値(正常文が比喩文DB文に極端にヒットしなくなる値)を使用することにより、ある程度の比喩文のみの判定が可能になると予想する。

(1) 入力

実験用のデータとして、閾値の時に用いた例文とDB文に使用した例文と同様でない文を、[7]より100文、[3]より100文、合計200文用意した。入力文は、付録に添付した。

(2) 判定方法

入力文と結合比喩DB文との類似度を4.1章の(式6)と同様の方法で求め、閾値 T を超えた場合その入力文は比喩と判定する(式7)。

$$T \leq \text{sim}(V_i, V_h) \quad (7)$$

但し \vec{V}_i は入力文 \vec{V}_h は結合比喩DB文の特性ベクトル

(3) 判定結果

比喩文の出力結果は、表3に示す。

表 1: 判定結果 1

入力	出力結果	
	正常文	比喩文
正常文	100	0
比喩文	82	18

$$\text{適合率} = \frac{\text{比喩と判定された文の中に比喩が含まれる数}}{\text{比喩と判定された数}} = \frac{18}{18} = 100\% \quad (8)$$

$$\text{再現率} = \frac{\text{判定された比喩の内、比喩であった数}}{\text{入力された比喩数}} = \frac{18}{100} = 18\% \quad (9)$$

4.2.2 類似文検索に基づく相対的な比喩判定

先の実験では、閾値を絶対値とした判定方法であったが、次の判定方法として入力文が正常文のベクトル空間と比喩文のベクトル空間においてどちらの距離が近いかによって判定する方法である。これは、相対的な判定方法であり、類似文検索では主に用いられる。

(1) 入力

先の閾値による判定実験と全く同一の 200 文を用意した。

(2) 判定方法

入力文が結合価パターン DB 文と比喩文 DB 文のとの 2 つの類似度を 4.1 節の (式 6) で求める。そして、比喩文の DB 文の方が類似度が高ければ比喩と判定する (式 10)。

$$\text{sim}(V_i, V_s) < \text{sim}(V_i, V_h) \quad (10)$$

但し \vec{V}_i は入力文 \vec{V}_s は結合価パターン DB 文 \vec{V}_h は結合比喩 DB 文の特性ベクトル

(3) 判定結果

比喩文の出力結果は、表 3 に示す。

表 2: 判定結果 2

入力	出力結果	
	正常文	比喩文
正常文	99	1
比喩文	76	24

判定結果より、適合率は 96% で再現率は 24% であった。

4.2.3 総合的評価

2 つの判定方法により、比喩が判定されたが、判定された比喩文を比較すると、一致した文が多数あるのだが、片方のみに検出された文もあり、それを総合的に見ると最終的に、比喩文は 27 文検出された。つまり再現率は、27% である。同様に、正常文も見つかる必要があるため、最終的な適合率は、96% である。判定結果は、表 3 に示す。

表 3: 判定された比喻文の一覧

	閾値による判定	相対的な判定	総合的な判定
No.01	考えが心に触れる。	胸のなかでわめく。	胸のなかでわめく。
No.02	眼に場面が甦る。	考えが心に触れる。	考えが心に触れる。
No.03	考えが見せる。	眼に場面が甦る。	眼に場面が甦る。
No.04	美が自分を護る。	不平が声をあげる。	不平が声をあげる
No.05	気持がさせる。	考えが見せる。	考えが見せる。
No.06	頭へものをぶちこむ	気持が捉える。	気持が捉える。
No.07	吃りを脱ぎ去る。	美が自分を護る。	美が自分を護る。
No.08	人々を吐き出す。	気持がさせる。	気持がさせる。
No.09	胸に情熱をかくす	吃りを脱ぎ去る。	吃りを脱ぎ去る。
No.10	足音を空に吸う。	人々を吐き出す。	人々を吐き出す。
No.11	ものが作用をよぶ	胸に情熱をかくす。	胸に情熱をかくす。
No.12	左右に心がふれる	月を雲に呑む。	月を雲に呑む。
No.13	皮膚に感触が蘇る。	足音を空に吸う。	足音を空に吸う。
No.14	帯が絹の叫びをあげる。	美がそこにおる。	美がそこにおる。
No.15	争奪欲がそうさせる。	左右に心がふれる。	左右に心がふれる。
No.16	行為が半生をひきまわす。	戦慄が肌を走る。	戦慄が肌を走る。
No.17	顔に味がある	ラッパが這う。	ラッパが這う。
No.18	気持が捉える	気がやってくる。	気がやってくる。
No.19		音がしみ入る。	音がしみ入る。
No.20		危篤が電話で騒ぐ。	危篤が電話で騒ぐ。
No.21		胸の底でわめきちらす。	胸の底でわめきちらす。
No.22		皮膚に感触が蘇る。	皮膚に感触が蘇る
No.23		帯が絹の叫びをあげる。	帯が絹の叫びをあげる。
No.24		争奪欲がそうさせる。	争奪欲がそうさせる。
No.25		行為が半生をひきまわす。	行為が半生をひきまわす。
No.25			頭へものをぶちこむ
No.26			ものが作用をよぶ
No.27			顔に味がある
No.28			気持が捉える

5 考察

5.1 結合価パターンのベクトル化における問題

予備実験として、結合価パターンと比喻文のマッチングを本特性ベクトルを用いて試みた。その結果、約6割がマッチしていた。その原因の内訳として、中村の結合比喻と語彙体系の収録用言の重複、判定方法の問題が考えられる。基本的には、結合価パターンでは、使用する格要素全ての条件が満たされなければならないが、ベクトルを用いた場合は、少なくとも一つの要素で条件が満たされれば一致したことになる。

今後の課題として、結合価パターンと比喻文の厳密なマッチングアルゴリズムの追加があげられる。

5.2 慣用句をDB文とした際の効果

慣用句の多くは、単語単独で眺めた際と異なる意味で単語が使われている。これは結合比喻に近い。したがって、[5]で紹介されている725文を含む慣用句のDBを作成する。

5.2.1 閾値調査

入力文として、[7]の例文と[3]の例文に今回[5]の慣用句725文を加える。4.1節の式(6)で、入力文とDB文との類似度を求める。図5に結果を示す。

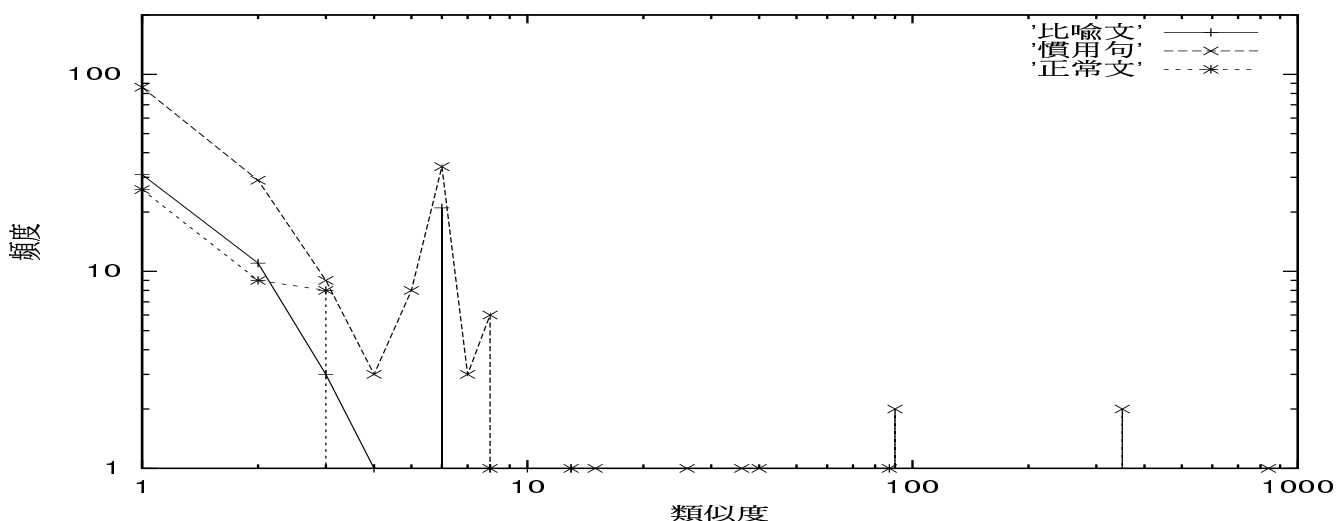


図 5: 内積統計 (慣用句)

比喻文の閾値調査の時と同様に、閾値3の地点において正常文が当たらなかった。

5.2.2 比喩判定実験

慣用句を加えた場合，同様に 2 つの実験を行い，総合的な判定を行う．

閾値による判定結果

(1) 入力

4.2 節の判定実験で使用された，合計 200 文を用意した．

(2) 判定方法

入力文と結合比喩 DB 文，慣用句 DB 文の類似度を 4.1 節の (式 6) の方法で求め，どちらかが閾値 T を超えた場合，その入力文は比喩と判定する (式 11,12) ．

$$T \leq sim(V_i, V_h) \quad (11)$$

$$T \leq sim(V_i, V_k) \quad (12)$$

但し \vec{V}_i は入力文 \vec{V}_h は結合比喩 DB 文 \vec{V}_k の特性ベクトル

(3) 閾値による判定結果

比喩文の出力結果は，表 6 に示す．

表 4: 判定結果 3

入力	出力結果	
	正常文	比喩文
正常文	100	0
比喩文	82	18

判定結果より，適合率が 100%，再現率が 18%であった．

類似文検索に基づく相対的な比喩判定

(1) 入力文

前節と同じ 200 文を用意した。

(2) 判定方法

入力文が結合価パターン DB 文と比喩文の DB 文と慣用句の DB 文との 3 つの距離を 4.1 節の (式 6) の方法でとり，比喩文 DB 文あるいは慣用句 DB 文が結合価パターン DB 文より内積が大きければ比喩と判定する (式 13,14)。

$$\text{sim}(V_i, V_s) < \text{sim}(V_i, V_h) \quad (13)$$

$$\text{sim}(V_i, V_s) < \text{sim}(V_i, V_k) \quad (14)$$

但し \vec{V}_i は入力文 \vec{V}_s は結合価パターン DB 文 \vec{V}_h は比喩文 DB 文 \vec{V}_k は慣用句 DB 文の特性ベクトルである

(3) 判定結果

比喩文の出力結果は，表 6 に示す。

表 5: 判定結果 4

入力	出力結果	
	正常文	比喩文
正常文	96	4
比喩文	75	25

判定結果より，適合率が 86%，再現率が 25%であった。

5.2.3 総合的評価

閾値による判定による文と相対的な判定による文の論理和をとることによる総合的な評価は，適合率は 87%であり再現率は 27%であった。これは，慣用句の位置付けを正常文と比喩文の中間的なものとしており，その結果が適合率を下げたと考えられる。今回，慣用句の DB 文を用いても判定率の向上はみられなかった。

表 6: 判定された比喩文の一覧 (慣用句の類似性を考えた場合)

	閾値による判定	相対的な判定	総合的な判定
No.01	考えが心に触れる。	彼は旗を左右に振った。	彼は旗を左右に振った。
No.02	眼に場面が甦る。	胸のなかでわめく。	胸のなかでわめく。
No.03	考えが見せる。	彼女は魚に包丁を入れた。	彼女は魚に包丁を入れた。
No.04	美が自分を護る。	顔におできが出来た。	顔におできが出来た。
No.05	気持がさせる。	考えが見せる。	考えが見せる。
No.06	頭へものをぶちこむ	気持が捉える。	気持が捉える。
No.07	吃りを脱ぎ去る。	美が自分を護る。	美が自分を護る。
No.08	人々を吐き出す。	気持がさせる。	気持がさせる。
No.09	胸に情熱をかくす	吃りを脱ぎ去る。	吃りを脱ぎ去る。
No.10	足音を空に吸う。	人々を吐き出す。	人々を吐き出す。
No.11	ものが作用をよぶ	胸に情熱をかくす。	胸に情熱をかくす。
No.12	左右に心がふれる	月を雲に呑む。	月を雲に呑む。
No.13	皮膚に感触が蘇る。	足音を空に吸う。	足音を空に吸う。
No.14	帯が絹の叫びをあげる。	美がそこにおる。	美がそこにおる。
No.15	争奪欲がそうさせる。	左右に心がふれる。	左右に心がふれる。
No.16	行為が半生をひきまわす。	戦慄が肌を走る。	戦慄が肌を走る。
No.17	顔に味がある	ラッパが這う。	ラッパが這う。
No.18	気持が捉える	気がやってくる。	気がやってくる。
No.19		音がしみ入る。	音がしみ入る。
No.20		危篤が電話で騒ぐ。	危篤が電話で騒ぐ。
No.21		胸の底でわめきちらす。	胸の底でわめきちらす。
No.22		皮膚に感触が蘇る。	皮膚に感触が蘇る
No.23		帯が絹の叫びをあげる。	帯が絹の叫びをあげる。
No.24		争奪欲がそうさせる。	争奪欲がそうさせる。
No.25		行為が半生をひきまわす。	行為が半生をひきまわす。
No.26		考えが心に触れる。	考えが心に触れる。
No.27		眼に場面が甦る。	眼に場面が甦る
No.28		不平が声をあげる	不平が声をあげる
No.29		頭へものをぶちこむ	頭へものをぶちこむ
No.30			ものが作用をよぶ
No.31			顔に味がある

6 おわりに

自動的な比喩判定の研究において、概念同士の関係性つまり名詞と名詞における比喩性の判定が行われている中、本研究では、名詞と動詞の関係における比喩の自動判定という新たな課題に取り組んだ。そのために、用言と格要素の関係および既存の比喩文との類似性に着目した点が本手法の特徴である。

類似性の判定方法として、類似文検索の一つである、ベクトル空間方法を用いた。そしてその特性ベクトルに名詞と格要素をとり入れ結合比喩における語と語の異常性をとらえた。

比喩の判定実験では、2つの方法を行った。1つは、類似性判定の閾値をさだめた絶対的な方法、もう1つは、相対的な方法として、入力文が正常文のベクトル空間に近いか比喩文のベクトル空間に近いかによる方法である。その結果、前者の方法は、適合率100%、再現率18%となり、後者の方法は、適合率は96%、再現率は24%になった。2つを総合的に判定すると、適合率は96%、再現率は27%となった。以上により、名詞と動詞の関係における比喩において、一つの解を示すことができた。

謝辞

本研究において一年間指導して頂いた鳥取大学知能情報工学科計算機C研究室の池原教授, 村上助教授と徳久助手にお礼を申し上げます。ALT-JAWSはNTTとの共同研究の下で使用させて頂きました。そして, 卒論を進めるにあたり, 数々のアドバイスを頂いた計算機C研究室の皆さんに感謝します。

参考文献

- [1] 榎井, 福本, 椎野, 河合: 確率的判定尺度を用いた比喩性検出手法, 自然言語処理, Vol.9, No.5, pp.72-92, 2002.
- [2] 内海: 比喩理解モデルの諸相, 電子情報通信学会, Vol.9, No.5, pp.72-92, 2002.
- [3] 中村: 比喩表現の理論と分類, 共立出版, 1977.
- [4] 池原, 宮崎, 白井, 横尾, 中岩, 小倉, 大山, 林: 日本語語彙体系, 岩波書店, 1997.
- [5] 宮地: 慣用句の意味と用法, 明治書院, 1982.
- [6] 木本: 単語意味属性を用いたベクトル空間法, 鳥取大学修士論文, 2000.
- [7] 計算機用日本語名詞辞書 IPAL 解説編, 情報処理振興事業協会技術センター, 1996.
- [8] 山梨: 比喩と理解, 東京大学, 1988.
- [9] 芳賀: メタファーの心理学, 誠信書房, 1990.
- [10] 瀬戸: メタファー思考, 講談社現代新書, 1995.
- [11] 中村: 比喩表現辞典, 角川書店.
- [12] 辻: メタファーの基本用語, 言語, 大修館書店, pp.24-25, 2002.
- [13] 岩山: 比喩を含む言語理解における顕現性の役割, 人工知能学会誌, pp.674-681, 1991.