

Title	名詞の属性に着目した比喩理解に関する研究
Author(s)	三島, 達夫
Citation	
Issue Date	2007-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/3608
Rights	
Description	Supervisor: 白井 清昭, 情報科学研究科, 修士

修 士 論 文

名詞の属性に着目した比喻理解に関する研究

北陸先端科学技術大学院大学
情報科学研究科情報処理学専攻

三島 達夫

2007 年 3 月

修 士 論 文

名詞の属性に着目した比喻理解に関する研究

指導教官 白井 清昭

審査委員主査 白井 清昭 助教授

審査委員 島津 明 教授

審査委員 東条 敏 教授

北陸先端科学技術大学院大学
情報科学研究科情報処理学専攻

510096 三島 達夫

提出年月: 2007 年 2 月

概 要

計算機による理解が難しい言語現象の一つに比喻表現がある。我々の使用する言語はほとんどと言っていいほど比喻的な表現が含まれており、計算機による比喻理解は自然言語処理にとって避けて通れない問題である。だが従来の技術を単純に適用するだけではその解析は難しい。例えば「リンゴのような頬」といった表現を「赤い頬」と理解するためには「リンゴ」の典型的な性質として「赤い」があるという知識が必要になる。

本論文では基本的な比喻である「名詞Aのような名詞B」という直喩表現の理解のための計算モデルを提案する。名詞とその属性値の組の共起頻度をコーパスから獲得することで名詞が持つ性質についての知識ベースを作り、それに基づいて直喩表現で暗示されている意味的に適切な属性値を比喻解釈として選ぶモデルを構築する。名詞と属性値の共起頻度に基づいてスコアを計算し一番高いスコアとなった候補を解釈として出力する。

実験では実際にコーパスから抽出した直喩表現に対して、精度は62%であった。特に他の異なる手法と提案した場合、本研究の主な3つのアイデアである、属性名毎に属性値のスコアを計算する、顕現性の高い属性値を抽出しやすいパターンで得た共起頻度についてスコアを高くする、名詞Bと属性値の関連性を考慮に入れることが精度の向上に有効であることが、異なる手法と比べることで確認できた。

目次

第1章	序論	1
1.1	本研究の背景と目的	1
1.2	本論文の構成	2
第2章	関連研究	3
2.1	比喩理解モデルについての研究	3
2.2	名詞の属性値自動分類と確率の自動獲得	4
2.3	「名詞Aのような名詞B」の比喩性判定	4
2.4	ウェブ検索エンジンを用いた比喩理解	5
2.5	本研究の特色	8
第3章	比喩理解	9
3.1	本研究で扱う表現	9
3.2	直喩の理解	9
3.3	属性名、属性値の定義	10
第4章	名詞と属性値の共起頻度の獲得	12
4.1	コーパスデータについて	12
4.2	単純な共起パターン	12
4.3	顕現性の高い属性値の抽出	14
4.4	顕現性抽出パターンの自動獲得	17
第5章	直喩理解の計算モデル	18
5.1	直喩理解についての考え	18
5.1.1	喩詞の属性値	18
5.1.2	被喩詞の属性値	19
5.2	比喩理解の計算モデル	20
5.2.1	手法1	20
5.2.2	手法2（提案手法）	21
第6章	実験	26
6.1	テストデータ	26
6.2	実験	27

6.2.1	パラメータ調整	27
6.2.2	実験	28
6.3	考察	30
第7章 まとめ		33
付 録 A テストデータ		37
A.1	クローズドテストデータ	37
A.2	オープンテストデータ	38
付 録 B 本研究で定義した属性名・属性値		41

目 次

2.1	概念（りんご）（類）の表現形式の記述例	4
2.2	田添らによる比喻性判定モデル	6
2.3	確率的な概念記述における特徴集合の顕現性落差	7
3.1	属性名 a_i , 属性値集合 V_i , 属性値 v_{ij} で表された名詞 N の概念	10
3.2	分類語彙表の表記例	11
4.1	単純共起パターン Pa^{co}	12
4.2	名詞と顕現属性値を含む文の kwic 表示例	15
4.3	顕現性抽出パターン Pa^{sa}	15
5.1	比喻理解計算モデルの概要	22

表 目 次

2.1	「りんご」「頬」の検索ヒット数上位 10 形容詞	5
2.2	「名詞 A のような名詞 B」の分類	7
2.3	概念対「砂」「雪」のクローズアップ属性値の適合性判定	8
4.1	単純抽出パターン Pa^{co} で獲得された属性値の例	13
4.2	例文収集に用いた名詞とその顕現属性値	14
4.3	名詞 N とその顕現属性値 A の間に見つかった文字列パターン	15
4.4	顕現性の高いパターンで抽出された属性値の例	16
4.5	名詞 N と顕現属性値 A を含む部分文字列上位 10 位	17
5.1	名詞「花」と一部の属性値の共起頻度	19
5.2	顕現性落差のない属性値が正解の比喻の例	20
6.1	新聞記事と小説の「ような」の比較	26
6.2	テストデータの一部	27
6.3	パラメータ α, β を変化させた時の正解数	27
6.4	各手法の比較 (クローズドテスト)	29
6.5	各手法の比較 (オープンテスト)	29
6.6	提案手法と $\beta = 0$ の比較	30
6.7	1 位と 2 位が同じような属性値であった例	31
6.8	不正解となった直喩表現	32

第1章 序論

1.1 本研究の背景と目的

計算機による自然言語処理の技術として代表的なものは構文解析や意味解析、格解析などの技術があげられる。また、これらの技術を支える知識源として、辞書やタグ付きコーパスなど様々な言語資源が存在する。

しかしながら、こういった従来技術では対応しにくい言語現象として比喩表現がある。我々が使用する言語にはほとんどといっていいほど比喩的な表現が含まれている。日常会話から新聞記事のようなものまで様々な表現が見られ、言語表現のほとんどは比喩的であり比喩は人間の概念体系に深く関わっているという指摘もある。したがって計算機による自然言語処理にとって、比喩表現の理解は避けて通れない問題であるが、既存の技術のように文の内容を解析することや既存の言語資源を適用するだけでは解析が難しい。

例えば「氷のような心」といった直喩表現があった場合、それは「冷たい心」といった意味が連想され、それらは単純な意味解析などでは理解できない。ここで必要なのは「氷」のがもつ典型的な性質として「冷たい」があるという知識である。

比喩の研究は従来から認知心理学の分野で発展しており、比喩の生成や比喩表現として適切かどうかなど多くの研究が行われている。同様に比喩の理解についての研究も行われているが、基本的にそれらの研究は心理実験によるものが主流であるために、被験者の確保や同一の状況を揃えることなどを考えると大規模な比喩理解の実験は単純な直喩の表現に限ったとしても難しいといわれている。

一方、自然言語処理の分野でも、比喩理解は心理実験により名詞の属性値を獲得する方法が多かった。しかし、近年ウェブの発展やコーパスの整備により、大規模な知識の自動獲得が比較的容易になってきている。比喩理解のための知識をコーパスから自動獲得する土台ができつつある。

そういった背景を踏まえ、本論文では基本的な比喩である直喩表現の理解のための計算モデルを提案する。特に直喩理解のための名詞とその属性値の組の共起頻度をコーパスから獲得することで名詞についての知識ベースを作り、それに基づいた直喩表現で暗示される、意味的に適切な属性値を比喩の解釈として選ぶモデルを構築する。

1.2 本論文の構成

本論文の構成は以下のとおりである。まず第2章では本研究の関連研究について述べる。第3章では本研究で扱う直喩の表現と理解の定義、および名詞の性質を示す属性名と属性値の説明をする。第4章では第3章で述べた名詞の属性値をコーパスから自動的に獲得する手法を説明する。第4章で獲得した名詞の属性値の知識を用いた直喩理解のモデルについて第5章で説明し、それらについての実験を第6章で述べる。最後に第7章で結論と今後の課題などについて述べる。

第2章 関連研究

2.1 比喩理解モデルについての研究

比喩理解モデルの研究として岩山らの研究がある [1]。岩山らは概念の性質の顕現性を基にした「視点表現」による比喩表現の理解の枠組みを提案している。視点表現は「概念 A (target 概念) を概念 B (source 概念) という視点から眺める」ことを表現する道具である。比喩では、target 概念は例えられる概念 (被喩詞)、source 概念が例える概念 (喩詞) に相当する。視点表現を用いた比喩理解というのは target 概念へ移される source 概念の性質を決定することである。

岩山らの研究では移される source 概念の性質の決定のために顕現性とよばれる尺度を用いている。顕現性とは概念の性質の典型性を表す尺度であり、顕現性が高いほど target 概念に移されやすい性質だといえる。この顕現性を求めるには図 2.1 のような概念の属性名とそれに含まれる属性値集合、そしてこの属性値の現れやすさを表す確率、という概念についての知識が必要である。図 2.1 を例に挙げるとリンゴの性質として「色」や「外形」「表面」といった属性名がある。それぞれの属性名は属性値集合からなっており例えば「色」であれば「赤」や「緑」といった属性値とそれらに対応する現れやすさの確率「0.8」や「0.15」が付与されている。

元々認知心理の分野での研究が多かった比喩理解の研究であるが、既存の研究において多くのものが心理実験を行い図 2.1 のような概念の知識を獲得している [2]。それらの実験のための概念の属性値集合も人手で決定していることや、実験は被験者に対するテストによって知識を得る手法であるため、

- 被験者の状態によって属性値の確率が左右される
- 大規模な知識構築のためにはコストがかかりすぎる

といった問題点がある。

また、岩山らのモデルでは比喩の理解として source 概念から target 概念に属性値が移されるわけであるが、その際 source 概念の全ての属性値集合についてそれぞれ最も顕現性の高い属性値が target 概念に移されることになる。したがって特に本研究で述べるような自動的な知識構築を行った場合に、移されるべきでない属性値集合の属性値についても target 概念に移されてしまう可能性がある。

それ以外にも被喩詞である target 概念の性質については基本的に考慮されていないという点も問題があると考えられる。なぜなら喩詞から移される属性値は、結果として被喩

詞を修飾することになるためその属性値による修飾が自然かどうかは、比喩の理解のためには非常に重要であると考えられるからである。

$$\begin{aligned}
 *(\text{りんご}) &= \left\{ \begin{array}{l} \text{色:} \left\{ \begin{array}{l} \text{赤} \# 0.8 \\ \text{緑} \# 0.15 \\ \text{茶色} \# 0.05 \end{array} \right\} \\ \text{外形:} \left\{ \begin{array}{l} \text{球状} \# 0.95 \\ \text{円柱状} \# 0.05 \end{array} \right\} \\ \text{表面:} \left\{ \begin{array}{l} \text{滑らか} \# 0.9 \\ \text{ざらざら} \# 0.1 \end{array} \right\} \\ \vdots \end{array} \right\} \\
 *(\text{頬}) &= \left\{ \begin{array}{l} \text{色:} \left\{ \begin{array}{l} \text{白い} \# 0.5 \\ \text{赤い} \# 0.3 \\ \text{青白い} \# 0.2 \end{array} \right\} \\ \text{外形:} \left\{ \begin{array}{l} \text{丸い} \# 0.95 \\ \text{四角い} \# 0.05 \end{array} \right\} \\ \vdots \end{array} \right\}
 \end{aligned}$$

図 2.1: 概念 (りんご)、(頬) の表現形式の記述例

2.2 名詞の属性値自動分類と確率の自動獲得

三島らはコーパスとウェブ検索エンジンを用いて名詞と形容詞の共起頻度を獲得、その頻度に基づいたクラスタリングにより形容詞を属性値集合に自動的に分類し、また確率付けも同時に行うという研究を行っている。この研究では名詞の属性値として既存辞書から形容詞を抽出し、それらの形容詞と比喩表現に使われている名詞を用い「形容詞 + 名詞」のクエリでウェブ検索エンジンによるヒット件数を獲得している。そのヒット件数を用いて各形容詞をクラスタリング、属性値のグループとして分類している。

共起頻度の例としてあげられているものを表 2.1 にあげる。図 2.1 と比較して、特に「赤い」と「青い」の差など、単純に「形容詞 + 名詞」の共起によって得られた属性値の頻度情報は名詞が持つ典型的な属性に対して必ずしも高いわけでは無いことがわかる。

また属性値の自動分類はやはり難しく、ハードクラスタリングによる手法では複数の属性名に分類することの出来る形容詞もどちらかに分けることしか出来ていないなどという問題点もある。したがって、本研究では、属性値の設定はシソーラスを用いて人手で設定する。また、名詞と形容詞の共起頻度を測る際、ウェブ検索エンジンを用いて表層的な表現の出現回数を得るのではなく、新聞記事および小説を形態素解析と構文解析をして名詞が持つ属性の頻度情報をより正確に得ることを試みる。

2.3 「名詞 A のような名詞 B」の比喩性判定

「名詞 A のような名詞 B」という表現の全てが比喩であるわけではない。田添らによると「名詞 A のような名詞 B」の用法は大きく分けて 2 つに分けることが出来る [4]。1

表 2.1: 「りんご」「頬」の検索ヒット数上位 10 形容詞

りんご		頬	
形容詞	ヒット数	形容詞	ヒット数
美味しい	1230	白い	859
赤い	852	冷たい	783
青い	752	赤い	478
甘い	557	柔らかい	172
甘酸っぱい	292	青白い	135
良い	273	丸い	122
高い	186	熱い	111
大きい	161	美しい	103
多い	135	可愛い	73
小さい	134	薄い	64

つは「リンゴのような頬」を例とする直喩であり、もう 1 つは「アメリカのような国」を例とする喩喩ではない用法である。この場合の用法は例示を意味している。他にも指示（例えば「次のような点」）、人の判断（「当たり前のようなこと」）などがあり、それらをまとめて『リテラルな表現』としている。田添らは「名詞 A のような名詞 B」は図 2.2 のチャートにより 6 つのパターンに分類されると述べている。（表 2.3）

これらのパターンのうち喩喩性がある表現はパターン 1 とパターン 3 であり、他のパターンは基本的にリテラルな表現と言える。またパターン 3 の表現については「名詞 A のような名詞 B のような X の名詞 B」という意味である。このとき喩詞が「名詞 A の名詞 B」、被喩詞「X の名詞 B」となるが、「名詞 A のような名詞 B」を抽出するだけでは被喩詞である X を特定出来ないことが指摘されている。

2.4 ウェブ検索エンジンを用いた喩喩理解

梶井らはウェブ検索エンジンを用いた喩喩解釈の研究を行っている [5][6]。梶井らの研究ではまずコーパスから概念と属性値の共起データを獲得し、喩喩に用いられている 2 つの概念との間で共通に現れ、かつ顕現性の落差がある属性値をクローズアップ属性値とする。

例として「子供」と「顔」の 2 つの概念を比較した場合を図 2.3 に示す。この場合「子供」の概念において顕現性の高い「幼い」「たくましい」が、「顔」という概念において顕現性が低いため顕現性の落差が働き、これらの 2 つの概念間では「幼い」「たくましい」という属性値が喩喩性を高める属性値としてクローズアップされ「子供のような顔」という喩喩理解の際の解釈の候補となる。

図 2.2: 田添らによる比喩性判定モデル

表 2.2: 「名詞 A のような名詞 B」の分類

パターン 1	名詞 A と名詞 B が直接対比され、異種概念であるもの (後のパターン 2 ~ 6 に該当しないもの)	例: リンゴのような頬 氷のような風
パターン 2	名詞 A と名詞 B が直接対比され、名詞 B が名詞 A の 上位概念であるもの	例: アメリカのような国 湾岸戦争のような戦争
パターン 3	名詞 B が名詞 A の静的な属性 (部分・形状) であるもの	例: ピラミッドのような形
パターン 4	名詞 B が名詞 A の動的な属性 (状態、状況) であるもの	例: 田舎のような自然 ボスニアのような問題
パターン 5	名詞 B が特定の名詞 (抽象名詞であるもの)	例: バールのようなもの 刃物のようなもの
パターン 6	名詞 A が特定の名詞 (人称代名詞、時制名詞、文中の 場所を指す名詞、事物の評価を表す名詞) であるもの	例: 私のような選手 従来のような勢い

そのクローズアップ属性値とされた属性値は、必ずしも比喩の解釈としては適当ではない。したがって「“ source 概念 ”のように“ クローズアップ属性値 ”と「“ クローズアップ属性値 ” + “ target 概念 ”」の 2 つの表現のウェブ検索を行い、両方とも閾値を超えるヒット数を得た場合にそのクローズアップ属性値は比喩の解釈として適合性があるとしている。

例えば喩詞として「砂」、被喩詞として「雪」という概念対に対して「新しい」「白い」「細かい」といった属性値がクローズアップされる。これらの属性値について上記のパターンによるウェブ検索を行う (表 2.3)。そのヒット件数に対して閾値 α を適用し、(この場合 $\alpha = 15$) 両方ともに閾値以上の件数が獲得できている「白い」「細かい」は「砂のような雪」という比喩表現の解釈として妥当であると判断している。梶井らはこの手法によって人間の判断を約 80% の正確さでシミュレートできたと述べている。

$$*(\text{子供}) = \left\{ \begin{array}{cc} \boxed{\text{幼い}} & 0.222 \\ \text{小さい} & 0.162 \\ \boxed{\text{たくましい}} & 0.030 \\ \text{可愛い} & 0.030 \\ \text{健康な} & 0.030 \\ \vdots & \vdots \end{array} \right\} \quad \left\{ \begin{array}{cc} \text{渋い} & 0.070 \\ \text{いい} & 0.055 \\ \text{涼しい} & 0.052 \\ \vdots & \vdots \\ \boxed{\text{たくましい}} & 0.008 \\ \boxed{\text{幼い}} & 0.003 \end{array} \right\} = * \text{顔}$$

図 2.3: 確率的な概念記述における特徴集合の顕現性落差

適合性判定の間違いの原因として単純にウェブ検索のヒット数が閾値以下だというデータの過疎性が問題である場合や、ウェブ検索において意味の異なる表現の一部との一致

表 2.3: 概念対「砂」「雪」のクローズアップ属性値の適合性判定

クエリ	頻度	判定
砂のように細かい	106	○
細かい雪	405	○
クエリ	頻度	判定
砂のように新しい	1	×
新しい雪	496	○

クエリ	頻度	判定
砂のように白い	15	○
白い雪	30709	○

により、本来のヒット数よりも高いヒット数が得られてしまうといった場合などがあげられている。また、その他の問題点として属性値の属性名がについては、考慮していない（例えば、「温度」という属性名に対する「冷たい」という属性値と「人の態度」という属性名に対する「冷たい」が区別されていない）こと、比喩に使われる喩詞と被喩詞の両方に共通の属性値がない直喩表現にはこの手法は適用できないといったことがあげられる。

2.5 本研究の特色

本研究では直喩の理解を目的とした計算モデルの構築を目指す。基本的には岩山らの手法をベースに、属性値の確率をコーパスから学習することで様々な比喩を理解できるよう大規模な名詞の知識を獲得する。特に顕現属性値を獲得しやすいようなパターンを発見することで、直喩の理解に最適な人間の直感に近い属性値の知識を得る方法を考える。また被喩詞の属性値を考慮に入れることで、特に比喩表現で移される属性の属性名の判定精度を上げることも考える。

第3章 比喩理解

本章では本研究で扱う比喩とその理解についての概要を説明する。

3.1 本研究で扱う表現

本研究では「名詞Aのような名詞B」という直喩表現を扱う。ただし2.3節で述べたように全ての「名詞Aのような名詞B」が必ずしも比喩的な表現ではない。「名詞Aのような名詞B」には比喩的な表現以外に、例示、指示などといったリテラルな表現が存在する。本研究では田添らの比喩性判定モデルで、パターン1に分類される比喩的な表現のみを扱うこととし「名詞Aのような名詞B」の分類などは行わない。また、同じ比喩的な表現であってもパターン3に分類される表現は、実際の意味として「名詞Aの名詞BのようなXの名詞B」であり、単純に「名詞Aのような名詞B」を抽出するだけでは被喩詞のXを特定することが出来ない。したがって本研究では扱わないこととする。

3.2 直喩の理解

本研究では直喩の理解を

- 喩詞から被喩詞へ移される属性名および属性値の決定

と定義する。また、名詞の性質としての属性名・属性値は岩山らの研究と同様に以下のように定義する。

- 名詞の性質は属性名 a_i と属性値集合 V_i の対であり、属性値集合は属性値 v_{ij} の集合である（図3.1）。

具体的に属性名とは「色」や「形」などという属性値の分類であり、属性値とは「赤い」や「青い」、「丸い」といった具体的な名詞の性質のことである。本研究における直喩の理解では名詞の性質を図3.1のような形式で学習し、喩詞が持つ属性名と属性値の中から被喩詞を移されるものを選択する。例えば「りんごのような頬」という直喩表現の理解は、移される属性名、属性値が「色：赤い」とであると決定することである。

$$*(N) = \left\{ \begin{array}{l} a_1 : \left\{ \begin{array}{l} v_{11} \# p_{11} \\ v_{12} \# p_{12} \\ \vdots \end{array} \right\} \\ a_2 : \left\{ \begin{array}{l} v_{21} \# p_{21} \\ v_{22} \# p_{22} \\ \vdots \end{array} \right\} \\ \vdots \\ a_i : \left\{ \begin{array}{l} v_{i1} \# p_{i1} \\ \vdots \\ v_{ij} \# p_{ij} \\ \vdots \end{array} \right\} \\ \vdots \end{array} \right\}$$

図 3.1: 属性名 a_i , 属性値集合 V_i , 属性値 v_{ij} で表された名詞 N の概念

3.3 属性名、属性値の定義

名詞の性質を図 3.1 のような属性名と属性値で表現するとき、どのように属性名や属性値を定義するかが問題となる。これを自動的に定義する試みもあるが [3]、本研究では既存の言語資源を用いて以下のように人手で決めることとした。

- 属性値は形容詞と形容動詞とする。
- 属性名は属性値の意味クラスとする。

具体的には属性値は形態素解析において形容詞・形容動詞語幹となる語句とし、属性名には分類語彙表 [7] における意味クラスを用いた。分類語彙表には図 3.2 のように語句が分類されている。例えば「赤い」と「青い」はどちらも「色」という意味クラスを持つ。このとき「色」という属性名の属性値として「赤い」「青い」があるとする。また「青い」という語は「人柄やすることが未熟である」という意味で「才能」という分類としても表記されている。このような場合は「才能」という属性名の属性値としても「青い」を分類する。この点で属性値集合 V_i の属性値が互いに排反であるとしている岩山らの研究と属性値に対する考えが異なる。「平和」という語句のように「体」と「相」という分類の両方にあらわれる語句もあるが、本研究においては属性値としての分類が必要であるため、体言を示す「体」の分類ではなく、「相」の分類である「人生・禍福」を「平和」の属性名とする。ただし属性名・属性値としてふさわしくないものについてはあらかじめ除外している。属性名・属性値としてふさわしくないものとは、それらが単一の名詞の属性として考えにくいものである。例えば「ない」という形容詞は何かの存在がないことを示す形容詞であるため属性値としては不適である。また「同一な」「同等な」といった語が属す

099131,94186,A,相,自然,自然,色,3.5020,04,01,01,赤い,赤い,あかい,いかあ
099141,94196,A,相,自然,自然,色,3.5020,06,01,01,青い,青い,あおい,いおあ
098145,93265,A,相,活動,行為,才能,3.3421,16,01,02,青い,青い,あおい,いおあ
041660,39928,A,体,活動,交わり,平和・治乱,1.3550,02,01,01,平和,平和,へいわ,わいへ
097428,92590,A,相,活動,生活,人生・禍福,3.3310,04,01,02,平和,平和,へいわ,わいへ

図 3.2: 分類語彙表の表記例

る意味クラス「異同・類似」は2つの物事を比較した際に得られる性質であり、単一の名詞の属性値としては適切ではない。100個の属性名と、延べ3006個(うち異なり2313個)の属性値となった。定義した全ての属性名と属性値を付録Bに示す。

第4章 名詞と属性値の共起頻度の獲得

直喩理解のためには属性値の顕現性を計算するための、名詞と属性値の知識ベースの構築が必要である。

従来からある心理実験による方法では妥当性の高い知識は得られるものの、多くの直喩に適用できるような大規模な知識の構築は極めて困難である。従って本研究では図 2.1 のような名詞の知識を構築するために、コーパスから名詞と属性値 (形容詞および形容動詞) の共起頻度を獲得した。共起頻度が高いということは、それだけその名詞と属性値の結びつきが強く顕現属性値であると言えるであろうという考えに基づく。

4.1 コーパスデータについて

本研究では名詞とその属性値である形容詞・形容動詞の共起頻度を獲得するための、コーパスとして毎日新聞 13 年分の新聞記事と青空文庫よりダウンロードした新字新仮名の小説 3,466 編を用いた。また、名詞と属性値の共起データは 4.2 節、4.3 節で述べる抽出パターンを用いて獲得する。そのパターンマッチのためにはコーパスの文の文節係り受け解析を行う必要がある。本研究では文節の係り受け解析に南瓜 [8] を用いた。

4.2 単純な共起パターン

一般的に名詞の属性値はそれを修飾する表現、例えば「“ 赤い ”りんご」や「“ 美しい ”髪」というように現れる。また、「りんごは“ 赤い ”」や「髪が“ きれい ”だ」のように「主語・述語」関係としても現れる。このような名詞と直接的に修飾 - 被修飾関係にある単純なパターンとして図 4.1 のようなパターンを考え、コーパスから共起頻度を得た。パターンの例において「→」は係り受け関係を示し、属性値 A は形容詞または形容動詞である。コーパス中から延べ数で $Pa^{co}(A \rightarrow N)$ は 894279 組、 $Pa^{co}(\text{主} \rightarrow \text{述})$ は 1385265 組

$$\begin{aligned} Pa^{co}(A \rightarrow N) &: \text{属性値 } A \rightarrow \text{名詞 } N \\ Pa^{co}(\text{主} \rightarrow \text{述}) &: \text{名詞 } N + \text{が・は (主語)} \rightarrow \text{属性値 } A \text{ (述語)} \end{aligned}$$

図 4.1: 単純共起パターン Pa^{co}

の名詞と属性値の共起を抽出できた。

いくつかの名詞に対する各パターンによる抽出結果の一部が表 4.1 である。

表 4.1: 単純抽出パターン Pa^{co} で獲得された属性値の例

名詞N	属性値A	$Pa^{co}(A \rightarrow N)$	$Pa^{co}(\text{主} \rightarrow \text{述})$
氷	熱い	0	1
	厚い	27	9
	冷たい	0	0
	白い	10	1
名詞N	属性値A	$Pa^{co}(A \rightarrow N)$	$Pa^{co}(\text{主} \rightarrow \text{述})$
花	赤い	146	6
	白い	530	7
	美しい	240	36
	優しい	11	0
名詞N	属性値A	$Pa^{co}(A \rightarrow N)$	$Pa^{co}(\text{主} \rightarrow \text{述})$
火	熱い	2	2
	冷たい	0	2
	赤い	41	2
	青い	39	0
名詞N	属性値A	$Pa^{co}(A \rightarrow N)$	$Pa^{co}(\text{主} \rightarrow \text{述})$
雪	冷たい	12	3
	温かい	0	0
	白い	89	10
	黒い	7	0

表 4.1 を見ると、必ずしも直感的に顕現性の高い属性値の頻度が高いわけではないことが分かる。例えば「雪」などは「冷たい」や「白い」といった直感的にも顕現性の高い属性値の頻度が高くなっていることがわかるが、「火」の「色」については「青い」の頻度が高い。しかし確かに「青い」火も存在するが、一般的に火の色を考えた場合は「赤い」を連想することが多いと思われる。また「氷」では「冷たい」という属性値が全く得られていないことが分かる。したがってこのような単純なパターンによる抽出のみに頼った属性値頻度の獲得は直喩の理解という点においては適切ではないと考えられる。

また「花」の属性値についてはある程度妥当な共起頻度が得られているが、属性名を考慮に入れない場合「美しい」よりも「白い」の共起頻度が大きいので「美しい」以上に「白い」ことの方が「花」の顕現属性値となってしまうことが考えられる。そこで属性値を一様に扱うのではなく、属性名毎に顕現性の高いものを考える必要があるということが 4.1 のデータから分かる。

4.3 顕現性の高い属性値の抽出

単純な共起頻度では名詞の顕現属性値を獲得することが難しいことがわかった。従って単純な共起頻度ではなく、顕現属性値を抽出しやすいパターンを発見することを試みた。名詞と顕現属性値を含む例文を収集し、それらに顕現属性値であることを示すような典型的パターンの発見を試みた。

表 4.2 は名詞とその顕現属性値を含む例文を収集するために用いた名詞とその顕現属性値の組である。名詞は青空文庫から抽出した直喩表現の中に喩詞として見つけることの出来た名詞である。また、必ずしも一つの名詞に対して一つの顕現属性値を組み合わせるのではなく、その名詞が直喩として使われた際に連想されうる全ての顕現属性値をあげている。

表 4.2: 例文収集に用いた名詞とその顕現属性値

名詞	属性値	名詞	属性値
雪	白い、冷たい	リンゴ	赤い、丸い
雲	白い	嵐	激しい
氷	冷たい	玉	美しい、丸い
火	熱い、赤い、暖かい、明るい	血	赤い
太陽	明るい、熱い、暖かい、眩しい	岩	固い、硬い、堅い
炎	熱い、赤い、暖かい、明るい	糸	細い
粉	細かい	花	優しい、美しい、可愛い、明るい
宝石	美しい	神様	偉い
鬼	怖い、恐ろしい	絵	美しい
重油	黒い	雷	激しい

表 4.2 にある名詞とその顕現属性値の組を含む例文を 4.1 節で述べた青空文庫のコーパスから 4144 文獲得した。名詞の前後にあるパターンを発見しやすくするために、獲得した例文を名詞を中心とした kwic 形式で表示した例が図 4.2 である。

図 4.2 を見てみると「火のように熱い」というパターンが多く見られる。

収集した例文から顕現属性が現れると思われるパターンを人手で探し出した。結果、図 4.3 のパターンが顕現性の高い属性値を抽出するパターンとした。

パターン $Pa^{sa}(Aさ)$ における「A(さ)」は、「形容詞語幹+さ」という形容詞の派生名詞を表す。また $Pa^{sa}(よونا)$ は、「Nのよونا」という文節と「A」という文節が同じ名詞 N' を含む文節に係ることを表す。 $Pa^{sa}(よウに)$ 、 $Pa^{sa}(よونا)$ については直喩的な表現であり顕現性が高いと考えるのは妥当である。それに対して $Pa^{sa}(ANよウに)$ 、 $Pa^{sa}(Aさ)$ のパターンは今回の調査で新しく発見できたパターンであるといえる。実際

図 4.2: 名詞と顕現属性値を含む文の kwic 表示例

表 4.3: 名詞Nとその顕現属性値Aの間に見つかった文字列パターン

名詞と顕現属性値のパターン	頻度
NのAさ	220
NのようにA	104
A Nのように	77
NのようなA	43

$Pa^{sa}(\text{ように})$: Nのように $\rightarrow$ A 例) 氷のように冷たい
 $Pa^{sa}(\text{ような})$: (Nのような)(A) $\rightarrow$ N' 例) 氷のような冷たい心
 $Pa^{sa}(\text{A Nように})$: A $\rightarrow$ Nのように 例) 冷たい氷のように
 $Pa^{sa}(\text{Aさ})$: Nの $\rightarrow$ A (さ) 例) 氷の冷たさ

図 4.3: 顕現性抽出パターン Pa^{sa}

にはもっと多くのパターンがあるのではないかと考えていたが、有効な顕現性抽出パターンは以上の4つしか発見できなかった。

上記のパターンを利用して獲得した属性値の頻度の一部を表 4.4 に示す。これを見ると

表 4.4: 顕現性の高いパターンで抽出された属性値の例

名詞	属性値	パターン 1. 2.	パターン 3.	パターン 4.
氷	熱い	0	0	0
	厚い	0	0	29
	冷たい	43	2	2
	白い	0	0	0
名詞	属性値	パターン 1. 2.	パターン 3.	パターン 4.
花	赤い	0	0	0
	白い	0	4	5
	美しい	16	1	59
	優しい	0	1	3
名詞	属性値	パターン 1.2.	パターン 3.	パターン 4.
火	熱い	15	0	1
	冷たい	0	0	0
	赤い	4	0	0
	青い	0	1	0
名詞	属性値	パターン 1. 2.	パターン 3.	パターン 4.
雪	冷たい	1	0	12
	温かい	0	0	0
	白い	31	3	18
	黒い	0	0	0

直喩の喩詞として使われた際に被喩詞に移されると解釈できるような属性値が表 4.1 と比べてより多く獲得できている。また顕現属性値でない属性値との差も出ているように思える。ただし 4.2 節のような単純なパターンに比べ獲得できた頻度は少なく、これだけを用いるのでは多くの名詞についての直喩解釈は行えない。従って顕現性の高いパターンを主とし、データの過疎性を補う形で単純なパターンで獲得された頻度を利用する方法を考える必要がある。単純なパターンで得られる頻度は必ずしも顕現性が高いことを示すわけではないが、十分なデータ量を持っているからである。

4.4 顕現性抽出パターンの自動獲得

前節で述べたような顕現性の高い属性値の抽出パターンを自動的に獲得できれば知識構築に非常に役に立つと考えた。そこで前節で獲得した例文に対して、名詞と顕現属性値をそれぞれN, Aというシンボルに置き換え、文章中のN, Aを含む部分文字列の頻度をprefixspan[10]によりカウントした。それらのうち頻度の高いものが顕現性抽出パターンになるのではないかという考えである。表 4.5 に prefixspan でカウントした部分文字列の上位 10 位をあげる。

パターン	頻度
A N	1573
N A	1379
A N の	1031
A N 、	940
A N 。	933
の N A	825
。 N A	750
、 N A	745
A N を	702
A N が	687

表 4.5: 名詞Nと顕現属性値Aを含む部分文字列上位 10 位

表 4.5 から分かるパターンとして実用的なものは得られていない。今回は単純に文字レベルで区切ってしまうため特に意味のある単位で区切れていないといった点が問題があったと考えられる。したがって、自立語か機能語かを考慮に入れることや、単語区切りや文節区切りなどを考え頻度をカウントすることで、もう少し意味としてまとまりのある単位で部分文字列をカウントした場合、パターンとして意味のあるものが獲得できるかもしれない。

次章では両方のパターンによって獲得した共起頻度を用いた直喩理解の計算モデルについて述べる。

第5章 直喩理解の計算モデル

本章では名詞と属性値の共起頻度に基づいた直喩理解の計算モデルについて述べる。

5.1 直喩理解についての考え

本研究において「名詞Aのような名詞B」を解釈するとき、名詞Aの顕現性の高い属性値を1つ選ぶが、その属性値の顕現性は以下の考えに基づいて求める。

1. 喩詞である名詞Aの中で共起頻度の高い属性値
 - (a) 顕現性の高いパターンで抽出されたものは顕現性を高く計算する
 - (b) 同じ属性名の他の属性値と比較して頻度の差が大きいものは顕現性を高くする
2. 被喩詞である名詞Bを修飾することが多い属性名の属性値は顕現性を高くする

以下、それらの理由を根拠となるデータをあげて説明する。

5.1.1 喩詞の属性値

直喩表現「名詞Aのような名詞B」の解釈において、基本的には喩詞であるAの属性値の中で顕現性の高いものが解釈として相応しいと考えられる。ここで、Aの顕現性の高さを計算するにあたって、単純に各属性値の共起頻度のみでは考えることが出来ないということが言える。例として名詞「花」の属性値の各パターンの共起頻度の一部を表5.1に示す。単純な頻度のみを考慮した場合「色：白い」が顕現属性値として考えることができる。しかしながら「花」は一般的に「白い」ばかりのものではない。実際に「赤い」や「黄色い」といった属性値も「白い」より少ないものの共起頻度は比較的多い。一方、「美醜：美しい」は「花」を喩詞とする直喩の解釈としては妥当であると考えることができ、顕現属性値であるといってよい。しかし共起頻度だけをみると「色：白い」より少ない。従って各属性名ごとに共起頻度を考え、その中でどれだけ他の属性値に比べて名詞に対して妥当な属性値であるかを計算する必要がある。ここで、異なる属性名に属する属性値の頻度を比較するために、属性名に対する属性値集合で正規化した相対頻度を用いる必要がある。

また Pa^{sa} と Pa^{co} で獲得した頻度は、その性質の違いから単純に頻度を合計してしまうと他の属性値との比較の際にパターン毎に分けた意味がなくなってしまう。しかしながら

表 5.1: 名詞「花」と一部の属性値の共起頻度

属性名	属性値	Pa^{co} による頻度	Pa^{sa} による頻度
美醜	美しい	276	75
美醜	優しい	11	3
色	白い	537	5
色	赤い	153	0
色	黄色い	129	0
色	黒い	11	0

知識として信頼性が高いと考えられる Pa^{sa} のみを用いただけではデータの過疎性が生じやすく、逆に Pa^{co} のみでは共起頻度の総数は多いが信頼性が低い低い。したがって両者の頻度はともに使うべきである。

5.1.2 被喩詞の属性値

直喩は「喩詞により被喩詞をたとえる」という表現であるが、それを解釈すると「喩詞の属性値によって被喩詞を修飾する」ということになる。例えば「春のような冬」という表現であれば「暖かい冬」というように「暖かい」という「春の属性値」によって「冬」が修飾されることになる。「暖かい」は「熱」「寒暖」などいくつかの属性名に属する属性値であるが、解釈として妥当なのはどちらかといえば「寒暖」の意味としてである。同じような意味でも「熱」に属する「温かい冬」という表現は不適切で、それは被喩詞である「冬」という語句がどのような属性名に修飾されやすいかが関わってくる。実際に「冬」と共起する頻度を見ると、「寒暖」という属性名に属する属性値と共起する回数が474、「熱」という属性名に属する属性値と共起する回数が82と明らかに差が出ている。従って移される属性値がどの属性名であるかどうかを判断するために、被喩詞と共起する属性値の属性名の考慮に入れることは重要であると考えられる。

また被喩詞の属性値の扱いについて、喩詞との間の顕現性の落差が移される属性値を決める際に必要になるという点が既存研究で言われている [1][5]。顕現性の落差とは喩詞では顕現性が高いが、被喩詞では顕現性が低いということである。例えば「りんごのような頬」という表現では「りんご」の属性値である「色：赤い」が、「頬」の「色：赤い」の顕現性と比べて差があるため、それが比喩的な意味を強める方向に働くという考え方である。

しかしながら実際の比喩表現では表 5.2 のような直喩が存在する。例えば「絵のような景色」という表現において顕現性の落差がある属性値のみを正解の候補とするのであれば、「美しい景色」という解釈は行えない。「炎のような夕陽」についても同様である。実際に獲得できた直喩の表現の中でも喩詞と被喩詞の両方の顕現性が高いことでより強い意

味合いを持たせるような表現が多く確認できた。従って、必ずしも喩詞と被喩詞の間で顕現性の落差がある属性値が解釈となるわけではないと考えられるため、本研究では顕現性の落差については計算を行っていない。

表 5.2: 顕現性落差のない属性値が正解の比喩の例

絵のような景色			炎のような夕陽		
属性名：属性値	絵	景色	属性名：属性値	炎	夕陽
美醜：美しい	0.85	0.88	色：赤い	0.65	0.80
美醜：きれい	0.05	0.09			

5.2 比喩理解の計算モデル

5.2.1 手法 1

本節では各パターンによって得られた共起頻度の重み付き和に基づいて、各属性値の顕現性を計算する手法について述べる。

まず属性名 a_i について式 (5.1) で定義される値 $O_A(a_i, v_{ij})$ を最大にする v_{ij} を選び、そのような属性名の組 (a_i, v_{ij}) をそれぞれ解釈の候補とする。

$$O_A(a_i, v_{ij}) = \sum_{\forall pattern} w_{pattern} \cdot O_A(a_i, v_{ij}, pattern) \quad (5.1)$$

$pattern$ は各抽出パターンを示し、 $O_A(a_i, v_{ij}, pattern)$ は、各パターンでの名詞 A と v_{ij} の共起頻度、 $w_{pattern}$ は各パターンで得た共起頻度に対する重みである。

そのように得られた (a_i, v_{ij}) の全ての組に対して式 (5.2) で定義される顕現性 $S(a_i, v_{ij})$ 計算し、一番高い値となる (a_i, v_{ij}) が直喩の解釈として出力される。

$$S(a_i, v_{ij}) = \alpha \left(\frac{O_A(a_i, v_{ij})}{\sum_{\forall j} O_A(a_i, v_{ij})} - \frac{O_A(a_i, v_{ir_2})}{\sum_{\forall j} O_A(a_i, v_{ij})} \right) + \beta \frac{\sum_{\forall j} O_A(a_i, v_{ij})}{\sum_{\forall i} \sum_{\forall j} O_A(a_i, v_{ij})} + \gamma \cdot S_B(a_i) \quad (5.2)$$

ただし $\alpha + \beta + \gamma = 1$ を満たすような値をとる

ここで v_{ir_2} は a_i の属性値集合 V_i 内で 2 番目に高い $O_A(a_i, v_{ij})$ の値をとる属性値を示す。式の第 1 項は 1 位と 2 位の相対頻度の差を示しており、2 位との差が大きい顕現属性値であればスコアが高くなる。また第 2 項は属性名 a_i が名詞 A と共起する属性値全体の中でどれくらい名詞 A と共起しやすい属性名であるかを示す指標である。これが高ければ属性名 a_i は名詞 A とよく共起すると言え、 a_i に属する v_{ij} の顕現性が高くなる。第 3 項である $S_B(a_i)$ は式 (5.3) で定義され、属性名 a_i がどれくらい名詞 B に対して共起しやすいかの指

標である。

$$S_B(a_i) = \frac{\sum_{\forall j} O_B(a_i, v_{ij})}{\sum_{\forall i} \sum_{\forall j} O_B(a_i, v_{ij})} \quad (5.3)$$

$O_B(a_i, v_{ij})$ は式 (5.1) と同様に各抽出パターンから得られる名詞 B と属性値 v_{ij} の共起頻度の重み付き和である。この値が大きければ名詞 B に対して属性名 a_i は共起しやすいといえるため、候補である (a_i, v_{ij}) の顕現性は高くなる。

予備実験

上記の手法をによる比喩理解の予備実験を行った。テストデータとなる比喩表現は 6.1 節で用いるものと同じものを使用した。47 の正解データのうち、属性値が正解した例は 24 であった。

実験の結果、顕現性抽出パターンによって得られた顕現属性値が単純共起パターンに吸収される形で失われたことで、不正解になってしまったものがいくつか見つかった。例えば「春のような冬」という表現にたいして、「春のように暖かい」「春の暖かさ」といった表現が多くあり、顕現性抽出パターンで得られた「春」と「暖かい」の共起頻度は 28 であった。また「暖かい春」「春は暖かい」のように単純パターンで得られた頻度は 38 であった。一方、「遅い」という属性値は顕現属性抽出パターンでは一つも発見されず「遅い春」「春が遅い」という表現だけで 84 の頻度が得られた。しかし、この結果本手法では出力として「遅い」が選択されている。つまり、顕現属性抽出パターンでは「暖かい」が「春」の顕現属性値として学習できているが、式 (5.2) のように 2 つの抽出パターンから得られる頻度の和によって顕現性を計っているため、より頻度の多い単純共起パターンに吸収される形で「暖かい」の顕現性がスコアに反映されていない。これは重み付きの和を求めることでパターンの違いを無視している点が問題であることの現れであると考えた。

従って本研究では獲得パターンによって顕現性の計算を別々に行うことにし、最終的に両方での計算結果を考慮に入れるモデルを構築した。

5.2.2 手法 2 (提案手法)

本研究の提案手法の概要を図 5.1 に示す。まず、名詞 A の各属性名 a_i について確率が 1 位である属性値 v_{ij} を直喩解釈の候補として選ぶ。その後、各 (a_i, v_{ij}) の組について名詞 A についての顕現性を示す値となる S_A と属性名 a_i がどれだけ名詞 B について関連性があるかを示す値である S_B をそれぞれ計算し、最終的なその属性名・属性値の組に対する顕現性のスコア $S(a_i, v_{ij})$ を計算し、それについてもっとも高いスコアとなった組 (a_i, v_{ij}) を直喩の正解として出力するということが基本的な考え方である。

図 5.1: 比喩理解計算モデルの概要

ここで、顕現性抽出パターンならびに Pa^{sa} と単純共起パターン Pa^{co} から得られる共起頻度を以下のように定義する。

$$\begin{aligned}
O_A^{sa}(v_{ij}) & \quad Pa^{sa} \text{ での名詞 A と属性値 } v_{ij} \text{ の共起頻度} \\
p_A^{sa}(v_{ij}) &= \frac{O_A^{sa}(v_{ij})}{\sum_{v_{ij} \in V_i} O_A^{sa}(v_{ij})} \quad O_A^{sa}(v_{ij}) \text{ の属性値集合 } V_i \text{ 内での相対頻度} \\
O_A^{co}(v_{ij}) & \quad Pa^{co} \text{ での名詞 A と属性値 } v_{ij} \text{ の共起頻度} \\
p_A^{co}(v_{ij}) &= \frac{O_A^{co}(v_{ij})}{\sum_{v_{ij} \in V_i} O_A^{co}(v_{ij})} \quad O_A^{co}(v_{ij}) \text{ の属性値集合 } V_i \text{ 内での相対頻度} \\
O_B^{sa}(v_{ij}) & \quad Pa^{sa} \text{ での名詞 B と属性値 } v_{ij} \text{ の共起頻度} \\
O_B^{co}(v_{ij}) & \quad Pa^{co} \text{ での名詞 B と属性値 } v_{ij} \text{ の共起頻度}
\end{aligned}$$

出力候補 (a_i, v_{ij}) の選択

各属性名 a_i に対して以下の条件を満たすような v_{ij} を選び、属性名と属性値の組 (a_i, v_{ij}) を直喩解釈の候補とする。

- 式 (5.4) の条件を満たす場合、 V_i の中から最大の $p_A^{sa}(v_{ij})$ を持つ v_{ij} を選ぶ。

$$\sum_{v_{ij} \in V_i} O_A^{sa}(v_{ij}) \geq 5 \quad (5.4)$$

- 式 (5.5) の条件を満たす場合、 V_i の中から最大の $p_A^{co}(v_{ij})$ を持つ v_{ij} を選ぶ。

$$\sum_{v_{ij} \in V_i} O_A^{co}(v_{ij}) \geq 15 \quad (5.5)$$

式 (5.4)、式 (5.5) の条件はパターン毎の閾値であり、ある一定以上の頻度がコーパスから得られていない場合は推定した $P_A^{sa}(v_{ij})$ や $P_A^{co}(v_{ij})$ の信頼性が低いと判断し、属性名と属性値の組を解釈の候補として加えないという考えに基づいて設定した。ある属性名に対し、式 (5.4)、式 (5.5) の両方ともに閾値を上回らない場合は、解釈の候補として適切な属性値がないと判断しシステムはその属性名に対しては出力の候補を生成しない。

このようにして得られた全ての候補 (a_i, v_{ij}) に対し式 (5.6) を計算し、それが最大になる組が名詞 A から名詞 B に移される属性名と属性値の組とする。

$$S(a_i, v_{ij}) = (1 - \beta) \cdot S_A(a_i, v_{ij}) + \beta \cdot S_B(a_i) \quad (5.6)$$

ここで $S_A(a_i, v_{ij})$ は属性名と属性値の組 (a_i, v_{ij}) の喩詞である名詞 A に対する顕現度を測る指標である。一方、 $S_B(a_i)$ は被喩詞である名詞 B と属性名 a_i の関連度を測る指標である。また β は両者に対する重みである。ここで、最適な重み β をいかに決めるかが問題になるが、本研究ではこれを実験的に定める。

名詞 A の顕現度 $S_A(a_i, v_{ij})$

顕現度 $S_A(a_i, v_{ij})$ は式 (5.7) で定義する。

$$S_A(a_i, v_{ij}) = \alpha \cdot S_A^{sa}(a_i, v_{ij}) + (1 - \alpha) S_A^{co}(a_i, v_{ij}) \quad (5.7)$$

$S_A^{sa}(a_i, v_{ij})$ および $S_A^{co}(a_i, v_{ij})$ はそれぞれ $p_A^{sa}(v_{ij})$ 、 $p_A^{co}(v_{ij})$ を基に計算される v_{ij} の顕現度である。また α はコーパスから共起頻度を抽出するパターンである Pa^{sa} と Pa^{co} に対してどちらを重視するかを示す重みである。最適な α をいかに決めるかが問題となるが、これも式 (5.6) の β と同様に実験的に定める。

$S_A^{sa}(a_i, v_{ij})$, $S_A^{co}(a_i, v_{ij})$ の定義を式 (5.8) , 式 (5.9) に示す。

$$S_A^{sa}(a_i, v_{ij}) = \begin{cases} (p_A^{sa}(v_{ir_1}) - p_A^{sa}(v_{ir_2})) \cdot \frac{\sum_{v_{ij} \in V_i} O_A^{sa}(v_{ij})}{\sum_{\forall v_{mn}} O_A^{sa}(v_{mn})} & (j = r_1) \\ 0 & (j \neq r_1) \end{cases} \quad (5.8)$$

$$S_A^{co}(a_i, v_{ij}) = \begin{cases} (p_A^{co}(v_{ir_1}) - p_A^{co}(v_{ir_2})) \cdot \frac{\sum_{v_{ij} \in V_i} O_A^{co}(v_{ij})}{\sum_{\forall v_{mn}} O_A^{co}(v_{mn})} & (j = r_1) \\ 0 & (j \neq r_1) \end{cases} \quad (5.9)$$

r_1 と r_2 は確率 $p_A^X(v_{ij})$ の大きい上位 1 位と 2 位の属性値の番号を示す。つまり式 (5.8) , 式 (5.9) 第 1 項は 1 位の属性値の確率と 2 位の属性値の確率の差が大きければ大きいほど、 v_{ij} は名詞 A の典型的な属性値であると考え高いスコアを与えるという意味である。一方、積の第 2 項は名詞 A の属性名 a_i の属性値集合 V_i に属する属性値が、名詞 A 全体の属性値の頻度に対してどの程度の割合であるかを表している。すなわち V_i に属する属性値が名詞 A とよく共起するのであれば属性名 a_i に対する候補 (a_i, v_{ij}) に高いスコアを与えることになる。また式 (5.8) , 式 (5.9) の条件部の意味であるが、本手法において候補 (a_i, v_{ij}) は抽出パターン Pa^{sa} および Pa^{co} についてそれぞれ異なる条件で別々に決定される。したがってある属性名 a_i に対して確率 $p^{sa}(v_{ij})$ に基づく候補と $p^{co}(v_{ij})$ に基づく候補で異なる場合がありうる。このように 1 位となる属性値がパターンによって異なる場合においては顕現性の計算の際に異なるパターンに基づく顕現度を 0 にする、というのが条件部の意味である。つまり、ある出力の候補 (a_i, v_{ij}) が Pa^{sa} 、 Pa^{co} の両方ともに 1 位である場合には式 (5.8) , 式 (5.9) の両方の計算結果を用いて式 (5.7) の計算が行われ、どちらか一方のみが 1 位で候補になっている場合はもう一方のパターンでの顕現度を 0 にして式 (5.7) の計算が行われる。

このように抽出パターンによって独立に顕現度を計算するのは 5.1.1 項で述べたように、2 つのパターンから得られる共起頻度は性質の異なるデータであるため単純にまとめて計算することで各々の性質が失われてしまう恐れがあるためである。単純なパターンは顕現性の面ではやや信頼性にかけるものの、多くの頻度を獲得できる。一方で顕現性抽出パ

ターンは顕現性の高い属性値の頻度が大きくなるが、獲得できる総頻度は少なく、データの過疎性が問題となる。従って両者の重み付き和を計算することで最終的な名詞Aに対する (a_i, v_{ij}) の顕現性 $S_A(a_i, v_{ij})$ を計算している。

$S_B(a_i)$ の計算

5.1.2 節で述べたように、被喩詞である名詞Bにどのような属性名の属性値が修飾しやすいかということが名詞Aから属性値が移される際にも重要になってくると考えられる。本研究では名詞Bに対する属性名 a_i の修飾のしやすさを式 (5.10) として定義した。

$$S_B(a_i) = \frac{\sum_{v_{ik} \in V_i} O_B^{sa}(v_{ik}) + O_B^{co}(v_{ik})}{\sum_{\forall v_{mn}} O_B^{sa}(v_{mn}) + O_B^{co}(v_{mn})} \quad (5.10)$$

$S_B(a_i)$ は被喩詞である名詞Bと共起する属性値の総和に対する、属性値集合 V_i に属する属性値の頻度の割合である。いなわちBがある属性名 a_i の属性値とよく共起していれば、その属性名に対する顕現属性値に対するスコアが高くなる。この式では属性値の共起パターンを区別していない。なぜなら、名詞Bとの共起のしやすさを計るため、その属性値が名詞Bにとって顕現性が高いか低いかというのは関係がないからである。

第6章 実験

6.1 テストデータ

本章では実際に文章中から抽出した直喩表現に対して、本研究での提案手法を適用する。まず毎日新聞 13 年分の新聞記事と青空文庫からダウンロードした小説 (新字新仮名)^{3,466} 編に対して茶筌 [9] で形態素解析を行い、「名詞 A のような名詞 B」というパターンにマッチした表現を抽出した。

新聞記事から抽出した例と青空文庫から得た例を表 6.1 に示す。これらの例から分かる

表 6.1: 新聞記事と小説の「ような」の比較

毎日新聞	青空文庫
事務のような仕事	花のような姫君
明治座のような大劇場	氷のような風
駅伝のようなスポーツ	棒のような慎重
宇宙服のような服	絹糸のような髪の毛
ベトナムのような国々	皿のような目

ように新聞記事では主に「例示」を示すリテラルな表現が多く逆に青空文庫から獲得した表現では比喩的な表現が多い。本研究では比喩的な表現の理解を目的としているため、リテラルについては考慮に入れない。また、3 節で述べたように、表 2.3 においてパターン 3 に分類されるものについては本研究では扱わない。従って青空文庫から獲得した「名詞 A のような名詞 B」について、2.3 節で述べた田添らの手法でパターン 1 となる比喩的な表現のうち 47 個の表現をクローズドテスト、50 個の表現をオープンテストとして用いた。本研究におけるクローズドテストとは、式 (5.6)、式 (5.7) のパラメータ α 、 β の最適な値を決めるためのデータと同じデータをを使って比喩理解の評価を行うこと、オープンテストとは α 、 β を決めるためのデータと異なるテストデータを使って比喩理解の評価を行うことを指す。すなわち、47 個の比喩表現でパラメータ α 、 β の調整を行う。またオープンテストのデータには喩詞として顕現属性値抽出パターンを調べた際に用いた名詞を含まないようにしている。

テストデータの一部を図 6.2 に示す。また、全てのテストデータを付録 A に示す。正解は人手によって付与しているが、それらを選ぶ際には実際に抽出できた比喩表現が含まれ

る文の前後の文脈も考慮に入れている。また、妥当であると思われる解答が複数存在する場合は正解を複数与えている。

表 6.2: テストデータの一部

名詞Aのような名詞B	正解 (a_i, v_{ij})
絵のような景色	美醜：美しい
春のような冬	寒暖：暖かい
氷のような声	待遇・礼など：冷たい
氷のような汗	熱：冷たい
花のような姫君	美醜：美しい、美醜：優しい

6.2 実験

6.2.1 パラメータ調整

5.2.2 項の式 (5.6)、式 (5.7) における α 、 β について最適な値を求める予備実験をクロードテストのテストデータに対して行った。 α は顕現性抽出パターンに対する重み、 β は名詞Bの属性名の頻度が全体へ与える影響の重みである。表 6.3 は α 、 β をそれぞれ 0 から 1 の範囲で 0.1 きざみで変動させたときの結果である。また、ここで評価している正解数は属性値についてのみ評価し、属性名が正しいかどうかは考慮していない。表 6.3 の結

表 6.3: パラメータ α, β を変化させた時の正解数

$\alpha \backslash \beta$	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.0	11	10	14	17	17	17	16	16	15	13
0.1	15	16	19	19	16	17	16	15	14	14
0.2	18	18	21	21	21	19	19	17	16	16
0.3	22	20	23	23	22	20	20	17	16	16
0.4	22	21	23	25	22	22	19	17	17	16
0.5	23	25	27	26	24	23	19	18	17	16
0.6	28	29	30	29	25	23	19	18	17	16
0.7	28	29	29	29	28	24	22	18	17	17
0.8	28	30	29	29	27	25	22	19	18	17
0.9	28	30	29	27	28	27	24	19	18	17
1.0	20	26	25	24	25	24	22	17	15	14

果から $\alpha = 0.6$ 、 $\beta = 0.2$ とした。

6.2.2 実験

実験において比較した手法は以下の通りである。

- ベースライン（以下 B L）
各抽出パターンによる共起頻度を区別せず、また属性名による属性値の分類も考慮にいれずに、単純に共起頻度の高い属性値を入力された直喩に対する解釈として出力する。
- 提案手法
5.2.2 項で述べたモデルで直喩に対する解釈を決定する。なお 6.2.1 項で述べたように $\alpha = 0.6$ 、 $\beta = 0.2$ とした。
- 顕現性の高いパターンを用いない。（ $\alpha = 0$ とする）
単純共起パターン Pa^{co} で獲得した共起頻度のみを用いた場合。これと提案手法を比較することで顕現属性値抽出パターン Pa^{sa} の有効性を確認する。
- 名詞 B の属性名の頻度を考慮しない。（ $\beta = 0$ とする）
名詞 B にどのような属性名の形容詞が修飾しやすいかを考慮に入れない場合。これと提案手法を比較することで、名詞 B と属性名との共起頻度を考慮に入れることが直喩理解に有効であることを確認する。
- 単純共起パターン Pa^{co} と顕現属性値抽出パターン Pa^{sa} を区別しない手法
単純共起パターン Pa^{co} と顕現属性値抽出パターン Pa^{sa} を区別しないで両方の共起頻度を合計しスコアを計算する。具体的には名詞 A の顕現度 $S_A(a_i, v_{ij})$ を以下のように定義する。

$$O_A(v_{ij}) = O_A^{sa}(v_{ij}) + O_A^{co}(v_{ij}) \quad (6.1)$$

$$p_A(v_{ij}) = \frac{O_A(v_{ij})}{\sum_{v_{ij} \in V_i} O_A(v_{ij})} \quad (6.2)$$

$$S_A(a_i, v_{ij}) = (p_A(v_{ir_1}) - p_A(v_{ir_2})) \cdot \frac{\sum_{v_{ij} \in V_i} O_A(v_{ij})}{\sum_{\forall v_{mn}} O_A(v_{mn})} \quad (6.3)$$

クローズドテストおよびオープンテストの結果を表 6.4、表 6.5 に示す。表中の“ 正解 ” は属性名・属性値ともに正解であったものの個数である。それぞれ精度と再現率も示す。また“ 属性のみ ” は属性値のみ評価した場合の正解数であり属性名については考慮しない場合である。“ 不正解 ” は属性名・属性値ともに正解とは異なった出力がされた場合である。“ 出力なし ” は全ての属性名についてが式 (5.4) および式 (5.5) における条件を満たさず、解釈の候補 (a_i, v_{ij}) が存在しなかった場合である。これは名詞 A と属性値の共起データがコーパスから十分に得られなかったことを意味する。なお B L では全ての直喩表現に対して解釈が出力されるため、出力なしということは起こらない。また、名詞 A から名詞 B に移される属性値は決定するが、その属性名は決定しない。

表 6.4: 各手法の比較 (クローズドテスト)

	B L	提案手法	$\alpha = 0$	$\beta = 0$	区別なし
正解	-	26	13	21	15
正解 (精度)	-	62%	31%	50%	36%
正解 (再現率)	-	55%	28%	45%	32%
属性値のみ	15	30	14	28	20
属性値のみ (精度)	32%	71%	33%	67%	43%
属性値のみ (再現率)	32%	64%	30%	60%	43%
不正解	32	12	28	14	27
出力なし	-	5	5	5	-

表 6.5: 各手法の比較 (オープンテスト)

	B L	提案手法	$\alpha = 0$	$\beta = 0$	区別なし
正解	-	20	14	15	17
正解 (精度)	-	46%	33%	35%	33%
正解 (再現率)	-	40%	28%	30%	34%
属性値のみ	22	26	20	23	24
属性値のみ (精度)	44%	60%	46%	53%	48%
属性値のみ (再現率)	44%	52%	40%	46%	48%
不正解	28	15	23	20	26
出力なし	-	7	7	7	-

6.3 考察

まずB Lと提案手法を比較する。B Lでは属性名を考慮に入れていないため属性値のみの正解の場合において比較すると、クローズドテストでは精度は32%から71%、再現率は32%から64%、オープンテストでは精度は44%から60%、再現率は44%から52%と両方ともにクローズドテスト、オープンテストともに提案手法の正解率が勝っている。このことから属性名を考慮に入れること、また属性値の抽出パターンについて Pa^{co} と Pa^{sa} を異なる頻度として考える提案手法の効果が現れているといえる。

また顕現属性値抽出パターン Pa^{sa} を考慮に入れない $\alpha = 0$ の場合を見ると属性値の精度だけでもクローズドテストで33%、オープンテストで46%となってしまう、B Lと同じ程度の精度まで下がってしまっている。ここでB Lで用いた顕現性の頻度には Pa^{sa} の頻度が Pa^{co} と同じ扱いであっても含まれていることを考えると、顕現属性抽出パターンで得られた共起頻度を単純に他のパターンと足した場合にはそれを考慮に入れなかった場合と大差ないということも言える。

$\beta = 0$ の場合の“正解”の精度が下がっている点を考えると、被喩詞である名詞Bの属性名の頻度を考慮に入れることが比喩理解において効果的であるという点が実証されたとと言える。具体例を表6.6に示す。

「氷のような声」という表現について、 $\beta = 0$ の場合は正解として「熱：冷たい」が選ば

表 6.6: 提案手法と $\beta = 0$ の比較

直喩	提案手法の出力	$\beta = 0$	名詞Bの総頻度
氷のような声	待遇・礼など：冷たい	熱：冷たい	9857
炎のような息	生理・病気など：熱い	好悪・愛憎：熱い	560
炎のような夕陽	色：赤い	好悪・愛憎：熱い	16
氷のような白刃	待遇・礼など：冷たい	熱：冷たい	3

れる。属性値はあっているが属性名が間違っている。ここで $\beta = 0.2$ (提案手法) の場合、「声」に関連性の高い属性名「待遇・例など」が選ばれているため属性名・属性値ともに正解が選ばれている。また、「炎のような息」と「炎のような夕陽」で両方ともに正解が選ばれている点も名詞Bの頻度を考慮したことが正解につながっているという意味で非常に注目すべき点であると考えられる。ただし「氷のような白刃」のように逆に $\beta = 0$ で正解である場合もある。これらの場合はほとんど名詞Bの頻度が少ない場合だと考えられる。従って名詞Bに関する頻度の獲得方法を改善することでこういった間違いを減らせる可能性がある。

その他に、表6.7に挙げるような間違いの例があげられる。

「鬼のような女」という表現は属性名「恐れ・怒り・悔しさ」の中で正解といえる属性値である「恐ろしい」や「怖い」が1位と2位になっているが、式(5.8)、式(5.9)で1位と2位の差をとってスコアを計算しているために顕現性が低く見積もられてしまう。その

表 6.7: 1 位と 2 位が同じような属性値であった例

直喩	出力 ($S(a_i, v_{ij})$)	正解の a_i	a_i の v_{ir1}	a_i の v_{ir2}
鬼のような女	程度：恐ろしい	恐れ・怒り・悔しさ	恐ろしい	怖い
重油のような湖面	音：高い	色	黒い	どす黒い

ため 2 位以下の属性値の頻度が 0 に近い属性名「程度」の顕現性が高くなっているということが原因である。このような属性値に付いてはある程度意味が同じであれば頻度をマージするという手法をとるべきであるかもしれない。表 6.8 の「重油のような湖面」についても「黒い」「どす黒い」という属性値の差を取ってしまっているという点で同じ原因での間違いである。

パターンを区別しなかった場合と提案手法を比較した場合について提案手法がクローステスト、オープンテストともに結果がよかったことから、単純共起パターンと顕現属性抽出パターンによる共起頻度を分けてスコアを計算することの有用性が見られる。

続いて完全に不正解となった表現を表 6.8 に示す。

基本的にはほとんどの表現において顕現属性値が的確に学習されていないことが原因である。例えば「粉のような雪」という表現については「白い粉」の共起頻度 215 であり、この場合の顕現属性値となるべき「細かい」に対する共起頻度 5.0 を大幅に上回っている。

また文脈を考慮していないという点も考えられる。「眠りのような死」という表現については「深い」という属性値それ自体は不適切ではないが、文脈上「安らか」「心地よい」といった意味が適当であるため不正解となっているというものである。

また「NのAさ」というパターンが程度を表している場合などもあると考えられ、「髪の毛の長さ」などという使い方）顕現性の計算に悪影響を与えている場合がある。例としては「玉の堅さ」という表現が多いため「堅い」を出力してしまっている「玉のような顔」という表現などである。ただし「玉」については顕現属性値である「丸い」の頻度がそもそも他より低いため、上記の問題を解決しても正解は得られなかったと思われる。

そのほかには属性名の分類の点もあげることができる。本研究では分類語彙表の意味クラスを基準に「色」と「長短・高低・深浅・厚薄・遠近」という属性名の分類を対等に扱っているが、実際に後者の属性名はより細かく分類することが出来るとも考えられる。例えば「氷のような髪」という表現でシステムの出力は「髪」は属性名「長短・高低・深浅・厚薄・遠近」に関する属性値の共起頻度が 410 と非常に高いため、この属性名が出力として選ばれているがそのうちのほとんどが「長短」に関する属性値の頻度である。これが原因で「厚い」という属性値に対するスコアがたかくなるのは問題があると考えられる。

属性値の定義による問題として形容詞・形容動詞だけではカバーできていない解釈が存在した点も挙げられる。「琥珀のような酒」という表現は「琥珀色の酒」を意味するが、「琥珀色」というのは形容詞・形容動詞ではないため本研究で属性値として定義されていない。こういった問題も属性名の定義と同様に見直す必要があると考えられる。

表 6.8: 不正解となった直喩表現

直喩	正解 (属性名 : 属性値)	出力 (属性名 : 属性値)
粉のような雪	弛緩・粗密・繁簡 : 細かい	色 : 白い
雷のような気合	行為・活動 : 激しい	多少 : 多い
眠りのような死	快・喜び : 心地よい	長短・高低・深浅・厚薄・遠近 : 深い
重油のような湖面	色 : 黒い	音 : 高い
電光のような眼	力 : 強い	速度 : 速い
神様のような犬	威厳・行儀・品行 : 偉い	色 : 白い
涙のような時雨	苦悩・悲哀 : 悲しい	好悪・愛憎 : 熱い
氷のような髪	熱 : 冷たい	長短・高低・深浅・厚薄・遠近 : 厚い
嵐のような息	行為・活動 : 激しい	色 : 白い
玉のような汗	形 : 丸い	安心・焦燥・満足 : 堅い
玉のような顔	美醜 : 美しい	安心・焦燥・満足 : 堅い
焰のような愛	行為・活動 : 激しい	長短・高低・深浅・厚薄・遠近 : 高い

第7章 まとめ

本論文では直喩表現「名詞Aのような名詞B」という表現の理解を目標とし、名詞についての知識をコーパスベースで獲得する手法とそれに基づいて直喩の理解を行う計算モデルを提案した。特に比喻理解の際に顕現性が高いと思われるパターンで得られた共起頻度について高い重みを与えることで属性値の正解率を、名詞Bの属性名毎の頻度を考慮に入れることで属性名の正解率をそれぞれあげることが出来た。名詞Bの属性名毎の頻度は属性名の正解率だけでなく、属性値の正解率の向上にも貢献している。テストデータに対してスコアの重みを最適化しているために厳密なオープンテストによる評価ではないが、提案手法の有効性はある程度認められた。

最後に今後の課題について述べる。まず提案手法がオープンなテストセットに対してどの程度有効なのか確認しなければならない。これについては青空文庫から獲得できた比喻表現を用いてテストすることが出来る。

また解決すべき提案手法の問題点の一つに、出現頻度の低い名詞を用いた表現については正解となる属性値を出力できない点が挙げられる。これについては知識源となるコーパスの量だけでなく、その種類に関しても様々なものを多く使うことが必要なのではないかと考える。そういった意味でウェブ上の言語資源は非常に有用なリソースであると考えられるが、妥当性の高い知識の獲得のために全てのウェブページについて形態素解析や係り受け解析を行うということは現実的ではない。ウェブ上に存在する膨大な言語資源を利用するためには係り受けに頼らない属性値の抽出パターンや手法を考える必要がある。

また本研究で顕現性の高い属性値を抽出するためのパターンを考慮に入れているが、やはり顕現性の高い属性値は当たり前であるがために表記されないという問題点もあるため、必ずしもそれらを獲得できているわけではない。本研究において顕現属性値抽出パターンの有効性はある程度認められたため、さらに多くのパターンを発見し利用することで知識の妥当性を上げることが出来るのではないかと考えられる。

言葉自体は日々増えていくため、計算機による比喻理解のためには自動的に名詞の知識を獲得することは不可欠であるため、今後もウェブ上の言語資源や既存コーパス等を用いてよりよい名詞と属性値についての知識を得られるような手法を考えることが必要不可欠である。

本手法では「名詞Aのような名詞B」という表現の文脈を考慮に入れてないことも問題である。これについては比喻表現の前後の語句から関連する属性値についてはスコアをあげるなどといった方法が考えられる。しかし実際の文章を読んで比喻を理解することは人間でも意見がわかれるような問題であるため、非常に難しいことであると考えられる。

また寺井らが述べているように喩詞とは直接共起しないような属性値も解釈としては妥当であるような場合も存在する [11]。例えば「氷のような心」という表現に対して「氷」の顕現属性値である「冷たい」から「酷い」「無関心な」といった属性値が創発特徴として顕現し、解釈となりうるということである。このような典型的な顕現属性値から他の属性値が連想されるということをどうやって知識として獲得するかという点も非常に重要な課題である。

謝辞

最後に、本研究を進めるにあたり、終始丁寧に御指導くださいました主テーマ指導教員の白井清昭助教授に心から感謝します。また主指導教員の島津明教授にはお忙しい中頻繁に声をかけて頂き、研究の助言などを頂きました。心より感謝します。

そして自然言語処理講座の中村誠助手、研究室の同期、後輩のメンバーの方々には多くの場所で相談にのって頂き、有益な助言を多く与えてくださりました。この場を借りて深く感謝の意を示します。

参考文献

- [1] 岩山真, 徳永健伸, 田中穂積. 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, Vol6, No5, pp.674-681, 1991.
- [2] Smith, E. E., Osherson, D. N., Rips, L. and Keane, M. Combining prototypes: a selective modification model. Cognitive Science, Vol.12, No.4, pp.485-527, 1988
- [3] 三島達夫, 徳永健伸, 田中穂積. 比喩理解のための名詞の持つ属性値分類に関する研究. 東京工業大学, 学士論文, 2005.
- [4] 田添丈博, 椎野努, 榊井文人, 河合敦夫. “ 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, Vol10, No.2, pp.43-57, 2003.
- [5] 榊井文人, 福本淳一, 椎野努, 河合敦夫. 確率的尺度を用いた比喩性判定手法. 自然言語処理, Vol9, No.5, pp.71-92, 2000.
- [6] 榊井文人, 福本淳一, 荒木健治. 比喩解釈を目的とする World Wide Web を利用した属性値の適合性判定 NLP2005, C2-2.
- [7] 国立国語研究所 (編) . 分類語彙表. 大日本図書, 2004.
- [8] CaboCha/南瓜 <http://chasen.org/~taku/software/cabocha/>
- [9] 形態素解析システム茶筌 <http://chasen.naist.jp/hiki/ChaSen/>
- [10] prefixspan <http://www.cs.sfu.ca/~jpei/publications/span.pdf>
- [11] 寺井あすか, 中川正宣, 徳永健伸. 比喩理解過程における創発特徴の心理実験による検証 日本認知科学会第 23 回大会発表論文集, pp.388-389, 2006.

付 録 A テストデータ

A.1 クローズドテストデータ

直喩表現	正解 (a_i, v_{ij})
火のような情熱	行為・活動：激しい、感動・興奮：熱い、信念・努力・忍耐：熱い
粉のような雪	弛緩・粗密・繁簡：細かい
野猪のような男	力：荒い、行為・活動：荒い、行為・活動：荒っぽい
雷のような気合	行為・活動：激しい
鐘のような声	音：低い
絹糸のような髪の毛	美醜：美しい、広狭・大小：細い
眠りのような死	快・喜び：心地よい
地獄のような呵責	恐れ・怒り・悔しさ：怖い、恐れ・怒り・悔しさ：恐ろしい
氷のような白刃	熱：冷たい
絵のような景色	美醜：美しい
重油のような湖面	色：黒い
火のような女	行為・活動：激しい、感動・興奮：熱い
春のような冬	寒暖：暖かい、寒暖：暖か
白刃のような絶望	力：強い
氷のような汗	熱：冷たい
嵐のような変化	行為・活動：激しい
電光のような眼	恐れ・怒り・悔しさ：恐ろしい、形：鋭い
神様のような犬	威厳・行儀・品行：偉い
熱病のような精神	行為・活動：激しい
宝石のような花	美醜：美しい
涙のような時雨	苦悩・悲哀：悲しい
雷のような一喝	行為・活動：激しい
蝶のような花	美醜：美しい、好悪・愛憎：かわいい
花のような姫君	美醜：美しい、人柄：優しい
氷のような髪	熱：冷たい
炎のような息	生理・病気など：熱い、生理・病気など：熱っぽい
火のような言葉	感動・興奮：熱い、力：強い

絹糸のような水藻	広狭・大小：細い、長短・高低・深浅・厚薄・遠近：長い
黒鉛のような夕暮	色：黒い
夕陽のような火	色：真っ赤、色：赤い
氷のような風	寒暖：冷たい
漆のような髪の毛	色：真っ黒、色：黒い
雪のような落花	色：真っ白、色：白い
嵐のような息	行為・活動：激しい
雪のような指先	色：白い
玉のような汗	形：丸い
氷のような声	待遇・礼など：冷たい
鬼のような女	恐れ・怒り・悔しさ：恐ろしい、恐れ・怒り・悔しさ：怖い
化け物のような女	恐れ・怒り・悔しさ：恐ろしい、恐れ・怒り・悔しさ：怖い
玉のような顔	美醜：美しい、形：丸い
宝石のような心もち	美醜：美しい
火花のような恋	行為・活動：激しい、好悪・愛憎：熱い
焰のような愛	行為・活動：激しい、好悪・愛憎：熱い
花のような笑み	好悪・愛憎：優しい、好悪・愛憎：かわいい
珠のような容貌	美醜：美しい
炎のような夕陽	色：真っ赤、色：赤い
鉄のような意志	材質：堅い、材質：固い

A.2 オープンテストデータ

直喩表現	正解 (a_i, v_{ij})
林檎のような頬	色：赤い、形：丸い
鉄のような胴体	材質：固い、材質：硬い、材質：堅い、軽重：重い
山のような高波	広狭・大小：大きい、長短・高低・深浅・厚薄・遠近：高い
小児のような怪物	広狭・大小：小さい
棒のような指	長短・高低・深浅・厚薄・遠近：長い、広狭・大小：細い
影のような目的	長短・高低・深浅・厚薄・遠近：薄い、浅い、特徴：つまらない
皿のような眼	形：丸い、形：円い、形：まるい
髪の毛のような蔓	広狭・大小：細い、長短・高低・深浅・厚薄・遠近：長い
蠟のような肉体	色：白い、材質：滑らか
針金のような顎	広狭・大小：細い、形：鋭い
水のような風	寒暖：涼しい、気象：さわやか
白熊のような犬	色：白い、広狭・大小：大きい

塔のような頭巾	長短・高低・深浅・厚薄・遠近：高い
暗闇のような男	光：暗い 色：黒い
バラのような頬	色：赤い
石のような女	材質：固い、材質：堅い、人柄：頑固
虹のような希望	広狭・大小：大きい、快・喜び：明るい
清水のような空気	気象：さわやか、気象：すがすがしい、気象：すずやか
焰のような宝石	色：真っ赤、色：赤い
月光のような霜	美醜：美しい、色：青白い
炭のような目	色：黒い
石のような沈黙	意味・問題・趣旨など：重い
矢のような催促	速度：速い
嘘のような景色	美醜：美しい
黒山のような人出	多少：多い
蚊のような声	音：細い、広狭・大小：小さい
水晶のようなしずく	美醜：美しい
綿のような雲	軽重：軽い、色：白い
悪夢のような室内	恐れ・怒り・悔しさ：恐ろしい、恐れ・怒り・悔しさ：怖い
熱湯のような血	熱：熱い
錐のような山	形：鋭い
鎧のような雨戸	軽重：重い
雨のような汗	熱：冷たい
満月のような乳房	形：丸い、美醜：美しい
鏡のような池	音：静か
鉛のような飯	意味・問題・趣旨など：重い
泥のような眠り	長短・高低・深浅・厚薄・遠近：深い
山のような氷	多少：多い
棒のような両手	広狭・大小：細い、長短・高低・深浅・厚薄・遠近：長い
大黒柱のような神経	自身・誇り・恥・反省：図太い、広狭・大小：太い
夢のような絵	美醜：美しい
野菊のような人	美醜：美しい
剃刀のような男	才能：鋭い
地獄のような悲鳴	恐れ・怒り・悔しさ：恐ろしい
箱のような建物	形：四角い
蜜のような歓楽	感動・興奮：甘い
熊のような山男	広狭・大小：大きい
琥珀のような酒	琥珀色の（形容詞・形容動詞に無い）

真珠のような月	美醜：美しい、形：丸い
夢のような恐怖	程度：恐ろしい

付 録 B 本研究で定義した属性名・属性値

におい	かぐわしい、かび臭い、きな臭い、フルーティー、汗臭い、血生臭い、香ばしい、酒臭い、臭い、熟柿臭い、焦げ臭い、生臭い、青臭い、泥臭い、乳臭い、芳しい、芳醇、抹香臭い、薬臭い
安心・焦燥・満足	おぼつかない、じれったい、せわしない、たまらない、のどか、のんき、もどかしい、ハングリー、安楽、安気、安心、楽、堪能、気ぜわしい、気安い、気楽、気掛かり、気詰まり、気強い、気軽、気遣わしい、気丈夫、窮屈、堅い、堅苦しい、歯がゆい、伸びやか、心もとない、心安い、心強い、心細い、心丈夫、身軽、辛気臭い、沈静、沈着、不安、不服、不平、不満、不満足、物足りない、平気、平静、豊か、末頼もしい、満足、悠長、憂い、頼もしい、力強い、冷静
威厳・行儀・品行	あくどい、いたずら、えげつない、がさつ、こすい、さもしい、ずるい、ぞんざい、ふしだら、みだら、やんちゃ、ストイック、プラトニック、悪質、悪性、悪辣、偉い、意地汚い、淫蕩、汚い、下品、義、凶暴、狂暴、好色、罪、罪作り、罪深い、自堕落、純潔、女好き、上品、性悪、浅ましい、疎略、粗暴、粗野、粗略、他人行儀、恥さらし、道徳的、罰当たり、卑しい、卑劣、非道、非礼、不義理、不敬、不行跡、不作法、不貞、不道德、不届き、不法、不倫、腹黒い、暴虐、無垢、無法、野蛮、乱暴、力強い、老獪、腕白、狡猾、猥褻

意志	いい加減、うかつ、おろそか、ちゃらんぽらん、でたらめ、むやみ、むら、やみくも、わがまま、わざとらしい、移り気、横暴、気まぐれ、気まま、気随、苦し紛れ、自在、自分勝手、自由、自由自在、手前勝手、縦横無尽、初心、勝手、身勝手、随意、専横、適当、任意、不行き届き、不心得、不注意、不本意、不用意、浮気、放縦、放漫、奔放、本意、無計画、無節操、無造作、野放図、優柔不断
意味・問題・趣旨など	でたらめ、めちゃくちゃ、もっての外、テクニカル、リーズナブル、一義的、奥深い、肝心、肝要、緊要、支離滅裂、主要、重い、重要、詳細、深遠、深長、大事、大切、的、同義、批判的、非合理、非常識、不合理、不条理、不毛、無意義、無意味、無軌道、問題、有意義、論外
一般・全体・部分	一般的、全面的、部分的、きれい、そっくり、てんでんばらばら、ばらばら、個別、孤独
因果	あらたか、偶然、至当、自業自得、駄目、当たり前、当然、必至、必定、無効、無駄、有効
音	うるさい、かさかさ、かしましい、かちかち、かまびすしい、からから、がさがさ、がたがた、がらがら、けたたましい、こちこち、しめやか、だぶだぶ、にぎやか、ひそやか、やかましい、アコースティック、マイナー、メジャー、甘い、閑寂、閑静、喧騒、甲高い、高い、高らか、細い、渋い、声高、静か、静寂、静粛、騒がしい、騒々しい、太い、低い、忍びやか、物騒がしい、変
快・喜び	うれしい、おかしい、おもしろい、さわやか、すがすがしい、つまらない、つまんない、コミカル、快い、快適、楽しい、寒い、喜ばしい、興味深い、軽快、豪快、渋い、小気味よい、上機嫌、心地よい、晴れやか、壮快、爽快、痛快、不快、不機嫌、不愉快、明るい、愉快、朗らか
寒暖	あつい、うそ寒い、ぬくい、ぬくとい、ぬるい、ほやほや、クール、温暖、温和、寒い、寒冷、暑い、暑苦しい、蒸し暑い、生ぬるい、生暖かい、暖か、暖かい、熱い、薄ら寒い、肌寒い、涼しい、冷たい、冷やっこい、冷ややか

感覚	かゆい、くすぐったい、こそばゆい、まばゆい、まぶしい、むずがゆい、ショック、煙い、煙たい、思い出深い、切実、痛い、痛切
感動・興奮	エキサイティング、ショッキング、ショック、ドラマチック、感慨深い、感心、感無量、甘い、甘美、劇的、熱い、涙ぐましい
関係	ダイレクト、縁遠い、遠い、緊密、近い、親密、疎い、疎遠、直接的、没交渉、密、密接、無縁、無関係
気象	うっとうしい、からから、さわやか、すがすがしい、のどか、陰気、陰鬱、穏やか、厳しい、晴れやか、陽気、涼やか
飢渴・酔い・疲労・睡眠など	かったるい、くたくた、くちい、ぐたぐた、けだるい、だるい、ひだるい、ひもじい、へとへと、ぺこぺこ、寝苦しい、退屈、眠い、眠たい、無聊
恐れ・怒り・悔しさ	おぞましい、おっかない、おどろおどろしい、すごい、ひどい、ものすごい、遺憾、臆病、悔しい、気味が悪い、気味悪い、恐ろしい、恐縮、業腹、苦苦しい、空恐ろしい、口惜しい、恨めしい、惨め、散散、残り惜しい、残り多い、残念、心外、心残り、凄惨、凄絶、惜しい、投げやり、破れかぶれ、薄気味悪い、悲惨、不気味、怖い、腹立たしい、物恐ろしい、未恐ろしい、未練、未練がましい、無惨、無残、無念、名残惜しい、悽愴
苦悩・悲哀	ありがた迷惑、いたたまれない、うっとうしい、うら寂しい、うら悲しい、うるさい、おっくう、しんどい、たまらない、つらい、はた迷惑、やり切れない、やるせない、わびしい、哀れ、哀切、暗鬱、陰惨、陰鬱、気重、狂おしい、狂わしい、胸苦しい、苦い、苦しい、苦痛、寂しい、重苦しい、情けない、人恋しい、切ない、息苦しい、大儀、嘆かわしい、沈痛、沈鬱、悩ましい、煩い、煩わしい、悲しい、悲壮、悲痛、物狂おしい、物寂しい、物悲しい、物憂い、迷惑、面倒、面倒臭い、厄介、憂鬱

形	いかつい、いびつ、くたくた、ごつい、しりすぼみ、つぶら、ぺしゃんこ、ぺちゃんこ、まっすぐ、まるい、まるやか、よれよれ、シャープ、ジグザグ、鋭い、鋭利、円い、横長、角い、緩い、丸い、細長い、四角、四角い、縦長、真ん丸、中高、鈍い、平たい、平ら、平らか、平坦、平板、無傷、扁平
敬意・感謝・信頼など	ありがたい、かたじけない、もったいない、悪い、恐れ多い、恐縮、孝行、済まない、心苦しい、親孝行、親不孝、不孝、不忠
経済・収支	くだらない、けち、せんない、たくさん、たつとい、ちやち、つましい、つまらない、とうとい、ぼろい、もったいない、リーズナブル、安い、安価、安上がり、安直、格安、割安、割高、簡素、貴い、貴重、結構、高い、高価、質素、惜しい、尊い、駄目、低廉、得難い、得用、入用、必須、必要、不可欠、不経済、不必要、不用、不要、不利、無益、無駄、無用、有益、有用、有利、廉価
軽重	スポーティー、荷重、過重、軽い、軽やか、軽快、重い、重げ、重たい、身軽、莊重、鈍重
見聞き	ミクロ、遠い、見づらい、見にくい、見苦しい、耳ざとい、耳遠い、耳障り、聞きづらい、聞きにくい、聞き苦しい、目ざとい、目障り
言語活動	うるさい、がたがた、くだくだしい、けんもほろろ、ざっくばらん、しちくどい、しどろもどろ、ちゃらんぽらん、ちんぷんかんぷん、でたらめ、やかましい、オーバー、寡黙、回りくどい、滑らか、嫌み、誇大、御大層、口うるさい、口さがない、口やかましい、口汚い、口下手、口軽、口上手、口達者、口不調法、口幅ったい、高飛車、耳障り、手短、上手、辛辣、声高、舌たるい、舌足らず、多弁、大げさ、単刀直入、直截、痛烈、能弁、無口、雄弁、理屈っぽい、率直、流暢、婉曲、訥弁、饒舌
限定・優劣	みすばらしい、圧倒的、限り、限定、秀抜、上手、真つ青、絶倫、卓抜、段違い、低劣、拔群、並びない、無類、優秀、有数

限度	ぎりぎり、すれすれ、たくさん、パーフェクト、果てしない、完全、完璧、極まりない、極端、結構、十全、十分、精一杯、徹底的、不完全、不十分、不全、不徹底、満足、無限、無制限、有限
固定・傾き・転倒など	なだらか、安静、安定、緩やか、急、不動
交わり	くみしやすい、なれなれしい、むつまじい、温かい、家庭的、気まずい、近しい、懇意、親しい、水臭い、熱い、不和、平和、和やか、慇懃
光	あからさま、あらわ、かすか、さやか、つややか、てかてか、てらてら、ほの暗い、まばゆい、まぶしい、ダーク、ビジュアル、ビビッド、暗い、一目瞭然、輝かしい、顕著、小暗い、真っ暗、澄明、鮮やか、鮮明、鮮烈、淡い、著しい、定か、透明、鈍い、薄い、薄暗い、不透明、不明瞭、明るい、明白、明瞭
公式・公平	オフィシャル、シビア、パブリック、フェア、フリー、プライベート、リベラル、潔白、厳格、厳正、公式、公正、公的、公平、公明、公明正大、合法、私的、自在、自由、正規、非公式、非合法、不公平、不正、不当、不平等、平等、法的、無私、僭越、僭上
好悪・愛憎	あほらしい、いけ好かない、いじらしい、いたいけ、いたわしい、いとおしい、いとしい、いとわしい、うらやましい、かわいい、かわいそう、かわいらしい、しおらしい、ぞっこん、たまらない、ちゃんちゃらおかしい、ねたましい、のろわしい、ばかばかしい、ばからしい、むつまじい、ゆかしい、チャーミング、ヒューマン、哀れ、愛くるしい、愛らしい、因業、奥ゆかしい、温かい、可憐、懐かしい、忌まわしい、気の毒、気安い、近しい、嫌、嫌い、嫌み、厚い、好き、好ましい、好ましい、恨めしい、懇意、懇切、子煩悩、手厚い、女好き、小憎らしい、小面憎い、小癪、笑止、笑止千万、情け深い、心安い、心憎い、親しい、親愛、親身、親切、親密、人嫌い、疎ましい、憎い、憎たらしい、憎らしい、他愛、大嫌い、大好き、痛ましい、痛痛しい、毒毒しい、熱い、薄情、非情、不親切、不人情、不愍、不憫、物好き、片腹痛い、慕わしい、望ましい、無惨、無残、無慈悲、無情、面憎い、優しい、冷血、恋しい

広狭・大小	かすか、か細い、きつい、きゃしゃ、ささやか、せせこましい、だだっ広い、だぶだぶ、ちっぽけ、でかい、でっかい、ぶかぶか、ジャンボ、スリム、ビッグ、ワイド、果てしない、華奢、過小、過大、窮屈、巨大、狭い、狭苦しい、狭小、狭隘、極小、極大、極微、軽微、広い、広大、広範、広闊、豪壮、些細、細い、手狭、手広い、小さい、小規模、小作り、太い、大きい、大掛かり、大振り、長大、微細、微弱、微小、幅広、幅広い、膨大、雄大、零細、瑣末、矮小
行為・活動	たけだけしい、むちゃ、りりしい、ヒロイック、ラフ、ワイルド、因業、押し付けがましい、果敢、果断、過激、活発、頑強、気丈、強引、強硬、矯激、激しい、元気、姑息、荒い、荒っぽい、豪胆、差し出がましい、手荒、手荒い、女らしい、女女しい、乗り気、精悍、前向き、壮絶、壮烈、大胆、男らしい、痛烈、卑怯、悲壮、不敵、放胆、無為、無抵抗、無鉄砲、無謀、勇ましい、勇敢、勇壮、勇猛、雄偉、剽悍、獍猛
才能	あざとい、あどけない、いとけない、うまい、おしゃま、おませ、ぎこちない、こざかしい、さかしい、さとい、そそっかしい、たどたどしい、つたない、とろい、のろい、のろま、はしこい、ばか、ぶきっちょ、ぼんくら、まずい、スマート、マニアック、悪賢い、暗愚、迂愚、英明、鋭い、鋭敏、鋭利、下手、下手くそ、堪能、間抜け、頑迷、器用、器用貧乏、強い、愚か、愚かしい、愚直、愚鈍、慧眼、見巧者、賢い、賢明、口下手、巧み、巧妙、手早い、俊敏、小利口、詳しい、上手、新鋭、生意気、生半可、精鋭、精妙、青い、青臭い、拙速、拙劣、絶妙、浅はか、疎い、聡明、多才、多能、達者、知的、稚拙、遅鈍、頭でっかち、得意、鈍い、鈍重、乳臭い、博学、博識、非常識、敏活、敏腕、不慣れ、不器用、不得意、不得手、不能、不明、物見高い、物好き、分別臭い、凡庸、未熟、無学、無知、無能、無能力、無筆、無分別、明るい、明敏、有能、幼稚、利口、利発、理知的、伶俐、魯鈍、老巧、老練

材質	かたい、かちかち、かりかり、がさがさ、きめ細か、きれい、ぎざぎざ、ぐにゃぐにゃ、こちこち、こわい、さくい、しなやか、ふにゃふにゃ、ぶよぶよ、へなへな、もろい、やわらかい、エコロジカル、ソフト、ハード、フレキシブル、ラフ、汚い、汚らしい、滑らか、頑丈、希薄、堅い、堅固、堅牢、固い、硬い、黒い、細密、柔らか、柔らかい、柔軟、純粹、純正、小汚い、丈夫、清い、清らか、清潔、清澄、清冽、粗い、軟らか、軟らかい、粘い、粘っこい、濃い、濃厚、濃密、薄い、薄汚い、不潔、不浄、無垢、緻密
作用・変化	めまぐるしい、オートマチック、コンスタント、ダイナミック、フレキシブル、リズムカル、安定、円滑、旺盛、可変、活発、自然、自動的、自由、自由自在、柔軟、盛ん、静的、前向き、大変、当意即妙、動的、不安定、不易、不自由、不定、不変、流動的
時間	あっけない、一刻、永い、永続的、継続的、持続的、遠い、暇、久しい、悠久、悠遠
時機	タイムリー、急、出し抜け、唐突、定期的、周期的、規則的、慢性的
自信・誇り・恥・反省	いい気、おこがましい、くすぐったい、ずうずうしい、つつましい、ふてぶてしい、みっともない、やましい、恐縮、謙譲、謙遜、謙抑、誇らしい、後ろめたい、光栄、厚かましい、厚顔、照れ臭い、図太い、恥ずかしい、恥知らず、得意、不遜、不体裁、不名誉、不面目、名誉、面はゆい、面目、野太い、有頂天、僭越、僭上
自然	自然、物的
趣・調子	あっぱれ、うらやましい、お粗末、すてき、すばらしい、とんでもない、べらぼう、まずい、異常、一律、快調、結構、堅調、見事、好調、好都合、殺伐、殺風景、順調、勝手、上向き、上出来、絶好、絶好調、大味、単調、低調、頓狂、軟調、不具合、不首尾、不出来、不順、不振、不成績、不調、不都合、目覚ましい、立派
出没	あからさま、あけすけ、あらわ、おおっぴら、ひそか、隠密、極秘、顕著、公然、内密、秘密、露骨

詳細・正確・不思議	あやふや、いい加減、いぶかしい、おかしい、おぼつかない、おもしろい、けげん、つまびらか、ミステリアス、ラフ、ルーズ、易しい、怪しい、怪しげ、怪奇、確か、確実、甘い、奇異、奇怪、奇態、奇妙、疑い深い、疑わしい、興味深い、決定的、厳密、言うまでもない、巧緻、荒っぽい、克明、細かい、細心、細緻、雑、雑駁、自明、周到、周密、臭い、詳しい、詳細、慎重、推して知るべし、正確、精確、精細、精密、精緻、疎略、疎漏、粗い、粗雑、粗大、粗放、粗末、粗略、粗笨、大ざっぱ、大まか、大事、大丈夫、大切、大味、丹念、丁重、丁寧、的確、適当、難しい、難解、入念、念入り、濃い、濃厚、不可解、不可思議、不確か、不思議、不審、不正確、不定、不徹底、不透明、不得要領、不明、不明確、不明朗、分明、平たい、平易、平明、明らか、明快、明確、明細、明白、明瞭、明朗、明晰、綿密、用心深い、曖昧
色	どぎつい、どす黒い、グリーン、グレー、ゴールド、シルバー、ブルー、ホワイト、メタリック、黄色、黄色い、黒い、黒ずんだ、黒っぽい、純白、色っぽい、色濃い、真っ黒、真っ青、真っ赤、真っ白、真白、真白い、生白い、青い、青黒い、青白い、赤い、赤黒い、赤茶けた、浅黒い、鮮やか、蒼白、多彩、毒毒しい、白い、白っぽい、白茶けた
信念・努力・忍耐	あながち、いこじ、いちず、かいがいい、がむしゃら、ぐうたら、けなげ、こまめ、しつこい、しぶとい、ずばら、たまらない、なまくら、ねちっこい、ひたむき、まめ、まめまめしい、ルーズ、安逸、一意、一筋、一所懸命、一心不乱、横着、我慢強い、奇特、強情、勤勉、懸命、御苦労、自堕落、執念深い、執拗、手堅い、殊勝、真剣、真摯、神妙、辛抱強い、怠惰、怠慢、大わらわ、地道、着実、忠実、入魂、熱い、熱っぽい、熱心、熱烈、粘り強い、必死、筆まめ、不精、片意地、本気、未練がましい、夢中、無精、無理、几帳面

心	いざとい、うかつ、こちこち、さとい、そぞろ、デリケート、意気、鋭い、鋭敏、何気ない、過敏、虚心坦懐、強気、元気、散漫、耳ざとい、弱気、乗り気、正気、多感、多情、鈍感、敏感、不覚、無意識、無感覚、無関心、無気力、無自覚、無心、無神経、無頓着、無念、目ざとい
新旧・遅速	いとけない、うら若い、じじむさい、たけなわ、はやい、ほやほや、みずみずしい、ナウい、ホット、モダン、画期的、希代、旧式、古い、古めかしい、古臭い、斬新、事新しい、耳新しい、若い、若若しい、初初しい、小さい、尚早、新しい、新た、新鮮、真新しい、清新、生新しい、生鮮、早い、早急、速い、大きい、遅い、陳腐、年長、目新しい、幼い、幼少
真偽・是非	ほんと、まっとう、まとも、オーソドックス、ノーマル、フォーマル、リアル、公式、公正、自然、邪、真っ赤、真正、真面目、神秘、正しい、正式、正常、正当、生、切実、素、尊厳、端的、当然、不思議、不正、不良、変則、本式、本当、略式
身上	しがない、たっとい、とうとい、やんごとない、縁遠い、貴い、極道、孤独、高貴、尊い、天涯孤独、卑しい、富貴、末恐ろしい、末頼もしい、有為、有望、頼もしい
身体	きゃしゃ、なまめかしい、のっぼ、ふくよか、もじゃもじゃ、グラマー、スリム、華奢、魁偉、丸裸、血まみれ、血みどろ、骨太、弱体、小作り、小柄、色っばい、色白、赤裸、大柄、頭でっかち、肉感的、福福しい、豊満、毛むくじゃら、毛深い
進行・過程・経由	のっぴきならない、はかばかしい、スムーズ、一途、円滑
人生・禍福	ありがたい、いかめしい、まがまがしい、めでたい、シニカル、ラッキー、忌まわしい、気高い、厳か、厳肅、好運、幸い、幸せ、幸運、幸福、困難、仕合わせ、森厳、神聖、神秘、数奇、荘厳、荘重、騒がしい、尊厳、多難、薄幸、皮肉、非常、不運、不穩、不吉、不遇、不幸、不幸せ、不用心、物騒、物騒がしい、平和、豊穰、面倒、靈的、靈妙

人柄	あどけない、いこじ、うぶ、おおらか、おちゃめ、おつちよこちょい、おとなしい、かたくな、がちがち、さくい、しとやか、せっかち、そそっかしい、たおやか、つつましい、ひょうきん、ふがいない、まじめ、まっすぐ、まめ、まめやか、オープン、センチ、センチメンタル、ソフト、ドライ、ナイーブ、ヒステリック、フランク、ユーモラス、偉大、一徹、一本気、陰気、臆病、温厚、溫柔、温順、温良、穏やか、穏健、快活、楽天的、滑稽、頑固、頑迷、気さく、気軽い、気高い、気弱、気長、気難しい、義理堅い、強気、強情、怯懦、狭量、謹厳、謹直、苦勞性、軽はずみ、軽薄、軽率、軽佻、潔い、潔癖、堅気、堅実、嫌、控えめ、硬骨、高潔、高邁、剛健、剛直、豪気、豪放、骨っぽい、質朴、実直、弱気、従順、柔弱、柔順、柔和、重厚、純潔、純情、純真、純朴、勝ち気、小心、小胆、慎重、真率、神経質、酔狂、崇高、性急、正直、清純、清廉、誠実、昔かたぎ、浅薄、粗忽、素直、素朴、太っ腹、情弱、大様、鷹揚、短気、忠実、沈着、貞淑、鉄火、天衣無縫、伝法、度し難い、内気、軟弱、難しい、熱っぽい、薄っぺら、卑屈、不まじめ、不謹慎、不屈、不正直、不誠実、浮薄、負けず嫌い、物堅い、物静か、文弱、偏狭、偏屈、忘れっぽい、朴訥、奔放、無気力、無邪気、無心、無神経、無責任、無頓着、明るい、明朗、優しい、優柔不断、悠長、陽気、律義、率直、廉潔、朗らか、洒脱、洒落、磊落、闊達、飄逸
水・乾湿	かさかさ、からから、ぐしょぐしょ、ぐちゃぐちゃ、しとど、びしょびしょ、ドライ、陰湿、高燥、湿っぽい、湿潤、水っぽい、多湿、露
生	みずみずしい、フレッシュ、健在、後天的、若い、若若しい、初初しい、新鮮、生、生々しい、青い、先天的、早熟、長寿、長命、未熟
生活・起臥	いぎたない、しどけない、小食、少食、食いしん坊、寝坊、清潔
生物	男っぽい、女っぽい、セクシュアル、性的、無害、無毒、有害、有毒

生理・病気など	あつい、か弱い、たくましい、はれぼったい、ひ弱、ひ弱い、まめ、エネルギー、タフ、ヘルシー、旺盛、頑健、虚弱、強い、強健、健やか、健康、健全、堅固、元気、弱い、弱弱しい、丈夫、盛ん、青白い、壮健、足弱、多病、達者、篤い、熱い、熱っぽい、病弱、病的、不健康、不健全、不死身、無傷、勇健、羸弱
説・論・主義	アングラ、エロチック、シュール、ユーモラス、ロマンチック、楽天的、滑稽、詩的、人道的、道德的、美的、法的、抹香臭い
相対	あべこべ、パラレル、互い違い、正反対、相いれない、相応、相当、対等、平行、裏腹
走り・飛び・流れなど	のたりのたり、上滑り、足早
即時	いち早い、てきめん、急
速度	すばしこい、のろい、はしこい、まだるっこい、めまぐるしい、スピーディー、スロー、緩やか、緩慢、機敏、急、急激、急速、軽快、慌ただしい、手っ取り早い、手早い、俊敏、迅速、拙速、素早い、足早、速い、速やか、遅い、遅鈍、鈍い、敏捷、悠長
存在	あえない、うつろ、がらがら、ない、はかない、ニヒル、皆無、外的、空しい、空っぽ、空虚、空疎、絶え絶え、絶無、中空、特有、不滅、無人
多少	おびただしい、たくさん、たわわ、ふんだん、まれ、もじゃもじゃ、フル、寡作、過少、過多、過大、希少、巨額、僅少、控えめ、高額、些細、些少、残り少ない、手薄、十分、潤沢、小口、少ない、少なめ、少量、人少な、甚大、多い、多め、多額、多才、多事、多情、多大、多難、多能、多弁、多忙、多量、大量、莫大、微少、品薄、貧しい、貧弱、福福しい、豊か、豊潤、豊富、豊満、乏しい、膨大、無尽蔵、無数、量的、零細

待遇・礼など	きめ細かい、けんもほろろ、しかつめらしい、すげない、つれない、とげとげしい、なれなれしい、ぬるい、ひどい、ぺこぺこ、むごい、むごたらしい、やかましい、よそよそしい、シビア、ソフト、ビジネスライク、意地悪、陰険、陰湿、遠慮ない、横柄、苛酷、過酷、寛仁、寛大、寛容、甘い、緩やか、居丈高、恭しい、権高、権柄づく、謙虚、謙譲、謙遜、謙抑、厳しい、厳格、厳酷、厳重、控えめ、高慢、酷、酷薄、懇切、細やか、罪、罪作り、残虐、残酷、残忍、失敬、失礼、邪険、邪慳、手ぬるい、手厳しい、柔らか、柔軟、峻厳、峻烈、親切、辛い、人懐っこい、水臭い、尊大、他人行儀、大風、丁重、丁寧、軟らか、不敬、不親切、傍若無人、無遠慮、無礼、冷たい、冷厳、冷酷、冷淡、傲慢、慇懃、驕慢
地	果てしない、険しい、険阻、殺風景、山深い、寂しい、雪深い、地味、肥沃、不毛、物寂しい、豊沃、豊饒、明媚、木深い
弛緩・粗密・繁簡	いろいろ、がさがさ、がらがら、きちきち、きつい、ぎゅうぎゅう、しどけない、すかすか、たわいない、とりどり、ばらばら、まばら、ややこしい、コンパクト、シンプル、タイト、ピュア、ラフ、ルーズ、過疎、過密、甘い、簡潔、簡素、簡単、簡明、簡略、緩い、緩やか、巧緻、荒っぽい、細かい、細密、雑、雑多、至純、寂しい、手短、純、純一、純粹、純正、冗長、冗漫、精巧、精密、千差万別、粗い、粗雑、粗末、素、多様、単一、単純、濃い、濃密、薄い、煩雑、貧弱、不純、複雑、密、稠密、緻密
調和・混乱	がたがた、きれい、ぎこちない、ごっちゃ、ちぐはぐ、まぜこぜ、まちまち、むちゃくちゃ、むら、めちゃくちゃ、めちゃめちゃ、アンバランス、スムーズ、一樣、一律、怪異、滑らか、均一、均質、均等、系統的、合法、雑、支離滅裂、破格、非合法、不ぞろい、不規則、不均衡、不調和、変則、無秩序、乱雑、乱脈

長短・高低・深浅・厚薄・遠近	すれすれ、つんつるてん、のっぽ、はるか、ひょろ長い、迂遠、遠い、遠浅、遠大、奥深い、間近、間近い、近い、厚い、厚ぼったい、高い、高度、高率、根深い、最高、細長い、手近、小高い、小幅、冗長、深い、身近、浅い、大幅、短い、短め、長い、長たらしい、長め、長大、低い、低率、程遠い、薄い、薄っぺら、分厚い、目深、遼遠
程度	あんまり、いい加減、えらい、きちきち、けっこう、この上ない、こよない、したたか、すばらしい、とてつもない、とんでもない、どぎつい、べらぼう、むやみ、めちゃくちゃ、もっての外、やみくも、よっぽど、下らない、恐ろしい、極端、極度、高度、最高、散散、至極、手ごろ、小幅、深甚、人一倍、甚だし、甚大、絶大、相当、存外、大まか、大層、大抵、大幅、大変、著しい、底抜け、程よい、適度、適当、非常、微妙、法外、滅相、余計
程度・限度	ぎりぎり、過度、完全、完璧、極端、極度、軽度、限り、最高、最上、十全、重度、人一倍、適度、不全、不徹底、満足、無限、無制限、有限
等級・系列	ハイレベル、下等、高位、高級、高等、上等、低級、劣等
特徴	いかがわしい、おかしい、この上ない、こよない、つまらない、とっぴ、またとない、まれ、ゆゆしい、アブノーマル、エキセントリック、オリジナル、ユニーク、異色、異様、異例、乙、格段、格別、奇異、奇矯、奇警、奇想天外、奇抜、奇妙、苦手、型破り、際どい、重大、常套、新奇、深刻、身近、人間並み、尋常、粹、大抵、珍しい、珍奇、珍妙、珍無類、当たり前、特異、特殊、特別、特有、独自、独特、破天荒、非常、非凡、不自然、不備、普通、風変わり、物珍しい、平凡、変、凡庸、未完成、妙、目覚ましい、瑣末

難易・安危	きな臭い、けんのん、たやすい、ちょろい、つつがない、やすい、やばい、ゆゆしい、イージー、スリリング、ハード、安らか、安易、安穩、安全、安泰、安直、易しい、円満、穏やか、穩便、可能、楽、堪え難い、完全、甘い、簡易、簡単、簡便、危うい、危ない、危なげない、危なっかしい、危険、急、緊急、輕便、健全、見やすい、見よい、険しい、険悪、厳しい、好都合、困難、際どい、至難、手輕、十全、重宝、生易しい、静穩、絶体絶命、耐え難い、大丈夫、調法、度し難い、難い、難しい、不穩、不可能、不自由、不能、不便、平たい、平安、平易、平穩、平靜、便宜、便利、万全、無事、無難、無理、厄介、容易
熱	あたたかい、ぐたぐた、ぬくい、ぬくとい、ぬるい、ぬるまっこい、ほやほや、ホット、温か、温かい、暖か、暖かい、熱い、冷たい、冷やっこい
判断・推測・評価	案外、意外、意想外、推定、存外、望外、有望
美醜	あでやか、きらびやか、きれい、はで、はでやか、みっともない、ハンサム、ビューティフル、汚い、汚らしい、華やか、華美、華麗、見よい、見苦しい、秀丽、醜い、醜悪、純粹、小汚い、身ざれい、清い、清らか、清潔、清楚、繊細、端正、端麗、薄汚い、美しい、美的、美妙、不格好、不器量、不潔、不細工、不純、無垢、優しい、優雅、優美、妖艶、流麗、麗しい
必然性	あながち、せんない、やむを得ない、よんどころない、偶然、必定、不可避、余儀ない
表情・態度	からから、かりかり、すごい、にこやか、ぶっきらぼう、クール、陰気、陰気臭い、我が物顔、嫌、険しい、心得顔、難しい、白白しい、怖い、無愛想、無表情、憂鬱、陽気、冷徹、朗らか
評判	ポピュラー、高名、著名、不人気、名高い、有名
貧富	しがない、にぎにぎしい、にぎやか、にぎわしい、ゴージャス、リッチ、閑散、器用貧乏、豪華、豪勢、寂しい、盛ん、繁華、貧しい、貧乏、不景気、不振、不如意、富貴、富裕、豊か、裕福、隆盛、贅沢
分割・分裂・分散	ばらばら、ぼろぼろ、ぼろぼろ、切れ切れ、不可分

文化・歴史・風俗	いき、おしゃれ、お粗末、かっこいい、きざ、きざっばい、きゃしゃ、くだらない、けばけばしい、たおやか、つまらない、はしたない、ぼろぼろ、みすばらしい、みっともない、むさい、やぼ、やぼったい、わびしい、エキゾチック、エスニック、エレガント、エロ、エロチック、クラシック、グロテスク、シック、スポーティー、スマート、センチメンタル、デカダン、デラックス、ハイカラ、バタ臭い、モダン、安っぽい、安価、淫靡、汚らしい、乙、下劣、華奢、雅、怪奇、簡素、輝かしい、愚劣、軽妙、嫌み、古雅、古風、高雅、高尚、豪華、豪勢、豪壮、渋い、身ざれい、粹、生硬、盛大、繊細、粗末、粗野、素朴、壮大、壮麗、草深い、俗、俗っぽい、俗悪、地味、通俗、低俗、泥臭い、土臭い、当世風、卑俗、卑猥、尾籠、貧相、不格好、不潔、不粹、不体裁、不風流、風雅、風流、文弱、未開、無骨、無粋、無風流、野蛮、優雅、雄大、雄勁、洋風、瀟洒、猥雑、猥褻
包み・覆いなど	ほこりまみれ、血まみれ、血みどろ、汗みどろ、灰だらけ、傷だらけ、穴だらけ
方向	まとも、斜め、垂直、水平、多方面
防止・妨害・回避	邪魔、目障り
本体・代理	めばしい、スタンダード、主、主要、主流、重要、二義的、副次的
毎日・毎度	ちょくちょく、まれ、散散、頻繁
味	うまい、えがらっぽい、えぐい、えごい、おいしい、くどい、こてこて、しつこい、しょっぱい、ほろ苦い、まずい、まるやか、スパイシー、マイルド、リッチ、塩辛い、甘い、甘ったるい、甘酸っぱい、甘辛い、甘美、苦い、酸い、酸っぱい、脂っこい、渋い、辛い、大味、淡泊、濃厚、美味、芳醇、無味
未来	間近、間近い、近い、程近い
欲望・期待・失望	がめつい、けち、しわい、ぺしゃんこ、ぺちゃんこ、みみっちい、やぶさか、ハングリー、勘定高い、願わしい、強欲、思わしい、世知辛い、切、待ち遠しい、淡泊、不景気、望ましい、本望、無欲、欲しい、欲深、欲張り、吝嗇、貪欲、貪婪

理由・目的・証拠	もっとも、わりない、白白しい、不都合、無根、理不尽
良不良・適不適	いい、いい加減、うってつけ、うまい、カッコいい、そぐわない、ぴったり、ふさわしい、まし、まずい、よい、よろしい、悪い、悪質、悪性、穏当、過分、究竟、凶悪、極悪、結構、健全、好ましい、好適、控えめ、合目的的、最悪、最適、最良、似つかわしい、耳障り、失当、邪悪、邪魔、手ごろ、醜悪、純良、順当、上乘、生半可、絶好、善良、粗悪、相応、妥当、駄目、中途半端、宙ぶらりん、程よい、適格、適正、適切、適度、適当、適任、半端、不穏当、不格好、不均衡、不健康、不健全、不向き、不相応、不体裁、不適、不適切、不適當、不適任、不当、不良、分相応、無難、目障り、優良、良好、良質、劣悪
力	か弱い、きつい、きゃしゃ、こっぴどい、すごい、すさまじい、たくましい、ひどい、ひ弱、ひ弱い、ものすごい、ものすさまじい、エネルギー、ダイナミック、ハード、パワフル、苛烈、華奢、過激、虚弱、強い、強固、強大、強力、強烈、強靱、屈強、劇烈、激しい、激越、激烈、堅固、堅牢、荒い、根強い、弱い、弱弱しい、弱小、弱体、手ごわい、手ぬるい、手ひどい、手荒い、手痛い、精強、脆弱、足弱、大力、痛烈、軟弱、薄弱、非力、微力、貧弱、怖い、無敵、無力、猛烈、優勢、有力、揺るぎない、力強い、熾烈
礼	失敬、失礼、不敬、無礼、慇懃
連続・反復	とぎれとぎれ、間遠、小刻み
労働・作業・休暇	せわしい、暇、慌ただしい、多事、多忙、大忙し、忙しい