

Title	名詞の属性に着目した比喻理解に関する研究
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

名詞の属性に着目した比喩理解に関する研究

三島 達夫 (510096)

北陸先端科学技術大学院大学 情報科学研究科

2007 年 2 月 8 日

キーワード: 比喩理解, 属性名, 属性値, 顕現性, コーパス.

計算機による理解が難しい言語現象一つに比喩表現がある。我々の使用する言語にはほとんどと言っていいほど比喩的な表現が含まれており、人間の概念体系に比喩が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は通れない問題であるが、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」といった直喩表現があった場合、それは「赤い頬」といった意味が連想される。しかしそのような意味は構文解析や意味解析だけでは理解できず、「リンゴ」が持つ典型的な性質（顕現属性値）として「赤い」があるという知識が必要になる。自然言語処理の分野では、心理実験により名詞の性質を獲得して比喩理解を行う手法が多かった。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。

そのような背景を踏まえ、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。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喩解釈として選ぶモデルを構築する。

本研究における直喩の理解とは、「名詞Aのような名詞B」といった表現があった場合、喩詞「名詞A」から被喩詞「名詞B」に移される属性名と属性値の組を決定することである。属性名とは「色」や「形」といった属性値の分類であり、属性値とは「赤い」や「丸い」といった具体的な名詞の性質である。前述の「リンゴのような頬」の場合、「リンゴ」から「頬」へ「色：赤い」という性質が移されると判定することが本研究の目的である。属性名はシソーラスにおける形容詞または形容動詞の意味クラス、属性値は意味クラスに属する形容詞または形容動詞とした。シソーラスとして分類語彙表を用いた。また属性値や属性名としてふさわしくないものはあらかじめ人手で除いた。最終的に 100 個の属性名と 3006 個の属性値を用意した。

直喩表現の理解において移される属性を選ぶ際に重要になるのが、その性質がどれだけ名詞にとって典型的な性質であるかを示す顕現性である。顕現性を計算するためには名詞が持つ属性値に関する知識が必要である。本研究では様々な比喩表現に対応するために、

大規模な名詞の知識ベースとして、コーパスから名詞と属性値の共起頻度をパターンマッチにより獲得した。しかしながら、名詞と顕現属性値は典型的であるが故にテキストで明示されないという特徴もある。したがって単純な係り受けなどの共起パターンを用いるだけでは人間の直感と一致するように顕現属性値に対して高い共起頻度が得られるわけではない。したがって、名詞に対して顕現属性値を抽出しやすい共起パターンをいくつか発見し、これらのパターンによって名詞と属性値の共起頻度を求めることで、直喩理解の精度を上げることを試みた。

以上のようにして得られた名詞の知識ベース（名詞と属性値の共起頻度）に基づいて直喩理解を行うモデルを構築した。喩詞のもつ属性名、属性値の組に対してスコアを計算し、それが一番高くなるものを直喩の解釈の結果として出力する。スコアは、同じ属性名をもつ他の属性値と比較して頻度の差が大きい場合はスコアを高くする、顕現属性値を抽出しやすい共起パターンによって得られた頻度情報を単純な共起パターンで得られた頻度情報より重視する、被喩詞を修飾しやすい属性名の属性値に対してスコアを高くする、といったことを考慮に入れて設計した。また、名詞と共起する属性値の頻度の総和が低い属性名は、信頼できる頻度情報が得られなかったとみなし、「名詞A」から「名詞B」に移される属性名の候補に加えないようにした。

実際にコーパスから抽出した「名詞Aのような名詞B」という直喩表現に対して本研究の提案手法を適用した。その結果、属性名・属性値ともに適切な場合を正解としたときの精度が62%、属性値のみ適切な場合を正解としたときの精度が71%であった。一方、名詞Aとの共起頻度が最大の属性値を選択するベースラインの属性値のみを評価の対象とした精度は32%であった。これにより提案手法の有効性が確認された。また、異なる手法を比較する実験の結果、本研究の基本的な3つのアイデア、属性名毎に顕現性を計算すること、顕現属性値を抽出しやすいパターンを用いること、名詞Bと属性名の関連性を考慮に入れることの有効性が確認された。

最後に今後の課題を述べる。まず、顕現属性値を正確に抽出するパターンをさらに発見することで、名詞の顕現属性値をより正確に学習する必要がある。また、現在は「黒い」と「どす黒い」のようにほとんど同じ意味を持つ属性値を別の属性値とし、これらに対する共起頻度を個別に取り扱っている、ある名詞が「黒い」ためという顕現属性値をもつことが捉えられないことがある。「黒い」と「どす黒い」は同じ属性値とみなすなど、属性値の定義の見直しが必要である。