

“‘名詞’のような‘名詞’”の分類と比喩性の判定モデル

田添丈博

鈴鹿工業高等専門学校 電子情報工学科

榊井文人 権野努

三重大学 工学部 情報工学科

本研究報告は、“‘名詞’のような‘名詞’”表現の比喩性を判定するモデルを提案する。この表現には比喩（直喩）とリテラル（例示など）の2つの用法があり、また省略によってこの表現だけではどちらも判断がつかないあいまいなものもある。我々はそれらを機械的に判定するために、省略パターンを分類し、また比喩性を決定する語の存在を明らかにして、比喩性の判定モデルを構築した。実用例文より抽出した表現77個を用いて動作検証した結果を、人間が比喩性を判定した結果と比較したところ、典型的な比喩・リテラルについては正確に判定できることが示された。厳密に評価すれば、F値で比喩が79.1%、リテラルが89.9%という値を得た。

The Metaphorical Judgment Model

in case of a Phrase "'a Noun' like 'a Noun'"

TAZOE Takehiro

Department of Electronic and Information Engineering,

Suzuka National College of Technology

MASUI Fumito SHIINO Tsutomu

Faculty of Engineering, Mie University

We propose the metaphorical judgment model in case of a phrase "'a noun' like 'a noun.'" This phrase is classified into two usages, a simile and a literal one, but it has an ambiguous one we can't judge because of its ellipsis. We classified into its elliptical patterns, showed the existence of metaphorical words, and constructed the machine judgment model. On 77 practical phrases, comparing the judgment result by a system with by a human, this model can judge correctly on a simile usage (F-measure: 79.1%) and a literal one (89.9%).

1. はじめに

近年、“IT革命”という言葉に代表されるように、情報分野の認知度・重要度は急速に増し、それに伴って自然言語処理分野の研究もさらに活発なものとなっている。形態素論から順に構文論へと研究は進み、現在は意味論に関する研究がその中心となっている。比喻表現に関する研究はその代表的なもの1つであり、さらに詳しく分類された表現について、様々なアプローチによる研究が精力的に進められている[4-12]。

本研究は日本語文章の比喻表現、その中でも直喩・隠喩的な比喻について、その認識・抽出を目的としている。内山ら[9]は比喻表現を大きく直喩・隠喩的な比喻と換喩的な比喻とに分類し、また比喻の研究についても検出と解釈の2つがあるとしている。内山らは換喩的な比喻を研究対象としているが、我々は直喩・隠喩的な比喻を対象としている。これは喩詞と被喩詞との類似性に基づいた比喻である。また検出と解釈については、認識・抽出を目的とする場合、主体は検出である。しかし直喩・隠喩的な比喻の検出と解釈には密接に関係する部分があり、解釈にはまったく取り組まないという訳にはいかない。

我々は確率的なプロトタイプモデル[12]に対して、コーパスから知識を取り出すことによって比喻認識に用いる大規模な知識ベースを自動構築する手法を提案し[5]、動作に基づく属性に注目した観点からの比喻認識を提案してきた[6][7]。これにより喩詞と被喩詞とからなる表現の定量的な比喻性判断が可能となった。しかし、比喻認識の精度に影響を与える課題の1つに、喩詞と被喩詞とを入力として明確に与えることができるかという問題がある。これは比喻認識の前処理という位置付けで構文パターンやシソーラスを用いる手法で研究を進めてきた[4]。今回、直喩の代表的な表現である“‘名詞’のような‘名詞’”について、省略解析や比喻表現のための知識を取り入れることによって喩詞と被喩詞とを明確に与えることが可能となり、比喻性を判定するモデルを構築した。

まず、2章で“‘名詞’のような‘名詞’”表現についてその特徴を述べる。3章で我々が提案する比喻性判定モデルを説明し、4章でそれに基づく実用例文の判定実験を、5章で判定結果を人間のそれと比較する。6章ではあいまいについて考察を述べる。

2. “‘名詞’のような‘名詞’”表現について

2.1 用法の分類

比喻表現の中の直喩の代表的な表現形式に“‘名詞’のような‘名詞’”がある。しかし逆に、“‘名詞’のような‘名詞’”がすべて直喩であるとは限らない。この表現の用法は大きく2つに分けられ、1つは「じゅうたんのような芝」に代表される直喩であり、この用法を単純に『比喻』と呼ぶことにする。もう1つは「中国のような国」に代表される、比喻ではない用法であり、この場合は例示を意味する。このような比喻ではない用法は、他にも、指示、人の判断などがあり、それらをまとめて『リテラル(字義通り)』と呼ぶことにする。

我々は、これら2つの用法をコンピュータで自動判定するモデルを提案する。明らかにこの判定は構文情報だけでは実現できず、意味情報や概念情報を扱う必要がある。

2.2 省略パターンの分類

“‘名詞’のような‘名詞’”表現が比喻であるかりテラルであるかは、2つの名詞の意味的な関係によって決まる。しかしながら、2つの名詞を単純に比較判定してはいけない場合もある。省略が含まれる場合がそれで、省略のあるなしの典型的なパターンを説明する。

省略のないパターンは先ほどの「じゅうたんのような芝」「中国のような国」などがそれにあたり、『AのようなB』と表現することにする。これは“A”と“B”との意味的な関係を考慮すれば、用法の判定ができるものと考える。

これに対して省略のあるパターンは、例えば「針

のような形」など、『A (のc) のような (B) のc』と表現できる。この例では“針”と“形”を比較判定するのは間違いで、本来「針の形のようなBの形」であり、“針”と“B”とを“形”という観点で比較しなければならない。しかし一般的には、“B”はこの表現だけでは補完できず、よって判定にもあいまい性が存在する。

もう1つ特殊な省略パターンとして、「パールのようなもの」など、『Aのようなもの』という表現を考える。これは“もの”が代名詞的な働きをし、本来の“B”は前後の文脈に現れるものと考えられる。このパターンもそれだけでは“B”を補完することができず、『Aのようなc』と同様に正確に判定することができない。

2. 3 比喩性を決定する語の存在

判定には2つの名詞の意味的關係を比較すればよいと述べたが、名詞の中には意味的な影響力が強く、2つの名詞を比較するまでもなく比喩・リテラルを判定できる場合がある。そのような名詞を、比喩に対して『比喩語』、リテラルに対して『リテラル語』と呼ぶことにする。

比喩語は、例えば、前の名詞(『名詞1』と呼ぶ)が“夢”や“魔法”である場合で、あえて2つの名詞を比較しなくとも比喩と判定することができる。比喩語は名詞1に多く、慣用的な比喩表現に使われている語の辞書となる。

リテラル語はいくつかのカテゴリーで説明できる。まずは名詞1が“私”や“君”など代名詞となる場合で、名詞1の顕現性が低く比喩的な効果が得られない。同様に特定でない時間を表す語(“現在”や“従来”など)もそうである。次に名詞1が文の場所を表す語(“次”や“以下”など)の場合で、これは指示となる。そして名詞1が人の気持ちを表す語(“当たり前”や“快感”など)の場合で、これは人の判断や感情となる。最後に後の名詞(『名詞2』)が“例”となる場合で、これは文字どおり例示となる。表2. 1にこれらをまとめる。

表2. 1 リテラル語のカテゴリー

対象	カテゴリー	具体例
名詞1	代名詞	私、君、…
名詞1	特定でない時間	現在、従来、…
名詞1	文の場所	次、以下、…
名詞1	人の気持ち	当たり前、快感、…
名詞2		例

3. 比喩性の判定モデルの提案

“‘名詞’のような‘名詞’”表現を入力として、比喩あるいはリテラルを判定するモデルを提案する。

まず、比喩語・リテラル語が存在するかどうかをチェックする。比喩語が存在すれば比喩、リテラル語が存在すればリテラルと判定する。存在しなければ、次の処理へ渡す。

次は“AのようなB”、“Aのようなc”、“Aのようなもの”のパターンに分類する。“Aのようなc”、“Aのようなもの”と分類されたものは、“B”がわからない以上正確に判定できないとして、『あいまい』と判定する。“AのようなB”と分類されたものは、次の処理へ渡す。

最後に、AとBの意味的關係を比較して、比喩・リテラルを判定する。簡便に、AとBが異種カテゴリーの關係にあれば、それは比喩と判定する。AがBにカテゴリーとして含まれる關係であるとき、それは例示を表すとしてリテラルと判定する。

4. 実用例文での検証

実用例文として、日本経済新聞1994年1月分(約11万文)より、“‘名詞’のような‘名詞’”表現を抜き出した。その際に名詞に修飾語句がないことを条件として、81個が抽出できた。その中にはまったく同じ表現がいくつか存在したので、重複表現を取り除いて最終的に77個となった。

4. 1 比喩語・リテラル語が存在するか?

まず比喩語・リテラル語が存在するかどうかチェックした。比喩語については、我々が比喩語かどうかを判断しているが、複数人で判断している

こともあり概して正しいものであると考えている。リテラル語については、先に述べたカテゴリーに基づいて判断している。

77個の表現のうち、比喩語が存在したものは19個、リテラル語が存在したものは29個、存在しなかったものは29個と分類できた。比喩語の19個は比喩、リテラル語の29個はリテラルと判定する。残りの29個は、次のパターン分類の処理に渡す。

4. 2 省略パターンの分類

比喩語・リテラル語が存在しない29個の表現について、省略パターンの分類を行なった。その結果、“AのようなB”が8個、“Aのようなc”が9個、“Aのようなもの”が12個と分類できた。“Aのようなc”と“Aのようなもの”を合わせた21個は、この表現だけではBがわからないので、あいまいと判定する。“AのようなB”については、AとBとの意味的關係を比較する。

4. 3 AとBの意味的關係の比較

“AのようなB”表現8個について、AとBの意味的關係によって比喩かリテラルかを判定する。AとBが異種カテゴリーのものであれば、それは比喩と判定され、8個中3個であった。AがBにカテゴリーとして含まれる場合は例示を表すリテラルとなり、残りの5個がすべてそれに該当した。

4. 4 比喩性の判定結果

以上の結果をまとめると、“名詞”のような‘名詞’表現の77個中、比喩が22個(28.6%)、リテラルが34個(44.2%)、その表現だけでは比喩ともリテラルとも判定できないあいまいが21個(27.3%)となった(表4.1参照)。

表4.1 システムの判定結果

	比喩	リテラル	あいまい	計
システム	22 (28.6%)	34 (44.2%)	21 (27.3%)	77

5. 人間による判定との比較

システムの性能を検証するために、人間が比喩性を判定し、判定結果を比較する。三重大学情報工学科椎野研究室のM1の学生3人に、同じ77個の“名詞”のような‘名詞’表現について、比喩・リテラル・あいまいを判定してもらった。3人の判定結果をもとに、人間による判定結果をまとめる。そのまとめ方は、3人のうち2人以上が選んだものを判定結果とする。3人ともが別のものを選んだ場合は、それは比喩・リテラル・あいまいと分かれるはずなので、比喩にもリテラルにもなり得るあいまいを判定結果とする。

5. 1 人間による判定結果

まとめた判定結果より、77個中、比喩が21個、リテラルが35個、あいまいが21個となった(表5.1参照)。

表5.1 人間による判定結果

	比喩	リテラル	あいまい	計
人間	21	35	21	77

5. 2 システムと人間との判定結果の比較

システムの判定結果と人間による判定結果を比較するために、マトリックス型で表にしたものを表5.2に示す。

表5.2 システムと人間との判定結果の比較

		人間			
		比喩	リテラル	あいまい	計
システム	比喩	17	0	5	22
	リテラル	0	31	3	34
	あいまい	4	4	13	21
	計	21	35	21	

システムが比喩で人間がリテラル、システムがリテラルで人間が比喩と判定したものは1つも存在しなかった。このことから、典型的な比喩・リテ

ラルについては、システムは正確に判定できる。

厳密に比喩・リテラル・あいまいを区別すると
して、各表現の再現率・適合率・F値を計算して、
表5.3に示す。

表5.3 各表現の再現率・適合率・F値

	比喩	リテラル	あいまい
再現率 recall	81.0% (17/21)	88.6% (31/35)	61.9% (13/21)
適合率 precision	77.3% (17/22)	91.2% (31/34)	61.9% (13/21)
F値 F-measure	79.1%	89.9%	61.9%

リテラルが最もよく、比喩についても十分に満足
できる値である。このことから、比喩・リテラ
ルと判定する部分については、その基本的な考え
方は正しいものと言える。あいまいについては、
やや低い値となっている。

6. あいまいについての考察

システムがあいまいと判定したものには、パタ
ーンとして“Aのようなc”と“Aのようなもの”
とがある。“Aのようなc”パターンについては、
さらに比喩性の高低というレベルで分類するこ
とができる。

6.1 比喩性の高いもの

“Aのようなc”について、cが次のような場
合、比喩性は高いと考える。

- ① cがAの部分、付属物、所有物を表す言葉
シャトル (の羽根) のような羽根
- ② cが静的属性 (構造、形状など) を表す言葉
針 (の形) のような形
航空機 (の構造) のような構造
- ③ cが感覚 (五感) を表す言葉
杉 (の感じ) のような感じ
- ④ cが動的属性 (動作、態度など) を表す言葉
雑誌 (のペース) のようなペース

6.2 比喩性の低いもの

“Aのようなc”について、cが次のような場
合、比喩性は低いと考える。

- ① cが状態/性質を表す言葉

東京のような混乱

- ② cがAに付随するものを表す言葉

ボスニアのような問題

米国のような意識

ニュージーランドのような自然

6.3 判定結果の再比較

比喩性が高いものは比喩、低いものはリテラル
と判定するようにして、システムの判定結果と人
間による判定結果を再度比較する。比較結果をマ
トリックス型で表にしたものを表6.1に示す。

表6.1 システムと人間との判定結果の再比較

		人間			
		比喩	リテラル	あいまい	計
シ	比喩	21	0	6	27
	リテラル	0	34	4	38
テ	あいまい	0	1	11	12
	計	21	35	21	

人間が比喩・リテラルと判定するものは、システ
ムでもほぼ同じ結果を出すことが可能となった。

各表現の再現率・適合率・F値を再計算して、
表6.2に示す。

表6.2 各表現の再現率・適合率・F値 (再)

	比喩	リテラル	あいまい
再現率 recall	100.0% (21/21)	97.1% (34/35)	52.4% (11/21)
適合率 precision	77.8% (21/27)	89.5% (34/38)	91.7% (11/12)
F値 F-measure	87.5%	93.1%	66.7%

比喩・リテラルの再現率、あいまいの適合率が飛躍的に向上した。このシステムは人間よりも比喩・リテラルをより明確に判定する傾向があることがわかる。F値については、すべて向上している。あいまいについてはまだ低い値となっている。

6. 4 システムがあいまいと判定できないもの
あいまいの中で、システムが比喩と判定してしまう代表例を以下に示す。

屋根裏のような部屋

結晶のようなもの

古典のようなもの

システムは、“屋根裏”、“結晶”、“古典”を比喩語と判断して、比喩と判定した。しかし、少ない場合ではあるが、これらの表現はリテラルとなり、定義に従えばあいまいと分類される。このようにほとんどの場合は比喩、少ない場合にリテラルとなる表現は、現定義に従えばあいまいと判定されるが、普通のあいまいと同じカテゴリーでよいのか問題は残る。

次に、システムがリテラルと判定してしまう代表例を以下に示す。

昔のような姿

システムは“昔”を特定でない時間を表すリテラル語と判断して、リテラルと判定した。これも少ない場合によっては比喩性をおびることがあり、あいまいとなる。リテラル語はカテゴリーとして表現されているが、中には少数の例外もあり、さらなる精緻化が必要と考えられる。

7. おわりに

本研究報告では、これまでに我々の比喩認識手法の前処理として行なわれてきた構文パターンやシソーラスを用いて喩詞と被喩詞を抽出する手法を発展させて、省略解析や比喩表現に関する知識を取り入れて、“名詞”のような“名詞”表現の比喩性を判定するモデルを提案した。システムで得られた判定結果と人間とのそれを比較して、比喩についてはF値でおおよそ80%、リテラルについては90%と良好な結果を得ることができた。今後

は判定ルールのみさらなる精緻化を進める。

比喩表現は自然言語の中でもより人間らしい表現であり、これらを機械的に扱う研究を進めることで、自然言語処理のレベルを上げるうえでの重要な知見を得ることができると確信している。

最後に、三重大学情報工学科椎野研究室の学生のみなさんには、ディスカッションでの有益な意見交換、検証データの整理などで非常なお世話になった。ここで感謝の意を表したい。

参考文献

- [1] 中村, “比喩表現の理論と分類,” 共立出版, 1977.
- [2] 山梨, “比喩と理解,” 東京大学, 1988.
- [3] 中村, “比喩表現辞典,” 角川書店, 1995.
- [4] 田添 他, “比喩認識システムにおける比喩候補の抽出と分類,” 信学技報, TL-97-3, pp.17-24, May, 1997.
- [5] 榊井 他, “比喩認識システムのための知識獲得と比喩性判定,” 信学技報, TL-97-4, pp.25-32, May, 1997.
- [6] 榊井 他, “統計的手法を用いた比喩認識,” 信学技法, TL98-7, pp.25-32, Jul., 1998.
- [7] 榊井 他, “コーパスに基づく比喩性検出手法,” 信学技報, TL99-12, pp.41-46, Dec., 1999.
- [8] 内海 他, “関連性理論を用いた文脈のなかの隠喩解釈の計算モデル,” 情報処理学会誌, vol.37, no.6, pp.1017-1029, 1996.
- [9] 内山 他, “統計的手法による換喩の解釈,” 自然言語処理, Vol.7, No.2, Apr., 2000.
- [10] 村田 他, “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, 15巻3号, pp.503-510, 2000年5月.
- [11] 今井 他, “比喩理解のための顕現属性の幾何学的抽出法,” 人工知能学会誌, 15巻2号, 2000年3月.
- [12] 岩山 他, “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, 6巻5号, pp.674-681, 1991年9月.