

Web からのメタファ抽出方式

間中 宏修[†] 山田 剛一[†] 絹川 博之[†]

東京電機大学大学院工学研究科[†]

1. はじめに

メタファ（隠喩）とは、日常的に使用される言葉の至る所に存在する比喩表現のひとつである。ある概念を既存の知識やイメージの概念領域を用いて理解することで、理解しやすく簡潔な言葉の表現が可能である。このような表現は、あまり聞きなれない新語や理解の困難な言葉の伝達手段に有効である。よって、このような表現を収集し、新しい表現の言葉を見つけることは有益である。

本研究では、Web 情報からメタファを自動的に抽出することを目的とし、その処理方法を考察する。

2. 直喩表現の利用

直喩は「あたかも」、「ようだ」などの定型語を使用し、その言葉が比喩であることを明示している表現である。それに対してメタファは、定型語を使用せず、比喩であることが明示されていない表現であり、表現構成語の特徴からメタファを抽出することは困難である。

メタファの中には、直喩表現で表されているものも存在する。よって、直喩表現の解析結果から得られたメタファのパターンを元にして、そのパターンに一致するものをメタファとして抽出する。

3. メタファ抽出の処理過程

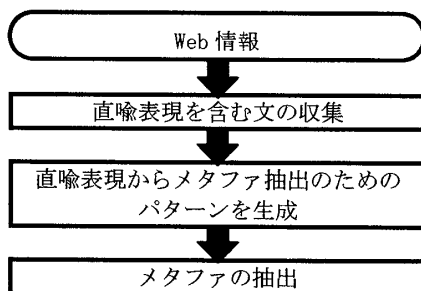


図 1. メタファ抽出の処理過程

図 1 にメタファ抽出の処理過程を示す。はじめに処理の対象となる「Web 情報」から「直喩表現を含む文を収集」する。そして、「直喩表現からメタファ抽出のためのパターンを生成」し、最後に「メタファの抽出」を行う。

3.1 Web 情報

ウェブ検索エンジン Google[1]より、3.2 に示す直喩表現の構文パターンを検索クエリとして直喩を含む文を収集する。

3.2 直喩表現を含む文の収集

インターネットの Web 情報に存在する直喩表現は、典型的な形式として「A は B のよう」のように連辞的な形式をしており、被喩辞、助詞、喩辞、定型語から構成されている。被喩辞はたとえられる語、喩辞はたとえる語であり、名詞同士を繋ぐ連辞的な意味を示す助詞は「は」であることが多い。定型語としては、ここでまず直喩表現の中でも数多く用いられる「のような」を対象とする。Web 情報内で多く使用される定型語を図 2 に示す。図 3 に収集の対象となる直喩表現の例を示す。

「のような」「のように」「のようだ」「のようで」、「みたいな」、「みたいに」、「みたいね」、「みたいや」、「みたいよ」、「みたいだ」、「みたいで」

図 2. 直喩表現を明示する定型語

言葉	は	マジック	のような
[被喩辞]	[助詞]	[喩辞]	[定型語]

図 3. 連辞的な直喩の例

上記のような直喩表現のパターンを検索クエリとし、Google にて検索を行い、その結果を収集する。ただし、被喩辞と喩辞は任意の文字列とする。

3.3 被喩辞と喩辞の特徴の分析

3.2 の処理によりパターンに一致した文について、被喩辞と喩辞の特徴から、直喩表現と比喩表現でないもの

Extraction of Metaphor from Web Information.

Hironobu MANAKA[†], Koichi YAMADA[†], Hiroshi KINUKAWA[†]

[†] Graduate School of Engineering, Tokyo Denki University

を判別する。

3.3.1 直喩表現の被喩辞と喩辞の特徴

今回は、連辞的な形式の直喩表現「(名詞)は(名詞)のような」の形式の文を対象とする。ただし、以下のものについては直喩表現ではないとする。

(1) 指示語

言葉の対象が何を指示しているのかが明記されていない指示語は、被喩辞と喩辞の関係が成立していない。

例：その使用形態は以下のようなものです。

(2) 代名詞

代名詞も指示語と同様である。代名詞の判定は形態素解析エンジン MeCab[2]を使用して判別する。

例：これは常識のような気がします。

(3) 時相名詞

「今日は嵐のような日だ」のように被喩辞が名詞の副詞可能である場合、「今日」という時間的な制限を持つ。

また、「日本は昔のような」は、「昔」が「いつ」なのかが明確ではない。よって直喩表現ではないとする。副詞可能も代名詞と同様、MeCabにより判別される。

(4) その他

「自分」、「中」等、被喩辞と喩辞において具体的な意味を持たない抽象的な言葉が存在する。

例：課長は自分のような若輩者に勤まるかどうか…

表1に連辞表現の分類結果を示す。

表1. 連辞表現の分類結果

項目	対応する直喩の数
直喩表現	115
指示語	107
代名詞	14
時相名詞	9
その他	55
合計	300

3.3.2 メタファ表現になりえない直喩表現

下記の条件に一致する直喩はメタファになりえない表現である。

(1) 慣用語

「夢」、「奇跡」、「天国」等、喩辞が慣用的に使用される表現である場合。

例：ラスベガスは夢のような場所だ。

(2) 既知概念

メタファの喩辞は、人間が持つ既知の概念で喩えられる。よって、喩辞が形態素解析によって未知語と判断された場合はメタファでないものとする。

例：水力発電はKERSのようだ。

(3) 隣接関係 (換喩)

被喩辞と喩辞の関係が概念的に近い比喩表現である場合。

例：手紙ははがきのようなものである。

(4) 包含関係 (提喩)

被喩辞と喩辞の関係が上位一下位で表される場合。

例：その桜は八重桜のような花を咲かせる。

3.3.3 被喩辞と喩辞の共通の性質

直喩表現の中でメタファにもなりうるものは、被喩辞と喩辞の間に何らかの共通の性質がある。今回の調査の結果、共通の性質は形容詞や動詞で表されることが多かった。

例：マンステールは納豆のようなニオイだ。

表2に直喩表現の分類結果を示す。

表2. 直喩表現の分類結果

項目	対応する直喩の数	
メタファになりうる表現	89	
メタファになりえない表現	慣用語	11
	既知概念	8
	隣接関係	2
	包含関係	5
合計	115	

4. おわりに

本論文では、直喩に用いられる比喩表現のパターンを解析し、そのパターンを元にメタファを抽出する処理方法を考察した。

今後は、「のように」、「みたいだ」など他の直喩表現の定型語について調査する。調査結果からメタファ抽出のためのパターンを生成し、メタファ抽出プログラムを作成する。作成プログラムについて実験評価し、改良していく予定である。

5. 謝辞

本研究は、形態素解析エンジン MeCab[2]を使用しており、それを開発された方々に感謝致します。

参考文献

- [1]Google ウェブ検索,
<http://www.google.com>
- [2]形態素解析エンジン MeCab,
<http://mecab.sourceforge.net>
- [3]山梨正明, 岩田純一:比喩と理解, 東京大学出版会