

二字熟語を構成する漢字間における比喩関係の扱い

3R-5

矢野 啓二郎 瀧口 伸雄 小谷 善行
 (東京農工大学 工学部 電子情報工学科)

1.はじめに

自然言語処理システムにおいて、未知語の処理を行うことは必須である。我々は処理対象を二字熟語に限定した未知語処理の研究を行っている。その方法は、形態素解析中の二字熟語の未知語の意味分類を推定し、その結果をフレーム構造によって表現するものである。これにより、二字熟語の意味を表現する。

これまでに、二字の漢字のそれぞれが持つ意味分類、格支配構造から規則を作り、二字熟語の意味構造を推定する方法を提案した[1]。

さらに本稿では、二字の漢字の間の関係が比喩関係にある二字熟語についての考察を行う。

2. 二字熟語を構成する漢字間の関係

二字熟語を構成する漢字間の関係を、次の項目により分類した。

(1) 漢字の間になんらかの語句（例えば、活用語尾、助詞など）を補うことによって、文に変形できるもの

(1-1) カテゴリー対応規則[1]により推定できるもの

熟語を構成する漢字の意味分類、格支配構造により、文法的に構造が決定される。

例 「美人」は「美しい人」のように「美」→「美しい」という直接的対応を使って、文章に変形できる。

(1-2) 比喩関係にあるもの

熟語を構成する漢字間の関係が比喩として理解される。

例 「狼男」は「狼のような男」のように「のような」を補うことによって文章に変形できる。

(1-3) その他

例 「赤恥」は「赤い恥」であるが、「赤い」という語に、「恥」を修飾する特殊な機能が

ある。

(2) 漢字の間に語句を補っても、文に変形できないもの

(2-1) 一つ一つの漢字から意味を推定できるもの
 例 格段

(2-2) 一つ一つの漢字から意味を推定することができないもの

例 横柄

文献[1]では、(1-1)に分類される未知語の処理について述べた。本稿では、(1-2)に分類されるものについて述べる。

3. 比喩関係にある二字熟語の顕現性処理

本稿では、扱う比喩を属性比喩[3]に限定する。二字熟語では、当然比喩を表現するために利用できる語は二語しかない。したがって、二字熟語で考えられる比喩の形は、直喩（または隠喩）である。諷喩、提喩、声喩は二字熟語では表現できないと考えられる。活喩においては、少なくとも概念が三つ必要になると考えられるので、二字熟語で表現するのは難しいと思われる。さらに、直喩（または隠喩）を、構造比喩、関係比喩、属性比喩という分類により分けると、二字熟語の漢字間の比喩関係の多くが、属性比喩に含まれると考えられる。

3.1 顕現性の算出方法

二字熟語で、一文字目の漢字をA、二文字目の漢字をBとする。これが、「AのようなB」あるいは、「BのようなA」となると考えて処理をする。AとBの順番を考える場合に、[2]の顕現性の式を利用する。すなわち、比喩で例えられている概念が、AからBに移されやすいか、BからAに移されやすいかを求めるために、顕現性を用いる

また、二字熟語の場合、漢字1字の持つ意味分類によって、AとBの順番に傾向が見られる。この傾向を顕現性の式に加味し、計算することで熟語の意味構造を求める。この傾向は、意味分類に、比喩になる場合は、前後どちらになりやすいかの割合（以後、“位置重み”と呼ぶ）を与えて

Processing of Metaphorical Relation

Between Kanji Characters

Keijirou YANO, Nobuo TAKIGUCHI, Yoshiyuki KOTANI

Tokyo University of Agriculture and Technology

おくことで実現する。

処理アルゴリズムを以下に示す。

①「AのようなB」で、Aの持つすべての性質に対して熟語の顕現性の集合(i_sal_1)を以下の式により計算する。

$$i_sal_1 = \text{saliency}(*(\text{A}), S_{i, \max}) \times \text{weight}(\text{A}, \text{前})$$

$*(\text{C}) = \{S_1, S_2, \dots, S_n\}$
: Cは性質 S_i の集合

$$\text{saliency}(*(\text{C}), S_{i, \max})$$

: [2] による顕現性の算出結果

$S_{i, \max}$: 概念 $*(\text{C})$ の最尤性質

$\text{weight}(\text{C}, \text{X})$: CのXの位置における意味分類の位置重み
($0 < \text{weight}(\text{C}) \leq 1$)

②「BのようなA」で、Bの持つすべての性質に対して顕現性の集合を以下の式により計算する。

$$i_sal_2 = \text{saliency}(*(\text{B}), S_{j, \max}) \times \text{weight}(\text{B}, \text{後})$$

③ ①②の算出結果より、二字熟語の顕現性の集合(idiom_saliency)を以下の式により算出する。

$$\text{idiom_saliency} = \text{set_from_value}(\max(i_sal_1 \cup i_sal_2))$$

$\text{set_from_value}(\text{X})$: 値Xを持つ集合をすべて返す

3.2 算出例

熟語「狼男」を例に顕現性の算出例を示す。

(1) システムが持っている知識

システムが持っている知識として、以下を定義する。

表1 位置重みの知識

意味分類	位置重み	
	前	後
動物	0.8	0.4
仲間	0.1	0.1

表2 漢字1字の知識

狼		
意味分類	value	動物
性質	value	獍猛(属性値=0.6)
性質	value	冷酷(属性値=0.4)
犬		
意味分類	value	動物
性質	value	獍猛(属性値=0.4)
性質	value	温厚(属性値=0.6)

男		
意味分類	value	仲間
性質	value	獍猛(属性値=0.3)
性質	value	温厚(属性値=0.2)
性質	value	冷酷(属性値=0.3)
性質	value	柔和(属性値=0.2)

人		
意味分類	value	仲間
性質	value	獍猛(属性値=0.25)
性質	value	温厚(属性値=0.25)
性質	value	冷酷(属性値=0.25)
性質	value	柔和(属性値=0.25)

(2) 「狼男」の顕現性の算出例

① i_sal_1

$$= \{\text{saliency}(*(\text{狼}), \text{性質: 獍猛}) \times \text{weight}(\text{狼}, \text{前})\}$$

$$= \{0.0232\}$$

② i_sal_2

$$= \{\text{saliency}(*(\text{男}), \text{性質: 獍猛}) \times \text{weight}(\text{男}, \text{後}),$$

$$\text{saliency}(*(\text{男}), \text{性質: 冷酷}) \times \text{weight}(\text{男}, \text{後})\}$$

$$= \{0.00145, 0.00145\}$$

③ ①②の算出結果より、

$$\text{idiom_saliency}$$

$$= \text{set_from_value}(\max(i_sal_1 \cup i_sal_2))$$

$$= \{0.0232\}$$

すなわち、「狼のような男」と、男の獍猛さを狼にたとえて表現した比喩である。

4. おわりに

二字熟語を構成する漢字が、比喩関係にあるものの顕現性の算出方法について述べた。なお本稿では、位置重み、属性値を筆者らが適当に定義した。

本稿で示した顕現性の算出方法により、比喩関係にある二字熟語の意味構造を推定することができる。

5. 参考文献

[1] 矢野ほか：二字熟語意味推定のためのカテゴリ-対応規則，第46回全国大会講演論文集(3)，pp.3-103,104(1993)

[2] 岩山ほか：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 Vol.6 No.5，pp.46-53(1991)

[3] 諏訪ほか：比喩の計算モデル，情報処理 Vol.34 No.5，pp.556-575(1993)

[4] 大野ほか：類語国語辞典，角川書店(1990)