

コア概念に基づく多義動詞のクラスタリング

3H-9

宮脇武久

宮崎正弘

新潟大学大学院工学研究科

1 はじめに

意味解析用の機械辞書を構築する際の語義記述のデータ構造として類語弁別ネットワークが提案されている[1]。動詞の類語弁別ネットワークをIPAL動詞辞書[2]を用いてボトムアップに下位語から上位語を見つけて構築していく[3]際に、下位語や上位語に多くの多義が存在する場合がある。多くの多義が存在する場合、それぞれの意味を持つ動詞についてネットワークを張ることは効率が悪く、体系が複雑になってくる。本稿では、このような問題の解決策として動詞のコア概念に基づいたクラスタリング法を提案する。

2 多義の種類

IPAL動詞辞書では、一つの見出し語（エントリ）はその意味および統語的特徴に基づいて下位区分（サブエントリ）されている。それら多義の存在する動詞を調べてみるとその動詞を構成する漢字より大まかに次の三種類に区別される。

1. 同音意義語として別エントリとして記載されている動詞
2. 異表記であるとして同じエントリに記載されている動詞
3. 一種類の表記で多義がある動詞

多義1の場合、「合う」「会う」がそれぞれ別のエントリとなっており、「会う」の異表記として「逢う」「遭う」が記載されている。

Clustering of Ambiguous Verb Based on Core Concept

Takehisa Miyawaki, Masahiro Miyazaki
Niigata University

表 1: 多義の具体例

	動詞	表記
1	あう	合、会
	うつす	移、写
	きく	聞、効
2	あがる	上、挙、揚
	あける	開、空、明
	うつ	打、撃、討
3	くる	来
	でる	出
	はいる	入

多義2の場合、「あがる」という動詞は一種類のみがエントリとして記載されており、文脈によって「上がる」「挙がる」「揚がる」の表記が用いられる。

多義3の場合、「でる」は「出る」という一種類の表記で32の多義が存在している。

3 クラスタリング

3.1 格の要素によるクラスタリング

IPAL動詞辞書では、多義すべてに書き換え可能な異表記がある動詞と多義の中である意味についてのみ異表記がある動詞の二種類がある。前者の場合には、後で述べるコア概念を用いてクラスタリングを行なう。後者の場合、動詞の主語や目的語の違いにより異表記があり、意味が微妙に違ってくるため、主語の要素や目的語の要素を調べることによって分類することが可能である。

例えば、「うつ」という語は、ほとんど「打つ」という表記であるが「鉄砲をうつ」や「親の仇をうつ」といった限られた用法ではそれぞれ「撃つ」

「討つ」といった他の漢字が用いられる。つまり、「鉄砲などを発射する」の意味や「敵を攻撃する」の意味を持つ場合に用いられる「うつ」には、「打つ」の表記を用いる場合と微妙に意味が異なっていると考えられる。

3.2 コアを用いたクラスタリング

コア理論 [4] では、多義を持つ動詞には抽象度の高い語義が存在しており、文脈により語義の具体化が生じていると考える。そこで、多義をある程度含む抽象的なコア概念を仮定し、多くの多義が存在する動詞を抽象的ないくつかのまとまりに分類する。

ある程度似た意味を持つ多義の場合には、上位語や反義語、類義語が等しくなると考え、最初にそれらによってまとめる。次に、IPAL 動詞辞書に記載されている角川類語新辞典の分類名称や意味記述を参照して似た意味を持つものをまとめていく。以上の作業を多義が多く存在する動詞について行ない、クラスタリングする。

例えば、「でる」には“脱出・取り出し”の意識を持つ概念と“出現”の意識を持つ概念の二種類の概念があると考えることができる。(図 1)

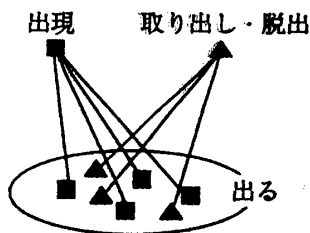


図 1: 「出る」の多義

二種類に分類した概念の中でも文型や格要素により意味が違ってくる。それら文型情報や格要素情報を付加すれば、さらなる使い分けが可能である。

4 コア概念を用いたネットワーク

クラスタリングを行なった動詞についてそれを下位語または上位語としてネットワークを張る。

下位語である場合、その上位語はクラスタリングされた概念について複数存在する場合がある。また、上位語である場合にはそれぞれのクラスタリングされた概念に下位語がくる。(図 2)

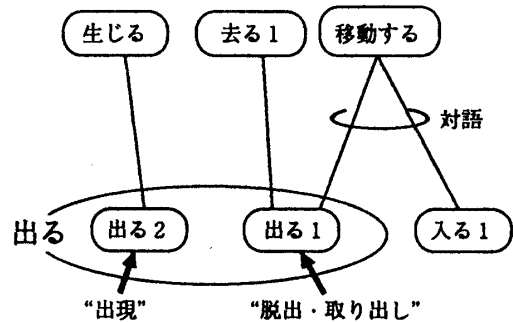


図 2: ネットワークの具体例

図 2 のようにクラスタリングを行なった動詞を用いて類語弁別ネットワークを構成していく。

5 おわりに

多くの多義が存在する動詞にコアを持ちいたクラスタリングを行ない類語弁別ネットワークを構成する方法を提案した。今後は、クラスタリングされた動詞を用いて類語弁別ネットワークを作成していく。最後に、IPAL 基本動詞辞書データの使用を許可された情報処理振興事業協会技術センターに深謝する。

参考文献

- [1] 中田、宮崎：類語弁別ネットワークの半自動生成、第 47 回情処全大、No.1M-2(1993)
- [2] 情報処理振興事業協会 (IPA) 技術センター：「計算機日本語基本動詞辞書 IPAL (Basic verbs)」(1987)
- [3] 宮脇、宮崎：類語・対語に着目した動詞類語弁別ネットワークの構築法、第 49 回情処全大、No.3G-1(1994)
- [4] 田中茂範：認知意味論 英語動詞の多義の構造、三友社出版 (1990)