

3C-3

二字漢字の語構成と概念説明文を用いた サ変動詞類語弁別ネットワークの半自動生成法

高坂光俊

宮崎正弘

新潟大学大学院自然科学研究科

1 はじめに

自然言語処理用の機械辞書を構築するさいに語義記述法が大きな問題となっている。人間用の辞書では、語義を他の語で置換し、その語の語義を定義していることが多い。その結果、語義記述の循環や曖昧性が起こり、機械辞書として人間用の辞書を用いることは不適當である。この様な問題点を解消するものとして、類語弁別ネットワークが提案されている^[1]。

本稿ではEDR概念辞書^[2]を用いて、二字漢字の語構成と概念説明文に着目して、サ変動詞の動詞類語弁別ネットワークの半自動生成法を提案し、その有効性を示す。

2 二字漢字

本稿では、上位語が和語動詞、下位語はサ変動詞という組合せに注目し、このサ変動詞についての類語弁別ネットワークの構成法を提案していく。ここで、和語動詞とは漢語の渡来以前からあった日本古来の動詞であり、最も基本的な動詞のことである。サ変動詞とは、主に漢語に「する」を加えたもので、「二字漢字」+「する」という形をとる。

表1：和語動詞とサ変動詞の例

和語動詞	書く・読む・知る・食べる・・・
サ変動詞	移動する・表現する・分裂する・・・

サ変動詞に注目した理由を次に示す。

サ変動詞は二字漢字より成り立っているが、それぞれの漢字の持つ性質により、大きく分けて次の2種類がある。

- 二字漢字の両方が動詞性のサ変動詞
- 二字漢字の片方が動詞性のサ変動詞

このことから、和語動詞の下位にあたるサ変動詞は、その和語動詞と類義の漢字を含んでいる可能性がかなり高いので、上位語の和語動詞をどの様に修飾しているのかによって、様々なサ変動詞が生成されているということになる。つまりサ変動詞間における修飾の仕方の違いによって分類すればいいわけであり、すなわち弁別特性を付与し易くなるということにつながる。

サ変動詞を構成する一字漢字間の結合パターンをもっと細かく分類すると、次のように分類できる。この分類をする際、二字漢字の字面だけではどのグループに属するか、完全には決定できないので、EDR概念辞書に記述されている概念説明文（語義記述文）を参照することにより、どのグループに属するかを決定した。

- N + V（ガ格）例：地震、国立、日没
- N + V（ヲ格、ニ格）例：水防、敵視、前進
- N + V（デ格）例：指斥、図解、手筆
- V + N（ガ格）例：降雨、積雪、落雷
- V + N（ヲ格、ニ格）例：読書、開店、登山
- V + N（デ格）例：執筆
- V + V（修飾）例：競泳、飛来、焼死
- V + V（並列）例：増加、破壊、断絶
- V + V（対立）例：売買、往復、進退

Construction of Verb Discrimination Network Based on Morphological Analysis and Concept Explication of Compound Verb

Mitsutoshi kousaka, Masahiro Miyazaki

Niigata University

- Adv + V 例：全壊、再発、予感

実際に、EDR 概念辞書から抽出した「書く」の下位語のサ変動詞 14 語について、それぞれの概念説明文を参照しながら、どのグループに属するかを検討してみたところ次のようになった。括弧内は、EDR 概念辞書に記述されている概念説明文である。

- N + V (ガ格)
例：直書する (自ら筆をとって書く)
- N + V (ヲ格、ニ格)
例：注記する ((本や記事に) 注を書き入れる)、親書する (手紙を自分自身で書く)、落筆する (書画を書くこと)、図取りする (物の形を図に写し取る)
- N + V (デ格)
例：染筆する (筆で書く)、手筆する (自らの手で書く)
- V + V (修飾)
例：記入する (記入する)、追記する (あとから書き加える)
- Adv + V
例：自記する (機械が自動的に記録する)、寸描する (簡単に描くこと)、清書する (ていねいに書き直す)、別記する (本文のほかに書き添える)、明記する (はっきり書く)

3 ネットワークの構成

先ほど説明した分類方法を用いてグループ分けをしてから、そのグループ内でさらにお互いの差分を見つける。そして、それを弁別特性として付与することにより、ネットワークを構築していく。前章の例を用いると、N + V (デ格) の「染筆する」と「手筆する」の違いは、「筆」で書くか「手」で書くかの違いであり、この差分を弁別特性とする。

このネットワークを構築することによって、上位語の語義を記述しておけば、その差分から下位語の語義記述を導くことができる。

4 問題点

二字漢字を構成している漢字間の結合パターンの分類をしたが、分類の仕方かなりの曖昧性がある。

実際に「書く」の下位語のサ変動詞について適用してみたが、その中の「手筆する」を例にとって、その曖昧性について検討してみる。「手筆する」の概念説明は「自らの手で書く」となっているが、字面からすると先ほどのように N + V (デ格) に当てはまる。しかし、意味を考えると「自分が書く」と解釈できるので、N + V (ガ格) に当てはまる。この様に解釈の仕方によって、どのグループに属するのかが曖昧になってくることがあるので、この曖昧性の解消が今後の課題と思われる。

また、弁別特性の付与の仕方についても、より明確な方法を見つけていかなければならないだろう。

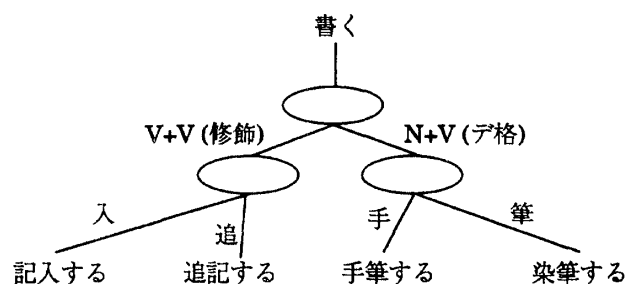


図1：サ変動詞類語弁別ネットワークの一例

5 おわりに

和語動詞を上位語とする場合の、サ変動詞類語弁別ネットワークを半自動生成する方法について述べた。そして、ネットワークを構築する際に必要となる弁別特性を、概念説明文を参照しながら、二字漢字を構成している漢字間の結合パターンを分類することにより、それを得る方法を示した。

現段階では、ネットワークを半自動で生成するのは少々苦しいと思われるが、二字漢字を構成している漢字間の結合パターンをより明確にしていけることが、半自動化のために必要である。

最後に、EDR 概念辞書の使用を許可された日本電子化辞書研究所に深謝する。

参考文献

- [1] 中田、宮崎：類語弁別ネットワークによる語義記述法、自然言語処理の基本問題シンポジウム論文集、pp.1-8 (1992)
- [2] 日本電子化辞書研究所：「EDR 電子化辞書」