

語を種々の観点から分類した多次元シソーラス

3Q-2

川村 和美 宮崎 正弘
新潟大学大学院工学研究科

1. はじめに

単語（概念）間の関係を表すシソーラスについては、上位/下位関係を中心に作成されたものが多い。しかし、単語間の関連には全体/部分、機能などの観点もある。また、上位/下位関係でも詳しく見ると、種々の観点が存在している。そこで、複数の観点から単語間の関連を表す多次元シソーラスが必要である。本稿では、名詞について、多次元シソーラスの有効性とその構成法について述べる。

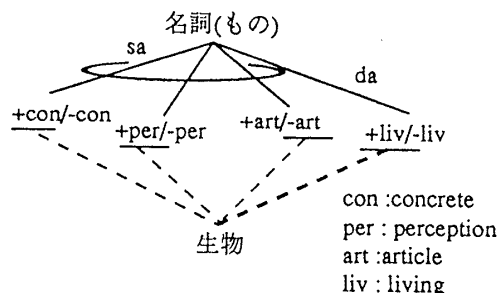


図1: 多次元シソーラスの上位の分類観点

2. 現在のシソーラスの問題点

(1) 単語（概念）間には、上位/下位関係と同時に種々の観点とそれによる関係が存在している。例えば、「鳥」と「飛行機」は上位/下位関係でみれば、それぞれ「動物」と「人工物」であり、シソーラス上かなり「遠い（離れている）」ことがわかる。しかし、「飛ぶ」という観点でみれば両者は非常に「近い」といえる。このように、語と語の関係はその観点によって、「遠く」なったり「近く」なったり変化する。語と語の共通点を見いだしたり、それらのある観点からクラスタリングすることは、自然言語理解において、非常に有効な手段である。それにもかかわらず、現在のシソーラスで表現することは困難である。そこで、観点による関係の変化を表現するために、語を種々の観点から記述するのが多次元シソーラスである。

(2) 現在のシソーラスで上位/下位関係は、すべてが同じ観点で分類されているわけではなく、分類の性質が異なっている。例えば、「動物」を「鳥」「魚」「獣」と分類する観点と、「人工物」を「乗りもの」「家具」と分類する観点は異なっている。前者は、形態や、生態などによる分類であるのに対して、後者は、用途や、機能による分類である。そこで、多次元シソーラスでは、二項間の関係を明確に表現する。

(3) シソーラスは階層的であるため、語の分類に順序が介入する。しかし、順序付けを行なうのが適当ではない分類観点もある。分類の順序によって、語と語の間の適当な関係を得られない場合がある。つまり、「鳥」「飛行機」を上位で、自然物、人工物と大別してしまうために、「近い」関係が表現できなくなってしまうのである。そこで、多次元シソーラスでは、このような上位の分類観点から順序を排除し、素性として扱う。（図1参照）

3. 語の分類観点

分類観点は、語によって、その数も内容も異なる。種々の観点をまとめずに羅列するだけでは語と語の関係を得るために、すべての観点を網羅的に組合せなければならぬ。そこで、観点をグループごとにまとめておくことが望ましい。観点をまとめる際には、そのグループ内での観定の比較で語と語の関係を簡単に表現できるようにする必要がある。そのため、それぞれが互いに排他的であるような以下の分類観点グループを設定する。

- 上位/下位 isa
- 静的属性 sa (static attribute)
- 動的属性 da (dynamic attribute)
- 利用・用途 use
- 全体/部分 hasa

hasa は、上位概念の性質が下位概念に伝搬しない点で他と異なる。hasa 以外の関係は、上位概念が何らか

の形で、下位概念に継承される。sa、daは、それぞれ、そのものが持つ静的、動的な性質を表し、useは、人間が何に利用するか、そのものの用途や、目的である。これらは、上位概念から下位概念に新たに加わる制約となる。一方、isaはそれとは性質が異なる。isaの観点から分類するものは互いに排他的であり、その分類観点は、sa、da、use、hasaが複雑に絡み合っており、上記のような制約を簡単に設定しにくい。

また、上記の基本的な分類観点グループに分割できない観点も存在する。それらがグループとしてまとめれば、新しい分類観点グループを設け、拡張していく。

4. 構成法

図2に多次元シソーラスの例を示す。前述の分類観点グループから語を分類する。ここで、use(漁)のように、乗りもの全般ではなく、「船」特有の観点は、その下位で記述する。さらに、「漁船」のようにそれを表す語が存在する場合は、これを表記する。また、それぞれの語と分類観点の間にポインタを張ることにより、語の定義が明確になる。

具体的には、図2のようなシソーラスを既存の上位/下位関係シソーラスをもとに半自動生成する。それぞれの語の観点は、漢字シソーラス¹⁴⁾を利用して抽出する。また、その他の既存の動詞、形容詞の例文データを利用することも考えている。

5. 有効性

多次元シソーラスによって、前述の「鳥」と「飛行機」の「飛ぶ」という共通点は次のように得ることができる。以下のように定義されているとする。(図2参照)

- 動物 = 生物 {da(動く), ...}
- 鳥 = 動物 {da(動く(空中)), use(食物, ペット),
 hasa(翼, くちばし), ...}
- 乗りもの = 人工物 {da(動く), use(移動), ...}
- 飛行機 = 乗りもの {da(動く(空中)), hasa(翼), ...}

ここで、da、useなどをそれぞれ比較する。useは、大きく異なるが、daを比較するとどちらも「動く(空中)」つまり、「飛ぶ」という共通点を得ることができる。その他、翼を持つという共通点も得ることができる。このように、分類観点別の関係を表現できる。

また、多次元シソーラスでは、二項間の関係を明確に記述する。その差異を弁別特性と考え、これを用いて語の明確な定義付けを行なうことができる。これにより、

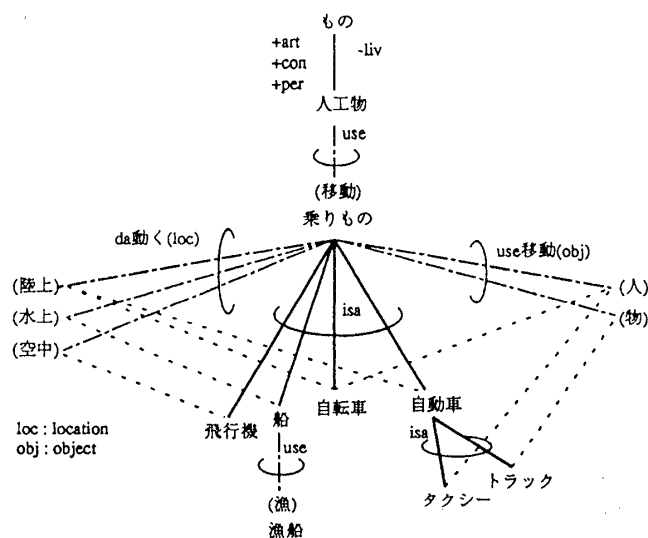


図2: 多次元シソーラス例

名詞の語義記述も可能である。

6. おわりに

語を種々の観点から分類した名詞の多次元シソーラスを提案し、その有効性を述べた。今後は、拡張性のあるデータ構造を検討し、既存のシソーラス¹⁴⁾を用いて具体的に語(概念)の観点を抽出した上で多次元シソーラスを半自動生成する予定である。

謝辞

名詞意味属性体系データ(名詞シソーラス)を提供して下さったN T T情報通信網研究所・知識処理研究部の池原悟氏に深謝する。

参考文献

[1] 池原、宮崎、横尾：日英機械翻訳のための意味解析用の知識とその分解能，情報処理学会論文誌，vol.34，no.8，pp1692-1704(1993)

[2] 川村、宮崎：既存のシソーラスを利用した漢字シソーラスの半自動生成法，信学技報，NLC93-59，pp37-44(1993)