3H－8

既存シソーラスを利用した多次元シソーラスの半自動生成法

片桐康裕　　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。しかし単語間の関係には種々の観点が存在する。そこで、観点によって変化する語と語の関係（近さ）を表現するために、複数の観点から語を分類する多次元シソーラス[1][2]が提案されている。本稿では、既存シソーラスを用いた多次元シソーラスの半自動生成法について述べる。

2　多次元シソーラスの構成

多次元シソーラスでは、名詞の概念を「実体」と実体が存在したり、生起する「場」「時」に分類している。また、実体では語と語の関係（共通点）を簡単に表現できるように、それぞれが互いに排他的である以下の分類観点グループを設定している。

上位／下位 isa
全体／部分 hasa
動的属性 da(dynamic attribute)
静的属性 sa(static attribute)
利用・用途 use
構成要素 com(component)
原因 cau(cause)
所属主 own(owner)

「場」や「時」は、実体が存在する場や、変化する時のように、実体と相互に関係しあっている場合が多い。このほかに、語全体に共通する分類観点として「位相」が用意されている。

A Semi-Automatic Construction of Multi-Demensional Thesaurus from Existing Thesaurus
Yasuhiro Katagiri, Masahiro Miyazaki
Niigata University

3　生成法

多次元シソーラスの構築のベースには、名詞意味属性体系データ（名詞シソーラス）[3]と角川類語新辞典[4]の対応をとることによって細分化されたデータを利用する。観点の抽出には一文字漢字に着目する。[5]

3.1　分類観点グループへの割り当て

語がある観点によって分類されている時には、先に挙げた多次元シソーラスのいずれかの分類観点グループに相当すると考えられる。以下に、どの分類観点グループに割り当てるかのルールについて述べる。また、漢字情報を得るために一文字漢字を意味分類した漢字シソーラス[6]を利用する。

上位/下位関係（ｉｓａ）

基本的には、他の分類観点グループに分類することができない語をｉｓａに分類するが、主に種を表わす語に対して用いる。例えば、角川類語新辞典でノードの語義が「～のいろいろ」と書いてある場合、種を表している語が多く含まれる。

全体/部分関係（ｈａｓａ）

名詞意味属性体系データでは、ｈａｓａで分類している箇所には、その旨の印が付けられているので、それをそのまま用いることにする。例えば、＜山＞を＜山（全体）＞と＜山（部分）＞に分類している。多次元シソーラスではこれらを一つにまとめ、＜山＞のｈａｓａとして＜山（部分）＞のデータを下位にぶら下げる。しかし、「車」における「窓」などすべてをカバーはできない。

動的属性（ｄａ）・静的属性（ｓａ）

ｄａは、「流氷」のように、基本的には主名詞が動詞によって修飾（分類）されている場合に用い

る。主名詞は具体物に限らない。ｓａは、「小川」のように主名詞が形容詞によって修飾されてる場合に用いるが、「性別」のような名詞で表される分類観点も含む。

利用・用途（ｕｓｅ）

ｕｓｅの形態は、「商船」「漁船」などのように、主名詞が「もの」であり、それを「人間活動」や「事象」の語によって修飾するものが非常に多く存在する。また、「花瓶」「薬草」なども目的物をとるｕｓｅと考えられるので、修飾する語は他に「生物」「人工物」がある。

そのほかに「動物」をペットとして飼うなど、ｕｓｅは字面には現れないことが多く、自動抽出するのは困難なものが多い。

構成要素（ｃｏｍ）

主名詞としては「石器」のように「もの」が多くみられるが、「氷山」「少年団」などのようにその他に「場」や「組織」も主名詞となりうる。修飾する語は「具体物（生物、自然物、人工物）」「人」が考えられる。

ｕｓｅとｃｏｍはどちらも「もの」を「もの」で修飾する場合があり、どちらに分類するかを自動で判断するのは難しい場合が多い。

原因（ｃａｕ）

「事故死」や「病死」などのように主名詞に「事（人間活動、事象、自然現象）」の語をとり、また、修飾する語にも「事」がくると考えられる。この典型的な原因のほかに、「足音」「水力」などもｃａｕに含まれるが、これらを自動で抽出することは難しい。

所属主（ｏｗｎ）

主名詞が「人」「組織」「機関」によって修飾されている場合、分類観点グループとしてｏｗｎを適用する。例えば、「県道」「国宝」などがこれに相当する。人工物に多くみられるが、「校則」「私財」など幅広い範囲で用いられる。

場（ｌｏｃ）・時（ｔｉｍｅ）

主名詞が「場」によって分類される時にｌｏｃを用いる。例えば、「山風」「海風」のように『～の起こる場』といったｄａとともに用いられる。

ｔｉｍｅは主名詞が「時」によって分類されている場合に用いる。「時」も「場」と同様に、ｄａを伴って分類する。

3.2 観点の付与

主名詞を修飾する分類観点が、字面として記述されている場合にはその語を、または、その語が収録されているノード名を分類観点名とする。しかし、「場」や「時」などによって分類する時には、ｄａなど複数の観点が絡み合っていることがある。このような場合には、ローカルなルールを設定することによって観点を付与する。例えば「生物」が「場」によって分類されている場合は、ｄａ（生息（ｌｏｃ））とする。

4 おわりに

既存シソーラスを利用し、主名詞を修飾する語を分類観点グループへ割り当てることにより多次元シソーラスを半自動生成する方法について述べた。今後は、ルールの設定をし、多次元シソーラスのサブセットを構築していく予定である。

最後に、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ＮＴＴコミュニケーション科学研究所の池原悟氏に深謝する。

参考文献

[1] 川村、宮崎：語を種々の観点から分類した多次元シソーラス、第４８回情報処理学会全国大会、No.3Q-2(1994)

[2] 片桐、川村、宮崎：多次元シソーラスにおける分類観点の体系化、第４９回情報処理学会全国大会、No.1R-6(1995)

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

[4] 大野、浜西：角川類語新辞典、角川書店、(1981)

[5] 片桐、宮崎：シソーラスの多次元化のための観点の半自動抽出法, 第 49 回情報処理学会全国大会,No.3G-9(1994)

[6] 川村、宮崎:既存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)