

多次元シソーラスにおける分類観点の体系化

1R-6

片桐 康裕 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

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⁾が提案されている。分類観点は、語によってその数も内容も異なるため、グループ化して扱い易くする必要がある。本稿では、多次元シソーラスが用意する分類観点グループについて述べる。

2. 分類観点グループ

名詞の概念を「実体」と実体が存在したり、生起する「場」「時」に分類する。「実体」を分類する観点と、「場」「時」を分類する観点は大きく異なっている。以下に「実体」「場」「時」についてその体系を述べる。これらは、「実体」が存在する「場」や、変化する「時」のように相互に関連し、互いに分類観点となっている場合もある。また、語全体に共通する分類観点として設定した「位相」について述べる。

2.1 実体の分類観点

実体を分類する複数の観点を、語と語の関係が簡単に求められるようにグループ化する。同じグループ内の観点を比較により、単語間の近さを求めることができる。多次元シソーラスでは、以下の分類観点グループを設定する。上位5つが^[1]で提案した基本の分類観点グループである。

- 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
- 構成要素 com(component)
- 原因 cau(cause)
- 所属主 own(owner)

分類観点を複数の観点から分類し、分類に順序が介入しないという、多次元シソーラスの基本概念に基づいて体系化する。多次元シソーラスは名詞について構成して

いるが、分類観点グループは名詞に限らず形容詞、動詞も同時に体系付けている点異なる。以下に、新しく追加した com、cau、own について述べる。

構成要素 com

この分類観点は人工物に多く見られる観点であり、そのものが何で構成されているか、またはその成分、材料などを表す。例えば、「鉄器」「銅器」のような分類である。この観点で分類された場合、「鉄器」「銅器」は「鉄」「銅」の性質の一部を継承する。

原因 cau

この分類観点は、それを引き起こすもとになっているものを表す。例えば、「病死」「事故死」という分類の観点である。他に「物音」「水力」などを「物」によって引き起こされる「音」、「水」によって引き起こされる「力」ととらえた場合の分類観点も含まれる。

所属主 own

この分類観点は所属主、所有者による分類であり、特に人工物に多く見られる。「私財」「私案」のような個人による分類だけではなく、「国宝」や「校則」のように組織による分類も含む。

2.2 場 loc(location)

「場」には、絶対/相対、つまり、特定の場を表すか、場と場の関係を表すかという分類観点と具体的場/抽象的場という大きな分類観点がある。これらの分類には順序性はなく、分類観点は並列に扱われる。具体的場には「実体」が存在することによって提供される場が含まれる(図1)。「場」という分類観点は「実体」の分類観点になる場合もある。特に、da とともに用いられ、自然現象の分類観点となっている場合が多い。例えば、「飛行機」が「移動する場」や、「風」が「吹く場」などである。

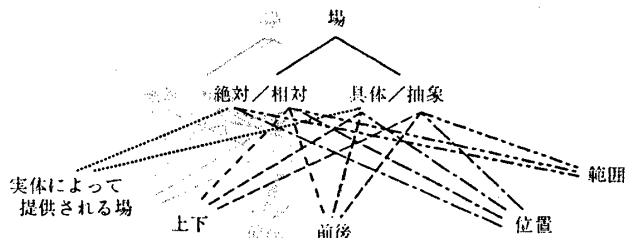


図1 場の分類例

2.3 時 time

「時」にも、絶対／相対という分類観点がある。また、時点／期間という分類観点もある(図2)。「時」も「実体」の分類観点となる場合があり、「場」と同様、daとともに用いられ、自然現象に多く見られる。例えば、「風」が「吹く時」のように分類する。

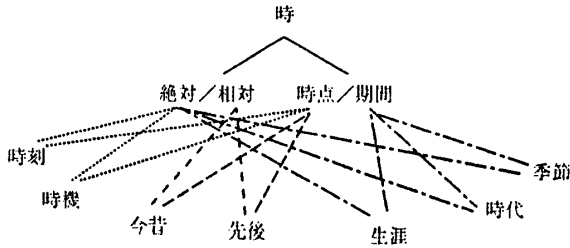


図2 時の分類例

2.4 位相 ph(phase)

この分類観点はすべての概念に共通の観点であり、その語を使用する人物や、場面を表す。例えば、「春風」は日常語、「春風」は文章語、「野分」は雅語のように分類される。

3. 分類観点グループの評価

「鳥」と「飛行機」は、既存のシソーラスでは“遠い”関係にあるが、「da 飛ぶ」という観点から見れば“近い”関係にある(図3)。多次元シソーラスでは、このような分類観点別の語と語の近さを求めることができる。分類観点グループは、特定の概念に限らず、名詞全体を通して比較的多い分類観点をまとめたものである。分類観点は、名詞の二字漢字熟語について、後方の漢字を主名詞、前方の漢字を主名詞の観点と考えて抽出し、グループに分類する。

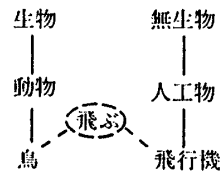


図3 観点別の語の近さ

グループに分類されない観点は、二字漢字熟語全体の一割程度である(ただし、あて字など明らかに分類観点とならない場合を除く)。それらはある概念に特有な分類観点である場合が多く、上述したようなグループと同様に扱うのは適当ではない。「ある概念」に特有の分類観点は、他の概念には現れない。そのため、他の概念との近さを求める場合には影響がない。例えば、「人間」の分類は上述の分類観点以外にも「職業」や「関係(親族関係、仲間)」による分類などがある。これらの分類観点は、「人間」という共通の概念の中で、さらに近い語の判定を行なう際には重要であるが、「人間」に限った観点であるために、他の語との近さを求める場合には影

響がない。しかし、これらの分類観点は、語義を記述する場合には重要である。そこで、語(概念)に特有な分類観点はグループ化を行わずに、その概念のもとだけで記述する。

多次元シソーラスを用いて、既存のシソーラスでは表すことのできない語と語の近さを表現するためには、上述した分類観点グループではほぼ十分であると考えられる。

4. 語と語の近さの評価例

分類観点をグループ化することで、語と語の近さの判定が可能となる。さらに、各グループごとに体系を整理すると近さの程度をより詳しく求めることができる。例えば、saを、視覚／聴覚／味覚／嗅覚／触覚という観点で分類する。視覚でとらえられる属性の中には、「色」や「形」がある。「色」の属性値として、「赤」や「白」などがある(図4)。この場合に、「赤」という属性値、「色」という属性、「視覚的属性(観点)」を持つ点のどの段階で共通なのかという情報により、近さの程度を求めることが可能となる。分類観点の近さが語と語の近さに反映する。このように、分類観点グループの体系を整理し、語と語の近さの評価として用いることは有用である。

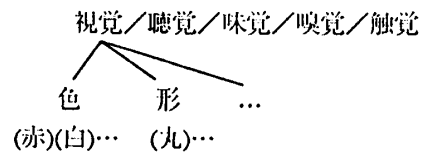


図4 saの体系(一部)

5. おわりに

多次元シソーラスにおける分類観点グループを挙げ、それぞれの定義を述べた。また、用意した分類観点グループの評価を行ない、これらを利用した語と語の近さの評価例を示した。現在は、分類観点や素性を記号化し、扱い易くすることを考察中である。また、上に述べた分類観点に基づいて語を半自動で分類し、多次元シソーラスを構成していく予定である。

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.3QG-2(1994)
 [2] 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.3QG-9(1994)