

格関係に基づく概念体系の検証と
構成語解析による分類前処理

5B-2

奥村登貴子 井原浩子

(株)日本電子化辞書研究所

0. はじめに

前回定義した概念体系 (ref.1参照) のもつ格関係について、基本和語 (第3節参照) の語義、500文の例文解析結果および既存シソーラス語彙からの射像によって集めた語彙にもとずいて分析した。語彙を基本体系と補助体系の組み合わせで表現することで格関係を整理できることが確認できた。またこの結果を構成語の分析によって分類前処理の可能性について検討した。

1. 概念体系と格関係

コト部分は以下の必須格で定義される。但し自由格としての時空間格、因果格、感受者は最上位体系で定義される。以下のリストは体系のかなり上位レベルであり、@以降は下位概念を示す。[]内は体系のもつコト属性である (体系内での「属性」とは限らない)。コト属性 [評価、可能] は最上位体系で与える。[...] としてあるものは下位体系で定義される。

- ・基本体系
- ・存在： @存在, 生死, 結成
対象具体物 [(対象のもつ属性)]
- ・情報関係： @伝達, 思考, 感情, 見聞
対象内容/テーマ、情報位置 (思考者)
[詳細、明確、難易、秩序]
- ・空間関係： @方向, 移動, 接触, 食べる, 距離..
対象具体物、関係位置 (場所/具体物)
[形、方向、距離、秩序]
- ・所有関係： @授受, 所有, 貧富
対象物、所有位置 (所有者) []
- ・抽象関係： @支配, 親類, 友好, 参加
対象物事、相手物事 [秩序]
- ・属性： @形, 能力, 階級, 詳細, 秩序, 評価..
対象物事 [...]
- ・補助体系
- ・変化性： @移動性, active, 継続, 進行, 完結
対象物事 [時間的属性、勢力、(速度)]
- ・意志性： (有意志体)、受益者、対象物事
[評価・感情・態度、能力]
- ・程度性： @比較, 割合, 範囲, 種類, 数量
対象物事、(格1)、格2 [程度]
- ・同伴性、直接性：
対象物事、(格1)、格2 [(格1、2のとりうる関係)]
- ・相互性、前提性、否定： @直接, 相互, BACK
対象物事、(格1、格2) []
- ・補助体系の格関係への影響：
変化性 コト属性格に source, goal, active を派生する
(ex. 移動性空間属性値における形 = 移動の形、非移動性

の場合 = 結果状態の形)。その他の格への影響についてまとめる。

- 1) 移動性：「情報、空間、所有」の位置格に source, goal, active を派生する。移動性と基本体系の組合せは体系内に別建てした。
- 2) active/start：変化性については変化する物事、関係については対象格に source, goal を派生する。属性については属性値を含むコト属性にのみ派生する。
- 3) stop：activeの下位であるが、コト属性格に否定と同じ影響を与える (否定の start)。

否定 ここで「否定」と組み合わせで扱ったものは「関係、接触、存在、変化性、思考、情報移動」についてであった。その他のものは極性/程度性として扱うか、体系に別建てしたか、語彙がなかった。

ある体系の否定にとったものは、その体系でのコト属性がつかなくなる。その他の格と、上位概念や組合わさった補助体系のもつコト属性はそのままつく。

程度性など 基本体系に組合わさった程度の意味は次の3通りに分けられる。

- ①基本体系がコト属性として持つ程度。
- ②基本体系がコト属性として持つ程度性属性に関する程度。(ex. すぐく移動する->すぐい距離移動する)
- ③基本体系がもつ格値の数量。(ex. すごい人数/場所数/回数移動する)

どの場合であるかは、程度性のもつ格1と基本体系がもつ格のshiftの記述で示される (ref.2参照)。格2は比較/割合の場合に派生して格1の値域を継承する。直接性 (-を介して)、同伴性 (-と共に) も同様である。相互性 (互いに-)、前提性 (-返す) の格1、2は移動性概念の位置 source, goal、関係概念の対象と相手に対応する。

意志性 有意志体と対象物事のもつ格との対応がある場合それを記述する。

・コト属性、因果格の限定：

1) 体系がもつコト属性：
例として空間関係について述べる。 ” | ” 以降は + active/start/stop の場合にさらにつくもの、 ” & ” 以降は体系以下のレベルで +- に依存してつくものを示す。

@移動	方向、距離、移動形、音
@方向	方向、距離
@接触	方向 音 & 結果形、大きさ、秩序 = + つなぐ = - 洗う

@端	方向、距離
@並ぶ	方向、位置、結果形、大きさ、秩序
@切断	方向、位置、結果形、大きさ、秩序
@刺す	方向、位置、形
@包囲	方向、距離 (半径)、密度、角度

@中外 距離、密度
&結果形=+入れる、=-汲む

@覆う 方向、結果形

なお変化性と組合わさった時コト属性が派生しえる場合は因果でなく属性にとっている。動きや動作で変化する性質はほぼ必ずかかるとわかる。

コト属性に限らず格は語彙特有の限定を持つ (ex. 乗り物に乗る、座布団を敷く)。特に構成語はもとの語のもつ限定を保存しつつ組織的に限定を生成する。これを利用してどの程度細分類の前処理ができるかの検討結果を第3節で述べる。

2. 派生する格

第1節で定義した以外の格の派生現象は、構成語による場合と、もともとの語彙のもつ性質による場合があるが、いずれも体系においては組合せによって表す (ref. 2参照)。

派生現象は以下のように分類できた。(体系@.とその必須格については第1節のリスト参照)。

・体系のもつ目的概念のもつ格が派生する場合:

ex. 体系@接触に属する語彙「拍手」は、体系@伝達を目的にもつため、@伝達のもつ必須格である「情報位置goal」を必須格にもつ (ex. 演者に拍手)。

目的が@伝達である@動き(旗を振る、狼煙を上げる)、@接触(拍手、捧げ銃)、@勢力(にらみがきく)、@形(お辞儀、表情)。当然人間の行為である。

目的が@接触である@移動(追う)、@形(腕をまくる、座る)

目的が@移動である@接触(蹴る、押しやる)、@所有(捨てる)

目的が上位コトである@形(身構える)

・その他の因果格による派生:

@移動(食い逃げする、曳航する)。構成語に多い。

・時間化:

ex 学校の「帰り」に寄る。

いまのところ@移動のみからしか出ていない。

・本来の格の値のpart-of、elementによる派生格:

ex @接触=拍手 → 人が拍手

@存在(禿げる) @接触、@開閉(まばたく)、@形(怪我)、@異常(病気)など、動物の部分が対象になっている場合が多い。

materialもmade-ofなら外観属性を、made-fromなら味臭などを継承する場合がある(ex. お菓子の家の味)が、コトの方だけでは決められなかった。

3. 分類前処理の実験

以上の現象を語彙に個別に付与してゆくのは大変なので、漢字の表意性を他のキーとうまく組み合わせて自動的に分類の前処理を行えないか検討したい。

なおここで使用できる語彙の情報は以下のとおりである。

- ・モノ概念多語義語の語義=上位語
- ・コト概念多語義語の語義=限定句+見出し語または上位語。ex あげる1=芸者-を-よぶ。
- ・品詞、活用、表層格情報。

まず漢字の表意性をとらえるキーファイルを作成した。漢字一字とそれに続くかな文字からなる用言のうち、漢字毎に最多語義のものを選び基本和語と考えた。これで得られた漢字約1300ではほぼすべての漢語の少なくとも一方の漢字を網羅しており、のこりはほとんどがモノ概念相当漢字であった(つまり見出し語解析は殆どの構成語について行える)。このうち漢字数約300にあた

る基本和語約1000語義を解析し、かつ体系を付与した。まず、この漢字2字からつくられる漢語約1000について実験した。

・見出し語のみについての解析:

体系の中/大2レベルの分類と、それに多少の格関係を付与したデータを用いて、以下のグループに分けた。

①2つの漢字のもつ中レベル体系が重複するもの: 約1/3がこれに属した。うちほとんどがuniqueな体系を与えた。約4/5がどれかの語義について正解を与えるようにキーファイルを修正するのは容易であった。(10漢字について各1-2体系の修正)

②2つの漢字がある格関係で結ばれる得るもの: 約1/3が属し、正解率も高いが、補助体系との格関係が多く、基本体系の推定範囲が単一の漢字よりあまりせばまらない。

③同一の大レベル体系に属するもの: 約1/10が属し、正解率が高い。

④②と③の重なり部分: 約1/5が属した。これの解析は解釈が難しいので、さらに語義解析と組合わせて考える。

⑤どれもでないもの: 約1/10。これについては、自由格性が高い為にわざとデータファイルから外しておいた格関係をこの段階で適用するとほとんどが正解を与えることがわかった。

いずれの場合も、組織的な間違いについて、補助体系との組合せと品詞から体系毎の表層格パターンに組織性がある場合はそれを使って排除できた。

・語義の分析:

完全なidiomをはじく為と、上位でしか決まらなかった体系を限定するためと、多語義の場合にリンク先を分けるために語義も検討する必要がある。上位語のもつ漢字情報は見出し語より増えない場合が半分であった。そこで限定句の意味情報を与える為に、モノ概念(ここでは名詞からコトを表す語義語尾のものを省いただけのもの)の方の体系化も検討した。モノ概念については上位語がかなり集束していることがわかった(8000語義中5000語義が多くとも5000上位語に集束)。

4. まとめと課題

語彙の組合せ体系としての表現が格関係の整理に有効であることが実際のデータで確認できた。問題点としては、現在「否定/変化性」ととっているものと外からつく否定/アスペクトの関係がまだ分析しきれていない。

大量作業への見通しについて、少なくともこの機械処理にほとんどすべての語義見出し語がかかることがわかった。結果の成績は見出し語+上位語までの範囲でも半分はかなりの精度で使えることがわかった。今後限定句についても解析し、漢字解析はむしろ補助的に見出し語/語義上位語について行うことにする。解析の視点について今回は「中レベル体系への予備分類が行えるか。」について行ったが、今後は以下の点について検討したい。

①体系レベルに分類されたものについて、漢字などをキーに派生格や格値限定についての細分類が行えるか。

②構成語(および語義)の分析から因果性の格をどのくらい収集できるか。

reference)1)井原: "概念の体系化に関する考察"

第36回情処全大

2)奥村: "概念辞書における知識表現" 同上