

用言の意味マーカと名詞の意味素性の関係のギャップを part-of 関係および名詞の意味情報を用いて解決する手法

関洋平 飯島正 原田賢一

慶應義塾大学理工学部

e-mail: yohei@hara.cs.keio.ac.jp

1 はじめに

用言と名詞句の間に成立する関係を記述する方法の一つに、意味マーカを用いた選択制限 (Selectional Restriction)[1]がある。より詳しくは、ある用言に対して格構造を設定し、表層格の組に対応させて意味マーカおよび用言全体の意味を与える。同様に、名詞句に対しても意味素を与えて、用言の意味マーカに名詞句の意味素との関係を調べることで、多義的な用言を名詞句の意味に応じて取り扱うことができるようになる。この関係の調べ方は、実際には、is-a 関係に基づく包摂関係のチェックが用いられることが多い。

しかし、用言によっては、名詞句の属性(名詞句とは attribute-of 関係にある)または名詞句の部分(名詞句とは part-of 関係にある)を意味的には取りながら、名詞句を補語として取る場合がある。前者の例としては「バラは美しい」(属性は色・形など)などの形容詞、後者の例としては、「彼が僕を殴った」(部分は僕の顔・頭など)などの体の一部分を対象に取るが、補語としては人間が来るような動詞がある。attribute-of 関係は実体に対する抽象的属性なのに対して、part-of 関係は実体の構成要素(所有物も含む)であり、具体的概念である。オブジェクト指向風に言うと、属性は状態変数であり、部分は部品オブジェクトに相当する。

上記の例を見てもわかるように、そのような用言が現

われる文一つだけを見ていては、用言と名詞句の関係が決まらない場合がある。また、「私は美しい」に対する「私の体の上半身の胸の形」など、部分は部分の部分や部分の属性などを次々に取ることができ、どの部分まで切れば終わりといった原子的な構成単位が存在しない。このような問題に対しては、「必要になったら情報を絞り込むが、必要にならない場合はそのままにしておく」といった基本姿勢を取ることが大切だと思われる。現実には、オブジェクト指向設計の分野において、構造的粒度をどこまで細かくするかは問題に依存して決まることが指摘されている[2]。

また、名詞句と用言の制約から意味の値を与えて、更に次の文で与えられる制約条件を満足するように値を絞り込んでいく過程は、制約伝播の手法[3]の一種と考えることができる。本研究では、用言と名詞句については、制約が強いと思われる順序で評価していき、一方の制約の評価が終わった段階で、次の制約を評価するという遅延評価に基づいたシンプルな制約解消を行う。具体的な方法は2章で説明する。

本研究では以上の指針に則り、以下の章でこの問題に対する一つのアプローチを示す。本研究の利点として、用言と名詞句の間の意味情報を詳細化することにより、省略補充や指示対象の同定などの文脈解析への情報の提供が期待できる。

2 本研究で提案する枠組

本研究では、意味を表現するためのデータ構造として、基本的に素性構造(feature structure)を用いる。ただし、part-of と attribute-of 関係を表すために、素性値として二項関数、三項関数を取ることにする。

An Approach for bridging the semantical gap between semantic marker of verb and semantic feature of noun with part-of relation and semantic attributes of noun.

Yohei SEKI, Tadashi IJIMA and Ken'ichi HARADA
Faculty of Science and Technology, Keio University
3-14-1 Hiyoshi, Kouhoku-ku, Yokohama 223, JAPAN

2.1 基本となる枠組

1. 部分表現の抽出方法

まず、部分表現の抽出方法について述べる。

(a) 「太郎が次郎を殴った」

(a) のような文章があった場合を考える。「殴る」という動詞が一回的に用いられた場合¹。「殴る」対象として、ただ一つの < 面 > を取る。これに対して、「殴る」補語として、「次郎」という < 構造 > を持つ具体物を指示する固有名 > が来ている。(用言の辞書項目は3章で詳しく説明する)

「次郎」の < 部分 > には < 面 > を性質として持つものが存在する。よって、以下の意味表現を抽出する。

(a) 「殴る [対象:att_of_p²(面, part_of(□, 次郎))]

なお、□は変数であり、「次郎」(意味素: 人間)と part-of 関係を取り、< 面 > の属性を持つ要素が、集約階層の探索(2.2節で後述)および単一化操作を用いることで代入される。

この□は、前の文か次の文に関連する名詞句が出現した時点で評価される³。例えば、1の文の次に(b)のような文が現われたと仮定する。

(b) 「顔から鼻血が出た」

すると、意味構造は以下のように詳細化される。

(b) 「殴る [対象:att_of_p(面, part_of(part_of(□, 顔), 次郎))]

このことを判断するためには、「鼻血が出る」対象と「殴る」対象の間に因果関係があることがわからなくてはならないが、このような辞書を作るとはむずかしい。本研究では、□が作り出した文に隣接した文の用言が、□を指定している用言と同じ意味マーカを指定している場合に、その名詞句および、同じ文中でその名詞句と部分-全体関係にある名詞句をチェックするという方法を用いる⁴。

¹ 「ぼこぼこに殴る」のような複数回数の事態ではないということ

² 明示された part-of 関係を取る要素のうち、最も細かい粒度の要素の attribute の意味

³ 探す範囲を特定することはむずかしい。現在のところは、判断する理由も無いので、隣接した文の名詞句だけを考えることにする。

⁴ 簡略化しているが、表現「…を殴る」の対象「…」と「…から鼻血が出た」の意味マーカが一致して、かつ「鼻」を形態素として認定

概念的粒度の限界は「面」をその部分的要素が持てなくなる場合であり、その場合には□が無くなるが、「顔」は「鼻」を部分に持ち、「鼻」は < 面 > を性質として持っているため、変数は維持され、同じ文中の「鼻」によって、最終的には以下のような意味構造が抽出される。

「殴る [対象:att_of_p(面, part_of(part_of(part_of(□, 鼻), 顔), 次郎))]

2. 属性表現の抽出方法

続いて、属性表現の抽出方法について説明する。属性は、用言が対象として指定している意味マーカに包摂される要素を属性値として持つような、補語の名詞句の属性を選ぶ。その属性と属性値を意味として抽出するため、att-of 関数は三項関数になる。

(a) 「太郎はCDを聞いた」

(b) 「太郎はカーペンターズを聞いた」

上記の文を考えると、まず「聞く」という用言は「音」を対象として指定する。実際の補語は、その音に is-a 関係で包摂されるような要素を属性値として持つ属性を持っている。すなわち、「CD」は < 音源 > という属性が < 音楽 > という属性値を持っており、「カーペンターズ」は < 歌手 > という属性が < 歌声 > という属性値を持っている。 < 音楽 > と < 歌声 > は < 音 > に包摂される。

以上の関係から、「聞く」に対しては、以下のような意味表現が与えられる。

(a) 「聞く [対象:att_of(音楽, 音源, CD)]」

(b) 「聞く [対象:att_of(歌声, 歌手, カーペンターズ)]」

3. 部分表現と属性表現の関係

属性は、名詞に対して与えられるものを、用言の制約から導き出すため、「属性の属性」のような入れ子構造は、名詞の属性値にあらかじめは属性構造が埋め込まれている場合に限り、考えることにする。また、「部分の属性」を取るような動詞はあるが、「属性の部分」は無いものと思われる。よって、ど

し、「顔」との部分関係を判定するといったプロセスはかなり複雑である

ちらを先に調べるかということ言えば、用言と名詞句の関係から、まず「部分」関係を取るかどうかを調べて、次に名詞句の「属性」関係を取るかどうかを調べる。ただし、3章で示すように、それぞれの特徴を示す動詞の種類・分類は異なる。

2.2 意味素に基づく変数の制約

この節では part-of 関係と attribute-of 関係についてより詳細に論じることにする。

2.1 節で述べたように、本研究では、「次郎」の part-of、もしくは「顔」の part-of 要素として、意味素を指定することによって、部分要素を制限している⁵。part-of 関係は、先の例では、顔が出てこなくて、いきなり「鼻血が出た」と来る場合もある。この場合、part-of 関係は、「顔」を介して部分の部分という形で繋がっている。

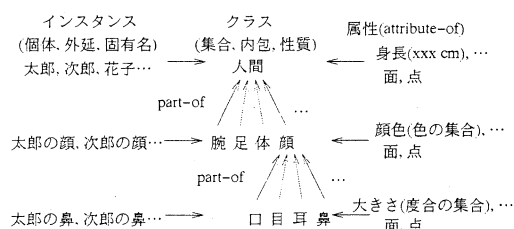


図 1: クラス間の集約階層ネットワーク

このために、「集約階層ネットワーク」のような構造をあらかじめ用意しておく。ネットワークは抽象的なクラスレベルで定義される。実際の文章には、インスタンスが出てくるので、対応づけが必要となる。属性関係については、各クラスに対して保持している属性と属性値(の集合)を指定しておく。属性値は必ずしも集合である必要は無く、2.1 節の例のように「音楽」と固定されていても良い。また、3章で詳しく論じるが、「点」「面」「立体」などの属性についても記述しておく。

先ほどの「鼻血」の例に対応するためには、(探索条件に該当した)名詞句が出て来る度に、集約階層を根に向けて辿り、最初の文が要請している意味マーカと交わるかどうか調べるといった操作が必要になる。交差した場合は、途中の part-of 関係も含めて意味表現に反映することになる。

⁵このことは、本研究が実質的には型付き単一化文法 (typed unification-based grammar) に基づいていることを示している。

3 日本語における対象

3.1 補語の部分または全体的要素を対象として取る用言の特徴

日本語の意志的動作動詞は、「発見する」⁶のように抽象的名称を意味の対象とするものと、「叩く」⁷のように、接触する具体的対象を取る用言に分類できる。ただし、意味的に具体物を取りながら、実際には「名前」(抽象名でも固有名でも良い)を取る場合もある。

また、具体的対象を取るものは、「点」⁸を取るものと、「面」⁹「立体」¹⁰を取るものがある。もちろん、「点」といっても完全に一つの点になるわけではなく、構造的粒度の問題が関係する。

1. 太郎君が次郎君を殴る (補語: 次郎君 → 対象: 顔 < の面 >)
2. タイソンがホリフィールドにかみつく (補語: ホリフィールド → 対象: 耳 < の面 >)
3. 車にガソリンを入れる (補語: 車 → 対象: ガソリンタンク)

この部分的関係には以下に示す包含関係がある。

点 ⊆ 面 ⊆ 立体 (構成要素) ⊆ 立体 (構造をもつ全体)

構造をもつ立体と構成要素の部分は入れ子になることはある。著者等は、この包含関係のうち、どの部分まで辿るべきかは、述語の種類によって区別できる場合もあると考えている。ただし、ここで注意すべき点は、点や面が、立体に対して構成要素となることもあれば、属性となることもあるというところである。

また、具体的対象を取るが、「点」「面」「立体」のいずれを取るかがとくに指定されていない用言¹¹の場合、関係は名詞の属性で決まる。とくに、慣用句的なものは、対象と関連の深い名詞のうち目立つ要素を補語に取る。この種の要素に対しては「頭」の部分ではなく、「頭」の属性として辞書登録しておく。この話題は提喻 (synecdoche)[4] との関連が興味深い。

1. 頭を分ける。頭を切る (補語: 頭 → 対象: 髪)
2. 頭が良い (補語: 頭 → 対象: 人間の知力)

⁶他に、「燃やす」「殺す」など

⁷他に、「触る」「点火する」「殴る」など

⁸「刺す」「つく」など

⁹「打つ」「蹴る」「押す」「切る」など

¹⁰「入る」「出る」「突っ込む」など

¹¹「分ける」など

3.2 補語の属性的要素を対象として取る用言の特徴

補語の属性的要素を対象として取る用言としては、以下の様なものがある。

1. 補語の性質を対象としているような形容詞¹²
 - (a) トマトが赤い(補語: トマト → 対象: 色)
 - (b) この計算機は便利だ(補語: 計算機 → 対象: 道具の性質)
2. 補語の状態を対象としているような動詞¹³
 - (a) 車が止まる(補語: 車 → 対象: 動き)
 - (b) 英語のテストに失望した(補語: テスト → 対象: 内容)
3. 人間を補語として取るが、実際には人間の持つ考え、性質を対象としているような用言
 - (a) 太郎君が次郎君に賛成する(補語: 次郎君 → 対象: 考え)
 - (b) 史子は明るい(補語: 史子 → 対象: 性格)
4. ある非接触概念に対する源泉を補語として取るが、実際には非接触概念を対象としているような動詞
 - (a) 太鼓が聞こえた(補語: 太鼓 → 対象: 音)
 - (b) 懐中電灯が太郎を照らした(補語: 懐中電灯 → 対象: 光)

4 他の関連研究との関係

部分-全体関係を記載した概念体系辞書としては、プリンストン大学で開発された WordNet において、英語の名詞の部分/全体関係についての情報が記述されている。さらに、構成要素、自立要素、メンバーの3種類に下位分類している[5]。また、IPA STaX(意味分類構成法)[6]は、is-a 関係の包摂と用言に関連したキーワード¹⁴が示されている。述語「XはYの一部分である」といった部分名称体系(meronymy)の検索もできる。

また、計算機科学の視点からの関連研究として、オブジェクト指向設計における複合オブジェクト(Composite Object)の取り扱いがある。設計の途中で、オブジェクトの構成要素をオブジェクトとして考え直すことが必要になった場合に、サブタイプ継承を参考にした集約継承[7]や置き換え[8]などの概念によって取り扱

うというアプローチがあるが、性質の継承などの点で柔軟性に欠けるという指摘がある[2]。

5 おわりに

本研究では名詞句の属性情報、および部分-全体関係を用いて、従来の is-a 関係に基づく概念階層だけでは捉え切れない用言と名詞句の関係を結びつける目的を、集約階層および制約解消という枠組に基づいて試みた。

実際の対話では、この他にも所有関係、名前づけなどに対して、話者と聴者の間で共通の認識を持っていた場合、人間は結びつけることが可能である。このような場合、語用論的な観点を考慮に入れる必要がある。

しかし、本研究の目的は、そこに繋げるまでの段階を計算機上で取り扱うといったところにあり、その目的に対して、ヒューリスティックな立場に基づく一つの見解は示せたものと思われる。

謝辞 George Lakoff の著作について御紹介して下さい。慶應義塾大学理工学部の小原京子先生に深く感謝致します。

参考文献

- [1] 長尾 真 編: 自然言語処理, 岩波講座 ソフトウェア科学 15, 岩波書店 (1996).
- [2] 飯島正: “内包的オブジェクト識別性とその応用,” 電子情報通信学会論文誌 D-I Vol.J79-D-I NO.10(1996), pp.644-655.
- [3] 淵 一博監修, 溝口文雄・古川康一・J-L.Lassez 編: 制約論理プログラミング, 共立出版 (1989).
- [4] G.レイコフ, M. ジョンソン著, 渡部昇一他訳: レトリックと人生, 大修館書店 (1986).
- [5] 松本裕治, 影山太郎, 永田昌明, 斎藤洋典, 徳永健伸: 単語と辞書, 岩波講座 言語の科学 3, 岩波書店 (1997).
- [6] 緒方典裕, 橋本三奈子: “意味分類の言語学的構成法とWWW上のシソーラス構築,” 情報処理学会研究報告 97-NL-117-7(1997), pp.45-50.
- [7] L.Liu: “Exploring semantics in aggregation hierarchies for object-oriented databases,” Proc. of the 8th ICDE'92(92CH3097-3), pp.116-125, IEEE, 1992.
- [8] G.Moerkotte and A. Zachmann: “Multiple substitutability without affecting the taxonomy,” Proc. of the EBDT'92, (Advances in Database technology - EBDT'92) LNCS-580, pp.120-135, Springer-Berlag, 1992.

¹² 属性形容詞など

¹³ 無意志動作動詞など

¹⁴ カテゴリー一致述語, カテゴリー固有述語