2I5-OS-08b-1

述語項構造シソーラスによる述語と名詞の構造化

Discussion on Semantic Structure of Nouns from the view of Predicate Arugument Structure Thesaurus

竹内 孔一　石原 靖弘　竹内 奈央[*1]
Koichi Takeuchi　Yasuhiro Ishihara　Nao Takeuchi

岡山大学大学院自然科学研究科
Graduate School of Natural Science and Technology, Okayama University

[*1]フリー言語アナリスト
Freelance Language Analyist #1

This manuscript discusses how semantic structure of nominal noun should be composed. From the view of previous work in linguistic analysis of noun phrases, natural language understanding and modeling scheme of function for design, we conclude that feature structure should be the scheme of describing noun semantics for information extraction task. Construction of noun semantic structure must be a part of IE system, then we depirct how the total IE system should be composed with constructed noun semantic structure.

1. はじめに

自然言語で書かれた文書からなにがしか必要な情報を取り出すというタスクにおいて，文中に現れる名詞の表現はどのように構造化すれば良いであろうか? 著者らは述語に関しては項構造をベースとしてシソーラス形式を提案し，述語の意味分類を行ってきた．その結果から述語には状態変化や存在，活動といった分類だけでなく,「できる」などに見られる可能性や期待，見通しなどモダリティに関する部分も含まれていることを事例とともに明らかにした[*1]．一方で人がイメージするモノに対する参照の情報は名詞句が持っていることから名詞の意味の取り扱いを明らかにすることが出来れば文書から情報を取り出すシステムの基本枠組みができあがると考えられる．

では名詞の意味構造は先行研究においてどのように扱われているであろうか? 言語学の分析，自然言語処理，自然言語理解，人工知能におけるオブジェクト記述の分野の展望した結果，意味構造を考えた際，1) 先行研究の提案する構造は全て属性に分けて名詞の意味を記述する構造に集約できることを示す．さらに我々が述語シソーラス構築における分析から，2) 属性による状態変化の導入の必要性，3) 人の評判も名詞のオブジェクトに記述する必要性があることを指摘する．

以降ではまず名詞の構造化の前に言語から情報抽出を行うシステムを構築する上で問題となる点について整理し，次に名詞の先行研究分析を行い，その結果を踏まえた名詞の意味構造のあるべき形について議論する．

2. 辞書 (先験的な知識) を構築しながら文書から情報抽出を行う際に問題となる点

本研究でのアプローチは先験的な言語知識を構築しながら自然言語から情報を抽出する方法を採用している．しかしながらこのアプローチには下記のような問題が考えられる．

a1(背景構造不要) 言語処理は言葉という記号と記号の間の処理であるから，背景構造を人手で与えるのではなく単に記号間として必要とする回答 (文字列) が示せればよい．

a2(記述の不完全さ) 人間は言語の背景にある理解世界における豊かな知識があり，ラベルなどに極度に集約した表現で構造化を行っても追いつく見込みが無い．

a3(分野依存) 静的な辞書情報だけではとけず，必ず分野依存の言語的な知識が必要

a4(背景知識の存在) 情報抽出の際，言語には表れない生活や人の行動に関する知識が関わっていると考えられる情報が必要な時がある．こうした背景知識はどう扱うか．

これらの問題は相互に関係しているが，まず **a1** について考えてみたい．**a1** はアプローチの違いであるが，人手で構築する静的な知識の有効性と構造化の有効性の 2 つの論点に分けて議論したい．まず前者であるが，静的な言語知識である WordNet が質問応答システム Watson で有効であったことから [ベイカー 11] 人手による背景知識は統計的手法などで取り込めば有効であると考えられる[*2]．ただし，この場合，単語対の類似度など簡素化した関係で有り，いわゆる言語理解 [Winograd 72] で仮定された意味構造を各単語に持たせているわけではない．よって背景知識の構築には一定の効果が予測されるが，複雑な意味構造が必要かどうかは解くべき問題に依存すると考えられる．

これに関連して **a1** と関係する **a4** の問題を考えてみたい．我々は言語処理が必要となる情報抽出の場面において，実際には背景知識がかなり必要であり，その補完として意味構造 (多重の単語対類似，上位下位他関係を含む構造) が必要であると考えられる．例えば日本語能力試験 N2[田代 11] の情報検索の問題を取り上げてみる．下記のようなお知らせのとき[*3]，「日曜日の午後 8 時頃に高熱を出した．どうすればよいか?」とい

■救急診療所
【診療項目】内科，小児科，外科
○日曜日・祝日・年末年始 = 午前 9 時〜午後 10 時
■歯科急病診療所
○日曜日・祝日・年末年始 = 午前 9 時〜午後 5 時

図 1: 情報検索の問題例

連絡先: 竹内孔一，岡山大学，岡山県岡山市北区津島中 3-1-1，koichi@cl.cs.okayama-u.ac.jp

∗1 http://vsearch.cl.cs.okayama-u.ac.jp/.

∗2 この他，NTCIR RITE2 含意認識タスクでも同様である．

∗3 ここでは簡略化した例である．

う質問に対して (選択肢から) 答える問題である．この時,「内科」と「高熱」との関係 (つまり背景知識) が分からなければこの問題を解くことはできない．もちろんこれは文書ではなく項目化されたお知らせであるが，文書であったとしても同様であると我々は考えている．つまり，認知言語学的な意味的関係 (WordNet) だけでなく，日常的な言語にまつわる知識は必須であり，どう構築するかは別として複雑な意味構造を取り込む必要がある．

さらに問題 **a2** と **a3** を考えてみよう．問題 **a2** の指摘は人間の理解世界は上記の例を出すまでもなく，名詞-述語の組み合わせの世界をかなり知っているという点である．例えば「ジャガイモを植える」と「ジャガイモを食べる」は表層は同じ「ジャガイモ」でも指しているものの状態 (つまり機能 (食べられる? 植えられる?)) はかなり異なる．つまり人間の理解レベルの知識を具現化しようとするとかなり詳細な記述が必要で，意味構造として構築することはコストの面からも曖昧生解消の面からも現実的では無いように見受けられる．

しかしながら一方で，問題 **a3** の指摘にもあるように人間でも背景知識が不足している分野の文書を読んだ場合，その文書に関する質問を受けても答えることはできない．つまり，情報抽出を行う際には逆に分野依存でよいので，詳細な意味構造が必要となる．例えば動詞の語義を例に取れば

- A 社の B 車に決めた/購入した

は述語項構造辞書構築の観点から「決める」という行為と「購入する」という行為は同じと見なすには距離があると考えられる．しかしながら，自動車購入関連の文書を集める際，上記の 2 文は「購入を決めた事例」の文書として集められた方が好ましい．つまり分野に適応して詳細な意味関係 (ここでは類義かどうかであるが) を (どう抽出するかの議論はさておき) 構築することが必要である．

上記の問題点を踏まえてまとめると，WordNet などの静的な言語知識の構築は有効である一方で，情報抽出までには，分野依存情報の獲得，背景知識の獲得が必要であると言える．よって静的な言語知識，分野依存知識，背景知識，含意認識エンジンを切り分けて構築することで，情報抽出システムの失敗があった場合，問題を切り分けて漸進的に改善が可能であると考えられる [竹内 14]．よってこのような開発枠組みの中で，名詞の構造化を考える．

3. 名詞の構造にまつわる先行研究

本稿では名詞を単に言語表現の名詞だけでなく，知識工学におけるモノの意味構造まで含めてそのモデル化についての先行研究を概観したい．まず言語表現に関する先行研究について下記に示す．

名詞句の分析

名詞句 (「X の Y」) および名詞述語文 (「A は B だ」) の分析から西山 [西山 03] は名詞の中に飽和名詞と非飽和名詞の 2 種類が存在することを指定している．非飽和名詞とはその名詞だけでは意味が理解できず，名詞が属する主体的な何かを必要とする名詞である．

- 飽和名詞: 俳優，作家，建築家，政治家，首飾り
- 非飽和名詞: 主役，著者，本場，友人，上司，蓋

この違いは「X の Y」の場合，飽和名詞ならば「の」の意味関係は文脈でしか決まらないが非飽和名詞の場合はその名詞が属する主体であることがわかる．下記に例を示す．

b1 太郎はこの芝居の俳優/北海道の俳優

b2 太郎はこの芝居の主役/?北海道の主役

この例の **b1** では「の」は様々な関係が考えられる (「北海道で有名な俳優」「北海道出身の俳優」) が，**b2** では「主役」のお芝居を指しており，それ以外の名詞が来ると解釈ができなくなるか[*4]，比喩的な意味となる．

これは意味構造の観点からは名詞にも項 [影山 11] があり，文書や文脈情報から項を埋めることで意味を完成させると考えられる．これを別の観点から見れば，ある主体の名詞に対する，属性と考えられる．つまり「お芝居」の属性として「主役」であり，属性値はその対象である．例えば上記の例文 **b** の場合

```
[お芝居
   属性：主役：太郎]
```

のようになる．言い換えればある名詞 (「芝居」) と名詞 (「太郎」) を結び付ける意味的な関係 (タイプ) とも考えられる．

高木の言語理解モデル

言語学ではないが言語に対する深い洞察から，高木らは [高木 87] 名詞節や名詞句，名詞述語文に関する表現の言い換えを集約する方法を提案している．例えば「あの車の色は赤い」は

```
車 (    の   ) 色 (        ( 赤い      ))
○=●=>◎<-*=○=●=>◎<-○=●=>◎<-[赤]
CAR  POSS COLOR  POSS HUE    EQ
```

と記述し,「赤い色をした車」は

```
車 (  した を 色 (        (  赤い     )))
○=●=>◎ <- ○=●=>◎<-○=●=>◎<-[赤]
CAR  POSS COLOR  POSS HUE    EQ
```

と表現する (詳細は [高木 87] 参照)．○は名詞，●が関係代名詞を表しており，英語の関係節を利用した構文に規格化している．よって上記の 2 つの文の意味は変換後の構造に反映され，構造がほとんど同じとなり (*=の記号の部分のみ異なる) 意味であることが示唆される．表層の単語ベースからの変換で構築することを目標としており興味深い．

ここで高木らの手法で注目すべき点は「X の Y」の意味処理，ならびに名詞述語や連体修飾節に関する処理において，名詞の属性を定義して扱っている部分である．上記の構造では「色」が属性で「赤」が属性値であり，主体の名詞「車」に対して係っている構造を文から生成している．高木らはこうした属性を色，形，重さなど約 20 種類程度 (高木ら 87:86) 定義して，数学の文章題まで解くシステムを提案している．図示してみると下記のようになる．

```
[車
   属性：色：赤]
```

つまり，上記の西山の分析から主体名詞に対して属性を仮定する必要性がうかがえたが，高木のモデルも同様であり，属性として色，形，重さなど整理しておけば，文書で書かれた世界モデル (高木らの例では数学の課題の世界や視覚の世界) の計算を行うことが出来る．言語は媒体であることから，媒体の先にある情報が処理できれば良いわけで，基本的な処理モデルの枠組みであると考えられる．

*4 ?印は文の意味が取りにくいことを示す．

Generative Lexicon ベース

Pustejovsky[Pustejovsky 95] は名詞の意味構造において特質構造 (qualia structure) を仮定し formal, constitutive, telic, agentive role という 4 つの基本的な属性を仮定することで名詞まわりの表現を柔軟に生成できる枠組みを提案した．

さらに影山 [影山 11] ではこのアイデアをさらに拡張し formal (外的分類)，constitutive (内的構成)，telic (目的・機能)，agentive (成り立ち) と考え，「はちまき」や「手ぬぐい」の違い，答案における「白紙」は単に白いという意味では無いことなど意味構造で記述することを提案している．また上記の西山の分析を受けて，飽和名詞の「俳優」と非飽和名詞の「主役」を下記のように整理し直している．

	「俳優」	「主役」
外的分類	人間 (x)	人間 (y)
目的・機能	x が芝居や映画で劇中の人物を演じる	y が芝居や映画で劇中の人物を演じる
成り立ち		y が [w] の主要人物の役をつとめる

ここで取り上げたいのは，次の 2 点である．まず 1 つ目は GL では 4 つの役割として名詞を特徴を属性にわけて記述していることである．高木らの分析における「色」「形」「重さ」といった属性は内的構成に位置づけて記述されると考えられるため，構造に矛盾が無い．つまり名詞の意味構造は属性 (特徴量のタイプ分け) として分解して記述するという点である．こうした属性を各個別に設定するのは大変であるため，外的分類で「人間 (x)」など上位概念，つまり，オブジェクト指向プログラミングで言えば，属性の継承関係を示しており，default の意味関係があれば省略できるという枠組みを入れられることを示唆している．

2 つ目の特徴として名詞の意味構造に対して動詞を記述している点である．これは名詞と動詞はある特定の組が特別な意味を持っていることを記述する必要があることを示しており，名詞と動詞の意味構造をそれぞれ独立に記述しただけでは成立しない意味関係があることを示している．目的と成り立ちは大きな分類で有用であるが，後の節では我々の分析から情報抽出では動的に人の認識に関する属性を記述する必要があることを指摘する．

知識工学のアプローチ

一方，言語から離れて名詞の参照先である実世界のモノの意味構造の記述でも属性によるモノの特徴が整理され，動作的な内容も取り込まれている．文献 [冨山 98] では，部品を知識構造で記述し，設計に役立てたりコピー機での故障の際の自己診断，機能の拡張に応用している．

興味深いのは，モノの意味構造も上記の先行研究同様，属性に分解して記述するオブジェクトとして記述し，モノの挙動をシュミレーションしている点である．モノの意味構造を上記の言語表現での名詞意味構造を包含しつつ，新たに (1) 見方の異なりによる機能の異なりと (2) 状態遷移，を導入しており，より実世界をシミュレーションできるように拡張されている点である[*5]．この見方によるモノの異なりと，異なった見方での動作状態の伝搬を扱う状態遷移を取り込んだアプローチを FBS モデリング [Umeda 95] として提案しており，この枠組みによってコピー機の故障診断や機能拡張による故障機能の補完といった高度の機能を実現した実システムを販売するに至っている [冨山 98]．

この見方による機能の異なりと状態遷移の関係について簡単に説明する．例えばある部品「電気回路」の場合，電気回路としての機能の見方以外に，単に熱を発生する部品としての機能の見方があり，それぞれにおいて，システムの中で役割が異なるモノとして定義される．その見方に応じて，どのような機能があるか，ある機能を作り出すために，どのような機能が必要かの連鎖をメタモデルとして記述しておき，ある製品がどいう機能の組み合わせ (そしてそれを構成する部品の組) で構築されているかを計算機に持たせる仕組みである．

これまでの GL までの議論を重ねると，見方の異なりは結局，意味構造のどの属性に着目するかであり，状態変化はその属性がどう状態変化するかである．よって GL の構造をそのまま拡張することが出来ると考えられる．

以上の結果をまとめると名詞，名詞句，ならびにモノの意味構造の記述では

- 属性に分けて記述する
- 述語も名詞に取り込んで記述する
- ある属性に注目して状態変化構造を取り込むことで名詞 (モノ) どうしの時間展開を記述することができる

という共通点が見受けられた．以降の節で我々が分析した結果からこれらをさらに拡張していく．

4. 述語の分析からの名詞意味構造の拡張

4.1 状態変化の取り込み

前節までの議論で，状態変化を名詞の意味構造に取り込こめる点を指摘したが，本研究では既に文献 [竹内 13] で示したとおり，状態変化は名詞のある属性の書き換えとして整理することで，移動や状態変化の起点・着点のペアが見通しよく整理できることを提案している．例えば「塩を手元に持ってきた/移動した」の表現はどちらも「塩」という物体に対してその位置を「手元」に変えたことを意味しており，下記のような状態変化であると考えられる．

```
[塩                                    [塩
 属性：位置 (a)]  =>位置変化=>    属性：位置 (手元)]
```

こうした状態遷移は FBS モデリングに対応しており記述枠組みは既に提案されている．よってこれからの問題は，FBS モデリングではすべての属性と属性値は定義された範囲の値であるのに対して，言語表現では用意できない新たな属性値が必ず現れるという点である．これに対処するため，例えば基本的な属性値 (例えば「手元」は身体付近であり，いつでも利用可能という範囲) を設定しておき，さまざまな表現に対して設定した値にどう集約するか，またどう基本属性値を設定するかなどが問題となる．

この課題は 2. 節で述べたように単に名詞の意味構造の記述では閉じない問題であるため，情報抽出システムを構築して，さまざまな実問題を解きながらシステムを更新することで明らかにする問題であると考えられる．

*5 GL でも意味構造に対して時間を取り込んでおり，影山も名詞の中に隠れた時間の概念があることを GL で記述している (影山 2011:46)．ただ状態遷移として処理できる形まで提案されていない．

状態遷移が有効である具体例として 2. 節の図 1 の情報検索の問題を取り上げてみよう．この例では「診療所」の開いている時間が示されているが，これは診療所の機能 (患者から見た場合) がその時間のみが ON でそれ以外が OFF という状態を指示しており，質問文の要求する時間 (「日曜午後 8 時」) で機能が ON かどうかを調べるというものである．これにより，その診療所で受診できるかどうか適切に回答することが出来る．

このように状態変化は文書情報からの情報抽出について必須の機能であり一見簡単な情報抽出でも名詞における状態遷移の処理が必要になる．

4.2 人間の活動や認識における名詞の位置付けも記述

述語の意味分析ならびに，情報抽出のタスクを分析すると 3. 節で取り上げた telic，agentive role だけではなく様々な述語 (すなわち人間の活動の中での名詞の位置付け) を記述しておく必要があるように見える．

具体的には図 1 の例で示したように「診療所」には開いている時間が記述されていたが，それは患者からみた機能であり (telic に相当?)，一方，勤務する医者からすれば，勤務時間という別のタイムテーブルが存在する (agentive role に相当?)．さらに，その医療機関の評判や評価や認識といった情報は人から発生するものであるが，モノに対して総合して記述しておくことが処理の観点から扱いやすいと考えられる．例えば「この診療所は安心できる」「腕が良い」「子供にはとても良い診療所」などの評価や認識である． ここまで含めて今まで議論し

```
[名詞概念
    外的要因:  is_kind_of (x)
    内的要因:  色:
               形:
               面積:
               部分:
               …
    目的・機能:
    成り立ち:
    その他: 評判.. ]
```

図 2: 抽象的な名詞の意味構造

た名詞の意味構造をまとめると図 2 に示すような構造になる．

言語学における語彙意味論の立場からすれば，その名詞の意味構造に記載すべき内容は語を成立させる最低限の要素に限るというのが基本的な立場であろう．これは本研究の枠組みでいうならば情報抽出システムをソフトウェアと捉えた場合，システム辞書が持つ基本オブジェクトデータと捉えることが出来る．つまり，評判やその診療所 (インスタンス) に関する人からみた認識などは，既存のオブジェクトデータに対して動的に加えられた属性項目と考えられる．人の言語表現はまさに発話者の認識において，抽象的に聞き手と共有するオブジェクトに対して個別の情報を加えることで新たな情報を提供していると捉えるならば，こうした名詞意味構造の属性の拡張は取り込むべき機能であり，ソフトウェアにおけるオブジェクト指向の枠組み*6で情報抽出システムを構成していく必要がある．

5. まとめ

名詞の意味構造について従来の言語学および知識工学における先行研究を踏まえて情報抽出という具体的なタスクの視点に立ち議論した．その結果，(1) 属性として項目を分けて記述すること，(2) 述語との関係を名詞に記述すること，(3) 状態遷移モデルを導入する必要があることを明らかにした．さらに，情報抽出にあたり (4) 人の評判や評価，認識といったものも動的に加えられる名詞の意味構造の属性として必要であることを議論した．また，こうした名詞の意味構造を情報抽出というタスクで具現化するために分野依存知識，背景知識の構築が不可欠であり，これらの部分処理を切り分けた上で情報抽出システムを構築する必要性があることを主張した．また，名詞の意味構造における属性の取り扱いは，ソフトウェアにおけるオブジェクト指向の考え方と類似しており，こうした考え方による文書処理システムが提案されつつあり [竹内 14][山田 14] 今後の発展が期待される．

今後具体的に辞書と同時に情報抽出システムを構築しながら，背景知識記述，文の規格化による情報抽出システムの部分処理システムを詳細化する予定である．

*6 Minsky のフレーム理論そのものである．

参考文献

[Pustejovsky 95] Pustejovsky, J.: *The Generative Lexicon*, MIT Press (1995)

[Umeda 95] Umeda, Y., Tomiyama, T., and Yoshikawa, H.: FBS modeling: modeling scheme of function for conceptual design, in *Proc. of the 9th Int. Workshop on Qualitative Reasoning*, pp. 271–278 (1995)

[Winograd 72] Winograd, T.: *Understanding Natural Language*, Academic Press (1972)

[ベイカー 11] ベイカー スティーブン：IBM 奇跡の“ワトソン”プロジェクト: 人工知能はクイズ王の夢をみる, 早川書房 (2011)

[影山 11] 影山 太郎：日英対照 名詞の意味と構文, 大修館書店 (2011)

[高木 87] 高木 朗, 伊東 幸宏:自然言語の理解, 丸善出版 (1987)

[山田 14] 山田 隆弘：語彙概念構造のオブジェクト指向化について, 言語処理学会第 20 回年次大会 (2014)

[西山 03] 西山 佑司：日本語名詞句の意味論と語用論, ひつじ書房 (2003)

[竹内 13] 竹内 孔一, 竹内 奈央, 石原 靖弘：述語項構造のシソーラス分類と意味役割の設計について, 人工知能学会全国大会, pp. 2D4–OS–03a–1 (2013)

[竹内 14] 竹内 孔一, 竹内 奈央, 石原 靖弘：言語学の知見に基づく関数オブジェクトを利用した言語理解システムの構成, 言語処理学会第 20 回年次大会 (2014)

[田代 11] 田代 ひとみ, 中村 則子, 初鹿野 阿れ, 清水 知子, 福岡 理恵子：新完全マスター読解日本語能力試験 N2, スリーエーネットワーク (2011)

[冨山 98] 冨山 哲男, 桐山 孝司, 梅田 靖, 下村 芳樹, 吉岡 真治：第 5 章モデルに重点を置いたアプローチ, 工学知識のマネージメント, pp. 180–229, 朝倉書店 (1998)