

述語項構造を意識した名詞の意味構造アノテーションのための 名詞意味構造の検討

竹内 孔一 (岡山大学大学院自然科学研究科)¹

Consideration of Semantic Structure for Nominal Noun on Predicate-Argument Thesaurus

Koichi Takeuchi (Graduate School of Natural Science and Technology, Okayama University)

要旨

動詞, 形容詞, 形容動詞といった述語の状態変化性を主に取り上げて述語項構造シソーラスを構築してきた。一方で, 文の意味は述語だけで無く, 名詞が概念を持つためこれらの意味を記述して取り込む必要がある。名詞概念の意味記述は様々な分野で提案されているが, 本稿では言語学で分析されてきた西山の名詞句の分析 (西山 (2003, 2013)), ならびに影山の名詞句に関する分析 (影山 (2011)) を元にした項構造をベースの名詞意味記述を仮定してアノテーションのためのフレームワークを考察する。具体的には西山で提案されてきた非飽和名詞という考え方をベースに, 必要とする要素を生成語彙論 (Generative Lexicon) の枠組みを利用して記述する影山の方法を採用する。さらに周辺の情報として, 名詞述語文との類義データの構築, 機能語表現と類義データの構築を一貫した項構造レベルで統一して記述し, 述語項構造シソーラスの一部として拡張することを提案する。

1 はじめに

述語の持つ項構造を一様な方法で記述することで, 述語間の交替関係だけでなく, 状態変化と結果状態といった語の概念に特徴的な含意関係を記述してきた。例えば, 「育てる」「鍛える」という語について下記の例文のように, 述語の係り元 (項) に対して意味役割を付与し, 述語に対しては語義概念を付与している。

- 母が [動作主] 僕を [対象 (人)] 一人前に [着点 (状態)] 育てる [成長]
- コーチが [動作主] 彼女を [対象 (人)] 一流選手に [着点 (状態)] 鍛える [成長]

ここで, 名詞句における [] 内は意味役割を現し, 述語における [] 内は語義概念を現す。

語義概念には語彙概念構造 (影山 (1996)) を拡張した構造を記述しており, 状態変化と状態といった含意関係を記述している。現在公開しているのは動詞だけであるが², 内部では, 形容詞, 形容動詞についても付与を行っている (竹内他 (2013))。例えば, 形容詞, 形容動詞は下記のような例がある。

- アントシアニンは [対象] 目に [目的] 良い/効果的だ [可能]

意味役割の種類は現在, 基本種類が 29 種類に対して属性との組み合わせとなり, 約 90 種類程度定義されている。また述語の概念は約 1000 種類ある。各概念には述語と例文が登録されおり, 約 1.1 万語に対して約 2 万例文が存在し, 各例文には上記の例に示したように, 意味役割と概念が付与されている。また, この体系に従って日本語書き言葉均衡コーパス (BCCWJ) (約 6000 文) の動詞に対して付与を行っている (竹内・上野 (2013))。

一方で, こうした述語項構造を言語理解の課題, 例えば NTCIR-10, RITE-2 含意認識タスク³ に利用しようと考えたとき, 名詞の意味構造がほとんど課題の中心であることが分かる。

- (t1) BLT サンドイッチとは, サンドイッチの一種であり, パンに挿む食材として, ベーコン, レタス, トマト が用いられることから, それぞれの頭文字を取って名づけられた。

¹koichi@cl.cs.okayama-u.ac.jp

²動詞項構造シソーラス <http://vsearch.cl.cs.okayama-u.ac.jp/>.

³<http://research.nii.ac.jp/ntcir/ntcir-10/index-ja.html>.

- (t2) サンドイッチの略称として具材となるベーコン、レタス、トマトの頭文字 BLT が用いられるものがある。

含意認識タスクとは、t2の文の内容がt1に含まれているなら真、含まれていないなら偽と判断するタスクで、専門知識が無くてもネイティブならば読んだだけで判断できるペアが構築され公開されている。この文では、「サンドイッチ」という物のタイプを表す名詞、「一種」、「略称」、「具材」、「頭文字」など名詞の中でも特別な機能を持つものが存在する。この例では「名づける」という動詞と「略称」との名前の関係、「頭文字」が何を指しているかなどが解けないと、含意認識を正しく判定することは出来ない。

こうした名詞の意味の記述にはどのような研究がなされてきたであろうか? 分野を超えて調べたところでは、言語学の他に、既に人工知能の分野で Winograd (1972) と高木・伊東 (1987) の研究が有り、人工物の概念として設計論の立場から 冨山他 (1998) がある。結局のところ形式的には Minsky のフレーム理論を拡張した形(素性構造をオブジェクトとして概念を記述)に集約できる(竹内他 (2014))と考えられる。よって形式的な側面からは素性構造で記述することになるが、何をどう書くかが問題である。

上記の事例を整理してみると抽象的な名詞に対して2つの関係が存在すると考えられる。

- X の 一種/略称/頭文字 は Y

このような名詞に対する分析は日本語では西山 (2003, 2013) が非飽和名詞として「A の B」の分析から明確な分類を提案おり、さらに影山 (2011) は名詞まわりの項(上記 X と Y)を生成語彙論(Generative lexicon (Pustejovsky (1995))), 以下 GL)の枠組みを利用して、意味構造で記述することを提案している。

そこで、本研究では影山の意味構造を参考に付与可能な名詞の意味構造について考察する。言語資源としては、既に、日本語では名詞格フレーム 笹野他 (2005) として自動構築され、公開されている。名詞格フレームは大規模である一方で、人手による整備がないため、関係名や例文が整理がなされていない。英語では既に NomBank が名詞の項について、意味役割的な関係と、例文を付与して公開している (Meyers et al. (2004))。こうした先行研究を受けて、我々は、既に構築している述語項構造シソーラスの意味役割体系を利用することで一貫した意味的關係を持ちつつ、例文ベースの名詞の項構造と事例を構築する。

さらに名詞周辺の表現として、名詞を含む述語表現と動詞との関係、ならびに機能語表現の集約も行い、述語項構造辞書に統合する。これらの構造について下記に考察を行う。

2 付与可能な名詞の意味構造

まず先行研究での記述法(日本語と英語について)を紹介してから、どのような記述が可能かについて考察する。影山 (2011) は GL の qualia structure を、【外的分類】、【内的構成】、【目的・機能】、【成り立ち】、と解釈し直して、「俳優」と「主役」の項の取り方の違いを概略、下記のような構造で記述することを提案している。これは「主役」という名詞は「そのお芝居 [w] の主役は太郎 [y] だ」

	「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] の主要人物の役を務める

のように何か主体になるもの [w] に対して成り立つ言葉で有り、さらに、主役そのものである項 [y] を取ることを意味している。さらに、外的分類で y は人間であることを示している。

一方で「俳優」の場合、「Xの俳優はY」という文では特に「俳優」が取り込む項としてXが存在するわけではない。この例文でのXと「俳優」との意味的關係は西山(2003)にあるように談話から与えられる関係であり、例えば「北海道の俳優」ならば「北海道で活躍する俳優」なのか「北海道で生まれた俳優」なのかは名詞の意味構造からではなく背景知識や文脈で決まる。

この構造は「主役」などの名詞の意味を記述した構造としては機能するが一方で言語処理の観点から「Xの主役はY」のXが例えば「昨日の主役はY」のように項ではなく、西山(2013)が示す時間に関する表現の場合もある⁴。こうした曖昧性解消には事例が必要であり、さらにその関係を一貫した意味関係で記述できることが望ましい。これを具現化したのがNomBankである。

例えば knowledge という単語に対しては、

- students' knowledge of two-letter consonant sounds

ARG0 = students, REL = knowledge, ARG1 = of two-letter consonant sounds

このように、knowledge の内容部分を ARG1, knowledge を所有する人を ARG0 と PropBank と同様の意味役割の定義を利用して記述する⁵。さらに、辞書内では、ARG0 は“KNOWER”で AGENT, ARG1 は“THING KNOWN OR THOUGHT”で THEME と各語に依存した詳細な意味役割が付与されている。よってこうした事例と一貫した意味役割を付与した構造は記述として参考にした。

これまでの議論では名詞と項との関係に着目してきたが、名詞と項の關係に大きなタイプ分けが存在して分類するという観点がある。例えば西山(2013)では譲渡不可能名詞という「なべの蓋」や「鼻」「耳」といった必ず主体を必要とする名詞の分類を提案している。こうした全体-部分関係は、述語表現でも観測されることから述語項構造ソーラスでは、意味役割に組み込んで記述してきた。

- 彼が[動作主]姿を[身体部分]現す

この例の場合では、「現す」の直接的な対象は「姿」であるが、それは「彼」の一部であることを指している。

名詞の意味構造記述において、西山や影山では明示的な提案がなされていないが、言語理解の観点からは、こうした全体-部分の関係をより一般化したものとして属性(Attribute)を分けて、素性として記述することが行われる(Winograd(1972); 高木・伊東(1987))。例えば「車の色は赤だ」「赤い色の車」「赤い車」はどれも「車の色が赤」であることを述べているが、こうしたものを言語理解として扱う場合、「色」という属性名詞を意味構造内で特別扱いして(属性として取り上げ)、その属性値が「赤」として扱うと、属性どうしの比較など処理が行いやすい。

上記のNomBankでは名詞を20種類以上分けており、その中で「色」「重さ」など ATTRIBUTE という分類で記述されている。例えば color の場合

- the colors and emblems of both teams

ATTRIBUTE, REL = color ARG1 = of both teams

と事例が記述されている。意味役割として ARG1 が THEME, ARG2 が VALUE と記述されており、事例には無いが「赤」など色がある場合も記述可能な枠組みである。こうした分類は日本語でも同様と考えられるため、言語理解を目標にすると取り込む必要がある。ただし、NomBank の事例は PennTreeBank を利用して半自動で構築されていることから、名詞にまつわる項構造の全ての要素があるわけではなく、事例として不足がある。よって例文はコーパスだけでなく作例を許し、なるべく想定される名詞の項が埋まった事例を記録していく必要がある。

さて、上記の分析を踏まえて、名詞の意味構造を下記の重要度順で構築したい。

1 名詞と名詞が取る項の例文を作成し、述語項構造ソーラスの意味役割を付与する

⁴例えば「昭和60年代の東京」、「あの頃の自分」など。

⁵ただしこの例はアノテーションマニュアルの事例であり、version 1.0 内には ARG0 が含まれる事例は無い。

2 名詞を属性, 譲渡不可能名詞, 他, カテゴリ分けする

3 BCCWJで事例を探し, 名詞の項構造を同様に付与する

4 GLの特質構造に基づく影山らが提案する構造を日本語文混じりで記述する
可能な限り述語項構造辞書を利用して記述する

ここで4については説明が必要である。上述のほとんどの議論は1の例文の重要性について記述したが、最初に記した影山らのGLベースの記述では名詞の意味そのものを記述するため、名詞が動詞と明確な関係がある場合、言語理解において重要な含意関係を扱うことが出来る。

例えば、「伊坂幸太郎は『重力ピエロ』の著者である」は「伊坂幸太郎が書いた『重力ピエロ』」の意味を含意しているが、これらの関係は「著者」の特質構造における【成り立ち】内で「書く」と

	「著者」
外的分類	人間 (x)
成り立ち	書く (動作主 [x], 対象 [w])

いった記述が成されていれば用意に結び付けることが出来る。また述語項構造シソーラス内に記述されている概念を利用すれば「記す」など「書く」の類義表現も含意関係を識別することが出来る。

3 名詞を含む述語表現の統合

例えば、「異なる」は「相違がある」といった名詞を含む述語表現と類義関係であると考えられる。述語項構造シソーラスには既に、「異なる」は意味役割とともに記述されているので、同様の形式で、こうした名詞を含む類義表現を類語辞典などから見つけ出し、登録する。

- 事実が [対象] 報道と [相互] 異なる [相違]
- 事実が [対象] 報道と [相互] 相違がある [相違]
- 報道と事実に [対象] 差がある [相違]

これらは一種の連語であるがNomBankにもSupport Verbという分類で、名詞が述語の一部として振る舞う場合の項構造が記述されている。これにより述語項構造シソーラスに名詞を含む類義表現も扱うことが出来る。

4 機能語に関する類義表現の統合

機能表現辞書として「つつじ」⁶が構築されているが、こうした機能表現との類義語は動詞で存在し、述語項構造シソーラスで既に登録されている。そのため、機能語に対して例文と意味役割を付与することが出来れば、機能語表現との言い換えを扱うことが出来る。

例えば、「予測」の意味で「～だろう、～かもしれない」では

- 樹齢は [対象] 千年 [補語相当] かもしれない/だろう
- 樹齢を [対象] 千年と [補語相当 (を格)] 予測する/推定する

と予測や推測といった分類の動詞と対応する。また、「放置」の意味で

- 宿題を [対象] し [連語] ないでおく
- 宿題を [対象] 放置する/ほったらかす

「～しないでおく」は「ほったらかす」など関係がある。こうした機能表現を例文付きで構築することで、品詞を超えた類義表現を幅広く集約することが出来る。

⁶<http://kotoba.nuee.nagoya-u.ac.jp/tsutsuji/>.

5 まとめ

本稿では、既に構築している述語項構造シソーラスの語義概念と意味役割の体系を利用して、非飽和名詞に対する名詞項構造事例構築のための構造を検討した。その結果、まず対象とする名詞と意味的關係がある項を含む例文の作成から着手し、意味役割の付与や名詞の分類を行うことが重要であることを明らかにした。さらに、BCCWJへの付与や Generative Lexicon ベースの意味構造が付与できれば、含意認識タスクでの扱える言い換えの範囲が広まることを示した。また名詞を含む類義語表現、ならびに機能表現の統合についても検討を加えた。

今後の課題として、意味構造を付与するための付与システムを構築し、実際に名詞の意味構造を付与しながら、付与できる範囲の意味構造を明確にする必要がある。

謝辞

本研究は、文部科学省科学研究費補助金基盤研究(C)「言語処理及び言語分析を指向した大規模コーパスを利用した述語シソーラスの拡張」(平成26~28年度, 代表者: 竹内孔一)による補助を得ています。

文献

- Adam Meyers, Ruth Reeves, and Catherine Macleod (2004) "NP-External Arguments: A Study of Argument Sharing in English," in *Proceedings of the Workshop on Multiword Expressions: Integrating Processing*, pp. 96–103.
- J. Pustejovsky (1995) *The Generative Lexicon*: MIT Press.
- Terry Winograd (1972) *Understanding Natural Language*: Academic Press.
- 影山太郎 (1996) 動詞意味論, くろしお出版.
- 影山太郎 (2011) 日英対照 名詞の意味と構文, 大修館書店.
- 高木朗、伊東幸宏 (1987) 自然言語の理解, 丸善出版.
- 笹野遼平、河原大輔、黒橋禎夫 (2005) 「名詞格フレーム辞書の自動構築とそれを用いた名詞句の關係解析」, 自然言語処理, 第12巻, 第3号, pp.129–144.
- 西山佑司 (2003) 日本語名詞句の意味論と語用論, ひつじ書房.
- 西山佑司 (編) (2013) 名詞句の世界, ひつじ書房.
- 竹内孔一、上野真幸 (2013) 「日本語コーパスに対する動詞項構造シソーラスの概念と意味役割のアーキテクチャ」, 言語処理学会第19回年次大会, pp.162–165.
- 竹内孔一、竹内奈央、石原靖弘 (2013) 「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, 2D4-OS-03a-1.
- 竹内孔一、竹内奈央、石原靖弘 (2014) 「述語項構造シソーラスによる述語と名詞の構造化」, 人工知能学会全国大会, 2I5-OS-08b-1.
- 富山哲男、桐山孝司、梅田靖、下村芳樹、吉岡真治 (1998) 「第5章モデルに重点を置いたアプローチ」, 工学知識のマネジメント, pp.180–229, 朝倉書店.

付録

意味役割の種類

現段階での意味役割の種類を記述する。大きく4つの類にわけて、中分類で29種類を定義した。さらにこれらに対して属性タイプが付与される形である。まず29種類を以下に示す。

構文類 ● 連語, 外の関係, 補語相当

対象類 ● 経験者, 被使役者, 対象, 基準, 相互, 起点, 着点, 起点・着点, 通過点, 経路, 方向

動作主類 ● 使役, 原因, 動作主, 使役者, 手段

条件周辺類 ● 限界, 領域, 場所, 時間, 条件, 様態, 程度, 目的, 順接, 逆接

これらに対して、属性タイプが一部付与される。例えば、「村民が悪者を懲らしめる」の場合、「懲らしめる」の対象は基本的には人であるはずである。こうした項の属性タイプが語義の概念から決まる場合にはタイプを付与している。

ここでは対象類について記述すると下記のように行列で現すことが出来る。列に意味役割、行に属性タイプを示す。下記の表における記号“○”印のところが、その属性タイプで限定される意味役割があることを示している。

	経験者	対象	基準	相互	起点	着点	起点・着点	通過点	経路	方向
人		○	○	○	○	○				○
操作対象	○									
程度						○				
場所					○	○				
時					○				○	
感情		○								
身体部分		○			○	○				
状態							○			
抽象						○		○	○	
材料					○					
生成物		○				○				
動作		○								
事態		○								