

## 名詞の項構造データの構築

竹内 孔一 (岡山大学大学院自然科学研究科)<sup>1</sup>

### Construction of Japanese Noun Argument Structure Data

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

#### 要旨

本研究では言語処理に利用可能な名詞の項構造データの構築を行っている。名詞の持っている情報を記述する方法として、述語と意味的關係の強い名詞や述語を関係タイプを付与して整理する素性構造が提案されており、本研究でも最も単純な構造を仮定して、テキストデータに付与を行っている。しかしながら、「以外」や「頭文字」など参照先の概念に対して操作を要求する語(ここでは抽象名詞とする)があり、従来の素性構造でもよくわからない。そこで本稿では、既に構築している名詞項構造データについて整理した上で、集められた抽象名詞に対して分析を行う。どの程度のタイプ分けが可能かについて言語処理における含意認識タスクを意識した分析を提示する。

#### 1 背景

本研究者は既に述語の項の関係を整理し意味役割を付与した述語項構造シソーラスを構築し、規則ベースによる意味役割付与システム公開してきた。しかしながら一方で含意認識タスクなど、言語処理において、ある表現が他の表現に含まれているかどうかを判定するには、述語の情報だけでなく、名詞の情報が必要である(竹内(2014))。名詞の持つ意味における先行研究として近年では、項構造を基にした分析(庵(2007))、非飽和名詞に基づく分析(西山(2003, 2013))、GL(Pustejovsky(1995); 影山(2011))、データ構築(A. Meyers and R. Reeves and C. Macleod and R. Szekely and V. Zielinska and B. Young and R. Grishman(2004))などがあげられる。

本研究ではこれらの成果を受けて、西山(2003)から非飽和名詞を集めて例文の作成と項構造の付与を行っている(竹内他(2015))。さらに素性構造に基づく名詞の意味構造の付与(竹内他(2014))を検討している。これにより、「著者」や「作者」といった表現の異なりと関連する名詞(著者名や作品名)との関係付けを行うことが可能であるが、一方で、「以外」「一種」など項構造の関係だけでなく、意味的な操作を要求する抽象名詞があり、扱うことができていない。また同じ抽象名詞でも「こと」や「場合」などは既存の項構造の枠組で捉えることで問題が無いように見える。つまり言語処理を踏まえた上での抽象名詞まで含めた名詞の意味構造の記述枠組がまだ確定できていない。

そこで本論文ではNTCIRの含意認識タスクRITE-2<sup>2</sup>を言語処理の応用例として設定して、名詞や抽象名詞の意味構造について考察を行う。まず、既に提案している名詞の項構造によってどのような意味的關係が記述できるかをまず整理した後、抽象名詞の分類とその意味記述の試案を行う。抽象名詞の収集方法は明らかでは無かったが、近年、田邊(2008)による連体修飾の方法で一部収集できることが明らかになってきた。よって収集された抽象名詞ならびに含意認識タスクの分析から集めた抽象名詞を分類しどのような意味構造記述が可能か検討する

#### 2 構築中の名詞項構造データと含意認識タスク

本節では、現在構築中の名詞データを整理すると共に、含意認識タスクにどのように寄与するかについて明らかにする。構築する名詞項構造データは下記のものがある。

- 1 名詞項構造事例データ
- 2 名詞項構造アノテーションデータ
- 3 名詞辞書データ

<sup>1</sup>koichi [at] cl.cs.okayama-u.ac.jp

<sup>2</sup><http://www.cl.ecei.tohoku.ac.jp/rite2/>.

まず1名詞項構造事例データであるが、普通名詞に対して「XのYはZだ」という形式で例文を人手で作成し、XやYに対する意味的な関係を意味役割で付与したデータである。現在2500例文を作成している。例えば、「上司」ではその部署の名前と人そのものが現れるので下記ようになる。

- ・ [主体 私の勤める会社] の上司は [対象 (人) 田中さん] だ

【主体】や【対象(人)】は意味役割を表しており、述語ソーラスで72種類定義されている<sup>3</sup>。この事例から、「上司」の周辺には2つの項が出現すること、またその具体的な例が提示されていることから後に名詞の項構造に対する機械学習の適用が期待できる。また、意味役割は今のところ述語から定義した意味的な関係を付与しているが、それだけでは関係が捉えられないことが予測される。そこで、PropBankやNomBank同様、ARG0、ARG1など数字として項を固定して、これに対して【主体】や【対象(人)】を付与することにする。これにより異なる理論が現れても、まずある名詞のARG0、ARG1などのどれかを固定することができるため、同じ土台で議論や実験が行える見通しである。

一方で、項構造事例データはどのように役立つであろうか。例えば下記のような含意認識の事例の際に役立つと考えられる。

t1 サファリジャケットの特徴は、一般的に軽量コットン生地または、より軽いポプリンで縫製されたジャケットで、伝統的にカーキ色、そして付属ベルトや、肩章と4つ以上のプリーツ付きポケット（ベローズポケット）が付いている。

t2 サファリジャケットは、4つ以上あるポケットが特徴だ。

含意認識タスクとは2つ目の文の内容が1つ目の文に含まれているかどうかを判定するタスク（最も単純な場合）である。テスト事例として含意される場合とされない場合の両方があり、含意認識システムはこれらを正確に分類する必要がある。上記の例は含意する例であり、「特徴」という名詞のもつ意味的な関係が解くポイントである。この例文では[t1]では「Xの特徴はY」という構文（construction）であり、[t2]の方は「XはYが特徴だ」という構文である。ここでXは「特徴」を持つ主体であり、Yはその特徴そのもの（つまり意味役割では【対象】）となる。よって「特徴」という項構造の事例とX、Yといった項との関係が同定されれば、これらの含意関係は正しく取り出せる可能性が高くなる。

次に2名詞項構造アノテーションデータは普通名詞に対して意味的な関係のある名詞が文中に出た場合、項として意味関係を記述するタスクである。対象とする文はRITE-2タスクの文であり、これによりRITE-2の名詞の項構造付与システムを作成した場合に、直接精度評価を行うことができる。また作例では無いため、気がつかない項や語義を見つける可能性がある。基本的にはタスク1と同じだがタスク2では文に合わせて項を付与する。基本的にはNomBankと同様で、例文に対して名詞の項に意味関係を付与する。

最後に3名詞辞書データは上記の項構造データに加えてさらに影山(2011)が提案している素性構造を付与する。全ての名詞について記述するのは不必要に難しくなるため含意認識に必要なと思われる名詞に対して付与する。例えば「著者」ならば項構造例文と合わせて表1のような構造となる。表1では「書く」は動詞の語義[生成]といった述語項構造ソーラスの語義概念を付与する。

一方、こうした素性構造は「著者」と著作物との関係を掴むことができる。どう役立つか下記のRITE-2の含意関係にある2文で説明する。

t1 『世界征服者の歴史』とは、『集史』などと並び、モンゴル帝国の政治家・歴史家のアラーウッディーン・アターマリク・ジュヴァイニーによって1260年に完成されたモンゴル帝国を語る上で重要な歴史書である。

t2 アラーウッディーン・アターマリク・ジュヴァイニーは歴史書『世界征服者の歴史』の著者として知られている。

<sup>3</sup>以降ここでは【】は意味役割を表す。

表 1: 「著者」の素性構造

「著者」	
項構造	(ARG0, ARG1)
項構造例文	[ARG1/対象 重力ピエロ]の著者は [ARG0/動作主 伊坂幸太郎]だ
外的分類 (Formal)	人間 (ARG0)
成り立ち (Agentive)	書く [生成](ARG0, ARG1)

この例文では [t1] では「Xによって完成された Y」という構文があり, [t2] では「Xは Yの著者」という構文が記述されている. 上記の素性構造では, 「著者」と「書く [生成]」という動詞が結び付けられており, 述語ソーラスの分類では「完成」も3層目で「生成」と分類されている. こうした情報から項構造が同定できれば, 素性構造を通して含意関係にあることが予測されやすくなる.

このように現在構築中の名詞項構造データは含意認識タスクでの貢献が期待されるが一方でこれら意味構造では捉えられない名詞が存在する. 次節では取り上げて有効な構造について論じる.

### 3 抽象名詞の事例と名詞の意味構造の検討

抽象名詞にどのような種類があるか言語学的な視点というよりも, 言語処理の観点から意味構造を検討してみたい. まず田邊 (2008) が抽出した抽象名詞を分類した後, 意味構造が素性構造だけでは捉えられない抽象名詞について検討する.

本研究での言語処理の視点は項構造に基づく意味構造に分解することである. 述語が取る項の意味的関係の分類 (竹内 (2014)) から, 大きくわけて構文類 (【補語相当】など構文的な関係をしめすもの), 対象類 (【対象】などイベントの状態変化や状態に関わるもの), 動作種類 (【動作主】や【原因】といった述語が指すイベントを発生させる要因), 条件周辺類 (【時間】や【場所】, 【条件】といったイベントの存在条件に関係するもの) の4つに分類される. こうした構造と上記の名詞の項構造データにより文同士の意味的関係を捉えんとする. この視点から, 田邊 (2008) の抽象名詞を分類すると表2のようになる.

表 2: 連体修飾で得られた名詞の分類

意味役割	時空間	「中」「際」「地域」「ところ」「とき」「うち」「間」
	手段	「方法」
	原因	「わけ」「ため」「理由」
	条件	「場合」「限り」
	逆接	「一方」
分類 (名詞の属性)		「人」「もの」「声」「情報」「点」「こと」「事」 「言葉」「方針」「女性」「気」「形」「調査」「意味」
モダリティ		「必要」「つもり」「はず」「予定」
機能的		「ほか」「前」

表2における意味役割に分類されているものは抽象名詞であり, 述語に対するその名詞句の意味役割を指定する働きがあると考えられる. また名詞の属性では「人」や「こと」といった言葉でその名詞句が人間か出来事なのか分類を明確にして述語との関係を明確にしようとしている. また, モダリティは確信度や必須か義務かなどを表すが, 英語の名詞項構造の分析 (NomBank) で既にモダリティ情報を付与することが提案されており日本語でも同様であることがわかる.

一方で、機能的な名詞であるが表2で示した「ほか」「前」のほかに RITE-2 データから「以外」、「頭文字」「略称」「一種」「候補」などがある。これらの機能的な言葉は、なにか元の概念に対して論理的な操作を加えると考えられる。論理的な操作の観点から分類すると、(1) 集合操作(「以外」「ほか」「一種」), (2) 順序操作(「前」「後」「候補」), (3) 対象と操作(「略称」「頭文字」)の3つに分けられる。(1)は集合を仮定する必要がある、含意認識タスクでは「日本・ドイツ以外のほぼすべての国」のように国の集合が必要になる。よって(1)を解くには文で求められる集合を獲得する手法が必要であると考えられる。また(2)から順序構造を扱う必要があることがわかる。時間や場所以外にも、必要に応じた順序の計算が求められる。最後の(3)であるが両方とも文字に関する操作であるが、ということが「略する」ことかは人には理解できるが記述は容易ではない。このあたりは事例を収集してより分析したい。

#### 4 おわりに

構築中の名詞の項構造データの含意認識タスクにおける効果について検討し、見通しを明らかにした。また抽象名詞について分析を行い、ほとんどのものは既存の意味役割関係や名詞の分類であることを示した。さらに、抽象名詞について論理的な操作の観点から分類を行い、集合的な操作や順序操作だけでなく自由度の高い操作を指定する抽象名詞があることを明らかにした。

#### 謝辞

本研究は、基盤研究(C) 課題番号 26370485 (研究代表者: 竹内孔一)の補助を得ている。ここに記して深く感謝する。

#### 文献

- A. Meyers and R. Reeves and C. Macleod and R. Szekely and V. Zielinska and B. Young and R. Grishman (2004) "Annotating Noun Argument Structure for NomBank," in *Proceedings of LREC2004*, pp. 803-806.
- J. Pustejovsky (1995) *The Generative Lexicon*: MIT Press.
- 庵功雄 (2007) 日本語におけるテキストの結束性の研究, 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.
- 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 (2014) 「述語項構造シソーラス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」, 第6回コーパスワークショップ予稿集, pp.51-56.
- 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.
- 竹内孔一, 宮田周, 河村一希 (2015) 「述語項構造シソーラスを意識した名詞データの構築」, 第7回コーパスワークショップ予稿集, pp.143-146.
- 田邊和子 (2008) 「BCCWJ に拠る名詞別格外連体修飾形の形成傾向の分析」, 第7回コーパス日本語ワークショップ予稿集, pp.165-174.