

# 国立国語研究所学術情報リポジトリ

## Construction of Argument Structure Data for Japanese Noun Utilizing Predicate-Argument Structure

メタデータ	言語: jpn 出版者: 公開日: 2018-03-20 キーワード (Ja): キーワード (En): 作成者: 竹内, 孔一, TAKEUCHI, Koichi メールアドレス: 所属:
URL	<a href="https://doi.org/10.15084/00001487">https://doi.org/10.15084/00001487</a>

## 名詞項構造付与データの構築

竹内 孔一 (岡山大学大学院自然科学研究科) \*

# Construction of Argument Structure Data for Japanese Noun Utilizing Predicate-Argument Structure

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

### 要旨

含意認識タスクなど言語処理での文間の表現を取り扱う際、名詞の意味的な関係を捉える必要がある。言語学の分析から名詞の中には名詞の意味を補完する外部情報が必要なものが分かっており、生成語彙における特質構造(クオリア構造)として記述することが提案されている。また言語資源ではNomBankに代表されるように名詞の項構造を事例とともに構築されている。本研究では、先行研究で提案された特質構造を利用した名詞の項構造データを基に言語処理の観点からより形式化した構築法を提案する。具体的には名詞の項構造の例文を構築するとともに、項を同定し、述語との関係を項構造を通して結び付ける記述枠組である。述語のデータとして述語項構造ソーラスを利用し、NTCIRのRITE-2で出現した名詞を対象に項構造の例文および対応する述語と項の関係を記述したデータを構築した。本稿では、記述枠組、および具体的に構築した名詞項構造データの事例を説明すると共に、付与での問題点や現状について記述する。

### 1. はじめに

含意認識タスクなど言語処理において、文同士の意味的な関係を捉えるには名詞の項構造の関係を捉える必要がある。例えば下記の2つの例文は<sup>(1)</sup>文(t2)の内容を文(t1)が含意する例である。

(t1) シアターで上映される映像作品「もしも月がなかったら」は、米国の天体物理学者ニール・F・カミンズ教授の同名の著書をもとにした科学エンターテインメントである。

(t2) 「もしも月がなかったら」は、ニール・F・カミンズ著の書籍である。

ここでは名詞として「作品」「著書」「書籍」といった非飽和名詞が出現し、作品名と作者との関係を結び付ける役割を果たしている。例えば「Xの著書はY」という表現があった場合、「X」の部分に人に関する表現があり、「Y」に何かの名称が現れていれば、Xが著者でありYがその著作物であることを表す。こうした構造は名詞「著書」が誰の著書なのかといった情報を外部に必要とするために生まれると考えられ、名詞の構文に関する研究が行われてきた(西

\* koichi 'at' cl.cs.okayama-u.ac.jp

<sup>(1)</sup> NTCIR (<http://research.nii.ac.jp/ntcir/index-ja.html>) の含意認識タスク (RITE2) より引用。

山 2003, Meyers et al. 2004, 庵 2007, 西山 2013, 影山 2011)。

そこで本研究では先行研究の分析を基に名詞の項構造データの構築を行う。基本的な記述枠組として、影山 (2011) が提案する生成語彙におけるクオリア構造による手法を利用する。表 1 に「作者」「主役」のクオリア構造の例を記す。

表 1 「作者」「主役」のクオリア構造

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

表 1 では「外的分類」で名詞の分類を記述し、「目的・機能」でその名詞に役割的な機能を記述する。「成り立ち」で名詞の意味の中である事態を引き起こした時のみその名詞の意味が成立する場合の意味を記述する。この構造の中で、特に「成り立ち」の部分は含意認識タスクで有効に利用できると思われる。例えば「作者」の場合、『重力ピエロ』の作者は伊坂幸太郎である」という表現は「伊坂幸太郎は『重力ピエロ』を書いた」の言い換えと考えられるが、この時「作者」という言葉と、動詞の「書く」との関係が処理できなければ、これらの含意関係を言語処理で扱うことは難しい。クオリア構造では「作者」と「書く」の関係が記述されているため、これを利用することができれば名詞に関連した言い換えを扱うことが可能となる。

しかしながら一方で、上記のクオリア構造をデータとして構築するには処理の観点から難しい点がある。例えば「外的分類」など名詞に対してどのような分類を仮定すべきかは処理タスクの応用によって異なると思われる。さらに「目的・機能」および「成り立ち」において、述語との関係を記述する際、記号 (y や w) を埋め込んだ文による表現では、直接名詞と動詞の表現間の関係を取り出すことができないため、利用が難しい。

そこで本研究では名詞の項構造を例文として記述し、例文を基に述語との項関係を構造化する手法を提案する。こうした名詞の項構造の形式化は異なるタスクではあるが質問応答システム Watson における言語パターンとして prolog で規則化されており<sup>(2)</sup>既に言語処理として利用されている (Lally et al. 2012)。次節で具体的な構造化の手法について記述する。

## 2. クオリア構造を基にした名詞項構造データの枠組

クオリア構造は名詞の意味を記述することが目的であるため、「外的分類」や意味に合わせた記述を行うため、多くの情報を記述できる。例えば表 1 では「作者」は著者が作品を書かないと作者とは言わないため「成り立ち」では「書いた」と過去形で記述する。しかしながら本研究では言語処理の言い換えに焦点を置くため、名詞の意味構造を網羅的に扱うのではなく、

<sup>(2)</sup> authorOf(Author, Composition) という述語で構文情報や品詞情報を基に文からパターンを同定する。

名詞の項構造を述語で言い換える部分に着目して項構造を記述する。つまり名詞と動詞のように品詞をまたいだ表現を構造的に収集することとする。これらを実現するために構築手順を下記のように示し、事例および各手順での背景と問題点について記述する。

### 名詞項構造データの構築手順

- 1) RITE-2 から名詞を取り出し付与する名詞を決める。
- 2) 名詞 (Y) に対して「X の Y は Z だ」という例文を作成する。
- 3) 例文の X や Z の項に対して番号を付与 (arg0, arg1 など) して固定する。
- 4) 各項に対して、項どうしを表現できる述語を述語項構造シソーラスから選択する。さらに項に対して、シソーラス上で定義されている意味役割を付与する。

例えば、1) で RITE-2 に現れる名詞で「店員」を選択し、2) で例文として「そのコンビニの店員は太郎だ」を作成し、3) で項構造として「そのコンビニ」を arg1、「店員」を arg0 と定義する。次に 4) で、この例文の項に対して、意味が対応する述語を述語シソーラスから探し、「働く」(「労働の語義」) を結び付ける。対応する意味役割として「そのコンビニ」を [場所]<sup>(3)</sup> とし、「太郎」に [動作主] を付与する。この様子を図 1 に示す。

「店員」に対する例文

[arg1 そのコンビニ]の 店員 は[arg0 太郎] だ

述語項構造シソーラスの  
述語(の事例)と結び付け

9807 働く 状態変化なし(活動)-身体的動作-活動・労働  
例文: [動作主 彼が] 働く

意味役割を付与

働く: [動作主 arg0: 太郎]  
[場所 arg1: そのコンビニ]

図 1 「店員」に対する名詞の項構造付与の事例

この名詞の項構造により、「そのコンビニの店員は太郎だ」と「太郎はそのコンビニで働く」の言い換え関係を取り出すことが可能になる。こうした名詞の項構造データを構築するための上記の手順の背景と問題点を下記に示す。

まず 1) の手順を置く理由は、作成した名詞項構造データを処理として利用することを仮定しているためである。これにより含意認識タスクを例に、有効な名詞項構造データの構造を洗練できる。付与作業の問題点としては、非飽和名詞や相対名詞に作業者が気がつかず、付与を取りこぼす点である。処理の観点から付与すべき名詞を見直す枠組が必要となる。

手順 2) の理由は、例文があると機械学習等を利用したアプローチが期待できるためである。また後の項構造を付与して項同士の関係を結び付ける際に、単なる記号のリンクだけでなく、事例があることで、人手の付与作業の判断に貢献すると考えられる。また RITE-2 の文そのも

<sup>(3)</sup> 意味役割ラベルを [] で示す。

ので名詞の項構造を付与しなかったのは、項が出現していない場合が多いことが理由である。項構造データとして名詞の項を表す典型的な例を構築するために作成する手法を選んだ。

一方、付与作業での問題点としては項構造の例文がどのようなものか名詞の構文に関する先行研究を作業者が理解する必要がある。例えば、「X の Y は Z だ」の構文の X は Y の項でなくてはならない。例えば「あの頃の著者」など時間を表す表現が X に来た場合は Y の項では無い(西山 2003)。また Z の項は名詞でなくてはならない。しかしながら、項を持つと考えられる名詞でも Z が表現できない場合がある。例えば「性格」はの項は「彼女」など人や動物などが考えられるが Z にあたる表現は形容詞か形容動詞が普通である(「彼女の性格は穏やかだ」)。このように Z が名詞で無い場合は現段階ではあてはまる表現で記述するに留めている。

手順 3) では作成した例文に対して項を同定して、番号ベースの意味役割を付与する。この時、PropBank (Kingsbury et al. 2002) に従って、人にあたる意味役割を arg0、それ以外の項は直接的なものと考えられる物から順に 1、2、と番号をつけて付与する。これは名詞の項は上記の例で示すように複雑であるため、「動作主」など一貫した分類の見通しが立たないためである。一方で、例文に対して arg0、arg1 と固定することで、後のデータ化において、扱いを固定化できる。よって構文が異なっても番号ベースの項による意味役割で項を同定できるため言い換えなどの処理が可能となる。例えば「主役」に対して「[arg1 その映画]の主役は [arg0 太郎] だ」のように名詞項構造データに記録されていれば「主役」の項が arg1 で、主役そのものが arg0 であることがわかる。さらに異なる構文であっても、「[arg0 彼] が [arg1 この舞台]の主役だ」のように意味的な関係を意味役割で関係付けすることができる。

手順 4) では対象とする名詞の意味に関係し、例文の項をとる述語を語義の違いを考慮して述語項構造ソーラスから選択する。その際、項の役割や付加詞も含めた 72 種類の意味役割の中から選択して、対応付けする。これにより動詞と名詞を同定できる。一方、問題点としては述語ソーラスに登録されている述語は約 1.1 万語 (2.2 万事例) であり、登録されていないものは記述できない。また、名詞に対する述語を付与するのは容易ではない作業である。

次節では上記の作業手順による付与作業を行った結果と問題点について明らかにする。

### 3. 名詞項構造付与作業

#### 3.1 付与作業と付与データ

付与対象のデータは含意認識タスク RITE-2(バイナリタスク)の文を MeCab<sup>(4)</sup> を利用して形態素解析を行い、その中から名詞を全て取り出した (3713 語)。次に項を持つ名詞に対して例文を経済学部 1 年生 2 名が作成した (1299 文)。法学部の 4 年生 1 名 (A)、言語学系研究室所属の修士の学生 2 人 (B、C) の計 3 名が意味役割と述語の同定を行った。1299 文に対して項が付与された箇所は 2554 箇所である<sup>(5)</sup>。項の付与件数の内訳は作業者 (A) が 297 件、(B) が 1019 件、(C) が 2017 件で合計 3333 件付与されている。付与箇所では部分的に複数の作業

<sup>(4)</sup> 辞書は Ipadic を利用した。

<sup>(5)</sup> これらの例文と項を付与した事例は先行研究の名詞項構造データ (竹内ほか 2015) とは別の新たなデータ集合である

者が付与している

そのコンビニの店員は太郎だ											
id	述語	項	arg意味役割	述語(vth)	対応意味役割	qualia	ユーザ	date	time	comment	Actions
65	店員	そのコンビニ	arg1	働く	場所	agentive	okada	2016-01-13	11:53:42		Delete Edit
66	店員	太郎	arg0	働く	動作主	agentive	okada	2016-01-13	11:53:35		Delete Edit
186	店員	そのコンビニ	arg1	働く	場所	agentive	okitsu	2016-01-18	10:12:17		Delete Edit
187	店員	太郎	arg0	働く	動作主	agentive	okitsu	2016-01-18	10:28:41		Delete Edit
3472	店員	そのコンビニ	arg1	働く	場所	telic	koichi	2017-02-04	12:12:44		Delete Edit
958	店員	そのコンビニ	arg1	働く	場所		matsushima	2016-09-21	13:52:30		Delete Edit
959	店員	太郎	arg0	働く	動作主		matsushima	2016-09-21	13:53:31		Delete Edit
3473	店員	太郎	arg0	働く	動作主	telic	koichi	2017-02-04	12:13:07		Delete Edit

Pred

検索

Actions	id	見出し語	yomi	大分類1	大分類2	中分類	小分類1	小分類2	pos
選択	9807	働く	ハタラク	状態変化なし(活動)	身体的動作	身体的動作	活動・労働		動詞
		彼が [動作主] 働く					[1]が活動・労働をする		

図2 CakePHPによる名詞の項構造付与ツール

名詞に対する例文作成、項の同定と述語項構造ソーラスからの述語の選択、および意味役割の付与はCakePHPを利用した付与ツールで付与した。図2に「店員」の例文に対する付与例を示している<sup>(6)</sup>。作業者は名詞の例文を見ながら、検索の部分で関連する述語を述語ソーラスから検索し、適合する述語があれば選択できるようにしている。また、意味役割についても同様で、意味役割の一覧から各項(arg0、arg1)に対して付与できる。図2中のqualiaの項は、各項が生成語彙におけるagentiveかtelicのどちらに分類されるか分かる場合は任意で付与できるようにしている。ただし、図に示すように、付与が無い場合や著者と判断が異なる作業もあるため現段階では必須のデータにしていない。意味役割付与については3333件の項のうち、3001件に対して付与されている。意味役割は図にもあるように述語ソーラスで記載されている必須項だけではないため付加詞など72種類の中から付与が行われている。全体的には作業者にとって名詞の項構造付与は容易ではなく、付与の取りこぼしが多く見られた。また、名詞に関する述語を選択する部分が容易ではない場合があることが分かってきた。次節で問題点を挙げる。

### 3.2 付与作業からの考察

付与作業を通して、名詞の項構造付与の難しい点がいくつか明らかになってきた。大きく分けると、(1)名詞項構造例文作成、(2)項構造を有するかどうかの判定、(3)対応する述語の選択の3つである。

まず、(1)では名詞の構文(「XのYはZだ」)が合わないことが挙げられる。例えば「作品」という名詞では「太郎の作品はこれだ」という例文が付与された。「作品」は項を持つ名詞

<sup>(6)</sup> 作業者 koichi は著者であり、上記の付与数からは外している。

で、「の」の前に作者が来る事例で正しい。しかしながら作者と作品の関係を具体的に入れるとおかしな例文となる。「伊坂幸太郎の作品は『重力ピエロ』だ」では作品が1つに限定されてしまい不自然に感じるため例文として成立しない<sup>(7)</sup>。名詞項構造データとしてできれば具体的な事例が入った名詞の項構造の例文が求められるが、今回の作業で指定した構文では捉えられない名詞が存在することが明らかになってきた。これに対しては今後異なる名詞の構文を用意する必要があると考えられる。

(2) では作業者が例文に対して項構造を付与しない例が少なからず見受けられた。例えば「作品」や「小説」など項を持つと考えられる名詞の項構造が付与されていない。これはひとつには項を持つ名詞の理解が難しいこともあるが、ほとんどの場合上記(1)のように例文が「太郎の小説」など「XのYはZだ」の構文にならなかった場合に作業者が項を付与しなかったことが原因である。よって(1)に加えてツールに例文をより容易に修正できる機能を加える必要があると考えられる。

(3) では名詞に対応する述語の選択が容易ではないことが作業から明らかになった。例えば「罰」の例文「[arg1 盗み食い]の罰は[arg2 反省文]だ」に対して、動詞「罰する」を選択した。項の対応としてはarg1が[原因]、arg2が[手段]が考えられる。つまり文で表現すると「[[原因]盗み食いで][手段]反省文で]罰する」である。「罰」の意味を考えると「罰する」が対応すると考えられるが、一方で作業者の中には、「書く」を選択したものもいた。これは「[[原因]盗み食いで][対象]反省文を]書く」と考えたためである。「反省文」という名詞自身が「罰」に関連した意味を持つため「反省文を書く」ことが例文の言い換えとして外れていないようにも考えられる。つまり、名詞の意味を生成語彙の枠組で述語の項として形式化した場合、名詞の概念との組合せで複数の述語が可能であることが考えられる。例えば「著者」の場合のagentiveに対して「本を書く」以外に「本を出す」という表現も「書く」の意味で使われることがある<sup>(8)</sup>。よって事例をベースとした名詞との結びつきで複数の述語との対応を許す枠組が必要である。

また、(3)では述語そのものを考えるのが難しい場合がある。例えば「候補」の例文「[arg1 次の生徒会長]の候補は[arg0 田中さん]だ」に対して、「選ぶ」という述語を選択した。つまり「[[対象]田中さんを][補語相当(を)]次の生徒会長に]選ぶ」と対応づけた。しかしながら、「候補」の意味としてはまだ「生徒会長」に選ばれたわけでは無い。その可能性があるだけである。よって「可能性がある」などモダリティに関する情報の付与も検討される。これは2節で述べた枠組にも関連する。名詞の意味に関連した述語表現を考えると過去形や機能表現が必要になることが考えられる。言い換え処理として利用する際、どこまでの粒度が必要か応用の観点からの見直しが必要と考えられる。

#### 4. まとめ

本稿では述語に結び付けた名詞の項構造データの構築法について記述した。先行研究の生成語彙を利用した記述枠組を元に、RITE-2 含意認識タスクに現れた名詞を中心に例文を作成し、

<sup>(7)</sup> 倒置指定文と考えられるが「私がつとも好きな」など文脈を与えると自然な例文になる。

<sup>(8)</sup> 例えば質問応答で著者を文から取り出す場合、「XがY(作品)を出した」という構文は有効である。

項構造の同定と対応する述語および意味役割を述語項構造ソーラスを基に付与した。1299文に対して2554箇所の項を同定した事例を作成した。項構造データ構築作業から、名詞の項構造例文作成の難しさや、対応する述語の選択での難しさを明らかにした。今後、付与した事例データを基に名詞項構造付与の枠組を整理する予定である。

## 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

- 西山佑司 (2003). 『日本語名詞句の意味論と語用論』 ひつじ書房.
- A. Meyers, R. Reeves, C. Macleod, R. Szekely, V. Zielinska, B. Young, and R. Grishman (2004). “Annotating Noun Argument Structure for NomBank.” *Proceedings of LREC2004*, pp. 803–806.
- 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.
- 西山佑司 (編) (2013). 『名詞句の世界』 ひつじ書房.
- 影山太郎 (2011). 『日英対照 名詞の意味と構文』 大修館書店.
- A. Lally, J. M. Prager, M. C. McCord, B. K. Boguraev, S. Patwardhan, J. Fan, P. Fodor, and J. Chu-Carroll (2012). “Question analysis: How Watson reads a clue.” *IBM Journal Research and Development*, 56:3/4, pp. 2:1–2:14.
- P. Kingsbury, M. Palmer, and M. Marcus (2002). “Adding Semantic Annotation to the Penn TreeBank.” *Proceedings of the Human Language Technology Conference*.
- 竹内孔一・宮田周・河村一希 (2015). 「述語項構造ソーラスを意識した名詞データの構築」第7回コーパス日本語学ワークショップ予稿集, pp. 143–146.