

PropBank スタイルの意味役割タグを導入した 述語項構造シソーラスとNPCMJへの付与計画

竹内 孔一 バトラー アラスデア† 長崎 郁‡ ホーン スティーブン ライト‡
岡山大学大学院 弘前大学† 国立国語研究所‡
koichi@cl.cs.okayama-u.ac.jp

1 はじめに

日本語の述語に意味役割付与を付与したデータとして述語項構造シソーラスを構築している [5]¹。日本語の動詞、サ変名詞、形容詞、形容動詞、約 1.1 万語について語彙概念構造 [6, 7] の考え方をもとにシソーラス上に展開し、粒度を上げて概念フレームとその意味役割体系をたてて付与事例を構築している。しかしながら付与されている意味役割ラベルには他の意味役割体系とのリンク情報は付与されていなかった。

一方で、英語圏で PropBank スタイルの Arg0, Arg1 といったデジタル化された意味役割が展開されている [3]。これは概念フレームに対して、必須項を番号で指定し事例により、どの項であるか指すことで、名前の違いによる体系の異なりを吸収する記述枠組である²。表 1 に PropBank スタイルの意味役割の利点と欠点を整理する。どのような体系も吸収できる一方で、Arg1, Arg2 といった意味役割のタグが単語の語義ごとに異なるため最終的な意味的關係が語義曖昧性解消 (概念フレームの同定) を行うまで、決定できない問題がある。しかしながら、述語項構造シソーラスでは語義 (概念フレーム) を事例ごとに分けて記述しているため曖昧性がなく PropBank スタイルの意味役割タグを導入することができる。

そこで本研究では述語項構造シソーラスに PropBank スタイルの意味役割を付与した。本稿では対応関係をどのように構築したか、また従来のバージョンからバグの修正など変更した部分について記述する。

さらに、この体系を利用して国立国語研究所が進めているフリーで Download 可能な日本語の統語解析情報付きコーパス (ツリーバンク) NPCMJ(NINJAL Parsed Corpus of Modern Japanese)[4]³に意味役割と概念フレームを付与する構想について述べる。

表 1: PropBank スタイルの意味役割タグの利点と欠点

利点	欠点
どんな意味役割にも対応できる	各事例の Arg0, Arg1 等の意味役割の中身は概念フレームによって異なる
分析が間に合わないときにもまず付与が可能	概念フレームが決まらないうちの意味的關係がわからない

2 意味役割と PropBank スタイルの意味役割の特徴と利用

意味役割は概念フレームに対してどういう意味的關係のタイプがあるかを記述したものである。大きく分けて 2 種類あり、述語に対して特定の関係にあるかか先 (項)⁴とどの述語に対しても付加可能なかか先 (付加詞) がある。意味役割は名前の付け方や粒度の捉え方で複数の体系が存在する可能性があるが、事例をもとに、数値化した意味役割を定義しておき、それらと名前の対応をとることで意味役割を定義してデータに対するアノテーションが可能になる。下記に「拡大する」の事例を示す。

- (1) [Arg0:動作主 フランスが] [Arg1:対象 領土を]
[程度の変化/多い 拡大する]
- (2) [Arg1:対象 領土が] [程度の変化/多い 拡大する]

ここでは各項に対して Arg0, Arg1 タイプの意味役割を付与する。同時に参考のために述語項構造シソーラスの意味役割体系である「動作主」「対象」を付与する。また、述語「拡大する」に対して、シソーラスで定義している概念フレーム (5 階層) の一部を記述する。

⁴従来、項から見て係り元が項でかか先が述語であるが、近年、述語から項の主辞に対して矢印を貼る Universal Dependency の考え方が広まりつつあるため、本稿もこの考え方に合わせる [8]。

¹<http://pth.cl.cs.okayama-u.ac.jp/>

²また「動作主」や「対象」といった意味役割の名前そのものに違和感がある場合などにもこうした名前を与えないタグは有効であると考えられる。

³<http://npcmj.ninjal.ac.jp/>

事例 (1) と (2) のそれぞれの項において、数字の意味役割ラベルを付与して関係だけを記述することで「動作主」などの名前を決めなくても、項の交替を捉えて、アノテーションすることができる。しかしながら、述語には複数の語義がある場合、これらは語義 (つまり概念) ごとに異なる意味役割になる。下記に「走る」の例を示す。

- (3) [*Arg*₂:着点 (身体部分) 背筋に] [*Arg*₁:対象 (身体部分) 悪寒が] [表現の出現 走る]
- (4) [*Arg*₀:動作主 太郎が] [*Arg*₁:経路 大通りを] [経路の移動 走る]

これらの事例からわかるように *Arg*₁ とに対して「走る」に対する同じ *Arg*₁ であっても意味的關係が大きく異なる。つまり *Arg*₁ と識別しても、概念フレームが異なれば全くことなる意味的關係であり、概念の識別がないと、同じ単語の *Arg*₁ と同様かどうかは明らかでない。PropBank スタイルでは概念フレームも番号化し、全体系への概念フレームの統一を省略し、語義の違いとして定義することで、辞書構築をさせて、アノテーションを優先した体系であるといえる⁵。

よって、述語項構造シソーラスに対して *Arg*₀, *Arg*₁ スタイルの意味役割を付与することで、「動作主」「対象」といった意味役割にリンクが張れるのと同時に、概念フレームも付与されているためどの語義に対する付与かも明らかになる。これにより、従来の約 70 種類に近い意味役割タイプだけでなく、約 30 程度の抽象的な意味役割として利用することが可能になる。

3 PropBank スタイルの意味役割との対応関係の構築

PropBank で定義されている [1] 数字化した意味役割、および、付加詞に対する意味役割を利用して、公開されている述語項構造シソーラスの 2.2 万事例に対して並行して意味役割を付与した。その付与作業の中で、出現した項と意味役割を付与した事例を表 2 に示す。

PropBank では *Arg*₀ は概ね動作主、*Arg*₁ は対象に付与するように配慮されている。そこで、述語項構造シソーラスでも同様の方針で付与した。ここで問題となるのが「経験者」(Experiencer) タグである。経験者は人間である気持や感情、認識をもつ主体であり VerbNet での事例をみると「admire」⁶ などの主語が Experiencer として付与されている。

⁵PropBank の体系から VerbNet や FrameNet へのリンク (SemLink) [2] を構築しており、このリンクを通して単なる語義ではなく全体を包括した概念体系が利用できるように工夫されている。

⁶<http://verbs.colorado.edu/verb-index/vn3.3.2-test-uv/vn/admire-31.2.php#admire-31.2.php>

(VerbNet) [*Experiencer* The tourists] admired [*Stimulus* the paintings].

これに対応させて PropBank-1.7 で「admire」の意味役割を確認すると下記のように *Arg*₀ で付与されている。

(PropBank) [*Arg*₀] admired [*Arg*₁ *trace*-1]⁷.

他の「feel」など認識に関する主語なども *Arg*₀ で付与されていることから述語項構造シソーラスでは「経験者」を下記のように *Arg*₀ で付与している。

(5) [*Arg*₀:経験者 社員が] [*Arg*₁:対象 給与を] 受給する

述語シソーラスでは主に項の事例を中心にアノテーション事例を構築しているため、例えば「時間」をあわらず *Arg*_M-TMP や「様態」を表す *Arg*_M-MNR などの事例は付与されていない。こうした意味役割に対応するために、例文に対する意味役割付与データの構築を予定している。既に、著者は日本語書き言葉均衡コーパス BCCWJ に対して述語項構造シソーラスの意味役割と概念フレームを付与した BCCWJ-PT⁸ に意味役割を付与している。こうした付加詞の意味役割 (*Arg*_M) はおそらく述語シソーラスの付加詞の意味役割と対応関係をとることのできるかなりの部分が人手を介さず変換で付与できる見通しである。そこで、5 節で示すように国立国語研究所が開発する統語解析情報が付与されている NPCMJ に対して付与することで、意味役割の付与事例の拡大を考察する。

4 BCCWJ-PT における意味役割データとの統一

PropBank スタイルの意味役割を導入すると同時に「時間」に対する意味役割を整理する。BCCWJ-PT では従来の意味役割体系で「時間 (毎)」「時間 (点)」「時間 (間)」と 3 種類の詳細化したタグを付与していた。しかしながら意味役割付与が難しいことから「時間」に統一する。元々、述語項構造シソーラスの辞書上では「時間」で扱ってきたが一方で BCCWJ-PT の旧版 (version 0.90) では分けて付与していた。そこで、今回のシソーラスの更新とともに、BCCWJ-PT も 0.91 版を作成して「時間」タグを統一し、国立国語研究所の中納言サイトで公開している。

⁷ここで trace は空範疇を表し、称賛する内容を指している。

⁸現バージョンは 0.91。http://pth.cl.cs.okayama-u.ac.jp/

表 2: PropBank スタイルの意味役割タグの事例

意味役割タグ	意味	事例
Arg0 から Arg3	必須項	[Arg1:対象 二人が][Arg2:補語相当 (ガ) 不仲に] [Arg3:経験者 人に] 映る
ArgM-DIR	方向	[ArgM-DIR:方向 こっちに] [Arg1:対象 (人) 子供を] 招く
ArgM-LOC	場所	[ArgM-LOC:場所 現場から] [Arg1:対象 (人) 住民を] 閉め出す
ArgM-EXT	程度	[Arg0:動作主 エースが] [ArgM-EXT:様態 一球] 遊ぶ
ArgM-REC	相互	[ArgM-REC:相互 古着と] [Arg1:対象 ちり紙を] 交換する
ArgM-CAU	原因	[ArgM-CAU:原因 飲酒が] [Arg1:(対象 (生成物) 事故を] 引き起こす
ArgM-PRP	目的	[ArgM-PRD:目的 通勤に] [Arg1:対象 車を] 使う

5 NPCMJ に対する述語項構造シソーラスの意味役割と概念フレームの付与

ツリーバンクである NPCMJ は PropBank と同様に構文解析結果が付与されているため、空範疇など参照先が省略されている情報も付与されている。そこで、NPCMJ に意味役割と概念フレームを付与することで、辞書を拡張するとともに意味役割付与データの拡大を目指す。図 5 に NPCMJ の構文解析情報が付与された事例を示す。

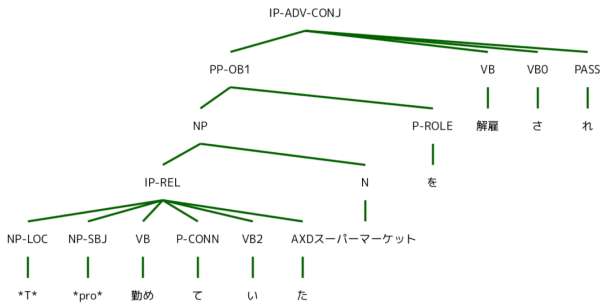


図 1: NPCMJ の統語情報の例 (一部)

元の文は「パンクの青年オッターは勤めていたスーパーマーケットを解雇され、家族からは疎んじられ、悶々とした日々を過ごしていた。」である。ツリーバンクでは、「勤めている」の項として「*pro*」と「*T*」が付与されており、「*pro*」は「パンクの青年オッター」であり「動作主」である。また「*T*」が勤め先の「スーパーマーケット」(着点)である。こうした意味役割情報を付与することでツリーバンクから検索により文例を取り出すことや、言語処理として PropBank と同様に統語情報が付与された意味役割付与データを利用することが可能になる。

6 おわりに

本稿では意味役割と概念フレームを付与した述語項構造シソーラスに PropBank スタイルを導入して事例に対して付与する際の対応関係について記述した。また、シソーラスの更新に関連して意味役割付与コーパス BCCWJ-PT の更新について記述した。さらにツリーバンクである NPCMJ に意味役割を付与する構想について記述した。

参考文献

- [1] Claire Bonial, Olga Babko-Malaya, Jinho D. Choi, Jena Hwang, and Martha Palmer. *PropBank Annotation Guidelines Version 3.0*, 2012.
- [2] Martha Palmer. Semlink: Linking propbank, verbnet and framenet. In *Proceedings of the Generative Lexicon Conference*, 2009.
- [3] Martha Palmer, Daniel Gildea, and Nianwen Xue. *Semantic Role Labeling*. Morgan & Claypool Publishers, 2010.
- [4] 吉本啓, 周振, 小菅智也, 大友瑠璃子, Alastair Butler. 日本語ツリーバンクのアノテーション方針. 言語処理学会第 19 回年次大会, 2013.
- [5] 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, 2D4-OS-03a-1, 2013.
- [6] 影山太郎. 動詞意味論. くろしお出版, 1996.
- [7] 竹内孔一. 語彙概念構造による動詞辞書の作成. 第 10 回言語処理学会年次大会, pp. 576-579, 2004.
- [8] 金山博, 宮尾祐介, 田中貴秋, 森信介, 浅原正幸, 植松すみれ. 日本語 Universal Dependencies の試案. 言語処理学会第 21 回年次大会発表論文集, pp. 505-508, 2015.