

PropBank 形式を考慮した NPCMJ に対する意味役割付与 ～ 態の違いと経験者の付与 ～

竹内 孔一 バトラー アラスデア† 長崎 郁‡ パルデシ プラシャント‡‡
 岡山大学 弘前大学† 名古屋大学‡ 国立国語研究所‡‡
 takeuc-k@okayama-u.ac.jp

1 はじめに

日本語に対して Penn History Corpus の付与方法を参考に構文木を付与するコーパス (ツリーバンク) NPCMJ(NINJAL Parsed Corpus of Modern Japanese)[7]¹に述語項構造シソーラス [9]² の意味役割と概念フレームを昨年付与している。意味役割として、番号による PropBank 形式の意味役割 (Arg0, Arg1 など)[4] と名前による意味役割 (動作主, 対象など) を採用する。番号による意味役割を利用して、動詞の自他の違いによる表現の違い, 使役態や受動態など構文の違いがあっても同じ項を能動態の表現から追跡できることを目指す。また、名前による意味役割を人が検索する際のタグとして利用することで、格交替の事例の検索などが容易にできるようにすることを目指す。

本研究では、述語項構造シソーラス同様、述語が所属する概念フレームに合わせた意味役割付与を行う。よって番号による意味役割も、PropBank では語義毎に意味役割を付与しているが、本枠組では概念フレームに合わせた番号による意味役割付与を行う。PropBank スタイルの意味役割は必須項 (Arg0 他) か付加詞か (ArgM など) の 2 種類に分ける必要がある。これらをどのように付与するかについて自動詞, 他動詞における使役態, 受動態を例に現段階の付与方法について記述する。また、現状 NPCM に対する付与事例数と展望について述べる。

2 日本語に対する PropBank 形式の意味役割付与

意味役割付与は PropBank[3] など英語に対して下記のように自他の異なりがあっても同じ役割の項を同定するラベルとして言語処理では提案されている。マニュアル [1, 2] や解説書 [4] も整備されて、一見理解しやすいが、事例を観測するとマニュアルの付与方法よりさらに詳細な付与と方針があることが付与事例

から推察できる。これらを勘案して、日本語に対する PropBank 形式の意味役割付与方法を作り上げる。

まず事例として英語での代表的な例を示す。

(1) [Arg0 Taro] [open.01 opened] [Arg1 the door].

(2) [Arg1 The door] [open.01 opened].

英語の動詞「open」は他動詞の場合は開ける主語 (動作主)(opener) が開けられる物 (thing opening) でありそれぞれ意味役割は Arg0 と Arg1 と付与する。一方、自動詞の場合は主語が開けられる物であり、Arg1 である。また、動詞の語義によってこれらの意味役割集合は変わるため、open.01 という語義の場合であることを指定している。

これにより、同じ主語であっても述語の意味的な関係が異なることから数字による意味役割を利用して名前をつける代わりに、事例と数字化した意味役割ラベルで関係を規定していくという手法である。意味役割は役割名の考え方で、様々な見方 (理論) が存在するので、それらと対応づけができるように設定している。

また、数字化した意味役割は Arg0, Arg1 と数字を順番に設定するが、自動詞の場合は Arg1 からスタートして良いことを示している。動作主を Arg0 と設定したので、意味的に使役などある動作をさせる主語がある場合には番号がずれてしまうため、特殊な意味役割が付与される。例えば、使役の ArgA が付与される場合は下記ようになる (PropBank-1.7)³。

(3) [ArgA Mr. Dinkins] would [march.01 march] [Arg0 his staff] [ArgM-*DIR* out of board meetings...]

ここでは「歩かせる」という使役化した動詞の例である。また、必須でない意味役割は ArgM で表すため、必須か付加詞かを明確にわけて付与する必要がある。以上の特徴をまとめてみると下記ようになる。

(a) Arg0, Arg1 は順に設定するが、他動詞など動作主があるかないかで Arg1 から初めて良い

(b) 構文的に必須項が増える場合、数字の序列に影響がないようにタグを増やす

¹<http://npcmj.ninjal.ac.jp/>

²<http://pth.cl.cs.okayama-u.ac.jp/>

³以降、英語の事例は PropBank-1.7 の frame から引用している。

(c) 語義毎に付与すべき意味役割集合は異なる

(d) 各語義で必須項を辞書で決めておく必要がある

これらの設定方針を日本語に当てはめた場合、方針(c)と(d)は語義では上記に対応する自他の違いがうまく記述できない。なぜならば、日本語の場合、和語動詞では自他で異なる単語になることが多いからである。そこで、自他を同じ概念フレームとして設定することで、意味役割を付与すると、英語の場合と同様の付与が可能である。

(4) [Arg_0 太郎が] [Arg_1 ドアを] [形状 (開閉)/開ける 開けた].

(5) [Arg_1 ドアが] [形状 (開閉)/開ける 開いた].

この例では「開けた」と「開いた」で同様に Arg_1 で開いた対象である「ドア」を指している。また概念フレームとして、自他をまとめてきた述語シソーラスの概念フレーム【形状 (開閉)/開ける】を利用している⁴。よって概念フレームごとに、 Arg_0 , Arg_1 といった数値化した意味役割を設定して付与することとする。以降、これらの設定化で、経験者と使役や間接受身での付与について考察する。

3 経験者に対する PropBank の意味役割

意味役割で議論となる関係の一つとして【経験者】(Experiencer)がある。ある動作主体であるが、その動作をコントロールできておらず、意図せずに行っている場合の主体である。例えば experience では経験者はマニュアルに指示されているとおり Arg_0 で付与されている。

(6) [Arg_0 Suppliers and subcontractors] are [$experience.01$ experiencing] [Arg_1 heightened pressure to support delivery schedules].

しかしながら、感情など動詞では、英語の場合は経験者は目的語 (Experiencer Object)、日本語の場合は主語 (Experiencer Subject) になり形が異なる。この影響から、PropBank-1.7 の frame をみると、感情の動詞だけ Experiencer は全て Arg_1 で付与されている。理由は、感情を引き起こす原因が Arg_0 であり、動作主として捉える方が自然だからである。例えば surprise では下記のように付与されている。

(7) [Arg_0 BellSouth's revised proposal] [$surprise.01$ surprised] [Arg_1 many industry analysts], ...

⁴最大 5 階層のシソーラス形式になっており、階層をすべて書く【状態変化あり/対象の変化/物理的变化/形状 (開閉)/開ける】である。

frame の注釈には Arg_0 は Cause であり、 Arg_1 は Experiencer であることが記載されている。よって日本語の場合も同様に述語シソーラスで【経験者】として付与してきた意味役割に対して感情的な動詞の場合は Arg_1 を付与することにする。

(8) [Arg_1 太郎が] [Arg_0 爆発音に] [感情変化/驚き 驚いた]

このようにマニュアルには明示されていないが言語学で議論されてきた内容を取り込んで、PropBank 形式の意味役割は付与されているため、日本語の付与でも同様に行う方針とする。

これらを踏まえて、使役態、受動態についての意味役割のタグ付与について記述する。

4 使役態

意味役割付与で期待される利点は、使役文など態の変更の際に、元々の項がどこに移動したのか追跡できることである。これは人が言語を学習する場合でも、言語処理においても有効な機構であると考えられる。よって能動態の意味役割を元に態が変わる場合の意味役割付与を考える。

使役形の場合、元の文の動詞の種類 (自動詞か他動詞かなど) によって付与する意味役割タグを変更する必要があると考えられる。下記に例を示す⁵。

● 自動詞 1 の場合

(9) [Arg_1 街が] 発展する

(10) [Arg_0 新市長が] [Arg_1 街を] 発展させる

● 自動詞 2 の場合

(10) [Arg_0 申請者が] 入国する

(11) [Arg_A 審査官が] [Arg_0 申請者を] 発展させる

● 他動詞の場合

(12) [Arg_0 太郎が][Arg_1 本を] 書く

(13) [Arg_A 出版社が][Arg_0 太郎に] [Arg_1 本を] 書かせる

まず自動詞 1 の場合、主語が Arg_1 であり、動作主体がない動詞 (非対格動詞 [10]) である。この使役文の主語は Arg_0 で良いと考えられる。一方で、自動詞 2 の場合 (非能格動詞 [10])、既に動作主 Arg_0 が付与されているため、数字がずれないように Arg_A を付与することになる。他動詞の場合も同様である。これにより、能動態と使役態の間での項の関係が維持できる。

次に感情的な動詞の場合の使役化について記述する。感情的な動詞の場合は使役のタイプが複数あると考えられる。下記に例を示す。

⁵概念フレームは省略した。

- 感情的な動詞の場合

- (14) [Arg1 母が] [Arg0 私の無茶な行動に] 泣いた
 (15) [Arg0 私の無茶な行動が] [Arg1 母を] 泣かせた
 (16) [ArgA 私が] [Arg1 母を] [Arg0 私の無茶な行動で] 泣かせた

例 (15) の場合は元々の Arg0 が主語になった例であるが、著者は例 (16) の使役も可能であると考えられる。この場合「私」は能動文 Arg0 の「の」で示唆されているが、現段階では別の要素として ArgA を付与することを考えている。また、追跡のために「行動」に対して Arg0 を付与することを考えている。

5 受動態

受動態の場合も動詞の種類で分けて考える。日本語では、被害受身 [6] や間接受身 [11] といった特徴的な構文が存在するため、下記に取り上げて説明する。

- 他動詞 1 の場合

- (17) [Arg0 妹が] [Arg1 ぞうきんを] [Arg2 ポロポロに] する
 (18) [Arg1 ぞうきんが] [Arg0 妹によって] [Arg2 ポロポロに] された
 (19) [ArgE 姉は] [Arg0 妹に] [Arg1 ぞうきんを] [Arg2 ポロポロに] された

- 他動詞 2 の場合

- (20) [Arg0 花子が] [Arg1 手荷物を] 忘れた
 (21) [ArgE 太郎は] [Arg0 花子に] [Arg1 手荷物を] 忘れられた

- 心理動詞の場合

- (22) [Arg1 子供が] [Arg0 おもちゃを忘れたことに] 泣いた
 (23) [ArgE 太郎は] [Arg1 子供に] [Arg0 おもちゃを忘れたことで] 泣かれた
 (24) [ArgA 二郎が] [Arg1 子供を] 泣かせた
 (24) [ArgE 太郎は] [ArgA 二郎に] [Arg1 子供を] 泣かされた

受動態の場合、動作の実行に対して間接的な影響を受ける主語を新たな追加することができる。そこで、上記の例では ArgE (Experiencer 経験者) として新たな項を設定し、数字の項に影響がないように付与した。受動態では使役とともに「させられる」といった複合した表現が可能であるので、それらも考慮して、能動態から項を追跡できるように設定している。

表 1: 付与事例数

付与例文数	32,044
付与した述語数	9,878
付与した項の数	21,454
付与した述語の種類	2,868

表 2: PropBank 形式の意味役割割上位 10 件

Arg	A1	A0	A2	ADV	TMP
頻度	8357	5413	3733	1209	767
Arg	LOC	M	CAU	A3	PRP
頻度	711	353	204	187	157

6 現在の付与データの方法

述語項構造シソーラスの付与事例をもとに、NPCMJ に対して意味役割と概念フレームの付与を作業を約 1 年行った。3 人のアノテータによる現状の作業結果を表 1 に示す。

PropBank 形式の意味役割割の上位 10 件の内訳を表 2 に示す⁶。付与結果から Arg1 と Arg0 が多く、次いで必須項の Arg2、副詞 (ADV) や時間 (TMP)、場所 (LOC) と続くことが分かる。Arg0, Arg1, Arg2 は概略、日本語ではガラニ格に対応することが多く頻度としては名前の意味役割で付与した BCCWJ-PT [5, 8] と大きく相違がないと考えられる。

7 おわりに

本稿では述語項構造シソーラスの概念フレームをベースに PropBank スタイルの意味役割割を導入した意味役割付与の方法について議論した。PropBank で付与されている方針を整理し、日本語にどのように適用するかについて述べた。その結果、自他が別の単語になる日本語では、これらをまとめて扱うと対応が良いことを明らかにした。また経験者についての扱いを整理して、PropBank と対応しつつ、使役態、受動態における意味役割タグの付与方法について提案した。現段階で 3 万文程度の事例について付与作業を行ったことを明らかにした。今後、付与事例の見直しを進めるとともに、新規の事例に対する付与と述語シソーラスの拡張を進める予定である。

⁶名前による意味役割割の数は省略した。

8 謝辞

本研究は国立国語研究所機関拠点型基幹研究プロジェクト「統語・意味解析コーパスの開発と言語研究」および科研費（課題番号 15H03210）と（課題番号 19K00552）の助成を受けたものである。

参考文献

- [1] Olga Babko-Malaya. *PropBank Annotation Guidelines*, 2005.
- [2] Claire Bonial, Olga Babko-Malaya, Jinho D. Choi, Jena Hwang, and Martha Palmer. *PropBank Annotation Guidelines Version 3.0*, 2010. (http://clear.colorado.edu/compsem/documents/propbank_guidelines.pdf accessed 2019/9/6).
- [3] Martha Palmer, Daniel Gildea, and Paul Kingsbury. The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, Vol. 31, No. 1, pp. 71–105, 2005.
- [4] Martha Palmer, Daniel Gildea, and Nianwen Xue. *Semantic Role Labeling*. Morgan & Claypool Publishers, 2010.
- [5] Koichi Takeuchi, Masayuki Ueno, and Nao Takeuchi. Annotating Semantic Role Information to Japanese Balanced Corpus. In *Proceedings of MAPLEX 2015*, 2015.
- [6] Anna Wierzbicka. Are Grammatical Categories Vague or Polysemous? (The Japanese ‘Adversative’ Passive in a Typological Context). *Paper in Linguistics*, Vol. 12, No. 1-2, pp. 111–162, 1979.
- [7] 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.
- [8] 竹内孔一. 語彙概念と語彙概念構造. 松本裕治, 奥村学 (編), コーパスと自然言語処理, pp. 94–113. 朝倉書店, 2017.
- [9] 竹内孔一, 竹内奈央, 石原靖弘. 述語項構造の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.
- [10] 影山太郎. 動詞意味論. くろしお出版, 1996.
- [11] 日本語記述文法研究会. 現代日本語文法 2 第 3 部格と構文 第 4 部ヴォイス. くろしお出版, 2009.