

NPCMJ に対する PropBank 形式の意味役割と 名前の意味役割付与における考察

竹内 孔^{1,a)}

概要：文内の係り関係のある要素に対して、述語から意味的な関係を付与する意味役割ラベルを付与する研究について記述する。意味役割ラベルには「動作主」や「対象」といった名前をつける形での関係付けをする手法がある一方で、Arg0, Arg1, ArgM-ADV などのように数字と記号で表す意味役割が PropBank で提案されている。こうした意味役割ラベルは、自動詞や他動詞の違いで、なにが動いているのか(対象)、動かしているのか(動作主)を追跡するためのラベルとして利用され、文の抽象的な構造化として利用されている。英語圏には、こうした意味役割の辞書と意味役割付与データが存在するが、日本語では PropBank 形式の意味役割辞書と付与事例は存在しない。そこで本研究では著者が従来名前での意味役割で作成してきた辞書をベースに拡張して、PropBank 形式の意味役割辞書の構築と、テキストに対する付与について報告する。テキストは国立国語研究所が開発を進めている日本語統語・意味解析コーパス NPCMJ (NINJAL Parsed Corpus of Modern Japanese) を対象にする。NPCMJ は Web で公開している日本語構文木付与テキストデータであり、現在約 4 万文付与されている。本稿では日本語に対して PropBank 形式と名前形式の意味役割をどのように付与してデータを構築しているかについて記述する。

Annotating Semantic Role Labels of Both PropBank Style Roles and Conventional Named Roles to NPCMJ

1. はじめに

本研究では文の意味的な構造化として述語を中心とした係り元に対するラベルを付与する意味役割ラベルの体系と事例データの構築を行っている。例えば、英語の open, 日本語の「開ける(あける)/開く(あく)」に対して、下記のようなラベルを付与することである。

- (1) [Arg0 Taro] [open.01 opened] [Arg1 the door].
- (2) [Arg1 The door] [open.01 opened].
- (3) [Arg0 太郎が] [Arg1 ドアを] [開ける 開けた]
- (4) [Arg1 ドアが] [開ける 開いた].

英語の事例では、open は自動詞でも他動詞でもどちらも表現があり、開く対象である「the door」は自動詞の場合は主語となり、他動詞の場合は目的語になる。このように「ドアが開いた」という意味を表現する場合に、誰が開けたのか

を明言する場合(他動詞)と省略する場合(自動詞)がある。このように、構文は異なるがほとんど同じことを述べる表現同士の関係を交替と呼び^{*1}、述語の意味を分析する手法として語彙意味論で研究されてきた(例えば [6], [7], [17])。開く対象である「ドア」をなるべく同じ意味的なラベルで指して関係を取り出そうとするのが意味役割ラベルである。文(1)と文(2)ではどちらも「the door」に対して Arg1 という意味役割ラベルを付与している。日本語の場合は、自動詞と他動詞で動詞の形が異なるが、英語と同様の表現が可能である。よって文(3)と文(4)では日本語の場合も「ドア」に対して Arg1 というラベルを付与している。

この Arg0, Arg1 というのは PropBank[2] が提案した記述方法で「動作主」や「対象」といった名前を使わない分、名前に影響されることなく、意味的な関係で、係り元のタイプ分けを記述する。例えば上記の場合の意味役割を記述すると

- Arg0: 開ける動作をするもの
- Arg1: 開けられるもの

^{*1} 言い換えの一種である。

¹ 岡山大学大学院自然科学研究科
Graduate School of Natural Science and Technology,
Okayama University, 3-1-1, Tsushima-naka, Kita-ku,
Okayama, 700-8530, Japan

^{a)} takeuc-k@okayama-u.ac.jp

と考えられる。こうした定義を辞書として記述しておくことで、一貫した意味的な関係を記述できる。英語圏では PropBank として意味役割の定義辞書と新聞記事などに付与した事例を構築して配付している*2。さらに、これらのデータを元に意味役割を自動で付与するシステムの構築と、意味役割をベースとした文の構造化が AMR (Abstract Meaning Representation) として提案されている (例えば [1], [10], [12], [13])。

一方で、こうした意味役割に関連した日本語の言語資源 (言語データ) を取り上げると、大規模なものでは EDR 電子化辞書が挙げられる [19]。約 30 種類の名前による意味役割 (agent, object) を定義し、約 20 万文の日本語について係り受け解析に意味役割が付与されている。しかしながら、名前の意味役割は PropBank 形式のデジタル化した意味役割と対応が異なること、また、述語の各語義ごとに意味役割集合 (roleset) を記述した辞書が存在しないことから利用することは難しい。

意味役割集合とは上記の事例に示したように代表例文に対してどういうものが Arg0 や Arg1 であるかを定義したものである。複数語義がある述語の場合は基本的に意味役割集合は語義毎に異なる。よって語義毎の例文に対して意味役割集合を付与した定義ファイルが必要である。本研究では意味役割集合辞書を概念フレーム辞書として記述する。

上記の事例で open.01 と記述しているのは open という動詞に対する語義の番号であり、open.02 では begin の意味で意味役割集合が定義されている。また、こうした辞書は PropBank では公開されているため事例が確認できて意味役割ラベルの理解が可能であるが、EDR 電子化辞書は有料であるため、付与した物を Web 上で公開して共有することは不可能である。

著者は日本語の述語について Lexeed の語義分類データをもとに、述語の意味を概念フレームと意味役割で分類した述語項構造シソーラスを構築し Web 上で公開してきた*3。日本語の動詞、形容詞、サ変名詞、形容動詞、約 1 万語に対して、語義毎に例文を作成し、係り元と述語との意味的關係を「動作主」「対象」など名前による意味役割ラベルを付与している。意味役割ラベルの形式は PropBank と異なるが、語義毎に例文が整理されているため、PropBank の形式に従った意味役割を併記することで、フレーム辞書として利用できる*4。

さらに、近年、国立国語研究所で言語学および言語学習者を意識した構文木付与テキストデータ NPCMJ (NINJAL Parsed Corpus of Modern Japanese) が開発されつつある。

Web 上で掲載可能な日本語のテキストを対象に構文木を付与することで幅広く利用されることを狙っている。多数のジャンルに対して現在約 4 万文に構文木が付与されている。

そこで、本研究では、NPCMJ に述語項構造シソーラスをベースとした意味役割 (名前形式と PropBank 形式の両方) と概念フレーム (各述語の語義を共通概念としてまとめたもの) を付与することで、英語の PropBank に近い意味役割が付与された日本語の言語資源構築を目指す。述語項構造シソーラスは例文をすべて著者側で作成しているため、Web 上で公開し、download して利用することも可能である。NPCMJ も同様であるため、海外の言語資源 (WordNet*5, FrameNet*6, PropBank の概念フレーム) で見られるように、Web 上で付与事例が確認できる言語資源を構築することができる。NPCMJ も Web 上で確認できる上、研究利用のために download して利用することも可能であるため、広く利用されることが期待される。

PropBank 形式の意味役割付与はアノテーションマニュアルが存在するが [2] 実際の付与例を確認すると言語学における語彙意味論の背景が考慮されており単純ではない。また英語と日本語の文法的な異なりを考慮する必要がある。本論文ではそれらの異なりを踏まえて、PropBank 形式の意味役割を日本語に適用し、どのように述語シソーラスに記述したか、また、どのような情報抽出が可能かについて明らかにする。

2. 意味役割ラベルの基本的な考え方

PropBank 形式の意味役割には大きく分けて 2 種類のラベルが存在する。一つは上記の例にも挙げた Arg0, Arg1 という数字による意味役割である。これは述語の語義毎に異なる必須項といわれるもので、述語と密接に関連する係り元 (語彙意味論の分野語学では項と呼ばれる) との意味的關係を捉えるために記述する。もう一つの形式が ArgM-ADV (副詞) など付加詞と呼ばれる要素にあたるものである。以下では、付加詞と必須項について現段階で設定している内容について記述する。

2.1 PropBank 形式における付加詞の意味役割

(5) [Arg0 太郎が] [Arg1 駅に] [ArgM-ADV 車で]
[着点への移動 行く]

(6) [Arg1 ドアを] [ArgM-TMP 6時に] [開ける 開けた].

付加詞は個別の述語に依存せず述語に付加できる要素で、1 文中に複数出てきても構わない。付加詞のタイプを何種類設定するかは議論がわかれるが、本研究では現在、表 1 に示すものを仮定している*7。表 1 の「連語」であるがこれ

*2 <https://propbank.github.io/> (2020/3/2 アクセス)

*3 <http://pth.cl.cs.okayama-u.ac.jp/> (2020 年 3 月 2 日アクセス)

*4 日本語の事例 (3) と (4) では「開ける」が概念を表す。実際には述語項構造シソーラスによる階層構造の概念フレーム名であり、省略せずに記述すると「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。

*5 <https://wordnet.princeton.edu/> (2020 年 3 月 2 日アクセス)

*6 <https://framenet.icsi.berkeley.edu/fndrupal/> (2020 年 3 月 2 日アクセス)

*7 意味役割の表記として ArgM-TMP は長いので TMP と省略して記述する。

表 1 PropBank 形式の付加詞に対する意味役割

意味役割	意味	事例
ArgM-TMP	時間	[TMP6 時に] 出発する
ArgM-LOC	場所	[LOC 駅で] [Arg1 本を] 買う
ArgM-MNS	手段	[MNS 車で] [Arg2 駅に] 行く
ArgM-DIR	方向	[DIR 右方向に] [Arg1 車が] 曲がる
ArgM-CAU	原因	[CAU 妹の悪行に] 困る
ArgM-CND	条件	[CND 晴れなら] 出掛ける
ArgM-EXT	程度	[EXT5 割] 減少する
ArgM-PRP	目的	[PRP 旅行のために] 働く
ArgM-REC	相互	[REC 古着と] [Arg1 紙を] 交換する
ArgM-MNR	様態	[MNR だらだらと] 走る
ArgM-MDF	修飾	[Arg1 魚を] 焼く [MDF 匂い]
ArgM-SCP	領域	[SCP 学年で] ずば抜ける
ArgM-CMP	補語	[Arg1 焼きそばを] [CMP おやつとして] 食べる
ArgM-CRT	基準	[Arg1 常識で] 判断する
ArgM-PRX	連語	[Arg1 彼が] [PRX 嵩に] 掛ける

は、付加詞というよりはどちらかという構文に関係したラベルである。例文にあるとおり、「嵩に掛ける」はこれの一つの表現であるため、「嵩に」は係り元というより述語の一部の役割である。述語項構造シソーラスでは従来、こうした連語情報も記録していたため、PropBank 形式での意味役割も容易に辞書に取り込むことができている。また、ArgM-MDF は言語学で分析されている外の関係と呼ばれるもので、連体修飾の際に出てくる特殊な関係である。下記に例を示す。

(7) [Arg1 さんまを] 焼いた [Arg0 太郎]

(8) [Arg2 落ち葉を] 焼いた [LOC 公園]

連体修飾の場合、修飾されている名詞が述語に関連する何らかの関係を持っていて平叙文で記述することができる。

(9) [Arg0 太郎が] [Arg1 さんまを] 焼いた

(10) [LOC 公園で] [Arg2 落ち葉を] 焼いた

しかしながら、表 1 の「匂い」というのは、こうした平叙文に戻すことができない。このように連体修飾でしか出てこない関係について、意味的なラベルを付与しているのが ArgM-MDF である。

これらの例から明らかなように表 1 の付加詞に対する意味役割は英語で定義されているものから拡張している。日本語の述語周辺で現れる意味的な関係を考慮して新たに設定している。この他、文書を付与するにあたり ArgM-AND (順接)、ArgM-BUT (逆接) など文の接続に関係するラベルを設定してテキストに意味役割を付与する。

2.2 PropBank 形式における数字の意味役割 (必須項)

述語と係り関係にある要素のうち、述語の意味に密接に関連して、交替などでも関与する要素を必須項として取り扱うことが提唱されてきた。必須の要素を項とよび、複数の項からなるので項構造とばれる (例えば [4], [9])。項構造

は述語の意味を統語構造から捉えるアプローチであるため、主語、直接目的語、間接目的語などを対象としてきた。しかしながら、何を必須にするかは指針はあるものの分析の見方 (概念フレーム) に依存することが考えられる。例えば時間は通常付加詞であるが、下記のように、「延期する」など時間の変更を述べる動詞では我々は必須項と考えている。

(11) [Arg1 会議を] [Arg2 3 時に] [延期 延期する]

(12) [Arg1 試合を] [Arg2 明日に] [延期 延ばす]

日本語の特徴として「ガラニ」格をとる場合は必須項と考えられる場合が多いが「デ格」であっても必須項と考える例がある。下記は言語学でよく取り扱われている構文である。

(13) [Arg1:対象 ペンキを] [Arg2:着点 壁に] [覆う 塗る]

(14) [Arg1:手段 ペンキで] [Arg2:対象 壁を] [覆う 塗る]

文 (13) と文 (14) はそれぞれ「ペンキ」がヲ格かデ格の違いがあるが、述語の意味としてはペンキを壁に塗ることである*8。文 (14) はデ格であるがこの場合は必須項として意味役割を付与する。

よって必須項かどうかは、どこまでを必須項とするか決めるものであると考えられる。つまり必須項の範囲を先に定義するオントロジーと同様の扱いと考えられる。PropBank 以外の言語資源として、FrameNet が挙げられる。FrameNet は述語を含むある表現に対してフレーム (Frame) を仮定し、登場要素である FrameElement に対して中心的な要素である Core を仮定している。概念フレームに対してその概念に登場する必須の要素を定義しており、数字の意味役割に対応すると考えられる*9。そこで、本研究では概念フレーム (述語を横断した形で定義する抽象的な意味) に対して、Arg0 から Arg5 までの意味役割を定義する。上記の「塗る」の例では、「塗る」以外に、「塗りつぶす」「舗装する」「覆う」などの述語が所属する「覆う」フレームに一貫して数字の意味役割を定義する。

一方で、PropBank 形式における数字の意味役割は名前による意味役割と表現によって異なる対応になる。既に文 (13) と文 (14) に示しているとおり、数字の意味役割では「ペンキ」は Arg1 であるが、文 (13) では名前の意味役割は「対象」であり、文 (14) では「手段」と付与されている。これは名前による意味役割はその文において話者が表現した文法的な構造も含めた解釈に従って付与されているためである。よって名前による意味役割では、文法的な取り扱いを形式化している一方で、数字の意味役割は表現の異なりを無視して、意味的に同一のものを付与する。本研究では両方記載するため、必要に応じて文から情報を取り出す

*8 文 (14) の場合はさらに壁を覆うぐらい塗ることを意味している。これらを分けた概念フレームで記述することもある。

*9 正確には PropBank の意味役割は述語の各語義に定義されているので、複数の述語を横断して定義されるフレームと異なる。

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	FrameID	Pos
325	通す [動作主]Arg0 客を [対象(人)]Arg1 居間に [着点]Arg2	トオス	状態変化あり	位置変化	位置変化(物理)	着点への移動	導き	15	動詞
326	案内する 主人が [動作主]Arg0 登山客を [対象(人)]Arg1 お花畑へ [着点]Arg2	アンナイスル	状態変化あり	位置変化	位置変化(物理)	着点への移動	導き	15	サ変名詞
327	先導する ガイドが [動作主]Arg0 旅行者を [対象(人)]Arg1 目的地に [着点]Arg2	センドウスル	状態変化あり	位置変化	位置変化(物理)	着点への移動	導き	15	サ変名詞
328	導く 秘書が [動作主]Arg0 来客を [対象(人)]Arg1 応接室に [着点]Arg2	ミチビク	状態変化あり	位置変化	位置変化(物理)	着点への移動	導き	15	動詞
329	誘導する スタッフが [動作主]Arg0 観客を [対象(人)]Arg1 会場に [着点]Arg2	ユウドウスル	状態変化あり	位置変化	位置変化(物理)	着点への移動	導き	15	サ変名詞
330	エスコートする 男性が [動作主]Arg0 女性を [対象(人)]Arg1 席に [着点]Arg2	エスコートスル	状態変化あり	位置変化	位置変化(物理)	着点への移動	導き	15	サ変名詞

図 1 述語項構造シソーラス

ことが可能である．より多くの具体例は既に Web 上の述語項構造シソーラスに試験的に掲載している*10．

述語に対して使役化により新たな必須項が付与されることが知られている．その場合には ArgA というタグを付与することが PropBank のガイドラインに示されている．

(15)[Arg0 登山者が] [Arg1 ザイルを] つかむ

(16)[ArgA ガイドが] [Arg0 登山者に] [Arg1 ザイルを] つかませる

文(15)では「つかむ」は他動詞であり、「登山者」は動作主を示す Arg0 である．この使役化した場合の例が文(16)である．文(16)には新たな使役者「ガイド」が現れて、「登山者」にその動作をするように仕向ける．この場合に「ガイド」に対して ArgA という意味役割を付与する．

これは英語も同様であるが、日本語の場合さらに受動態でも新規の項が現れる．自動詞の受動態である間接受動態である [11], [18]．下記に例を示す．

(17)[Arg0 記者が] [Arg0 記事を] 書いた

(18)[ArgE 太郎は] [Arg0 記者に] [Arg1 記事を] 書かれた
文(17)に対する受動態として文(18)では新たに影響を受ける「太郎」が項として付加されている．そこで上記に示すように ArgE という意味役割ラベルを新たに設定して付与する．こうした自動詞に対して受動態がとる構文は英語には存在しないため、PropBank のガイドラインにはなく、日本語に対する PropBank 形式の意味役割の設定で新たに設定したものである．

3. 概念フレーム辞書

PropBank 形式の意味役割をテキストデータに付与するには、各述語の各語義において、数字化した意味役割の説明と事例が掲載された概念フレーム辞書が必要となる．本研究では、もともと名前の意味役割で日本語に対して構築してきた述語項構造シソーラスの概念フレームに対して、数字形式の意味役割を併記することで、PropBank 形式の概念フレーム辞書を構築する．図 1 に現段階で Web 上に

*10 <http://pth.cl.cs.okayama-u.ac.jp/testp/pth/Vtths> (2020 年 3 月 3 日アクセス)

Arg0: 突破する人 Arg1: 突破する内容 Arg2: 突破する元	状態変化あり	主体の変化	状況変化	突破	261
Arg0: 諦める人 Arg1: 諦めた内容	状態変化あり	主体の変化(判断・認識の変化)	判断(決心)	諦め	262
Arg0: 決心した人 Arg1: 決心した内容 Arg2: 決心した内容をどう見なすか	状態変化あり	主体の変化(判断・認識の変化)	判断(決心)	決心	263
Arg0: 仮定する人 Arg1: 仮定する対象 Arg2: 仮定した内容	状態変化あり	主体の変化(判断・認識の変化)	判断(認識)	仮定	264
Arg0: 確認する人 Arg1: 確認する内容	状態変化あり	主体の変化(判断・認識の変化)	判断(認識)	確認	265
Arg0: 誤認識する人 Arg1: 誤認識する対象 Arg2: 誤認識した内容	状態変化あり	主体の変化(判断・認識の変化)	判断(認識)	誤認識	266

図 2 PropBank 形式の意味役割の定義ファイル

掲載している述語項構造シソーラスを示す．

図 1 は、「状態変化あり/状態変化/位置変化(物理)/着点への移動/導き」という概念フレームの事例を表している*11．2 行毎に 1 つの事例が記述されており、この概念フレームに属する動詞として、「通す」「案内する」「先導する」「導く」「誘導する」「エスコートする」といった動詞が存在し、それぞれの動詞の例文が「導き」フレームに属することを示している．

例えば frame.id が 325 番の「通す」の場合、「客を居間に通す」が例文であり、「客を」に対して「対象(人)」と Arg1 の意味役割が付与されている．また、「居間に」に対して「着点」と Arg2 の意味役割が付与されている．さらにこの例には出現していないが、「動作主」(Arg0)つまり、この動作を行う要素が存在可能であることが示されている．ここで「対象(人)」というのは他からの作用で状態変化するものであり、かつ基本的に人であることを示している．こうした意味の分類によって、例えば、「X を居間に通す」という文があった場合に、X が人である可能性が高いことがわかる．

図 2 に現在整理中の意味役割定義ファイルを示す．最も左の列が各概念フレームで一貫した数字意味役割の定義を言葉で記述している．こうして意味を記述することで概念が理解しやすくなり、付与誤りを減らすことが期待される*12．意味役割定義ファイルを含む概念フレーム辞書を検

*11 概念フレームの番号 (frame.id) は 15 番である．

*12 今後、意味役割定義ファイルは公開する予定である．

索可能な状態にして内部で公開して付与作業を行う。

4. NPCMJ に対する意味役割と概念フレーム付与

NPCMJ に対して、述語の概念フレームと意味役割を付与する。下記では NPCMJ について記述した後、付与作業について記述する。

4.1 NPCMJ について

日本語のタグ付きコーパスは既に新聞記事を中心に言語資源データとして多くのものが構築されている (例えば BCCWJ[8], EDR コーパス [19] や GDA コーパス [16])。しかしながら、テキストデータに権利があるため、購入や利用契約を結ぶ必要があり、契約なしで付与内容を Web 上で検索することはできない。

そこで NPCMJ では Web 上での検索ならびに研究利用での download が可能なテキストを集めて、構文木や論理的な意味情報を付与したコーパスを構築する [5], [14]。これにより国内外の日本語学習者や言語学者が構文解析済みの日本語例文を Web 上で確認できたり付与された構文や品詞タグなどで検索して事例を確認することが容易になる*13*14。

現在 NPCMJ は 14 の出典 (小説, 聖書, 会話, ニュースなど) に対して構文木を付与しており、最終的に 6 万文まで増やす予定である。付与されている構文木情報には非終端記号の他にゼロ代名詞やトレースまで付与されている。例えば, wikipedia.RepoMAN における 1 文は, 形態素レベルで「*T* パンクの 青年 オットー は *T* pro 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、 *PRO* 悶々とした 日々 を 過ごしていた。」のように記述される。この文の構文木の一部を図 3 に示す。最初の「*T*」は「青年」のトレースである。こ

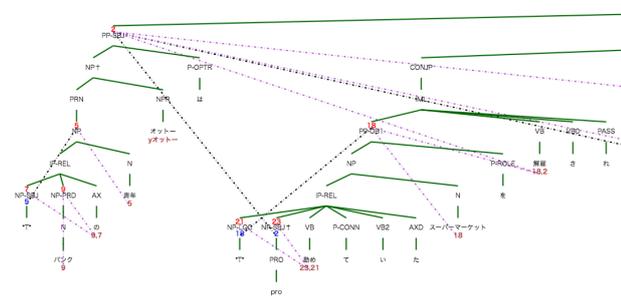


図 3 NPCMJ における構文木の例 (wikipedia.RepoMAN)

の場合の「の」は同格を表す「です/である」と考えており

*13 tree 構造を検索するツール, および構文木付与のマニュアルも公開している <http://npcmj.ninjal.ac.jp/> (2020 年 3 月 3 日アクセス)

*14 漢字圏ではない学習者・研究者のためにテキストはローマ字化したデータも構築している。

「*t* パンクである青年 (t)」つまり「青年はパンクです」が連体修飾の構文になって現れていることを示している。また、「*T* pro 勤めていたスーパーマーケット」における pro はゼロ代名詞であり, ここでは「オットー」のことを指している。連体修飾表現を平叙文に戻すと, 「スーパーマーケットに (*T*) pro が 勤めていた」となる。

現在, このデータは Web に掲載されており, 付与されている構文タグに対して xpath を利用した柔軟な検索が可能である。次節では, NPCMJ に対して概念フレームと意味役割を付与する手法および付与事例について記述する。

4.2 意味役割と概念フレームの付与

NPCMJ は詳細な構文木情報が付与されており, これらを利用して述語と項を組合わせた論理形式の表示機能を備えている [3]。よって構文木から自動で付与する対象の述語とその係り元が抽出されて, 作業者は述語とその係り元に対して, 概念フレームと意味役割を付与する作業に集中できる。

係り元と述語のペアを取り出した結果を図 4 に示す。図

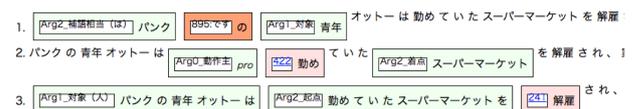


図 4 NPCMJ における係り受け表示の例 (wikipedia.RepoMAN)

4 は全て同じ文であるが, 対象とする述語に対して係り元を同定して表示している。例えば 1 行目は同格の「の」が対象の述語であり, これに対して, 作業者は「です」という動詞として言い換えて付与している。「です」は多義であり, その概念フレームとして frame.id が 895 番 (コンピュータの場合の概念フレーム) を選択している。コンピュータの概念フレームは基本的に下記の様な構文である。

(19) [Arg1:対象 この本は] [Arg2:補語相当 (は) 彼のもの] です
つまり Arg2 は補語相当であり, Arg1 を説明している内容であることを示している (この本=彼のもの)。これと対応させて「青年」の意味役割は「対象」かつ Arg1 であり, 「パンク」は「補語相当 (は)」かつ Arg2 である。

3 行目の事例では「解雇」が対象とする述語として示されているが, 後続する「され」という受動態を表す助動詞まで含めて述語として意味役割を付与する。つまり「解雇される」として係り元の意味役割を決定する。能動態の「解雇する」を含む概念フレーム 241 番「引退」の例は下記の通りである。

(20) [Arg0:動作主 店が] [Arg1:対象 パートを] 解雇する

(21) [Arg0:動作主 彼が] [Arg2:起点 大学を] 退学する

また概念フレーム「引退」の必須項は下記の通りである。

Arg0:動作主, Arg1:引退する人, Arg2:引退する元

これらを基に意味役割を付与する。基本的に, 能動態の

「解雇する」と同様に意味役割を付与するため大きく変わらない。上記の例文では解雇されたのは「オットー」であり、意味役割は「対象(人)」かつ Arg1 である。勤めていたところが「スーパーマーケット」であり、「起点」かつ Arg2 である。

付与作業の手順として、作業者は事例をもとに概念フレーム辞書を調べ、付与できない場合、1) 適切な述語の事例が辞書に無い場合は、事例不足を共通データシートに報告する、2) 述語の事例はあるがどう付与するか自信がない場合にはその事例についても共通データシートに報告する。管理者(本著者)は共通データシートから、事例の登録が必要なものは辞書に登録する。同時に、作業者が自信のない事例については直接、作業者と打ち合わせて事例を解決する。

述語シソーラスに登録されている語彙数は約 1.1 万語である。未登録の述語が作業員から報告されているが、図 4 に示したように、対応する述語がない場合でも、言い換えできる述語と概念フレームが存在する場合、置きかえて付与を継続している。例えば「参画する」という述語が付与対象であるが、概念フレーム辞書に語彙項目が無い場合、同様の意味で言い換えられる述語と概念フレームを探す。この場合は「加わる」(概念フレーム 242「所属」)に書き換えて付与する。付与の仕方は「の」に対して「です」を付与した方法と同様に記述する。つまり辞書には掲載がなくても、NPCMJ の付与の中で出現した述語に対して、概念フレームと意味役割を付与することが可能である。これにより、アノテーションされたテキストによって辞書項目が拡張されることになる。

4.3 付与作業の現状と課題

現在、3 人の作業員で意見交換しながら付与を進めている。疑問点に対して管理者は作業員と議論して解消している。現段階で約 3 万文に対して 1 度目の付与を実施しており、見直しを進めている段階である。名詞の項構造 [15] など概念フレーム辞書にない語彙の追加が現在の課題である。

5. まとめ

本稿では日本語の述語と係り元に対して PropBank 形式の意味役割を導入した意味役割付与の枠組を提案し、NPCMJ に対して意味役割付与コーパスを構築する手法について記述した。係り元に対して付加詞の意味役割、および必須項に対する意味役割について記述し、現状について明らかにした。付与した結果は NPCMJ を通して公開する予定である。

謝辞 本研究は国立国語研究所の共同研究プロジェクト「NPCMJ コーパスに対する日本語意味役割分析」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] Banarescu, L., Bonial, C., Cai, S., Georgescu, M., Grif-fitt, K., Hermjakob, U., Knight, K., Koehn, P., Palmer, M. and Schneider, N.: Abstract Meaning Representation for Sembanking, *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pp. 178–186 (2013).
- [2] Bonial, C., Babko-Malaya, O., Choi, J. D., Hwang, J. and Palmer, M.: PropBank Annotation Guidelines Version 3.0, *Proceedings of the Joint Conference on EMNLP and CoNLL: Shared Task*, pp. 1–40 (2012).
- [3] Butler, A.: Treebank Semantics, Technical report, Hirosaki University (2019). (<http://www.compling.jp/ajb129/tsdoc.html> accessed 2019/11/30).
- [4] Grimshaw, J.: *Argument Structure*, MIT Press (1990).
- [5] Horn, S. W., Nagasaki, I., Butler, A. and Yoshimoto, K.: *Annotation Manual for the NPCMJ*, National Institute of Japanese Language and Linguistics (2019).
- [6] Levin, B.: *English Verb Classes and Alternations*, University of Chicago Press (1993).
- [7] Levin, B. and Hovav, M. R.: *Unaccusativity*, MIT Press (1995).
- [8] Maekawa, K., Yamazaki, M., Ogiso, T., Maruyama, T., Ogura, H., Kashino, W., Koiso, H., Yamaguchi, M., Tanaka, M. and Den, Y.: Balanced Corpus of Contemporary Written Japanese, *Language Resources and Evaluation*, Vol. 48, pp. 345–371 (2014).
- [9] Pinker, S.: *Learnability and Cognition*, MIT Press (1989).
- [10] Wang, C. and Xue, N.: Getting the Most out of AMR Parsing, *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, pp. 1257–1268 (2017).
- [11] Wierzbicka, A.: 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).
- [12] Zhang, S., Ma, X., Duh, K. and Durme, B. V.: AMR Parsing as Sequence-to-Graph Transduction, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (2019).
- [13] Zhou, J. and Xu, W.: End-to-end Learning of Semantic Role Labeling Using Recurrent Neural Networks, *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics* Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, pp. 1127–1137 (2015).
- [14] 吉本 啓, 周 振, 小菅智也, 大友瑠璃子, Butler, A.: 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).
- [15] 竹内孔一: 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 回コーパス日本語学ワークショップ予稿集, pp. 51–56 (2014).
- [16] 橋田浩一: GDA 意味的修飾に基づく多用途の知的コンテンツ, 人工知能学会論文誌, Vol. 13, No. 4, pp. 528–535 (1998).
- [17] 影山太郎: 動詞意味論, くろしお出版 (1996).
- [18] 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).
- [19] 日本電子化辞書研究所: EDR 電子化辞書使用説明書 (第 2 版) (1995).