

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

竹内 孔^{1,a)} Alastair Batler² 長崎 郁³ Prashant Pardeshi⁴

概要：国立国語研究所では日本語のテキストに対して統語・意味解析情報付き現代日本語コーパス (NPCMJ) の構築を進めており、2019年3月の段階で約3万文の構文木データを付与している。本研究では NPCMJ で構築されている構文木に対して、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するプロジェクトをスタートした。意味役割と概念フレームを付与する際に、シソーラスに不足している事例がある場合には、同時にシソーラスも更新しながら付与を進める。これにより、事例だけでなく、概念フレームと意味役割の辞書を提供する。本発表では、アノテーションの枠組と体制、意味役割および概念フレーム付与における NPCMJ での特徴的な点を取り上げて事例を説明するとともに、今後の見通しについて説明する。

Annotating NPCMJ with Semantic Role Labels and Frames of Predicate Thesaurus

1. はじめに

国立国語研究所では日本語のテキストデータに対して、Penn Historical Corpora のアノテーション [2] を参考に構文木情報を付与する NPCMJ (NINJAL Parsed Corpus of Modern Japanese) の構築をすすめている [1]。付与した木構造を検索できるツールを開発し、Web サイト上でユーザがツールを利用して文法規則を利用した詳細な事例検索ができるようにしている^{*1}。これにより言語学研究者および日本語学習者が文法的なパターンを考慮した事例検索ができる。対象とするテキストデータおよび構文木情報が無償でダウンロード可能であり、ユーザが手元で加工することが可能である。よって、言語学研究だけでなく、自然言語処理での利用も期待される。

2019年3月現在で約3万文の構文木を構築しており、今

後3年にわたって毎年1万文の構文木を追加する予定である。新聞記事など許諾の必要な文ではなく、公開を許されている文を利用しているため、Web サイト上でテキストおよび構文木情報が確認できる点が利点である。また、コーパス単位で付与されているため、文の前後につながりがある。よって文間関係を分析する事例として利用することも期待できる。

図1に「..長女の愛は厳しく育てたが..」の NPCMJ の構文木の例を示す。構文木情報とともに痕跡や直接係り

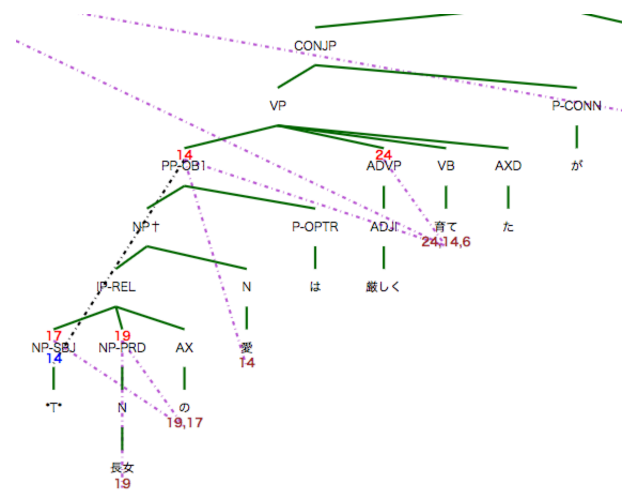


図1 「妹の真理」の部分の NPCMJ の構文木

¹ 岡山大学
Okayama University, 3-1-1, Tsushimanaka, Kita-ku, Okayama, 700-8530, Japan

² 弘前大学
Hirosaki University

³ 名古屋大学
Nagoya University

⁴ 国立国語研究所
National Institute for Japanese Language and Linguistics

a) takeuc-k@okayama-u.ac.jp

*1 <http://npcmj.ninjal.ac.jp/>

関係のない項との関係についても情報が付与されていることが分かる。

本研究ではNPCMJの構文木情報に対して述語項構造シソーラスに基づく意味役割と概念フレームを付与する。意味役割とは言語学で提案されている述語と係り関係にある係り先との意味的な関係を示すタグであり、動作主 (Agent) や対象 (Theme) といった各述語に直接関与する要素 (必須項) から、時間、場所、条件のような修飾的な要素 (付加詞) まで提案されている [3], [5], [7]。述語項構造シソーラスでは述語の語義を細分化するのではなく、WordNet と同様に共通特性の概念として5階層のシソーラス状にまとめ、各概念に対して、ある述語のある語義に対する例文を付加している。各例文の係り先に対して意味役割を付与しているため、述語項構造シソーラスは述語に対する概念フレームと意味役割の辞書データ (リポジトリ) として利用できる。概念フレームと意味役割を NPCMJ に付与することで、下記のような異なる助詞表現や述語を検索することが期待できる。

- 神戸港を:起点 出港する:起点からの移動
- 神戸港から:起点 出港する:起点からの移動
- 神戸港を:起点 出る:起点からの移動
- 神戸港から:起点 出る:起点からの移動

ここでコロン「:」記号の右側が意味役割および簡略化した概念フレームを示している。

NPCMJ に対する意味役割付与では述語項構造シソーラスの体系と同時に PropBank スタイルの数字による意味役割も付与する [4]。構文木情報には*speaker*や*pro*といったゼロ代名詞および痕跡 (*T*) が付与されており、述語に対する項もしくは係り関係にある要素として意味役割を付与する。述語項構造シソーラスでは BCCWJ に対して、出現した項についてのみ意味役割と概念フレームを付与した BCCWJ-PT を構築してきている [6]。しかしながら、ゼロ代名詞などを記述していなかったため、今回の研究で初めてゼロ照応を含めた形で意味役割を付与することになる。よって意味役割が付与されたコーパスは Penn Treebank に意味役割を付与した PropBank に近い形式を持つことになる。

本稿では意味役割および概念フレームの付与枠組について記述する。さらに NPCMJ の例文に付与する中で現れた特徴的な事例について取り上げ、今後の見通しについて記述する。

2. 意味役割と概念フレームの付与作業のフレームワークと体制

本研究プロジェクトでは単に NPCMJ の例文に対して意味役割と概念フレームを付与するだけでなく、述語項構造シソーラスに事例が不足している場合は、シソーラスを拡張 (および修正) する。これにより、事例データとシソー

ラスの同時構築を目指す。よって、作業フレームワークとして例文付与作業とシソーラス更新の2つの作業が発生する。それぞれについて説明する。

NPCMJ への付与作業

全体の流れを図2に示す。まず NPCMJ から付与対象とする述語と係り先の組合せを取り出す。NPCMJ は構文解析済みであるため、構文情報を元にプログラムで付与すべき述語と係り先を抽出する。次に、作業者が述語項構造シソーラスの検索システム*2を利用して付与対象の述語の事例を検索する。検索結果が出た場合は、事例の語義と付与対象文の意味が一致しているかどうかを確認し、一致している場合は事例に付与されている概念フレームを付与し、シソーラスの事例を元に意味役割を付与する。事例の意味役割が合っていない場合は、意味役割の全種類の一覧*2から候補を探し、その事例をシソーラスの検索システムで確認して、最も近い意味役割を付与する。

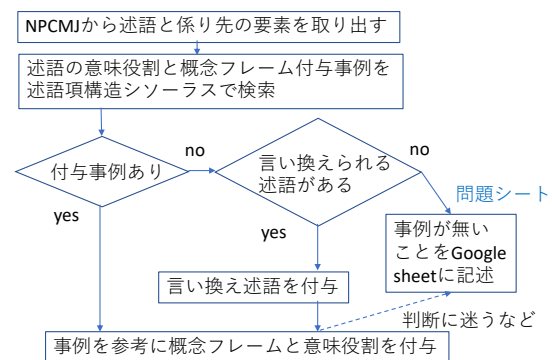


図2 NPCMJ に対する付与作業の流れ

一方、述語が見つからない場合でも、シソーラス中にある別の述語で言い換え可能な場合がある (例えば 3.1 節参照)。その場合、言い換えた述語の事例で概念フレームと意味役割を付与し、注釈として言い換えた述語を記録する。

また、述語が辞書で見つかったも、付与対象文の語義が合っていない場合は作業員間で共通して問題を記述する問題シート (Google sheet) に内容を記述する。また上記の意味役割付与の時にタグについて迷った場合も同様に問題シートに記述する。

述語項構造シソーラスの更新作業

管理者 (第一著者) は問題シートの内容を確認し、述語項構造シソーラスに必要な事例を追加する。また、付与作業で記述の曖昧性などが見つかった場合は辞書を修正して更新する。

作業体制

NPCMJ への付与作業として複数人の作業員を雇用し、付与を続けている。作業員は岡山大学の研究室だけでなく、外部の組織に属する作業員も含まれている。それゆえ、付

*2 <http://pth.cl.cs.okayama-u.ac.jp>

与対象文データや問題シート，シソーラスは全て，ネットワークを介して，どこの組織に所属しても付与できるように構築している．また，管理者の作業として意味役割や概念フレーム選択の際に作業者が迷った事例に対して，一貫した付与方針で説明する必要が生じる．問題シートに対して，回答を書くと同時に同じ研究室の作業員には直接指示し，他の組織の作業員には，インターネットを利用した会議を開き，作業員間で情報が共有されるように配慮する．

3. NPCMJ に対する付与での特徴的な構造

NPCMJ では能動態を中心とした辞書の例文と異なり，話し言葉や書き言葉の中にうめこまれた表現に対して述語項構造を付与する．本節ではその中で特徴的な例を取り上げて，現段階でどのように付与しているか記述する．

3.1 格助詞の「の」

NPCMJ では連体修飾の「の」について述語の場合に意味役割と概念フレームを付与する．例えば misc コーパスにおいて，「.. 妹の真理は..」という例文がある．この部分の木の構造を取り上げると 図3 のように，「妹の」が「真

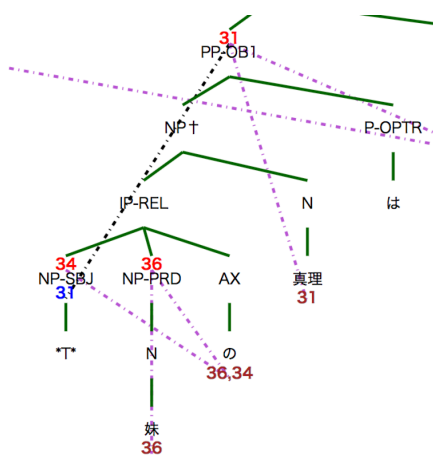


図3 「妹の真理」の部分の NPCMJ の構文木

理」に対する連体修飾節として付与されている．日本語の「の」は文脈に応じて複数の意味を取ることが指摘されている [8] が，ここでは，「妹である真理」と言い換えることができる同格表現である．つまり，「です/である」の述語を用いた連体修飾節である．よってこの場合に「です」による意味役割を付与する．

述語項構造シソーラスではコピュラ文を作成する「です/である」に対する概念フレームと意味役割付与事例が登録されていなかった．そこで，辞書を拡張して「です」に対してコピュラ概念フレームを設定し，意味役割として，能動文に戻して下記のように付与する．

真理は:Arg1_対象 妹:Arg2_補語相当(は)です:
コピュラ (895)

ここで，「です」の概念フレームの()内の数字は概念フレーム ID を示す．

付与対象文では能動態ではなく，連体修飾で現れており，かつ述語に相当する部分は「の」であるため，上記の結果をそのまま付与すると，どの述語に対する概念フレームの付与が明らかでない．そこで，文の表現に合わせる際に，述語の対象となる「の」に対して概念フレーム ID の他にシソーラス上でどの述語と解釈して付与しているか同時に併記してタグを付与する．この場合は「です」として付与するので結局，図4 に示すようなアノテーションとなる．



図4 「妹の真理」の部分の意味役割と概念フレームを付与した事例

3.2 ゼロ代名詞も含めた付与

NPCMJ ではゼロ代名詞が付与されているため，直接係り関係にない項であっても，ゼロ代名詞を介して述語の項に対して意味役割を付与することができる．例えば wikipedia コーパスの「.. 勤めていたスーパーマーケット..」について構文木は図5 のようになる*3．図5 の*T*

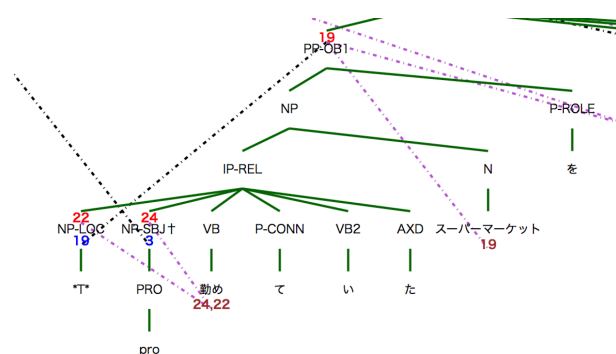


図5 「勤めていたスーパーマーケット」の部分の NPCMJ の構文木

はスーパーマーケットを表しており，pro は「勤める」の主語であることを示している．図には描かれていないが，pro は「オットー」であることが構文木情報で記されている．よって「勤める」に対する概念フレームと意味役割は能動文で記すと下記のようなになる．

pro は:Arg0_動作主 スーパーマーケットに:Arg2_着点 勤める:習慣的移動動作 (422)

これを付与対象文の連体修飾の形に直して付与した結果，図6 に示すアノテーションとなる．

*3 全体の文はゼロ代名詞や痕跡を入れて表すと「*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。」である．

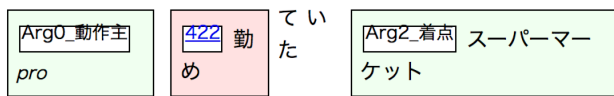


図 6 「勤めていたスーパーマーケット」の部分の意味役割と概念フレームを付与した事例

4. 現状と今後の展開

2019年の4月からスタートさせた現在の付与状況を表1に示す。表1に示すとおり、現在約4000箇所の述語に対する意味役割と概念フレームの付与を行っている。作業者は言語学の学生3名、文学部など文系出身の作業者が3名である。作業者のスタート時期と1日の作業時間が全く異なるため、平均の付与速度は出していない。また1回目の付与であり、見直しなどを行う予定である。

表 1 NPCMJ に対する意味役割と概念フレーム付与状況

コーパスの種類	コーパスの数	付与した事例数
book	51	1594
diet	1	158
fiction	2	488
misc	8	84
news	16	107
nonfiction	4	423
ted	5	819
wiki	11	568
合計		4241

4月から言語学出身でない作業者を中心に構文木と意味役割、概念フレームについてほぼ毎日、意見交換しながら作業を進めてきた。初期に比べて付与が安定しているように見受けられる。現状の付与量は文数に直すと約1800文程度である。この作業は基本的に今年度行う予定であり、付与精度を確認しながらデータ量を拡張していく予定である。

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）の助成を受けたものである。

参考文献

- [1] Stephen Wright Horn, Iku Nagasaki, Alastair Butler, and Kei Yoshimoto. *Annotation Manual for the NPCMJ*. National Institute of Japanese Language and Linguistics, 2019.
- [2] Anthony Kroch, Beatrice Santorini, and Ariel Diertani. The Penn-Helsinki Parsed Corpus of Modern British English (PPCMBE). Technical report, Department of Linguistics, University of Pennsylvania, 2010.
- [3] 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.

- [4] 竹内孔一, バトラーアラスティア, 長崎郁, ホーン スティーブンライト. 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. 言語処理学会第25回年次大会, pp. 136–138, 2019.
- [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] 上野真幸, 竹内孔一. 動詞語義及び意味役割付与作業システムの構築. 第2回日本語コーパスワークショップ, pp. 69–76, 2012.
- [7] 影山太郎. 動詞意味論. くろしお出版, 1996.
- [8] 西山佑司(編). 名詞句の世界. ひつじ書房, 2013.

正誤表

下記の箇所に誤りがございました。お詫びして訂正いたします。

訂正箇所	誤	正
1 ページ 3 行目	Alastair Batler	Alastair Butler