

# 常識推論エンジン開発に向けた 事態性名詞のアノテーション基準について

米本 悠 藤田 響平 永田 基樹 森 遼太 小代 義行  
株式会社 pluszero

{yonemoto.haruka, fujita.kyohei, nagata.motoki,  
mori.ryota, ojiro.yoshiyuki}@plus-zero.co.jp

## 概要

自然言語の高度な意味理解を伴うシステムとして有効な汎用的常識推論エンジンを構築するには事象の状態変化を正確に捉えることが重要である。以前提案した常識推論エンジン [1] においては、述語に関連する状態変化を捉えることができること、また、それらが常識推論に有効であることを示した。しかしながら、状態変化を捉えるためには述語だけでは不十分で、動詞と同等の役割を果たす名詞（事態性名詞）も扱う必要があり、これらは我々のエンジンでは対応できていない。本稿では、常識推論エンジンで事態性名詞を扱うにあたり既存コーパスで見られたアノテーションの不足について明らかにし、それらを補うようなアノテーション基準を提案する。

## 1 はじめに

機械が自然言語の高度な意味理解を行うためには、文章中に明示されている情報以外に、常識を用いて推論をする必要がある。例えば、「リビングに新聞がある」「リサが新聞をとってキッチンへ歩いて行った」という2つの文があるとき、新聞がどこからどこへ移動したかを理解するには文章中の表現だけでは足りず、人間が解釈する際に用いるような常識、この場合、「手に取って歩いていくと手に取ったものも一緒に移動する」という常識を必要とする。我々は以前、このような常識推論の問題に取り組むため述語項構造ソーラスと Event Calculus (EC) を用いた汎用的常識推論エンジンを構築し提案を行った [1]。

この推論エンジンは、大きく下記のステップに分かれている。

1. 事前処理として常識を意味表現に変換しておく

2. 自然言語文に対して意味役割付与を行う
3. 付与された意味役割を意味表現に変換する
4. 自然言語文から変換した意味表現と常識から変換した意味表現を用いて推論を行う

このうち、2では日本語意味役割付与システム (ASA) [2] を用いた。これは、自然文に対して述語項構造解析を行い表層格の項を同定し、係り受け関係にある項に意味役割を付与するものである。この推論エンジンにより、述語に関連する状態変化を捉えることができること、また、それらが常識推論に有効であることを示した。しかしながら、常識推論に必要な状態変化を捉えるためには、述語以外にも、動詞と同等の役割を果たす名詞（事態性名詞）についても捉える必要があり、提案した常識推論エンジンではこれらの表現に対応できていない。我々は事態性名詞に関しても意味役割を付与することで、より汎用的な常識推論エンジンを構築することを目指している。

日本語に対する比較的規模の大きい事態性名詞のアノテーションデータに NAIST テキストコーパス [3] がある。事態性名詞と考えられる名詞に、表層格であるガ・ヲ・ニ格の項が付与されている。また、名詞の中には文脈上事態を現しているのか、物を現しているのかを判断することが難しい名詞があり、それらをどちらとしたかを示すためのラベルも付与されている。事態性名詞に着目したデータセットは少なく、NAIST テキストコーパスは貴重な言語資源であるが、常識推論エンジンの開発を目的として利用する際には、下記のような問題がある。

- 作業員によりアノテーションにずれがある

著者らによっても指摘されていることであるが [3]、作業員間でアノテーションが一致しない問題がある。これは、事態性名詞の基準が明文化されておらず、ある単語に事態性があると

捉える基準が人によって異なることが1つの原因だと推測される。

- 項をとらない状態変化を表す名詞を捉えられない

NAIST テキストコーパスにおける事態性名詞は表層格の項をとることを前提としている。そのため、項をとることはないが一般的に事態性があると想起させるような単語、例えば、雨、運動会といった単語は事態性名詞の対象とならない。

- 状態変化を伴わない名詞が捉えられてしまう

常識推論エンジンでは、状態変化を捉えることで高度な文章理解を行うが、NAIST テキストコーパスでは状態変化を伴わないような名詞にもアノテーションが付与されている。1つは、複合名詞を構成素に分解したときに前方に位置する構成素を事態性名詞としている場合である。例えば、NAIST テキストコーパスでは「消費税」という複合名詞の「消費」を事態性名詞として捉えているが、我々は、このような名詞は後方の名詞を修飾するだけと考え、事態性名詞としてみなさないこととした。もう1つは、文脈上状態変化の情報を持たないような単語を事態性名詞としている場合である。例えば、「この街の暮らしにも慣れた」「自由貿易は暮らしの質を高める手段」という文中における「暮らし」という単語は、NAIST テキストコーパスでは両者とも事態性名詞として捉えているが、我々は、前者においては、具体的な事象であり状態変化の情報を持つと解釈し、後者は一般的な事柄について述べているだけで具体性がなく、状態変化の情報を持たないとして事態性名詞の対象から除外することとした。

1つ目の問題点に取り組むため、事態性があるとみなせる基準を可能な限り明文化し、事態性名詞のアノテーション基準をフローチャート化した。2つ目の問題点に対しては、項をとらず、なおかつ事態を表す名詞の特徴を明文化し基準を作成した。3つ目の問題点に対しては、複合名詞の取り扱いについての基準に加え、実際に状態変化を伴う表現かどうかに関する基準の作成を行った。

## 2 判定基準詳細

### 2.1 作成方針

判定基準は、作業者が判断しやすいようにフローチャート形式で作成した。また、他のアプリケーションで利用する際に基準の追加や削除を行いやすいよう、目的ごとに基準をまとめ下記3つのブロックとした。

1. 事態性名詞であり得ないものを除外する（除外ブロック）
2. 事態性名詞であり得るものを取得する（取得ブロック）
3. 事態性を判定をする（事態性判定ブロック）

除外ブロックは、過剰なアノテーションを避けるための基準、取得ブロックは事態性名詞の候補となる単語を選び出すための基準、事態性判定ブロックは候補に対して状態変化の情報を持つかを判断するための基準をそれぞれまとめたものである。

### 2.2 全体像

図1は、事態性名詞を判断するためのフローチャートである。丸四角で表現されているのは個々の判断基準であり、四角で表されているのは付与されるラベルである。矢印は、各基準に該当した(Yes)、該当しない(No)場合に次に進む方向を示している。色付きの四角は第2.1章で述べたブロックを示しており、右上に該当する名前を記載している。以下の章では、各ブロックに含まれる基準の詳細を説明する。

### 2.3 事態性名詞であり得ないものを除外する

#### 2.3.1 複合名詞の前方にある

NAIST テキストコーパスでは複合名詞の場合、分解してそれぞれの構成素に対して事態性の判別を実施していたが、文脈上状態変化を伴わない名詞が数多くアノテーションされていた。そのため複合名詞の前方の語は後方の語を修飾するだけと考え、事態性を考慮する対象としては除外した。例えば、「消費税」という単語の場合、前方の「消費税」に注目している場合、この単語は事態性の候補外、後方の「税」に注目している場合、この単語は事態性名詞の候補とし、後段の基準でさらに判定を続ける。

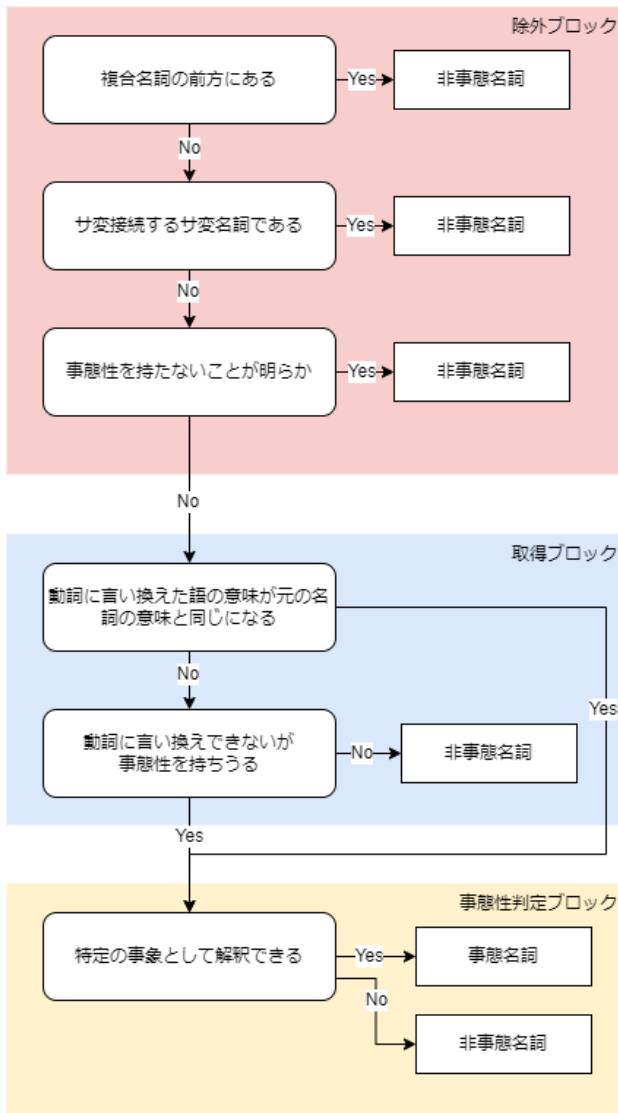


図1 事態性名詞の判定基準

### 2.3.2 サ変接続するサ変名詞である

サ変名詞はその性質上、事態性名詞の有力な候補となるが、サ変接続する場合には接続語も含めて述語として捉えることができるため、事態性名詞としては除外する。例えば、「家庭での学習が基本だ」の「学習」は事態性名詞の候補となるが、「各家庭で学習する」のようにサ変接続する場合には事態性名詞の候補から除外する。

### 2.3.3 事態性を持たないことが明らか

人の名前や机、花瓶といった物理的なものなど、一般的に明らかに事態として想起しない単語を除外する。以下に例を挙げる。

- 物理的なもの・・・机，花瓶，車など
- 役割・役職・・・部長，お父さんなど
- 場所・位置・方角・・・バス停，前方，南など
- 時間・期間・・・2時，奈良時代など
- 固有名詞（地名，組織名，サービス名，人名，商品名，建物名）・・・東京都，大学など
- 形容動詞として使われている名詞・・・正常，不能など

## 2.4 事態性名詞であり得るものを取得する

### 2.4.1 動詞に言い換えた語の意味が元の名詞の意味と同じになる

動詞に言い換えられる名詞は状態変化を表現する候補として有力なため、動詞的要素を持つ名詞を事態性名詞の候補とする。名詞に動詞的な役割があるかどうかを判断するために、動詞として言い換え可能な名詞中の語を動詞として言い換え、その意味を元の名詞と比較し、同じ意味になる場合は該当するものとみなす。以下に言い換えの例を挙げる。

- サ変名詞をサ変接続する  
「被災者の苦悩は察するに余りある」→「被災者が苦悩する」
- 末尾が動詞の連用形と同形の名詞を終止形にする  
「両者がせめぎ合いを続ける」→「両者がせめぎ合う」
- 末尾が動詞の連用形と同形の名詞を、下記のルールで言い換える  
- 用言を修飾する語+動詞

「六十五歳の男性がダンスの下敷きになって」→「ダンスが下に敷く」

– 目的語+動詞

「国道176号が通行止めとなった」→  
「国道176号の通行を止める」

• 熟語を下記のルールで言い換える

– 目的語+動詞

「40代の男性が殺人の容疑で逮捕された」→「男性が人を殺す」

– 主語+動詞

「千畳敷カールで四日発生した雪崩に」  
→「雪が崩れる」

• 接尾辞をサ変接続するもしくは、動詞の終止形にする

「地球温暖化が原因で海面が上昇する」→「地球温暖化する」

#### 2.4.2 動詞に言い換えできないが事態性を持ちうる

一般的に事態として想起される名詞の中には、運動会、台風といった必ずしも動詞に言い換えできない(項をとることがない)ものも含まれる。これらの名詞は、動詞で表現されるような状態変化が複数同時に起こるような事態だと捉えることができる。例えば、火事という名詞には、「建物が燃える」、「人が逃げる」といった複数の状態変化が含まれている。以下に例を挙げる。

- 行事, 催事, 競技名・・・野球, 東京オリンピック
- 事件, 災害, 人災・・・火災, 地盤沈下, 事故
- 自然現象, 社会現象・・・雨, ラグビーブーム

## 2.5 事態性を判定をする

### 2.5.1 時間と場所の情報を含み特定の事象と解釈できる

状態変化を推論するには、対象としている事象が実際に事物の変化を起こすものであるかという具体性が重要である。そのような具体性を持つ表現を捉えるには時間や場所などの情報が重要な要素となるため、ここでは時間や場所の情報が具体的に文脈上推定できる場合、事態性を持つと判定する。例えば、「この街の暮らしにも慣れた」「自由貿易は暮らしの質を高める手段」という文中における「暮らし」と

いう単語は、前者では時間や場所の情報を含む具体的な事象としてみなし事態性を持つと判定し、後者は時間や場所などの情報を含まないとみなし事態性を持たないと判定する。

## 3 考察

本稿で作成したフローチャートは、大きく3つの部分からなる。除外ブロックは事態性名詞となり得ないものを除外することを目的とした基準の集合で、取得ブロックは事態性名詞となり得るものを選択することを目的とした基準の集合である。事態性判定ブロックは、先2つのブロックで得られた事態性名詞の候補に対し、事態性を判断することを目的とした基準の集合である。それぞれのブロックに含まれる基準は常識推論エンジンの開発に向けた判定基準であり、アプリケーションが変われば必要となる基準も異なると考えられる。そのような場合においても、ブロック毎の目的に合わせて個々の基準を組み替えることでより柔軟に対応が可能である。例えば、項をとらないような単語を事態性名詞と扱いたくない場合には、第2.4.2章の基準を除外ブロックへ移動させることで対応が可能であるし、事態性の定義を変更したいような場合には、事態性判定ブロックについてのみ基準を追加・削除することにより対応が可能である。

また、第1章で、フローチャート形式による基準の明文化には、作業間隔のアノテーションのずれを軽減する狙いがあると述べたが、実際にどの程度有効であるかは定量的な調査を行う必要がある。それらの調査で確認される揺れの傾向を把握し、基準のさらなる改定を行う予定である。

本稿では、常識推論エンジンの開発を目的としているため、状態変化の情報を含む名詞を事態性名詞としてアノテーションする基準を作成した。状態変化の情報を持つかどうかの指標として、場所や時間の表現が含まれており、具体的な事象かどうかを判断するとしたが、この基準のみで状態変化の情報をすべて捉えられるかどうかには検討の余地がある。例えば、動作主体が明示されている場合に、時間や場所の情報を必ずしも含まなくても具体事象として解釈することが自然な事例が存在する可能性がある。このような点を踏まえ、さらに基準に関して検討する必要がある。



---

## 参考文献

- [1] 杉本智紀, 藤田響平, 井坂諭紀雄, 永田基樹, 森遼太. 述語項構造ソーラスと EC を用いた汎用的常識推論エンジンの構築. 言語処理学会第 26 回年次大会, pp. 597–600, 2020.
- [2] 竹内孔一, 土山傑, 守屋将人, 森安祐樹. 類似した動作や状況を検索するための意味役割及び動詞語義付与システムの構築. 電子情報通信学会技術研究報告. NLC, 言語理解とコミュニケーション, Vol. 109, No. 390, pp. 1–6, 2010.
- [3] 飯田龍, 小町守, 乾健太郎, 松本裕治ほか. Naist テキストコーパス: 述語項構造と共参照関係のアノテーション. 情報処理学会研究報告自然言語処理 (NL), Vol. 2007, No. 7 (2007-NL-177), pp. 71–78, 2007.