

事態間関係の獲得のための動詞語釈文の構造化

青山桜子 阿部修也 大西良明 乾健太郎 松本裕治
{hanako-a, shuya-a, yoshiaki-o, inui, matsu}@is.naist.jp

奈良先端科学技術大学院大学 情報科学研究科

1 はじめに

言語処理を高度化するためには、大規模な言語知識による頑健な言語解析の上に、大量の世界知識に基づく高度な推論を実現する必要がある。そうした世界知識には、モノだけでなくコト、すなわち事態に関する上位下位関係や部分全体関係、因果関係など知識が含まれる。本稿では、そうした事態に関する基本知識の一部を国語辞典の動詞語釈文から収集する試みを報告する。

我々が、事態に関する知識を使って実現したい推論は大きく2種類ある。一つは、近年研究者の注目を集めつつあるテキスト間の含意関係を認識する推論 [4] である。例えば、次の文 (1a) が成り立つならば、文 (1b) も成り立つ。

(1) a. 村上春樹は1982年の夏に『羊をめぐる冒険』を雑誌「群像」に発表した。

b. 『羊をめぐる冒険』の著書は村上春樹である。

このようなテキスト間の含意関係を判断する推論は、質問応答や情報抽出、機械翻訳など、自然言語処理の広い範囲で重要な役割を果たすと期待されるが、これには例えば (2) のような事態間の含意関係の知識があれば実現できると考えられる。

(2) a. 「Xが著作物Yを雑誌Zに発表した」なら、それ以前に「人Xは著作物Yを書いている」はずである。

b. 「人Xが著作物Yを書いた」ならば、「人Xは著作物Yの著者である」。

もう一つの推論は、人間の行動やそのプランに関する推論である。例えば、「庭に洗濯物を干したとたんに雨が降ってきた」という文が望ましくない事態であることを推論するためには、

(3) a. 「洗濯物を干した」行為の目的が「洗濯物を乾かす」ことであって、

b. その目的が「雨」のために達成されなかった、といったことを認識する必要がある。

(3a) のような推論はプラン認識と呼ばれ、対話プランニングや人の行動支援に有用である。こうした推論を実現するには、人の行動に関する目的、手段、条件などの知識が必要と考えられる。

こうした事態間の関係は膨大な量にのぼると予想されるので、人手で網羅的に記述するのは現実的でない。そのため、大規模コーパスからこうした関係を自動獲得する研究が進んでおり、一定の成果を上げている [7, 10, 3, 1]。ただし、これらの研究はいずれもコーパスにおける述語間の共起情報に頼るものであり、知識獲得を補助する基本的な知識がなければ、Webサイズのデータを以てしても過疎性の問題から逃れ得ない。

こうした背景から本稿では、事態を表現する言語表現の基本単位として述語項構造を考え、述語項構造間の多様な含意関係を国語辞典の動詞語釈文から部分的に収集することを考える。これによって得られる知識は、上述のような推論のための基本知識になるとともに、コーパスからの知識獲得のためのシード知識として利用できることを期待する。

2 動詞語釈文の構造化

本稿では便宜的に、国語辞典の動詞見出し語が表す事態を「見出し語事態」、語釈文中に含まれる各述語項構造が表す事態を「定義語事態」と呼ぶ。また、見出し語の述語項構造を「見出し語項構造」、語釈文の各述語の項構造を「定義語項構造」と呼ぶ。また、行為者の意志的に活動する事態を「行為」と呼び、それ以外の事態を「出来事」と呼んで区別する。

2.1 概要

本稿で考える動詞語釈文の構造化は基本的には、(i) 語釈文を定義語項構造に分解し、(ii) 各見出し語項構造と定義語項構造の関係を記述という2つの作業からなる。定義語項構造の認定は、見出し語項構造の事態が定義語項構造の事態を含意するという関係が成り立つかどうかによって決める。また、見出し語項構造と定義語項構造の関係は、(a) 項同士の対応関係と (b) 見出し語事態と定義語事態の意味関係からなる。以下、2.2 で構造化作業の具体的内容について述べ、3節で意味関係の分類の詳細を述べる。

2.2 作業内容

国語辞典の動詞の各語義の語釈文について以下のような注釈付けを行う。(例-1)は「置く」の語義の語釈文から一部抜粋したもので、これに対して(例-2)のような注釈を人手で付与する。

(1) 動詞語釈文から見出し語が物か事態かを判断する。
(2) (1)で事態と判定されたものについて、語義に対応する見出し語の文型を(例-2-b)のような項のリストの形で指定する。

(3) 語釈文中の各定義語項構造について以下を行う。
(a) 定義語項構造の項と見出し語項構造の項の対応関係を記述する。たとえば(例-2-d)「重ねる」の格要素「誰かA」は見出し語「置く」の格要素と対応している。
(b) 語釈文中に格要素が出現している場合はそれを記述する。(例-3-e)「おとす」の格要素が「水・湯・薬品」であることを記述する。

(c) 定義語項構造と見出し語項構造の項の対応関係を記述する。「折る」のヲ格要素「広がっているもの」は〈何か A〉と対応する。また(例-3)の〈何か A〉と「汚れ」のように、いわば間接照応のような、「の」を介する間接的な対応関係が成り立つ場合は(例-3-e)のように記述する。

(d) 定義語事態と見出し語事態の意味関係を記述する。現状では意味関係として、上位・同義、同義・言い換え、結果状態、不可分、前提条件、付帯状況、手段、目的、反義語の9種類を仮定している。これらの関係については次節で詳述する。例えば、(例-2-d)「折る」は(例-2-b)「畳む」の〈手段〉と考える。

(4) 語義が見出し語の慣用句、機能動詞、名詞的用法を指している場合は、見出し語の格要素を指定する(例-3-c)。

(例-1)

- a. 畳む (31542-0-0-1-0)
- b. 折り返して重ねる。広がっているものを折って小さくする。「洗濯物を畳む」

(例-2)

- a. 語義 ID : 31542-0-0-1-0
- b. 見出し語 : 〈誰か A〉ガ〈何か A〉ヲ 畳む
- c. 手段 : 〈誰か A〉ガ〈何か A〉ヲ 折り返す
- d. 上位・同義 : 〈誰か A〉ガ〈何か A = もの〉ヲ 重ねる
- e. 前提条件 : 〈何か A = もの〉ガ 広がっている
- f. 手段 : 〈誰か A〉ガ〈何か A = もの〉ヲ 折る
- g. 上位・同義 : 〈誰か A〉ガ〈何か A = もの〉ヲ 《小さく》する
- h. 結果状態 : 〈何か A = もの〉ガ 小さい

(例-3)

- a. 洗う (1256-0-0-1-0)
- b. 水・湯・薬品で不要なもの特に汚れをおとす。
- c. 「足を洗う」(良くない仕事をやめて正業につく意)
- d. 見出し語 : 〈誰か A〉ガ〈何か A〉ヲ 洗う
- e. 同義・言い換え : 〈誰か A〉ガ〈何か A〉ノ 汚れヲ 水・湯・薬品デ おとす
- f. 道具 : 水・湯・薬品

意味的含意関係の詳細は次節で述べる。

3 述語項構造間の意味関係

1節で述べたように、我々の目的は、事態間の含意関係の認識や人の行動に関する推論に必要な知識の一部を、事態に関する基礎知識の一部として収集することである。そこで見出し語事態と定義語事態の意味関係を(a)両事態の時間的關係、(b)含意関係の必然性の強さ、(c)事態の意思性の区別(行為と出来事)の3つの観点から以下で述べるように分類した。

3.1 見出し語事態と同時に起こる定義語事態

見出し語事態と同時の起こる定義語事態について、定義文の主辞になっている定義語事態を主要な事態と考え、見出し語事態との間に〈上位・同義〉関係を認める。一方、そうした主要事態に対して付帯的に起こる事態を〈付帯状況〉と呼ぶ。ただし、〈上位・同義〉は「〈見出し語〉{を行う、が起こる}ことは〈定義語〉{を行う、が起こる}ことである」という文を意味的に成り立たせるような組でなければならない。同様に、〈付帯状況〉は「〈見出し語〉{を行う、が起こる}ときには

たいてい同時に〈定義語〉{を行う、が起こる}」のような言語テストをパスするものに限られる。

たとえば、次の「安息する」の語釈文(4a)に含まれる2つの定義語事態(4b)と(4c)はともに見出し語事態と同時に起こる事態である。このような場合、我々は語釈文の主辞になっている(4b)の方の見出し語事態の〈上位・同義〉と見なし、(4c)の方の事態を〈付帯状況〉と考える。

- (4) a. X ガ安息する = 何の心配・苦痛もなく静かに休む。
- b. X ニ心配・苦痛がない(上位・同義)
- c. X ガ静かに休む(付帯状況)

付帯状況のうち見出し語事態と定義語事態両方が行為の場合に、行為-手段の関係が認められる場合がある。その場合は特別に〈手段〉関係に分類する。すなわち、〈手段〉は〈付帯状況〉の特殊な場合と考える。たとえば「圧縮する」の語釈文(5a)に含まれる2つの定義語事態は(5b)の方が見出し語事態の〈上位・同義〉、(5c)は〈手段〉と考える。

- (5) a. X ガ気体や固体に圧力をかけてその体積を小さくする。
- b. X ガ体積を小さくする(上位・同義)
- c. X ガ気体や固体に圧力をかける(手段)

3.2 見出し語事態の後に起こる定義語事態

見出し語事態よりも後に起こる定義語事態の多くは、見出し語事態の結果として生じる事態、あるいは見出し語事態が行為のときにその効果として期待される事態のいずれかである。前者を〈結果(状態)〉、後者を〈目的〉に分類した。〈結果(状態)〉は見出し語事態の完了とともに必然的に成立する事態、たとえば、(6a)に含まれる2つの定義語事態は(6b)の方が見出し語事態の〈上位・同義〉、(6c)は〈結果(状態)〉と考える。

- (6) a. X ガ気体や固体に圧力をかけてその体積を小さくする。
- b. X ガ体積を小さくする(上位・同義)
- c. (気体や固体)の体積が小さい(結果状態)

一方〈目的〉は見出し語事態の行為を行っても、成立するとは限らない点異なる。たとえば、(7a)に含まれる2つの定義語事態は(7b)の方が見出し語事態の〈上位・同義〉、(7c)は〈目的〉と考える。

- (7) a. X ガ(物事がうまくゆくように)あれこれと、体や気を使う。
- b. X ガ体や気を使う(上位・同義)
- c. 物事がうまくゆく(目的)

したがって〈結果(状態)〉は「〈見出し語〉{を行う、が起こる}と、その結果として必ず〈定義語〉{を行う、が起こる}」という言語テストをパスするもの、〈目的〉は「〈定義語〉{を行う、が起こる}ために/が目的で〈見出し語〉{を行う}」のような言語テストをパスするものに限られる。

3.3 見出し語事態の前に起こる定義語事態

見出し語事態よりも前に成立している定義語事態は多くの場合、見出し語事態の成立の〈前提条件〉と見なすことができる。〈前提条件〉については、見出し語事態、定義語事態ともに行為、出来事を問わない。たとえば、(8a)に含まれる2つの定義語事態は、(8b)の方が見出し語事態の〈上位・同義〉、(8c)は〈前提条件〉と考える。

表 1: 見出し語事態と定義語事態の関係

時間関係	見出し語事態	定義語事態	関係
同時	行為	行為	手段
	行為	出来事	
	出来事	行為/出来事	付帯状況
定義語が後	行為/出来事	出来事+必然性アリ	結果状態
	行為	行為/出来事+必然性ナシ	目的
定義語が前	行為/出来事	出来事+必然性アリ	前提

表 2: 付与された関係とその割合

関係	上位・同義	結果状態	不可分	前提条件	付帯状況	手段	目的	反義語	道具	不明
個数	2840	346	18	527	256	527	147	72	39	3

- (8) a. X が隠されていることや秘密をはっきりさせる。
 b. X が Y をはっきりさせる (上位・同義)
 c. Y が隠されている (前提条件)
 つまり、〈前提条件〉は「〈見出し語〉{を行う、が起こる}ためには、その前に〈定義語事態〉が成立し{ていなければならない、ているはずだ}」のような言語テストをパスするものに限られる。

詳細は <http://cl.naist.jp/nltools/kinakochan> を参照されたい。

4 作業経過

現在「岩波国語辞典コーパス 2004」として入手可能な岩波国語辞典第 4 版には動詞 11469 語 (17104 語義) の語釈文が収録されており、我々はこれらすべてについて上述の構造化作業を行う予定である。現在、コーパスにおける出現頻度の高い動詞を対象に作業を進めており、作業の速度や安定性、仕様の問題点などが少しずつ明らかになってきた。以下、これまでに得られた知見について述べる。

4.1 作業速度と結果

現在、作業者 1 人、週 30 時間の作業時間で約 1000 語義のペースで進んでおり、上述の 1 万 7 千語義をすべて処理する時間は見直しの時間を含めても半年程度で収まる見込みである。このことから、上述の仕様は法外にコストの高い作業ではないことがわかる。

作業済みデータのうち、最新の仕様に基づいて見直した 1925 語、2756 語義について、得られた含意関係の分布を調べたところ、表 2 のようになった (定義文が理解しにくいと判断された 57 語義を除く)。〈上位・同義〉が多数を占めるのは予想されたことであるが、その他〈前提条件〉、〈手段〉、〈結果状態〉、〈付帯状況〉など、いずれも含意関係認識に有用と考えられる関係が多数得られていることがわかる。

4.2 作業結果の一致率

作業者間でどの程度揺れなく作業可能かを調査するため 2 名の作業者が 26 語 (50 語義) に対し、上述した注釈付けを行った。作業者間の一致率を次の 3 項目について評価した。

- 項目 1: 語釈文から定義語事態として選択した述語が一致するかどうか 113 個の述語中 105 個の述語が一致していた。一致していなかった 8 個は、定義文が理解しにくい場合に付与する「意味不明タグ」が 4 個、形容詞を述語的扱いにしていなかったものが 1 個、語釈文中の括弧内を無視したものが 1 個、述語の未抽出が 2 個あった。

- 項目 2: 1 で一致した定義述語のうち、見出し語構造との項の対応関係が完全に一致するかどうか 述語が一致した 105 個のうち項の選択が一致したものは 98 個あった。一致していなかった 7 個のうち 2 個がミスで、5 個が作業者間で何を項とするかの認識の違いだった。
- 項目 3: 1 で一致した定義述語のうち、見出し語構造との意味関係が一致するかどうか 述語が一致した 105 個のうち関係の選択が異なったものは 2 個あった。1 つは〈付帯状況〉と〈手段〉の 2 つの事態が行為か出来事かの曖昧な間違い、もう 1 つは見出し語「楽しむ」の定義文「楽しいと感ずることをして喜ぶ」の「楽しむ - 楽しい」の関係を一人は〈結果状態〉を選択し、もう一人は〈前提条件〉を選択する間違いだった。

以上のように、現在の作業仕様に関する限り、項の対応関係、意味関係ともに十分な一致率を見ている。ただし、意味関係の一致率が高かった理由の一つは、現在の関係分類が〈上位〉と〈同義〉を区別していないことにある。両者の明確な区別がかなり困難であることが予備調査から分かっていたためである。しかしながら、実際の含意関係認識においては両者の区別が必要になる場合も出てくると予想されるので、これについては今後の課題としたい。

4.3 問題点と今後の対応

作業者間である程度一致して作業ができる枠組みになってはいるが、いくつかの問題点が明らかになった。第 1 に、見出し語項構造と語釈文項構造との対応関係の記述が難しい例が存在する。たとえば、「(進学を) 希望する」の語釈文は「こうなればよい、なってほしいと願うこと」とあるが、見出し語の項を埋める「進学」のような事態名詞はこの語釈文中のどの述語の項にも対応しない。あえて言えば、「進学する」という事態の成就と「こうなる」が対応するが、こうした対応関係を首尾一貫した表現で記述するのは容易ではない。このように語釈文が補文を持つ場合の扱いについて今後検討を進める必要がある。

また、語釈文中に出現する動詞で、辞書中に定義されていない語が存在する。そのような語の多くは、辞書に定義されていない表記が語釈文中で使用されていることにある。これらの表記の揺れに関しても人手で統一的な表記にする必要がある。

さらに、次の例のように、見出し語が登録されていない動詞の出現も見られる。特に複合語に関しては、定

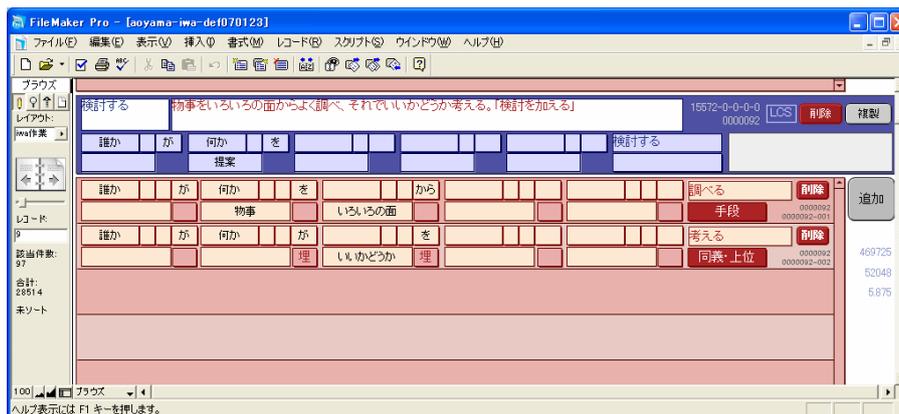


図 1: 作業用インターフェース

義されている語と定義されていない語が混在している。以下に定義されていない語の例を示す。

- (9) a. 動詞 + 動詞 尽きはてて、なくなる(枯渇)
 b. 動詞 + 補助的な動詞 同じ宿舎に 泊りこむ こと(合宿), からみ合って、入り乱れる(纏れる)
 c. 名詞 + 動詞による複合語 形式ばら ないで話し合うこと(座談)

構造化を行う上で、これらの非定義語を現在の定義語に留めることが必要である。(a)の動詞 + 動詞型に関しては分解可能な複合語もあるが、(b)(c)の動詞 + 補助的な動詞型に関しては見出し語が双方を含意していることは少ない。さらに、(d)のように名詞 + 動詞による複合語に関しては、分解することができないものもある。今後はこれらの取り扱いをどのようにするか、定式化していくことが必要となる。また、今後の展望としては構造化した動詞語釈文を再帰的に利用することで、語釈文に出現した事態間の関係だけでなく、より多くの事態間の関係を獲得したいと考えている。

5 関連研究

辞書の語釈文から述語間の関係を抽出する研究は、MindNet[9] や鍛治ら [8] の言い換えモデルなど、すでにいくつか報告がある。こうした研究はいずれも語釈文を自動解析し、全自動で知識を抽出することを試みている点では我々の試みに比べて有利である。ただし、例えば MindNet では動詞間の関係はある程度分類できているものの、項の対応関係までは捉えられていない。鍛治らの手法は項の対応づけにアプローチしているが、扱っている意味関係は言い換え(または上位)に限定されている。

述語間の意味関係を人手で記述した代表的な資源に WordNet[5] と FrameNet[2] がある。WordNet には本稿で述べたような分類に近い意味関係が記述されている。FrameNet はフレーム間にも限定的ではあるが同様の意味関係が記述されている。ただし、WordNet には項の対応関係の記述は一切なく、FrameNet でもフレーム間をまたぐ述語間については項の対応関係は原則として特定されていない。

これら既存の資源に対し、我々が作成中の資源は、述語項構造の対を関係の単位として項の対応関係を記述

するとともに、従来よりも詳細に意味関係を分類している点に資源としての価値があると考えている。

6 おわりに

本稿では、述語項構造で表現される事態間の含意関係を (a) 述語が持つ項の対応関係、(b) 関係のクラス(上位・同義、前提条件、結果状態、目的、手段など)の情報とともに国語辞典の動詞語釈文から人手で抽出する試みについて報告した。これにより、事態間の語彙的含意関係について基本的な知識が得らるとともに、手段、前提条件といった推論的含意関係の一部を収集できると期待される。得られた知識は、今後、コーパスからの事態間関係獲得 [1] に利用していく予定である。開発中の資源はいずれも順次公開することを予定している。詳細は <http://cl.naist.jp/nltools/kinakochan> を参照されたい。

参考文献

- [1] 阿部修也, 乾健太郎, 松本裕治. 事態含意名詞の利用と共起パターンの学習による事態間関係知識の獲得. 言語処理学会第 13 回年次大会, 2007.
- [2] Collin F. Baker, Charles J. Fillmore, and John B. Lowe. The berkeley framenet project. In *Proceedings of COLING-ACL*, pp. 86–90, 1998.
- [3] Timothy Chklovski and Patrick Pantel. Verbocean: Mining the web for fine-grained semantic verb relations. In *Proceedings of EMNLP*, pp. 33–40, 2004.
- [4] Ido Dagan, Oren Glickman, and Bernardo Magnini. The pascal recognising textual entailment challenge. In *Proceedings of Pascal Challenge Workshop on Recognising Textual Entailment*, 2005.
- [5] Christiane Fellbaum, editor. *WordNet-An Electronic Lexical Database*. The MIT press, 1998.
- [6] 乾孝司, 乾健太郎, 松本裕治. 接続標識「ため」に基づく文書集合からの因果関係知識の自動獲得. 情報処理学会論文誌, Vol. 45, No. 3, pp. 919–933, 2004.
- [7] Takashi Inui, Kentaro Inui, and Yuji Matsumoto. Acquiring causal knowledge from text using the connective marker tame. *ACM Transactions on Asian Language Information Processing (TALIP)*, Vol. 4, No. 4, pp. 435–474, 2005.
- [8] 鍛治伸裕, 河原大輔, 黒橋禎夫. 国語辞典とコーパスを用いた用言の言い換え規則の学習. 言語処理学会 第 8 回年次大会, 2002.
- [9] 鈴木久美, Gary Kacmarcik, Lucy Vanderwende, Arul Menezes. Mindnet/mnex:意味関係データベースの自動構築と解析のためのツール. 言語処理学会 第 11 回年次大会 発表論文集, 2005.
- [10] Kentaro Torisawa. Acquiring inference rules with temporal constraints by using japanese coordinated sentences and noun-verb co-occurrences. In *Proceedings of HLT-NAACL*, 2006.