

動詞語義を推定するための語義付与コーパスの作成

竹内 孔一, 下村拓也
岡山大学大学院自然科学研究科
koichi@cl.cs.okayama-u.ac.jp

1 はじめに

動詞の語義の推定ならびに動詞間関係を整理することで項構造レベルでの表現で包含される意味関係をうまく処理したい。例えば「太郎が A 社に就職する」ならば「太郎は A 社に所属する」ことが推測される。こうした推論を行うための基礎情報として動詞の辞書を何を記述して、どのように推論で行うかは確立されていないが、先行研究 [3] の分析から下記の 2 つの問題を解くことがまず重要であると考えられる。(1) 動詞の意味はそれが直接的な関係をもつ名詞 (項) との組み合わせで大きく変わることから項構造を表層関係とともに記録しつつ、ある表現が他の動詞と同じ概念を共有するかを分類して記述しておく必要がある。例えば、「彼は子供を抱える」は「子供を抱く」という動詞と共通する概念を持つと考えられるが「会社がパートを抱える」は「パートを雇用する」という動詞と共通するであろう。このような項構造ベースで動詞間の概念を整理して基本データとして記録しておく必要がある。(2) もう一つは動詞の語義あいまい性解消である。これはさらに次の 2 つの問題に分解できる。1 つは項に存在する名詞の違いによるあいまい性解消 (例えば上記の「子供/パートを抱える」での「抱える」の違い)。もう一つは動詞の概念 (語義) が持つ潜在的な項と表層との対応関係である。例えば

- 会社が太郎を雇う
- 会社がアルバイトを雇う

前者では、「太郎が会社に所属すること」が含意されるが後者では、「誰かがアルバイトとして会社に所属すること」が含意される。つまり表層格と動詞と語義が全く同じでも名詞によって解釈が異なり、言い換え関係も異なってくる。つまり、動詞の語義あいまい性解消というタスクは単なる動詞の分類を当てる問題では収まらず、項の意味分類 (以降、意味役割と呼ぶ) まで解析できなければ、より意味に即した言語処理としては十分ではない¹。

上記 (1) に関してはこれはずでに、先行研究において 4425 語の動詞について人手による分析を行い、項構造に基づく動詞間関係を概念体系とともに整理したデータを構築している [3]。しかしながら (2) の問題

は動詞と名詞の語義、ならびに意味役割の組み合わせの問題であるため意味役割が付与された事例が十分に無ければモデル化することが難しい。

そこで本研究では意味役割と動詞、名詞の語義を付与したタグ付コーパスを作成することで、こうした動詞の語義曖昧性解消に対するメカニズムを明らかにするための基礎データを構築することを目標とする。意味役割は先行研究 [9] [1][6][2][10] [5][4] で様々な視点から議論されているが、付与可能でかつ言語処理として利用可能という観点から意味役割集合とその付与の問題点は整理されていない。本稿では大規模コーパス構築に向けた予備実験として意味役割と動詞、名詞の語義を 96 語の動詞について付与を行い問題点を整理した。その結果意味役割集合の付与方針を明らかにするとともに、意味役割付与対象は格助詞の関係だけでなく付加詞 (「6 時に」など) も考慮する必要があることを明らかにする。

2 語義付与コーパスの構成

語義付与コーパス構築の主な目的は動詞の項構造と表層表現との対応事例の収集および語義付与モデルのための基本データを提供することである。しかし単に表層と項構造を結ぶだけでなく項構造に対応した語義を体系化した辞書として整理することで、他の動詞との言い換え関係や含意関係の推論まで行うことが求められるべきである。こうした考察から、作成する語義付与コーパスは下記のような構成をとる必要があると考えられる。

- (1) コーパスに付与する意味役割集合と語義を整理した動詞辞書を用意し動詞間の項の関係を記述する
- (2) 語義ごとの例文を複数用意する
- (3) 体系化された名詞の語義を付与する

(1) の辞書に関して本研究では竹内らの動詞辞書 [3] を利用する。他の機械処理用の辞書として EDR [10] が存在するが EDR は各単語の語義に対して項構造が付与されていないため利用できない。また IPAL [5] は動詞の語義に対応して格フレームと意味役割を記述しているが意味役割の粒度が荒いこと (20 程度)、分析対象語数が約 2000 語と上記の動詞辞書より少ないことから利用を見送っている。(2) に関して語義付与データとして RWCP コーパス [7] があるが文書に付与され

¹動詞だけでなく名詞はさらに重要である。「社長が太郎を雇う」の場合、「太郎は (社長の属する会社) に所属する」ことが推測されるがここでの「会社」名詞の意味の中に埋め込まれている。

ているため、1つの語義に対して多くの事例を獲得することが難しい。よって新聞記事を対象に人手で語義を調べて事例を集める必要がある。(3)に関しては分類語彙表 [12] と日本語語彙大系 [13] を利用する。しかしどのような名詞の語義が動詞の語義に関与するかはまだ明らかではない。後の 4.2.4 節では具体的な付与例から現状の名詞の語義の問題点について考察する。

語義付与コーパスに関連する研究として英語では FrameNet があげられる。FrameNet では語義を Frame という概念で記述し、語義ごとに意味役割まで付与したコーパスを作成し公開している。一方、日本語に対しては国立国語研究所が中心になって進めている日本語コーパス [8] プロジェクトで約 100 万語の文書に対して語義を付与することが始められている。

3 意味役割集合の構築方針

意味役割の定義集合はさまざまな提案が先行研究でなされている。例えば VerbNet では意味役割は 31 種類であるが FrameNet では 746 種類仮定している。これは設計の違いであり、FrameNet では語義の単位である Frame(884 種類) に対して意味役割を独立に構築し、異なる Frame の意味役割の対応は別立てのリンク関係で記述する。一方、VerbNet では最初から動詞間を横断的に意味役割を付与している。また、黒田ら [4] は意味役割を 1 文に対して多重に付与することで様々な文の解釈の仕方に対する意味役割を提案している。

これに対して本研究では過剰な意味役割数は押さえて、動詞どうしの項の関係をラベルとして記述することを目指す。なぜならば意味役割を利用して処理を行う観点から複数のラベルは扱いが煩雑になることが予想されるためである。さらに、個別に複数のラベルを立てると意味役割付与において一貫した基準で付与することが難しいと考えられることが理由である。

そこで意味役割の機能から整理して付与方針を考察する。大きく分けて、意味役割には 2 つの側面があると考えられる。一つは動詞の概念を現わす要素としてどれだけの組み合わせがあるのかを記述することである。これが記述できれば省略された動詞の項の補完が可能になる。もう一つの側面は動詞間の項の対応付けを明確にすることである。例えば「寒気が北から降りる」と「漁船が港から出発する」「降りる」と「出発する」は動詞の意味としては異なるがカラ格は動作の [起点] ² として考えられるだろう。こうした意味役割の機能から下記の 2 つの方針を立てることで意味役割のラベル数の参考にする。

(1) 一文の意味役割の種類は重ならないようにする

(2) 同じ要素に複数の意味役割がつく数をなるべく少なくする

(1) と (2) は相反した方針である。つまり (1) を詳細に記述していくと意味役割の種類が増加するが、(2) の方針によって意味役割の数を制限する。ここで (2) は全く 0 になることは無い。例えば [動作主] と [起点] という意味役割を仮定すれば一つの要素に多重に付与されることが起こる。(「彼が [動作主/起点] 彼女にプレゼントを渡す」) よって (1) と (2) の方針に沿うことで適度な意味役割集合が得られると考えられる。

4 語義付与コーパス作成における問題点の整理

竹内らの動詞辞書 [3]³には動詞の語義と暫定的に構築された意味役割集合 (87 種類) がある。出発点としてこの語義と意味役割ラベルを新聞記事コーパスに対して付与する。これと同時に名詞の意味役割も人手で付与する。どのように例文を集めたか、またこれらの作業を通してどのような問題に直面したかを記述して考察する。

4.1 基礎付与データの作成

意味役割の付与には係り受け解析された文が必要である。そこで人手で係り受け情報が修正された京都大学コーパス⁴を利用する。例文を決定するには 2 つのプロセスが必要である。一つ目はコーパスと辞書を参照して頻度が高く、かつ語義が多すぎない動詞を選択する。ここでは語義が 2 つ以上 3 つ以下のものを選ぶ。二つ目は各語義に対して複数例文を集めることである。これは人間が文書を読まなければ判断できず時間のかかる作業である。この際、新聞記事ベースであるため、ほとんど同じ語義が使用される傾向が強く幅広く複数の例文を獲得するのは容易な作業ではなかった。

また付与対象の例文は (a) 基本的に能動態を選択する、(b) 取り立て助詞の「は」「も」や「彼の運転」といった助詞「の」による関係、「運転した車」のような外の関係に対しては例文の対象として除外する。これは現段階では意味役割の体系の構築を中心に作業を進めるため標準的な文を集めることが理由である。

以上のような制約の中で京都大学コーパスから例文を獲得した。その結果 96 種類の動詞について例文が 1107 文であった⁵。しかしながら 1 動詞に対して 1 語

³<http://cl.cs.okayama-u.ac.jp/rsc/lex/>

⁴<http://nlp.kuee.kyoto-u.ac.jp/nl-resource/corpus.html>

⁵大学院生 5 人で 2ヶ月の作業であった。

²以降、意味役割のラベルは □ で示す。

義しか例文が集まらなかった。この点においては、(1) 新聞記事以外のコーパスの利用を考えること、(2) 抽出は半自動で獲得するなど工夫が必要であることを次回の課題にしたい。特に(1)は大規模な日本語書き言葉コーパスが構築されているため利用を検討したい [8]⁶。

4.2 付与における問題点と整理

前節で得られた例文に対して語義の付与と意味役割の付与を行った。そこで主な問題としてあがった問題について以下に取り上げて整理する。

4.2.1 ガ格に関する整理

表層格のガ格は意味役割として [使役主], [動作主], [経験主], [原因] [対象] が考えられている。これらの違いをどのように考えるべきであろうか。以下例を基に説明する。

- (1) 彼が [使役主] 太郎を笑わせる
- (2) 雨が [原因] かさをぬらす
- (3) 彼が [動作主] 太郎を雇う
- (4) 彼が [経験主] 悲しむ
- (5) 列車が [対象] 到着する

[使役主] はある主体に対して動作を行わせるものであり、意志性があるものとする。「命令する」「お願いする」のガ格も同様である。一方で [動作主] も同様であるが操作対象に対して直接ある操作を行うものとする。よって [動作主] に対しては [使役主] をさらに付与することが可能である。(例えば「社長が彼に太郎を雇わせる」)。[経験主] は意志性はあるものの不本意な状況に陥る主体であり、[対象] は意志性が無いものである。

4.2.2 意味役割の重なり

意味役割の重なりについて例をあげて説明する。

- (6) 砲撃主が [動作主] 大砲から [起点] 弾を発射する
- (7) 大砲が [動作主/起点] 弾を発射する
- (8) 大砲で [道具] 弾を発射する

例(6)では「大砲」は [起点] と考えられるが例(7)では動作を行う主体のように表現されている。しかし同時に「発射する」という概念を考えると(7)の場合ガ格が [起点] でもある。よってこのような場合には意味役割を多重に付与することにする。また、意味役割を付与する際には表層格も考慮して付与を行う。例えば(8)に示すように「大砲」は「発射する」に対しては [道具] とも考えられる。しかしそれはデ格との

結びつきではじめて指示されるものであり、ガ格に対して [道具] の意味役割まで多重に付与することは行わない。

4.2.3 対象とする助詞の範囲

利用する動詞辞書 [3] では意味役割は必須項のみに限定して付与することが試みられてきた。動詞の意味分類の構築という目的からは誤りではないが動詞の多義性解消という観点からは下記の事例に示すように付加詞も考慮する必要がある。

(9) 4年間 [期間], 学校に [場所] 行く

(10) 6時に [時間点] 学校に [組織] 行く

例(9)では「通う」の意味であるのに対して(10)では「移動する」の意味である。これらの違いは時間副詞の表現の仕方で異なることから動詞の概念として付加詞に対して意味役割を仮定する必要がある⁷。

今回の付与作業を通して付加詞に対する意味役割を以下のように仮定する。

- Time (時間, 期間は Time-Line, 時間点は Time-Point)
- Reason (理由・原因, e.g. 雨で試合に負ける)
- Purpose (目的, e.g. 子供のために絵本を買う)
- Location (起点、着点に含まれない場所 e.g. 学校でサッカーをする 基本的に助詞「で」を伴う)
- Citation (引用、基本的に助詞「と」で引用される)
- Instrument (道具, e.g. バットでボールを打つ)
- Scene (状況、時間と場所が掛け合わせ e.g. 夏祭り、ワールドカップ)
- Premise (前提, e.g. 基に、基盤として、仮に)
- As (用途, e.g. 叔父を仲介者として会った この場合 叔父=仲介者 ある人・物の用途を示す場合に使う)
- Limit (制限, e.g. 50km/h で巡航する)
- Manner (マナー, e.g. 順番に)
- Around (対象の周辺, e.g. 将来について考える 子供のことを考える)
- Modificand (被修飾要素, 連体修飾として使われる動詞に限定)

これらの修飾要素に対する意味役割の設定はEDR電子化辞書で定義されている。そこでEDRの「事象・事実表現のための関係子」と比較すると良く対応しており、上記のうちPremise, As, Aroundに対応する関係子が無かった、逆にEDRではcooccurrence, sequence, and (連結関係), or (選択関係)といった我々が設定し

⁶<http://www.tokuteicorpus.jp/>

⁷名詞の意味役割の異なりについては文献 [11] を参照した。

ていない関係子も見受けられたが動詞の語義判別という観点からはこれらは特に必要ないと考えられる。

4.2.4 名詞の語義付与の問題

分類語彙表と日本語語彙大系を基に名詞の語義付与を行った。そこで問題となったのは付与単位と有効な語義とのずれである。

(11) 二人が川崎に移動する

上記の事例の「二人」は直感的判断からすれば『人間』⁸に関連する名詞の語義を付与すべきであろう。しかしながら、日本語語彙大系では「二」は『数』であり、「人」は接尾辞であるので『単位』となってしまう人に関する情報を付与することができない。後の動詞語義あいまい性解消モデルの構築の際には名詞の語義は自動的な計算で求まる必要があることから「二人」を ad hoc に『人』と付与するのは正しくない。現在では、これは構文解析の結果に即して『単位』として付与しているが名詞の単位と概念とのマッチングについては別に考察する必要があると考えられる。

5 まとめ

本稿では動詞の語義を推定するための基本データの構築として、動詞の語義、名詞の語義、項構造の意味役割を付与した語義付与コーパスの構築を行った。96動詞について各語義ごとに例文を京都大学コーパスから集めた。獲得した1107例文を基に先行研究の動詞辞書 [3] を参考に意味役割付与のセットに対する考察、ならびに名詞の語義付与に関する考察を行った。その結果、付与可能であり、かつ言語処理に有効と考えられる意味役割の粒度についてある一定の基準で評価できることを明らかにした。さらに意味役割について定義を見直しを行い、付与範囲として付加詞まで付与する必要があることが明らかになった。

現段階では語義の統計量などを出せる規模では無いが、今回明らかになった問題点をふまえてコーパス構築を行いたい。付与したコーパスは Web サイト⁹で公開する予定である。

謝辞

本研究は、文科省科研費基盤研究(C)「情報集約に必要な文脈と名詞の属性を考慮した動詞の多義性解消モデルの構築」(19500122, 代表: 竹内孔一)の支援を受けている。記して深く感謝する。

参考文献

- [1] Baker, C. F., Fillmore, C. J. and Lowe, J. B.: The Berkeley FrameNet project, *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics*, pp. 86–90 (1998).
- [2] Kipper, K., Dang, H. T. and Palmer, M.: Class-Based Construction of a Verb Lexicon, *Proceedings of the Seventh National Conference of Artificial Intelligence (AAAI-2000)* (2000).
- [3] 竹内孔一, 乾健太郎, 竹内奈央, 藤田篤: 意味の包含関係に基づく動詞項構造の細分類, 言語処理学会第14回年次大会発表論文集 (2008).
- [4] 黒田航, 李在鎬, 渋谷良方, 井佐原均: 複層意味フレーム分析 (の簡略版) を使った意味役割タグづけの現状: タグづけデータから派生する言語資源の紹介を中心に, 第14回言語処理学会発表論文集 (2008).
- [5] 情報処理振興事業協会技術センター情報処理振興事業協会 技術センター: 計算機用日本語動詞辞書 IPAL (1986).
- [6] 肥塚真輔, 岡本紘幸, 斎藤博昭, 小原京子: 日本語フレームネットに基づく意味役割推定, *自然言語処理*, Vol. 14, No. 1, pp. 43–66 (2007).
- [7] 白井清昭, 柏野和佳子, 橋本三奈子, 徳永健伸, 有田英一, 井佐原均, 荻野紫穂, 小船隆一, 高橋裕信, 長尾確, 橋田浩一, 村田真樹: 岩波国語辞典を利用した語義タグ付きテキストデータベースの作成, *情報処理学会自然言語処理研究会*, No. 9, pp. 117–122 (2001).
- [8] 山崎誠, 前川喜久雄, 田中牧郎, 小椋秀樹, 柏野和佳子, 小磯花絵, 間淵洋子, 丸山岳彦, 山口昌也, 秋元祐哉, 稲益佐知子, 吉田谷幸宏: 代表性を有する現代日本語書き言葉コーパスの設計, *言語処理学会第12回年次大会発表論文集*, pp. 440–443 (2006).
- [9] チャールズ・J. フィルモア (田中春美, 船城道雄訳): *格文法の原理*, 三省堂 (1975).
- [10] 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).
- [11] 青山文啓, 橋本三奈子: 名詞の辞書記述, *情報処理学会自然言語処理研究会*, pp. 9–16 (1994). NL-104.
- [12] 国立国語研究所編: 分類語彙表, 大日本図書 (2004).
- [13] 池原 悟他編: 日本語語彙大系, 岩波出版 (1997).

⁸名詞の語義は『』で表現する。

⁹<http://cl.cs.okayama-u.ac.jp/rsc/lex/role/>