

### 3.3 述語項構造解析

松林 優一郎 (東北大学)

#### 述語項構造解析とは

述語項構造は、文章内の述語とその項との関係を表す構造である。たとえば次の文、

(1) [太郎<sub>ガ</sub>]は[手紙<sub>ヲ</sub>]を書く。

では、「書く」という表現が述語であり、「太郎」と「手紙」がこの述語の項である。項には述語に対する意味機能を表すラベルが付与される。上の例では、主語を表す「ガ格」や目的語を表す「ヲ格」が採用されている。単語と単語の関係解析する類似の技術として構文解析があるが、構文解析が単語同士の文法構造上の主従関係を解析するのに対して、述語項構造は述語（出来事や状態を表現する単語）を中心として各単語の文章中での意味的な役割関係を表す。文法上の主従関係に基づく解析では、文脈上繰り返す必要がなく主語や目的語を省略する場合などで、意味的に関係のある単語間で直接の主従関係が失われてしまう。一方、述語項構造は出来事や状態に関係する単語を述語と直接結びつけるため、複雑な文や文章構造を持った文章においても「誰が、何を、どうした」といった文章理解に重要な情報が明示的に表現される。このため、述語項構造解析は機械翻訳、情報抽出、要約などの複雑な文構造・意味関係を扱う必要のある処理において有効な基礎解析処理とされている (図-1)。

#### 解析手法

図-2 に最も基本的な解析の手順を示す。解析は(1)述語の特定、(2)項候補の枝刈り、(3)項の特定、(4)役割の分類という手順で行われる。(1)では、各単語に二値分類器をかけ、述語か否かを判定する。(2)以降では、(1)で特定された各述語に対して、探索範囲内の各単語がその述語の項となるかを



図-1 構文解析と述語項構造解析。述語項構造解析では「次郎」と「踊る」の関係が明示される。

入力文: 太郎 が 次郎 に タンゴ を 踊る よう 勤めた

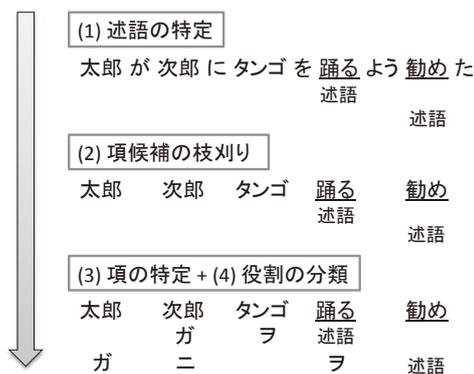


図-2 述語項構造解析の基本的な手順

判定し、項となる場合はその役割を分類する<sup>☆1</sup>。(2)では、高い適合率を達成するルールを用いて、項となる可能性がきわめて低い単語を枝刈りする。一般には、通常述語の項となり得ない品詞の単語を候補から除外する。(3)の項の特定では、二値分類器を用いて各単語がその述語の項となるかを判別する。(4)の役割の分類では、(3)で項と判別された単語について、多値分類器を用いて述語に対する役割を分類する。項の特定・分類のための特徴量としてこれまでに数十種類のもものが検討されてきたが、そのうち特に重要なものとして、「統語パス」と「述語と項の共起に関する情報」がある。統語パスは、文中

☆1 一般に項は句や節を形成しているが、解析時は便宜上「主辞」と呼ばれる句の中で文法・意味的に最も重要な一語（日本語の場合は句や節の最後の単語）を項の範囲として解析する。

の単語間にある文法上の主従関係のうち、項候補と述語の間を結ぶ主従関係の系列を取り出したもので、特定の文法構造が項とその役割を決めることを捉えるために用いられる。述語と項の共起情報は、たとえば「食べる」のヲ格（目的語）には食べ物がきやすいといった、項になりやすい単語の種類を捉えるための情報で、表層文字列や単語クラスターの共起に関する統計量が使われる。分類器の学習は一般に人手により述語項構造が注釈付けられたコーパスに基づく教師つき学習で行われるが、近年では大規模データからの教師なし学習の手法も研究されている。

### 課題点と分析・展望

述語項構造解析における当面の課題は、省略された項の補完にある(図-1の「踊る」に対して「次郎ガ」を補う処理)。既存の解析器では述語と項の間に文法上の直接的な主従関係がある比較的容易な事例においては適合率と再現率の調和平均で90%弱と高い性能が得られているものの、省略を伴う事例においては、省略された項が同一文内に出現している場合で50%弱、文をまたいで補わなければならない場合は20%前後と低い水準にとどまっている。この現状は日本語の省略解析の難易度の高さを物語っているが、一方で、日本語において省略の現象は全述語-項ペアの約40%を占めるため、きわめて重要な課題となっている。

省略された項が同一文内に出現している事例に焦点を当て詳細に分析した結果からは、省略された項Aを含む述語Pと、Aと直接的に係り受け関係を持つ述語Dとの間でAを項として共有している事例が項省略事例全体の58%を占めることが分かっており、項の共有関係の特定が重要な手がかりとなることが示唆されている。PとDの間に並列構造や機能動詞構文などの構文構造上の強い手がかりが

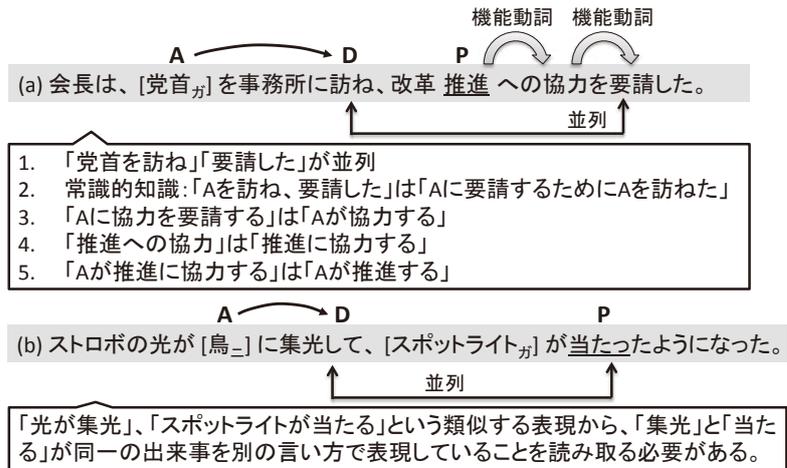


図-3 解析対象述語Pに対して省略された項Aを見つけるための手がかりの一例。DはAと直接係り受け関係にある語で、解析の起点となる。

ある場合は、既存の解析器においても先の調和平均で65%前後と相対的に高い性能を示しており、項の共有に関する一部の特徴はとらえられていると考えられる。一方で、PとDの間には、図-3(a)のように並列構造や機能動詞構文などの局所的パターンの多重な重ね合わせによって最終的な関係が導かれる複雑な事例が一定数存在することも分かっている。そのほか、図-3(a)の手がかり2や(b)の例のように文脈や世界知識に基づく推論が必要なものも多く存在している。

加えて、述語項構造解析の文脈ではあまり取り組まれていない世界知識や文脈を読み解く必要がある事例や、その他のいまだ一般化されていない雑多な手がかりを用いる事例の解析は低い性能にとどまっており、省略解析で高精度を実現するためには、世界知識を用いた推論や談話解析などの技術を取り込むか、もしくはそのような後段の処理につなげるための適切な問題設定やインタフェースを今後模索していく必要がある。

(2015年10月31日受付)

松林 優一郎 (正会員) y-matsu@ecei.tohoku.ac.jp  
 2010年東京大学大学院情報理工学系研究科博士課程修了。博士(情報理工学)。同年より国立情報学研究所・特任研究員。2012年より東北大学大学院情報科学研究科・特任助教。意味解析の研究に従事。言語処理学会、人工知能学会、ACL各会員。