

複文を対象とした自然言語文の拡張型述語論理への自動変換法

上條 敦史[†] 石川 勉[‡]拓殖大学大学院工学研究科[†]拓殖大学工学部情報工学科[‡]

1 はじめに

最近の Web 情報の爆発的な増加,普及に伴い,電子化情報に対する高度な意味処理の適用が求められ,文章を一定の形式の知識に変換する必要性が出てきている。

我々は自然言語文の述語知識化について研究している。この研究ではこれまで単文を対象とし,一般的な平叙文の他に,受身文や丁寧文等も変換してきた[1][2]。さらに,“ある”や“する”等の多くの意味を有する特殊動詞を含む文についても変換してきた[3][4]。このような単文において文法的に正しく,省略のない文については約 90%程度の変換精度を実現している。

本報告では,新たに複文を対象とした述語知識への変換法について提案する。

2 知識表現法

2.1 基本的な知識表現法

ここでは,一つの文を以下のような一つの述語式で表す。

$$L = s_1, s_2, \dots, s_n P(r_1: t_1, r_2: t_2, \dots, r_n: t_n) \quad (1)$$

ここで, P は述語部であり,基本的にはその文の述部を構成する単語(動詞,形容詞,名詞のいずれか)とする。 P が,動詞,形容詞の場合は終止形を,名詞の場合(述部が「～である」の場合)はそのまま用いる。 $r_i: t_i$ はラベル付き引数(項)であり, t_i が引数本体, r_i がそれと述語の関係を表すラベルである。引数本体は基本的には,その文の述部と関連する名詞または名詞句である。ラベル r_i は述語部により異なり,動詞の場合は agt (主格), obj (対象格), plc (場所格)等の深層格(主に EDR 辞書[5]で用いられるラベルを想定)を,形容詞や名詞の場合は sbj , inst のような新たに設定したラベルを用いる[6]。また, s_i は文の態や様相を表す識別子であり,過去(*),否定(!),受身(u)等とする。

2.2 複文の知識表現法

複文は,名詞節,連体節,連用節,並列節に分類される従属節と主節よりなる。従属節が名詞節,連体節,連用節(条件文を除く)の場合は文全体を一つの述語式で表現する。すなわち,(1)式で主節を表し,従属節は同式中の P や t に埋め込んで表現する。具体的には従属節を L' ,それにより修飾される語を a としたとき $[a][L']$ という形で埋め込む。ここで L' は(1)式の L と構造上は基本的に同一である。ただし,(1)式が真偽を表すのに対し, L' は事象やアクションを表す。なお, L' では L の述語部 P とその a の間に格関係が存在する場合には該当するラベルの引数部に記号“#”を付与する。

また,並列節や連用節(条件文)の場合は,一つの述語式では表現できないため,主節と従属節をそれぞれ独立の文と

A Translating Method of Natural Language Sentences to Extended Predicate Logic Formula for Complex Sentences.

[†] Atsushi Kamijo, Graduate School of Electronic Information Engineering, Takushoku University.

[‡] Tsutomu Ishikawa, Electronic Information Engineering, Takushoku University.

して扱い,節毎に述語式を用いて表し,それぞれ連言($\wedge$)や含意($\Rightarrow$)の論理記号を用いて連結する。ただし,条件文の場合,文が一般的な概念でなく固有名詞など個体に関する文の場合はラベル cnd (条件)で表現する。以下に表現例を示す。

- ① 刑事は犯人が隠れている家を探している
 $\Rightarrow$ 探す(agt : 刑事, obj : [家]{隠れる(agt : 犯人, gol : #)})
- ② 花子は太郎が犯人であることを知っている
 $\Rightarrow$ 知る(agt : 花子, obj : [こと]{犯人(inst : 太郎)})
- ③ 花子は太郎が経営している会社の社員である
 $\Rightarrow$ [[会社]{経営する(agt : 太郎, obj : #)}]の社員(inst : 花子)]
- ④ 太郎は花子が帰宅するときに会社へ行く
 $\Rightarrow$ 行く(agt : 太郎, gol : 会社, tme : [とき]{帰宅する(agt : 花子, tme : #)})
- ⑤ 太郎は本屋に行って本を買った
 $\Rightarrow$ *行く(agt : 太郎, gol : 本屋) $\wedge$ *買う(agt : 太郎, obj : 本)
- ⑥ 秋が来れば葉が落ちる
 $\Rightarrow$ 来る(obj : 秋) $\rightarrow$ 落ちる(obj : 葉)
- ⑦ 太郎が来るならば次郎は来ない
 $\Rightarrow$!来る(agt : 次郎, cnd : [条件]{来る(agt : 太郎)})

①は従属節が連体節の場合である。連体節は,被修飾名詞と連体節の述語との間に何らかの格関係が存在する。この場合,節述語と被修飾名詞との間に格関係があるためそこに#を付している。②は名詞節の場合であり,名詞節は基本的には上記の格関係が存在しないため節述語内に#は付さない。③は従属節が述語式の引数部でなく述語部で表現される例であり,④,⑤はそれぞれ条件文以外の連用節,並列節の場合である。また,⑥,⑦は連用節(条件文)の場合であり,⑥は一般的な概念について述べられているため,普遍的な事実と捉え,含意を用いて表している例である。⑦は特定の個体について述べられているため,ラベル cnd を用い主節に埋め込む形式で表現する。これらの判定には,品詞情報やシンソーラス[7]のカテゴリ情報を用いて行っている。具体的には,品詞が固有名詞である場合や動詞の動作主が主体となりうる特定の個体であるかをチェックし,判定する。

3 知識変換法

ここでは,自然言語文を前章で述べた述語論理に変換する方法について説明する。

まず,自然言語文を形態素・構文解析(それぞれ茶釜[8], 南瓜[9] を利用)し,単語の品詞情報,文節の係り受け,格助詞,構文番号などを獲得する。これらから複文か単文かを判断し,文が複文であった場合は,単文に分割し,それぞれに述語知識化を行い,結果を連結する。

単文の変換は,先の情報と EDR 辞書(動詞共起副辞書と単語辞書)を基に必須格を定める。任意格については,格助詞とその直前の名詞のカテゴリ情報を基に決定する。これについては文献[10]を参照されたい。図 1 に,“太郎は五時に自宅に帰った”を例とした述語知識生成までの流れを示す。

