

語尾変化・カテゴリ情報を利用した受身・使役・丁寧文の述語知識への変換法

上條 敦史†
拓殖大学大学院工学研究科‡

石川 勉‡
拓殖大学工学部情報工学科‡

1 はじめに

我々は、自然言語インターフェースのコンサルティングシステムや Web 情報等の電子化文書を利用した質問応答システムについて研究している。ここでは、自然言語文を一定の知識表現に変換し、それをを用いて推論することが不可欠となる。これまで平叙文の単文・複文に対して自然言語文の述語知識化について研究してきた [1]。本報告では、平叙文以外の文態 (単文) の述語知識への変換法について提案する。

2 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。

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

ここで、 P は述語部であり、その文の主節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 [2] 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 [3]。また、 s_i は文の態や様相を表す識別子であり、過去 (*), 受身 (u), 使役 (s), 丁寧 (t) 等とする。すなわち、ここでは平叙文以外は従来の一階述語論理における述語にこれら識別子を付加することで表現する。

この場合、受身文の扱いが問題となる。受身文は基本的には格の置き換えにより能動文に変換可能となるが、ここでは文の主題を重視することとし、受身文のまま扱う。すなわち、これらの位置を置換せず識別子を付加して表現する。以下に具体的な表現例を示す。

“太郎は次郎に蜜柑を食べられた”
⇒ *u 食べる (agt: 太郎, sou: 次郎, obj: 蜜柑)
“先生が学生に本を読ませる”
⇒ s 読む (agt: 先生, gol: 学生, obj: 本)
“社長がフランス料理を召し上がる。”
⇒ t 食べる (agt: 社長, obj: フランス料理)

3 判別法

文型の判定には、まず、形態素、構文解析 (それぞれ茶筌 [4], 南瓜 [5] を利用) を行い、主動詞 (主節の動詞), 助詞および動詞の後に付く「れる」「せる」など

の語尾を獲得する。この語尾を平叙文との差異とする。次に、獲得した主動詞に対して EDR 辞書 (動詞共起副辞書および概念辞書) を利用し、その格情報 (助詞とそれに付属する名詞のカテゴリ情報) を取得する。これらの情報からそれぞれの文型の判定を行う。なお、文によっては後述するように態を確定できない場合があるが、この場合には可能性のある態の数だけ述語式を生成することとする。

3.1 受身文

文末に語尾「れる」が付く文が受身文に該当する。ただし、「れる」には四つの用法がある [6]。各意味と例文を以下に示す。

[受身]: 太郎は次郎に蜜柑を食べられた
[可能]: この実は食べられる
[自発]: 吉報が待たれる
[尊敬]: 天皇は式典に参加された

ここで、自発については受身文として扱うものとする。これは推論においては識別子が同一なことを前提に同一ラベルごとに照合がとられるため、ラベルが正確に獲得されていれば問題ないからである。この例では“u 待つ (obj: 吉報)”と表され、何が“待たれる”かが明確である。また、尊敬については 4 節で述べるように識別する。

従って、受身、可能の識別が重要となるが、ここでは以下の場合に受身と限定する。

- i) 「に、から、より、によって」の助詞が付く名詞が動作主 (主動詞の動作主体) である場合
- ii) 「は」格および「が」格がない場合
 - i) には、先の表現例が該当する。この例では、“次郎”が主動詞“食べる”の動作主となっている。
 - ii) では、一般的に他の態である可能性が低く、ここでは受身文とする。“太郎に財布を盗まれた”や“車に轢かれた”などが該当する。

これらに該当しない場合、例えば“太郎は蜜柑を食べられた”のような文は、「に」格が省略された受身文とも考えられるし、単純に可能文とも考えられる。すなわち、受身文と可能文の識別は上記以外は極めて困難といえる。従って、ここでは両方の述語式を生成する。

3.2 使役文

使役文には次の特性がある。

- i) 動詞に語尾「せる」が付随
- ii) 動詞自体に使役の意味が含まれる
 - i) では、「を、に」格の助詞が付く名詞が動作主である場合、その文を使役文とする。動作主の判定は前節と同様に行う。

ii) では、“芸人は客を笑わす”のような文が該当する。この文を形態素解析すると文の動詞は“笑わす”となり、それ自体に使役の意味が含まれている。この場合、述語式は“笑わす (agt: 芸人, gol: 客)”となるが、本手法では、2 章のような述語式を想定しているため、

Transforming method from passive, causative and polite sentences to predicate knowledge using inflection and category of words

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

‡ Tsutomu Ishikawa, Electronic Information Engineering, Takushoku University.

“s 笑う (agt:芸人,gol:客)”と表現したい。そこで、この種の動詞を登録しておき、終止形に置換することとした。具体的には、形態素解析の際に使用される辞書 (Ipadic) 中の「～す」で終わる動詞 2335 語からこの種の動詞 107 語を抜き出し対処した。また、“泣かせる”など「～せる」で終わる動詞 1052 語から使役の意味が含まれている 41 語を抜き出し同様に対処した。

3.3 使役受身文

使役受身文では「される」や「せられる」という語尾が付く。この語尾は「れる」を含んでいるため、受身文と同じ特性がある。すなわち、この文態には 3.1 節でも述べたように使役受身文と使役可能文の両方が考えられる。例えば、“太郎は花子に牡蠣を食べさせられる”等の文である。従って、以下の場合のように識別可能な文を除き、基本的には両者に対する述語式を生成する。

[使役受身文]

文の主動詞が自動詞でかつ目的格が「に」格のとき。例えば“太郎は花子に走らせられる”のような文。

[使役可能文]

文の主動詞が自動詞でかつ目的格が「を」格のとき。例えば、“太郎は花子を走らせられる”のような文。

3.4 丁寧文

丁寧文は次のような特性を持つ。

- i) 語尾「れる」が動詞に付随
- ii) 動詞に「になる」が付随
- iii) 動詞が語源の名詞に「になる」が付随
- iv) サ変名詞に「なさる・くださる」が付随
- v) 文の主動詞がそれ自体で丁寧の意味を持つ動詞
 - i) については、3.1 節の受身文等の場合もあり、丁寧文の識別は極めて難しい。例えば、“社長は料理を食べられた”は丁寧、可能、受身のいずれとも解釈できる。従って、ここでは以下の場合のみ丁寧文に限定する。
 - a) “様”などの敬意を表す接尾語が主語の名詞に付加している場合。
 - b) 主語が動作主であり、かつ地位、職業等にあたる概念(それが名詞に付加した場合を含む)であり、かつ「を、に」格が動作主になり得ない場合。
 - a) は例えば、“山田様が行かれた”のような文、b) は 3.1 節の尊敬の例文等が該当する。
 - ii) は“先生が年賀状をお書きになる”などが該当する。この場合、形態素解析結果は「“書く(動詞)”+“に(助詞)”+“なる(動詞)”となり、助詞の前部にある動詞を主動詞とする。
 - iii) は“先生は和室でお休みになる”などが該当する。この場合、形態素解析では“休み”が名詞となる。これについては、この種の名詞を動詞に置き換えるための辞書を作成し、対処した。辞書は日本語語彙体系 [7] および IPAL 動詞辞書から名詞と動詞を抜き出し、漢字部が一致する語を対とした(先の例では、“休み”と“休む”)。獲得した対は全部で 390 対である。
 - iv) は“先輩が車を運転なさる”などが該当する。この場合、“なさる”を“する”に置換し、対処する。
 - v) は“社長が今年の抱負をおっしゃる”などが該当する。これについては、この種の動詞と基本動詞を対応させるテーブルを作り、対処した(この例では、“言う”に置換)。

4 評価

インターネット等から先に述べた語尾変化を含む単文(一部編集)を取得した。これらに対して態が一つに特定できる文を対象とし、それぞれ評価した。表 1 のように比較的良い結果が得られた。

表 1. 評価結果

文態	受身	使役	使役受身	丁寧
例文数	100	100	43	30
正解率 (%)	90	100	88.4	93.3

なお、誤りの主な原因は以下の 2 点であった。

i) 3.4 節 iii) で述べた名詞と動詞の対が不足

ii) 述語式のラベルが不的確

i) は、例えば、“先生が絵画をご覧になる”のような文で生じた。この例では、“ご覧”が一つの名詞として判断されたが、このような名詞に接頭詞が付く語を一語として登録していなかったためである。これについては辞書の拡充で対処可能である。

ii) は、動詞共起副辞書で特定できない深層格のラベル付けは、助詞とそれが付く名詞のカテゴリにより一定のアルゴリズムに基づき行っているが [1]、これがまだ不正確なためである。これについては今後改善していく予定である。

5 考察

文に複数の態が考えられる場合どの態が適切かが問題となる。例えば、3.3 節の例の場合、牡蠣を食べるのは受身文なら太郎であり、可能文なら花子となる。これらが正確に判定できないと当然推論では、誤った結論を導いてしまうことになる。しかし、3.1 節等で述べたようにこの判定は極めて難しく、文単体では原理的に不可能であり、その文が使われた状況に基づき、決定する必要がある。具体的には、前文の意味解析結果を適切にフィードバックして判定していくことが必要となるが、これについては今後の課題である。

なお、現時点では、可能文は「～できる」という表現が付随する場合が多いと考えられ、受身、可能(丁寧)の順で優先順位を付け推論に利用することを想定している。

6 まとめ

本報告では、平叙文以外の受身、可能等を表す文をその文態を表す識別子を付加した述語式に変換する手法について提案した。文態が明確に特定できる文については、提案手法は比較的高い精度で変換できた。

参考文献

- [1] 佐々木智彦, 石川勉: “連結定数で結合された素式群による複文の述語知識表現法とそれへの変換法”, FIT2004, E-017
- [2] EDR 電子化辞書: http://www2.nict.go.jp/r/r312/EDR/J_index.html
- [3] 石川勉: “日常語をベースとした順序ソート論理による知識表現法とその推論処理法”, 人工知能学会論文誌 23 巻 6 号 F(2008)
- [4] 形態素解析システム茶釜 (Chasen): <http://chasen-legacy.sourceforge.jp>
- [5] 日本語係り受け解析器南瓜 (CaboCha): <http://chasen.org/~taku/software/cabocho/>
- [6] 新村出, 編: 広辞苑第四版, 岩波書店 (1997)
- [7] 池原悟, 他: 日本語語彙体系, 岩波書店 (1997)