E-005

名詞のカテゴリ情報と格助詞の特性を用いた任意格の推定法

Determining the Optionally Case Using Meaning of Particle and Category Information of Noun

上條 敦史† 石川 勉‡
Atsushi Kamijo Tsutomu Ishikawa

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，より広範囲の自然言語文を扱うため，一般的な平叙文の他に，受身文や丁寧文等も対象としている[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].

本報告では，この研究の一環として格助詞とその前方に出現する名詞によって単語間の意味的な関係が異なる任意格の推定法を提案する．

２．基本的な知識表現法と変換法

2.1 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．

$$s_1,s_2,\cdots,s_nP(r_1:t_1,r_2:t_2,\cdots,r_n:t_n)$$

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[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(!)，使役(l)等とする．以下に具体的な表現例を示す．

①“太郎はカレーを食べた”
⇒ *食べる(agt：太郎，obj：カレー)

②“太郎はシルクでドレスを仕立てた”
⇒ *仕立てる(agt：太郎，mat：シルク，obj：ドレス)

2.2 知識変換法

入力文を，形態素・構文解析し(それぞれ，茶筌[7]，南瓜[8]を利用)，単語の品詞情報，文節の係り受け情報等を獲得する．それら情報と EDR 辞書(動詞共起副辞書および単語辞書)を基にまず必須格を定める．先の表現例①の場合，形態素解析結果から述部を特定し，その述部ごとに格助詞の照合処理を行う．動詞共起副辞書では動詞“食べる”にかかる主体(agent)には“30f6b0(人間)”，“30f6bf(動物)”等の概念関係子が記載されている．また，単語辞書では“太郎”は“30f6b0(人間)”のカテゴリに属している．これらからラベル(この例では agt)が特定される．また，“カレー”についても同様にラベルを特定する．次に，次章で述べる任意格を特定していく．

3．任意格の推定法

3.1 基本的な考え方

文に任意格が出現している場合(先の表現例では②が該当し，「で」格が任意格となる)，各助詞に対して考えうるラベル候補を設定し(表 1)，候補ごとに次の 2 つのチェックを行う．

1) 必須格との重複
2) 格助詞の前方名詞に対するカテゴリ情報との照合

1)では，EDR 辞書により既に決定した必須格と重複していないかをチェックする．先の表現例②では agt および obj が必須格で決定するので，「で」格ではこれらと重複しないかをチェックする．

2)では，深層格ごとに該当しうる範囲を日本語語彙大系[9]のシソーラス上のカテゴリ番号を基に設定し(この範囲を以下，照合範囲と呼ぶ)，チェックする．このチェックは，カテゴリ番号の照合範囲を厳格に指定したチェックとその照合範囲を拡大したチェックの 2 段階で行う．

表 1．格助詞ごとのラベル順位

格助詞	第１候補	第２候補	第３候補	第４候補	第５候補	第６候補	第７候補	第８候補
で	plc*	imp	scn	mat	cau*	tme	cnd	
に	gol	tme	plc*	scn	cau*	mat	imp	mnr
を	obj	gol	scn	plc*				
から	tfr	sou	cau*	mat				
まで	plc*	gol	tto					
へ	gol	plc*						
が	agt	tme	obj					
と	注)「と」に関しては別手法で処理する							
より	sou							
にて	plc*	imp	scn	mat	cau*	tme	cnd	
として	as							
副詞	mnr							
によって	cau*	imp						

具体的には，これらのチェックを第 1 候補から順に行い，先の 2 つの条件を満たした場合はそのラベルを付与し，そうでなければ次の候補に移る．最終候補において条件を満たさない場合は，第 1 候補に戻り 2)のカテゴリ番号の照合範囲を拡大し，再度最終候補までチェックしていく．ここで，一度目のチェックを第 1 チェック(番号が存在すればほぼ確実にそのラベルといえる)，再チェックを第 2 チェック(そのラベルの可能性が高い)と呼ぶ．第 2 チェックで，どのラベルも不適合と判断された場合は各助詞の最も頻度の高いラベルを割り当てる(最終候補が条件(cnd)の場合は強制的にそのラベルを付与する)．ただし，格助詞「として」および「より」は候補となるラベルが 1 つしかないため，前者はチェックを行わないものとし，後者および格助

†拓殖大学大学院工学研究科
‡拓殖大学工学部情報工学科

詞「と」は 3.5 節で述べるようなチェックを行う．また，ラベル mnr は品詞情報を利用してチェックし，格助詞の前方の品詞が副詞，形容詞，形容動詞語幹であった場合はこのラベルを付与する．なお，表 1 で * が付加しているラベルは 3.4 節で述べるようなチェックを行う．

3.2 第 1 チェック

前節で述べたようにカテゴリ番号の照合範囲は厳格なものとする．例えば場所格(plc)であれば，日本語語彙大系にある“場所”や“組織”等の上位概念ではなく，その下位概念にあるカテゴリを照合範囲とする．例えば“会社”などの語は組織カテゴリに含まれ場所格(plc)がラベルとして適切であるが，この組織カテゴリの下位には家庭カテゴリがあり，その中には“家”等の語が含まれている．しかし，このカテゴリには“名門”等の語も含まれており，この語は場所格として不適切である．第 1 チェックではこのようなカテゴリを照合範囲から除外している．

3.3 第 2 チェック

第 1 チェックよりカテゴリの照合範囲を広げる．前節の例では“場所”や“組織”等のカテゴリ内には“名門”等の語が含まれるため対象外としていたが，第 2 チェックではこのような語が含まれていてもそのカテゴリを照合範囲に加える．すなわち“場所”等の上位概念まで照合範囲を拡大する．

3.4 否定のチェック

表 1 で * が付加しているラベルでは格助詞の前方名詞が指定したカテゴリに含まれないかをチェックする．これは本来拾うべきでない概念番号を拾ってしまう可能性を考慮し，設けたチェックである．例えば場所格であれば主体や生き物本体(顔などの部位ではなく犬や人間などの語)に該当することはないものと考えられ，これを除外するために否定のチェックを行う．このチェックは先述した第 1 および第 2 チェックにおいて行う．

3.5 格助詞「と」と「より」の処理

格助詞「と」では，前方名詞が主体ならばラベルを agt とし(I)，そうでなければ南瓜の係り受け先の情報を用いる(II)．(I)では“太郎は次郎と野球した”などが該当し，格助詞「と」は係助詞「は」と同様に主体(agt)となる．(II)では次のような条件を設定する．

a) 「と」の前方名詞の係り先が主節の動詞でない
b) 「と」の前方名詞が副詞
c) 上記のいずれにも該当しない

a)の場合，「と」の係り先と同じラベルを割り当てる．例えば“太郎はパンとカレーを食べた(1)”の構文解析結果は図 1 のようになり，“パン”の係り先は“カレー”であるので述語知識は“*食べる(agt：太郎，obj：パン，obj：カレー)”とする．

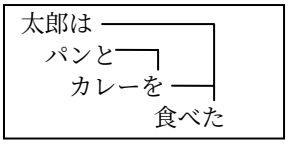

図 1．(1)の構文解析結果

b)の場合はラベル mnr を割り当てる．例えば“彼女はゆっくりと喋る”等が該当し，これは形態素解析結果の品詞情報から決定できる．

c)の場合はラベル gol を割り当てる．例えば“キャベツを塩と交換する(2)”等が該当し“交換する(obj：キャベツ，gol：塩)”と表現する．以下に(2)の構文解析結果を示す．

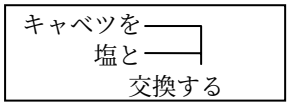

図 2．(2)の構文解析結果

格助詞「より」では，直後に形容詞が出現しかつその後に「を」格が出現している場合に比較的用法とし，そうでなければラベル sou を付与する．前者の例としては“彼は彼女より高い賃金を貰っている”などが該当し，これを“貰う(agt：彼，obj：彼女より高い賃金)”と表現する．後者の例としては“硫黄島より手紙が届く”などが該当し，“届く(obj：手紙，sou：硫黄島)”と表現する．

4．評価

インターネット等から，任意格が出現する単文 100 文(一部編集)を取得し評価した．正解率は 92%であった．

失敗した主な例文は“女子の前にて恥をかかせる”等が該当し，“l かく(agt：女子，obj：恥，mnr：前)”と変換されてしまった．これは形態素解析結果で「にて」が「に」＋「て」に分割されてしまい格助詞「にて」を正しく判別できなかったことが原因として挙げられる．

5．まとめ

本報告では，文に任意格が存在する場合のその深層格の決定を，格助詞とその前方名詞のカテゴリ情報を用いて段階的に行う手法を提案した．具体的には，助詞ごとに深層格の候補を設定し，これに対してカテゴリの照合を行う．この照合は，シソーラスの上位下位の概念に基づいた広狭 2 つの範囲を設定し，これらを照合範囲として行う．評価の結果，比較的高い精度で推定できた．

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]上條 敦史，石川 勉：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，情報処理学会第 71 回全国大会，2S-7 (2009)
[3]永田 拓，石川 勉：動詞「ある」を含む自然言語文の述語知識変換法，電子情報通信学会 (2008)
[4]上條 敦史，石川 勉：動詞「する」を含む自然言語文の述語知識変換法，人工知能学会第 23 回全国大会 (2009)
[5]EDR 電子化辞書：EDR 電子化辞書第 1.5 版，日本電子化辞書研究所 (1996)
[6]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)
[7]形態素解析システム茶筌：http://chasen-legacy.sourceforge.jp
[8]日本語係り受け解析器南瓜：http://chasen.org/~taku/software/cabocha/
[9]池原，宮崎，白井，横尾，中岩，小倉，大山，林：日本語語彙大系，岩波書店 (1997)