

意味判定が容易な述語知識表現法とそれへの自然言語文の自動変換法

野中 昌行 花田 真理 石川 勉

本稿では、対話文などの比較的簡単な自然文を対象とした、一階述語論理ベースの知識表現形式を提案する。本形式は、電子化文書等を利用して容易に変換可能な形式であり、素式を構成する述語、引数は基本的には単語をベースとする。また、一般的な文型と考えられる述部が動詞となる文を対象とした、本形式への変換法を提案する。本手法では、動詞の深層格を引数とし、これに EDR 共起辞書の概念関係子をラベルとして付加する。本手法について、ライトハウス英和辞典等の例文を用いて評価した結果、約 88% の変換率を得た。また、述語知識から自然文へ自動的に変換する手法も検討した。これについては、前述の手法で変換した述語知識を対象に実験した結果、約 99% の正解率を得た。

A Knowledge Representation Scheme for Predicate Logic and Transformation of Japanese Sentence to the Scheme

MASAYUKI NONAKA , MARI HANADA and TSUTOMU ISHIKAWA

This paper proposes a knowledge representation scheme of first-order predicate logic, into which simple Japanese sentences can be easily translated by using available electronic dictionaries and tools for natural language processing. In this scheme, predicate symbol and terms of atomic formula are basically represented by using ordinary Japanese words. Terms are labeled using case grammar for atomic formula to have only one meaning. A method to transform simple sentences into this scheme is also proposed, in which EDR dictionaries are used to decide deep cases of individual constants. An evaluation of this method using 300 example sentences shows that about 80% are correctly transformed in to predicate logical formula.

1 はじめに

我々は、不完全な知識を類似知識で補い、概略的な解を導く概略推論法 [1] の研究を行なっている。この推論法では、将来的には述語知識で表現された大量の常識知識の活用を目指しており、これらは電子化文書から自動的に獲得することを想定している。また、この推論法を問題解決機能部に利用する、自由対話システム [2] の開発を行っている。このシステムでは、ユーザーから入力された情報から事実知識を取り出し、それを利用して問題解決を行なう。従って、このシステムでは、入力文を述語知識形式に自動的に変換する必要がある。

ここでは、以上の推論法やその応用システムで重要な課題となる、自然言語文の述語知識表現法およ

び、それらの間の自動変換法について提案する。これらに関しては、自然文を意味解析し語彙知識を獲得することについては研究されている [3] が、これを論理式に変換する研究は一部 [4] を除いてあまりなされていない。本稿では、まず、多様な自然文を一意に表現することを前提として、さらに以下の条件を考慮した述語知識形式を提案する。

- i) 対話文などの比較的簡単な文を対象とする
- ii) 電子化辞書などの利用可能なツールを用いて容易に変換可能な形式とする
- iii) 現実的な推論メカニズムを前提に、一階述語論理の枠組みを越えない表現法とする
- iv) 推論を日常的な問題に適應することを想定し、知識の構成要素は単語をベースとする

次に、最も一般的な文型と考えられる動詞が述部となる場合について、自然文をこの形式に自動的に

¹ 拓殖大学工学部情報工学科
Department of Computer Science, Takushoku University

変換する手法を示す。この方法では、引数に深層格を用い、EDR 共起辞書の概念関係子をラベルとして付加する。次に、述語知識から自然文へ変換する手法について示し、最後に、これらの手法の有効性を確認するために行った実験・評価についてふれる。

2 述語知識表現法

2.1 基本的な表現形式

2.1.1 引数へのラベル付け

従来の述語論理では、1つの論理式に複数の解釈が存在する場合がある。例えば、“診察する(太郎,花子)”から、以下の2つの解釈が考えられる。

太郎が花子を診察する

花子が太郎を診察する

このため、異なった解釈で作成された知識を用いれば誤った結論を導く可能性がある。ここでは、自然文を一意に表現し、論理式を利用する人が異なっても解釈が1つに定まるようにラベルをつけることとする。すなわち、述語知識のフォーマットを

述語(ラベル₁:引数₁, ..., ラベル_n:引数_n, ...)

とする。例えば、主体を表すラベルを“agt”とし、対象を表すラベルを“obj”とした場合、先の例では“診察する(agt:太郎,obj:花子)”のように表現され、これは、上記の 2つの解釈のみを持つことになる。

2.1.2 述語と引数の表現

基本的には、述語は単語か複合語を想定する。なお、複合語の場合は、単語を“・”で連結することとする。また、引数は単語や複合語、変数、定数(スコールム定数)を想定する。複合語の場合、具体的には複数の名詞が助詞を用いず連結されている場合、以下のように構成する名詞間を“・”で連結して表現する。

例) 大型自動車 ⇒ 大型・自動車

なお、引数に一般名詞を用いる場合、その一般名詞に代表される個体(それを表す記号)を引数とする素式との関係が問題となる。例えば、“医者”が患者を治す”を述語知識表現する場合、

治す(agt:医者,obj:患者)

治す(agt:x,obj:y)∧医者(inst:x)∧患者(inst:y)の2通りが考えられる。ここでは、広範な質問へ対応可能とするため、一般名詞で扱える場合には 変数で表すこととし、一般名詞で扱いきれない場合に個体を利用することとする。なお、これらが混在している場合は、推論処理の過程で適宜対応することとする。例えば、

“治す(agt:太郎,obj:花子)”

“治す(agt:医者,obj:患者)”

を、単一化する場合、事実知識の中に、“医者(inst:太郎)”と“患者(inst:花子)”があれば、これらは同一とみなすこととする。

2.2 述部が動詞の場合

2.2.1 基本形

語と語の間の意味関係を動詞中心に捉えたフィルモアの格文法[8]の深層格を利用し、引数に格情報を持ったラベルを付けることとする。このための情報としては、日本語の主要動詞について、動詞の格に関連する情報を約13,000レコード記述した、EDR電子化辞書の日本語動詞共起パターン副辞書[9](以後、共起辞書)があり、ここでは、この共起辞書を利用する。この辞書には各動詞の各概念について、表層格に対応する深層格の種類と、その深層格を取る場合の格助詞、意味情報構成要素(深層格を満たす概念の範囲を示した概念識別子群)が記述されている。図1に共起辞書の一部を示す。

概念関係子 意味情報構成要素		
<語1>		
が	agent	30f6b0:30f6bf
<語2>		
を	object	30f6e5:3f9639
食べる		

図1: “食べる”の共起情報

これらを用い、各概念関係子を使い、引数には表1のラベルを付加する。具体的には、時制が現在形の肯定文で、かつ、格が単純な単語のみの、最も基本的な場合は以下のように表す。

“太郎は東京から京都へ行く”⇒

行く(agt:太郎,sou:東京,gol:京都)

表1: 格と概念関係子

格	概念関係子	ラベル	説明
主体	agent	agt	動作を行う主体
対象	object	obj	動作思考の対象
源泉	source	sou	動作の起点
目標	goal	gol	動作の終点
場所	place	plc	事象の成立する場所
場面	scene	sce	事象の成立する場面
道具	implement	imp	動作における手段道具
材料	material	mat	材料
条件	condition	con	条件

2.2.2 時制、否定等を考慮した表現

動詞には、過去形や否定形、助動詞で修飾されている場合など、多様な状態がある。このような場合、述語の直前に表2のような記号を付けることとする。

例えば“ 太郎は林檎を食べた ”は“ *食べる (agt:太郎,obj:林檎) ”と表す。

表 2: 動詞の状態を表す記号

状態	記号	日本語	英語
否定	!		
過去	*		
希望	w	～したい	want
可能	c	～できる	can
提案	s	～すべき	should
許可	m	～してもよい	may
推測	p	～かもしれない	probably

2.2.3 引数が名詞句の場合

(1) “ A の B ”の場合

“ A の B ”型の名詞句における“ A ”と“ B ”の意味的關係は多数ある。これらの判定に関しては、これまで多くの研究がなされている [5][6] が、自動的かつ正確に判定することはかなり困難である。ここでは、文献 [5] のように、日本語語彙体系 [7] のカテゴリーを利用して比較的簡単に判定できる範囲を対象とし、2つの引数を用いて表現する。具体的には、帰属や所有等の“ の ”によって結ばれた名詞句は定数“ S ”で表し、その定数を“ A ”、“ B ”を用いた素式で限定する。すなわち、以下のようにその素式を元の素式と連言で結ぶ。

“ 元の素式 (S を含む) ∧ B (rel: S, co-rel: A) ”

以下に例を示す。

例) “ 花子は太郎の叔父と会う ”⇒

“ 会う (agt:花子,obj:S) ”

∧ 叔父 (rel: S, co-rel: 太郎) ”

これは本来、“ ∃x 会う (agt:花子,obj:x) ∧ 叔父 (rel:x, corel: 太郎) ”であり、x をスコールム定数 S に置き換えたものである。また、引数のラベル“ rel ”、“ co-rel ”については2.3節で説明する。

なお、容易に判定できない“ A の B ”の場合には、“ A ”と“ B ”を“ · ”で連結し以下のように複合語として扱う。

例) 大型の船 ⇒ 大型・船

(2) 形容詞を含む名詞句

“ 形容詞 ”+“ 名詞 ”の文型の場合である。この場合は、形容詞と形容詞で修飾された名詞句全体を定数“ S ”で表し、その定数を述語が形容詞と名詞の2つの素式で限定し、それらの素式を元の素式と連言で結ぶ。以下に例を示す。

例) “ 太郎は美しい鳥を飼う ”⇒

飼う (agt:太郎,obj:S) ”

∧ 美しい (sub:S) ∧ 鳥 (inst:S) ”

(3) 動詞を含む名詞句

この場合、動詞を含む名詞句全体を定数“ S ”で表し、その定数を述語が動詞と名詞の2つの素式で限定し、それらの素式を連言で結ぶ。以下に例を示す。

例) “ 太郎は飛んでいる鳥を撃った ”⇒

*撃つ (agt:太郎,obj:S) ”

∧ 飛ぶ (agt:S) ∧ 鳥 (inst:S) ”

2.2.4 引数が名詞節の場合

引数が名詞節の場合、1つの文で複数の事象を表していることとなる。そこで、1つの事象毎に素式を作り、それら素式を連言で結ぶこととする。引数が修飾される場合と、述語が修飾される場合がある。それぞれの表現について示す。

(1) 引数が修飾される場合

“ 太郎は花子が経営している会社の社員です ”を例にして説明する。この文は、“ 太郎は会社 S の社員です ”と“ 花子は会社 S を経営している ”に分けられ、それぞれを述語知識にする。ここで会社 S は2つの知識で共通なので、会社 (inst:S) とし、それぞれの知識を連言で結合する。よって、例文は

社員 (rel: 太郎, co-rel: S) ”

∧ 経営する (agt:花子,obj:S) ∧ 会社 (inst:S) ”

と表現される。

(2) 述語が修飾される場合

“ 太郎は営業部で働く日芝の社員です ”を例にして説明する。この文は、“ 太郎は日芝の社員です ”と“ 太郎は営業部で働く ”に分けられる。従って、それぞれを、述語知識にして連言で結合する。よって、この例文は

社員 (rel: 太郎, co-rel: 日芝) ”

∧ 働く (agt:太郎, plc: 営業部) ”

と表現される。

2.3 述部が名詞の場合

“ A は B である ”のような、述部が名詞となる文型については、ラベルとしては、“ inst ”、“ rel ”、“ co-rel ”を使った表現とする。名詞が述語となる素式では、一般には引数は1つであるが、その名詞が関係を表す場合には引数は関係の対象となる個体であり2つとなる。引数が1つの場合は“ inst ”を、2つの場合は“ rel ”、“ co-rel ”を用いる。

2.3.1 基本的な形式

述語知識では、引数に変数や定数が入り、述語は引数の特徴を表す。一般にBが普通名詞となる場合は、Bが変数や定数の特徴を表すことになるので、以下の形式とする。

Bが普通名詞の場合：B (inst:A)

例えば、“太郎は医者である”の場合は“医者 (inst:太郎)”と表す。

一方、変数や定数を強調する場合は、Bが固有名詞となる。この場合は、Bの特徴をAが表すことになるので、

Bが固有名詞の場合：A (inst:B)

とする。例えば、“犯人は太郎である”の場合は、“犯人 (inst:太郎)”と表す。

2.3.2 名詞句を含む場合

(1) Bが“ B_1 の B_2 ”のとき

B_2 が普通名詞の場合は、 B_2 はAの特徴を表す語となるので、 B_2 を述語とする。また、 B_1 は B_2 を限定するものなので引数とする。ただし、Aが主引数であるのでAを“rel”， B_1 を“co-rel”とする。すなわち、述語形式を

$B_2(\text{rel:A,co-rel:}B_1)$

とする。例えば、“太郎は日芝の社長である”の場合は“社長 (rel:太郎,co-rel:日芝)”と表す。

なおこの場合、述語形式として、“ $B_1 \cdot B_2(\text{inst:A})$ ”の形式が考えられる。しかし、 B_1 が変数になることがある(上記の例ではx社の社長)ので、一階述語論理を前提としている場合、この形式は適切とはいえない。

(2) Aが“ A_1 の A_2 ”のとき

この場合は、AとBの部分でそれぞれ1つの素式を作り、共通の定数Sを用いてそれらを連言で結ぶ。すなわち、以下のような述語形式とする。

$B_2(\text{rel:S,co-rel:}B_1) \wedge A_2(\text{rel:S,co-rel:}A_1)$

以下に例文と変換例を示す。

例)“太郎の息子は日芝の社長である”

⇒社長 (rel:S,co-rel:日芝)

∧息子 (rel:S,co-rel:太郎)

ここで、社長 (rel:太郎・息子,co-rel:日芝)の述語形式も考えられるが、“太郎”の部分を変数xとなる場合もあり、推論処理の複雑化をまねく。そこで太郎・息子の代わりに変数を置き、太郎・息子で一つの素式を作り、それを連言で結合することとした。

(3) Aが“ A_1 と $\sim A_2$ ”の場合

“ A_1 ”と“ A_2 ”が格助詞“と”で結ばれた場合、基本的には、Bを述語、“ A_1 ”と“ A_2 ”を、それぞれ引数とし、以下のように2つの素式にする。

B (inst: A_1)

B (inst: A_2)

以下に例を示す。

例)花子と良子は学生です

⇒学生 (inst:花子)

学生 (inst:良子)

なお、Bが関係を表す名詞の場合は引数を2つにして表現する。その判定には日本語語彙体系のカテゴリーを利用する。この場合引数の関係は対等となるので、ラベルは同じものを使う。よって、述語形式は

B (rel: A_1 ,rel: A_2)

とする。例を以下に示す。

例)“太郎と花子は友達である”

⇒友達 (rel:太郎,rel:花子)

なお、この場合、意味的にまったく同一である“太郎は花子の友達である”を変換した“友達 (rel:太郎,co-rel:花子)”と異なった表現になってしまう。これについては推論処理の過程で“rel”と“co-rel”を同一とみなすこととする。

2.4 述部が形容詞の場合

述語が形容詞となる述語形式を定める。ラベルとしては、“sbj”，“co-sbj”を使った表現形式とする。以下に、表現形式を示す。

(1) 基本形

“AはB”という文型である。その場合、形容詞は個体の特徴を表すので、形容詞のBが述語となり、Aが引数となる。よって述語形式は

B (sbj:A)

とする。例を以下に示す。

例)“花子は美しい” ⇒美しい (sbj:花子)

(2) 比較形

“ A_1 は A_2 よりB”という文型である。このような比較形では、比較される個体を引数とする。主語に当たる A_1 の個体のラベルを“sbj”とし、比較される方のラベルを“co-sbj”とする。述語形式は

B (sbj: A_1 ,co-sbj: A_2)

とする。例を以下に示す。

例) 花子は良子より美しい
 ⇒ 美しい (sbj:花子,co-sbj:良子)

(3) “ ~は~が~ ” の場合
 “ A₁ は A₂ が B ” という文型である。その場合、述語形式は A₁ が A₂ を所有しているものとみなし、

B (sbj:A₁,co-sub:A₂)

とする。例を以下に示す。

例) “ 花子は目が綺麗 ”
 ⇒ 綺麗 (sbj:花子,co-sub:目)

3 述語知識への変換

述部が動詞となる文を対象とし、文中の必須格を引数として取り出し、それにラベルを付ける方法について述べる。

3.1 述語知識への変換の流れ

基本的な変換の手順を図2に示す。この方法では、文の動詞を述語、名詞又は名詞句、複合語を引数とする。まず動詞の共起情報を EDR 共起辞書から取り出す。取り出した共起情報を利用して引数に深層格を用いたラベル付けをする。次に、名詞が意味的にラベルの引数となるか EDR 単語辞書 [10](以後、単語辞書) を使いチェックする。全てのラベルの引数が適切だった場合、述語知識に変換する。

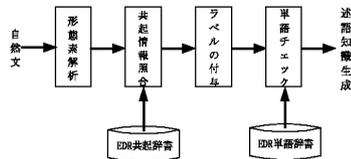


図 2: 述語知識変換の流れ

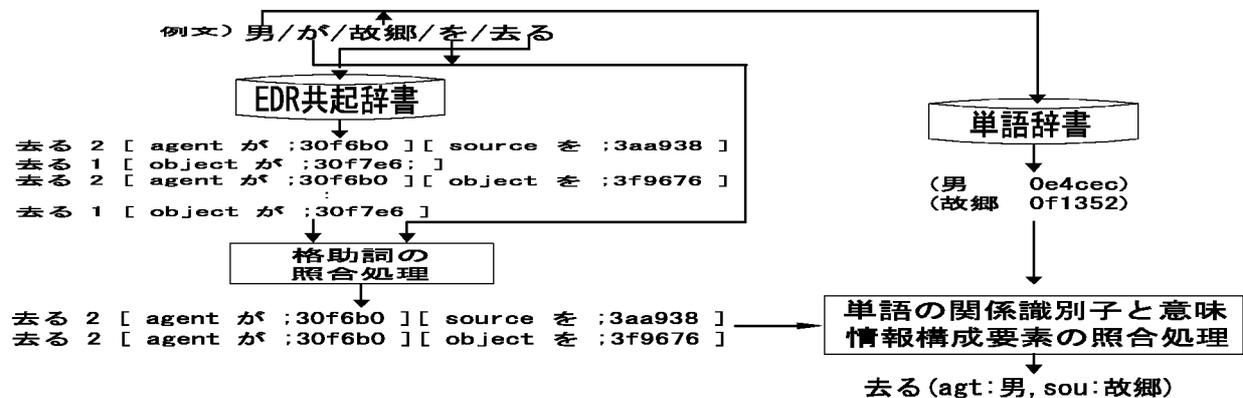


図 3: 述語知識変換の流れ

3.2 ラベルの付与

ラベルの付与は、EDR 共起辞書の共起情報と形態素解析の結果とのパターン照合で決定する。具体的には、動詞の全ての共起情報を取り出す。さらに、入力された文の格助詞を全て含む共起情報を特定し、同じ格助詞の前の語句をラベルの引数とする(図3)。

文が受動態の場合は、それを能動態にしてからラベルを付与する。受動態の検出は、形態素解析結果から“ れる ”または“ られる ”が検出された場合、その文を受動態とする。変換対象とする文型は以下とする。

i) Aは(が) Bに(によって) ~される。

ii) Aは(が) ~される。

これらの文型の場合、まず能動態になおしてから述語知識に変換する。

文型 i) : Aは(が) Bに(によって) ~される

格助詞の数が2つで、格助詞が“ は ”または“ が ”と、“ に ”または“ によって ”の場合、文型 i) とする。この文型は、“ A ”と“ B ”を入れ換え、格助詞“ に ”または“ によって ”を、“ を ”に置き換える。以下に例を示す。

例) “ 花子は太郎によって助けられた ”

⇒ 太郎は花子を助けた

⇒*助ける (agt:太郎,obj:花子)

文型 ii) : Aは(が) ~される

格助詞の数が1つで、格助詞が“ は ”または“ が ”の場合、文型 ii) とする。この文型は、主格を“ 未定 ”の意味とし全角の“ # ”を仮の引数として置く。さらに、“ A ”を目的格とし、その後ろに格助詞“ を ”つけ、能動態に置き換え変換する。以下に例を示す。

例) 機密文書は破棄される

⇒ #は機密文書を破棄する

⇒ 破棄する (agt:#,obj:機密・文書)

3.3 単語のチェック

3.3.1 基本的な処理

格助詞が一致しても、ラベルが意味的に候補の単語を取らないことがある。そこで、入力された単語の概念識別子が、意味情報構成要素の下位概念と一致したとき、ラベルの引数とする(図3)。全ての格助詞についてこの処理を行う。この処理において、共起情報が複数ある場合は、全てのラベルが候補の単語を引数とした共起情報を使う。

3.3.2 補足処理

共起辞書は複数の人により人手で作成されているため、不完全な部分がある。例えば、ラベルが候補の単語を取り得る場合でも、概念識別子が意味情報構成要素と一致しない場合がある。また、動詞が共起辞書に収録されておらず、共起情報が取り出せない場合もある。そこでこの取りこぼしを少なくするために、以下のルールを用いた。

ルール1)

候補の単語の概念識別子が、意味情報構成要素の1つ上の概念と同じ概念識別子を持つならば引数とする。

“男が傘を電車に忘れる”という文を例にして説明する。この文の共起情報は、

”忘れる [agent 30f6b0 が][object 3cf123 を][goal 3aa930 に]”となる。この共起情報の“object”の意味情報構成要素は“3cf123(生産物)”である。一方、“object”を候補とする例文の単語“傘”の概念識別子は“3cf8f5”である。図4から、“傘”は“生産物”を上位概念に持たないが、“生産物”の1つ上の概念“静物”は、“傘”を下位概念として持つ。よって、“object”は“傘”を引数としてとることとする。

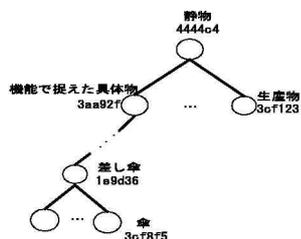


図4: 補足処理1

ルール2)

動詞が共起辞書に登録されていなかった場

合や、取り出した共起情報の格パターンが一致しなかった場合、IPAL 動詞辞書を使い、類義語の共起情報を利用する。IPAL 動詞辞書には、861語の動詞について、その上位語、類義語、対義語等が記載されている。

4 自然言語文への変換法

対話システムは、問題解決規則や常識知識あるいは推論結果から回答文や質問文を生成する。ここでは述語知識から自然言語文への変換法について提案する。

この方法は基本的には、3章で述べた自然文から述語知識への変換法の逆の手順で行なう。述語知識から自然言語文へ変換する手順を図5に示す。

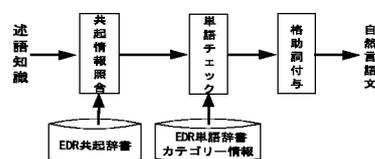


図5: 自然言語文への変換の流れ

この処理において、変換対象の一致する共起情報が複数ある場合、どの共起情報により正しい自然文に変換できるか判定する必要がある。このため、ラベル照合の後、対象の引数をチェックし、使用する共起情報を決定する。

例) 変換対象: 通る (agt:彼,obj:試験)

共起情報:

通る [agent 30f746 世間に][object 3d017c 作品が]

通る [agent 30f6b0 息子が][object 30f83e 試験に]

を用いれば、“彼に試験が通る”となり、を用いれば、“彼が試験に通る”となる。変換対象中の“彼”、“試験”の概念識別子がの方と一致するため、を用いる。

以上の方法は、基本的には3.3.1項と同様の方法で単語のチェックを行なう方法であるが、ここではもう一つの方法として共起辞書中の例文を利用する方法を提案する。すなわち、EDR共起辞書の共起情報は例文を基に作成されている。そこで、入力文とEDR共起情報の例文のカテゴリーを比較することで、入力文と似ている共起情報を使用する。比較は次のように行なう。対象とする引数と、例文の単語のカテゴリーを調べる。これらの単語間のシソーラス上での距離を求める。各単語間の距離の平均が最も短かった共起情報を用いて変換を行なう。

先の例を用いて説明する。引数と共起情報の単語

[述語知識]	[自然文]	
言う (agt:彼,obj:冗談)	⇒	彼が冗談を言う
晴れる (obj:霧)	⇒	霧が晴れる
行く (agt:彼女,gol 大学:)	⇒	彼女が大学に行く

[9] 日本電子化辞書研究所:EDR 電子化辞書, 日本語動詞共起パターン副辞書,1996.

[10] 日本電子化辞書研究所:EDR 電子化辞書, 日本語単語辞書,1996.

[11] 竹林滋, 小島義郎:ライトハウス英和辞典:研究社 (1993).

6 おわりに

対話文のような, 比較的簡単な自然文を対象とした, 一階述語論理の枠組みを越えない知識表現形式の提案を行った. この形式は, 電子化文書等を利用して容易に変換可能であり, 知識は全て単語ベースで表現される. さらに, 自然文を EDR 共起辞書を利用して自動的にこの表現形式に変換する手法と, 逆に述語知識から自然文に変換する手法を提案した. また, 述語知識への変換手法に対し, 英和辞典等の例文を利用して評価した結果, 約 86%の変換率が得られた. また, 自然文への変換手法に関しては, 100%の正解率が得られた.

今後は, より複雑な文を対象とした変換法について検討していく.

参考文献

- [1] Nguyen Viet Ha, 石川勉, 阿部明典: 知識の類似性を利用した概略推論法, 電子情報通信学会論文誌 D ,Vol.J84-D-1, No. 4, pp.389-400 (2001).
- [2] 小林 高志:感情表現と質問応答能力を備えた知的自由対話システム, 人工知能学会研究会資料 SIG-SLUD-A003-1,pp.1-8(2001)
- [3] 大石 享, 松本 裕治:格パターン分析に基づく動詞の語彙知識獲得, 情報処理学会論文誌,Vol.36,No.11,pp.2597-2610(1995)
- [4] 富浦 洋一, 中村 貞吾, 日高 達:名詞句「NP の NP」の意味構造, 情報処理学会論文誌,Vol.36,No.6,pp.1441-1448(1995)
- [5] 池原 悟, 村上 仁一, 宮本 健司:「A の B」型名詞句の日英翻訳規則について, 情報処理学会論文誌,Vol.43,No.7,pp.2300-2308(2002)
- [6] 横山 昌一, 加藤 貴子, 廣重 拓司:格助詞「の」の分類と解析, 情報処理学会第 1 回年次大会,pp.137-140(1995)
- [7] NTT コミュニケーション科学研究所:日本語語彙体系,1 意味体系,1997
- [8] 黒石禎夫他:自然言語処理:岩波講座ソフトウェア化学 15:pp200-230 (1996).