

## 推論ルールを用いた 日本語文末表現の言い換え

雨崎翔悟<sup>†</sup> 田辺利文<sup>†</sup> 首藤公昭<sup>†</sup>

本論文では、推論ルールを用いた日本語文末表現の言い換えについて述べる。我々は、これまで約 1,450 個の複単語表現を含む日本語文末表現を約 1,500 個収集しており、収集した表現のそれぞれに意味関数を与えている。推論ルールはこれまで約 200 個収集している。例えば、推論ルール 過去[過剰性[x]]→否定[願望[x]] を用いることで、「私は飲みすぎた」から「私は飲みたくない」などへの言い換えが実現できる。

### Paraphrasing of Japanese Sentence-Final Expressions Using Inference Rules

Shougo Amazaki,<sup>†</sup> Toshifumi Tanabe<sup>†</sup> and kosho Shudo<sup>†</sup>

This paper presents a method of paraphrasing of Japanese sentence-final expressions using inference rules. We have extracted approximately 1,500 sentence-final functional expressions including approximately 1,450 multiword expressions, assigning semantic functions to each expression. We have devised about 200 inference rules. Using an inference rule, such as PAST-TENSE[EXCESSIVE[x]]→NEGATION[ASPIRATION[x]], makes the sentence "I drank too much." is paraphrased into "I don't want to drink."

### 1. はじめに

近年、Web に代表される文書情報の爆発的な増加に伴い、より精度の高い自然言語処理の必要性が認識されている。特に、意味を考慮した処理システムの需要が増加していることは Semantic Web 研究等の発展を見ても明らかである。筆者らは、意味を考慮した言語データ処理においては、複単語表現(Multiword Expression: MWE)を組み込んだシステム構築が不可欠であると考えている。MWE の研究動向は、2002 年の論文「Multiword Expressions: A Pain in the Neck for NLP」(Sag et al. 2002)を皮切りに(国際)計算言語学会(Association for Computational Linguistics: ACL)を中心に MWE workshop がほぼ毎年開催されるなど、世界的に重要性が認識されてきた。MWE は頻繁に自然言語に現れることも報告されており、(Sag et al. 2002)では、WordNet1.7 での見出しの約 41%が MWE であること、また(田辺ら 2006)では、日本語の述部における文末表現を構成する助動詞、終助詞相当の MWE の出現比率が約 42%であることなどから、MWE の適切な取り扱いが自然言語処理の質を向上させる上で必要不可欠であることを示している。

これまで筆者らは、非構成(イディオム)性、要素語間の強い共起性、のうち少なくとも一方の性質を持つ単語列を MWE として収集してきた(首藤ら 2010)。ここでの非構成性とは、概略、表現を構成している単語の通常の意味から表現全体の意味を構成するのが難しい性質、要素語間の強い共起性とは、要素単語相互の確率的な縛りが強い性質を意味する。

日本語では、自立語性 MWE として「油を売る」「ぐっすり眠る」「手をこまぬく」などが、機能語性 MWE として、「によって」「に関して」「なければならない」「かもしれない」などがある。本論文では、日本語述部の文末表現として現れる機能語性 MWE に着目する。このような表現は、必要性を表す「なければならない」、欲求を表す「たい」、様態を表す「ようだ」のように、発話者の主観を表す表現がバラエティに富んでおり、かつ「なければならない/ようだった」のように複数個連続して用いられることも多く、発話者の主観を推定するような意味処理を行うためには、機能語性 MWE を意識した適切な文末表現の取り扱いが必要不可欠であると考えている。

一方近年では、含意関係認識(Recognizing Textual Entailment: RTE)が盛んに研究されている。含意関係認識においては、「(X を)洗う→(X が)きれいになる」などのような自立語間で定義される含意関係を取り扱う研究がほとんどのようであるが、機能語も重要な役割を担うものと考えている。本論文では、含意関係認識のリソースとして日本語文末表現に着目し、従来の研究にはあまり見られなかった機能語間レベルで定義できる含意関係、具体的には日本語文末表現に対応する意味構造間で定義できる含意

<sup>†</sup> 福岡大学工学部  
Fukuoka University, Faculty of Engineering

関係(推論ルールと呼ぶ)について報告する。

本論文では、第2章で非命題的意味構造を紹介し、日本語文末の特徴を挙げながら非命題的意味構造の枠組みの妥当性を示す。第3章では非命題的意味構造間で定義される推論ルールの概要および適用例について説明する。第4章で関連研究として、モデルティ、言い換え、含意関係認識などについて動向を概観する。第5章では今後の課題をリストアップし、最後の第6章でまとめを述べる。

## 2. 非命題的意味構造

一般的に文の意味は、命題的意味と非命題的意味からなっていると考えられ、日本語においては、非命題的意味を表す部分は述語に後接した、態、時制、相(アスペクト)、否定、話し手の態度など、広義の様相情報を与える助動詞、終助詞およびそれらに相当する機能語性 MWE により構成される場合が多い。本論文ではそれらの表現を助述表現と呼び、これまで約 1,450 個の機能語性 MWE と約 50 個の助動詞、終助詞を収録した辞書を作成している[a]。機能語性 MWE としては、例えば、「て」と「いる」をまとめた「ている」、「かも」と「しれ」と「ない」をまとめた「かもしれない」などを収録している。辞書作成に関しては網羅性の高さを目標の一つと考えており、例えば、(森田ら 1989)の収録表現のうち該当表現はすべてカバーしている。機能語性 MWE を適切に設定することで、日本語文の構造の大枠は次の生成規則で表すことができる。

- (1)  $S_0 \rightarrow BP^* \text{ PRED}$
- (2)  $S_i \rightarrow S_{i-1} \cdot e_i \quad (1 \leq i \leq n)$

生成規則(1)において、 $S_0$ ,  $BP^*$ ,  $*$ ,  $PRED$ ,  $e_i$  はそれぞれ骨格文、文節、閉包演算子、述語、助述表現を表す。(2)は、 $S_{i-1}$  と  $e_i$  とで文  $S_i$  が構成されること、及び、文  $S_i$  の構造が左分岐であることを表している。助述表現に対応する意味を意味関数として捉え、意味関数の並びに対応する形で、非命題的意味構造を作成することができる。現時点で意味関数は約 130 種類設定している。非命題的意味構造は、おおよそ(3)に示すように入れ子型構造として表現できる。

- (3)  $M_m [M_{m-1} \dots [M_2 [M_1 [S]]] \dots]$

ここで  $M_i$  ( $1 \leq i \leq m$ ) は一種の意味関数である。例えば、日本語文『彼は動き始めていないかもしれない』では、日本語文の構造の概形および対応する非命題的意味構造は

a) これらの数値は、漢字-かな などの表現のゆれをまとめて 1 見出しとした値である。

図1のようになる[b]。(1),(2)に示される構文構造と,(3)に示される意味構造の間には一種の同型性があると言える。助述表現が文の述部にいくつも並ぶ複雑な文末の場合でも、意味関数との対応をとることにより非命題的意味構造を求めることが可能である。日本語文末の特徴をつかむため、(田辺ら 2006)では日本語文の述部にどの程度助述表現が表れるか、EDR コーパス(EDR 1996)からランダムに抽出した 9,210 個の文を対象にし、述語の品詞が動詞、形容詞の場合に限定して調査している。その結果を表1に示す[c]。ここで  $n$  は述語に連続して現れた助述表現の個数である。表1から確認できることは、(a)日本語文末に少なくとも1個の助述表現が含まれる割合は  $47\% = (9,210 - 4,899) / 9,210 * 100$  であること、(b)助述表現は連続して現れることが少ないこと、(c)助述表現が複単語である割合は 42%であったこと、などが挙げられる。これらの調査結果を踏まえると、機能語性 MWE を適切に設定しつつ入れ子型で表現した、非命題的意味構造の枠組みの妥当性が示されたといえる。このように(3)で表現される非命題的意味構造の枠組みは、構造のシンプルさと同時に対応可能な表現の多様さから工学的にも重要な性質を示していると考えられ、言語依存性も無いとされるため[d]、言い換えや機械翻訳を行う際の間言言語として有効であると考えられる。

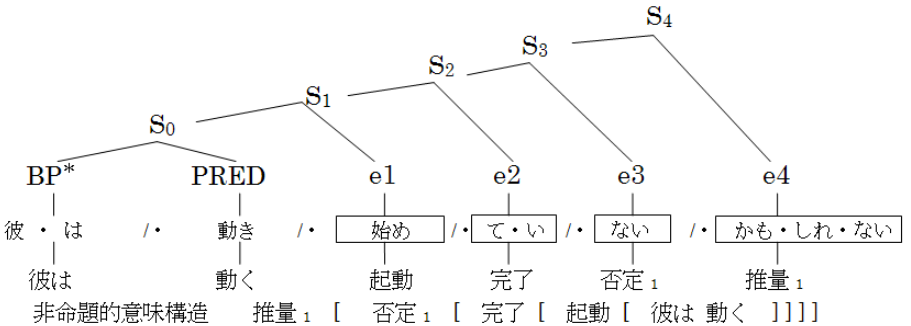


図1 日本語文の概形及び非命題的意味構造

b) 図1の記号「・」は通常の単語境界を表し、「/」はMWEを意識した単語境界を表す。また、動詞に後接する「はじめる」は動詞ととらえる考え方もあるが、アスペクト情報を含むと考えられるため本論文では助述表現としている。

c)  $n=5$  の例として「左右/さ/れ/ない/よう/に/し/たい/もの/だ」「発揮/でき/ない/状況/に/ある/から/だ/と/いう」が観測された。ここで下線が付与されている表現は助述表現であることを表す。

d) (Cinque 1999; Cinque 2006)は、非命題的意味構造における意味関数の生起順序には、言語に依存しない規則性が存在するのではないかと報告している。

表1 助述表現の生起数

助述表現 n	述部の 生起	単一の単語助述表現 の生起(A)	複単語助述表現 の生起(B)	A+B	B/(A+B)
0	4,899	-	-	-	-
1	3,131	1,852	1,279	3,131	0.41
2	966	1,128	804	1,932	0.42
3	178	276	258	534	0.48
4	34	63	73	136	0.54
5	2	7	3	10	0.30
合計	9,210	3,326	2,417	5,743	0.42

### 3. 推論ルール

ここでは機能語間レベルで定義できると思われる含意関係について考える。第2章で述べた非命題的意味構造は、助述表現列を意味的に抽象化したものであると捉えられる。そのため、非命題的意味構造レベルで含意関係を取り扱うことができれば含意関係知識をコンパクトにできるものと仮定し、非命題的意味構造間で含意関係を定義することを試みる。

特定の文脈で、非命題的意味構造 X を、別の非命題的意味構造 Y に含意(推論)によって書き換えることが出来るとき、推論ルール  $X \rightarrow Y$  が成り立っていると定義する。推論ルールには、例えば“不足性[x]→願望[x]”や“過去[過剰性[x]]→否定[願望[x]]”などがある[e]。推論プロセスは、例えば『飲み/たりない』は、非命題的意味構造“不足性[飲む]”に変換された後、推論ルール“不足性[x]→願望[x]”が適用された場合には別の非命題的意味構造“願望[飲む]”に書き換えて『飲み/たい』を生成する。一方、『飲み/すぎ/た』に対しては、非命題的意味構造“過去[過剰性[飲む]]”に変換された後、推論ルール“過去[過剰性[x]]→否定[願望[x]]”が適用された場合には別の非命題的意味構造“否定[願望[飲む]]”に書き換えて『飲み/たく/ない』を生成する。

推論ルールの適用には曖昧さが生じる場合がある。例えば、非命題的意味構造“不足性[x]”に関しては、適用できる推論ルールとして“不足性[x]→願望[x]”だけでなく“不足性[x]→命令[x]”を設けている。推論ルール適用の曖昧さを軽減するため、このケースでは主語が1人称、例えば『(私は)飲み/たりない』の場合には非命題的意味構造“不足性[(私は)飲む]”に変換された後に、推論ルール“不足性[x]→願望[x]”を優

e) 推論ルール中の x は命題的意味を表すものとする。

先して適用し、“願望[(私は)飲む]”に言い換えて『(私は)飲み/たい』を生成する。一方、このケースで主語が2人称、例えば『(お前は)飲み/たりない』の場合には非命題的意味構造“不足性[(お前は)飲む]”に変換された後に、推論ルール“不足性[x]→命令[x]”を優先して適用し、“命令[(お前は)飲む]”に言い換えて『(お前は)飲み/なさい』を生成する。このように、曖昧さがある場合にどの推論ルールを適用するかを適用条件として設定することが必要である[f]。適用条件は一般的には文脈に依存することが考えられるが、文脈情報を網羅的に抽出することは非現実的であると考えており、現時点では、推論ルールごとにルールの適用条件を粗く設定している。具体的な適用条件としては発話文の主語の人称や、自動詞や他動詞などの動詞の粗いカテゴリを設定している。推論ルールの一部および推論ルールの適用例を表2に示す。推論ルールおよび適用条件はこれまで人手で作成している。現時点での推論ルールの総数は約200種類であるが、さらに網羅性を向上させるため、複数人での推論ルールの追加を行っている。

### 4. 関連研究

#### 4.1 モダリティに関するもの

推量を表す「だろう」や「かもしれない」、義務を表す「なければならない」などの日本語文末表現は一般にモダリティの範疇に位置づけられているが、例えば(Narrog 2009; Narrog 2010)のように、一般にモダリティとは言い切れない、時制、相(アスペクト)を表す表現、「よ」「ね」のような終助詞、「食べ/たい」の「たい」などの表現もモダリティとして取り扱う言語学分野での研究が存在する。Narrog が提案するモダリティ体系は、本論文における助述表現の体系と類似している部分が多い[g]。

また工学的分野での研究としては、情報発信者の主観的な態度(モダリティ)情報を情報抽出や含意認識などの自然言語処理の応用タスクに組み込み、精度向上を目指した(江口ら 2010)がある。(江口ら 2010)でも、一般にモダリティとは言えないような表現を加えた拡張モダリティ体系を提案し、この体系に基づいたコーパスを作成している[h]。これらの研究は、従来のモダリティの枠組みでは言語現象を包括的に説明できず、また、深い自然言語処理、いいかえると精密な意味処理を行うためには不十分であることを示唆しているように見うけられる[i]。

f) 非適用条件、具体的には適用しない場合を表す適用条件も含む。

g) (Narrog 2010)は、日本語による非命題的意味構造の意味関数の生起順序に何らかの規則性があることを報告している。

h) 拡張モダリティには、「望む」のような動詞、副詞、形容詞も含んでいるが、本論文では自立語相当表現は文末表現として扱わない。

i) 本研究では、その表現がモダリティの範疇に属しているかどうかは立ち入らない。

表2 推論ルールの一例と適用例(下線が付与されている箇所は助述表現, 推論ルール中の x は命題の意味を表す)

例	書き換え後	推論ルール		適用条件	推論ルールの特徴
		書き換え前	書き換え後		
彼が誘う/のだもの	彼に誘われ/たく/なかつた	理由付け[x]	過去[否定[願望[受動態 <sub>1</sub> [x]]]]	x: 他動詞	voice が変わる
ご飯/を/食べる/とも	ご飯/を/食べる/つもりだ	同意[x]	方針・予定[x]	x: 他動詞	自動詞か他動詞で書き換える推論ルールが変わる
成績/は/上がる/とも	成績/は/上がる/はずだ		推量 <sub>2</sub> [x]	x: 自動詞	
食べ/にくい	食べ/たく/ない	困難性[x]	否定[願望[x]]	x: 他動詞	
上がり/にくい	上げ/たい		願望[x]	x: 自動詞	
食べ/て/くれる/な	食べる/な	禁止[受動態 <sub>2</sub> [x]]	禁止[x]	x: 他動詞	「てくれる」がなくなる
	食べ/ない/でほしい		願望[否定[x]]	x: 他動詞	
うどん/を/食べ/なければならぬ	うどん/を/食べ/て/いい/ない	必要性 <sub>1</sub> [x]	否定[進行[x]]	x: 継続動詞 or 瞬間動詞	モダリティとアスペクトが絡む
ナマコ/を/食べ/なければならなかつた	ナマコ/を/食べ/たく/なかつた	過去[必要性 <sub>1</sub> [x]]	過去[否定[願望[x]]]	x: 他動詞	モダリティと主観表現が絡む
飲み/すぎる	飲む/な	過剰性[x]	禁止[x]	x: 他動詞, 主語: 2 人称	基本的には, 発話文の主語の人称によって適用できる推論ルールが異なる. また, 上司と部下の関係などによっても, 適用に制約を付加させることができる
	飲む/けいこうがある		傾向[x]	x: 他動詞	
	飲み/すぎる/な		禁止[過剰性[x]]	x: 他動詞, 主語: 2 人称	
	飲み/すぎ/ない/ほうがいい		必要性 <sub>4</sub> [否定[過剰性[x]]]	x: 他動詞, 主語: 2 or 3 人称	
飲み/すぎ/た	飲み/たく/ない	過去[過剰性[x]]	否定[願望[x]]	主語 1 人称	
飲み/たりない	飲み/たい	不足性[x]	願望[x]	x: 他動詞, 主語: 1 人称	<不足性>と<過剰性>は関連した意味関数であり, 主語の人称によって適用できる推論ルールが異なる点は同じだが, 書き換え後の非命題的意味構造の体系は, <過剰性>には禁止があるなど, 異なった体系になるようである
	飲み/なさい		命令[x]	x: 他動詞, 主語: 2 人称	
	飲む/ほうがいい		必要性 <sub>4</sub> [x]	主語: 2 or 3 人称	
	飲ん/だ		過去[x]	x: 他動詞	
	飲ま/せろ		使役態 <sub>1</sub> [x]	x: 他動詞	
飲み/たりない/です	飲み/ましょう	丁寧[不足性[x]]	勧誘[x]	x: 他動詞	

#### 4.2 言い換え, 含意関係認識などに関するもの

言い換え技術全体の研究動向をサーベイしたものと(乾ら 2004)が挙げられる。(乾ら 2004)では, 機能語相当表現の言い換えについて, 語彙的な性格が比較的強く局

所的な情報を参照するだけで言い換えられるものも多いとし, この類の言い換えを実現するためには同義の機能語相当表現をグループ化して辞書を整備することが課題であると述べている。

(松吉ら 2007)は、見出し語総数約 17,000 の日本語機能表現辞書「つつじ」を編纂した[j]. 「つつじ」は見出し体系が 9 つの階層レベルで記述されており、含まれる文末表現には約 100 種の意味分類が定義されているようである。「つつじ」を言い換えに用いた研究として(Matsuyoshi et al. 2008)があげられるほか、(梶田ら 2008)では、独自に作成した準機能表現辞書と「つつじ」を組み合わせた言い換えシステムを作成している。ここでの準機能表現とは、自立語を含んだ機能語と意味的に等価な表現であるが従来の機能語相当表現には含まれない表現であり、例えば「～以外に方法はない」などが挙げられる。さらに(Izumi et al. 2010)では、日本語文末表現を標準形に言い換える枠組みを提案しており、その 1 つとして、例えば「買ったかった」を「買った」に言い換えるような、複雑な表現から簡素な表現への言い換えを行っている。一方(伊佐治ら 2005)は、独自に機能語相当表現辞書を作成し、日本語文末を標準的な表現、例えば「でしょう」を「だろう」に置換する機能を含む構文解析システム ibukiC を構築している[k]. これらの機能語相当表現の言い換えは、基本的に表現から抽出された意味タグと同じ意味タグをもつ別の表現に言い換えるものがほとんどであり、異なった意味タグに言い換える研究は少ないようである。異なった意味タグに言い換える研究としては(首藤ら 2001; Tanabe et al. 2001; Shudo et al. 2004)などが挙げられる。これらは日本語文末に位置する機能語相当表現の言い換え規則を類似性規則として意味タグレベルで記述し、例えば“否定 [否定 [X]] ⇐ [X]”などの論理的規則、“疑問[否定[X]] ⇐ 勧誘[X]”などの語用論的規則の 2 種類の類似性規則を挙げており、異なった意味タグ列間で言い換えても意味はほとんど変わらない場合があることを示している[l].

一方近年では、応用範囲の広さなどから含意関係認識(Recognizing Textual Entailment: RTE)が盛んに研究されている(Lin et al. 2001; Dagan et al. 2006; Szpektor et al. 2007; Berant et al. 2011). RTE とは、一対のテキストが与えられたときに一方が他方の記述から含意(あるいは推論)されるか否かを判別する問題である。一例として(阿部ら 2010)では因果関係として“(X を)洗う → (X が)きれいになる”などを挙げており、高精度の含意関係認識を実現するためにはこのような関係を多く収集しておくことが必要であるとされている[m]. これらの含意関係認識においては自立語を対象とした研究がほとんどのようであるが、機能語も重要な役割を担っていると思われる。(乾 2008)では、含意関係認識では、個々のテキストからどれくらいリッチな意味的情報を引き

j) <http://kotoba.nuee.nagoya-u.ac.jp/tsutsuji/>

k) <http://www.ikd.info.gifu-u.ac.jp/ibukiC/>

l) 例に挙げた論理的規則によって「行かないことはない」と「行く」が、また語用論的規則によって「見ないか」と「見ようよ」がそれぞれ互いに言い換えることができる。

m) 含意関係認識で収集される関係は広義の言い換え、言い換えると単方向の言い換えと見なすことができる。

出せるかがキーの一つであるとしている。そのためには、自立語情報だけでなく、本論文で示した機能語情報に基づいた含意関係知識を組み合わせることなどが、含意関係認識の精度向上に貢献できるものと我々は考えている。

機能語情報に基づく含意関係知識を扱っている研究として、(本田ら 2008)では、日本語文末表現意味体系をもとに、類似以外の関係として、話し手の発話を受けた聞き手が推論しうる関係として、“願望[X が Y する] → 否定[完了[X が Y する]]”などを提案している。この関係は「X が Y したい」ならば「X が Y していない」はず、を表している。しかし(本田ら 2008)が扱っている関係は少数の日本語文末表現に対してのみであり、日本語文末表現全般に対し網羅的に調査したものは現時点では見当たらないようである。

本論文の研究は、日本語文末表現間で推論の関係になり得る日本語文末表現対(推論ルール)を意味タグレベルで網羅的に収集、記述するもので、これまでの研究ではあまり見られなかった観点での意味的情報の抽出を試みるものである。

## 5. 今後の課題

今後の課題として、次のようなタスクを考えている。

- ・推論ルールのカバレッジ向上
- ・推論ルールのカテゴリライズ
- ・推論ルールの適用条件の詳細化
- ・推論ルール適用後の非命題の意味構造のブロッキング
- ・性能評価
- ・推移律の検証

以下、それぞれのタスクを説明する。

### 推論ルールのカバレッジ向上

現段階では約 200 種類作成しているが、ルールの網羅性をさらに向上させるため、現時点では多人数による推論ルールの作成を念頭に考えている。

### 推論ルールのカテゴリライズ

推論ルールをいくつかのカテゴリに分類することを検討しているが、現時点では 1 例として、大まかに感情抽出型と事実抽出型とに分類できるのではないかと考えている。感情抽出型としては推論ルール適用結果に願望や命令、禁止が含まれるもの、事実抽出型はそうでないもの、のような分類ができそうであるが、その他の観点でのカテゴリライズも含めて検討したい。

### 推論ルールの適用条件の詳細化

やみくもに推論ルールを適用すれば再現率は上がるが適合率は極端に下がるため、適用できる場合とできない場合を区別することが重要である。そのためには推論ルー

ルの細分化と適用条件の付与が必要である。例えば、動詞は、自動詞-他動詞レベルから、アスペクトの概念などを用いてさらに細分化する必要があると考えている。また適用条件には文脈情報を取り入れなければならない。「殺すつもりだった」では、『最初から殺すつもりだった』の場合には「殺している」が、また『本当は殺すつもりだった』では「殺していない」が推論されなければならない。このケースでは、「最初から」が文に存在すれば「殺している」が推論されるような適用条件を設定すればよい。しかし、一般的には、発話者と聞き手の関係が上司と部下の関係にあるかどうかが要求されるなど、推論ルールごとに、どのような文脈情報が適用条件として必要であるかを整理し、かつ抽象化することが必要になってくる[n]。本論文で示した助述表現辞書および推論ルールそのもののサイズは、自立語間で定義される含意関係知識のサイズに比べてかなりコンパクトになることが予想されるが、推論ルールに与えられる適用条件に関するリソース部分が肥大化してしまえば、トータルで考えると機能語による推論部分がコンパクトであるとは一概には言えなくなってしまう。そのため、適用条件は詳細にしたまま、かつ、適用条件を記述したリソースのサイズをできる限りコンパクトにすることが必要であると考えている。

#### 推論ルール適用後の非命題的意味構造のブロッキング

適用条件を満たした推論ルールを適用した後に、ありえない非命題的意味構造が生成されることも考えられる。ありえない非命題的意味構造として、例えば“願望[過去[x]]”、“必要性[命令[x]]”、“過剰性[疑問[x]]”などが挙げられる。このような、ありえない非命題的意味構造を生成してしまった場合にはその推論ルールの適用を取り消す。このような、ありえない非命題的意味構造のパターンを網羅的に調査する[o]。

#### 性能評価

推論ルールを実際の日本語文に適用した場合に、どの程度、推論された結果が妥当であるかを実験によって確認する必要がある。

一般的な(ルール適用の方向が単方向ではなく双方向の)言い換え規則であれば、意味の等価性が判断できればよいため実験対象として単文が集まったコーパスを選ぶことができる。また、一般的には単方向である推論ルールの場合でも、自立語の場合に

n) 適用条件の詳細化の問題は一般的な含意関係認識における関係にも存在する。例えば、関係“(Xを)洗う→(Xが)きれいになる”を言い換え規則として用いる場合、例えば「テキトーに服を洗う」が「服がきれいになる」に言い換えられてしまう。この場合には「テキトー」が文中に存在すれば言い換え規則として適用しない、などが考えられる。

o) (Cinque 1999; Cinque 2006)や(Narrog 2010)が述べている非命題的意味構造の意味関数の生起順序の規則性に関して、それに違反したものを、ここでのありえない非命題的意味構造として捉えることも考えている。意味関数の生起順序としては、日本語の述部においては、アスペクトを表す表現より態を表す表現がより先頭側に位置する、などがある。例えば「ている」と「れる」が共に述部に混在しているケースで、「ている」はアスペクトであるため、「食べていられる」の場合には「れる」は態の解釈ではなく可能の解釈になり、「食べられている」の場合には「れる」は態の解釈になり、曖昧さの解消が可能になる。

は正誤の判断が比較的容易に出来るものと考えている。しかし、本研究で提案した推論ルールによる日本語文末表現の言い換えの妥当性を判断する際には、一般的なコーパスなどの単文の集まりを用いるのではなく、いくつかの文がまとまったデータベースが必要になり、判断を人手で行うことは避けられないと考えている。データベースとして考えられるものはブログ、ニュースなどのインターネット上のデータなどが量的に適切であると思われるが、著作権などの問題も絡むこと等から考えるとそのままでの使用は難しい。そのため実験対象として現時点で検討しているものが青空文庫[p]に収録された小説である。青空文庫とは著作権の切れた小説をテキストデータにインターネット上に公開している Web ページである。青空文庫は、(a)リソースが小説であるため文脈情報がかかめること、(b)公開されている小説は無料で自由にダウンロードでき、テキスト化も容易であるため計算機処理に適していること、(c)アクセスランキング[q]を参照することで著名な小説の選択が可能であること、などの性質を持つことから、推論ルールの言い換えの妥当性判断のために適切であると判断している。さらに、どの小説を実験対象として選定するかも重要であり、例えば、児童向けであるとみなされている小説については繰り返し表現が多いことが予想されるため一部の推論ルールのみ適用にとどまる可能性があること、古語が多用されている小説は形態素解析の失敗が多発し文末表現から非命題的意味構造が適切に生成されないと考えられること、などが考えられるため一般向けの口語で記述された小説が適切ではないかと考えている。

また、推論ルールは非命題的意味構造間で定義しており、助述表現列間で推論ルールを定義したものに比べて意味的に抽象化したものであるため、推論ルールの抽象表記の妥当性について、コンパクト性とトレードオフについても検討しなければならない。

#### 推移律の検証

推論ルールは  $X \rightarrow Y$  の形をしている一方、推移律が成り立たないケースがある。推移律とは、 $X \rightarrow Y$  かつ  $Y \rightarrow Z$  が成り立っている場合には  $X \rightarrow Z$  も成り立つという性質である。しかし、単純な推論ルールの枠組み、例えば、(Berant et al. 2011)でも言及しているように“ $X \text{ buy } Y \rightarrow X \text{ acquire } Y$ ”と“ $X \text{ acquire } Y \rightarrow X \text{ learn } Y$ ”に推移律を適用した“ $X \text{ buy } Y \rightarrow X \text{ learn } Y$ ”は一般的に成り立たない。この解決策としてはルールの細分化を行うことなどが考えられる。なおこのケースは自立語間の推論ルールであるが、本論文で提案した推論ルールでも推移律が成り立つかどうかの検証が必要であると考えている。

p) <http://www.aozora.gr.jp/>

q) [http://www.aozora.gr.jp/access\\_ranking/](http://www.aozora.gr.jp/access_ranking/)



## 6. おわりに

非命題の意味を適切に抽出することは、対話理解、文脈モデルや話者の態度の推定など、近未来の自然言語処理研究で重要な役割を果たすものと考えている。本論文では、日本語文末に位置する助述表現列による非命題の意味構造間で定義できる含意関係(推論ルール)について述べた。

現時点での助述表現辞書には、意味関数を約 130 種定義している。例えば《疑問》を表す意味関数は 4 種あり、「食べる/か」の「か」には<疑問<sub>1</sub>>、「食べる/んだって」の「んだって」には<疑問<sub>2</sub>>、「食べる/んだよね」の「んだよね」には<疑問<sub>3</sub>>、「食べる/のではないか」の「のではないか」には<疑問<sub>4</sub>>のように細分化している。それぞれの意味関数を含む非命題の意味構造の推論ルールは異なっており、例えば、“疑問<sub>2</sub>[x]→推量<sub>2</sub>[否定<sub>1</sub>[x]]”、“疑問<sub>3</sub>[x]→願望[x]”、“疑問<sub>4</sub>[x]→推量<sub>2</sub>[x]”を作成している。これらの推論ルールを適用することで、それぞれ「食べ/ないはずだ」、「食べ/てほしい」、「食べる/はずだ」などが推論される。意味分類を細かくするにつれてより多様な推論ルールが記述できると思われるが、推論を行うために必要な最小限の意味分類は現時点でも確保できていると考えている。

本研究によってもたらされる主なメリットとしては、

- ・高精度な主観感情情報の処理の実現
- ・通信トラフィックの削減

などが挙げられる。推論ルールを適切に用いることで、相手の言いたいことを早期に、かつ適切につかむことができるため、相手の意図を再確認するなどの無駄な対話を省略でき、通信トラフィックの削減などが期待できる。

## 参考文献

- 1) 阿部修也, 乾健太郎, 松本裕治: 項の共有関係と統語パターンを用いた事態間関係獲得, 自然言語処理, Vol.17, No.1, pp.121-139 (2010).
- 2) Jonathan Berant, Ido Dagan and Jacob Goldberger: Global Learning of Typed Entailment Rules, Proceedings of the 49th ACL, pp.610-619 (2011).
- 3) Guglielmo Cinque: Adverbs and Functional Heads, OXFORD UNIVERSITY PRESS (1999).
- 4) Guglielmo Cinque: Restructuring and Functional Heads, OXFORD UNIVERSITY PRESS (2006).
- 5) Ido Dagan, Oren Glickman and Bernardo Magnini: The PASCAL Recognising Textual Entailment Challenge, Lecture Notes in Computer Science, 3944 (XIII), Springer, pp.177-190. (2006).
- 6) EDR(日本電子化辞書研究所): EDR 電子化辞書, <http://www.ijnet.or.jp/edr/> (1996).
- 7) 江口萌, 松吉俊, 佐尾ちとせ, 乾健太郎, 松本裕治: モダリティ, 真偽情報, 価値情報を統合した拡張モダリティ解析, 言語処理学会第 16 回年次大会発表論文集, pp.852-855 (2010).
- 8) 本田聖晃, 田辺利文, 吉村賢治, 首藤公昭: 非命題の意味解析のための日本語文末表現意味体系, 主観表現処理の最前線シンポジウム, 電子情報通信学会言語理解とコミュニケーション研究会 (2008).
- 9) 乾健太郎: 言語情報間の含意・矛盾関係の認識, 月刊言語, pp.30-37 (2008).
- 10) 乾健太郎, 藤田篤: 言い換え技術に関する研究動向, 自然言語処理, Vol.11, No.5, pp.151-198 (2004).
- 11) 伊佐治和也, 山田佳裕, 石原吉晃, 高松大地, 松本忠博, 池田尚志: 文節構造解析システム ibukiC, 言語処理学会第 11 回年次大会発表論文集 (2005).
- 12) Tomoko Izumi, Kenji Imamura, Genichiro Kikui, Satoshi Sato: Standardizing Complex Functional Expressions in Japanese Predicates: Applying Theoretically-Based Paraphrasing Rules, The Proc. of the Workshop on Multiword Expressions(MWE2010), pp.63-71 (2010).
- 13) Dekang Lin and Patrick Pantel: Discovery of Inference Rules for Question Answering, Natural Language Engineering, Vol.7, pp.343-360 (2001).
- 14) 榊田達也, 松吉俊, 佐藤理史: 日本語機能表現から多様な類義表現への言い換え, 言語処理学会第 14 回年次大会発表論文集 (2008).
- 15) 松吉俊, 佐藤理史, 宇津呂武仁, 日本語機能表現辞書の編纂, 自然言語処理, Vol.14, No.5 (2007).
- 16) Suguru Matsuyoshi and Satoshi Sato: Automatic Paraphrasing of Japanese Functional Expressions Using a Hierarchically Organized Dictionary, Proceedings of the 3rd International Joint Conference on Natural Language Processing (IJCNLP), Vol.1, pp.691-696 (2008).
- 17) 森田良行, 松木正恵: 日本語表現文型用例中心・複合辞の意味と用法, アルク (1989).
- 18) Heiko Narrog: Modality in Japanese The layered structure of the clause and hierarchies of functional categories, John Benjamins Publishing Company (2009).
- 19) Heiko Narrog: The order of meaningful elements in the Japanese verbal complex. Morphology, 20(1), pp.205-237 (2010).
- 20) Ivan. A. Sag, T. Baldwin, F. Bond, A. Copestake and D. Flickinger: Multiword Expressions; A Pain in the Neck for NLP, Proceedings of the 3rd CICLING (2002).
- 21) 首藤公昭, 田辺利文: 日本語の複単語表現辞書: JDMWE, 自然言語処理, Vol.17, No.5, pp.51-74 (2010).
- 22) Kosho Shudo, Toshifumi Tanabe, Masahito Takahashi, Kenji Yoshimura: MWEs as Non-propositional Content Indicators, Proceedings of the 2nd ACL Workshop on MWE, pp.32-39 (2004).
- 23) 首藤公昭, 田辺利文, 吉村賢治: 日本語モダリティ表現とその言い換え, 言語処理学会第 7 回年次大会ワークショップ論文集 (2001).
- 24) Idan Szpektor, Eyal Shnarch and Ido Dagan: Instance-based Evaluation of Entailment Rule Acquisition, The Proc. of ACL, pp.456-463 (2007).
- 25) 田辺利文, 本田聖晃, 高橋雅仁, 小山泰男, 吉村賢治, 首藤公昭: 日本語文末表現の取り扱いについて, FIT2006, pp.241-244 (2006).
- 26) Toshifumi Tanabe, Kenji Yoshimura and Kosho Shudo: Modality Expressions in Japanese and Their Automatic Paraphrasing, Proceedings of the 6th Natural Language Processing Pacific Rim Symposium (NLPRS), pp.507-512 (2001).