

## 日本語文末表現による推論について

田辺利文<sup>†</sup> 首藤公昭<sup>‡</sup><sup>†</sup>福岡大学工学部 <sup>‡</sup>福岡大学大学院工学研究科

{tanabe, shudo}@fukuoka-u.ac.jp

## 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 に関しては、2002 年の論文「Multiword Expressions: A Pain in the Neck for NLP」(Sag et al. 2002) を皮切りに、重要性が世界的に認識されてきた。実際 (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% であることが報告されており、MWE の適切な処理が自然言語処理の質を向上する上で必要不可欠であることを示唆している。

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、構成している単語の通常の意味から全体の意味を構成するのが難しい性質を意味する。日本語では、自立語性 MWE として『油を売る』『計画を立てる』『手をこまねく』などが、機能語性 MWE として「によって」「かもしれない」などがある。本論文では、日本語述部の文末に存在しうる機能語性 MWE にスポットを当てる。日本語文末には、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、バラエティに富んだ主観を表す表現が含まれ、かつ「なければならないようだ」のように複数個連続して用いられることも多く、発話者の主観などを推定するには機能語性 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関係(推論ルールと呼ぶ)について報告する。

## 2. 非命題の意味構造

一般的に文の意味は、命題の意味と非命題の意味からなっていると考えられ、日本語においては、非命題の意味を表す部分は述語に後接した、時制、判断、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。筆者ら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<sup>1</sup>。助述表現に含まれる機能語性 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。助述表現は、発話者の主観に基づく表現であるとみなすことができる。筆者ら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  $S_i$  が左分岐的に構成され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、並びを逆順にすることで、非命題の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。非命題の意味構造は、一般的に入れ子型構造として表現できる。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の意味構造は図 1 のようになる<sup>2</sup>。(1)、(2)に示される構文構造と、非命題の意味構造と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んだ複雑な文末の場合でも、意味関数との対応をとることで比較的容易に非命題の意味構造を求めることが可能である。

<sup>1</sup> 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。

<sup>2</sup> 図 1 の記号  $\cdot$  は通常の単語境界を表し、 $\cdot$  は MWE による単語境界を表す。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。

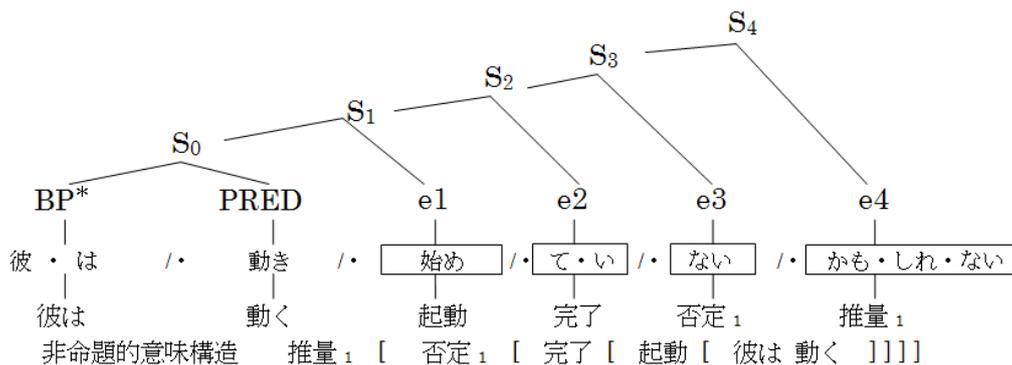


図1 日本語文の構造の概形及び非命題の意味構造

このように非命題の意味構造は、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<sup>3</sup>、言い換えや機械翻訳を行う際の間言語として有効であると考えている。

### 3. 推論ルール

第3章では機能語間レベルで定義できるとされる含意関係について考える。第2章で述べた非命題の意味構造は、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、非命題の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、非命題の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、非命題の意味構造Xを別の非命題の意味構造Yに、含意(推論)によって書き換えることができる場合、推論ルールX→Yが成り立っていると定義する<sup>4</sup>。推論ルールには、例えば“不足性[x]→願望[x]”や“不足性[x]→命令[x]”などがあり、現時点でのルール総数は約200種類である<sup>5</sup>。推論プロセスは、例えば『飲み足りない』は、非命題の意味構造“不足性[飲む]”に変換された後、推論ルール“不足性[x]→願望[x]”を適用し別の非命題の意味構造“願望[飲む]”に書き換え、『飲みたい』を生成する。一方、『飲み足りない』に対し別の推論ルール“不足性[x]→命令[x]”が適用された場合には『飲みなさい』を生成する。どの推論ルールが適用できるかは文脈に大きく依存するため、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。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。なお、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。推論ルールの一部および推論ルールの適用例を表1に示す。

<sup>3</sup> (Cinque 1999; Cinque 2006)は、非命題の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と報告している。

<sup>4</sup> 推論ルールは非命題の意味構造の集合間で定義される関係であり、反射律および推移律を満たすものと考えている。

<sup>5</sup> 推論ルール中のxは命題の意味を表すものとする。

### 4. 関連研究

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』のような「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<sup>6</sup>。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<sup>7</sup>。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、いっ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られる<sup>8</sup>。

(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<sup>9</sup>。「つつじ」に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」の応用として、(Izumi et al. 2010)では、複雑な日本語文末をシンプルにするための言い換え規則を提案している。

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)。

<sup>6</sup> (Narrog 2010)は、日本語による非命題の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。

<sup>7</sup> 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。

<sup>8</sup> 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。

<sup>9</sup> <http://kotoba.nuee.nagoya-u.ac.jp/tsumuji/>

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 <sup>1</sup> なかつた	理由付け <sup>x</sup>	過去 <sup>1</sup> 否定 <sup>1</sup> 願望 <sup>1</sup> 受動態 <sup>1</sup> <sup>x</sup> ]	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 <sup>x</sup>	方針・予定 <sup>x</sup>	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 <sup>1</sup> とも	成績は上がるはずだ		推量 <sup>2</sup> <sup>x</sup> ]	x: 自動詞	
食べにくい	食べたく <sup>1</sup> ない	困難性 <sup>x</sup>	否定 <sup>1</sup> 願望 <sup>x</sup> ]	x: 他動詞	
上がりにくい	上げたい		願望 <sup>x</sup> ]	x: 自動詞	
食べ <sup>1</sup> てくれる <sup>1</sup> な	食べる <sup>1</sup> な 食べない <sup>1</sup> でほしい	禁止 <sup>1</sup> 受動態 <sup>2</sup> <sup>x</sup> ]	禁止 <sup>x</sup> ] 願望 <sup>1</sup> 否定 <sup>x</sup> ]	x: 他動詞 x: 他動詞	「てくれる」がなくなる
うどんを <sup>1</sup> 食べなければならぬ	うどんを <sup>1</sup> 食べてい <sup>1</sup> ない	必要性 <sup>1</sup> <sup>x</sup> ]	否定 <sup>1</sup> 進行 <sup>x</sup> ]	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 <sup>1</sup> 食べなければならぬ	ナマコを <sup>1</sup> 食べたく <sup>1</sup> なかつた	過去 <sup>1</sup> 必要性 <sup>1</sup> <sup>x</sup> ]	過去 <sup>1</sup> 否定 <sup>1</sup> 願望 <sup>x</sup> ]	x: 他動詞	モダリティと主観表現が絡む
飲みすぎる	飲み <sup>1</sup> な	過剰性 <sup>x</sup>	禁止 <sup>x</sup> ]	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
	飲み <sup>1</sup> けいこうがある		傾向 <sup>x</sup> ]	x: 他動詞	
	飲み <sup>1</sup> すぎる <sup>1</sup> な		禁止 <sup>1</sup> 過剰性 <sup>x</sup> ]	x: 他動詞, 主語 2 人称	
	飲み <sup>1</sup> すぎ <sup>1</sup> ない <sup>1</sup> ほうがいい		必要性 <sup>1</sup> <sup>x</sup> ]	x: 他動詞, 主語: 2 or 3 人称	
飲み <sup>1</sup> すぎ <sup>1</sup> た	飲み <sup>1</sup> たく <sup>1</sup> ない	過去 <sup>1</sup> 過剰性 <sup>x</sup> ]	否定 <sup>1</sup> 願望 <sup>x</sup> ]	主語 1 人称	
飲み <sup>1</sup> たりない	飲み <sup>1</sup> たい	不足性 <sup>x</sup>	願望 <sup>x</sup> ]	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 <sup>1</sup> 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 <sup>1</sup> なさい		命令 <sup>x</sup> ]	x: 他動詞, 主語 2 人称	
	飲み <sup>1</sup> ほうがいい		必要性 <sup>1</sup> <sup>x</sup> ]	主語: 2 or 3 人称	
	飲み <sup>1</sup> だ		過去 <sup>x</sup> ]	x: 他動詞	
	飲み <sup>1</sup> せろ		使役態 <sup>1</sup> <sup>x</sup> ]	x: 他動詞	
飲み <sup>1</sup> たりない <sup>1</sup> です	飲み <sup>1</sup> ましょう	丁寧 <sup>1</sup> 不足性 <sup>x</sup> ]	勧誘 <sup>x</sup> ]	x: 他動詞	

表1 推論ルールと適用例(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 は命題的意味を表す)

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。含意関係認識の例として、(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”などを挙げている。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。例えば、(中川ら 1995)では、アスペクトを表す日本語文末表現である「テイル」「テアル」「テオク」「テミル」「テシマウ」について、省略された主語などの推定ができることを言っており、日本語は省略が多いことから、特に日英機械翻訳の精度向上に応用可能であることが示されている。また、(本田ら 2008)では、日本語文末表現意味体系を提案し推論ルールをいくつか紹介している。しかしいずれの研究も少数の日本語文末表現に対してのみであり、数多くの日本語文末表現に対し推論ルールを作成したものは現時点では見当たらないようである。(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考えている。

## 5. おわりに

非命題的意味を適切に抽出することで、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的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的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<sup>1</sup> <sup>x</sup>]”、“疑問<sub>3</sub>[x]→願望<sup>x</sup>”、“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<sup>1</sup>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メリットを整理してみる。

- ・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

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。

一方、これからの課題は次の通りである。

- ・推論ルールのカバレッジ向上
- ・推論ルールの適用条件の詳細化
- ・性能評価

**推論ルールのカバレッジ向上**：现阶段では約 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。

**推論ルールの適用条件の詳細化**：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や要求されるなど、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。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ふ。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。

**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。

## 謝辞

2010 年度 NLP 若手の会(YANS)におきまして、コメントを下さった方々に対し、この場を借りて感謝いたします。

## 参考文献

阿部修也, 乾健太郎, 松本裕治. 2010. 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

Dekang Lin and Patrick Pantel. 2001. Discovery of Inference

Rules for Question Answering. Natural Language Engineering. Vol.7. pp.343-360.

江口萌, 松吉俊, 佐尾ちとせ, 乾健太郎, 松本裕治. 2010. 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

Guglielmo Cinque. 1999. Adverbs and Functional Heads. OXFORD UNIVERSITY PRESS.

Guglielmo Cinque. 2006. Restructuring and Functional Heads. OXFORD UNIVERSITY PRESS.

Heiko Narrog. 2009. Modality in Japanese The layered structure of the clause and hierarchies of functional categories. John Benjamins Publishing Company.

Heiko Narrog. 2010. The order of meaningful elements in the Japanese verbal complex. Morphology, 20(1). pp.205-237.

本田聖晃, 田辺利文, 吉村賢治, 首藤公昭. 2008. 非命題的意味解析のための日本語文末表現意味体系. 電子情報通信学会主観表現処理の最前線シンポジウム. pp.39-44.

Idan Szpektor, Eyal Shnarch and Ido Dagan. 2007. Instance-based Evaluation of Entailment Rule Acquisition. The Proc. of ACL, pp.456-463.

Ido Dagan, Oren Glickman and Bernardo Magnini. 2006. The PASCAL Recognising Textual Entailment Challenge. Lecture Notes in Computer Science, 3944 (XIII), Springer, pp.177-190.

乾健太郎. 2008. 言語情報間の含意・矛盾関係の認識. 月刊言語. pp.30-37.

Ivan A. Sag, Timothy Baldwin, Francis Bond, Ann Copestake and Dan Flickinger. 2002. Multiword Expressions: A Pain in the Neck for NLP. The Proc. of the 3rd CICLING. pp. 1-15.

松吉俊, 佐藤理史, 宇津呂武仁. 2007. 日本語機能表現辞書の編纂. 自然言語処理. Vol.14. No.5. pp.123-146.

森田良行, 松木正恵. 1989. 日本語表現文型用例中心・複合辞の意味と用法. アルク.

中川裕志, 森辰則. 1995. 日本語マニュアル文におけるテイル、テアル、テオク、テミル、テシマウの語用論. 自然言語処理. Vol.2. No.4. pp.19-35.

首藤公昭, 田辺利文. 2010. 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.

田辺利文, 本田聖晃, 高橋雅仁, 小山泰男, 吉村賢治, 首藤公昭. 2006. 日本語文末表現の取り扱いについて. FIT2006. pp.241-244.

Tomoko Izumi, Kenji Imamura, Genichiro Kikui and Satoshi Sato. 2010. Standardizing Complex Functional Expressions in Japanese Predicates: Applying Theoretically-Based Paraphrasing Rules. The Proc. of the Workshop on Multiword Expressions(MWE2010). pp.63-71.