

2012年度 卒業論文

日本語文における論理式変換の実装と課題調査

2013年3月31日

情報知能システム総合学科

(学籍番号: A9TB2022)

稲田 和明

東北大学工学部

## 概要

近年、自然言語処理において、係り受け解析や述語項構造解析など、様々な解析が高精度化し、その結果は、含意関係認識や談話解析などの、より高度なタスクにも使われてきている。しかし、意味的に高度な理解を要求されるこれらのタスクでは、世界知識との照合や推論を行う必要があるが、係り受け解析や述語項構造解析の出力は、それらの処理と親和性が高いとは言えない。そこで我々は、一つの解決策として、従来の解析器の出力を、推論を扱いやすい論理式へ変換することを考える。本稿では、日本語文を論理式へ変換する処理についてその実装を模索し、また、いくつかの例を分析することで、実装の際に生じるであろう問題点を整理する。また、調査結果を踏まえ、現状の解析器で実現可能な論理式変換器を実装する。

# 目次

第1章	序論	1
第2章	研究背景	3
2.1	論理式への変換器	3
2.1.1	論理式変換器 Boxer	4
2.2	neo-Davidsonian 論理形式	4
2.3	構文解析	5
2.4	拡張モダリティ解析	6
2.5	正規化への取り組み	7
第3章	日本語論理式変換器の実装へ向けての調査	8
3.1	調査方法	8
3.1.1	変換処理	8
3.1.2	調査対象データ	9
3.2	調査結果	10
3.2.1	変換の問題	10
3.2.2	正規化の問題	15
第4章	日本語文の論理式変換器の実装	19
4.1	使用した構文・意味解析器	19
4.2	実装した変換プロセス	20
第5章	まとめ	21

# 第1章 序論

本稿では、自然言語文の意味構造を表現する手段としての論理式に着目し、現状では存在しない、日本語文を論理式へ変換する処理についてその実装を模索し、また、いくつかの例を分析することで、実装の際に生じるであろう問題点を整理し、現状で実現できる部分に関して、日本語文の論理式変換器を作成した。

近年、Web などにより電子化されたテキストが広く一般的に使用され、また言語処理を行うための情報が書き加えられたデータが増加している。自然言語処理に利用可能なこれらのデータが増大してきたことを受け、係り受け解析や述語項構造解析など、様々な解析が高精度で行えるようになってきており、その結果は、含意関係認識や談話解析などの、より高度な自然言語処理に使われてきている。

しかし、意味的に高度な理解を要求されるこれらのタスクでは、文の表層上の繋がりだけでなく、人間の持つ一般的な常識の集合である世界知識との照合や、その知識から得られる情報を用いて推論を行う必要があるが、現状の係り受け解析や述語項構造解析の出力は、一般的に木構造をしているため、部分的な書き換えや演算を行う必要があるそれらの処理と親和性が高いとはいえない。

そこで我々は、一つの解決策として、従来の解析器の出力を、推論を扱いやすい論理式へ変換することを考える。実際に、英語圏では既に文を論理式へと変換するツール [1, 2] が公開されており、この結果を含意関係認識や談話解析などの応用的なタスクへ利用した研究が報告されている。しかし、日本語に関しては、文を論理式へと変換するためのツールは存在しておらず、また、文法の異なり度合いなどの言語的な差異から、日本語に特有の問題が存在することも容易に想像できる。そこで、日本語における基本的な言語解析器が高精度化してきたことを受け、現状の解析器を組み合わせることによって、適切な意味表現を持つ論理式へと変換することができるのではないかと考え、現状の解析器で実現可能かを調査した。

このような背景から、我々は、日本語の文を、推論を取り扱い易い論理式へと変換することを目標にし、その初期段階として、人手による変換結果をもとに、変換に伴う問題の調査と整理を行った。調査では、現状の解析器と極力単純なアルゴリズムを使用することで、現状の解析器により適切な意味を持つ論理式を得ることが可能か、もし適切な論理式が得られない場合には、どのような追加の処理、解析が必要になるのかを調査した。

本稿では、妥当な論理式へと変換するための仕様を作成するため、以下の2点について考える。

1. 形態素、統語係り受け、述語項構造、拡張モダリティが正しく解析できると仮定し、その結果から論理式へ変換できるかを吟味し、問題点を洗い出す。
2. 推論を行いやすくするため、同義な表現をある程度吸収しながら論理式に変換する方法を模索する。

1では、文に人手で解析結果の正解を与え、意味的に正しい論理式が得られるプロセスが設計できるかを調査した(以降、変換の問題と呼ぶ)。2では、表層が似ている同義な文対を人手で論理式に変換し、類似表現を吸収しようとする際の問題点を調査した(以降、正規化の問題と呼ぶ)。

調査の結果、変換の問題では、精密な仕様を設けることによって、半数近くは現状の解析器の結果から、意味表現として適切な論理式を得ることができたが、現状の解析器では対応できない文や、一部の意味は汲み取れているが、意味的に厳密な論理式を得るために、より深い解析が望まれる文も見られた。また、正規化の問題では、正確な述語項構造解析により、半数近くの文対を統一の論理式へと吸収することができそうだが、そのほかの文では、正規化するための一般的な方法を見つけることはできなかった。

また、調査の結果を受けて、現状の解析で実現可能な部分を論理式へと変換する論理式変換器を実装した。

本稿では、まず論理式変換の関連研究と論理式による意味表現、及び現状の構文・意味解析について概観する(2章)。次に論理式変換の実装へ向けて実施した調査の方法と結果を述べ(3章)、調査結果を反映した日本語論理式変換器の概要について説明する(4章)。

## 第2章 研究背景

### 2.1 論理式への変換器

英語圏では、統語解析器の結果を論理式へと変換し、その論理式に対して、世界知識との照合や、推論を行うことで、質問応答や含意関係認識を行う研究が実施されており、論理式による意味表現が、高度な自然言語処理に有用であることが示されている [9, 10]. しかし、これらの研究では、それぞれ質問応答と含意関係認識に使用先を限定して論理式変換を行っており、さらに、論理式への変換の際に、主に統語構造のみの情報から変換を行っているため、文内の情報全てを読み取れないのは明らかである。日本語圏でも同様に、論理式を用いた応用的なタスクへの取り組みが行われており、その一つとして、論理式を用いた日英間の翻訳を行う ATLAS というソフトウェアが富士通ソフトより販売されている。

また、統語構造の強力な記述能力を持つ枠組みの一つである Combinatory Categorical Grammar (CCG, 組合せ範疇文法) を統語解析、意味表現として一階述語論理よりも表現能力の優れた高階動的論理を採用した日本語文法理論が構築されている [15, 16]. 現状では、まだこの文法理論を取り入れた統語解析器は実装されていないが、解析器が完成したあかつきには、日本語文から高階動的論理として意味表現を得ることができる変換器となる。しかし、意味表現として採用している高階動的論理は、一階述語論理よりも精密な意味を捉え、表現することができるが、その代償として、得られた論理式による推論は、現実的な計算時間で処理できなくなってしまう。そこで本稿では、論理式変換後に効率の良い推論を行うことを想定して、一階述語論理へ変換することを目標に調査を行った。

一方、吉本らは、論理式による意味表現を伴う日本語のツリーバンクとして、櫛コーパスを作成している [19]. 櫛コーパスは、ペン通時コーパスの解析規約に従って、フラットな文構造を採用し、統語構造と文の意味を一階述語論理で記述した情報を保持した大規模なツリーバンクである。彼らの主張では、日本語文をツリーバンクが構成する統語情報で解析することができれば、Butler のスコープ制御理論 [4] により、意味表現として一階述語論理へと変換することができると述べている。しかし、櫛コーパスは現状の統語、品詞体系と異なるため、現在の体系で開発されている意味解析を導入するために、余分なコストが必要となることが予想される。さらに、現状存在する各解析器は高い性能を持っているため、新たな文法体系と論理式の組を用いた新たなツリーバンクに頼る必要はなく、意味表現として適切な論理式を得られる可能性は十分にあると考えられる。このような仮説から、文の持つ意味情報を論理式へと変換するためには、どの程度の意味解析を行う必要があるのかを調査することが、本稿の目的の一つであり、統語構造・意味情報を持つツリーバンク作成とは少し興味が異なる。

### 2.1.1 論理式変換器 Boxer

英語圏では、統語解析に CCG、意味表現に Discourse Relation Structure (DRS) を用いた「自然言語文 論理式」の変換を行うツール Boxer が Bos らにより公開されている [1, 2]. このツール内での意味表現は DRS を用いているが、DRS へ記述する論理式には、一階述語論理、neo-Davidsonian 形式を用いており、簡単に一階述語論理の枠組みへと変換可能なため、得られた論理式を用いて効果的な推論を行うことができる。以下、図 2.1 に、Boxer の Graphical 形式の出力例を示す。

Boxer から得られる論理式の利用先として、Bos らは含意関係認識の研究を行っている [3]. さらに近年の研究では、Boxer の出力と大規模な世界知識を用いて、Cost Based Abduction (重み付き仮説推論) による含意関係認識や談話解析の研究が行われており、それなりの成果を上げている [6, 13]. このような点からも、「自然言語文 論理式」変換器の存在は自然言語の応用的な解析の発展にとって、一つの重要な要素であると言え、実際に日本語における論理式変換器の実現を望む声も多いが、現在、日本語圏では Boxer のような「自然言語文 論理式」の変換を行うツールは存在していない。

## 2.2 neo-Davidsonian 論理形式

自然言語処理において、言語の意味を表現する方法は、論理式、意味ネットワーク、フレームベースなどの様々な表現方法があるが、今回は論理式の中でも表現がシンプルで、効率の良い推論が期待できる一階述語論理を用いる。論理式には、一階述語論理よりも詳細な表現が可能な、二階述語論理や動的述語論理などの形式も存在するが、それらの高度な表現では、現状、推論を効率的に行うことができないため、本稿では、変換先の論理式として一階述語論理を採用する。以下では、一階述語論理による意味表現の基本的な概念や手法について述べる。

自然言語には、文中に主語や補語といった概念があり、その間の関係は文の意味を規定する重要な役割を持っているが、述語論理は、そのような自然言語文の構成関係を表現するための記述能力を持つ論理である。中でも一階述語論理は、論理公理と推論規則を適切に導入することで、現実的な時間での推論が可能となる論理体系である。自然言語を一階述語論理上で表現するときは、個々の物体には変数、及び定数を与え、動詞や形容詞などの動作や状態を示す事象は述語として表現し、また事象間の関係を関数で表す。以下に、一階述語論理による自然言語の表現の例を示す。以降、自然言語文に対応する論理式表現を LF と記述する。

「日本人はおにぎりを食べる」

LF :  $\exists x$  日本人(x)  $\wedge$  食べる(x, おにぎり)

一階述語論理では「食べる(ガ格, ヲ格)」のように、述語や関数の持つ引数の数を予め決めておく必要がある。しかし、自然言語では、動作や状態の持つ引数、即ち「誰が、何を、どこで、いつ」など関連する事象の数は、操作や状態ごとに固有の数ではないため、動作や状態が持つ引数のパターンごとに別の述語を用意しなければならない。

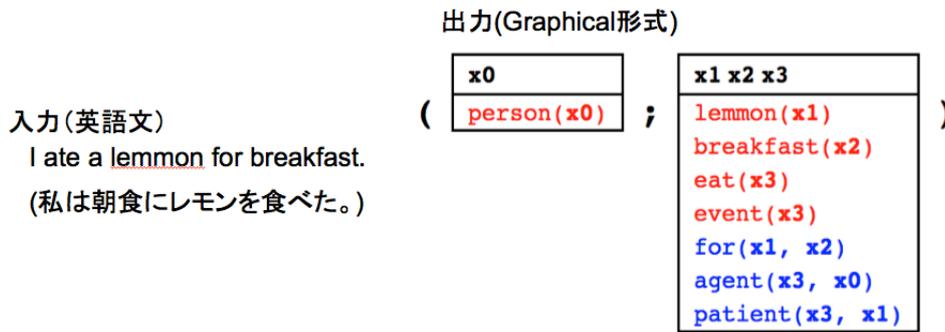


図 2.1: Boxer の出力例

- A: 「日本人はおにぎりを食べる」 ⇒ 食べる (ガ格, ヲ格)  
 B: 「日本人はおにぎりを教室で食べる」 ⇒ 食べる (ガ格, ヲ格, デ格)  
 C: 「日本人は昼におにぎりを教室で食べる」 ⇒ 食べる (ガ格, ヲ格, デ格, 二格)

このように「食べる」だけでも多様な述語を用意する必要があるため、日本語すべての動作や状態を示す語に対し、予め全ての引数パターンを書くことは困難である。また、推論を行う際にも、別々に定義した述語を同一の意味であることを示すために、余分な演算が必要となり不便である。

そこで、1967年、哲学者 Donald Davidson が提唱した Davidsonian 論理形式を、Parsons が部分的に書き換えた neo-Davidsonian 論理形式という表現方法を導入する [11]。neo-Davidsonian 論理形式では、動詞や形容詞などの動作や状態を示す事象に特別な変数  $e$  を与え、変数  $e$  と関連する事象一つを引数とする述語で関係を表現し、その関係を論理積 (AND,  $\wedge$ ) で繋ぎ合わせることで文の意味を表す。以下に neo-Davidsonian 論理形式に従い、意味表現を行った例を示す。

「日本人はおにぎりを教室で食べる」

LF:  $\exists x, e$  日本人 ( $x$ )  $\wedge$  食べる ( $e$ )  $\wedge$  ガ ( $e, x$ )  $\wedge$  ヲ ( $e$ , おにぎり)  $\wedge$  デ ( $e$ , 教室)

neo-Davidsonian 論理形式では、引数を予め定める必要がないため、動詞の持つ格や修飾語の数が一定でないなど、自然言語の特徴を表現するのに適する。そこで、本稿ではこの neo-Davidsonian 論理形式で記述された一階述語論理へと変換することを目標に、変換器の実装へ向けての調査を行った。

## 2.3 構文解析

本稿では、構文解析器として、形態素、統語係り受け、述語項構造が理想的に解析できると仮定した上で調査を行った。以下でそれぞれの解析に関して、概要を説明する。

形態素解析とは、文を形態素に分離し、その形態素がどのような基本形を持ち、何の品詞であるかなどを解析することである。日本語では、英語などの言語のようにスペースで形態素が区切られていないため、形態素解析は単語を認識するための処理として、最も基本的、かつ重要な言語処理である。論理式へと変換する際にも、文を形態素単位に分離することで、論理式の「述語」や「変数」を振り分けることが可能となり、また品詞解析により、その形態素の文法的機能がわかるため、助詞などの機能的な働きをする語か、名詞や動詞などの何らかの実態を持つ語かを判

断し、変数を割り振るかどうかを決定することができる。

統語係り受け解析とは、形態素を文節にまとめ上げ、それらの文節がどの文節と係り受け関係にあるかを解析することである。形態素を文節にまとめ上げ、文節間の係り受け関係を調べることは、文中の事象間の関係を知るための一つの指標となる。係り受け関係は事象間に何らかの関係があることを示し、係り受け元の末尾に存在する助詞などの機能語によって、その関係の意味が表現されている。また、文節にまとめ上げることで、名詞句の判定などの処理が行えるため、より明確に変数の割り振りが可能となる。

述語項構造解析とは、文章内の動作や状態を見つけ、それらについての「誰が、何を、どこで、いつ」などの関係する事象を探し、その役割を判断することである。動作や状態と事象の関係を認識することで、係り受け関係だけでは捉えられない意味的な関係性を捉えることができる。日本語においては、主に格助詞を伴い、述語と補語の関係を表現することが多いが、格助詞の省略、必須項の省略、文をまたぐ関係など様々な問題がありうるため、それらを適切に取り扱う処理を想定した上で、論理式変換を考慮する必要がある。

## 2.4 拡張モダリティ解析

本稿では、拡張モダリティの枠組みによる意味解析を導入し、その解析結果を意味情報として論理式に反映する。以下で拡張モダリティに関しての概要を述べる。

拡張モダリティとは、文章中の事象に対する情報発信者の主観的な態度(モダリティ)に加え、真偽判断や価値判断などを統合した事象の情報である[14]。この拡張モダリティの解析は、文章に記述されている事象が、実際に成立した事実であるのか、それとも、成立しなかったことであるのか、もしくは、成立を望んでいるだけであるのかなど、文章中に書き記された意味情報を、より深く読み取ることができる。

A:「この薬品の使用を中止したい」

B:「この薬品の使用しないこともない」

⇒ A、B両方とも事象「この薬品を使用」は成立している。

上記の例のように、「この薬品を使用」という事象が成立しているという情報は、含意関係認識などの高度な言語処理を解く上で、大変重要な意味情報となる。人間であれば、A、Bどちらも成立していることを簡単に判断できるが、否定表現がそれぞれAには1個、Bには2個入っている点からも分かるように、機械的に否定表現の数を数え上げるなどの単純な方法では事象の成立、不成立を判断できない。

このような拡張モダリティの情報を解析するシステムとして、江口らは周囲の文脈の情報と、手がかり表現辞書を用いた解析器を作成しており[12]、それなりの精度を上げていることから、本稿では江口らが提案している拡張モダリティ体系を参考にしている。江口らは拡張モダリティの構成要素として、態度表明者、時制、仮想、態度、真偽判断、価値判断、焦点の七つを提案している。今回は、文中に二つ以上の事象が存在するときに、それらの時間関係を知る手がかりと成り得るであろう「時制」、また、単純な否定形を含め、事象が実際に起きたかどうかを示す情報となる

真偽判定の二つのタグが、文の意味表現として重要かつ有用であると考え、論理式変換器に拡張モダリティ解析を取り入れることとした。

## 2.5 正規化への取り組み

高度な言語処理に向けての正規化への取り組みとして、日本語文に構文・格・省略解析を行い、同義な表現を一つの表現へと吸収した後、含意関係認識を行う研究が存在する [18]。しかし、「大きい」と「小さくない」は意味的に差があるのは明らかだが、これらを統一の表記「大きい」へと変換しているなど、厳密には意味の異なる表記を正規化している例が見受けられるため、含意関係以上に精密な意味を捉える必要がある高度な言語処理においては、彼らの手法を用いることができない。本稿が目指す論理式変換器の出力となる論理式は、含意関係認識よりも、さらに厳密な意味を必要とするタスクでの使用も想定しているため、彼らの研究よりも意味的な厳密さを考慮した上で、統一の表記へと吸収できないかを調査した。

そのほかの正規化への取り組みとして、文末の機能語表現の同義語辞書を作成し、同義な文末表現を一つの表記へと正規化する研究が行われている [21]。ただし、この文末機能語表現の正規化は、今回の意味解析として採用した拡張モダリティ解析の一部であると考えられるため、機能語表現の正規化は行わずに調査した。

## 第3章 日本語論理式変換器の実装へ向けての調査

本稿では、論理式へ変換する際、現状の構文・意味解析器を用いることで、意味的に正しい論理式へと変換することができるか(変換の問題)と、推論を行いやすくするため、論理式変換時に同義な表現を一つの論理式へ吸収することができるか(正規化の問題)の二つの問題の調査を行った。この章では、論理式へと変換する手順の設定や調査の対象としたデータの特徴など、調査方法について説明し(3.1節)、変換の問題と正規化の問題の調査結果を述べる(3.2節)。

### 3.1 調査方法

現状の構文・意味解析が理想的にできると仮定した上で、変換アルゴリズムに従い、解析結果を論理式へと変換する。初期段階の調査のため、変換アルゴリズムは、構文・意味解析器から得られた情報を直接的に論理式で表現する簡潔なものとした。多様な文に対応できるアルゴリズムは、今回の調査をふまえた上で吟味し、実装のときに取り入れることとする。実装の際に用いたアルゴリズムに関する詳細は、第4章を参照されたい。

#### 3.1.1 変換処理

論理式変換の最初のステップとして、以下の簡潔な方法で変換を試みる。変換処理の全体像は、図3.1を参照されたい。

1. 文を、形態素解析、係り受け解析、述語項構造解析、拡張モダリティ解析にかける。
2. 形態素解析における品詞タグ付けの結果から、各形態素に存在限量された変数を割り当てる。述語となる動詞や形容詞には変数  $e$  を、名詞などの内容語には変数  $x$  を割り当て(例えば、花( $x$ ))、助詞などの機能語には変数を割り当てない。なお、述語、及び関数名には原型を使用する。
3. 述語項構造解析の結果から、格助詞名を述語とし、格関係を表す論理式を加える(例えば、ガ( $e, x1$ ))。
4. 拡張モダリティ解析の結果から、拡張モダリティタグ名を述語とし、その解析結果を表す論理式を加える(例えば、真偽( $e, 成立$ ))。
5. 文節と係り受け関係解析を、係り受け元の文節末に存在する助詞を述語として、係り関係の論理式を加える(例えば、の( $e, x1$ ))。ただし、すでに述語項構造解析の結果で、論理式に書き加えられている関係は除く。

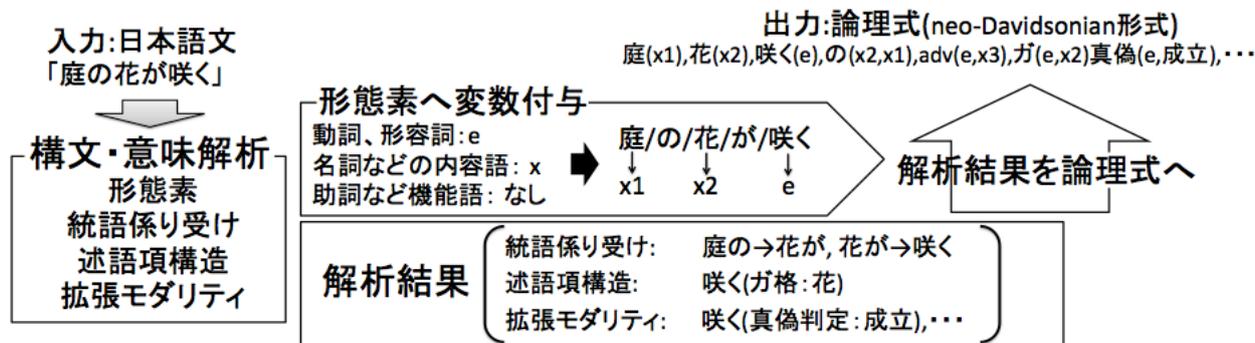


図 3.1: 変換プロセスの全体像

以下に、上記の流れで変換した論理式の例を示す。

「私はケーキを食べる」

LF: 私 ( $x_1$ )  $\wedge$  ケーキ ( $x_2$ )  $\wedge$  食べる ( $e$ )  $\wedge$  ガ ( $e, x_1$ )  $\wedge$  ヲ ( $e, x_2$ )

「羽田の空港へ行った」

LF: 羽田 ( $x_1$ )  $\wedge$  空港 ( $x_2$ )  $\wedge$  行く ( $e$ )  $\wedge$  の ( $x_2, x_1$ )  $\wedge$  ニ ( $e, x_1$ )  $\wedge$  相対時 ( $e$ , 非未来)  $\wedge$  真偽 ( $e$ , 成立)

以降では、この方法では対応できない文や、表現として問題となる点を調査するため、実際の日本語文を分析する。また同時に、意味的に同義な表現を一つの論理式へと変換できないかを調査する。

### 3.1.2 調査対象データ

本稿が目指す論理式変換器の出力は、実際の含意関係認識などの応用的なタスクで使用することを想定するため、京都大学の黒橋・河原研究室が公開している RTE 評価データセットを調査対象として用いた [17]。また、このデータセットは、比較的簡単な日本語で構成されているため、初期の調査対象として適当であると言える。以下にそのデータの一例を示す。

ID	カテゴリ:サブカテゴリ	推論判定	テキスト (t)	仮説 (h)
15	語彙 (体言):下位 上位		あの人は呼吸器専門医だ	あの人は医者だ。

評価セットは 2,472 セットからなり、それぞれ包含、語彙 (体言)、語彙 (用言)、構文、推論の五つのカテゴリに分類されている。またそれぞれの評価セットには、 $\wedge$ 、 $\vee$ 、 $\rightarrow$ 、 $\times$  の四段階の推論判定が付与されている。 $\rightarrow$  は、 $t$  が真であったとき  $t$  の情報から  $h$  が必ず真であるといえる場合であり、論理的な推論において、演繹と呼ばれるものである。以下、 $\rightarrow$ 、 $\vee$  の順に、 $t$  が真であったとき  $h$  が真となる可能性が下がっていき、 $\times$  は  $t$  が真であったとき  $h$  が全くの誤りだと分かるときや  $t$  と  $h$  に全く関連性がない場合に付与される。

## 3.2 調査結果

### 3.2.1 変換の問題

変換の問題では、データセット内の  $t$  の文全体を観察しながら、3.1.1 で示した変換方法に従い、手動で変換を行うことで、論理式へと変換する際に問題となる部分を洗い出した。

#### 副詞の修飾 (115 文)

係り受け関係が存在するが、副詞を含む係り受け元の末尾に助詞がないため、現状の解析結果を直接的に変換する簡潔なアルゴリズムでは、係り受け関係が失われてしまう。そのため、対応策の一つとして、特別な述語 ( $adv:adverb$ ) を用いて係り受け関係を表現することが考えられる。

「彼はすぐ怒る」

LF:  $\text{すぐ}(x) \wedge \text{怒る}(e) \wedge \text{adv}(e, x), \dots$

一方で、副詞は動詞に割り当てられた変数  $e$  の一つの性質とも考えられるため、係り先の動詞と同一の変数を副詞に与えることで、関係を表現することもできる。

「彼はすぐ怒る」

LF:  $\text{すぐ}(e) \wedge \text{怒る}(e), \dots$

上記では「副詞の修飾」における二つの対応策を挙げたが、動詞や形容詞に特別な変数を与えることを考慮すると、動詞や形容詞のみへ修飾する副詞には、二つ目の方法で副詞自身にも修飾先と同じ動詞や形容詞の変数を与える方が、意味的に自然であると考えられる。

#### 連体節の修飾 (194 文)

連体節では係り受け関係が存在するが、係り受け元の文節末尾に助詞がないため、副詞と同様の問題が発生する。しかし連体節では、修飾元と修飾先が述語と補語の関係となる「内の関係」の場合、述語項構造解析により格助詞関係を持つことがある。

「タバコを吸っている人」

LF:  $\text{吸う}(e) \wedge \text{人}(x) \wedge \text{ガ}(e, x), \dots$

一方、内の関係を持たない場合は、修飾する節全体が、修飾先の体言の内容を表す「外の関係」となる。そのため、外の関係では内の関係のように格助詞関係がなく、また係り受け元の文節末に助詞がないため、特別な述語 ( $atc:attribute\ clause$ ) を用いて、以下のように係り受け関係を表現する。

「教育を受ける義務」

LF: 受ける ( $e$ )  $\wedge$  義務 ( $x$ )  $\wedge$   $atc(x, e), \dots$

上記のように、内の関係と外の関係で修飾の意味合いが異なるため、論理式上での意味表現を分離すべきである。しかし、内の関係は格助詞関係を持ち、外の関係は格助詞関係を持たないため、理想的に述語項構造解析を行うことができれば、内の関係と外関係を分離することができる。従って、格助詞関係を持つ場合は、内の関係として格助詞関係を論理式に反映し、一方、格助詞関係を持たない場合には、外関係として係り受け関係を論理式へと変換することで対応する。ただし、外関係を論理式で表現する際、係り受け元は動詞のため、述語となる助詞が存在しない。そのため外関係では、上記のように特別な述語  $atc$  を用意し、係り受け関係を論理式として表現する必要がある。

### 連体詞の修飾 (105 文)

連体詞は、係り受け関係の修飾元に助詞が存在しないため、係り受け関係の表現には、特別な述語 ( $att:attribute$ ) を用意する必要がある。

「大きな川のそばだ」

LF: 大きな ( $x_1$ )  $\wedge$  川 ( $x_2$ )  $\wedge$   $att(x_2, x_1), \dots$

しかし連体詞の中には、以前に出現した語（先行詞）を指し示す「指示連体詞」が存在する。指示連体詞には、先行詞と指示連体詞及びその修飾先が、同義表現や上位下位関係であるなど、直接的な指示の関係にある場合を「直接照応」といい、先行詞と指示連体詞及びその修飾先が部分全体関係や属性関係であるなど、間接的な指示の関係にある場合を「間接照応」といい、二種類の照応関係が存在する。

直接照応: 「図書館で資料を手に入れた。このデータは掲載される」

⇒ 「このデータ」が「資料」を指し示す。

間接照応: 「広瀬川の調査を行った。このデータは掲載される」

⇒ 「この」は「調査」を指し、「データ」を修飾する。

上記のように、直接照応と間接照応で修飾の振る舞いが異なるため、論理式上ではこれらを区別して表現すべきである。しかし、指示連体詞と係り受け先の語の関係からのみでは、先行詞と指示連体詞及びその修飾先が同一のものを指す直接照応か、ただ関連しているだけの間接照応なのかが判断できない。

直接照応は、指示連体詞の先行詞と係り受け先が同一の事象を指し示すことを考慮し、以下のような論理式による意味表現が考えられる。

「図書館で資料を手に入れた. このデータは掲載される。」

LF : 資料 ( $x$ )  $\wedge$  データ ( $x$ ), ...

間接照応は、指示連体詞の先行詞が係り受け先の文節を修飾していると考えられるため、以下のような、特別な述語 (att: attribute) を用いて論理式による意味表現を行う。

「広瀬川の調査を行った. このデータは掲載される」

LF : 調査 ( $x_1$ )  $\wedge$  データ ( $x_2$ )  $\wedge$  att( $x_2, x_1$ ), ...

上記のように指示連体詞を含む場合、意味的に正しい論理式を得るためには、直接照応と間接照応を判断した上での照応解析が必要となる。しかし、照応詞がどの先行詞を指すかを当てる解析は既に研究されている [5] ため、直接照応か間接照応かを照応詞から判断することで、解決できると考えられる。

### コピュラ文の表現 (145 文)

コピュラ文とは「X は Y だ」のように、主語 X とその後に置かれる語 Y を結ぶ表現を指す。コピュラ文には、 $X=Y$ ,  $Y=X$  と交換可能で、X と Y が一致する「指定」、 $X=Y$  とはできず、Y が X の属性を表す「措定」の二種類がある。しかし、現状の解析器では、述語項構造解析により、これらのコピュラ文はどれも格助詞関係として出力され、区別できない。

「オリオン座は冬の星座だ」

LF : オリオン座 ( $x_1$ )  $\wedge$  星座 ( $x_2$ )  $\wedge$  ガ ( $x_2, x_1$ ), ...

より厳密な意味を得るためには、コピュラ文の種類を表す特別な述語を用することで、統一の表記へと変換する方法が考えられるが、そのためには指定と措定を区別するための解析器を取り入れる必要がある。しかし、コピュラ文の種類を区別するためには、外部知識の導入や文脈情報の捕捉が重要であると考えられるが、これらの処理を論理式変換後に行う推論に担当させるべきか、論理式変換前により詳しい解析を行うべきかは悩ましい問題である。

### 数量表現 (95 文)

文内に数量表現を伴うとき、以下のように文節内に助詞を伴わない上、数量とその数量が指し示すもの間に直接係り受け関係がない場合がある。

「栗を 7 つ拾った」 係り受け : 栗を 拾った, 7 つ 拾った

しかし、単純に関係を表現できないだけでなく、数量が何を指し示しているかといったスコープの問題が生じている。例えば上記の文では、「7 つ」は「栗」を指し示しているが、厳密に考えると「拾った栗」が「7 つ」であり、落ちていた栗は 7 つ以上だったとも考えられる。そして、このような情報を無視すると、t : 「昨日会った学生が 3 人来た」と h : 「昨日会った学生は 3 人だ」

で、t が持つ「昨日会った学生の内、3人が来た」という情報が失われてしまい、t が h を完全に含意していると見なしてしまう。一方で、「拾った栗」が「7つ」や「来た学生」が「3人」という情報は、一階述語論理上での表現能力を超えているため、論理式を一階述語論理から拡張しなければならなくなる。そのため、数量と事象の情報をどこまで解析し、どこまで論理式上で表現するかを、深く考えなければならない。

仮に数量がどの事象を指し示すのかを解析することができ、一階述語論理を意味表現として採用すると、以下のような表現方法が考えられる。ただし、数量と動詞間の係り受け関係を表す特別な述語 (qtv : quantity-verb) を用いている。

「栗を7つ拾った」

LF :  $\exists x, e$  栗 ( $x$ )  $\wedge x \geq 7 \wedge$  拾う ( $e$ )  $\wedge \forall (e, x) \wedge qtv(e, 7)$

### 名詞句表現 (629 文)

論理式では、文節内に含まれる名詞それぞれに別々の変数を与えてしまうため、名詞句の構成情報が失われてしまう問題が発生する。そこで考えられる一つ対応策として、名詞句全体を単一の事象とみなし、変数を一つ与える方法が挙げられる。

「チョコレート工場の見学へ行く」

LF : チョコレート工場 ( $x$ ), …

一方、形態素単位に分解された名詞それぞれに変数を与え、名詞句の構成情報を特別な述語 (NN : Noun-Noun) を用いて表現する方法も考えられる。

「チョコレート工場の見学へ行く」

LF : チョコレート ( $x_1$ )  $\wedge$  工場 ( $x_2$ )  $\wedge NN(x_1, x_2)$ , …

現状の解析器の結果からは、上記の二通りの対応策が考えられるが、全ての名詞句表現をどちらか一方の規則で表現して良いかを判断するのは難しい。名詞句を名詞単位に分割し、構成情報を保持することは、後々のタスクにおいて有用であると考えられる。例えば、t:「これは爆発物です」と h:「これは爆発する」は、明らかに t が h を含意しており、名詞句「危険物」を「危険」と「物」という名詞単位で保持しておくことで、「爆発する」との対応が取りやすくなる。

しかし、「物」など特に実態を持たない名詞に対して、変数を付与して良いかという問題もある。変数  $x$  は何かしら実態のある内容語に対し付与している点から考えると、「危険物」の「物」はそれ自身で何か意味を持っている訳ではないため、「危険物」という一つの実態に対して変数を一つ与える方が自然と考えられる。

現状の解析で名詞句を表現する際には、名詞句単位に分解して構成情報持つ方が、より多くの情報を論理式上で保持できるため、実装の際は分解する方向で検討する。ただし、形態素解析の結果、機能語的な名詞によって名詞句が構成される場合 (例えば、「物」などの接尾) は、一般の名詞句とは分けて表現することで、対処したい。

## その他考慮すべき事象

1. 動詞と助動詞、助動詞と助動詞の関係
2. 文節末尾以外の助詞の処理
3. 形容動詞、事態性名詞の扱い
4. 接頭詞と名詞の関係
5. 括弧やアルファベットなどの記号の処理
6. 量化記号の特定

今回の調査対象中ではあまり観察されなかったが、1や2などの否定や時制などの助動詞が果たす機能を、拡張モダリティ解析でどれだけカバーできるかが、今後の実装での問題となる。特に、拡張モダリティにおける時制は非未来と未来を判断するのみに留まっているため、文章内で言及された事態や状態変化の時間的順序関係の同定など、より深い解析が求められる。

さらに、最も深刻な問題は各変数に与える適切な量化記号の特定である。今回の調査では、「すべて」などの語が係り受け関係にあるような、明らかに全称量子子を使用すべき場合を除き、全て存在限量子を変数に割り当てることを仮定し、調査していたが、実際には名詞の性質（総称・非総称、定・不定など）や文脈に応じて量子子を変化させる必要がある。以下に「犬」に対して、存在限量子を用いるべき場合、及び全称量子子を用いるべき場合の二つの例を示す。

- ・存在限量子が望ましい例

「私は犬を飼っています」

LF: 私( $x_1$ )  $\wedge$  犬( $x_2$ )  $\wedge$  飼う( $e$ )  $\wedge$  ガ( $e, x_1$ )  $\wedge$  ヲ( $e, x_2$ )

- ・全称量子子が望ましい例

「犬は動物です」

LF:  $\forall x$  犬( $x$ )  $\rightarrow$  動物( $x$ )

このような例からも、各変数に適切な量化記号を与えるためには、文内の情報を全て活用しても、判断できない場合があるため、何らかの外部情報を用いる必要があると考えられる。単純な方法としては、対象の語と「すべて」等の全称量子子を示す表現との共起頻度を取ることが考えられる。

## 考察

簡潔なアルゴリズムでは、全データセット 2,472 文中の 1,100 文以上で変換の際に問題が生じた。しかし、簡潔なアルゴリズムで生じた問題のうち、約 3 割はアルゴリズムを工夫することにより、現状の解析で対応が可能だった。残りの問題の内、約 2 割は現状の解析器では十分に解析できず、適切な意味表現を得られなかった。また、残りの 5 割は、文をより詳しく解析した後、その結果を

論理式へと変換すべきか、現状で解析できている情報を論理式に変換し、変換後の推論に託すべきか、どちらの処理を取るべきかが自明でない問題であった。

今回は現状の解析器から理想的な結果を得られると仮定したが、意味表現として適切な論理式を得るために最も重要な解析器は、基本的な役割関係を得ることができる述語項構造解析だった。述語項構造解析は、連体節の問題の解決に欠かせなく、コンピュータ文の表現の解決にも重要な要素となるからである。ただし、現在の述語項構造解析器では、ガ格、ヲ格、ニ格のみしか格助詞関係を判断しないため、高性能な論理式変換器のためには、より優れた述語項構造解析器が必要となる。

### 3.2.2 正規化の問題

正規化の問題では、推論を行いやすくするため、同義な表現を一つの論理式へ吸収することができるかを調べた。その方法として、RTE 評価データセットから、推論や上位・下位などの等価になり得ないカテゴリに属する文対を除いた上で、含意関係の推論判定が<sup>1</sup>の文対に対して、人手で見て正規化できそうな文対があるかを調査し、表現形式の一般化を検討した。ただし、「床屋」と「理髪店」など、内容語が別の内容語に書き変わっている文対は、全く同一の論理式へ変換するのは不適当と考えられるため、今回は調査の対象から外した。

上記の抽出の結果、データセットの全 2,472 文対中、44 文対が正規化の条件に該当した。

#### 受動態 (10/44 文対)

今回想定している述語項構造解析は、受動態のときも、無標の場合の格助詞で解析できることを想定している。そのため格助詞関係に関しては、文対から同一の解析結果を得ることができるが、結果として、変換後の格助詞と受動態を表す「れる・られる」が混在してしまう。そこで、受動態を表す「れる・られる」を伴う場合は、「れる・られる」を省略し、述語項構造解析の結果を反映させることで、正規化を試みる<sup>1</sup>。

t: 「太郎が次郎をナイフで刺した」  
h: 「次郎が太郎にナイフで刺された」  
LF: ガ (刺す, 太郎) ∧ ヲ (刺す, 次郎), …

ただし、場所や組織などが受動態に変化する際、ガ格とニ格が入れ替わるような特別な事例も存在した。以降、正規化のための規則を「正規化」と記述する。

t: 「貨幣は造幣局で作られる」  
h: 「造幣局が貨幣を作る」  
正規化: デ (作る, 造幣局) ⇔ ガ (作る, 造幣局)

このような文対の正規化には、動作や状態と事象の単純な格助詞関係のみでなく、格助詞関係以上の深い意味役割を捉える解析を行う必要があると考えられる。

<sup>1</sup>以降では、論文上で論理式を読みやすさのため、論理式表現に変数として直接単語を使用している

## 強調構文 (11/44 文対)

強調構文とは、元の文のある要素 A が、文末へ移動し「…のは、A だ」などの形で A の部分が強調される構文である。しかし、強調される語 A は、文によってその役割が変わるので、正規化のためには A の持つ役割を正確に解析する必要がある。

今回の調査では、格助詞関係が強調された例として、ガ格、ヲ格、デ格の役割を持つ文節が強調されている文対が見られた。このような格助詞関係が強調されている場合は、強調構文「のは… A だ」における A の格助詞関係と「のは」を省略する規則を設けることによって、正規化できる。以下に格助詞関係を持つ文節が強調された文対の例を示す。

### ・ガ格の強調

t:「化粧が女を化かす」

h:「女を化かすのが化粧だ」

### ・ヲ格の強調

t:「彼は必死に窓を叩いた」

h:「彼が必死に叩いたのは窓だった」

### ・デ格の強調

t:「更衣室でお金がなくなった」

h:「お金がなくなったのは、更衣室でだった」

強調される要素は格助詞を持つ文節だけでなく、以下のように、より複雑な強調構文も存在した。

t:「国民は教育を受ける義務がある」

h:「教育を受けるのが国民の義務だ」

正規化: ガ(受ける, 国民)  $\wedge$  atc(義務, 受ける)  $\Leftrightarrow$  atc(の, 受ける)  $\wedge$  ガ(の, 義務)  $\wedge$  の(義務, 国民)

この文対では、「受ける」のガ格である「国民」と、「受ける」が連体節の外の関係として修飾している「義務」の二つが、「国民の義務」と連結した上で、強調されている。これは、国民と義務の間に所持の関係があるため、このような構文的な変換が発生していると考えられる。現状の構文・意味解析では、所持の関係などのより深い意味役割を考慮した関係解析を十分に行えないため、正規化することは難しいと言える。

## 動詞の位置交換 (5/44 文対)

受動態や強調構文などによる動詞の構文的な位置の移動が生じる際、動詞に付随する助詞、助動詞、非自立語などの機能語が変化する。このような動詞の位置交換に伴い機能語の変化が生じる場合は、受動態、及び強調構文の正規化規則を用いた上で、さらなる規則を適応することによ

り、正規化できる.

t:「金太郎は斧を担いで熊に乗っている」

h:「斧は熊に乗った金太郎に担がれている」

正規化: tで「で(乗る,担ぐ)」削除することにより、統一の論理式になる

t:「情報が流出して困るのは、我々ではなく彼らの方だ」

h:「情報が流出すると、我々ではなく、彼らの方が困る」

正規化: て(困る,流出する) ⇔ と(困る,流出する)

上記の例では、「で」や「と」などの順接の接続助詞を省略することで、正規化できそうだが、事例数が少ないため、現状では判断できない。そのため、上記の規則を正規化規則として一般的に用いるためには、逆助詞や助動詞を省略すると問題になる例が含まれるデータでの調査や、真偽判断や時制などの拡張モダリティ解析が助詞や助動詞の意味を捉えることができるかを検証する必要がある。

#### 主語の変換 (6/44 文対)

主語が文中の他の要素と入れ替わる場合にも、意味的に同義な文対が見受けられた。

t:「太郎は次郎のいとこだ」

h:「次郎は太郎のいとこだ」

正規化: ガ(いとこ,太郎) ∧ ノ(いとこ,次郎) ⇔ ガ(いとこ,次郎) ∧ ノ(いとこ,太郎)

この文対では、係り受けと述語項構造の解析結果により、正規化する規則が得られそうに感じられる。しかし、上記の例は、「太郎」と「次郎」が同じ人名のカテゴリに属し、「いとこ」が二項関係となる語のため、係り受けと述語項構造による規則で正規化することができるが、同様の性質を持たない文の場合は元の文の意味から変化してしまう。例えば、「オリオン座は冬の星座だ」という同様の構造を持つ文に対し、上記の規則を適応すると、「冬はオリオン座の星座だ」といった日本語として不自然な文と同一視してしまう。このように、主語の変換の際には、単に文の構造や構成情報のみではなく、単語がどのカテゴリに属しているのか、特定の格を取る項の入れ替えを認める単語かなどの外部知識を加える必要がある。

#### 固有の変換 (12/44 文対)

特定のカテゴリにまとめられなかったが、意味的に同義で、統一の表記へと変換できそうな例が見受けられた。以下では固有の正規化規則が必要な文対とその規則の一例を示すが、これらが一般に適用できるかは、より多くの文対で検証が必要となる。

・機能語と機能語の入れ替え (8 文対)

t:「ビューラーはまつげをカールさせるための道具だ」

h:「まつげをカールさせる道具をビューラーという」

正規化: ための(道具, カールする) ⇔ *atc*(道具, カールする)

ガ(道具, ビューラー) ⇔ ヲ(いう, 道具) ∧ ト(いう, ビューラー)

t:「リャオヤンは中国遼寧省の都市である」

h:「中国遼寧省にリャオヤンという都市がある」

正規化: ガ(都市, リャオヤン) ⇔ という(都市, リャオヤン)

ノ(都市, 中国遼寧省) ⇔ ガ(ある, 都市) ∧ ニ(ある, 中国遼寧省)

t:「夏野菜は体を冷やす効果がある」

h:「夏野菜の効果は体を冷やすことだ」

正規化: ガ(ある, 効果) ∧ ハ(ある, 夏野菜) ⇔ の(効果, 夏野菜)

*atc*(効果, 冷やす) ⇔ *atc*(こと, 冷やす) ∧ ガ(こと, 効果)

・機能語と内容語の入れ替え(4文対)

t:「カツオは海の生き物である」

h:「海の生き物としては、カツオが挙げられる」

正規化: ガ(生き物, カツオ) ⇔ として(挙げる, 生き物) ∧ ヲ(挙げる, カツオ)

t:「ミャンマーはかつてビルマと呼ばれていた」

h:「ミャンマーはかつてビルマだった」

正規化: ヲ(呼ぶ, ミャンマー) ∧ ト(呼ぶ, ビルマ) ⇔ ガ(ビルマ, ミャンマー)

t:「彼は財産を守るために遺言書を書くことにした」

h:「彼は財産を守るために遺言書を書く予定だ」

正規化: *atc*(こと, 書く) ∧ ニ(する, こと) ⇔ *atc*(予定, 書く)

## 考察

含意関係認識のデータでは、今回の調査対象となる文対は少なかった。調査した文対のうちの半数は、正確な述語項構造解析と簡単な規則により、正規化することができた。一方、残りの文対は、複雑すぎる構文変化、外部の知識が必要となる、一般的に適応できるか分からないなどの原因により、正規化のための規則を発見することはできなかつたため、さらに多くの文対での調査が必要となる。

## 第4章 日本語文の論理式変換器の実装

### 4.1 使用した構文・意味解析器

論理式変換に適応する構文・意味解析として、形態素解析、係り受け解析、述語項構造解析、拡張モダリティ解析を行ったが、それらの解析のために使用したソフトウェア、参考にした研究を以下で紹介する。

今回は、MeCab[8] を形態素解析器として用いた。MeCab は Conditional Random Fields (CRF) を利用して入力された文を形態素ごとに分解、及び各形態素へ品詞タグ付けを行っている。

また、統語係り受け解析として、CaboCha[7] を使用した。CaboCha は毎日新聞記事に係り受け関係を人手で付与したデータを教師データとし、Support Vector Machine を用いて、入力された文の形態素を文節にまとめ上げた後、係り受け関係を付与している。

一方、述語項構造解析器として、渡邊らの研究で用いられた解析器を使用している [20]。この述語項構造解析器は、入力として CaboCha の結果を用いており、NAIST テキストコーパスに人手で正解データを付与したものを教師データとして学習することにより、文中の動作や状態、及びそれに関係する項を認識する。

最後に、拡張モダリティ解析器として、江口らの解析器を使用している [12]。モダリティタグを付与したコーパスと、拡張モダリティ用に作成された手がかり表現辞書から、入力として与えられた述語項構造解析で述語やイベントと判断された事象に対し、拡張モダリティタグを付与する。

以下図 4.1 に、MeCab、CaboCha、渡邊らの述語項構造解析、江口らの拡張モダリティ解析の結果を xml 形式へ統合、変換した出力を示す。

入力：「庭の花が咲く」

```
<Sentence id="null" role="none" text="庭の花が咲く">
  <Chunk func="0/1" head="c1" id="c0" score="1.522522" type="D">
    <Token id="t0" orig="庭" pos="名詞" pos1="名詞-一般" pos2="" pos3="" pron="ニワ" surf="庭"/>
    <Token id="t1" orig="の" pos="助詞" pos1="助詞-連体化" pos2="" pos3="" pron="/" surf="の"/>
  </Chunk>
  <Chunk func="0/1" head="c2" id="c1" score="0.000000" type="D">
    <Token id="t2" orig="花" pos="名詞" pos1="名詞-一般" pos2="" pos3="" pron="ハナ" surf="花"/>
    <Token id="t3" orig="が" pos="助詞" pos1="助詞-格助詞-一般" pos2="" pos3="" pron="ガ" surf="が"/>
  </Chunk>
  <Chunk func="0/0" head="c-1" id="c2" score="0.000000" type="D">
    <Token id="t4" orig="咲く" pos="動詞" pos1="動詞-自立" pos2="五段・力行イ音便" pos3="基本形" pron="サク" surf="咲く">
      <Modality assumptional="0" authenticity="成立" focus="0" sentiment="0" source="wr" tense="非未来" type="叙述"/>
      <Predicate ga="t2" type="pred"/>
    </Token>
  </Chunk>
</Sentence>
```

図 4.1: 解析器出力例

## 4.2 実装した変換プロセス

日本語文の論理式変換器の変換プロセスには、4.1 に示したとおり、形態素解析に MeCab、統語係り受け解析に CaboCha、述語項構造解析に渡邊らの解析器、拡張モダリティ解析に江口らの解析器を用い、基本的に 3.1.1 と同じアルゴリズムを使用した。ただし、変数付与と係り受け関係の表現、及び同一文節内の表現において、調査結果を反映した規則を導入した。以下に、導入した規則を示す。

### 1. 変数の割り振り方

基本的に 3.1.1 と同じ処理を行うが、副詞への変数付与は、係り受け解析の結果を参照し、係り受け先の動詞や形容詞と同じ変数を割り振る。

### 2. 係り受け関係

係り受け元の文節末に助詞が存在しない場合、以下の表 4.1 のように、係り受け元、係り受け先の品詞などの情報から、特別な述語を決定し、係り受け関係を表現する。

表 4.1: 係り受け元に助詞がない場合に使用する述語

使用する述語	係り受け元の情報	係り受け先の情報
atc	動詞	名詞
att	連体詞	名詞
qtv	数量	動詞
rel	上記以外	上記以外

### 3. 名詞句

名詞句は、構成する名詞それぞれに対して変数を付与し、名詞句の構成情報を特別な述語 NN を用いることで表現する。ただし、数字が連続する場合は数字全体に単一の変数を与え、また接尾などの機能的な名詞が含まれる場合、述語 NN の一方を小文字 n にすることで表す (例えば、「危険物」ならば、危険 ( $x_1$ )  $\wedge$  物 ( $x_2$ )  $\wedge$   $Nn(x_1, x_2)$ )。

### 4. 名詞 + 「する」

名詞が「する」を伴い動詞化する場合は、名詞に変数  $x$ 、「する」に変数  $e$  を付与し、名詞+「する」の構成情報を特別な述語 ( $Ns$ : Noun-suru) を用いることで表現する (例えば、「運動する」ならば、運動 ( $x$ )  $\wedge$  する ( $e$ )  $\wedge$   $Ns(x, e)$ )。

以下の図 4.2 に実装した変換器により出力される論理式の例を示す。

入力: 「庭の花が再び咲く」

$x_0, x_1, e_0, \text{庭}(x_0), \text{花}(x_1), \text{再び}(e_0), \text{咲く}(e_0), \text{ga}(e_0, x_1), \text{の}(x_1, x_0), \text{fact}(e_0, \text{成立}), \text{tense}(e_0, \text{非未来})$

図 4.2: 論理式変換器出力例

## 第5章 まとめ

本稿では、文の意味表現を論理式で表現するときに生じる問題を、以下の二点に焦点を当てて調査した。

第一に、形態素、統語係り受け、述語項構造、拡張モダリティが正しく解析できると仮定し、その結果から簡潔なアルゴリズムで論理式へ変換できるかを調査した。調査の結果、現状の解析器と簡潔なアルゴリズムでは、今回のデータセット中で生じた問題の内、約3割は変換アルゴリズムへ簡単な規則を追加することにより、対処することができたが、残りの約2割は現状の意味解析器に加えて、新たな解析器が必要であることがわかった。さらに残りの約5割は、より正確な意味を表現する論理式を得るために更なる意味解析を行うか、現状の解析結果から得られる情報を直接的に論理式で表現しておき、変換後の推論に更なる意味解析を託すべきかが、自明でない問題であった。

第二に、同義な表現を吸収しながら論理式に変換するために、今後必要となる処理について整理した。統一の表現へと吸収できそうな文対中、半数近くの文対は、正確な述語項構造解析により、正規化できることがわかった。しかし、残りの半数の文対では、複雑な構造・意味役割の解析や外部知識の使用が必要となるなどの理由により、統一の表現へと正規化するための一般的な方法を発見することはできなかった。

また、調査の結果にもとづいて作成した論理式変換アルゴリズムを用いて、日本語文の論理式変換器を作成し、公開へ向けて準備を進めている。

今後の課題として、他のデータセットでの調査が挙げられる。今回の調査対象としたデータセットは、比較的簡単な文のみで構成されていたため、より複雑な分で変換可能かを調査する予定である。また、正規化の問題の調査対象となる文対が少量しか見つからなかったため、さらに多くの文対でも調査を行うことを検討している。また、本稿で実装した論理式変換器に関して、論理式変換性能の評価や実際の言語処理タスクへの応用を行うことも検討している。

## 謝 辞

本研究を進めるにあたり、ご指導を頂いた乾健太郎教授、岡崎直観准教授に感謝いたします。  
調査の実施、及び本論文の作成にあたり、終始適切な助言を賜った松林優一郎研究特任助教、井之上直也氏に感謝いたします。

日常の議論を通じて多くの知識や示唆を頂いた乾・岡崎研究室の皆様にも感謝いたします。

## 参考文献

- [1] Johan Bos. Wide-Coverage Semantic Analysis with Boxer. In Johan Bos and Rodolfo Delmonte, editors, *Semantics in Text Processing. STEP 2008 Conference Proceedings*, Research in Computational Semantics, pp. 277–286. College Publications, 2008.
- [2] Johan Bos, Stephen Clark, Mark Steedman, James R. Curran, and Julia Hockenmaier. Wide-Coverage Semantic Representations from a CCG Parser. In *COLING '04*, 2004.
- [3] Johan Bos and Katja Markert. Recognising Textual Entailment with Logical Inference. In *HLT/EMNLP*, Vol. 2012-NL-206, pp. 628–635, 2005.
- [4] Alastair Butler. *The Semantics of Grammatical Dependencies*. Emerald, 2010.
- [5] Naoya Inoue, Ryu Iida, Kentaro Inui, and Yuji Matsumoto. Resolving direct and indirect anaphora for japanese definite noun phrases. *Journal of Natural Language Processing*, Vol. 17, No. 1, pp. 221–246, January 2010.
- [6] Naoya Inoue, Ekaterina Ovchinnikova, Kentaro Inui, and Jerry R. Hobbs. Coreference resolution with ilp-based weighted abduction. In *Proceedings of the 24th International Conference on Computational Linguistics*, pp. 1291–1308, 2012.
- [7] Taku Kudo and Yuji Matsumoto. Japanese dependency analysis using cascaded chunking. In *COLING-02 proceedings of the 6th conference on Natural language learning*, Vol. 20, pp. 1–7, 2002.
- [8] Taku Kudo, Kaoru Yamamoto, and Yuji Matsumoto. Applying conditional random fields to japanese morphological analysis. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP-2004)*, pp. 230–237, 2004.
- [9] Dan Moldovan, Christine Clark, Sanda Harabagiu, and Steve Maiorano. COGEX: A Logic Prover for Question Answering. In *HLT/NAACL*, pp. 87–93, 2003.
- [10] Rajat Raina, Andrew Y. Ng, and Christopher D. Manning. Robust textual inference via learning and abductive reasoning. In *American Association for Artificial Intelligence*.
- [11] Parsons Terence. *Events in the Semantics of English*. MIT Press, 1990.
- [12] 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.
- [13] 杉浦純, 井之上直也, 乾健太郎. 談話関係認識への連想情報の応用. NLP 若手の会第7回シンポジウム, 2012.
- [14] 松吉俊, 佐尾ちとせ, 乾健太郎, 松本裕治. 拡張モダリティタグ付与コーパスの設計と構築. 言語処理学会第17回年次大会発表論文集, pp. 147–150, 2011.

- [15] 戸次大介. CCG による日本語文法記述の進捗と展望～活用体系・統語構造・意味合成～. 言語処理学会第 16 回年次大会発表論文集, pp. 1078–1081, 2010.
- [16] 戸次大介. 日本語文法の形式理論-活用体系・統語構造・意味合成-. くろしお出版, 2010.
- [17] 小谷通隆, 柴田和秀, 中田貴之, 黒橋禎夫. 日本語 Textual Entailment のデータ構築と自動獲得した類義表現に基づく推論関係の認識. 言語処理学会第 14 回年次大会発表論文集, pp. 1140–1143, 2008.
- [18] 小谷通隆, 柴田和秀, 黒橋禎夫. 言い換え表現の述語項構造への正規化とテキスト含意関係認識での利用. 言語処理学会第 15 回年次大会発表論文集, pp. 260–263, 2009.
- [19] 吉本哲, 周振, 小菅智也, 大友瑠璃子, Alastair Butler. 日本語ツリーバンクのアノテーション方針. 言語処理学会第 19 回年次大会発表論文集, pp. 928–931, 2013.
- [20] 渡邊陽太郎, 浅原正幸, 松本裕治. 述語語義と意味役割の結合学習のための構造予測モデル. 人工知能学会論文誌, 第 25 巻, pp. 252–261.
- [21] 松木 久幸 佐藤理史, 駒谷和範. 文末機能表現シソーラスの網羅性の検証. 情報処理学会全国大会 講演論文集, 第 75 巻, pp. 1Q–2, 2013.