

効率的な推論処理のための日本語文の論理式変換に向けて

稲田 和明 松林 優一郎 井之上 直也 乾 健太郎

東北大学

{kazuaki.inada, y-matsu, naoya-i, inui}@ecei.tohoku.ac.jp

1 はじめに

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する。

近年、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、より高度な処理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、文の表層上の繋がりだけでなく、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。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 [3, 4, 8]。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未だ存在せず、また、文法の異なり度合いから、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。このような背景から、我々は、日本語の文を、推論を取り扱いやす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う。

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. 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、同義な表現の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、文に人手で正解を与え、我々が提案する手法で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、含意関係認識の評価データにおける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、汎化の問題と呼ぶ)。

2 関連研究

英語圏においては、質問応答や含意関係認識を解く過程において、統語解析の結果を基に文を論理式に変換し、推論を行う手法が提案され、その有用性が示されている [5, 6]。しかし、統語解析の結果のみでは、文内の情報全てを読み取れないことは明らかである。また、より高度な変換の枠組みとして、Combinatory Categorical Grammar (CCG) による統計的統語解析器を用いて、Discourse Relation Structure 形式へと変換する Boxer と呼ばれるシステムが公開されており [1, 2]、推論を用いる意味解析の応用タスクにおいて広く用いられている。¹

日本語においては、統語構造の記述に CCG、意味構造の記述に高階動的述語論理を用いた文法理論の構築が進んでおり [10, 11]、この理論を用いて、実用的に日本語文を論理式に変換出来る可能性があるが、現段階では、この文法理論は実用できる形で実装されていない。また、日本語の表現を汎化する試みの一つとして、構文・格・省略解析の結果を、統一の表記へと変換する研究がある [13]。この研究では、文の解析結果と外部の知識を組み合わせることで、1つの表現へと汎化しているが、明らかに意味が異なる「小さくない」と「大きい」を汎化してしまっているなど、荒さが見受けられる。我々は、物事の事実性や相対的な時間関係など、統語的解析では捉えられない情報を扱う、拡張モダリティ [7, 9] の解析も取り入れた論理式への変換を試みる。

3 論理式への変換

我々は論理式の表現形として、表現がシンプルで、効果的な推論演算が期待できる一階述語論理を用いる。一階述語論理よりも詳細な表現が可能な、二階述語論理、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、現状で推論を効率的に行うことができないため、本稿では一階述語論理を採用する。また、一階述語論理の形式として特に neo-Davidsonian を用いる。neo-Davidsonian は、予め述語が持つ引数の数を定義しない表現形式であるため、動詞が持つ格の数や修飾語の数が一定ではないなど、自然言語の特徴を表現するには適当な形式であると言える。

我々は、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。

¹<http://svn.ask.it.usyd.edu.au/trac/candc>

1. 文を、品詞タグ付け、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. 品詞が助詞以外の品詞を持つ形態素に変数を与える。変数名は単語原型とし、述語項構造解析において、格助詞関係を取る述語には変数 e を、それ以外には変数 x を用いる。
3. 述語項構造解析の結果から、格関係を表す論理式を加える。
4. 拡張モダリティ解析の結果から、その解析結果を表す論理式を加える。
5. 文節と係り受け関係解析、及び助詞から、係り関係の論理式を加える。

以下に neo-Davidsonian に従い、上述の流れで変換した 論理式 (LF) の例を示す。

- (a) 私はケーキを食べる。
LF: 私 (x_1), ケーキ (x_2), 食べる (e), ガ (e, x_1), ヲ (e, x_2)
- (b) 羽田の空港に行った。
LF: 羽田 (x_1), 空港 (x_2), 行く (e), ノ (x_2, x_1), ニ (e, x_1), 相対時 (e , 非未来), 真偽 ($e, +CT$)

以降では、この方法で対応できない文や、表現として問題になりそうな所が無いかを調査するため、実際の日本語文を分析する。また、同義な文対を同じ論理式へ変換する方法を検討し、その際の問題点を述べる。

4 分析対象データ

我々が最終的に出力する論理式は、実際の含意関係認識等の応用タスクで利用することを想定することから、調査対象のデータには、京都大学 RTE 評価データを使用した [12]。

単文の問題では、データセットに含まれる文全体を観察しながら、Boxer における論理式の表現方法を元に、日本語文の変換の際に問題となる部分を洗い出した。文対の問題では、まず、データセットから意味がほぼ等価な文対を次の方法で取り出し、表現形式の一般化を検討した。(1) 推論や上位・下位などカテゴリ的に等価になり得ないもの、推論判定が最高評価でないものを除外する。(2) 残りを人手で見、意味的に等価な文対を抽出する。ただし、意味的に等価な文とみなせる対のうち、ある内容語が別の内容語に書き変わっているものは、全く同一の論理式へ変換するのは不相当と考えられるため、今回は調査の対象から外した。

以下にデータセットから抽出した文対の例を示す。

- (c) 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。
- (d) t: 貨幣は造幣局で作られる。
h: 造幣局が貨幣を作る。

4.1 変換の問題

4.1.1 副詞の修飾

- (1) 彼は すぐ怒る。 LF: すぐ (e), 怒る (e)

下線部は係り受け関係を持つが、論理式の述語となる助詞が無いので、この関係を表現できない。そこで副詞「すぐ」に、その係り先である動詞「怒る」と同様の変数を割り当てることで、修飾関係を表す。ただし、このような処理を行うためには、3節で定義したパイプライン的な処理ではなく、全ての解析結果を総合的に見て、変数を割り当てる必要がある。

4.1.2 連体詞の修飾

- (2) 大きな川のそばだ。
LF: 大きな (x_1), 川 (x_2), mod(x_2, x_1)

連体詞は、助詞を必要とせずに修飾するが、形容詞と異なり格助詞を取らないので、その修飾関係を示す方法が問題となる。そこで今回は連体詞の関係を、特別な述語 mod を用いて表現しておくが、将来的には形容詞と同等の意味とみなし、一つの表現へと吸収したい問題である。

- (3) この本はベストセラーだ。

さらに、「この」などの連体詞は、直接照応するか間接照応するかで、修飾時の振る舞いに変化する。照応詞が直接照応していれば、照応詞とその修飾語は同一のものを指すが、間接照応の時は、照応詞が修飾先と同一のものを指していないため、直接照応と間接照応で同一の表現を用いることはふさわしくない。よって、連体詞の修飾では、直接照応と間接照応で場合分けした上で、論理式へと変換しなければならないと言える。

4.1.3 連体節の修飾

- (4) 国民は教育を受ける義務がある。
LF: 受ける (e), 義務 (x), con(x, e)

連体節では、連体節を構成している最後の語に、その直前の語が修飾する形を取るが、助詞を伴わないため、この関係を表現することができない。そこで連体節の関係を、特別な述語 con により表現しておく。

- (5) タバコを吸っている人が肺癌になる率は高い
LF: 吸う (e), 人 (x), con(x, e), ガ (e, x)

さらに、連体節を成す語の間で格関係が生じる場合がある。これは連体節における外の関係、内の関係と呼ばれる問題であり、述語項構造解析で得られるこれらの結果も考慮した上で、論理式へと変換することが望ましい。

4.1.4 コピュラ文の表現

- (6) オリオン座は冬の星座である。
LF: オリオン座 (x_1), 星座 (x_2), eq(x_2, x_1)

この文では、「オリオン座」が「星座」のガ格を取るが、名詞同士の格構造をそのまま論理式として用いても、文の意味を表す適切な表現とは言いがたい。そこで特別な述語 eq を用意し、論理式への変換を行う。

4.1.5 並列表現

- (7) 太郎と花子は勉強した。
- (8) 太郎と花子は仲直りした。

体言が並列になっている文では、述語に変数を与える際に問題が生じる。(7)では、太郎と花子が勉強という1つの事象が一緒に行われたのではなく、それぞれ別々の事象として表現した方が自然である。一方(8)では、太郎と花子が「仲直り」という同一の事象が行われたと考えるのが普通である。すなわち、正しい論理式を得るためには、並列表現での詳細なイベント情報を得るための解析を行う必要があると言える。

4.1.6 数量表現

- (9) みかんを 7つ 食べた。

「7つ 食べる」の係り受け関係から副詞と同様に扱えそうだが、実際に「7つ」が示す対象はみかんである。従って、正しい論理式を表現するためには、数量が指し示す事象が何であるかの解析しなければならないことが分かる。

4.1.7 名詞句表現

- (10) チョコレート工場の見学へ行った。
LF: チョコレート (x1), 工場 (x2), nm(x1, x2)。

論理式上では、名詞句の意味を判断するとき、名詞句を名詞単位に分解してあることが望ましいが、元の名詞句の構成情報も保持しておかなければならない。そこで、特別な述語 nm を用意し、論理式上で名詞句を表現する。

4.1.8 その他の表現

助動詞や非自立語など、上記以外の関係性を表現する方法として、特別な述語 rel を用いる。助動詞や非自立語は、述語項構造解析や拡張モダリティ解析の結果に表現されている部分もあるため、省略しても問題ないかもしれないが、断定はできないので、汎化の問題では省略せずに論理式変換を考える。

4.2 汎化の問題

4.2.1 受動態

受動態では、正確な述語項構造解析により、汎化することができる。

- 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。

格関係が正しく解析できたとすると、助詞「れる」の処理が問題となる。しかし、述語項構造解析において、受動態を考慮した格を出力すると定めることで、助詞「れる」を省略することができる。

- t: 貨幣は造幣局で作られる。
- h: 造幣局が貨幣を作る。

しかし上記のように、組織や場所、状況などは、ガ格とデ格を交換しても同じ意味をさす場合があることを考慮しなければならない。

4.2.2 強調構文

強調構文とは「…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Aは、文によってその役割が変わるので、述語項構造などの正確な解析が必要となる。以下にいくつかの例を示す。

- ・ガ格の強調
 - t: 化粧は女を化かす。
 - 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。

さらに、上記のような格以外の関係も強調されることがある。

- 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。

上記の例では、受けるのガ格と連体節「受ける義務」の「義務」が強調されている。このように述語項構造解析だけでは捉えられない関係が存在するため、強調構文を正しい論理式へ変換し、汎化するの難しいと考えられる。

4.2.3 動詞の位置交換

受動態や強調構文などによる構文的な位置の変化に加えて、助詞や助動詞や非自立が変化する文対が見られる。その例と汎化のための規則を示す。

- t: 金太郎はまさかりを担いで熊に乗っている。
 - h: まさかりは熊に乗った金太郎に担がれている。
- テ(いる、乗る) デ(乗る、担ぐ) ⇔ rel(乗る、た) テ(いる、担ぐ)

- t: 彼は功績が認められて、表彰された。
- h: 彼が表彰されたのは、功績が認められたからだ。

rel(表彰する、た) て(表彰する、認める) ⇔ から(だ、認める)
rel(認める、た) は(認める、の) con(表彰する、の)

上記の例では、動詞の接続助詞と非自立語の動詞が、汎化の際の問題となっているため、省略などの規則を設ける必要がある。

4.2.4 主語の変換

- t: 花子は陽子の親友だ。

h: 陽子は花子の親友だ。

eq(花子, 親友) の (親友, 陽子) \Leftrightarrow eq(陽子, 親友) の (親友, 花子)

上記のように主語が入れ替わる場合にも、汎化が可能な文対がある。しかし、(6) の文では上記の規則による汎化を適用すると、元の文の意味を保持しなくなることから、格や助詞を見るだけでは不十分だとわかる。即ち、この問題では、文の構造のみでなく、単語自身の持つ意味も必要と言える。

4.2.5 固有の変換

上述のカテゴリに分類されない汎化規則も存在した。以下で可能な限り例を示す。

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。

ための (道具, カールする) \Leftrightarrow con(カールする, 道具)
eq(ビューラー, 道具) \Leftrightarrow 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。

eq(リャオヤン, 都市) \Leftrightarrow という (都市, リ)
ノ (都市, 中国遼寧省) \Leftrightarrow 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。

ガ (ある, 効果) ハ (ある, 夏野菜) \Leftrightarrow ノ (効果, 夏野菜)
con(冷やす, 効果) \Leftrightarrow con(冷やす, こと) eq(こと, 効果)

さらに、以下の例に見られる機能語と内容語の書き換えにも、特別の規則が必要と考えられる。)

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。
eq(生き物, カツオ) \Leftrightarrow として (挙げる, 生き物), ヲ (挙げる, カツオ)

t: 風で帽子がとんだ。
h: 帽子が飛んだのは、風のせいだ。
で (とぶ, 風) \Leftrightarrow eq(の, せい) の (せい, 風) con(とぶ, の)

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。
eq(こと, 書く) ニ (する, こと) \Leftrightarrow eq(予定, 書く)

このように、汎化には様々な書き換え規則が必要だが、これらが一般的な規則として適用できるかは判断できない。他のデータセットを見て、吟味しなければならない。

5 おわりに

本稿では、文の意味表現を論理式で表現する時に生じる問題を、以下の2点に焦点を当てて調査した。第一に、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吟味し、不足している部分を洗い出した。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。

今後の課題として、今回の結果を受けてのアルゴリズムの精密化に加えて、他のデータセットでの調査が挙げられる。今回の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、より複雑な文で変換可能かを調査しなければならない。また、汎化の調査対象となる文対が少量しか見つからなかったの、さらに多くの文対でも調査する必要がある。

謝辞 本研究は、文部科学省科研費課題 22・9719 および課題 23240018 の一環として行われた。

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] 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.
- [5] Dan Moldovan, Christine Clark, Sanda Harabagiu, and Steve Maiorano. COGEX: A Logic Prover for Question Answering. In *HLT/NAACL*, pp. 87–93, 2003.
- [6] Rajat Raina, Andrew Y. Ng, and Christopher D. Manning. Robust textual inference via learning and abductive reasoning. In *American Association for Artificial Intelligence*.
- [7] 江口晴, 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] 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.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] 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.