

# 法令文の論理式への変換 -原子文について-

北田 安希雄 江尻 暁 島津 明

北陸先端科学技術大学院大学 情報科学研究科

## 1 はじめに

我々の社会の構造や機能の基本的部分は各種の法律や法規によって明示的に記述されている。したがって法規や法律は、社会の構造や機能を使う情報システムを規定する一種の仕様と見ることができる。したがって、これを形式的に表現することができるなら法推論等により、情報システムを検証することができる。このためには自然言語の法令文は計算機が推論することのできる論理表現で表される必要がある [1]。本研究は、自然言語で書かれた法令文書を入力として、その法令文書に書かれた内容を述語論理式に変換することを目的とする。

本研究では、法令文を論理表現に変換する方法として、法令文全体の論理構造への変換、要素の原子文への変換という段階的な方式を考え、本報告では原子文への変換について報告する。法令文全体の論理構造を捉えるシステムは別稿 [2] で述べる。このシステムと本研究で開発するシステムを併用することにより、法令文を論理式に変換する (図 1)。

原子文は、述語動詞と名詞との意味的な関係、すなわち深層格を表現するものである。それぞれの述語動詞に対する格解析が必要となる。また、法令文に出現する表現、例えば、「改善するよう努める」や「有罪であると認める」といったように、「(述語動詞) + よう + (述語動詞)」や「(述語動詞) と (述語動詞)」といった構造においては述語動詞が述語動詞の対象や目的となることがあり、論理式には、これらことを表現する必要がある。そこで、本研究で行う格解析では、これらの構造の解析も行う。さらに、「区民に対する警察署の協力」といった句では、「協力」の動作主格は「区」、対象格は「区民」となっている。このようなサ変名詞がとる深層格も、論理式で表現する必要がある。

本報告では、実際の法令文を基にして構築した格フレーム辞書 (3.1)、構築した辞書と JUMAN [3]、KNP [4] の解析結果を用いて格解析を行うシステム (3.2)、格解析の結果を用いて原子文生成をするシステム (4)、システムの評価実験 (5) について述べる。

## 2 関連研究

### 2.1 格解析

自然言語文から原子文を生成するには、述語の決定、述語に対する項を決定する格構造解析が必須となる。そのため、格フレーム辞書が必要となる。河原 [5] らは新聞記事のコーパスから、表層格の格フレーム辞書の自動構築を行っている。本研究では、原子文を生成すること、および法令文特有の性質を考慮し深層格の情報まで載せた格フレーム辞書を構築した。

### 2.2 法令文の論理表現

吉野 [6] は法令文書を表すための論理表現について議論している。田中 [7] は法律文の構造を分析している。長野ら [8] は、法令文が前提条件となる要件部とその要件に対して帰結となる効果部とから成るという性質を考慮することにより、法令文の構文構造を解析するシステムを開発している。

## 3 法令文の格解析

### 3.1 格フレーム辞書

本研究では、千代田区生活環境条例 全 28 条と富山県条例第 54 号「情報通信技術の利用に関する条例」全 10 条に出現する 129 種類、計 431 個の述語動詞から、

- 1.
- (a) 土地、建物又は工作物を所有する者は、
- (b) それらの清潔を保ち、
- (c) 良好な生活環境を保全するよう努めなければならない。



- 2.
- (a) 所有 (e1)  $\wedge$  人 (x1)  $\wedge$  agt(e1, x1)  $\wedge$  ( 土地 (x2)  $\vee$  建物 (x2)  $\vee$  工作物 (x2) )  $\wedge$  obj(e1, x2)
- (b) 保つ (e2)  $\wedge$  agt(e2, x1)  $\wedge$  obj(e2, e3)  $\wedge$  清潔だ (e3)  $\wedge$  obj(e3, x2)
- (c) 努める (e4)  $\wedge$  agt(e4, x1)  $\wedge$  obj(e4, e5)  $\wedge$  保全 (e5)  $\wedge$  agt(e5, x1)  $\wedge$  obj(e5, x4)  $\wedge$  生活環境 (x4)  $\wedge$  良好 (e6)  $\wedge$  obj(e6, x4)



- 3.
- $\forall x1, x2, e1$  所有 (e1)  $\wedge$  人 (x1)  $\wedge$  agt(e1, x1)  $\wedge$  ( 土地 (x2)  $\vee$  建物 (x2)  $\vee$  工作物 (x2) )  $\wedge$  obj(e1, x2)
- 
- $\exists e2, e3, e4, e5, e6, x3$  保つ (e2)  $\wedge$  agt(e2, x1)  $\wedge$  obj(e2, e3)  $\wedge$  清潔 (e3)  $\wedge$  obj(e3, x2)  $\wedge$  O ( 努める (e4)  $\wedge$  agt(e4, x1)  $\wedge$  obj(e4, e5)  $\wedge$  保全 (e5)  $\wedge$  agt(e5, x1)  $\wedge$  obj(e5, x3)  $\wedge$  生活環境 (x3)  $\wedge$  良好 (e6)  $\wedge$  obj(e6, x3) )

1 から 2 へは、意味的に切り分けられた部分からそれぞれ原子文を生成する。

2 から 3 へは、様相演算子、限量子、含意等を用いて生成された原子文を統合している。

なお、agt は動作主格、obj は対象格を表す。O は義務の様相演算子である。

図 1: 法令文から論理式への変換

表 1: 格フレーム辞書の例 「図る」

深層格	表層格	名詞	頻度
agt	ガ	区	2
		県	2
		者	1
obj	ヲ	連携	2
		合理化	1
		向上	1
		改善	2
		推進	1

格フレーム辞書を構築した。辞書に載せた内容は、これらの述語動詞がどのような名詞を深層格としてとっているか、その名詞が深層格として取られた頻度、深層格の名前、そして付随する表層格である。「図る」の辞書の例を表 1 に示す。この辞書の構築は、深層格を判断する必要があるため人手で行った。この結果、129 種類の述語動詞の情報を持つ格フレーム辞書ができた。

### 3.2 格解析の手法

本研究における格解析は次のような手順である。

1. 法令文を JUMAN [3]、KNP [4] により、形態素解析、構文解析を行う。
2. 格フレーム辞書を参照して、格解析の対象とする述語動詞がどのような深層格をとりうるのかチェックする。
3. とりうる深層格としてのスコア付けを、述語動詞に直接係っている各文節および、被連体修飾名詞に対して行う。
4. スコアが閾値を超え、最も高い文節を深層格として決定する。閾値は 10 点とした。
5. 4. でスコアが閾値を超える文節がないならば、3. と 4. を対象としている述語動詞に直接係っていない文節に対しても行う<sup>1</sup>。ただし、対象としている述語動詞より後方にある、被連体修飾名詞を除く文節は除外する。
6. 5. でもスコアが閾値を超える文節がないならば、その深層格となる文節は存在しないということに

<sup>1</sup>正確には、ここにも制約があるが字数の制限のため、省略する。

する。

各文節へのスコアの付け方は、以下のとおりである。

- 法令文に出現する多くの述語動詞が、文頭にある「～は、」という文節にある名詞を動作主格としていたので、この文頭の「～は、」という文節には動作主格としてのスコアを 10 点加えている。また、「～は、」が、対象格となる場合も数例あったので、対象格としてのスコアには 5 点を加えている。
- 格フレーム辞書の表層格に一致する格助詞を持つならば、+10 点。
- 副助詞、接続助詞を持つならば、+ 5 点。
- 文頭の「～は、」以外の文節においてスコア付けの対象としている文節から格解析の対象としている述語動詞の間にある読点の数を 3 倍した数をスコアから引いている。これは、文頭の「～は、」以外の文節は、読点を越えたところにある述語動詞の格にはなりにくいという性質を考慮したものである。
- スコア付けの対象としている文節の名詞に対して、格フレーム辞書内の深層格となっている名詞群との意味の類似度によって、スコアを付ける。

スコア付けの対象としている文節の主辞  $w$  と、格フレーム辞書にある名詞群  $C$  との意味の類似スコア  $sim\_score(w, C)$  は以下のように求める。まず、単語  $e_1, e_2$  間の類似度  $sim(e_1, e_2)$  は、日本語語彙大系のソーラスを利用する。

$$sim(e_1, e_2) = \frac{2L}{l_x + l_y}$$

$l_x, l_y$  はそれぞれ単語  $e_1, e_2$  のソーラスの根からの階層の深さ、 $L$  は  $e_1, e_2$  の意味属性で一致している階層の深さを表す。スコア付けの対象としている文節の主辞の名詞を  $w$ 、格フレーム辞書にある名詞群  $C$  のそれぞれの名詞を  $c_1, c_2 \dots c_n$ 、その名詞それぞれの頻度を  $f_1, f_2 \dots f_n$  とすると、意味の類似スコア  $sim\_score(w, C)$  は

$$sim\_score(w, C) = \frac{\sum_{i=1}^n sim(w, c_i) \times f_i}{\sum_{i=1}^n f_i} \times 10$$

とする。

図 2: 原子文生成の例

工事の施工者は、工事により生じる土砂、がれき、廃材等が、公共の場所に飛散し、又は流出しないよう、適正に管理しなければならない。

↓ 変換

管理 (e1) ∧ 施工者 (x1) ∧ agt(e1, x1) ∧ ( 飛散 (e2) ∨ 流出 (e2) ) ∧ goal (e1, e2) ∧ ( 土砂 (x3) ∨ がれき (x3) ∨ 廃材等 (x3) ) ∧ agt(e2, x3) ∧ 場所 (x4) ∧ loc(e2, x4) ∧ 生じる (e3) ∧ obj(e3, x3) ∧ 工事 (x5) ∧ way(e3, x5)

## 4 原子文の生成

格解析結果から述語動詞や名詞に変数を過不足なく割り振り、それらの関係を宣言する原子文を生成する。並列句と解析された語句を、選言によって原子文にすることも行う。例えば、「妊婦、障害者、又はけが人」を「妊婦 (x1) ∨ 障害者 (x1) ∨ けが人 (x1)」にするといった具合である。

## 5 評価実験

### 5.1 実験例

上記に基づくシステムが、法令文から原子文を生成した例を図 2 に示す。

### 5.2 解析結果

本研究で開発したシステムで、格フレームを構築する際に基にした「千代田区生活環境条例」の 3 条～12 条に出現する 71 個の述語動詞に対して格解析を行った。また、「広島市ばい捨て等の防止に関する条例」全 20 条に出現する 67 個の述語動詞に対しても行った。解析結果を表 2 に示す。表 2 における正解とは、正しい原子文を生成するのに必要な格解析を正確に行えたこととしている。すなわち、ある述語動詞に対して格となりうる文節を過不足なく発見できたということで

ある。千代田区条例に対する解析誤りの主な原因は構文解析の誤りと、格となる名詞がシソーラスになかったことである。広島市条例に対する解析誤りの原因をいくつか以下に挙げる。

- 原因の7割以上が、述語動詞が格フレーム辞書にないことであった。
- 千代田区条例に現れた際にとっていた深層格の名詞と、広島市条例に現れた際にとっていた深層格の名詞の類似度が低いために解析を誤った例が6例あった。例えば「防止する」の格フレーム辞書にある対象格の名詞群は「駐車」と「散乱」であるが、これらの語と広島条例で現れた際に「防止する」の対象格となっていた「行為」は類似度が低い。
- 辞書を構築する際には現れなかった深層格が解析の際に述語動詞に付随し、辞書の情報が不十分となったため正しく解析できなかった例が4例あった。例えば、「啓発する」は千代田区条例に現れた際には、動作主格と対象格しかとっていないが、広島市条例に現れた際には目標格もとっていた。
- 法令文には使役形や受動態はあまり現れないので、実装システムではこれらを考慮していない。このため、解析できなかった例が4例あった。
- 述語動詞の用法の多様性を考慮していない。このため「する」や「ある」といった動詞において、その用法の多様性のため解析できない例が3例あった。例えば、辞書には「活動をする」といった意味の「する」しか載っていないので、「公園を禁煙区域とする」といった意味の「する」を正しく解析できなかった。

## 6 まとめ

本研究では、格フレーム辞書を実際の法令文を基に構築し、それをを用いた格解析を行うシステム、およびその解析結果から原子文を生成するシステムを開発した。

今後は、格フレーム辞書のカバレッジを高くすることと、述語動詞の用法の多様性、使役形、受動態へ対処する必要がある。また、法令文の意味をより正確に

表 2: 解析結果

		正解	誤り
千代田区 生活環境条例 (3条～14条)	普通動詞	66	5
	連体修飾	20	5
	サ変名詞	16	10
広島市 ばい捨て防止条例 (3条～20条)	普通動詞	42	24
	連体修飾	10	22
	サ変名詞	7	18

論理式で表現するため、複合名詞を解析する必要がある。さらに、「(名詞) + の + (名詞)」といった名詞句の解析も今後の課題である。

## 参考文献

- [1] 片山. 検証進化可能電子社会 -情報科学による安心な電子社会の実現-, 情報処理, V46N5, pp.515-521, 2005.
- [2] 江尻, 北田, 島津. 法令文の論理式への変換 -論理構造について- 第12回言語処理学会, 2006.
- [3] 黒橋, 河原. 日本語形態素解析システム JUMAN version 4.0 使用説明書, 2003.
- [4] 黒橋. 日本語構文解析システム KNP version 2.0 b6 使用説明書, 1998.
- [5] 河原, 黒橋. 格フレーム辞書の漸次的自動構築, 自然言語処理, V12N2, pp.109-131 2005.
- [6] 吉野. 法律エキスパートシステムの基礎, ぎょうせい, pp.12-24, 1986.
- [7] 田中. 法律効果規定部の意味機能について, NL98-21, pp1-8, 1998.
- [8] 平松, 永井, 中村貞吾, 野村浩郷. 要件効果構造に基づく法律文統語構造解析, NL97-118, pp41-48, 1997.