JAIST Repository
https://dspace.jaist.ac.jp/

Title	法令文の論理式への変換 -原子文について-
Author(s)	北田, 安希雄
Citation	
Issue Date	2006-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1954
Rights	
Description	Supervisor:島津　明, 情報科学研究科, 修士

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

法令文の論理式への変換
-原子文について-

北田 安希雄 (410037)

北陸先端科学技術大学院大学 情報科学研究科

2006 年 2 月 9 日

キーワード: 法令文, 論理式, 原子文, 格解析, 格フレーム辞書.

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。

我々は、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、本研究は原子文への変換を行う。

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ので、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。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「警察署」、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。本研究では以上に述べた構造の解析や格解析を行う。

まず、実際の法令文を基にして格フレーム辞書を構築する。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、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。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。

Copyright © 2006 by Kitada Akio

次に、構築した辞書を用いて格解析を行う。格解析の手法は、まず、法令文を JUMAN [2]、KNP [3] により、形態素解析、構文解析を行う。次に、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し、とりうる深層格としてのスコア付けを、格の候補の文節に対して行う。そして、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。スコアは基本的には、その文節が格フレーム辞書の表層格に一致する格助詞を持つかどうか、主辞となる名詞と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、その文節と述語動詞との間にある読点の数などによって決める。加えて、法令文の特徴も考慮する。例えば、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るので、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高く付ける。

最後に、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。

上記に基づいて開発したシステムで、「千代田区生活環境条例」の３条〜１２条に出現する述語動詞に対して格解析を行ったところ正しく解析できたのは、７１個の普通動詞のうち６６個、２６個のサ変名詞のうち１６個、２５個の連体修飾語となる述語動詞のうち２０個であった。なお、ここでいうところの「正しく解析できた」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。主な誤りの原因は、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。また、「広島市ぽい捨て等の防止に関する条例」全２０条に出現する述語動詞に対しても格解析を行ったところ正しく解析できたのは、６６個の普通動詞のうち４２個、２５個のサ変名詞のうち７個、３２個の連体修飾語となる述語動詞のうち１０個であった。ここでの誤りの原因の約 6 割が、述語動詞が格フレーム辞書にないことであった。その他の原因としては、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。

これらの実験によって、上記の格フレーム辞書、および格解析により、法令文からその意味を表現する原子文がある程度生成できることが確かめられた。今後は、格フレーム辞書のカバレージを高くすることによって格解析の精度を向上する必要がある。

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] 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.

[3] 黒橋. 日本語構文解析システム KNP version 2.0 b6 使用説明書, 1998.