

表層パターンに基づく法律文の構文解析

Legal Sentence Parsing Based on Syntactic Patterns

大原誠¹
Makoto Ohara

松原茂樹²
Shigeki Matsubara

稲垣康善¹
Yasuyoshi Inagaki

名古屋大学大学院工学研究科¹
Graduate School of Engineering, Nagoya University

名古屋大学情報連携基盤センター²
Information Technology Center, Nagoya University

1 はじめに

法律、判例、議会関係資料など日本の法律文書のデジタル化が盛んに行われている。法律文書は大量に存在し、制定や追加、改正も頻繁に行われるため、翻訳や要約など計算機による文書加工技術の活用が望まれる。これらを実現する上で構文解析技術、特に、法律文書の特徴をとらえた高精度の解析手法が必要である。

そこで、本稿では、表層パターンと構文カテゴリに基づく構文規則を用いた法律文の構文解析手法を提案する。本解析手法は、法律文書に頻出する特有の言い回しに注目し、パターンマッチングとボトムアップチャート法をベースとする。表層パターンを構文規則として利用することにより、曖昧性の少ない解析の実現が期待できる。

2 法律文の構文的特徴

最も代表的かつ基本的な法律の1つである民法を用いて、法律文の特徴分析を行った。分析には、全1978文を用いた¹。平均文長、及び、異なり形態素数を表1に示す。比較対象として、毎日新聞(1995年版)の1978文を用いた。文長は、文中に含まれる形態素の数とした。異なり形態素数として、基本形が異なる単語を計数した。

表1 法律文と新聞記事の比較

	法律文	新聞記事
平均文長	30.0	22.7
異なり形態素数	1672	6617

民法に含まれる法律文は、新聞記事中の文より長い傾向がある。その一方で、法律文の異なり形態素数は、新聞記事の約25%に過ぎない。固有名詞の頻出する新聞記事と違い、法律文は、ある一定の語彙のパリエーションで構成されていると考えられる。

法律などの文章に現れる用語には、意味や使用方法が定まっているものがある[1]。また、法律文には、ある種の決まった言い回しが表層パターンとして存在し、文体や文の構造にもある一定の決まりがある。構文規則の作成においては、これらの特徴を考慮する必要がある。

3 法律文の構文解析手法

3.1 表層パターンに基づく構文規則

本手法では、表層パターンと構文カテゴリから構成される構文規則を用いる。構文規則の例を図1下に示す。このような構文規則の採用には、以下の利点がある。

- 構文規則を表層的まとまりとして表現することにより、小さな構文単位の解析が不要となり、曖昧性の少

¹民法は、総則、物権、債権、親族、相続の5編から成る。本来、第一編から三編までは文語体で記されているが、本研究ではそれらを口語体に変換した文章を用いた。

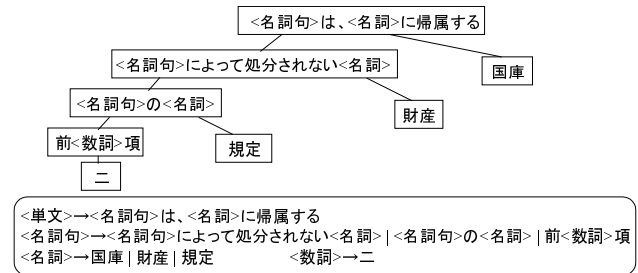


図1 構文木と構文パターンの例

ない高精度の解析が可能となる。

- 構文規則の集合は文脈自由文法の範疇に入るため、従来の構文解析アルゴリズムを利用できる。
- 対応する英訳法律文にも同様の構文規則を定めることにより、両言語間のフレーズレベルでの対応づけが容易となり、法律文書の翻訳にも利用できる。

本手法の構文パターンを用いて、「前二項の規定によって処分されない財産は、国庫に帰属する」(民法第七十二条第三項)を構文木で表すと、図1上のようなになる。木の各節点が、構文規則となる。

3.2 解析アルゴリズム

解析手法は、パターンマッチングとボトムアップチャート解析[2]に基づく。本手法は、文中の形態素が構文規則中の形態素列(表層パターン)にマッチするときのみ弧を張り、不活性弧を最左項に持つ活性弧を張る処理は行わない。このため、通常の解析より、弧を張る回数を抑えることができる。

民法の200文を用いて解析実験を行った。各文に構文木を付与し、そこから抽出した構文規則を使用した。構文規則の数は1037(うち、非終端記号を含まないものが382)である。実験の結果、186文(93%)の解析に成功し、本手法の利用可能性を確認した。

4 まとめ

本稿では、法律文に特有のパターンに基づいた構文解析手法を提案した。現在、民法の条文、及びその英訳文への構文木付与を進めており、今後、これらを活用した法律文書の多言語化手法について検討する予定である。

本研究を進めるにあたり、法情報学の観点から有益な御示唆を頂く名古屋大学法学部の松浦好治教授、フランクベネット助教授、角田篤泰助教授に感謝致します。

参考文献

- [1] 長谷川彰一, “法令解釈の基礎,” ぎょうせい, 2000.
- [2] M. Kay, “Algorithm schemata and data structures in syntactic processing”, *Technical Report CSL-80-12*, Xerox PARC, 1980.