

## 平成 26 年度 情報工学コース卒業研究報告要旨

外山 研究室	氏 名	水尻 隼人
卒業研究題目	法令文の並列構造解析のための CFG の構築	

法令文には、特有の言い回しや専門用語、複雑な並列構造が頻出するため、一般の人が正しく理解することは容易ではない。そのため、法令文の理解を助けることを目的に、法令文に現れる用語のシソーラスを構築する研究や、法令文を読みやすく表示する研究が行われている。これらの研究では、基盤技術として、法令文の並列構造解析技術の開発が強く求められている。シソーラスの構築では、並列関係の利用により、語の類義関係や上位下位関係の同定が期待できる。また、表示方法の開発では、並列関係を二次元上に表示することが読みやすさに繋がる。

一方、並列構造解析は、一般の文を対象とした場合でも困難なタスクとして知られている。それに加え、法令文は複雑な並列構造が頻出するため、法令文の並列構造解析に取り組んだ先行研究は、いずれも十分な精度を達成できていない。しかし、複雑な並列構造から生じる解釈の曖昧性を排除するため、法令文の記述方法には詳細な規則が存在している。並列構造を解析する上での文法制約としてそれらの規則を用いることができれば、高精度な解析を実現できる可能性がある。

そこで本研究では、法令文の高精度な並列構造解析を実現することを目指し、そのための一步として、法令文の並列構造に関する記述規則を形式化した文脈自由文法 (CFG) を構築する。法令文の並列構造に関する記述規則について、等位接続詞「及び」と「並びに」を例に説明する。両者とも、ある事物と他の事物を併合的に結び付ける語であるが、並列構造を階層的に表すために、それらの使い分けは明確に定められている。具体的には、「及び」は結びつきが最も強い並列関係を、「並びに」は「及び」よりも弱い結びつきの並列関係を表す際に用いられる。これらが用いられた階層的な並列構造の例を図 1 に示す。

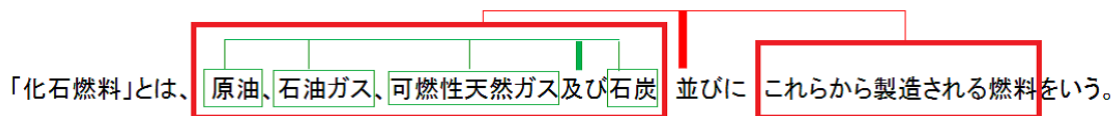


図 1: 「及び」と「並びに」による並列構造の例

この文の場合、「原油」と「石油ガス」、「可燃性天然ガス」、「石炭」の四つを最も強く結びつける並列関係が「及び」によって示され、「原油～石炭」と「これらから製造される燃料」を二番目に強く結びつける並列関係が「並びに」によって示される。このような規則は、他の等位接続詞（「又は」、「若しくは」、「かつ」）にもそれぞれ存在する。これらの規則を法制執務に関する複数の書籍に基づいて精査し、それらを形式化するために 30 種類の非終端記号、104 種類の生成規則からなる CFG を構築した。

構築した CFG により生成できる並列構造が、実際の法令文の並列構造をどの程度カバーしているのかについて、実験により評価した。実験には、人手で係り受け関係が付与された 414 法令文を用いた。これらには、719 個の並列構造（図 1 の場合、並列構造が 2 個あるとして集計）が存在していた。ボトムアップチャートパーザを用い、並列構造木の選択時にはオラクルが与えられるとして解析した結果、構築した CFG の生成規則により、全体の 90.1% (=648/717) の並列構造が生成できることを確認した。