

| | |
|--------------|---|
| Title | Recognition of Parallel Multiple Context-Free Grammars and Finite State Translation Systems |
| Author(s) | 楫, 勇一 |
| Citation | 大阪大学, 1994, 博士論文 |
| Version Type | VoR |
| URL | https://doi.org/10.11501/3075134 |
| rights | |
| Note | |

Osaka University Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

Osaka University

| | |
|------------|---|
| 氏名 | 梶 勇 一 |
| 博士の専攻分野の名称 | 博士(工学) |
| 学位記番号 | 第 11413 号 |
| 学位授与年月日 | 平成6年3月25日 |
| 学位授与の要件 | 学位規則第4条第1項該当 基礎工学研究科物理系専攻 |
| 学位論文名 | Recognition of Parallel Multiple Context-Free Grammars and Finite State Translation Systems (並列多重文脈自由文法および有限状態変換系の認識問題について) |
| 論文審査委員 | (主査) 教授 高 忠雄 (副査) 教授 都倉 信樹 教授 谷口 健一 教授 菊野 亨 助教授 関 浩之 |

論文内容の要旨

文脈自由言語のクラス (CFL) より真に大きく文脈規定言語のクラス (CSL) より真に小さい言語クラスについて、多くの研究がなされている。自然言語処理の観点、および計算理論の観点から見て、この「隙間」にあるいくつかの言語クラス間の関係、性質を明らかにすることは非常に重要である。

伝統的な形式文法を用いて自然言語の構文を記述しようとした場合、文脈自由文法 (CFG) では生成能力が弱過ぎ、例えば *respectively* 構文のような互いに割り込んだ句構造を簡潔に表現できないことが指摘されている。一方、文脈規定文法 (CSG) では生成能力が強過ぎ、多項式時間認識アルゴリズムが知られていないことからわかり、効率的な処理が困難である。これらの問題を解決するために、CFG より真に強い生成能力を持ち、かつ、効率的な取り扱いが可能であるような新しい文法、例えば *head grammar* (HG), *tree adjoining grammar* (TAG) や *generalized context-free grammar* (GCFG) などが提案されている。とくに GCFG は、CFG の自然な拡張であること、また句構造が簡潔に定義できることなどから、非常に興味深い定式化である。しかし残念ながら、GCFG の生成能力は 0 型文法と等しくなることが示されている。

著者の属する研究グループでは、GCFG の部分クラスとして並列多重文脈自由文法 (*parallel multiple context-free grammar*, PMCFG) および多重文脈自由文法 (*multiple context-free grammar*, MCFG) を提案している。PMCFG および MCFG の非終端記号 A には次元と呼ばれる正整数 $d(A)$ が付与されており、 A は終端記号の $d(A)$ -項組を生成する。CFG は、全ての非終端記号についてその次元が 1 であるような、特殊な MCFG と考えることが出来る。Vijay-Shanker らによって提案された *linear context-free rewriting systems* は、本質的には MCFG と同様の定式化である。PMCFG, MCFG の生成する言語はそれぞれ、並列多重文脈自由言語 (PM CFL), 多重文脈自由言語 (MCFL) と呼ばれ、どちらも多項式時間で認識可能であることが知られている。また、MCFL のクラスは TAG や HG によって生成される言語クラスを真に含み、PM CFL のクラスに真に含まれること、PM CFL のクラスは CSL に真に含まれることが示されている。すなわち、以下の包含関係

$$CFL \subseteq TAL = HL \subseteq MCFL \subseteq PM CFL \subseteq CSL$$

が示されている。

PMCFG, MCFG をはじめ、これら新しく提案された文法については数多くの研究がなされている。とくに、さまざまな文法クラスに対して一般認識問題の計算量が議論されており、それらの結果を比較検討することは、各々の文法

の「簡潔さ」を測るという意味で非常に興味深いものがある。

本論文第3章では、PMCFG, MCFG およびその部分クラスに対する一般認識問題の計算量について議論する。文法クラス C に対する一般認識問題とは、 C に属する文法 G と終端記号列 w を入力とし、 G が w を生成するか否かを決定する問題である。文法 G が入力の一部となっていることから、その文法の「簡潔さ」が同問題の計算量に影響してくる。例えば、複雑な言語をコンパクトに記述することが出来るような文法クラスに対する一般認識問題の計算量は、文法サイズに比して非常に大きくなる。一方、同じ言語を生成するのに非常に大きな記述量を必要とするような文法クラスに対しては、同問題の計算量は文法サイズに比して小さくなる。すなわち、一般認識問題の計算量を明らかにすることにより、その文法クラスの「簡潔さ」を定量的に測ることが出来る。第3章では、PMCFG, MCFG に対する同問題がどれも多項式指数時間完全であることを示す。また、情報無損失条件を満たす PMCFG, MCFG, 次元が定数であるような PMCFG, MCFG および自由度が定数であるような PMCFG, MCFG に対する同問題が、各々 PSPACE 完全, NP 完全, P 完全となることを示す。さらにこれらの理論的結果に基づき、一般認識問題の計算量と、各クラスの簡潔さ、および各クラスを用いて言語の構文を獲得するときの難しさとの関係について議論する。

本論文第4章では、変形文法のモデルとして提案された有限状態変換系 (FSTS), および FSTS の部分クラスの生成能力について議論する。FSTS の部分クラス化による生成能力の階層については、過去 Engelfriet らによって研究がなされてきた。しかし、その階層と他の言語クラスとの関係については十分な研究がなされているとはいえない。本論文第4章では、決定性 FSTS の生成能力が PMCFG と等しいことを示す。決定性 FSTS の生成する言語クラス L (FSTS) については、従来からクラス P の部分クラスとなることが知られていたが、その認識に必要な計算量については知られていなかった。第4章では、PMCFG の自由度に相当する概念を決定性 FSTS に導入することにより、 L (FSTS) に属する言語は $O(n^{e+1})$ 時間で認識可能であることも示す。ここで n は入力系列長、 e は決定性 FSTS の自由度である。本論文ではさらに、state-bound 2 の非決定性 monadic FSTS の生成する言語クラス中に、NP 完全な言語が存在することを示す。

論文審査の結果の要旨

並列多重文脈自由文法 (PMCFG) は、文脈自由文法 (CFG) の拡張として導入された形式文法であり、言語演算に対する閉包性など、CFG の主性質を受け継いでいる。また PMCFG は、自然言語の構文記述を目的として提案された形式文法である Tree Adjoining Grammar や Head Grammar よりも生成能力が真に大きいことが知られている。一方、有限状態変換系 (FSTS) は、変形文法や構文指向変換の代表的モデルであり、Engelfriet らによって盛んに研究されてきた。

本論文は、これら PMCFG および FSTS の基本的性質、特に認識問題と生成能力について研究を行なった結果がまとめられている。

第3章では、PMCFG およびその部分クラスに対する一般認識問題の計算量を明らかにしている。文法に対して全く制約を課さない場合、同問題は多項式指数時間完全となること、文法のいくつかの部分クラスを考えた場合、同問題の計算量が PSPACE 完全, NP 完全, P 完全となることを示している。特に、自由度と呼ばれるパラメータを定数以下に限定した部分クラスに対しては、文法の記述長についても多項式時間認識可能であることを示している。これらは理論的に興味深い結果であるとともに、PMCFG の文法サイズが構文解析の効率にどのように影響するかを明らかにした点で意義あるものである。

第4章では、FSTS およびその部分クラスの生成能力について議論を行なっている。まず、決定性 FSTS の生成能力が PMCFG の生成能力と等しいことを示している。この結果より、PMCFG, FSTS の一方についての既知の結果を他方の性質の解明に利用できる場合がある。例えば、決定性 FSTS にも自由度と呼ばれるパラメータを定義することにより、決定性 FSTS の生成する言語は、オーダ n^{e+1} 時間 (n は入力系列長、 e は与えられた決定性 FSTS の自由度) で認識可能であることが直ちに導かれる。さらに、state-bound 2 の非決定性 monadic FSTS の生成する言語クラス中に、少なくとも一つ NP 完全な言語が属することを証明しており、非決定性構文指向変換の能力の高さを理論

的に示している。

これらの成果は形式言語理論に対する大きな貢献であり、本論文は博士論文として価値あるものと認める。