

平成 10 年 2 月 23 日

システム情報工学専攻	学籍番号	903719
申請者氏名	椎名 広光	

指導教官氏名	増山 繁 教授 中川 聖一 教授
--------	---------------------

論文要旨 (博士)

論文題目	構文解析の高速化に関する研究
------	----------------

(要旨 1,200 字程度)

自然言語処理やプログラム言語の処理は多くの場合、形式言語理論における文脈自由言語及びその部分クラスの言語を対象としており、既存の解析手法もそれに合わせて開発している。言い替えば、現状では文脈自由言語やその部分クラスの言語の処理方式に合わない部分を無視して処理している。

しかしながら、現実の言語現象には、文脈自由言語の枠組では表現できないものが少なからず存在する。また、コンピュータの並列化や処理速度の高速化によって、これまでの逐次処理より高速な構文解析や、文脈自由言語より広いクラスの言語に対する実用的な構文解析が実現できる可能性がでてきている。

本論文では、文脈自由文法より小さいクラスの文法である単純順位文法と LR 文法に対しては並列構文解析アルゴリズムを考案し、文脈自由文法より広いクラスに対しては高速化可能な文法 ($unrestricted\ LR(k)$ 文法) の定義やその文法に対する解析法の考案、および句構造文法に対する解析順を提示した構文解析法を提案した。

コンパイラの構文解析に部分的に使用されることが多い単純順位文法に対して、構文解析木の左右の子供の節点候補を計算するのに最大値と最小値を計算すれば良いことを利用して、文脈自由文法や LR 文法に対する構文解析の並列構文解析アルゴリズムに比べて使用するプロセッサ数が $O(n^2)$ 個と少なく、また $O(\log^2 n)$ 時間で実行されるアルゴリズムを提案した。

コンパイラなどの構文解析で良く使用される LR 文法に対して、 LR 状態遷移図の状態と文法記号間の関係を前処理として計算しておくことによってループの回数を削減し、 $O(n^3)$ 個のプロセッサを用いて $O(\log n)$ 時間で処理する並列アルゴリズムを提案した。先読み文字列を入力文字列から非終端記号に変更することによって、文脈依存文法のクラスより部分的に広いクラスをカバーする $unrestricted\ LR(k)$ 文法を提案した。また、その文法を利用した構文解析が出力される構文解析木の大きさに対して定数倍で処理できることを明らかにした。

そして最後に構文解析を行なう対象とする言語のクラスを、句構造文法の一つである $unrestricted$ 文法が生成する言語のクラスとし、その言語を構文解析する方法として GLR 構文解析を拡張した $UGLR$ 構文解析を新たに提案した。また、この $UGLR$ 構文解析器が受理できる言語のクラスが文脈自由言語のクラスを含むことを、 $UGLR$ 構文解析に与える文法に制限をつけると、 $UGLR$ 構文解析の動作が Earley 法と一致することによって明らかにした。