

## 論文の内容の要旨

論文題目	木接合言語のクラスの特徴付けについて
学位 申請者	松原 俊一

本論文は、木接合言語と呼ばれる形式言語のクラスを研究対象とし、このクラスに関する二つの特徴付けを新たに示している。木接合言語のクラスは、自然言語の形式的な言語のクラスとして、近年の計算言語学において広く支持を集めている。本研究の一つ目の特徴付けは、文脈自由言語に関する表現定理としてよく知られている Chomsky-Schützenberger の定理の文字列から木への一般化であり、二つ目は、自然言語の処理系での応用を目的とした拡張範疇文法と呼ばれる新しい形式文法による特徴付けである。

計算機科学では、文脈自由文法と呼ばれる形式文法が、プログラミング言語の記述などの目的でよく用いられる。しかし自然言語には、文脈自由文法で扱うことの難しい現象がよく現れることから、文脈自由文法の生成する言語のクラスは、自然言語の形式的な言語のクラスとしては妥当ではないと考えられている。一方で、木接合言語は、多項式時間の認識アルゴリズムが考案されている点などで、文脈自由言語と類似した性質を持ちながらも、自然言語の統語現象を表現するのに十分な記述能力を持っている。こうした点から木接合言語が自然言語の形式化として支持を集めている。

本論文の一つ目の成果は、任意の木接合言語  $L$  に対して、認識可能集合  $R$  が存在し、拡張 Dyck 言語と認識可能集合  $R$  の共通集合の準同型像として、 $L$  が表されるという定理の証明である。この結果は、文脈自由言語についての表現定理として有名な Chomsky-Schützenberger の定理を拡張することで得られたものである。木接合言語が持っている文脈自由言語との類似性に注目し、文字列上で定義された括弧構造などの概念を、木の上へと一般化することにより、Chomsky-Schützenberger の定理を拡張している。Chomsky-Schützenberger の定理とは、任意の文脈自由言語  $L$  に対して、正則集合  $R$  が存在し、Dyck 言語と正則集合  $R$  の共通集合の準同型像として  $L$  が表されるという定理である。Dyck 言語とは、括弧を含んでいるアルファベット上の文字列のうち、全ての括弧が整合している文字列の全体からなる言語のことである。括弧の整合性の概念を文字列から木へと拡張することにより、拡張 Dyck 言語を新たに導入している。認識可能集合は、文字列上の正則集合の木言語への拡張で、

基本的な木言語の一つとして知られている。

二つ目の成果は、従来の範疇文法の拡張として、拡張範疇文法と呼ばれる文法モデルを新たに導入し、この文法モデルの定義する言語のクラスと木接合言語のクラスとの一致を証明したことである。この文法モデルの導入の動機は、自然言語の処理系への応用にある。構文の情報を型と呼び、型の集合を各単語に割当てることにより、構文の規則を辞書に分散させる点や、単語だけで構文木を簡潔に表せる点が、範疇文法から継承した特長である。範疇文法と拡張範疇文法のいずれの文法も、各単語について型を選択し、隣り合った二つの型を一つに書換える還元と呼ばれる演算を繰返していくことにより、文の構造を決定する。範疇文法が、一つの型の列の上で型の書換えを行うのに対し、拡張範疇文法は、補助的な型の列を新たに導入することで、二つの列を連動させながら還元していく。この操作は、新しい型とそれに対応した還元の規則を導入することによって実現されている。この仕組みにより、範疇文法では困難とされる「wh移動」などの自然言語の統語現象を明示的に捉えることができる。さらに拡張範疇文法は、転送スタック付きプッシュダウンオートマトンと呼ばれるオートマトンモデルと対応付けられているため、処理系を実現しやすいという利点を持っている。木接合言語のクラスを特徴付ける形式的な文法モデルには、既に、木接合文法など様々な文法が考案されてきた。しかし、それらの文法モデルでは、「wh移動」などの統語現象を直接的に捉えることが容易ではない。さらに、それらの文法モデルは、構文木を簡潔に表すことが難しいという欠点を持っている。今回新たに導入した拡張範疇文法は、こうした問題を解決している点で、既存の文法モデルにはない大きな利点を持っている。

本論文の構成は次の通りである。第1章では、木接合言語のクラスに関する先行研究を取り上げ、本論文の研究背景を述べている。第2章では、本論文全体を通して必要な概念や記法などを定義している。第3章は、一つ目の研究成果である拡張Dyck言語による特徴付を示し、第4章は、二つ目の研究成果である拡張範疇文法について述べている。第5章では、研究結果をまとめたうえで、今後の展望や未解決問題について述べている。

## 論文審査の結果の要旨

学位申請者氏名 松原 俊一

審査委員主査 笠井 琢美

委員 岩田 茂樹

委員 阿部 公輝

委員 小林 聡

委員 西野 哲朗

委員

委員

木接合言語は、従来研究されてきた文脈自由言語よりも強力な表現能力を持ち、かつ、効率のよい処理アルゴリズムが提案されており、最近注目を集めている言語である。また言語学者からも、自然言語の形式的な言語として幅広く支持されている。

本論文は、木接合言語と呼ばれる形式言語のクラスを二つの側面から特徴付けている。一つ目は、木接合言語のクラスについての新たな表現定理によるものである。文脈自由言語に関する定理としてよく知られているChomsky-Schützenbergerの定理を文字列から木へと一般化することにより、新たな定理を示している。二つ目は、新たに導入した拡張範疇文法と呼ばれる文法モデルによる特徴付けである。この文法モデルは、転送と呼ばれる概念の導入により、従来の範疇文法を拡張したものであり、自然言語の処理系で用いることを目的としている。

本論文の第2章では、任意の木接合言語  $L$  に対して認識可能集合  $R$  が存在し、拡張 Dyck 言語と認識可能集合  $R$  の共通集合の準同型像として  $L$  が表されるという定理が証明されている。この定理は、文脈自由言語についての表現定理として有名な Chomsky-Schützenberger の定理の文字列から木への一般化である。木接合言語と文脈自由言語の類似性に注目し、文字列上で定義された概念を木の上へと一般化することによって、文脈自由言語の場合と類似した定理が木接合言語についても成り立つことを示している。Chomsky-Schützenberger の定理とは、任意の文脈自由言語  $L$  に対して正則集合  $R$  が存在し、Dyck 言語と正則集合  $R$  の共通集合に関する準同型像として  $L$  が表されるという定理で、文脈自由言語が本質的に括弧構造を持っていることを明らかにしている基本的かつ重要な定理である。本論文では、文字列上で考えられている括弧構造の概念を木の上へと拡張することによって、拡張 Dyck 言語と呼ばれる木言語を導入している。新たに得られた定理は、木接合言語の本質的な構造を明らかにしているという点で有意義なものと言える。また、木接合言語よりも簡単に表すことのできる認識可能言語を適当に定めることで、任意の木接合言語を表すことができるという点でも興味深い結果であり、注目に値する。

第3章では、従来の範疇文法を拡張することにより、自然言語の処理系への応用を目的とした拡張範疇文法という文法モデルを導入している。そして、拡張範疇文法の定義する言語のクラスと木接合言語のクラス的一致を証明している。型と呼ばれる構文の情報の集合を各単語に割当てることにより、構文の規則を辞書に分散させる点や、単語だけで構文木を簡潔に表せる点が、範疇文法から継承した特長である。一方で、従来の範疇文法が、文脈自由文法と等価であったのに対して、拡張範疇文法は、文脈自由文法よりも強い能力を持っており、木接合言語を生成することができる。これを自然言語の形式化という視点から見た場合、生成文法理論で発見された「移動」と呼ばれる統語現象を、範疇文法で記述するのが難しいのに対し、拡張範疇文法では十分に記述できるという点に現れている。さらに拡張範疇文法は「移動」を形式的に表現できるというだけでなく、処理系の実現といった計算機での応用を考慮した設計となっている点で有用な文法モデルであると言える。転送スタック付きプッシュダウンオートマトンと呼ばれるオートマトンモデルとの対応付けがなされている点も、処理系を実現するうえで利点となる。木接合言語のクラスを特徴付ける形式的な文法モデルとしては、これまでも様々なものが提案されてきた。しかし、生成文法理論で展開されている文法論を、計算機上で実現しやすい形で形式化したモデルは、拡張範疇文法の導入以前にはなかった。この点で拡張範疇文法は評価に値する。また、言語学という異なる研究分野の成果を積極的に取り入れ、有効に活用できるという点でも価値ある研究であると考えられる。

以上のように本論文は、Chomsky-Schützenbergerの定理の一般化によって、木接合言語の本質的な性質を明らかにしている。さらに、拡張範疇文法と呼ばれる形式文法を導入し、その文法によって木接合言語を特徴付けることで、木接合言語の表現能力を計算機上で有効に活用するうえで重要な結果を示している。

よって、本論文は博士（理学）の学位論文として十分な価値を有するものと認める。