審 査 の 結 果 の 要 旨

氏　　名　　宮尾　祐介

本論文は、文法理論に基づく深い統語解析を実用化する研究についてまとめたものである。具体的には、実世界の文を網羅する大規模な文法を開発するための方法論、および、曖昧性解消のための確率モデルを提案している。さらに、現代文法理論の一つである主辞駆動句構造文法（HPSG）に基づく英語解析器を実際に開発し、実世界テキストの統語解析実験により、提案の有効性を実証している。このように、本論文は統語解析の実用化のための手法の提案、および、実際に統語解析器を構築することによる実証に主たる貢献がある。以下に、各章について説明する。

第１章では、文法理論に基づく統語解析の必要性、および、統語解析を実用化する際の困難について議論し、その解決案を提示している。第一に、文法の一貫性・無矛盾性を維持することが大規模文法開発の主たる困難であるとし、そのための解決案として、コーパス指向の文法開発を提案している。第二に、実用アプリケーションで統語解析器を利用するためには最適な解析結果一つを選択する必要があることを指摘し、統語解析の曖昧性解消のための素性森モデルを提案している。第２章では、文法理論と統語解析に関する既存研究と、曖昧性解消によく用いられる確率モデルとして最大エントロピーモデルについて簡潔にまとめている。

第３章では、大規模な文法を効率的に開発するための方法論として、コーパス指向の文法開発について議論している。従来の文法開発においては、文法規則と辞書項目を開発することが主目的であった。ところが、文法が大規模化するにつれてこれらの一貫性を維持するのが困難になる。そこで、文法開発過程に解析済みコーパス（ツリーバンク）の構築を持ち込むことを提案している。ツリーバンクは、実際のテキストに対する統語構造の実例である。これを文法開発中に構築することで、文法の修正・拡張が実際の文の解析にどう影響するかを検証することができる。さらに、ツリーバンクは辞書項目の情報を真に含んでいることから、実際には、文法規則とツリーバンクを開発することを文法開発の主目的とすればよいということになる。本論文では、効率的にツリーバンクを構築するため、既存の CFG ツリーバンク Penn Treebank を HPSG ツリーバンクに変換する手法を提案している。実際にツリーバンクを変換するパターンルールを詳述し、HPSG ツリーバンクが妥当なコストで開発できることを示している。

第４章、第５章では、曖昧性解消のために統語解析の確率モデルを構築する手法として、素性森モデルについて議論している。文法理論に基づく統語解析に関する既存研究では、最大エントロピーモデルを適用して統語構造の確率モデルを構築していた。しかし、モデルのパラメタを推定するには文法が導出する全ての統語構造を列挙する必要があるが、一般に統語構造は文長に対して指数爆発するので、実用化は不可能であった。第４章では、指数個の木構造の集合を多項式サイズで表現するデータ構造として素性森を定義し、その上に最大エントロピーモデルを定義することで、パラメタが現実的コストで推定できることを証明している。第５章においては、HPSG 文法が導出する構文木および述語項構造が素性森を用いて表現できることを示している。これらにより、統語構造の確率モデルを実用化するための理論的基盤を確立している。

第６章では、本論文で開発した HPSG に基づく英語解析器を、新聞テキストのコーパスである Penn Treebank の統語解析実験において評価している。まず、文法の被覆率（正しい統語構造を導出できる割合）を測定し、85%近くの文を被覆できることを示した。これは既存の文法よりはるかに高い結果であり、コーパス指向の文法開発が実用的な大規模文法の開発に有効であることが実証されている。次に、素性森モデルを適用することで曖昧性解消を行い、解析精度を測定している。解析器が出力する述語項関係の精度は 86〜

87%を達成しており、実用レベルに達していることを実証している。さらに、統語解析エラーの原因などの詳細な分析を行っており、英語解析器の実用性と、さらなる性能向上に向けての指針を示している。

以上のように、本研究は文法理論に基づいた深い統語解析を実用化するという目的に対して、文法開発の方法論と曖昧性解消モデルを提案し、その有効性を確認した世界で最初の研究となっている。本研究で開発された英語解析器の性能も世界水準に達するものであり、今後の自然言語処理研究の基盤となる貴重なものとなっている。

よって本論文は博士（情報理工学）の学位請求論文として合格と認められる。