

## 審査の結果の要旨

氏名 原 忠 義

本論文では、自然言語の構文解析器の性能を向上させるための二つの方策が提案されている。一つは構文解析器の既存の統計モデルを新規の対象分野に適応させる方策である。もう一つは構文解析のエラーを体系的に分析する方策であり、エラーの分析結果により構文解析器の改良に要する作業を軽減することを目標としている。二つの方策は互いに相補的な関係にあり、二つを組み合わせる試みについても報告されている。

本論文は5章から成り立っている。

第1章では、本研究の目的と概要および本論文の構成について述べられている。

第2章では本論文の背景知識について説明されている。特に、本研究で扱う構文解析器について解説されている。本研究で扱う構文解析器は、HPSG（主辞駆動句構造文法）と呼ばれる言語形式に基づいた深い構文解析器である。深い構文解析器の出力は、浅い構文解析器が出力する単純な木よりも豊富な表現力を有しているが、深い文法が持つ複雑さのために、その性能の向上は困難かつ時間を要するものとなっている。

第3章において最初の方策が詳述されている。この方策は、既存の統計モデルをもとに、対象分野のコーパスを用いて統計モデルを再学習する手法から成り立っている。実際に、Penn Treebankから得られた統計モデルが、GENIA（生医学分野コーパス）およびBrown（文学分野コーパス）によって再学習され、本手法が効率的かつ効果的であることが実証されている。特に、学習にかかる時間が従来手法の7.5倍少なく済むという高い効率が示されている。

第4章では二番目の方策について解説されている。この方策は二つの手法から成り立っている、最初の手法は、アノテーションされた対象分野のコーパスを利用して、構文解析のエラーを言語的に意味のあるパターンに分類する。具体的に、40%以上のエラーを14パターンに分類できたことが報告されている。二つ目の手法は、エラーの間の依存関係を捉えるものである。特定のエラーを強制的に修正することにより別のエラーが消滅するとき、後者は前者に依存すると考えられる。このような依存関係を用いて、より本質的なエラーを同定することができる。二つの手法を組み合わせ、エラーのパターンの間の依存関係を求めることもできる。この章の最後では、第3章の再学習に際してエラーがどのように低減されるかが解析されており、本論文の二つの方策を組み合わせる試みが成されている。

第5章では、本論文の結論と将来の課題が述べられている。

以上をまとめると、本研究はHPSGに基づく深い構文解析器の性能を向上させるために、既存の統計モデルを新規の対象分野のコーパスを用いて再学習する方策と、エラーを分類するとともに相互の依存関係を求める方策を提案し、それらの有効性を実証している。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。