

## 論文審査の結果の要旨

氏名 二宮 崇

本論文は、単一化文法の一つである主辞駆動句構造文法 (Head-driven Phrase Structure Grammar) を用いる高効率かつ頑健な自然言語処理システムを実現するための効率化に関する手法および頑健化に関する手法について論じている。本論文は7章から成り立っており、2章と3章において並列処理を活用した高効率な自然言語処理システムについて、4章から6章にかけて文法規則の獲得による自然言語処理システムの頑健化について述べられている。1章は序論、7章は結論にあてられている。

2章において並列構文解析に関して述べられている。本研究では、まず、素性構造を並列計算機上で高効率に処理するためのプログラミング環境を提供し、その上で高効率な並列構文解析器を実現する。プログラミング環境はエージェントモデルに基づいており、素性構造を処理するための特殊なエージェントを提供することにより、素性構造処理のカプセル化、および高効率化が同時に実現される。素性構造処理の効率化は、i) 共有メモリを通して素性構造を転送する、ii) オンデマンドで素性構造をコピーする、および、iii) 複数のエージェントが同時にメモリに対し読み書きできる機構を用意することにより実現される。

3章において、以上の特徴を持つプログラミング環境上で開発された並列構文解析アルゴリズムについて述べられている。このアルゴリズムはCKYアルゴリズムを基にしており、CKY表と呼ばれるデータ構造の各要素に対しエージェントを割り当て、各エージェントは割り当てられた要素を計算する単純なタスクを処理することにより構文解析全体が行われる。以上の並列化によって本研究の構文解析器は非常に高速に稼動し、50台用いた場合に、その解析時間はEDR日本語コーパスを用いて一文あたり78ミリ秒であり、台数効果は13.2倍に達した。また、並列構文解析アルゴリズムの計算量についても言及されている。理論的な計算量は逐次計算機による解析も並列計算機による解析も共に指数オーダーであるが、経験的計算量においてはほぼ線形時間で解析されることを実験を通して確認されている。

4章では、素性構造に存在するいくつかのパスの値をインデックスとみなすことにより高速に単一化可能な文法規則を求める手法が提案されている。本論文の方法で獲得される文法規則は数千から数万のオーダーに

なる。この手法は、実行時構文解析において単一化可能な文法規則を高速に検索するために用いられる。

5章では、次章の文法規則の獲得手法において用いられるデフォルト単一化について述べられている。本研究で提案されている単一化手法は *lenient default unification* と呼ばれ、捕捉ルールと単一化可能な範囲において、単一化結果に含まれる情報を極大化するように定義されている。

6章において、文法規則の獲得手法について述べられている。本研究では「オフラインの頑健な構文解析」を行い、その解析結果から被覆率を向上するための文法規則を新たに獲得するという手法をとっている。「オフラインの頑健な構文解析」とは、非実行時に静的に訓練コーパスに対して構文解析を行う構文解析を行い、獲得した文法規則を現実の応用先で適用しつつ構文解析を行うことをいう。オフライン得られた文法規則にはオフラインの解析時の (1) 適用条件、および (2) 適用結果が反映されているため、これらの文法規則を用いた構文解析では過剰な解析結果を出力することない。

また、「オフラインの頑健な構文解析」は次にあげる三つの利点をもつ。(a) 実行時解析ではないので多少解析時間がかかってもかまわない。(b) 括弧付きコーパスなどの正解コーパスを参照することにより誤った補足ルールの適用を抑制できる。すなわち、探索空間をより狭くすることを可能にし、正解コーパスと矛盾しないという意味において正しい解析結果を返すことを可能とする。(c) 最終的には人手により解析結果を確認できる。

「オフラインの頑健な構文解析」における補足ルールの適用方法は、(i) 元来与えられていた文法規則を汎化するか、(ii) 非主辞に相当する素性構造を汎化するかによって二つの方法に分かれる。

(i) の方法による解析結果から文法規則を獲得する場合は、汎化された文法規則に適用条件を付加することにより新しい文法規則が得られる。この方法により得られた文法規則には、元の文法規則に記述されていた構造共有の情報も残されるため、構文木中の子ノードに記述された情報を親ノードへ伝播する能力が保たれる。

(ii) の方法により得られた解析結果から学習する場合は、得られた構文木の断片そのものを文法規則とみなすことにより文法規則を獲得する。このような文法規則だけを用いて一般的な自然言語現象を記述するのは困難であるが、これらの文法規則は例外的な親ノードの導出規則を容易に

表現するため、元々用意されていた文法規則とあわせて用いることにより、元の文法規則だけでは捉えることが困難であった例外的自然言語現象の捕捉が可能になる。

6章では、以上二つの手法を用いて得られた文法規則を用い、テストコーパスに対する解の数および被覆率を各々実験を通して確認している。

なお、本論文は、辻井潤一氏、田浦健次朗氏、鳥澤健太郎氏との共同研究であるが、論文提出者が主体となって分析及び検証を行ったもので、論文提出者の寄与が十分であると判断する。

以上の結果により、博士(理学)を授与できると認める。