

## 論文の内容の要旨

論文題目 Efficient HPSG Parsing with Supertagging and CFG-filtering  
(Supertagging と CFG-フィルタリングを用いた効率的な HPSG 構文解析)

氏名 松崎 拓也

本論文では、決定性の曖昧性解消器をスーパータギングおよび CFG フィルタリングという 2 つの前処理技術と組み合わせた、高効率な HPSG 構文解析手法を提案する。近年の研究により、語彙化文法の構文解析を高速かつ精度よく行うために、スーパータギングと呼ばれる手法が有効であることが明らかになっている。スーパータギングとは、入力文の各単語に対し少数の語彙項目をあらかじめ選択することで、構文解析の際の探索空間を狭めるという手法である。スーパータギングを用いた既存の解析手法のうち多くのもでは、スーパータギングの際の誤りによる構文解析の失敗を防ぐためにチャートパーズングが用いられており、これが最も計算時間を要する部分となっている。スーパータギングの際の誤りによる構文解析の失敗は HPSG 文法の場合に特に問題になる。これは以下の 2 つの理由による。一つ目は、HPSG 文法における語彙項目の数が一般に多数であることで、これは大規模解析済みコーパスから獲得した文法の場合に特に顕著である。このことは、他の比較的少数の語彙項目を持つ文法に比べ、HPSG におけるスーパータギングを難しくしている。二つ目は、チャートパーズングによる HPSG 構文解析はそれ自体が一般に高コストであることで、これは文法で用いられているデータ構造が複雑であることによる。このことは、各単語に割り当てる語彙項目数の増加が解析時間を急激に増大させる原因となる。

本論文では、スーパータギングに CFG フィルタリングの手法を組み合わせ、さらに強力な前処理を行うことで、非常に高速な構文解析が可能であることを示す。CFG フィルタリングは、CFG による HPSG の近似に基づく前処理手法である。CFG フィルタリングとスーパータギングを組み合わせることで、ほぼ確実に解析可能であるような語彙項目列を高速に列挙することが可能となる。具体的には、以下のように構文解析を行う。まず、CFG フィルタリングを用いて、入力単語列に割り当てる語彙項目列のうち大域的に一貫性のあるものを、スーパータギングによるスコアの順に列挙する。次に、それらの語彙項目列をひとつずつ曖昧性解消器へ入力し、最初に得られた構文木を出力とする。この手法においては、既存手法におけるチャートパーザを決定性の曖昧性解決器に置き換えることができ、非常に高速な構文解析を実現できる。曖昧性解消器は決定性のシフト還元型構文解析器として実装され、分類器を用いて構文解析動作を順次選択していくことで構文木をひとつ選び出す。この構文解析動作の際にも、近似 CFG による入力文上の構文森を利用する。大規模 HPSG 文法を用いた実験の結果から、提案手法は、同じ文法を用いた既存手法に比べ同程度の解析精度をもち、約 10 倍(20 ミリ秒/文)の速度を達成することが明らかになった。