

論文の内容の要旨

論文題目 Extensions and Restrictions of Abstract Categorical Grammars

(抽象的範疇文法の拡張と制限)

氏名：吉仲 亮

<序論>

いわゆる範疇文法の重要な特長として、自然言語の文とその意味表現が、ある形式論理体系の中の1つの証明図から同時に得られることがある。Lambek 計算上に構築される Lambek 文法は最も代表的な範疇文法であるが、しかし Lambek 文法には説明が困難な自然言語上の現象も見つかっており、そのような問題を解決するべく多様な改善案が提示されてきた。その流れの中で、2001年に de Groote は単純型付き線形ラムダ項に基づく文法形式である「抽象的範疇文法 (Abstract Categorical Grammar)」を提唱した。抽象的範疇文法は、Lambek 文法の枠組みを高度に抽象化したものであり、文中の量子子の作用域の曖昧性などの自然言語上の困難な現象を洗練された仕方でも説明できるといった特長を持っている。その生成する言語は、単純型付き線形ラムダ項の集合である。単純型付きラムダ項は、文字列やモンタギュー流の意味表現のみならず、木やリストなど多様なデータ構造も扱える抽象的な記述であるから、したがって抽象的範疇文法の扱えるデータ構造も、文と意味に限らず多岐にわたることになる。

しかしながら、抽象的範疇文法の数理的性質については、典型的な範疇文法よりもむしろ、いわゆる弱文脈依存文法との関連が指摘されている。文脈自由文法が自然言語を記述するためには不十分であることがわかって以来、弱い文脈依存性を扱える、弱文脈依存文法と呼ばれる多様な文法形式が提唱され、研究が盛んになっている。これらの文法は、それぞれ、文字列を直接

扱ったり、木構造を用いて生成する言語を定義している一方、抽象的範疇文法は、文字列も木構造も特殊なラムダ項として扱うから、従来の文法形式と同列に比較し競合すべきものではない。むしろそれらを一般化し、統一的な枠組みの中で記述する装置たることが期待されている (de Groote & Pogodalla 2004)。実際、De Groote & Pogodalla (2004) による先行研究は、弱文脈依存文法の代表的ないくつかのものが、抽象的範疇文法の特殊な形として記述可能であることを示している。彼らの結果は、単に抽象的範疇文法の言語クラスがこれらの文法よりも広いことのみを意味しているのではない。弱文脈依存文法から抽象的範疇文法への変換は直接的であり、前者における言語要素の導出規則および導出過程は、後者においても維持され直接的な対応関係をもっている。De Groote は、ラムダ項の複雑さに関する指標を用いた抽象的範疇文法の階層を提案しているが、この階層と、これらの弱文脈依存文法形式の間の関係も明らかにされてきている。

本学位論文では、抽象的範疇文法に対して、いくつかの自然な拡張と制限を導入した場合を中心に、その数理的性質を考察する。特に、抽象的範疇文法が従来の文法形式の一般化であるといったときに、いかなる意味で一般化と見做せるのか、追究している。

<本論>

第一に、抽象的範疇文法の基本的一般的な数理的性質について考察する。特に、抽象的範疇文法の一般所属性判定問題および非空集合生成問題が決定可能なら、未解決問題である線形論理の乗法的指数的断片の決定性を含意すること、semilinear でない言語を生成する抽象的範疇文法が存在することを示す。これらの結果は抽象的範疇文法が自然言語のモデルとしては複雑かつ強力すぎることを示唆しており、抽象的範疇文法への適切な制限を追究することが必要であると言える。なお、これらの結果は従来の線形論理に関する研究から比較的容易に導かれるものであり、抽象的範疇文法と論理学との関係の深さを反映している。

第二に、非線形なラムダ項を許す抽象的範疇文法の拡張について考察する。抽象的範疇文法は単純型付き線形ラムダ項に基づくが、非線形なラムダ項の導入によって、自然言語の現象の記述がより自然になったり簡潔になったりする場合があります。このような拡張について考察する意義がある。また一方で、いくつかの弱文脈依存文法に関する先行研究は、導出規則中に表れる引数の消去を許す場合と許さない場合とで、言語の生成能力に差が生じないことを示していた。すなわち、Multiple Context-Free Grammar と Linear Context-Free Rewriting System の等価性や、引数複製規則を持たない monadic な Context-Free Tree Grammar から引数消去規則が除去可能なことが知られていた。線形ラムダ項は、BCK ラムダ項において引数消去を許さないように制限したものであり、線形ラムダ項に基づく標準の抽象的範疇文法に対して、BCK ラムダ項に基づく BCK 抽象的範疇文法も自然に定義できる。本稿

では、BCK 抽象的範疇文法から非線形なラムダ項を取り除き、等価な標準の抽象的範疇文法を得る方法を提示する。しかもこの変換手法は、Multiple Context-Free Grammar から Linear Context-Free Rewriting System への変換と、引数複製規則の存在しない Context-Free Tree Grammar から引数消去規則を除去する手法を系として導出する。従って、抽象的範疇文法は、個々の弱文脈依存文法を模倣し一般化するというのみならず、これらの文法における変形をも一般化しているといえる。

第三に、抽象的範疇文法に「語彙化」という制限を加えた場合について考察する。一般に文法が語彙化されているとは、文法中の全ての語彙項目が、最終的な導出結果に明示的に現れる要素を必ず含んでいることをいう。言語現象は各単語に内在する情報によって説明される、とする語彙主義の考え方からは、文法が語彙化されていることは自然な要請である。本稿は、まず語彙化抽象的範疇文法の一般所属判定問題が NP 完全の複雑さを持つことを示す。この結果は非制限的な抽象的範疇文法に関する結果と対照的であり、語彙主義の要請からのみならず、計算の複雑さにおいても、抽象的範疇文法において語彙化が（充分ではないにせよ）望ましい制限であることを示している。一方で、Salvati は、語彙化の制限を緩めた「半語彙化」という制限を定義し、半語彙化抽象的範疇文法の定義する言語が NP に含まれることを示していた。このことは、半語彙化抽象的範疇文法の生成能力は、語彙化抽象的範疇文法に比べてさほど強力ではないことを意味する。本稿では、このような半語彙化抽象的範疇文法を等価な語彙化抽象的範疇文法に変換する語彙化の手続きを提示する。この語彙化手順は、ほぼ de Groote 階層について閉じている。一方で、文脈自由文法に対応する階層については閉じておらず、文脈自由文法を表現する抽象的範疇文法を語彙化すると、木接合文法 (Tree Adjoining Grammar) の文字列言語を生成する抽象的範疇文法の形状に変換される。このことは、有限に曖昧な文脈自由文法を木接合文法として語彙化した先行研究の結果に対応する。

第四に、2次元抽象的範疇文法の生成能力について議論する。Lambek 文法が文と意味の対の集合を言語とする「2次元文法」であるのに対し、抽象的範疇文法では、文も意味もラムダ項として扱い区別を要しないから、形式的には抽象的範疇文法は単にラムダ項の集合を言語とする1次元文法として定義されている。もとより抽象的範疇文法は多次元的拡張が容易なように設計されているものの、これまで2次元抽象的範疇文法の生成能力についての議論はほとんどなかった。まず、1次元抽象的範疇文法の性質のいくつかを2次元抽象的範疇文法でも成り立つことを示す。また、2次元抽象的範疇文法の生成能力を評価するため、言語間の関係を定義するいくつかの既存のフォーマリズムを2次元抽象的範疇文法で模倣する方法を示す。特に、de Groote による1次元抽象的範疇文法の階層分類を2次元抽象的範疇文法にも適用した場合、階層中の1つのクラスが、引数複製規則や引数消去規則を持たない

Macro Tree Transducer のクラスに完全に一致することがわかる。さらに、同期的木接合文法 (Synchronous Tree Adjoining Grammar) の生成する言語も 2次元抽象的範疇文法で生成可能であることがわかった。これらの結果は、1次元抽象的範疇文法が諸々の弱文脈依存文法を一般化したように、2次元抽象的範疇文法もまた、従来のよく知られたフォーマリズムの一般化として捉えられることの例証である。また、Salvati は、文字列を生成する抽象的範疇文法について、階層がある部分でつぶれることを示したが、2次元抽象的範疇文法においても同様の結果が成り立つことを、決定性 Tree Walking Transducer の2次元抽象的範疇文法での模倣を通じて証明する。

第五に、抽象的範疇文法における構文解析と関係の深い、線形ラムダ計算における高階単一化問題の特殊な場合である Interpolation について議論する。この問題が NP に属することは知られていたが、NP 困難性については証明されていなかった。そこでまず、線形ラムダ計算における Interpolation が NP 困難であることを証明する。ところで、線形ラムダ項では変数が複数回出現することは許されないが、定数項の出現回数には制限がないことから、線形ラムダ項中の定数項の複数回出現が NP 困難性に本質的な意味を持つのかという疑問が起こる。本稿は、定数項の異同を束縛変数の型割り当ての異同に置き換えることにより、定数項を排除した線形ラムダ計算における Interpolation が NP 困難であることを証明する。

<結論>

以上のように、本稿は、抽象的範疇文法に対する自然な拡張と制限を導入し、その数理的性質を明らかにしている。特に、既存のよく知られた文法形式で知られていたいいくつかの結果について、同様の定理が抽象的範疇文法でも成り立つことが示された。すなわち文法中の引数消去規則の除去が可能なこと、ある条件下で文法の語彙化が可能なことである。これらの結果は、既存の個々の文法が抽象的範疇文法で表現可能なだけでなく、その文法に対する変形手続もまた抽象的範疇文法の一般化された枠組みで記述可能なことを示している。また、抽象的範疇文法を2次元に拡張した場合の生成能力を議論し、1次元の場合と同様に、2次元抽象的範疇文法が従来のよく知られたフォーマリズムを記述可能なことを示した。このように、多様な意味において抽象的範疇文法が既存のフォーマリズムの一般化になっていることを明らかにした。

なお、抽象的範疇文法は線形ラムダ計算に基づいており、本稿の多くの結果も、陰に陽に線形論理やラムダ計算に関する定理を援用して得られている。また、線形ラムダ計算における Interpolation の問題のように、抽象的範疇文法の研究が契機となって、論理学や理論計算機科学に新たな問題を提起することがある。本稿の議論は、抽象的範疇文法の、数理言語学と論理学や理論計算機科学との橋渡しとしての性格を反映していると言える。

主要参考文献

- Philippe de Groote. 2001. Towards abstract categorial grammars. In *Association for Computational Linguistics, 39th Annual Meeting and 10th Conference of the European Chapter, Proceedings of the Conference*, pages 148–155.
- Philippe de Groote and Sylvain Pogodalla. 2004. On the expressive power of abstract categorial grammars: Representing context-free formalisms. *Journal of Logic, Language and Information*, **13**(4):421–438, 2004.
- Sylvain Salvati. 2005. *Problèmes de Filtrage et Problèmes d’analyse pour les Grammaires Catégorielles Abstraites*. PhD thesis, L’Institut National Polytechnique de Lorraine.