

Title	組合せ範疇文法による漸進的構文解析
Author(s)	谷口, 雅弥
Citation	
Issue Date	2023-03
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/18423
Rights	
Description	Supervisor: 東条 敏, 先端科学技術研究科, 博士

氏名	谷口 雅弥
学位の種類	博士 (情報科学)
学位記番号	博情第 497 号
学位授与年月日	令和 5 年 3 月 24 日
論文題目	Incremental Parsing in Combinatory Categorical Grammar
論文審査委員	東条 敏 北陸先端科学技術大学院大学 教授 Nguyen Minh Le 同 教授 廣川 直 同 准教授 井之上 直也 同 准教授 乾 健太郎 東北大学大学院情報科学研究科 教授 峯島 宏次 慶應義塾大学文学部人文社会学科 准教授

論文の内容の要旨

Combinatory Categorical Grammar (CCG) includes combinators in addition to Categorical Grammar (CG) to accommodate linguistic phenomena. For example, the type-raising rule is realized by a combinator in CCG to exchange the argument-functor relation, and such a rule is generalized as the Continuation-Passing Style (CPS) transformation. However, there is a concern that CPS may excessively accept ungrammatical sentences. In this thesis, we investigate the expanded grammar rules of CCG in terms of Lambek Calculus (LC). First, we show that Barker's CPS transformation is provable in LC, but Plotkin's CPS transformation is not. Second, we show a provable subset of Plotkin's CPS transformations. Due to the complexity of proving unprovability, we formalize the proof in Isabelle/HOL and verify it. We show that this subset is a grammatical class represented in LC and call it type-restricted CPS transformation.

The grammar defined in the formal system, the notion of normative grammar, is helpful for tagging sentences of a large corpus, i.e., to annotate each word by a part of speech (POS). In this research, we aim to obtain categorial grammar rules, where the category is a generalized notion of POS. However, finding a proper set of grammar rules is computationally exponential regarding the length of the sentence, and thus, a reliable but exhaustive search method is in demand. Here, we present a support system for annotation by the CCG parser based on the bidirectionality and non-determinism in logic programming. Contrary to the common usage of the parser, we extract a set of grammar rules from a syntax tree and retrieve all the probable readings.

In grammar extraction, the problem is keeping a head-dependency relation in a treebank, which is a collection of trees representing sentence constituency and dependency relation. We are motivated to extract grammar rules from the treebank, i.e., to decompose the tree data structure and to find grammar rules. After the extraction, we need to validate the adequacy of the grammar so that we inspect the generative power of the obtained grammar. In this phase, the syntactic head is a significant feature; however, the head information is missing in the obtained grammar. Hence, we propose supplementing the lost head information with the type-raising rule of the categorial grammar (CG). We extend the same issue to Combinatory Categorical Grammar (CCG) and solve it using generalized type-raising. Furthermore, we verify our grammar by the formal proof written in the proof assistant system, Isabelle/HOL.

Finally, we show the application to parse a sentence using our grammar system. Although a simple

grammar system is well-inspected in computational linguistics, the extension of CG is not so concerning the computational order problem and the proof-theoretic properties. In this thesis, we reveal the computational complexity of CG, its variants, and the form of the parsing tree as a proof tree in substructural logic, especially Lambek Calculus. Our contribution is three-fold: (a) we showed the new grammar rule named type-restricted CPS transformation; (b) we gave the new grammar extraction system keeping the head-dependency relation; (c) we showed constructive (partial) proof of the left-branching derivation in categorial grammar and Lambek Calculus.

Keywords: Combinatory Categorial Grammar, Lambek Calculus, Continuation-Passing Style Transformation, Incremental Parsing, Theorem Proving

論文審査の結果の要旨

カテゴリー文法(Category Grammar)は、文の生成過程・解析過程をランベック計算(Lambek Calculus)のシーケント計算による証明過程とする、論理的・計算科学的に厳密な体系である。同文法はチョムスキーの文脈自由文法と同様の文生成能力を有し、自然言語の文法を記述する際にも有用である。ところが、これに **type-raising** のコンビネータを導入し、カテゴリーの型の変換を行うようにすると、より広汎な文生成を可能にする代わりに、計算の複雑さが増大し、かつ自然言語の枠に収まらない過剰な文生成を可能にしてしまう。したがって、**type-raising** の可不可を適切に制限し、自然言語として認められる範囲の生成能力を持つ文法を形成できるようにする必要がある。この過程で制約条件として用いられるのが **Continuation Passing Style (CPS)**変換である。もともと **CPS** 変換はプログラミング言語にある概念であり、コードを左から右へ順に読んでいく過程が計算の進行と一致しない場合、値が返ってくるタイミングに合わせて計算を遅延実行する概念である。自然言語にも同様に構文木の生成が文の左から右へと書かれた文字面に沿わない場合があり、これに適切な **CPS** 変換を導入することにより、より自然言語の枠内に近い言語生成を試みたのが本論文である。まず、文を左端から読み込み、動的に構文木を生成していく過程が逐次的パーサ(**Incremental Parser**)である。これを可能にするためには、構文木がいつでも左分岐(**left-branching**)の形に変形できる必要がある。逐次的パーサは他人の発話を聞きながら同時並行して意味を汲み取ろうとする(木構造を組み立てる)人間の認知過程をそのまま具現するものであり、将来の **AI** におけるロボットとの柔軟な対話などに必須となる技術である。本論文では、任意の構文木が左分岐の形に変換できることをコンビネータの書き換えによって証明した。この中で、先行研究の **Plotkin** と **Barker** の提唱する **CPS** 変換のちょうど中間に位置する変換を提案し、いつでもその生成過程が証明可能であるクラスを提案した。特に、**Plotkin** の **CPS** 変換には証明可能でないものが存在することを定理証明器 **Isabelle/HOL** を使って網羅的に検証したことは、理論研究が計算機検証と結びついた顕著な業績である。本研究には、付随して文法獲得の成果もある。ここでは、構文木を割り振られたコーパスから **prolog** の両方向の変数充足性を用いて、木と文(文字列)を充足する規則を文法として獲得した。この際、二分岐の構文規則にヘッドの概念を選択する方法論も提示された。最後に、**type-raising** を含む **CG** が決定可能である条件を明示し、文生成が計算論的にも妥当な過程であることを示した。以上を要するに、本論文は博士学位論文として十分にふさわしいであると判断する。