Abstract

Combinatory Categorial Grammar (CCG) includes combinators in addition to Categorial 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 Categorial 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