Combinatory Categorial Grammars as Acceptors of Weighted Tree Languages

Lena Katharina Schiffer*
Institute of Computer Science, Universität Leipzig, Germany.
E-mail: schiffer@informatik.uni-leipzig.de

Combinatory Categorial Grammar (CCG) is an extension of categorial grammar that is well-established in computational linguistics. It is mildly context-sensitive, so it is efficiently parsable and reaches an expressiveness that is conjectured to be sufficient to describe natural languages. The basis for CCG is provided by a lexicon and a rule system. The lexicon assigns syntactic categories to the symbols of the input and the rule system describes how adjoining categories can be combined to eventually obtain a (binary) derivation tree. The seminal work by Vijay-Shanker and Weir showed weak equivalence of CCG, Tree-Adjoining Grammar (TAG), and Linear Indexed Grammar (LIG), i.e. they accept the same string languages.[1] However, in a linguistic setting we are also interested in the underlying structure of an input sentence in the form of a constituency parse tree. In our previous work we related three different variants of CCG to classes of tree languages.[2] Here we extend our results to a weighted scenario using arbitrary commutative semirings as a weight structure. The tree language accepted by a CCG is defined as a relabeling of its derivation trees. In the weighted case, each rule is associated with a weight. To obtain the weight of a given tree, we sum over the weights of all derivation trees that are relabeled to that tree, where the weight of a single derivation tree is the product of the weights of the applied rules. Our main result is that the weighted tree languages accepted by weighted CCGs are included in the weighted simple monadic context-free tree languages (wsCFTG). When restricting the rule system to application rules and composition rules of first degree, the accepted weighted tree languages are exactly the weighted regular tree languages. When only application rules are allowed, a proper subset of the weighted regular tree languages is accepted.

*Supported by the German Research Foundation (DFG) Research Training Group GRK 1763 'Quantitative Logics and Automata'

[1] K. Vijay-Shanker and D. J. Weir, "The Equivalence of Four Extensions of Context-Free Grammars," in *Mathematical Systems Theory*, vol. 27, no. 6, pp. 511–546. Springer, 1994.

[2] M. Kuhlmann, A. Maletti and L. K. Schiffer, "The Tree-Generative Capacity of Combinatory Categorial Grammars," in *Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2019)*, vol. 150 of *LIPIcs*, pp. 44:1–44:14. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2019.