

Chomsky-Schützenberger Characterisation for Multiple Context-Free Languages

Tobias Denking

The classical Chomsky-Schützenberger result [CS63, Proposition 2] states that every context-free language can be expressed as the homomorphic image of the intersection of a regular language and a Dyck language. A similar result has recently been used to derive an efficient parsing algorithm for context-free grammars [Hul11].

Multiple context-free languages are a proper superclass of context-free languages. They are researched in the natural language processing community [Kal10; Kuh13] because their parsing complexity is polynomial and they are capable of expressing cross-serial dependencies.

We obtain a characterisation similar to the classical Chomsky-Schützenberger result, i.e. that a multiple context-free language can be expressed as the homomorphic image of the intersection of a regular language and the inverse homomorphic image of (what we call) a multiple Dyck language.

Similar to Hulden [Hul11, Section 4.1] the result is extended to weighted languages by assigning the weights of the original multiple context-free grammar rules to specific rules in the constructed regular grammar.

References

- [CS63] N. Chomsky and M.-P. Schützenberger. “The algebraic theory of context-free languages”. In: *Computer Programming and Formal Systems* (1963), pp. 118–161.
- [Hul11] M. Hulden. “Parsing CFGs and PCFGs with a Chomsky-Schützenberger Representation”. In: *Human Language Technology. Challenges for Computer Science and Linguistics*. Ed. by Z. Vetulani. Vol. 6562. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2011, pp. 151–160. ISBN: 978-3-642-20094-6.
- [Kal10] L. Kallmeyer. *Parsing beyond context-free grammars*. Springer, 2010.
- [Kuh13] M. Kuhlmann. “Mildly Non-projective Dependency Grammar”. In: *Computational Linguistics* 39 (2013).