

# Context-free grammars and finite-state automata over categories

Noam Zeilberger

November 2, 2024

I propose to talk about joint work with Paul-André Melliès developing a categorical perspective on context-free grammars and finite-state automata [MZ24], which grew out of our line of work on type refinement systems. The starting point of that earlier work was to consider type systems fibrationally as functors from a category of typing derivations to a category of terms [MZ15], so that typing is reduced to a “lifting problem” along the functor. In the talk I will explain how both parsing by a context-free grammar (CFG) and recognition by a nondeterministic finite-state automaton (NFA) may be similarly expressed as lifting problems along certain kinds of functors, and argue that viewing CFGs and NFAs this way seems to have some real explanatory power.

A generalized CFG is a functor from a free colored operad (aka multicategory) generated by a pointed finite species into an arbitrary base operad: this encompasses classical CFGs by taking the base to be a certain operad constructed from a free monoid, as an instance of a more general construction of an *operad of spliced arrows*  $\mathcal{WC}$  for any category  $\mathcal{C}$ . A generalized NFA is a functor from an arbitrary bipointed category or pointed operad satisfying the unique lifting of factorizations and finite fiber properties: this encompasses classical word automata and tree automata without  $\epsilon$ -transitions, but also automata over non-free categories and operads. We show that generalized context-free and regular languages satisfy suitable generalizations of many of the usual closure properties, and in particular we give a simple conceptual proof that context-free languages are closed under intersection with regular languages.

## References

- [MZ15] Paul-André Melliès and Noam Zeilberger. Functors are type refinement systems. In *Proceedings of the 42nd Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages*, pages 3–16. ACM, 2015. doi:10.1145/2676726.2676970.
- [MZ24] Paul-André Melliès and Noam Zeilberger. The categorical contours of the Chomsky-Schützenberger representation theorem. *Logical Methods in Computer Science*, 2024. To appear. URL: <https://arxiv.org/abs/2405.14703>.