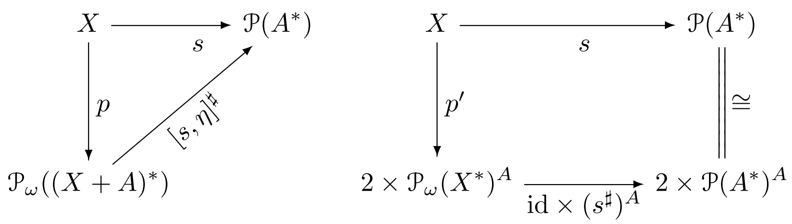
On Language Equations and Grammar Coalgebras for Context-free Languages[*]

Jurriaan Rot[1,2] (jrot@liacs.nl) and Joost Winter[2,1] (winter@cwi.nl)

[1] LIACS – Leiden University
[2] Formal Methods – Centrum Wiskunde & Informatica (CWI)

Introduction. In [3], a coalgebraic presentation of context-free grammars and languages was given based on behavioural differential equations, yielding final coalgebra semantics via the functor $2 \times (-)^A$ of deterministic automata. There, the correctness was given via the well-known notion of derivations in a grammar. Another classical approach to the semantics of context-free grammars (which generalizes to algebraic power series in noncommuting variables) is through systems of (language) equations (see, e.g., [2]). In this note, we explore the connection between systems of behavioural differential equations and these classical systems, and give an elementary proof of their correspondence.

Solutions by initiality. A map s is a *solution* to a grammar p (over an alphabet A and a finite set X of nonterminals) if it makes the left diagram below commute:



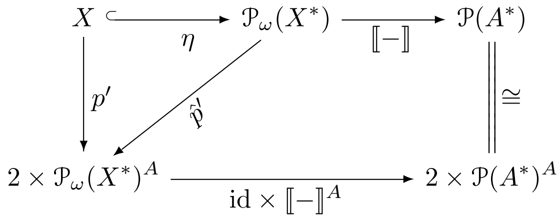
Here $[s, \eta]$ is the co-tupling of the solution mapping s and the injection η of alphabet symbols into languages, and $\mathcal{P}_\omega(-)$ is the finite powerset functor. The $^\sharp$-symbol denotes the inductive extension based on language union and concatenation. Classically, the (denotational) *semantics* of a grammar p is the (pointwise) least solution. Whenever p is in Greibach normal form, it has a unique solution s, and there is an isomorphism

$$\mathcal{P}_\omega((X + A)^*)_{\mathrm{GNF}} \;\cong\; 2 \times \mathcal{P}_\omega(X^*)^A$$

transforming p into an equivalent mapping p'. Moreover, the final coalgebra structure on $\mathcal{P}_\omega(A^*)$ also is an isomorphism. We thus obtain the diagram to the right of the previous diagram, and observe that s makes the left diagram commute iff it makes the right diagram commute.

[*] Supported by the NWO project CoRE: Coinductive Calculi for Regular Expressions.

Solutions by finality. We recall from [3] that context-free grammars can be represented using systems of behavioural differential equations: given a grammar p, the corresponding system is (by definition) identical to p' as defined above. Now p' can be extended inductively to a mapping $\hat{p}'$ by defining union and concatenation using behavoural differential equations. As $\hat{p}'$ defines a $2 \times (-)^A$-coalgebra on $\mathcal{P}_\omega(X^*)$, we now obtain a unique mapping $[\![-]\!]$ into the final coalgebra, as in the following diagram:



It is easy to prove that the mapping $[\![-]\!]$ is an algebra morphism for the monad $\mathcal{P}_\omega(-^*)$: this can be done either directly by induction, or by making use of a distributive law [1]. In both cases the crux is that the behavioural differential equations 'correctly' define union and concatenation.

Equivalence. We will now show that, given a system $p' : X \to 2 \times \mathcal{P}_\omega(X^*)^A$, the (extended) solutions $s^\sharp$ and $[\![-]\!]$, both uniquely defined, necessarily coincide. First, we obtain the following lemma as a direct consequence of the fact that $[\![-]\!]$ is an $\mathcal{P}_\omega(-^*)$-algebra morphism:

Lemma. $([\![-]\!] \circ \eta)^\sharp = [\![-]\!]$.

Hence the mapping $[\![-]\!] \circ \eta$, when substituted for s (with $[\![-]\!]$ substituting $s^\sharp$ by the lemma), is a solution to the first diagram. Since there is a unique such solution, we obtain our main result:

Theorem. *A mapping $s : X \to \mathcal{P}(A^*)$ is a solution to a (classical) system of equations iff $s = [\![-]\!] \circ \eta$.*

Generalizations. As a final remark, we note that the preceding lemma and theorem can directly be generalized from languages to formal power series over any commutative semiring K: $[\![-]\!]$ now is an algebra morphism for the monad $K\langle - \rangle$, the semiring of (noncommuting) polynomials, generalizing the monad $\mathcal{P}_\omega(-^*)$; and the final coalgebra now is $K\langle\langle A \rangle\rangle$, the semiring of formal power series over noncommuting variables in A, generalizing $\mathcal{P}(A^*)$.

References

1. Bonsangue, M., Hansen, H.H., Kurz, A., Rot, J.: Presenting distributive laws. In: CALCO (2013), *to appear.*
2. Petre, I., Salomaa, A.: Algebraic systems and pushdown automata. In: *Handbook of Weighted Automata*, pp. 257–289. Springer (2009).
3. Winter, J., Bonsangue, M.M., Rutten, J.J.M.M.: Context-free languages, coalgebraically. In: CALCO. LNCS 6859, pp. 359–376. Springer (2011).