

Towards a Coalgebraic Chomsky Hierarchy

Sergey Goncharov, Stefan Milius, Alexandra Silva

CMCS 2014, Grenoble, 6.04.2014



FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG

TECHNISCHE FAKULTÄT

Short History of Coalgebraic Invasion to Automata Theory

- Deterministic automata as coalgebras [**Rutten, 1998**].
- Generalized regular expressions and Kleene's theorem for (Kripke-)polynomial functors [**Silva, 2010**].
- Generalized powerset construction [**Silva et al., 2010**].
- Regular expressions for equationally presented functors and monads [**Myers, 2013**].
- Context-free languages, coalgebraically [**Winter et al., 2013**].

Base Case: Moore Automata

Moore automaton with input alphabet A and output alphabet B is given by

$$t^m : X \times A \rightarrow X \quad (\text{transition}) \quad \text{and} \quad o^m : X \rightarrow B \quad (\text{output})$$

Thus, a Moore automaton is a coalgebra $m : X \rightarrow B \times X^A$ on **Set**.

We shall assume finite

Base Case: Moore Automata

Moore automaton with input alphabet A and output alphabet B is given by

$$t^m : X \times A \rightarrow X \quad (\text{transition}) \quad \text{and} \quad o^m : X \rightarrow B \quad (\text{output})$$

Thus, a Moore automaton is a coalgebra $m : X \rightarrow B \times X^A$ on **Set**.

We shall assume finite

Base Case: Moore Automata

Moore automaton with input alphabet A and output alphabet B is given by

$$t^m : X \times A \rightarrow X \quad (\text{transition}) \quad \text{and} \quad o^m : X \rightarrow B \quad (\text{output})$$

Thus, a Moore automaton is a coalgebra $m : X \rightarrow B \times X^A$ on **Set**.

A final $L_{A,B}$ -coalgebra is carried by the set B^{A^*} of formal power series on B with coalgebra structure $\langle o, t \rangle : B^{A^*} \rightarrow B \times (B^{A^*})^A$

$$o(\sigma : A^* \rightarrow B) = \sigma(\varepsilon) \quad \text{and} \quad t(\sigma : A^* \rightarrow B, a) = \lambda w. \sigma(a \cdot w).$$

$$L_{A,B}X := B \times X^A$$

We shall assume finite

Base Case: Moore Automata

Moore automaton with input alphabet A and output alphabet B is given by

$$t^m : X \times A \rightarrow X \quad (\text{transition}) \quad \text{and} \quad o^m : X \rightarrow B \quad (\text{output})$$

Thus, a Moore automaton is a coalgebra $m : X \rightarrow B \times X^A$ on **Set**.

A final $L_{A,B}$ -coalgebra is carried by the set B^{A^*} of formal power series on B with coalgebra structure $\langle o, t \rangle : B^{A^*} \rightarrow B \times (B^{A^*})^A$

$$o(\sigma : A^* \rightarrow B) = \sigma(\varepsilon) \quad \text{and} \quad t(\sigma : A^* \rightarrow B, a) = \lambda w. \sigma(a \cdot w).$$

$$L_{A,B}X := B \times X^A$$

We shall assume finite

Base Case: Moore Automata

Moore automaton with input alphabet A and output alphabet B is given by

$$t^m : X \times A \rightarrow X \quad (\text{transition}) \quad \text{and} \quad o^m : X \rightarrow B \quad (\text{output})$$

Thus, a Moore automaton is a coalgebra $m : X \rightarrow B \times X^A$ on **Set**.

A final $L_{A,B}$ -coalgebra is carried by the set B^{A^*} of formal power series on B with coalgebra structure $\langle o, t \rangle : B^{A^*} \rightarrow B \times (B^{A^*})^A$

$$o(\sigma : A^* \rightarrow B) = \sigma(\varepsilon) \quad \text{and} \quad t(\sigma : A^* \rightarrow B, a) = \lambda w. \sigma(a \cdot w).$$

Derivatives: given $w \in A^*$,

$$\partial_\varepsilon(\sigma) = \sigma \quad \text{and} \quad \partial_{a \cdot w}(\sigma) = t(\partial_w(\sigma), a)$$

If $B = 2$ then $B^{A^*} \simeq \mathcal{P}(A^*)$ is the set of all formal languages over A .

Rationality of Formal Power Series

By the the universal property of the final coalgebra for any $m : X \rightarrow B \times X^A$ there exists a unique $L_{A,B}$ -coalgebra homomorphism \widehat{m} such that:

$$\begin{array}{ccc}
 X & \xrightarrow{\widehat{m}} & B^{A^*} \\
 m \downarrow & & \downarrow \iota \\
 B \times X^A & \xrightarrow{\text{id} \times \widehat{m}^A} & B \times (B^{A^*})^A
 \end{array}$$

Given $x \in X$, $\llbracket x \rrbracket_m := \widehat{m}(x)$ is the “language” recognized by m at x .

A formal power series $\sigma : A^* \rightarrow B$ is **rational** if $\{\partial_w(\sigma) \mid w \in A^*\}$ is finite.

Theorem. A formal power series σ is rational iff $\sigma = \llbracket x \rrbracket_{\widehat{m}}$ for some $m : X \rightarrow B \times X^A$ and $x \in X$.

Generalized Powerset Construction and \mathbb{T} -automata

Definition: Let \mathbb{T} be a monad. A **\mathbb{T} -automaton** is a triple of maps

$$o^m : X \rightarrow B, \quad t^m : X \times A \rightarrow TX, \quad a^m : TB \rightarrow B$$

where a^m is a \mathbb{T} -algebra.

Essentially, a \mathbb{T} -automaton is a coalgebra $m : X \rightarrow B \times (TX)^A$.

$$\begin{array}{ccccc}
 X & \xrightarrow{\eta} & TX & \xrightarrow{\widehat{m}^\sharp} & B^{A^*} \\
 \downarrow m & & \swarrow m^\sharp & & \downarrow \iota \\
 B \times (TX)^A & \xrightarrow{\text{id} \times (\widehat{m}^\sharp)^A} & & & B \times (B^{A^*})^A
 \end{array}$$

Factorization $m = m^\sharp \eta$ is unique and we put $\llbracket x \rrbracket_m = \llbracket \eta(x) \rrbracket_{m^\sharp}$.

Theorem: If B is finite then $\llbracket x \rrbracket_m$ is rational.

Generalized Powerset Construction and \mathbb{T} -automata

Definition: Let \mathbb{T} be a monad. A **\mathbb{T} -automaton** is a triple of maps

$$o^m : X \rightarrow B, \quad t^m : X \times A \rightarrow TX, \quad a^m : TB \rightarrow B$$

where a^m is a \mathbb{T} -algebra.

Essentially, a \mathbb{T} -automaton is a coalgebra $m : X \rightarrow B \times (TX)^A$.

$$\begin{array}{ccccc}
 X & \xrightarrow{\eta} & \mathcal{P}X & \xrightarrow{\widehat{m}^\sharp} & \mathcal{P}A^* \\
 \downarrow m & & \swarrow m^\sharp & & \downarrow \iota \\
 2 \times (\mathcal{P}X)^A & \xrightarrow{\text{id} \times (\widehat{m}^\sharp)^A} & & & 2 \times (\mathcal{P}A^*)^A
 \end{array}$$

Factorization $m = m^\sharp \eta$ is unique and we put $\llbracket x \rrbracket_m = \llbracket \eta(x) \rrbracket_{m^\sharp}$.

Theorem: If B is finite then $\llbracket x \rrbracket_m$ is rational.

Monads as Theories

An **algebraic theory** is given by a signature Σ and a set of equations.

Any algebraic theory \mathcal{E} defines a monad:

- $T_{\mathcal{E}}X$ = ‘set of Σ -terms over X modulo \mathcal{E} ’;
- η coerces a variable to a term;
- $\sigma^*(t)$ applies substitution $\sigma : X \rightarrow T_{\mathcal{E}}Y$ to $p : T_{\mathcal{E}}X$.

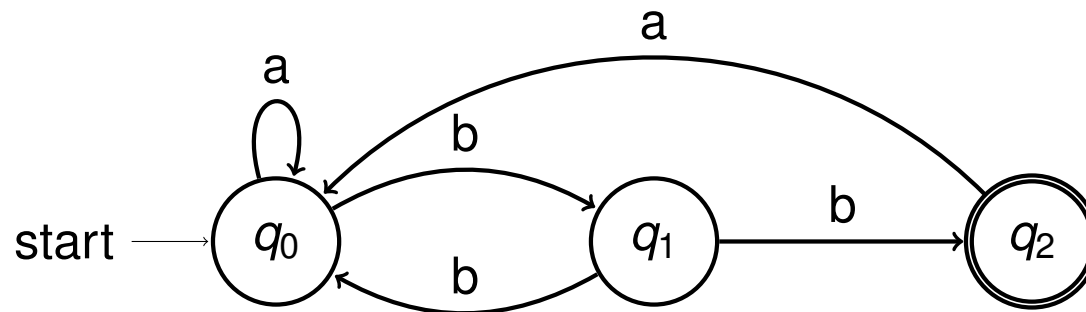
Example: Finite powerset monad $\mathcal{P}_{\omega} \iff$ join semilattices with bottom.

Example: Finite store monad $TX = (X \times S)^S \iff$ finite **mnemoids**
 ($lookup_l : X^V \rightarrow X$, $update_{l,v} : X \rightarrow X$ where $S \cong L \rightarrow V$).

Expressions for \mathbb{T} -automata (by Example)

$$B = \{\top, \perp\}$$

$$\mathbb{T} = \mathcal{P}_\omega$$



$$\begin{cases}
 e_0 = a.e_0 \dot{\cup} b.e_1 \dot{\cup} \perp \\
 e_1 = a.\emptyset \dot{\cup} b.(e_0 + e_2) \dot{\cup} \perp \\
 e_2 = a.e_0 \dot{\cup} b.\emptyset \dot{\cup} \top
 \end{cases}$$

Equivalently,

$$e_0 = \mu x. (a.x \dot{\cup} b.\mu y. (a.\emptyset \dot{\cup} b.(x + \mu z. (a.x \dot{\cup} b.\emptyset \dot{\cup} \top))) \dot{\cup} \perp) \dot{\cup} \perp).$$

Stack \mathbb{T} -automata

Stack monad is a submonad of the store monad $\Gamma^* \rightarrow (X \times \Gamma^*)$:

$\langle r, t \rangle : \Gamma^* \rightarrow (X \times \Gamma^*)$ is in TX iff

$$r(u \cdot w) = r(u) \qquad t(u \cdot w) = t(u) \cdot w$$

whenever $|u| > k$ for some k .

Stack theory is given by **pop** : $X^{n+1} \rightarrow X$ and **push** _{i} : $X \rightarrow X$ ($i \leq n$):

$$\mathbf{push}_i(\mathbf{pop}(x_1, \dots, x_n, y)) = x_i$$

$$\mathbf{pop}(\mathbf{push}_1(x), \dots, \mathbf{push}_n(x), x) = x$$

$$\mathbf{pop}(x_1, \dots, x_n, \mathbf{pop}(y_1, \dots, y_n, z)) = \mathbf{pop}(x_1, \dots, x_n, z)$$

Stack \mathbb{T} -automaton is a \mathbb{T} -automaton $m : X \rightarrow \mathcal{B}(\Gamma) \times (TX)^A$ where $\mathcal{B}(\Gamma)$ are predicates over Γ^* such that $p \in \mathcal{B}(\Gamma)$ iff $p(u \cdot w) \iff p(u)$ once $|u| > k$ with some k .

Stack \mathbb{T} -automata

$\Gamma = \{\gamma_1, \dots, \gamma_n\}$ is stack alphabet

Stack monad is a submonad of the store monad $\Gamma^* \rightarrow (X \times \Gamma^*)$:

$\langle r, t \rangle : \Gamma^* \rightarrow (X \times \Gamma^*)$ is in TX iff

$$r(u \cdot w) = r(u)$$

$$t(u \cdot w) = t(u) \cdot w$$

whenever $|u| > k$ for some k .

Stack theory is given by **pop** : $X^{n+1} \rightarrow X$ and **push_i** : $X \rightarrow X$ ($i \leq n$):

$$\mathbf{push}_i(\mathbf{pop}(x_1, \dots, x_n, y)) = x_i$$

$$\mathbf{pop}(\mathbf{push}_1(x), \dots, \mathbf{push}_n(x), x) = x$$

$$\mathbf{pop}(x_1, \dots, x_n, \mathbf{pop}(y_1, \dots, y_n, z)) = \mathbf{pop}(x_1, \dots, x_n, z)$$

Stack \mathbb{T} -automaton is a \mathbb{T} -automaton $m : X \rightarrow \mathcal{B}(\Gamma) \times (TX)^A$ where $\mathcal{B}(\Gamma)$ are predicates over Γ^* such that $p \in \mathcal{B}(\Gamma)$ iff $p(u \cdot w) \iff p(u)$ once $|u| > k$ with some k .

\mathbb{T} -automata and Context-free Languages

Theorem. If m is a stack \mathbb{T} -automaton then for any $x \in X$ and any $s \in \Gamma^*$,

$$\{w \in A^* \mid \llbracket x \rrbracket_m(w)(s)\}$$

is a **deterministic real-time CFL**; and conversely: any deterministic real-time CFL can be obtained in this way.

Example: Expression $\mu x. (a.\mathbf{push}(x) \uplus b.\mathbf{pop}(x, \top) \uplus \perp)$ corresponds to context-free grammar: $X \rightarrow aXX, X \rightarrow b$.

Further results:

- **Nondeterministic stack \mathbb{T} -automata** capture CFL;
- **Nondeterministic $(m > 2)$ -stack \mathbb{T} -automata** capture NTIME.

Tape Monad

Tape monad is a submonad of the store monad $\mathbb{Z} \times \Gamma^{\mathbb{Z}} \rightarrow (X \times \mathbb{Z} \times \Gamma^{\mathbb{Z}})$:
 $\langle r, z, t \rangle : \mathbb{Z} \times \Gamma^{\mathbb{Z}} \rightarrow (X \times \mathbb{Z} \times \Gamma^{\mathbb{Z}})$ is in TX iff there is k such that for any i, j ,
 σ, σ' if $\sigma =_{i \pm k} \sigma'$ then

$$\begin{aligned} t(i, \sigma) &=_{i \pm k} t(i, \sigma'), & t(i, \sigma_{+j}) &= t(i + j, \sigma)_{+j}, & t(i, \sigma) &=^{i \pm k} \sigma, \\ z(i, \sigma) &= z(i, \sigma'), & z(i, \sigma_{+j}) &= z(i + j, \sigma) - j, & |z(i, \sigma) - i| &\leq k, \\ r(i, \sigma) &= r(i, \sigma'), & r(i, \sigma_{+j}) &= r(i + j, \sigma). \end{aligned}$$

where $\sigma_{+j}(i) = \sigma(i + j)$; $\sigma =_{i \pm k} \sigma'$ iff $\sigma(j) = \sigma'(j)$ for all j such that $|i - j| \leq k$; $\sigma =^{i \pm k} \sigma'$ iff $\sigma(j) = \sigma'(j)$ for all j such that $|i - j| > k$.

Tape theory is given over signature of operations: **read** : $n \rightarrow 1$,
write_i : $n \rightarrow 1$ ($1 \leq i \leq n$), **Imove** : $1 \rightarrow 1$, **rmove** : $1 \rightarrow 1$.

Theorem: Tape theory is not finitely axiomatizable.

Tape \mathbb{T} -automata

Tape \mathbb{T} -automaton is a \mathbb{T} -automaton $m : X \rightarrow \mathcal{C}(\Gamma) \times (TX)^A$ where

- \mathbb{T} is the tape monad over Γ ;
- $\mathcal{C}(\Gamma)$ is the set of predicates over $\mathbb{Z} \times \Gamma^{\mathbb{Z}}$ such that $p \in \mathcal{C}(\Gamma)$ iff there is a k such that $p(i, \sigma) = p(i, \sigma')$ and $p(i, \sigma_{+j}) = p(i + j, \sigma)$ if $\sigma =_{i \pm k} \sigma'$.

Conjecture: Tape \mathbb{T} -automata capture precisely linear-time languages.

Theorem: Let $\tau \in A$ and let \mathcal{L} be the class of languages over A captured by \mathbb{T} -automata. Then $\{L \setminus \tau \mid \mathcal{L}\}$ are exactly all r.e. languages where $L \setminus \tau$ is the result of removing τ from L .

Conclusions

What did not fit this talk:

- Kleene theorem for \mathbb{T} -automata.
- Algebraic expressions (fixpoint expressions with sequential composition).
- Lower expressivity bounds for \mathbb{T} -automata as functions of \mathbb{T} .
- τ -elimination by CPS-transformations.

Further work:

- (Complete) equational calculi of fixpoint expressions.
- Chomsky-Schützenberger theorem for \mathbb{T} -automata.
- Identifying further language and complexity classes.

The top portion of the slide features a dark blue background with a silhouette of a classical building facade on the left, showing several statues on the roofline. On the right, a large, faint circular seal is visible, containing the word 'ACADEMIA' and a profile of a person's head.

Thank You for Your Attention!



**FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG**

TECHNISCHE FAKULTÄT

📄 Myers, R. (2013).

Rational Coalgebraic Machines in Varieties: Languages, Completeness and Automatic Proofs.

PhD thesis, Imperial College London.

📄 Rutten, J. J. M. M. (1998).

Automata and coinduction (an exercise in coalgebra).

In Sangiorgi, D. and de Simone, R., editors, *CONCUR*, volume 1466 of *Lecture Notes in Computer Science*, pages 194–218. Springer.

📄 Silva, A. (2010).

Kleene coalgebra.

PhD thesis, Radboud Univ. Nijmegen.

- 📄 Silva, A., Bonchi, F., Bonsangue, M. M., and Rutten, J. J. M. M. (2010).
Generalizing the powerset construction, coalgebraically.
In *FSTTCS*, volume 8 of *LIPICs*, pages 272–283.
- 📄 Winter, J., Bonsangue, M. M., and Rutten, J. J. M. M. (2013).
Coalgebraic characterizations of context-free languages.
LMCS, 9(3).