Local state classifier for algebraic language theory

Ryuya Hora

University of Tokyo

May 16, 2025

Slides

- *Topoi of automata I* [Hora 2024b]
- *Topoi of automata II* [Hora n.d.(b)]
- *Internal Parameterization of Hyperconnected Quotients* [Hora 2024a]

Table of Contents

Table of Contents

Motivating example: Continuous actions form a topos (1/3)

For any topological group (G, τ), it's well-known that

Fact

$\mathbf{Cont}(G, \tau)$ *is a Grothedieck topos.*

By definition, $\mathbf{Cont}(G)$ is a full subcategory of $\mathbf{PSh}(G)$. A right G-set X belongs to $\mathbf{Cont}(G)$ if and only if it satisfies **the open stabilizer condition**:

Open stabilizer condition

For any $x \in X$, its stabilizer subgroup $\mathrm{Stab}_X(x) \subset G$ is open.

Motivating example: Continuous actions form a topos (2/3)

Two topoi $\mathbf{PSh}(G)$ and $\mathbf{Cont}(G,\tau)$ are related by a **hyperconnected geometric morphism**.

Definition (Hyperconnected geometric morphism [Johnstone 1981])

A geometric morphism $f\colon \mathcal{E} \to \mathcal{F}$ is said to be **hyperconnected** if f^* is fully faithful, and its counit $\epsilon\colon f^*f_* \to \mathrm{id}_{\mathcal{E}}$ is monic.

Example (Topological groups)

For a topological group (G,τ), the inclusion $\mathbf{Cont}(G,\tau) \hookrightarrow \mathbf{PSh}(G)$ defines a hyperconnected geometric morphism $\mathbf{PSh}(G) \twoheadrightarrow \mathbf{Cont}(G,\tau)$.

Informally speaking, we will observe that **all hyperconnected geometric morphisms are given by a (generalized) open stabilizer condition!**

Motivating example: Continuous actions form a topos (3/3)

The open stabilizer condition is equivalent to **the lifting condition**

$$\begin{tikzcd} & F \arrow[d, tail] \\ X \arrow[r, "\mathrm{Stab}_X"'] \arrow[ur, dashed, "\exists"] & \Xi \end{tikzcd} \qquad \text{in } \mathbf{PSh}(G)$$

where

- Ξ denotes the set of all subgroups of G, and
- F denotes the set of open subgroups of (G, τ).

Motivating example: Continuous actions form a topos (3/3)

The open stabilizer condition is equivalent to **the lifting condition**

$$\begin{tikzcd} & F \arrow[d, tail] \\ X \arrow[ur, dashed, "\exists"] \arrow[r, "\mathrm{Stab}_X"'] & \Xi \end{tikzcd} \qquad \text{in } \mathbf{PSh}(G)$$

where

- Ξ denotes the set of all subgroups of G, and
- F denotes the set of open subgroups of (G, τ).

Question

But how can we define Ξ and Stab in general topoi?

Local state classifier

Definition (LSC: local state classifier [Hora 2024a])

A **local state classifier** Ξ of a category $\mathcal{E}$ is a colimit of all monomorphisms:

$$\Xi = \mathrm{colim}(\mathcal{E}_{\mathsf{mono}} \rightarrowtail \mathcal{E}).$$

The associated cocone is referred to as $\{\xi_X : X \to \Xi\}_{X \in \mathrm{ob}\mathcal{E}}$.

$$\begin{tikzcd} U \arrow[rr, tail, "\iota"] \arrow[dr, "\xi_U"'] & & X \arrow[dl, "\xi_X"] \\ & \Xi & \end{tikzcd}$$

Example (In the topos **PSh**(G))

- Ξ is the set of all subgroups of G, and
- each $\xi_X : X \to \Xi$ is the stabilizer morphism Stab_X.

Induced Full subcategory (1/2)

Definition ([Hora n.d.(a)])

For a category $\mathcal{E}$ with a LSC Ξ and a subobject $F \rightarrowtail \Xi$, we define $\mathcal{E}_F \hookrightarrow \mathcal{E}$ as the full subcategory that consists of objects X satisfying the lifting property

$$\begin{tikzcd} & F \arrow[d, tail] \\ X \arrow[ur, dashed, "\exists"] \arrow[r, "\xi_X"'] & \Xi. \end{tikzcd}$$

Induced Full subcategory (2/2)

Even if $\mathcal{E}$ is a topos, $\mathcal{E}_F$ is not a topos in general.

Example

Let G be a group and $\mathcal{E} = \mathbf{PSh}(G)$.

- If (G, τ) is a topological group and $F \rightarrowtail \Xi$ is the set of all open groups, then $\mathcal{E}_F = \mathbf{Cont}(G, \tau)$.
- If $F = \{\text{the trivial subgroup}\}$, then $\mathcal{E}_F$ is the full subcategory that consists of all free G-actions. This is far from being a topos. (cf. [Fiore, Galal, and Paquet 2024])

To ensure $\mathcal{E}_F$ is a topos, we need an **order-theoretic** criterion on $F \rightarrowtail \Xi$.

Internal semilattice structure

Proposition (Semilattice structure [Hora 2024a])

If a cartesian closed category (in particular, a topos) $\mathcal{E}$ has an LSC Ξ, then Ξ admits a unique internal $\wedge$-semilattice structure

- $\wedge \colon \Xi \times \Xi \to \Xi$
- $\top \colon 1 \to \Xi$

that satisfies $\xi_X \wedge \xi_Y = \xi_{X \times Y}$.

$$\begin{tikzcd} & X \times Y \arrow[dl, "\xi_X \times \xi_Y"'] \arrow[dr, "\xi_{X\times Y}"] & \\ \Xi \times \Xi \arrow[rr, "\wedge"'] & & \Xi \end{tikzcd}$$

Example (In the topos $\mathbf{PSh}(G)$,)

The $\wedge$-operator $\wedge \colon \Xi \times \Xi \to \Xi$ coincides with the intersection, since $\mathrm{Stab}_X(x) \cap \mathrm{Stab}_Y(y) = \mathrm{Stab}_{X \times Y}(x, y)$

Internal filters

Recall that a **filter** F of a $\wedge$-semilattice A is an upward closed subset $F \subset A$ that is closed under finite infimum $\top, \wedge$.

Definition

An **internal filter** of an internal semilattice Ξ is a subobject $F \rightarrowtail \Xi$ that induces a filter $\mathcal{E}(X, F) \rightarrowtail \mathcal{E}(X, \Xi)$ to each object X

In a presheaf topos $\mathbf{PSh}(\mathcal{C})$, an internal filter is just a subobject $F \rightarrowtail \Xi$ that is a filter $F(c) \rightarrowtail \Xi(c)$ for each object $c \in \mathcal{C}$.

Example (Topological group)

For any topological group (G, τ), open subgroups define an internal filter of Ξ.

Parameterization of hyperconnected geometric morphisms

Theorem (Main Theorem of [Hora 2024a])

Every Grothendieck topos $\mathcal{E}$ has a LSC Ξ, and the construction $F \mapsto \mathcal{E}_F$ provides a one-to-one correspondence between

- *internal filters $F \rightarrowtail \Xi$, and*
- *hyperconnected geometric morphisms from $\mathcal{E}$.*

The left adjoint $h^* \colon \mathcal{E}_F \to \mathcal{E}$ is just the embedding. The right adjoint $h_* \colon \mathcal{E} \to \mathcal{E}_F$ is given by pullback:

$$\begin{tikzcd} h_*X \arrow[r] \arrow[d, tail, "\epsilon_X"] \arrow[dr, phantom, "\lrcorner", very near start] & F \arrow[d, tail, "\iota_F"] \\ X \arrow[r, "\xi_X"'] & \Xi. \end{tikzcd}$$

Digression (1/3): Lawvere's open problems

This was motivated by one of Lawvere's open problems:

> ***Is there a Grothendieck topos for which the number of these quotients is not small?*** *[Lawvere 2009]*

Yes, if Σ is infinite, Σ-**Set** has! [Kamio and Hora 2024]

> ***At the other extreme, could they be parameterized internally, as subtoposes are?*** *[Lawvere 2009]*

For hyperconnected quotients, yes! [Hora 2024a]

Digression (2/3): Similarity with LT-topology

Corollary

For any Grothendieck topos $\mathcal{E}$, there is a one-to-one correspondence between

- *semilattice homomorphisms $j : \Xi \to \Omega$, and*
- *hyperconnected geometric morphisms from $\mathcal{E}$.*

Proposition (LSC of presheaf topos is "dual" to Ω)

For any small category $\mathcal{C}$, the local state classifier of $\mathbf{PSh}(\mathcal{C})$ is given by $\Xi(c) := \{\text{quotient objects of } y(c)\}$

Digression (3/3): classes of topoi

Corollary (Localic topoi [Hora 2024a])

*For a Grothendieck topos $\mathcal{E}$, its local state classifier is terminal $\Xi = 1$ if and only if $\mathcal{E}$ is **localic**.*

Proposition (Local topoi [Menni 2025])

*For a **local** topos $\mathcal{E}$, the LSC has exactly one global section $1 \to \Xi$.*

Conjecture (étendue w/ M.Menni)

a Grothendieck topos $\mathcal{E}$ is an **étendue** if and only if Ξ has the minimum element $\perp : 1 \to \Xi$

This conjecture holds for presheaf topoi.

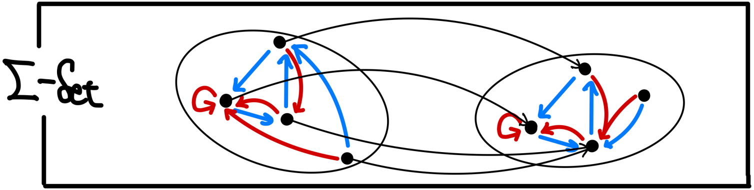
Table of Contents

Recall: Topos of Σ-sets

Definition

Σ-$\mathbf{Set}$ is the presheaf topos Σ-$\mathbf{Set} := \mathbf{PSh}(\Sigma^*)$.

$\Sigma = \{a, b\}$



LSC in Σ-**Set** (1/2)

Example (LSC in Σ-Set [Hora n.d.(b)])

The LSC of Σ-**Set** is **the Σ-set of right congruences** Ξ equipped with the Σ^*-action

$$u\ (\sim * w)\ v \iff wu \sim wv.$$

Each component $\xi_{(Q,\delta)}$ sends q to

$$u\ \xi_{(Q,\delta)}(q)\ v \iff qu = qv.$$

The $\wedge$-operation coincides with the logical one

$$u\ (\sim \wedge \sim')\ v \iff u \sim v \text{ and } u \sim' v$$

LSC in Σ-**Set** (2/2)

A set of right congruences $F \subset \Xi$ is an internal filter in Σ-**Set** if and only if

1. F is closed under the Σ^*-action.

$$\sim \in F \implies \sim * w \in F$$

2. F is closed under finite infimums $\top, \wedge$.

$$\top \in F$$

$$\sim, \sim' \in F \implies \sim \wedge \sim' \in F$$

3. F is upward closed.

$$\sim \leq \sim' \text{ and } \sim \in F \implies \sim' \in F$$

The topos of orbit-finite Σ-**Set**

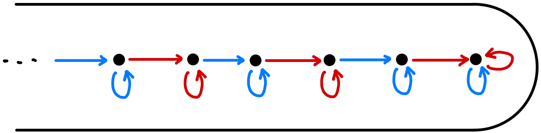
Let us consider the internal filter

$$F_{\mathrm{o.f.}} := \{\sim \in \Xi \mid \Sigma^*/\sim \text{ is finite}\}.$$

The corresponding hyperconnected quotient $h_{\mathrm{o.f.}} : \Sigma\text{-}\mathbf{Set} \twoheadrightarrow \Sigma\text{-}\mathbf{Set}_{\mathrm{o.f.}}$ is the topos of orbitwise-finite Σ-set.

Definition

A Σ-set $(Q, \delta \colon Q \times \Sigma \to Q)$ is said to be **orbit-finite** if, for any state $q \in Q$, its orbit $\{qw \mid w \in \Sigma^*\}$ is a finite set.



Recall: Languages form an internal Boolean algebra in Σ-**Set**

Example (The canonical point[1])

The forgetful-cofree adjunction defines a point of Σ-**Set**.

$$p_* : \mathbf{Set} \underset{S \mapsto S^{\Sigma^*}}{\overset{Q \leftarrow\!\shortmid (Q,\delta)}{\rightleftarrows}} \Sigma\text{-}\mathbf{Set} : p^* \quad (\perp)$$

We have $p_*(S) = S^{\Sigma^*}$. In particular, we have $p_*(\{\top, \bot\}) = \mathbf{Lan}$

Lemma

Lan *is an internal Boolean algebras in* Σ-**Set***.*

[1] This terminology is borrowed from [Rogers 2023]

Σ-**Set**$_{\mathrm{o.f.}}$ captures four definitions of regular languages (1/4)

Σ-**Set**$_{\mathrm{o.f.}}$ also has the canonical point

$$p_{\mathrm{o.f.}} : \mathbf{Set} \xrightarrow{p} \Sigma\text{-}\mathbf{Set} \xrightarrow{h_{\mathrm{o.f.}}} \Sigma\text{-}\mathbf{Set}_{\mathrm{o.f.}}.$$

Proposition ([Hora 2024b])

The internal Boolean algebra $(p_{\mathrm{o.f.}})_*(\{\top, \bot\})$ coincides with the Boolean algebra of regular languages **Reg**.

Σ-**Set**$_{\text{o.f.}}$ captures four definitions of regular languages (1/4)

Σ-**Set**$_{\text{o.f.}}$ also has the canonical point

$$p_{\text{o.f.}} : \mathbf{Set} \xrightarrow{p} \Sigma\text{-}\mathbf{Set} \xrightarrow{h_{\text{o.f.}}} \Sigma\text{-}\mathbf{Set}_{\text{o.f.}}.$$

Proposition ([Hora 2024b])

The internal Boolean algebra $(p_{\text{o.f.}})_*(\{\top, \bot\})$ coincides with the Boolean algebra of regular languages **Reg**.

This is nothing other than **the Myhill-Nerode theorem,** since $\xi_{\mathbf{Lan}}(L)$ is the Nerode-congruence:

$$u\ \xi_{\mathbf{Lan}}(L)\ v \iff L * u = L * v$$

$$\begin{tikzcd} \mathbf{Reg} \arrow[r] \arrow[d, hook, "\epsilon_{\mathbf{Lan}}"] \arrow[dr, phantom, "\lrcorner", very near start] & F_{\text{o.f.}} \arrow[d, hook, "\iota_F"] \\ \mathbf{Lan} \arrow[r, "\xi_{\mathbf{Lan}}"'] & \Xi, \end{tikzcd}$$

Σ-$\mathbf{Set}_{\text{o.f.}}$ captures four definitions of regular languages (2/4)

Definition

A language $L \in \mathbf{Lan}$ is **regular** if and only if

1. L is recognized by a **finite automaton.**
2. there exists a **finite monoid** M, a subset $S \subset M$, and a monoid homomorphism $\phi \colon \Sigma^* \to M$ such that $L = \phi^{-1}(S)$.
3. Its Nerode-**congruence** $\sim_L$ satisfies $\Sigma^*/\sim_L$
4. There exists a clopen subset of the **profinite monoid** of profinite words $U \subset \widehat{\Sigma^*}$ such that $L = \Sigma^* \cap U$.

If you read the proof of their equivalence, you may notice that it is the proof of **Morita equivalences between** 4 **ways to define the single boolean ringed topos** $(\Sigma$-$\mathbf{Set}_{\text{o.f.}}, \mathbf{Reg})$!

Σ-$\mathbf{Set}_{\mathrm{o.f.}}$ captures four definitions of regular languages (3/4)

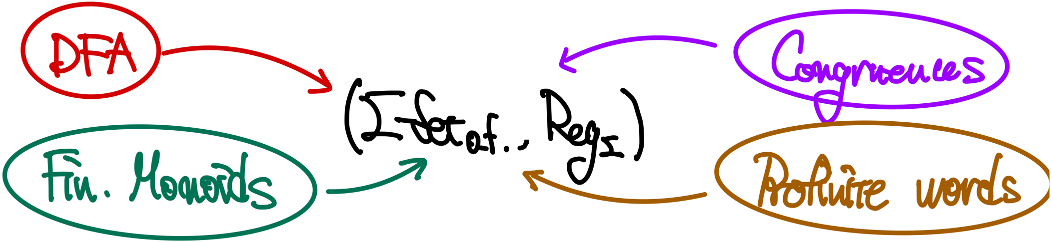
Theorem (Hora 2024b preprint (and my ongoing paper))

All of the following boolean ringed topoi are equivalent to $(\Sigma\text{-}\mathbf{Set}_{\mathrm{o.f.}}, \mathbf{Reg})$.

- $(\mathbf{Sh}(\Sigma\text{-}\mathbf{FinSet}, J), \mathrm{DFA})$: *The sheaf topos over the (Boolean algebra-ed) site of* ***DFA***.
- $(\mathbf{Sh}(\Sigma\text{-}\mathbf{FinMon}, J), \mathcal{P})$: *The sheaf topos over the (Boolean algebra-ed) site of* ***finite (Σ-) monoids.***
- $(\Sigma\text{-}\mathbf{Set}_{F_{\mathrm{o.f.}}}, h_{\mathrm{o.f.}*}(\mathbf{Lan}))$ *The hyperconnected quotient topos of* Σ-$\mathbf{Set}$ *corresponding to finite orbit right congruences. (**Myhill-Nerode theorem**)*
- $(\mathbf{Cont}(\widehat{\Sigma^*}), \mathrm{Clopen}(\widehat{\Sigma^*}))$: *The continuous action topos of the profinite monoid* $\widehat{\Sigma^*}$ *of* ***profinite words*** *with the Boolean algebra of clopen subsets.*

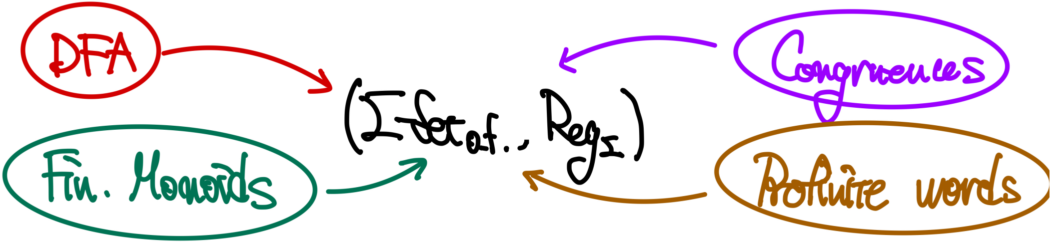
Σ-**Set**$_{\text{o.f.}}$ captures four definitions of regular languages (4/4)

This enriches the (conditional) equivalence between four definitions of regular languages to the (structural) equivalence between four boolean topoi!



Σ-$\mathbf{Set}_{\text{o.f.}}$ captures four definitions of regular languages (4/4)

This enriches the (conditional) equivalence between four definitions of regular languages to the (structural) equivalence between four boolean topoi!



Question

Can we define Σ-$\mathbf{Set}_{\text{o.f.}}$ by the syntactic system of (gen.) ***regular expressions****?*

Table of Contents

The complete lattice $\mathrm{HQ}(\mathcal{E})$

Corollary

For any Grothendieck topos $\mathcal{E}$, its hyperconnected quotients form a (small) complete lattice $\mathrm{HQ}(\mathcal{E})$ with the inclusion relations.

Due to the internal parameterization theorem, we have

$$\mathrm{HQ}(\mathcal{E}) \cong \mathrm{Filt}(\Xi).$$

Example ($\mathrm{HQ}(\mathbf{PSh}(G))$)

For a discrete group G, the lattice $\mathrm{HQ}(\mathbf{PSh}(G))$ is the complete lattice of "action topologies." [Rogers 2023]
If G is finite, we have $\mathrm{HQ}(\mathbf{PSh}(G)) \cong \mathrm{NorSub}_{\mathrm{Grp}}(G)^{\mathrm{op}}$

Galois connection for an object (1/2)

We start from the familiar Galois connection

$$\begin{tikzcd} \mathrm{Sub}_{\mathcal{E}}(\Xi) \arrow[r, phantom, "\perp"] & \mathrm{Sub}_{\mathcal{E}}(A) \arrow[l, bend right, "\exists_{\xi_A}"'] \arrow[l, leftarrow, bend left, "\xi_A^{-1}"] \end{tikzcd} .$$

Its right adjoint ξ_A^{-1} is restricted to another right adjoint (due to GAFT)

$$\begin{tikzcd} \mathrm{HQ}(\mathcal{E}) \cong \mathrm{Filt}(\Xi) \arrow[r, phantom, "\perp"] & \mathrm{Sub}_{\mathcal{E}}(A) \arrow[l, bend right, "?"'] \arrow[l, leftarrow, bend left, "\xi_A^{-1}"] \end{tikzcd} .$$

Here, the composite $\mathrm{HQ}(\mathcal{E}) \to \mathrm{Sub}_{\mathcal{E}}(A)$ sends h to $h^* h_* A \rightarrowtail A$.

Galois connection for an object (2/2)

Lemma (Galois connection for each object)

For any Grothendieck topos $\mathcal{E}$ and any object A, we have the Galois connection

$$\mathrm{HQ}(\mathcal{E}) \underset{h \mapsto h^* h_* A}{\overset{?}{\underset{\longrightarrow}{\overset{\longleftarrow}{\perp}}}} \mathrm{Sub}_{\mathcal{E}}(A).$$

Example (Finite group)

For a finite group G and a G-set X, the Galois connection is

$$\mathrm{NorSub}_{\mathrm{Grp}}(G)^{\mathrm{op}} \cong \mathrm{HQ}(\mathbf{PSh}(G)) \underset{N \mapsto X^N}{\underset{\longrightarrow}{\overset{\longleftarrow}{\perp}}} \mathrm{Sub}_{\mathbf{PSh}(G)}(X)$$

Galois connection for an algebra (1/2)

Suppose that A is an internal $\mathbb{T}$-algebra for an algebraic theory $\mathbb{T}$. Then the Galois connection

$$\begin{tikzcd} \mathrm{HQ}(\mathcal{E}) \arrow[r, shift right=1.5ex, "h \mapsto h^* h_* A"'] \arrow[r, phantom, "\perp"] & \mathrm{Sub}_{\mathcal{E}}(A) \arrow[l, shift right=1.5ex, "?"'] \end{tikzcd} .$$

further restricts to

$$\begin{tikzcd} \mathrm{HQ}(\mathcal{E}) \arrow[r, shift right=1.5ex, "h \mapsto h^* h_* A"'] \arrow[r, phantom, "\perp"] & \mathrm{Sub}_{\mathbb{T}}(A) \arrow[l, shift right=1.5ex] \end{tikzcd}$$

since $h^* h_* A \rightarrowtail A$ is always a sub$\mathbb{T}$-algebra.

Galois connection for an algebra (2/2)

Theorem (The Galois connection for any internal algebra)

For any Grothendieck topos $\mathcal{E}$, any algebraic theory $\mathbb{T}$, and any internal $\mathbb{T}$-algebra A, we obtain the following Galois connection

$$\mathrm{HQ}(\mathcal{E}) \underset{h \mapsto h^* h_* A}{\overset{\longleftarrow}{\underset{\longrightarrow}{\perp}}} \mathrm{Sub}_{\mathbb{T}}(A)$$

Here, $\mathrm{Sub}_{\mathbb{T}}(A)$ consists of **internal** sub$\mathbb{T}$-algebras.

Example (For a finite-dimensional Galois extension $K/\mathbb{Q}$,)

$$\mathrm{NorSub}_{\mathrm{Grp}}(\mathrm{Gal}(K/\mathbb{Q}))^{\mathrm{op}} \cong \mathrm{HQ}(\mathbf{PSh}(\mathrm{Gal}(K/\mathbb{Q}))) \underset{N \mapsto K^N}{\overset{\longleftarrow}{\underset{\longrightarrow}{\cong}}} \mathrm{Sub}_{\mathbb{Q}\text{-alg}}(K)$$

Table of Contents

The Galois connection between topoi and language classes

Let $\mathrm{Sub}_{\mathrm{bool}}(\mathbf{Lan})$ denotes the complete lattice of all Σ-**Set**-internal[2] subboolean algebras of **Lan**.

Corollary (Galois connection for language varieties)

There is a Galois connection

$$\begin{tikzcd} \mathrm{HQ}(\Sigma\text{-}\mathbf{Set}) \arrow[r, shift right=1.5ex, "\mathbb{L}: h \mapsto h^* h_* \mathbf{Lan}"'] \arrow[r, phantom, "\perp"] & \mathrm{Sub}_{bool}(\mathbf{Lan}) \arrow[l, shift right=1.5ex, "\mathcal{T}"'] \end{tikzcd} .$$

$\mathbb{L}$ denotes the right adjoint $h \mapsto h^* h_* \mathbf{Lan}$, and $\mathcal{T}$ denotes its left adjoint.

[2]Here, "internal" means "Σ^*-action closed": $L \in C \implies L * w = \{u \in \Sigma^* \mid wu \in L\} \in C$.

Right adjoint $\mathbb{L}\colon \mathrm{HQ}(\Sigma\text{-}\mathbf{Set}) \to \mathrm{Sub}_{\mathrm{bool}}(\mathbf{Lan})$

Theorem (General Myhill-Nerode theorem)

For any hyperconnected geometric morphism $h\colon \Sigma\text{-}\mathbf{Set} \to \mathcal{F}$, $\mathbb{L}_{\mathcal{F}} \subset \mathbf{Lan}$ *coincides with all of the following:*

comonad: $\mathbb{L}_{\mathcal{F}} := h^* h_* \mathbf{Lan} \rightarrowtail \mathbf{Lan}$

int. filter: $\mathbb{L}_{\mathcal{F}} = \{L \in \mathbf{Lan} \mid \xi_{\mathbf{Lan}}(L) \in F\}$, *i.e.,*

The pullback $\mathbb{L}_{\mathcal{F}} = \mathbf{Lan} \times_{\Xi} F$*:*

$$\begin{tikzcd} \mathbb{L}_{\mathcal{F}} \arrow[r] \arrow[d, tail, "\epsilon_{\mathbf{Lan}}"] \arrow[dr, phantom, "\lrcorner", very near start] & F \arrow[d, tail, "\iota_F"] \\ \mathbf{Lan} \arrow[r, "\xi_{\mathbf{Lan}}"'] & \Xi, \end{tikzcd}$$

automaton: $\mathbb{L}_{\mathcal{F}}$ *is the set of languages recognized by a* Σ*-set in* $\mathcal{F} \subset \Sigma\text{-}\mathbf{Set}$.

Right adjoint $\mathbb{L}\colon \mathrm{HQ}(\Sigma\text{-}\mathbf{Set}) \to \mathrm{Sub}_{\mathrm{bool}}(\mathbf{Lan})$

Example (Original Myhill-Nerode theorem)

For the hyperconnected geometric morphism $h\colon \Sigma\text{-}\mathbf{Set} \to \Sigma\text{-}\mathbf{Set}_{\text{o.f.}}$, we have $\mathbb{L}_{\Sigma\text{-}\mathbf{Set}_{\text{o.f.}}} = \mathbf{Reg}$ by two reasons

int. filter: $\mathbb{L}_{\Sigma\text{-}\mathbf{Set}_{\text{o.f.}}} = \{L \in \mathbf{Lan} \mid \xi_{\mathbf{Lan}}(L) \in F_{\text{o.f.}}\}$, i.e.,

The pullback $\mathbb{L}_{\Sigma\text{-}\mathbf{Set}_{\text{o.f.}}} = \mathbf{Lan} \times_{\Xi} F_{\text{o.f.}}$:

$$\begin{tikzcd} \mathbb{L}_{\mathcal{F}} \arrow[r] \arrow[d, hook, "\epsilon_{\mathbf{Lan}}"] \arrow[dr, phantom, "\lrcorner", very near start] & F_{\text{o.f.}} \arrow[d, hook, "\iota_F"] \\ \mathbf{Lan} \arrow[r, "\xi_{\mathbf{Lan}}"'] & \Xi, \end{tikzcd}$$

automaton: $\mathbb{L}_{\Sigma\text{-}\mathbf{Set}_{\text{o.f.}}}$ is the set of languages recognized by a Σ-set in $\Sigma\text{-}\mathbf{Set}_{\text{o.f.}} \subset \Sigma\text{-}\mathbf{Set}$.

Right adjoint $\mathbb{L}$ subsumes recognition by monoid

Algebraic language theory (in a narrow sense) is a monoid-theoretic approach to (regular) language theory.

Definition (Language Recognition)

A language L is said to be **recognized** by a surjective monoid homomorphism $\phi\colon \Sigma^* \twoheadrightarrow M$ if there is a subset $S \subset M$ such that $L = \phi^{-1}(S)$.

Proposition ($\mathbb{L}$ subsumes recognition by monoid)

A surj. monoid hom. $\phi\colon \Sigma^* \twoheadrightarrow M$ recognizes L iff $L \in \mathbb{L}_{\mathbf{PSh}(M)}$ for the corresponding hyperconnected quotient $\phi\colon \Sigma\text{-}\mathbf{Set} \twoheadrightarrow \mathbf{PSh}(M)$.

Left adjoint $\mathcal{T} : \mathrm{Sub}_{\mathrm{bool}}(\mathbf{Lan}) \to \mathrm{HQ}(\Sigma\text{-}\mathbf{Set})$

Compared to the right adjoint $\mathbb{L}$, the left adjoint $\mathcal{T}$ is trickier and more interesting.

Example (Other examples)

For general Σ,

- $\mathcal{T}_{\mathbf{Lan}} = \Sigma\text{-}\mathbf{Set}$
- $\mathcal{T}_{\mathbf{Reg}} = \Sigma\text{-}\mathbf{Set}_{\mathrm{o.f.}}$
- $\mathcal{T}_{\mathbf{Grp}} = \mathbf{Cont}(\widehat{F_\Sigma})$
- $\mathcal{T}_{\mathbf{Com}} = \mathbf{PSh}(\mathbb{N}^{\oplus\Sigma})$

For $\Sigma = \{*\}$,

- $\mathcal{T}_{\mathbf{Lan}} = \mathbf{PSh}(\mathbb{N}) = \sigma\text{-}\mathbf{Set}$
- $\mathcal{T}_{\mathbf{Reg}} = \mathbf{Cont}(\mathbb{N} \cup \widehat{\mathbb{Z}})$
- $\mathcal{T}_{\mathbf{star-free}} = \mathbf{Cont}(\mathbb{N} \cup \{\infty\})$
- $\mathcal{T}_{\mathbf{Grp}} = \mathbf{Cont}(\widehat{\mathbb{Z}})$
- $\mathcal{T}_{p\text{-}\mathbf{adL}} = \mathbf{Cont}(\mathbb{Z}_p)$

Left adjoint $\mathcal{T}$ subsumes Syntactic monoid (1/3)

Remark (Domain of $\mathcal{T}$)

Although the domain of $\mathcal{T}$ is a priori $\mathrm{Sub}_{\mathrm{bool}}(\mathbf{Lan}) \subset \mathcal{P}(\mathbf{Lan})$, we can naturally extend this to $\mathcal{P}(\mathbf{Lan})$ by

$$\begin{tikzcd} \mathrm{HQ}(\Sigma\text{-}\mathbf{Set}) \arrow[r, shift right, "\mathbb{L}"'] \arrow[r, phantom, "\perp"] & \mathrm{Sub}_{\mathrm{bool}}(\mathbf{Lan}) \arrow[l, shift right, "\mathcal{T}"'] \arrow[r, hook, shift right] \arrow[r, phantom, "\perp"] & \mathcal{P}(\mathbf{Lan}) \arrow[l, two heads, shift right, "\langle - \rangle"'] \end{tikzcd}$$

In particular, we can consider $\mathcal{T}_{\langle L \rangle}$, **the topos associated to a language L**!

Left adjoint $\mathcal{T}$ subsumes Syntactic monoid (2/3)

For a language L, its **syntactic monoid** $\pi_L \colon \Sigma^* \twoheadrightarrow M_L$ is the minimum monoid that recognizes L.

Proposition (Syntactic topos subsumes syntactic monoid)

For any regular[3] language L, we have

$$\mathcal{T}_{\langle L \rangle} = \mathbf{PSh}(M_L).$$

(Here is an interesting word combinatorics. This holds for a language L s.t. $\xi_{\equiv}(\xi_{\mathbf{Lan}}(L)) \in F_{\text{o.f.}}$.)

[3] This works for a broader class.

Left adjoint $\mathcal{T}$ subsumes Syntactic monoid (3/3)

But in general, $\mathcal{T}_{\langle L \rangle}$ is a topological monoid (action topos), which has more information than the syntactic monoid M_L!

Example (Well-parenthesized words has prodiscrete syntactic monoid)

For $\Sigma = \{<, >\}$ and the language $L = \{w \in \Sigma^* \mid w \text{ is well-parenthesized.}\}$,

- $\mathcal{T}_{\langle L \rangle} \subsetneq \mathbf{PSh}(M_L)$ for $M_L = \langle a, b \mid ab = 1 \rangle$, and
- it is a prodiscrete monoid action topos $\mathcal{T}_{\langle L \rangle} = \mathbf{Cont}(M_L \sqcup \{\infty\})$.

(cf. Algebraic language theory try to connect properties of languages and those of syntactic monoids. It goes well for regular languages, but ...)

Left adjoint $\mathcal{T}$: Local variety theorem

The lattice $\mathrm{Sub}_{\mathrm{bool}}(\mathbf{Reg})$ has been of interest in the algebraic language theory as the lattice of **local varieties of regular languages**.

Theorem (Local varieties)

The Galois connection for the boolean ringed topos $(\Sigma\text{-}\mathbf{Set}_{\text{o.f.}}, \mathbf{Reg})$

$$
\begin{tikzcd}
\mathrm{HQ}(\Sigma\text{-}\mathbf{Set}_{\text{o.f.}}) \arrow[r, phantom, "\perp"] \arrow[r, bend right, "\mathbb{L}"'] & \mathrm{Sub}_{\mathrm{bool}}(\mathbf{Reg}) \arrow[l, bend right, "\mathcal{T}"']
\end{tikzcd} .
$$

is an isomorphism of complete lattices.

This reminds me of the $\mathbb{Q}$-algebra-ed topos $(\mathbf{PSh}(\mathrm{Gal}(K/\mathbb{Q})), K)$.

Digression: Module structure of $\mathrm{HQ}(\mathcal{E})$

Proposition

For any Grothendieck topos $\mathcal{E}$, the complete lattice of hyperconnected quotients $\mathrm{HQ}(\mathcal{E})$ is a left module of the semiring $\mathrm{End}_\wedge(\Omega)$.

The lattice of subtopoi is the lattice of idempotents of $\mathrm{End}_\wedge(\Omega)$.

Example (Σ-**Set**$_{\text{o.f.}}$ when $\Sigma = \{*\}$)

If $\Sigma = \{*\}$ and $\mathcal{E} = \Sigma$-**Set**$_{\text{o.f.}}$,

- the semiring $\mathrm{End}_\wedge(\Omega)$ is the **min-plus algebra** $\{0 < 1 < \cdots < \infty < \infty'\}$, and
- $\mathrm{HQ}(\mathcal{E}) \cong \overline{\mathbb{N}}^{\mathrm{Spec}(\mathbb{Z})}$.

References

- Fiore, Marcelo, Zeinab Galal, and Hugo Paquet (2024). "Stabilized profunctors and stable species of structures". In: *Logical Methods in Computer Science* 20.
- Hora, Ryuya (2024a). "Internal parameterization of hyperconnected quotients". In: *Theory and Applications of Categories* 42.11, pp. 263–313.
- — (2024b). "Topoi of automata I: Four topoi of automata and regular languages". In: *arXiv:2411.06358*.
- — (n.d.[a]). *Normalization of a subgroup, in a category, and of a word-congruence (ongoing)*.
- — (n.d.[b]). *Topoi of automata II: Hyperconnected geometric morphisms, syntactic monoids, and language classes (ongoing)*.
- Johnstone, Peter T. (1981). "Factorization theorems for geometric morphisms, I". In: *Cahiers de topologie et géométrie différentielle catégoriques* 22.1, pp. 3–17.
- Kamio, Yuhi and Ryuya Hora (2024). "A solution to the first Lawvere's problem A Grothendieck topos that has a proper class many quotient topoi". In: *arXiv:2407.17105*.
- Lawvere, F. William (Apr. 2009). *open problems in topos theory*. One can find this article via `https://ncatlab.org/nlab/show/William+Lawvere`.
- Menni, Matias (2025). "Non-singular maps in toposes with a local state classifier". In: *arXiv:2505.07131*.
- Rogers, Morgan (2023). "Toposes of topological monoid actions". In: *Compositionality: the open-access journal for the mathematics of composition* 5.

Algebraic Language theory

Definition (Syntactic Congruence)

For a language L, its **syntactic congruence** $\equiv_L$ is defined by

$$u \equiv_L v \iff (\forall w, w' \in \Sigma^*,\ wuw' \in L \iff wvw' \in L)\,.$$

The projection $\pi_L \colon \Sigma^* \twoheadrightarrow \Sigma^*/\cong_L$ induces a monoid structure on $\Sigma^*/\cong_L$. This monoid is called **the syntactic monoid** of L, and denoted by M_L.

Proposition

The syntactic monoid $\pi_L \colon M_L$ provides the minimum monoid that recognizes L.

Examples of Syntactic monoid (1/2): ABA

The syntactic monoid M_{ABA} has 6 elements:

- $[\varepsilon] = \{\varepsilon\}$
- $[a] = \{a, aba, ababa, \dots\}(= \mathrm{ABA})$
- $[b] = \{b, bab, babab, \dots\}$
- $[ab] = \{ab, abab, ababab, \dots\}$
- $[ba] = \{ba, baba, bababa, \dots\}$
- $[aa] = [bb] = \{aa, bb, aaa, aab, abb, baa, bba, bbb, \dots\}$

We can conclude many properties of ABA from the properties of M_{ABA}. (e.g. Since the syntactic monoid M_{ABA} is acyclic, the Schützenberger theorem implies that ABA is definable by first-order logic.)

$(\exists x\ (\forall y\ \neg(y < x)) \wedge a(x)) \wedge (\exists x\ (\forall y\ \neg(x < y)) \wedge a(x)) \wedge (\forall x \forall y\ (x < y \wedge (\forall z, \neg(x < z < y))) \to ((a(x) \wedge b(y)) \vee (b(x) \wedge a(y))))$

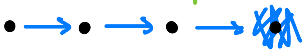
Cohomology of language class, the easiest case

As every topos does, the topos $\mathcal{T}_C$ has geometry, in particular, **cohomology.**
Even in the easiest case, where $\Sigma = \{*\}$ and $C = \{L\}$, we have:

L is	$\cdots$	$H^3(\mathcal{T}_L, A)$	$H^2(\mathcal{T}_L, A)$	$H^1(\mathcal{T}_L, A)$	$H^0(\mathcal{T}_L, A)$	dim
star-free	0	0	0	0	A^f	0
reg. but non-s.f.	$A_{\text{odd/even}}$	A_{odd}	A_{even}	A_{odd}	A^f	∞
non-regular	0	0	0	$A/(1-f)$	A^f	1

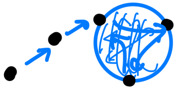
L: Star-free

like point



L: regular, not star-free

like $\mathbb{RP}^\infty$



L: non-regular

like S^1