Topoi of automata (IRIF)

Ryuya Hora

University of Tokyo

May 6, 2025

Slides

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

Table of Contents

1. Motivation: Why topos?
2. Preliminaries: What's topos?
3. Σ-**Set**: Languages emerge from a point
4. Σ-**Set**$_{\mathrm{o.f.}}$: The topos of regular languages
5. $\mathcal{T}_C$: Syntactic topos of a language class
6. How to calculate $\mathcal{T}$: the Local state classifier Ξ

Table of Contents

1. Motivation: Why topos?
2. Preliminaries: What's topos?
3. Σ-**Set**: Languages emerge from a point
4. Σ-**Set**$_{\mathrm{o.f.}}$: The topos of regular languages
5. $\mathcal{T}_C$: Syntactic topos of a language class
6. How to calculate $\mathcal{T}$: the Local state classifier Ξ

Overview of this section

We study **topos of automata**, not automata in a topos (cf. [Iwaniack 2024]).
In this section, I will explain **three motivations**:

Multi-facets Both automata theory and topos theory are **multi-faceted in a quite similar way**.

Galois-Geometry Both theories deal with **profinite Galois theory**.

New invariants Towards some open problems (like gen. Star-height one), I want to provide new **geometric invariants**.

First motivation: Multi-facets

Topos theory and automata theory are **multi-faceted in a similar way.**

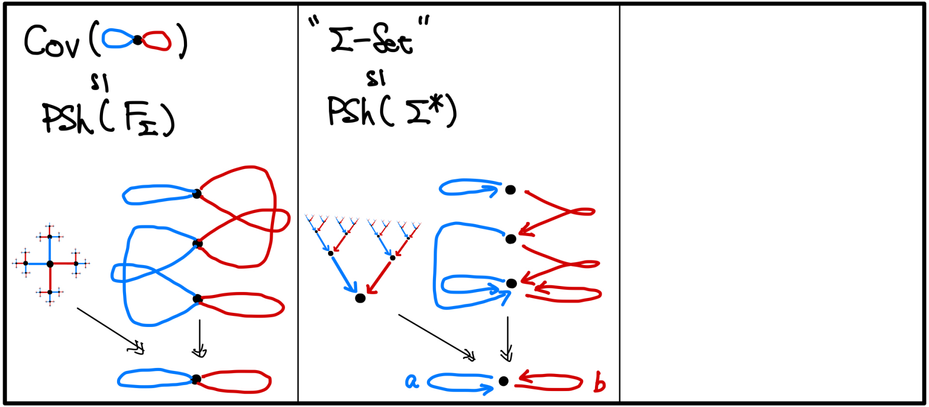
Logic	first-order, and higer order logic, ...
Topology	Stone duality, fibration and profinite topological monoids, ...
Algebra	Monoids, and Grothendieck (semi-)Galois theory, ... (cf. [Uramoto 2017])
Category	Coalgebra, fibration, Kan extension, factorization ... (cf. [Melliès and Zeilberger 2023],[Colcombet and Petrişan 2020])

Question

Can we understand the multi-facetedness of regular languages as that of a Grothendieck topos?

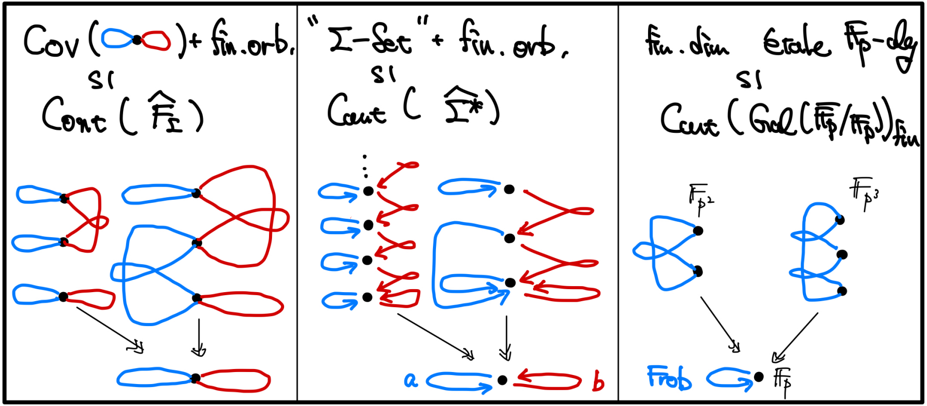
$\rightarrow$ The topos of regular languages Σ-**Set**$_{\mathrm{o.f.}}$ with 4 definitions.

Second motivation: Galois geometry



[cf. Melliès and Zeilberger 2023]

Second motivation: Galois geometry



[cf. Uramoto 2016]

Third motivation: New invariants

- Motivated by Weil conjecture, Grothendieck defined **topoi** (sites, and schemes) to introduce continuous/infinite geometric methods into number theory, which is traditionally thought of as a finite/discrete subject.

Third motivation: New invariants

- Motivated by Weil conjecture, Grothendieck defined **topoi** (sites, and schemes) to introduce continuous/infinite geometric methods into number theory, which is traditionally thought of as a finite/discrete subject.
- Can we imitate Grothendieck in automata theory, which is “usually" regarded as finite/discrete subject?

Third motivation: New invariants

- Motivated by Weil conjecture, Grothendieck defined **topoi** (sites, and schemes) to introduce continuous/infinite geometric methods into number theory, which is traditionally thought of as a finite/discrete subject.
- Can we imitate Grothendieck in automata theory, which is "usually" regarded as finite/discrete subject?

Question

Can we construct a space (= topos) for each language class and extracts its topological invariants?

For each $C \subset \mathbf{Lan}$, we construct a (connected and locally connected) topos $\mathcal{T}_C$

Third motivation: New invariants

Are topoi really spaces? in *Topo-logie* [Anel and Joyal 2021]

[...] the intuition of space is in fact forged in a set of specific operations on spaces

- (e.g., covering, pasting, quotienting, localizing, intersecting, crossing, deforming, direct image, inverse image, homotopy, **(co)homology,** etc.),

which leads to distinguishing some classes of spaces

- (compact, connected, contractible, etc.)

and some classes of maps

- (open immersions, étale maps, submersions, proper maps, bundles, etc.).

So far, all of these notions and the structural relations they have between them have been **successfully generalized to topoi.**

This talk in one slide

Section 2 Preliminaries on Topos.

Section 3 **Languages emerge from a point**: exemplified by a topos Σ-**Set**.

Section 4 Regular language theory "$\simeq$" geometry of a topos Σ-**Set**$_{\mathrm{o.f.}}$.

Section 5 **Syntactic topos** $\mathcal{T}_C$ for each language class $C \subset$ **Lan**.

- $\mathcal{T}_{\{L\}} \simeq$ **PSh**$(\Sigma^*/\cong_L)$ for a regular language L and its **syntactic monoid** $\Sigma^*/\cong_L$.
- $\mathcal{T}_{\mathbf{Reg}} \simeq$ **Cont**$(\widehat{\Sigma^*})$ for the profinite monoid of **profinite words** $\widehat{\Sigma^*}$.

$\to$ the n-**th cohomology** $H^n(\mathcal{T}_C, A)$ for each $C \subset$ **Lan**

Section 6 The key tool is **the colimit of all monomorphisms** Ξ in Σ-**Set**, which consists of all **right congruences**.

Table of Contents

1. Motivation: Why topos?
2. Preliminaries: What's topos?
3. Σ-**Set**: Languages emerge from a point
4. Σ-**Set**$_{\mathrm{o.f.}}$: The topos of regular languages
5. $\mathcal{T}_C$: Syntactic topos of a language class
6. How to calculate $\mathcal{T}$: the Local state classifier Ξ

Formal definition of topos, which we won't use.

A topos is a category satisfying very strong conditions so that we can do geometry.

Definition (topos)

A **topos**[1] $\mathcal{E}$ is a category $\mathcal{E}$ such that

- there exsists a small category $\mathcal{C}$, and
- a fully faithful functor $\iota \colon \mathcal{E} \hookrightarrow \mathbf{PSh}(\mathcal{C})$ with a left exact left adjoint.

What to remember:
A topos is just a(n extremely nice) category (without any additional structures).

[1] In this talk, 'topos' always means a Grothendieck topos

Typical Examples of topoi (1/2)

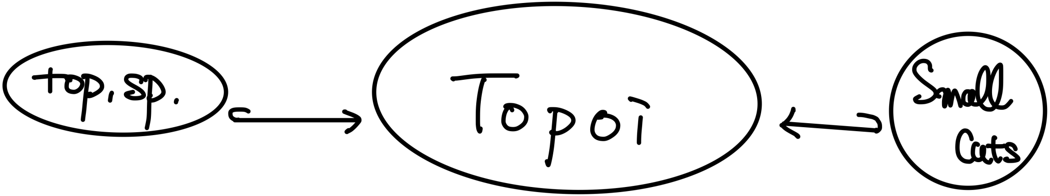
Example (Topological example, which we won't use.)

For a topological space X, we can construct its **sheaf topos Sh**(X).

Example (Presheaf topos, which we will use.)

For a small category $\mathcal{C}$, its **presheaf category PSh**$(\mathcal{C})$ is a topos.

Especially, **Set** $=$ **Sh**(1) $=$ **PSh**(1) is a topos.



Typical Examples of topoi (2/2)

Example (Cont(M) [Rogers 2023])

For a topological monoid M, a **continuous** M**-action** is a pair (X, τ) of a set ($=$ discrete space) X and a continuous action $\tau \colon X \times M \to X$.
For any topological monoid M, the category of continuous M-actions **Cont**(M) is a topos.

- Schanuel topos **Cont**$(\mathrm{Aut}(\mathbb{N}))$
- **Cont**$(\widehat{\Sigma^*})$

Geometric morphism (morphism of topoi)

Definition

A **geometric morphism** f from a topos $\mathcal{E}$ to another topos $\mathcal{F}$ is an adjunction

$$\begin{tikzcd} \mathcal{E} \arrow[r, bend right, "f_*"'] \arrow[r, phantom, "\perp"] & \mathcal{F} \arrow[l, bend right, "f^*:\ \text{lex}"'] \end{tikzcd},$$

such that the left adjoint f^* preserves finite limits.

Geometric morphisms generalize continuous maps:

Example

A conti. map $f\colon X \to Y$ induces a geometric morphism $f\colon \mathbf{Sh}(X) \to \mathbf{Sh}(Y)$.

What to remember

- **Topoi** are very nice categories (without any additional structures)
  - For any small category $\mathcal{C}$, $\mathbf{PSh}(\mathcal{C})$ is a topos.
  - For any topological monoid M, $\mathbf{Cont}(M)$ is a topos.
- A **geometric morphism** $f\colon \mathcal{E} \to \mathcal{F}$ is an adjunction

$$\begin{tikzcd} \mathcal{E} \arrow[r, bend right, "f_*"'] \arrow[r, phantom, "\perp"] & \mathcal{F} \arrow[l, bend right, "f^*\colon \text{ lex}"'] \end{tikzcd}.$$

This talk in one slide

Section 2 Preliminaries on Topos.

Section 3 **Languages emerge from a point**: exemplified by a topos Σ-**Set**.

Section 4 Regular language theory "$\simeq$" geometry of a topos Σ-**Set**$_{\text{o.f.}}$.

Section 5 **Syntactic topos** $\mathcal{T}_C$ for each language class $C \subset$ **Lan**.

- $\mathcal{T}_{\{L\}} \simeq \mathbf{PSh}(\Sigma^*/\cong_L)$ for a regular language L and its **syntactic monoid** $\Sigma^*/\cong_L$.
- $\mathcal{T}_{\mathbf{Reg}} \simeq \mathbf{Cont}(\widehat{\Sigma^*})$ for the profinite monoid of **profinite words** $\widehat{\Sigma^*}$.

$\rightarrow$ the n**-th cohomology** $H^n(\mathcal{T}_C, A)$ for each $C \subset$ **Lan**

Section 6 The key tool is **the colimit of all monomorphisms** Ξ in Σ-**Set**, which consists of all **right congruences**.

Table of Contents

1. Motivation: Why topos?
2. Preliminaries: What's topos?
3. Σ-**Set**: Languages emerge from a point
4. Σ-**Set**$_{\mathrm{o.f.}}$: The topos of regular languages
5. $\mathcal{T}_C$: Syntactic topos of a language class
6. How to calculate $\mathcal{T}$: the Local state classifier Ξ

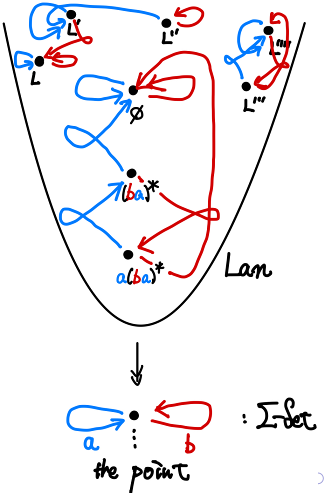
Overview of this section

Slogan: Languages emerge from a point

$$\mathbf{Lan} = p_*(\{\top, \bot\})$$

In this section, we observe this slogan via the prototypical example Σ-**Set**:

- The topos Σ-**Set** has the canonical point p
- With this, we obtain languages, recognition, minimal automaton, and coalgebraic automata.



Notations and terminologies

- We fix a finite set Σ, which we will call an **alphabet**.
- Σ^* denotes the (free) monoid of finite words.
- A **language** is a subset of Σ^*.

Σ-set

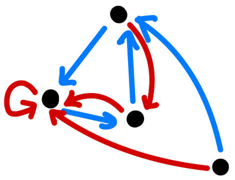
Definition (Σ-set)

A Σ-**set** is a pair (Q, δ) of

- a set Q (of states) and
- a (transition) function $\delta \colon Q \times \Sigma \to Q$.

Σ-set = right Σ^*-action = presheaf over the one-object category Σ^* = (contravariant) functor from Σ^* to **Set** = discrete fibration over the bouquet $\mathbb{B}_\Sigma$.

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



: Σ-set

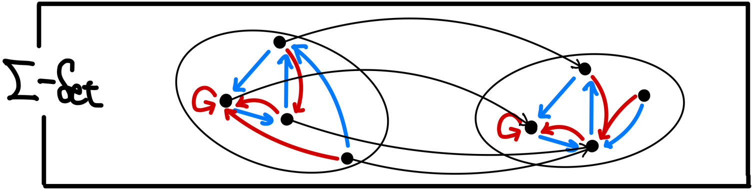
Topos of Σ-sets

Definition

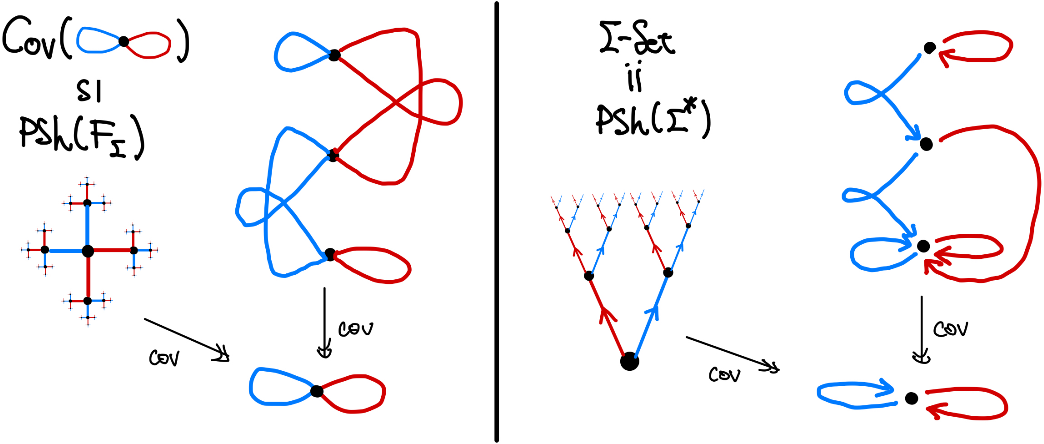
Σ-**Set** is the category of Σ-sets defined by Σ-**Set** $:=$ **PSh**(Σ^*).

Σ-**Set** is a (presheaf) topos.

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



Geometric idea: Σ-**Set** = directed bouquet



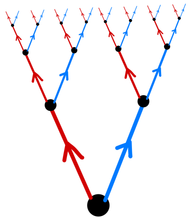
(cf. Nielsen-Schreier thm, Semi-Galois cat., fibrational approach)

Two objects of Σ-sets (1/2): Words

The topos Σ-**Set** has three important objects:

- The Σ-set of words Σ^*,
- The Σ-set of languages **Lan**.
- The Σ-set of right congruences Ξ.

Here, we will see first two.



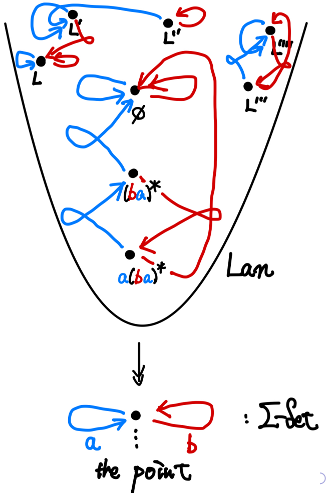
Example (The Σ-set of words Σ^*.)

Elements	Words.
Σ-action	Concatenation $v * w = vw$.
Universality	Yoneda lemma $\Sigma\text{-}\mathbf{Set}(\Sigma^*, (Q, \delta)) \cong Q$.

Two objects of Σ-sets (2/2): Languages

Example (The Σ-set of languages Lan)

Elements	Languages
Σ-action	$L * w := \{v \in \Sigma^* \mid wv \in L\}$ (Brzozowski derivative)
Universality	We will see soon!



Point of a topos

Definition (Points)

A **point** of a topos $\mathcal{E}$ is a geometric morphism $p\colon \mathbf{Set} \to \mathcal{E}$

$$\begin{tikzcd} \mathbf{Set} \arrow[r, shift right=2ex, "p_*"'] \arrow[r, phantom, "\perp"] & \mathcal{E} \arrow[l, shift right=2ex, "p^*\colon\ \mathrm{lex}"'] \end{tikzcd}$$

Example (Point of top.sp. is the stalk functor)

For a "sober" topological space X, we have

$$\text{Points of } X \overset{1:1}{\longleftrightarrow} \text{Points of } \mathbf{Sh}(X).$$

The canonical point of Σ-**Set** induces **Lan**

Example (The canonical point[2])

The forgetful-cofree adjunction defines a point of $p \colon \mathbf{Set} \to \Sigma\text{-}\mathbf{Set}$.

$$
\begin{tikzcd}
p_* : \mathbf{Set} \arrow[r, bend right, "S \mapsto S^{\Sigma^*}"'] \arrow[r, phantom, "\perp"] & \Sigma\text{-}\mathbf{Set} : p^* \arrow[l, bend right, "Q \mapsfrom (Q,\delta)"']
\end{tikzcd}
$$

Proposition (Universality of **Lan**)

- $\mathbf{Lan} = p_*(\{\top, \bot\})$
- $\Sigma\text{-}\mathbf{Set}((Q, \delta), \mathbf{Lan}) \cong \mathbf{Set}(Q, \{\top, \bot\}) \cong \mathcal{P}(Q)$.

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

Corollaries (1/3): Language recognition = composition

Proposition

These data are in bijective correspondence:

1. *An automaton* $(Q, \delta, q \in Q, F \subset Q)$.
2. *A diagram* $\Sigma^* \to (Q, \delta) \to \mathbf{Lan}$ *in the topos* Σ-**Set**.

Thus, an automaton

$$\Sigma^* \xrightarrow{\text{pick } q} (Q, \delta) \xrightarrow{\chi_F^\sharp} \mathbf{Lan}.$$

recognizes a language:

$$\Sigma^* \xrightarrow{\text{the recognised language}} \mathbf{Lan}.$$

Corollaries (2/3): Minimal automaton = epi-mono factorization

Conversely, any language L

$$\Sigma^* \xrightarrow{\quad \lceil L \rceil \quad} \mathbf{Lan}$$

induces the **minimal automaton**

$$\Sigma^* \twoheadrightarrow \mathrm{Im}(\lceil L \rceil) \rightarrowtail \mathbf{Lan}.$$

This also provides the concrete construction

$$\mathrm{Im}(\lceil L \rceil) = \Sigma^* / \ker\lceil L \rceil$$

since $\ker\lceil L \rceil \subset \Sigma^* \times \Sigma^*$ is the **Nerode congruence.**
(cf. Functorial approach [Colcombet and Petrișan 2020].)

Corollaries (3/3): Automata form a topos

Proposition

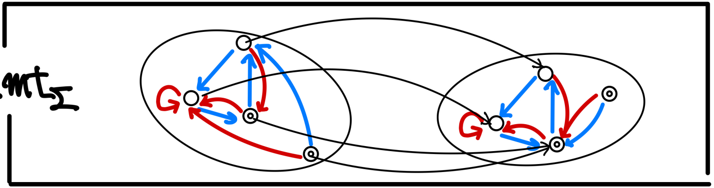
The category of automata $\mathbf{Atmt}_\Sigma = \mathbf{Coalg}_{2\times\Sigma}$ *is a presheaf topos.*

Proof.

$\mathbf{Atmt}_\Sigma \simeq \Sigma\text{-}\mathbf{Set}/\mathbf{Lan} \simeq \mathbf{PSh}\left(\int \mathbf{Lan}\right)$ $\square$

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

Atmt_Σ



Languages emerge from a point: the canonical boolean algebra

For any pointed topos $p\colon \mathbf{Set} \to \mathcal{E}$, p_* preserves all internal algebras, in particular, the boolean algebra of truth values $\{\top, \bot\}$.

Definition (canonical internal boolean algebra)

We call $p_*(\{\top, \bot\})$ **the canonical internal boolean algebra** of a pointed topos $p\colon \mathbf{Set} \to \mathcal{E}$.

Example (Summary of this section)

For the pointed topos $p\colon \mathbf{Set} \to \Sigma\text{-}\mathbf{Set}$, the canonical internal boolean algebra is the boolean algebra **Lan** of all languages (with derivatives).

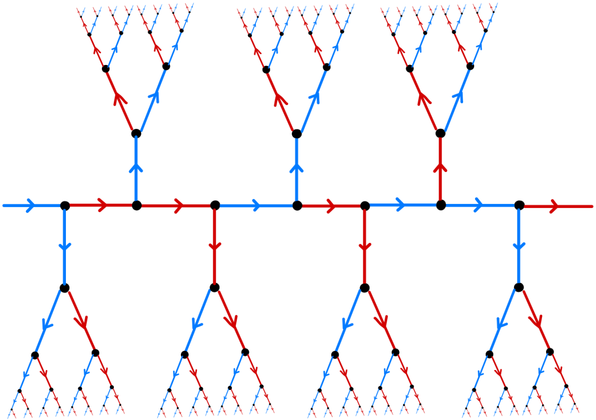
Digression: Other points = infinite words recognition!? (1/2)

We can apply the same argument to all other points of Σ-**Set**.

Proposition

*Other points of Σ-**Set** are in bijective correspondence with **infinite words** up to finite edit distance.*

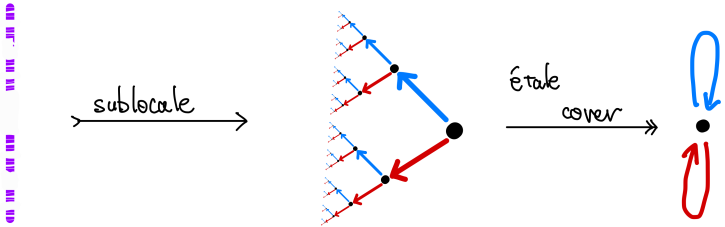
(cf. Büchi automata, [Leinster 2024])



Digression: Other points = infinite words recognition!? (2/2)

There is a canonical geometric morphism from the **Cantor space** Σ^ω!

$$\gamma\colon \mathbf{Sh}(\Sigma^\omega) \xhookrightarrow{\text{sublocale}} \mathbf{Sh}(\Sigma^*, \leq) \xrightarrow{\text{étale cover}} \Sigma\text{-}\mathbf{Set}$$



In particular, we obtain $\gamma^*\colon \Sigma\text{-}\mathbf{Set} \to \mathbf{Sh}(\Sigma^\omega)$, which captures the **eventual behavior** of (Q, δ).

This talk in one slide

Section 2 Preliminaries on Topos.

Section 3 **Languages emerge from a point**: exemplified by a topos Σ-**Set**.

Section 4 Regular language theory "$\simeq$" geometry of a topos Σ-**Set**$_{\mathrm{o.f.}}$.

Section 5 **Syntactic topos** $\mathcal{T}_C$ for each language class $C \subset$ **Lan**.

- $\mathcal{T}_{\{L\}} \simeq \mathbf{PSh}(\Sigma^*/\cong_L)$ for a regular language L and its **syntactic monoid** $\Sigma^*/\cong_L$.
- $\mathcal{T}_{\mathbf{Reg}} \simeq \mathbf{Cont}(\widehat{\Sigma^*})$ for the profinite monoid of **profinite words** $\widehat{\Sigma^*}$.

$\to$ the n-**th cohomology** $H^n(\mathcal{T}_C, A)$ for each $C \subset$ **Lan**

Section 6 The key tool is **the colimit of all monomorphisms** Ξ in Σ-**Set**, which consists of all **right congruences**.

Table of Contents

1. Motivation: Why topos?
2. Preliminaries: What's topos?
3. Σ-**Set**: Languages emerge from a point
4. Σ-**Set**$_{\mathrm{o.f.}}$: The topos of regular languages
5. $\mathcal{T}_C$: Syntactic topos of a language class
6. How to calculate $\mathcal{T}$: the Local state classifier Ξ

Overview of this section

Slogan: Regular language theory "$\simeq$" geometry of a topos Σ-Set$_{\mathrm{o.f.}}$.

DFA

Congruences

(Σ-Set$_{\mathrm{o.f.}}$, Reg$_\Sigma$)

Fin. Monoids

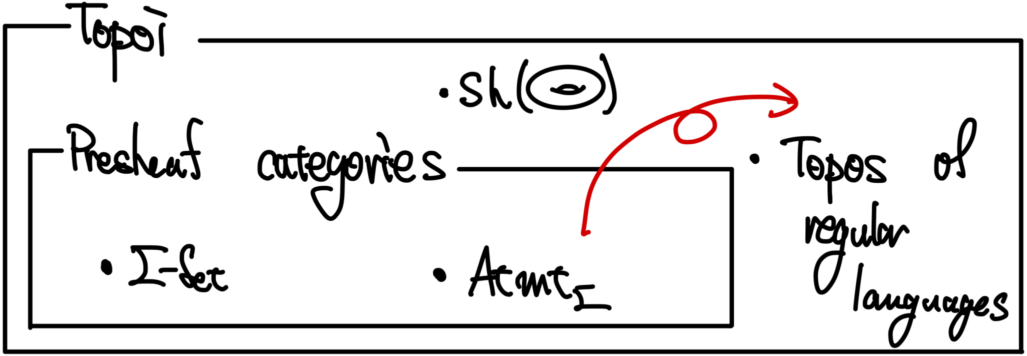
Profinite words

- The category of orbit-finite Σ-sets Σ-**Set**$_{\mathrm{o.f.}}$ is a pointed topos because it is a **hyperconnected quotient** of Σ-**Set**.
- Many aspects of regular languages defines the same topos Σ-**Set**$_{\mathrm{o.f.}}$.

We need finiteness, we need topoi!

So far,

- we did not care **finiteness**, and
- we only consider presheaf categories ($\subsetneq$ topoi).

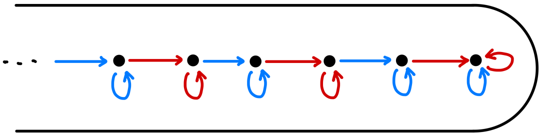


Orbit-finiteness

Instead of considering finiteness[3], we consider orbit-finiteness.

Definition (Σ-Set$_{\text{o.f.}}$)

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 $q\Sigma^* := \{qw \mid w \in \Sigma^*\}$ is a finite set. Let Σ-**Set**$_{\text{o.f.}}$ denote the full subcategory of Σ-**Set** that consists of all orbit finite Σ-sets.



[3] Σ-**FinSet** is nice [Uramoto 2017], but doesn't contain Σ^*, **Lan**, . . . , and it's not a topos.

Σ-**Set**$_{\mathrm{o.f.}}$ is a topos since it's a hyperconnected quotient (1/3)

But why is Σ-**Set**$_{\mathrm{o.f.}}$ a topos? Σ-**Set**$_{\mathrm{o.f.}}$ has nice closure properties in Σ-**Set**.
$\cdots$ A Σ-set is orbit-finite iff it recognizes only regular languages.

Lemma

*The full subcategory Σ-**Set** $\hookleftarrow$ Σ-**Set**$_{\mathrm{o.f.}}$ are closed under*

1. *finite products,* $\cdots$ *boolean operations*
2. *small coproducts, and* $\cdots$ *don't change language classes*
3. *subquotients.* $\cdots$ *don't change language classes*

Σ-**Set**$_{\mathrm{o.f.}}$ is a topos since it's a hyperconnected quotient (2/3)

Definition (Hyperconnected quotient)

A **hyperconnected quotient** of a topos $\mathcal{E}$ is a full subcategory of $\mathcal{E}$ closed under

1. finite products,
2. small coproducts, and
3. subquotients.

Example

Σ-**Set**$_{\mathrm{o.f.}}$ is a hyperconnected quotient of Σ-**Set**.

Σ-**Set**$_{\mathrm{o.f.}}$ is a topos since it's a hyperconnected quotient (3/3)

Fact (Hyperconnected quotient is a topos)

Any hyperconnected quotient $\mathcal{F}$ of any topos $\mathcal{E}$ is a topos. The embedding $\mathcal{E} \hookleftarrow \mathcal{F}$ is a lex left adjoint, so it defines a geometric morphism $h\colon \mathcal{E} \twoheadrightarrow \mathcal{F}$. (cf. [Johnstone 1981])

Proposition

Σ-**Set**$_{\mathrm{o.f.}}$ *is a topos that admits a geometric morphism* $h\colon \Sigma\text{-}\mathbf{Set} \twoheadrightarrow \Sigma\text{-}\mathbf{Set}_{\mathrm{o.f.}}$.

We got another pointed topos $p_{\mathrm{o.f.}}\colon \mathbf{Set} \to \Sigma\text{-}\mathbf{Set}_{\mathrm{o.f.}}$!

$$\begin{tikzcd} \mathbf{Set} \arrow[r, "p"] \arrow[rr, bend right, "p_{\mathrm{o.f.}}"'] & \Sigma\text{-}\mathbf{Set} \arrow[r, two heads, "\text{hyperconn.}"] & \Sigma\text{-}\mathbf{Set}_{\mathrm{o.f.}} \end{tikzcd}$$

The pointed topos Σ-$\mathbf{Set}_{\mathrm{o.f.}}$ induces regular languages (1/2)

In order to calculate the canonical internal boolean algebra $p_{\mathrm{o.f.}*}(\{\top, \bot\}) = h_*(p_*(\{\top, \bot\}))$, let's see the geometric morphism $h \colon \Sigma\text{-}\mathbf{Set} \twoheadrightarrow \Sigma\text{-}\mathbf{Set}_{\mathrm{o.f.}}$ in detail.

Example (Hyperconnected geometric morphism $h \colon \Sigma\text{-}\mathbf{Set} \twoheadrightarrow \Sigma\text{-}\mathbf{Set}_{\mathrm{o.f.}}$)

$$h : \Sigma\text{-}\mathbf{Set} \underset{h_*}{\overset{h^*}{\rightleftarrows}} \Sigma\text{-}\mathbf{Set}_{\mathrm{o.f.}} \quad (h^* \dashv h_*)$$

$h^* \colon \Sigma\text{-}\mathbf{Set}_{\mathrm{o.f.}} \hookrightarrow \Sigma\text{-}\mathbf{Set}$ is just the embedding. Its right adjoint h_* sends a Σ-set (Q, δ) to its **maximum orbit-finite subobject**:

$$h_*(Q, \delta) = \{q \in Q \mid \text{ the orbit of } q \text{ is finite}\}.$$

The pointed topos $\Sigma\text{-}\mathbf{Set}_{\text{o.f.}}$ induces regular languages (2/2)

Theorem ($\Sigma\text{-}\mathbf{Set}_{\text{o.f.}}$ induces regular languages)

The canonical internal boolean algebra $p_{\text{o.f.}*}(\{\top, \bot\})$ for the pointed topos $p_{\text{o.f.}} : \mathbf{Set} \to \Sigma\text{-}\mathbf{Set}_{\text{o.f.}}$ is the boolean algebra of regular languages $\mathbf{Reg}$.

$$p_{\text{o.f.}*}(\{\top, \bot\}) \cong \mathbf{Reg}$$

Proof.

We have $p_{\text{o.f.}*}(\{\top, \bot\}) = h_*(p_*(\{\top, \bot\})) = h_*(\mathbf{Lan}) = \mathbf{Reg}$. The last equation is due to the Myhill-Nerode theorem (or the next theorem!). $\square$

It has 4 definitions! (1/3)

Theorem ([Hora 2024b] preprint)

*All of the following **pointed topoi** are mutually equivalent:*

- *The pointed topos* Σ-**Set**$_{\mathrm{o.f.}}$
- **Sh**$(\Sigma\text{-}\mathbf{FinSet}, J)$: *The sheaf topos over **finite** Σ**-sets.***
- **Sh**$(\Sigma\text{-}\mathbf{FinMon}, J)$: *The sheaf topos over **finite (**Σ**-)monoids.***
- **Cont**$(\widehat{\Sigma^*})$: *The continuous action topos of the profinite monoid* $\widehat{\Sigma^*}$ *of **profinite words**.*

These equivalences induce the isomorphisms between their canonical internal boolean algebras.

It has 4 definitions! (2/3)

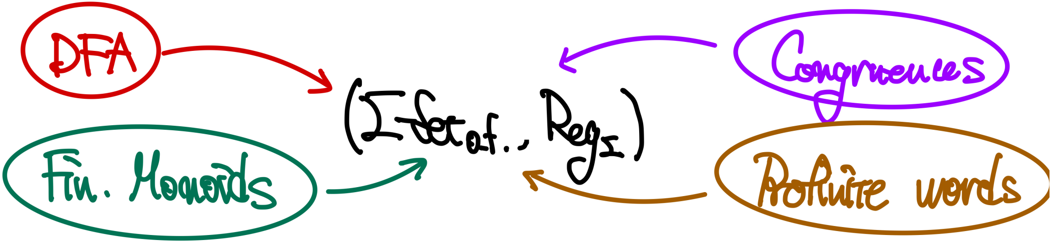
Corollary ([Hora 2024b] preprint)

All of the following ***Boolean algebra-ed topoi*** *are mutually equivalent:*

- $(\mathbf{\Sigma\text{-}Set}_{\mathrm{o.f.}}, \mathbf{Reg})$ *The orbit-finite automata with* ***orbit-finite (= regular) languages.***
- $(\mathbf{Sh}(\mathbf{\Sigma\text{-}FinSet}, J), \mathrm{DFA})$*: The sheaf topos over the (Boolean algebra-ed) site of* ***DFA****.*
- $(\mathbf{Sh}(\mathbf{\Sigma\text{-}FinMon}, J), \mathcal{P})$*: The sheaf topos over the (Boolean algebra-ed) site of* ***subsets of finite (Σ-) monoids.***
- $(\mathbf{Cont}(\widehat{\Sigma^*}), \mathrm{Clopen}(\widehat{\Sigma^*}))$*: The continuous action topos with the* ***clopen subsets of profinite words****.*

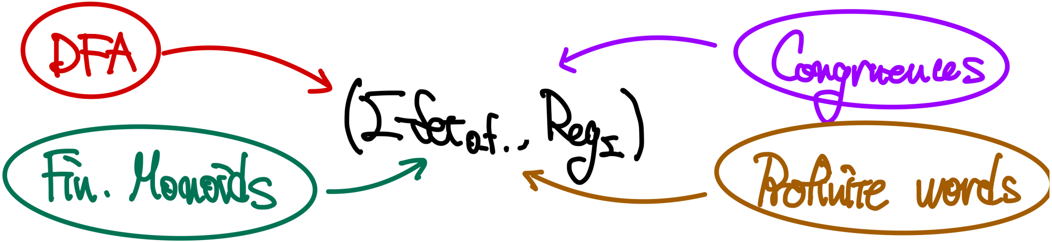
It has 4 definitions! (3/3)

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



It has 4 definitions! (3/3)

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



Question

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

This talk in one slide

Section 2 Preliminaries on Topos.

Section 3 **Languages emerge from a point**: exemplified by a topos Σ-**Set**.

Section 4 Regular language theory "$\simeq$" geometry of a topos Σ-**Set**$_{\mathrm{o.f.}}$.

Section 5 **Syntactic topos** $\mathcal{T}_C$ for each language class $C \subset$ **Lan**.

- $\mathcal{T}_{\{L\}} \simeq \mathbf{PSh}(\Sigma^*/\cong_L)$ for a regular language L and its **syntactic monoid** $\Sigma^*/\cong_L$.
- $\mathcal{T}_{\mathbf{Reg}} \simeq \mathbf{Cont}(\widehat{\Sigma^*})$ for the profinite monoid of **profinite words** $\widehat{\Sigma^*}$.

$\rightarrow$ the ***n*-th cohomology** $H^n(\mathcal{T}_C, A)$ for each $C \subset$ **Lan**

Section 6 The key tool is **the colimit of all monomorphisms** Ξ in Σ-**Set**, which consists of all **right congruences**.

Table of Contents

1. Motivation: Why topos?
2. Preliminaries: What's topos?
3. Σ-**Set**: Languages emerge from a point
4. Σ-**Set**$_{\mathrm{o.f.}}$: The topos of regular languages
5. $\mathcal{T}_C$: Syntactic topos of a language class
6. How to calculate $\mathcal{T}$: the Local state classifier Ξ

Overview of this section

We define a language class $\mathbb{L}_{\mathcal{F}}$ **recognized by a hyperconnected quotient** $\mathcal{F}$.
The **Syntactic topos** $\mathcal{T}_C$ is defined by the Galois connection:

$$\begin{tikzcd} \mathrm{HQ}(\Sigma\text{-}\mathbf{Set}) \arrow[r, shift right=2, "\mathbb{L}"'] \arrow[r, phantom, "\perp"] & \mathcal{P}(\mathbf{Lan}) \arrow[l, shift right=2, "\mathcal{T}"'] \end{tikzcd}$$

This construction $\mathcal{T}$ subsumes

- previous sections $\mathcal{T}_{\mathbf{Lan}} \simeq \Sigma\text{-}\mathbf{Set}$, $\mathcal{T}_{\mathbf{Reg}} \simeq \Sigma\text{-}\mathbf{Set}_{\mathrm{o.f.}}$, and
- the **syntactic monoid** of a regular language $\mathcal{T}_{\{L\}} \simeq \mathbf{PSh}(\Sigma^*/\cong_L)$.

We can start geometry here!
$\rightarrow$ the n**-th cohomology** $H^n(\mathcal{T}_C, A)$ for each $C \subset \mathbf{Lan}$

Idea: Hyperconnected quotients subsume monoid surjections

In algebraic language theory, we use surjective monoid homomorphisms.

Lemma (HQ subsumes surj. monoid hom. 1)

A surjective monoid homomorphism $\phi \colon \Sigma^* \twoheadrightarrow M$ *induces a hyperconnected quotient* $\phi \colon \Sigma\text{-}\mathbf{Set} \twoheadrightarrow \mathbf{PSh}(M)$.

$$\begin{tikzcd} \Sigma^* \arrow[r] \arrow[d, two heads, "\phi"'] & \mathrm{End}(Q) \\ M \arrow[ur, dashed] & \end{tikzcd}$$

We will do **algebraic language theory** replacing monoids **with hyperconnected quotients!**

Language recognition by hyperconnected quotient (1/3)

We define the notion of **language recognition** by a hyperconnected quotient.

Definition (A concrete definition of Language recognition by HQ)

A hyperconnected quotient $h \colon \Sigma\text{-}\mathbf{Set} \twoheadrightarrow \mathcal{F}$ **recognizes** a language L if there is an object $(Q, \delta) \in \mathrm{ob}(\mathcal{F})$ that recognizes L. We write $\mathbb{L}_{\mathcal{F}} \subset \mathbf{Lan}$ for the set of all languages recognized by $\Sigma\text{-}\mathbf{Set} \twoheadrightarrow \mathcal{F}$.

Example (Regular)

$\Sigma\text{-}\mathbf{Set} \twoheadrightarrow \Sigma\text{-}\mathbf{Set}_{\mathrm{o.f.}}$ recognizes L iff L is regular.

$$\mathbb{L}_{\Sigma\text{-}\mathbf{Set}_{\mathrm{o.f.}}} = \mathbf{Reg}$$

Language recognition by hyperconnected quotient (2/3)

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

Proposition (HQ subsumes surj. monoid hom. 2)

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

Language recognition by hyperconnected quotient (3/3)

What is the relationship with the point?

$$\begin{tikzcd} \mathbf{Set} \arrow[r, "p"] \arrow[rr, bend right, "p_{\mathcal{F}}"'] & \Sigma\text{-}\mathbf{Set} \arrow[r, two heads, "\text{hyperconn.}"] & \mathcal{F} \end{tikzcd}$$

Proposition (Myhill-Nerode theorem for each $\mathcal{F}$)

For any hyperconnected quotient $h\colon \Sigma\text{-}\mathbf{Set} \twoheadrightarrow \mathcal{F}$, the canonical internal Boolean algebra coincides with $\mathbb{L}_{\mathcal{F}}$

$$\mathbb{L}_{\mathcal{F}} = p_{\mathcal{F}*}(\{\top, \bot\}) = h_*\mathbf{Lan} \subset \mathbf{Lan}$$

This proves that $\mathbb{L}_{\mathcal{F}}$ is closed under derivatives and boolean operations.

The construction $\mathcal{T}$ (1/3): Definition

Definition (Syntactic topos $\mathcal{T}_C$ as the minimum topos)

The **syntactic topos**[4] $\mathcal{T}_C$ of a language class $C \subset \mathbf{Lan}$ is the minimum hyperconnected quotient that recognizes all languages in C.

Proposition (Existence)

For any language class $C \subset \mathbf{Lan}$, its syntactic topos $\mathcal{T}_C$ exists.

Thus we obtain the Galois connection:

$$\begin{tikzcd} \mathrm{HQ}(\Sigma\text{-}\mathbf{Set}) \arrow[rr, shift right, "\mathbb{L}"'] \arrow[rr, phantom, "\perp"] & & \mathcal{P}(\mathbf{Lan}) \arrow[ll, shift right, "\mathcal{T}"'] \end{tikzcd} .$$

[4] This can be defined syntactically with LSC.

The construction $\mathcal{T}$ (2/3): Syntactic monoid

Proposition (Syntactic topos subsumes syntactic monoid)

For any regular[5] *language L, we have*

$$\mathcal{T}_{\{L\}} = \mathbf{PSh}(\Sigma^*/\cong_L),$$

where $\Sigma^*/\cong_L$ is the syntactic monoid of L.

The proof is easier with the **local state classifier**, which is the theme of the next sectoion.

[5] This works for a broader class.

The construction $\mathcal{T}$ (3/3): Other examples

Example (Other examples)

For general Σ,

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

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

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

(To do: comparison with the free pro-**V** monoid)

Future work: Local variety theorem

Theorem (hyperconnected quotients of Σ-$\mathbf{Set}_{\mathrm{o.f.}}$)

There is an isomorphism of complete lattices between

- $\mathrm{HQ}(\Sigma\text{-}\mathbf{Set}_{\mathrm{o.f.}})$: *the complete lattice of hyperconnected quotients of* $\Sigma\text{-}\mathbf{Set}_{\mathrm{o.f.}}$.
- *The complete lattice of regular language classes* $C \subset \mathbf{Reg}$ *that are closed under*
  - *Derivative, and*
  - *Boolean operations.*

The latter is just **internal subalgebras** of **Reg**. This correspondence is in the same framework with the Galois correspondence for abelian extensions of fields.

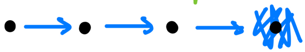
Future work: cohomology of language class, the easiest case

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.	$A_{\mathrm{odd/even}}$	A_{odd}	A_{even}	A_{odd}	A^f	∞
non-regular	0	0	0	$A/(1-f)$	A^f	1

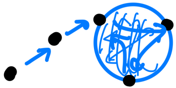
L: Star-free

like point



L: regular, not star-free

like $\mathbb{RP}^\infty$



L: non-regular

like S^1

This talk in one slide

Section 2 Preliminaries on Topos.

Section 3 **Languages emerge from a point**: exemplified by a topos Σ-**Set**.

Section 4 Regular language theory "$\simeq$" geometry of a topos Σ-**Set**$_{\text{o.f.}}$.

Section 5 **Syntactic topos** $\mathcal{T}_C$ for each language class $C \subset$ **Lan**.

- $\mathcal{T}_{\{L\}} \simeq \mathbf{PSh}(\Sigma^*/\cong_L)$ for a regular language L and its **syntactic monoid** $\Sigma^*/\cong_L$.
- $\mathcal{T}_{\mathbf{Reg}} \simeq \mathbf{Cont}(\widehat{\Sigma^*})$ for the profinite monoid of **profinite words** $\widehat{\Sigma^*}$.

$\to$ the ***n*-th cohomology** $H^n(\mathcal{T}_C, A)$ for each $C \subset$ **Lan**

Section 6 The key tool is **the colimit of all monomorphisms** Ξ in Σ-**Set**, which consists of all **right congruences**.

Table of Contents

1 Motivation: Why topos?

2 Preliminaries: What's topos?

3 Σ-**Set**: Languages emerge from a point

4 Σ-**Set**$_{\mathrm{o.f.}}$: The topos of regular languages

5 $\mathcal{T}_C$: Syntactic topos of a language class

6 How to calculate $\mathcal{T}$: the Local state classifier Ξ

Overview of this section

There are so many things to be calculated about $\mathcal{T}$. But how can we (and did I) calculate hyperconnected quotients?

- We will define a **local state classifier** Ξ of a topos as the **colimit of all monomorphisms**.
- Ξ has full information about hyperconnected quotients.
- In the topos Σ-**Set**, the LSC Ξ is the Σ-set of **right congruences**.
- Ξ connects hyperconnected quotients and word combinatorics.

How can we calculate $\mathcal{T} \dashv \mathbb{L}$? (1/2): LSC

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

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

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

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

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

How can we calculate $\mathcal{T} \dashv \mathbb{L}$? (2/2)

Theorem (Internal Parameterization of Hyperconnected Quotients [Hora 2024a])

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

- *hyperconnected quotients of $\mathcal{E}$, and*
- *Internal filters of the local state classifier Ξ.*

With this theorem, we can reduce our problem of **classifying all hyperconnected quotients** to a problem of **studying a single object** Ξ.

Right congruences form LSC Ξ (1/3)

Non-trivially, all topoi (including Σ-**Set**) have local state classifiers.
The local state classifier of Σ-**Set** is the third important Σ-set (following Σ^* and **Lan**).

Example (The Σ-set of right congruences Ξ)

Elements	Right congruences (= equivalence relations on Σ^* in Σ-**Set**)
Σ-action	$u\ (\sim * w)\ v \iff wu \sim wv.$
Universality	The local state classifier = colimit of all monomorphisms!

Right congruences form LSC Ξ (2/3)

For any Σ-set (Q, δ), we define a morphism

$$\xi_{(Q,\delta)} : (Q, \delta) \to \Xi$$

by sending $q \in Q$ to a right congruence $\xi_{(Q,\delta)}(q) \in \Xi$:

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

Theorem (LSC of Σ-Set)

*The Σ-set of right congruences Ξ, together with the cocone $\{\xi_{(Q,\delta)} : (Q, \delta) \to \Xi\}_{(Q,\delta)\in \mathrm{ob}(\Sigma\text{-}\mathbf{Set})}$ is the local state classifier of Σ-**Set**.*

Especially, $\xi_{\mathbf{Lan}}(L)$ coincides with the Nerode congruence!

Right congruences form LSC Ξ (3/3)

Corollary (Classification of hyperconnected quotients of Σ-**Set**)

There is a one-to-one correspondence between

- *hyperconnected quotients of* Σ-**Set**, *and*
- *Non-empty sets of right congruences* $F \subset \Xi$ *such that*
  1. F *is closed under the* Σ^**-action:* $\sim \in F \implies \sim * w \in F$
  2. F *is closed under finite infimums:* $\sim, \sim' \in F \implies \sim \wedge \sim' \in F$
  3. F *is upward closed:* $\sim \leq \sim'$ *and* $\sim \in F \implies \sim' \in F$

Example

The hyperconnected quotient $h \colon \Sigma\text{-}\mathbf{Set} \twoheadrightarrow \Sigma\text{-}\mathbf{Set}_{\text{o.f.}}$ corresponds to

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

Ξ connects concrete and abstract

LSC Ξ is the bridge connecting word combinatorics—such as the Nerode congruence and syntactic congruence—with hyperconnected quotients.
e.g.) For any hyperconn. quotient $h\colon \Sigma\text{-}\mathbf{Set} \twoheadrightarrow \mathcal{F}$, $h_*\colon \Sigma\text{-}\mathbf{Set} \to \mathcal{F}$ is given by

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

which subsumes the **Myhill-Nerode theorem:**

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

This talk in one slide

Section 2 Preliminaries on Topos.

Section 3 **Languages emerge from a point**: exemplified by a topos Σ-**Set**.

Section 4 Regular language theory "$\simeq$" geometry of a topos Σ-**Set**$_{\text{o.f.}}$.

Section 5 **Syntactic topos** $\mathcal{T}_C$ for each language class $C \subset$ **Lan**.

- $\mathcal{T}_{\{L\}} \simeq \mathbf{PSh}(\Sigma^*/\cong_L)$ for a regular language L and its **syntactic monoid** $\Sigma^*/\cong_L$.
- $\mathcal{T}_{\mathbf{Reg}} \simeq \mathbf{Cont}(\widehat{\Sigma^*})$ for the profinite monoid of **profinite words** $\widehat{\Sigma^*}$.

$\to$ the n**-th cohomology** $H^n(\mathcal{T}_C, A)$ for each $C \subset$ **Lan**

Section 6 The key tool is **the colimit of all monomorphisms** Ξ in Σ-**Set**, which consists of all **right congruences**.

References

- Anel, Mathieu and André Joyal (2021). "Topo-logie". In: *New Spaces in Mathematics: Formal and Conceptual Reflections* 1, pp. 155–257.
- Colcombet, Thomas and Daniela Petrişan (2020). "Automata minimization: a functorial approach". In: *Logical Methods in Computer Science* 16.
- 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.). *Topoi of automata II: Hyperconnected geometric morphisms, syntactic monoids, and language classes (ongoing)*.
- Iwaniack, Victor (2024). "Automata in W-Toposes, and General Myhill-Nerode Theorems". In: *International Workshop on Coalgebraic Methods in Computer Science*. Springer, pp. 93–113.
- Johnstone, Peter T. (1981). "Factorization theorems for geometric morphisms, I". In: *Cahiers de topologie et géométrie différentielle catégoriques* 22.1, pp. 3–17.
- Leinster, Tom (2024). "The eventual image". In: *Theory and Applications of Categories* 42.9, pp. 180–221.
- Melliès, Paul-André and Noam Zeilberger (2023). "The categorical contours of the Chomsky-Sch\" utzenberger representation theorem". In: *arXiv preprint arXiv:2405.14703*.
- Rogers, Morgan (2023). "Toposes of topological monoid actions". In: *Compositionality: the open-access journal for the mathematics of composition* 5.
- Uramoto, Takeo (2016). "Semi-galois categories I: the classical Eilenberg variety theory". In: *Proceedings of the 31st Annual ACM/IEEE Symposium on Logic in Computer Science*, pp. 545–554.
- — (2017). "Semi-galois Categories I: The Classical Eilenberg Variety Theory". Version 4. In: *arXiv:1512.04389*.