Grothendieck toposes and the geometry of language elaborations

Laurent Lafforgue

(Lagrange Center and Huawei Boulogne-Billancourt Research Center, France)

London CSTT Workshop, British Library, June 3-4, 2024

Note: this presentation has been largely inspired by exchanges with O. Caramello.

The core question of language elaboration:

Starting remarks:

- Mathematics does not consist only in deriving consequences from axioms in a given language.
- It also consists in enlarging or changing the language in order to get new insights.

Striking examples:

- Descartes' analytic geometry:
  from the language of ancient geometry to algebra.
- Newton's physics:
  from the language of time series to differential calculus.
- Galois' theory:
  from the language of algebraic equations
  to the language of symmetry groups and group actions.

AI problems as problems of language changes:

- LLM: coding texts as vectors in a way which makes approximation techniques efficient.
- Image recognition: from pixels to words.
- Deep learning: moving from an input language to an output language through mysterious intermediate layers.

A key difficulty and an overlooked question:

The difficulty of jumps:

Usually, we don’t move from a language to another language

in a continuous way nor even through easy intermediate steps.

Ex: Descartes, Newton, Galois made genius jumps.

Consequence:

DNN systems which would be “meaningful” are hard to imagine.

$\longrightarrow$ Maybe intermediate languages between

an input language and an output language do not exist?

$\longrightarrow$ Maybe an approximation process

such as gradient descent backward propagation cannot be meaningful?

The overlooked question of choosing a starting description language:

Ex: The language of pixels for representing images should be open to question.

$\longrightarrow$ For instance, could images be represented in terms of more or less precise qualitative descriptions of distinguishable contours and the connected components of their complement?

The necessity of formalized languages and their elements:

If we want machines to deal with some languages,

they have to be formalized languages, i.e. the type of languages used in mathematics.

Elements of mathematical languages:

• Vocabulary:

– names of "objects" i.e. of "spaces of variables" $G, F, V, A, B, \cdots$
– names of maps in a family of variables $f : A_1 \cdots A_n \rightarrow B$
– names of relations in a family of variables $R \rightarrowtail A_1 \cdots A_n$

• Substitution:

– replacing a variable x^B by a function $f(x_1^{A_1} \cdots x_n^{A_n})$

• Logical symbols allowing to form first-order formulas:

– truth, finite and infinite conjunctions $\top, \wedge, \bigwedge$

– false, finite and infinite disjunctions $\perp, \vee, \bigvee$

– negation $\neg$

– implication $\Rightarrow$

– existential and universal quantifiers $\exists, \forall$

• Formation of quotients by equivalence relations.

• Second-order constructions: $(A, B) \longmapsto B^A = \mathcal{H}om(A, B)$

$$A \longmapsto \mathcal{P}(A) = \Omega^A$$

• Interpretations: They always exist in Set and, more generally, in any topos $\mathcal{E}$.

Geometrization of logic:

Theorem (dating back to the 1970's). – *For any first-order theory $\mathbb{T}$ wich is "geometric" (meaning its axioms only use the symbols $\top, \wedge, \bot, \bigvee, \exists$), there exists an associated "topos" (= generalized space) $\mathcal{E}_{\mathbb{T}}$ such that*

$$\left\{\begin{array}{c}\text{interpretations} = \text{"models"} \\ M \\ \text{of } \mathbb{T} \text{ in a topos } \mathcal{E}\end{array}\right\} \longleftrightarrow \left\{\begin{array}{c}\text{"topos morphisms"} \\ (= \text{ generalized continuous maps}) \\ \mathcal{E} \to \mathcal{E}_{\mathbb{T}}\end{array}\right\}$$

$$\left\{\begin{array}{c}\text{change of parameters} \\ \text{for models} \\ M \text{ in } \mathcal{E} \mapsto f^* M \text{ in } \mathcal{E}'\end{array}\right\} \longleftrightarrow \left\{\begin{array}{c}\text{composition with topos morphisms} \\ f : \mathcal{E}' \to \mathcal{E} \\ (\mathcal{E} \to \mathcal{E}_{\mathbb{T}}) \mapsto (\mathcal{E}' \xrightarrow{f} \mathcal{E} \to \mathcal{E}_{\mathbb{T}})\end{array}\right\}$$

$$\left\{\begin{array}{c}\text{interpretation of a formula} \\ \varphi(x_1^{A_1} \cdots x_n^{A_n}) \\ \text{in a model } M \text{ of } \mathbb{T} \text{ in } \mathcal{E}\end{array}\right\} \longleftrightarrow \left\{\begin{array}{c}\text{embedding in } \mathcal{E} \\ M\varphi \hookrightarrow MA_1 \times \cdots \times MA_n\end{array}\right\}$$

Remarks:

- In particular, set-valued models M of $\mathbb{T}$ correspond to "points": point topos $\mathrm{Pt} = \{\text{topos of sets}\} \to \mathcal{E}_{\mathbb{T}}$.
- For any model M of $\mathbb{T}$ in $\mathcal{E}$, any geometric formula φ, and any $f : \mathcal{E}' \to \mathcal{E}$,

$$f^*(M\varphi) \xrightarrow{\sim} (f^* M)\varphi \, .$$

- For more general formulas, there is only a natural morphism in $\mathcal{E}$

$$f^*(M\varphi) \longrightarrow (f^* M)\varphi \, .$$

Semantics and geometry:

Definition (originally introduced by Tarski). –
The semantics of a theory $\mathbb{T}$ (considered as a syntactic object) consists in its "models" in $\{$sets$\}$ (and, more generally, in an arbitrary topos $\mathcal{E}$).

Corollary. – *For any first-order geometry theory $\mathbb{T}$, its semantics is incarnated by its associated topos $\mathcal{E}_{\mathbb{T}}$.*

Remarks:

- In particular, two theories $\mathbb{T}$ and $\mathbb{T}'$ are semantically equivalent if and only if

$$\mathcal{E}_{\mathbb{T}} \cong \mathcal{E}_{\mathbb{T}'} \,.$$

- A geometric theory $\mathbb{T}$ is contradictory if and only if

$$\mathcal{E}_{\mathbb{T}} = \emptyset \,.$$

A miracle of Topos Theory:

The semantics of any first-order geometric theory $\mathbb{T}$

is incarnated by a well-defined mathematical object $\mathcal{E}_{\mathbb{T}}$

which is of topological nature and is amenable to computation.

Back from semantics to syntax:

Theorem. – *Any topos morphism* $f : \mathcal{E}' \to \mathcal{E}$ *uniquely factorizes as*

$$\mathcal{E}' \underset{\text{“surjective”}}{\twoheadrightarrow} \mathrm{Im}(f) \underset{\text{embedding}}{\hookrightarrow} \mathcal{E}\,.$$

Remarks:

- As a consequence, there is a well-defined push-forward map

$$f_* : \{\text{subtoposes } \mathcal{E}'_1 \hookrightarrow \mathcal{E}'\} \to \{\text{subtoposes } \mathcal{E}_1 \hookrightarrow \mathcal{E}\},$$
$$(\mathcal{E}'_1 \hookrightarrow \mathcal{E}') \mapsto \mathrm{Im}(\mathcal{E}'_1 \hookrightarrow \mathcal{E}' \xrightarrow{f} \mathcal{E}).$$

- One can prove that there exists also a pull-back map

$$f_* = f^{-1} : \{\text{subtoposes } \mathcal{E}_1 \hookrightarrow \mathcal{E}\} \to \{\text{subtoposes } \mathcal{E}'_1 \hookrightarrow \mathcal{E}'\}$$
$$(\mathcal{E}_1 \hookrightarrow \mathcal{E}) \mapsto (f^{-1}\mathcal{E}_1 \hookrightarrow \mathcal{E}')$$

characterized by $\quad \mathcal{E}_1 \supseteq f_*(\mathcal{E}'_1) \Leftrightarrow f^{-1}\mathcal{E}_1 \supseteq \mathcal{E}'_1.$

Theorem (O. Caramello). – *For any geometric theory* $\mathbb{T}$, *subtoposes* $\mathcal{E} \hookrightarrow \mathcal{E}_{\mathbb{T}}$ *correspond to “quotient” theories* $\mathbb{T}'$ *derived from* $\mathbb{T}$ *by adding extra axioms.*

Consequence: For any model M of $\mathbb{T}$ in a topos $\mathcal{E}$, corresponding to $\mathcal{E} \xrightarrow{m} \mathcal{E}_{\mathbb{T}}$,

$\mathrm{Im}(m) \hookrightarrow \mathcal{E}_{\mathbb{T}}$ corresponds to a quotient theory $\mathbb{T}'$

which can be called a syntactic description of M.

How to represent natural families of data?

• If we want to process natural families of data (ex: images),

we first need to figure out

to which type of mathematics objects they should correspond.

• On the basis of classical practice,

the first idea would be to represent data as points of some spaces,

in particular as vectors of some (high dim.) linear spaces.

Objection:

If we think in the more general terms of toposes,

points $\mathrm{Pt} \longrightarrow \mathcal{E}_{\mathbb{T}}$ or $\mathcal{E} \longrightarrow \mathcal{E}_{\mathbb{T}}$

correspond to "models" of geometric theories $\mathbb{T}$.

They are of semantic nature, whereas stored data should be syntactic.

Proposed alternative:

Represent natural families of data

as families of subtoposes of a given topos.

Reasons for representing data as subtoposes:

First reason: syntactic expression. –
For any equivalence $\mathcal{E} \cong \mathcal{E}_{\mathbb{T}}$, the purely geometric notion of subtopos $\mathcal{E}_1 \hookrightarrow \mathcal{E}$ corresponds one-to-one to the purely syntactic notion of "quotient" theory $\mathbb{T}_1$ of $\mathbb{T}$.

Second reason: topological expression. –
The notion of subtopos has other expressions.
For any equivalence $\mathcal{E} \cong \widehat{\mathcal{C}}_J =$ topos of "sheaves" on a small category $\mathcal{C}$ endowed with a "topology" J, subtoposes $\mathcal{E}_1 \hookrightarrow \mathcal{E}$ correspond one-to-one to "topologies" J_1 on $\mathcal{C}$ which refine J.

Third reason: amenability to geometric processing. –
Any "geometric" correspondence between toposes

$$\begin{array}{ccc} & \mathcal{E}_{\Gamma} & \\ {}^{p}\swarrow & & \searrow^{q} \\ \mathcal{E}_{\mathbb{T}_0} & & \mathcal{E}_{\mathbb{T}_1} \end{array}$$

transforms subtoposes $\mathcal{E}_0 \hookrightarrow \mathcal{E}_{\mathbb{T}_0}$
into subtoposes $q_*(p^{-1}\mathcal{E}_0) \hookrightarrow \mathcal{E}_{\mathbb{T}_1}$.

A possible general form of topos-theoretic deep learning:

Any chain of correspondences

$$\begin{array}{ccccccccc} & & \mathcal{E}_{\Gamma_1} & & & & \mathcal{E}_{\Gamma_n} & & \\ & \swarrow & & \searrow & & \swarrow & & \searrow & \\ \mathcal{E}_{\mathbb{T}_0} & & & & \mathcal{E}_{\mathbb{T}_1} \quad \cdots \quad \mathcal{E}_{\mathbb{T}_{n-1}} & & & & \mathcal{E}_{\mathbb{T}_n} \end{array}$$

induces a processing machine

$$\left\{\begin{matrix}\text{subtoposes} \\ \text{of } \mathcal{E}_{\mathbb{T}_0}\end{matrix}\right\} \rightarrow \left\{\begin{matrix}\text{subtoposes} \\ \text{of } \mathcal{E}_{\mathbb{T}_1}\end{matrix}\right\} \rightarrow \cdots \rightarrow \left\{\begin{matrix}\text{subtoposes} \\ \text{of } \mathcal{E}_{\mathbb{T}_n}\end{matrix}\right\}$$

or, equivalently

$$\left\{\begin{matrix}\text{quotient} \\ \text{theories} \\ \text{of } \mathbb{T}_0\end{matrix}\right\} \rightarrow \left\{\begin{matrix}\text{quotient} \\ \text{theories} \\ \text{of } \mathbb{T}_1\end{matrix}\right\} \rightarrow \cdots \rightarrow \left\{\begin{matrix}\text{quotient} \\ \text{theories} \\ \text{of } \mathbb{T}_n\end{matrix}\right\}.$$

Remark:

• In such a scheme, the bottom line $\mathcal{E}_{\mathbb{T}_0}, \mathcal{E}_{\mathbb{T}_1}, \cdots, \mathcal{E}_{\mathbb{T}_n}$
should be understood syntactically
while the upper line $\mathcal{E}_{\Gamma_1}, \cdots, \mathcal{E}_{\Gamma_n}$ should be understood semantically:
each $\mathcal{E}_{\Gamma_i}$ carries simultaneously two model structures of $\mathbb{T}_{1-i}$ and $\mathbb{T}_i$.

Fundamental questions:

First question: the starting description language

How to choose a starting description theory $\mathbb{T}_0$

for the family of data under consideration?

Remark:

If the data in such a natural family are to be represented as subtoposes of $\mathcal{E}_{\mathbb{T}_0}$,

$\mathbb{T}_0$ should not be a “theory of this type of data”

but a “theory of viewpoints” on this type of data.

Second question: geometric language elaboration

How to elaborate from an already constructed description language $\mathbb{T}_i$

a deeper (or better fitted for our objectives) description language $\mathbb{T}_{i+1}$

related to $\mathbb{T}_i$ through a double intertwined model structure:

$$\mathcal{E}_{\mathbb{T}_{i-1}} \longleftarrow \mathcal{E}_{\Gamma_i} \longrightarrow \mathcal{E}_{\mathbb{T}_i}$$

Tying series of data through a joint description vocabulary:

Basic facts relating formalized languages and geometry. –

(1) Any first-order geometric theory $\mathbb{T}$ *consists in*
a vocabulary Σ *and a family of axioms* $\varphi(x_1^{A_1}, \cdots, x_n^{A_n}) \vdash \psi(x_1^{A_1}, \cdots, x_n^{A_n})$.

(2) Any vocabulary Σ *defines a "category" (= "mathematical country" consisting in: cities + itineraries + composition law of itineraries)*
$\mathcal{C}_\Sigma$ *whose "cities" and "itineraries" are "formulas" (= sentences in the vocabulary* Σ*).*

(3) This category $\mathcal{C}_\Sigma$ *defines the topos* $\mathcal{E}_\Sigma = \widehat{\mathcal{C}}_\Sigma$.

(4) Choosing axioms to define $\mathbb{T}$ *from* Σ *is equivalent to*

$$\begin{cases} - & \text{choosing a subtopos } \mathcal{E}_\mathbb{T} \hookrightarrow \mathcal{E}_\Sigma = \widehat{\mathcal{C}}_\Sigma, \\ - & \text{choosing a "topology" } J_\mathbb{T} \text{ on the "category" } \mathcal{C}_\Sigma. \end{cases}$$

Suppose we want to introduce a starting description vocabulary Σ_0
for a natural family of data (ex. images, plane configurations, algebraic equations, $\cdots$)

$\rightarrow$ Start with a family of concrete instances $i \in I$, each represented by a description vocabulary V_i supplemented by conditions expressed in this vocabulary.

$\rightarrow$ The fact that all $i \in I$ belong to a natural family should allow to choose a "joint description vocabulary" Σ_0 endowed with "naming functors"

$$\mathcal{C}_{V_i} \longrightarrow \mathcal{C}_{\Sigma_0}, \quad \forall\, i \in I.$$

A principle for inductive reasoning and syntactic learning:

- Suppose we are given a series of concrete instances $i \in I$ of a natural family of data.
- Suppose each instance $i \in I$ is described by conditions expressed in a vocabulary V_i, which can equivalently be thought of as

$$\begin{cases} - & \text{a topology } J_i \text{ on } \mathcal{C}_{V_i}, \\ - & \text{a subtopos } \quad \widehat{(\mathcal{C}_{V_i})}_{J_i} \hookrightarrow \widehat{\mathcal{C}_{V_i}} = \mathcal{E}_{V_i}. \end{cases}$$

- Suppose the instances $i \in I$ are related by a "joint description vocabulary" Σ_0 and naming functors

$$\mathcal{C}_{V_i} \longrightarrow \mathcal{C}_{\Sigma_0}, \quad i \in I,$$

inducing topos morphisms $\quad \mathcal{E}_{V_i} = \widehat{\mathcal{C}_{V_i}} \xrightarrow{e_i} \widehat{\mathcal{C}_{\Sigma_0}}, \; i \in I.$

Principle of inductive reasoning:

The starting description theory $\mathbb{T}_0$ in the vocabulary Σ_0 should be "as economical as possible" under the constraint that

$$\begin{cases} \text{for any } i \in I, \text{ the pull-back map } e_i^{-1} \text{ by } \widehat{\mathcal{C}_{V_i}} \xrightarrow{e_i} \widehat{\mathcal{C}_{\Sigma_0}}, \\ \text{should verify} \quad e_i^{-1}(\mathcal{E}_{\mathbb{T}_0} \hookrightarrow \widehat{\mathcal{C}}_{\Sigma_0}) \subseteq (\widehat{(\mathcal{C}_{V_i})}_{J_i} \hookrightarrow \widehat{\mathcal{C}_{V_i}}) \end{cases}$$

Remark: Pull-back maps always respect finite unions.

A principle for elaborating chains of "higher" description languages:

Question: If a "description language" $\mathbb{T}_i$ is already constructed, how to derive from $\mathbb{T}_i$ "higher description languages" $\mathbb{T}_{i+1}$ related to $\mathbb{T}_i$ by geometric correspondences:

$$\mathcal{E}_{\mathbb{T}_i} \xleftarrow{\;p_i\;} \mathcal{E}_{\Gamma_i} \xrightarrow{\;q_i\;} \mathcal{E}_{\mathbb{T}_{i+1}}$$

Remark: It may happen that p_i, q_i or both are equivalences.

Even that case can be very deep.

Proposed process:

$\longrightarrow$ Consider different models Γ_i of $\mathbb{T}_i$ in toposes $\mathcal{E}_{\Gamma_i}$,
or equivalently different topos morphisms $\mathcal{E}_{\Gamma_i} \xrightarrow{p_i} \mathcal{E}_{\mathbb{T}_i}$.

$\longrightarrow$ Consider higher order constructions built from Γ_i in $\mathcal{E}_{\Gamma_i}$
(e.g. symmetry groups or, more generally, global invariants).

$\longrightarrow$ Recognize that these higher-order constructions
are models of some other first-order geometric theory $\mathbb{T}_{i+1}$.

$\longrightarrow$ Choose the model $\mathcal{E}_{\Gamma_i} \xrightarrow{p_i} \mathcal{E}_{\mathbb{T}_i}$ and the higher-order construction
so that the induced correspondence $\mathcal{E}_{\mathbb{T}_i} \xleftarrow{p_i} \mathcal{E}_{\Gamma_i} \xrightarrow{q_i} \mathcal{E}_{\mathbb{T}_{i+1}}$ is best fitted.

Examples:

- **Descartes' equivalence:**

  Start with the theory $\mathbb{T}$ of affine planes.

  Consider its "universal model" U in $\mathcal{E}_{\mathbb{T}}$.

  Consider the associated groups of "translations" and "dilatations" of U and the associated field structure on lines endowed with two points.

$\rightarrow$ This induces an equivalence $\quad \mathcal{E}_{\mathbb{T}} \cong \mathcal{E}_{\mathbb{T}'}$ $\quad$ if $\quad \mathbb{T}' =$ theory of fields.

- **Differential calculus:**

  Start with a theory $\mathbb{T}$ of numbers. Consider a "complete" model $\mathbb{R} : \mathcal{E}_{\mathbb{R}} \to \mathcal{E}_{\mathbb{T}}$.

  Construct in $\mathcal{E}_{\mathbb{R}}$ the inner space of functions $\mathcal{H}om(\mathbb{R}, \mathbb{R})$ and define subspaces of "differentiable" and "integrable" functions, yielding a topos morphism $\mathcal{E}_{\text{diff}} \to \mathcal{E}_{\mathbb{R}}$.

  Derive the algebraic rules of differential calculus (linearity, Leibnitz' rule, integration of derivatives, change of variables) defining a theory $\mathbb{T}'$ endowed with a model structure $\quad \mathcal{E}_{\text{diff}} \longrightarrow \mathcal{E}_{\mathbb{T}'}$.

Examples:

- **Galois' equivalence:**
  Start with the theory $\mathbb{T}$ of algebraic extensions of fields,
  endowed with $\mathcal{E}_{\mathbb{T}} \to \mathcal{E}_{\mathbb{B}}$ for $\mathbb{B} =$ theory of fields.
  Consider any model $k : \mathrm{Pt} \to \mathcal{E}_{\mathbb{B}}$, i.e. any field k, and the fiber product of toposes

$$\begin{array}{ccc} \mathcal{E}_{\mathbb{T}_k} & \longrightarrow & \mathcal{E}_{\mathbb{T}} \\ \downarrow & & \downarrow \\ \mathrm{Pt} & \xrightarrow{k} & \mathcal{E}_{\mathbb{B}} \end{array}$$

  where $\mathbb{T}_k =$ theory of algebraic extensions of k.

  Choose a separable closure $\overline{k}$ of k, considered as a point $\overline{k} : \mathrm{Pt} \longrightarrow \mathcal{E}_{\mathbb{T}_k}$.
  Consider the associated group G of symmetries of $\overline{k} : \mathrm{Pt} \to \mathcal{E}_{\mathbb{T}_k}$
  and the associated theory $\mathbb{T}_G$ of principal G-actions, yielding a topos embedding

$$\mathcal{E}_{\mathbb{T}_G} \hookrightarrow \mathcal{E}_{\mathbb{T}_k} .$$

- **An automatic system for analyzing time series inspired by O. Caramello's topos-theoretic ideas:**
  Starting with a theory $\mathbb{T}_0$ of "viewpoints" on some type of time series,
  a software company has constructed a chain of theories $\mathbb{T}_1, \cdots, \mathbb{T}_n$
  where each $\mathbb{T}_i$ is a theory of "higher order viewpoints" on $\mathbb{T}_{i-1}$.

Remark: So far, the length of the chain is already $n \geq 10$.