Ontologies, knowledge representations and Grothendieck toposes

Olivia Caramello[1] Laurent Lafforgue[2]

[1]University of Insubria - Como and IHÉS
[2]Huawei Paris Research Center

Semantics workshop, Lagrange Center, Huawei,
3-4 February 2022

Aim of the talk

The purpose of this double talk is to introduce Grothendieck toposes and
the duality between toposes and their presentations
starting from the article

Knowledge Representations and Ontologies: Logics, Ontologies and Semantic Web Languages

by S. Grimm, P. Hitzler and A. Abecker.

More specifically, we shall start from quotes of this text to propose an interpretation

"ontologies" $\longrightarrow$ toposes,

"knowledge representations" $\longrightarrow$ presentations of toposes (e.g. geometric ones (sites) or linguistic ones (theories))

Ontologies and toposes

"Ontologies are conceptual models of what "exists" in some domain, brought into machine-interpretable form by means of knowledge representation techniques."

We propose to mathematically embody

- ontologies (in this sense) by Grothendieck toposes,
- knowledge representation techniques by ways of defining and studying toposes (geometrically or linguistically).

The categorical and classifying nature of toposes

"Ontological categories provide a means to classify all existing things."

- Toposes are a special kind of categories:
  - those that are built in a certain way, or, equivalently,
  - those which satisfy certain axioms.
- Toposes are characterized by the fact that they classify objects of a certain kind: their "points" are the "models" of a certain language (= vocabulary + grammar rules), namely, all the particular objects to which a certain language applies.
- In this sense,
  - every topos is classifying for a language (in fact, for infinitely many languages),
  - every language is classified by a topos.

Toposes as universal categories

"In ontology, categories are also referred to as universals, and the concrete things that they serve to classify are referred to as particulars."

- It is possible to express the mathematical definition of topos by saying that toposes are the categories which, in a precise sense, are universal.
- The "points" of a topos can be seen as the concrete realisations, or the particularisations of some abstract language.
- A topos is classifying in the sense that no concrete realisation of that language, no particularisation, is forgotten: all appear as "points" of the topos. It is precisely in this sense that the qualification "universal" applies.

The categorical structure of toposes

"The systematic organisation of such categories allows to analyse the world that is made up by these things in a structured way."

- Every topos is a universe (that is, a whole, complete in a precise sense) which has the structure of a category (in the mathematical sense of the word).
- Different toposes also entertain mutual relations.
- Every topos is organized as a category, and toposes considered in their mutual relationships form an organisation which is defined and studied by means of the mathematical theory of categories.

The visual part of the notion of category

"A semantic network is a graph whose nodes represent concepts and whose arcs represent relations between these concepts."

The visual part of a category (in the mathematical sense of the word) is an oriented graph, consisting in

- "nodes" or "objects";
- "arcs" or "edges" or "arrows" relating the nodes:

- Every edge goes from a node to a node;
  it can go from a node to itself or to another node.
- Between two nodes there can be no edges,
  or a single edge, or several (possibly infinitely many) edges,
  in each of the two directions.

The non-visual part of the notion of category

"Not all the information in an ontology can be visualized in a graph."

A topos, and more generally a category (in the mathematical sense of the word), consists in:

- a visual part, which is an oriented graph, consisting of nodes and edges,

- a non-visual, deeper part, which is a way of associating to each pair of edges which follow one another

$$X \xrightarrow{f} Y \xrightarrow{g} Z$$

a "composed" edge

$$X \xrightarrow{g \circ f} Z .$$

The non-visual part of the notion of category

One requires that:

• for any triplet of edges

$$X \xrightarrow{f} Y \xrightarrow{g} Z \xrightarrow{h} W$$

which follow one another, one has the <u>"associativity" property</u>

$$h \circ (g \circ f) = (h \circ g) \circ f;$$

• every node X has a <u>unique "identity" edge</u>

$$\mathrm{id}_X \circlearrowright X$$

such that for any edges $f : Y \to X$ and $g : X \to Z$,

$$\mathrm{id}_X \circ f = f \qquad \text{and} \qquad g \circ \mathrm{id}_X = g\ .$$

The choice of a topos and of a geometric sketch of a topos

"The choice of ontological categories and particular objects determines the things about which knowledge can be represented in a complex system."

- The choice of a certain topos to study embodies the choice of a certain mathematical content which forms a world by itself.
- The choice of a family of particular objects of a topos defines a sort of geometrical sketch of this topos.
  This geometrical sketch is the category
  - whose nodes are the chosen objects of the topos,
  - whose edges are the arrows of the topos between the chosen objects,
  - the way of composing these edges is that of the topos.
- The choice of such a sketch allows one to represent, in a way more or less apt for computation,
  - a part at least of the other objects of the topos, and of its arrows,
  - a part at least of the information that can be theoretically extracted from the topos.

The choice of languages for describing a topos

"Ontology provides the labels for nodes and edges in a semantic network or the names for predicates and constants in rules or logical formulas, that constitute an ontological vocabulary."

- The choice of particular objects of a topos and of names for these chosen objects and for the edges which connect them defines a vocabulary which can be qualified as "ontological " in the sense that it is appropriate for describing the topos.
- The structure of the topos induces grammar rules on this vocabulary.
- The set consisting of an ontological vocabulary and the set of grammar rules induced on it by the structure of the topos defines a language.
- If the family of chosen objects of the topos is sufficiently rich, the classifying topos of that language is the original topos. In other words, this language determines the topos by means of the description of its points as "models".

Which languages do appear in the description of toposes

"The most prominent and fundamental logical formalism classically used for knowledge representation is first-order logic... Its roots can be traced back to the ancient Greek philosopher Aristotle."

- Choosing in any topos a sufficiently rich family of particular objects, with names for these objects and for the arrows connecting them in the topos, and with the grammar rules induced by the structure of the topos,
  defines a language which, according to the terminology of mathematical logic, is a "first-order theory".
- Conversely, every first-order theory (whose axioms are presented in a certain form - namely, within the framework of "geometric logic") has a unique "classifying topos", whose "points" can be interpreted as the "models" of this theory, that is, as the concrete realisations to which the vocabulary of the theory applies and which satisfy all the grammar rules.

Which relation between geometric sketches and languages?

"By defining "what exists", an ontology determines the things that can be predicated about. The terms of the ontological vocabulary are then used to represent knowledge, forming statements about the domain."

- For any "first-order theory" (whose axioms are presented in geometric form), any "first-order formula" expressed in the language of this theory defines an object of its classifying topos.
- Conversely, every object of the classifying topos can be obtained by "glueing" of finite or infinite families of such objects defined by formulas.
- Every object of the classifying topos can actually be presented in infinitely many ways (varying from more or less simple to very complex) in terms of the theory. The "glueing" rules and the equivalences between different presentations result from the "axioms" of the theory (which are the "grammar rules" of that language).

A geometry as flexible and expressive as logic

"First-order (predicate) logic is the prevalent and single most important knowledge representation system."

- We said that the choice in a topos of a (sufficiently rich) family of particular objects defines a "first-order theory" which suffices to determine the topos.
- On the contrary, the category associated with the family of objects as described above does not suffice in general to reconstruct the topos.
- To reconstruct it, one has to take into account a supplementary *datum*, called a "Grothendieck topology".
- A Grothendieck topology is a coherent family of "glueing rules".
- Conversely, every "site" consisting of
  - a category,
  - a Grothendieck topology on that category

defines a topos.

Knowledge representations by means of (logical) theories

"Rule-based knowledge representations systems are especially suitable for reasoning about concrete instance data."

- Every topos mathematically embodies a certain domain of reality, susceptible of becoming an object of knowledge.
- The instantiations of this reality, that is, its particular manifestations, appear geometrically as the points of that topos.
- Every presentation of this topos as classifying topos for a "first-order theory" provides a system for representing and elaborating knowledge about the reality embodied by the topos.
- In particular, the vocabulary of that theory applies to the points of that topos and allows one to reason about them: they satisfy the axioms of the theory and the properties which can be deduced from them by applying the general rules of logical reasoning.

The equivalence of syntax and semantics in toposes

"Logical consequence and universal truth can be described in terms of model-theoretic semantics."

- Every topos can be considered and studied syntactically, geometrically and semantically:
  - syntactically through any "first-order theory" which it classifies,
  - geometrically through its structure as a category, consisting of objects, arrows and a law of composition for arrows,
  - semantically through its "points", interpreted as the "models" of any theory which it classifies.
- In fact, one can define in the topos itself a "universal model" from which all the particular models of the theory can be obtained by instantiation.
- Any property expressed in the language of a theory classified by the topos is provable (i.e. it is a logical consequence of the axioms) if and only if it is semantically verified in the universal model lying in the topos.

Geometric interpretation of logical implications

"The most prominent type of relation in semantic networks is that of subsumption... Subsumption is associated with the notion of inheritance in that a specialized concept inherits all the properties from its more general parent concepts."

- Every time that a topos is presented as the classifying topos of a certain theory, the relations of implication between formulas written in the language of the theory have a geometric translation in the semantic context provided by the topos and the "universal model" of the theory inside it.
- In fact, the logical implications which are provable in the theory correspond precisely to the inclusion (or subsumption) relations in the sense of the categorical structure of the topos.

The articulation between the particular and the universal in toposes

"The semantic network... illustrates the distinction between general concepts... and individual concepts... The particular relation which links individuals to their classes is that of instantiation."

- A topos deserves the name of "semantic network" since it has the structure of a category, it contains a "universal model" of any theory which it classifies, and every formula (or sentence) expressed in the language of that theory interprets in that universal model.
- The relation between the "universal model" of such a theory and its particular models, which can be interpreted semantically as the "points" of the topos, is that of instantiation.

The articulation between reasoning and intuition in toposes

"The concepts and relations in an ontology can be intuitively grasped by humans, as they correspond to the elements in our mental model."

- The linguistic presentation of a topos as the classifying topos of one theory or another
  allows one to reason about the topos by applying to it the elements of vocabulary of the given theory and deducing, by using logic, consequences from its axioms or grammar rules.

- On the other hand, the geometric presentation of a topos by a "site", that is a category of semantic nature endowed with a Grothendieck topology (that is, a coherent family of glueing rules), is closer to our mental intuition.

The multiplicity of points of view

"Ontologies have been explored from different points of view."

- The multiplicity and diversity of the presentations of any topos by "first-order theories" which it classifies or by geometric "sites" of definition is a mathematical embodiment of the multiplicity and diversity of the ways to talk about a subject and the different points of view on it.
- There is no linguistic or geometric presentation which is better, in an absolute sense, than all the others. Certain presentations can be more convenient for extracting from the topos a certain kind of information and less adapted otherwise.

Toposes as incarnations of ontological invariants

"An ontology is a formal explicit specification of a shared conceptualization of a domain of interest...
An ontology reflects an agreement
on a domain conceptualization in a community."

- Every topos mathematically embodies a certain domain of common reality, susceptible of being described or seen by means of a multiplicity of different languages or geometric presentations. As such, it is the fundamental ontological invariant for that reality.
- Once it is discovered that different linguistic or geometric presentations define the same topos (up to equivalence), this common topos can be used as a "bridge" connecting them.
- This is the basic principle of the theory of "toposes as bridges" (developed by O.C. since 2010).

The duality between geometric presentations and linguistic descriptions

"Ontologies appear in different forms related to the forms of knowledge representation...
A knowledge engineer views an ontology by means of some graphical or formal visualisation while for storage or transfer it is encoded in an ontology language.
A reasoner, in turn, interprets an ontology as a set of axioms that constitute a logical theory."

- The choice of a presentation of a topos by a "site" (that is, by a geometric sketch consisting of a category endowed with a Grothendieck topology) is a way of visualizing the topos in part geometrically (the oriented graph underlying the category) and in part formally (the law of composition for edges).
- On the other hand, the choice of a description language, that is of a theory which the topos "classifies", is the determination of a sufficiently rich vocabulary for distinguishing the nodes and edges or the topos from each other, and of a list of grammar rules sufficiently complete to faithfully represent its categorical structure.

Some crucial elements of geometric visualizations

"The visualization... presents to the knowledge engineer a taxonomy, i.e. a subsumption hierarchy. In the visualisation, the knowledge engineer can also see conceptual relations as edges pointing from the domain concept to the range concept."

- The choice of a presentation of a topos by a "site" consisting of
  - chosen nodes (or objects) of the topos,
  - chosen edges (or arrows) of the topos,
  - the law of composition of these edges,
  - a coherent family of "glueing" rules,

  makes it appear very clearly the edges which occur in this presentation. Once interpreted in the context of a theory classified by the topos, they correspond to conceptual relations between concepts.
- Among the most important kinds of edges, there are in particular inclusions or subsumptions. Once interpreted logically in a theory classified by the topos, they correspond to relations of logical implicitation.

Knowledge representation techniques

"An ontology is expressed in a knowledge representation language that provides a formal semantics...
The specification of domain knowledge in an ontology is machine-representable and is being interpreted in a well-defined way. The techniques of knowledge representation help to realize this aspect."

- A topos is always concretely exhibited through one or more presentations, e.g. through a theory (= vocabulary + axioms) which it classifies or a site (= oriented graph + composition law + Grothendieck topology) with which it is associated.
- These presentations of linguistic or geometric nature are well-adapted for symbolic or formal computation.
- The knowledge representation techniques consist in expressing pieces of information that are susceptible of being extracted from the topos, in terms of theories or sites which describe the semantics or sketch the geometry of the topos.
- In this sense, topos theory can be regarded as a sort of "genetics of Mathematics".

The choice of the presentations according to the posed questions

"For some more detailed information, such as complex axioms and restrictions on concepts, there does not exist any appropriate visualisation paradigm other than expositing such fragments of the ontology in a formal language... Ontology engineering environments usually provide extra means for displaying and editing complex axiomatic information using a special-purpose ontology language or logical formal notation."

- If, for instance, one considers a fragment of a topos consisting in a family of objects and arrows, one can choose to present this topos by means of sites whose underlying oriented graphs contain this fragment.
- In the same situation, one can choose to describe the topos by means of theories whose vocabulary and axioms are sufficiently rich as for the elements of the fragment to be entirely expressible in terms of formulas written in the language of such theories.

Extraction and transfer of information, or the theory of "toposes as bridges"

"An ontology can be used as a schema for data-intensive instance retrieval on large knowledge or data bases."

- A topos mathematically embodies a certain "reality", which is a world by itself and hence cannot be entirely known.
- On the contrary, it is possible to try to know, that is, to express or calculate, topos-theoretic "invariants": the invariants supported by a topos are types of partial information which are susceptible of being extracted from it.
- Extracting from a topos the kind of partial information coded by a certain "invariant" consists in expressing or calculating that invariant in terms of its different presentations (e.g., by sites or theories).
- When different presentations of a given topos are into play, expressing or calculating an "invariant" in terms of such presentations realizes a transfer of information between them.
- This is the other key principle of the theory of "toposes as bridges".

Automatising transfer-of-information processes

"An application can make intensive use
of automated reasoning techniques
in order to derive implicit knowledge."

- Often, the expression or calculation of topos-theoretic invariants in terms of presentations of toposes of geometric nature (e.g. sites) or linguistic nature (e.g. theories) has an element of automatism inherent to it.

- In fact, in several situations, this expression or calculation of topos-theoretic invariants in the "concrete" terms of their presentations can be realised by means of algorithms.

The multiplicity of toposes and their relations

"The specifications in an ontology are limited to knowledge about a particular domain of interest. The explicit specification of domain knowledge can be modularised and expressed using different ontologies... An ontology is a piece of knowledge that can be used as a knowledge-based application among other pieces of knowledge."

- There are infinitely many different toposes which mathematically embody an infinite number of "parts of reality".
- Toposes are not isolated one from one another but related by means of a mathematical notion of "morphism" of toposes.
- A morphism from a topos to another can mathematically embody the natural relations between two "parts of reality".
- A network of toposes and morphisms relating them can serve for modelling and defining processes of treatment of information such as those arising in the context of the work of J.-C. Belfiore and D. Bennequin on Deep Neural Networks.

A mathematical morphogenesis

- The essential ambiguity given by the fact that any topos is associated in general with an infinite number of theories or different sites allows to study the relations between different theories, and hence the theories themselves, by using toposes as 'bridges' between these different presentations.

- Every topos-theoretic invariant generates a veritable mathematical morphogenesis resulting from its expression in terms of different representations of toposes, which gives rise in general to connections between properties or notions that are completely different and apparently unrelated from each other.

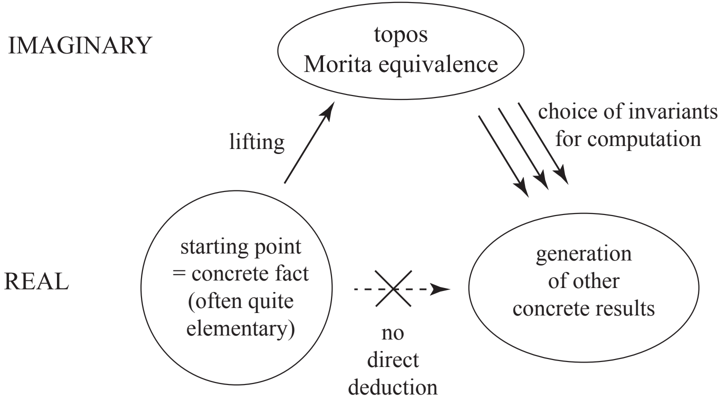
The duality between 'real' and 'imaginary'

- The passage from a site (or a theory) to the associated topos can be regarded as a sort of 'completion' by the addition of 'imaginaries' (in the model-theoretic sense), which materializes the potential contained in the site (or theory).

- The duality between the (relatively) unstructured world of presentations of theories and the maximally structured world of toposes is of great relevance as, on the one hand, the 'simplicity' and concreteness of theories or sites makes it easy to manipulate them, while, on the other hand, computations are much easier in the 'imaginary' world of toposes thanks to their very rich internal structure and the fact that invariants live at this level.

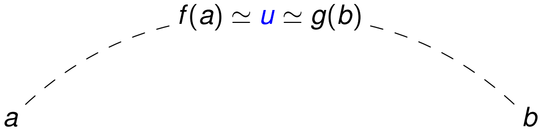
Leaps into the 'imaginary'

We can thus schematically represent the way of obtaining concrete results by applying the 'bridge' technique in the form of an ascent followed by a descent between two levels, the 'real' one of concrete mathematics and the 'imaginary' one of toposes:



Bridge objects

- A *bridge object* (in the sense of O.C.) connecting two objects a and b is an object u which can be <u>'built' from any of the two objects</u> and <u>admits two different representations</u> $f(a)$ and $g(b)$ related by some kind of equivalence $\simeq$, the former being in terms of the object a and the latter in terms of the object b:



- <u>Transfers of information</u> arise from the process of 'unraveling' properties of (resp. constructions on) the 'bridge object' u into properties of (resp. constructions on) the two objects a and b by using the two different representations $f(a)$ and $g(b)$ of u.

One or multiple?

- Any object can be thought of as the collection of all its presentations. A fundamental equivalence relation subsists between these presentations: that of presenting the same object.
- Any object can thus play the role of a 'bridge' across its different presentations.
- We 'access' an object by means of the multiplicity of its presentations, but the objects themselves are actually equivalence classes of presentations.

Invariants or dictionaries?

The method of bridges can be interpreted linguistically as a methodology for translating concepts from one context to another.

But which kind of translation is this?

In general, we can distinguish between two essentially different approaches to translation:

- The 'dictionary-oriented' or 'bottom-up' approach, consisting in a dictionary-based renaming of the single words composing the sentences;
- The 'invariant-oriented' or 'top-down' approach, consisting in the identification of appropriate concepts that should remain invariant in the translation, and in the subsequent analysis of how these invariants can be expressed in the two languages.

The 'bridge-based' translations are of the latter kind.

Unification and morphogenesis

- Bridges abound both in mathematics and in other scientific fields, and can be considered 'responsible' (at least abstractly) for the genesis of things and the nature of reality as we perceive it.
- Indeed, whenever we have an invariant, we can try to use it to build 'bridges' connecting its different manifestations.
- A 'bridge' is precisely the expression, and, in a sense, also the *explanation*, of the connection which exists between the different manifestations of a given invariant.
- Think, for instance, to the notion of *energy* in physics as an invariant: energy is in itself a very abstract concept, but the different forms in which it manifests itself can be very concrete (e.g., thermal energy, electromagnetic energy, mechanical energy, etc.); moreover, the possibility of transforming, as in a 'bridge', a form of energy into another is something very important.

Ideal = real?

- The idea of bridge is an abstraction (like that of invariant), but, interestingly, bridges arising in the experimental sciences can often be identified with actual physical objects (think, for instance, in biology, to the DNA, or, in astronomy, to the stars around which planets revolve).
- In fact, the most enlightening situations occur when these ideal objects admit 'concrete' representations, allowing us to contemplate the dynamics of 'differentiation from the unity' in a more direct and effective way.
- Topos theory allows us to materialize a tremendous number of ideal objects, and hence to establish effective bridges between a great variety of different contexts.
- In general, looking for 'concrete' representations of (or ways of realizing) imaginary concepts can lead to the discovery of more 'symmetric' environments in which phenomena can be described in natural and unified ways.

Contingent and universal

- Every language or point of view is partial (or 'holed'), and it is only through the integration of all possible points of view that one can capture the essence of things.
- There is no universal language that would be better (in an absolute sense) then all the others; every point of view enlightens certain aspects of a phenomenon by hiding others, and can be more or less convenient than others in relation to a certain goal.
- Universality should thus be researched not at the level of languages but at that of 'ideal' objects on which invariants are defined.
- It it therefore crucial to reason at two levels, that of invariants (and of objects on which they are defined) and that of their manifestations in the context of 'concrete' situations, and to study the duality between these two levels, a duality which can be thought of as that between a 'meaning' and the different ways to express it.

Completions and invariants

- To relate different languages or points of view with each other, we need in general to 'complete' them to objects which realize *explicitly* the *implicit* hidden in each of them.

- It is at the level of these completed objects that invariants, or symmetries, manifest themselves, and that we can understand the relations between our given objects thanks to the bridges induced by invariants.

- For example, the classifying topos of a theory is constructed by means of a process of completion of the theory itself, with respect, in a sense, to all the concepts that it is potentially capable to express.

- Thanks to the 'bridge' technique, different theories which describe the same mathematical content are put in relation with each other as if they were fragments of a unique object, partial languages which complete themselves by reflecting one into the other in the totality of points of view embodied by the classifying topos.

For further reading

- O. Caramello
  Grothendieck toposes as unifying 'bridges' : a mathematical morphogenesis,
  to appear in the Springer book *Philosophy of Mathematics. Objects, Structures, and Logics*.
- O. Caramello
  La "notion unificatrice" de topos, in
  Lectures Grothendieckiennes, Spartacus and SMF (2022).
- O. Caramello
  Grothendieck toposes as unifying 'bridges' in Mathematics,
  Mémoire d'habilitation à diriger des recherches,
  Université de Paris 7, 2016,
  available from www.oliviacaramello.com.
- O. Caramello.
  Theories, Sites, Toposes: Relating and studying mathematical theories through topos-theoretic 'bridges',
  Oxford University Press, 2017.