## Appendix E: Some challenges of a human- and computer-readable semantic representation of mathematics

For computers to deal intelligently with a mathematical result, the result must be codified in an unambiguous semantic manner. This does not imply that computer-readable forms must consist of human-unreadable sequences of non-alphabetic characters. On the contrary, modern programming languages can and do unite the precision and unique meaning of computer code with the readability of natural languages for definitions, theorems, and algorithms.

A dozen examples illustrative of previous attempts at capturing semantics (in Isabelle, Coq, Mizar, MathAbs, the Wolfram Language, AGDA, Russel, Nuprl, and ACL2), are presented in Appendix F.

Representing pure mathematics in a human- and computer-readable semantic form poses a not small number of unique challenges, some of which we highlight below.

- In mathematical structures (such as groups, rings, fields, spaces), we have mathematical operations. Most mathematical operations, from simple arithmetic operations to $q$-integrals and functional integrals, have abbreviated operator forms. In addition, mathematical properties or relationships have special notations (such as the unusual symbol for transverse intersections of submanifolds). Since such notations are often specific to a particular field of study, how can this be reflected in a formalized manner that keeps the compactness of the original notation while providing the precision and well-defined meaning needed in a computer language? [84]
- What is the fundamental nature of mathematical meta-structures (theorems, definitions, etc.) and objects [118] (e.g. function spaces) within a future semantic mathematical representation? Are they entities with properties in a four-category ontological sense [15]? How many

such structures does one need [212]? (For many "real-world" entities, a computational form of these concepts was pioneered in the Wolfram Language [19] through its `Entity[]`-concept.)

- A small theorem that nicely illustrates the need for a variety of structures can be found in [27]: " The conformal automorphism group of a countably connected circular region of connectivity of at least three is either a Fuchsian group or a discrete elementary group of Möbius transformations." What constellation of contexts is needed for the semantic interpretation of such a statement, and how is this to be represented?
- Classical pen-and-paper mathematics and mathematical publications use a large number of implicit assumptions and dependencies. For a precise and unambiguous representation, these implicit dependencies must be made explicit. How is this best done at a syntactic level without sacrificing human readability? An example of a fairly, but not fully, explicit theorem is drawn from [25]:

  "Assume $d \geq 2$, $\mathfrak{A} \in L^2_{\mathrm{loc}}(\mathbb{R}^d : \mathbb{R}^d)$, $\mathbf{p} := -i\nabla$, and $T_{\mathfrak{A}} := |\mathbf{p} + \mathfrak{A}|$, then $|\mathbf{x}|\ T_{\mathfrak{A}} + T_{\mathfrak{A}}\ |\mathbf{x}| \geq 0$ on $C_0^\infty(\mathbb{R}^d)$."
- What is the role of an ontology for mathematics within such a semantic representation [23]? Does one need a hierarchy of mathematical structures [30]? Does the overall philosophical stance on (or on the nature of) mathematics (of logicians, formalists, intuitionists, realists) have any influence on a semantic representation? If yes, what kind of influence [24], [207]?
- How can a semantic language for mathematics be best designed from a human-readability point of view?
- How general should the meaning of mathematical structures be? On the one hand, one wishes to minimize the number of constructs and names, but on the other hand, over-generalizations can potentially result in some unnatural and stilted usage. For example,

" How far does the definition of the structure of a hypergraph extend?" " On what structure ($R^n$, lattices in $R^n$, groups, etc.) does a random walk take place?" What role does polymorphism play in a semantic language for mathematics?

- What is the viewpoint to take on the hierarchy of mathematical structures? How generally and abstractly should mathematical structures be represented? How should substructures, superstructures, morphisms acting on, and endomorphisms of mathematical structures be represented?
- What is the relation between intrinsic and extrinsic views of mathematical objects [107], [120]? Example from [45]: Under what conditions is the better definition of an orthogonal basis in a finite-dimensional Hilbert space $H$ not the classic one, but rather the one viewing a basis as a commutative special †-Frobenius comonoid in the category **FdHilb**?
- How should the introduction of new operators, properties, and structures be allowed so that they do not duplicate previously defined structures and are as " unique as possible?"
- Example from [26] defining the (new) concept of $A$-monogeny. Definition: An $A$-valued function $u(x)$ is called $A$-monogenic if $u(x)$ is the solution of the Dirac equation in the algebra $A = (R^n, \text{o})$: $D \circ u(x) = 0$, where $D = (\partial/\partial x_1, \ldots, \partial/\partial x_n)$.
- How can a semantic representation, that is easily human-comprehensible, be defined so that it can become a computational representation in the future? [29]
- What role do computational effects play in a representation of pure mathematics? [43]
- What are the roles of type theory and dependent type theory within such a language? [46], [99], [102], [112], [115]
- How can a language be designed such that foundational (e.g. [123]) and undecidability issues (e.g. [124]) are " invisible" for most of mathematics, yet the language is coherent enough to

allow expression of mathematical statements about them?

- How faithfully can a given mathematical statement from the original literature be represented? How much canonicalization of variable names, argument orders, and function names should happen in a semantic encoding? How does one best capture false statements already in the literature?

Many more aspects and challenges of representing pure mathematics in a symbolic language are discussed in Stephen Wolfram's recent blog post [15] and Mohan Ganesalingam's recent book *The Language of Mathematics.* [98]

From a practical point of view, in approaching the definition and implementation of such a semantic representation, questions such as the following should also be discussed.

- What can be learned from past formalization initiatives? [51], [52], [53], [119]
- Is it best for one to start with the semantic representation of a random selection of mathematical results (from within one mathematics sub-discipline) or would it be better to start with a large coherent body of existing material (e.g. the stacks project for algebraic geometry [55])?
- How is one to balance experts' expertise and the wisdom of the crowd (harvested through collaborative community efforts, such as [56], [57], [59]) in the development of such a semantic representation?