A PROOF-THEORETICAL SEMANTICS FOR HOMOTOPY TYPE THEORY

MARCO BENINI AND ROBERTA BONACINA

1. Introduction

This work illustrates a semantics based on 1-categories for the Martin-Löf type theory which comprehends the structural rules and dependent product (Π-types). This semantics is oriented toward proof theory in the precise sense that its completeness allows to derive some fundamental results to support the proof-theoretical analysis, in an analogous way as [1] uses PER models. In particular, the semantics is **not** intended to provide a *meaning* to the formal system: its purpose is to have a soundness and completeness theorem, and an easy-to-manage classifying model, specifically devoted to analyse judgemental equality. In this respect, comparing with other semantics [2] is deviating since the respective purposes are different. This focus also justifies the absence of natural models, apart the canonical one.

This semantics and its consequences are easily extended to the canonical inductive types (dependent sum, coproducts, etc.) and to the subset of higher inductive types [4] which admits only first-level higher order constructors. This extension follows by rephrasing the inference rules for those types as introductions of constants and interpreting them as functors, as done for Π-types. The details are in [3].

2. Semantics for the basic system

The key idea to interpret types as endo-functors in a suitable class of categories. In turn, the structural elements (contexts, variables, universes, and judgemental equalities) define the class of possible models. In particular, contexts are interpreted in a special preorder, and terms are objects in a slice category over that preorder to represent the judgements having the context as hypotheses. The indecomposable arrows in slices model typing, and isomorphisms model judgemental equivalences.

Definition 2.1. An arrow f is *weakly indecomposable* when it cannot be factorised unless a factor is an isomorphism, and it is *indecomposable* when it is also not an isomorphism. An object A is *iso-stable* when the only arrow $A \to A$ is the identity id_A which is weakly indecomposable. A category is *iso-stable* when all its objects are iso-stable. An iso-stable category is *typical* when every object is the codomain of an indecomposable arrow. Given a typical category with a distinguished object u, v is a *weak universe* when there is $v \to u$ in $\mathbb{C}$. Also, $\langle \mathbb{C}; u\rangle$ *has weak universes* when for every object A there is $v \to A$ with v a weak universe.

Proposition 2.2. *If $\langle \mathbb{C}; u\rangle$ has weak universes, then there is a chain $\{u_i\}_{i\in\mathbb{N}}$ of non-isomorphic weak universes such that $u_{i+1} \to u_i$ exists and it is indecomposable.*

Definition 2.3. Let $\langle \mathbb{C}; u\rangle$ have weak universes and a terminal object, and let $\{u_i\}_{i\in\mathbb{N}}$ be a chain of weak universes as above. Then $\langle \mathbb{C}; \{u_i\}_{i\in\mathbb{N}}\rangle$ is a *category with universes*, and each u_i is said to be a (*strong*) *universe*.

Moeover, an object A is a *term* when A is non-terminal and there is an arrow $u_i \to A$; a term A is a *type* when $u_i \to A$ is indecomposable.

The class of ML-categories, deputed to interpret the basic system, is roughly composed by categories of the form $\mathfrak{M} = \langle \mathbb{M}; \mathbb{C}, F_\Pi, I_\Pi, E_\Pi \rangle$ where

- $\mathbb{C} \subseteq \mathbb{M}$ is a quasi-order;
- every slice in $\mathbb{M}$ over the objects in $\mathbb{C}$ is a category with universes, where universes, and thus types, are naturally shared among slices;
- each indecomposable arrow $c \to d$ in $\mathbb{C}$ identifies a variable which is reflected in a term of the slice over d;
- F_Π, I_Π and E_Π are families of endofunctors to interpret the dependent product by transforming $\mathbb{M}$ into itself, eventually relating terms and types in different slices (their domains and codomains are, in fact, suitable groupoids of terms). They are deputed to interpret, respectively, the formation, introduction and elimination rules.

  These functors have to satisfy a number of constraints: for example, F_Π must be conservative and must respect isomorphisms.

The precise definition is quite technical and complex, see [3]: it is omitted in this extended abstract while it will be presented during the talk. However, the technical aspect is not the central point here: this semantics is not directed to *explain* the basic system, or in its more general form, homotopy type theory; its purpose is to provide a suitable framework where it is simple to derive structural properties about judgemental equality, which is the focus of our contribution.

Definition 2.4. Fix a theory T as the set of derivable judgements. The set $\mathcal{O}_T$ is defined as the minimal one satisfying: if $(\Gamma\ \mathsf{ctx}) \in T$ then $[\Gamma\ \mathsf{ctx}] \in \mathcal{O}_T$; if $(\Gamma \vdash a : A) \in T$ then $[\Gamma \vdash a] = (a, [\mathsf{AV}(a)\ \mathsf{ctx}]) \in \mathcal{O}_T$, where $\mathsf{AV}(a)$ is the smallest subcontext of Γ in which $a : A$ can be defined.

Given a pair $(a, b) \in \mathcal{O}_T \times \mathcal{O}_T$, $\mathsf{dom}(a, b) = a$ and $\mathsf{cod}(a, b) = b$. Also, if $(a, b) \in \mathcal{O}_T \times \mathcal{O}_T$ and $(b, c) \in \mathcal{O}_T \times \mathcal{O}_T$, then $(b, c) \circ (a, b) = (a, c)$. The set $\mathcal{A}_T \subseteq \mathcal{O}_T \times \mathcal{O}_T$ is the reflexive and transitive closure of the minimal set such that

(1) if $(\Gamma, x : A\ \mathsf{ctx}) \in T$ then $([\Gamma\ \mathsf{ctx}], [\Gamma, x : A\ \mathsf{ctx}]) \in \mathcal{A}_T$;
(2) if $(\Gamma\ \mathsf{ctx}), (\Delta\ \mathsf{ctx}) \in T$ and Γ is a permutation of Δ, $([\Gamma\ \mathsf{ctx}], [\Delta\ \mathsf{ctx}]) \in \mathcal{A}_T$;
(3) if $(\Gamma \vdash a : A) \in T$, then $[\Gamma \vdash a : A] = ([\Gamma \vdash A], [\Gamma \vdash a]) \in \mathcal{A}_T$ and also $([\Gamma \vdash a], [\mathsf{AV}(a)\ \mathsf{ctx}]) \in \mathcal{A}_T$;
(4) if $(\Gamma \vdash a \equiv b : A) \in T$ then $[\Gamma \vdash a \equiv b : A] = ([\Gamma \vdash a], [\Gamma \vdash b]) \in \mathcal{A}_T$.

The *syntactical category* $\mathbb{M}_T$ has $\mathcal{O}_T$ as objects, and $\mathcal{A}_T$ as arrows, its composition is $\circ$, and has the following families of associated functors acting on groupoids of terms. Whenever $(\Gamma \vdash A : \mathcal{U}_i) \in T$,

$$\begin{aligned}&\mathcal{F}_T([\Gamma\ \mathsf{ctx}], [\Gamma \vdash A])(([\Gamma, x : A \vdash B], [\Gamma, x : A\ \mathsf{ctx}])) \\ &\quad = ([\Gamma \vdash \Pi x : A.B], [\Gamma\ \mathsf{ctx}]) \ ,\end{aligned}$$

and for each instance of a Π-type

$$\begin{aligned}&\mathcal{I}_T([\Gamma\ \mathsf{ctx}], [\Gamma \vdash A], [\Gamma, x : A \vdash B])(([\Gamma, x : A \vdash b], [\Gamma, x : A\ \mathsf{ctx}])) \\ &\quad = ([\Gamma \vdash \lambda x : A.\, b], [\Gamma\ \mathsf{ctx}]) \ , \\ &\mathcal{E}_T([\Gamma\ \mathsf{ctx}], [\Gamma \vdash A], [\Gamma, x : A \vdash B])(([\Gamma \vdash f], [\Gamma\ \mathsf{ctx}]), ([\Gamma \vdash a], [\Gamma\ \mathsf{ctx}])) \\ &\quad = ([\Gamma \vdash f\, a], [\Gamma\ \mathsf{ctx}]) \ .\end{aligned}$$

Finally, variables as terms in a context are identified by $[\Gamma, x : A \vdash x]$.

Definition 2.5. The canonical interpretation $[\![\cdot]\!]_T$ over T is defined as follows:

- if $(\Gamma\ \mathsf{ctx}) \in T$ then $[\![\Gamma\ \mathsf{ctx}]\!]_T = !_{[\Gamma\ \mathsf{ctx}]}$;
- if there is A such that $(\Gamma \vdash a : A) \in T$, then $[\![\Gamma \vdash a]\!]_T = ([\Gamma \vdash a], [\Gamma\ \mathsf{ctx}])$;
- if $(\Gamma \vdash a : A) \in T$ then $[\![\Gamma \vdash a : A]\!]_T = ([\Gamma \vdash A], [\Gamma \vdash a])$;
- if $(\Gamma \vdash a \equiv b : A) \in T$ then $[\![\Gamma \vdash a \equiv b : A]\!]_T = ([\Gamma \vdash a], [\Gamma \vdash b])$.

It can be easily shown that the syntactical category, equipped with the canonical interpretation, is a model. This model is, indeed, classifying.

Theorem 2.6 (Classifying model)**.** *If $\mathfrak{M} = \langle \mathbb{M}, [\![\cdot]\!] \rangle$ is a model for T, then there is a functor $\mathfrak{I} \colon \mathfrak{M}_T \to \mathfrak{M}$ which preserves the interpretation.*

Completeness immediately follows from the existence of a canonical model.

3. Application to proof theory

The semantics is cumbersome, counter-intuitive, and obscure, but being sound, complete, and, in particular, having a rather simple classifying model, is the reason to consider it. In fact, reasoning on the classifying model allows to derive structural properties of interest in a simpler way, in analogy with the PER model in [1]. In the following, we derive that Π is injective, and that no universe is judgementally equivalent to a functional space.

Proposition 3.1. *If $\Gamma \vdash \Pi x : A.\, B \equiv \Pi x : C.\, D : \mathcal{U}_i$ then $\Gamma \vdash A \equiv C : \mathcal{U}_i$ and $\Gamma, x : A \vdash B \equiv D : \mathcal{U}_i$.*

Proof. In the semantics, the hypothesis becomes $[\![\Gamma \vdash \Pi x : A.\, B]\!] \cong [\![\Gamma \vdash \Pi x : C.\, D]\!]$. Thus, by definition of interpretation

$$F_\Pi([\![\Gamma\ \mathsf{ctx}]\!], [\![\Gamma \vdash A]\!])([\![\Gamma, x : A \vdash B]\!]) \cong F_\Pi([\![\Gamma\ \mathsf{ctx}]\!], [\![\Gamma \vdash C]\!])([\![\Gamma, x : C \vdash D]\!])$$

Since F_Π is conservative,

$$\begin{aligned} [\![\Gamma, x : A \vdash B]\!] &= F_\Pi^{-1}([\![\Gamma\ \mathsf{ctx}]\!], [\![\Gamma \vdash A]\!])([\![\Gamma \vdash \Pi x : A.\, B]\!]) \\ &\cong F_\Pi^{-1}([\![\Gamma\ \mathsf{ctx}]\!], [\![\Gamma \vdash C]\!])([\![\Gamma \vdash \Pi x : C.\, D]\!]) = [\![\Gamma, x : C \vdash D]\!] \ . \end{aligned}$$

By definition of classifying model, $[\![\Gamma, x : A\ \mathsf{ctx}]\!] \cong [\![\Gamma, x : C\ \mathsf{ctx}]\!]$, so, $[\![\Gamma \vdash A]\!] \cong [\![\Gamma \vdash C]\!]$. Since F_Π respects isomorphisms, it follows $[\![\Gamma, x : A \vdash B]\!] \cong [\![\Gamma, x : A \vdash D]\!]$. By completeness, $\Gamma \vdash A \equiv C : \mathcal{U}_i$ and $\Gamma, x : A \vdash B \equiv D : \mathcal{U}_i$. □

Proposition 3.2. *In no case $\Gamma \vdash \Pi x : A.\, B \equiv \mathcal{U}_i : \mathcal{U}_h$.*

Proof. Suppose $\Gamma \vdash \Pi x : A.\, B \equiv \mathcal{U}_i : \mathcal{U}_h$. Then, $[\![\Gamma \vdash \Pi x : A.\, B]\!] \cong [\![\Gamma \vdash \mathcal{U}_i]\!]$.

Since F_Π is conservative, $F_\Pi^{-1}([\![\Gamma\ \mathsf{ctx}]\!], [\![\Gamma \vdash A]\!])([\![\Gamma \vdash \Pi x : A.\, B]\!]) = [\![\Gamma, x : A \vdash B]\!]$. Also, the functor F_Π is defined on $[\![\Gamma \vdash \mathcal{U}_i]\!]$: $F_\Pi^{-1}([\![\Gamma\ \mathsf{ctx}]\!], [\![\Gamma \vdash A]\!])([\![\Gamma \vdash \mathcal{U}_i]\!]) = [\![\Gamma, x : A \vdash C]\!]$ for some C. However, $F_\Pi([\![\Gamma\ \mathsf{ctx}]\!], [\![\Gamma \vdash A]\!])([\![\Gamma, x : A \vdash C]\!]) = [\![\Gamma \vdash \Pi x : A.\, C]\!] = [\![\Gamma \vdash \mathcal{U}_i]\!]$. In the classifying model, this imposes $\Pi x : A.\, C$ to be literally equal to $\mathcal{U}_i$, which is evidently false. □

References

1. Andreas Abel, Thierry Coquand, and Peter Dybjer, *Normalization by evaluation for Martin-Löf type theory with typed equality judgements*, Proceedings of the 22^{nd} Annual IEEE Symposium on Logic in Computer Science, IEEE Computer Society, 2007, pp. 3–12.
2. B. Ahrens, P. LeFanu Lumsdaine, and V. Voevodsky, *Categorical structures for type theory in univalent foundations*, Tech. report, ArXiV.org, 2018, url: `https://arxiv.org/abs/1705.04310`.
3. Roberta Bonacina, *Semantics for homotopy type theory*, Ph.D. thesis, PhD in Computer Science and Mathematics of Computation, University of Insubria, December 2019.
4. The Univalent Foundations Program, *Homotopy type theory: Univalent foundations of mathematics*, Institute for Advanced Study: `https://homotopytypetheory.org/book`, 2013.