A semantic account of the presheaf translation

Loïc Pujet

February 12, 2020

1 Introduction

In these notes, we will explain how the so-called "Prefascist Translation" from Pédrot[4] amounts to first getting a definition of presheaves that work in intentional type theory, and then unrolling the well-known presheaf model of type theory[2] using this definition. We hope that these notes will serve two purposes: on the one hand, helping the reader in building intuition for the sometimes abstruse syntactic translation, and on the other hand justifying the semantics of the translation.

2 Strict presheaves

The whole construction follows from our definition of presheaves. In *intentional* type theory, defining a type of presheaves over an arbitraty category $\mathcal{C}$ is a notoriously difficult problem, quickly leading to the infamous "coherence hell". However, under some mild assumptions on our type theory (namely, the existence of a hierarchy of proof-irrelevant sorts SProp), we can sidestep this issue by strictifying our gadgets : with sufficiently clever definitions, all coherence equations will be satisfied up to definitional equality.

First, in order to get around the intricacies of defining a category in type theory, we will restrict ourselves to a *strict* category $\mathcal{C}$ internal to Type that we fixed beforehand. By this, we mean that we assume the data of

- a type $\mathcal{C}_0 : \text{Type}$ of objects,
- a term $\mathcal{C}_1 : \mathcal{C}_0 \to \mathcal{C}_0 \to \text{Type}$ encoding morphisms,
- a term $\text{id} : \prod_{a : \mathcal{C}_0} \mathcal{C}_1\ a\ a$ of identity morphisms,
- a term $\text{comp}\ \{a\, b\, c : \mathcal{C}_0\} : \mathcal{C}_1\ b\ c \to \mathcal{C}_1\ a\ b \to \mathcal{C}_1\ a\ c$ defining composition (that we will also write $\circ$).

Moreover, we ask for these to satisfy the defining equations of a category up to *definitional* equality (that we write $\equiv$). That is, we require for the following terms to be convertible:

- comp (id a) $f \equiv f$,
- comp f (id a) $\equiv f$,
- comp (comp f g) $h \equiv$ comp f (comp g h)

This might seem like a very strong requirement, but one can show that if the theory has a sort of strict propositions (as in [1]), then any 1-category can be presented in such a way.

Our "intentional presheaves on $\widehat{\mathcal{C}}$" can be described as elements of the record

$$\mathcal{C}\text{Type} := \begin{Bmatrix} \mathsf{T} & : & \Pi\,(p : \mathcal{C}_0)\,.\,\text{Type} \\ \mathsf{R} & : & \Pi\,(p : \mathcal{C}_0)\,(\Pi\,(q \xrightarrow{\alpha} p)\,.\,\mathsf{T}\ q)\,.\,\text{SProp} \end{Bmatrix}$$

where the notation $\Pi\,(q \xrightarrow{\alpha} p)$ means $\Pi\,\{q : \mathcal{C}_0\}\,(\alpha : \mathcal{C}_1\ q\ p)$. We can recover a regular presheaf in $\widehat{\mathcal{C}}$ from any $A : \mathcal{C}\text{Type}$ as follows:

$$\mathcal{F}(A)_p = \sum_{s\,:\,\Pi\,(q \xrightarrow{\alpha} p)\,.\,A.\mathsf{T}\ q} \Pi\,(q \xrightarrow{\alpha} p)\,.\,A.\mathsf{R}\ q\ s|_\alpha$$

where the notation $s|_\alpha$ means $\lambda\, r\, \beta\,.\, s\ r\ (\alpha \circ \beta)$. Given an arrow $\alpha : \mathcal{C}_1\ q\ p$ and a section $(x_1, x_2) : \mathcal{F}(A)_p$, the functoriality is given by

$$\mathcal{F}(A)_\alpha\ x := (x_1|_\alpha, x_2|_\alpha).$$

which we will also write as $x|_\alpha$ for short. All of this can be defined in intentional type theory, and functoriality equations are satisfied definitionally:

$$x|_\alpha|_\beta \equiv x|_{\alpha\circ\beta}$$
$$x|_1 \equiv x$$

In that sense, we are justified in calling them strict presheaves. Then, the type of morphisms between two such presheaves A and B is defined with the record

$$\mathcal{C}\text{Arrow}\ (A\ B : \mathcal{C}\text{Type}) :=$$
$$\begin{Bmatrix} \mathsf{T} & : & \Pi\,(p : \mathcal{C}_0)\,(x : \mathcal{F}(A)_p)\,.\,B.\mathsf{T}\ p \\ \mathsf{R} & : & \Pi\,(p : \mathcal{C}_0)\,(x : \mathcal{F}(A)_p)\,.\,B.\mathsf{R}\ p\ (\lambda\, q\, \alpha\,.\,\mathsf{T}\ q\ x|_\alpha) \end{Bmatrix}.$$

It should be clear that an element $f : \mathcal{C}\text{Arrow}\ A\ B$ of this record is nothing but an encoding of a natural transformations from $\mathcal{F}(A)$ to $\mathcal{F}(B)$: one can define

$$\mathcal{F}(f)_p\ x := (\lambda\, q\, \alpha\,.\, f.\mathsf{T}\ q\ x|_\alpha,\ \lambda\, q\, \alpha\,.\, f.\mathsf{R}\ q\ x|_\alpha)$$

and readily check that naturality holds definitionally. It is therefore clear that our definitions make ($\mathcal{C}$Type, $\mathcal{C}$Arrow) into a subcategory of $\widehat{\mathcal{C}}$.

Theorem 1. *After adding enough extensionality to our type theory so that we can define the category of presheaves in the standard set-theoretic way, we can show that this subcategory is actually equivalent to $\widehat{\mathcal{C}}$.*

See next section for a proof.

3 Free presheaves

In this section, we explain the reasoning leading to this strange-looking definition of presheaves. In order to do that, we will work in extensional type theory, and use the usual set-theoretic definition of presheaves.

If we consider $\mathcal{C}_0$ as a discrete category, there is an obvious functor $\iota : \mathcal{C}_0 \hookrightarrow \mathcal{C}$, which induces a surjective geometric morphism from $\widehat{\mathcal{C}_0}$ to $\widehat{\mathcal{C}}$:

$$\widehat{\mathcal{C}} \underset{\iota_*}{\overset{\iota^*}{\underset{\bot}{\rightleftarrows}}} \widehat{\mathcal{C}_0}$$

Here, ι^* is the functor corresponding to precomposition with ι, and ι_* its uniquely determined right adjoint which can be computed (using the Yoneda lemma) as follows:

$$\begin{aligned}(\iota_* A)_p &= \mathrm{Hom}_{\widehat{\mathcal{C}_0}}(\iota^* \mathbf{y}_p, A) \\ &= \Pi(q \overset{\alpha}{\to} p) \,.\, A_q\end{aligned}$$

and, for every $\alpha : \mathcal{C}_1\ q\ p$ and $x : (\iota_* A)_p$, the functoriality is given by

$$(\iota_* A)\ \alpha\ x = x|_\alpha.$$

This is remarkable, since the type of presheaves of the form $\iota_* A$ for some $A : \widehat{\mathcal{C}_0}$ is easy to define without any extensionality, and functoriality equations hold definitionally for them. Unfortunately, not every presheaf can be written as a "free" presheaf. But pullbacks of free presheaves can also be defined easily, and it turns out that this will be sufficient.

Lemma 1. *Let X be a presheaf in $\widehat{\mathcal{C}}$. Then we can find A in $\widehat{\mathcal{C}_0}$ and (B, f) a subobject of $\iota^* \iota_* A$ such that X is equal to the following pullback:*

$$\begin{array}{ccc} \mathcal{F}(A, B, f) & \longrightarrow & \iota_* B \\ \big\downarrow & \lrcorner & \big\downarrow{\scriptstyle \iota_* f} \\ \iota_* A & \xrightarrow{\ \eta_{\iota_* A}\ } & \iota_* \iota^* \iota_* A \end{array}$$

where η is the unit of the adjunction.

Proof. ι^* is a surjective geometric morphism. This implies that $\widehat{\mathcal{C}}$ is comonadic over $\widehat{\mathcal{C}_0}$, for the comonad $\iota^* \iota_*$ [3]. Therefore, X is the equalizer of the pair in the following diagram:

$$X \xrightarrow{\ \eta_X\ } \iota_* \iota^* X \underset{\eta_{\iota_* \iota^* X}}{\overset{\iota_* \iota^* \eta_X}{\rightrightarrows}} \iota_* \iota^* \iota_* \iota^* X$$

It is now quite easy to rearrange this diagram into an adequate looking pullback diagram : η is mono and ι^* preserves monos, so taking $A = \iota^* X$ and $(B, f) = (\iota^* X, \iota^* \eta_X)$ will work. $\square$

This leads us to the following definition : let $\mathrm{Psh}(\mathcal{C})$ be the category whose objects are triples (A, B, f) with $A : \widehat{\mathcal{C}_0}$ and $(B, f) : \mathrm{Sub}(\iota^* \iota_* A)$; and whose morphisms are defined by

$$\mathrm{Hom}_{\mathrm{Psh}(\mathcal{C})}((A, B, f), (C, D, g)) := \mathrm{Hom}_{\widehat{\mathcal{C}}}(\mathcal{F}(A, B, f), \mathcal{F}(C, D, g)).$$

Now, note that in our definition of the objects of $\mathrm{Psh}(\mathcal{C})$, the pair (B, f) amounts to giving a proposition depending over $(\iota^* \iota_* A)_p$ for all p. This is exactly how we defined $\mathcal{C}$Type. Likewise, replacing monos in $\widehat{\mathcal{C}_0}$ with dependent propositions, the definition of morphisms of $\mathrm{Psh}(\mathcal{C})$ matches $\mathcal{C}$Arrow.

Lemma 2. $\mathrm{Psh}(\mathcal{C})$ *and* $\widehat{\mathcal{C}_0}$ *are equivalent categories.*

Proof. By construction, $\mathcal{F}$ is a full and faithful functor from $\mathrm{Psh}(\mathcal{C})$ to $\widehat{\mathcal{C}_0}$, and lemma 1 tells us that it is essentially surjective (the proof even gives an explicit section). □

For completeness, we unroll the definitions required to turn a presheaf X into $\mathcal{G}(X) : \mathcal{C}\mathrm{Type}$

$$\mathcal{G}(X).\mathsf{T} := \lambda\, p \,.\, X_p$$
$$\mathcal{G}(X).\mathsf{R} := \lambda\, p\ (s\ :\ \Pi\, (q \xrightarrow{\alpha} p) \,.\, X_q) \,.\, \Pi\, (q \xrightarrow{\alpha} p) \,.\, X(\alpha)\ (s\ 1_p) \equiv s\ \alpha.$$

Now that we are convinced that our definition encodes presheaves, let us replicate some standard constructions with our definitions.

4 Hofmann-Streicher universes

An internal universe in $\widehat{\mathcal{C}}$ is a representant $\mathcal{U}$ for the "slice category" functor $\widehat{\mathcal{C}}/\_$. That is, it is a presheaf $\mathcal{U}$ that satisfies

$$\widehat{\mathcal{C}}/X \cong \mathrm{Hom}_{\widehat{\mathcal{C}}}[X, \mathcal{U}]\,.$$

Lemma 3. *The image of an object p by $\mathcal{U}$ is $\mathcal{U}_p = \widehat{\mathcal{C}/p}$, and the image of a morphism is the obvious composition.*

Proof. By the Yoneda lemma, we know that for any object $p : \mathcal{C}_0$, the set $\mathcal{U}_p$ is the set of presheaves over $\mathbf{y}_p$, that is $\widehat{\mathcal{C}}/\mathbf{y}_p$. But this category is just $\widehat{\int \mathbf{y}_p}$, the presheaves over the Grothendieck construction of $\mathbf{y}_p$. Finally, one can show that $\int \mathbf{y}_p \simeq \mathcal{C}/p$. □

Now, using our intentional definition of presheaves, we can define $\widehat{\mathcal{C}/p}$ by replacing $\mathcal{C}$ with $\mathcal{C}/p$:

$$\mathbf{y}\mathrm{Type}\ (p : \mathcal{C}_0) := \left\{ \begin{array}{lll} \mathsf{T} & : & \Pi\, (q \xrightarrow{\alpha} p) \,.\, \mathrm{Type} \\ \mathsf{R} & : & \Pi\, (q \xrightarrow{\alpha} p)\, (\Pi\, (r \xrightarrow{\beta} q) \,.\, \mathsf{T}\ (\alpha \circ \beta)) \,.\, \mathrm{SProp} \end{array} \right\}$$

Using the functor $\mathcal{G}$ from previous section, we could define $\mathcal{U} : \mathcal{C}\text{Type}$ as follows:

$$\mathcal{U}.\mathsf{T} := \lambda\, p\, .\, \mathbf{y}\text{Type}\ p$$

$$\mathcal{U}.\mathsf{R} := \lambda\, p\ (s\ :\ \Pi\, (q \xrightarrow{\alpha} p)\, .\, \mathbf{y}\text{Type}\ q)\, .\, \Pi\, (q \xrightarrow{\alpha} p)\, .\, \begin{cases} (s\ 1_p).\mathsf{T}\ \alpha \equiv (s\ \alpha).\mathsf{T}\ 1_q \\ (s\ 1_p).\mathsf{R}\ \alpha \equiv (s\ \alpha).\mathsf{R}\ 1_q \end{cases}.$$

But it turns out that a slightly different, albeit equivalent definition has been chosen in [4] :

$$\mathbf{y}\text{Type}\ (p : \mathcal{C}_0) := \begin{Bmatrix} \mathsf{T} & : & \Pi\, (q \xrightarrow{\alpha} p)\, .\, \text{Type} \\ \mathsf{R} & : & \Pi\, (\Pi\, (q \xrightarrow{\alpha} p)\, .\, \mathsf{T}\ \alpha)\, .\, \text{SProp} \end{Bmatrix}$$

$$\mathcal{U}.\mathsf{T} := \lambda\, p\, .\, \mathbf{y}\text{Type}\ p$$

$$\mathcal{U}.\mathsf{R} := \lambda\, p\ (s\ :\ \Pi\, (q \xrightarrow{\alpha} p)\, .\, \mathbf{y}\text{Type}\ q)\, .\, \Pi\, (q \xrightarrow{\alpha} p)\, .\, (s\ 1_p).\mathsf{T}\ \alpha \equiv (s\ \alpha).\mathsf{T}\ 1_q.$$

Therefore, we will use this version in what follows.

5 Π-types

Given two presheaves $A\ B\ :\ \widehat{\mathcal{C}}$, the "exponential" presheaf B^A should verify functorially

$$\text{Hom}_{\widehat{\mathcal{C}}}(X \times A, B) \cong \text{Hom}_{\widehat{\mathcal{C}}}(X, B^A).$$

Therefore, using the Yoneda lemma, we get

$$\begin{aligned} (B^A)_p &= \text{Hom}_{\widehat{\mathcal{C}}}(\mathbf{y}_p, B^A) \\ &= \text{Hom}_{\widehat{\mathcal{C}}}(\mathbf{y}_p \times A, B) \end{aligned}$$

We could then just apply $\mathcal{G}$ to get a corresponding $\mathcal{C}$Type. However, it happens that the Yoneda presheaf $\mathbf{y}_p$ is already a strict intentional presheaf. Therefore, given $A\ :\ \mathcal{C}\text{Type}$, we can get a strict presheaf

$$(\mathbf{y}_p \times A)_q := (\mathcal{C}_1\ q\ p) \times \mathcal{F}(A)_q$$

with restriction along $\beta\ :\ \mathcal{C}_1\ r\ q$ given by

$$(\mathbf{y}_p \times A)_\beta\ (\alpha, x) := (\alpha \circ \beta, x|_\alpha)$$

Since $\mathcal{C}$Arrow only uses the fact that A has a strict presheaf structure, it readily extends to $\mathcal{C}\text{Arrow}\ (\mathbf{y}_p \times A)\ B$ and we get the following definition for $B^A\ :\ \mathcal{C}\text{Type}$

$$\begin{aligned} B^A.\mathsf{T} &:= \lambda\, p\, .\, \Pi\, (x\ :\ \mathcal{F}(A)_p)\, .\, B.\mathsf{T}\ p \\ B^A.\mathsf{R} &:= \lambda\, p\, (s\ :\ \Pi\, (q \xrightarrow{\alpha} p)\, (x\ :\ \mathcal{F}(A)_q)\, .\, B.\mathsf{T}\ q)\, . \\ &\qquad \Pi\, (x\ :\ \mathcal{F}(A)_p)\, .\, B.\mathsf{R}\ p\ (\lambda\, q\, \alpha\, .\, s\ q\ \alpha\ x|_\alpha) \end{aligned}$$

One can check that $\mathcal{F}(B^A)$ is equivalent to $\mathcal{C}\text{Arrow}\ (\mathbf{y}_p \times A)\ B$. This settles the case of nondependent function types. Dependent ones turn out to work in a very similar manner, except for one not-so-minor detail

TODO : dependency requires rewriting in type

6 Working inside the universe

A term $\Gamma \vdash t : A$ should be translated as a morphism from the presheaf $[\Gamma]$ corresponding with Γ to the presheaf $[A]$, or in other words a term of type $\mathcal{C}$Arrow $[\Gamma]$ $[A]$. Expanding the definition, we get:

$$\begin{array}{l} p : \mathcal{C}_0 \\ x_0 : \Pi\,(q \xrightarrow{\alpha} p)\,.\,[\Gamma].\mathsf{T}\ q \\ x_\varepsilon : \Pi\,(q \xrightarrow{\alpha} p)\,.\,[\Gamma].\mathsf{R}\ q\ x_0|_\alpha \end{array} \vdash \begin{array}{l} [t]_p^{\mathsf{T}} : [A].\mathsf{T}\ p \\ [t]_p^{\mathsf{R}} : [A].\mathsf{R}\ p\ (\lambda\, q\ \alpha\,.\,[t]_q^{\mathsf{T}}\ x_0|_\alpha\ x_\varepsilon|_\alpha) \end{array}$$

which looks very much like a parametricity translation. This tells us that

- a type A should be translated as a "type" $[A].\mathsf{T}$ and a "parametricity predicate" $[A].\mathsf{R}$;
- a term t should be translated as a "term" $[t]^{\mathsf{T}}$ and a "proof of parametricity" $[t]^{\mathsf{R}}$;
- a context should be a list of terms and parametricity proofs.

Then, unrolling the standard presheaf model with our definitions of the universe and Π-types, one obtains the exact same translation as in [4].

TODO : rewrite the translation with our notations

References

[1] Gaëtan Gilbert, Jesper Cockx, Matthieu Sozeau, and Nicolas Tabareau. Definitional Proof-Irrelevance without K. *Proceedings of the ACM on Programming Languages*, pages 1–28, January 2019.

[2] Martin Hofmann. *Syntax and semantics of dependent types*, pages 13–54. Springer London, London, 1997.

[3] P.T. Johnstone. *Sketches of an Elephant: A Topos Theory Compendium.* Number v. 2 in Oxford Logic Guides. Clarendon Press, 2002.

[4] Pierre-Marie Pédrot. Russian constructivism in a prefascist theory, 2020.