The next 700 syntactical models of type theory
Simon Boulier, Pierre-Marie Pédrot, Nicolas Tabareau

To cite this version:
Simon Boulier, Pierre-Marie Pédrot, Nicolas Tabareau. The next 700 syntactical models of type theory. Certified Programs and Proofs (CPP 2017), Jan 2017, Paris, France. pp.182 - 194, 10.1145/3018610.3018620. hal-01445835

HAL Id: hal-01445835
https://inria.hal.science/hal-01445835
Submitted on 25 Jan 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
The Next 700 Syntactical Models of Type Theory

Simon Boulier Pierre-Marie Pédrot Nicolas Tabareau
Inria, France
firstname.name@inria.fr

Abstract
A family of syntactic models for the calculus of construction with universes (CC\(_\omega\)) is described, all of them preserving conversion of the calculus definitionally, and thus giving rise directly to a program transformation of CC\(_\omega\) into itself. Those models are based on the remark that negative type constructors (e.g., dependent product, coinductive types or universes) are underspecified in type theory—which leaves some freedom on extra intensional specifications. The model construction can be seen as a compilation phase from a complex type theory into a simpler type theory. Such models can be used to derive (the negative part of) independence results with respect to CC\(_\omega\), such as functional extensionality, propositional extensionality, univalence or the fact that bisimulation on a coinductive type may not coincide with equality. They can also be used to add new principles to the theory, which we illustrate by defining a version of CC\(_\omega\) with ad-hoc polymorphism that shows in particular that parametricity is not an implicit requirement of type theory. The correctness of some of the models/program transformations have been checked in the Coq proof assistant and have been instrumented as a Coq plugin.

Categories and Subject Descriptors F.4.1 [MATHEMATICAL LOGIC AND FORMAL LANGUAGES]: Mathematical Logic

Keywords Dependent type theory, Program translation

1. Introduction
“Can we prove that \(\forall x. f x = g x \rightarrow f = g\)?”, “What is the difference between \(P \leftrightarrow Q\) and \(P = Q\), when \(P\) and \(Q\) are two propositions?”, “Can we prove that a term of type \(\forall A. A \rightarrow A\) is necessarily the identity function?”. Those simple questions are frequently asked to someone who starts teaching the use of a theorem prover based on type theory such as Coq (The Coq Development Team 2015) or Agda (The Agda Development Team 2015). However, the definitive answer to those questions is far from being simple as it involves proving the consistency of a type theory augmented with some logical principles or their negation.

Traditionally, given a type theory described as a syntactic object, there are two ways to justify it.

1. The most standard one, coming all the way back from model theory, consists in building a model of it, usually into set theory or a given categorical structure. This amounts formally to translating the syntax into some well-behaved quotient. See e.g. (Werner 1997).

2. The most syntactical one consists in showing some good properties of the syntax, usually strong normalization and subject reduction, so as to derive the desired properties on the type theory.

Each approach suffers from its own set of defects. Models of dependent type theory have been widely studied from a set theoretic point of view (see (Hoffman 1997) for a general overview), giving rise to several families of models, e.g., categories with families (Dybjer 1996), comprehension categories (Jacobs 1993), ... But all those models of a type theory are built into some other fundational theory, putting back the burden on the consistency on the latter. This is unavoidable, because of the incompleteness theorem, and may be actually seen as an advantage when targeting such a stress-tested theory as set theory. Nonetheless, the discrepancy between the source theory and the target theory can be at the root of subtle but very important mismatches. This is typically seen when considering the decidability of conversion which may hold in the source theory, but is lost as soon as objects are interpreted semantically.

Going the syntactical way is not without issues either, as the proofs are often very tedious, not modular, and keep repeating themselves between close flavours of type theory.

In this paper, we advocate for a third alternative approach, based on a proof-as-program interpretation of type theory, which can somehow be seen as a reconciliation of the two aforementioned kinds of justification.

“To justify a complex type theory, it suffices to compile it into a simpler—already justified—type theory.”
Essentially, we follow the model approach, but instead of

targeting an alien system, we use type theory itself and can
thus rely on the consistency of it, proved in a syntactical
manner. Furthermore, contrarily to recent models of type
theory into itself (Chapman 2009; Altenkirch and Kaposi
2016), we do not describe the source theory as a deep em-
bedding in the target one, but rather use a shallow embedding
where all source proofs are translated into the target sys-
tem in a way that preserves most of the structure. This is no
more than the dependently typed counterpart of more stan-
dard logical translations seen as program translations, e.g.
Lafont-Streicher-Reus CPS (Lafont et al. 1995) is the pro-
gramming language equivalent of some variant of Gödel’s
double-negation translation.

This way, several properties from the target theory are in-
herited for free, as long as the program transformation pre-
serves conversion of the calculus definitionally and that the
empty type ⊥ (encoded here as II A : □i. A) is translated
into some type equivalent to ⊥. This provides a very simple
notion of model of type theory, accessible to a broader audi-
ence than specialists of category theory and set theory, which
can be used to quickly answer if an axiom is derivable.

As said, such translations can not be used directly to
prove the consistency of the type theory under considera-
tion, and rely on the consistency of the base type theory itself.
But they can rather be used to study some properties of type
theory enriched with new principles. This approach is in
accordance with the recent work on forcing in type theory
(Jaber et al. 2012, 2016) and parametricity in type theory
(Bernardy et al. 2010; Bernardy and Moulin 2012), which can
both be seen as particular cases of the general picture
described in this paper.

We provide several translations in this paper. They are
based on the remark that negative type constructors (e.g.,
dependent product, coinductive types or universes) are un-
derspecified in type theory—which leaves a lot of freedom
on extra intensional specifications. Such translations can be
used to derive a new type theory where (the negative part of)
independence results can be proven, such as functional ex-
tensionality, propositional extensionality, univalence or the
fact that bisimulation on a coinductive type may not coincide
with equality (that will be called here “stream extensional-
ity”).

It has to be noticed that obtaining similar results using
set-theoretic models is very tedious. Indeed, although most
available models negate univalence, finding a model that
negates functional extensionality or propositional extension-
ality is much more difficult as those principles are hardcoded
in set theory. It becomes even harder when it comes to mod-
els integrating coinductive types, as such model are based
upon the notion of final coalgebra and analysing the connec-
tion between bisimulation and equality in this setting is a
very complex task.

Type-theoretic translations can also be used to add new
principles to the theory, which we illustrate by defining a
version of a type theory with ad-hoc polymorphism (i.e.,
enriched with the (quote) operator of Lisp) that shows in
particular that parametricity is not an implicit requirement
of type theory.

The systems we study are built upon CCω, a type theory
featuring only dependent products and a denumarable hier-
archy of universes □i.

**Contributions.** There are a handful of program translations
in type theory (starting from the subset model of Hofmann
1997), but they are usually not presented as such. This is
why we believe that giving a proper account of them is
worthwhile, as it permits to exhibit the nice properties of
those models, most notably the preservation of computation.
We also hope to popularize such a way of building models.

In this paper, we provide the following contributions,
which are summarized in Figure 1:

- we provide a general approach to defining translations of
  CCω into itself by adding more intensional properties to
  negative types
Coq Formalization. The correctness of the first three program translations have been checked in the Coq proof assistant and have been instrumented as a Coq plugin. They have been developed using the 8.5 release of Coq (The Coq Development Team 2015) and are available at:

https://github.com/CoqHott/Program-translations-CC-omega

2. Overview
The systems we study in this paper are built upon CCω, a type theory featuring only dependent products, also known as II-types. It features a denumerable hierarchy of universes □i, and is a close relative to Luo’s ECC (Luo 1989).

Definition 1 (Typing system). We define two statements mutually recursively as usual. The statement \( \Gamma \vdash M : A \) means that the environment \( \Gamma \) is well formed, while \( \Gamma \vdash M : A \) means that the term \( M \) has type \( A \) in environment \( \Gamma \). The typing rules are given in Figure 2. □ stands for any □i or the impredicative universe * when it is considered.

Before exposing specific translations between type theories, we sketch the general schema followed by those translations. Given two type theories, the source theory \( S \) and the target theory \( T \), we require that any term \( t \) in \( S \) is translated into a term \( [t] \) in \( T \) by induction over the syntax of \( t \), and that there is an internal operation \( \iota \) which coerces a translated type into a type of the target type theory. We write \([A]\) for \( \iota [A] \). Note that the induced type may live in a bigger universe, we will use this fact in Section 5.2. The context is then translated pointwise as:

\[
\begin{align*}
[\cdot] & := \cdot \\
[\Gamma, x : A] & := [\Gamma], x : [A]
\end{align*}
\]

The expected theorem is typing soundness which states that \( [\Gamma] \vdash [M] : [A] \) whenever \( \Gamma \vdash M : A \). This requires in particular to show a form of computational soundness which says that \( [M] \equiv [N] \) whenever \( M \equiv N \) in order to interpret the conversion rule transparently as another conversion. Another expected property of the translation is consistency preservation, saying that the translation of \( \Pi X : \square i, X \) is not inhabited. Typing soundness together with consistency preservation ensures consistency of the source theory provided consistency of the target theory.

We argue that this approach is strictly sharper than the various notions of models from the literature. First, the translation is only defined by induction on raw syntax, i.e. it does not depend on any typing assumption and is purely local. Furthermore, preservation of conversion gives a first-class treatment of computation even in an untyped setting. Finally, the target is pure type theory, so that every assumption needed to make the soundness proofs go through has to be internal, meaning that one cannot rely on pushing back side-conditions in the metatheory. We call this approach monistic, in the sense that it only requires type theory as foundations.

Although a few instances of models of type theory from the literature can be described in this formalism, most of them cannot, and even those which can are typically not presented this way and rely on an intermediate structure such as categories with families or sets. Examples of models which are syntactical but not program translations include the setoid models of (Hofmann 1997) and (Altenkirch 1999). Amongst the few members of our class of models, one can cite the subset type (Paulin-Mohring 1989) and parametricity translations (Bernardy and Lasson 2011), which are quite similar, as well as the call-by-name forcing translation (Jaber et al. 2016).

The source theory \( S \) may contain more type constructors or logical properties than \( T \) as long as there is a term in \( T \) corresponding to the translation of those new type constructors or logical properties. In practice, \( S \) is often \( T + \text{axiom} \) and the translation ensures that the axiom can be assumed consistently. In this case, we also have a trivial translation from \( T \) to \( T + \text{axiom} \) (the injection) and hence the two theories are equiconsistent. Computational soundness can sometimes be refined so that we can deduce strong normalization for \( S \) from strong normalization for \( T \), although we will not describe this here. This is done by analyzing the preservation of reduction steps rather than mere conversion.

In all the translations presented in this paper, we use the fact that inhabitants of negative types can only be partially observed from the outside. For instance, dependent functions can only be observed using application. In the same way, types can only be observed on a right-hand side of a typing judgment, which means that the translation only depends on \([A]\) and is oblivious to any additional data contained in \([A]\).

Figure 1 summarizes how this general scheme can be applied to various extensions of CCω to give a computational meaning to new principles available in those extensions. Those translations are presented in this paper, but for parametricity (already described in (Bernardy and Moulin 2012)) and for functional reactive programming and forcing which have already been described in (Jaber et al. 2012 2016)—the former being a particular case of the latter because stream \( A \cong \mathbb{N} \to A \) corresponds to forcing on the...
We do not use a more general form of dependent elimination to simplify
seen externally as the only way to observe a function is by
ery dependent function with a boolean value that cannot be

\[ \Gamma, \Delta ::= \cdot \mid \Gamma, x : A \]

\[ A, B, M, N ::= \Box_i \mid x \mid M \, N \mid \lambda x : A. \, M \mid \Pi x : A. \, B \]

\[ \Gamma \vdash i < j \]

\[ \Gamma \vdash \Box_i : \Box_j \]

\[ \Gamma, x : A \vdash M : B \]

\[ \Gamma \vdash \Pi x : A. \, B : \Box \]

\[ \Gamma \vdash \lambda x : A. \, M : \Pi x : A. \, B \]

\[ \Gamma \vdash : \Box_i \]

\[ \Gamma \vdash A : \Box_j \]

\[ \Gamma \vdash : \Box_{\text{max}(i,j)} \]

\[ \Gamma \vdash M : A \]

\[ \Gamma \vdash N : A \]

\[ \Gamma \vdash M : B \]

\[ \Gamma \vdash A : \Box \]

\[ \Gamma \vdash M : B \]

\[ (\lambda x : A. \, M) \, N \equiv M \{ x := N \} \]

Figure 2. Typing rules of CC\(_\omega\)

Extensions. Depending on the translation, we will use vari-
ous extensions of the base CC\(_\omega\) system, and we will refer
to them explicitly when needed. Those extensions are sum-
marized in Figure 3. They feature an impredicative sort *\(\ast\)
as it exists in CoQ, that will be used to distinguish between
propositional extensionality and univalence. To reason about
equality in CC\(_\omega\), we use Martin-Löf identity type, that is
typed in \(\Box_i\) (but it could be typed in *\(\ast\) as in CoQ). We may
need booleans and negative pairs, also known as \(\Sigma\)-types,
with projection \(\pi\) and, in one case, surjective pairing.

Plan of the paper. In Section 3 we give a first translation
that acts on dependent products in order to realize the nega-
tion of functional extensionality. This shows for instance that
this principle is not provable in CC\(_\omega\). In Section 4 we show
that a similar translation can be done on streams in order to
negate the principle of stream extensionality. In Section 5
we provide two translations, one that negates univalence and
even propositional extensionality and one that introduces a
notion of ad-hoc polymorphism in the form of a (quote) op-
erator à la Lisp. In Section 6 we show how some transla-
tions have been formalized in the CoQ proof assistant and
describe the CoQ plugins that implement those transla-
tions.

3. Intensional Functions

In this section, we present a translation that augments ev-
ery dependent function with a boolean value that cannot be
seen externally as the only way to observe a function is by
applying it, which forgets about the boolean value. In this
translation, every function coming from a translated term of
CC\(_\omega\) will be equipped with the value \(\text{true}\).

\[ (\lambda x : A. \, M) \, N \equiv M \{ x := N \} \]

\[ [\Box_i]_f := \Box_i \]

\[ [x]_f := x \]

\[ [\lambda x : A. \, M]_f := (\lambda x : [A]_f, [M]_f, \text{true}) \]

\[ [M \, N]_f := \pi_1 \, [M]_f \, [N]_f \]

\[ [\Pi x : A. \, B]_f := (\Pi x : [A]_f, [B]_f) \times \Box \]

\[ [A]_f := [A]_f \]

Note that \([A]_f := [A]_f\), which means that for this trans-
lation, \(\iota_f\) is the identity function, which is possible because
\([\Box_i]_f := \Box_i\).

Proposition 3 (Substitution lemma). For all terms \(M, N\)
and any variable \(x\) we have

\[ [M \{ x := N \}]_f \equiv [M]_f \{ x := [N]_f \} \]

Proof. By induction on \(M\).

Using the substitution lemma, it is not difficult to derive
the two main properties of the translation.

Theorem 4 (Computational soundness). If \(M \equiv N\) then
\([M]_f \equiv [N]_f\).

Proof. The only non-trivial case is the \(\beta\)-reduction rule,
which is proved by the following rewriting steps.

\[ [(\lambda x : A. \, M) \, N]_f \equiv (\lambda x : [A]_f, [M]_f) \, [N]_f \]

\[ \equiv [M]_f \{ x := [N]_f \} \]

All congruence rules are interpreted as-is.

Theorem 5 (Typing soundness). If \(\Gamma \vdash M : A\) then
\([\Gamma]_f \vdash [M]_f : [A]_f\).

\footnote{We do not use a more general form of dependent elimination to simplify
the theoretical development in this paper.}
Impredicative Universe

\[ A, B, M, N \quad ::= \quad \ldots \mid \ast \]

\[ \Gamma \vdash \Gamma \quad \Gamma \vdash A : B \quad \Gamma, x : A \vdash B : \ast \]

Negative Pairs

\[ A, B, M, N \quad ::= \quad \ldots \mid \Sigma x : A. B \mid \pi_1 M \mid \pi_2 M \mid (M, N) \]

\[ \Gamma \vdash A : B_i \quad \Gamma, x : A \vdash B_j : \square_i \quad \Gamma \vdash M : \Sigma x : A. B \]

\[ \Gamma \vdash \pi_1 M : A \quad \Gamma \vdash \pi_2 M : B[x := \pi_1 M] \]

\[ \pi_1 (M, N) \equiv M \quad \pi_2 (M, N) \equiv N \quad (\pi_1 M, \pi_2 M) \equiv M \]

Booleans

\[ A, B, M, N \quad ::= \quad \ldots \mid B \mid \text{true} \mid \text{false} \mid \text{if } M \text{ ret } P \text{ then } N_1 \text{ else } N_2 \]

\[ \Gamma \vdash M : B \quad \Gamma \vdash P : B \rightarrow B_i \quad \Gamma \vdash N_1 : P \text{ true} \quad \Gamma \vdash N_2 : P \text{ false} \]

\[ \Gamma \vdash \text{if } M \text{ ret } P \text{ then } N_1 \text{ else } N_2 : P \ast M \]

\[ \Gamma \vdash B : B_i \quad \Gamma \vdash \text{true} : B \quad \Gamma \vdash \text{false} : B \]

if \text{true} \text{ ret } P \text{ then } N_1 \text{ else } N_2 \equiv N_1 \quad \text{if false ret } P \text{ then } N_1 \text{ else } N_2 \equiv N_2

Equality

\[ A, B, M, N \quad ::= \quad \ldots \mid \text{refl}_x \mid x =_A y \mid J(P, e, t) \]

\[ \Gamma \vdash \Gamma \quad \Gamma \vdash A : B_i \quad \Gamma \vdash t, t' : A \quad \Gamma \vdash t : A \]

\[ \Gamma \vdash e : t =_A t' \quad \Gamma \vdash y : A \quad q : t =_A t' \vdash P : B_i \]

\[ \Gamma \vdash J(P, e, u) : P[y := t']\{q := \text{refl}_t\} \]

\[ J(P, \text{refl}_t, u) \equiv u \]

\[ \text{Theorem 6 (Consistency preservation). The translation preserves consistency.} \]

\[ \text{Proof. Indeed, we have} \]

\[ [\Pi A : B_i, A]_f := (\Pi A : B_i, A) \times B \]

\[ \text{which is inhabited if and only if } \Pi A : B_i, A \text{ is.} \]

While being quite simple, this syntactical model already implements strictly more than just CC\textsubscript{\omega}. Thanks to the additional boolean worn by functions, it becomes possible to

\[ \text{Proof. By induction on the typing derivation. The only non-trivial cases are the conversion rule which is interpreted by the above theorem, and rules manipulating products which are easily proved.}\]

This translation can account for various extensions of CC\textsubscript{\omega} in a straightforward fashion. For instance, we can interpret inductive datatypes functorially, even though we will not describe it here. In the end of this section, we assume we have at our disposal equality.
Thus a contradiction.

true

\( \lambda x \ (\pi_1 [M]_s) \), \( \pi_2 [M]_s ) 

\[ \text{stream corec } S M_0 M_1 N_s := \]

\( \text{bisim corec} \)

\( (\lambda s_0 \ s_1 \ [S]_s \ (s_0 , (\pi_2 [P]_s) \ s_1 , (\pi_2 [P]_s) \ s) \ s) \)

\( (\lambda s_0 \ s_1 \ [M]_s \ (s_0 , (\pi_2 [P]_s) \ s_1 , (\pi_2 [P]_s) \ s) \ s) \)

\( (\pi_1 [P]_s) \ (\pi_1 [P]_s) \ [N]_s \)

\[ \ldots \]

\[ [A]_s := [A]_s \]

\[ \text{funext} := \Pi(A : \Box) (B : \Box) (f g : A \rightarrow B), \]

\( \Pi x : A. f x = g x ) \rightarrow f = \lambda x . g x \)

Then there is, in the target system, a closed proof of \([\text{funext} \rightarrow \bot]_s\).

Proof. It is sufficient to instantiate in the premise \(A\) and \(B\) by some arbitrary type \(X\) and \(f\) and \(g\) respectively by \((\lambda x : X . x . \text{true})\) and \((\lambda x : X . x . \text{false})\). This immediately leads to a proof that \(\text{true} = \text{false}\) by injection and thus a contradiction.

Almost every model of \(\text{CC}_\omega\), if not all, validates \(\text{funext}\), so it is consistent with \(\text{CC}_\omega\). The previous theorem shows that it is not provable in \(\text{CC}_\omega\).

Corollary 8. \(\text{funext}\) is independent from \(\text{CC}_\omega\).

It is possible to define a similar translation by adding the boolean value “inside” the function \(\Pi x : A. [B]_f := \Pi x : [A]_f. [B]_f \times \mathbb{B}\). This translation is very similar so we do not present it in detail.

4. Intensional Coinductive Types

We now enrich our source theory with coinductive datatypes, and by using a translation similar to the one of intensional functions, we obtain in a straightforward way a variant for intensional streams. Using intensional streams, it is possible to prove that stream extensionality (the fact that bisimilarity coincides with equality for streams) is not provable in \(\text{CC}_\omega\).

Streams. We assume that our source and target theory are equipped with a type of streams and a relation of bisimilarity, defined in Figure 4. Given a stream \(M\), it is possible to get its first element (or head) by using the destructor \(\text{hd}\) and to get its tail by using the destructor \(\text{tl}\). The only way to create a stream is by coinduction using the constructor \(\text{stream corec}\) which produces, for any type \(S\), a map from \(S\) to stream \(A\) given a map \(S \rightarrow A\) and a map \(S \rightarrow S\). Bisimilarity of streams is defined similarly.

The translation for intensional streams works by adding a boolean value to every stream, throwing it away when accessing the head, and threading it when accessing the tail.

Definition 9. The intensional stream translation \([\cdot]_s\) from \(\text{CC}_\omega + \text{stream} \rightarrow \text{CC}_\omega + \Sigma + \mathbb{B} + \text{stream}\) is defined by induction on terms as defined in Figure 4 where unspecified cases are defined by commutation of the translation with the corresponding syntax constructor.

The properties satisfied by \([\cdot]_s\) are the same as for intensional functions, the proof arguments being similar.

Theorem 10 (Typing soundness). Assuming definitional surjective pairing in the target, if \(\Gamma \vdash M : A\) then \([\Gamma]_s \vdash [M]_s : [A]_s\).

Proof. Essentially the same as for the functional case. Conversion is once again interpreted as conversion, as the substitution lemma is trivial. We only need surjective pairing in order to show that \([N]_s\) is well-typed in the interpretation of \(\text{bisim corec}\), otherwise all rules are straightforward.

Remark 11. The requirement of surjective pairing can be lifted by defining a new dedicated coinductive type in the target system. We choose instead to reuse the same notion of bisimilarity for the sake of conceptual simplicity at the expense of a stronger requirement on the target system.

Even more directly than for dependent functions, falsity is translated into itself because the translation does not act on dependent products, so that we can derive consistency preservation directly.

Theorem 12 (Consistency preservation). The translation preserves consistency.

As coinductive types are negatives types, the only way to observe them is by using their destructors, which for the case of streams amounts to observing only the head and the tail of a stream (coinductively). Thus, intensional streams allow to

\(^2\) Coinductive types are defined categorically as terminal coalgebras, so their universal property can be used to construct maps going into them. This corresponds to the coinduction principle.
negate stream extensionality by comparing two streams that share the same elements, but differ on the additional boolean value. As previous, the translation extends to identity types and we have:

**Theorem 13.** We define stream extensionality as follows.

\[
\text{streamext} := \Pi(A : \square_i)(s_1 s_2 : \text{stream } A).
\]

Then there is in the target system a closed proof of \([\text{streamext} \rightarrow \bot]_s\).

*Proof.* Once again, as in the functional case, intensional equality observes the boolean in the translated streams but bisimilarity does not. \(\square\)

## 5. Intensional Types

In this section, we exploit the underspecification of types in \(\text{CC}_\omega\) to give more content to types. Our first translation builds upon the \(\times \mathbb{B}\) trick we used in the previous sections, but our second translation is much more subtle as it allows to implement ad-hoc polymorphism in vanilla type theory.

### 5.1 Ad-hoc intensional types

We define here a translation giving a very naive intensional content to types. Just as we did in the function case, we enrich every type with a boolean which is not observable in the source theory, except for intensional equality.

**Definition 14.** The ad-hoc intensional type translation \([\cdot]_t\) from \(\text{CC}_\omega\) to \(\text{CC}_\omega + \Sigma + \mathbb{B}\) is defined by induction on terms as follows.

\[
[\square_i]_t := (\square_i \times \mathbb{B}, \text{true})
\]

\[
[x]_t := x
\]

\[
[\lambda x : A. M]_t := \lambda x : [A]_t. [M]_t
\]

\[
[M N]_t := [M]_t [N]_t
\]

\[
[\Pi x : A. B]_t := (\Pi x : [A]_t. [B]_t, \text{true})
\]

\[
[A]_t := \pi_1 [A]_t
\]

As in previous sections, the translation verifies the substitution lemma and computational soundness, from which we derive:

**Theorem 15** (Typing soundness). If \(\Gamma \vdash M : A\) then \([\Gamma]_t \vdash [M]_t : [A]_t\).

*Proof.* By induction on the typing derivation of \(\Gamma \vdash M : A\). The only interesting cases are the two rules for type constructors.

1. **Case** \(\square_i : \square_j\). Direct from the fact that

\[(\square_i \times \mathbb{B}, \text{true}) : \square_j \times \mathbb{B}\]

2. **Case** \(\Pi x : A. B : \square_j\). Direct from the fact that

\[(\Pi x : [A]_t. [B]_t, \text{true}) : \square_i \times \mathbb{B}\]

\(\square\)
Theorem 16 (Consistency preservation). The translation preserves consistency.

Proof. This comes immediately from the fact that 
\[ \Pi A : \square_i A \] := \Pi A : \square_i \times \mathbb{B} \pi_1 A \]
which is inhabited if and only if \( \Pi A : \square_i A \)
is inhabited if and only if \( \Pi A : \square_i A \) is \( \square_i A \).

Proposition 17. The above translation can easily accommodate the presence of an impredicative universe \( \ast \) as long as there is one in the target system, by defining

\[ \ast \] := \( \ast \times \mathbb{B}, \text{true} \).

The rules for an impredicative universe from Figure 3 are valid with this interpretation.

More generally, the translation can be extended to any type former \( \Phi \) which is not a universe by defining

\[ \Phi (M_1, \ldots, M_n) \] := \( \Phi ([M_1], \ldots, [M_n]), \text{true} \).

In the remainder of this section, we assume that we have an equality and an empty type and their translations at hand.

Theorem 18. We define propositional extensionality as:

\[ \text{propext} := \Pi (A B : \ast). (A \rightarrow B) \rightarrow (B \rightarrow A) \rightarrow A = \ast B. \]

There is in the target system a proof of \( [\text{propext} \rightarrow \bot] \).

Proof. Just as in the functional extensionality proof, it is sufficient to take \( A \) and \( B \) to be the same underlying proposition \( X \), but with a different boolean.

Using the validity of propositional extensionality in the set model (Werner 1997), we get the independence of propositional extensionality with \( \omega \).

In the same way, it is possible that the negation of univalence holds in the source theory. Using the validity of univalence in the simplicial model of (Kapulkin et al. 2012), we conclude that univalence is independent from \( \omega \).

5.2 Ad-hoc polymorphism

In this section, we push the concept of type intensionality into its utmost consequences, namely by showing that there exists a syntactic model of type theory interpreting a quoting operator on types. Such an operator allows to do case-analysis on the normal form of any type in the theory, which proves that ad-hoc polymorphism is actually compatible with usual type theory. Although this result looks surprising, the model is actually quite straightforward to obtain as soon as the target theory is expressive enough.

We will assume here that our target theory features induction-recursion, and we will use it to define the type of codes (as in (Dybjer 2000)).

Definition 19 (Codes). We define a family of inductive-recursive definitions \( \text{TYPE} \) and \( \text{Elt} \) where \( i \) ranges over \( \mathbb{N} \) as follows, using a suggestive syntax.

\[ \text{Inductive \ TYPE} : \square_i 1 := \]
\[ | \Pi \ ] : \Pi A : \text{TYPE} (\text{Elt}_0 A \rightarrow \text{TYPE}_0) \rightarrow \text{TYPE}_0 \]
with \( \text{Elt}_0 : \text{TYPE}_0 \rightarrow \square_0 := \text{fun} \)
\[ | \Pi \ast \ A B \ ] := \Pi x : \text{Elt}_0 A. \text{Elt}_0 (B x) \]
\[ \ast \times \mathbb{B} \]
\[ \text{Inductive \ TYPE}_{i+1} : \square_{i+2} := \]
\[ | \Pi \ ] : \Pi A : \text{TYPE}_{i+1} (\text{Elt}_{i+1} A \rightarrow \text{TYPE}_{i+1}) \rightarrow \text{TYPE}_{i+1} \]
\[ | \Pi \ast \ast \ A B \ ] := \Pi x : \text{Elt}_0 A. \text{Elt}_{i+1} (B x) \]
\[ \ast \times \mathbb{B} \]

The typing and reduction rules generated by these definitions are formally given in Figure 3. These inductive-recursive types respect the usual positivity conditions and thus do not endanger the consistency of the target system.

There are several points to discuss about codes. First, the definition of codes makes a closed-world assumption on the type constructors from the source theory. Furthermore, as codes reflect faithfully the various operations at our disposal in the source theory, we actually need to hardwire the universe hierarchy in the inductive-recursive definitions by externally quantifying over the index \( i \in \mathbb{N} \). In order not to make the translation even more complex as it is, we simplify a bit the authorized type formers to prevent a combinatorial explosion. Most notably, the codes for product must be duplicated as many times as there are valid combinations of sorts for it.

It is interesting to remark that the resulting model suffers from an expressivity limitation directly linked to the inductive-recursive construction, that is, it cannot interpret an impredicative universe. This would obviously entail a non-wellfounded loop in the definition of codes, because \( \ast : \square_0 \) but all quantifications targeting \( \ast \) would have to refer to every \( \text{TYPE}_i \) defined afterwards.

In order to define the translation, we need to know the universe level at which a term is being typed at each inductive step. Therefore, we will use as a source system \( \omega \), a variant of \( \omega \), which is slightly less expressive but well stratified, defined in Figure 7. It would probably be possible to adapt the translation to the full-blown \( \omega \) system, at the cost of a more involved translation. Essentially, the main difference is that we annotate both the sequent and the variables from the context with the level of their expected sort. We will omit the indices when they are uniquely defined from the context.

Proposition 20. \( \omega \) is a subsystem of \( \omega \).

Proof. It is indeed sufficient to erase all level annotations on colons to translate a well-typed \( \omega \) term into a well-typed \( \omega \) term.
A, B, M, N ::= ... | Type | Elt \_i | \Pi \_i | \Pi^{+}\_i | TypeRec_i

\Gamma \vdash \Gamma \\
\Gamma \vdash Type_i : \Box_{i+1} \\
\Gamma \vdash Elt_i : Type_i \rightarrow \Box_i \\
\Gamma \vdash \Pi_i : \Pi_i (A : Type_i) \rightarrow Type_i \\
\Gamma \vdash \Pi_i : \Pi_i (A : Type_i) \rightarrow Type_i \\
\Gamma \vdash \Pi_i : \Pi_i (A : Type_i) \rightarrow Type_i \\
\Gamma \vdash \Pi_i : \Pi_i (A : Type_i) \rightarrow Type_i

\Gamma \vdash \Pi_i : \Pi_i (A : Type_i) \rightarrow Type_i \\
\Gamma \vdash \Pi_i : \Pi_i (A : Type_i) \rightarrow Type_i \\
\Gamma \vdash \Pi_i : \Pi_i (A : Type_i) \rightarrow Type_i \\
\Gamma \vdash \Pi_i : \Pi_i (A : Type_i) \rightarrow Type_i

\Gamma \vdash TypeRec_0 : \Pi P : Type_i \rightarrow \Box_j, P_{\Pi_0} \rightarrow \Pi A : Type_i, P A

where

\begin{align*}
\Pi_i & := \Pi (A : Type_i) (B : Elt_i A \rightarrow Type_i). (\Pi x : Elt_i A. P (B x)) \rightarrow P (\Pi_i A B) \\
\Pi_i^{+} & := \Pi (A : Type_i) (B : Elt_i A \rightarrow Type_i). (\Pi x : Elt_i A. P (B x)) \rightarrow P (\Pi_i A B)
\end{align*}

\begin{align*}
\Gamma, x \vdash A : Type_i & \quad \Rightarrow \quad \Gamma, x \vdash A : Type_i \\
\Gamma, x \vdash A : Type_i & \quad \Rightarrow \quad \Gamma, x \vdash A : Type_i \\
\Gamma, x \vdash A : Type_i & \quad \Rightarrow \quad \Gamma, x \vdash A : Type_i \\
\Gamma, x \vdash A : Type_i & \quad \Rightarrow \quad \Gamma, x \vdash A : Type_i
\end{align*}

(\lambda x : A. M)_N \equiv M(x := N)

(congruence rules omitted)

Contrarily to previous translations, this translation needs to be indexed by the universe level. This is because the code for dependent products is not polymorphic in the universe level, which must thus be inserted explicitly during the translation. Therefore, the index is here to keep track of the current universe level.

**Definition 21.** The ad-hoc polymorphism translation \([.]_{T_i \in N}\) from CC\_\omega to CC\_\omega + Type\_\omega is defined by induction on terms as follows.

\begin{align*}
[\Box]_{T_i} & := U_i \\
[x]_{T_i} & := x \\
[\lambda x : A. M]_{T_i} & := \lambda x : [A]_{T_i} \cdot [M]_{T_i} \\
[M_j N]_{T_i} & := [M]_{T_j} \cdot [N]_{T_j} \\
[\Pi x : A. B]_{T_i} & := \Pi_i^{+} \cdot [A]_{T_j} \cdot [B]_{T_i} \\
[A]_{T_i} & := Elt_i [A]_{T_{i+1}} \\
[I]_{T_i} & := \cdot \\
[\Gamma, x : A]_{T_i} & := [\Gamma]_{T_i} \cdot x : [A]_{T_i}
\end{align*}
In general, $\Pi^j_i$ may not be defined so that we arbitrarily fix $\Pi^j_i := \Pi^0_i$ whenever $i < j$. Note that the context translation is not indexed, as the corresponding level is hardcoded in each variable from the context.

**Proposition 22** (Substitution lemma). For all terms $M$, $N$ and any variable $x$ s.t. $\Gamma, x ; A, \Delta \vdash M : B$, we have

$$[M \{x := N\}]_{\Pi^j_i} \equiv [M]_{\Pi^j_i} \{x := [N]_{\Pi^j_i}\}$$

**Proof.** By induction on the typing derivation of $M$. Actually, we need a much weaker condition than typing on $M$ for the theorem to hold, which amounts to checking that variables are used at the correct universe level in the axiom case. \hfill $\square$

**Theorem 23** (typing soundness). If $\Gamma \vdash M : A$ then $[\Gamma]_{\Pi^j_i} \vdash [M]_{\Pi^j_i} : [A]_{\Pi^j_i}$.

**Proof.** By induction on the typing derivation of $M$. Conversion is interpreted by conversion itself, thanks to the substitution lemma, and the other rules are straightforward. \hfill $\square$

**Remark 24.** When the source theory only features $\Pi$-types and universes, it is not clear whether the translation preserves consistency. Indeed, we have

$$[\Pi A \vdash \Box_i A]_{\Pi^j_i} \equiv \Pi A : \text{TYPE}_i \cdot \text{Elt}_i A$$

but then there is no obvious closed code $A$ made out of $\Pi$-types and universes to give to a term of the above type in order to recover an inconsistency. For instance, trying to feed it with the code of the impredicative encoding of falsity in order to recover an inconsistency. For instance, trying to feed it with the code of the impredicative encoding of falsity simply gives back the same type.

This phenomenon does not occur as soon as the source theory features an empty type $\perp$, reflected in the codes as $F_i$ s.t. $\text{Elt}_i F_i \equiv \perp$. In this case, we can instantiate a proof of the above type with $F_i$ and we obtain immediately a proof of the empty type.

**Theorem 25** (Consistency preservation). If the source theory features an empty type $\perp$, the ad-hoc polymorphic translation preserves consistency.

We now turn to show that this translation allows to fully observe the normal form of types by doing case analysis on any term $A : \Box_i$. We define formally what we mean by this under the notion of type quoting below.

**Definition 26** (Type quoting). A type quoting operator is given in $\text{CC}_{\omega}$ by the data described in Figure 8.

**Theorem 27** (Type Quote). The $\equiv_{\Pi^j_i}$ translation gives a computational content to type quoting.

**Proof.** As all types are represented by codes, the various quote operators are simply defined in terms of the $\text{TYPE}_\text{rec}$ operators.

$$[\text{quote}_i P \ M]_{\Pi^j_i} := \text{TYPE}_\text{rec}_i (\lambda A : \text{TYPE}_i , \text{Elt}_i (\{P\}_{\Pi^j_i} A)) [M]_{\Pi^j_i}$$

Preservation of typing and reduction is straightforward, as everything has been done to obtain this property. \hfill $\square$

The type operator can be used to define non-parametric functions. For example, it becomes possible to prove

$$\Sigma f : (\Pi A : \Box_i A \rightarrow A) . f \equiv \Pi A : \Box_i A \rightarrow A \text{id} \rightarrow \perp$$

by simply passing a function that is the negation on $\mathbb{N}$ and the identity on other types. This shows that parametricity is not built-in in $\text{CC}_{\omega}$.

### 6. Formalization and Instrumentation

This section presents the framework for the Coq formalization of some of the translations. We then describe the instrumentation of such translations as a Coq plugin.

#### 6.1 Deep Embedding using De Bruijn Indices

We rely on the formalization of Pure Type Systems (PTS) given by Siles and Herbelin [Siles and Herbelin 2012]. It consists in first defining untyped terms and then defining the conversion and typing rules. Note that we could have also used another approach based on induction-recursion that defines directly well-typed terms as in the recent work of Altenkirch and Kaposi [Altenkirch and Kaposi 2016]. However, it appears that this approach is a bit too rigid as it forbids to use an untyped term during the translation, making the translation very difficult to define in practice.

Terms are defined by an inductive type, where variables are represented using De Bruijn indices and $\text{Sorts}$ is an inductive type representing the sorts (either $\Box_i$ or $\Box_i$ and $\ast$).

**Inductive Term :** Set :=
- $\textbf{Var} : \mathbb{N} \rightarrow \text{Term}$
- $\textbf{Sort} : \text{Sorts} \rightarrow \text{Term}$
- $\Pi : \text{Term} \rightarrow \text{Term} \rightarrow \text{Term}$
- $\lambda : \text{Term} \rightarrow \text{Term} \rightarrow \text{Term}$
- $\textbf{App} : \text{Term} \rightarrow \text{Term} \rightarrow \text{Term}$
- $\textbf{Eq} : (A \ t_1 \ t_2 : \text{Term}) , \text{Term}$
- $\textbf{refl} : \text{Term} \rightarrow \text{Term}$
- $\textbf{J} : \forall (A \ t_1 \ u t_2 : \text{Term}) , \text{Term}$
- $\textbf{var} : \forall (A \ A \ v : \text{Term} , \Gamma \vdash \nu \in \text{dom} \rightarrow \Gamma \vdash \text{Var} v : A$

The substitution of the $n$-th De Bruijn variable by $\nu$ in $\mathbb{M}$ is defined by recursion on $\mathbb{M}$. We write it $\mathbb{M} [n \mapsto \mathbb{N}]$. Conversion in the calculus, written $\mathbb{M} \equiv \mathbb{N}$, is defined as the reflexive, symmetric and transitive closure of the beta reduction, defined as an inductive family. The type system is defined by mutual induction with the well-formedness of contexts as

**Inductive wf :** Env $\rightarrow$ Prop :=
- $\textbf{wf}_\text{nil} : \text{nil} \rightarrow \perp$
- $\textbf{wf}_\text{cons} : \forall \Gamma A \ a , \Gamma \vdash A : \text{Sort} \rightarrow \text{Env} \rightarrow \Gamma \vdash \perp$

where "$\Gamma \vdash \perp := (\textbf{wf} \ \Gamma)$

**with** typ :** Env $\rightarrow$ Term $\rightarrow$ Term $\rightarrow$ Prop :=
- $\textbf{cVar} : \forall \Gamma A \ v , \Gamma \vdash \nu \in \text{dom} \rightarrow \text{A} \perp \nu \subset \text{dom} \rightarrow \Gamma \vdash \text{Var} v : A$
Theorem $\text{tsl\_correctness}$:

$$(\forall \Gamma, \Gamma' \vdash \Gamma' \iota \vdash \Gamma') \land (\forall \Gamma \Gamma' \Gamma'' A, \Gamma \vdash \Gamma' \vdash \Gamma'' \vdash A \rightarrow A' \equiv A''')$$

In the additional materials accompanying this paper, we have formalized the intensional functions translation, the intensional streams translation and ad-hoc intensional types translation. In each case, computational soundness, typing soundness and consistency preservation have been checked.

6.2 Instrumentation as a Coq Plugin

It is quite commonly agreed upon that type theory can be difficult to work with formally, which is why proof assistants were implemented in the first place. In particular, it is not very convenient to prove new logical or computational principles directly through the paper translation, let alone in the deep embedding formalization. Luckily, our translations rely on a rather generic target theory, so that we can use the one underlying the CoQ proof assistant.

This is where plugins shine in. A CoQ plugin is simply a program that, given a CoQ proof term $M$, produces the translation $[M]$ as another CoQ term. This is a shallow embedding, and this process is orthogonal to the CoQ formalization which uses deep embedding instead. Typically, the fact that $[M]$ is still well-typed relies on the soundness theorem that lives in the metatheory rather than in the object theory.

While there is no real point in implementing model translations in any formal foundations such as e.g. set theory, here, the fact that we have an actual proof assistant at hand makes the plugin quite useful. It allows to give the impression to the user that she is working transparently in the

---

3. Some bureaucratic lemmas about lifting of De Bruijn indices have been admitted for the type system extended with streams. See Readme.md.
source theory, while it turns out everything is translated into the target theory on the fly. In particular, typechecking remains decidable by construction, and one can take advantage of all of the facilities provided by the host language. This effectively prevents the necessity to reimplement a standalone proof assistant for the source theory.

Writing a plugin can be cumbersome, for two unrelated reasons. One has to know the internals of the Coq system, which strongly restricts the amount of people able to do it. Furthermore, a distinct term-translating program must be written for every translation one wishes to internalize, which implies that each translation requires a distinct plugin. For instance, the forcing plugin described in (Jaber et al. 2016) is quite different from the ones we implemented for this paper. As of today, there is no such thing as a meta-plugin that would allow to quickly prototype a term translation in Coq.

Remark 28. The type theory implemented by Coq and the various extensions of CCoq are slightly different, leading to small discrepancies between the previously described translations and their actual instrumentation as a plugin. First, parameters do not appear in dependent pairs from the PTS presentation, while they do in Coq. Second, and more importantly, the management of universes in Coq is more complex as it deals with universe variables and a graph of constraints.

In the remaining of this section, we describe the commands provided by a plugin. First, any Coq term also pertaining to the source theory can be translated into the target theory automatically. Assuming some constant \( c : A \), the plugin command

\[ \text{Translate } c. \]

produces a new constant \( c^f : [A] \) that is added to the current environment and whose body is the translation of the body of \( c \).

Moreover, it is possible to add new symbols to the system through a translation. To do so, it is sufficient to provide for any such symbol \( p : A \) its translation \( p^* : [A] \). This is reflected in the Coq plugin by the command

\[ \text{Implement } p : A. \]

which opens a new goal of type \( [A] \). When it is solved, the resulting term \( p^* \) is used to automatically extend the translation with a term \( p : A \) by defining \( [p] := p^* \).

Theoretically, this amounts to considering a typing relation \( \vdash \) in the translation layer defined by extending CIC with the axiom \( \Gamma \vdash p : A \). We easily get that if \( \Gamma \vdash M : B \) then \( [\Gamma] \vdash [M] : [B] \) using the typing correctness of the translation, which justifies the abovementioned equiconstancy result.

In practice, the plugin can thus be used to inhabit new axioms. For instance, for the case of the intensional functions translation, we can add the negation of functional extensionality in the translation layer by providing a term

\[ \text{Implement } \neg \text{fun_ext} : (\forall A B, (f g : A \rightarrow B), (\forall x, f x = g x) \rightarrow f = g) \rightarrow \text{False}. \]

Such a plugin has been defined for the intensional functions translation, the intensional streams translation and ad-hoc intensional types translation. Code and examples come with this article.

7. Future Work

As a first step, we wish to improve the plugins which are only prototypes for the moment. This requires generalizing the program transformations presented in this paper to the entire language of Coq—defining it on all inductive types, coinductive types and records. This way, it will be possible to work on the translation layer without restriction and still benefit from the new logical or computational principles available in this layer.

Regarding the type quote operator, an instrumentation as a Coq plugin would require to define induction-recursion in Coq. This is the subject of on-going work and has already been implemented in an experimental branch of Coq 8.5 developed by Matthieu Sozeau (https://github.com/mattam82/coq/tree/IR). However, the extension of the syntactic guard condition from inductive to inductive-recursive definitions is still the subject of ongoing research.

Another possible line of work would be on the development of a generic Coq plugin that allows users to define their own compilation phase with a minimum effort—e.g., by defining only the translation on untyped terms in the deep embedding—and generate automatically the associated plugin for this transformation. This way, it will be possible to experiment the frontier of what is provable in type theory at a low cost investment.

But the most important is the development of more translations in order to realize axioms and logic principles. For instance, we wonder if it is possible to find a presentation of the setoid model as a program translation.

References


