Growing a Proof Assistant

William J. Bowman
wjb@williamjbowman.com
Northeastern University

Abstract

Theoreticians often use sophisticated notation to communicate and reason about key ideas in their theories and models. Notation is often domain-specific or even invented on-the-fly when creating a new theory or model. Proof assistants aid theoreticians by rigorously checking formal models, but have poor support for allowing users to conveniently define and use sophisticated notation. For example, in a proof assistant like Coq or Agda, users can easily define simple notation like \( \Gamma \vdash e : t \), but to use BNF notation the user must use a preprocessing tool external to the proof assistant, which is cumbersome.

To support convenient and sophisticated extension, we can use language extension as a fundamental part of the design of a proof assistant. By starting from language extension we can not only facilitate convenient and sophisticated user-defined extensions, but also get a single, compositional system for writing all extensions to the core proof language.

We describe how to design a language-extension system that supports safe, convenient, and sophisticated user-defined extensions, and how to design a proof assistant based on language extension. We evaluate this design by building a proof assistant that features a small dependent type theory as the core language and implementing the following extensions in small user-defined libraries: pattern matching for inductive types, dependently typed staged meta-programming, a tactic language, and BNF and inference-rule notation for inductive type definitions.

1. Introduction

Notation is important to convey ideas quickly while ignoring uninteresting details, but notation is not typed. Each domain has its own notation used to hide the uninteresting details common in that domain. The notation commonly used in programming languages research differs from the notation commonly used in cryptography. Even in the same domain, individual results require new notation to suit the needs of each new model or proof. Every new programming language result may use common domain-specific notation, like BNF grammars, but may also define new notation to convey new ideas.

When working with models on papers, we may conveniently create arbitrarily sophisticated notation. We may define simple syntactic sugar by saying "we write \( \text{let } x = e_1 \text{ in } e_2 \) to mean \( (\lambda x : t. e_2) e_1 \). Or we may define sophisticated extensions that require the reader to perform computation: "we omit the type annotation and instead write \( \lambda x. e \) when the type of \( x \) can be inferred".

Creating these extensions is easy when developing models on paper, but not when using proof assistants. Proof assistants provide increased confidence in formal models and proofs, but lack support for allowing users to conveniently define sophisticated extensions. This lack of support has two downsides. First, formal models need to be reproduced in another medium (such as \( \LaTeX \)) to communicate them effectively, which duplicates effort and risks the two models falling out of sync. Second, it decreases confidence that the specification is correct since the specification must be manually encoded into the language of the proof assistant, rather than written in familiar notation.

Some proof assistants, like Agda, enable convenient user-defined extensions as long as the extension is not sophisticated. Agda’s \texttt{mix} x notation (Danielsson and Norell 2008) is convenient to use but only supports simple notation definitions, like writing a function named \( \_ \vdash \_ : \_ \), where \( \_ \) indicates the position of arguments. Other proof assistants, like Coq, support sophisticated extensions, but creating these extensions is inconvenient to the point that few users can do it. Writing a Coq plugin requires the developer to use a separate toolchain to compile against the Coq implementation, and requires the user to compile and link the plugin against their Coq installation. However, these plugins support sophisticated extensions like Mtac (Ziliani et al. 2013), a new tactic language for Coq. To support more convenient but less sophisticated extensions, Coq also provides other extensions systems like notations, Ltac, and extensible parsing, but with multiple systems comes an increased learning curve for users and the challenge of composing multiple extensions from difference extension systems.

We propose to design proof assistants by using language extension, in the style of Lisp and its descendants, as a fundamental feature. This not only supports convenient and sophisticated user-defined extensions, but provides a simple and compositional system for writing extensions. Informally, we can think of this design as follows: rather than start with a proof assistant and add user-defined extensions, we start with a core language plus a language-extension system from which we can “grow” a proof assistant. We explain the design in detail by:

- Describing a core language for expressing formal models and proofs. \textbf{section 2}. Our core language, called Curnel, is a dependently typed \( \lambda \)-calculus, and does not contain any features except those required for expressing encodings of formal models and proofs. The Curnel implementation is less than 700 lines of code.

- Describing the design and implementation of our language-extension system. \textbf{section 3}. We explain what it means for language-extension systems to enable safe, convenient, and sophisticated extensions, and how to build the "seed" of a proof assistant from a core language and a language-extension system.

We evaluate this design by implementing a proof assistant called \texttt{Cur} that supports safe, convenient, and sophisticated language extension as defined in \textbf{section 3}. To evaluate convenience, we rely partially on lines of code as a proxy, although it does not take into account automatic integration into the proof assistant or compositionality of the extension system. To evaluate the level of sophistication we support, we implement proof-of-concept versions of features provided by existing proof assistants and one feature that is...
only supported via external tools. Specifically, we demonstrate that Cur:

- Enables users to define syntactic sugar and surface language features that are primitive in the surface languages of other proof assistants [section 4]. We build a proof assistant by implementing a surface language as a user-de ned library. This library provides notation including let, non-dependent function arrows, automatic currying, pattern matching on inductive types, and dependently-typed staged meta-programming. This library is less than 400 lines of code.

- Enables users to define tactic languages for proof-writing, including interactive tactic-based proving [section 5]. While existing proof assistants like Coq and VeriML (Stapoulis and Shao 2010) provide tactic languages, they either do not support user-de ned tactic languages, or require users to use external toolchains. Our design enables writing the tactic language in a library, using the same extension system that provides syntactic sugar. Our implementation of this tactic system, excluding tactics, is less than 200 lines of code. By comparison, the Mtac plugin for Coq is over 1200 lines of code.

- Enables users to define domain-speci c languages for writing formal models [section 6]. In particular, we de ne a library that enables modeling programming languages using BNF and inference rule notation, and extracting the models to Coq and EfjX in addition to using them in Cur directly. This library is inspired by Ott (Sewell et al. 2007), an external tool that outputs les for multiple proof assistants from a single let with BNF and inference rule notation. While Ott is a mature tool with many more features so comparing directly is di cult, our library is less than 400 lines of code, while just the lexer for Ott is more than 400 lines of code. No other proof assistant supports BNF and inference-rule notation in the language, nor provides support for users to add the feature as a library in just 400 lines of code.

2. Core Language

Our choice of core language is not vital to our design. We choose a dependently-typed calculus because dependent types have proven to be useful not only for theorem proving but also for verifi ed programming. Whatever the choice, we want to restrict the core language to contain only features necessary to represent the models, proofs, and programs that we are interested in. Features that are not required for expressivity (Felleisen 1990), that only add convenience or reduce burden for the user, should be implemented using language extension.

Our core language, Curnel, is a dependently-typed $\lambda$-calculus with inductive families (Dybjør 1994), an in nite hierarchy of cumulative predicative universes, and an impredicative universe. Curnel is inspired by TT, the core language of Idris (Brady 2013), and similar to Luo’s UTT (Luo 1994). Curnel is implemented in Redex (Felleisen et al. 2009; Matthews et al. 2004) and all gures in this section are extracted from the Redex implementation, so some notation may be slightly non-standard. The top of Figure 1 presents the syntax of Curnel. A Curnel term is either a universe $U$, a function $\lambda x.t.e$, a variable, a dependent function type $\Pi x:t.e$, an application $e(e)$, or the elimination of an inductive type $\text{elim}_D$(methods $\ldots$) $e$. We write universes of level $i$ as $\text{Unv}_i$. We usually write variables using the meta-variable $x$, but we use $D$ for the name of declared inductive types and $c$ for the names of inductive-type constructors. We use application contexts $\Theta$ to represent nested application, and $\Xi$ to represent nested product contexts (telescopes).
Since the eliminator is applied to \((s \, z)\), the second constructor for Nat, we step to a use of the second method \(m_1\). We pass this method the argument to \(s\), and recursively eliminate that argument.

The type system of Curvel is a standard intuitionistic dependent-type theory, but we will present it brie y so that readers are aware of any differences between our system and existing proof assistants. Figure 2 presents the type system.

Starting at the top of the gure, the judgment \(\Delta; \Gamma \vdash e \Rightarrow t\) encodes the predicativity rules for universes and the types of function spaces. We use predicativity rules similar to Coq; functions are impredicative in Coq, and recursively eliminate that argument.

In the bottom of the gure, we de ne term typing. We separate term typing into two judgments to simplify algorithmic implementation of convertibility during type checking. The type-inference judgment \(\Delta; \Gamma \vdash e = t\) infers the type \(t\) for the term \(e\) under term environment \(\Gamma\) and declarations \(\Delta\). Note that this inference is trivial since Curvel terms are fully annotated. Without loss of generality, we assume the inference judgment can return a type in normal form if required.

The type-checking judgment \(\Delta; \Gamma \vdash e = t\) checks that term \(e\) has a type \(t_0\) that is convertible to the type \(t\). The typing for eliminators is based on CIC’s typing of case, but extended to handle recursive arguments. We omit the presentation as it is standard.

3. Designing for Language Extension

If we were following the design of other proof assistants, we would now describe a practical surface language. The surface language would include features that are not necessary for expressivity but simplify development. We would then write an elaborator that transforms the surface language into the core language. In this paper, we instead essentially develop an expressive and extensible elaborator and expose it to the user. This elaborator is the language-extension system.

As with our choice of core language, the choice of language-extension system is not vital to our design. We choose Racket’s language-extension system as it is the subject of active research and development and already supports convenient and sophisticated extension, so we need only ensure the extensions are safe. We imagine that if a proof assistant like Coq were implemented using our design, it would feature a language-extension system written in OCaml. Whatever the choice, we want the language-extension system to enable the users to conveniently write safe and sophisticated extensions.

We will refer to the “metallanguage” as the language in which users write extensions and the “object language” as the language in which users write proof and formal models. In fact, Cur supports an in nite hierarchy of metalanguages; we do not discuss this further in this paper, but refer the interested reader to the literature on multi-stage programming—Taha (2004), for example.

Safe Extension

For extensions to be safe in the context of a proof assistant, all extensions must be valid—in the sense that they do not introduce inconsistency—according to the core language. The simplest and most convenient way to ensure safety is to check the output of all extensions after elaboration into the core language. The biggest drawback to this approach is that type errors may arise in expanded code, rather than in the code written by the user.

To improve type errors, we enable extension authors to check subforms and report errors during expansion. While our design enables extension authors to overcome the problem, other approaches avoid the problem entirely by typing extensions. We discuss these alternatives in section 7.

Convenient Extension

To ensure extensions are convenient to write, we must allow extension authors to reuse language infrastructure and automatically integrate extensions into the object language. Extension authors should not need separate toolchains to write or integrate new extensions; the language-extension system should be an integral part of the proof assistant. Users should be free to write metalanguage and object language code in the same module. Users should only need to write the semantics of extension; the extensions should automatically integrate into the parser and receive parsed objects to compute over. All extensions should be automatically checked for safety.

By comparison, Coq plugins require users to compile against the Coq implementation using a separate toolchain. Users abandon their project when they nd they require a plugin.

Writing extensions as external preprocessors, such as Ott (Sewell et al. 2007), requires users to write a parser and compiler, and the resulting tool does not integrate into the object language. This adds a barrier to both developing and using such extensions.

Sophisticated Extension

To support sophisticated extension, we must allow extensions to perform computation in a general-purpose metalanguage. New extensions should integrate into the syntax of the object language as if they were native syntax. Users must be able to rede ne and extend existing syntax, including base syntax like \(\lambda\) and application.

Language-extension should also support extensions to non-syntactic features of the language. A user may want to extend the reader to parse new literals, rather than just perform rewrites on the AST of the object language. A user may want to extend the interpretation of a module to perform advanced type inference before the existing type checker runs.

Mix \(\times\) notation in Agda only supports de ning new functions whose arguments appear in non-standard positions. It does not support \(de\) ning a form whose subforms are not evaluated as object language expressions; we give an example of such a form in section 6.

Coq features notations and Idris (Brady 2013) features macros, but these are limited to simple syntactic rewrites. They do not support general-purpose computation nor rede ning existing syntax.

3.1 Racket Languages and Language Extension

To describe how we implement Cur and language extension in Cur, we must rst explain Racket’s language-extension facilities. Racket is both a language and a system for de ning languages (Tobin-Hochstadt et al. 2011). We use Racket as both: we implement Cur as an object language in Racket as a system, and write language extensions in Cur using Racket as the metalanguage. We use Racket because Racket’s existing language extension features support con-

There are many examples on the Coq-Club mailing list.
In **transform-lambda**, we ignore the first element of the syntax, as that will be the name of the macro. In the body of the clause, we use syntax quote # to create a template for a new syntax object. Pattern variables are bound inside the template. We simply preserve the pattern of the syntax object, and replace the macro identier with the unicode name λ. In a practical implementation, we would add additional conditions on the pattern variables, such as check that x is an identifier rather than an arbitrary expression.

Finally, outside the meta-language block, we declare the new syntactic form using **define-syntax**, and it is added to the object language.

The second stage of running a program is to run the macro expander. The expander recursively traverses the syntax and calls the associated transformer when it reaches a use of a macro identier. This recursive expansion enables macros to generate calls to other macros, and to build abstractions for de ning macros.

In Racket, we can even use macros to extend and redefine new language features that do not normally have an associated syntactic identier, like the semantics of application. Language features that do not normally have an associated identier have a secondary explicit name. For example, while we normally write application (f e), this is just a special syntax for (#%app f e). We can redefine application by redefining %app, exactly as we define lambda. Similarly, we can redefine the semantics of a module by redefining the %module-begin form. The ability to redefine new language features enables the most sophisticated language extensions and allows us to de ne new object languages in Racket.

We de ne a new library by de ning a library that provides the base syntactic forms of the language and a de nition for %module-begin to implement the semantics of a module. Each Racket module begins with a line of the form %lang name, where name is the name of the library. This causes Racket to use the library name to interpret the module.

We can also de ne extensions to the language reader which allow adding new literals or entirely new kinds of syntax for writing object and metalanguage programs. A reader mixin is an extension to the reader. Users can use reader mixins by adding them.

---

Figure 2: Cur’s Type System (excerpts)
before the language name in the #lang name line. For example, #lang sweet-exp cur adds the sweet-expression reader sweet-exp to Cur, allowing users to write Cur using sweet-expressions instead of Racket’s usual s-expressions. Using reader mixins does not affect how macros are written, since macros are defined on syntax objects which the reader returns, and not on, say, token streams.

3.2 Implementing Cur

We define Cur as a Racket language invoked using #lang cur. Cur uses Racket as the metalanguage, but replaces the object language. The base syntactic forms of Cur are the Curnel forms given in section 2. For example, we can write the identity function as:

```racket
#lang cur
(λ (A : (Type 0)) (λ (a : A) a))
```

To eliminate some of the syntactic noise of s-expressions, we can write the same code using sweet-expressions. Sweet-expressions are similar to s-expressions, but infer some structure from indentation, provide some support for in x notation, and support the $ operator used in Haskell for controlling precedence. S-expressions are also valid sweet-expressions, so we can still express structure manually when necessary. The rest of the example in this paper will use sweet-expression syntax.

```racket
#lang sweet-exp cur
(λ (A : (Type 0)) $ λ (a : A) a)
```

Cur also provides a define form for creating run-time value definitions and a data form for defining inductive types:

```racket
define id $ λ (A : (Type 0)) $ λ (a : A) a

data Nat : 0 (Type 0)
z : Nat
s : (λ (x : Nat) Nat)
```

The base forms plus define and data make up the default object language. The module semantics recursively expand all syntactic forms into base forms. The forms data and define generate no code, but affect the module’s environment. The data form extends the module’s inductive type declaration Δ. The define form extends the module’s value de nitions. As each syntactic form in the module is expanded, prior value definitions are inlined and added to the term environment Γ, the expanded term is type-checked, and top-level expression are normalized and printed.

We can write macros as we saw in the previous section to extend the syntax of Cur. The example of the ASCII lambda is a valid Cur extension. As in Racket, we can also build and use metalanguage abstractions to simplify de ning new language extensions. For instance, the previous example required a lot of code to simply add another name to an existing form. Instead, we could use the following metalinguage abstraction that generate transformers that just replace the macro ident er with another identifier:

```racket
define-syntax lambda
make-rename-transformer(# λ)
```

3.3 Reflection API

Some of the extensions we want to write check and infer types, and run Cur terms at compile-time. We can implement staged meta-programming by running Cur terms at compile-time. We can add type errors to extensions by type-checking during macro expansion. We therefore provide a reflection API—a metalinguage API to the Curnel implementation—for language extensions to use. We explain the API functions that we use in the rest of this paper. Note that these API functions are added to the metalanguage, not the object language, and can only be used by extensions during expansion and before object language code is checked in the core language.

```racket
(cur-expand syn id ...) → SyntaxObject
syn : SyntaxObject
id : Identifier
```

The cur-expand function runs the macro expander on syn until the expander encounters a base form or one of the identi ers in the id ... list. The resulting term is not a fully expanded Curnel term; expansion halts when the top-level form begins with any of the identi ers in the id list or any of the Curnel base forms. For instance, (cur-expand # (λ (x : t) e)) does not expand t or e since λ is a base form. This function lets users write extensions by only considering certain syntactic forms. Since users can arbitrarily extend the syntax of Cur, using cur-expand before pattern matching in a new extensions ensures all unexpected extensions are already expanded.

```racket
(cur-type-check? term type) → Boolean
term : SyntaxObject
type : SyntaxObject
```

The cur-type-check? function returns true when term expands to e, type expands to t, and Δ;Γ ⊢ e ⇒ t holds, and returns false otherwise. Note that this function is not meant to provide an error message; it is meant to be used by extensions that catch type errors in surface syntax and provide their own error messages.

```racket
(cur-type-infer term) → (Maybe SyntaxObject)
term : SyntaxObject
```

The cur-type-infer function returns a syntax representation of t when term expands to the term e and Δ;Γ ⊢ e ⇒ t holds, and false otherwise. This function allows users to build type inference into extensions and reduce annotation burden.

```racket
(cur-normalize term) → SyntaxObject
term : SyntaxObject
```

The cur-normalize function is Cur’s version of eval, but usable only by extensions at compile-time. It essentially calls the implementation of module semantics explained earlier: it expands all extensions, type checks the result, then normalizes the term in the Curnel. Speci cally, the function returns a syntax representation of e1 when term expands to a well-typed term e0 and e1 = reduce[e1, e0]. This lets users explicitly reduce terms or simplify proofs when the type system or other extensions might not evaluate far enough.

This is similar to a feature in Zombie (Casinghino et al. 2014); Zombie users can write potentially non-terminating program to compute proofs, but to do so must explicitly force evaluation and thus act as a termination oracle. As Cur is terminating, the similarity is seeing.

This function also enables us to implement staged meta-programming and run-time re ection without extending the core language.

4. Growing a Surface Language

Cur provides an object language with no more convenience than Curnel. It contains only features necessary for expressivity and nothing that is macro expressible in terms of other features (Felleisen 1990). By contrast, Gallina, the core language of Coq, includes extensions that are macro expressible in terms of other features, such as a non-dependent arrow and multi-arity functions.

Let us now take on the role of a Cur user (more precisely, an extension author) and begin implementing a surface language.
We begin by implementing the simple syntactic sugar like the non-dependent arrow notation. We then implement more sophisticated syntax sugar: we implement a \texttt{let} form with optional type annotations, and a pattern matching form for eliminating inductive types. We conclude with some extensions that go beyond syntax—extensions that add compile-time behaviors but do not necessarily generate code, like debugging features and staged metaprogramming.

4.1 Simple Syntax

Alias for \texttt{(Type 0)}

Writing \texttt{(Type 0)} for all these examples is somewhat tedious. We start with a simple example macro that elaborates \texttt{Type} to \texttt{(Type 0)}. Eventually, a more sophisticated extension could resolve all universe levels automatically, but this will do for now. First, let us import a renamed copy of the default form:

\begin{verbatim}
   require
   only-in curnel
   Type default-Type
\end{verbatim}

Next we define a simple macro with two syntaxes: one applied to an argument, one without any argument. In the first case, we simply expand to the default form. In the second case, we provide level 0 as a default. Note that the choice of surface syntax does not affect how we write syntax objects.

\begin{verbatim}
   define-syntax (Type syn)
   syntax-case syn ()
   [[(Type i) # (default-Type i)]
   [Type # (default-Type 0)]
\end{verbatim}

Multi-Arity Syntax

Cur provides only single-arity functions in the base language. As mentioned in section \textcolor{red}{3}, we can rede ne existing forms like function and application syntax, so we rede ne them to support multi-arity functions via automatic currying. First, let us import renamed copies of the default forms:

\begin{verbatim}
   require
   only-in curnel
   #\texttt{app default-app}
   \lambda default-\lambda
\end{verbatim}

With these renamed copies, we can rede ne \texttt{#\texttt{app}} and \lambda while still generating code that uses the original forms.

Next, we define a simple recursive macro for \lambda that carries all arguments using default-\lambda.

\begin{verbatim}
   define-syntax \lambda
   syntax-rules ()
   [[(\lambda b) b]
   [(\lambda (a : t) (ar : tr) ... b)
   (default-\lambda (a : t) (\lambda (ar : tr) ... b))]
\end{verbatim}

\begin{verbatim}
   define-syntax make-rename-transformer # \lambda
\end{verbatim}

The \texttt{syntax-rules} form is similar to \texttt{syntax-case}, but specialized to support writing simple syntactic rewrites rather than arbitrary metalanguage computation. The syntax \ldots is part of the pattern and template languages for syntax objects. In a pattern, it matches a list of the preceding pattern. In a template, it splices in the list indicated by the preceding pattern.

Next, we rede ne application. The macro automatically curries applications using default-app.

\begin{verbatim}
   define-syntax #\texttt{app}
   syntax-rules ()
   [[#\texttt{app e1 e2}]
   [(default-app e1 e2)]
   [[#\texttt{app e1 e2 e3} ...]
   (#\texttt{app (#\texttt{app e1 e2} e3 ...)]]
\end{verbatim}

These forms are automatically integrated into the object language, replacing the old forms, and are ready to use even later in the same module:

\begin{verbatim}
   define id $ \lambda (A : Type) (a : A) a
   id Nat z
\end{verbatim}

Non-dependent Arrow Syntax

Now let us define a non-dependent arrow form. We start by defining a single-arity arrow syntax \texttt{arrow}:

\begin{verbatim}
   define-syntax \texttt{arrow syn}
   syntax-case syn ()
   [[(arrow t1 t2)
   # ([\Pi (#, (gensym) : t1) t2])]
\end{verbatim}

Note that in Cur we must explicitly generate a fresh name using \texttt{gensym}, due to a limitation in the representation of names in our object language. We use syntax quasiquote # to create a syntax template that supports escaping to compute parts of the template, and use syntax unquote #, to escape the template and run metalanguage expressions.

Now we can easily define the multi-arity arrow, with both ASCII and unicode names.

\begin{verbatim}
   define-syntax \texttt{->}
   syntax-rules ()
   [[(-> a a) a]
   [(-> a a* ...) (arrow a (-> a* ...))]
\end{verbatim}

\begin{verbatim}
   define-syntax \texttt{-> make-rename-transformer(# ->)}
\end{verbatim}

; Usage:

\begin{verbatim}
   data Nat : 0 Type
   z : Nat
   s : (Nat -> Nat)
\end{verbatim}

Top-level Function Definition Syntax

Writing top-level function definitions using \texttt{lambda} is verbose. Most languages feature special syntax for conveniently defining top-level functions, so let us add this to Cur:

\begin{verbatim}
   define-syntax define
   syntax-rules ()
   [[(define (id (x : t) ...) body)
   (default-define id (lambda (x : t) ... body))]
   [(define id body)
   (default-define id body)]
\end{verbatim}

\begin{verbatim}
   define (id (A : Type) (a : A)) a
\end{verbatim}

Notation for Formal Models and Proofs

Recall that our original goal was to provide better support for user-defined notation in formal models and proofs. Thus far, we have merely defined a surface language, which users should not be expected to do. However, the same language-extension facilities serve both purposes. As language implementers, we use language extension to build a surface language. As users, we use language extension to define our own notation—and we have the same power as language implementers when defining new notation.

For instance, suppose we as a user model the simply-typed \lambda-calculus. After writing a small-step evaluation relation and a type-checking relation, we want to use standard notation while doing proofs about the model. We define this notation as follows:

\begin{verbatim}
   define-syntax-rule (\texttt{->} e1 e2) (steps-to e1 e2)
   define-syntax \texttt{->}
   syntax-rules ()
   [[(\Gamma e : t) (type-checks \Gamma e t)]
\end{verbatim}
Recall the addition function for natural number we defined in 
section 2 which we redefine as $+ \text{below}$:

``` regimes
define (+ [n1 : Nat] [n2 : Nat])
elim Nat
  lambda (x : Nat) Nat
  (n1
    lambda (x : Nat) (ih : Nat) $ s ih)
n1
```

This version of $+$ is easier to read and write now that we have 
multi-arity functions, but still requires a lot of annotations and other 
syntactic overhead. Instead, we would like to define $+$ plus using 
pattern matching and avoid writing obvious annotations, like the 
motive, when they can be inferred. We would like to define $+$ plus, 
for instance, like so:

``` regimes
define (+ [n1 : Nat] [n2 : Nat])
match n1
  z n2
  (s (x : Nat)) s $ recur x
```

In this definition we use `match`, which we present shortly. The 
`match` form automatically infers the motive, the annotations on 
`elim`, and inductive hypotheses. It also provides the `recur` form 
to allow users to refer to generated inductive hypotheses by the 
name of the recursive argument from which they are derived.

Figure 3 presents the definition of the `match` extension. This 
simplified version demonstrates how to coordinate two different 
extensions, namely the `match` and `recur` forms, and how to generate 
and compose multiple syntax templates. It makes use of nontrivial 
metalanguage features, like state and user-defined datatypes 
structures. We omit parts of the implementation that do not contribute 
to the goal of demonstrating these features, and features like error 
checking code. Recall that $+$ is an identity used in patterns and 
templates, so we use $+$ to indicate omitted code.

First, we transform the syntax representing a sequence of 
clauses into a list of syntax using `syntax->list`. We parse each 
clause into a structure using `clause-parse`. Each clause consists 
of a pattern and a body. The pattern must be either a constructor 
name for constructors that take no arguments, or a constructor 
named followed by names with type annotations for all arguments to 
the constructor. We store the list of arguments and the body of the 
clause in a clause structure, to be used when generating methods 
for the eliminator.

After parsing each clause, we compute the motive. The body of the 
motive is the type $R$ of the result of the `match`, and the 
argument to the motive has type $D$ of the discriminant. Note that 
in this implementation, we do not attempt to handle indexed or 
parameterized inductive types. The full implementation can infer 
parameters and some indexed inductive types and supports optional 
appearing syntax for when inference fails.

Finally, we generate a method for each clause using `clause->method`. While generating methods, we infer which arguments 
are recursive arguments and compute the inductive hypotheses. We 
update a compile-time dictionary `ih-dict` that associates the name 
of the recursive argument to the generated name for that inductive 
hypothesis. The `recur` form looks up its argument in `ih-dict`.

4.3 Beyond Syntax

Thus far, all our examples demonstrate syntactic transformations. 
Our sophisticated language-extension system also supports creating 
syntactic forms that have semantic behavior at compile-time and 
do not necessarily generate object language code.

For example, we can create a type assertion form that allows 
users to check that an expression has a particular type and receive 
a type error if not:

``` regimes
define-syntax (:: syn)
syntax-case syn ()
```
define-syntax (match syn)
syntax-case syn ()
  ; expects discriminant e and list of clauses clause*
  (_ e clause* ...)
  let* ([clauses (map clause-parse (syntax->list # (clause* ...)))]
  [R (infer-result clauses)]
  [D (cur-type-infer # e)]
  [motive # (lambda (x : #:D) #:R)]
  [U (cur-type-infer R)]
  # (elim #:D #:motive
  #: (map (curry clause->method D motive) clauses)
  e)

define-syntax (recur syn)
syntax-case syn ()
  (_ id)
  dict-ref ih-dict $ syntax->datum # id
begin-for-syntax
define-struct clause (args body)
define ih-dict (make-hash)
define (clause-parse syn)
syntax-case syn
  (pattern body)
  make-clause (syntax-case pattern (:)
  [c ()]
  [(c (x : t) ...) (syntax->list # ((x : t) ...))])
  # body
define (infer-result clauses)
  for/or ([clause clauses])
  cur-type-infer $ clause-body clause
define (infer-ihs D motive args-syn)
syntax-case args-syn () ....
  ; D needed to detect recursive arguments
  ; motive needed to compute type of inductive hypotheses
define (clause->method D motive clause)
  let* ([ihs (infer-ihs D motive (clause-args clause))])
  dict-for-each ihs
  lambda (k v)
  dict-set! ih k $ car v
  # (lambda
  #@clause-args(clause)
  #@(dict-map ihs (lambda (k v) #,(car v) : #,(cdr v))))
  #,clause-body(clause))

Figure 3: A Pattern Matcher for Inductive Types

We check during expansion that e has type t. If the check succeeds, we generate the no-op expression # (void). If the check fails, it reports an error similar to what we do for checking annotations in the let form. This form has no behavior in the object language but provides extra behavior at compile-time to support debugging.

We can also create a syntactic form that forces normalization at compile-time:

```
define-syntax (run syn)
syntax-case syn ()
  (_ expr) $ cur-normalize # expr
```

The run form does not provide new syntactic sugar, but transforms the syntax by normalization via the reflection API.

This can be used to simplify proofs or perform staged meta-programming. For example, we specialize the exponentiation function exp to the square function at compile-time:

```
define square $ run $ exp (s (s z))
```

5. Tactics

In this section we describe a tactic system called ntac implemented in Cur. We begin with an example of using the tactic system to prove a trivial theorem:

```
ntac $ forall (A : Type) (a : A) A
  by-intro A
  by-intro b
  by-assumption
```

This example shows the type of the polymorphic identity function written using tactics. We use ntac, a form that builds an expression given an initial goal followed by a tactic script. This is similar to Goal in Coq, which introduces an anonymous goal that can be solved using an Ltac script. In this example we use the by-intro tactic, which takes a single optional argument representing
the name to bind as an assumption in the local proof environment. Then we conclude the proof with \texttt{by-assumption}, which takes no arguments and searches the local environment for a term that matches the current goal. Since all goals are complete at this point, we end the proof.

\begin{verbatim}
define-theorem id $ forall (A : Type) (a : A) A by-obvious
\end{verbatim}

We can also use \texttt{define-theorem} to define a new identifier using an ntac script. The form \texttt{(define-theorem name goal script \ldots)} is simply syntax sugar for \texttt{(define name (ntac goal script \ldots))}. In this example, we use the \texttt{by-obvious} tactic which solves certain trivial theorems.

We begin implementing ntac by implementing the \texttt{ntac} form:

\begin{verbatim}
define-syntax (ntac stx)
syntax-case stx ()
  [(ntac # goal # script)]
\end{verbatim}

The \texttt{ntac} form runs, at compile-time, the metalanguage function \texttt{ntac-interp} to generate an object language term. The function \texttt{ntac-interp} takes syntax representing an object language type, \texttt{# goal} and syntax representing a sequence of tactics, \texttt{# script}.

In ntac, we use proof trees \texttt{ntt} to represent partial terms in the object language with multiple holes and contextual information such as assumptions, and then use the proof tree tactic \texttt{nttz} to navigate the tree and focus on a particular goal. Tactics are metalanguage functions on \texttt{ntt}s. We will not discuss this design or these data structures in more details here; the design is described in the Cur documentation.

Since tactics are just metalanguage functions, we can create syntactic sugar for defining tactics as follows:

\begin{verbatim}
define-syntax $ define-tactic syn
  syntax-rules ()
  [(begin-for-syntax
     [define e . . ])]
\end{verbatim}

The form \texttt{define-tactic} is simply a wrapper for conveniently defining a new metalanguage function. Note that this extension generates metalinguage code, by generating a new metalinguage block containing a metalinguage definition. Until now, we have only seen extensions that compute using the metalanguage and generate code in object language, but recall from section 3 that Cur supports an infinite hierarchy of language extension.

Now let us write the tactic script interpreter. We begin by defining the function \texttt{run-tactic}, which takes a proof tree zipper and a syntax object representing a call to a tactic.

\begin{verbatim}
begin-for-syntax
define $ run-tactic nttz tactic-stx
define tactic $ eval tactic-stx tactic nttz
\end{verbatim}

We use \texttt{eval} to evaluate the syntax representing the function name and get a function value. Then we simply apply the tactic to the proof tree zipper.

Finally, we define \texttt{ntac-interp} to interpret a tactic script and solve a goal.

\begin{verbatim}
begin-for-syntax
define $ ntac-interp goal script
define pt $ new-proof-tree $ cur-expand goal
define last-nttz for/fold [[nttz (make-nttz pt)]]
  [[(tactic-stx (syntax->list script))]
run-tactic nttz tactic-stx
proof-tree->term $ finish-nttz last-nttz
\end{verbatim}

We begin by generating a fresh proof tree which starts with one goal. The \texttt{for/fold} form folds \texttt{run-tactic} over the list of tactic calls with a starting proof tree zipper over the initial proof tree.

After running all tactics, we check that the proof has no goals left, then generate a term from the proof tree.

Many operations work directly on the current proof tree, so it is cumbersome to define each tactic by rst extracting the proof tree from the proof tree zipper. We introduce a notion of \texttt{tacticals}, metalanguage functions that take a context and a proof tree and return a new proof tree. We define a tactic \texttt{fill} to take a tactical and apply it at the focus of the proof tree zipper. With a notion of tacticals, we can easily define a new intro tactic as follows:

\begin{verbatim}
define-tactical (((intro [name #f]) env pt)
ntac-match $ ntt-goal pt
  [[(forall (x:id : P:expr) body:expr)
    make-ntt-apply
    goal
    λ (body-pf)
    λ (#,the-name : P) #,body-pf]
list
make-ntt-env
λ (old-env)
  (hash-set old-env the-name # P)
make-ntt-hole # body)
\end{verbatim}

We define a new intro tactic \texttt{intro}, which takes one optional argument from the user name, and will be provided the local environment and proof tree from the ntac interpreter. In \texttt{intro}, we start by extracting the current goal from the proof tree. To pattern match on the goal we use the form \texttt{ntac-match}, a simple wrapper around the Racket \texttt{match} form that hides some boilerplate such as expanding the goal into a Cur form and raising an exception if no patterns match.

If the goal has the form of a dependent function type, we make a new node in the ntac proof tree that solves goal by taking a solution for the type of the body of the function and building a lambda expression in the object language. This node contains a subtree that makes the solution of \texttt{# body} the new goal and adds the assumption that \texttt{name} has type \texttt{P} in the scope of this new goal.

To make the \texttt{intro} tactic easier to use at the top level, we define the \texttt{by-intro} tactic:

\begin{verbatim}
begin-for-syntax
define-syntax $ by-intro syn
  syntax-case syn ()
  [... # (fill (intro))]
  [... # (fill (intro # name))]
\end{verbatim}

We create a metalinguage macro \texttt{by-intro} that takes a name as an optional argument. This macro expands to an application of the \texttt{fill} tactic to the \texttt{intro} tactical. We define \texttt{by-intro} as a macro so the user can enter a name for the assumption as a raw symbol, like \texttt{(by-intro A)}, rather than as a syntax object like \texttt{(by-intro # A)}.

Since tactics are arbitrary metalinguage functions, we can define tactics in terms of other tactics, define recursive tactics, and even call to external programs or solvers in the metalinguage or even through the foreign-function interface of our metalinguage. Our next tactic, \texttt{by-obvious}, demonstrates these two features. It will solve any theorem that follows immediately from its premises.

\begin{verbatim}
define-tactical $ obvious env pt
ntac-match $ ntt-goal pt
  [(forall (a : P) body)
    (intro env pt)
    [a:id (assumption env pt)]
\end{verbatim}

We introduce another tactic \texttt{by-obvious} for \texttt{by-obvious} ptz.

\begin{verbatim}
define-tactic $ by-obvious ptz
define nttz $ (fill obvious) ptz
if $ nttz-done? nttz
\end{verbatim}
First we define the obvious tactical, which simply matches a goal and uses either intro or assumption to solve it. Then we define the by-obvious tactic which fills the hole using the obvious tactical and recurs until there are no goals left.

As we have the entire metalanguage available, we can define sophisticated tactics that do arbitrary metalanguage computation. For instance, since our metalanguage provides I/O and reflection, we can define interactivity as a user-defined tactic. We begin implementing interactive by rst implementing the print tactic. This tactic will print the state of the focus of the proof tree zipper and return it unmodified.

```
define-tactic $ print ptz
match $ ntz-focus ptz
[(ntt-hole _ goal)
  for [[[k v] (in-hash (ntt-context ntz))]
    printf "\~a : " (syntax->datum v)
  printf "\~a\n\n" (syntax->datum goal)]
[(ntt-done _ _ _)
  printf "Proof complete.\n"]
```

We rst match on the focus of proof tree zipper. If there is a goal, then we print all assumptions in the context followed by a horizontal line and the current goal. If the zipper indicates the proof has no goals left, then we simply print “Proof Complete”.

Now we define the interactive tactic. This tactic uses the print tactic to print the proof state, then starts a read-eval-print-loop (REPL).

```
define-tactic $ interactive ptz
  print ptz
  let (
    [cmd (read-syntax)]
    syntax-case cmd (quit)
  [tactic
  (define ntz (run-tactic ptz tactic))
  (interactive ntz)]
```

The REPL reads in a command and runs it via run-tactics if it is a tactic. The REPL also accepts quit as a command that exits the REPL and returns the nal proof state.

Now we define not only a user-defined tactic system, but a user-defined interactive tactic system. We can use the interactive tactic just like any other tactic:

```
define-language stlc
# : vars (x)
#: output-latex "stlc.tex"
val (v) ::= true false unit
type (A B) ::= boolty unitty (-> A B) (* A A)
term (e) ::= x v (lambda (#:bind x : A) e) (app e e) (cons e e) (let (#:bind x #:bind x) = e in e)
```

The rst argument to the form is a name for the language—stlc in this case. The next three lines are optional keyword arguments. The #:vars argument is a set of meta-variables that represent variables in the syntax. The #:output-coq argument is a string representing the Coq representation of the language syntax, written to the speci ed le during compilation. Similarly, the #:output-latex argument is a string representing the LaTeX name; when given, a Latex rendering of the BNF grammar is written to the speci ed le during compilation. After the optional arguments, define-language expects an arbitrary number of non-terminal declarations from which it generates inductive types.

To better understand define-language non-terminal clauses, let us rst look at the code generated for the term non-terminal.

```
data stlc-term : 0 Type
  Nat->stlc-term : (Nat -> stlc-term)
  stlc-val->stlc-term : (stlc-value -> stlc-term)
  stlc-lambda : (stlc-type -> stlc-term -> stlc-term)
  stlc-app : (stlc-term -> stlc-term -> stlc-term)
  stlc-cons : (stlc-term -> stlc-term -> stlc-term)
  stlc-let : (stlc-term -> stlc-term -> stlc-term)
```

References to other non-terminals, such as the reference to x, result in conversion constructors which simply inject one non-terminal into the other. The names of the conversion constructors are generated from the types of the non-terminals with the sigil -> between them, indicating conversion. For example, Nat->stlc-term is a stlc-term constructor that converts a Nat (representing a de Brujin index) to a stlc-term. Other constructor names are generated from the name of the language, stlc, and the name of the constructor given in the syntax. For example, the constructor name generated from the lambda syntax is stlc-lambda.

More formally, the syntax of a non-terminal de nition is (nt- name (meta-variables ...)) ::= syn-clause ...). As Cur does not currently support mutual inductive de nitions, all
non-terminal definitions must appear in order of their dependencies. Each syn-clause must be either a reference to a previously defined non-terminal, a terminal represented by a unique identifier, or an s-expression whose rst element is a unique identifier.

For each non-terminal, we generate a new inductive type. We generate constructors for the inductive type for each syn-clause. We prefix the name of each inductive type and each constructor by the language name. For references to previously defined non-terminals, we generate a constructor that acts as a tag and injects the previous non-terminal into the new one.

For terminals, we generate a constructor that takes no arguments and whose name is based on the terminal.

For s-expressions, we create a new constructor whose name is based on the identifier at the head of the s-expression and whose arguments’ types are computed from the meta-variables that appear in the rest of the s-expression. We only use the non-meta-variable symbols such as \( x \) in the Latex rendering of the BNF grammar. The syntax \$\texttt{:bind } x \$ declares \( x \) to be a binding position, so it is not treated as an argument. Since we use de Bruijn indices for binding, binding positions are erased.

The define-language form allows us to create the model using BNF notation, but working with the model requires using the generated constructor names. Instead of using the constructors directly, we can write an extension that parses the STLC syntax into the appropriate constructors. Figure 4 presents an excerpt of the parser. The form begin-stlc simply calls the metalinguage function parse-stlc with the syntax object e and a new hashtable. The parse-stlc function declares each of the constructor names and syntactic symbols as literals, and loops over the syntax object generating the constructors that correspond to the stlc syntax. It uses the hashtable to map variable names to de Bruijn indices. When parsing a lambda, it shifts each index in the hashtable. For convenience, the parser accepts the syntax (el e2) for application instead of (app el e2) and 1 for the unit type instead of unity.

Inference-rule Notation

Figure 5 presents an example of using the inference rule notation. We use the define-relation form to model a type system for stlc. The define-relation form takes as its rst argument a name for the new relation applied to the types of its arguments. This example declares the inductive type has-type which relates a list of stlc-types, an stlc-term, and an stlc-type. Like the define-language form, define-relation takes optional arguments for generating Coq and Latex output. The rest of the form is interpreted as a sequence of inference rules. Each inference rule is a list of assumptions, followed by a horizontal line represented by an arbitrary number of hyphens, a name for the rule that will become the name of the constructor, and a conclusion that must be the relation applied to its arguments.

Figure 4 presents an excerpt of the implementation for define-relation. To implement this form, we use syntax-parse (Culpepper 2012), but only present enough of syntax-parse to understand the key ideas in our implementation.

The syntax-parse form allows parsing syntax objects rather than merely pattern matching on them. The form allows specifying patterns as in syntax-case, but also supports refining the patterns using syntax classes that specify subpatterns and side-conditions. We can apply a syntax class to a pattern variable by using the syntax \( \texttt{pv: class} \), or by using the syntax \( \texttt{~var pv class} \). Both declare the pattern variable \( pv \) must match the syntax class \( \texttt{class} \), but the \( \texttt{~var} \) syntax is required when the syntax class takes arguments.

---

² It should be possible to generate this parser in the define-language extension, but we have not yet implemented that feature.
Our work enables advanced notation, but we do not consider the is-
notation of function application but gets strongly typed extensions.

Fore strongly typed. This enabled writing strongly typed tactics and
extended proof language. This enables sophisticated and safe ex-
pressions, and ensures type errors occur in the surface language, but
requires de ning a new proof checker to de ne macro expressible
syntactic transformations.

While we focus on notation in this work, language extension also
provides support for meta-programming. As we saw in Sec-
7. Related Work

Proof assistants remain an active area of study, but much of this
work focuses on aspects other than notation.

VeriML features a proof assistant with a small core over which
the user can use ML for writing extensions (Stampoulis and Shao
2010). All extensions in VeriML are ML functions, and are there-
fore strongly typed. This enabled writing strongly typed tactics and
decision procedures to manipulate or generate proof terms. Our
work and the VeriML work would complement each other. Extens-
ions in VeriML are ML functions, so the user is limited to the
notation of function application but gets strongly typed extensions.
Our work enables advanced notation, but we do not consider the is-
issue of static guarantees for notation. A language extensions system
designed with a strongly typed language such as ML would enable de-
ning strongly typed notation.

Fisler et al. (1999) study the linguistic issues related to designing a
plug-and-play theorem prover, in which users can select a set of logics and interfaces that can be that can be easily composed.

In their theorem prover, the users must write a plugin that is loaded by
the system. While their focus is on extending the logics of the prover, each extensions can de ne its own notation. However, no-
tations from one extension cannot be used within the notation from
another extension.

We ensure safe language extension by checking all code in the
core language after expansion. This approach can result in type
errors in expanded code that the user did not write, if the extension
author is not careful. Alternative approaches ensure all type errors
occur in the source language rather than the expanded language.

Lorenzen and Erdweg (2016) demonstrates how to de ne well-
typed extensions that justify new typing rules in terms of the old
type system. Since each extension is well-typed, type checking happ-
en in the extended language rather than after expansion. How-
ever, these extensions are currently limited to desugaring and do
not enable general purpose computation in a metalanguage.

Pombrio and Krishnamurthi (2015) develop techniques for re-
sugaring; rather than catch errors before expanding extensions,
these systems undo the expansion before reporting errors. Integrat-
ing this with a language-extension system could reduce the burden
of manually catching errors in new extensions.

The Milawa theorem prover (Myreen and Davis 2014) allows
the user to rede ne the proof-checking function, after establishing
that the new proof checker is valid. The new proof checker may
admit new syntax and new axioms can report errors in terms of the
extended proof language. This enables sophisticated and safe ex-
pressions, and ensures type errors occur in the surface language, but
requires de ning a new proof checker to de ne macro expressible
syntactic transformations.

While we focus on notation in this work, language extension also
provides support for meta-programming. As we saw in Sec-
8. Bibliography

Eric Allen, Ryan Culpepper, Janus Dam Nielsen, Jon Rafkind, and Suky-
oung Ryu. Growing A Syntax. In Proc. of Workshop on Foundations of
Object-Oriented Languages, 2009.

Edwin Brady and Kevin Hammond. Dependently Typed Meta-
Programming. In Proc. of 7th Symposium on Trends in Functional

Edwin C. Brady. Idris, A General-Purpose Dependently Typed Pro-
gramming Language: Design and Implementation. Journal of Func-
cambridge.org/article_S095679681300018X]


