Python: The Full Monty
A Tested Semantics for the Python Programming Language

Joe Gibbs Politiz  
Providence, RI, USA  
joe@cs.brown.edu

Alejandro Martinez  
La Plata, BA, Argentina  
amtriathlon@gmail.com

Matthew Milano  
Providence, RI, USA  
matthew@cs.brown.edu

Sumner Warren  
Providence, RI, USA  
jswarren@cs.brown.edu

Daniel Patterson  
Providence, RI, USA  
dbpatter@cs.brown.edu

Junsong Li  
Beijing, China  
ljs.darkfish@gmail.com

Anand Chitipothu  
Bangalore, India  
anandology@gmail.com

Shriram Krishnamurthi  
Providence, RI, USA  
sk@cs.brown.edu

Abstract

We present a small-step operational semantics for the Python programming language. We present both a core language for Python, suitable for tools and proofs, and a translation process for converting Python source to this core. We have tested the composition of translation and evaluation of the core for conformance with the primary Python implementation, thereby giving confidence in the fidelity of the semantics. We briefly report on the engineering of these components. Finally, we examine subtle aspects of the language, identifying scope as a pervasive concern that even impacts features that might be considered orthogonal.

Categories and Subject Descriptors J.3 [Life and Medical Sciences]: Biology and Genetics

Keywords serpents

1. Motivation and Contributions

The Python programming language is currently widely used in industry, science, and education. Because of its popularity it now has several third-party tools, including analyzers that check for various potential error patterns [2, 5, 11, 13]. It also features interactive development environments [1, 8, 14] that offer a variety of features such as variable-renaming refactorings and code completion. Unfortunately, these tools are unsound: for instance, the simple eight-line program shown in the appendix uses no “dynamic” features and confuses the variable renaming feature of these environments.

The difficulty of reasoning about Python becomes even more pressing as the language is adopted in increasingly important domains. For instance, the US Securities and Exchange Commission has proposed using Python as an executable specification of financial contracts [12], and it is now being used to script new network paradigms [10]. Thus, it is vital to have a precise semantics available for analyzing programs and proving properties about them.

This paper presents a semantics for much of Python (section 5). To make the semantics tractable for tools and proofs, we divide it into two parts: a core language, \( \lambda_{\pi} \), with a small number of constructs, and a desugaring function that translates source programs into the core. The core language is a mostly traditional stateful lambda-calculus augmented with features to represent the essence of Python (such as method lookup order and primitive lists), and should thus be familiar to its potential users.

The term desugaring is evocative but slightly misleading, because ours is really a compiler to a slightly different language. Nevertheless, it is more suggestive than a general term like “compiler”. We blame Arjun Guha for the confusing terminology.
Because desugaring converts Python surface syntax to the core language, when it is composed with an interpreter for $\lambda_\pi$ (which is easy to write), we have another implementation of Python. We can then ask how this implementation compares with the traditional CPython implementation, which represents a form of ground truth. By carefully adjusting the core and desugaring, we have achieved high fidelity with CPython. Therefore, users can build tools atop $\lambda_\pi$, confident that they are conformant with the actual language.

In the course of creating this high-fidelity semantics, we identified some peculiar corners of the language. In particular, scope is non-trivial and interacts with perhaps unexpected features. Our exposition focuses on these aspects.

In sum, this paper makes the following contributions:

• a core semantics for Python, dubbed $\lambda_\pi$, which is defined as a reduction semantics using PLT Redex [3];
• an interpreter, dubbed $\lambda_\pi$, implemented in 700LOC of Racket, that has been tested against the Redex model;
• a desugaring translation from Python programs to $\lambda_\pi$, implemented in Racket;
• a demonstration of conformance of the composition of desugaring with $\lambda_\pi$ to CPython; and,
• insights about Python gained from this process.

Presenting the semantics in full is neither feasible, given space limits, nor especially enlightening. We instead focus on the parts that are important or interesting. We first give an overview of $\lambda_\pi$‘s value and object model. We then introduce desugaring through classes. We then discuss generators, classes, and their interaction with scope. Finally, we describe the results of testing our semantics against CPython. All of our code, and an additional appendix including the full specification of the semantics, is available online at [http://cs.brown.edu/research/plt/di/lambda-py/](http://cs.brown.edu/research/plt/di/lambda-py/).

2. Warmup: A Quick Tour of $\lambda_\pi$

We provide an overview of the object model of $\lambda_\pi$ and Python, some of the basic operations on objects, and the shape of our small step semantics. This introduces notation and concepts that will be used later to explain the harder parts of Python’s semantics.

2.1 $\lambda_\pi$ Values

Figure 1 shows all the values and expressions of $\lambda_\pi$. The metavariables $v$ and $val$ range over the values of the language. All values in $\lambda_\pi$ are either objects, written as triples in ⟨⟩, or references to entries in the store $\Sigma$, written @ref.

Each $\lambda_\pi$ object is written as a triple of one of the forms:

\( \langle v, mval, \langle string: ref, \ldots \rangle \rangle \)

\( \langle x, mval, \langle string: ref, \ldots \rangle \rangle \)

These objects have their class in the first position, their primitive content in the second, and the dictionary of string-indexed fields that they hold in the third. The class value is either another $\lambda_\pi$ value or the name of a built-in class. The primitive content, or meta-val, position holds special kinds of built-in data, of which there is one per built-in type that $\lambda_\pi$ models: numbers, strings, the distinguished meta-none value, lists, tuples, sets, classes, and functions.

The distinguished $\bot$ (“skull”) form represents uninitialized heap locations whose lookup should raise an exception. Within expressions, this form can only appear in let-bindings whose binding position can contain both expressions and $\bot$. The evaluation of $\bot$ in a let-binding allocates it on the heap. Thereafter, it is an error to look up a store location containing $\bot$; that location must have a value $val$ assigned into it for lookup to succeed. Figure 2 shows the behavior of lookup in heaps $\Sigma$ for values and for $\bot$. This notion of undefined locations will come into play when we discuss scope in section 4.2.

Python programs cannot manipulate object values directly; rather, they always work with references to objects. Thus, many of the operations in $\lambda_\pi$ involve the heap, and few are purely functional. As an example of what such an

\[ \Sigma ::= \langle (\text{ref } v+\text{undef}) \ldots \rangle \]

\[ \text{ref} ::= \text{natural} \]

\[ v, \text{val} ::= \langle \text{val}, \text{mval}, \langle \text{string: ref}, \ldots \rangle \rangle \]

\[ \langle x, \text{val}, \langle \text{string: ref}, \ldots \rangle \rangle \]

\[ \langle \text{meta-const} \rangle \text{sym} \langle \text{string: ref}, \ldots \rangle \]

\[ \bot \text{ref} \langle \text{sym string} \rangle \]

\[ v+\text{undef} ::= v | \bot \]

\[ e+\text{undef} ::= e | \bot \]

\[ t ::= \text{global} | \text{local} \]

\[ mval ::= \langle \text{no-mr} \rangle | \text{number} | \text{string} | \text{meta-const} \]

\[ \langle \text{val} \ldots \rangle | \langle \text{val} \ldots \rangle | \langle \text{val} \ldots \rangle \]

\[ \langle \text{meta-class} \rangle \text{sym} \langle \text{string: ref}, \ldots \rangle \]

\[ \lambda \langle x \ldots \rangle \text{o} \]

\[ \text{o} ::= v | \text{ref} | \langle \text{fetch} \rangle | \langle \text{set!} \text{e} \rangle | \langle \text{alloc} \text{e} \rangle \]

\[ \langle \text{e}[\text{e}] \rangle | \langle \text{e}[\text{e}] := \text{e} \rangle \]

\[ \langle \text{let x = e+\text{undef}} \rangle \text{in e} \]

\[ \langle x | \text{e} \rangle ::= \langle \text{delete} \rangle \]

\[ \langle \text{while e} | \langle \text{loop e} \rangle | \langle \text{break} \rangle \text{continue} \rangle \]

\[ \langle \text{builtin-prim op (e \ldots)} \rangle \]

\[ \langle \text{fun (e \ldots)} \rangle \text{o} \]

\[ \langle \text{e,mval} \rangle | \langle \text{list (e,e \ldots)} \rangle \]

\[ \langle \text{tuple (e,e \ldots)} \rangle | \langle \text{set (e,e \ldots)} \rangle \]

\[ \langle \text{tryexcept e x e e} \rangle \]

\[ \langle \text{raise} e | \langle \text{err val} \rangle \rangle \]

\[ \langle \text{module e e} \rangle | \langle \text{construct-module e} \rangle \]

\[ \langle \text{in-module e e} \rangle \]

---

2 We express dictionaries in terms of lists and tuples, so we do not need to introduce a special mval form for them.
In the E-List rule, we also use evaluation contexts \( E \) to enforce an order of operations and eager calling semantics. This is a standard application of Felleisen-Hieb-style small-step semantics [4]. Saliently, a new list value is populated from the list expression via the \( \text{alloc} \) metafunction, this is allocated in the store, and the resulting value of the expression is a pointer \( \text{refnew} \) to that new list.

E-List is a good example for understanding the shape of evaluation in \( \lambda_\pi \). The general form of the reduction relation is over expressions \( e \), global environments \( \epsilon \), and heaps \( \Sigma \):

\[
(E \[\text{let } x = \text{v+undef in } e \] \epsilon \Sigma) \quad \quad \text{[E-LetLocal]}
\]

\[
\rightarrow (E[\text{x/\text{ref}}]e \] \epsilon \Sigma)
\]

where \( (\Sigma_i \text{ ref}) = \text{alloc}(\Sigma, \text{v+undef}) \)

\[
(E[\text{ref}] \epsilon \Sigma) \rightarrow (E[\text{val}] \epsilon \Sigma) \quad \quad \text{[E-GetVar]}
\]

where \( \Sigma = ((\text{ref}, \text{v+undef},) \ldots (\text{ref Val}, \text{v+undef},) \ldots) \)

\[
(E[\text{ref}] \epsilon \Sigma) \rightarrow (E[\text{raise } \langle \text{str}, \text{"Uninitialized local"},()\rangle] \epsilon \Sigma)
\]

where \( \Sigma = ((\text{ref}, \text{v+undef},) \ldots (\text{ref } \sharp) (\text{ref}, \text{v+undef},) \ldots) \)

2.2 Accessing Built-in Values

Now that we’ve created a list, we should be able to perform some operations on it, like look up its elements. \( \lambda_\pi \) defines a number of built-in primitives that model Python’s internal operators for manipulating data, and these are used to access the contents of a given object’s \text{mval}. We formalize these built-in primitives in a metafunction \( \delta \). A few selected cases of the \( \delta \) function are shown in figure 4. This metafunction lets us, for example, look up values on built-in lists:

\[
(\text{prim “list-getitem” } ((\text{list},[(\text{str}, \text{"first-elt"},())]),(\text{int},0,())))
\]

\[
\Rightarrow (\text{str}, \text{"first-elt"},())
\]

Since \( \delta \) works over object values themselves, and not over references, we need a way to access the values in the store. \( \lambda_\pi \) has the usual set of operators for accessing and updating mutable references, shown in figure 5. Thus, the real \( \lambda_\pi \) program corresponding to the one above would be:

\[
(\text{prim “list-getitem” } ((\text{fetch list}(\text{list}),[(\text{str}, \text{"first-elt"},()))]),(\text{fetch (\text{int},0,()))))
\]

Similarly, we can use \( \text{set!} \) and \( \text{alloc} \) to update the values in lists, and to allocate the return values of primitive operations.
We desugar to patterns like the above from Python’s actual surface operators for accessing the elements of a list in expressions like mylist[2].

2.3 Updating and Accessing Fields
So far, the dictionary part of $\lambda_\Pi$ objects have always been empty. Python does, however, support arbitrary field assignments on objects. The expression

$$\sigma_{obj}[\sigma_{str} := \sigma_{val}]$$

has one of two behaviors, defined in figure 6. Both behaviors work over references to objects, not over objects themselves, in contrast to $\delta$. If $\sigma_{str}$ is a string object that is already a member of $\sigma_{obj}$, that field is imperatively updated with $\sigma_{val}$. If the string is not present, then a new field is added to the object, with a newly-allocated store position, and the object’s location in the heap is updated.

The simplest rule for accessing fields simply checks in the object’s dictionary for the provided name and returns the appropriate value, shown in E-GetField in figure 6. E-GetField also works over reference values, rather than objects directly.

2.4 First-class Functions
Functions in Python are objects like any other. They are defined with the keyword `def`, which produces a callable object with a mutable set of fields, whose class is the built-in function class. For example a programmer is free to write:

```python
def f():
    return f.x

f.x = -1
f()  # ==> -1
```

We model functions as just another kind of object value, with a type of `mval` that looks like the usual functional $\lambda$:

$$\lambda(x \ldots) \ opt-var \cdot e$$

The `opt-var` indicates whether the function is variable-arity: if `opt-var` is of the form $(y)$, then if the function is called with more arguments than are in its list of variables $(x \ldots)$, they are allocated in a new tuple and bound to $y$ in the body. Since these rules are relatively unexciting and verbose, we defer their explanation to the appendix.

2.5 Loops, Exceptions, and Modules
We defer a full explanation of the terms in figure 1, and the entire reduction relation, to the appendix. This includes a mostly-routine encoding of control operators via special evaluation contexts, and a mechanism for loading new code via modules. We continue here by focusing on cases in $\lambda_\Pi$ that are unique in Python.

3. Classes, Methods, and Desugaring
Python has a large system with first-class methods, implicit receiver binding, multiple inheritance, and more. In this section we discuss what parts of the class system we put in $\lambda_\Pi$, and which parts we choose to eliminate by desugaring.

3.1 Field Lookup in Classes
In the last section, we touched on field lookup in an object’s local dictionary, and didn’t discuss the purpose of the class position at all. When an object lookup $(\langle val, mval, d \rangle \sigma_{str})$ doesn’t find $\sigma_{str}$ in the local dictionary $d$, it defers to a lookup algorithm on the class value $val$. More specifically, it uses the “$\_mro\_” (short for method resolution order) field of the class to determine which class dictionaries to search for the field. This field is visible to the Python programmer:

```python
class C(object):
    pass # a class that does nothing

print(C.__mro__)
# (class 'C'), (class 'object')
```

Field lookups on objects whose class value is $c$ will first look in the dictionary of $c$, and then in the dictionary of the built-in class object. We define this lookup algorithm within $\lambda_\Pi$ as class-lookup, shown in figure 7 along with the reduction rule for field access that uses it.

This rule allows us to model field lookups that defer to a superclass (or indeed, a list of them). But programmers don’t explicitly define “$\_mro\_” fields; rather, they use higher-level language constructs to build up the inheritance hierarchy the instances eventually use.

3.2 Desugaring Classes
Most Python programmers use the special class form to create classes in Python. However, class is merely syntac-
Figure 6: Simple field access and update in $\lambda_\pi$

\[
\begin{align*}
(E[@ref_{obj} [@ref_{str} := val_{str}]] & \in \Sigma) \quad \text{[E-SetFieldUpdate]} \\
\rightarrow (E[val_{str}] & \in \Sigma[@ref := val_{str}]) \\
& \quad \text{where} \quad \begin{cases}
(\text{any}_{cls}, mval, \{\text{string} := \text{ref}_{str}, \ldots, \text{string} := \text{ref}_{str}, \text{string} := \text{ref}_{str}, \ldots\}) = \Sigma(@ref_{obj}), \\
(\text{any}_{cls}, \text{string}, \text{any}_{dict}) = \Sigma(@ref_{str}), \\
\Sigma = E[@ref_{obj} := (\text{any}_{cls}, mval, \{\text{string} := \text{ref}_{new}, \text{string} := \text{ref}_{str}, \ldots\})], \\
(\text{not} (\text{member} \text{string} (\text{string}, \ldots)))
\end{cases}
\end{align*}
\]

\[
\begin{align*}
(E[@ref [@ref_{str}]] & \in \Sigma) \quad \text{[E-GetField]} \\
\rightarrow (E[@ref_{str}] & \in \Sigma) \\
& \quad \text{where} \quad \begin{cases}
(\text{any}_{cls}, \text{string}, \text{any}_{dict}) = \Sigma(@ref_{str}), \\
(\text{any}_{cls}, mval, \{\text{string} := \text{ref}_{str}, \ldots, \text{string} := \text{ref}_{str}, \text{string} := \text{ref}_{str}, \ldots\}) = \Sigma(@ref)
\end{cases}
\end{align*}
\]

Figure 7: Class lookup

The documentation states explicitly that the two following forms [sic] produce *identical* type objects:

```python
class X:
    a = 1

X = type('X', (object,), dict(a=1))
```

This means that to implement classes, we merely need to understand the built-in function `type`, and how it creates new classes on the fly. Then it is a simple matter to desugar class forms to this function call.

The implementation of `type` creates a new object value for the class, allocates it, sets the `__mro__` field to be the computed inheritance graph and sets the fields of the class to be the bindings in the dictionary. We elide some of the verbose detail in the iteration over `dict` by using the for syntactic abbreviation, which expands into the desired iteration:

```
@ref := E[ref] @ref := alloc(\Sigma, @ref) \\
\begin{align*}
\text{class-lookup[@ref_{obj}], (any}_{x}, \text{any}_{x} := \text{val}_x, \{\text{string} := \text{ref}_{str}, \ldots, \text{__mro__} := \text{ref}, \text{string} := \text{ref}_{str}, \ldots\}, & = (\Sigma \text{val}_{result}) \\
\text{string, } \Sigma] \\
\text{class-lookup-mro}[@ref_{x} \text{val}_x, \ldots, \text{string, } \Sigma] & = \Sigma(\text{ref}) \\
\text{class-lookup-mro}[@ref_{x} \text{val}_x, \ldots, \text{string, } \Sigma] & = \Sigma(\text{ref}) \\
\text{fetch-pointer[@ref, } \Sigma] & = \Sigma(\text{ref})
\end{align*}
```

\[\text{http://docs.python.org/3/library/functions.html#type}\]

\[\text{http://www.python.org/download/releases/2.3/mro/}\]

221
%type :=
fun (cls bases dict)
  let newcls = (alloc (Xtype,(meta-class cls),{})) in
  newcls[〈%str,"__mro__"〉:=
    (builtin-prim "type-buildmro" (newcls bases))]
  (for (key elt) in dict[〈%str,"__items__"〉] ()
    newcls[key := elt])
  (return newcls)

This function, along with the built-in type class, suffices for bootstrapping the object system in λπ.

### 3.3 Python Desugaring Patterns

Python objects can have a number of so-called magic fields that allow for overriding the behavior of built-in syntactic forms. These magic fields can be set anywhere in an object’s inheritance hierarchy, and provide a lot of the flexibility for which Python is well-known.

For example, the field accesses that Python programmers write are not directly translated to the rules in λπ. Even the execution of o.x depends heavily on its inheritance hierarchy. This program desugars to:

```
o[〈%str,"__getattribute__"〉] (o[〈%str,"x"〉])
```

For objects that don’t override the "__getattribute__" field, the built-in object class’s implementation does more than simply look up the "x" property using the field access rules we presented earlier. Python allows for attributes to implement special accessing functionality via properties which can cause special functions to be called on property access. The "__getattribute__" function of object checks if the value of the field it accesses has a special "__get__" method, and if it does, calls it:

```
object[〈%str,"__getattribute__"〉] :=
  fun (obj field)
    let value = obj[field] in
    if (builtin-prim "has-field?" (value[〈%str,"__get__"〉]))
      (return value[〈%str,"__get__"〉] ())
    (return value)

This pattern is used to implement a myriad of features. For example, when accessing function values on classes, the method object closed over c

```
self_is_5 = C.f.__get__(5)
```

Thus, very few object-based primitives are needed to create static class methods and instance methods.

Python has a number of other special method names that can be overridden to provide specialized behavior. λπ tracks Python this regard; it desugars surface expressions into calls to methods with particular names, and provides built-in implementations of those methods for arithmetic, dictionary access, and a number of other operations. Some examples:

```
o[p] desugars to... o[〈%str,"__getitem__"〉] (p)
n + m desugars to... n[〈%str,"__add__"〉] (m)
fun(a) desugars to... fun[〈%str,"__call__"〉] (a)
```

With the basics of type and object lookup in place, getting these operations right is just a matter of desugaring to the right method calls, and providing the right built-in versions for primitive values. This is the form of much of our desugaring, and though it is labor-intensive, it is also the straightforward part of the process.

### 4. Python: the Hard Parts

Not all of Python has a semantics as straightforward as that presented so far. Python has a unique notion of scope, with new scope operators added in Python 3 to provide some features of more traditional static scoping. It also has powerful control flow constructs, notably generators.

#### 4.1 Generators

Python has a built-in notion of generators, which provide a control-flow construct, yield, that can implement lazy or generative sequences and coroutines. The programmer interface for creating a generator in Python is straightforward: any function definition that uses the yield keyword in its body is automatically converted into an object with a generator interface. To illustrate the easy transition from function to generator, consider this simple program:

```
def f():
  x = 0
  while True:
    x += 1
    return x
  f() # ==> 1
  f() # ==> 1
  # ...
```

When called, this function always returns 1.

Changing return to yield turns this into a generator. As a result, applying f() no longer immediately evaluates the body; instead, it creates an object whose next method evaluates the body until the next yield statement, stores its state for later resumption, and returns the yielded value:

```
def f():
  x = 0
  while True:
    x += 1
    yield x
  f() # ==> 1
  f() # ==> 1
  # ...
```

3. [http://docs.python.org/3/library/functions.html#property](http://docs.python.org/3/library/functions.html#property)
def f():
    x = 0
    while True:
        x += 1
        yield x

gen = f()
gen.__next__() # ==> 1
gen.__next__() # ==> 2
gen.__next__() # ==> 3
# ...

Contrast this with the following program, which seems to perform a simple abstraction over the process of yielding:

def f():
    def do_yield(n):
        yield n
    x = 0
    while True:
        x += 1
        do_yield(x)

Invoking f() results in an infinite loop. That is because Python strictly converts only the innermost function with a yield into a generator, so only do_yield is a generator. Thus, the generator stores only the execution context of do_yield, not of f.

Failing to Desugar Generators with (Local) CPS

The experienced linguist will immediately see what is going on. Clearly, Python has made a design decision to store only local continuations. This design can be justified on the grounds that converting a whole program to continuation-passing style (CPS) can be onerous, is not modular, can impact performance, and depends on the presence of tail-calls (which Python does not have). In contrast, it is natural to envision a translation strategy that performs only a local conversion to CPS (or, equivalently, stores the local stack frames) while still presenting a continuation-free interface to the rest of the program.

From the perspective of our semantics, this is a potential boon: we don’t need to use a CPS-style semantics for the whole language! Furthermore, perhaps generators can be handled by a strictly local rewriting process. That is, in the core language generators can be reified into first-class functions and applications that use a little state to record which function is the continuation of the yield point. Thus, generators seem to fit perfectly with our desugaring strategy.

To convert programs to CPS, we take operators that can cause control-flow and reify each into a continuation function and appropriate application. These operators include simple sequences, loops combined with break and continue, try–except and try–finally combined with raise, and return. Our CPS transformation turns every expression into a function that accepts an argument for each of the above control operators, and turns uses of control operators into applications of the appropriate continuation inside the function. By passing in different continuation arguments, the caller of the resulting function has complete control over the behavior of control operators. For example, we might rewrite a try–except block from

```
try:
    raise Exception()
except e:
    print(e)
```

to

```
def except_handler(e):
    print(e)
except_handler(Exception())
```

In the case of generators, rewriting yield with CPS would involve creating a handler that stores a function holding the code for what to do next, and rewriting yield expressions to call that handler:

```
def f():
    x = 1
    yield x
    x += 1
    yield x

g = f()
g.__next__() # ==> 1
g.__next__() # ==> 2
g.__next__() # throws "StopIteration"
```

would be rewritten to something like\(^6\)

```
def f():
    def yielder(val, rest_of_function):
        next.to_call_next = rest_of_function
        return val

def next():
    return next.to_call_next()

def done():
    raise StopIteration()

def start():
    x = 1
    def rest():
        x += 1
        return yielder(x, done)
    return yielder(x, rest)

next.to_call_next = start
return { 'next' : next }```

\(^6\)This being a sketch, we have taken some liberties for simplicity.
g = f()
g['next']() # ==> 1
g['next']() # ==> 2
g['next']() # throws "StopIteration"

We build the yielder function, which takes a value, which it returns after storing a continuation argument in the to_call_next field. The next function always returns the result of calling this stored value. Each yield statement is rewritten to put everything after it into a new function definition, which is passed to the call to yielder. In other words, this is the canonical CPS transformation, applied in the usual fashion.

This rewriting is well-intentioned but does not work. If this program is run under Python, it results in an error:

```python
x += 1
UnboundLocalError: local variable 'x'
```

This is because Python creates a new scope for each function definition, and assignments within that scope create new variables. In the body of rest, the assignment x += 1 refers to a new x, not the one defined by x = 1 in start. This runs counter to traditional notions of functions that can close over mutable variables. And in general, with multiple assignment statements and branching control flow, it is entirely unclear whether a CPS transformation to Python function definitions can work.

The lesson from this example is that the interaction of non-traditional scope and control flow made a traditional translation not work. The straightforward CPS solution is thus incorrect in the presence of Python’s mechanics of variable binding. We now move on to describing how we can express Python’s scope in a more traditional lexical model. Then, in section 4.3 we will demonstrate a working transformation for Python’s generators.

### 4.2 Scope

Python has a rich notion of scope, with several types of variables and implicit binding semantics that depend on the block structure of the program. Most identifiers are local; this includes function parameters and variables defined with the = operator. There are also global and nonlocal variables, with their own special semantics within closures, and interaction with classes. Our core contribution to explaining Python’s scope is to give a desugaring of the nonlocal and global keywords, along with implicit local, global and instance identifiers, into traditional lexically-scoped closures. Global scope is still handled specially, since it exhibits a form of dynamic scope that isn’t straightforward to capture with traditional let-bindings.

We proceed by describing Python’s handling of scope for local variables, the extension to nonlocal, and the interaction of both of these features with classes.

#### 4.2.1 Declaring and Updating Local Variables

In Python, the operator = performs local variable binding:

```python
def f():
    x = 'local variable'
    return x
f() # ==> 'local variable'
```

The syntax for updating and creating a local variable are identical, so subsequent = statements mutate the variable created by the first.

```python
def f():
    x = 'local variable'
    x = 'reassigned'
    x = 'reassigned again'
    return x
f() # ==> 'reassigned again'
```

Crucially, there is no syntactic difference between a statement that updates a variable and one that initially binds it. Because bindings can also be introduced inside branches of conditionals, it isn’t statically determinable if a variable will be defined at certain program points. Note also that variable declarations are not scoped to all blocks—here they are scoped to function definitions:

```python
def f(y):
    if y > .5: x = 'big'
    else: pass
    return x
f(0) # throws an exception
f(1) # ==> "big"
```

Handling simple declarations of variables and updates to variables is straightforward to translate into a lexically-scoped language. \( \lambda \pi \) has a usual let form that allows for lexical binding. In desugaring, we scan the body of the function and accumulate all the variables on the left-hand side of assignment statements in the body. These are let-bound at the top of the function to the special \( \equiv \) form, which evaluates to an exception in any context other than a let-binding context (section 2). We use \( x := e \) as the form for variable assignment, which is not a binding form in the core. Thus, in \( \lambda \pi \), the example above rewrites to:
let (f = ⚪ in
  f :=
  fun (y) (no-var)
  let x = ⚪ in
  if (builtin-prim "num>" (y〈%float,0.5〉))
    x :=〈%str,"big"〉
    none
  (return x)
  f (〈%int,0〉)
  f (〈%int,1〉))

In the first application (to 0) the assignment will never happen, and the attempt to look up the ⚪-valued x in the return statement will fail with an exception. In the second application, the assignment in the then-branch will change the value of x in the store to a non-⚪ string value, and the string “big” will be returned.

The algorithm for desugaring scope is so far:

• For each function body:
  • Collect all variables on the left-hand side of = in a set locals, stopping at other function boundaries,
  • For each variable var in locals, wrap the function body in a let-binding of var to ⚪.

This strategy works for simple assignments that may or may not occur within a function, and maintains lexical structure for the possibly-bound variables in a given scope. Unfortunately, this covers only the simplest cases of Python’s scope.

### 4.2.2 Closing Over Variables

Bindings from outer scopes can be seen by inner scopes:

```python
def f():
  x = 'closed-over'

def g():
  return x

return g()
```

f() # ==> 'closed-over'

However, since = defines a new local variable, one cannot close over a variable and mutate it with what we’ve seen so far; = simply defines a new variable with the same name:

```python
def g():
  x = 'not affected'

def h():
  x = 'inner x'
  return x

return (h(), x)
```

g() # ==> ('inner x', 'inner x')

Thus, the presence or absence of a nonlocal declaration can change an assignment statement from a binding occurrence of a variable to an assigning occurrence. We augment our algorithm for desugaring scope to reflect this:

• For each function body:
  • Collect all variables on the left-hand side of = in a set locals, stopping at other function boundaries,
  • Let locals’ be locals with any variables in nonlocal declarations removed,
  • For each variable var in locals’, wrap the function body in a let-binding of var to ⚪.

Thus the above program would desugar to the following, which does not let-bind x inside the body of the function assigned to h.
def f(x, y):
    print(x); print(y); print(""")

class c:
    x = 4
    print(x); print(y)
    print("""
    def g(self):
        print(x); print(y); print(c)
    return c

f("x-value", "y-value")().g()

# produces this result:

x-value
y-value

4
y-value

x-value
y-value
<class '__main__.c'>

Figure 8: An example of class and function scope interacting

let g = ☠ in
g :=
fun ()
let x = ☠ in
let h = ☠ in
x := (%str,"not affected by h",{})
h :=
fun ()
x := (%str,"inner x",{})
(return x)
(return tuple(%tuple,(h () () x)))
g ()

The assignment to x inside the body of h behaves as a
typical assignment statement in a closure-based language
like Scheme, mutating the let-bound x defined in g.

4.2.3 Classes and Scope

We’ve covered some subtleties of scope for local and nonlo-
cal variables and their interaction with closures. What we’ve
presented so far would be enough to recover lexical scope
for a CPS transformation for generators if function bodies
contained only other functions. However, it turns out that we
observe a different closure behavior for variables in a class
definition than we do for variables elsewhere in Python, and
we must address classes to wrap up the story on scope.

Consider the example in figure 8. Here we observe an
interesting phenomenon: in the body of g, the value of the
variable x is not 4, the value of the most recent apparent
assignment to x. In fact, the body of g seems to "skip" the
scope in which x is bound to 4, instead closing over the outer
scope in which x is bound to "x-value". At first glance
this does not appear to be compatible with our previous
notions of Python’s closures. We will see, however, that the
correct desugaring is capable of expressing the semantics
of scope in classes within the framework we have already
established for dealing with Python’s scope.

Desugaring classes is substantially more complicated
than handling simple local and nonlocal cases. Consider the
element from figure 8, stripped of print statements:

def f(x, y):
    class c:
        x = 4
        print(x); print(y)
    def g(self): pass
    return c

In this example, we have three local scopes: the body of
the function f, the body of the class definition c, and the body
of the function g. The scopes of c and g close over the same
scope as f, but have distinct, non-nesting scopes themselves.
Figure 9 shows the relationship graphically. Algorithmically,
we add new steps to scope desugaring:

• For each function body:
  • For each nested class body:
    – Collect all the function definitions within the class
      body. Let the set of their names be defnames and
      the set of the expressions be defbodies,
    – Generate a fresh variable deflifted for each vari-
      able in defnames. Add assignment statements to the
      function body just before the class definition as-
4.3 Generators Redux

With the transformation to a lexical core in hand, we can return to our discussion of generators, and their implementation with a local CPS transformation.

To implement generators, we first desugar Python down to a version of λπ with an explicit yield statement, passing yields through unchanged. As the final stage of desugaring, we identify functions that contain yield, and convert them to generators via local CPS. We show the desugaring machinery around the CPS transformation in figure 11. To desugar them, in the body of the function we construct a generator object and store the CPS-ed body as a ____resume attribute on the object. The ____next__ method on the generator, when called, will call the ____resume closure with any arguments that are passed in. To handle yielding, we desugar the core yield expression to update the ____resume attribute to store the current normal continuation, and then return the value that was yielded.

Matching Python’s operators for control flow, we have five continuations, one for the normal completion of a statement or expression going onto the next, one for a return statement, one each for break and continue, and one for the exception throwing raise. This means that each CPS-ed expression becomes an anonymous function of five arguments, and can be passed in the appropriate behavior for each control operator.

We use this configurability to handle two special cases:

• Throwing an exception while running the generator
• Running the generator to completion

Figure 10: Full class scope desugaring
In the latter case, the generator raises a `StopIteration` exception. We encode this by setting the initial “normal” continuation to a function that will update `___resume` to always raise `StopIteration`, and then to raise that exception. Thus, if we evaluate the entire body of the generator, we will pass the result to this continuation, and the proper behavior will occur.

Similarly, if an uncaught exception occurs in a generator, the generator will raise that exception, and any subsequent calls to the generator will result in `StopIteration`. We handle this by setting the initial `raise` continuation to be code that updates `___resume` to always raise `StopIteration`, and then we raise the exception that was passed to the continuation. Since each `try` block in CPS installs a new exception continuation, if a value is passed to the top-level exception handler it means that the exception was not caught, and again the expected behavior will occur.

5. Engineering & Evaluation

Our goal is to have desugaring and $\lambda_\pi$ enjoy two properties:

- Desugaring translates all Python source programs to $\lambda_\pi$ (totality).
- Desugared programs evaluate to the same value as the source would in Python (conformance).

The second property, in particular, cannot be proven because there is no formal specification for what Python does. We therefore tackle both properties through testing. We discuss various aspects of implementing and testing below.

5.1 Desugaring Phases

Though we have largely presented desugaring as an atomic activity, the paper has hinted that it proceeds in phases. Indeed, there are four:

- Lift definitions out of classes (section 4.2.3).
- Let-bind variables (section 4.2.2). This is done second to correctly handle occurrences of `nonlocal` and `global` in class methods. The result of these first two steps is an intermediate language between Python and the core with lexical scope, but still many surface constructs.
- Desugar classes, turn Python operators into method calls, turn `for` loops into guarded `while` loops, etc.
- Desugar generators (section 4.3).

These four steps yield a term in our core, but it isn’t ready to run yet because we desugar to open terms. For instance, `print(5)` desugars to

```python
print (〈%int,5〉)
```

which relies on free variables `print` and `%int`.

---

Figure 11: The desugaring of generators
5.2 Python Libraries in Python

We implement as many libraries as possible in Python augmented with some macros recognized by desugaring. For example, the built-in tuple class is implemented in Python, but getting the length of a tuple defers to the \( \delta \) function:

```python
class tuple(object):
    def __len__(self):
        return ___delta("tuple-len", self)
...```

All occurrences of `___delta(str, e, ...)` are desugared to `(builtin-prim str (e ...))` directly. We only do this for library files, so normal Python programs can use `___delta` as the valid identifier it is. As another example, after the class definition of tuples, we have the statement

```python
___assign("%tuple", tuple)
```

which desugars to an assignment statement `%tuple := tuple`. Since %-prefixed variables aren’t valid in Python, this gives us a private namespace of global variables that are un-tamperable by Python. Thanks to these decisions, this project produces far more readable desugaring output than a previous effort for JavaScript [6].

5.3 Performance

\( \lambda_\pi \) may be intended as a formal semantics, but composed with desugaring, it also yields an implementation. While the performance does not matter for semantic reasons (other than programs that depend on time or space, which would be ill-suited by this semantics anyway), it does greatly affect how quickly we can iterate through testing!

The full process for running a Python program in our semantics is:

1. Parse and desugar roughly 1 KLOC of libraries implemented in Python
2. Parse and desugar the target program
3. Build a syntax tree of several built-in libraries, coded by building the AST directly in Racket
4. Compose items 1-3 into a single \( \lambda_\pi \) expression
5. Evaluate the \( \lambda_\pi \) expression

Parsing and desugaring for (1) takes a nontrivial amount of time (40 seconds on the first author’s laptop). Because this work is needlessly repeated for each test, we began caching the results of the desugared library files, which reduced the testing time into the realm of feasibility for rapid development. When we first performed this optimization, it made running 100 tests drop from roughly 7 minutes to 22 seconds. Subsequently, we moved more functionality out of \( \lambda_\pi \) and into verbose but straightforward desugaring, causing serious performance hit: running 100 tests now takes on the order of 20 minutes, even with the optimization.

5.4 Testing

Python comes with an extensive test suite. Unfortunately, this suite depends on numerous advanced features, and as such was useless as we were building up the semantics. We therefore went through the test suite files included with CPython, April 2012 and ported a representative suite of 205 tests (2600 LOC). In our selection of tests, we focused on orthogonality and subtle corner-cases. The distribution of those tests across features is reported in figure 12. On all these tests we obtain the same results as CPython.

It would be more convincing to eventually handle all of Python’s own `unittest` infrastructure to run CPython’s test suite unchanged. The `unittest` framework of CPython unfortunately relies on a number of reflective features on modules, and on native libraries, that we don’t yet cover. For now, we manually move the assertions to simpler if-based tests, which also run under CPython, to check conformance.

5.5 Correspondence with Redex

We run our tests against \( \lambda_\pi \), not against the Redex-defined reduction relation for \( \lambda_\pi \). We can run tests on \( \lambda_\pi \), but performance is excruciatingly slow: it takes over an hour to run complete individual tests under the Redex reduction relation. Therefore, we have been able to perform only limited testing for conformance by hand-writing portions of the environment and heap (as Redex terms) that the Python code in the test uses. Fortunately, executing against Redex should be parallelizable, so we hope to increase confidence in the Redex model as well.

---

Feature | # of tests | LOC
---|---|---
Built-in Datatypes | 81 | 902
Scope | 39 | 455
Exceptions | 25 | 247
(Multiple) Inheritance | 16 | 303
Properties | 9 | 184
Iteration | 13 | 214
Generators | 9 | 129
Modules | 6 | 58
Total | 205 | 2636

Figure 12: Distribution of passing tests

---

\( ^{8} \) We could not initially use existing implementations of these in Python for bootstrapping reasons: they required more of the language than we supported.

\( ^{9} \) http://www.python.org/getit/releases/3.2.3/
6. Future Work and Perspective

As section 5 points out, there are some more parts of Python we must reflect in the semantics before we can run Python’s test cases in their native form. This is because Python is a large language with extensive libraries, a foreign-function interface, and more.

Libraries aside, there are some interesting features in Python left to tackle. These include special fields, such as the properties of function objects that compute the content of closures, complex cases of destructuring assignments, a few reflective features like the metaclass form, and others.

More interestingly, we are not done with scope! Consider `locals`, which returns a dictionary of the current variable bindings in a given scope:

```python
def f(x):
    y = 3
    return locals()

f("val") # ==> {"x": 'val', 'y': 3}
```

This use of `locals` can be desugared to a clever combination of assignments into a dictionary along with variable assignments, which we do. However, this desugaring of `locals` relies on it being a strictly local operation (for lack of a better word). But worse, `locals` is a value!

```python
def f(x, g):
    y = 3
    return g()

f("x-val", locals)
# ==> {"x": 'x-val', 'y': 3,
#        'g': <builtins.function locals>}
```

Thus, any application could invoke `locals`. We would therefore need to deploy our complex desugaring everywhere we cannot statically determine that a function is not `locals`, and change every application to check for it. Other built-in values like `super` and `dir` exhibit similar behavior.

On top of this, `import` can splice all identifiers (*) from a module into local scope. For now, we handle only `imports` that bind the module object to a single identifier. Indeed, even Python 3 advises that `import *` should only be used at module scope. Finally, we do not handle `exec`, Python’s “eval” (though the code-loading we do for modules comes close). Related efforts on handling similar operators in JavaScript [6] are sure to be helpful here.

We note that most traditional analyses would be seriously challenged by programs that use functions like `locals` in a higher-order way, and would probably benefit from checking that it isn’t used in the first place. We don’t see the lack of full support for such functions as a serious weakness of λπ, or an impediment to reasoning about most Python programs. Rather, it’s an interesting future challenge to handle a few of these remaining esoteric features. It’s also useful to simply call out the weirdness of these operators, which are liable to violate the soundness of otherwise-sound program tools.

Overall, what we have learned most from this effort is how central scope is to understanding Python. Many of its features are orthogonal, but they all run afoot on the shoals of scope. Whether this is intentional or an accident of the convoluted history of Python’s scope is unclear (for example, see the discussion around the proposal to add `nonlocal` [15]), but also irrelevant. Those attempting to improve Python or create robust sub-languages of it—whether for teaching or for specifying asset-backed securities—would do well to put their emphasis on scope first, because this is the feature most likely to preclude sound analyses, correctness-preserving refactorings, and so on.

7. Related Work

We are aware of only one other formalization for Python: Smeding’s unpublished and sadly unheralded master’s thesis [9]. Smeding builds an executable semantics and tests it against 134 hand-written tests. The semantics is for Python 2.5, a language version without the complex scope we handle. Also, instead of defining a core, it directly evaluates (a subset of) Python terms. Therefore, it offers a weaker account of the language and is also likely to be less useful for certain kinds of tools and for foundational work.

There are a few projects that analyze Python code. They are either silent about the semantics or explicitly eschew defining one. We therefore do not consider these related.

Our work follows the path laid out by λJS [7] and its follow-up [6], both for variants of JavaScript.

Acknowledgments

We thank the US NSF and Google for their support. We thank Caitlin Santone for lending her graphical expertise to figure 9. The paper title is entirely due to Benjamin Lerner. We are grateful to the Artifact Evaluation Committee for their excellent and detailed feedback, especially the one reviewer who found a bug in iteration with generators.

This paper was the result of an international collaboration resulting from an on-line course taught by the first and last authors. Several other students in the class also contributed to the project, including Ramakrishnan Muthukrishnan, Bo Wang, Chun-Che Wang, Hung-I Chuang, Kelvin Jackson, Victor Andrade, and Jesse Millikan.

Bibliography

Appendix: Confusing Rename Refactorings

This program:

```python
def f(x):
    class C(object):
        x = "C’s x"
        def meth(self):
            return x + ', ' + C.x
        return C

f('input x')().meth()
```

confuses the variable rename refactoring of all the Python IDEs we tried. We present these weaknesses to show that getting a scope analysis right in Python is quite hard! We found these tools by following recommendations on StackOverflow, a trusted resource. Two of the tools we tested, PyCharm and WingWare IDE, are products that developers actually purchase to do Python development (we performed this experiment in their free trials).

For PyCharm, if we rename the x parameter to y, the class variable x also gets changed to y, but the access at C.x does not. This changes the program to throw an error. If we instead select the x in C.x and refactor to y, The class variable and use of x change, but the original definition’s parameter does not. This changes the behavior to an error again, as y is an unbound identifier in the body of meth.

PyDev has the same problem as PyCharm with renaming the function’s parameter. If we instead try rename the x in the body of C, it gets it mostly right, but also renames all the instances of x in our strings (something that even a parser, much less a lexical analyzer should be able to detect):

```python
def f(x):
    class C(object):
        y = "C’s y"
        # we highlighted the x before = above
        # and renamed to y
        def meth(self):
            return x + ', ' + C.y
        return C

f('input y')().meth()
```

WingWare IDE for Python is less obviously wrong: it pops up a dialog with different bindings and asks the user to check the ones they want rewritten. However, if we try to refactor based on the x inside the method, it doesn’t give us an option to rename the function parameter, only the class variable name and the access at C.x. In other cases it provides a list that contains a superset of the actual identifiers that should be renamed. In other words, it not only overapproximates (which in Python may be inevitable), it also (more dangerously) undershoots.