Metaprogramming with Macros

Eugene Burmako
LAMP, I&C, EPFL

***Abstract*—Macros realize the notion of *textual abstraction*. Textual abstraction consists of recognizing pieces of text that match a specification and replacing them according to a procedure.**

In the focus of the study is semantics of syntactic macros in lexically scoped programming languages. We highlight the problems of *hygiene* and *referential transparency* and describe the solutions employed Template Haskell [1], Nemerle [2,18] and Racket [3,20].

We discuss integration of hygienic macros into statically typed languages and plan to improve upon state of the art by providing a flexible type system for syntax templates and uncovering synergies with high-level language features such as path-dependent types and implicits [4].

***Index Terms*—metaprogramming, macros, quasiquotes, hygiene, referential transparency**

I. Introduction

PROCEDURAL abstraction is pervasive. Factoring out parameterized fragments of programs into procedures is a conventional best practice.

Modern programming languages integrate the notion of procedures into their semantics. Procedures are viewed as independent programs that can communicate with the main program. As of such they can be manipulated as units, and big procedures can be built from the smaller ones. This is a powerful way to manage complexity of software systems.

Proposal submitted to committee: September 3rd, 2012; Candidacy exam date: September 10th, 2012; Candidacy exam committee: Christoph Koch, Martin Odersky, Viktor Kuncak.

This research plan has been approved:

Date: _____

Doctoral candidate: _____
(name and signature)

Thesis director: _____
(name and signature)

Thesis co-director: _____
(if applicable) (name and signature)

Doct. prog. director: _____
(R. Urbanke) (signature)

EDIC-ru/05.05.2009

However procedural abstraction is sometimes not enough, because its manifestations are bound by language syntax and it operates within the semantics of the language.

For example, in most programming languages it is impossible to define short-circuiting logical operators as procedures, because procedures are usually not in control of operational semantics. Another example in this vein is a C-like `for` loop, which supports optional prologue that introduces variables visible in its body. Procedures typically cannot abstract over variable bindings, so they cannot express this language construct.

Textual abstraction consists of recognizing pieces of text that match a specification and replacing them according to a procedure. Matched fragments are referred to as macro calls or macro applications, and procedures that transform them are dubbed macros or macro transformers. The process of applying macros is called macro expansion [5].

Having means of textual abstraction in their toolbox, programmers can use a multitude of techniques, some of which are:

- reification (providing code as data),
- language virtualization (overloading/overriding semantics of the original programming language to enable deep embedding of DSLs),
- domain-specific optimization (application of optimizations such as inlining or fusion based on knowledge about the program being optimized),
- static verification (using reified representation of the program and, possibly, its contracts defined alongside the program, to verify invariants at compile time),
- algorithmic program construction (generation of code that is tedious to write with supported abstractions).

One possibility to implement a macro system is having it as a standalone tool operating on character streams. This gives rise to lexical macros. Such a design warrants simplicity of the implementation, but undermines robustness, because macros operating on lexical level have no mechanism that prevents generation of syntactically invalid programs.

Alternative approach is to integrate a macro expander into the compiler and have macros work with syntax trees, introducing *syntactic macros*. In this model macro application is a node in the program tree, and macro expansion produces a new node that replaces the macro application without distorting the structure of the program.

A significant body of recent research in textual abstraction has been done on syntax extensibility ([6–8] to name only a few). This paper however focuses on semantics of macro-enabled languages, i.e. on preventing macros from expanding into nonsense and finding synergies with existing language features.

```
(let ((x 40) (y 2)) (print (+ x y)))

((lambda (x y) (print (+ x y))) (40 2))
```

a) Examples of a macro application and a macro expansion of the `let` macro

```
(defmacro let args
  (cons
    (cons 'lambda
          (cons (map car (car args))
                (cdr args)))
    (map cadr (car args))))
```

b) Creates the result at low level (manipulates S-expressions with standard symbol and list processing functions)

```
(defmacro let (decls body)
 `(
    (lambda
      ,(map car decls)    ;; (x y)
      ,body               ;; (print (+ x y))
    ,@(map cadr decls))) ;; (40 2)
```

c) Uses quasiquotes [10] as a templating mechanism (backquote introduces a static code template, commas (also known as unquotes) insert dynamic values into the template)

```
(define-syntax let
  (syntax-rules ()
    ((let ((name expr) ...) body ...)
    ((lambda (name ...) body ...) expr ...))))
```

d) Macro-by-example notation [11] (uses a tree matching macro that can extract and reassemble fragments of syntax objects; ellipses capture recurrent parts of the input)

```
(define-syntax (let stx)
  (syntax-parse stx
    ((let ((name:identifier expr:expr) ...) body:expr ...)
     #:fail-when (check-duplicate #'(var ...))
                 "duplicate variable name"
     #'((lambda (name ...) body ...) expr ...))))
```

e) Macro-by-example notation augmented with a syntax specification [12] (colons denote syntax classes, which are first-class and can be built from the ground up).

Figure 1. Assorted implementations of the `let` macro in Lisp dialects

II. Examples

`let` is a language construct typical to functional programming. It introduces a scope for a computation and brings temporary variables with provided values into that scope. To implement `let` the compiler might wrap the computation in a lambda abstraction and apply it right away (Figure 1a).

This notion cannot be abstracted procedurally, because the body of the computation typically contains free variables. However textual abstraction fits the bill, because macros can manipulate the program on a level where bindings don't exist and therefore don't impose restrictions. To set up a stage for further chapters, let's implement the `let` macro in Lisp.

The most straightforward solution to the problem is a low-level macro transformer (Figure 1b). It takes an S-expression that represents a macro application, destructures it using standard list manipulation functions, such as `car` and `cdr`, and creates a new S-expression with `cons`. Even in this simple example this notation is very noisy. It's quite difficult to figure out expected shapes of input and output expressions from the imperative algorithm.

Quasiquotes [10] make it possible to reduce obscurity of the macro by providing a domain-specific language for syntax templates (Figure 1c). The quasiquote operator (`` ` ``) demarcates a static template. Quasiquoted code is inserted verbatim into the output (that's why there's no longer need in explicit `cons`'ing). Unquote operators (`,` and `,@`) interrupt a quasiquote, producing "holes" filled in with dynamically calculated data. For example, for `let` we statically know the shape of code to produce (an application of a lambda abstraction) - this makes up the static part of the quasiquote. On the other hand, body and parameters of the lambda as well as the arguments of the application may vary from expansion to expansion - this is the dynamic part.

```
(defmacro or (x y)
  '(let ((temp ,x))
     (if temp temp ,y)))

(or 42 "the result is 42")
```

a) A simple yet erroneous implementation of the `or` macro

```
(let ((temp "451 Fahrenheit"))
  (or null temp))

(let ((temp "451 Fahrenheit"))
  (let ((temp null))
    (if temp temp temp)))
```

b) Violation of hygiene [5]: binding established during expansion affects call site

```
(let ((if hijacked))
  (or true false))

(let ((if hijacked))
  (let ((temp true))
    (if temp temp false)))
```

c) Violation of referential transparency [9]: binding established during expansion is affected by call site

```
(define-syntax forever
  (syntax-rules ()
    ((forever body ...)
     (call/cc (lambda (abort)
                (let loop () body ... (loop)))))))

(forever (print 4) (print 2) (abort))
```

d) Intentional variable capture: infinite looping construct `forever` provides a predefined identifier to break from the loop. `abort` is introduced during macro expansion, but it should be visible to the body of the loop, which comes from the macro call site.

Figure 2. Intentional and unintentional variable capture

Another simplification of the macro can be achieved with MBE, the macro-by-example notation [11]. In their seminal work Kohlbecker and Wand came up with a specification of a pattern matcher that matches singular and repetitive parts of S-expressions. Identifiers which appear in the input pattern are treated as pattern variables, ellipses (`...`) used as a last element of a list that contains pattern variables denote repetition and, when nested, can capture lists of arbitrary depth. The revised version of the `let` macro is particularly minimalistic (Figure 1d).

A recent development of the MBE syntax has been proposed by Culpepper and Felleisen [12]. Their refinement addresses the need for principled input validation and error reporting. Indeed, MBE covers the success path, but doesn't help with detecting errors. For example, duplicate identifier names as in `(let ((x 40) (x 2)) (print (+ x y))` will go unnoticed until the compiler gets to the resulting lambda form, which will produce confusing error messages. Authors enhance MBE with both declarative and procedural means of validation (Figure 1e). Colons next to the names of pattern variables denote syntax classes, which put restrictions on the shape of the variables, `#:fail-when` clauses can contain validation code and error messages (this doesn't cover all the capabilities of syntax specifications, please, refer to [12] for details). These validation facilities can be packed into custom syntax classes, which can be built from the ground up.

III. Bindings

Syntactic macros operate on ASTs, so they cannot produce syntactically invalid code, but nothing prevents such macros from making semantic errors, i.e. expanding into syntactically valid nonsense. The most blatant mistakes will result in type errors, however there's an entire class of oversights that might silently change the behavior of the program.

When writing procedures in lexically scoped languages, programmers typically don't need to think about name clashes between variables declared inside procedures and at their call sites. Except for recursion, scopes of procedure bodies and procedure call sites are different, so neither variables defined inside the procedures can change the semantics at the call site, nor vice versa.

In a macro-enabled language, especially in presense of quasiquotes which make macros look like normal procedures, this intuition becomes compromised, because after macro expansion the scopes of macro definition and macro call site are mechanically merged by the expander.

Conider the `or` macro shown on Figure 2a. This macro picks one of its two arguments based on the truthiness of the first argument. To prevent double evaluation, the macro introduces a temporary variable `temp` that holds the result of evaluation of the first argument. The temporary variable is then tested with `if`, which selects the resulting value.

As simple as this code can be, it is also incorrect, having two potential bugs.

The first problem happens when the call site defines its own variable named `temp` and relies on it in the second argument of a call to `or` (Figure 2b). After the expansion, two `temp`s clash, which produces incorrect results if the first argument of the call is falsy. This problem is dubbed *hygiene violation*, and the macro is said to introduce the temporary variable *unhygienically*.

In his thesis [5] Kohlbecker defines the hygiene condition and presents a macro system that automatically prevents such naming collisions:

> Hygiene Condition for Macro Expansion.
> Generated identifiers that become binding instances in the completely expanded program must bind only identifiers that are generated during the same transcription step.

The second problem is in a sense dual to the first one. It happens when the call site redefines one of the procedures or macros used in the expansion. For example, on Figure 2c the call site hijacks the meaning of `if`, which destroys the original intent of the macro author. This is a *referential opaqueness* problem.

Dybvig et al. [9] build on Kohlbecker's work and devise a macro expander that automatically avoids this class of errors:

> Macros defined in [our] high-level specification language are referentially transparent in the sense that a macro-introduced identifier refers to the binding lexically visible where the macro definition appears rather than to the top-level binding or to the binding visible where the macro call appears.

This notion is also called *cross-stage persistence* in a sense that bindings in place at the compilation stage are persisted into the runtime stage. When viewed in this light, it becomes apparent that referential transparency for macros can only work for references to top-level definitions. Therefore common practice is to report cross-stage references to local variables as errors [9].

Despite the conveniency of automatic segregation of the scopes of macro definitions and macro call sites, at times it is necessary to have them melded. A looping macro `forever` from Figure 2d demonstrates the need for this. Expansion of this macro is an infinite loop that provides a magic identifier `abort` to exit the loop. In this case automatic hygiene facility will do harm, making the body of the loop unable to see the loop control lever.

Because of similar situations some programmers argue in favor of manual control over the scoping discipline, proposing manual (or macro-powered!) renaming of identifiers whose capture would be undesired [13,14].

Another popular line of thought favors further empowerment of macro systems to minimize the necessity for low-level tinkering. In fact, one of the macro systems reviewed below provides a design pattern [3] that solves the `abort` challenge in a fully hygienic way (i.e. without introducing identifiers with manually provided names).

IV. Typechecking

An important feature of macros is that macro expansion happens before the code is executed (below in the paper this phase is referred to as *compile-time*, despite that, strictly speaking, interpreted languages can also support macros). Therefore macro expansions can be checked for absense of semantic errors (to the extent supported by the language) in an automatic fashion, without requiring effort from the programmer.

What's even better is that there is another line of defense. Macros operate on syntax objects, which are snippets of the code to be, thus it might be a good idea to typecheck the code represented by these snippets in advance. This amounts to extending typechecking to the before-compile-time phase (as in compile-time of macros themselves).

One approach is to treat quasiquotes similarly to functions, requiring all free variables to be bound in the enclosing lexical scope and assigning quasiquotes a definitive type (Figure 3a). This strategy is used in MetaML [15], a statically typechecked language with special support for program generation.

All quasiquotes in MetaML are typechecked individually and receive one of the `<a>` types (which means "code that evaluates to a value of type `a`"). For example, on Figure 3a the quasiquote in the body of `pow` represents a function from `int` to `int`, so it has the `<int -> int>` type.

Programs in MetaML can generate and run programs at runtime, these programs can do the same, and so on. Quasiquotes here represent templates of the programs to be generated and delineate execution stages. Due to its specific nature, MetaML requires strong guarantees from the type system. At runtime, if program generation fails because of an error in the metaprogram, it's too late to report typechecking errors.

Unfortunately this means that MetaML has to give up the freedom of arbitrary manipulations of syntax objects, the most important of which are destructuring and introduction of new bindings. (e.g. in MetaML its impossible to express `let` as a macro).

MacroML [16,17], a macro system built atop of MetaML, provides a clever way to partially alleviate this restriction, letting the programmer introduce new binding constructs in a very controlled setting. This is achieved by having the type system track not only the regular environments containing declared macros and variables, but also a so called body parameter environment, which remembers which parts of quasiquotes see which variables introduced by those quasiquotes.

This still doesn't allow arbitrary manipulations, e.g. similarly to MetaML there is no way to pull apart a syntax object and reassemble it using a quasiquote as a template (as done in the `let` macro on Figure 1).

Therefore later research on macros focused on relaxing the typechecking discipline of MetaML.

```
-| fun pow n = <fn x =>
    ~(if n = 0 then <1> else <x * (~(pow (n - 1)) x)>)>;
val pow = fn : int -> <int -> int>

-| val cube = (pow 3);
val cube = <(fn a => x %* x %* x %* 1)> : <int -> int>

-| (run cube) 5;
val it = 125 : int
```

a) Strict typechecking [15]: each quasiquote is typechecked individually and is assigned a "code of something" type (`<a>` stands for code that evaluates to a value of type `a`).

```
[| 'a' + True |] -- rejected

printf :: String -> Expr -- allowed
$(printf "Error: %s on line %d") "urk" 341

f :: Q Type -> Q [Dec] -- rejected
f t = [d| data T = MkT $t; g (MkT x) = g+1 |]
```

b) Lenient typechecking [1]: quasiquotes are sanity checked to prevent blatant errors and are required to have their bindings established before macro expansions kick in. No additional checks are done, and all quasiquotes get the same type-agnostic `Expr` type.

```
macro using(name, expr, body) {
  <[
    def $name = $expr;
    try { $body } finally { $name.Dispose() }
  ]>
}
using(db, Database("localhost"), db.LoadData())
```

c) No typechecking [2,18]: quasiquotes are not typechecked at all until a macro that uses them expands, after which the result is typechecked as a whole. This provides maximum freedom, permitting even unquotes in binding positions (like in `def $name`).

Figure 3. Typechecking strategies for quasiquotes

Template Haskell [1] employs a lenient typechecking strategy for meta-programs (but not for expanded programs, which are typechecked with the usual rigor of Haskell).

Unlike in MetaML, the type of syntax objects is non-polymorphic - `Expr`, not `Expr a` - and Template Haskell doesn't try to provide up-front guarantees that meta-programs produce well-typed code. For example, well-typedness of the `printf` macro application (Figure 3b) is checked on the spot, immediately after the expansion, not ahead of time.

However Template Haskell does require the bindings of variables used in quasiquotes to be resolved statically. This is done to enforce the "static scope is absolute" design principle (which is a rehash of hygiene and referential transparency conditions outlined in the previous chapter) as stated in [19]:

> If an occurrence of a variable `x` looks as if it is lexically bound to an enclosing binding for `x`, then it is so bound, and no Template-Haskell transformation can threaten that binding.

Unfortunately this lightweight typecheck has proven to impose unnecessary limitations. For example, the meta-program `f` from Figure 3b that uses another meta-program `t` to insert a type declaration in a quasiquote is rejected. This happens because the compiler cannot prove that the usage of this type declaration (`MkT x`) is correct for all possible `t`s.

It is due to this reason that Nemerle [2,18], a macro-enabled language inspired by Template Haskell, has ceased typechecking quasiquotes altogether. As we will see below this doesn't prevent the macro expander in Nemerle from being hygienic, but with respect to typechecking this brings the evolution to square one.

V. State of the art

This chapters discusses macro systems in Template Haskell [1], Nemerle [2,18] and Racket [3,20], which to the best of our knowledge comprise up-to-date developments in the field of typechecking and hygienic expansion of hygienic macros.

References

[1] T. Sheard and S. Peyton Jones, Template meta-programming for Haskell. ACM SIGPLAN Notices, 2002.

[2] K. Skalski, M. Moskal and P. Olszta, Meta-programming in Nemerle. Generative Programming and Component Engineering, 2004.

[3] E. Barzilay, R. Culpepper and M. Flatt, Keeping it Clean with Syntax Parameters. Scheme and Functional Programming Workshop, 2011.

[4] M. Odersky, L. Spoon and B. Venners, Programming in Scala 2nd Edition. Artima, 2010.

[5] E. Kohlbecker, Syntactic Extensions in the Programming Language Lisp. PhD thesis, Indiana University, 1986.

[6] K. Atkinson and M. Flatt, Adapting Scheme-like macros to a C-like language. Scheme and Functional Programming, 2011.

[7] E. Allen, R. Culpepper, J-D. Nielsen, J. Rafkind and S. Ryu, Growing a Syntax. Foundations of Object-Oriented Languages, 2009.

[8] S. Erdweg, T. Rendel, C. Kästner and K. Ostermann, SugarJ: Library-based syntactic language extensibility, Object-Oriented Programming, Systems, Languages, and Applications, 2011.

[9] R. Dybvig, R. Hieb and C. Bruggeman, Syntactic Abstraction in Scheme. Lisp and Symbolic Computations, 1992.

[10] A. Bawden, Quasiquotation in Lisp. Partial Evaluation and Program Manipulation, 1999.

[11] E. Kohlbecker, M. Wand, Macro-by-Example: Deriving Syntactic Transformations from their Specifications. Principles of Programming Languages, 1987.

[12] R. Culpepper, M. Felleisen, Fortifying Macros. International Conference on Functional Programming, 2010.

[13] D. Hoyte, Let Over Lambda. Lulu, 2008.

[14] W. Clinger, Hygienic macros through explicit renaming. ACM SIGPLAN Lisp Pointers, 1991.

[15] W. Taha, Multi-Stage Programming: Its Theory and Applications. PhD Thesis, Oregon Graduate Institute, 1999.

[16] S. Ganz, A. Sabry, W. Taha, Macros as Multi-Stage Computations: Type-Safe, Generative, Binding Macros in MacroML. International Conference on Functional Programming, 2001.

[17] W. Taha and P. Johann, Staged Notational Definitions. Generative Programming and Component Engineering, 2003.

[18] K. Skalski, Syntax-Extending and Type-Reflecting Macros in an Object-Oriented Language. MsC Thesis, University of Wrocław, 2005.

[19] T. Sheard and S. Peyton Jones, Notes on Template Haskell Version 2. Glasgow Haskell Compiler, 2003.

[20] M. Flatt and PLT, Reference: Racket, PLT Inc., 2010.