

Sweet.js - Hygienic Macros for JavaScript

Tim Disney
UC Santa Cruz

Nate Faubion

David Herman
Mozilla

1. Introduction

Sweet.js is a new hygienic macro system for JavaScript.

Macros systems have a long history in the design of extensible programming languages going back at least to lisp as a tool to provide programmers syntactic flexibility.

While powerful macro systems have been used extensively in lisp derived languages there has been considerable less movement for macros systems for languages with an expression based syntax such as JavaScript. This is due to a variety of technical reasons that have held back macro systems in expression based languages which we address in this paper.

Recently the Honu project has shown how to overcome some of the existing challenges in developing a macro system for expression based language. The Honu technique was designed for an idealized JavaScript like language. In this paper we show how to extend the ideas of Honu for full JavaScript and present additional techniques that target expression based languages.

The design of sweet.js attempts to overcome the following technical challenges:

- a correct implementation of `read` that structures the token stream before expansion begins
- parser class annotation (e.g. `:expr`) in patterns to allow macro authors easier declaration of a macro shape
- operator overloading
- infix macros
- the `invoke` primitive to allow custom parser classes and more powerful matching

2. Overview

TODO: syntax, main features etc. . .

3. Read

The syntax of JavaScript presents a challenge to correctly implement the critical `read` function. This challenge is not present in Honu because their language is an idealized syntax that misses the problematic interaction of delimiters and regular expression literals.

TODO: motivate `read` with examples etc.

$$\begin{aligned} e &::= x \mid /x/ \\ &\mid e(e) \\ &\mid \text{if } (e)e \end{aligned}$$

3.1 Proof of read

We first define a simplified grammar that captures just the essential complexity we want to address, namely the interaction between the division symbol `/` and the regular expression literal `/x/`.

$$\begin{aligned} \text{Token} &::= x \mid / \\ \text{TokenTree} &::= x \mid / \mid /x/ \end{aligned}$$

We also define Expr' with the same productions as Expr but defined over TokenTree^* instead of Token^* . So the terminal `/x/` in the second production of Expr is three distinct tokens (`/`, `x`, and `/`) while it is a single token in Expr' .

We define the grammar of a Program_e as a function from Token to Expr .

$$\text{Program}_e :: \text{Token} \rightarrow \text{Expr}$$

Theorem 1 (Parse Equivalence).

$$\forall s. s \in \text{Program}_e \Leftrightarrow \text{read}(s, \epsilon, b) \in \text{Program}'_e$$

Proof.

By showing parse equivalence for each non-terminal in the grammar. Details in the appendix. \square

The grammar we present is a simplified version of the grammar specified in the ES5 specification. The simplification we are making do not affect the `read` algorithm, and only serve to shorten its presentation. In particular, functions can only have a single parameter, function bodies are not allowed to be empty and a few other minor changes. We only address the `=` assignment operator and only the `/` and `+` binary operators.

Figure 1: Simplified ES Grammar

$PrimaryExpr_x$	$::=$	x
$PrimaryExpr_{/x/}$	$::=$	$/ \cdot x \cdot /$
$PrimaryExpr_{\{x:e\}}$	$::=$	$\{x:AssignExpr_e\}$
$PrimaryExpr_{(e)}$	$::=$	$(AssignExpr_e)$
$MemberExpr_e$	$::=$	$PrimaryExpr_e$
$MemberExpr_e$	$::=$	$FunctionExpr_e$
$MemberExpr_{e.x}$	$::=$	$MemberExpr_e \cdot x$
$CallExpr_{e(e')}$	$::=$	$MemberExpr_e (AssignExpr_{e'})$
$CallExpr_{e(e')}$	$::=$	$CallExpr_e (AssignExpr_{e'})$
$CallExpr_{e.x}$	$::=$	$CallExpr_e \cdot x$
$BinaryExpr_e$	$::=$	$CallExpr_e$
$BinaryExpr_{e/e'}$	$::=$	$BinaryExpr_e / BinaryExpr_{e'}$
$BinaryExpr_{e+e'}$	$::=$	$BinaryExpr_e + BinaryExpr_{e'}$
$AssignExpr_e$	$::=$	$BinaryExpr_e$
$AssignExpr_{e=e'}$	$::=$	$CallExpr_e = AssignExpr_{e'}$
$StmtList_e$	$::=$	$Stmt_e$
$StmtList_{e e'}$	$::=$	$StmtList_e Stmt_{e'}$
$Stmt_{\{e\}}$	$::=$	$\{StmtList_e\}$
$Stmt_{x:e}$	$::=$	$x : Stmt_e$
$Stmt_e$	$::=$	$AssignExpr_e ; \quad \text{where lookahead} \neq \{ \text{or function} \}$
$Stmt_{\text{if}(e) e'}$	$::=$	$\text{if } (AssignExpr_e) Stmt_{e'}$
$Stmt_{\text{return}}$	$::=$	return
$Stmt_{\text{return } e}$	$::=$	$\text{return [no line terminator here] } AssignExpr_e ;$
$FunctionDecl_{\text{function } x(x') \{e\}}$	$::=$	$\text{function } x(x') \{SourceElements_e\}$
$FunctionExpr_{\text{function } x(x') \{e\}}$	$::=$	$\text{function } x(x') \{SourceElements_e\}$
$SourceElement_e$	$::=$	$Stmt_e$
$SourceElement_e$	$::=$	$FunctionDecl_e$
$SourceElements_e$	$::=$	$SourceElement_e$
$SourceElements_{e e'}$	$::=$	$SourceElements_e SourceElement_{e'}$
$Program_e$	$::=$	$SourceElements_e$
$Program$	$::=$	ϵ

Figure 2: Simplified Read Algorithm

$Punctuator ::= / + : ; = .$	$isExprPrefix : TokenTree^* \rightarrow Bool \rightarrow Int$
$Keyword ::= return function if$	$isExprPrefix(\epsilon, true, l) = true$
$Token ::= Punctuator Keyword$	$isExprPrefix(p \cdot /, b, l) = true$
$\quad x \{ () \}$	$isExprPrefix(p \cdot +, b, l) = true$
$TokenTree ::= Punctuator Keyword$	$isExprPrefix(p \cdot =, b, l) = true$
$\quad x /x/ (t) \{t\}$	$isExprPrefix(p \cdot :, b, l) = b$
$x \in Variable$	$isExprPrefix(p \cdot return^l, b, l') = false \text{ if } l \neq l'$
$r \in RegexBody$	$isExprPrefix(p \cdot return^l, b, l') = true \text{ if } l = l'$
$s \in Token^*$	$isExprPrefix(p, b, l) = false$
$t, p \in TokenTree^*$	
$read : Token^* \rightarrow TokenTree^* \rightarrow Bool \rightarrow TokenTree^*$	
$read(/ \cdot x \cdot / \cdot s, \epsilon, b)$	$= /x/ \cdot read(s, /x/, b)$
$read(/ \cdot x \cdot / \cdot s, p \cdot p', b)$	$= /x/ \cdot read(s, p \cdot p' \cdot /x/, b)$
$\quad \text{if } p' \in Punctuator \text{ or } Keyword$	
$read(/ \cdot x \cdot / \cdot s, p \cdot if \cdot (t), b)$	$= /x/ \cdot read(s, p \cdot if \cdot (t) \cdot /x/, b)$
$read(/ \cdot x \cdot / \cdot s, p \cdot p' \cdot function^l \cdot x \cdot (t) \cdot \{t'\}, b)$	$= /x/ \cdot read(s, p \cdot p' \cdot function^l \cdot x \cdot (t) \cdot \{t'\} \cdot /x/, b)$
$\quad \text{if false} = isExprPrefix(p', b, l)$	
$read(/ \cdot x \cdot / \cdot s, p \cdot \{t\}^l, b)$	$= /x/ \cdot read(s, p \cdot \{t\}^l \cdot /x/, b)$
$\quad \text{where false} = isExprPrefix(p, b, l)$	
$read(/ \cdot s, p \cdot x, b)$	$= / \cdot read(s, p \cdot x \cdot /, b)$
$read(/ \cdot s, p \cdot /x/, b)$	$= / \cdot read(s, p \cdot /x/ \cdot /, b)$
$read(/ \cdot s, p \cdot (t), b)$	$= / \cdot read(s, p \cdot (t) \cdot /, b)$
$read(/ \cdot s, p \cdot p' \cdot function^l \cdot x \cdot (t) \cdot \{t'\}, b)$	$= / \cdot read(s, p \cdot p' \cdot function^l \cdot x \cdot (t) \cdot \{t'\} \cdot /, b)$
$\quad \text{if true} = isExprPrefix(p', b, l)$	
$read(/ \cdot s, p \cdot \{t\}^l, b)$	$= / \cdot read(s, p \cdot \{t\}^l \cdot /, b)$
$\quad \text{where true} = isExprPrefix(p, b, l)$	
$read((\cdot s \cdot) \cdot s', p, b)$	$= (t) \cdot read(s', p \cdot (t), b)$
$\quad \text{where } s \text{ contains no unmatched })$	$\quad \text{where } t = read(s, \epsilon, false)$
$read(\{^l \cdot s \cdot \} \cdot s', p, b)$	$= \{t\}^l \cdot read(s', p \cdot \{t\}^l, b)$
$\quad \text{where } s \text{ contains no unmatched } \}$	$\quad \text{where } t = read(s, \epsilon, isExprPrefix(p, b, l))$
$read(x \cdot s, p, b)$	$= x \cdot read(s, p \cdot x, b)$
$read(\epsilon, p, b)$	$= \epsilon$

4. Enforestation

The core algorithm introduced by Honu is called *enforestation* which is basically responsible for expanding macros and building a partial syntax tree with enough structure to match on parse classes. Sweet.js implements this algorithm mostly as described with some additions to provide infix macros and invoke pattern classes described below.

4.1 Infix Macros

The macros we have described so far must all be prefixed by the macro identifier and syntax after the macro name is matched. This is sufficient for many kinds of macros but

some syntax forms require the macro identifier to sit between patterns.

Honu addresses this need in a limited way by providing a way to define new binary and unary operators which during expansion can manipulate their operators. However, those operators must be fully expanded and must match as an expression.

Sweet.js provides *infix macros* which allows a macro identifier to match syntax before it. For example, the following implements ES6-style arrow functions via infix macros:

```
macro (=>) {
  rule infix {
```

```

    ($params ...) | { $body ... }
  } => {
    function ($params ...) {
      $body ...
    }
  }
}

```

```
var id = (x) => { return x; }
```

The macros we have described so far must all be prefixed by the macro identifier and syntax after the macro name is matched. This is sufficient for many kinds of macros but some syntax forms require the macro identifier to sit between patterns.

Honu addresses this need in a limited way by providing a way to define new binary and unary operators which during expansion can manipulate their operators. However, those operators must be fully expanded and must match as an expression.

Sweet.js provides *infix macros* which allows a macro identifier to match syntax before it. For example, the following implements ES6-style arrow functions via infix macros:

```

macro (=>) {
  rule infix {
    ($params ...) | { $body ... }
  } => {
    function ($params ...) {
      $body ...
    }
  }
}
var id = (x) => { return x; }

```

This is accomplished by simply providing the state of previously expanded syntax to macro transformers. Macros may then consume from either ends as needed, yielding new previous and subsequent syntax. At first glance, this can appear brittle:

```
var foo = bar(x) => { return x; }
```

We've juxtaposed the => macro next to a function call, which we did not intend to be valid syntax. A naive expansion results in unparseable code:

```
var foo = bar function(x) { return x; }
```

In more subtle cases, a naive expansion can result in parsable code with incorrect semantics. [Example needed?] To preserve the integrity of previously expanded syntax, we verify that an infix macro only matches previous syntax on boundaries delimited by the partial syntax tree we've built. The syntax tree would show bar(x) as a complete function call term. Consuming the parentheses would result in a split term and is disallowed, failing the rule. In practice, this has proven to be an intuitive restriction. [Elaborate?]

The primary limitation of infix macros is they do not obey precedence and associativity. They stand as a complement to Honu operators, not as a replacement, for when an infix form does not adhere to operator semantics, much like the => macro. Its left-hand-side and right-hand-side are not arbitrary expressions but must match a specific form. Additionally, it would need a precedence of infinity so as to always bind tighter than other operators.

4.2 Invoke and Pattern Classes

[Something about default pattern classes and 'expr'...]

Pattern classes are extensible via the `invoke` class which is parameterized by a macro name.

```

macro color {
  rule { red } => { red }
  rule { green } => { green }
  rule { blue } => { blue }
}
macro colors_options {
  rule { ($opt:invoke(color) ...) } => { ... }
}

```

The macro is essentially inserted into the token stream. If the expansion succeeds, the result will be loaded into the pattern variable, otherwise the rule will fail. We've added sugar so that any non-primitive pattern classes are interpreted as `invoke` parameterized by the custom class. We've also added identity rules to shorten definitions of simple custom classes.

```

macro color {
  rule { red }
  rule { green }
  rule { blue }
}
macro colors_options {
  rule { ($opt:color (,) ...) } => { ... }
}

```

These macros may return an optional pattern environment which will be scoped and loaded into the invoking macro's pattern environment. This lets us define Honu-style pattern classes as simple macro-generating macros.

```

macro color {
  ...
}
macro number {
  ...
}
// 'pattern' is just a macro-generating macro
pattern color_value { $color:color $num:number }

macro color_options {
  rule { ($opt:color_value (,) ...) } => {
    var cols = [$opt$color (,) ...];
  }
}

```

```
    var nums = [$opt$num (,) ...];  
  }  
}
```

...?

5. Hygiene

Mostly straightforward implementation from scheme with some details to handle var.

6. Implementation

Sweet.js is written in JavaScript and runs in the major JS environments (i.e. the browser and node.js). This is in contrast to Honu which translates its code to Racket code and reuses the hygienic expansion machinery already built in Racket. While this simplifies the implementation of Honu it also requires an installation of Racket which in some cases is not feasible (e.g. sweet.js is able to run in mobile device browsers).

7. Related Work

- Scheme/Racket
- Honu
- Template Haskell
- Nemerle
- Scala
- Closure

8. Conclusion

A. Read Proof

Theorem 2 (Parse Equivalence for Program).

$$\forall s. \\ s \in \text{Program}_e \Leftrightarrow \text{read}(s, \epsilon, \text{false}) \in \text{Program}'_e$$

Proof. For the if direction, there are two cases:

- $s \in \text{Program}_e ::= \epsilon$. The result is immediate.
- $s \in \text{Program}_e ::= \text{SourceElements}_e$ this holds by Lemma 1.

A similar argument holds for the only if direction. \square

Lemma 1 (Parse Equivalence for SourceElements).

$$\forall s. \\ s \in \text{SourceElements}_e \Leftrightarrow \text{read}(s, \epsilon, \text{false}) \in \text{SourceElements}'_e$$

Proof. For the if direction, there are two cases:

- $s \in \text{SourceElements}_e ::= \text{SourceElement}_e$. This holds by Lemma 2.
- $s \in \text{SourceElements}_e \text{ by case:}$

$$\text{SourceElements}_e \text{ SourceElement}_{e'}$$

We have $s = s' \cdot s''$ where $s' \in \text{SourceElements}_e$ and $s'' \in \text{SourceElement}_{e'}$.

$$\begin{aligned} t &= \text{read}(s' \cdot s'', \epsilon, \text{false}) \\ &= \text{read}(s', \epsilon, \text{false}) \cdot \text{read}(s'', \text{read}(s', \epsilon, \text{false}), \text{false}) \end{aligned}$$

By induction $\text{read}(s', \epsilon, \text{false}) \in \text{SourceElements}'_e$ and by Lemma 2, $\text{read}(s'', \text{read}(s', \epsilon, \text{false}), \text{false}) \in \text{SourceElement}'_{e'}$ (since by Lemma 12, $\text{read}(s', \epsilon, \text{false}) \in \text{RegexPrefix}$). Thus $t \in \text{SourceElements}'_{e'}$.

The argument is similar for the only if direction. \square

Lemma 2 (Parse Equivalence for SourceElement).

$$\forall s, p \in \text{RegexPrefix}. \\ s \in \text{SourceElement}_e \Leftrightarrow \text{read}(s, p, \text{false}) \in \text{SourceElement}'_e$$

Proof. For the if direction there are two cases:

- $s \in \text{SourceElement}_e ::= \text{Stmt}_e$. This holds by Lemma 4 since $p \in \text{RegexPrefix}$.
- $s \in \text{SourceElement}_e ::= \text{FunctionDecl}_e$. This holds by Lemma 3.

\square

Lemma 3 (Parse Equivalence for FunctionDecl).

$$\forall s, p, b. \\ s \in \text{FunctionDecl}_e \Leftrightarrow \text{read}(s, p, b) \in \text{FunctionDecl}'_e$$

Proof. For the if direction we only have a single case:

$$s \in \text{FunctionDecl}_{\text{function } x (x') \{e\}} ::= \text{function } x (x') \{\text{SourceElements}_e\}$$

We have $s = \text{function } x (x) \{s'\}$ where $s' \in \text{SourceElements}_e$ so:

$$\begin{aligned} t &= \text{read}(\text{function } x (x) \{s'\}, p, b) \\ &= \text{function } x (x) \{t'\} \end{aligned}$$

where $t' = \text{read}(s', \epsilon, \text{false})$. Since by Lemma 1, $t' \in \text{SourceElements}'_e$ we have $t \in \text{FunctionDecl}'_{\text{function } x (x) \{e\}}$. \square

Lemma 4 (Parse Equivalence for Stmt).

$$\forall s, p \in \text{RegexPrefix}. \\ s \in \text{Stmt}_e \Leftrightarrow \text{read}(s, p, \text{false}) \in \text{Stmt}'_e$$

Proof. For the if direction we have several cases:

- $s \in \text{Stmt}_{\{e\}} ::= \{\text{StmtList}_e\}$. We have $s = \{s'\}$ where $s' \in \text{StmtList}_e$. Then:

$$\begin{aligned} t &= \text{read}(\{s'\}, p, \text{false}) \\ &= \{t'\} \end{aligned}$$

where $t' = \text{read}(s', \epsilon, \text{false})$. By Lemma 5, $t' \in \text{StmtList}'_e$ so $t \in \text{Stmt}'_{\{e\}}$.

- $s \in \text{Stmt}_e ::= \text{AssignExpr}_e ;$. We have $s = s' ;$ where $s' \in \text{AssignExpr}_e$. Then:

$$\begin{aligned} t &= \text{read}(s' ;, p, \text{false}) \\ &= \text{read}(s', p, \text{false}) ; \end{aligned}$$

Since $p \in \text{RegexPrefix}$ by Lemma 6, $\text{read}(s', p, \text{false}) \in \text{AssignExpr}'_e$ we have $t \in \text{Stmt}'_e$.

- $s \in \text{Stmt}_{\text{if } (e)} \text{ }_{e'} ::= \text{if } (\text{AssignExpr}_e) \text{ Stmt}_{e'}$. We have $s = \text{if} \cdot (s') \cdot s''$ where $s' \in \text{AssignExpr}_e$ and $s'' \in \text{Stmt}_{e'}$.

$$\begin{aligned} t &= \text{read}(\text{if} \cdot (s') \cdot s'', p, \text{false}) \\ &= \text{if} \cdot (t') \cdot t'' \end{aligned}$$

where

$$\begin{aligned} t' &= \text{read}(s', \epsilon, \text{false}) \\ t'' &= \text{read}(s'', p \cdot \text{if} \cdot (t'), \text{false}) \end{aligned}$$

By Lemma 6 $t' \in \text{AssignExpr}'_e$. By induction, $t'' \in \text{Stmt}'_{e'}$ (since $p \cdot (t') \in \text{RegexPrefix}$). Thus $t \in \text{Stmt}'_{\text{if } (e)} \text{ }_{e'}$.

- $s \in \text{Stmt}_{\text{return}} ::= \text{return}$. Since $s = \text{return}$ and $\text{read}(s, p, \text{false}) = \text{return} \in \text{Stmt}'_{\text{return}}$ we have our result directly.
- $s \in \text{Stmt}_{\text{return } e} ::= \text{return AssignExpr}_e ;$. We have $s = \text{return} \cdot s' ;$ where $s' \in \text{AssignExpr}_e$. Then:

$$\begin{aligned} t &= \text{read}(\text{return} \cdot s' ;, p, \text{false}) \\ &= \text{return} \cdot \text{read}(s', p \cdot \text{return}, \text{false}) ; \end{aligned}$$

By Lemma 6, $\text{read}(s', p \cdot \text{return}, \text{false}) \in \text{AssignExpr}'_e$ thus $t \in \text{Stmt}'_{\text{return } e}$.

- $s \in Stmt_{x:e} ::= x : Stmt_e$. We have $s = x : : s'$ where $s' \in Stmt_e$. Then:

$$\begin{aligned} t &= \text{read}(x : : s', p, \text{false}) \\ &= x : : \text{read}(s', p \cdot x : :, \text{false}) \end{aligned}$$

By induction, $\text{read}(s', p \cdot x : :, \text{false}) \in Stmt'_e$ so $t \in Stmt'_{x:e}$. \square

Lemma 5 (Parse Equivalence for StmtList).

$\forall s, p \in \text{RegexPrefix}$.
 $s \in StmtList_e \Leftrightarrow \text{read}(s, p, \text{false}) \in StmtList'_e$

Proof. For the if direction we have two cases:

- $s \in StmtList_e ::= Stmt_e$. This follows by Lemma 4 since $p \in \text{RegexPrefix}$.
- $s \in StmtList_{e'e'} ::= StmtList_e Stmt_{e'}$. So $s = s' \cdot s''$ where $s' \in StmtList_e$ and $s'' \in Stmt_{e'}$. Then,

$$\begin{aligned} t &= \text{read}(s' \cdot s'', p, \text{false}) \\ &= \text{read}(s', p, \text{false}) \cdot \text{read}(s'', p \cdot \text{read}(s', p, \text{false}), \text{false}) \end{aligned}$$

By induction $\text{read}(s', p, \text{false}) \in StmtList'_e$ and by Lemma 4, $\text{read}(s'', p \cdot \text{read}(s', p, \text{false}), \text{false}) \in Stmt'_{e'}$ (since by Lemma 13, $p \cdot \text{read}(s', p, \text{false}) \in \text{RegexPrefix}$). Thus, $t \in StmtList'_{e'e'}$. \square

Lemma 6 (Parse Equivalence for AssignExpr).

$\forall s, p \in \text{RegexPrefix}$. If $b = \text{false}$ then $s \neq \{ \cdot s' \}$. If $b = \text{true}$ then s is unconstrained.

$$s \in AssignExpr_e \Leftrightarrow \text{read}(s, p, b) \in AssignExpr'_e$$

Proof. The constraint on s when b is false is due to the lookahead check in the production $Stmt_e ::= AssignExpr_e ;$. Lemma 11 will make use of this constraint.

For the if direction we have two cases:

- $s \in AssignExpr_e ::= BinaryExpr_e$. This holds by Lemma 7.
- $s \in AssignExpr_{e=e'}$ by

$$CallExpr_e = AssignExpr_{e'}$$

Then $s = s' \cdot = \cdot s''$ where $s' \in CallExpr_e$ and $s'' \in AssignExpr_{e'}$. Then:

$$\begin{aligned} t &= \text{read}(s' \cdot = \cdot s'', p, b) \\ &= \text{read}(s', p, b) \cdot \text{read}(= \cdot s'', p \cdot \text{read}(s', p, b), b) \\ &= \text{read}(s', p, b) \cdot = \cdot \text{read}(s'', p \cdot \text{read}(s', p, b) \cdot =, b) \end{aligned}$$

Since $p \in \text{RegexPrefix}$ by Lemma 8, $\text{read}(s', p, b) \in CallExpr'_e$ and by induction $\text{read}(s'', p \cdot \text{read}(s', p, b) \cdot =, b) \in AssignExpr'_{e'}$ (since $p \cdot \text{read}(s', p, b) \cdot = \in \text{RegexPrefix}$) so $t \in AssignExpr'_{e=e'}$.

Lemma 7 (Parse Equivalence for BinaryExpr).

$\forall s, p \in \text{RegexPrefix}$. If $b = \text{false}$ then $s \neq \{ \cdot s' \}$. If $b = \text{true}$ then s is unconstrained.

$$s \in BinaryExpr_e \Leftrightarrow \text{read}(s, p, b) \in BinaryExpr'_e$$

Proof. To show the if direction we have three cases:

- $s \in BinaryExpr_e ::= CallExpr_e$. This holds by Lemma 8 since $p \in \text{RegexPrefix}$.
- $s \in BinaryExpr_{e/e'}$ by

$$BinaryExpr_e / BinaryExpr_{e'}$$

We have $s = s' \cdot / \cdot s''$ where $s' \in BinaryExpr_e$ and $s'' \in BinaryExpr_{e'}$. Then:

$$\begin{aligned} t &= \text{read}(s' \cdot / \cdot s'', p, b) \\ &= \text{read}(s', p, b) \cdot \text{read}(/ \cdot s'', p \cdot \text{read}(s', p, b), b) \\ &= \text{read}(s', p, b) \cdot / \cdot \text{read}(s'', p \cdot \text{read}(s', p, b) \cdot /, b) \\ &\quad (\text{since } p \cdot \text{read}(s', p, b) \in \text{DividePrefix}) \end{aligned}$$

By induction $\text{read}(s', p, b) \in BinaryExpr'_e$ and $\text{read}(s'', p \cdot \text{read}(s', p, b) \cdot /, b) \in BinaryExpr'_{e'}$ (since $p \cdot \text{read}(s', p, b) \cdot / \in \text{RegexPrefix}$) thus $t \in BinaryExpr'_{e/e'}$.

- $s \in BinaryExpr_{e+e'}$ by

$$BinaryExpr_e + BinaryExpr_{e'}$$

We have $s = s' \cdot + \cdot s''$ where $s' \in BinaryExpr_e$ and $s'' \in BinaryExpr_{e'}$. Then:

$$\begin{aligned} t &= \text{read}(s' \cdot + \cdot s'', p, b) \\ &= \text{read}(s', p, b) \cdot + \cdot \text{read}(s'', p \cdot \text{read}(s', p, b) \cdot +, b) \end{aligned}$$

By induction $\text{read}(s', p, b) \in BinaryExpr'_e$ and $\text{read}(s'', p \cdot \text{read}(s', p, b) \cdot +, b) \in BinaryExpr'_{e'}$ (since $p \cdot \text{read}(s', p, b) \cdot + \in \text{RegexPrefix}$) thus $t \in BinaryExpr'_{e+e'}$. \square

Lemma 8 (Parse Equivalence for CallExpr).

$\forall s, p \in \text{RegexPrefix}$. If $b = \text{false}$ then $s \neq \{ \cdot s' \}$. If $b = \text{true}$ then s is unconstrained.

$$s \in CallExpr_e \Leftrightarrow \text{read}(s, p, b) \in CallExpr'_e$$

Proof. For the if direction we have three cases:

- $s \in CallExpr_{e(e')}$ by

$$MemberExpr_e (AssignExpr'_{e'})$$

We have $s = s' \cdot (s'')$ where $s' \in MemberExpr_e$ and $s'' \in AssignExpr_{e'}$. Then

$$\begin{aligned} t &= \text{read}(s' \cdot (s''), p, b) \\ &= \text{read}(s', p, b) \cdot \text{read}((s''), p \cdot \text{read}(s', p, b), b) \end{aligned}$$

Since $p \in \text{RegexPrefix}$, by Lemma 9 we have $\text{read}(s', p, b) \in \text{MemberExpr}'_e$ and by Lemma 6 we have $\text{read}(s'', \epsilon, \text{false}) \in \text{AssignExpr}'_{e'}$, thus $t \in \text{CallExpr}'_{e \cdot (e')}$.

- $s \in \text{CallExpr}_{e \cdot x} ::= \text{CallExpr}_e \cdot x$. Then $s = s' \cdot \dots \cdot x$ where $s' \in \text{CallExpr}_e$. Then

$$\begin{aligned} t &= \text{read}(s' \cdot \dots \cdot x, p, b) \\ &= \text{read}(s', p, b) \cdot \dots \cdot x \end{aligned}$$

By induction $\text{read}(s', p, b) \in \text{CallExpr}'_e$. Thus $t \in \text{CallExpr}'_{e \cdot x}$. \square

Lemma 9 (Parse Equivalence for MemberExpr).

$\forall s, p \in \text{RegexPrefix}$. If $b = \text{false}$ then $s \neq \{ \cdot s' \}$. If $b = \text{true}$ then s is unconstrained.

$$s \in \text{MemberExpr}_e \Leftrightarrow \text{read}(s, p, b) \in \text{MemberExpr}'_e$$

Proof. For the if direction we have three cases:

- $s \in \text{MemberExpr}_e ::= \text{PrimaryExpr}_e$. This follows from Lemma 11 since $p \in \text{RegexPrefix}$.
- $s \in \text{MemberExpr}_e ::= \text{FunctionExpr}_e$. This follows from Lemma 10.
- $s \in \text{MemberExpr}_{e \cdot x} ::= \text{MemberExpr}_e \cdot x$. We have $s = s' \cdot \dots \cdot x$ where $s \in \text{MemberExpr}_e$. Then

$$\begin{aligned} t &= \text{read}(s' \cdot \dots \cdot x, p, b) \\ &= \text{read}(s', p, b) \cdot \dots \cdot x \end{aligned}$$

By induction $\text{read}(s', p, b) \in \text{MemberExpr}'_e$ thus $t \in \text{MemberExpr}'_{e \cdot x}$. \square

Lemma 10 (Parse Equivalence for FunctionExpr).

$\forall s, b, p \in \text{RegexPrefix}$.

$$s \in \text{FunctionExpr}_e \Leftrightarrow \text{read}(s, p, b) \in \text{FunctionExpr}'_e$$

Proof. Since $s = \text{function } x \ (x') \ \{s'\}$ where $s' \in \text{SourceElements}_e$ and

$$t = \text{read}(s, p, b) = \text{function } \cdot x \cdot (x') \cdot \{t'\}$$

where $t' = \text{read}(s', \epsilon, \text{false})$ and by Lemma 1, $t' \in \text{SourceElements}'_e$ we have $t \in \text{FunctionExpr}'_{\text{function } x \ (x) \ \{e\}}$. \square

Lemma 11 (Parse Equivalence for PrimaryExpr).

$\forall s, p \in \text{RegexPrefix}$. If $b = \text{false}$ then $s \neq \{ \cdot s' \}$. If $b = \text{true}$ then s is unconstrained.

$$s \in \text{PrimaryExpr}_e \Leftrightarrow \text{read}(s, p, b) \in \text{PrimaryExpr}'_e$$

Proof. For the if direction we have several cases:

- $s \in \text{PrimaryExpr}_x ::= x$. Then $s = x$ and $\text{read}(x, p, b) \in \text{PrimaryExpr}'_x$ directly.

- $s \in \text{PrimaryExpr}_{/x/} ::= / \cdot x \cdot /$. Then $s = / \cdot x \cdot /$ and $\text{read}(/ \cdot x \cdot /, p, b) = /x/ \in \text{PrimaryExpr}'_{/x/}$ since $p \in \text{RegexPrefix}$.

- $s \in \text{PrimaryExpr}_{\{x:e\}} ::= \{x:\text{AssignExpr}_e\}$. Then $s = \{ \cdot x \cdot : \cdot s' \cdot \}$ where $s' \in \text{AssignExpr}_e$. Then:

$$\begin{aligned} t &= \text{read}(\{ \cdot x \cdot : \cdot s' \cdot \}, p, \text{true}) \\ &= \{x \cdot : \cdot t'\} \end{aligned}$$

where $t' = \text{read}(s', x \cdot :, \text{true})$. By Lemma 6, $t' \in \text{AssignExpr}'_e$ and thus $t \in \text{PrimaryExpr}'_{\{x:e\}}$.

- $s \in \text{PrimaryExpr}_{(e)} ::= (\text{AssignExpr}_e)$. Then $s = (s')$ where $s' \in \text{AssignExpr}_e$. So,

$$\begin{aligned} t &= \text{read}((s'), p, b) \\ &= (t') \end{aligned}$$

where $t' = \text{read}(s', \epsilon, \text{false})$. By Lemma 6 $t' \in \text{AssignExpr}'_e$. So, $t \in \text{PrimaryExpr}'_{(e)}$. \square

Lemma 12 (SourceElement Prefix).

$\forall s, p \in \text{RegexPrefix}$.

$$s \in \text{SourceElement}_e \Rightarrow \text{read}(s, p, \text{false}) \in \text{RegexPrefix}$$

Proof. There are two cases:

- $s \in \text{Stmt}_e$. This holds by Lemma 13.
- $s \in \text{FunctionDecl}_e$. Follows directly. \square

Lemma 13 (Stmt Prefix).

$\forall s, p \in \text{RegexPrefix}$.

$$\text{read}(s, p, \text{false}) \in \text{Stmt}'_e \Rightarrow p \cdot \text{read}(s, p, \text{false}) \in \text{RegexPrefix}$$

Proof. We have several cases:

- $\text{read}(s, p, \text{false}) \in \{\text{StmtList}'_e\}$. Follows since $p \in \text{RegexPrefix}$.
- $\text{read}(s, p, \text{false}) \in \text{AssignExpr}'_e$; . Follows since $t \cdot ; \in \text{RegexPrefix}$ for any t .
- $t = \text{read}(s, p, \text{false}) \in \text{if } (\text{AssignExpr}'_e) \text{ Stmt}'_e$. Since $t = \text{if } \cdot (t') \cdot t''$ where $t'' \in \text{Stmt}'_e$ by induction $t'' \in \text{RegexPrefix}$ and thus $p \cdot t \in \text{RegexPrefix}$.
- $\text{read}(s, p, \text{false}) \in \text{return}$. Follows since $p \cdot \text{return} \in \text{RegexPrefix}$.
- $t = \text{read}(s, p, \text{false}) \in \text{return AssignExpr}'_e$; . Since $; \in \text{RegexPrefix}$ then $p \cdot t \in \text{RegexPrefix}$.
- $t = \text{read}(s, p, \text{false}) \in x : \text{Stmt}'_e$. Since $t = x \cdot : \cdot t'$ where $t' \in \text{Stmt}'_e$. Since $p \in \text{RegexPrefix}$ and by induction $t' \in \text{RegexPrefix}$ we have $p \cdot t \in \text{RegexPrefix}$. \square