Type-checking Linearity in Core: Semantic Linearity for a Lazy Optimising Compiler

Rodrigo Mesquita
Advisor: Bernardo Toninho

NOVA SCHOOL OF SCIENCE & TECHNOLOGY

Linear Haskell

Haskell has Linear Types!

Linear Haskell

Haskell has Linear Types!
A linear function $\multimap$ consumes its argument *exactly once*

Linear Haskell

Haskell has Linear Types!

A linear function $\multimap$ consumes its argument *exactly once*

$$
\begin{aligned}
&\mathit{bad} :: \mathit{Ptr} \multimap \mathit{IO}\ ()\\
&\mathit{bad}\ x = \mathbf{do}\\
&\quad \mathit{free}\ x\\
&\quad \mathit{free}\ x
\end{aligned}
$$

Linear Haskell

Haskell has Linear Types!

A linear function $\multimap$ consumes its argument *exactly once*

$$
\begin{aligned}
&\mathit{bad} :: \mathit{Ptr} \multimap \mathit{IO}\ ()\\
&\mathit{bad}\ x = \mathbf{do}\\
&\quad \mathit{free}\ x\\
&\quad \mathit{free}\ x
\end{aligned}
$$

$$
\begin{aligned}
&\mathit{ok} :: \mathit{Ptr} \multimap \mathit{IO}\ ()\\
&\mathit{ok}\ x = \mathit{free}\ x
\end{aligned}
$$

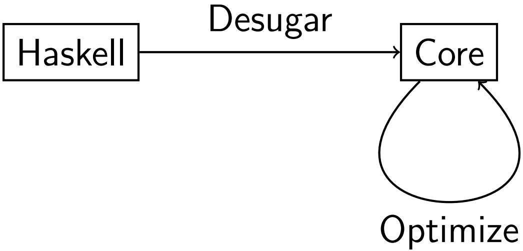
Linearity in the Glasgow Haskell Compiler (GHC)

$\boxed{\text{Haskell}}$

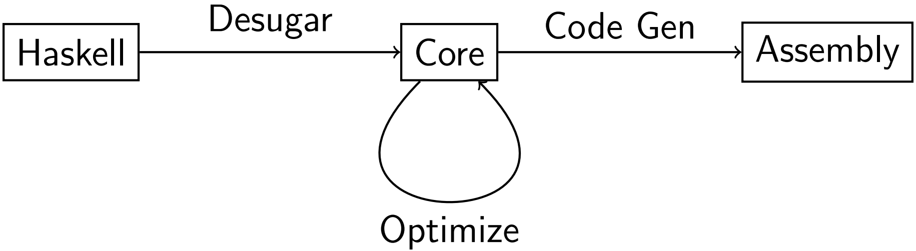
Linearity in the Glasgow Haskell Compiler (GHC)

$$\boxed{\text{Haskell}} \xrightarrow{\text{Desugar}} \boxed{\text{Core}}$$

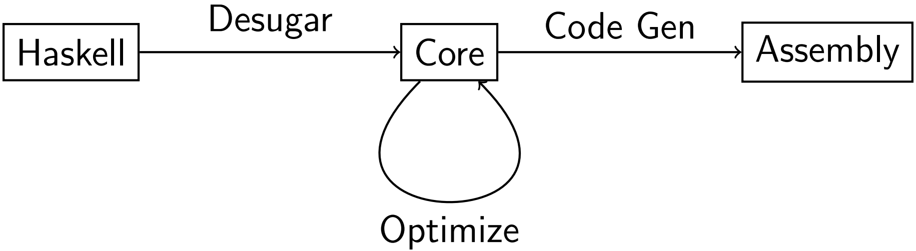
Linearity in the Glasgow Haskell Compiler (GHC)



Linearity in the Glasgow Haskell Compiler (GHC)

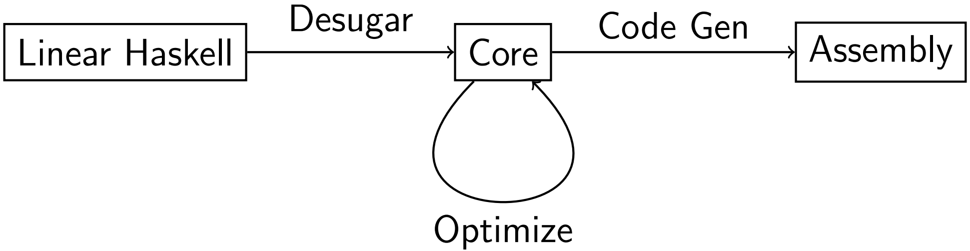


Linearity in the Glasgow Haskell Compiler (GHC)



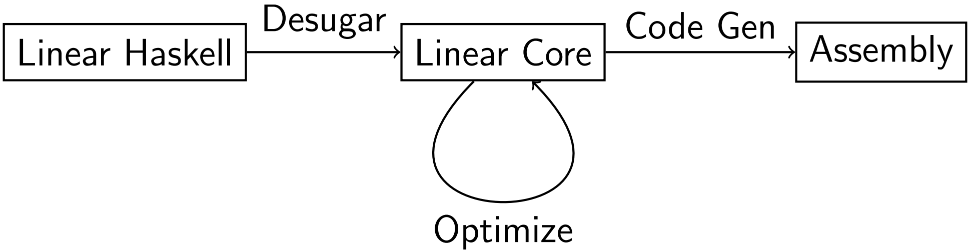
Core is both lazy and typed

Linearity in the Glasgow Haskell Compiler (GHC)



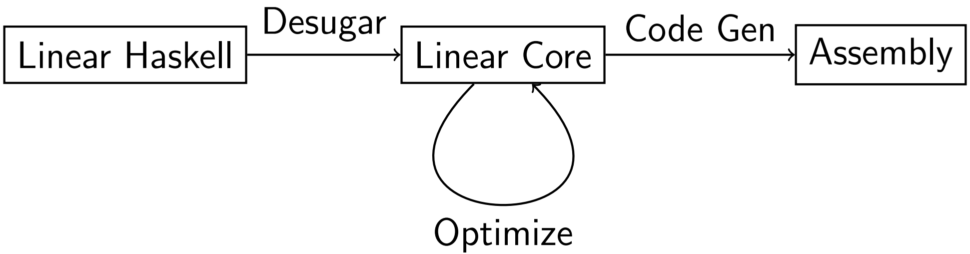
Core is both lazy and typed

Linearity in the Glasgow Haskell Compiler (GHC)



Core is both lazy and typed

Linearity in the Glasgow Haskell Compiler (GHC)



Core *should be* both lazy and *linearly* typed

So, why isn't Core linear?

Optimised programs stop *looking* linear

So, why isn't Core linear?

Optimised programs stop *looking* linear

$$\textbf{let } y = \mathit{free}\ x \textbf{ in } y$$
$$\Longrightarrow$$
$$\textbf{let } y = \mathit{free}\ x \textbf{ in } \mathit{free}\ x$$

So, why isn't Core linear?

Optimised programs stop *looking* linear

$$\textbf{let}\ y = \mathit{free}\ x\ \textbf{in}\ y$$
$$\Longrightarrow$$
$$\textbf{let}\ y = \mathit{free}\ x\ \textbf{in}\ \mathit{free}\ x$$

Linearity is ignored in Core, or most programs would be rejected

Semantic vs Syntactic Linearity

- Programs are still linear *semantically* because of laziness

Semantic vs Syntactic Linearity

- Programs are still linear *semantically* because of laziness
- **Key insight:** Under lazy evaluation,

$$\textit{syntactic}\ \text{occurrence} \not\Rightarrow \textit{consuming}\ \text{a resource}$$
$$\textit{syntactic}\ \text{linearity} \neq \textit{semantic}\ \text{linearity}$$

Semantic vs Syntactic Linearity

- Programs are still linear *semantically* because of laziness
- **Key insight:** Under lazy evaluation,

$$\textit{syntactic}\ \text{occurrence} \not\Rightarrow \textit{consuming}\ \text{a resource}$$
$$\textit{syntactic}\ \text{linearity} \neq \textit{semantic}\ \text{linearity}$$

- We type *syntactic* linearity in Core, but that is not enough

Semantic vs Syntactic Linearity

- Programs are still linear *semantically* because of laziness
- **Key insight:** Under lazy evaluation,

$$\textit{syntactic}\ \text{occurrence} \nRightarrow \textit{consuming}\ \text{a resource}$$
$$\textit{syntactic}\ \text{linearity} \neq \textit{semantic}\ \text{linearity}$$

- We type *syntactic* linearity in Core, but that is not enough
- Optimisations push laziness x linearity to the limit

Our Contributions

- Linear Core: a type system that understands semantic linearity in the presence of laziness

Our Contributions

- Linear Core: a type system that understands semantic linearity in the presence of laziness
- We proved Linear Core and multiple optimising transformations to be sound

Our Contributions

- Linear Core: a type system that understands semantic linearity in the presence of laziness
- We proved Linear Core and multiple optimising transformations to be sound
- We implemented Linear Core as a GHC plugin

Semantic Linearity, by example

Semantic Linearity: Lets

let $y = free\ ptr$
in if *condition*
  then y
  else *return ptr*

Semantic Linearity: Lets

let $y = free\ ptr$
in if *condition*
  then y
  else *return ptr*

Resources in lets are only consumed if the binder is evaluated

Semantic Linearity: Case

$$\textbf{case } (x, y) \textbf{ of}$$
$$\quad (a, b) \rightarrow \mathit{something}\ a\ b$$

Semantic Linearity: Case

$$\begin{array}{l}\textbf{case } (x, y) \textbf{ of}\\ \quad (a, b) \rightarrow \textit{something } a\ b\end{array}$$

$$\begin{array}{l}\textbf{case } (x, y) \textbf{ of}\\ \quad (a, b) \rightarrow \textit{something } x\ y\end{array}$$

Semantic Linearity: Case

$$\begin{array}{l}\textbf{case}\ (x, y)\ \textbf{of} \\ \quad (a, b) \rightarrow \mathit{something}\ a\ b\end{array}$$

$$\begin{array}{l}\textbf{case}\ (x, y)\ \textbf{of} \\ \quad (a, b) \rightarrow \mathit{something}\ x\ y\end{array}$$

$$\begin{array}{l}\textbf{case}\ \mathit{free}\ x\ \textbf{of} \\ \quad \mathit{Result}\ v \rightarrow \mathit{free}\ x\end{array}$$

Semantic Linearity: Case

$$\textbf{case } (x, y) \textbf{ of} \\ \quad (a, b) \rightarrow \textit{something } a\ b$$

$$\textbf{case } (x, y) \textbf{ of} \\ \quad (a, b) \rightarrow \textit{something } x\ y$$

$$\textbf{case } \textit{free } x \textbf{ of} \\ \quad \textit{Result } v \rightarrow \textit{free } x$$

$$\textbf{case } \textit{use } x \textbf{ of} \\ \quad \textit{Result } v \rightarrow ()$$

Semantic Linearity: Case

$$\textbf{case } (x, y) \textbf{ of}\\ \quad (a, b) \rightarrow \textit{something } a\ b$$

$$\textbf{case } (x, y) \textbf{ of}\\ \quad (a, b) \rightarrow \textit{something } x\ y$$

$$\textbf{case } \textit{free } x \textbf{ of}\\ \quad \textit{Result } v \rightarrow \textit{free } x$$

$$\textbf{case } \textit{use } x \textbf{ of}\\ \quad \textit{Result } v \rightarrow ()$$

Resources are *kind of* consumed if the expression is evaluated

Linear Core

Linear Core: Lets

$$
\begin{array}{l}
\textbf{let } y_{\{ptr\}} = \mathit{free}\ \mathit{ptr} \\
\textbf{in if } \mathit{condition} \\
\quad \textbf{then } y_{\{ptr\}} \\
\quad \textbf{else } \mathit{return}\ \mathit{ptr}
\end{array}
$$

Linear Core: Lets

$$\begin{array}{l}\textbf{let } y_{\{ptr\}} = \mathit{free}\ \mathit{ptr} \\ \textbf{in if } \mathit{condition} \\ \quad \textbf{then } y_{\{ptr\}} \\ \quad \textbf{else } \mathit{return}\ \mathit{ptr}\end{array}$$

Let-binders don't consume resources

- Annotate *let*-vars with linear resources (Δ) used in its body
- Using a *let*-var equates to using its annotated context (Δ)

Linear Core: Lets

$$
\begin{array}{l}
\textbf{let } y_{\{ptr\}} = \mathit{free\ ptr} \\
\textbf{in if } \mathit{condition} \\
\quad \textbf{then } y_{\{ptr\}} \\
\quad \textbf{else } \mathit{return\ ptr}
\end{array}
$$

$$\frac{}{\cdot; \mathit{ptr} \vdash \textbf{let } y = \mathit{free\ ptr} \textbf{ in } \dots}$$

Let-binders don't consume resources

- Annotate *let*-vars with linear resources (Δ) used in its body
- Using a *let*-var equates to using its annotated context (Δ)

Linear Core: Lets

$$\begin{array}{l}\textbf{let}\ y_{\{ptr\}} = \mathit{free}\ \mathit{ptr}\\ \textbf{in if}\ \mathit{condition}\\ \quad \textbf{then}\ y_{\{ptr\}}\\ \quad \textbf{else}\ \mathit{return}\ \mathit{ptr}\end{array}$$

$$\frac{\cdot; \mathit{ptr} \vdash \mathit{free}\ \mathit{ptr}}{\cdot; \mathit{ptr} \vdash \textbf{let}\ y = \mathit{free}\ \mathit{ptr}\ \textbf{in}\ \ldots}$$

Let-binders don't consume resources

- Annotate *let*-vars with linear resources (Δ) used in its body
- Using a *let*-var equates to using its annotated context (Δ)

Linear Core: Lets

$$
\begin{aligned}
&\textbf{let}\ y_{\{ptr\}} = \mathit{free}\ \mathit{ptr} \\
&\textbf{in if}\ \mathit{condition} \\
&\quad \textbf{then}\ y_{\{ptr\}} \\
&\quad \textbf{else}\ \mathit{return}\ \mathit{ptr}
\end{aligned}
$$

$$
\frac{\cdot; \mathit{ptr} \vdash \mathit{free}\ \mathit{ptr} \qquad y{:}_{\{ptr\}}; \mathit{ptr} \vdash \textbf{if}\ \mathit{condition} \ldots}{\cdot; \mathit{ptr} \vdash \textbf{let}\ y = \mathit{free}\ \mathit{ptr}\ \textbf{in}\ \ldots}
$$

Let-binders don't consume resources

- Annotate *let*-vars with linear resources (Δ) used in its body
- Using a *let*-var equates to using its annotated context (Δ)

Linear Core: Case

Case scrut evaluate to WHNF, unless they are already in WHNF

Linear Core: Case

Case scrut evaluate to WHNF, unless they are already in WHNF

$$\begin{array}{l}\textbf{case}\ (x, y)\ \textbf{of} \\ \quad (a, b) \to \mathit{something}\ x\ y\end{array}$$

Linear Core: Case

Case scrut evaluate to WHNF, unless they are already in WHNF

$$\begin{array}{l} \textbf{case } (x, y) \textbf{ of} \\ \quad (a, b) \rightarrow \mathit{something}\ x\ y \end{array}$$

$$\begin{array}{l} \textbf{case } \mathit{free}\ x \textbf{ of} \\ \quad \mathit{Result}\ v \rightarrow \mathit{free}\ x \end{array}$$

Linear Core: Case

Case scrut evaluate to WHNF, unless they are already in WHNF

$$\textbf{case}\ (x, y)\ \textbf{of}\\ \quad (a, b) \to \mathit{something}\ x\ y$$

$$\textbf{case}\ \mathit{free}\ x\ \textbf{of}\\ \quad \mathit{Result}\ v \to \mathit{free}\ x$$

Key idea: We need to branch on *WHNF-ness*

Linear Core: Case WHNF

$$\begin{array}{l}\textbf{case } (x, y) \textbf{ of} \\ \quad (a_{\{x\}}, b_{\{y\}}) \to \mathit{use}\ x\ y\end{array}$$

Linear Core: Case WHNF

$$\textbf{case } (x, y) \textbf{ of} \\ \quad (a_{\{x\}}, b_{\{y\}}) \rightarrow \mathit{use}\ x\ y$$

- Scrutinee resources are available in the body
- Pattern variables are annotated with corresponding scrutinee variables

Linear Core: Case WHNF

$$\begin{array}{l}\textbf{case } (x, y) \textbf{ of} \\ \quad (a_{\{x\}}, b_{\{y\}}) \rightarrow \mathit{use}\ x\ y\end{array}$$

$$\frac{}{\cdot; x, y \vdash \textbf{case } (x, y) \textbf{ of } (a, b) \rightarrow \ldots}$$

- Scrutinee resources are available in the body
- Pattern variables are annotated with corresponding scrutinee variables

Linear Core: Case WHNF

$$\begin{array}{l}\textbf{case}\ (x, y)\ \textbf{of} \\ \quad (a_{\{x\}}, b_{\{y\}}) \to \mathit{use}\ x\ y\end{array}$$

$$\frac{\cdot; x, y \vdash (x, y)}{\cdot; x, y \vdash \textbf{case}\ (x, y)\ \textbf{of}\ (a, b) \to \ldots}$$

- Scrutinee resources are available in the body
- Pattern variables are annotated with corresponding scrutinee variables

Linear Core: Case WHNF

$$\begin{array}{l}\textbf{case } (x, y) \textbf{ of} \\ \quad (a_{\{x\}}, b_{\{y\}}) \to \textit{use } x\ y\end{array}$$

$$\frac{\cdot; x, y \vdash (x, y) \qquad a{:}_{\{x\}}, b{:}_{\{y\}}; x, y \vdash \textit{use } x\ y}{\cdot; x, y \vdash \textbf{case } (x, y) \textbf{ of } (a, b) \to \ldots}$$

- Scrutinee resources are available in the body
- Pattern variables are annotated with corresponding scrutinee variables

Linear Core: Case Not-WHNF

$$\begin{array}{l}\textbf{case}\ \mathit{free}\ x\ \textbf{of}\\ \quad \mathit{Result}\ v_{\{[x]\}} \rightarrow \mathit{free}\ x\end{array}$$

Linear Core: Case Not-WHNF

$$\begin{array}{l}\textbf{case}\ \mathit{free}\ x\ \textbf{of}\\ \quad \mathit{Result}\ v_{\{[x]\}} \rightarrow \mathit{free}\ x\end{array}$$

Scrutinee resources are *irrelevant* in the body

- They cannot be instantiated with *Var*
- But must still be used exactly once

Linear Core: Case Not-WHNF

$$\begin{array}{l}\textbf{case}\ \mathit{free}\ x\ \textbf{of}\\ \quad \mathit{Result}\ v_{\{[x]\}} \rightarrow \mathit{free}\ x\end{array}$$

$$\frac{}{\cdot; x \vdash \textbf{case}\ \mathit{free}\ x\ \textbf{of}\ \ \dots}$$

Scrutinee resources are *irrelevant* in the body

- They cannot be instantiated with *Var*
- But must still be used exactly once

Linear Core: Case Not-WHNF

$$\begin{array}{l}\textbf{case}\ \mathit{free}\ x\ \textbf{of} \\ \quad \mathit{Result}\ v_{\{[x]\}} \to \mathit{free}\ x\end{array}$$

$$\frac{\cdot; x \vdash \mathit{free}\ x}{\cdot; x \vdash \textbf{case}\ \mathit{free}\ x\ \textbf{of}\ \ldots}$$

Scrutinee resources are *irrelevant* in the body

- They cannot be instantiated with *Var*
- But must still be used exactly once

Linear Core: Case Not-WHNF

$$\begin{array}{l}\textbf{case } \mathit{free}\ x \textbf{ of} \\ \quad \mathit{Result}\ v_{\{[x]\}} \to \mathit{free}\ x\end{array}$$

$$\frac{\cdot; x \vdash \mathit{free}\ x \qquad v{:}_{\{[x]\}}; [x] \vdash \mathit{free}\ x}{\cdot; x \vdash \textbf{case } \mathit{free}\ x \textbf{ of } \ldots}$$

Scrutinee resources are *irrelevant* in the body

- They cannot be instantiated with *Var*
- But must still be used exactly once

Metatheory: Linear Core

- Not obvious whether these rules make sense together
- We proved the system is type safe via preservation + progress

Metatheory: Linear Core

- Not obvious whether these rules make sense together
- We proved the system is type safe via preservation + progress
    - *Irrelevance* lemma
    - Linear-var substitution lemma
    - Δ-var substitution lemma
    - Unr-var substitution lemma

Metatheory: Optimising Transformations

- Inlining
- β-reduction
- β-reduction with sharing
- β-reduction for multiplicity abstractions
- Case-of-known-constructor
- Full laziness
- Local transformations (three of them)
- η-expansion
- η-reduction
- Binder swap
- Reverse binder swap (contentious!)
- Case-of-case

GHC Plugin: Linear Core Implementation

We implemented Linear Core as a GHC plugin

Library	Total Accepted	Total Rejected	Unique Rejected	Linear modulo Call-by-name	Linear Rejected	¬ Linear Rejected	Unknown Rejected
linear-smc	19438	4	1	1	0	0	0
priority-sesh	6781	19	1	0	0	0	1
linear-base	112311	538	87	10	8	2	67

Figure: Linear Core Plugin on Linear Libraries

Conclusion

- Linear Core is a suitable type system for Core, as it understands the interaction between linearity and laziness that the optimiser pushes to the limit

Conclusion

- Linear Core is a suitable type system for Core, as it understands the interaction between linearity and laziness that the optimiser pushes to the limit
- Not every single program is accepted by Linear Core
  - Future work: *multiplicity coercions*
  - Discuss linearity modulo call-by-name
  - Iron out quirks (rewrite rules, ...)

Conclusion

- Linear Core is a suitable type system for Core, as it understands the interaction between linearity and laziness that the optimiser pushes to the limit
- Not every single program is accepted by Linear Core
  - Future work: *multiplicity coercions*
  - Discuss linearity modulo call-by-name
  - Iron out quirks (rewrite rules, ...)
- Builds on the shoulders of Linear Haskell and Linear Mini-Core

Conclusion

- Linear Core is a suitable type system for Core, as it understands the interaction between linearity and laziness that the optimiser pushes to the limit
- Not every single program is accepted by Linear Core
  - Future work: *multiplicity coercions*
  - Discuss linearity modulo call-by-name
  - Iron out quirks (rewrite rules, ...)
- Builds on the shoulders of Linear Haskell and Linear Mini-Core
- There's much more in the thesis!

Fim

Semantic Linearity: Case of Var

$$(\lambda x.\ \textbf{case}\ x\ \textbf{of}\ \_ \to x)$$

Semantic Linearity: Case of Var

$$(\lambda x.\ \textbf{case}\ x\ \textbf{of}\ \_ \to x)$$
$$\Longrightarrow_{\text{call by name}}$$

Semantic Linearity: Case of Var

$$(\lambda x.\ \textbf{case}\ x\ \textbf{of}\ \_ \to x)$$

$$\Longrightarrow_{\text{call by name}}$$

$$\textbf{case}\ \mathit{free}\ x\ \textbf{of}\ \_ \to \mathit{free}\ x$$

Semantic Linearity: Case of Var

$$(\lambda x.\ \textbf{case}\ x\ \textbf{of}\ \_ \to x)$$

$$\Longrightarrow_{\text{call by name}}$$

$$\textbf{case}\ \mathit{free}\ x\ \textbf{of}\ \_ \to \mathit{free}\ x$$

$$\Longrightarrow_{\text{call by need}}$$

Semantic Linearity: Case of Var

$$(\lambda x.\ \textbf{case}\ x\ \textbf{of}\ \_ \rightarrow x)$$

$$\Longrightarrow_{\text{call by name}}$$

$$\textbf{case}\ \mathit{free}\ x\ \textbf{of}\ \_ \rightarrow \mathit{free}\ x$$

$$\Longrightarrow_{\text{call by need}}$$

$$\textbf{let}\ y = \mathit{free}\ x\ \textbf{in}\ \textbf{case}\ y\ \textbf{of}\ \_ \rightarrow y$$

System FC

- *System* F_C is a polymorphic lambda calculus with explicit type-equality coercions

System FC

- *System F_C* is a polymorphic lambda calculus with explicit type-equality coercions
- A coercion $\sigma_1 \sim \sigma_2$ can be used to safely *cast* an expression e of type σ_1 to type σ_2, written $e \blacktriangleright \sigma_1 \sim \sigma_2$.

System FC

Definition (Syntax)

$$
\begin{array}{lcll}
u & ::= & x \mid K & \text{Variables and data constructors} \\
e & ::= & u & \text{Term atoms} \\
 & \mid & \Lambda a{:}\kappa.\ e \ \mid \ e\ \varphi & \text{Type abstraction/application} \\
 & \mid & \lambda x{:}\sigma.\ e \ \mid \ e_1\ e_2 & \text{Term abstraction/application} \\
 & \mid & \textbf{let}\ x{:}\sigma = e_1\ \textbf{in}\ e_2 & \\
 & \mid & \textbf{case}\ e_1\ \textbf{of}\ \overline{p \to e_2} & \\
 & \mid & e \blacktriangleright \gamma & \text{Cast} \\
 & & & \\
p & ::= & K\ \overline{b{:}\kappa}\ \overline{x{:}\sigma} & \text{Pattern}
\end{array}
$$