An Experimental Ambiguity Detection Tool

Sylvain Schmitz

Laboratoire I3S, Université de Nice - Sophia Antipolis & CNRS

March 25, 2007

Standard ML

Milner et al. [1997]

Example

```
datatype 'a option = NONE | SOME of 'a

fun filter pred l =
  let
    fun filterP (x::r, l) =
          case (pred x)
            of SOME y => filterP(r, y::l)
             | NONE   => filterP(r, l)
     |  filterP ([], l) = rev l
  in
    filterP (l, [])
  end
```

The Issue

SML Compilers Behaviour

```
| filterP ([], l) = rev l
```

- MLton
- Moscow ML
- Poly/ML
- SML/NJ

```
Error: match.sml 9.25.
  Syntax error: replacing  EQUALOP with  DARROW.
```

The Issue

SML Compilers Behaviour

```
| filterP ([], l) = rev l
```

- MLton
- Moscow ML
- Poly/ML
- SML/NJ

```
! Toplevel input:
!       |  filterP ([], l) = rev l
!                           ^
! Syntax error.
```

The Issue

SML Compilers Behaviour

```
| filterP ([], l) = rev l
```

- MLton
- Moscow ML
- Poly/ML
- SML/NJ

```
Error: => expected but = was found
```

The Issue

SML Compilers Behaviour

```
| filterP ([], l) = rev l
```

- MLton
- Moscow ML
- Poly/ML
- SML/NJ

```
stdIn:7.24-7.29 Error: syntax error:
  deleting  EQUALOP ID
```

The Issue

Partial SML Grammar

$$\begin{array}{rcl}
\langle \mathit{dec} \rangle & \rightarrow & \textbf{fun}\ \langle \mathit{fvalbind} \rangle \\
\langle \mathit{fvalbind} \rangle & \rightarrow & \langle \mathit{sfvb} \rangle \mid \langle \mathit{fvalbind} \rangle\ '|'\ \langle \mathit{sfvb} \rangle \\
\langle \mathit{sfvb} \rangle & \rightarrow & \mathit{vid}\ \langle \mathit{atpats} \rangle\ =\ \langle \mathit{exp} \rangle \\
\\
\langle \mathit{exp} \rangle & \rightarrow & \textbf{case}\ \langle \mathit{exp} \rangle\ \textbf{of}\ \langle \mathit{match} \rangle \\
\langle \mathit{match} \rangle & \rightarrow & \langle \mathit{mrule} \rangle \mid \langle \mathit{match} \rangle\ '|'\ \langle \mathit{mrule} \rangle \\
\langle \mathit{mrule} \rangle & \rightarrow & \langle \mathit{pat} \rangle => \langle \mathit{exp} \rangle \\
\\
\langle \mathit{atpats} \rangle & \rightarrow & \langle \mathit{atpat} \rangle \mid \langle \mathit{atpats} \rangle\ \langle \mathit{atpat} \rangle \\
\langle \mathit{pat} \rangle & \rightarrow & \mathit{vid}\ \langle \mathit{atpat} \rangle \\
\langle \mathit{atpat} \rangle & \rightarrow & \mathit{vid}
\end{array}$$

The Issue

Shift/Reduce Conflict

GNU Bison

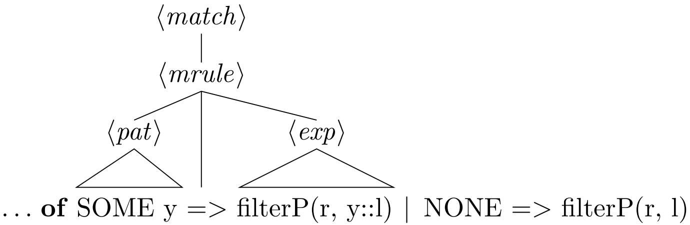
```
state 20
    6 exp: "case" exp "of" match .
    8 match: match . '|' mrule

    '|'  shift, and go to state 24
    '|'       [reduce using rule 6 (exp)]
```

The Issue

Shift/Reduce Conflict

Which action to choose?

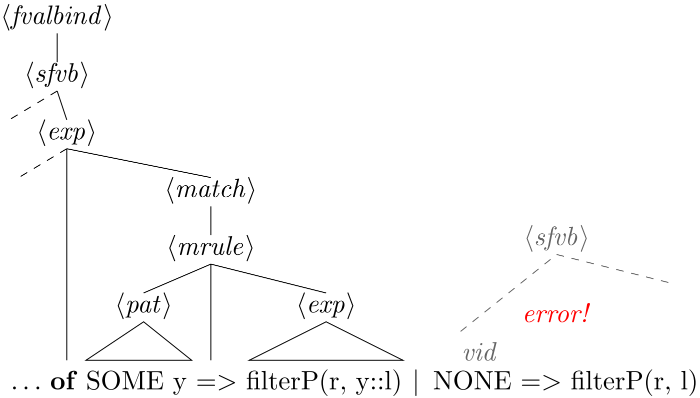


. . . **of** SOME y => filterP(r, y::l) | NONE => filterP(r, l)

The Issue

Shift/Reduce Conflict

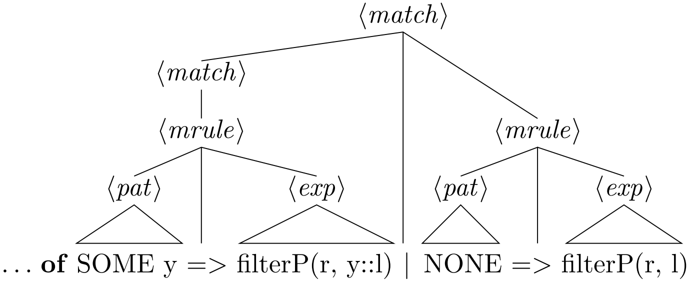
Which action to choose? Reduce?



The Issue

Shift/Reduce Conflict

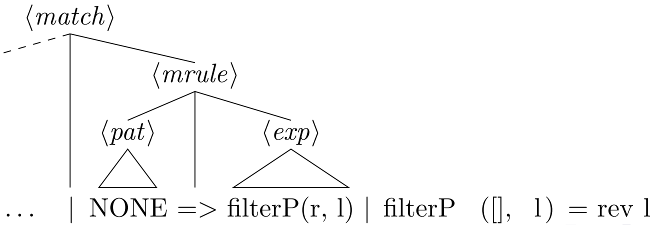
Which action to choose? Shift?



The Issue

Shift/Reduce Conflict

Which action to choose?

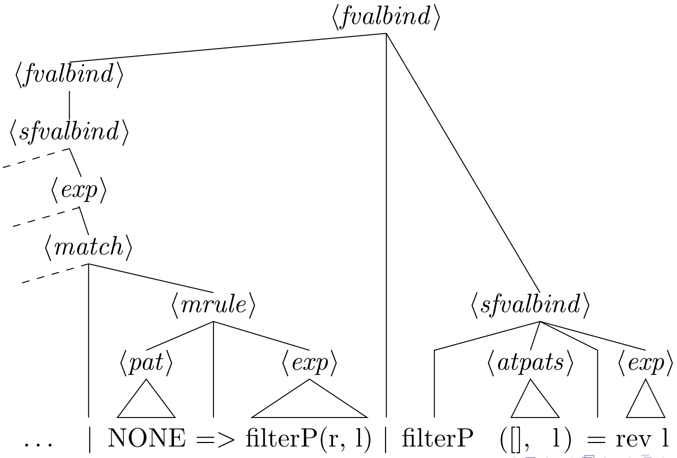


. . . | NONE => filterP(r, l) | filterP ([], l) = rev l

The Issue

Shift/Reduce Conflict

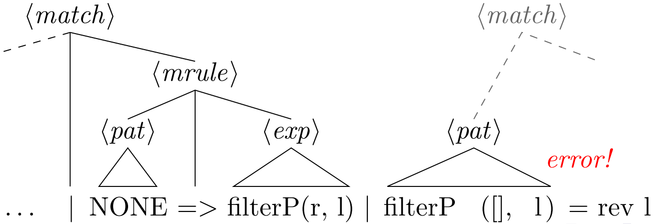
Which action to choose? Reduce?



The Issue

Shift/Reduce Conflict

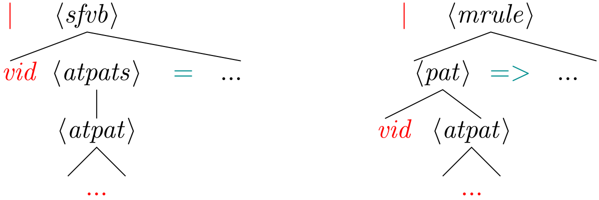
Which action to choose? Shift?



The Issue

Shift/Reduce Conflict

Which action to choose? Lookahead?



An unbounded lookahead is needed.

The Issue

Shift/Reduce Conflict

Which action to choose?

Example

Ambiguous contruct:

case a **of** b => **case** b **of** c => c | d => d

The Issue

Shift/Reduce Conflict

Which action to choose? Reduce?

Example

Ambiguous contruct:

```
case a of
     b => case b of
                   c => c
   | d => d
```

The Issue

Shift/Reduce Conflict

Which action to choose? Shift?

Example

Ambiguous contruct:

```
case a of
     b => case b of
                c => c
              | d => d
```

Here, the standard favors the shift.

Generalized LR

Tomita [1986]

- checks every possibility
- considered fast enough
- painless

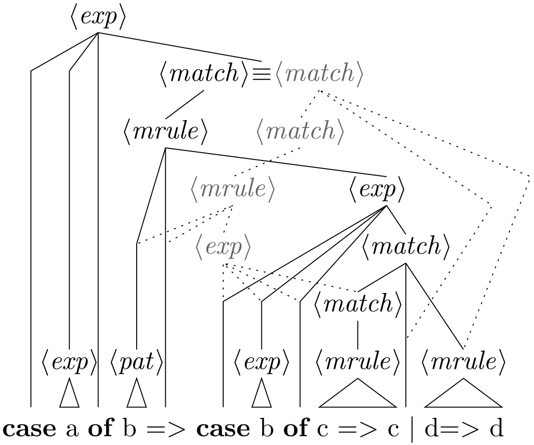
Generalized LR

Tomita [1986]

- checks every possibility
- considered fast enough
- painless. . . or is it?

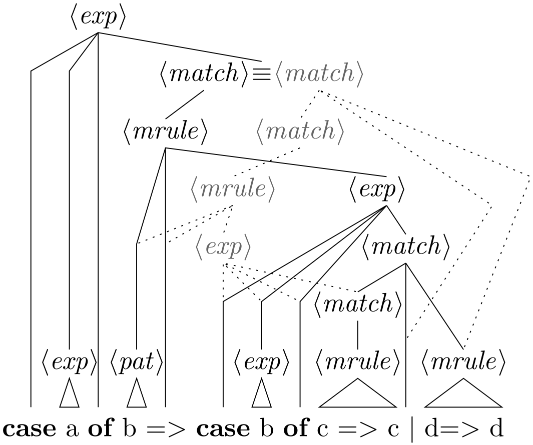
Generalized LR

- painless. . . or is it?
- need to consider ambiguity



Generalized LR

- need to consider ambiguity

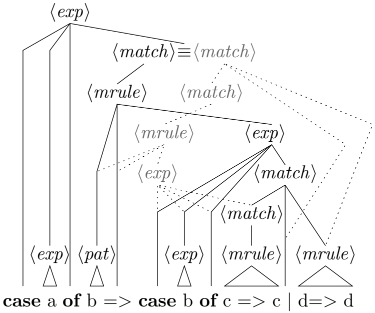


- select a single tree: disambiguation filters

Disambiguation Filters

Klint and Visser [1994]

Example



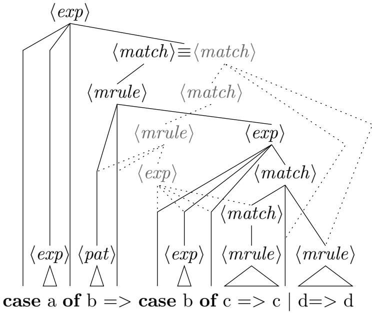
Select the tree according to the employed rule: ⟨*match*⟩→⟨*mrule*⟩ rather than ⟨*match*⟩→⟨*match*⟩′|′⟨*mrule*⟩.

- prefer and avoid in SDF2
- %dprec in GNU Bison

Disambiguation Filters

Klint and Visser [1994]

Example



Select the tree according to the employed rule: $\langle match\rangle \rightarrow \langle mrule\rangle$ rather than $\langle match\rangle \rightarrow \langle match\rangle '|' \langle mrule\rangle$.

- `prefer` and `avoid` in SDF2
- `%dprec` in GNU Bison

The Sword of Damocles

Do we catch all the ambiguities?

Example

case a **of** b => b | c => **case** c **of** d => d | e => e

The Sword of Damocles

Do we catch all the ambiguities?

Example

```
case a of
      b => b
    | c => case c of
                  d => d
                | e => e
```

The Sword of Damocles

Do we catch all the ambiguities?

Example

```
case a of
     b => b
   | c => case c of
                d => d
   | e => e
```

The Sword of Damocles

Do we catch all the ambiguities?

Example

case a **of** b => b | c => **case** c **of** d => d | e => e

The previous filter is inoperant.

- warnings in SDF2 and Elkhound, optionally aborts
- abort in GNU Bison

Detecting Ambiguities

How to find conservatively all the ambiguities in a grammar?

For the whole SML grammar:

- conflicts in the LALR(1) parser

  ```
  sml.y: conflicts: 223 shift/reduce, 35 reduce/reduce
  ```

- ?
- Our tool:

  ```
  89 potential ambiguities with LR(1) precision detected
  ```

Detecting Ambiguities

How to find conservatively all the ambiguities in a grammar?

For the whole SML grammar:

- conflicts in the LALR(1) parser

  ```
  sml.y: conflicts: 223 shift/reduce, 35 reduce/reduce
  ```

- ?
- Our tool:

  ```
  89 potential ambiguities with LR(1) precision detected
  ```

Detecting Ambiguities

How to find conservatively all the ambiguities in a grammar?
For the whole SML grammar:

- conflicts in the LALR(1) parser

  ```
  sml.y: conflicts: 223 shift/reduce, 35 reduce/reduce
  ```

- ?
- Our tool:

  ```
  89 potential ambiguities with LR(1) precision detected
  ```

Detecting Ambiguities

How to find conservatively all the ambiguities in a grammar?

For the whole SML grammar:

- conflicts in the LALR(1) parser

  ```
  sml.y: conflicts: 223 shift/reduce, 35 reduce/reduce
  ```

- ?
- Our tool:

  ```
  89 potential ambiguities with LR(1) precision detected
  ```

The Algorithm

Results on the SML grammar fragment

```
2 potential ambiguities with LR(0) precision detected:
    (match -> mrule . , match -> match . '|' mrule )
    (match -> match . '|' mrule , match -> match '|' mrule . )
```

Principle

- pairs of items
- synchronized to read the same prefix
- find pairs that might reach the end of input simultaneously

The Algorithm

Results on the SML grammar fragment

```
2 potential ambiguities with LR(0) precision detected:
    (match -> mrule . , match -> match . ’|’ mrule )
    (match -> match . ’|’ mrule , match -> match ’|’ mrule . )
```

Principle

- pairs of items
- synchronized to read the same prefix
- find pairs that might reach the end of input simultaneously

Example Run

$$(\langle \mathit{match} \rangle \longrightarrow \langle \mathit{match} \rangle \bullet \; '|' \; \langle \mathit{mrule} \rangle, \quad \langle \mathit{exp} \rangle \longrightarrow \textbf{case} \; \langle \mathit{exp} \rangle \; \textbf{of} \; \langle \mathit{match} \rangle \bullet)$$

Example Run

$$(\langle \mathit{match}\rangle \longrightarrow \langle \mathit{match}\rangle \bullet \; '|' \; \langle \mathit{mrule}\rangle, \quad \langle \mathit{exp}\rangle \longrightarrow \textbf{case} \; \langle \mathit{exp}\rangle \; \textbf{of} \; \langle \mathit{match}\rangle \bullet)$$

Example Run

$$(\langle match\rangle{\rightarrow}\langle match\rangle \bullet\ '|'\ \langle mrule\rangle,\quad \langle exp\rangle{\rightarrow}\textbf{case}\ \langle exp\rangle\ \textbf{of}\ \langle match\rangle \bullet)$$

$$(\langle match\rangle{\rightarrow}\langle match\rangle \bullet\ '|'\ \langle mrule\rangle,\quad \langle mrule\rangle{\rightarrow}\langle pat\rangle{=>}\langle exp\rangle \bullet)$$

$$(\langle match\rangle{\rightarrow}\langle match\rangle \bullet\ '|'\ \langle mrule\rangle,\quad \langle exp\rangle{\rightarrow}\textbf{case}\ \langle exp\rangle \bullet \textbf{of}\ \langle match\rangle)$$

$$(\langle match\rangle{\rightarrow}\langle match\rangle \bullet\ '|'\ \langle mrule\rangle,\quad \langle sfvb\rangle{\rightarrow}vid\ \langle atpats\rangle = \langle exp\rangle \bullet)$$

Example Run

$$(\langle \mathit{match}\rangle\rightarrow\langle \mathit{match}\rangle\,\bullet\,{}'|'\,\langle \mathit{mrule}\rangle,\quad \langle \mathit{mrule}\rangle\rightarrow\langle \mathit{pat}\rangle{=}{>}\langle \mathit{exp}\rangle\,\bullet)$$

$$(\langle \mathit{match}\rangle\rightarrow\langle \mathit{match}\rangle\,\bullet\,{}'|'\,\langle \mathit{mrule}\rangle,\quad \langle \mathit{exp}\rangle\rightarrow\textbf{case}\ \langle \mathit{exp}\rangle\,\bullet\,\textbf{of}\ \langle \mathit{match}\rangle)$$

$$(\langle \mathit{match}\rangle\rightarrow\langle \mathit{match}\rangle\,\bullet\,{}'|'\,\langle \mathit{mrule}\rangle,\quad \langle \mathit{sfvb}\rangle\rightarrow \mathit{vid}\ \langle \mathit{atpats}\rangle = \langle \mathit{exp}\rangle\,\bullet)$$

Example Run

$$(\langle \mathit{match}\rangle\rightarrow\langle \mathit{match}\rangle\bullet\ '|'\ \langle \mathit{mrule}\rangle,\quad \langle \mathit{mrule}\rangle\rightarrow\langle \mathit{pat}\rangle{=}{>}\langle \mathit{exp}\rangle\bullet)$$

$$(\langle \mathit{match}\rangle\rightarrow\langle \mathit{match}\rangle\bullet\ '|'\ \langle \mathit{mrule}\rangle,\quad \langle \mathit{exp}\rangle\rightarrow\textbf{case}\ \langle \mathit{exp}\rangle\bullet\textbf{of}\ \langle \mathit{match}\rangle)$$

$$(\langle \mathit{match}\rangle\rightarrow\langle \mathit{match}\rangle\bullet\ '|'\ \langle \mathit{mrule}\rangle,\quad \langle \mathit{sfvb}\rangle\rightarrow \mathit{vid}\ \langle \mathit{atpats}\rangle = \langle \mathit{exp}\rangle\bullet)$$

$$(\langle \mathit{match}\rangle\rightarrow\langle \mathit{match}\rangle\bullet\ '|'\ \langle \mathit{mrule}\rangle,\quad \langle \mathit{match}\rangle\rightarrow\langle \mathit{mrule}\rangle\bullet)$$

$$(\langle \mathit{match}\rangle\rightarrow\langle \mathit{match}\rangle\bullet\ '|'\ \langle \mathit{mrule}\rangle,\quad \langle \mathit{match}\rangle\rightarrow\langle \mathit{match}\rangle\ '|'\ \langle \mathit{mrule}\rangle\bullet)$$

Our two potential ambiguities

Example Run

$$(\langle match\rangle\rightarrow\langle match\rangle\bullet\ '|'\ \langle mrule\rangle,\quad \langle match\rangle\rightarrow\langle mrule\rangle\bullet)$$

$$(\langle match\rangle\rightarrow\langle match\rangle\bullet\ '|'\ \langle mrule\rangle,\quad \langle match\rangle\rightarrow\langle match\rangle\ '|'\ \langle mrule\rangle\bullet)$$

Our two potential ambiguities

Example Run

$$(\langle \mathit{match}\rangle \to \langle \mathit{match}\rangle \bullet\ '|'\ \langle \mathit{mrule}\rangle, \quad \langle \mathit{mrule}\rangle \to \langle \mathit{pat}\rangle {=>} \langle \mathit{exp}\rangle \bullet)$$

$$(\langle \mathit{match}\rangle \to \langle \mathit{match}\rangle \bullet\ '|'\ \langle \mathit{mrule}\rangle, \quad \langle \mathit{exp}\rangle \to \textbf{case}\ \langle \mathit{exp}\rangle \bullet \textbf{of}\ \langle \mathit{match}\rangle)$$

$$(\langle \mathit{match}\rangle \to \langle \mathit{match}\rangle \bullet\ '|'\ \langle \mathit{mrule}\rangle, \quad \langle \mathit{sfvb}\rangle \to \mathit{vid}\ \langle \mathit{atpats}\rangle = \langle \mathit{exp}\rangle \bullet)$$

Example Run

$$(\langle match\rangle\!\rightarrow\!\langle match\rangle \bullet \ '|' \ \langle mrule\rangle, \quad \langle mrule\rangle\!\rightarrow\!\langle pat\rangle\!=>\!\langle exp\rangle \bullet)$$

$$(\langle match\rangle\!\rightarrow\!\langle match\rangle \bullet \ '|' \ \langle mrule\rangle, \quad \langle exp\rangle\!\rightarrow\!\textbf{case} \ \langle exp\rangle \bullet \textbf{of} \ \langle match\rangle)$$

$$(\langle match\rangle\!\rightarrow\!\langle match\rangle \bullet \ '|' \ \langle mrule\rangle, \quad \langle sfvb\rangle\!\rightarrow\! vid \ \langle atpats\rangle = \langle exp\rangle \bullet)$$

Example Run

$$(\langle \mathit{match}\rangle \rightarrow \langle \mathit{match}\rangle \bullet\ '|'\ \langle \mathit{mrule}\rangle,\quad \langle \mathit{mrule}\rangle \rightarrow \langle \mathit{pat}\rangle {=}{>} \langle \mathit{exp}\rangle \bullet)$$

$$(\langle \mathit{match}\rangle \rightarrow \langle \mathit{match}\rangle \bullet\ '|'\ \langle \mathit{mrule}\rangle,\quad \langle \mathit{exp}\rangle \rightarrow \textbf{case}\ \langle \mathit{exp}\rangle \bullet \textbf{of}\ \langle \mathit{match}\rangle)$$

$$(\langle \mathit{match}\rangle \rightarrow \langle \mathit{match}\rangle \bullet\ '|'\ \langle \mathit{mrule}\rangle,\quad \langle \mathit{sfvb}\rangle \rightarrow \mathit{vid}\ \langle \mathit{atpats}\rangle = \langle \mathit{exp}\rangle \bullet)$$

$$(\langle \mathit{match}\rangle \rightarrow \langle \mathit{match}\rangle \bullet\ '|'\ \langle \mathit{mrule}\rangle,\quad \langle \mathit{fvalbind}\rangle \rightarrow \langle \mathit{sfvb}\rangle \bullet)$$

Example Run

$$(\langle \mathit{match}\rangle \to \langle \mathit{match}\rangle \bullet \ '|' \ \langle \mathit{mrule}\rangle, \quad \langle \mathit{fvalbind}\rangle \to \langle \mathit{sfvb}\rangle \bullet)$$

Example Run

$$(\langle match\rangle{\longrightarrow}\langle match\rangle\bullet\ '|'\ \langle mrule\rangle,\quad \langle fvalbind\rangle{\longrightarrow}\langle sfvb\rangle\bullet)$$

$$(\langle match\rangle{\longrightarrow}\langle match\rangle\bullet\ '|'\ \langle mrule\rangle,\quad \langle fvalbind\rangle{\longrightarrow}\langle fvalbind\rangle\bullet\ '|'\ \langle sfvb\rangle)$$

Example Run

$$(\langle \mathit{match} \rangle \rightarrow \langle \mathit{match} \rangle \bullet \; '|' \; \langle \mathit{mrule} \rangle, \quad \langle \mathit{fvalbind} \rangle \rightarrow \langle \mathit{fvalbind} \rangle \bullet \; '|' \; \langle \mathit{sfvb} \rangle)$$

Example Run

$$(\langle \mathit{match}\rangle{\rightarrow}\langle \mathit{match}\rangle \bullet \text{'|'}\ \langle \mathit{mrule}\rangle,\quad \langle \mathit{fvalbind}\rangle{\rightarrow}\langle \mathit{fvalbind}\rangle \bullet \text{'|'}\ \langle \mathit{sfvb}\rangle)$$

$$(\langle \mathit{match}\rangle{\rightarrow}\langle \mathit{match}\rangle\ \text{'|'} \bullet \langle \mathit{mrule}\rangle,\quad \langle \mathit{fvalbind}\rangle{\rightarrow}\langle \mathit{fvalbind}\rangle\ \text{'|'} \bullet \langle \mathit{sfvb}\rangle)$$

Example Run

$$(\langle match\rangle \rightarrow \langle match\rangle \ '|' \ \bullet \langle mrule\rangle, \quad \langle fvalbind\rangle \rightarrow \langle fvalbind\rangle \ '|' \ \bullet \langle sfvb\rangle)$$

Example Run

$$(\langle \mathit{match}\rangle \rightarrow \langle \mathit{match}\rangle\ '|'\ \bullet \langle \mathit{mrule}\rangle,\quad \langle \mathit{fvalbind}\rangle \rightarrow \langle \mathit{fvalbind}\rangle\ '|'\ \bullet \langle \mathit{sfvb}\rangle)$$

$$(\langle \mathit{mrule}\rangle \rightarrow \bullet \langle \mathit{pat}\rangle {=}{>} \langle \mathit{exp}\rangle,\quad \langle \mathit{fvalbind}\rangle \rightarrow \langle \mathit{fvalbind}\rangle\ '|'\ \bullet \langle \mathit{sfvb}\rangle)$$

Example Run

$$(\langle mrule\rangle\longrightarrow\bullet\langle pat\rangle{=}{>}\langle exp\rangle,\quad \langle fvalbind\rangle\longrightarrow\langle fvalbind\rangle\ '|'\ \bullet\langle sfvb\rangle)$$

Example Run

$$(\langle \mathit{mrule}\rangle \rightarrow \bullet \langle \mathit{pat}\rangle => \langle \mathit{exp}\rangle,\quad \langle \mathit{fvalbind}\rangle \rightarrow \langle \mathit{fvalbind}\rangle\ '|'\ \bullet \langle \mathit{sfvb}\rangle)$$

$$(\langle \mathit{mrule}\rangle \rightarrow \bullet \langle \mathit{pat}\rangle => \langle \mathit{exp}\rangle,\quad \langle \mathit{sfvb}\rangle \rightarrow \bullet\, \mathit{vid}\ \langle \mathit{atpats}\rangle = \langle \mathit{exp}\rangle)$$

$$(\langle \mathit{pat}\rangle \rightarrow \bullet\, \mathit{vid}\ \langle \mathit{atpat}\rangle,\quad \langle \mathit{sfvb}\rangle \rightarrow \bullet\, \mathit{vid}\ \langle \mathit{atpats}\rangle = \langle \mathit{exp}\rangle)$$

$$(\langle \mathit{pat}\rangle \rightarrow \mathit{vid} \bullet \langle \mathit{atpat}\rangle,\quad \langle \mathit{sfvb}\rangle \rightarrow \mathit{vid} \bullet \langle \mathit{atpats}\rangle = \langle \mathit{exp}\rangle)$$

$$(\langle \mathit{pat}\rangle \rightarrow \mathit{vid} \bullet \langle \mathit{atpat}\rangle,\quad \langle \mathit{atpats}\rangle \rightarrow \bullet \langle \mathit{atpat}\rangle)$$

$$(\langle \mathit{pat}\rangle \rightarrow \mathit{vid} \langle \mathit{atpat}\rangle \bullet,\quad \langle \mathit{atpats}\rangle \rightarrow \langle \mathit{atpat}\rangle \bullet)$$

$$(\langle \mathit{mrule}\rangle \rightarrow \langle \mathit{pat}\rangle \bullet => \langle \mathit{exp}\rangle,\quad \langle \mathit{atpats}\rangle \rightarrow \langle \mathit{atpat}\rangle \bullet)$$

Example Run

$$(\langle \mathit{mrule}\rangle \longrightarrow \bullet \langle \mathit{pat}\rangle {=>} \langle \mathit{exp}\rangle, \quad \langle \mathit{sfvb}\rangle \longrightarrow \bullet\, \mathit{vid}\ \langle \mathit{atpats}\rangle = \langle \mathit{exp}\rangle)$$

$$(\langle \mathit{pat}\rangle \longrightarrow \bullet\, \mathit{vid}\ \langle \mathit{atpat}\rangle, \quad \langle \mathit{sfvb}\rangle \longrightarrow \bullet\, \mathit{vid}\ \langle \mathit{atpats}\rangle = \langle \mathit{exp}\rangle)$$

$$(\langle \mathit{pat}\rangle \longrightarrow \mathit{vid} \bullet \langle \mathit{atpat}\rangle, \quad \langle \mathit{sfvb}\rangle \longrightarrow \mathit{vid} \bullet \langle \mathit{atpats}\rangle = \langle \mathit{exp}\rangle)$$

$$(\langle \mathit{pat}\rangle \longrightarrow \mathit{vid} \bullet \langle \mathit{atpat}\rangle, \quad \langle \mathit{atpats}\rangle \longrightarrow \bullet \langle \mathit{atpat}\rangle)$$

$$(\langle \mathit{pat}\rangle \longrightarrow \mathit{vid} \langle \mathit{atpat}\rangle \bullet, \quad \langle \mathit{atpats}\rangle \longrightarrow \langle \mathit{atpat}\rangle \bullet)$$

$$(\langle \mathit{mrule}\rangle \longrightarrow \langle \mathit{pat}\rangle \bullet {=>} \langle \mathit{exp}\rangle, \quad \langle \mathit{atpats}\rangle \longrightarrow \langle \mathit{atpat}\rangle \bullet)$$

$$(\langle \mathit{mrule}\rangle \longrightarrow \langle \mathit{pat}\rangle \bullet {=>} \langle \mathit{exp}\rangle, \quad \langle \mathit{sfvb}\rangle \longrightarrow \mathit{vid}\ \langle \mathit{atpats}\rangle \bullet = \langle \mathit{exp}\rangle)$$

Example Run

$$(\langle mrule\rangle \longrightarrow \bullet \langle pat\rangle => \langle exp\rangle, \quad \langle sfvb\rangle \longrightarrow \bullet\, vid\ \langle atpats\rangle = \langle exp\rangle)$$

$$(\langle pat\rangle \longrightarrow \bullet\, vid\ \langle atpat\rangle, \quad \langle sfvb\rangle \longrightarrow \bullet\, vid\ \langle atpats\rangle = \langle exp\rangle)$$

$$(\langle pat\rangle \longrightarrow vid \bullet \langle atpat\rangle, \quad \langle sfvb\rangle \longrightarrow vid \bullet \langle atpats\rangle = \langle exp\rangle)$$

$$(\langle pat\rangle \longrightarrow vid \bullet \langle atpat\rangle, \quad \langle atpats\rangle \longrightarrow \bullet \langle atpat\rangle)$$

$$(\langle pat\rangle \longrightarrow vid \langle atpat\rangle \bullet, \quad \langle atpats\rangle \longrightarrow \langle atpat\rangle \bullet)$$

$$(\langle mrule\rangle \longrightarrow \langle pat\rangle \bullet => \langle exp\rangle, \quad \langle atpats\rangle \longrightarrow \langle atpat\rangle \bullet)$$

$$(\langle mrule\rangle \longrightarrow \langle pat\rangle \bullet => \langle exp\rangle, \quad \langle sfvb\rangle \longrightarrow vid\ \langle atpats\rangle \bullet = \langle exp\rangle)$$

Example Run

$$(\langle mrule\rangle \rightarrow \bullet \langle pat\rangle {=}{>} \langle exp\rangle, \quad \langle sfvb\rangle \rightarrow \bullet\, vid\ \langle atpats\rangle = \langle exp\rangle)$$

$$(\langle pat\rangle \rightarrow \bullet\, vid\ \langle atpat\rangle, \quad \langle sfvb\rangle \rightarrow \bullet\, vid\ \langle atpats\rangle = \langle exp\rangle)$$

$$(\langle pat\rangle \rightarrow vid \bullet \langle atpat\rangle, \quad \langle sfvb\rangle \rightarrow vid \bullet \langle atpats\rangle = \langle exp\rangle)$$

$$(\langle pat\rangle \rightarrow vid \bullet \langle atpat\rangle, \quad \langle atpats\rangle \rightarrow \bullet \langle atpat\rangle)$$

$$(\langle pat\rangle \rightarrow vid \langle atpat\rangle \bullet, \quad \langle atpats\rangle \rightarrow \langle atpat\rangle \bullet)$$

$$(\langle mrule\rangle \rightarrow \langle pat\rangle \bullet {=}{>} \langle exp\rangle, \quad \langle atpats\rangle \rightarrow \langle atpat\rangle \bullet)$$

$$(\langle mrule\rangle \rightarrow \langle pat\rangle \bullet {=}{>} \langle exp\rangle, \quad \langle sfvb\rangle \rightarrow vid\ \langle atpats\rangle \bullet = \langle exp\rangle)$$

Example Run

$$(\langle \mathit{mrule}\rangle \to \bullet \langle \mathit{pat}\rangle {=}{>} \langle \mathit{exp}\rangle,\quad \langle \mathit{sfvb}\rangle \to \bullet\, \mathit{vid}\ \langle \mathit{atpats}\rangle = \langle \mathit{exp}\rangle)$$

$$(\langle \mathit{pat}\rangle \to \bullet\, \mathit{vid}\ \langle \mathit{atpat}\rangle,\quad \langle \mathit{sfvb}\rangle \to \bullet\, \mathit{vid}\ \langle \mathit{atpats}\rangle = \langle \mathit{exp}\rangle)$$

$$(\langle \mathit{pat}\rangle \to \mathit{vid} \bullet \langle \mathit{atpat}\rangle,\quad \langle \mathit{sfvb}\rangle \to \mathit{vid} \bullet \langle \mathit{atpats}\rangle = \langle \mathit{exp}\rangle)$$

$$(\langle \mathit{pat}\rangle \to \mathit{vid} \bullet \langle \mathit{atpat}\rangle,\quad \langle \mathit{atpats}\rangle \to \bullet \langle \mathit{atpat}\rangle)$$

$$(\langle \mathit{pat}\rangle \to \mathit{vid} \langle \mathit{atpat}\rangle \bullet,\quad \langle \mathit{atpats}\rangle \to \langle \mathit{atpat}\rangle \bullet)$$

$$(\langle \mathit{mrule}\rangle \to \langle \mathit{pat}\rangle \bullet {=}{>} \langle \mathit{exp}\rangle,\quad \langle \mathit{atpats}\rangle \to \langle \mathit{atpat}\rangle \bullet)$$

$$(\langle \mathit{mrule}\rangle \to \langle \mathit{pat}\rangle \bullet {=}{>} \langle \mathit{exp}\rangle,\quad \langle \mathit{sfvb}\rangle \to \mathit{vid}\ \langle \mathit{atpats}\rangle \bullet = \langle \mathit{exp}\rangle)$$

Example Run

$$(\langle mrule\rangle \rightarrow \bullet \langle pat\rangle {=>} \langle exp\rangle,\quad \langle sfvb\rangle \rightarrow \bullet\, vid\ \langle atpats\rangle\ =\ \langle exp\rangle)$$

$$(\langle pat\rangle \rightarrow \bullet\, vid\ \langle atpat\rangle,\quad \langle sfvb\rangle \rightarrow \bullet\, vid\ \langle atpats\rangle\ =\ \langle exp\rangle)$$

$$(\langle pat\rangle \rightarrow vid \bullet \langle atpat\rangle,\quad \langle sfvb\rangle \rightarrow vid \bullet \langle atpats\rangle\ =\ \langle exp\rangle)$$

$$(\langle pat\rangle \rightarrow vid \bullet \langle atpat\rangle,\quad \langle atpats\rangle \rightarrow \bullet \langle atpat\rangle)$$

$$(\langle pat\rangle \rightarrow vid \langle atpat\rangle \bullet,\quad \langle atpats\rangle \rightarrow \langle atpat\rangle \bullet)$$

$$(\langle mrule\rangle \rightarrow \langle pat\rangle \bullet {=>} \langle exp\rangle,\quad \langle atpats\rangle \rightarrow \langle atpat\rangle \bullet)$$

$$(\langle mrule\rangle \rightarrow \langle pat\rangle \bullet {=>} \langle exp\rangle,\quad \langle sfvb\rangle \rightarrow vid\ \langle atpats\rangle \bullet\ =\ \langle exp\rangle)$$

Example Run

$$(\langle \mathit{mrule}\rangle \to \bullet \langle \mathit{pat}\rangle => \langle \mathit{exp}\rangle,\quad \langle \mathit{sfvb}\rangle \to \bullet\, \mathit{vid}\ \langle \mathit{atpats}\rangle\ =\ \langle \mathit{exp}\rangle)$$

$$(\langle \mathit{pat}\rangle \to \bullet\, \mathit{vid}\ \langle \mathit{atpat}\rangle,\quad \langle \mathit{sfvb}\rangle \to \bullet\, \mathit{vid}\ \langle \mathit{atpats}\rangle\ =\ \langle \mathit{exp}\rangle)$$

$$(\langle \mathit{pat}\rangle \to \mathit{vid} \bullet \langle \mathit{atpat}\rangle,\quad \langle \mathit{sfvb}\rangle \to \mathit{vid} \bullet \langle \mathit{atpats}\rangle\ =\ \langle \mathit{exp}\rangle)$$

$$(\langle \mathit{pat}\rangle \to \mathit{vid} \bullet \langle \mathit{atpat}\rangle,\quad \langle \mathit{atpats}\rangle \to \bullet \langle \mathit{atpat}\rangle)$$

$$(\langle \mathit{pat}\rangle \to \mathit{vid}\langle \mathit{atpat}\rangle \bullet,\quad \langle \mathit{atpats}\rangle \to \langle \mathit{atpat}\rangle \bullet)$$

$$(\langle \mathit{mrule}\rangle \to \langle \mathit{pat}\rangle \bullet => \langle \mathit{exp}\rangle,\quad \langle \mathit{atpats}\rangle \to \langle \mathit{atpat}\rangle \bullet)$$

$$(\langle \mathit{mrule}\rangle \to \langle \mathit{pat}\rangle \bullet => \langle \mathit{exp}\rangle,\quad \langle \mathit{sfvb}\rangle \to \mathit{vid}\ \langle \mathit{atpats}\rangle \bullet\ =\ \langle \mathit{exp}\rangle)$$

Example Run

$$(\langle mrule\rangle \to \bullet \langle pat\rangle {=}{>} \langle exp\rangle, \quad \langle sfvb\rangle \to \bullet\, vid\ \langle atpats\rangle = \langle exp\rangle)$$

$$(\langle pat\rangle \to \bullet\, vid\ \langle atpat\rangle, \quad \langle sfvb\rangle \to \bullet\, vid\ \langle atpats\rangle = \langle exp\rangle)$$

$$(\langle pat\rangle \to vid \bullet \langle atpat\rangle, \quad \langle sfvb\rangle \to vid \bullet \langle atpats\rangle = \langle exp\rangle)$$

$$(\langle pat\rangle \to vid \bullet \langle atpat\rangle, \quad \langle atpats\rangle \to \bullet \langle atpat\rangle)$$

$$(\langle pat\rangle \to vid \langle atpat\rangle \bullet, \quad \langle atpats\rangle \to \langle atpat\rangle \bullet)$$

$$(\langle mrule\rangle \to \langle pat\rangle \bullet {=}{>} \langle exp\rangle, \quad \langle atpats\rangle \to \langle atpat\rangle \bullet)$$

$$(\langle mrule\rangle \to \langle pat\rangle \bullet {=}{>} \langle exp\rangle, \quad \langle sfvb\rangle \to vid\ \langle atpats\rangle \bullet = \langle exp\rangle)$$

Overview of the Algorithm

Schmitz [2006]

Computation of (**mac** $\cup$ **mas** $\cup$ **mae**)* from the initial item pair

$$(S' \rightarrow .S\$, \quad S' \rightarrow .S\$)$$

mac from a pair in conflict to its upwards closure,

mas by shifting a symbol,

mae by downwards closure.

Complexity: $\mathcal{O}(|R|^2)$ with $|R|$ the number of relations for individual items.

LR(0), SLR(1) $\mathcal{O}(|\mathcal{G}|^2)$

LR(1) $\mathcal{O}(|\mathcal{G}|^2\,|T|^4)$

Overview of the Algorithm

Schmitz [2006]

Computation of (**mac** $\cup$ **mas** $\cup$ **mae**)* from the initial item pair

$$(S' \rightarrow .S\$, \quad S' \rightarrow .S\$)$$

mac from a pair in conflict to its upwards closure,

mas by shifting a symbol,

mae by downwards closure.

Complexity: $\mathcal{O}(|R|^2)$ with $|R|$ the number of relations for individual items.

LR(0), SLR(1) $\mathcal{O}(|\mathcal{G}|^2)$

LR(1) $\mathcal{O}(|\mathcal{G}|^2\,|T|^4)$

Overview of the Algorithm

Schmitz [2006]

Computation of (**mac** $\cup$ **mas** $\cup$ **mae**)* from the initial item pair

$$(S' \rightarrow .S\$, \quad S' \rightarrow .S\$)$$

mac from a pair in conflict to its upwards closure,

mas by shifting a symbol,

mae by downwards closure.

Complexity: $\mathcal{O}(|R|^2)$ with $|R|$ the number of relations for individual items.

LR(0), SLR(1) $\mathcal{O}(|\mathcal{G}|^2)$

LR(1) $\mathcal{O}(|\mathcal{G}|^2\,|T|^4)$

Overview of the Algorithm

Schmitz [2006]

Computation of (**mac** ∪ **mas** ∪ **mae**)* from the initial item pair

$$(S' \rightarrow . S\$, \quad S' \rightarrow . S\$)$$

mac	from a pair in conflict to its upwards closure,
mas	by shifting a symbol,
mae	by downwards closure.

Complexity: $\mathcal{O}(|R|^2)$ with $|R|$ the number of relations for individual items.

LR(0), SLR(1) $\mathcal{O}(|\mathcal{G}|^2)$

LR(1) $\mathcal{O}(|\mathcal{G}|^2\,|T|^4)$

Overview of the Algorithm

Schmitz [2006]

Computation of (**mac** $\cup$ **mas** $\cup$ **mae**)* from the initial item pair

$$(S' \rightarrow .S\$, \quad S' \rightarrow .S\$)$$

mac from a pair in conflict to its upwards closure,

mas by shifting a symbol,

mae by downwards closure.

Complexity: $\mathcal{O}(|R|^2)$ with $|R|$ the number of relations for individual items.

LR(0), SLR(1) $\mathcal{O}(|\mathcal{G}|^2)$

LR(1) $\mathcal{O}(|\mathcal{G}|^2\,|T|^4)$

Some Experiments

Number of conflicts reported by Bison, and number of initial LR(0) conflicting pairs remaining with the LR, LRR and NU tests employing successively LR(0), SLR(1) and LR(1) items.

Grammar	Bison	LR(0) items			SLR(1) items			LR(1) items		
		LR	LRR	NU	LR	LRR	NU	LR	LRR	NU
Pascal	1	243	56	56	5	5	5	1	1	1
Mini C	1	300	16	15	5	5	4	1	1	1
ANSI C	1	377	14	3	14	14	3	1	1	1
ANSI C'	38	387	43	32	43	43	32	29	-	-
Standard ML	258	477	299	271	165	163	135	127	125	104
Small Elsa C++	58	1379	278	226	66	63	58	64	-	-
Elsa C++	115	2094	323	308	91	88	77	-	-	-

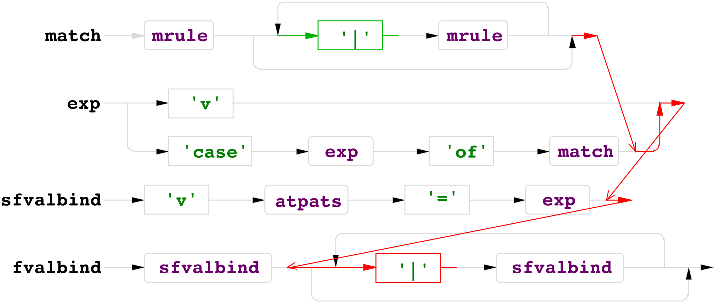
Limitations

Ambiguity Report

- grambiguity [Brabrand et al., 2006]

```
*** horizontal ambiguity at E[plus]: Exp <--> '+' Exp
    ambiguous string: "x+x+x"
```

- ANTLRWorks [Parr, 2007]



match mrule '|' mrule

exp 'v'

'case' exp 'of' match

sfvalbind 'v' atpats '=' exp

fvalbind sfvalbind '|' sfvalbind

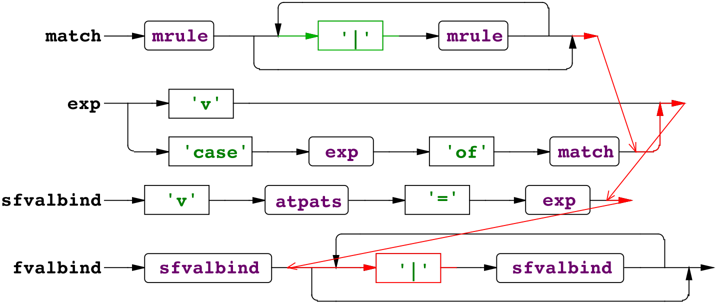
Limitations

Ambiguity Report

- grambiguity [Brabrand et al., 2006]

```
*** horizontal ambiguity at E[plus]: Exp <--> '+' Exp
    ambiguous string: "x+x+x"
```

- ANTLRWorks [Parr, 2007]



Other Limitations

- memory requirements: a solution could be a NLALR test
- dynamic disambiguation: inverse problem, some means to deciding equivalence needed

Closing Comments

- ambiguity hampers the reliability of grammarware
- conservative detection techniques can be improved
- prototype tool
- many, many open questions left

References

C. Brabrand, R. Giegerich, and A. Møller. Analyzing ambiguity of context-free grammars. Technical Report RS-06-09, BRICS, University of Aarhus, May 2006. URL http://www.brics.dk/~brabrand/grambiguity/.

C. Donnely and R. Stallman. Bison version 2.1, Sept. 2005. URL http://www.gnu.org/software/bison/manual/.

P. Klint and E. Visser. Using filters for the disambiguation of context-free grammars. In G. Pighizzini and P. San Pietro, editors, ASMICS Workshop on Parsing Theory, Technical Report 126-1994, pages 89–100. Università di Milano, 1994. URL http://citeseer.ist.psu.edu/klint94using.html.

S. McPeak and G. C. Necula. Elkhound: A fast, practical GLR parser generator. In E. Duesterwald, editor, CC'04, volume 2985 of Lecture Notes in Computer Science, pages 73–88. Springer, 2004. ISBN 3-540-21297-3. doi: 10.1007/b95956.

R. Milner, M. Tofte, R. Harper, and D. MacQueen. The definition of Standard ML. MIT Press, revised edition, 1997. ISBN 0-262-63181-4.

T. J. Parr. The Definitive ANTLR Reference: Building Domain-Specific Languages. The Pragmatic Programmers, 2007. ISBN 0-9787392-5-6.

S. Schmitz. Conservative ambiguity detection in context-free grammars. Technical Report I3S/RR-2006-30-FR, Laboratoire I3S, Université de Nice - Sophia Antipolis & CNRS, Sept. 2006. URL `http://www.i3s.unice.fr/~mh/RR/2006/RR-06.30-S.SCHMITZ.pdf`.

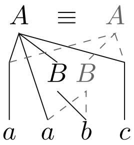
M. Tomita. Efficient Parsing for Natural Language. Kluwer Academic Publishers, 1986. ISBN 0-89838-202-5.

M. van den Brand, J. Scheerder, J. J. Vinju, and E. Visser. Disambiguation filters for scannerless generalized LR parsers. In R. N. Horspool, editor, CC'02, volume 2304 of Lecture Notes in Computer Science, pages 143–158. Springer, 2002. ISBN 3-540-43369-4. URL http://www.springerlink.com/content/03359k0cerupftfh/.

E. Visser. Syntax Definition for Language Prototyping. PhD thesis, Sept. 1997. URL http://citeseer.ist.psu.edu/visser97syntax.html.

Dynamic Disambiguation Filters

$$A \longrightarrow aBc \mid aaBc, \; B \longrightarrow ab \mid b \qquad (\mathcal{G}_1)$$


$A \equiv A$
B B
a a b c