

## Analysis of the Given Grammar

(February 18, 2016)

*COLORS in this document*

*Black: for original rules in the grammar in Language Specification document*

*Red: specifies needs for modifications*

*Green: rules do not require modification*

### Grammar:

*The nonterminal <program> is the start symbol of the given grammar.*

1. **<program>**  $\Rightarrow$  <otherFunctions> <mainFunction>

The LL(1) property for the nonterminal <program> is satisfied trivially due to single production for it. The following FOLLOW set will be required for verifying LL(1) compatibility for the nonterminal <otherFunctions> in (3).

$\text{FOLLOW}(\text{<otherFunctions>}) = \text{FIRST}(\text{<mainFunction>}) = \{\text{TK\_MAIN}\}$

Note that all function definitions precede the main function definition and the language does not have constructs for function prototype declarations.

2. <mainFunction>  $\Rightarrow$  TK\_MAIN <stmts> TK\_END

The LL(1) property for the nonterminal <mainFunction> is satisfied trivially due to single production for it.

$\text{FIRST}(\text{<mainFunction>}) = \text{FIRST}(\alpha) = \{\text{TK\_MAIN}\}$

where  $\alpha$  represents the right hand side of the production i.e. TK\_MAIN <stmts> TK\_END.

3. <otherFunctions>  $\Rightarrow$  <function><otherFunctions> | eps

The nonterminal <otherFunctions> need special care for verifying the LL(1) compatibility due to a nullable production. Let us first verify that the sets  $\text{FIRST}(\text{<function><otherFunctions>})$  and  $\text{FOLLOW}(\text{<otherFunctions>})$  are disjoint.

$\text{FIRST}(\text{<function><otherFunctions>}) = \text{FIRST}(\text{<function>}) = \{\text{TK\_FUNID}\}$

Note that <function> has no nullable production.

Also from 1, we have  $\text{FOLLOW}(\langle \text{otherFunctions} \rangle) = \{\text{TK\_MAIN}\}$

i.e.

$\text{FIRST}(\langle \text{function} \rangle \langle \text{otherFunctions} \rangle) \cap \text{FOLLOW}(\langle \text{otherFunctions} \rangle) = \phi$

The given productions for  $\langle \text{otherFunctions} \rangle$  are LL(1) compatible.

NOTE: Other properties such as ambiguity, left recursiveness, left factoring needs etc. causing violation of LL(1) compatibility of the rules will be discussed only if one or more of them exist. Otherwise I am focussing on violations due to epsilon productions. Also I will highlight the introduction of new nonterminals to incorporate the precedence of arithmetic operators and handling operations on record variables.

4.  $\langle \text{function} \rangle ::= \text{TK\_FUNID } \langle \text{input\_par} \rangle \langle \text{output\_par} \rangle \text{TK\_SEM } \langle \text{stmts} \rangle \text{TK\_END}$

This has no issue of violations in LL(1) compatibility due to epsilon productions.  $\langle \text{input\_par} \rangle$  is an essential construct to be part of the function definition while a function may or may not return values. Hence the  $\langle \text{output\_par} \rangle$  can have a syntax of (6) .

$\text{FIRST}(\langle \text{function} \rangle) = \{\text{TK\_FUNID}\}$

5.  $\langle \text{input\_par} \rangle ::= \text{TK\_INPUT } \text{TK\_PARAMETER } \text{TK\_LIST } \text{TK\_SQL } \langle \text{parameter\_list} \rangle \text{TK\_SQR}$

Single production having no conflict has its FIRST set as given follows:

$\text{FIRST}(\langle \text{input\_par} \rangle) = \{\text{TK\_INPUT}\}$

6.  $\langle \text{output\_par} \rangle ::= \text{TK\_OUTPUT } \text{TK\_PARAMETER } \text{TK\_LIST } \text{TK\_SQL } \langle \text{parameter\_list} \rangle \text{TK\_SQR} \mid \text{eps}$

Presence of epsilon production makes us verify whether  $\text{FIRST}(\alpha) \cap \text{FOLLOW}(\langle \text{output\_par} \rangle) = \phi$  or not, where  $\alpha$  represents the right hand side

$\text{TK\_OUTPUT } \text{TK\_PARAMETER } \text{TK\_LIST } \text{TK\_SQL } \langle \text{parameter\_list} \rangle \text{TK\_SQR}$

$\text{FIRST}(\alpha) = \{\text{TK\_OUTPUT}\}$

Refer rule 4 to compute the  $\text{FOLLOW}(\langle \text{output\_par} \rangle)$  as below

$\text{FOLLOW}(\langle \text{output\_par} \rangle) = \{\text{TK\_SEM}\}$

This implies that  $\text{FIRST}(\alpha) \cap \text{FOLLOW}(\langle \text{output\_par} \rangle) = \phi$

Hence the given rules for  $\langle \text{output\_par} \rangle$  conform to the LL(1) specifications.

7.  $\langle \text{parameter\_list} \rangle ::= \langle \text{dataType} \rangle \text{TK\_ID } \langle \text{remaining\_list} \rangle$

The single production for  $\langle \text{parameter\_list} \rangle$  is trivially LL(1) compatible.

$\text{FIRST}(\langle \text{parameter\_list} \rangle) = \text{FIRST}(\langle \text{dataType} \rangle)$

$$\begin{aligned}
&= \text{FIRST}(\langle \text{primitiveDatatype} \rangle) \cup \text{FIRST}(\langle \text{constructedDatatype} \rangle) \\
&= \{ \text{TK\_INT}, \text{TK\_REAL}, \text{TK\_RECORD} \}
\end{aligned}$$

$$\begin{aligned}
8. \quad \langle \text{dataType} \rangle &\Rightarrow \langle \text{primitiveDatatype} \rangle \mid \langle \text{constructedDatatype} \rangle \\
&\text{FIRST}(\langle \text{primitiveDatatype} \rangle) \cap \text{FIRST}(\langle \text{constructedDatatype} \rangle) \\
&= \{ \text{TK\_INT}, \text{TK\_REAL} \} \cap \{ \text{TK\_RECORD} \} \\
&= \phi
\end{aligned}$$

Since there is no nullable production for  $\langle \text{dataType} \rangle$ , there is no need for computing  $\text{FOLLOW}(\langle \text{dataType} \rangle)$ .

$$\begin{aligned}
9. \quad \langle \text{primitiveDatatype} \rangle &\Rightarrow \text{TK\_INT} \mid \text{TK\_REAL} \\
&\text{FIRST}(\text{TK\_INT}) \cap \text{FIRST}(\text{TK\_REAL}) = \{ \text{TK\_INT} \} \cap \{ \text{TK\_REAL} \} = \phi \\
&\text{Hence the rules for } \langle \text{primitiveDatatype} \rangle \text{ are LL(1) compatible.} \\
&\text{FIRST}(\langle \text{primitiveDatatype} \rangle) = \{ \text{TK\_INT}, \text{TK\_REAL} \}
\end{aligned}$$

$$\begin{aligned}
10. \quad \langle \text{constructedDatatype} \rangle &\Rightarrow \text{TK\_RECORD TK\_RECORDID} \\
&\text{FIRST}(\langle \text{constructedDatatype} \rangle) = \{ \text{TK\_RECORD} \}
\end{aligned}$$

$$11. \quad \langle \text{remaining\_list} \rangle \Rightarrow \text{TK\_COMMA} \langle \text{parameter\_list} \rangle \mid \text{eps}$$

There exist a nullable production for the non-terminal  $\langle \text{remaining\_list} \rangle$ . The RHS (right hand side) of the other production is such that the null string  $\text{eps}$  (epsilon) is not derivable from it.

Also

$\text{FIRST}(\text{TK\_COMMA} \langle \text{parameter\_list} \rangle) \cap \text{FOLLOW}(\langle \text{remaining\_list} \rangle)$  should be empty.

$$\text{FIRST}(\text{TK\_COMMA} \langle \text{parameter\_list} \rangle) = \{ \text{TK\_COMMA} \}$$

and

$$\begin{aligned}
\text{FOLLOW}(\langle \text{remaining\_list} \rangle) &= \text{FOLLOW}(\text{parameter\_list}) && \text{.....from (7)} \\
&= \{ \text{TK\_SQR} \}
\end{aligned}$$

Hence,

$$\begin{aligned} & \text{FIRST}(\text{TK\_COMMA } \langle \text{parameter\_list} \rangle) \cap \text{FOLLOW}(\langle \text{remaining\_list} \rangle) \\ &= \{ \text{TK\_COMMA} \} \cap \{ \text{TK\_SQR} \} \\ &= \phi \end{aligned}$$

Therefore both the given rules for the non-terminal  $\langle \text{remaining\_list} \rangle$  conform to LL(1) specifications.

## 12. $\langle \text{stmts} \rangle ::= \langle \text{typeDefinitions} \rangle \langle \text{declarations} \rangle \langle \text{otherStmts} \rangle \langle \text{returnStmt} \rangle$

The nonterminal  $\langle \text{stmts} \rangle$  specifies the grammar for the body of the function. The ordering of other nonterminals such as  $\langle \text{typeDefinitions} \rangle$ ,  $\langle \text{declarations} \rangle$  and  $\langle \text{returnStmt} \rangle$  have fixed positions within the body of the function code.

There are no epsilon productions and a single non nullable production for  $\langle \text{stmts} \rangle$  is LL(1) compatible.

$$\text{FIRST}(\langle \text{stmts} \rangle) = \text{FIRST}(\langle \text{typeDefinitions} \rangle) \cup \text{FIRST}(\langle \text{declarations} \rangle) \cup \text{FIRST}(\langle \text{otherStatements} \rangle) \cup \text{FIRST}(\langle \text{returnStmt} \rangle)$$

.....using 13, 18,21 and 42

$$= \{ \text{TK\_RECORD} \} \cup \{ \text{TK\_TYPE} \} \cup \{ \text{TK\_ID}, \text{TK\_RECORD\_ID}, \text{TK\_WHILE}, \text{TK\_IF}, \text{TK\_READ}, \text{TK\_WRITE}, \text{TK\_SQL}, \text{TK\_CALL} \} \cup \{ \text{TK\_RETURN} \}$$

$$= \{ \text{TK\_RECORD}, \text{TK\_TYPE}, \text{TK\_ID}, \text{TK\_RECORD\_ID}, \text{TK\_WHILE}, \text{TK\_IF}, \text{TK\_READ}, \text{TK\_WRITE}, \text{TK\_SQL}, \text{TK\_CALL}, \text{TK\_RETURN} \}$$

Since the nonterminal  $\langle \text{stmts} \rangle$  does not have a nullable production, we need not compute FOLLOW( $\langle \text{stmts} \rangle$ ) for populating the parsing table.

Recall the construction of the parsing table. The columns corresponding to the tokens in FIRST( $\langle \text{stmts} \rangle$ ) corresponding to the row  $\langle \text{stmts} \rangle$  are populated with the rule number 12.

## 13. $\langle \text{typeDefinitions} \rangle ::= \langle \text{typeDefinition} \rangle \langle \text{typeDefinitions} \rangle \mid \text{eps}$

The first of the above productions is not nullable as  $\text{FIRST}(\langle \text{typeDefinition} \rangle \langle \text{typeDefinitions} \rangle) = \text{FIRST}(\langle \text{typeDefinition} \rangle) = \{ \text{TK\_RECORD} \}$ .

Also

$$\text{FOLLOW}(\langle \text{typeDefinitions} \rangle) = \text{FIRST}(\langle \text{declarations} \rangle) = \text{FIRST}(\langle \text{declaration} \rangle) = \{ \text{TK\_TYPE} \}$$

Next,

$$\text{FIRST}(\langle \text{typeDefinition} \rangle \langle \text{typeDefinitions} \rangle) \cap \text{FOLLOW}(\langle \text{typeDefinitions} \rangle)$$

$$= \{ \text{TK\_RECORD} \} \cap \{ \text{TK\_TYPE} \} = \phi$$

Hence the give two productions for the non terminal  $\langle \text{typeDefinitions} \rangle$  conform to the LL(1) specifications.

$\text{FIRST}(\langle \text{typeDefinitions} \rangle) = \text{FIRST}(\langle \text{typeDefinition} \rangle) = \{\text{TK\_RECORD}\}$

Recall the construction of the Predictive Parsing table.

The rule  $\langle \text{typeDefinitions} \rangle \Rightarrow \text{eps}$  is populated as the table entry  $T[\langle \text{typeDefinitions} \rangle, \text{TK\_TYPE}]$  and

the rule  $\langle \text{typeDefinitions} \rangle \Rightarrow \langle \text{typeDefinition} \rangle \langle \text{typeDefinitions} \rangle$  is populated as the table entry  $T[\langle \text{typeDefinition} \rangle, \text{TK\_RECORD}]$

14.  $\langle \text{typeDefinition} \rangle \Rightarrow \text{TK\_RECORD TK\_RECORDID} \langle \text{fieldDefinitions} \rangle \text{TK\_ENDRECORD TK\_SEM}$

Here on, I am ignoring the check for LL(1) if the nonterminal has a single non nullable production.

$\text{FIRST}(\langle \text{typeDefinition} \rangle) = \{\text{TK\_RECORD}\}$

15.  $\langle \text{fieldDefinitions} \rangle \Rightarrow \langle \text{fieldDefinition} \rangle \langle \text{fieldDefinition} \rangle \langle \text{moreFields} \rangle$

As two consecutive occurrences of  $\langle \text{fieldDefinition} \rangle$  are there at the RHS of the above production, this imposes a requirement of atleast two fields in a record. There can be one or more fields afterwrds. As the above rule is single and non nullable, it is LL(1) compatible.

$\text{FIRST}(\langle \text{fieldDefinitions} \rangle) = \text{FIRST}(\langle \text{fieldDefinition} \rangle) = \{\text{TK\_TYPE}\}$

16.  $\langle \text{fieldDefinition} \rangle \Rightarrow \text{TK\_TYPE} \langle \text{primitiveDatatype} \rangle \text{TK\_COLON TK\_FIELDID TK\_SEM}$

$\text{FIRST}(\langle \text{fieldDefinition} \rangle) = \{\text{TK\_TYPE}\}$

17.  $\langle \text{moreFields} \rangle \Rightarrow \langle \text{fieldDefinition} \rangle \langle \text{moreFields} \rangle \mid \text{eps}$

$\text{FOLLOW}(\langle \text{moreFields} \rangle) = \phi \cup \text{FOLLOW}(\langle \text{fieldDefinitions} \rangle) = \{\text{TK\_ENDRECORD}\}$

.....using 15 & 14

$\text{FIRST}(\langle \text{fieldDefinition} \rangle \langle \text{moreFields} \rangle) \cap \text{FOLLOW}(\langle \text{moreFields} \rangle)$

$= \{\text{TK\_TYPE}\} \cap \{\text{TK\_ENDRECORD}\}$

$= \phi$

Hence, the above two rules for the non terminal  $\langle \text{moreFields} \rangle$  are LL(1) compatible.

18.  $\langle \text{declarations} \rangle \Rightarrow \langle \text{declaration} \rangle \langle \text{declarations} \rangle | \epsilon$

$$\begin{aligned} \text{FOLLOW}(\langle \text{declarations} \rangle) &= \text{FIRST}(\langle \text{otherStmts} \rangle) \cup \text{FOLLOW}(\langle \text{otherStmts} \rangle) \\ &= \{ \text{TK\_ID}, \text{TK\_RECORD\_ID}, \text{TK\_WHILE}, \text{TK\_IF}, \text{TK\_READ}, \text{TK\_WRITE}, \text{TK\_SQL}, \text{TK\_CALL} \} \end{aligned}$$

$$\text{FIRST}(\langle \text{declaration} \rangle \langle \text{declarations} \rangle) = \text{FIRST}(\langle \text{declaration} \rangle) = \{ \text{TK\_TYPE} \}$$

We observe that both of the above sets are disjoint. Therefore both the above productions for the non terminal  $\langle \text{declarations} \rangle$  conform to LL(1) specifications.

19.  $\langle \text{declaration} \rangle \Rightarrow \text{TK\_TYPE} \langle \text{dataType} \rangle \text{TK\_COLON} \text{TK\_ID} \text{TK\_COLON} \langle \text{global\_or\_not} \rangle \text{TK\_SEM}$

This rule has a correction as follows

$$\langle \text{declaration} \rangle \Rightarrow \text{TK\_TYPE} \langle \text{dataType} \rangle \text{TK\_COLON} \text{TK\_ID} \langle \text{global\_or\_not} \rangle \text{TK\_SEM}$$

Also to incorporate the optional existence of global keyword prefixed with a colon is reflected as modification in rule 20.

$$\text{FIRST}(\langle \text{declaration} \rangle) = \{ \text{TK\_TYPE} \}$$

Rule conforms to LL(1) specifications.

20.  $\langle \text{global\_or\_not} \rangle \Rightarrow \text{TK\_GLOBAL} | \epsilon$

This rule is also modified as follows

$$\langle \text{global\_or\_not} \rangle \Rightarrow \text{TK\_COLON} \text{TK\_GLOBAL} | \epsilon$$

$$\text{FOLLOW}(\langle \text{global\_or\_not} \rangle) = \{ \text{TK\_SEM} \}$$

.....using 19

and

$$\text{FIRST}(\text{TK\_COLON} \text{TK\_GLOBAL}) = \{ \text{TK\_COLON} \}$$

As

$$\text{FOLLOW}(\langle \text{global\_or\_not} \rangle) \cap \text{FIRST}(\text{TK\_COLON} \text{TK\_GLOBAL}) = \{ \text{TK\_SEM} \} \cap \{ \text{TK\_COLON} \} = \phi$$

Therefore the new rule for  $\langle \text{global\_or\_not} \rangle$  is LL(1) compatible.

21.  $\langle \text{otherStmts} \rangle \Rightarrow \langle \text{stmt} \rangle \langle \text{otherStmts} \rangle | \epsilon$

$$\text{FOLLOW}(\langle \text{otherStmts} \rangle) = \text{FIRST}(\langle \text{returnStmts} \rangle)$$

$$\text{FIRST}(\langle \text{otherStmts} \rangle) = \text{FIRST}(\langle \text{stmt} \rangle)$$

[ Note:  $\langle \text{stmt} \rangle$  does not derive an empty string]

$$= \{ \text{TK\_ID}, \text{TK\_RECORDID}, \text{TK\_WHILE}, \text{TK\_IF}, \text{TK\_READ}, \text{TK\_WRITE}, \text{TK\_SQL}, \text{TK\_CALL} \}$$

.....using 22

22.  $\langle \text{stmt} \rangle ::= \langle \text{assignmentStmt} \rangle \mid \langle \text{iterativeStmt} \rangle \mid \langle \text{conditionalStmt} \rangle \mid \langle \text{ioStmt} \rangle \mid \langle \text{funCallStmt} \rangle$

Referring rules 23, 28, 29, 31 and 25, we conclude that the five productions given above for the nonterminal  $\langle \text{stmt} \rangle$  are not nullable.

We must check the LL(1) property that the FIRST sets of the RHSs of all productions above are disjoint.

$$\text{FIRST}(\langle \text{assignmentStmt} \rangle) \cap \text{FIRST}(\langle \text{iterativeStmt} \rangle) \cap \text{FIRST}(\langle \text{conditionalStmt} \rangle) \cap \text{FIRST}(\langle \text{ioStmt} \rangle) \cap \text{FIRST}(\langle \text{funCallStmt} \rangle)$$

$$= \{ \text{TK\_ID}, \text{TK\_RECORDID} \} \cap \{ \text{TK\_WHILE} \} \cap \{ \text{TK\_IF} \} \cap \{ \text{TK\_READ}, \text{TK\_WRITE} \} \cap \{ \text{TK\_SQL}, \text{TK\_CALL} \}$$

$$= \phi$$

Hence, the above five productions for the nonterminal  $\langle \text{stmt} \rangle$  conform to LL(1) specifications.

and ,

$$\text{FIRST}(\langle \text{stmt} \rangle) = \{ \text{TK\_ID}, \text{TK\_RECORDID}, \text{TK\_WHILE}, \text{TK\_IF}, \text{TK\_READ}, \text{TK\_WRITE}, \text{TK\_SQL}, \text{TK\_CALL} \}$$

23.  $\langle \text{assignmentStmt} \rangle ::= \langle \text{SingleOrRecId} \rangle \text{TK\_ASSIGNOP} \langle \text{arithmeticExpression} \rangle \text{TK\_SEM}$

$$\text{FIRST}(\langle \text{assignmentStmt} \rangle) = \text{FIRST}(\langle \text{singleOrRecId} \rangle) = \{ \text{TK\_ID}, \text{TK\_RECORDID} \} \quad \text{.....using 24}$$

24.  $\langle \text{singleOrRecId} \rangle ::= \text{TK\_ID} \mid \text{TK\_RECORDID} \text{TK\_DOT} \text{TK\_FIELDID}$

The token TK\_RECORDID is replaced with TK\_ID as the identifier of record type expands with dot followed by the name of the field and not the record identifier name. You should left factor the new grammar rule to make it LL(1) compatible.

$$\langle \text{singleOrRecId} \rangle ::= \text{TK\_ID} \mid \text{TK\_ID} \text{TK\_DOT} \text{TK\_FIELDID}$$

Both of the above productions for the nonterminal  $\langle \text{singleOrRecId} \rangle$  are non nullable and their first sets are not disjoint. This requires left factoring

The above two rules are left factored using new non terminal  $\langle \text{new\_24} \rangle$

$$\langle \text{singleOrRecId} \rangle ::= \text{TK\_ID} \langle \text{new\_24} \rangle$$

$$\langle \text{new\_24} \rangle ::= \text{eps} \mid \text{TK\_DOT} \text{TK\_FIELDID}$$

Therefore, the above productions are LL(1) compatible.

Also  $FIRST(<singleOrRecId>) = \{TK\_ID\}$

25.  $<funCallStmt> \Rightarrow <outputParameters> \quad TK\_CALL \quad TK\_FUNID \quad TK\_WITH \quad TK\_PARAMETERS \quad <inputParameters>$

The rule must have at its end  $TK\_SEM$  to represent semicolon

$FIRST(<funCallStmt>) = FIRST(<outputParameters>) \cup FOLLOW(<outputParameters>)$   
 $= \{TK\_SQL\} \cup \{TK\_CALL\} = \{TK\_SQL, TK\_CALL\}$

The above rule conforms to LL(1) specifications.

26.  $<outputParameters> \Rightarrow TK\_SQL \quad <idList> \quad TK\_SQR \quad TK\_ASSIGNOP \mid \epsilon$

$FOLLOW(<outputParameters>) = \{TK\_CALL\}$

Due to the presence of the nullable production for the non terminal  $<outputParameters>$ , we check the following LL(1) property

$FIRST(TK\_SQL \quad <idList> \quad TK\_SQR \quad TK\_ASSIGNOP) \cap FOLLOW(<outputParameters>)$   
 $= \{TK\_SQL\} \cap \{TK\_CALL\} = \emptyset$

Hence, the productions in 26 are LL(1) compatible.

$FIRST(<outputParameters>) = \{TK\_SQL\}$

27.  $<inputParameters> \Rightarrow TK\_SQL \quad <idList> \quad TK\_SQR$

LL(1) compatible with  $FIRST(<inputParameters>) = \{TK\_SQL\}$

28.  $<iterativeStmt> \Rightarrow TK\_WHILE \quad TK\_OP \quad <booleanExpression> \quad TK\_CL \quad <stmt> <otherStmts> \quad TK\_ENDWHILE$

LL(1) compatible with  $FIRST(<iterativeStmt>) = \{TK\_WHILE\}$

29.  $<conditionalStmt> \Rightarrow TK\_IF \quad <booleanExpression> \quad TK\_THEN \quad <stmt> <otherStmts> \quad TK\_ELSE \quad <otherStmts> \quad TK\_ENDIF$

30.  $<conditionalStmt> \Rightarrow TK\_IF \quad <booleanExpression> \quad TK\_THEN \quad <stmt> <otherStmts> \quad TK\_ENDIF$



The two rules (29) and (30) for the nonterminal  $\langle \text{conditionalStmt} \rangle$  need modifications as they originally form ambiguous grammar. We need to perform the left factoring as follows

$\langle \text{conditionalStmt} \rangle \Rightarrow \text{TK\_IF TK\_OP} \langle \text{booleanExpression} \rangle \text{TK\_CL TK\_THEN} \langle \text{stmt} \rangle \langle \text{otherStmts} \rangle \langle \text{elsePart} \rangle$

.....number rule as 29

$\langle \text{elsePart} \rangle \Rightarrow \text{TK\_ELSE} \langle \text{stmt} \rangle \langle \text{otherStmts} \rangle \text{TK\_ENDIF} \mid \text{TK\_ENDIF}$

.....number rules as 30

Notice that I also added TK\_OP and TK\_CL in new rule 29 to enclose boolean expression . These two tokens were missing earlier.

Also I am putting a constraint of atleast one statement within else part.

Now there is a single rule for  $\langle \text{conditionalStmt} \rangle$  which is non nullable and its first set is {TK\_IF}

But the newly introduced nonterminal  $\langle \text{elsePart} \rangle$  has two non nullable productions and their corresponding FIRST sets are disjoint.

$\text{FIRST}(\text{TK\_ELSE} \langle \text{otherStmts} \rangle \text{TK\_ENDIF}) \cap \text{FIRST}(\text{TK\_ENDIF}) = \{\text{TK\_ELSE}\} \cap \{\text{TK\_ENDIF}\} = \phi$

Therefore the new set of rules for  $\langle \text{elsePart} \rangle$  is also LL(1) compatible.

And,

$\text{FIRST}(\langle \text{conditionalStmt} \rangle) = \{\text{TK\_IF}\}$

$\text{FIRST}(\langle \text{elsePart} \rangle) = \{\text{TK\_ELSE}, \text{TK\_ENDIF}\}$

31.  $\langle \text{ioStmt} \rangle \Rightarrow \text{TK\_READ TK\_OP} \langle \text{allVar} \rangle \text{TK\_CL TK\_SEM} \mid \text{TK\_WRITE TK\_OP} \langle \text{allVar} \rangle \text{TK\_CL TK\_SEM}$

This rule needs a minor modification. The identifier and record identifier are the only arguments of the read operation.

$\langle \text{ioStmt} \rangle \Rightarrow \text{TK\_READ TK\_OP} \langle \text{singleOrRecId} \rangle \text{TK\_CL TK\_SEM}$

Also the write statement is not expected to take as argument a record id. Any number, real number, identifier or constructed record identifier can be the arguments of write statement.

$\langle \text{ioStmt} \rangle \Rightarrow \text{TK\_WRITE TK\_OP} \langle \text{allVar} \rangle \text{TK\_CL TK\_SEM}$

Both the above productions are non nullable.

and,

$\text{FIRST}(\text{TK\_READ TK\_OP } \langle \text{singleOrRecId} \rangle \text{ TK\_CL TK\_SEM}) \cap \text{FIRST}(\text{TK\_WRITE TK\_OP } \langle \text{allVar} \rangle \text{ TK\_CL TK\_SEM})$   
 $= \{\text{TK\_READ}\} \cap \{\text{TK\_WRITE}\} = \emptyset$

Therefore, the above productions are LL(1) compatible.

and,

$\text{FIRST}(\langle \text{ioStmt} \rangle) = \{\text{TK\_READ}, \text{TK\_WRITE}\}$

Notice that the arguments for a write may be a number, real number, id, or a constructed record id variable. A record identifier is not the argument of WRITE directly. This is relaxed here to simplify the code generation equivalent of printing all field values of the record type variable.

32.  $\langle \text{allVar} \rangle ::= \langle \text{var} \rangle \mid \text{TK\_RECORDID TK\_DOT TK\_FIELDID}$

Trivially, the two non nullable productions above conform to LL(1) and

$\text{FIRST}(\langle \text{allVar} \rangle) = \text{FIRST}(\langle \text{var} \rangle) \cup \{\text{TK\_RECORDID}\}$   
 $= \{\text{TK\_NUM}, \text{TK\_RNUM}, \text{TK\_ID}, \text{TK\_RECORDID}\}$

33.  $\langle \text{arithmeticExpression} \rangle ::= \langle \text{arithmeticExpression} \rangle \langle \text{operator} \rangle \langle \text{arithmeticExpression} \rangle$

I had left a major correction for you to do in the arithmetic expression grammar. You were required to impose precedence of operators and the operations on record variables.

34.  $\langle \text{arithmeticExpression} \rangle ::= \text{TK\_OP } \langle \text{arithmeticExpression} \rangle \text{ TK\_CL} \mid \langle \text{var} \rangle$

A replacement of  $\langle \text{var} \rangle$  with a new nonterminal  $\langle \text{all} \rangle$  is required to incorporate the record addition and subtraction

$\langle \text{arithmeticExpression} \rangle ::= \text{TK\_OP } \langle \text{arithmeticExpression} \rangle \text{ TK\_CL} \mid \langle \text{all} \rangle$

where  $\langle \text{all} \rangle ::= \text{TK\_ID} \mid \text{TK\_NUM} \mid \text{TK\_RNUM} \mid \text{TK\_RECORDID } \langle \text{temp} \rangle$

and  $\langle \text{temp} \rangle ::= \text{eps} \mid \text{TK\_DOT TK\_FIELD}$

An arithmetic expression is expected to take as argument an identifier, statically available number, real number, a record identifier and a record id (constructed by expanding the name using dot with a field name)

Now we reformulate the rules 33, 34(a) and 35 to impose precedence.

$\langle \text{arithmeticExpression} \rangle \Rightarrow \langle \text{arithmeticExpression} \rangle \langle \text{lowPrecedenceOperators} \rangle \langle \text{term} \rangle \mid \langle \text{term} \rangle$  .....(A1)  
 $\langle \text{term} \rangle \Rightarrow \langle \text{term} \rangle \langle \text{highPrecedenceOperators} \rangle \langle \text{factor} \rangle \mid \langle \text{factor} \rangle$  .....(A2)  
 $\langle \text{factor} \rangle \Rightarrow \text{TK\_OP} \langle \text{arithmeticExpression} \rangle \text{TK\_CL} \mid \langle \text{all} \rangle$  .....(A3)  
 $\langle \text{highPrecedenceOperator} \rangle \Rightarrow \text{TK\_MUL} \mid \text{TK\_DIV}$  .....(A4)  
 $\langle \text{lowPrecedenceOperators} \rangle \Rightarrow \text{TK\_PLUS} \mid \text{TK\_MINUS}$  .....(A5)  
 Also we have  
 $\langle \text{all} \rangle \Rightarrow \text{TK\_ID} \mid \text{TK\_NUM} \mid \text{TK\_RNUM} \mid \text{TK\_RECORDID} \langle \text{temp} \rangle$  .....(A6)  
 and  $\langle \text{temp} \rangle \Rightarrow \text{eps} \mid \text{TK\_DOT} \text{TK\_FIELD}$  .....(A7)

[Note : I have given new numbers for referring to the rules, but you can have your own numbering scheme for the rules]

Now let us verify the rules A1-7 for the nonterminals  $\langle \text{arithmeticExpression} \rangle$ ,  $\langle \text{term} \rangle$ ,  $\langle \text{factor} \rangle$ ,  $\langle \text{lowPrecedenceOperators} \rangle$ ,  $\langle \text{highPrecedenceOperators} \rangle$ ,  $\langle \text{all} \rangle$  and  $\langle \text{temp} \rangle$

The rule A1 needs left recursion elimination and the new set of productions are obtained by introducing new nonterminal, say  $\langle \text{expPrime} \rangle$

$\langle \text{arithmeticExpression} \rangle \Rightarrow \langle \text{term} \rangle \langle \text{expPrime} \rangle$   
 $\langle \text{expPrime} \rangle \Rightarrow \langle \text{lowPrecedenceOperators} \rangle \langle \text{term} \rangle \langle \text{expPrime} \rangle \mid \text{eps}$

Similarly rule A2 also needs left recursion elimination. The new set of productions are obtained by introducing new nonterminal, say,  $\langle \text{termPrime} \rangle$

$\langle \text{term} \rangle \Rightarrow \langle \text{factor} \rangle \langle \text{termPrime} \rangle$   
 $\langle \text{termPrime} \rangle \Rightarrow \langle \text{highPrecedenceOperators} \rangle \langle \text{factor} \rangle \langle \text{termPrime} \rangle \mid \text{eps}$

Now we have two more nonterminals  $\langle \text{expPrime} \rangle$  and  $\langle \text{termPrime} \rangle$ . The complete set of non left recursive, unambiguous arithmetic expression grammar then becomes with new numbers A1-A9.

- A1)  $\langle \text{arithmeticExpression} \rangle \Rightarrow \langle \text{term} \rangle \langle \text{expPrime} \rangle$
- A2)  $\langle \text{expPrime} \rangle \Rightarrow \langle \text{lowPrecedenceOperators} \rangle \langle \text{term} \rangle \langle \text{expPrime} \rangle \mid \text{eps}$
- A3)  $\langle \text{term} \rangle \Rightarrow \langle \text{factor} \rangle \langle \text{termPrime} \rangle$
- A4)  $\langle \text{termPrime} \rangle \Rightarrow \langle \text{highPrecedenceOperators} \rangle \langle \text{factor} \rangle \langle \text{termPrime} \rangle \mid \text{eps}$

- A5)  $\langle \text{factor} \rangle \Rightarrow \text{TK\_OP} \langle \text{arithmeticExpression} \rangle \text{TK\_CL} \mid \langle \text{all} \rangle$   
 A6)  $\langle \text{highPrecedenceOperator} \rangle \Rightarrow \text{TK\_MUL} \mid \text{TK\_DIV}$   
 A7)  $\langle \text{lowPrecedenceOperators} \rangle \Rightarrow \text{TK\_PLUS} \mid \text{TK\_MINUS}$   
 A8)  $\langle \text{all} \rangle \Rightarrow \text{TK\_ID} \mid \text{TK\_NUM} \mid \text{TK\_RNUM} \mid \text{TK\_RECORDID} \mid \langle \text{temp} \rangle$   
 A9)  $\langle \text{temp} \rangle \Rightarrow \text{eps} \mid \text{TK\_DOT} \text{TK\_FIELDID}$

Analysis of A1: As there is only single rule available for the nonterminal  $\langle \text{arithmeticExpression} \rangle$ , we must see that is not nullable  
 $\text{FIRST}(\langle \text{arithmeticExpression} \rangle) = \text{FIRST}(\langle \text{term} \rangle) = \text{FIRST}(\langle \text{factor} \rangle) = \{\text{TK\_ID}, \text{TK\_NUM}, \text{TK\_RNUM}, \text{TK\_RECORDID}, \text{TK\_OP}\}$   
 There is no need to go further to check LL(1) compatibility. The grammar for  $\langle \text{arithmeticExpression} \rangle$  is LL(1) compatible.

Analysis of A2: One of the productions for the nonterminal  $\langle \text{expPrime} \rangle$  is nullable.

Let us verify  $\text{FIRST}(\langle \text{lowPrecedenceOperators} \rangle \langle \text{term} \rangle \langle \text{expPrime} \rangle)$  and  $\text{FOLLOW}(\langle \text{expPrime} \rangle)$  are disjoint.

$$\begin{aligned} \text{FIRST}(\langle \text{lowPrecedenceOperators} \rangle \langle \text{term} \rangle \langle \text{expPrime} \rangle) &= \text{FIRST}(\langle \text{lowPrecedenceOperators} \rangle) \\ &= \{\text{TK\_PLUS}, \text{TK\_MINUS}\} \end{aligned} \quad \text{.....using analysis of A7}$$

and,

$$\begin{aligned} \text{FOLLOW}(\langle \text{expPrime} \rangle) &= \text{FOLLOW}(\langle \text{arithmeticExpression} \rangle) \quad \text{.....using A1} \\ &= \{\text{TK\_SEM}\} \quad \text{.....using 23} \end{aligned}$$

This implies that

$$\begin{aligned} &\text{FIRST}(\langle \text{lowPrecedenceOperators} \rangle \langle \text{term} \rangle \langle \text{expPrime} \rangle) \cap \text{FOLLOW}(\langle \text{expPrime} \rangle) \\ &= \{\text{TK\_PLUS}, \text{TK\_MINUS}\} \cap \{\text{TK\_SEM}\} \\ &= \phi \end{aligned}$$

Therefore, A2 is also LL(1) compatible.

Analysis of A3:

As there is only single rule available for the nonterminal  $\langle \text{term} \rangle$ , we must see that is not nullable

$$\text{FIRST}(\langle \text{term} \rangle)$$

= FIRST(<factor> )

= {TK\_ID, TK\_NUM, TK\_RNUM, TK\_RECORDIID, TK\_OP}

.....using analysis of A5

There is no need to go further to check LL(1) compatibility. The grammar for <term> is LL(1) compatible.

#### Analysis of A4:

One of the productions for the nonterminal <termPrime> is nullable.

Let us verify FIRST(<highPrecedenceOperators><factor> <termPrime>) and FOLLOW(<termPrime>) are disjoint.

FIRST(<highPrecedenceOperators><factor> <termPrime>) = FIRST(<highPrecedenceOperators>)

= { TK\_MUL, TK\_DIV }

.....using analysis of A6

And,

FOLLOW(<termPrime>)

= FOLLOW(<term>)

.....using A3

= FIRST(<expPrime>)  $\cup$  FOLLOW(<expPrime>)

.....using A2

= FIRST(<lowPrecedenceOperators>)  $\cup$  FOLLOW(<expPrime>)

.....using A2

= {TK\_PLUS, TK\_MINUS}  $\cup$  FOLLOW(<expPrime>)

.....using analysis of A7

= {TK\_PLUS, TK\_MINUS}  $\cup$  FOLLOW(<arithmeticExpression>)

.....using A1

= {TK\_PLUS, TK\_MINUS}  $\cup$  {TK\_SEM}

.....using 23

= {TK\_PLUS, TK\_MINUS, TK\_SEM}

Therefore,

FIRST(<highPrecedenceOperators><factor> <termPrime>)  $\cap$  FOLLOW(<termPrime>)

= { TK\_MUL, TK\_DIV }  $\cap$  {TK\_PLUS, TK\_MINUS, TK\_SEM}

=  $\phi$

Hence the rules defined by A4 conform to LL(1) specifications.

Analysis of A5: There are two productions for the non terminal <factor>.

$\text{FIRST}(\text{TK\_OP} \langle \text{arithmeticExpression} \rangle \text{TK\_CL}) = \{\text{TK\_OP}\}$

and,

$\text{FIRST}(\langle \text{all} \rangle) = \{\text{TK\_ID}, \text{TK\_NUM}, \text{TK\_RNUM}, \text{TK\_RECORDIID}\}$

..... using analysis of A8

It is now obvious that no production for <factor> is a nullable production and,

$\text{FIRST}(\text{TK\_OP} \langle \text{arithmeticExpression} \rangle \text{TK\_CL}) \cap \text{FIRST}(\langle \text{all} \rangle) = \phi$

Therefore the grammar for <factor> is LL(1) compatible.

and,

$\text{FIRST}(\langle \text{factor} \rangle) = \{\text{TK\_ID}, \text{TK\_NUM}, \text{TK\_RNUM}, \text{TK\_RECORDIID}, \text{TK\_OP}\}$

Analysis of A6: The two productions of the nonterminal <highPrecedenceOperators> are trivially LL(1) compatible

$\text{FIRST}(\langle \text{highPrecedenceOperators} \rangle) = \{\text{TK\_MUL}, \text{TK\_DIV}\}$

Analysis of A7: The two productions of the nonterminal <lowPrecedenceOperators> are also trivially LL(1) compatible

$\text{FIRST}(\langle \text{lowPrecedenceOperators} \rangle) = \{\text{TK\_PLUS}, \text{TK\_MINUS}\}$

Analysis of A8: The two productions of the nonterminal <all> are also trivially LL(1) compatible.

$\text{FIRST}(\langle \text{all} \rangle) = \{\text{TK\_NUM}, \text{TK\_RNUM}, \text{TK\_ID}, \text{TK\_RECORDID}\}$

Analysis of A9: There is an epsilon production for the nonterminal <temp> and it is essential to check that

$\text{FIRST}(\text{TK\_DOT} \text{TK\_FIELDID}) \cap \text{FOLLOW}(\langle \text{temp} \rangle)$  is empty.

Now computing  $\text{FOLLOW}(\langle \text{temp} \rangle)$ ,

$\text{FOLLOW}(\langle \text{temp} \rangle)$

$= \text{FOLLOW}(\langle \text{all} \rangle)$

.....using A8

$= \text{FOLLOW}(\langle \text{factor} \rangle)$

.....using A5

$= \text{FIRST}(\langle \text{termPrime} \rangle) \cup \text{FOLLOW}(\langle \text{termPrime} \rangle)$

.....using A4

$= \text{FIRST}(\langle \text{highPrecedenceOperators} \rangle) \cup \text{FOLLOW}(\langle \text{termPrime} \rangle)$

.....using A4

$= \{\text{TK\_MUL}, \text{TK\_DIV}\} \cup \text{FOLLOW}(\langle \text{termPrime} \rangle)$

.....using analysis of A6

$$\begin{aligned}
 &= \{TK\_MUL, TK\_DIV\} \cup FOLLOW(<term>) && \text{.....using A3} \\
 &= \{TK\_MUL, TK\_DIV\} \cup FIRST(<expPrime>) \cup FOLLOW(<expPrime>) && \text{.....using A2} \\
 &= \{TK\_MUL, TK\_DIV\} \cup FIRST(<lowPrecedenceOperators>) \cup FOLLOW(<expPrime>) && \text{.....using A2} \\
 &= \{TK\_MUL, TK\_DIV\} \cup \{TK\_PLUS, TK\_MINUS\} \cup FOLLOW(<expPrime>) && \text{.....using analysis of A7} \\
 &= \{TK\_MUL, TK\_DIV\} \cup \{TK\_PLUS, TK\_MINUS\} \cup FOLLOW(<arithmeticExpression>) && \text{.....using A1} \\
 &= \{TK\_MUL, TK\_DIV\} \cup \{TK\_PLUS, TK\_MINUS\} \cup \{TK\_SEM\} && \text{.....using 23} \\
 &= \{TK\_MUL, TK\_DIV, TK\_PLUS, TK\_MINUS, TK\_SEM\}
 \end{aligned}$$

Therefore,

$$\begin{aligned}
 &FIRST(TK\_DOT TK\_FIELDID) \cap FOLLOW(<temp>) \\
 &= \{TK\_DOT\} \cap \{TK\_MUL, TK\_DIV, TK\_PLUS, TK\_MINUS, TK\_SEM\} \\
 &= \phi
 \end{aligned}$$

This proves that the rules defined in A9 are LL(1) compatible. [I again insist that I am avoiding proving every rule to be unambiguous. This is left to be proved by you.]

35.  $\langle \text{operator} \rangle \implies TK\_PLUS \mid TK\_MUL \mid TK\_MINUS \mid TK\_DIV$

This rule is discarded now.

36.  $\langle \text{booleanExpression} \rangle \implies TK\_OP \langle \text{booleanExpression} \rangle TK\_CL \langle \text{logicalOp} \rangle TK\_OP \langle \text{booleanExpression} \rangle TK\_CL$

$FIRST(TK\_OP \langle \text{booleanExpression} \rangle TK\_CL \langle \text{logicalOp} \rangle TK\_OP \langle \text{booleanExpression} \rangle TK\_CL) = \{TK\_OP\}$

and the production is not nullable.

37.  $\langle \text{booleanExpression} \rangle \implies \langle \text{var} \rangle \langle \text{relationalOp} \rangle \langle \text{var} \rangle$

$FIRST(\langle \text{var} \rangle \langle \text{relationalOp} \rangle \langle \text{var} \rangle) = FIRST(\langle \text{var} \rangle)$

$= \{TK\_ID, TK\_NUM, TK\_RNUM\}$

and the production is not nullable.

38.  $\langle \text{booleanExpression} \rangle \Rightarrow \text{TK\_NOT } \langle \text{booleanExpression} \rangle$

Considering the above rules 36-38 for analysis, we observe that none of the three productions for the nonterminal  $\langle \text{booleanExpression} \rangle$  is nullable (using First set of  $\langle \text{var} \rangle$  from 39)

and, The first sets of the right hand sides of the three productions are disjoint.

$$\begin{aligned} & \text{FIRST}(\text{TK\_OP } \langle \text{booleanExpression} \rangle \text{ TK\_CL } \langle \text{logicalOp} \rangle \text{ TK\_OP } \langle \text{booleanExpression} \rangle \text{ TK\_CL}) \\ & \cap \text{FIRST}(\langle \text{var} \rangle \langle \text{relationalOp} \rangle \langle \text{var} \rangle) \\ & \cap \text{FIRST}(\text{TK\_NOT } \langle \text{booleanExpression} \rangle) \\ & = \{\text{TK\_OP}\} \cap \{\text{TK\_ID}, \text{TK\_NUM}, \text{TK\_RNUM}\} \cap \{\text{TK\_NOT}\} \\ & = \emptyset \end{aligned}$$

Hence, the three productions for the nonterminal  $\langle \text{booleanExpression} \rangle$  are LL(1) compatible.

39.  $\langle \text{var} \rangle \Rightarrow \text{TK\_ID} \mid \text{TK\_NUM} \mid \text{TK\_RNUM}$

Trivially the three non nullable productions above for the non terminal  $\langle \text{var} \rangle$  are LL(1) compatible.

and,

$$\text{FIRST}(\langle \text{var} \rangle) = \{\text{TK\_ID}, \text{TK\_NUM}, \text{TK\_RNUM}\}$$

40.  $\langle \text{logicalOp} \rangle \Rightarrow \text{TK\_AND} \mid \text{TK\_OR}$

I have not defined the precedence of AND over OR or vice versa. We have the support to enclose the boolean expression in parentheses.

Trivially the productions for  $\langle \text{logicalOp} \rangle$  conform to LL(1) specifications

41.  $\langle \text{relationalOp} \rangle \Rightarrow \text{TK\_LT} \mid \text{TK\_LE} \mid \text{TK\_EQ} \mid \text{TK\_GT} \mid \text{TK\_GE} \mid \text{TK\_NE}$

All six productions above for the nonterminal  $\langle \text{relationalOp} \rangle$  are trivially LL(1) compatible and

$$\text{FIRST}(\langle \text{relationalOp} \rangle) = \{\text{TK\_LT}, \text{TK\_LE}, \text{TK\_EQ}, \text{TK\_GT}, \text{TK\_GE}, \text{TK\_NE}\}$$

42.  $\langle \text{returnStmt} \rangle \Rightarrow \text{TK\_RETURN } \langle \text{optionalReturn} \rangle \text{ TK\_SEM}$

$$\text{FIRST}(\langle \text{returnStmt} \rangle) = \{\text{TK\_RETURN}\}$$

and the production is not nullable, hence the rule is LL(1) compatible.

43.  $\langle \text{optionalReturn} \rangle \Rightarrow \text{TK\_SQL } \langle \text{idList} \rangle \text{ TK\_SQR} \mid \text{eps}$



$\text{FIRST}(\text{TK\_SQL } \langle \text{idList} \rangle \text{ TK\_SQR}) = \{ \text{TK\_SQL} \}$

and,

$\text{FOLLOW}(\langle \text{optionalReturn} \rangle) = \{ \text{TK\_SEM} \}$  .....using 42

Since both of the above sets are disjoint, the grammar for the nonterminal  $\langle \text{optionalReturn} \rangle$  is LL(1) compatible.

44.  $\langle \text{idList} \rangle \Rightarrow \text{TK\_ID } \langle \text{more\_ids} \rangle$

$\text{FIRST}(\langle \text{idList} \rangle) = \{ \text{TK\_ID} \}$  and the production is not nullable, hence the rule is LL(1) compatible.

45.  $\langle \text{more\_ids} \rangle \Rightarrow \text{TK\_COMMA } \langle \text{idList} \rangle \mid \text{eps}$

$\text{FIRST}(\text{TK\_COMMA } \langle \text{idList} \rangle) = \{ \text{TK\_COMMA} \}$

and,

$\text{FOLLOW}(\langle \text{more\_ids} \rangle) = \text{FOLLOW}(\langle \text{idList} \rangle)$  .....using 44

$= \{ \text{TK\_SQR} \}$  .....using 43

This implies that

$\text{FIRST}(\text{TK\_COMMA } \langle \text{idList} \rangle) \cap \text{FOLLOW}(\langle \text{more\_ids} \rangle)$

$= \{ \text{TK\_COMMA} \} \cap \{ \text{TK\_SQR} \}$

$= \phi$

Therefore, the rules for the nonterminal  $\langle \text{more\_ids} \rangle$  are LL(1) conformable.

*Vandana*