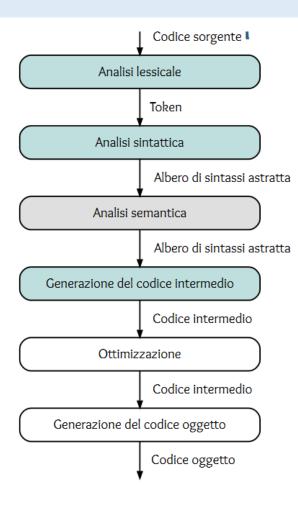
## Laboratorio di Linguaggi Formali e Traduttori LFT lab T2, a.a. 2022/2023

Analisi sintattica

## Analisi sintattica



#### Analisi sintattica:

- Fase successiva a quella dell'analisi lessicale.
- Input: la sequenza di token (l'output dell'analisi lessicale) che corrisponde al programma in input.
- Se il programma in input corrisponde alla grammatica del linguaggio: costruisce un albero di sintattico/produce una derivazione.
- Se il programma in input *non* corrisponde alla grammatica del linguaggio: fa output di un messaggio di errore.

#### Questa lezione:

 Realizzare un parser (riconoscitore) a discesa ricorsiva per una grammatiche LL(1) per la generazione di espressioni aritmetiche semplici.

- Una procedura associata ad ogni variabile.
- Procedura di una variabile: codice derivato dalle produzioni associate con la variabile.
- Dato un input da analizzare, l'albero di chiamate delle procedure ricorsive corrisponde all'albero sintattico.

```
\langle expr \rangle EOF
  \langle start \rangle
                                    \langle term \rangle \langle exprp \rangle
\langle exprp \rangle ::= + \langle term \rangle \langle exprp \rangle
                                     -\langle term \rangle \langle exprp \rangle
  \langle term \rangle ::= \langle fact \rangle \langle termp \rangle
\langle termp \rangle ::=
                                    \star \langle fact \rangle \langle termp \rangle
                                     / \langle fact \rangle \langle termp \rangle
                                      (\langle expr \rangle) \mid NUM
    \langle fact \rangle
```

```
public void start() {
    // ... completare ...
    expr();
    match (Tag.EOF);
    // ... completare ...
private void expr() {
    // ... completare ...
private void exprp() {
    switch (look.tag) {
    case '+':
    // ... completare ...
private void term() {
    // ... completare ...
private void termp()
    // ... completare ...
private void fact() {
    // ... completare ...
```

- Una procedura associata ad ogni variabile.
- Procedura di una variabile: codice derivato dalle produzioni associate con la variabile.
- Dato un input da analizzare, l'albero di chiamate delle procedure ricorsive corrisponde all'albero sintattico.

```
\langle expr \rangle EOF
    |start\rangle
                                    \langle term \rangle \langle exprp \rangle
\langle exprp \rangle ::=
                                 +\langle term \rangle \langle exprp \rangle
                                     -\langle term \rangle \langle exprp \rangle
  \langle term \rangle ::= \langle fact \rangle \langle termp \rangle
\langle termp \rangle ::=
                                    \star \langle fact \rangle \langle termp \rangle
                                     / \langle fact \rangle \langle termp \rangle
                                      (\langle expr \rangle) \mid NUM
    \langle fact \rangle
```

```
public void start() {
    // ... completare ...
    expr();
   match (Tag.EOF);
    // ... completare ...
private void expr() {
    // ... completare ...
private void exprp() {
    switch (look.tag) {
    case '+':
    // ... completare ...
private void term() {
    // ... completare ...
private void termp()
    // ... completare ...
```

private void fact() {

// ... completare ...

- Una procedura associata ad ogni variabile.
- Procedura di una variabile: codice derivato dalle produzioni associate con la variabile.
- Dato un input da analizzare, l'albero di chiamate delle procedure ricorsive corrisponde all'albero sintattico.

```
\langle expr \rangle EO\overline{F}
   \langle start \rangle
                                        \langle term \rangle \langle exprp \rangle
    \langle expr \rangle
\langle exprp \rangle ::=
                                       + \langle term \rangle \langle exprp \rangle
                                        -\langle term \rangle \langle exprp \rangle
                                     \langle fact \rangle \langle termp \rangle
  \langle term \rangle ::=
\langle termp \rangle ::=
                                       \star \langle fact \rangle \langle termp \rangle
                                        /\langle fact \rangle \langle termp \rangle
                                        (\langle expr \rangle) \mid NUM
     \langle fact \rangle
```

```
public void start() {
    // ... completare ...
    expr();
    match (Tag.EOF);
    // ... completare ...
private void expr() {
    // ... completare ...
private void exprp() {
    switch (look.tag) {
    case '+':
    // ... completare ...
private void term() {
    // ... completare ...
private void termp()
    // ... completare ...
private void fact() {
    // ... completare ...
```

- Una procedura associata ad ogni variabile.
- Procedura di una variabile: codice derivato dalle produzioni associate con la variabile.
- Dato un input da analizzare, l'albero di chiamate delle procedure ricorsive corrisponde all'albero sintattico.

```
\langle expr \rangle EOF
   \langle start \rangle
                                       \langle term \rangle \langle exprp \rangle
                                       + \langle term \rangle \langle exprp \rangle
 \langle exprp \rangle
                                       -\langle term \rangle \langle exprp \rangle
                                      \langle fact \rangle \langle termp \rangle
  \langle term \rangle ::=
\langle termp \rangle ::=
                                       \star \langle fact \rangle \langle termp \rangle
                                       /\langle fact \rangle \langle termp \rangle
                                        (\langle expr \rangle) \mid NUM
     \langle fact \rangle
```

```
public void start() {
    // ... completare ...
    expr();
    match (Tag.EOF);
    // ... completare ...
private void expr() {
    // ... completare ...
private void exprp() {
    switch (look.tag) {
    case '+':
    // ... completare ...
private void term() {
    // ... completare ...
private void termp()
    // ... completare ...
private void fact() {
    // ... completare ...
```

- Una procedura associata ad ogni variabile.
- Procedura di una variabile: codice derivato dalle produzioni associate con la variabile.
- Dato un input da analizzare, l'albero di chiamate delle procedure ricorsive corrisponde all'albero sintattico.

```
\langle expr \rangle EOF
   \langle start \rangle
                                      \langle term \rangle \langle exprp \rangle
\langle exprp \rangle ::=
                                      + \langle term \rangle \langle exprp \rangle
                                       -\langle term \rangle \langle exprp \rangle
                                       \langle fact \rangle \langle termp \rangle
   \langle term \rangle
\langle termp \rangle ::=
                                       \star \langle fact \rangle \langle termp \rangle
                                       /\langle fact \rangle \langle termp \rangle
     \langle fact \rangle
                                        (\langle expr \rangle) \mid NUM
```

```
public void start() {
    // ... completare ...
    expr();
   match (Tag.EOF);
    // ... completare ...
private void expr() {
    // ... completare ...
private void exprp() {
    switch (look.tag) {
    case '+':
    // ... completare ...
private void term() {
    // ... completare ...
private void termp()
    // ... completare ...
private void fact() {
    // ... completare ...
```

- Una procedura associata ad ogni variabile.
- Procedura di una variabile: codice derivato dalle produzioni associate con la variabile.
- Dato un input da analizzare, l'albero di chiamate delle procedure ricorsive corrisponde all'albero sintattico.

```
\langle expr \rangle EOF
   \langle start \rangle
                                    \langle term \rangle \langle exprp \rangle
 \langle exprp \rangle ::=
                                  + \langle term \rangle \langle exprp \rangle
                                      -\langle term \rangle \langle exprp \rangle
                                    \langle fact \rangle \langle termp \rangle
  \langle term \rangle ::=
                                      \star \langle fact \rangle \langle termp \rangle
\langle termp \rangle ::=
                                      /\langle fact \rangle \langle termp \rangle
                                      (\langle expr \rangle) \mid NUM
     \langle fact \rangle
```

```
public void start() {
    // ... completare ...
    expr();
   match (Tag.EOF);
    // ... completare ...
private void expr() {
    // ... completare ...
private void exprp() {
    switch (look.tag) {
    case '+':
    // ... completare ...
private void term() {
    // ... completare ...
private void termp() {
    // ... completare ...
private void fact() {
    // ... completare ...
```

- Una procedura associata ad ogni variabile.
- Procedura di una variabile: codice derivato dalle produzioni associate con la variabile.
- Dato un input da analizzare, l'albero di chiamate delle procedure ricorsive corrisponde all'albero sintattico.

```
\langle expr \rangle EOF
  \langle start \rangle
                                     \langle term \rangle \langle exprp \rangle
\langle exprp \rangle ::= + \langle term \rangle \langle exprp \rangle
                                      -\langle term \rangle \langle exprp \rangle
  \langle term \rangle ::= \langle fact \rangle \langle termp \rangle
\langle termp \rangle ::=
                                      \star \langle fact \rangle \langle termp \rangle
                                      /\langle fact \rangle \langle termp \rangle
     \langle fact \rangle
                                          \langle expr \rangle
```

```
public void start() {
    // ... completare ...
    expr();
   match (Tag.EOF);
    // ... completare ...
private void expr() {
    // ... completare ...
private void exprp() {
    switch (look.tag) {
    case '+':
    // ... completare ...
private void term() {
    // ... completare ...
private void termp()
    // ... completare ...
private void fact() {
    // ... completare ...
```

- Esempio: variabile  $\langle start \rangle$ .
- Grammatica:
  - Una singola produzione è associata con \(\start\).
  - La produzione consiste di una variabile  $\langle expr \rangle$ , seguito da un simbolo terminale EOF.

```
\langle start \rangle \quad ::= \quad \langle expr \rangle \; \mathrm{EOF}
```

#### • Codice:

- Una chiamata alla procedura expr...
- ... seguita da un controllo rispetto a EOF.

```
public void start() {
    // ... completare ...
    expr();
    match(Tag.EOF);
    // ... completare ...
}
```

- Esempio: variabile  $\langle start \rangle$ .
- Grammatica:
  - Una singola produzione è associata con \(\start\).
  - La produzione consiste di una variabile (expr), seguito da un simbolo terminale EOF.

```
\langle start \rangle ::= \langle expr \rangle EOF
```

#### • Codice:

- Una chiamata alla procedura expr...
- ... seguita da un controllo rispetto a EOF.

```
public void start() {
    // ... completare ...
expr();
match(Tag.EOF);
    // ... completare ...
}
```

- Esempio: variabile  $\langle start \rangle$ .
- Grammatica:
  - Una singola produzione è associata con \(\start\).
  - La produzione consiste di una variabile  $\langle expr \rangle$ , seguito da un simbolo terminale EOF.

```
\langle start \rangle ::= \langle expr \rangle EOF
```

#### Codice:

- Una chiamata alla procedura expr...
- ... seguita da un controllo rispetto a EOF.

```
public void start() {
    // ... completare ...
    expr();
    match(Tag.EOF);
    // ... completare ...
}
```

 Pseudo-codice per la parsificazione a discesa ricorsiva (slide 4 del file «parsing\_rd.pdf» con titolo «4.2 Parsing ricorsivo discendente»).

```
var w: string
                                                          ||w\rangle è la stringa da riconoscere con \$ in fondo
                                                      ||i\rangle| è l'indice del prossimo simbolo di w da leggere
var i:int
procedure match(a: symbol)
  if w[i] = a then i \leftarrow i + 1 else error
procedure parse(v: string)
                                                                            ||v\rangle è la stringa da riconoscere
  w \leftarrow v$
  i \leftarrow 0
  S()
                                                                \parallel S è il simbolo iniziale della grammatica
  match($)
                                                                   // controlla di aver letto tutta la stringa
procedure A()
                                                         ||A \rightarrow \alpha_1| \cdots |\alpha_n| sono le produzioni per A
  if w[i] \in 	ext{GUIDA}(A 	o lpha_1) then
  else if w[i] \in 	ext{GUIDA}(A 	o lpha_k) then
     for X \in \alpha_k do
        if X è un terminale then match(X) else X()
                                         ||w[i]| non è nell'insieme guida di nessuna produzione per A
  else error
```

```
var w: string
                                                         ||w\rangle è la stringa da riconoscere con \$ in fondo
var i:int
                                                    \mid\mid i è l'indice del prossimo simbolo di w da leggere
procedure match(a:symbol)
  if w[i] = a then i \leftarrow i + 1 else error
procedure parse(v:string)
                                                                          ||v\rangle è la stringa da riconoscere
  w \leftarrow v$
  i \leftarrow 0
                                                               \mid\mid S è il simbolo iniziale della grammatica
  S()
  match($)
                                                                 // controlla di aver letto tutta la stringa
procedure A()
                                                       ||A \rightarrow \alpha_1| \cdots |\alpha_n| sono le produzioni per A
  if w[i] \in 	ext{GUIDA}(A 	o lpha_1) then
  else if w[i] \in 	ext{GUIDA}(A 	o lpha_k) then
    for X \in \alpha_k do
       if X è un terminale then match(X) else X()
                                        ||w[i]| non è nell'insieme guida di nessuna produzione per A
  else error
```

```
void move() {
    look = lex.lexical_scan(pbr);
    System.out.println("token = " + look);
}

void error(String s) {
    throw new Error("near line " + lex.line + ": " + s);
}

void match(int t) {
    if (look.tag == t) {
        if (look.tag != Tag.EOF) move();
    } else error("syntax error");
}
```

```
var w: string
                                                         ||w\rangle è la stringa da riconoscere con \$ in fondo
var i:int
                                                    \mid\mid i è l'indice del prossimo simbolo di w da leggere
procedure match(a:symbol)
  if w[i] = a then i \leftarrow i + 1 else error
procedure parse(v:string)
                                                                          ||v\rangle è la stringa da riconoscere
  w \leftarrow v$
  i \leftarrow 0
  S()
                                                               \mid\mid S è il simbolo iniziale della grammatica
  match($)
                                                                 // controlla di aver letto tutta la stringa
procedure A()
                                                       ||A \rightarrow \alpha_1| \cdots |\alpha_n| sono le produzioni per A
  if w[i] \in 	ext{GUIDA}(A 	o lpha_1) then
  else if w[i] \in 	ext{GUIDA}(A 	o lpha_k) then
    for X \in \alpha_k do
       if X è un terminale then match(X) else X()
                                        ||w[i]| non è nell'insieme guida di nessuna produzione per A
  else error
```

```
void move() {
    look = lex.lexical_scan(pbr);
    System.out.println("token = " + look);
}

void error(String s) {
    throw new Error("near line " + lex.line + ": " + s);
}

void match(int t) {
    if (look.tag == t) {
        if (look.tag != Tag.EOF) move();
    } else error("syntax error");
}
```

Pseudo-codice	Codice
w[i]	look
a	t
i ← i+1	move()
error	error()

```
||w\rangle è la stringa da riconoscere con \$ in fondo
var w: string
var i:int
                                                    \parallel i è l'indice del prossimo simbolo di w da leggere
procedure match(a:symbol)
  if w[i] = a then i \leftarrow i + 1 else error
procedure parse(v:string)
                                                                          ||v\rangle è la stringa da riconoscere
  w \leftarrow v$
  i \leftarrow 0
  S()
                                                               \mid\mid S è il simbolo iniziale della grammatica
  match($)
                                                                 // controlla di aver letto tutta la stringa
procedure A()
                                                       ||A \rightarrow \alpha_1| \cdots |\alpha_n| sono le produzioni per A
  if w[i] \in 	ext{GUIDA}(A 	o lpha_1) then
  else if w[i] \in 	ext{GUIDA}(A 	o lpha_k) then
    for X \in \alpha_k do
       if X è un terminale then match(X) else X()
                                        ||w[i]| non è nell'insieme guida di nessuna produzione per A
  else error
```

```
void move() {
    look = lex.lexical_scan(pbr);
    System.out.println("token = " + look);
}

void error(String s) {
    throw new Error("near line " + lex.line + ": " + s);
}

void match(int t) {
    if (look.tag == t) {
        if (look.tag != Tag.EOF) move();
    } else error("syntax error");
}
```

Pseudo-codice	Codice
w[i]	look
а	t
i ← i+1	move()
error	error()

```
procedure A() ||A \rightarrow \alpha_1| \cdots |\alpha_n sono le produzioni per A if w[i] \in \text{GUIDA}(A \rightarrow \alpha_1) then \vdots else if w[i] \in \text{GUIDA}(A \rightarrow \alpha_k) then for X \in \alpha_k do if X è un terminale then match(X) else X() \vdots else error ||w[i]| non è nell'insieme guida di nessuna produzione per A
```

```
\langle start \rangle ::= \langle expr \rangle EOF
```

```
public void start() {
    // ... completare ...
    expr();
    match(Tag.EOF);
    // ... completare ...
}
```

```
procedure A() ||A \rightarrow \alpha_1| ||A \rightarrow \alpha_1| sono le produzioni per A if w[i] \in \text{GUIDA}(A \rightarrow \alpha_1) then else if w[i] \in \text{GUIDA}(A \rightarrow \alpha_k) then for X \in \alpha_k do if X è un terminale then match(X) else X() else error ||w[i]| non è nell'insieme guida di nessuna produzione per A
```

```
\langle start \rangle \quad ::= \quad \langle expr \rangle \; \mathrm{EOF}
```

```
public void start() {
    // ... completare ...
    expr();
    match(Tag.EOF);
    // ... completare ...
}
```

```
procedure A() ||A \rightarrow \alpha_1| ||A \rightarrow \alpha_1| sono le produzioni per A if w[i] \in \operatorname{GUIDA}(A \rightarrow \alpha_1) then else if w[i] \in \operatorname{GUIDA}(A \rightarrow \alpha_k) then for X \in \alpha_k do if X è un terminale then match(X) else X() else error ||w[i]| non è nell'insieme guida di nessuna produzione per A
```

```
\langle start \rangle \quad ::= \quad \langle expr \rangle \; \mathrm{EOF}
```

```
public void start() {
    // ... completare ...
    expr();
    match(Tag.EOF);
    // ... completare ...
}
```

```
procedure A() ||A \rightarrow \alpha_1| ||A \rightarrow \alpha_1| sono le produzioni per A if w[i] \in \text{GUIDA}(A \rightarrow \alpha_1) then else if w[i] \in \text{GUIDA}(A \rightarrow \alpha_k) then for X \in \alpha_k do if X \in \text{Im}(A) if X \in \text{Im}(A) else X(i) else error ||w[i]| non è nell'insieme guida di nessuna produzione per A
```

```
\langle start \rangle \quad ::= \quad \langle expr \rangle \; \mathrm{EOF}
```

```
public void start() {
    // ... completare ...
expr();
match(Tag.EOF);
    // ... completare ...
}
```

```
procedure A() ||A \rightarrow \alpha_1| ||A \rightarrow \alpha_1| sono le produzioni per A if w[i] \in \text{GUIDA}(A \rightarrow \alpha_1) then else if w[i] \in \text{GUIDA}(A \rightarrow \alpha_k) then for X \in \alpha_k do if X è un terminale then match(X) else X() else error ||w[i]| non è nell'insieme guida di nessuna produzione per A
```

```
\langle start \rangle ::= \langle expr \rangle EOF
```

```
public void start() {
    // ... completare ...
expr();
match(Tag.EOF);
    // ... completare ...
}
```

```
procedure A() ||A \rightarrow \alpha_1| ||A \rightarrow \alpha_1| sono le produzioni per A if w[i] \in \operatorname{GUIDA}(A \rightarrow \alpha_1) then else if w[i] \in \operatorname{GUIDA}(A \rightarrow \alpha_k) then for X \in \alpha_k do if X è un terminale then match(X) else X() else error ||w[i]| non è nell'insieme guida di nessuna produzione per A
```

```
\langle start \rangle ::= \langle expr \rangle EOF
```

```
public void start() {
    // ... completare ...
    expr();
    match(Tag.EOF);
    // ... completare ...
}
```

```
\begin{array}{ll} \operatorname{procedure} A() & || A \to \alpha_1 \mid \cdots \mid \alpha_n \text{ sono le produzioni per } A \\ \text{if } w[i] \in \operatorname{GUIDA}(A \to \alpha_1) \text{ then} \\ \vdots \\ \text{else if } w[i] \in \operatorname{GUIDA}(A \to \alpha_k) \text{ then} \\ \text{ for } X \in \alpha_k \text{ do} \\ \text{ if } X \text{ è un } \underbrace{\operatorname{terminale}}_{\text{then match}} \text{ then match}(X) \text{ else } X() \\ \vdots \\ \text{else error} & || w[i] \text{ non è nell'insieme guida di nessuna produzione per } A \end{array}
```

```
 \begin{array}{ccc} \langle exprp \rangle & ::= & + \langle term \rangle \langle exprp \rangle \\ & | & - \langle term \rangle \langle exprp \rangle \\ & | & \varepsilon \end{array}
```

```
private void exprp() {
    switch (look.tag) {
    case '+':
    // ... completare ...
    }
}
```

```
procedure A() ||A \rightarrow \alpha_1| \alpha_2 \alpha_3 sono le produzioni per A if w[i] \in \operatorname{GUIDA}(A \rightarrow \alpha_1) then \vdots else if w[i] \in \operatorname{GUIDA}(A \rightarrow \alpha_k) then for X \in \alpha_k do if X è un terminale then match(X) else X() \vdots else error ||w[i]| non è nell'insieme guida di nessuna produzione per A
```

```
\begin{array}{cccc} \langle \mathit{exprp} \rangle & ::= & + \langle \mathit{term} \rangle \, \langle \mathit{exprp} \rangle & \alpha_1 \\ & | & - \langle \mathit{term} \rangle \, \langle \mathit{exprp} \rangle & \alpha_2 \\ & | & \varepsilon & \alpha_3 \end{array}
```

```
private void exprp() {
    switch (look.tag) {
    case '+':
    // ... completare ...
    }
}
```

```
procedure A() ||A \rightarrow \alpha_1 | \alpha_2 | \alpha_3 sono le produzioni per A if w[i] \in \text{GUIDA}(A \rightarrow \alpha_1) then \vdots else if w[i] \in \text{GUIDA}(A \rightarrow \alpha_k) then for X \in \alpha_k do if X è un terminale then match(X) else X() \vdots else error ||w[i]| non è nell'insieme guida di nessuna produzione per A
```

```
\begin{array}{cccc} \langle \mathit{exprp} \rangle & ::= & + \langle \mathit{term} \rangle \, \langle \mathit{exprp} \rangle & \alpha_1 \\ & | & - \langle \mathit{term} \rangle \, \langle \mathit{exprp} \rangle & \alpha_2 \\ & | & \varepsilon & \alpha_3 \end{array}
```

```
private void exprp() {
    switch (look.tag) {
    case '+':
    // ... completare ...
    }
}
```

```
procedure A() ||A \rightarrow \alpha_1| \alpha_2 \alpha_3 sono le produzioni per A if w[i] \in \operatorname{GUIDA}(A \rightarrow \alpha_1) then ||A \rightarrow \alpha_1| \alpha_2 \alpha_3 sono le produzioni per A else if w[i] \in \operatorname{GUIDA}(A \rightarrow \alpha_k) then for X \in \alpha_k do if X \in \alpha_k do if X \in A if X \in A
```

```
\begin{array}{cccc} \langle \mathit{exprp} \rangle & ::= & + \langle \mathit{term} \rangle \, \langle \mathit{exprp} \rangle & \alpha_1 \\ & | & - \langle \mathit{term} \rangle \, \langle \mathit{exprp} \rangle & \alpha_2 \\ & | & \varepsilon & \alpha_3 \end{array}
```

# Messaggi di errore

```
void error(String s) {
    throw new Error("near line " + lex.line + ": " + s);
}
```

- Argomento s: utilizzare per dare informazione utile all'utente del programma quando l'input non corrisponde alla grammatica.
- Consiglio: segnalare la procedura che invoca error (ad esempio, error ("Error in term")).
- Tutte le procedure devono avere dei meccanismi per rilevare input erroneo, in modo tale che eventuali errori sono segnalati appena possibile.

Input	Procedura in cui è rilevato l'errore
) 2	start
2+(	expr
5+)	term
1+2(	termp
1*+2	fact
1+2)	<pre>match (in questo caso il messaggio di errore è semplicemente "syntax error")</pre>

- main: simile a parse; entrambe chiamano la procedura della variabile iniziale della grammatica (\(\start\)).
- Si nota che la fine del input è rappresentato esplicitamente con il terminale EOF quindi, in nostro caso, non c'è la necessità di controllare `\$' nel main (è controllato nella procedura start con match (Tag.EOF)).

```
public static void main(String[] args) {
    Lexer lex = new Lexer();
    String path = "...path..."; // il percorso del file da leggere
    try {
        BufferedReader br = new BufferedReader(new FileReader(path));
        Parser parser = new Parser(lex, br);
        parser.start();
        System.out.println("Input OK");
        br.close();
    } catch (IOException e) {e.printStackTrace();}
}
```