

# LA SINTASSI DEI LINGUAGGI DI PROGRAMMAZIONE

Ivan Lanese

# argomenti

- Grammatiche BNF
- Varianti:
  - BNF +  $\epsilon$
  - EBNF

# Remind: cos'è una grammatica

- Una grammatica è uno strumento linguistico per definire insiemi di stringhe infiniti

Es: CIFRA ::= 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9  
NATURALE ::= CIFRA | CIFRA NATURALE

- Ci interessano le grammatiche per definire la sintassi dei linguaggi di programmazione

Es: PROGRAMMA = int main(){ DICHIARAZIONI COMANDI }

- Le grammatiche danno una definizione precisa e inequivocabile della sintassi
- Le grammatiche sono alla base di molti tool relativi ai linguaggi di programmazione, a partire dal compilatore

# Remind: una grammatica, formalmente

- Una tripla  $(N, T, R)$ 
  - $N$ : insieme di simboli non terminali (Es: NATURALE)
  - $T$ : insieme di simboli terminali (Es: 0)
  - $R$ : insieme di regole (Es: NATURALE ::= CIFRA)
- Una regola ha la forma  $N ::= \dots$  dove  $N$  è un non terminale e  $\dots$  una stringa di terminali e non terminali
- Quando non sarà ambiguo scriveremo solo le regole, usando lettere maiuscole per i non terminali e simboli e lettere minuscole per i terminali
- Possiamo unire più produzioni per uno stesso non terminale usando l'operatore di alternativa  $|$

# Cosa ci si attende da voi

- Che impariate a usare le grammatiche, cioè
- Data una grammatica dovreste sapere che linguaggio genera
- Data una descrizione informale di un linguaggio dovreste saperlo formalizzare tramite una grammatica

# BNF – Esercizio 1: Dalla grammatica al linguaggio

- Grammatica:

$N ::= NCIFRA \mid NCARATTERE NCIFRA$

$NCIFRA ::= CIFRA \mid CIFRA NCIFRA$

$NCARATTERE ::= CARATTERE \mid CARATTERE NCARATTERE$

$CIFRA ::= 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9$

$CARATTERE ::= a \mid b \mid c \mid d \mid e \mid f \mid \dots \mid x \mid y \mid z$

- Qual è il linguaggio generato?

il linguaggio delle stringhe che iniziano con un certo numero (anche 0) di caratteri e terminano con delle cifre

a5, fg3, wzsg745, 12345

## BNF – Esercizio 2

- Grammatica:

$N ::= NCIFRA \mid NCARATTERE \ NCIFRA$

$NCIFRA ::= CIFRA \mid CIFRA \ NCIFRA$

$NCARATTERE ::= CARATTERE \mid CARATTERE \ NCARATTERE$

$CIFRA ::= 0 \mid PARI \mid DISPARI$

$PARI ::= 2 \mid 4 \mid 6 \mid 8$

$DISPARI ::= 1 \mid 3 \mid 5 \mid 7 \mid 9$

$CARATTERE ::= a \mid b \mid c \mid d \mid e \mid f \mid \dots \mid x \mid y \mid z$

- Qual è il linguaggio generato?

Lo stesso linguaggio dell'esempio precedente:

uno stesso linguaggio può essere generato da grammatiche diverse

# BNF – Esercizio 3: dal linguaggio alla grammatica

- Linguaggio: il linguaggio delle stringhe composte da un numero dispari di coppie miste (cifra carattere) oppure (carattere cifra) intervallate da “.”

Es:        a4.5b.2c            7s            a1.2b.2c.6r.0e

- Grammatica:

$N ::= \text{COPPIA} \mid \text{COPPIA} . \text{COPPIA} . N$

$\text{COPPIA} ::= \text{CIFRA} \text{ CARATTERE} \mid \text{CARATTERE} \text{ CIFRA}$

$\text{CIFRA} ::= 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9$

$\text{CARATTERE} ::= a \mid b \mid c \mid d \mid e \mid f \mid \dots \mid x \mid y \mid z$



# BNF – Esercizio 4: dal linguaggio alla grammatica

- Linguaggio: il linguaggio delle stringhe palindrome su un alfabeto composto solo da “a” e “b”

a b aa bb aba abba ababa bab

- Grammatica:

$N ::= a \mid b \mid aa \mid bb \mid a N a \mid b N b$

# BNF – Epsilon

- Il simbolo  $\varepsilon$  viene usato per indicare la stringa vuota.

$N ::= \varepsilon \mid a N$

genera il linguaggio delle sequenze anche nulle di  $a$

$\{\varepsilon, a, aa, aaa, \dots\}$

- Come si può riscrivere la grammatica dell'Esempio 1 usando il simbolo  $\varepsilon$  ?
- Come si può riscrivere la grammatica dell'Esempio 4 usando il simbolo  $\varepsilon$  ?

# BNF – Esercizio 5: epsilon

- La grammatica dell'Esempio 1:

$N ::= NCIFRA \mid NCARATTERE NCIFRA$

$NCIFRA ::= CIFRA \mid CIFRA NCIFRA$

$NCARATTERE ::= CARATTERE \mid CARATTERE NCARATTERE$

$CIFRA ::= 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9$

$CARATTERE ::= a \mid b \mid c \mid d \mid e \mid f \mid \dots \mid x \mid y \mid z$

- diventa:

$N ::= NCARATTERE NCIFRA$

$NCIFRA ::= CIFRA \mid CIFRA NCIFRA$

$NCARATTERE ::= \varepsilon \mid CARATTERE NCARATTERE$

$CIFRA ::= 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9$

$CARATTERE ::= a \mid b \mid c \mid d \mid e \mid f \mid \dots \mid x \mid y \mid z$

# BNF – Esercizio 6: epsilon

- La grammatica dell'Esempio 4:

$$N ::= a \mid b \mid aa \mid bb \mid a N a \mid b N b$$

- diventa:

$$N ::= \varepsilon \mid a \mid b \mid a N a \mid b N b$$

- Notare che il linguaggio generato dalla seconda grammatica è leggermente diverso: contiene anche la stringa vuota

# EBNF

- EBNF: Estendono le grammatiche BNF con dei simboli aggiuntivi per la definizioni di parti opzionali o ricorrenti: parentesi tonde, quadre e graffe.
- La sintassi del linguaggio C++ è espressa con grammatiche EBNF

# EBNF – parentesi tonde

- Parentesi tonde ( ) : indicano una sola occorrenza, tipicamente si usano per alternative annidate:

$$N ::= (a|b) c$$

produce stringhe della forma  $ac$  oppure  $bc$

equivale alla grammatica

$$N ::= M c$$
$$M ::= a | b$$

# EBNF – parentesi quadre

- Parentesi quadre [ ] : indicano 0 o 1 occorrenza

$$N ::= [a] c$$

produce stringhe della forma  $c$  oppure  $ac$

equivale alla grammatica  $N ::= M c$

$$M ::= \varepsilon \mid a$$

- Anche nelle parentesi quadre possiamo inserire alternative annidate

$$N ::= [a \mid b] c$$

produce stringhe della forma  $c$  oppure  $ac$  oppure  $bc$

equivale alla grammatica  $N ::= M c$

$$M ::= \varepsilon \mid a \mid b$$

# EBNF – parentesi graffe

- Parentesi graffe  $\{ \}$  : indicano 0, 1 o più occorrenze:

$$N ::= \{a\} c$$

- produce stringhe della forma  $c, ac, aac, aaac \dots$
- equivale alla grammatica  $N ::= M c$

$$M ::= \varepsilon \mid a M$$

- Anche con le parentesi graffe possiamo esprimere alternative annidate:

$$N ::= \{a \mid b\} c$$

- produce stringhe della forma  $c, ac, bc, abc, baC, abac \dots$
- equivale alla grammatica  $N ::= M c$

$$M ::= \varepsilon \mid a M \mid b M$$



# EBNF – Esercizio 7

- Rivediamo la grammatica BNF dell'Esempio 1:

$N ::= NCIFRA \mid NCARATTERE NCIFRA$

$NCIFRA ::= CIFRA \mid CIFRA NCIFRA$

$NCARATTERE ::= CARATTERE \mid CARATTERE NCARATTERE$

$CIFRA ::= 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9$

$CARATTERE ::= a \mid b \mid c \mid d \mid e \mid f \mid \dots \mid x \mid y \mid z$

- Come si può riscrivere la grammatica precedente usando le parentesi?

# EBNF – Esercizio 7

- Ecco la EBNF che genera lo stesso linguaggio dell'Esempio 1:

$N ::= \{ \text{CARATTERE} \} \text{CIFRA} \{ \text{CIFRA} \}$

$\text{CIFRA} ::= 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9$

$\text{CARATTERE} ::= a \mid b \mid c \mid d \mid e \mid f \mid \dots \mid x \mid y \mid z$

- Trasformare in EBNF le grammatiche degli Esempi 3 e 4

# EBNF – Esercizio 8

- Ecco la BNF dell'Esempio 3:

$N ::= \text{COPPIA} \mid \text{COPPIA} . \text{COPPIA} . N$

$\text{COPPIA} ::= \text{CIFRA} \text{ CARATTERE} \mid \text{CARATTERE} \text{ CIFRA}$

$\text{CIFRA} ::= 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9$

$\text{CARATTERE} ::= a \mid b \mid c \mid d \mid e \mid f \mid \dots \mid x \mid y \mid z$

- La relativa EBNF:

$N ::= \text{COPPIA} \{ . \text{COPPIA} . \text{COPPIA} \}$

$\text{COPPIA} ::= \text{CIFRA} \text{ CARATTERE} \mid \text{CARATTERE} \text{ CIFRA}$

$\text{CIFRA} ::= 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9$

$\text{CARATTERE} ::= a \mid b \mid c \mid d \mid e \mid f \mid \dots \mid x \mid y \mid z$

# EBNF – Esercizio 9

- Ecco la BNF dell'Esempio 4:

$N ::= a \mid b \mid aa \mid bb \mid a N a \mid b N b$

- La relativa EBNF:

$N ::= a \mid b \mid a [N] a \mid b [N] b$

# EBNF – Altri esercizi

Quali linguaggi generano le seguenti grammatiche?

- $N ::= \{a \mid b\}$ 
  - Linguaggio generato:  $\{\varepsilon, a, b, ab, ba, aa, bb, aaa, aab, \dots\}$
  - una sequenza di “o a o b”
- $N ::= \{a\} \mid \{b\}$ 
  - Linguaggio generato:  $\{\varepsilon, a, aa, aaa, \dots, b, bb, bbb, \dots\}$
  - o una sequenza di a, o una sequenza di b
- $N ::= \{a\}\{b\}$ 
  - Linguaggio generato:  $\{\varepsilon, a, b, aa, ab, aab, \dots, aaaabbb, \dots\}$
  - una sequenza di a seguita da una sequenza di b
- $N ::= \{a[b]\}$ 
  - Linguaggio generato:  $\{\varepsilon, a, aa, aaa, \dots, ab, aaba, aaabaabaa, \dots\}$
  - una sequenza di “a seguito opzionalmente da un b”
  - non ci possono essere due b di fila né b iniziali