Grammatiche

- Grammatiche libere da contesto
- Grammatiche regolari
- Potenza delle grammatiche libere e regolari
- Struttura di frase: Alberi di derivazione

Esempio dei numeri interi

- Si consideri il linguaggio per dei numeri interi, costituito cioè da tutte le sequenze di simboli che rapresentano numeri interi:
  - es: +38
    -567
    +456
    0

Esempio di Grammatica

- Vediamo come può essere definita una grammatica per i numeri:

<{0,1,2,3,4,5,6,7,8,9,+,-},

{cifra, nat, int} ,

int,

{cifra::=0|1|2|3|4|5|6|7|8|9,

nat ::= cifra cifra*,

int::= (+|-) nat | nat}

>

Attenzione alla notazione

Definizione di Grammatica

- Una grammatica è una quadrupla $<\Lambda, V, s \in V, P>$ dove:

  Λ = insieme finito di **terminali** (alfabeto)

  V = insieme finito di **non terminali** (categorie sintattiche o grammaticali)

  s = simbolo **iniziale** (categoria sintattica del linguaggio)

  $P = \{i ::= e \mid i \in V,\ e \in E_{\Lambda \cup V}\}$ insieme finito di **produzioni.**

  Le produzioni hanno una parte sinistra (i) e una parte destra separate da una freccia ($\rightarrow$ oppure ::=)

- Le grammatiche regolari hanno produzioni in cui a sinistra del ::= c'è un simbolo non terminale a destra un'espressione regolare su $\Lambda \cup V$.

Espressioni regolari

Espressioni regolari su un alfabeto A: E_A

- $\varepsilon \in E_A$ (ε è un'espressione regolare)
- $\forall a \in A,\ a \in E_A$ (a è un'espressione regolare)
- $\forall e, e' \in E_A,\ ee' \in E_A$ (giustapposizione)
- $\forall e, e' \in E_A,\ e \mid e' \in E_A$ (alternativa)
- $\forall e \in E_A,\ [e] \in E_A$ (opzione)
- $\forall e \in E_A,\ e^* \in E_A$ (chiusura)

Significato delle produzioni di una grammatica

- Le produzioni di una grammatica permettono di derivare insiemi di sequenze di simboli terminali, definite dall'espressione a destra $A=\Lambda \cup V$:
  - $\varepsilon \in E_A$ $\quad$ $\mathbf{S}(\varepsilon)= \{\varepsilon\}$
  - $\forall a \in \Lambda,$ $\quad$ $\mathbf{S}(a)=\{a\}$
  - $\forall a \in V,$ $\quad$ $\mathbf{S}(a)=\mathbf{S}(e)\quad a::=e \in P$ $\quad$ Una sola produzione
  - $\forall e, e' \in E_A,\ ee' \in E_A$ $\quad$ $\mathbf{S}(ee')= \mathbf{S}(e)\times \mathbf{S}(e')=$
    $\{ss' | s \in \mathbf{S}(e),\ s' \in \mathbf{S}(e')\}$
  - $\forall e, e' \in E_A,\ e \mid e' \in E_A$ $\quad$ $\mathbf{S}(e \mid e')=\mathbf{S}(e) \cup \mathbf{S}(e')$

Significato delle produzioni di una grammatica -continua

- Le produzioni di una grammatica permettono di derivare insiemi di sequenze di simboli terminali, definite dall'espressione a destra $A = \Lambda \cup V$:
  - $\forall e \in E_A$, $[e] \in E_A$ $\qquad$ $\mathbf{S}([e]) = \{\mathbf{S}(e)\} \cup \{\varepsilon\}$
  - $\forall e \in E_A$, $e^* \in E_A$ $\qquad$ $\mathbf{S}(e^*) = \{\mathbf{S}(e)\}^* = \cup_{n \geq 0} \mathbf{S}(e)^n$

$$\mathbf{S}(e)^0 = \{\varepsilon\},\ \mathbf{S}(e)^1 = \mathbf{S}(e),$$

$$\mathbf{S}(e)^2 = \mathbf{S}(ee)$$

$$\mathbf{S}(e)^n = \mathbf{S}(\underbrace{e \ldots e}_{\text{n volte}})$$

Linguaggio di una grammatica

- Le sequenze che possono essere derivate da s (simbolo iniziale) costituiscono il linguaggio definito dalla grammatica. L(G)= S(s)
- Esempio:
  - G=<{a,b},{A,B,S},S,{A::=a|aa, B::=bb,
    S::=AB |BA

    L(G)=?

L’operatore *

- L’operatore interessante è * (stella):
  - permette di derivare tutte le sequenze comunque lunghe di simboli della base
  - introduce insiemi infiniti

– G=<{a,b},{A,S},S,{A::=aa*, S::=Ab | b*}>

L(G)=?

Osservazioni

- La semantica ($\mathcal{S}$) definisce un metodo effettivo di calcolo del linguaggio.
  - generazione dell'insieme anche infinito
  - appartenenza di una sequenza?
    - quando termino? se termino
- Esercizi:
  - lettura di una grammatica
  - definizione di grammatiche.
  - sottolinguaggi (sottoinsiemi)

Esempio dei numeri

- Si consideri il linguaggio per i numeri, costituito cioè da tutte le sequenze di simboli che rapresentano numeri, sia interi che frazionari, rappresenti in virgola fissa e anche in virgola mobile:
  - es: +38
    -567
    +456.34
    -1239.02
    +0.289E2

Esempio di Grammatica

- Vediamo come può essere definita una grammatica per i numeri:

<{0,1,2,3,4,5,6,7,8,9,+,-,E,.},

{cifra, nat, int,fract,exp,num} ,

num,

{cifra::=0|1|2|3|4|5|6|7|8|9,

nat ::= cifra cifra*,

int::= ([+|-]) nat,

fract::= int.nat,

exp::= fract E int,

num::= nat | fract | exp>

Esempio delle espressioni aritmetiche

- Ci sono molte grammatiche equivalenti per uno stesso linguaggio, ad esempio:

  <{0,1,2,3,4,5,6,7,8,9,+,-,E,.},
  {nat,int fract,exp,num}
  num,
  {nat ::= (0|1|2|3|4|5|6|7|8|9) (0|1|2|3|4|5|6|7|8|9)*,
  int::=(+|-) nat
  fract::= int.nat,
  exp::= fract E int,
  num::= nat | fract | exp>

  È stato eliminato il non terminale cifra con un peggiorameno della leggibilità della grammatica

Osservazioni

- Le grammatiche regolari possono sempre essere riscritte con un unico non terminale.
  - Definizione di S
  - Peggioramento della leggibilità della grammatica

Grammatiche regolari

- Lo studio dei linguaggi e delle grammatiche è argomento di corsi più avanzati noi vediamo solo:
  - le grammatiche Libere da Contesto (LC) che ci permettono di definire la sintassi dei linguaggi di programmazione e
  - le grammatiche regolari incluse in quelle LC.
- Nelle grammatiche libere da contesto le produzioni possono avere a sinistra della freccia solo un simbolo non terminale. (Gli esempi visti rientrano tutti in questo caso).
- Le grammatiche regolari (GR) sono LC ed inoltre esiste un ordinamento totale sui non terminali tale che j>i se j compare a destra nella produzione di i cioè $i::=e \in P \Rightarrow e \in E_{\{j \in V \mid j>i\} \cup \Lambda}$

Potenza delle grammatiche

- Si consideri l'insieme di sequenze su {a,b,c} costituite da n occorrenze di 'a' seguite da altrettante occorrenze di 'b', cioè:

$$\{a^n c b^n \mid n \geq 0\}$$

- Si definisca una grammatica G tale che $L(G)=\{a^n c b^n \mid n \geq 0\}$

Grammatiche libere da contesto

- Le produzioni delle grammatiche LC sono così definite $P = \{i ::= \alpha \mid i \in V, \alpha \in (\Lambda \cup V)^{+}\}$ senza nessuna restrizione sui non terminali a destra, come invece avveniva per le GR (ordinamento).
- L’alternatore(|), la [] e l’operatore * non sono ammessi nelle produzioni.A destra c’è una concatenazione di simboli (terminali e non).
- Le produzioni possono essere ricorsive: Non è possibile definire un ordinamento su V.
- Si dimostra (non in questo corso) che le grammatiche LC includono le GR.

Caratteristiche delle grammatiche LC

- La produzione

  S := a S b | c

  è una produzione corretta ma...
  - Cosa significa?
  - Quale linguaggio denota L(S)?
- Il significato delle produzioni e quindi il linguaggio definito dalla grammatica possono essere ottenuti con la relazione $\rightarrow^*$.

Esempi di derivazioni

produzione

S ::= aSb

derivazioni

S → aSb

aSb → aaSbb

aaSbb →aaaSbbb

e ancora...

produzione

S ::= c

derivazioni

S → c

aSb → acb

aaSbb → aacbb

continua1

Esempi di $\longrightarrow^*$

$S ::= a\ S\ b$

$S ::= c$

$S \longrightarrow^* S$

$S \longrightarrow^* aSb$

$aSb \longrightarrow^* aaaSbbb$

$S \longrightarrow^* c$

$S \longrightarrow^* acb$

$S \longrightarrow^* aacbb$

$S \longrightarrow^* aaacbbb$

Ottenuto per induzione sui passi di $\longrightarrow^*$

Linguaggio definito da una grammatica G

Sia G = $<\Lambda, V, S, P>$, $A \in V$

Definiamo linguaggio L(A) denotato in G da A:

$$L(A) = \{ \alpha \in \Lambda^* \mid A \rightarrow^* \alpha \}$$

Sia G = $<\Lambda, V, S, P>$,

Definiamo linguaggio L(G) denotato dalla grammatica

$$L(G) = L(s) = \{ \alpha \in \Lambda^* \mid s \rightarrow^* \alpha \}$$

Esempio: G=<{a,b}, {S}, S, {S::=a S b, S::= ε }>

allora $L(G) = L(S) = \{a^n b^n \mid n \geq 0\}$

Esempio delle espressioni aritmetiche

La seguente grammatica definisce il linguaggio delle espressioni aritmetiche:

<{0,1,2,3,4,5,6,7,8,9,*,+,-,/ ,(,)}, {Int,Op,Exp}, Exp,

{Int::= 0|1|2|3|4|5|6|7|8|9,

Op::= (+|-|*|/),

Exp::= Int, Exp::= Exp Op Exp, Exp::=(Exp)}, >

Struttura delle frasi

- Ad esempio L(G) contiene: {3+4*6,78-34,... ecc.}
- Le sequenze di simboli non rappresentano la struttura della frase che è invece espressa nella grammatica. Non c’è traccia in una sequenza di simboli di come sia stata derivata.
- Quali sono i modi per derivare una sequenza?
- Vediamo alcuni esempi di derivazioni diverse.

Struttura delle frasi

- Ad esempio 3+4*6 potrebbe essere stata derivata:

  Exp → Exp Op Exp → Int Op Exp → Int Op Exp Op Exp →

  3 Op Exp Op Exp → 3 + Exp Op Exp →3 + Int Op Exp →

  3 + 4 Op Exp → 3 + 4 * Exp →3 + 4 * Int → 3 + 4 * 6

  oppure

  Exp → Exp Op Exp → Exp Op Int → Exp Op Exp Op Int → Int Op Exp Op Int → Int Op Int Op Int→3 Op Int Op Int →3 + Int Op Int → 3 + 4 Op Int →3 + 4 * Int → 3 + 4 * 6

Definizione di grafo

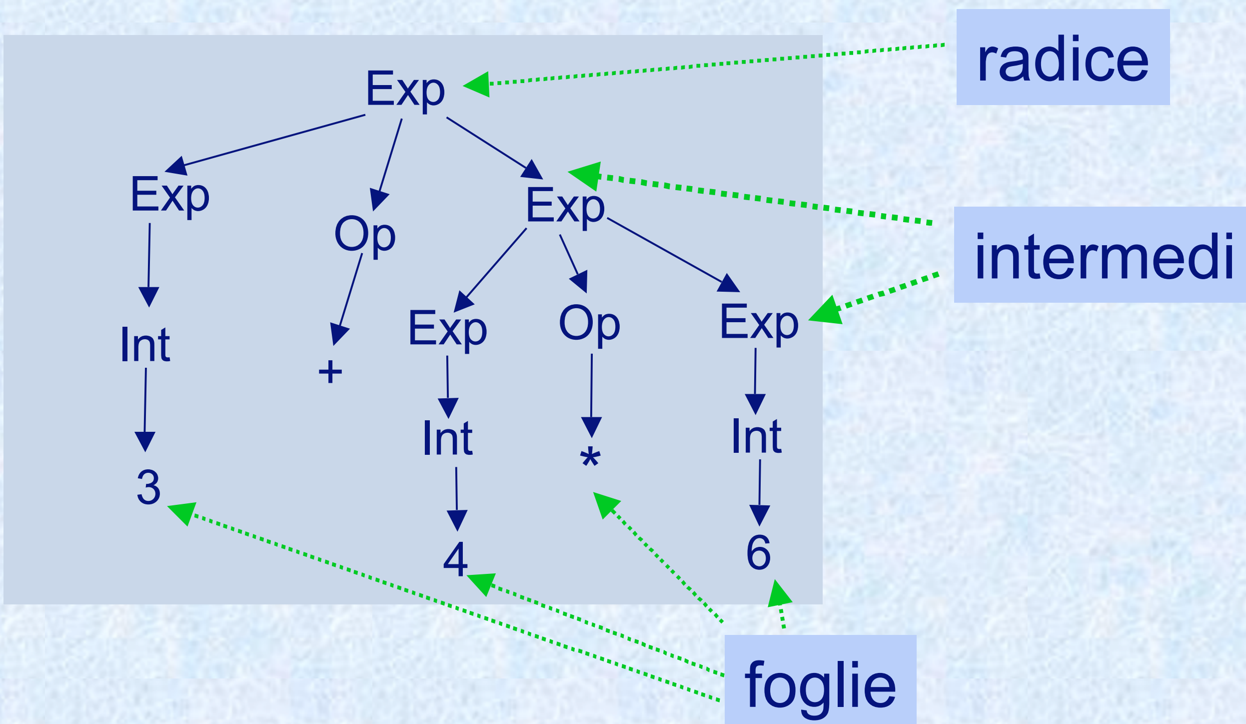
- Un grafo è una coppia di insiemi <N_L,E> dove N_L sono i nodi e E gli archi.
- I nodi hanno associata un'etichetta - L: N_L -> L(N_L) *(L è una funzione che dato un nodo mi calcola la sua etichetta).*
- Gli archi sono coppie <s,t> o triple <s,t,c> dove s e t sono nodi (di partenza e di arrivo se il grafo è orientato) mentre c è l'eventuale etichetta.

Alberi

- Un albero è un grafo orientato in cui ogni nodo ha al più un arco entrante e non ci sono cicli.
- Esiste un unico nodo detto **radice** che non ha archi entranti.
- I nodi che non hanno archi uscenti sono detti **foglie.**
- I nodi B_1, B_2, ..., B_k raggiunti da un arco uscente dal nodo A sono detti **figli** di A mentre A è detto **padre** di B_1, B_2, ..., B_k
- Si possono definire:
  - la **profondità** di un albero data dal cammino più lungo dalla radice ad una foglia.
  - l'**ampiezza** data dal massimo numero di archi uscenti da un nodo (ovvero dal numero massimo di figli per ogni nodo)

Esempio

Gli alberi sono utilissimi per strutturare le informazioni. Nella sintassi servono per rappresentare la struttura della frase, che altrimenti si perde nella sequenza di simboli.



Visita e frontiera di un albero

- La **visita** di un albero è un operazione in cui si esaminano tutti i nodi dell'albero:
  - l'accesso è dalla radice scendendo sui figli
  - esistono vari tipi di visite (non le trattiamo) a seconda dell'ordine con cui si esaminano i nodi.
- La **frontiera** di un albero è costituita dai nodi foglia dell'albero.

Alberi di derivazione sintattica parse tree

- Le produzioni di una grammatica possono essere utilizzate per costruire un albero, detto **albero di derivazione sintattica** (ADS), per rappresentare ogni frase del linguaggio.
- Sia G = <Λ,V,s,P> e sia A= <N,E> un ADS derivante da G. Valgono le seguenti proprietà:
  - L(N) $\subseteq \Lambda \cup$ V in particolare:
  - la radice ha etichetta s
  - le foglie sono etichettate su Λ
  - i nodi intermedi su V
- Per ogni nodo non foglia n con etichetta a, lo indicheremo n_a, compresa la radice, se esiste un arco $<n_a,n_b>$ $\Rightarrow$ esiste a::= $\gamma b \beta \in$ P

Costruzione degli ADS a partire da una grammatica G

- Sia $G = <\Lambda, V, s, P>$ una grammatica,

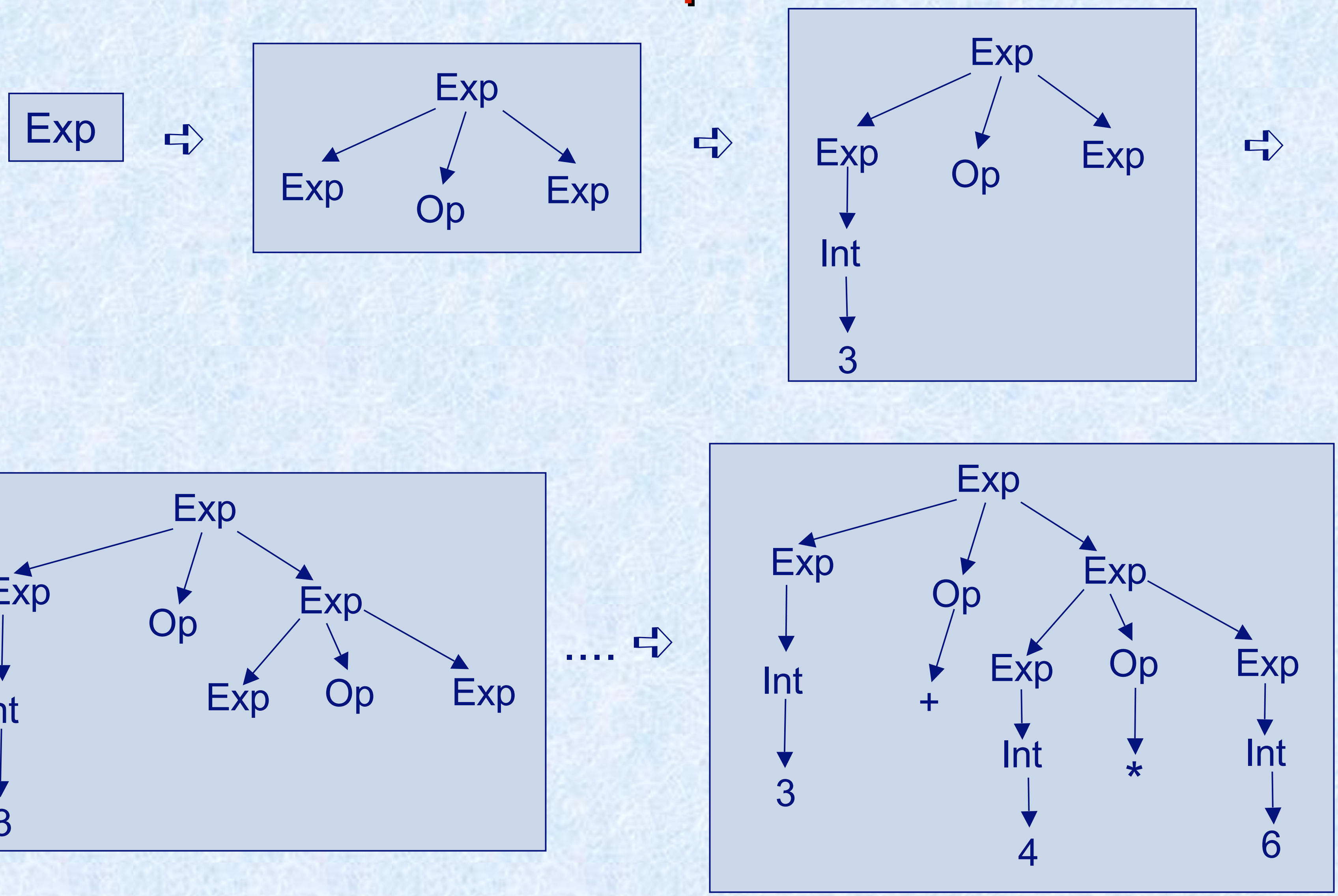
  1. si inizia costruendo un albero che ha un solo nodo etichettato con s cioè $A=<N,E>$ tali che $N=\{n_s\}, E=\{\}$
  2. Finchè la frontiera di A contiene un simbolo non terminale a si sceglie una produzione $a::=x_1 \ldots x_k \in P$, (si noti che $x_i \in \Lambda \cup V$) e si costruisce $A'=<N',E'>$ trasformando (riscrivendo) A nel seguente modo:
     - $N' = N \cup \{n_{x_i} \mid i \in [1,...k]\}$
     - $E' = E \cup \{<n_a, n_{x_i}> \mid i \in [1,...k]\}$.

     il nodo n_a che era una foglia in A, diviene in A' un nodo intermedio con k nodi figli, n_{x_i}

Produzioni come riscritture di alberi

- In sostanza si parte costruendo la radice e si espande l’albero sostituendo le foglie etichettate con simboli non terminali, con sottoalberi corrispondenti ad una produzione che abbia il simbolo non terminale a sinistra.
- Gli alberi che rappresentano frasi del linguaggio sono quelli che hanno sulla frontiera solo simboli terminali.
- La frase del linguaggio rappresentata da un albero si ottiene concatenando etichette della sua frontiera.

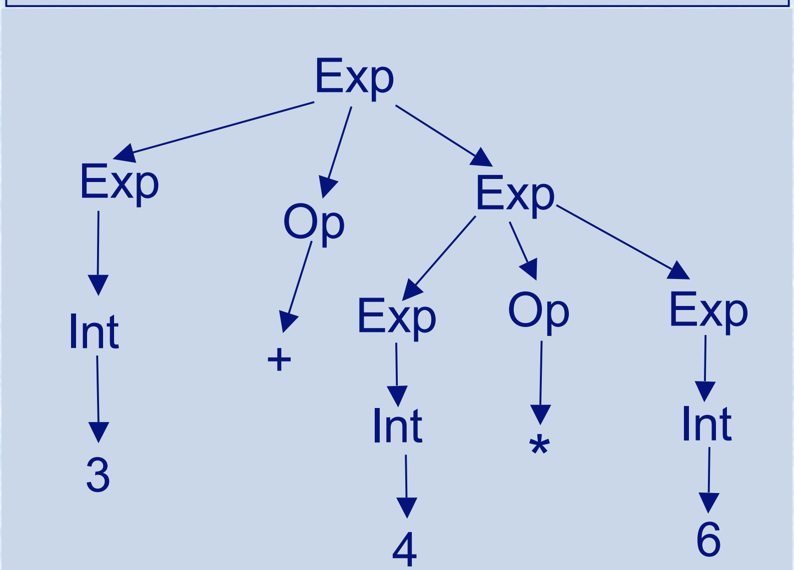
Esempio


Exp
Exp
Exp Op Exp
Exp
Exp Op Exp
Int
3
Exp
Exp Op Exp
Int Exp Op Exp
3
....
Exp
Exp Op Exp
Int + Exp Op Exp
3 Int * Int
4 6

Grammatiche ambigue

Una grammatica è ambigua se esiste più di un albero di derivazione sintattica con una stessa frontiera. Es: la grammatica delle espressioni aritmetiche definita precedentemente.

3+4*6 l'albero rappresenta 3+(4*6)



3+4*6, l'albero rappresenta (3+4)*6