Linguaggi Formali e Traduttori

3.2 Alberi sintattici, derivazioni canoniche e ambiguità

Sommario

Problema

- Per tradurre un programma è importante riconoscerne la **struttura**, che è codificata nella derivazione che genera quel programma.

- Tuttavia, possono esserci derivazioni differenti che generano lo stesso programma, anche imponendo certe discipline sull'ordine delle riscritture nelle derivazioni.

In questa lezione

- Definiamo gli **alberi sintattici** come alternativa alle derivazioni per ragionare sulla struttura delle stringhe generate da una grammatica astraendo dall'**ordine** delle riscritture.

- Mostriamo che certe grammatiche sono **ambigue**, in quanto ammettono alberi sintattici diversi.

- Definiamo due forme **canoniche** di derivazione, a **sinistra** e a **destra**, per disciplinare l'ordine delle riscritture e caratterizzare l'ambiguità di una grammatica (anche) in termini di derivazioni canoniche.

Alberi sintattici

Definizione

Data una grammatica $G = (V, T, P, S)$, gli **alberi sintattici** di G sono alberi con le seguenti caratteristiche:

- Ogni nodo interno (diverso da una foglia) è etichettato con una variabile in V.
- Ogni foglia è etichettata con una variabile in V, o da un terminale in T, o da ε.
- Se una foglia è etichettata con ε, è anche l'unico figlio del suo genitore.
- Se un nodo interno è etichettato con A e i suoi figli sono etichettati (da sinistra a destra) con $X_1, X_2, \ldots, X_n$, allora $A \to X_1 X_2 \cdots X_n$ è una produzione in P.

Definizione

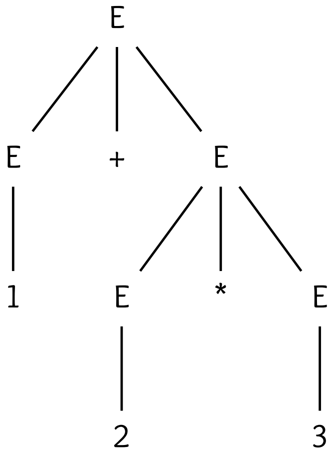
Il **prodotto** di un albero sintattico è la stringa ottenuta concatenando, da sinistra verso destra, le etichette di tutte le foglie dell'albero.

Teorema

$A \Rightarrow^*_G \alpha$ se e solo se esiste un albero sintattico di G con radice A e prodotto α.

Esempio

$$G = (\{E\}, \{0, 1, \dots, 9, +, *, (,)\}, \{E \to 0 \mid \cdots \mid 9 \mid E + E \mid E * E \mid (E)\}, E)$$



$$\begin{aligned} E &\Rightarrow E + E \\ &\Rightarrow 1 + E \\ &\Rightarrow 1 + E * E \\ &\Rightarrow 1 + 2 * E \\ &\Rightarrow 1 + 2 * 3 \end{aligned}$$

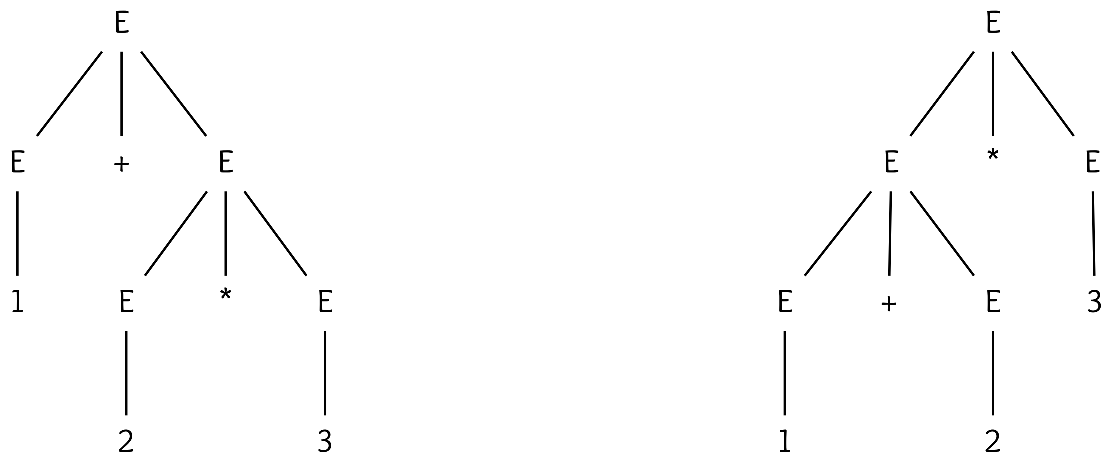
Grammatiche ambigue

Definizione

Una grammatica è **ambigua** se ammette <u>più alberi sintattici distinti</u> con lo <u>stesso prodotto</u>.

Esempio

La grammatica delle espressioni in forma infissa è ambigua.



- Questi alberi sottintendono due significati molto diversi per l'espressione $1 + 2 * 3$ che certamente vorremmo tradurre in modi diversi

Ambiguità e derivazioni

Attenzione

- Basta trovare **due** alberi sintattici **distinti** con lo stesso prodotto per dire che una grammatica ambigua
- **Non** basta trovare due derivazioni distinte che generano la stessa stringa per dire che una grammatica ambigua

Esempio

- $E \Rightarrow E + E \Rightarrow \underline{1 + E} \Rightarrow 1 + 2$
- $E \Rightarrow E + E \Rightarrow \underline{E + 2} \Rightarrow 1 + 2$

Queste due derivazioni sono evidentemente distinte (si noti la parte sottolineata), hanno lo stesso prodotto ($1 + 2$), ma corrispondono allo stesso albero sintattico!

Osservazioni

- L'ordine in cui vengono riscritte variabili diverse è irrilevante (esempio in questa slide)
- Quello che conta è se la stessa variabile viene riscritta in modi diversi (slide precedente)
- Imponendo un particolare ordine di riscrittura delle variabili possiamo individuare l'ambiguità guardando le derivazioni

Derivazioni canoniche

Definizione

Una derivazione $X \Rightarrow^* \alpha$ si dice **derivazione a sinistra** se (a ogni passo) viene riscritta la variabile che si trova più a sinistra. Useremo $\Rightarrow_{lm}$ e $\Rightarrow_{lm}^*$ come simboli per derivazioni a sinistra, dove lm abbrevia leftmost.

In maniera speculare si definiscono le **derivazioni a destra** e la notazione associata $\Rightarrow_{rm}$ e $\Rightarrow_{rm}^*$, dove rm abbrevia rightmost.

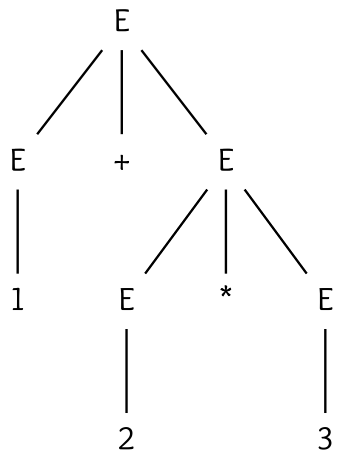
Proposizione

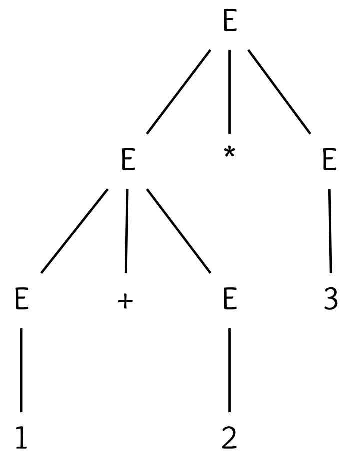
Se esistono due derivazioni canoniche distinte di G (entrambe leftmost o entrambe rightmost) per derivare la stessa stringa allora G è ambigua

Esempio

$$
\begin{aligned}
E &\Rightarrow_{lm} E + E \\
&\Rightarrow_{lm} 1 + E \\
&\Rightarrow_{lm} 1 + E * E \\
&\Rightarrow_{lm} 1 + 2 * E \\
&\Rightarrow_{lm} 1 + 2 * 3
\end{aligned}
$$

$$
\begin{aligned}
E &\Rightarrow_{lm} E * E \\
&\Rightarrow_{lm} E + E * E \\
&\Rightarrow_{lm} 1 + E * E \\
&\Rightarrow_{lm} 1 + 2 * E \\
&\Rightarrow_{lm} 1 + 2 * 3
\end{aligned}
$$





Esercizi

Mostrare che le seguenti grammatiche sono ambigue

1. $S \to aS \mid aSbS \mid \varepsilon$
2. $B \to t \mid f \mid B \wedge B \mid B \vee B \mid \neg B \mid (B)$

trovando per ciascuna:

- due alberi sintattici distinti con lo stesso prodotto;
- due derivazioni canoniche a sinistra distinte per la stessa stringa;
- due derivazioni canoniche a destra distinte per la stessa stringa.