Satzformen sind die Wörter aus $(N \cup T)^*$.

Notation:

Wir verwenden oft

- $a, b, c, \ldots$ für Terminalsymbole
- $A, B, C, \ldots$ für Nonterminale
- $u, v, w, \ldots$ für Terminalwörter aus T^*
- $\alpha, \beta, \gamma, \ldots$ für Satzformen

Beispiel

Eine kontextfreie Grammatik:

$$G = (\{E, A\}, \{0, 1, +, -, (,)\}, P, E)$$

mit

$$P = \{E \to (E+E), E \to (E-E), E \to A, A \to 0, A \to 1, A \to AA\}$$

Ableitungen

Wir definieren die Relation

$$\Rightarrow_G : (N \cup T)^* \to (N \cup T)^*$$

vermöge

$$\alpha A \beta \Rightarrow \alpha \gamma \beta \text{ falls } A \to \gamma \in P.$$

$\Rightarrow_G$ heißt Ableitungsrelation.

$\overset{*}{\Rightarrow}_G$ ist die reflexiv-transitive Hülle von $\Rightarrow_G$

Beispiel

$G = (\{E, A\}, \{0, 1, +, -, (,)\}, P, E)$

mit
$P = \{E \to (E+E), E \to (E-E), E \to A, A \to 0, A \to 1, A \to AA\}$

$E \Rightarrow (E + E) \Rightarrow (E + A) \Rightarrow (A + A) \Rightarrow (AA + A) \Rightarrow (A0 + A) \Rightarrow$
$(A0 + 1) \Rightarrow (10 + 1) \in T^*$

Wir geben in Zukunft oft nur die Produktionen an.

(Wenn klar ist was N, T und das Startsymbol sind.)

Sprache einer Grammatik

Definition

Es sei $G = (N, T, P, S)$ eine kontextfreie Grammatik.

Wir definieren $L(G) = \{ w \in T^* \mid S \overset{*}{\Rightarrow} w \}$.

$L(G)$ ist die von G *erzeugte Sprache*.

Wir nennen $\alpha \in \beta$ einen *Ableitungsschritt* und
$S \Rightarrow \alpha_1 \Rightarrow \cdots \Rightarrow \alpha_{n-1} \Rightarrow w$ eine *Ableitung* von w.

Eine Ableitung ist eine Linksableitung, falls in jedem Schritt das am weitesten links stehende Nonterminal ersetzt wird. (Rechtsableitung analog).

Beispiel

CFG G mit $S \to aSSb$, $S \to \epsilon$.

$S \Rightarrow aSSb \Rightarrow aaSSbSb \Rightarrow aaSbSb \Rightarrow aabSb \Rightarrow aabaSSbb \Rightarrow aabaSbb \Rightarrow aababb$

ist eine Linksableitung von $aababb$.

Ableitungsbäume

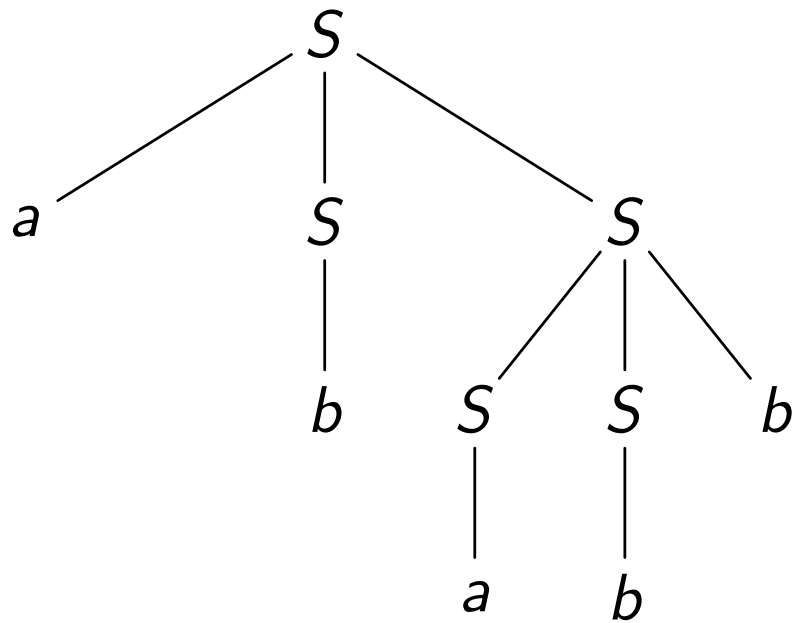
Definition

Ein Ableitungsbaum ist ein Baum:

1. er hat eine Wurzel
2. er ist orientiert
3. innere Knoten sind Nonterminale
4. Blätter sind Terminalsymbole
5. Kinder eines inneren Knotens A sind rechte Seite einer Produktion $A \to \alpha$

Beispiel

$S \to aSS \mid SSb \mid a \mid b$



Zugehörige Linksableitung:

$S \Rightarrow aSS \Rightarrow abS \Rightarrow abSSb \Rightarrow abaSb \Rightarrow ababb$

Theorem

Es sei $G = (N, T, P, S)$ eine CFG und $w \in T^*$.
Dann gilt $w \in L(G)$ genau dann, wenn es einen Ableitungsbaum gibt

- *mit Wurzel S,*
- *mit Blättern w.*

Beweis.

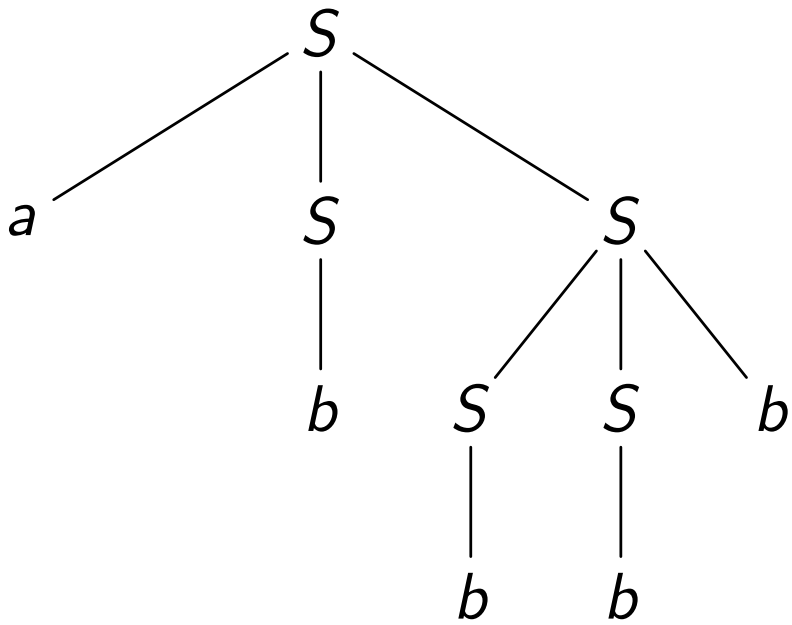
(Idee) Allgemeinere Aussage:
$A \overset{*}{\Rightarrow} \alpha$ genau dann, wenn Ableitungsbaum mit Wurzel A und Blättern α.

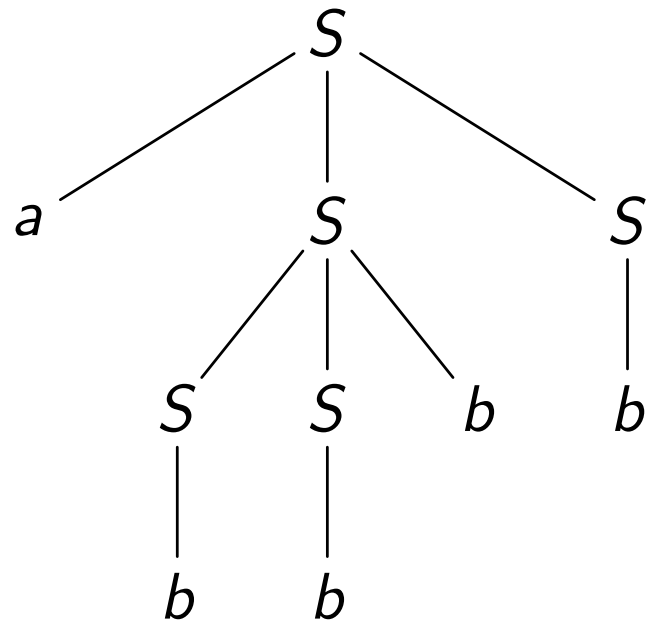
Beweis mit Induktion über Länge einer Ableitung. □

Zu jedem Ableitungsbaum gibt es eine *eindeutige* Linksableitung.

Eindeutigkeit

$S \to aSS \mid SSb \mid a \mid b$



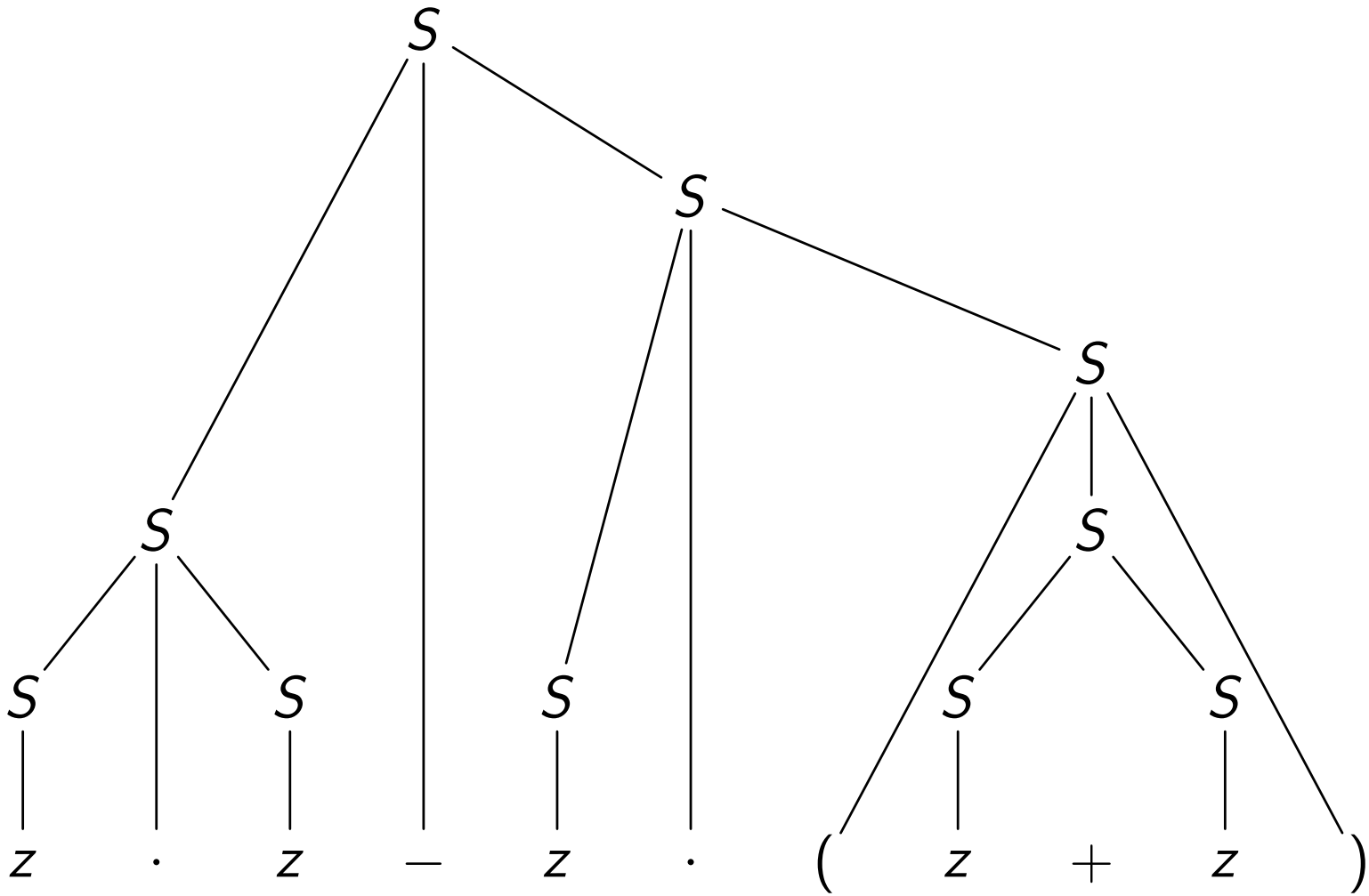


Es gibt zwei verschiedene Linksableitungen.

Beispiel 1

$S \to S + S \mid S - S \mid S \cdot S \mid S/S \mid (S) \mid \mathbf{z}$

Betrachte: $\mathbf{z} \cdot \mathbf{z} - \mathbf{z} \cdot (\mathbf{z} + \mathbf{z})$



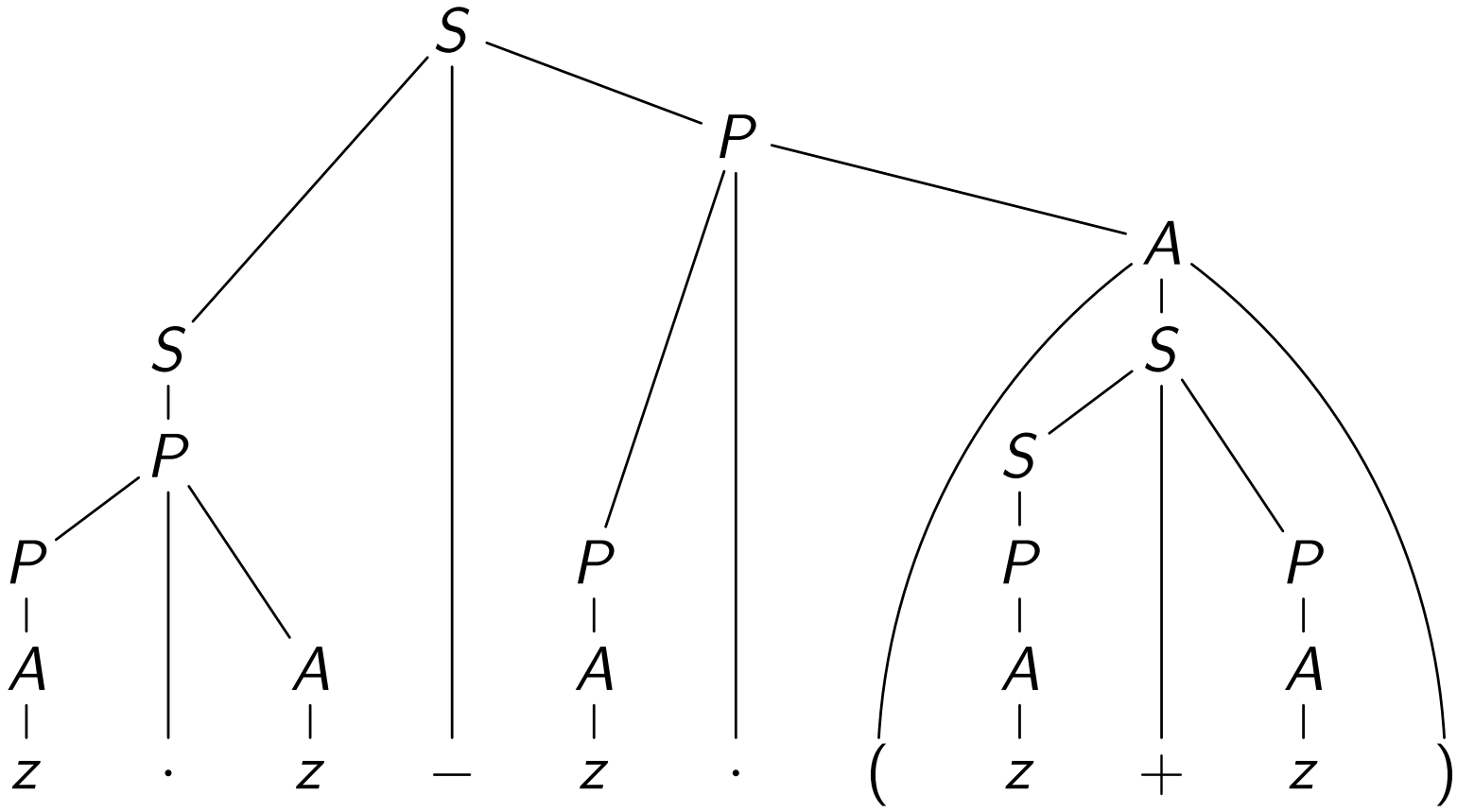
Beispiel 2

$S \to S + P \mid S - P \mid P$

$P \to P \cdot A \mid P/A \mid A$

$A \to (S) \mid \mathbf{z}$

Betrachte: $\mathbf{z} \cdot \mathbf{z} - \mathbf{z} \cdot (\mathbf{z} + \mathbf{z})$



Welchen Vorteil hat diese Grammatik?

RWTH AACHEN UNIVERSITY

Beispiel 3

$S \to P + S \mid P - S \mid P$
$P \to A \cdot P \mid A/P \mid A$
$A \to (S) \mid \mathbf{z}$

Betrachte: $\mathbf{z} \cdot \mathbf{z} - \mathbf{z} \cdot (\mathbf{z} + \mathbf{z})$

Welchen Vorteil hat diese Grammatik?

```c
#include ⟨stdlib.h⟩
#include ⟨stdio.h⟩

char * s;
void A(), P(), Z();

void S() {
  P();
  if(*s == '+') s++, S();
  if(*s == '-') s++, S();
}

void P() {
  A();
  if(*s == '*') s++, P();
  if(*s == '/') s++, P();
}
```

```c
void A() {
  if(*s == '(') {
    ++s;
    S();
    if(*s == ')')++s;
    else exit(1);
  }
  else Z();
}

int main(int argc, char * argv[])
{
  if(argc ≠ 2) exit(1);
  s = argv[1];
  S();
  if(*s ≠ 0) exit(1);
  printf("okay!\n");
  return 0;
}
```