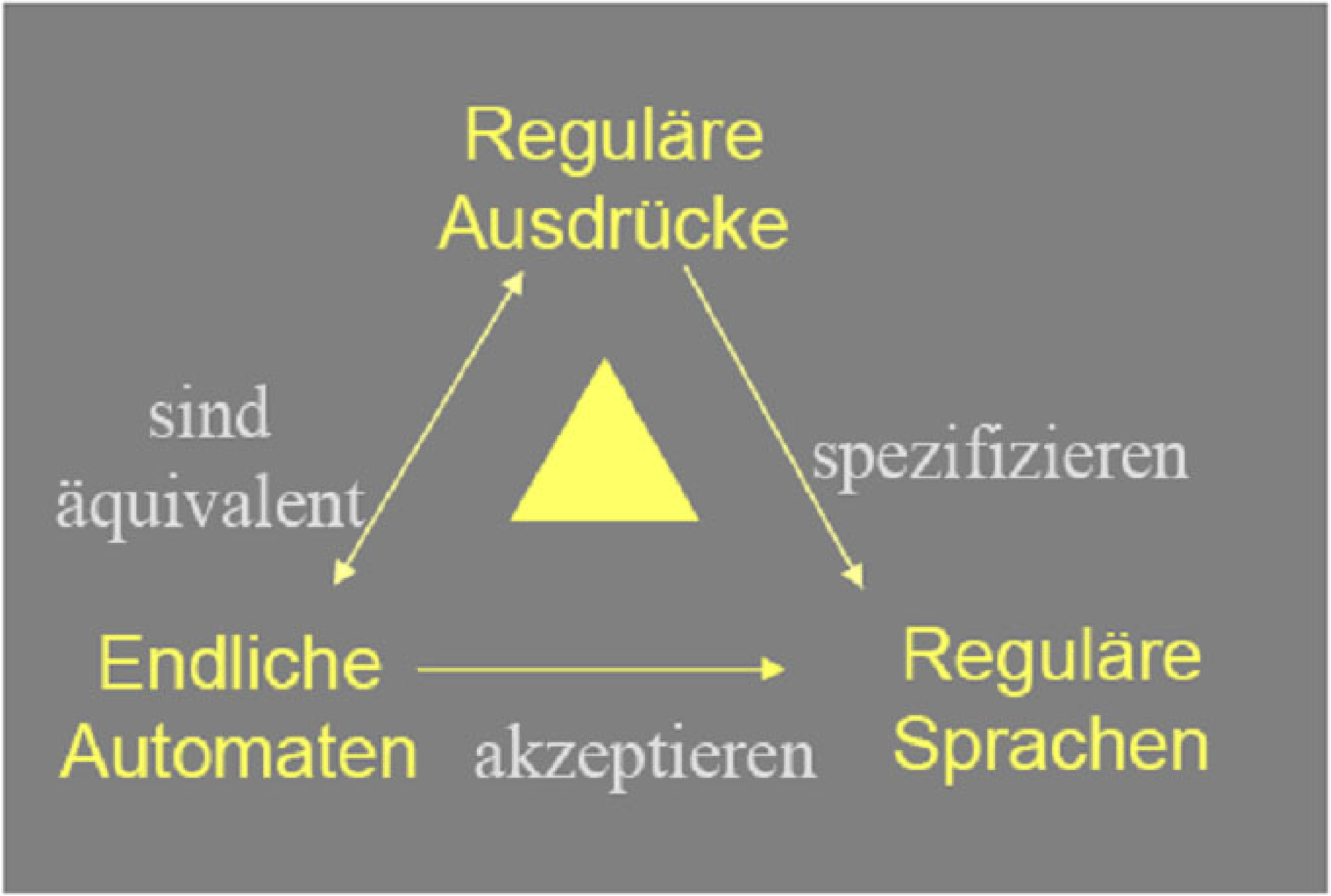
Einführung in die Computerlinguistik

Kontextfreie Sprachen und Pushdown-Automaten

Dozentin: Wiebke Petersen

WS 2004/2005

Wiederholung



©2004 Karin Haenelt

zu den Hausaufgaben

Welche Aussagen kann man mit Hilfe der Abschlußeigenschaften der regulären Sprachen und dem Pumping-Lemma über die Komplexität folgender formaler Sprachen machen:

1. $L_1 = \{w \in \{a,b\}^* : w$ enthält eine ungerade Anzahl von $b's\}$.
   L_1 ist regulär (L_1 akzeptierenden Automaten / L_1 erzeugende rechts-lineare Grammatik / regulärer Ausdruck)

2. $L_2 = \{w \in \{a,b\}^* : w$ enthält die gleiche Anzahl von $b's$ und $a's\}$.
   L_2 ist nicht regulär (Schnitt mit regulärer Sprache und Pumping-Lemma), obwohl für alle $v \in L_2$ auch $v^i \in L_2$

3. w^R ist das Wort w in umgekehrter Reihenfolge.
   $L_3 = \{ww^R : w \in \{a,b\}^*\}$.
   L_3 ist nicht regulär (Schnitt mit regulärer Sprache und Pumping-Lemma)

kontextfreie Sprachen

Definition 1. *Eine Grammatik* (N, T, S, P) *heißt* **kontextfrei**, *wenn alle Regeln/Produktionen die folgende Form haben:*

$$A \rightarrow \alpha, \text{ wobei } A \in N \textit{ und } \alpha \in (T \cup N)^*.$$

Eine durch eine kontextfreie Grammatik erzeugte Sprache heißt **kontextfrei**.

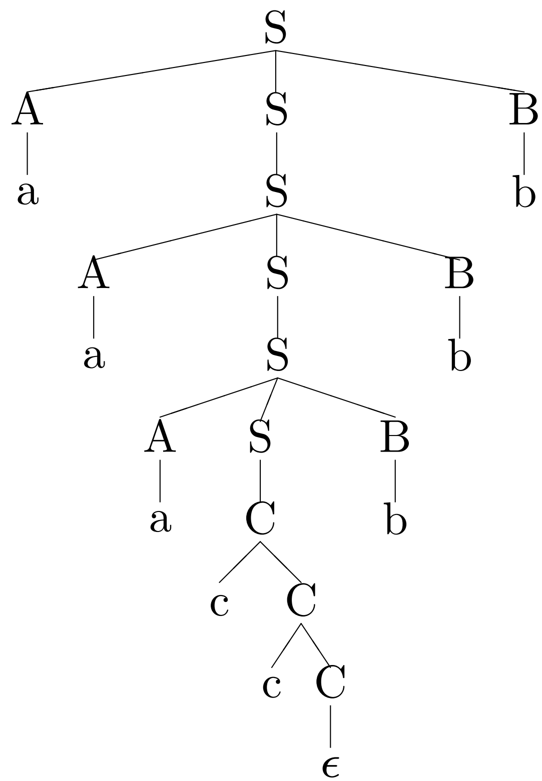
Satz 2. *Die Menge der kontextfreien Sprachen ist eine echte Obermenge der Menge der regulären Sprachen*

Beweis: Jede reguläre Sprache ist per Definition auch kontextfrei und es gibt mindestens eine kontextfreie Sprache, nämlich $a^n b^n$, die nicht regulär ist. $(S \rightarrow aSb, S \rightarrow \epsilon)$

Beispiel einer kontextfreien Sprache

$$G = \langle \{\mathsf{S}, \mathsf{A}, \mathsf{B}, \mathsf{C}\}, \{a, b, c\}, \mathsf{S}, P \rangle$$

$$P = \left\{ \begin{array}{cccccc} \mathsf{S} & \longrightarrow & \mathsf{ASB} & \mathsf{S} & \longrightarrow & \mathsf{C} \\ \mathsf{A} & \longrightarrow & a & \mathsf{B} & \longrightarrow & b \\ \mathsf{C} & \longrightarrow & cC & \mathsf{C} & \longrightarrow & \epsilon \end{array} \right\}$$



weitere kontextfreie Sprachen

- $L_1 = \{ww^R : w \in \{a,b\}^*\}$
- $L_2 = \{a^i b^j : i \geq j\}$
- $L_3 = \{w \in \{a,b\}^*$: Zahl der $a's$ größer als Zahl der $b's\}$
- $L_4 = \{w \in \{a,b\}^*$: Zahl der $a's$ gleich Zahl der $b's\}$

$$\left\{ \begin{array}{lllllllll} S & \rightarrow & aB & A & \rightarrow & a & B & \rightarrow & b \\ S & \rightarrow & bA & A & \rightarrow & aS & B & \rightarrow & bS \\ & & & A & \rightarrow & bAA & B & \rightarrow & aBB \end{array} \right\}$$

ambige Grammatiken und ambige Sprachen

Definition 3. *Gegeben eine kontextfreie Grammatik G. Eine Ableitung bei der stets das am weitesten links stehende nichtterminale Symbol ersetzt wird, heißt* **Linksableitung**

Definition 4. *Eine kontextfreie Grammatik G heißt* **ambig** *genau dann, wenn es für ein $w \in L(G)$ mehrere Linksableitungen mit $S \to^* w$ gibt. Sonst heißt G* **eindeutig***.*

Definition 5. *Eine kontextfreie Sprache L heißt* **ambig** *genau dann, wenn jede kontextfreie Grammatik mit $L(G) = L$ ambig ist.*

Ableitungsbäume und Linksableitungen

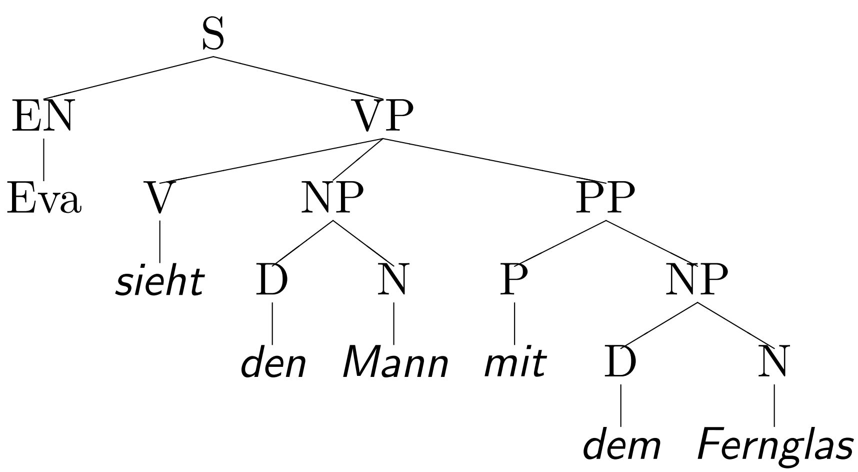
Zu jeder Linksableitung gibt es genau einen Ableitungsbaum
und
zu jedem Ableitungsbaum gibt es genau eine Linksableitung.

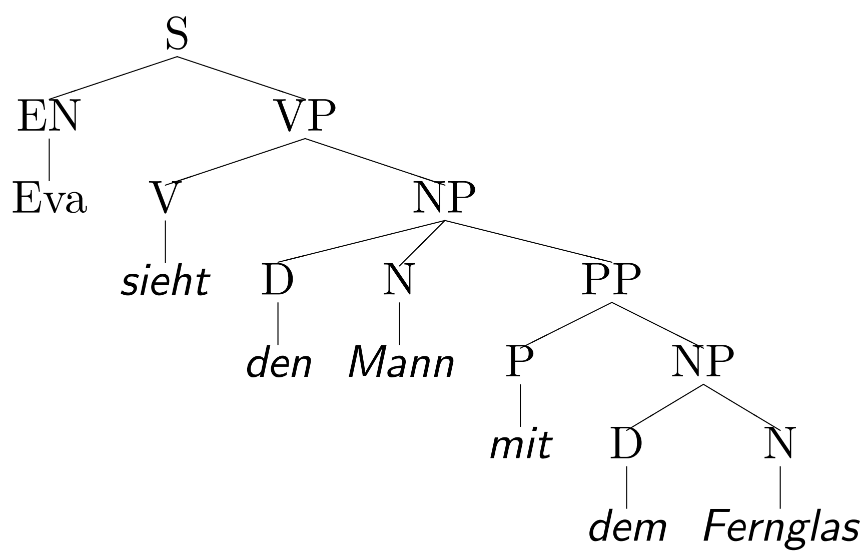
Beispiel einer ambigen Grammatik

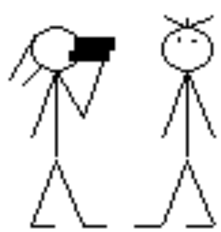
$$G = (N, T, \mathsf{NP}, P) \text{ mit } N = \{\mathsf{S}, \mathsf{EN}, \mathsf{NP}, \mathsf{VP}, \mathsf{PP}, \mathsf{D}, \mathsf{N}, \mathsf{P}\},$$

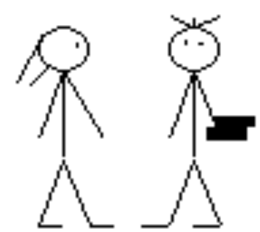
$$T = \{\text{Eva}, \text{sieht}, \text{den}, \text{Mann}, \text{mit}, \text{dem}, \text{Fernglas}\},$$

$$P = \left\{ \begin{array}{rcl rcl rcl} \mathsf{S} & \longrightarrow & \mathsf{EN\ VP} & \mathsf{VP} & \longrightarrow & \mathsf{V\ NP} & \mathsf{VP} & \longrightarrow & \mathsf{V\ NP\ PP} \\ \mathsf{NP} & \longrightarrow & \mathsf{D\ N} & \mathsf{NP} & \longrightarrow & \mathsf{D\ N\ PP} & \mathsf{PP} & \longrightarrow & \mathsf{P\ NP} \\ \mathsf{EN} & \longrightarrow & \text{Eva} & \mathsf{P} & \longrightarrow & \text{mit} & \mathsf{V} & \longrightarrow & \text{sieht} \\ \mathsf{D} & \longrightarrow & \text{den} & \mathsf{D} & \longrightarrow & \text{dem} & \mathsf{N} & \longrightarrow & \text{Mann} \\ \mathsf{N} & \longrightarrow & \text{Fernglas} & & & & & & \end{array} \right\}$$





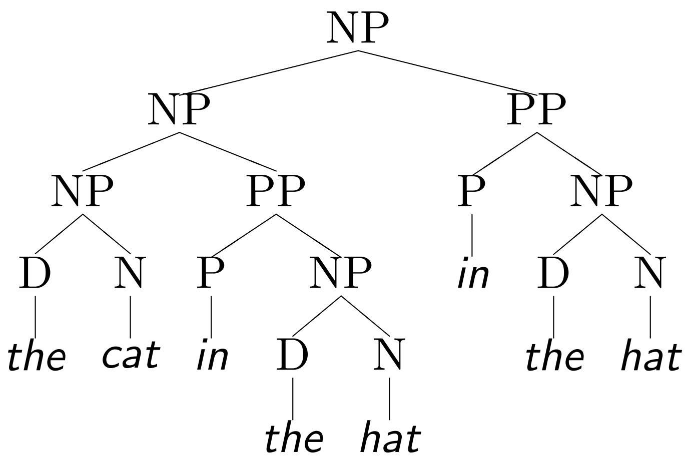


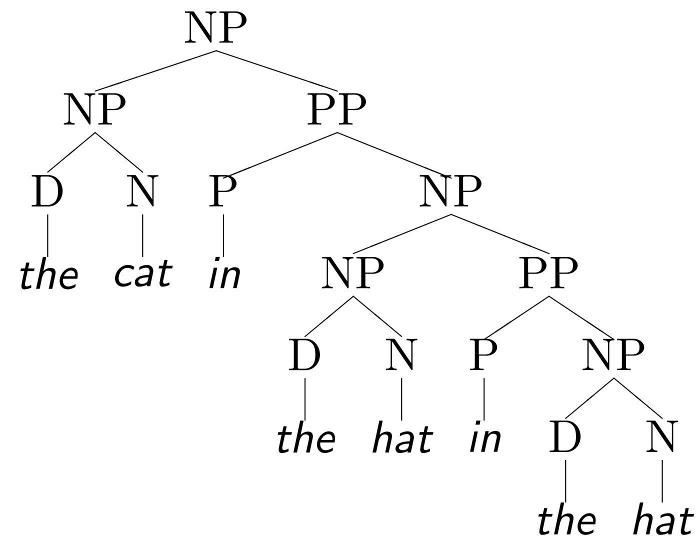


2. Beispiel einer ambigen Grammatik

$G = (N, T, \mathsf{NP}, P)$ mit $N = \{\mathsf{D}, \mathsf{N}, \mathsf{P}, \mathsf{NP}, \mathsf{PP}\}$, $T = \{\mathsf{the}, \mathsf{cat}, \mathsf{hat}, \mathsf{in}\}$,

$$P = \left\{ \begin{array}{lllllllll} \mathsf{NP} & \rightarrow & \mathsf{D\ N} & \mathsf{D} & \rightarrow & \mathsf{the} & \mathsf{N} & \rightarrow & \mathsf{hat} \\ \mathsf{NP} & \rightarrow & \mathsf{NP\ PP} & \mathsf{N} & \rightarrow & \mathsf{cat} & \mathsf{P} & \rightarrow & \mathsf{in} \\ \mathsf{PP} & \rightarrow & \mathsf{P\ NP} & & & & & & \end{array} \right\}$$





Chomsky-Normalform

Definition 6. *Eine Grammatik ist in* **Chomsky-Normalform** *(CNF), wenn alle Regeln die Gestalt*

1. $A \longrightarrow a$
2. $A \longrightarrow BC$

   mit $A, B, C \in T$ *und* $a \in \Sigma$ *haben (und gegebenenfalls* $S \longrightarrow \epsilon$*, dann aber ohne* S *auf den rechten Regelseiten).*

Satz 7. *Jede kontextfreie Sprache kann durch eine Grammatik in Chomsky-Normalform erzeugt werden.*

$$\textbf{Beispiel:} \left\{ \begin{array}{lllllllll} S & \longrightarrow & aB & A & \longrightarrow & a & B & \longrightarrow & b \\ S & \longrightarrow & bA & A & \longrightarrow & aS & B & \longrightarrow & bS \\ & & & A & \longrightarrow & bAA & B & \longrightarrow & aBB \end{array} \right\}$$

$$\left\{ \begin{array}{lllllllll} S & \longrightarrow & A'B & A & \longrightarrow & a & A' & \longrightarrow & a \\ S & \longrightarrow & B'A & B & \longrightarrow & b & B' & \longrightarrow & b \\ A & \longrightarrow & A'S & B & \longrightarrow & B'S & & & \\ A & \longrightarrow & B'A_2 & B & \longrightarrow & A'B_2 & & & \\ A_2 & \longrightarrow & AA & B_2 & \longrightarrow & BB & & & \end{array} \right\}$$

13 Kellerautomaten (Pushdown-Automaten)

Ziel:

Einführung eines Automatenmodells, mit dem genau die kontextfreien Sprachen akzeptiert werden können.

Wir benötigen eine echte Erweiterung des Modells des endlichen Automaten.

Ansatz: Hinzunahme eines unbeschränkten Speichers.

Speicherinhalt ist jeweils ein Wort.

Zugriff auf Speicherinhalt über die Spitze: Der Wortanfang kann jeweils gelesen und modifiziert werden.

Wir sprechen von Kellerspeicher oder Stack oder Pushdown Stack.

Informelles Beispiel

Wir betrachten die Sprache $\{a^i b^i \mid i > 0\}$.

Das Akzeptieren eines Eingabewortes geschieht wie folgt:

1. Aufbau des Kellers: für jedes gelesene a lege ein Symbol auf dem Keller ab (wir nehmen das Symbol Z)
2. Abbau des Kellers: für jedes gelesene b nehme ein Symbol Z vom Keller herunter
3. durch zwei Kontrollzustände sorge dafür, dass Aufbau und Abbau nur in dieser Reihenfolge möglich sind (Aufbau mit q_0, Abbau mit q_1, keine Rückkehr nach q_0)
4. Akzeptiere, wenn am Ende des Eingabewortes der Kellerboden erreicht ist.

Der Keller realisiert in diesem Fall eine Zählervariable.

Kellerautomaten

Kellerautomaten entstehen aus endlichen Automaten durch

- Hinzunahme eines Kelleralphabets
- Erweiterung der Transitionen (es muss das Lesen und Ersetzen der Kellerspitze realisiert werden)

Kellerautomaten: Definition

Definition 8. *Ein* **Kellerautomat** *ist ein 6-Tupel* $\langle \Phi, \Sigma, \Gamma, \Delta, q_0, F \rangle$ *bestehend aus:*

1. *einem* **Zustandsalphabet** Φ
2. *einem* **Eingabealphabet** Σ
3. *einem* **Kelleralphabet** Γ
4. *einer* **Übergangsrelation / Transitionsrelation** $\Delta \subseteq \Phi \times \Sigma \cup \{\epsilon\} \times \Gamma \cup \{\epsilon\} \times \Gamma^* \times \Phi$
5. *einem* **Startzustand** q_0 *und*
6. *einer Menge von* **Endzuständen** $F \subset \Phi$.

Wirkung einer Transition

Bedeutung der Transition (p, a, Z, v, q):

In Zustand p, bei Eingabebuchstabe a und Symbol Z auf Kellerspitze:

> **Ersetze Z an Kellerspitze durch das Wort v, gehe zum nächsten Eingabebuchstaben und in den Zustand q über.**

Bei zweiter Komponente ε:

> **analog, ohne auf der Eingabe voranzuschreiten und ohne den Eingabebuchstaben zu lesen.**

Transition

oberstes Symbol auf dem Stack
(wird entfernt)
- pop up -

neuer Zustand

(p, a, Z, v, q)

aktueller Zustand

zu lesendes Eingabesymbol

Wort, das auf den Stack gelegt wird
- push down -

Transition

oberstes Symbol auf dem Stack
(wird entfernt)
- pop up -

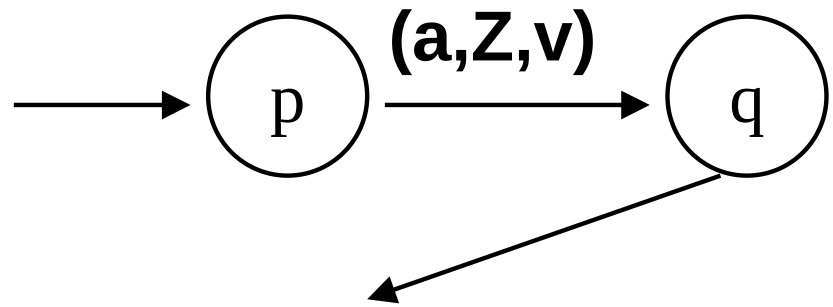
neuer Zustand

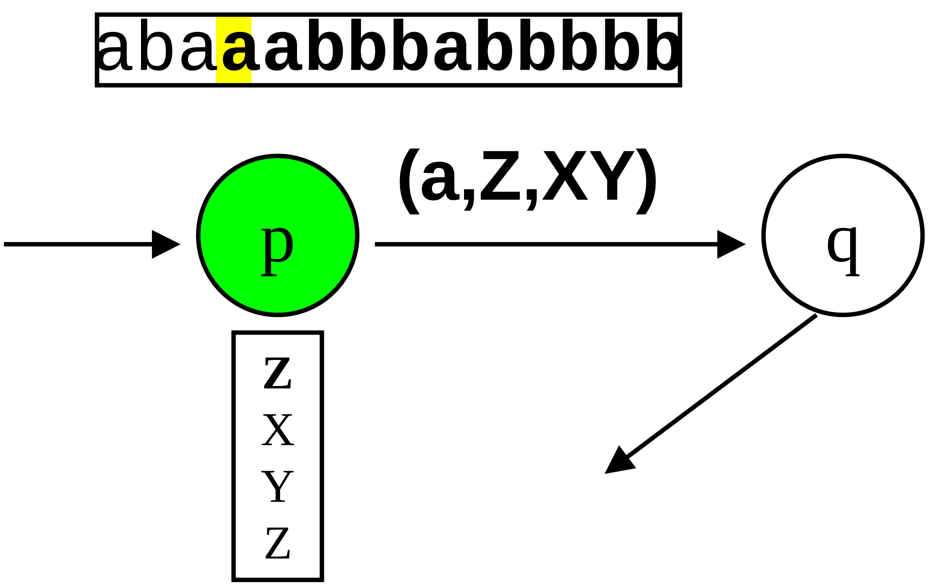
(p, a, Z, v, q)

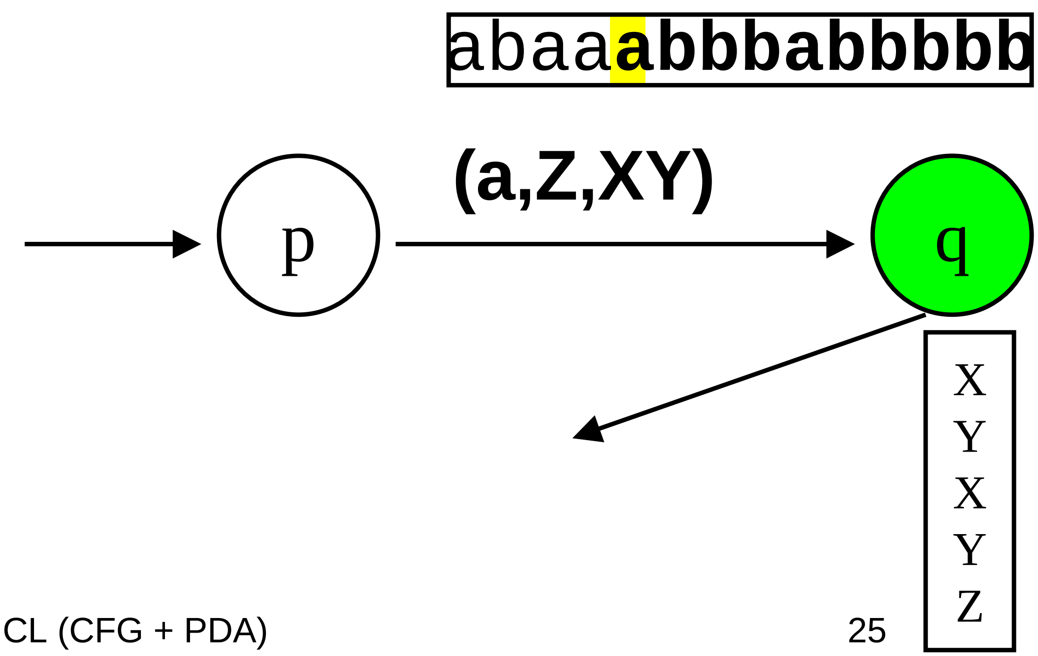
aktueller Zustand

zu lesendes Eingabesymbol

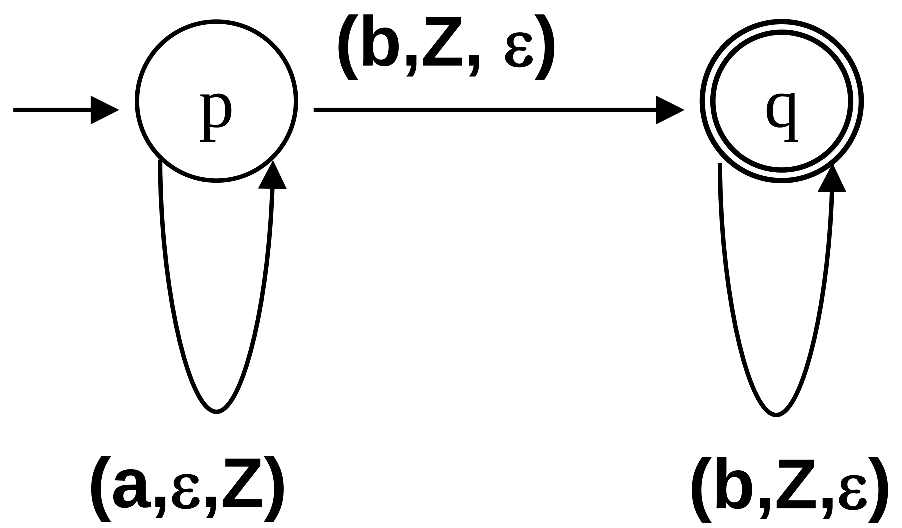
Wort, das auf den Stack gelegt wird
- push down -





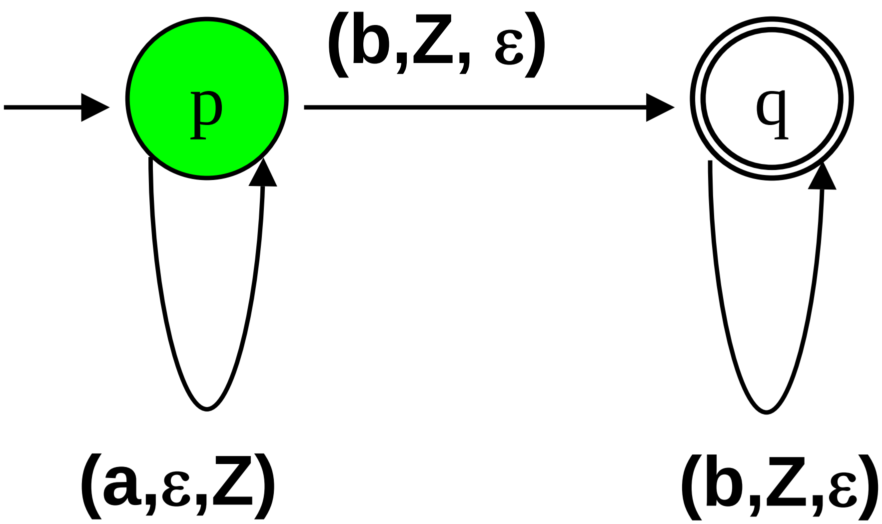


Beispiel eines Kellerautomaten



dieser Kellerautomat akzeptiert
die Sprache $a^n b^n$

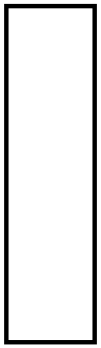
Arbeitsweise eines Kellerautomaten



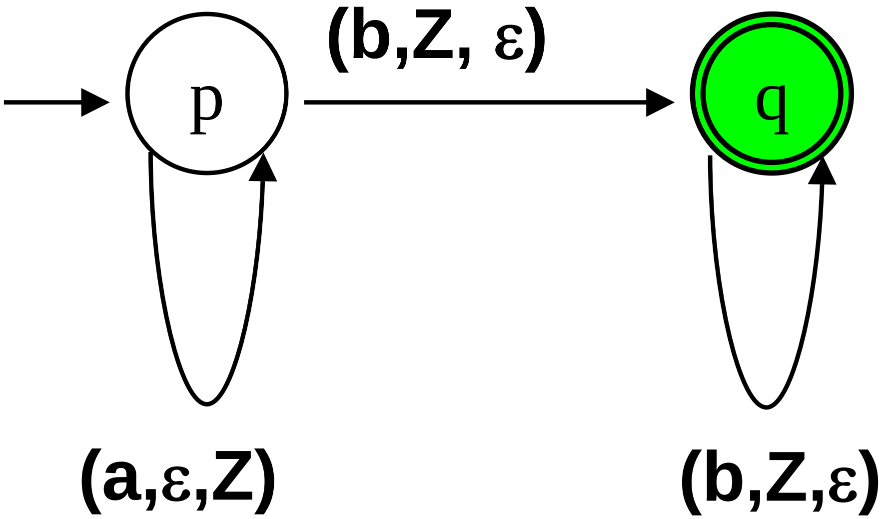
(noch) zu lesendes Wort:

a	**a**	**a**	**b**	**b**	**b**

aktueller Stack:



Arbeitsweise eines Kellerautomaten



Der Automat befindet sich in einem Endzustand!

Darum akzeptiert der Kellerautomat das Wort!

(noch) zu lesendes Wort:

Das Wort ist abgearbeitet!

aktueller Stack:

Der Stack ist leer!

Von Grammatiken zu Kellerautomaten

Satz 9. *Zu einer kontextfreien Grammatik G kann man einen Kellerautomaten A konstruieren, der die von G generierte Sprache akzeptiert.*

Konstruktionsvorschrift:
Sei $G = (N, T, S, P)$ die gegebene Grammatik. Konstruiere den Automaten $A = \langle \Phi, \Sigma, \Gamma, \Delta, q_0, F \rangle$ wie folgt:

- $\Phi = \{q_0, q_1\}$, $\Sigma = T$, $\Gamma = T \cup N$, $F = \{q_1\}$,
- Δ enthält die folgenden Transitionen:
  - $(q_0, \epsilon, \epsilon, S, q_1)$
  - für jede Regel $B \rightarrow \beta$ der Grammatik eine Transition $(q_1, \epsilon, B, \beta, q_1)$
  - für jedes Symbol $a \in T$ eine Transition $(q_1, a, a, \epsilon, q_1)$.

Wirkungsweise der Konstruktion

- Für das Akzeptieren einer kontextfreien Sprache genügt ein Kellerautomat mit nur zwei Zuständen, wobei die einzige Aufgabe des Startzustands darin besteht, das Startsymbol S der Grammatik in den Keller zu legen.
- Während der eigentlichen Rechnung befindet sich der Automat permanent in dem selben Zustand, die Rechnung findet nur in dem Keller statt.
- Der konstruierte Automat vollzieht zwei verschiedene Arbeitsschritte:
  - Nicht-Leseschritt mit Bezug auf Grammatikregel $B \longrightarrow \beta$:
    Ersetze die Kellerspitze B mit β
  - Leseschritte:
    Lies ein $a \in T$ der Eingabekette und entferne a von der Kellerspitze.

Hausaufgaben

1. Beschreiben Sie die Sprache, die von der Grammatik auf Folie 4 generiert wird.
2. Geben Sie die Grammatik von Folie 4 in der Chomsky-Normalform wieder.
3. Zeichnen Sie alle Ableitungsbäume für das Wort $aababb$ der Sprache L_4 von Folie 5.
4. Gebe eine Grammatik an, die die Sprache der wohlgeformten arithmetischen Terme über dem Alphabet $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, (,), +, -, \cdot, :\}$ generiert.
5. Geben Sie kontextfreie Grammatiken für die Sprachen L_1, L_2 und L_3 von Folie 5 an.
6. w^R ist das Wort w in umgekehrter Reihenfolge. Geben Sie eine kontextfreie Grammatik für die Sprache $L = \{ww^R : w \in \{a, b\}^*\}$ an. Konstruieren Sie einen Kellerautomaten, der L akzeptiert.
7. Gegeben sei folgende Regelmenge einer Grammatik G:

$$\begin{array}{llllllllll} S & \rightarrow & aB & S & \rightarrow & bA & & & \\ A & \rightarrow & a & A & \rightarrow & aS & A & \rightarrow & bAA \\ B & \rightarrow & b & B & \rightarrow & bS & B & \rightarrow & aBB \end{array}$$

   Welche Sprache wird von der Grammatik generiert?
   Geben Sie einen Kellerautomaten an, der die Sprache akzeptiert.