Das Pumping-Lemma für kontextfreie Sprachen

Chomsky-Hierarchie

Typ	Name	Erlaubte Produktionen	Akzeptierende Maschine	Beispiel
3	Regulär	N → wM $w \in T^*$	Endlicher Automat	a^n
2	Kontextfrei	N → w $w \in (N \cup T)^*$	Kellerautomat	$a^n b^n$
1	Kontext-sensitiv	uNv → uwv $\mathrm{u}, \mathrm{v} \in (N \cup T)^*$ $w \in (N \cup T)^+$ S → eps	Linear gebundener Automat	$a^n b^n c^n$
0	Rekursiv aufzählbar	u → v $\mathrm{u} \in V^*\mathrm{N}V^*, \mathrm{v} \in V^*$ $\mathrm{V} = (N \cup T)$	Turing Maschine	Halteproblem

Abgrenzung kontextfreier Sprachen

- Die Zugehörigkeit einer Sprache zu dieser Sprachklasse kann durch Angabe einer Typ-2-Grammatik oder eines Kellerautomaten beweisen werden.

- Wie beweist man die Nicht-Zugehörigkeit?

Rückblick

- Die Nichtzugehörigkeit einer Sprache zur Klasse der regulären Sprachen können wir mit Hilfe des Pumping-Lemmas für reguläre Sprachen zeigen.

- Gibt es so etwas auch für kontextfreie Sprachen?

- Das Pumping-Lemma für reguläre Sprachen basiert auf Schleifen beim Ablauf eines endlichen Automaten.

- Welche Schleifen finden wir bei kontextfreien Grammatiken?

Schleifen bei kontextfreien Grammatiken

- NonTerminals im Ableitungsbaum können sich wiederholen

- Dadurch können sich Teile der Ableitung wiederholen

Beispiel Schleife

Produktionsregeln:

$S \rightarrow UA$

$U \rightarrow u$

$A \rightarrow WY$

$Y \rightarrow y$

$W \rightarrow w$

Welche Sprache wird durch diese Produktionsregeln erzeugt?

Beispiel Schleife

Produktionsregeln:

$S \rightarrow UA$

$U \rightarrow u$

$A \rightarrow WY$

$Y \rightarrow y$

$W \rightarrow w$

Weitere Produktionsregeln

$W \rightarrow VB$

$V \rightarrow v$

$B \rightarrow WX$

$X \rightarrow x$

Welche Sprache wird durch diese Produktionsregeln erzeugt?

Beispiel

- Sei
  $G = (\{A, B, C\}, \{0, 1,2\}, A, \{A \rightarrow 0B0|2, B \rightarrow C|CC, C \rightarrow 1A1\})$

- Eine lange Ableitungsfolge:

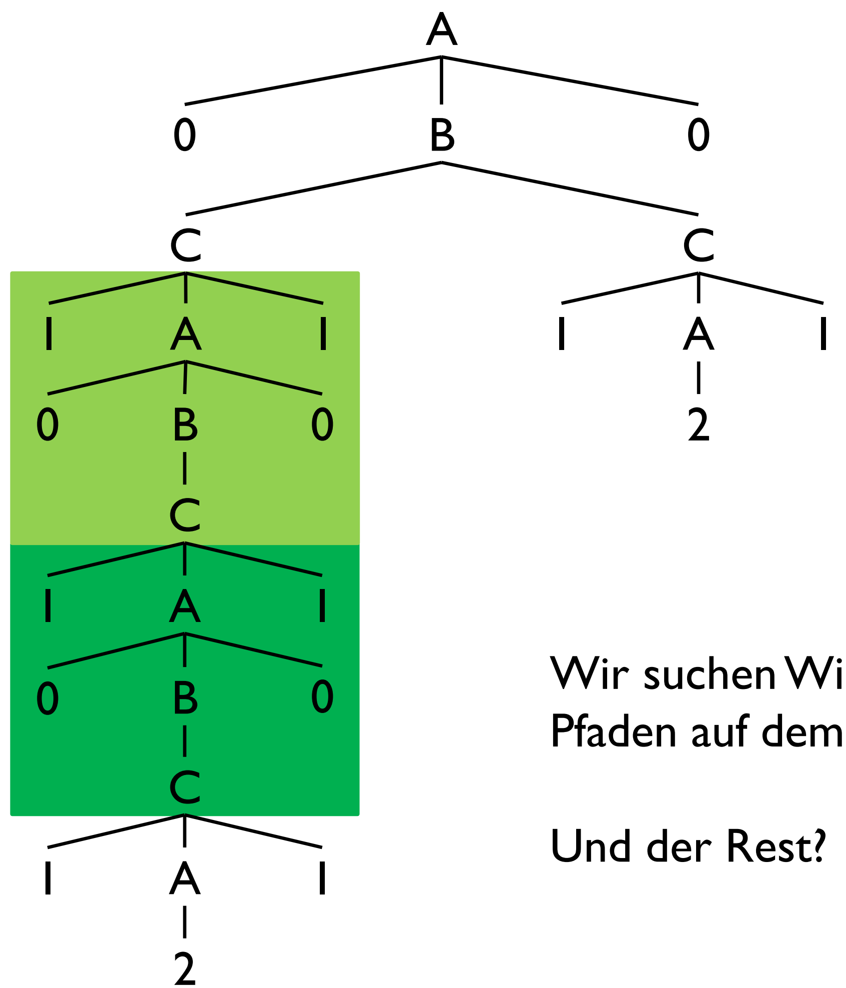
$$A \overset{A \to 0B0}{\Rightarrow} 0B0 \overset{B \to CC}{\Rightarrow} 0CC0 \overset{C \to 1A1}{\Rightarrow} 01A1C0$$

$$01A1C0 \overset{A \to 0B0}{\Rightarrow} 010B01C0 \overset{B \to C}{\Rightarrow} 010C01C0 \overset{C \to 1A1}{\Rightarrow} 0101A101C0$$

$$0101A101C0 \overset{A \to 0B0}{\Rightarrow} 01010B0101C0 \overset{C \to 1A1}{\Rightarrow} 01010B01011A10 \overset{A \to 2}{\Rightarrow}$$
$$01010B01011210 \underset{B \to C}{\Rightarrow} 01010C01011210 \underset{C \to 1A1}{\Rightarrow} 010101A101011210$$

$$010101A1010111210 \underset{A \to 2}{\Rightarrow} 010101210101210$$

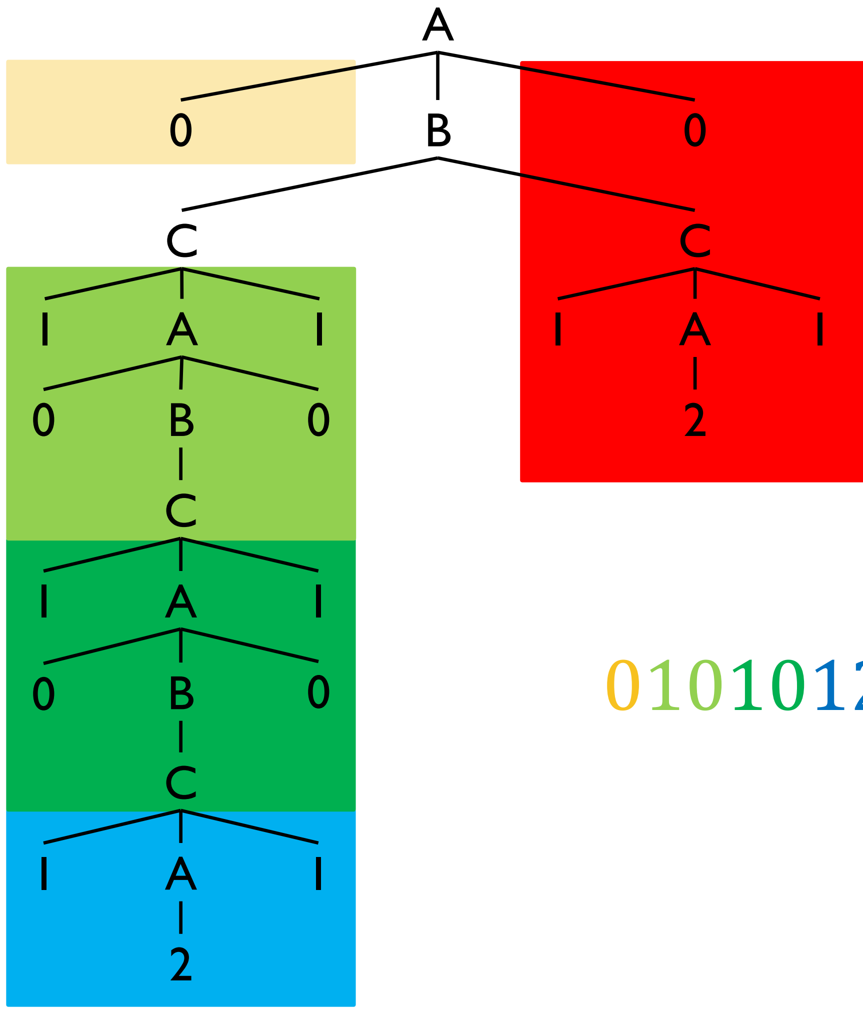
Beispiel



Wir suchen Wiederholungen in den Pfaden auf dem Ableitungsbaum!

Und der Rest?

Beispiel



010101210101011210

Beispiel

- $0101012101011210 \in L$
- $010121011210 \in L$
- $01010101210101011210 \in L$
- $010101010121010101011210 \in L$
- $0101010101012101010101011210 \in L$
- …

Schleifen bei kontextfreien Grammatiken

- Wenn die Tiefe des Ableitungsbaums größer ist als die Zahl der NonTerminals in der Grammatik, dann muss sich ein NonTerminal im Ableitungsbaum wiederholen
- Wenn wir eine Mindesttiefe des Ableitungsbaums garantieren können, dann können wir auch die Existenz einer Schleife garantieren
- Können wir einen Zusammenhang zwischen der Länge eines Wortes und der Tiefe des Ableitungsbaums herstellen?

Pfadlänge bei Grammatiken in Chomsky-Normalform

- Sei $G = (N, T, S, P)$ eine Grammatik in Chomsky-Normalform.
- Sei dazu B ein Ableitungsbaum zu einem Wort $w \in L(G)$ und n der längste Pfad in B.

- Dann gilt: $|w| \leq 2^{n-1}$.

Beweis

- Induktion über die Länge des längsten Pfades.
- Induktionsanfang $n = 1$:
  - Nur ein Ableitungsschritt möglich.
  - Dieser muss von der Form $S \Rightarrow x, x \in T$ sein.
  - Also gilt $|w| = |x| = 1 = 2^{1-1}$.
- Induktionsvoraussetzung:
  - Für alle Grammatiken in CNF gilt: wenn der längste Pfad eines Ableitungsbaums zu einem Wort w die Länge n hat, dann gilt $|w| \leq 2^{n-1}$.
- Induktionsschritt:
  - Sei G eine Grammatik in CNF und $\mathrm{w} \in L(G)$ und B ein Ableitungsbaum für w, so dass der längste Pfad in B die Länge $n + 1$ hat. $|w| \leq 2^{(n+1)-1}$

Beweis

- Induktionsschritt:
  - Sei G eine Grammatik in CNF und w $\in L(G)$ und B ein Ableitungsbaum für w, so dass der längste Pfad in B die Länge $n + 1$ hat.
  - Sei O.b.d.A $S \rightarrow UV, UV \in N$ die Produktion für die obersten Kanten im Ableitungsbaum.
  - Dann gibt es $u, v \in T^*$ mit $w = uv$ und $U \Rightarrow^* u$ und $V \Rightarrow^* v$
  - Dann hat der Teilbaum unterhalb von U als maximale Tiefe n.
  - Dann hat der Teilbaum unterhalb von V als maximale Tiefe n.
  - Dann gilt $|u| \leq 2^{n-1}$ und $|v| \leq 2^{n-1}$.
  - Dann gilt $|w| = |u|+|v| \leq 2^{n-1} + 2^{n-1} = 2^n = 2^{(n+1)-1}$.

Schleifen bei kontextfreien Grammatiken

- Wir können das Wiederholen eines NonTerminals im Ableitungsbaum garantieren

- Eine Ableitungsteilbaum der Form N → vNx wiederholt sich

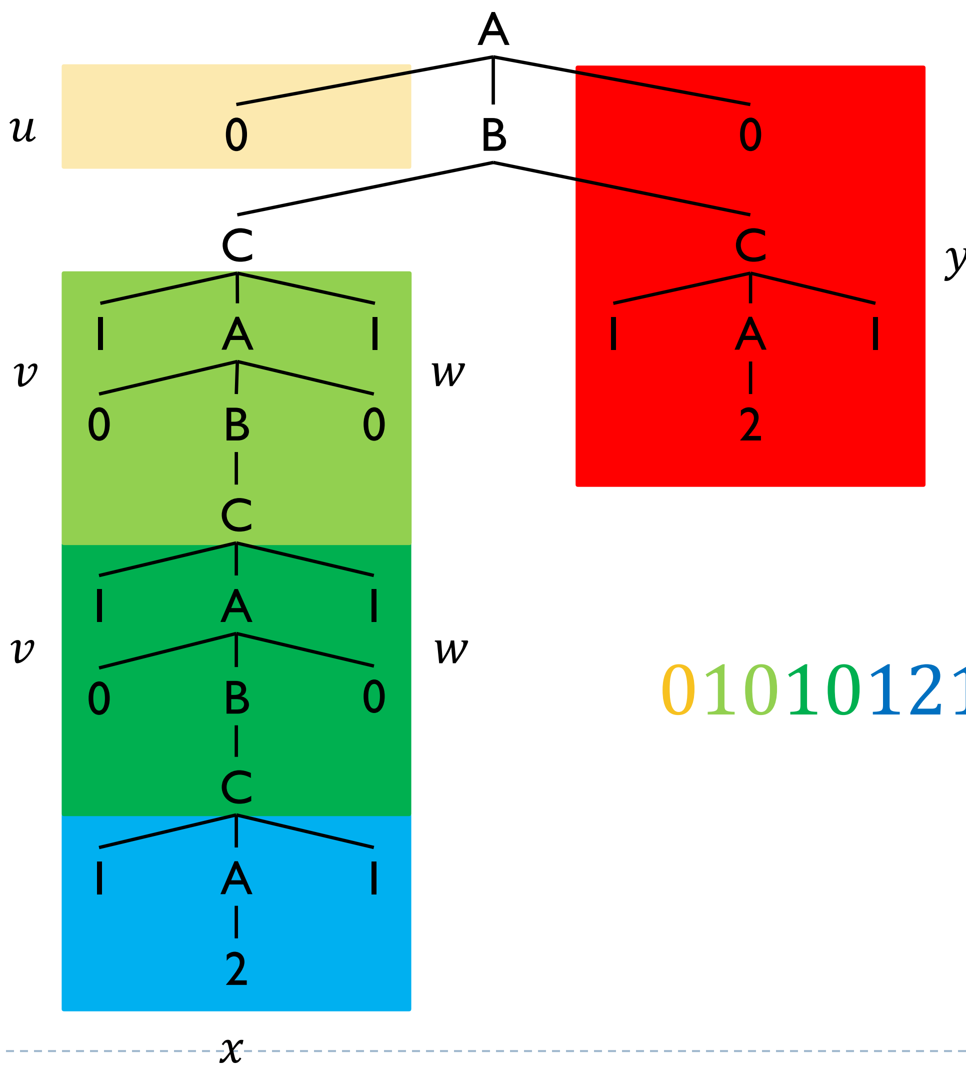
- Am Schleifenende wird N → w abgeleitet

- Dazu kommen noch ein Präfix u und ein Postfix y

- Unsere Ableitung mit Schleife produziert

$$uv^kwx^ky$$

Beispiel



0101012101011210

Pumping-Lemma für kontextfreie Sprachen

- Sei L eine kontextfreie Sprache über einem Alphabet A.

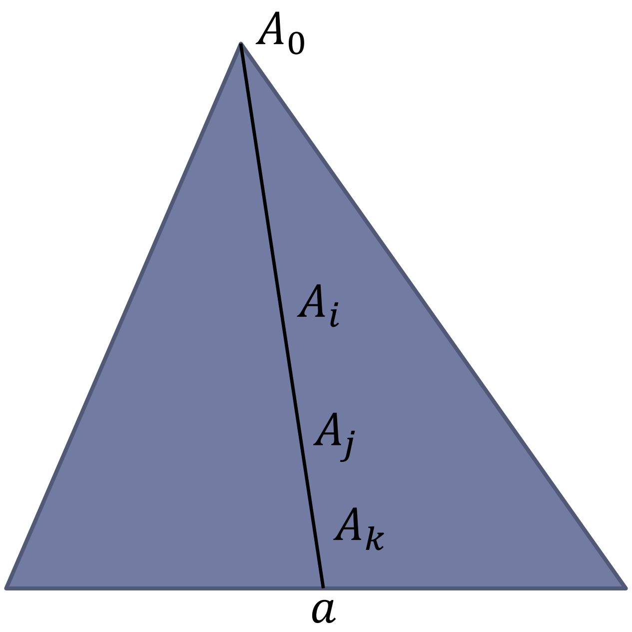
- Dann gibt es eine Konstante n derart, so dass für alle Wörter $z \in L$ mit $|z| \geq n$ gilt:
- Es gibt Wörter $\mathrm{u}, \mathrm{v}, \mathrm{w}, x, y \in A^*$ mit $\mathrm{z} = uvwxy$, für die gilt:
  - $vx \neq \varepsilon$
  - $|vwx| \leq n$
  - $\forall k \in \mathbb{N}_0: uv^k wx^k y \in L$

Beweis

- Sei $G = (N, T, S, P)$ eine Grammatik in Chomsky-Normalform, mit $L(G) = L$.
  - Da L kontextfrei ist, existiert eine Grammatik (4.7) und daher auch eine in CNF (4.15).
- Sei $k = |N|$, sei $z \in L$ mit $|z| \geq 2^k = n$.
- Dann hat der längste Pfad für einen Ableitungsbaum von z in G mindestens die Länge $k + 1$ (4.20).

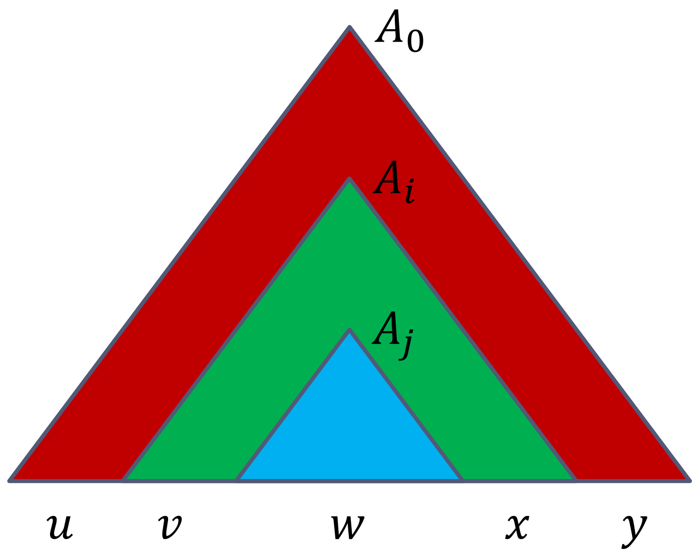
Beweis

- Seien $A_0 A_1 \dots A_k$ die Nichtterminale auf diesem Pfad.
  - Das sind k Ableitungsschritte.
  - Mit einem weiteren Ableitungsschritt wir aus dem letzten Nichtterminal ein Terminal.
- Das sind $k + 1$ Nichtterminale.
- In N sind aber nur k verschiedene Nichtterminale enthalten.
- Unter den $k + 1$ Nichtterminalen aus $A_0 A_1 \dots A_k$ muss es mindestens zwei gleiche geben.
- Seien i und j die Indizes dieser beiden Nichtterminalen, also $A_i = A_j$



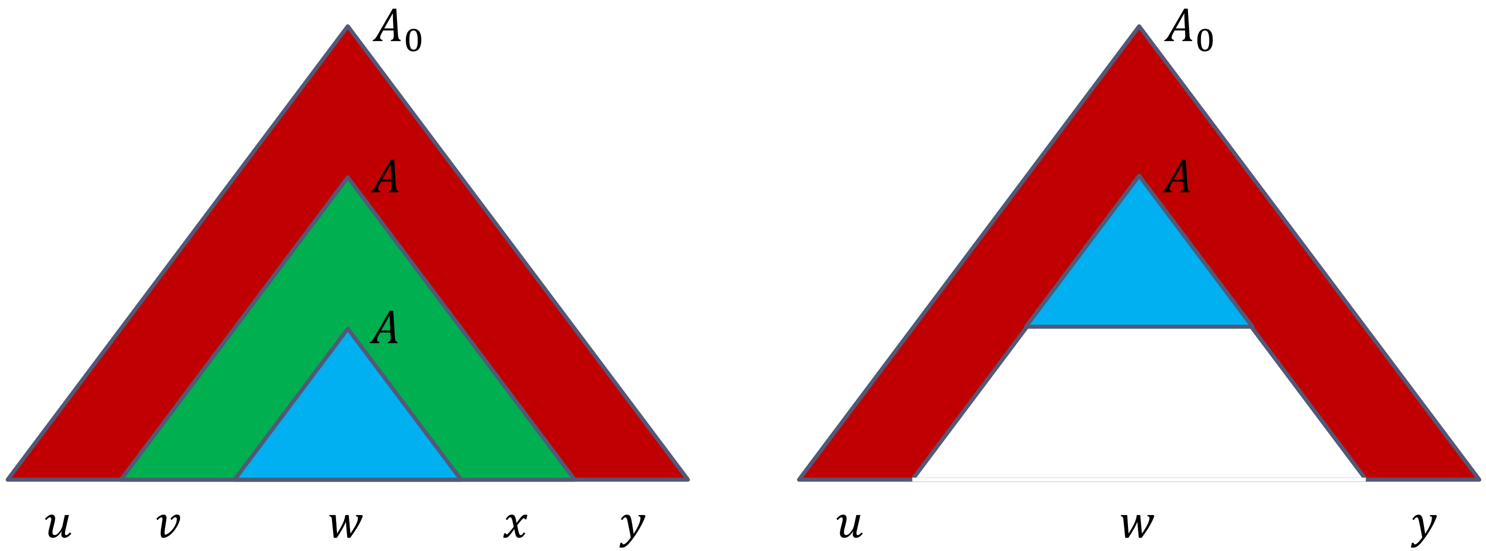
Beweis

- Was wird aus diesen Nichtterminalen erzeugt?



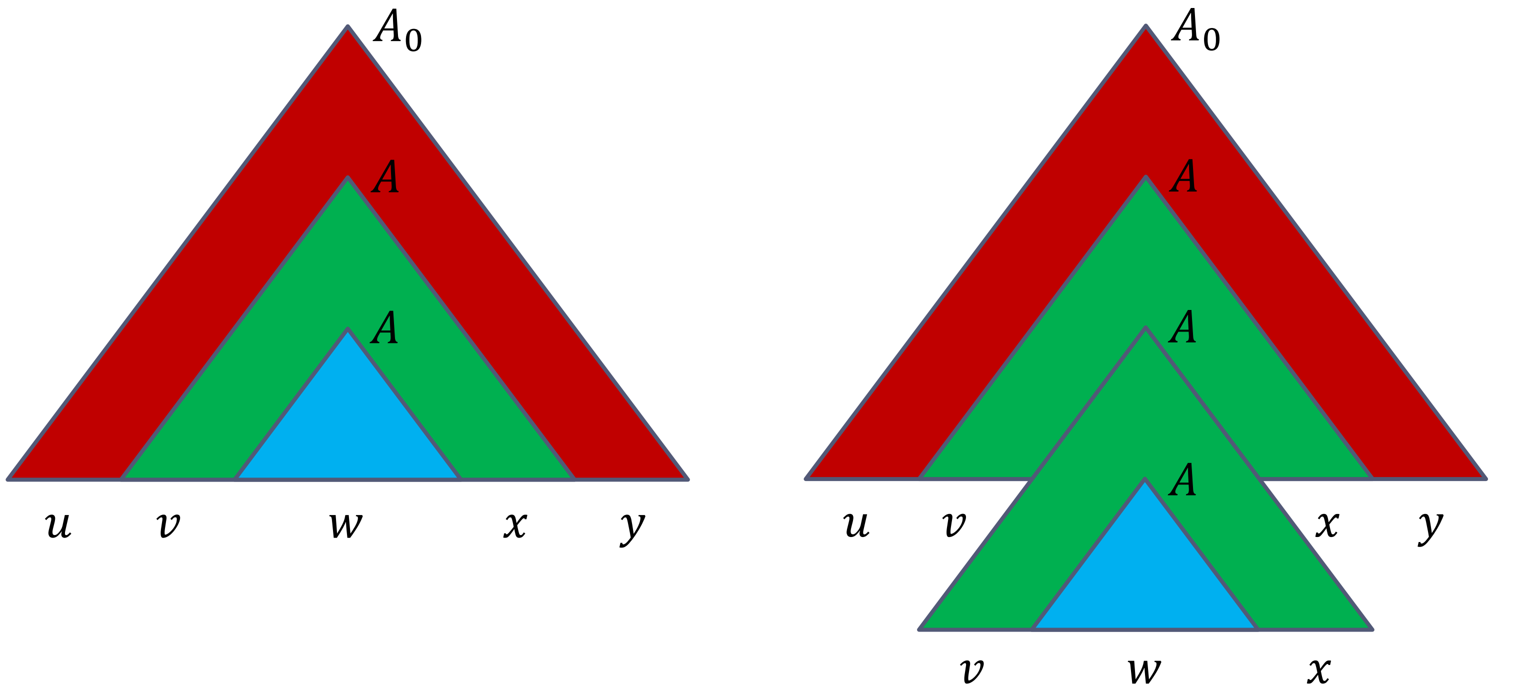
Beweis

- Was wird aus diesen Nichtterminalen erzeugt?
  Sei $A = A_i = A_j$



Beweis

- Was wird aus diesen Nichtterminalen erzeugt?
  Sei $A = A_i = A_j$



Beweis

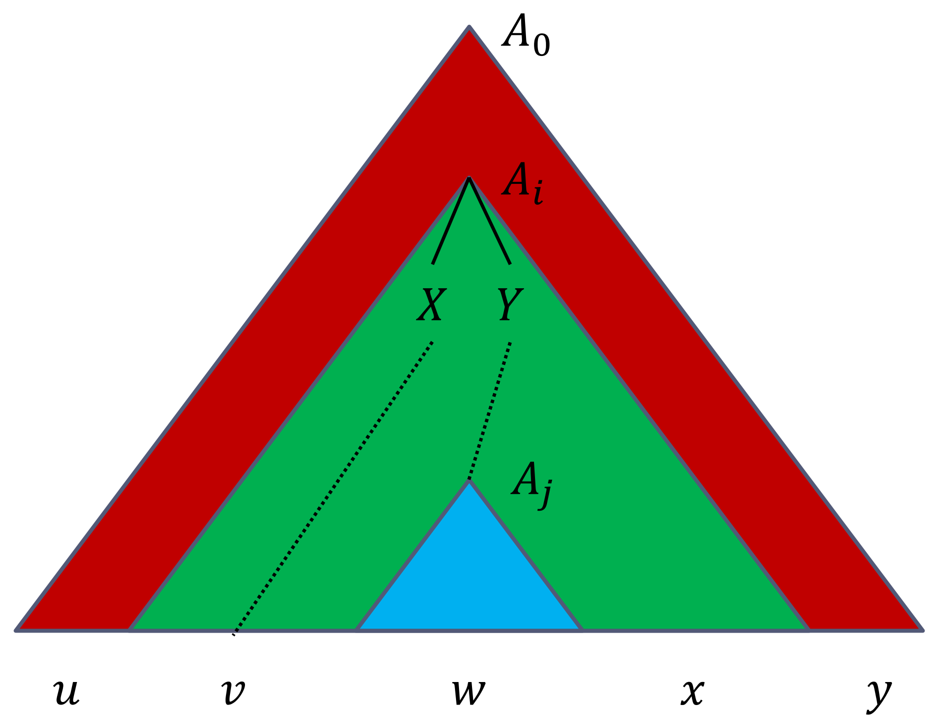
- Es kann im Ableitungsbaum also der unter A_i stehende Teilbaum durch denjenigen ersetzt werden, der unter A_j steht.
- Es kann im Ableitungsbaum also der unter A_j stehende Teilbaum beliebig häufig durch den unter A_i stehenden ersetzt werden.

- So entstehen Ableitungsbäume für Wörter alle Wörter der Form uv^kwx^ky mit $k \in \mathbb{N}_0$.
- **Es gilt also:** $\forall \boldsymbol{k} \in \mathbb{N}_{\boldsymbol{0}}: \boldsymbol{uv^kwx^ky} \in \boldsymbol{L}$

- Es bleibt zu zeigen: $vx \neq \varepsilon$ und $|vwx| \leq n$

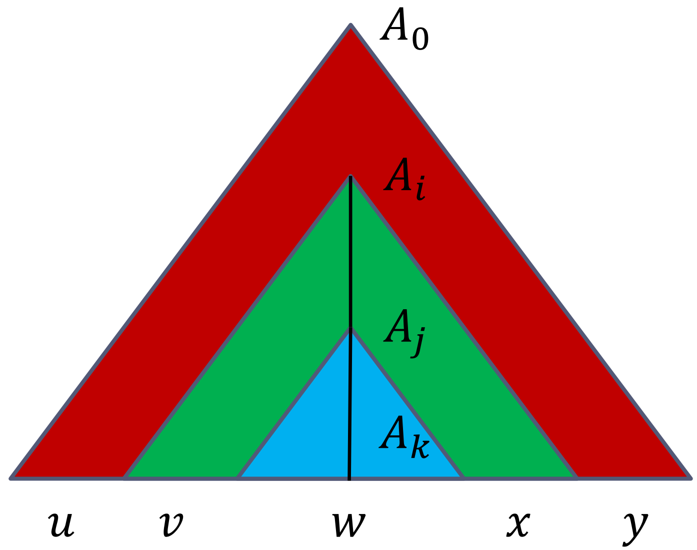
Beweis

- $vx \neq \varepsilon$:
  - Es gilt $i < j$.
  - Daher wird auf dem Pfad zwischen A_i und A_j mindestens eine Produktion der Form $\mathrm{N} \rightarrow XY$ verwendet.
  - Nur einer von X oder Y wird im folgenden Pfad zu A_j abgeleitet.
  - G ist in CNF, daher gibt es keine Produktionen der Form $N \rightarrow \varepsilon$ (außer Startsymbol).
  - Aus dem anderen aus X und Y wird mindestens ein Terminal für v oder x.



Beweis

- $|vwx| \leq n$:
  - A_i und A_j sind die letzte Doppelung von Nichtterminalen auf dem längsten Pfad.
  - Der Pfad von A_i bis A_k hat daher maximal die länge k, bis zum Terminal also $k + 1$.
  - Das unter A_i erzeugte Wort vwx hat also maximal die Länge $2^{(k+1)-1} = 2^k = n$.



Anwendung des Pumping-Lemmas für kontextfreie Sprachen

- Annahme eine Sprache L **sei** kontextfrei.
- Zeige, dass für **beliebiges** n das Folgende gilt.
- Es gibt **ein** Wort $z \in L$ (im Allgemeinen abhängig von n), so dass
- es **keine** Zerlegung $\mathrm{z} = uvwxy$ gibt so dass gleichzeitig gilt:
- $vx \neq \varepsilon$ und
- $|vwx| \leq n$ und
- $\forall k \in \mathbb{N}_0: uv^k wx^k y \in L$

Anwendung des Pumping-Lemmas für kontextfreie Sprachen

- Annahme eine Sprache L sei kontextfrei.
- Zeige, dass für **beliebiges** n das Folgende gilt.
- Es gibt **ein** Wort $\mathrm{z} \in L$ (bei Bedarf abhängig von n), so dass
- **Für alle** Zerlegungen $\mathrm{z} = uvwxy$ mit $vx \neq \varepsilon$ und $|vwx| \leq n$ gilt:
- **Es gibt** ein k, so dass $uv^kwx^ky \notin L$.

Beispiel

- Sei $L = \{0^t 1^t 2^t \mid t \in \mathbb{N}_0\}$.
- Behauptung: L ist nicht kontextfrei.
- Beweis:
  - Sei $n \in \mathbb{N}$ beliebig.
  - Wir wählen $\mathrm{z} = 0^n 1^n 2^n$.
  - Sei $uvwxy = z$ eine beliebige Zerlegung von z mit $vx \neq \varepsilon$ und $|vwx| \leq n$.
  - Dann kann vwx nicht gleichzeitig Nullen und Zweien enthalten, da zwischen den Nullen und Zweien n Einsen liegen und $|vwx| \leq n$.
  - Damit bricht $v^k w x^k$ die Gleichheit zwischen Nullen, Einsen und Zweien

Pumping Lemma: Lernziele

- Sprachen in Typ 3, 2, 1 einteilen können

- Mathematische Beweismethoden
  Schubfachprinzip
  Widerspruchsbeweis
  Vollständige Induktion

Pumping Lemma: Mögliche Klausuraufgaben

Mithilfe der Pumping Lemma Beweisen dass eine Sprache einen gegebenen Typ nicht hat

Pumping Lemma: Survival Guide

- Definition des Pumping Lemma aufschreiben
- Beispielwort aus der Sprache wählen
- Beispielzerlegung erstellen
- Durch Aufpumpen der Beispielzerlegung ein Wort erzeugt, dass die Regeln der Sprache bricht
- Was muss die Zerlegung enthalten, damit Aufpumpen die Regeln der Sprache bricht?
- Zeigen dass ab einer bestimmten Wortlänge, jede Zerlegung mit einer Maximallänge einen solchen Aufpump-Regelbrecher enthält

Formale Sprachen: Lernziele

- Wissen
  - Definition Sprache
  - Reguläre Ausdrücke
  - Kontextfreie Grammatiken
  - Ableitung
  - Chomsky Hierarchie
- Können
  - Umgang mit formalen Definitionen und Regeln
  - Verwenden von Hilfskonstrukten
  - Keller und Rekursion

Wiederholungssession 13.12

- Online im BBB Raum
- 8:30 – 10:00 Marco
- 10:15 – 11:45 Markus
  Wiederholungsübung im Moodle
  Fragen für die Wiederholungssession könnt ihr ebenfalls im Moodle stellen