2.4 Topdown Parsing

Idee:

- Benutze den Item-Kellerautomaten.
- Benutze die nächsten k Zeichen, um die Regeln für die Expansionen zu bestimmen ;-)
- Eine Grammatik heißt $LL(k)$, falls dies immer eindeutig möglich ist.

2.4 Topdown Parsing

Idee:

- Benutze den Item-Kellerautomaten.
- Benutze die nächsten k Zeichen, um die Regeln für die Expansionen zu bestimmen ;-)
- Eine Grammatik heißt $LL(k)$, falls dies immer eindeutig möglich ist.

Wir definieren:

Eine reduzierte Grammatik heißt dann $LL(k)$, falls für je zwei verschiedene Regeln $A \to \alpha$, $A \to \alpha' \in P$ und jede Ableitung $S \to_L^* u\, A\, \beta$ mit $u \in T^*$ gilt:

$$\mathsf{First}_k(\alpha\,\beta) \cap \mathsf{First}_k(\alpha'\,\beta) = \emptyset$$

Beispiel 1:

$$
\begin{array}{lcl}
S & \rightarrow & \mathsf{if}\ (\ E\ )\ S\ \mathsf{else}\ S \quad | \\
 & & \mathsf{while}\ (\ E\ )\ S \quad | \\
 & & E\ ; \\
E & \rightarrow & \mathsf{id}
\end{array}
$$

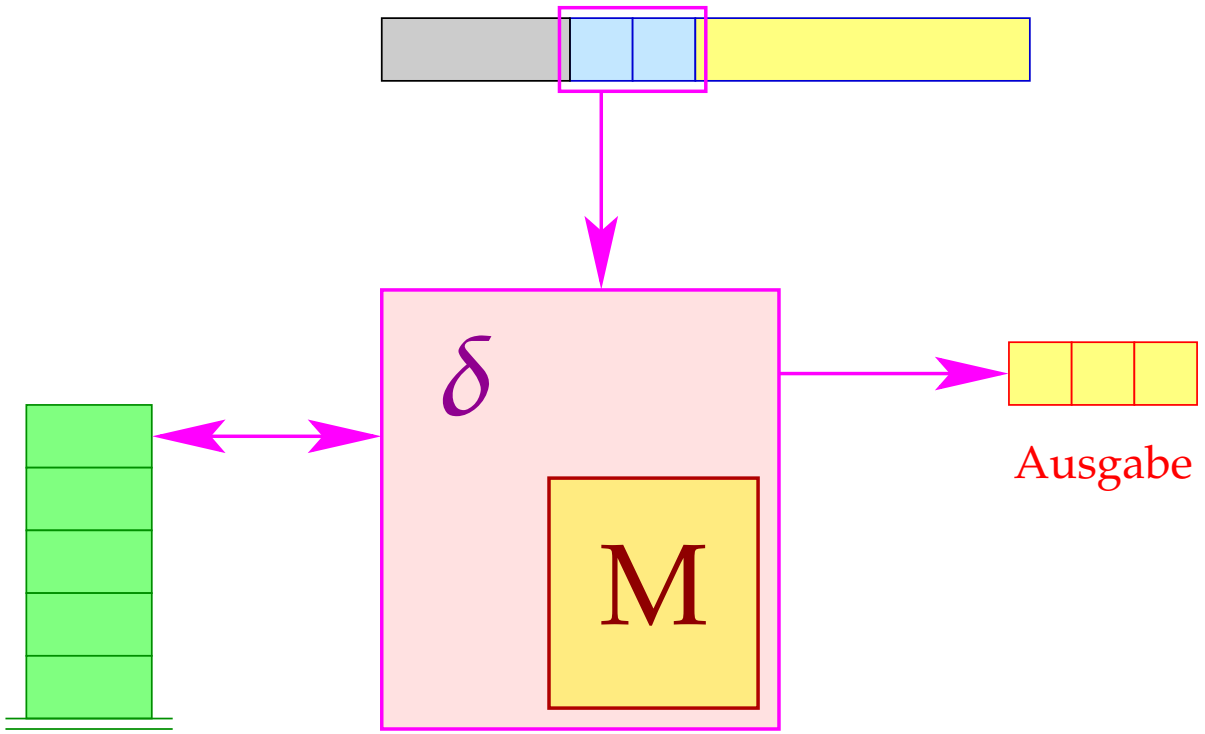
ist $LL(1)$, da $\quad \mathsf{First}_k(E) = \{\mathsf{id}\}$:-)

Beispiel 2:

$$
\begin{array}{lcl}
S & \rightarrow & \mathsf{if}\ (\ E\ )\ S\ \mathsf{else}\ S \quad | \\
 & & \mathsf{if}\ (\ E\ )\ S \quad | \\
 & & \mathsf{while}\ (\ E\ )\ S \quad | \\
 & & E\ ; \\
E & \rightarrow & \mathsf{id}
\end{array}
$$

... ist nicht $LL(k)$ für jedes $k > 0$.

Struktur des $LL(k)$-Parsers:



- Der Parser sieht ein Fenster der Länge k der Eingabe;
- er realisiert im Wesentlichen den Item-Kellerautomaten;
- die Tabelle $M[q, w]$ enthält die jeweils zuwählende Regel :-)

... im Beispiel:

$$
\begin{array}{lcl}
S & \rightarrow & \mathsf{if}\ (\ E\ )\ S\ \mathsf{else}\ S\ {}^{0} \quad | \\
 & & \mathsf{while}\ (\ E\ )\ S\ {}^{1} \quad | \\
 & & E\ ;\ {}^{2} \\
E & \rightarrow & \mathsf{id}\ {}^{0}
\end{array}
$$

Zustände: Items

Tabelle:

	if	while	id
$[\ldots \rightarrow \ldots \bullet S \ldots]$	0	1	2
$[\ldots \rightarrow \ldots \bullet E \ldots]$	–	–	0

Im Allgemeinen ...

- ist die Menge der möglichen nächsten k Zeichen gegeben durch:

$$\mathsf{First}_k(\alpha\,\beta) \;=\; \mathsf{First}_k(\alpha) \odot \mathsf{First}_k(\beta)$$

  wobei:

  (1) α die rechte Seite der passenden Regel;

  (2) β ein möglicher rechter Kontext von A ist :-)

- $\mathsf{First}_k(\beta)$ müssen wir dynamisch akkumulieren.

$\Longrightarrow$ Wir erweitern Items um Vorausschau-Mengen ...

Ein erweitertes Item ist ein Paar: $[A \to \alpha \bullet \gamma\,,\ L] \quad (A \to \alpha\,\gamma \in P,\ L \subseteq T^{\leq k})$

Die Menge L benutzen wir, um $\mathsf{First}_k(\beta)$ für den rechten Kontext β von A zu repräsentieren :-)

Ein erweitertes Item ist ein Paar: $[A \to \alpha \bullet \gamma\,,\ L] \quad (A \to \alpha\,\gamma \in P,\ L \subseteq T^{\leq k})$

Die Menge L benutzen wir, um $\mathsf{First}_k(\beta)$ für den rechten Kontext β von A zu repräsentieren :-)

Konstruktion:

Zustände: erweiterte Items

Anfangszustand: $[S' \to \bullet\ S,\ \{\epsilon\}]$

Endzustand: $[S' \to S\ \bullet,\ \{\epsilon\}]$

Übergänge:

Ein erweitertes Item ist ein Paar: $[A \to \alpha \bullet \gamma\,,\ L] \quad (A \to \alpha\,\gamma \in P,\ L \subseteq T^{\leq k})$

Die Menge L benutzen wir, um $\mathsf{First}_k(\beta)$ für den rechten Kontext β von A zu repräsentieren :-)

Konstruktion:

Zustände: erweiterte Items

Anfangszustand: $[S' \to \bullet\ S,\ \{\epsilon\}]$

Endzustand: $[S' \to S\ \bullet,\ \{\epsilon\}]$

Übergänge:

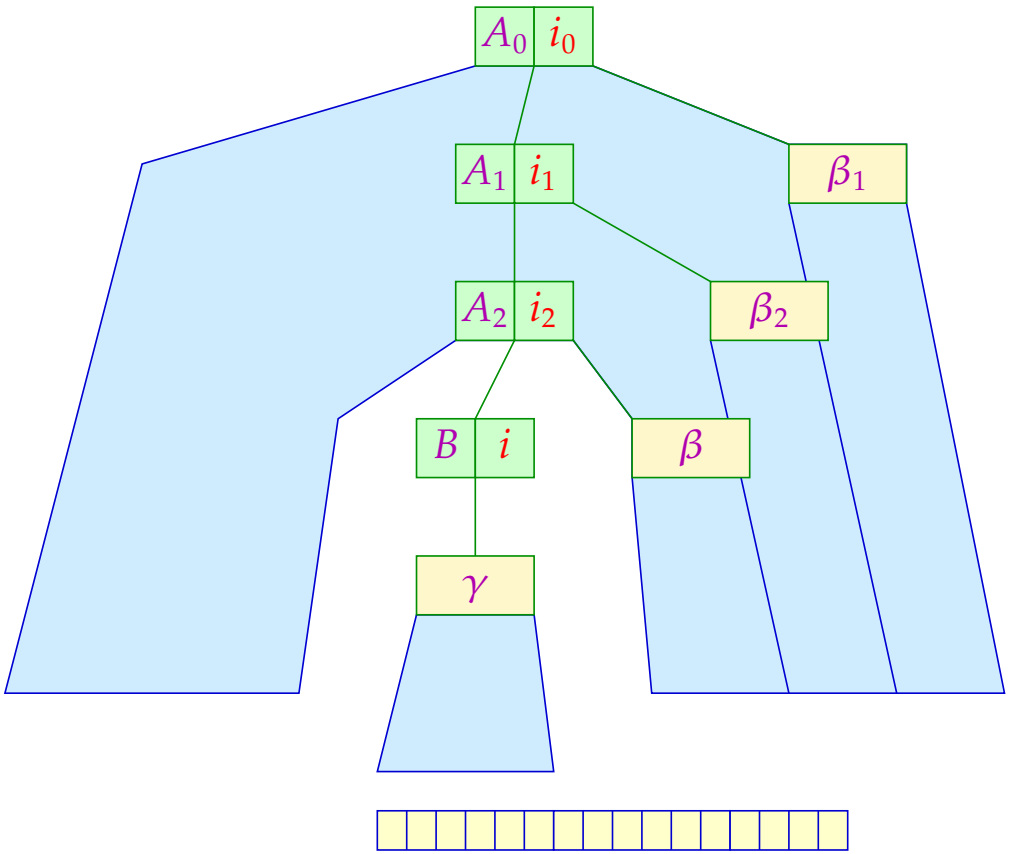
Expansionen: $([A \to \alpha \bullet B\,\beta\,,\ L], \epsilon, [A \to \alpha \bullet B\,\beta\,,\ L]\ [B \to \bullet\,\gamma\,,\ \boxed{\mathsf{First}_k(\beta) \odot L}])$

für $A \to \alpha\,B\,\beta,\ B \to \gamma \in P$

Shifts: $([A \to \alpha \bullet a\,\beta\,,\ L], a, [A \to \alpha\,a \bullet \beta\,,\ L])$ für $A \to \alpha\,a\,\beta \in P$

Reduce: $([A \to \alpha \bullet B\,\beta\,,\ L]\ [B \to \gamma\bullet\,,\ L'], \epsilon, [A \to \alpha\,B \bullet \beta\,,\ L])$ für

$A \to \alpha\,B\,\beta,\ B \to \gamma \in P$



Die Vorausschau-Tabelle:

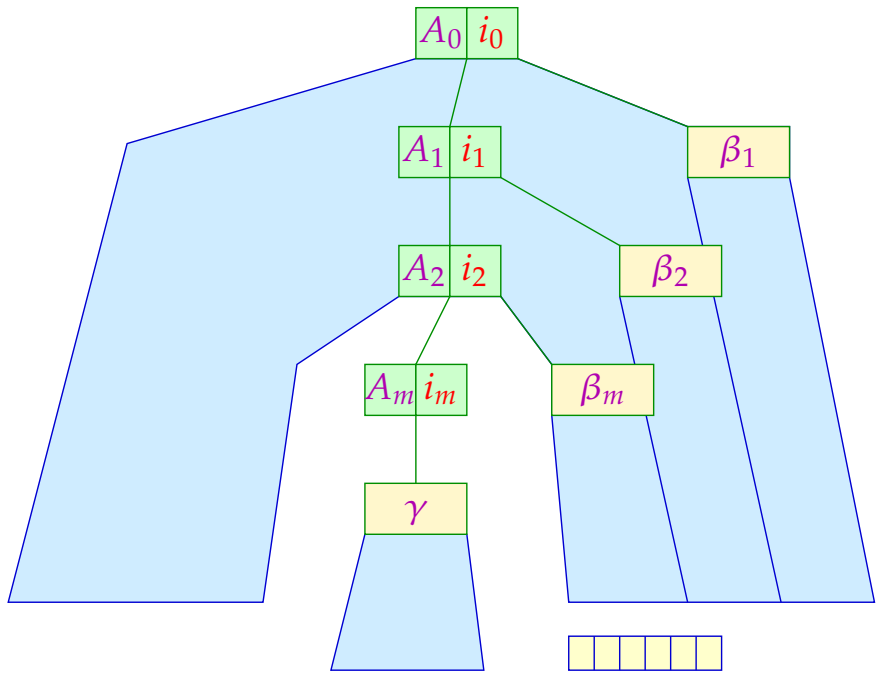
Wir setzen $M[\,[A \to \alpha \bullet B\,\beta\,,\,L],\,w] \;=\; i$ genau dann wenn (B, i) die Regel $B \to \gamma$ ist und: $w \in \mathsf{First}_k(\gamma) \odot \mathsf{First}_k(\beta) \odot L$

$$([A_0 \to \bullet\, \alpha_1\, A_1\, \beta_1,\ L_1],\ uv) \vdash^* ([A_0 \to \alpha_1 \bullet A_1\, \beta_1,\ L_1] \ldots [A_{m-1} \to \alpha_m \bullet A_m\, \beta_m,\ L_m],\ v)$$
$$\vdash^* ([A_0 \to \alpha_1\, A_1\, \beta_1 \bullet,\ L_1],\ \epsilon) \quad \ldots \quad \text{gilt genau dann wenn:}$$

(1) $\alpha_1 \ldots \alpha_m \to^* u$

(2) $A_m\, \beta_m \ldots \beta_1 \to^* v$

(3) $L_m = \mathsf{First}_k(\beta_{m-1}) \odot \ldots \odot \mathsf{First}_k(\beta_1) \odot L_1$



Satz

Die reduzierte kontextfreie Grammatik G ist $LL(k)$ genau dann wenn die k-Vorausschau-Tabelle für alle benötigten erweiterten Items wohl-definiert ist.

Diskussion:

- Der erweiterte Item-Kellerautomat zusammen mit einer k-Vorausschau-Tabelle erlaubt die deterministische Rekonstruktion einer Links-Ableitung :-)
- Die Anzahl der Vorausschau-Mengen L kann sehr groß sein :-(
- ...

Beispiel: $S \to \epsilon \mid a\,S\,b$

Die Übergänge des erweiterten Item-Kellerautomat $(k = 1)$:

0	$[S' \to \bullet\, S, \{\epsilon\}]$	ϵ	$[S' \to \bullet\, S, \{\epsilon\}]\,[S \to \bullet, \{\epsilon\}]$
1	$[S' \to \bullet\, S, \{\epsilon\}]$	ϵ	$[S' \to \bullet\, S, \{\epsilon\}]\,[S \to \bullet\, a\,S\,b, \{\epsilon\}]$
2	$[S \to \bullet\, a\,S\,b, \{\epsilon\}]$	a	$[S \to a \bullet S\,b, \{\epsilon\}]$
	$[S \to \bullet\, a\,S\,b, \{b\}]$	a	$[S \to a \bullet S\,b, \{b\}]$
3	$[S \to a \bullet S\,b, \{\epsilon\}]$	ϵ	$[S \to a \bullet S\,b, \{\epsilon\}]\,[S \to \bullet, \{b\}]$
	$[S \to a \bullet S\,b, \{b\}]$	ϵ	$[S \to a \bullet S\,b, \{b\}]\,[S \to \bullet, \{b\}]$
4	$[S \to a \bullet S\,b.\{\epsilon\}]$	ϵ	$[S \to a \bullet S\,b.\{\epsilon\}]\,[S \to \bullet\, a\,S\,b, \{b\}]$
	$[S \to a \bullet S\,b.\{b\}]$	ϵ	$[S \to a \bullet S\,b.\{b\}]\,[S \to \bullet\, a\,S\,b, \{b\}]$
5	$[S \to a \bullet S\,b, \{\epsilon\}]\,[S \to \bullet, \{b\}]$	ϵ	$[S \to a\,S \bullet b, \{\epsilon\}]$
	$[S \to a \bullet S\,b, \{b\}]\,[S \to \bullet, \{b\}]$	ϵ	$[S \to a\,S \bullet b, \{b\}]$
	...		...

	...		...
6	$[S \to a \bullet S\, b, \{\epsilon\}]\, [S \to a\, S\, b \bullet, \{b\}]$	ϵ	$[S \to a\, S \bullet b, \{\epsilon\}]$
	$[S \to a \bullet S\, b, \{b\}]\, [S \to a\, S\, b \bullet, \{b\}]$	ϵ	$[S \to a\, S \bullet b, \{b\}]$
7	$[S \to a\, S \bullet b, \{\epsilon\}]$	b	$[S \to a\, S\, b \bullet, \{\epsilon\}]$
	$[S \to a\, S \bullet b, \{b\}]$	b	$[S \to a\, S\, b \bullet, \{b\}]$
8	$[S' \to \bullet\, S, \{\epsilon\}]\, [S \to \bullet, \{\epsilon\}]$	ϵ	$[S' \to S \bullet, \{\epsilon\}]$
9	$[S' \to \bullet\, S, \{\epsilon\}]\, [S \to a\, S\, b \bullet, \{\epsilon\}]$	ϵ	$[S' \to S \bullet, \{\epsilon\}]$

Die Vorausschau-Tabelle:

	ϵ	a	b
$[S' \to \bullet\, S, \{\epsilon\}]$	0	1	–
$[S \to a \bullet S\, b, \{\epsilon\}]$	–	1	0
$[S \to a \bullet S\, b, \{b\}]$	–	1	0

Beobachtung:

- Die auszuwählende Regel hängt hier ja gar nicht von den Erweiterungen der Items ab !!!
- Unter dieser Voraussetzung können wir den Item-Kellerautomaten ohne Erweiterung benutzen :-)
- Hängt die auszuwählende Regel nur von der aktuellen Vorausschau w ab, nennen wir G auch stark $LL(k)$ ...

Beobachtung:

- Die auszuwählende Regel hängt hier ja gar nicht von den Erweiterungen der Items ab !!!
- Unter dieser Voraussetzung können wir den Item-Kellerautomaten ohne Erweiterung benutzen :-)
- Hängt die auszuwählende Regel nur von der aktuellen Vorausschau w ab, nennen wir G auch stark $LL(k)$ ...

Wir definieren:

$$\mathsf{Follow}_k(A) \;=\; \bigcup\{\mathsf{First}_k(\beta) \mid S \rightarrow_L^* u\, A\, \beta\}\;.$$

Beobachtung:

- Die auszuwählende Regel hängt hier ja gar nicht von den Erweiterungen der Items ab !!!
- Unter dieser Voraussetzung können wir den Item-Kellerautomaten ohne Erweiterung benutzen :-)
- Hängt die auszuwählende Regel nur von der aktuellen Vorausschau w ab, nennen wir G auch stark $LL(k)$ ...

Wir definieren: $\mathsf{Follow}_k(A) = \bigcup\{\mathsf{First}_k(\beta) \mid S \to_L^* u\, A\, \beta\}$.

Die reduzierte kontextfreie Grammatik G heißt stark $LL(k)$, falls für je zwei verschiedene $A \to \alpha$, $A \to \alpha' \in P$:

$$\mathsf{First}_k(\alpha) \odot \mathsf{Follow}_k(A) \cap \mathsf{First}_k(\alpha') \odot \mathsf{Follow}_k(A) = \emptyset$$

... im Beispiel:

$S \to \epsilon \quad | \quad a\,S\,b$

$$\begin{array}{lcl} \mathsf{Follow}_1(S) & = & \{\epsilon, b\} \\ \\ \mathsf{First}_1(\epsilon) \odot \mathsf{Follow}_1(S) & = & \{\epsilon\} \odot \{\epsilon, b\} = \{\epsilon, b\} \\ \mathsf{First}_1(a\,S\,b) \odot \mathsf{Follow}_1(S) & = & \{a\} \odot \{\epsilon, b\} = \{a\} \end{array}$$

Wir schließen: Die Grammatik ist in der Tat stark $LL(1)$:-)

Ist G eine starke $LL(k)$-Grammatik, können wir die Vorausschau-Tabelle statt mit (erweiterten) Items mit Nichtterminalen indizieren :-)

Wir setzen $M[B, w] = i$ genau dann wenn (B, i) die Regel $B \to \gamma$ ist und: $w \in \mathsf{First}_k(\gamma) \odot \mathsf{Follow}_k(B)$.

... im Beispiel: $S \to \epsilon \mid a\,S\,b$

	ϵ	a	b
S	0	1	0

Ist G eine starke $LL(k)$-Grammatik, können wir die Vorausschau-Tabelle statt mit (erweiterten) Items mit Nichtterminalen indizieren :-)

Wir setzen $M[B, w] = i$ genau dann wenn (B, i) die Regel $B \to \gamma$ ist und: $w \in \mathsf{First}_k(\gamma) \odot \mathsf{Follow}_k(B)$.

... im Beispiel: $S \to \epsilon \mid a\,S\,b$

	ϵ	a	b
S	0	1	0

Satz

- Jede starke $LL(k)$-Grammatik ist auch $LL(k)$:-)
- Jede $LL(1)$-Grammatik ist bereits stark $LL(1)$:-))

Beweis:

Sei G stark $LL(k)$.

Betrachte eine Ableitung $S \rightarrow_L^* u\, A\, \beta$ und Regeln $A \rightarrow \alpha\,,\ A \rightarrow \alpha' \in P$.

Dann haben wir:

$$\begin{aligned} \mathsf{First}_k(\alpha\,\beta) \cap \mathsf{First}_k(\alpha'\,\beta) &= \mathsf{First}_k(\alpha) \odot \mathsf{First}_k(\beta) \cap \mathsf{First}_k(\alpha') \odot \mathsf{First}_k(\beta) \\ &\subseteq \mathsf{First}_k(\alpha) \odot \mathsf{Follow}_k(A) \cap \mathsf{First}_k(\alpha') \odot \mathsf{Follow}_k(A) \\ &= \emptyset \end{aligned}$$

Folglich ist G auch $LL(k)$:-)

Sei G $LL(1)$.

Betrachte zwei verschiedene Regeln $A \to \alpha$, $A \to \alpha' \in P$.

Fall 1: $\epsilon \in \mathsf{First}_1(\alpha) \cap \mathsf{First}_1(\alpha')$.

Dann kann G nicht $LL(1)$ sein :-)

Sei G $LL(1)$.

Betrachte zwei verschiedene Regeln $A \to \alpha$, $A \to \alpha' \in P$.

Fall 1: $\epsilon \in \mathsf{First}_1(\alpha) \cap \mathsf{First}_1(\alpha')$.

Dann kann G nicht $LL(1)$ sein :-)

Fall 2: $\epsilon \notin \mathsf{First}_1(\alpha) \cup \mathsf{First}_1(\alpha')$.

Sei $S \to_L^* u\, A\, \beta$. Da G $LL(1)$ ist, gilt:

$$\begin{aligned}
\mathsf{First}_1(\alpha) \odot \mathsf{Follow}_1(A) \cap \mathsf{First}_1(\alpha') \odot \mathsf{Follow}_1(A) & \\
&= \mathsf{First}_1(\alpha) \cap \mathsf{First}_1(\alpha') \\
&= \mathsf{First}_1(\alpha) \odot \mathsf{First}_1(\beta) \cap \mathsf{First}_1(\alpha') \odot \mathsf{First}_1(\beta) \\
&= \emptyset
\end{aligned}$$

Fall 3: $\epsilon \in \mathsf{First}_1(\alpha)$ und $\epsilon \notin \mathsf{First}_1(\alpha')$.

Dann gilt:

$$
\begin{aligned}
&\mathsf{First}_1(\alpha) \odot \mathsf{Follow}_1(A) \cap \mathsf{First}_1(\alpha') \odot \mathsf{Follow}_1(A) \\
&\qquad = \mathsf{First}_1(\alpha) \odot \mathsf{Follow}_1(A) \cap \mathsf{First}_1(\alpha') \\
&\qquad = \mathsf{First}_1(\alpha) \odot (\bigcup\{\mathsf{First}_1(\beta) \mid S \rightarrow_L^* u\, A\, \beta\}) \cap \mathsf{First}_1(\alpha') \\
&\qquad = (\bigcup\{\mathsf{First}_1(\alpha) \odot \mathsf{First}_1(\beta) \mid S \rightarrow_L^* u\, A\, \beta\}) \cap \mathsf{First}_1(\alpha') \\
&\qquad = \bigcup\{\mathsf{First}_1(\alpha) \odot \mathsf{First}_1(\beta) \cap \mathsf{First}_1(\alpha') \mid S \rightarrow_L^* u\, A\, \beta\} \\
&\qquad = \bigcup\{\emptyset \mid S \rightarrow_L^* u\, A\, \beta\} \\
&\qquad = \emptyset
\end{aligned}
$$

Fall 3: $\epsilon \in \mathsf{First}_1(\alpha)$ und $\epsilon \notin \mathsf{First}_1(\alpha')$.

Dann gilt:

$$
\begin{aligned}
&\mathsf{First}_1(\alpha) \odot \mathsf{Follow}_1(A) \cap \mathsf{First}_1(\alpha') \odot \mathsf{Follow}_1(A) \\
&\quad = \mathsf{First}_1(\alpha) \odot \mathsf{Follow}_1(A) \cap \mathsf{First}_1(\alpha') \\
&\quad = \mathsf{First}_1(\alpha) \odot (\bigcup\{\mathsf{First}_1(\beta) \mid S \to_L^* u\, A\, \beta\}) \cap \mathsf{First}_1(\alpha') \\
&\quad = (\bigcup\{\mathsf{First}_1(\alpha) \odot \mathsf{First}_1(\beta) \mid S \to_L^* u\, A\, \beta\}) \cap \mathsf{First}_1(\alpha') \\
&\quad = \bigcup\{\mathsf{First}_1(\alpha) \odot \mathsf{First}_1(\beta) \cap \mathsf{First}_1(\alpha') \mid S \to_L^* u\, A\, \beta\} \\
&\quad = \bigcup\{\emptyset \mid S \to_L^* u\, A\, \beta\} \\
&\quad = \emptyset
\end{aligned}
$$

Fall 4: $\epsilon \notin \mathsf{First}_1(\alpha)$ und $\epsilon \in \mathsf{First}_1(\alpha')$: analog :-)

Beispiel:

$$\begin{array}{lcl} S & \rightarrow & a\,A\,a\,a^{\,0} \mid b\,A\,b\,a^{\,1} \\ A & \rightarrow & b^{\,0} \mid \epsilon^{\,1} \end{array}$$

Offenbar ist die Grammatik $LL(2)$:-) Andererseits gilt:

$$\begin{array}{rcl} \mathsf{First}_2(b) \odot \mathsf{Follow}_2(A) \cap \mathsf{First}_2(\epsilon) \odot \mathsf{Follow}_2(A) & & \\ = & \{b\} \odot \{a\,a, b\,a\} \cap \{\epsilon\} \odot \{a\,a, b\,a\} & \\ = & \{b\,a, b\,b\} \cap \{a\,a, b\,a\} & \\ \neq & \emptyset & \end{array}$$

Folglich ist die Grammatik nicht stark $LL(2)$:-(

Beispiel:

$$\begin{aligned} S &\rightarrow a\,A\,a\,a^{\,0} \mid b\,A\,b\,a^{\,1} \\ A &\rightarrow b^{\,0} \mid \epsilon^{\,1} \end{aligned}$$

Offenbar ist die Grammatik $LL(2)$:-) Andererseits gilt:

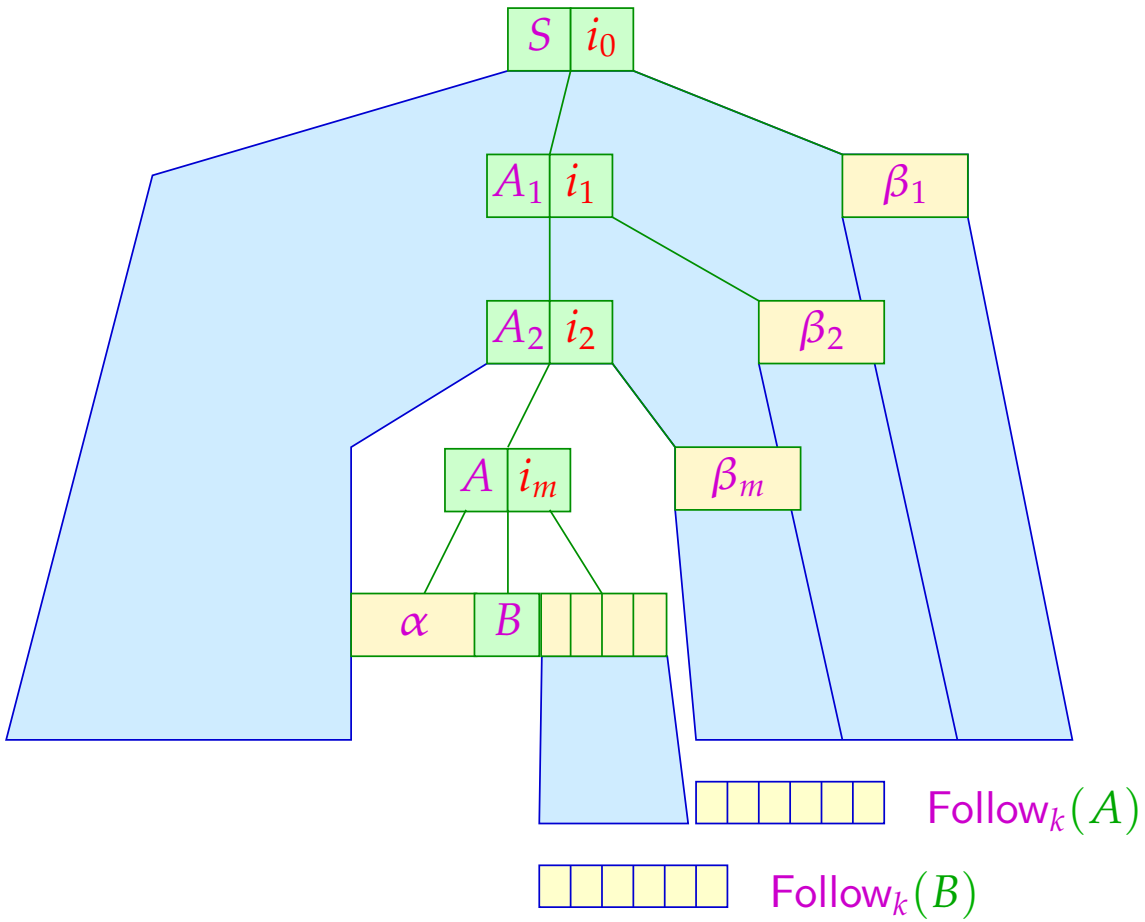
$$\begin{aligned} \mathsf{First}_2(b) \odot \mathsf{Follow}_2(A) \cap \mathsf{First}_2(\epsilon) \odot \mathsf{Follow}_2(A) & \\ = \{b\} \odot \{a\,a, b\,a\} \cap \{\epsilon\} \odot \{a\,a, b\,a\} & \\ = \{b\,a, b\,b\} \cap \{a\,a, b\,a\} & \\ \neq \emptyset & \end{aligned}$$

Folglich ist die Grammatik nicht stark $LL(2)$:-(

Wir schließen:

- Für $k > 1$ ist nicht jede $LL(k)$-Grammatik automatisch stark $LL(k)$.
- Zu jeder $LL(k)$-Grammatik kann jedoch eine äquivalente starke $LL(k)$-Grammatik konstruiert werden $\Longrightarrow$ Übung!

Berechnung von $\mathsf{Follow}_k(B)$:



Berechnung von $\mathsf{Follow}_k(B)$:

Idee:

- Wir stellen ein Ungleichungssystem auf :-)
- ϵ ist ein möglicher rechter Kontext von S :-)
- Mögliche rechte Kontexte der linken Seite einer Regel propagieren wir ans Ende jeder rechten Seite ...

... im Beispiel: $S \to \epsilon \mid a\,S\,b$

$$\begin{array}{lcl} \mathsf{Follow}_k(S) & \supseteq & \{\epsilon\} \\ \mathsf{Follow}_k(S) & \supseteq & \{b\} \odot \mathsf{Follow}_k(S) \end{array}$$