Earley für TAGs

- Earley 1970 für kontextfreie Grammatiken

- Schabes 1988 für TAGs
  allgemein: $O(n^6)$
  mit korrekter-Präfix-Eigenschaft: $O(n^9)$

- Nederhof 1997 für TAGs
  mit korrekter-Präfix-Eigenschaft: $O(n^6)$

Korrekte Präfix-Eigenschaft

- Ein Parsingalgorithmus hat die
  korrekte Präfix-Eigenschaft
  wenn es nach Einlesen (ohne Reject)
  von $x_1...x_i$
  eine Verlängerung $x_{i+1}...x_k$ gibt,
  so daß
  $x_1..x_k$ akzeptiert werden muß.

- *Der Algorithmus bricht so früh wie möglich ab.*

Top-Down vs. Bottom-Up Parsing

- **Bottom-up:**

- Ausgehend von den Terminalen werden kleine Teile zu größeren zusammengesetzt.
  Die Regeln werden „*rückwärts*“ angewandt.

- **Top-down:**

- Ausgehend vom Startsymbol wird versucht, Regeln so anzuwenden, daß die gefundenen Terminale produziert werden.
  Die Regeln werden „*vorwärts*“ angewandt.

Der Earley-Algorithmus

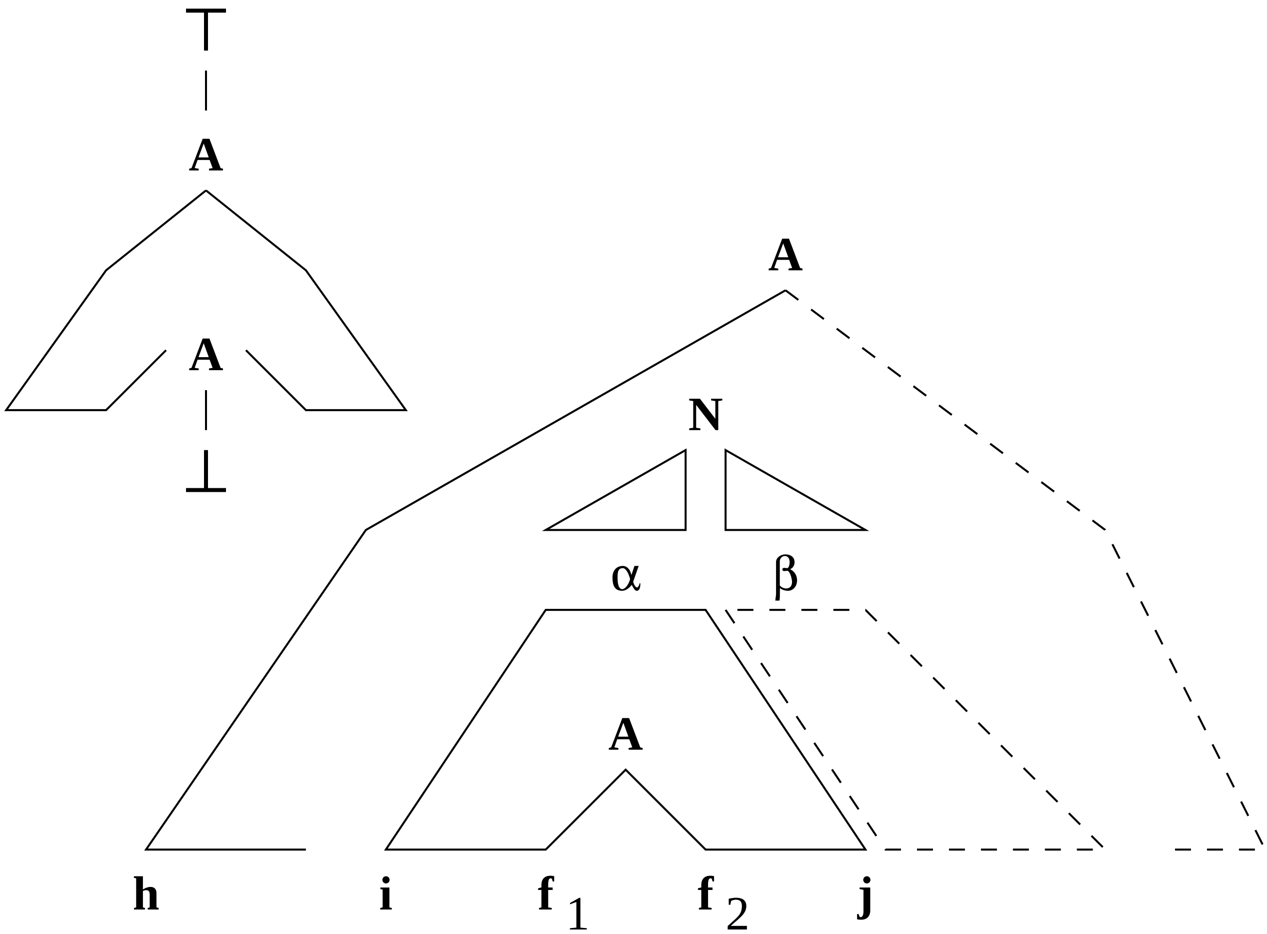
- Idee: **gepunktete Regeln**
  $A \to cB \bullet DfE$
  „Alles bis zum $\bullet$ ist abgeleitet...“

- **Items**:
  $[A \to cB \bullet DfE, i, j]$
  „...und deckt $x_{i+1}...x_j$ ab.“

Earley für TAGs

- Items für TAGs:

  $[h, N \rightarrow \alpha \bullet \beta, i, j, f_1, f_2]$



Earley für TAGs

- 5 Klassen von Schritten:
  *Inital*schritte
  Scanner
  Predictor
  Completor
  Adjunctor (neu für TAGs!)

Earley für TAGs

- **Initalisierung:**

- $\alpha \in I$
  $\vdash$
  $[0, \top \rightarrow \bullet R_\alpha, 0, 0, -, -]$

- *Jede Ableitung beginnt mit der Wurzel eines Initialbaums*

Earley für TAGs

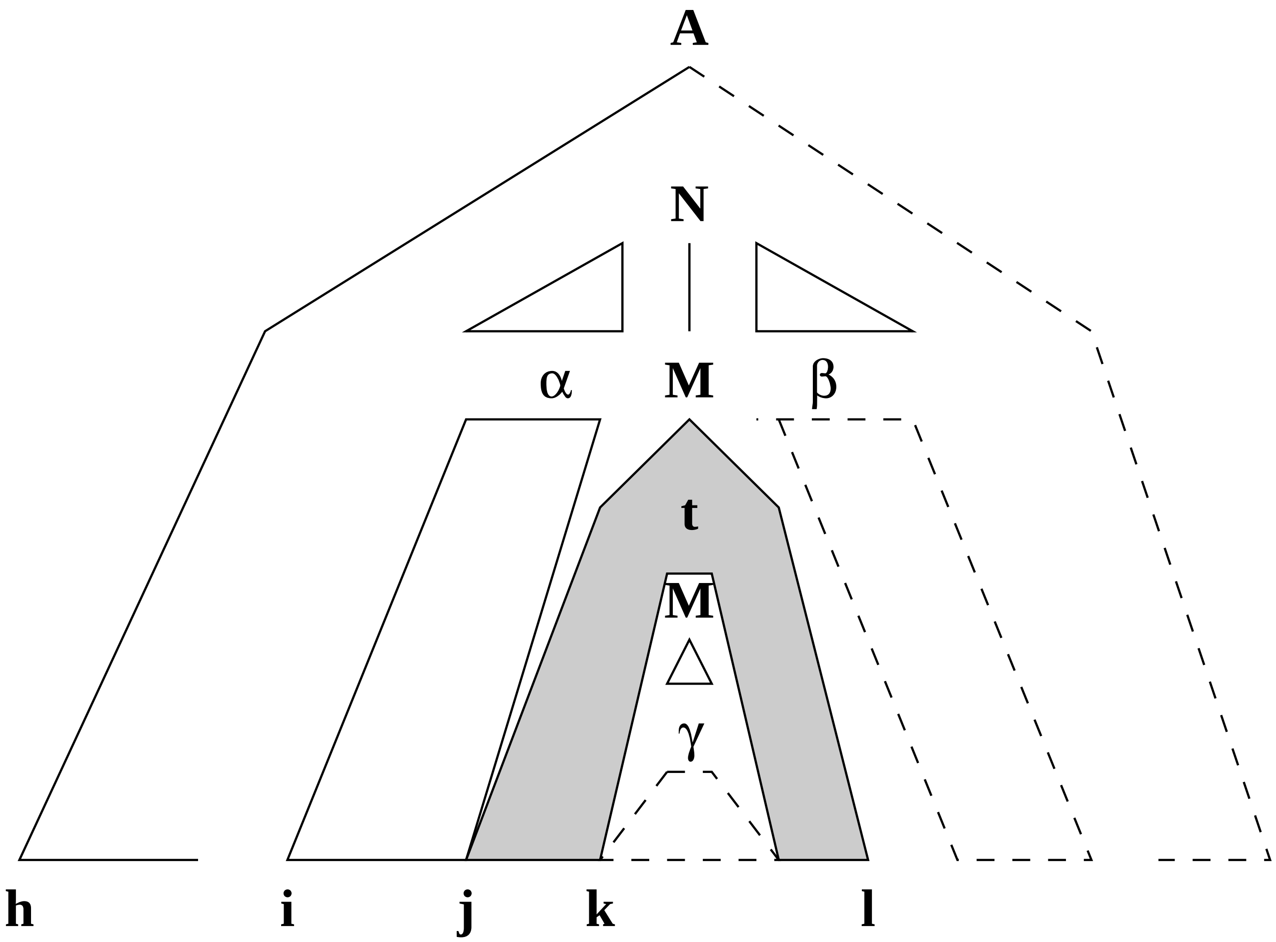
- **Scanner**

- Scan1:
  $[h, N \rightarrow \alpha \bullet x_{j+1}\beta, i, j, f_1, f_2]$,
  $\vdash$
  $[h, N \rightarrow \alpha x_{j+1} \bullet \beta, i, j+1, f_1, f_2]$,

- Scan2:
  $[h, N \rightarrow \alpha \bullet \epsilon\beta, i, j, f_1, f_2]$,
  $\vdash$
  $[h, N \rightarrow \alpha\epsilon \bullet \beta, i, j, f_1, f_2]$,

Earley für TAGs

- Adjunktion in Predictor und Completor



Earley für TAGs

- **Predictor**

- Pred1:
  $[h, N \rightarrow \alpha \bullet M\beta, i, j, f_1, f_2]$,
  $t \in \textit{Adjunct}(M)$
  $\vdash$
  $[j, \top \rightarrow \bullet R_t, j, j, -, -]$

- Pred2:
  $[h, N \rightarrow \alpha \bullet M\beta, i, j, f_1, f_2]$,
  $\textit{Children}(M) = \gamma$, kein OA an M
  $\vdash$
  $[h, M \rightarrow \bullet\gamma, j, j, -, -]$,

- Pred3:
  $[h, N \rightarrow \alpha \bullet M\beta, i, j, f_1, f_2]$,
  $[j, F_t \rightarrow \perp, k, k, -, -]$,
  $t \in \textit{Adjunct}(M), \textit{Children}(M) = \gamma$
  $\vdash$
  $[h, M \rightarrow \bullet\gamma, k, k, -, -]$

Earley für TAGs

- **Completor**

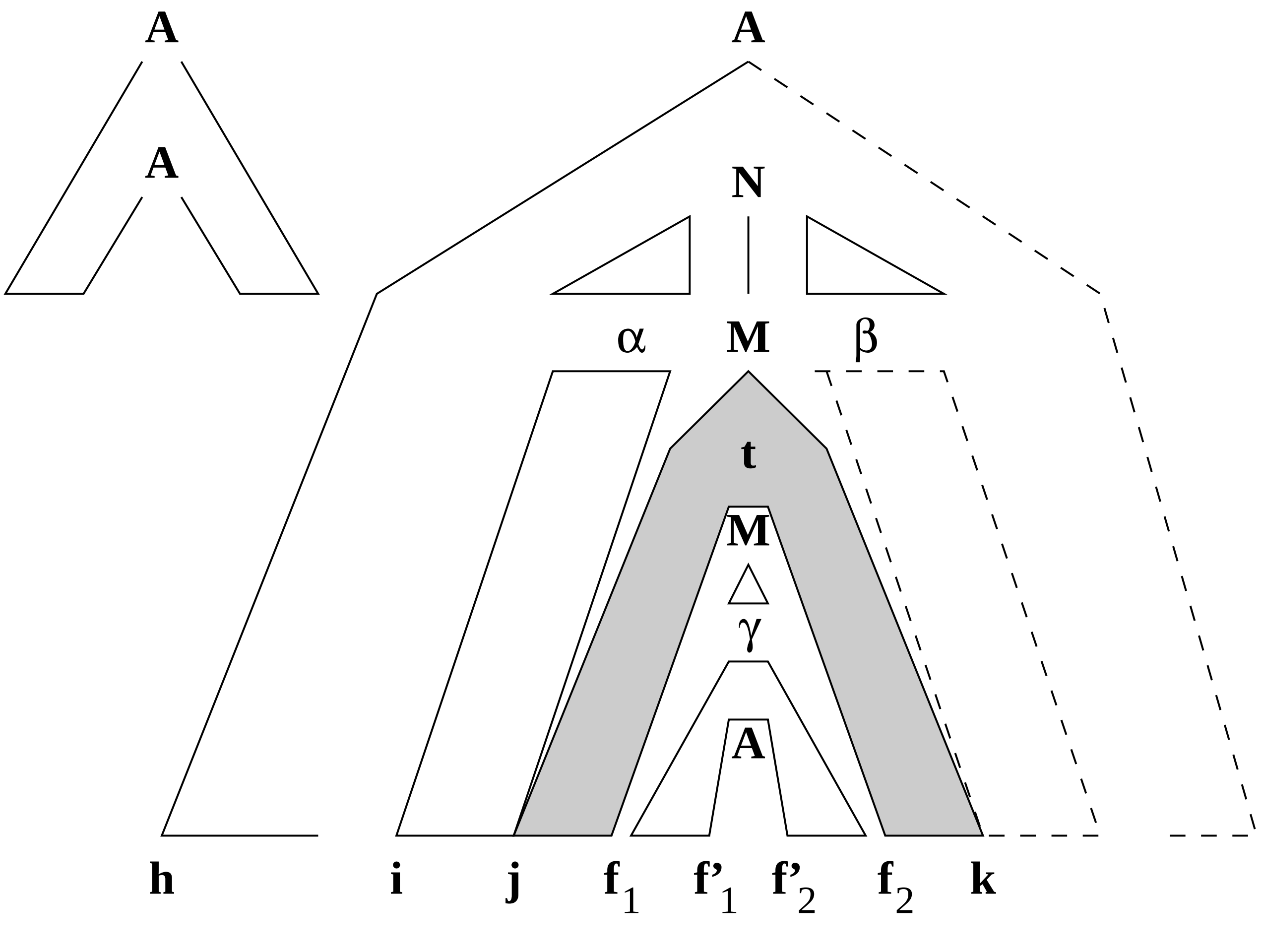
- Comp1:
  $[h, N \to \alpha \bullet M\beta, i, j, f_1, f_2]$,
  $[h, M \to \gamma\bullet, k, l, f_1', f_2'], t \in \mathit{Adjunct}(M)$,
  $[j, F_t \to \bullet\perp, k, k, -, -]$
  $\vdash$
  $[j, F_t \to \perp\bullet, k, l, k, l]$

- Comp2:
  $[h, N \to \alpha \bullet M\beta, i, j, -, -]$,
  M dominiert den Fußknoten,
  $[h, M \to \gamma\bullet, k, l, f_1, f_2]$
  $\vdash$
  $[h, N \to \alpha M \bullet \beta, i, j, f_1, f_2]$

- Comp3:
  $[h, N \to \alpha \bullet M\beta, i, j, f_1, f_2]$,
  M dominiert den Fußknoten nicht,
  $[h, M \to \gamma\bullet, j, k, -, -]$
  $\vdash$
  $[h, N \to \alpha M \bullet \beta, i, j, f_1, f_2]$

Earley für TAGs

- Adjunctor



Earley für TAGs

- **Adjunctor**

- Adj0:
  $[j, \top \to R_t\bullet, j, k, f_1, f_2]$,
  $[h, M \to \gamma\bullet, f_1, f_2, f_1', f_2']$,
  $t \in \mathit{Adjunct}(M)$
  $\vdash$
  $[M \to \gamma\bullet, j, k, f_1', f_2']$ neuartiges Item!

- Adj1:
  $[M \to \gamma\bullet, j, k, f_1', f_2']$,
  M dominiert Fußknoten von t',
  $[h, F_{t'} \to \bot\bullet, f_1', f_2', f_1', f_2']$,
  $[h, N \to \alpha \bullet M\beta, i, j, -, -]$
  $\vdash$
  $[h, N \to \alpha M \bullet \beta, i, k, f_1', f_2']$

- Adj2:
  $[M \to \gamma\bullet, j, k, -, -]$,
  $[h, N \to \alpha \bullet M\beta, i, j, f_1', f_2']$
  $\vdash$
  $[h, N \to \alpha M \bullet \beta, i, k, f_1', f_2']$

Earley für TAGs

- **Zeitkomplexität**

- grobe Abschätzung für Schritt **Comp1:**
  $[h, N \to \alpha \bullet M\beta, i, j, f_1, f_2]$,
  $t \in \mathit{Adjunct}(M)$,
  $[h, M \to \gamma\bullet, k, l, f_1', f_2']$,
  $[j, F_t \to \bullet\perp, k, k, -, -]$
  $\vdash$
  $[j, F_t \to \perp\bullet, k, l, k, l]$

  9 Variablen $\Rightarrow O(n^9)$

- bessere Abschätzung:
  Variablen i, f_1, f_2, f_1', f_2' erscheinen nur einmal und können eliminiert werden
  $\Rightarrow O(n^4)$

Earley für TAGs

- Zeitkomplexität:
  $O(n^6)$ Eingabelänge,
  $O(|G|^2)$ Grammatikgröße

- Vorteil:
  korrekte Präfix-Eigenschaft,
  keine Normalform notwendig