Tree Transducers

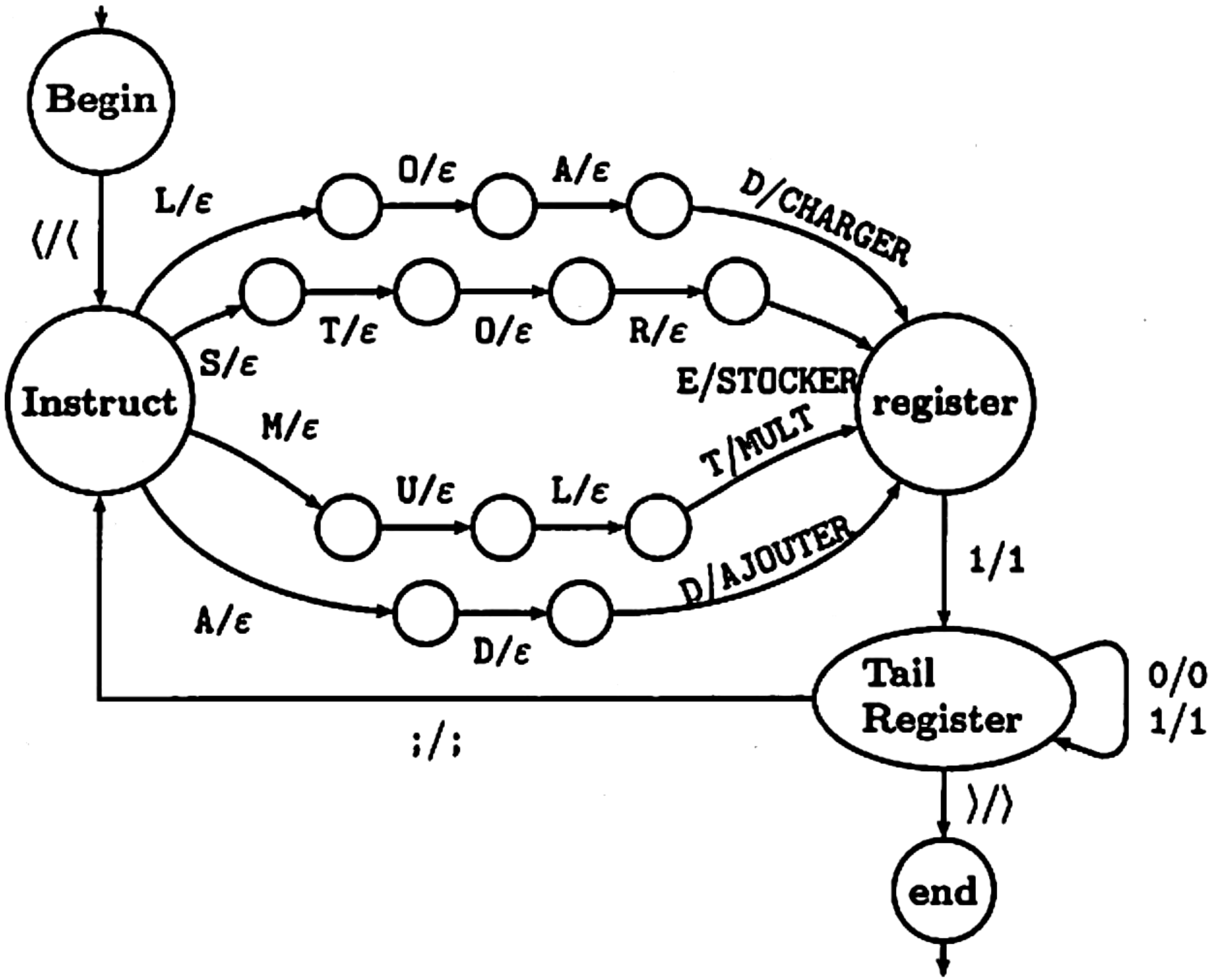
- Transformation system for trees

  Ex.: Conversion of LaTeX to HTML

  - Bottom-up tree transducer
  - Top-down tree transducer
  - Homomorphic approach

String (word) case

- **Rational Transducers**



- **Rational Transducers:** $R = (Q, \mathcal{F}, \mathcal{F}', Q_i, Q_f, \Delta)$
  - $\mathcal{F}$ ($\mathcal{F}'$): **Set of finite input (output) symbols**
  - Q_i (Q_f): **Set of finite initial (final) states**
  - Δ : **Set of rules** $q \overset{f/m}{\to} q'$,

    where, $f \in \mathcal{F} \cup \{\varepsilon\}$, $m \in \mathcal{F}'^*$, $q, q' \in Q$
- **Transition relation** $\to_R$ ($\subseteq \mathcal{F}^* \times Q \times \mathcal{F}'^*$)

  $(ft, q, u) \to_R (t, q', um)$ **for** $q \overset{f/m}{\to} q' \in \Delta$
- **Relation** T_R **recognized by** R:

  $T_R = \{(t, u) \mid (t, q, \varepsilon) \to_R^* (\varepsilon, q', u), q \in Q_i, q' \in Q_f\}$

- **Definition by two homomorphisms:**
  - $B = (\Phi, L, \Psi)$, **where**

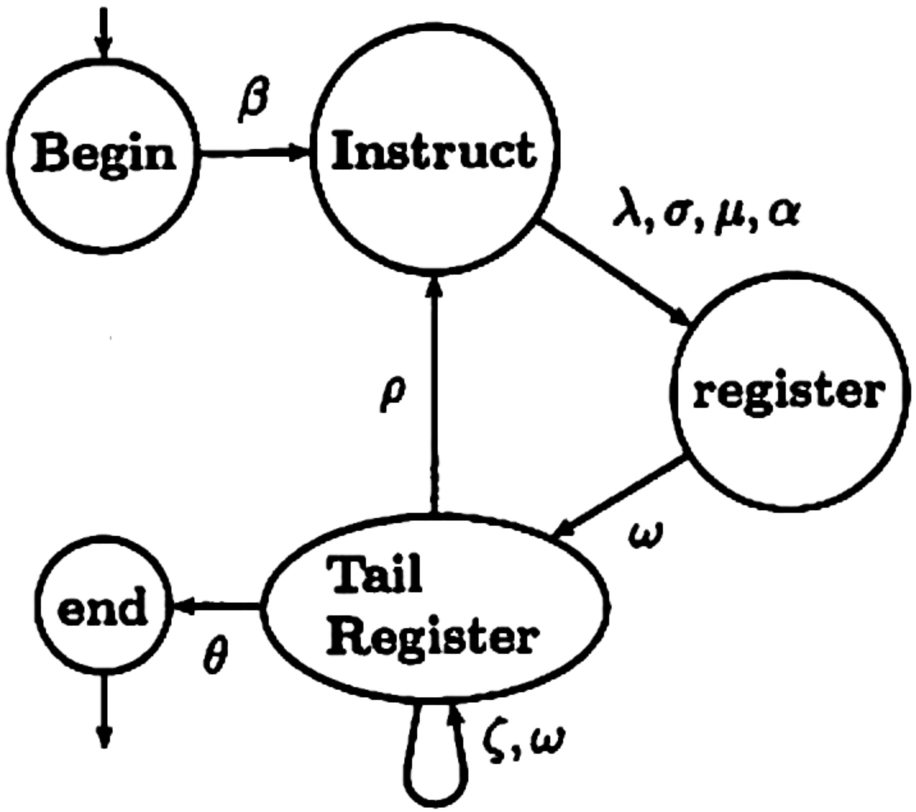
    $\Phi : \mathcal{F}''^{*} \to \mathcal{F}^{*}$

    $\Psi : \mathcal{F}''^{*} \to \mathcal{F}'^{*}$

    L **($\subseteq \mathcal{F}''^{*}$): regular language**
  - $T_B = \{(\Phi(w), \Psi(w)) \mid w \in L\}$
- ε**-free: no** $a \in \mathcal{F}''$ **such that** $\Phi(a) = \varepsilon$

• Ex.: Transducer defined by homomorphisms:

$$\begin{array}{llll}
\Phi(\beta) = \langle & \Phi(\lambda) = \mathtt{LOAD} & \Phi(\sigma) = \mathtt{STORE} & \Phi(\mu) = \mathtt{MULT} \\
\Phi(\alpha) = \mathtt{ADD} & \Phi(\rho) =; & \Phi(\omega) = 1 & \Phi(\zeta) = 0 \\
\Phi(\theta) =\rangle & & & \\
\Psi(\beta) = \langle & \Psi(\lambda) = \mathtt{CHARGER} & \Psi(\sigma) = \mathtt{STOCKER} & \Psi(\mu) = \mathtt{MULT} \\
\Psi(\alpha) = \mathtt{ADD} & \Psi(\rho) =; & \Psi(\omega) = 1 & \Psi(\zeta) = 0 \\
\Psi(\theta) =\rangle & & &
\end{array}$$



Tree transducers

- Ex.: Transform syntax tree of $(a+b)\times c$ to $\times(+(a,b),c)$
- A grammar for arithmetic not suitable for parsing

$$\begin{array}{rclcrcl} E & \to & M \mid M+E & \quad & M & \to & F \mid F\times M \\ F & \to & I \mid (E) & \quad & I & \to & a \mid b \mid \cdots \mid z \end{array}$$

- A LL(1)-parsable grammar G

$$\begin{array}{rclcrcl} E & \to & ME' & \quad & E' & \to & +E \mid \varepsilon \\ M & \to & FM' & \quad & M' & \to & \times M \mid \varepsilon \\ F & \to & I \mid (E) & \quad & I & \to & a \mid b \mid \cdots \mid z \end{array}$$

- **Bottom-up tree transducer NUTT:**

  $A = (Q, \mathcal{F}, \mathcal{F}', Q^f, \Delta)$, **where Δ is rules**

  $f(q_1(x_1), \ldots, q_n(x_n)) \to q(u)$

  for $f \in \mathcal{F}, u \in \mathsf{T}(\mathcal{F}', \mathcal{X}_n)$

  - **Transition function A:**

    $$s \to_A t \quad \stackrel{\text{def}}{\Longleftrightarrow}$$

    $$\exists C, \sigma, l \to r \in \Delta.\ s = C[l\sigma] \wedge t = C[r\sigma]$$

  - **Conversion: for $T \subseteq \mathsf{T}(\mathcal{F})$,**

    $$A(T) = \{s \in \mathsf{T}(\mathcal{F}') \mid t \in T,\ t \to_A^* q(s),\ q \in Q^f\}$$

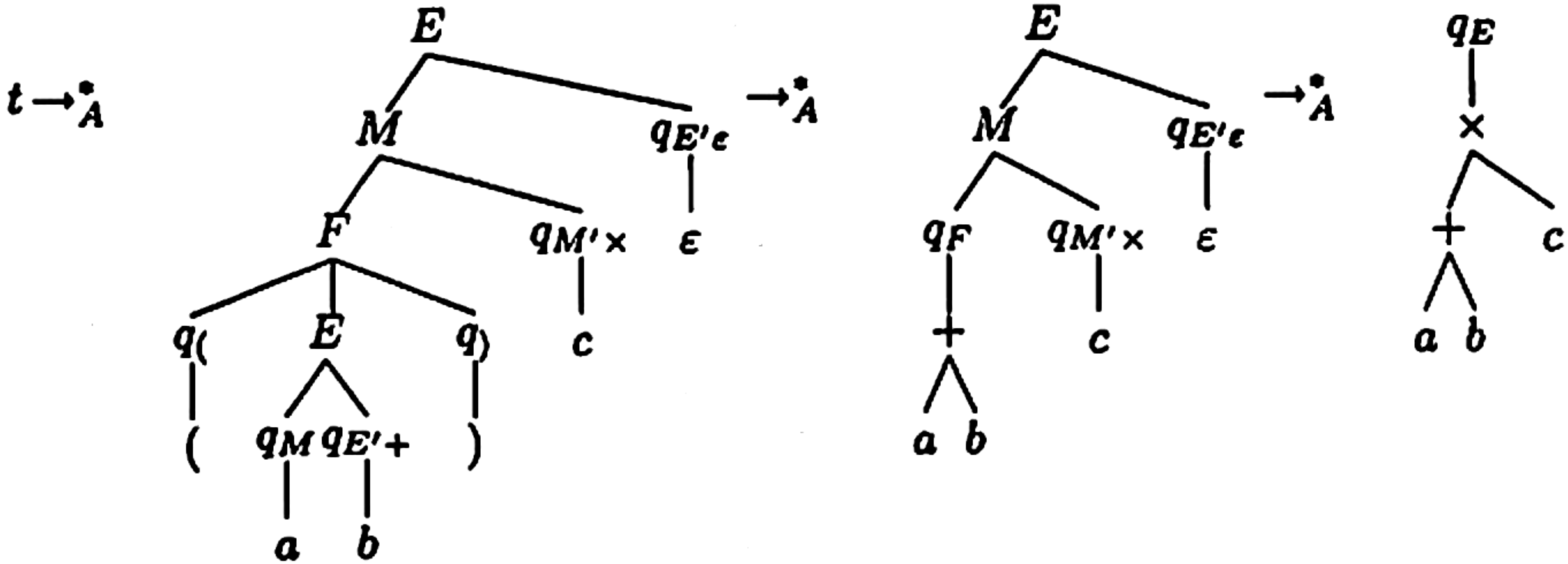
  - **linear: linear variables in RHS**
  - **non-erasing: RHS has $\mathcal{F}'$ symbols**
  - **non-deleting: LHS-varible appears in RHS**

- **Ex.: the final state q_E**

$$
\begin{aligned}
a &\rightarrow q(a) \\
c &\rightarrow q(c) \\
) &\rightarrow q_{)}()) \\
+ &\rightarrow q_{+}(+) \\
I(q(x)) &\rightarrow q_I(x) \\
M'(q_\varepsilon(x)) &\rightarrow q_{M'\varepsilon}(x) \\
M(q_F(x), q_{M'\varepsilon(y)}) &\rightarrow q_M(x) \\
M'(q_\times(x), q_M(y)) &\rightarrow q_{M'\times}(y) \\
E'(q_+(x), q_E(y)) &\rightarrow q_{E'+}(y) \\
F(q_{(}(x), q_E(y), q_{)}(z)) &\rightarrow q_F(y)
\end{aligned}
$$

$$
\begin{aligned}
b &\rightarrow q(b) \\
\varepsilon &\rightarrow q_\varepsilon(\varepsilon) \\
(&\rightarrow q_{(}(() \\
\times &\rightarrow q_\times(\times) \\
F(q_I(x)) &\rightarrow q_F(x) \\
E'(q_\varepsilon(x)) &\rightarrow q_{E'\varepsilon}(x) \\
E(q_M(x), q_{E'\varepsilon}(y)) &\rightarrow q_E(x) \\
M(q_F(x), q_{M'\times}(y)) &\rightarrow q_M(\times(x, y)) \\
E(q_M(x), q_{E'+}(y)) &\rightarrow q_E(+(x, y))
\end{aligned}
$$

- **NUTT (cont.)**
  t **is the** G**-syntax tree of** $(a + b) \times c$.



- **Ex.** U_1**: non-erasing and non-deleting**
$\mathcal{F} = \{f(), a\}$, $\mathcal{F}' = \{g(,), f(), f'(), a\}$, $q' \in Q^f$

$$
\begin{array}{ll}
a \rightarrow q(a) & f(q(x)) \rightarrow q(f(x)) \\
f(q(x)) \rightarrow q(f'(x)) & f(q(x)) \rightarrow q'(g(x,x))
\end{array}
$$

$U_1(\{f(f(f(a)))\})$ **is the set of**

$$
g(f(f(a)), f(f(a))),\ g(f(f'(a)), f(f'(a))),
$$
$$
g(f'(f(a)), f'(f(a))),\ g(f'(f'(a)), f'(f'(a)))
$$

- **Top-down tree transducer: NDTT**

  $A = (Q, \mathcal{F}, \mathcal{F}', Q^i, \Delta)$, **where Δ is rules**

  $$q(f(x_1, \ldots, x_n)) \to C[q_1(x_{i_1}), \ldots, q_p(x_{i_p})]$$

  $$f \in \mathcal{F}, C \in \mathsf{T}(\mathcal{F}', \mathcal{X}_p), x_{i_1}, \ldots, x_{i_p} \in \mathcal{X}_n$$

- **Transition function A:**

  $$s \to_A t \quad \stackrel{\text{def}}{\Longleftrightarrow}$$

  $$\exists C, \sigma, l \to r \in \Delta.\ s = C[l\sigma] \wedge t = C[r\sigma]$$

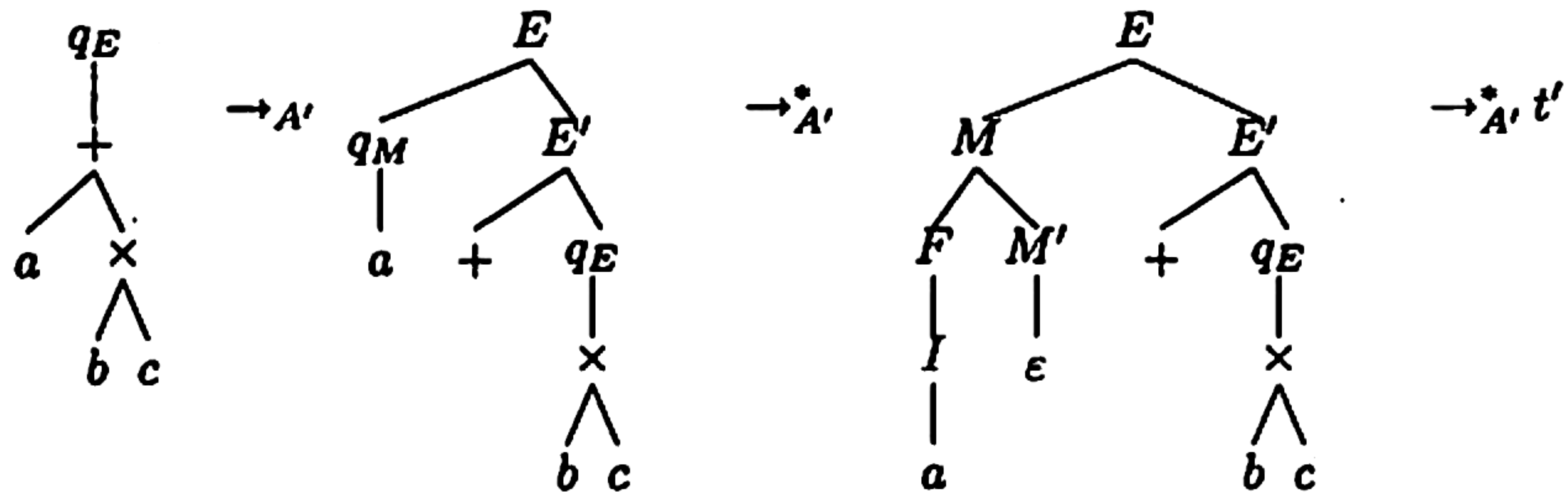
- **Conversion: for $T \subseteq \mathsf{T}(\mathcal{F})$,**

  $$A(T) = \{s \in \mathsf{T}(\mathcal{F}') \mid t \in T,\ q \in Q^i,\ q(t) \to_A^* s\}$$

- **Ex.: with an initial state** q_E

$$
\begin{array}{rlrl}
q_E(x) & \to E(q_M(x), E'(\varepsilon)) & q_E(+(x,y)) & \to E(q_M(x), E'(+, q_E(y))) \\
q_M(x) & \to M(q_F(x), M'(\varepsilon)) & q_M(\times(x,y)) & \to M(q_F(x), M'(\times, q_M(y))) \\
q_F(x) & \to F((, q_E(x),)) & q_F(a) & \to F(I(a)) \\
q_F(b) & \to F(I(b)) & q_F(c) & \to F(I(c))
\end{array}
$$

- **NDTT (cont.)**



For tree $s \in \mathsf{T}(\mathcal{F}')$, the tree $t' \in \mathsf{T}(\mathcal{F})$ is obtained by

$$q_E(s) \to^*_{A'} t',$$

where t' is the G-syntax tree of $a + b \times c$.

- **Ex.** D_1**:**
  $\mathcal{F} = \{f(), a\}$, $\mathcal{F}' = \{g(,), f(), f'(), a\}$, $q \in Q^i$

$$
\begin{array}{ll}
q(f(x)) \to g(q'(x), q'(x)) & q'(f(x)) \to f(q'(x)) \\
q'(a) \to a & q'(f(x)) \to f'(q'(x))
\end{array}
$$

$D_1(\{f(f(f(a)))\})$ **is the set of 16 trees**

$$
\begin{array}{l}
g(f(f(a)), f(f(a))),\ \ g(f(f(a)), f(f'(a))), \\
g(f(f(a)), f'(f(a))),\ \ g(f(f(a)), f'(f'(a))), \\
\quad \vdots \\
g(f'(f'(a)), f'(f(a))),\ \ g(f'(f'(a)), f'(f'(a)))
\end{array}
$$

- Relation between classes
  - NUTT $\neq$ NDTT (e.g. U_1 and D_1)
  - linear NDTT $\subseteq$ linear NUTT
  - NDTT $=$ NUTT on linear and non-deleting
- Closure property by composition
  - Not closed: NUTT, NDTT
  - Closed: linear NUTT, determ. NUTT
  - Composition of determ. NDTTs has an equivalent composition of a non-deleting NDTT and a transducer of linear homomorphism
- Domain of tree transducers are regular
- Regularity is preserved by linear transducer

- **Tree transducer by homomorphism: similar to string case**
  - $B = (\Phi, L, \Psi)$, **where**

    $\Phi : \mathsf{T}(\mathcal{F}'') \rightarrow \mathsf{T}(\mathcal{F})$

    $\Psi : \mathsf{T}(\mathcal{F}'') \rightarrow \mathsf{T}(\mathcal{F}')$

    L **(**$\subseteq \mathsf{T}(\mathcal{F}'')$**): regular tree language**
- **NUTTs** U **has the same power as tree transducers** $B = (\Phi,\ L,\ \Psi)$ **by homomorphism**
  - U **is linear** $\iff$ Ψ **is linear**
  - U **is non-deleting** $\iff$ Ψ **in non-deleting**
  - U **is** ε**-free** $\iff$ Ψ **is** ε**-free**