Chomsky-Schützenberger-
parsing is
Chart parsing

Thomas Ruprecht

Freitagsseminar 2018-06-01

Prerequisites: The C-S-Theorem (unweighted, CF)

Language L, t.f.a.e.

- L is CF
- $\exists$ string homomorphism h, regular language R, Dyck language D s.t.

$$L = h(R \cap D)$$

$$w \in L \iff \exists u \in R \cap D : h(u) = w$$

$$\iff \exists u \in R \cap D : u \in h^{-1}(w)$$

$$\iff \exists u \in R \cap h^{-1}(w) \cap D$$

cf. Chomsky & Schützenberger, 1963

Prerequisites: C-S parsing (unweighted, CF)

$$w \in L \iff \exists u \in R \cap h^{-1}(w) : u \in D$$

- construct fsas for R, $h^{-1}(w)$
- product construction of fsas
- extract Dyck words
- read off derivation trees of some CFG

Hulden, 2011

Prerequisites:

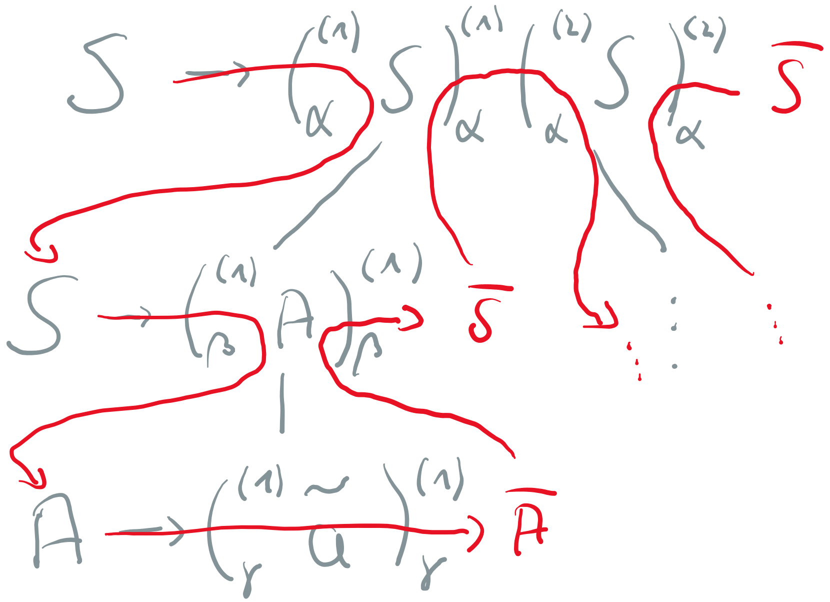
- construct fsas for R, $h^{-1}(w)$

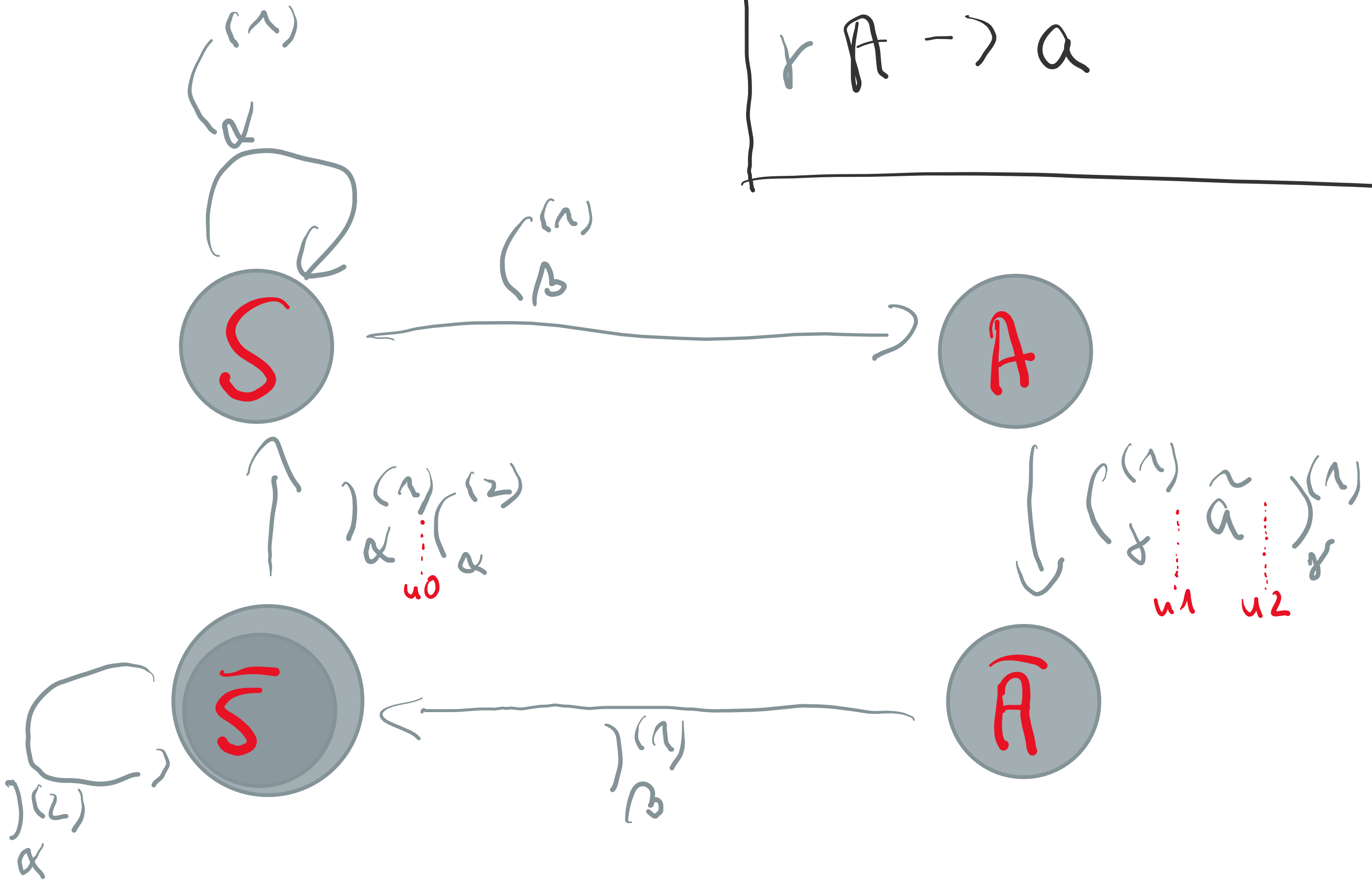
Grammar

$\alpha \quad S \to SS$

$\beta \quad S \to A$

$\gamma \quad A \to a$

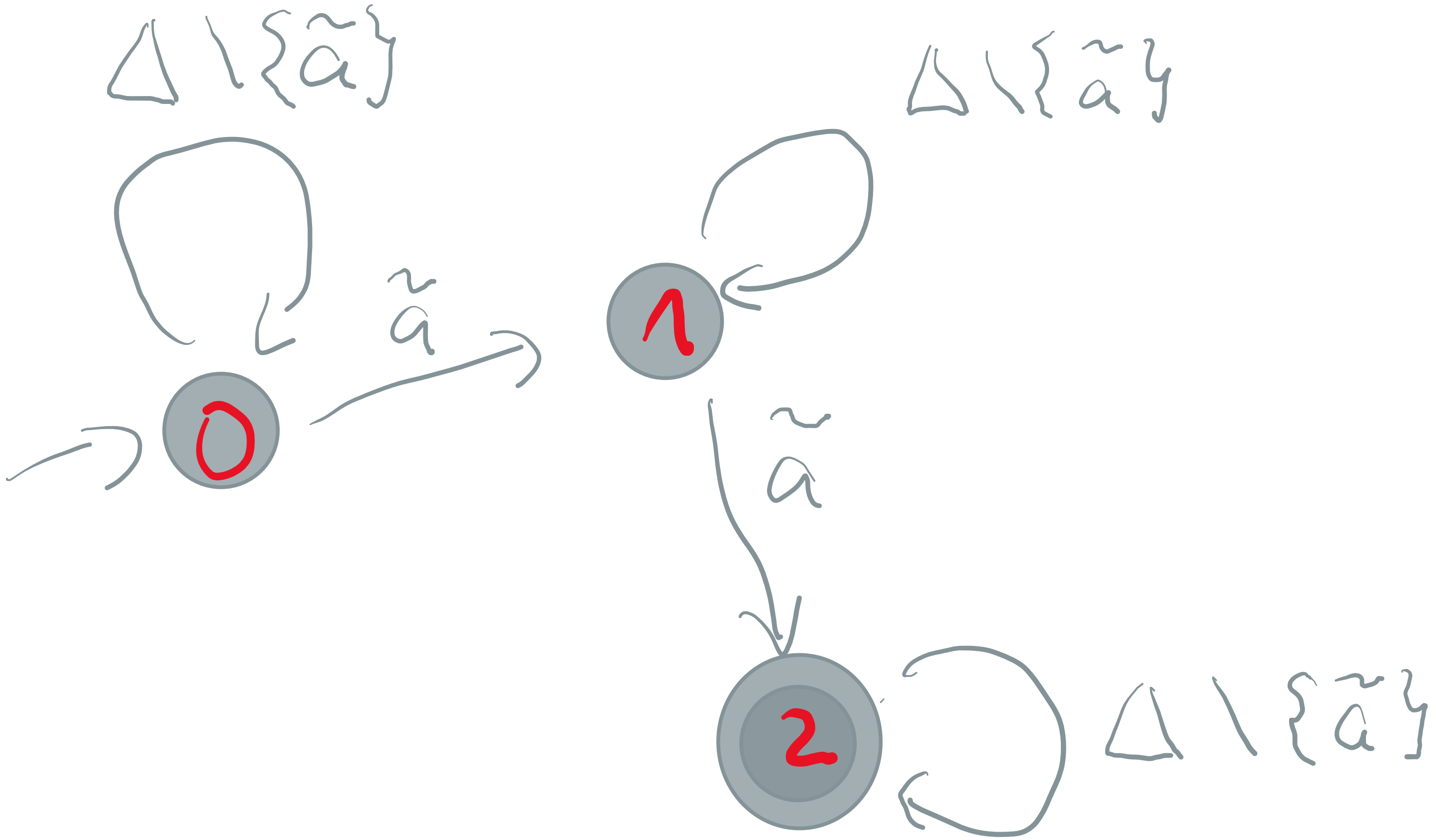




cf. Hulden, 2011

Prerequisites :

∘ construct fsas for R, $h^{-1}(w)$



Grammar

α $S \to SS$

β $S \to A$

γ $A \to a$

Word

$w = {}_0 a_1 a_2$

$$\Delta = \{ (^{(n)}_{\alpha},)^{(n)}_{\alpha} \mid \alpha \in P, n \in [\![rhs(\alpha)]\!] \} \cup \{ \tilde{\sigma} \mid \sigma \in \Sigma \}$$

Hulden, 2011

Prerequisites:

- extract Dyck words from product automaton

→ put $p \xrightarrow{\tilde{a}} q \in T$ in agenda

→ for $p \xrightarrow{v} q$ in agenda

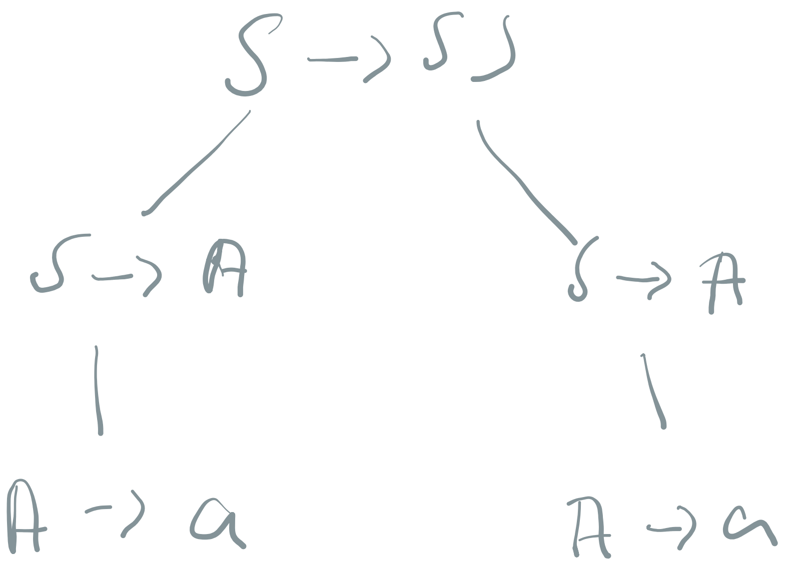
- put $p \xrightarrow{v} q$ into results
- for each $p' \xrightarrow{(} p$, $q \xrightarrow{)} q' \in T$:
  put $p' \xrightarrow{(v)} q'$ in agenda
- for each $p' \xrightarrow{w} p$ in results
  put $p' \xrightarrow{wv} q$ in agenda
- for each $q \xrightarrow{w} q'$ in results
  put $p \xrightarrow{vw} q'$ in agenda
- if $p = q_0$, $q = q_f$: return v

Holden, 2011

Prerequisites: Chart parsing

- build up chart (using deduction system)
  as relation $(N \times \mathbb{N} \times \mathbb{N}) \times (P \times (\mathbb{N} \times \mathbb{N})^*)$

NT + Range	Backtrace
(S, 0, 2)	S → S S, (0,1), (1,2)
(S, 0, 1)	S → A, (0, 1)
(A, 0, 1)	A → a
(S, 1, 2)	S → A, (1, 2)
(A, 1, 2)	A → a



cf. Huang & Chiang, 2005

So, how is Chomsky-Schützenberger-parsing the same as

Chart parsing ?

unweighted Chomsky-Schützenberger-parsing for CFG is
unweighted chart parsing for CFG

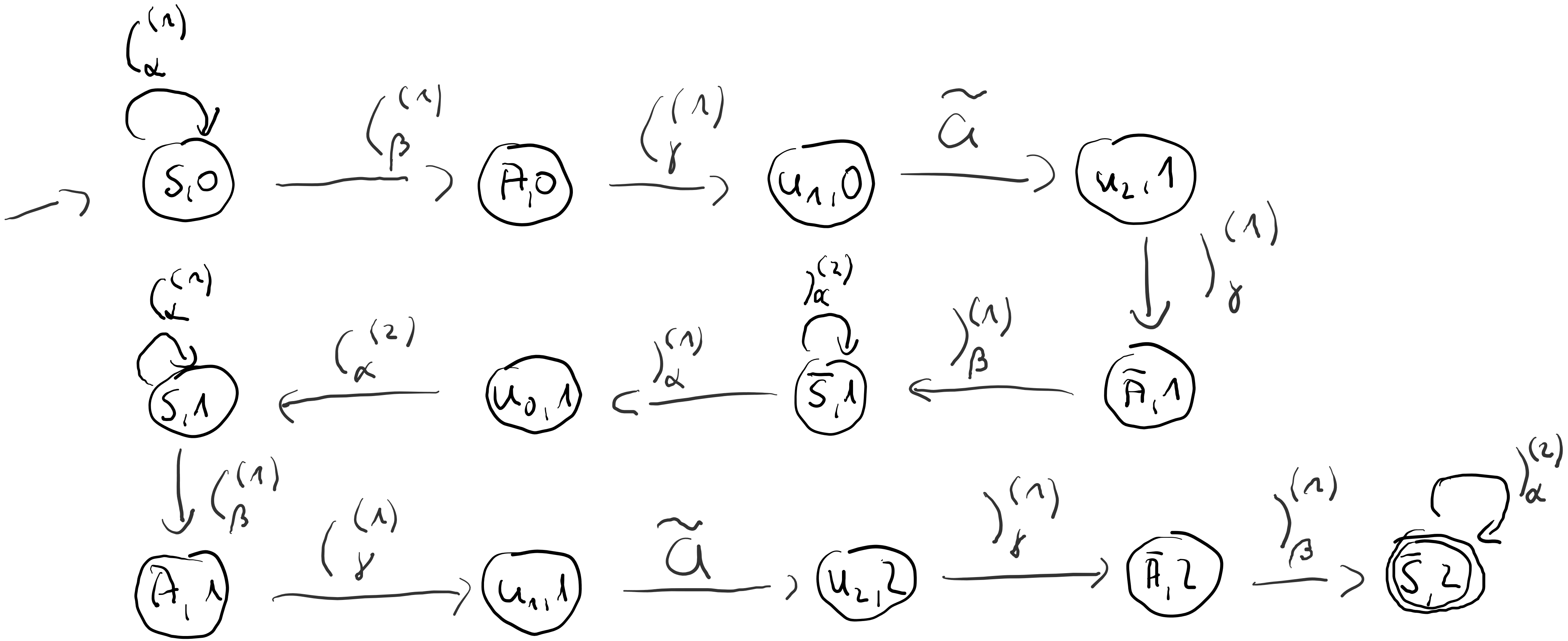
A closer look at the product fsa

- R w/ states $\{A, \bar{A} \mid A \in N\}$ (and some uniques)
- $h^{-1}(w)$ w/ states $[0, |w|]$

$\longrightarrow$ consider runs from (A, i) to $(\bar{A}, j)$ as NT A spanning range $\langle i, j \rangle$

(if brackets are Dyck)

A closer look at the product fsa



A closer look at the extraction of Dyck words

- efficient extraction w/ dynamic programming
  - → store information about how to produce intermediate results
    - initial $(p \xrightarrow{\tilde{a}} q)$
    - by concatenation $(p \xrightarrow{u} p',\ p' \xrightarrow{v} q \Rightarrow p \xrightarrow{uv} q)$
    - by enbracketing $(p \xrightarrow{(} p',\ p' \xrightarrow{u} q',\ q' \xrightarrow{)} q) \Rightarrow (p \xrightarrow{(u)} q)$

A closer look at the extraction of Dyck words

- build up chart

State × State	Backtrace
(p, q)	$\text{init}(\tilde{a})$
	$\text{concat}(p, p', q)$
	$\text{enbracket}(\text{“(“}, p', q', \text{“)“})$

- evaluate chart starting w/ (q_0, q_f)

Putting both together

	State × State	Backtrace
$S, \langle 0,2\rangle$	$(S,0),(\bar{S},2)$	$\mathrm{concat}((u0,1))$
	$(S,0),(u0,1)$	$\mathrm{enbracket}(\ (^1_\alpha,\ (S,0),\ (\bar{S},1),\ )^1_\alpha\ )$
	$(u0,1),(\bar{S},2)$	$\mathrm{enbracket}(\ (^2_\alpha\ (S,1),\ (\bar{S},2),\ )^2_\alpha\ )$
$S,\langle 0,1\rangle$	$(S,0),(\bar{S},1)$	$\mathrm{enbracket}(\ (^1_\beta,\ (A,0),\ (\bar{A},1),\ )^1_\beta\ )$
$A,\langle 0,1\rangle$	$(A,0),(A,1)$	$\mathrm{enbracket}(\ (^1_\gamma,\ (u1,0),(u2,1),\ )^1_\gamma\ )$
	$(u1,0),(u2,1)$	$\mathrm{init}(\tilde{a})$
	⋮	⋮

$(^1_\alpha\ S,\langle 0,1\rangle)^1_\alpha\ (^2_\alpha\ S\langle 1,2\rangle)^2_\alpha$

$(^1_\beta A,\langle 0,1\rangle)^1_\beta$

$(^1_\gamma\ \hat{a})^1_\gamma$

Putting both together

$N \times Range$	$(\Delta \cup (N \times Range))^*$
$S, \langle 0,2 \rangle$	${(}^1_\alpha \; S, \langle 0,1 \rangle \; {)}^1_\alpha \; {(}^2_\alpha \; S, \langle 1,2 \rangle \; {)}^2_\alpha$
$S, \langle 0,1 \rangle$	${(}^1_\beta \; A, \langle 0,1 \rangle \; {)}^1_\beta$
$A, \langle 0,1 \rangle$	${(}^1_\gamma \; \tilde{a} \; {)}^1_\gamma$
$S, \langle 1,2 \rangle$	${(}^1_\beta \; A, \langle 1,2 \rangle \; {)}^1_\beta$
$A, \langle 1,2 \rangle$	${(}^1_\gamma \; \tilde{a} \; {)}^1_\gamma$

Putting both together

$N \times \mathbb{N} \times \mathbb{N}$	$P \times (\mathbb{N} \times \mathbb{N})^*$
$S, 0, 2$	$\alpha, (S, 0, 1), (S, 1, 2)$
$S, 0, 1$	$\beta, (A, 0, 1)$
$A, 0, 1$	γ
$S, 1, 2$	$\beta, (A, 1, 2)$
$A, 1, 2$	γ

But wait, there's more!

	CFG	LCFRS
un-W	✓	CFG approx
W	✓	WCFG approx

Chomsky-Schützenberger-
parsing for CFG is
Chart parsing for CFG.