ERATO セミナ 2010 - No.03

Chomsky-Schützenberger-Type Characterization of Multiple Context-Free Languages

吉仲亮
JST ERATO 湊離散構造処理系プロジェクト

2010 年 6 月 4 日

概要

It is a well-known theorem by Chomsky and Schẗzenberger (1963) that every context-free language can be represented as a homomorphic image of the intersection of a Dyck language and a regular language. This paper gives a Chomsky-Schẗzenberger-type characterization for multiple context-free languages, which are a natural extension of context-free languages, with introducing the notion of multiple Dyck languages, which also are a generalization of Dyck languages.

吉仲研究員

Chomsky-Schützenberger-Type Characterization of Multiple Context-Free Languages

Ryo Yoshinaka[1], Yuichi Kaji[2], Hiroyuki Seki[2]

[1]Hokkaido University (⟶ Japan Science and Technology Agency)

[2]Nara Institute of Science and Technology

Japan

1. Introduction

Multiple Context-Free Languages (mcfls)

- Multiple Context-Free Grammars (mcfgs)

Multiple Context-Free Languages (mcfls)

- Multiple Context-Free Grammars (mcfgs)
- Extension of Context-Free Grammars (cfgs)

Multiple Context-Free Languages (mcfls)

- Multiple Context-Free Grammars (mcfgs)
- Extension of Context-Free Grammars (cfgs)
- Equivalent to several other formalisms:
  Linear context-free rewriting systems, Context-free hypergraph grammars, Multicomponent tree-adjoining grammars, Local unordered scattered context grammars, etc.

Multiple Context-Free Languages (mcfls)

- Multiple Context-Free Grammars (mcfgs)
- Extension of Context-Free Grammars (cfgs)
- Equivalent to several other formalisms:
  Linear context-free rewriting systems, Context-free hypergraph grammars, Multicomponent tree-adjoining grammars, Local unordered scattered context grammars, etc.
- Inherit virtues from CFL

Multiple Context-Free Languages (mcfls)

- Multiple Context-Free Grammars (mcfgs)
- Extension of Context-Free Grammars (cfgs)
- Equivalent to several other formalisms:
  Linear context-free rewriting systems, Context-free hypergraph grammars, Multicomponent tree-adjoining grammars, Local unordered scattered context grammars, etc.
- Inherit virtues from CFL
  - Polynomial time parsing

Multiple Context-Free Languages (mcfls)

- Multiple Context-Free Grammars (mcfgs)
- Extension of Context-Free Grammars (cfgs)
- Equivalent to several other formalisms:
  Linear context-free rewriting systems, Context-free hypergraph grammars, Multicomponent tree-adjoining grammars, Local unordered scattered context grammars, etc.
- Inherit virtues from CFL
  - Polynomial time parsing
  - Substitution closed full AFL, i.e., closed under
    homomorphism, inverse homomorphism, intersection with regular languages, union, concatenation, Kleene plus, substitution

Multiple Context-Free Languages (mcfls)

- Multiple Context-Free Grammars (mcfgs)
- Extension of Context-Free Grammars (cfgs)
- Equivalent to several other formalisms:
  Linear context-free rewriting systems, Context-free hypergraph grammars, Multicomponent tree-adjoining grammars, Local unordered scattered context grammars, etc.
- Inherit virtues from CFL
  - Polynomial time parsing
  - Substitution closed full AFL, i.e., closed under
    homomorphism, inverse homomorphism, intersection with regular languages, union, concatenation, Kleene plus, substitution
  - Pumping lemma — open problem

Multiple Context-Free Languages (mcfls)

- Multiple Context-Free Grammars (mcfgs)
- Extension of Context-Free Grammars (cfgs)
- Equivalent to several other formalisms:
  Linear context-free rewriting systems, Context-free hypergraph grammars, Multicomponent tree-adjoining grammars, Local unordered scattered context grammars, etc.
- Inherit virtues from CFL
  - Polynomial time parsing
  - Substitution closed full AFL, i.e., closed under
    homomorphism, inverse homomorphism, intersection with regular languages, union, concatenation, Kleene plus, substitution
  - Pumping lemma — open problem
  - Representation Theorem — this talk

Representation Theorem

- A language L is context-free, if and only if

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

for a homomorphism h, a Dyck language D and a regular language R.

(Chomsky & Schützenberger 1963)

Representation Theorem

- A language L is context-free, if and only if

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

for a homomorphism h, a Dyck language D and a regular language R.

(Chomsky & Schützenberger 1963)

- A language L is *multiple* context-free, if and only if

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

for a homomorphism h, a *multiple* Dyck language D and a regular language R.

Outline

2. Multiple Context-Free Grammars

Multiple Context-Free Grammars

- CF rules:
  bottom-up recursive definition of the languages of nonterminals
- $A \to aBC$ (cf rule):
  $B \stackrel{*}{\Rightarrow} u$ and $C \stackrel{*}{\Rightarrow} v$ implies $A \stackrel{*}{\Rightarrow} auv$.

Multiple Context-Free Grammars

- CF rules:
  bottom-up recursive definition of the languages of nonterminals
- $A \to aBC$ (cf rule):
  $B \stackrel{*}{\Rightarrow} u$ and $C \stackrel{*}{\Rightarrow} v$ implies $A \stackrel{*}{\Rightarrow} auv$.
- $A\langle axy\rangle$:– $B\langle x\rangle, C\langle y\rangle$ (mcf rule notation).

Multiple Context-Free Grammars

- CF rules:
  bottom-up recursive definition of the languages of nonterminals
- $A \to aBC$ (cf rule):
  $B \stackrel{*}{\Rightarrow} u$ and $C \stackrel{*}{\Rightarrow} v$ implies $A \stackrel{*}{\Rightarrow} auv$.
- $A\langle axy\rangle$:– $B\langle x\rangle, C\langle y\rangle$ (mcf rule notation).
- Each nonterminal generates
  - a set of strings (cfg)
  - a set of *tuples* of strings (mcfg)

Multiple Context-Free Grammars

- CF rules:
  bottom-up recursive definition of the languages of nonterminals
- $A \to aBC$ (cf rule):
  $B \stackrel{*}{\Rightarrow} u$ and $C \stackrel{*}{\Rightarrow} v$ implies $A \stackrel{*}{\Rightarrow} auv$.
- $A\langle axy\rangle$:– $B\langle x\rangle, C\langle y\rangle$ (mcf rule notation).
- Each nonterminal generates
  - a set of strings (cfg)
  - a set of *tuples* of strings (mcfg)
- $D\langle bx_1, y_1cx_2\rangle$:– $E\langle x_1, x_2\rangle, F\langle y_1\rangle$,
  $E \stackrel{*}{\Rightarrow} \langle u_1, u_2\rangle$ and $F \stackrel{*}{\Rightarrow} \langle v_1\rangle$ implies $D \stackrel{*}{\Rightarrow} \langle bu_1, v_1cu_2\rangle$.

Multiple Context-Free Grammars

$G = \langle \Sigma, N, P, S \rangle$

- Σ: alphabet of terminal symbols
- N: *indexed* alphabet of nonterminal symbols ($\dim(A) \geq 1$)
  - A generates $\dim(A)$-tuples of strings
- $S \in N$: distinguished symbol ($\dim(S) = 1$)
- P: production rules

Multiple Context-Free Grammars

$G = \langle \Sigma, N, P, S \rangle$

- Σ: alphabet of terminal symbols
- N: *indexed* alphabet of nonterminal symbols ($\dim(A) \geq 1$)
  - A generates $\dim(A)$-tuples of strings
- $S \in N$: distinguished symbol ($\dim(S) = 1$)
- P: production rules
  - Each rule has the form

$$A\langle t_1, \ldots, t_{\dim(A)} \rangle :- B_1\langle x_{1,1}, \ldots, x_{1,\dim(B_1)} \rangle, \ldots, B_n\langle x_{n,1}, \ldots, x_{n,\dim(B_n)} \rangle$$

  - $t_i \in (\Sigma \cup \{ x_{i,j} \mid 1 \leq i \leq n, 1 \leq j \leq \dim(B_i) \})^*$
  - each variable $x_{i,j}$ occurs exactly once in $t_1 \ldots t_{\dim(A)}$.

Example

$\star\ S\langle x_1 y_1 x_2 y_2 \rangle :- A\langle x_1, x_2 \rangle, B\langle y_1, y_2 \rangle,$

$\star\ A\langle a, c \rangle :- ,\quad \star\ A\langle a x_1, c x_2 \rangle :- A\langle x_1, x_2 \rangle,$

$\star\ B\langle b, d \rangle :- ,\quad \star\ B\langle b y_1, d y_2 \rangle :- B\langle y_1, y_2 \rangle,$

Example

$\star\ S\langle x_1 y_1 x_2 y_2 \rangle :- A\langle x_1, x_2 \rangle, B\langle y_1, y_2 \rangle,$

$\star\ A\langle a, c \rangle :- ,\quad \star\ A\langle a x_1, c x_2 \rangle :- A\langle x_1, x_2 \rangle,$

$\star\ B\langle b, d \rangle :- ,\quad \star\ B\langle b y_1, d y_2 \rangle :- B\langle y_1, y_2 \rangle,$

$$\langle a, c \rangle \in L(A) \Longrightarrow \langle aa, cc \rangle \in L(A) \ldots \Longrightarrow \langle a^m, c^m \rangle \in L(A)$$

Example

$\star\ S\langle x_1 y_1 x_2 y_2 \rangle :- A\langle x_1, x_2 \rangle, B\langle y_1, y_2 \rangle,$

$\star\ A\langle a, c \rangle :- ,\quad \star\ A\langle a x_1, c x_2 \rangle :- A\langle x_1, x_2 \rangle,$

$\star\ B\langle b, d \rangle :- ,\quad \star\ B\langle b y_1, d y_2 \rangle :- B\langle y_1, y_2 \rangle,$

$$\langle a, c \rangle \in L(A) \Longrightarrow \langle aa, cc \rangle \in L(A) \ldots \Longrightarrow \langle a^m, c^m \rangle \in L(A)$$

$$L(A) = \{\langle a^m, c^m \rangle \mid m > 0\}$$
$$L(B) = \{\langle b^n, d^n \rangle \mid n > 0\}$$

Example

$\star\ S\langle x_1 y_1 x_2 y_2 \rangle :- A\langle x_1, x_2 \rangle, B\langle y_1, y_2 \rangle,$

$\star\ A\langle a, c \rangle :- ,\quad \star\ A\langle a x_1, c x_2 \rangle :- A\langle x_1, x_2 \rangle,$

$\star\ B\langle b, d \rangle :- ,\quad \star\ B\langle b y_1, d y_2 \rangle :- B\langle y_1, y_2 \rangle,$

$$\langle a, c \rangle \in L(A) \Longrightarrow \langle aa, cc \rangle \in L(A) \ldots \Longrightarrow \langle a^m, c^m \rangle \in L(A)$$

$$L(A) = \{\langle a^m, c^m \rangle \mid m > 0\}$$
$$L(B) = \{\langle b^n, d^n \rangle \mid n > 0\}$$

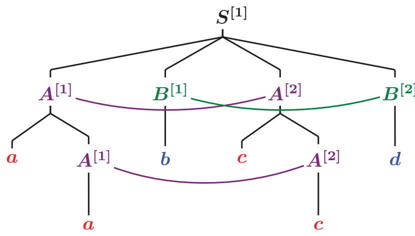
$$L(G) = \{ a^m b^n c^m d^n \mid m, n > 0 \}$$

Example

$\star\ S\langle x_1y_1x_2y_2\rangle :\!-\ A\langle x_1, x_2\rangle, B\langle y_1, y_2\rangle,$

$\star\ A\langle a, c\rangle :\!-\ ,\quad \star\ A\langle ax_1, cx_2\rangle :\!-\ A\langle x_1, x_2\rangle,$

$\star\ B\langle b, d\rangle :\!-\ ,\quad \star\ B\langle by_1, dy_2\rangle :\!-\ B\langle y_1, y_2\rangle,$

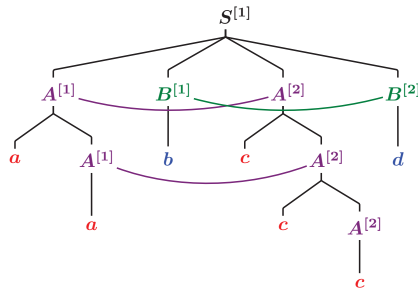


Example

$\star\ S\langle x_1y_1x_2y_2\rangle :\!-\ A\langle x_1, x_2\rangle, B\langle y_1, y_2\rangle,$

$\star\ A\langle a, c\rangle :\!-\ ,\quad \star\ A\langle ax_1, cx_2\rangle :\!-\ A\langle x_1, x_2\rangle,$

$\star\ B\langle b, d\rangle :\!-\ ,\quad \star\ B\langle by_1, dy_2\rangle :\!-\ B\langle y_1, y_2\rangle,$

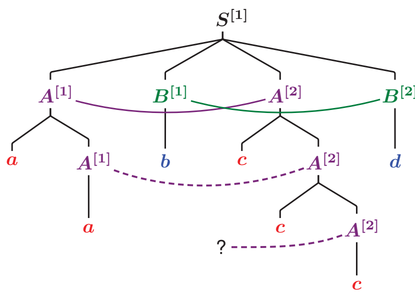


Example

$\star\ S\langle x_1y_1x_2y_2\rangle :\!-\ A\langle x_1, x_2\rangle, B\langle y_1, y_2\rangle,$

$\star\ A\langle a, c\rangle :\!-\ ,\quad \star\ A\langle ax_1, cx_2\rangle :\!-\ A\langle x_1, x_2\rangle,$

$\star\ B\langle b, d\rangle :\!-\ ,\quad \star\ B\langle by_1, dy_2\rangle :\!-\ B\langle y_1, y_2\rangle,$



MCFL Hierarchy

- m-MCFG(r): collection of mcfgs such that
  - $\dim(A) \leq m$ for all nonterminals A,
  - Rules have at most r nonterminals on right,
    i.e., $A\langle\vec{t}\rangle :\!-\ B_1\langle\vec{x}_1\rangle, \ldots, B_n\langle\vec{x}_n\rangle$ implies $n \leq r$.

MCFL Hierarchy

- m-MCFG(r): collection of mcfgs such that
  - $\dim(A) \leq m$ for all nonterminals A,
  - Rules have at most r nonterminals on right,
    i.e., $A\langle\vec{t}\rangle :\!-\ B_1\langle\vec{x}_1\rangle, \ldots, B_n\langle\vec{x}_n\rangle$ implies $n \leq r$.
- m-MCFL(r): family of languages corresponding to m-MCFG(r).

MCFL Hierarchy

- m-MCFG(r): collection of mcfgs such that
  - $\dim(A) \leq m$ for all nonterminals A,
  - Rules have at most r nonterminals on right,
    i.e., $A\langle\vec{t}\rangle :\!-\ B_1\langle\vec{x}_1\rangle, \ldots, B_n\langle\vec{x}_n\rangle$ implies $n \leq r$.
- m-MCFL(r): family of languages corresponding to m-MCFG(r).
  - 1-MCFL$(2) = \bigcup_{r\geq 1}$ 1-MCFL(r): CFL

MCFL Hierarchy

- m-MCFG(r): collection of mcfgs such that
  - $\dim(A) \le m$ for all nonterminals A,
  - Rules have at most r nonterminals on right, i.e., $A\langle\vec{t}\rangle :- B_1\langle\vec{x}_1\rangle, \ldots, B_n\langle\vec{x}_n\rangle$ implies $n \le r$.
- m-MCFL(r): family of languages corresponding to m-MCFG(r).
  - 1-MCFL$(2) = \bigcup_{r\ge 1}$ 1-MCFL(r): CFL
  - m-MCFL(r) $\subsetneq$ m-MCFL($r+1$) for $m \ge 2$ unless $m = r = 2$
  - m-MCFL(r) $\subsetneq$ ($m+1$)-MCFL(r) for $r \ge 1$

$\{ a_1^n b_1^n a_2^n b_2^n \ldots a_m^n b_m^n \mid n \ge 0 \} \in m$-MCFL$(1) \setminus (m-1)$-MCFL$(1)$

MCFL Hierarchy

- m-MCFG(r): collection of mcfgs such that
  - $\dim(A) \le m$ for all nonterminals A,
  - Rules have at most r nonterminals on right, i.e., $A\langle\vec{t}\rangle :- B_1\langle\vec{x}_1\rangle, \ldots, B_n\langle\vec{x}_n\rangle$ implies $n \le r$.
- m-MCFL(r): family of languages corresponding to m-MCFG(r).
  - 1-MCFL$(2) = \bigcup_{r\ge 1}$ 1-MCFL(r): CFL
  - m-MCFL(r) $\subsetneq$ m-MCFL($r+1$) for $m \ge 2$ unless $m = r = 2$
  - m-MCFL(r) $\subsetneq$ ($m+1$)-MCFL(r) for $r \ge 1$

$\{ a_1^n b_1^n a_2^n b_2^n \ldots a_m^n b_m^n \mid n \ge 0 \} \in m$-MCFL$(1) \setminus (m-1)$-MCFL$(1)$

- m-MCFL(r): substitution closed full AFL

Representation Theorem of MCFL

- A language L is multiple context-free in m-MCFL(r), if and only if

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

for a homomorphism h, a multiple Dyck language D in m-MCFL(r) and a regular language R.

Representation Theorem of MCFL

- A language L is multiple context-free in m-MCFL(r), if and only if

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

for a homomorphism h, a multiple Dyck language D in m-MCFL(r) and a regular language R.

- "if" part follows from that m-MCFL(r) is a substitution closed full AFL.

3. Multiple Dyck Languages

Multiple Dyck Languages

- Dyck Grammar on an alphabet Δ defines a language over $\Delta \cup \overline{\Delta}$
  - Unique nonterminal S,
  - $S \to aS\bar{a}$ where $a \in \Delta$, $\bar{a} \in \overline{\Delta}$,
  - $S \to SS \mid \varepsilon$.

Multiple Dyck Languages

- Dyck Grammar on an alphabet Δ defines a language over $\Delta \cup \overline{\Delta}$
  - Unique nonterminal S,
  - $S\langle a x \bar{a} \rangle$:– $S\langle x \rangle$ where $a \in \Delta$, $\bar{a} \in \overline{\Delta}$,
  - $S\langle x_1 x_2 \rangle$:– $S\langle x_1 \rangle, S\langle x_2 \rangle$ and $S\langle \varepsilon \rangle$:– .

Multiple Dyck Languages

- Dyck Grammar on an alphabet Δ defines a language over $\Delta \cup \overline{\Delta}$
  - Unique nonterminal S,
  - $S\langle a x \bar{a} \rangle$:– $S\langle x \rangle$ where $a \in \Delta$, $\bar{a} \in \overline{\Delta}$,
  - $S\langle x_1 x_2 \rangle$:– $S\langle x_1 \rangle, S\langle x_2 \rangle$ and $S\langle \varepsilon \rangle$:– .
- *Multiple Dyck Grammar* in m-MCFG(r) on an *indexed* alphabet Δ such that $\dim(a) \leq m$ for all $a \in \Delta$ defines a language over

$$\widehat{\Delta} = \{\, a^{[i]}, \bar{a}^{[i]} \mid 1 \leq i \leq \dim(a),\ a \in \Delta \,\}$$

Multiple Dyck Languages

- Dyck Grammar on an alphabet Δ defines a language over $\Delta \cup \overline{\Delta}$
  - Unique nonterminal S,
  - $S\langle a x \bar{a} \rangle$:– $S\langle x \rangle$ where $a \in \Delta$, $\bar{a} \in \overline{\Delta}$,
  - $S\langle x_1 x_2 \rangle$:– $S\langle x_1 \rangle, S\langle x_2 \rangle$ and $S\langle \varepsilon \rangle$:– .
- *Multiple Dyck Grammar* in m-MCFG(r) on an *indexed* alphabet Δ such that $\dim(a) \leq m$ for all $a \in \Delta$ defines a language over

$$\widehat{\Delta} = \{\, a^{[i]}, \bar{a}^{[i]} \mid 1 \leq i \leq \dim(a),\ a \in \Delta \,\}$$

  - Nonterminals: $S_1, \ldots, S_m$ where $\dim(S_k) = k$,

Multiple Dyck Languages

- Dyck Grammar on an alphabet Δ defines a language over $\Delta \cup \overline{\Delta}$
  - Unique nonterminal S,
  - $S\langle a x \bar{a} \rangle$:– $S\langle x \rangle$ where $a \in \Delta$, $\bar{a} \in \overline{\Delta}$,
  - $S\langle x_1 x_2 \rangle$:– $S\langle x_1 \rangle, S\langle x_2 \rangle$ and $S\langle \varepsilon \rangle$:– .
- *Multiple Dyck Grammar* in m-MCFG(r) on an *indexed* alphabet Δ such that $\dim(a) \leq m$ for all $a \in \Delta$ defines a language over

$$\widehat{\Delta} = \{\, a^{[i]}, \bar{a}^{[i]} \mid 1 \leq i \leq \dim(a),\ a \in \Delta \,\}$$

  - Nonterminals: $S_1, \ldots, S_m$ where $\dim(S_k) = k$,
  - $S_k\langle a^{[1]} x_1 \bar{a}^{[1]}, \ldots, a^{[k]} x_k \bar{a}^{[k]} \rangle$:– $S_k\langle x_1, \ldots, x_k \rangle$ where $a \in \Delta$, $\dim(a) = k$,
  - Any terminal-free rules allowed in m-MCFG(r), i.e., $S_k\langle \vec{t} \rangle$:– $S_{k_1}\langle \vec{x}_1 \rangle, \ldots, S_{k_n}\langle \vec{x}_n \rangle$ where $\vec{t}$ is terminal-free and $n \leq r$.

Multiple Dyck Languages

- Elements of multiple Dyck language on Δ are those of a (cf) Dyck language on $\{\, a^{[i]} \mid a \in \Delta,\ 1 \leq i \leq \dim(a) \,\}$
- Occurrences of bracket-pairs are partitioned into groups
- Each consists of $\langle a^{[1]}, \bar{a}^{[1]} \rangle, \ldots, \langle a^{[\dim(a)]}, \bar{a}^{[\dim(a)]} \rangle$ for $a \in \Delta$

Multiple Dyck Languages

- Elements of multiple Dyck language on Δ are those of a (cf) Dyck language on $\{\, a^{[i]} \mid a \in \Delta,\ 1 \leq i \leq \dim(a) \,\}$
- Occurrences of bracket-pairs are partitioned into groups
- Each consists of $\langle a^{[1]}, \bar{a}^{[1]} \rangle, \ldots, \langle a^{[\dim(a)]}, \bar{a}^{[\dim(a)]} \rangle$ for $a \in \Delta$
- If a member of a group X occurs inside another Y, all members of X occur inside Y, too.

(^_^) $a^{[2]}\ b^{[1]}\ a^{[1]}\ \bar{a}^{[1]}\ a^{[2]}\ \bar{a}^{[2]}\ \bar{b}^{[1]}\ \bar{a}^{[2]}\ a^{[1]}\ b^{[2]}\ \bar{b}^{[2]}\ \bar{a}^{[1]}$

where $\dim(a) = \dim(b) = 2$.

Multiple Dyck Languages

- Elements of multiple Dyck language on Δ are those of a (cf) Dyck language on $\{\, a^{[i]} \mid a \in \Delta,\ 1 \le i \le \dim(a) \,\}$
- Occurrences of bracket-pairs are partitioned into groups
- Each consists of $\langle a^{[1]}, \bar{a}^{[1]} \rangle, \ldots, \langle a^{[\dim(a)]}, \bar{a}^{[\dim(a)]} \rangle$ for $a \in \Delta$
- If a member of a group X occurs inside another Y, all members of X occur inside Y, too.

  (^␣^) $a^{[2]}\ b^{[1]}\ a^{[1]}\ \bar{a}^{[1]}\ a^{[2]}\ \bar{a}^{[2]}\ \bar{b}^{[1]}\ \bar{a}^{[2]}\ a^{[1]}\ b^{[2]}\ \bar{b}^{[2]}\ \bar{a}^{[1]}$

  (x_x) $b^{[1]}\ a^{[1]}\ \bar{a}^{[1]}\ \bar{b}^{[1]}\ a^{[2]}\ b^{[2]}\ \bar{b}^{[2]} \bar{a}^{[2]},$

  where $\dim(a) = \dim(b) = 2$.

Multiple Dyck Languages

- Elements of multiple Dyck language on Δ are those of a (cf) Dyck language on $\{\, a^{[i]} \mid a \in \Delta,\ 1 \le i \le \dim(a) \,\}$
- Occurrences of bracket-pairs are partitioned into groups
- Each consists of $\langle a^{[1]}, \bar{a}^{[1]} \rangle, \ldots, \langle a^{[\dim(a)]}, \bar{a}^{[\dim(a)]} \rangle$ for $a \in \Delta$
- If a member of a group X occurs inside another Y, all members of X occur inside Y, too.

  (^␣^) $a^{[2]}\ b^{[1]}\ a^{[1]}\ \bar{a}^{[1]}\ a^{[2]}\ \bar{a}^{[2]}\ \bar{b}^{[1]}\ \bar{a}^{[2]}\ a^{[1]}\ b^{[2]}\ \bar{b}^{[2]}\ \bar{a}^{[1]}$

  (x_x) $b^{[1]}\ a^{[1]}\ \bar{a}^{[1]}\ \bar{b}^{[1]}\ a^{[2]}\ b^{[2]}\ \bar{b}^{[2]} \bar{a}^{[2]},\quad a^{[1]}\ a^{[2]}\ \bar{a}^{[2]}\ \bar{a}^{[1]},$

  where $\dim(a) = \dim(b) = 2$.

Multiple Dyck Languages

- Elements of multiple Dyck language on Δ are those of a (cf) Dyck language on $\{\, a^{[i]} \mid a \in \Delta,\ 1 \le i \le \dim(a) \,\}$
- Occurrences of bracket-pairs are partitioned into groups
- Each consists of $\langle a^{[1]}, \bar{a}^{[1]} \rangle, \ldots, \langle a^{[\dim(a)]}, \bar{a}^{[\dim(a)]} \rangle$ for $a \in \Delta$
- If a member of a group X occurs inside another Y, all members of X occur inside Y, too.

  (^␣^) $a^{[2]}\ b^{[1]}\ a^{[1]}\ \bar{a}^{[1]}\ a^{[2]}\ \bar{a}^{[2]}\ \bar{b}^{[1]}\ \bar{a}^{[2]}\ a^{[1]}\ b^{[2]}\ \bar{b}^{[2]}\ \bar{a}^{[1]}$

  (x_x) $b^{[1]}\ a^{[1]}\ \bar{a}^{[1]}\ \bar{b}^{[1]}\ a^{[2]}\ b^{[2]}\ \bar{b}^{[2]} \bar{a}^{[2]},\quad a^{[1]}\ a^{[2]}\ \bar{a}^{[2]}\ \bar{a}^{[1]},\quad b^{[2]}\ \bar{b}^{[2]}$

  where $\dim(a) = \dim(b) = 2$.

Multiple Dyck Languages

- Elements of multiple Dyck language on Δ are those of a (cf) Dyck language on $\{\, a^{[i]} \mid a \in \Delta,\ 1 \le i \le \dim(a) \,\}$
- Occurrences of bracket-pairs are partitioned into groups
- Each consists of $\langle a^{[1]}, \bar{a}^{[1]} \rangle, \ldots, \langle a^{[\dim(a)]}, \bar{a}^{[\dim(a)]} \rangle$ for $a \in \Delta$
- If a member of a group X occurs inside another Y, all members of X occur inside Y, too.

  (^␣^) $a^{[2]}\ b^{[1]}\ a^{[1]}\ \bar{a}^{[1]}\ a^{[2]}\ \bar{a}^{[2]}\ \bar{b}^{[1]}\ \bar{a}^{[2]}\ a^{[1]}\ b^{[2]}\ \bar{b}^{[2]}\ \bar{a}^{[1]}$

  (x_x) $b^{[1]}\ a^{[1]}\ \bar{a}^{[1]}\ \bar{b}^{[1]}\ a^{[2]}\ b^{[2]}\ \bar{b}^{[2]} \bar{a}^{[2]},\quad a^{[1]}\ a^{[2]}\ \bar{a}^{[2]}\ \bar{a}^{[1]},\quad b^{[2]}\ \bar{b}^{[2]}$

  where $\dim(a) = \dim(b) = 2$.

- Ways of aligning brackets are restricted by available means in m-MCFG(r)

4. Representation Theorem of CFL

Representation Theorem

- A language L over Σ is context-free, if and only if

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

  for a homomorphism $h : (\Delta \cup \overline{\Delta})^* \to \Sigma^*$,
  the Dyck language D on Δ and
  a regular language $R \subseteq (\Delta \cup \overline{\Delta})^*$.

Representation Theorem

- A language L over Σ is context-free, if and only if

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

for a homomorphism $h : (\Delta \cup \overline{\Delta})^* \to \Sigma^*$,
the Dyck language D on Δ and
a regular language $R \subseteq (\Delta \cup \overline{\Delta})^*$.

- Elements of Δ are written as $\llbracket_X$ for some X and those of $\overline{\Delta}$ are $\rrbracket_X$ instead of $\overline{\llbracket_X}$

Example

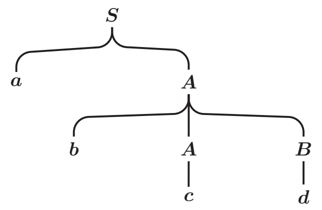
$L = h(D \cap R)$

- Original cfg (L)
$\pi_1 : S \to aA,\quad \pi_2 : A \to bAB,\quad \pi_3 : A \to c,\quad \pi_4 : B \to d$

Example

$L = h(D \cap R)$

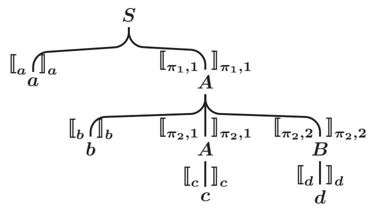
- Original cfg (L)
$\pi_1 : S \to aA,\quad \pi_2 : A \to bAB,\quad \pi_3 : A \to c,\quad \pi_4 : B \to d$



Example

$L = h(D \cap R)$

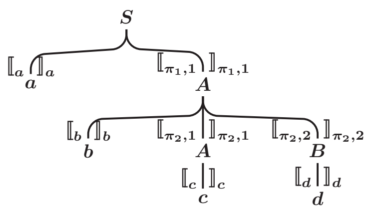
- Original cfg (L)
$\pi_1 : S \to aA,\quad \pi_2 : A \to bAB,\quad \pi_3 : A \to c,\quad \pi_4 : B \to d$



Example

$L = h(D \cap R)$

- Original cfg (L)
$\pi_1 : S \to aA,\quad \pi_2 : A \to bAB,\quad \pi_3 : A \to c,\quad \pi_4 : B \to d$

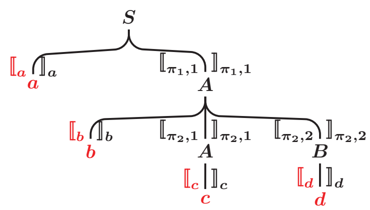


- $\llbracket_a \rrbracket_a \llbracket_{\pi_1,1} \llbracket_b \rrbracket_b \llbracket_{\pi_2,1} \llbracket_c \rrbracket_c \rrbracket_{\pi_2,1} \llbracket_{\pi_2,2} \llbracket_d \rrbracket_d \rrbracket_{\pi_2,2} \rrbracket_{\pi_1,1} \in D \cap R$

Example

$L = h(D \cap R)$

- Original cfg (L)
$\pi_1 : S \to aA,\quad \pi_2 : A \to bAB,\quad \pi_3 : A \to c,\quad \pi_4 : B \to d$



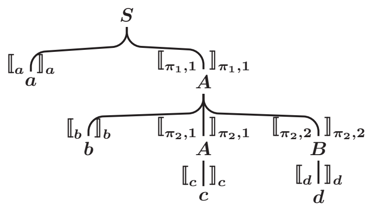
- $\llbracket_a \rrbracket_a \llbracket_{\pi_1,1} \llbracket_b \rrbracket_b \llbracket_{\pi_2,1} \llbracket_c \rrbracket_c \rrbracket_{\pi_2,1} \llbracket_{\pi_2,2} \llbracket_d \rrbracket_d \rrbracket_{\pi_2,2} \rrbracket_{\pi_1,1} \in D \cap R$

- $h(X) = \begin{cases} z & \text{if } X = \llbracket_z \text{ with } z \in \{a, b, c, d\} \\ \varepsilon & \text{otherwise} \end{cases}$

Example

$L = h(D \cap R)$

- Original cfg (L)
  $\pi_1 : S \to aA,\ \ \pi_2 : A \to bAB,\ \ \pi_3 : A \to c,\ \ \pi_4 : B \to d$



- $\llbracket_a \rrbracket_a \llbracket_{\pi_{1,1}} \llbracket_b \rrbracket_b \llbracket_{\pi_{2,1}} \llbracket_c \rrbracket_c \rrbracket_{\pi_{2,1}} \llbracket_{\pi_{2,2}} \llbracket_d \rrbracket_d \rrbracket_{\pi_{2,2}} \rrbracket_{\pi_{1,1}} \in D \cap R$

Example

$L = h(D \cap R)$

- Original cfg (L)
  $\pi_1 : S \to aA,\ \ \pi_2 : A \to bAB,\ \ \pi_3 : A \to c,\ \ \pi_4 : B \to d$
- Dyck-decorated cfg $(D \cap R)$
  $\hat{\pi}_1 : S \to \llbracket_a \rrbracket_a \llbracket_{\pi_{1,1}} A \rrbracket_{\pi_{1,1}}$,
  $\hat{\pi}_2 : A \to \llbracket_b \rrbracket_b \llbracket_{\pi_{2,1}} A \rrbracket_{\pi_{2,1}} \llbracket_{\pi_{2,2}} B \rrbracket_{\pi_{2,2}}$,
  $\hat{\pi}_3 : A \to \llbracket_c \rrbracket_c$, $\hat{\pi}_4 : B \to \llbracket_d \rrbracket_d$

Example

$L = h(D \cap R)$

- Original cfg (L)
  $\pi_1 : S \to aA,\ \ \pi_2 : A \to bAB,\ \ \pi_3 : A \to c,\ \ \pi_4 : B \to d$
- Dyck-decorated cfg $(D \cap R)$
  $\hat{\pi}_1 : S \to \llbracket_a \rrbracket_a \llbracket_{\pi_{1,1}} A \rrbracket_{\pi_{1,1}}$,
  $\hat{\pi}_2 : A \to \llbracket_b \rrbracket_b \llbracket_{\pi_{2,1}} A \rrbracket_{\pi_{2,1}} \llbracket_{\pi_{2,2}} B \rrbracket_{\pi_{2,2}}$,
  $\hat{\pi}_3 : A \to \llbracket_c \rrbracket_c$, $\hat{\pi}_4 : B \to \llbracket_d \rrbracket_d$
- Control right-linear grammar (R)
  $(\pi_1)\ S \to \llbracket_a \rrbracket_a \llbracket_{\pi_{1,1}} A,\ T \to \rrbracket_{\pi_{1,1}} T,$
  $(\pi_2)\ A \to \llbracket_b \rrbracket_b \llbracket_{\pi_{2,1}} A,\ T \to \rrbracket_{\pi_{2,1}} \llbracket_{\pi_{2,2}} B,\ T \to \rrbracket_{\pi_{2,2}} T,$
  $(\pi_3)\ A \to \llbracket_c \rrbracket_c T$, $(\pi_4)\ B \to \llbracket_d \rrbracket_d T$,
  (aux.) $T \to \varepsilon$.
- $A \overset{*}{\underset{G_L}{\Rightarrow}} w$ iff $A \overset{*}{\underset{G_R}{\Rightarrow}} wT$ and $w \in D$.

5. Representation Theorem of MCFL

MCFL's Case (naive application)

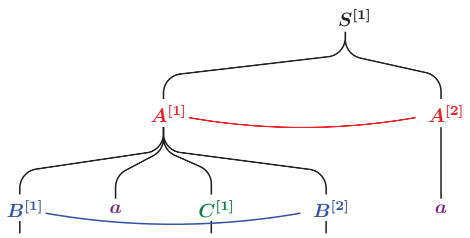
$L = h(D \cap R)$

- Original mcfg (L):
  $\rho : S\langle z_1 z_2 \rangle :\!-\ A\langle z_1, z_2 \rangle,$
  $\pi : A\langle x_1 a y_1 x_2,\ a \rangle :\!-\ B\langle x_1, x_2 \rangle, C\langle y_1 \rangle$

MCFL's Case (naive application)

$L = h(D \cap R)$

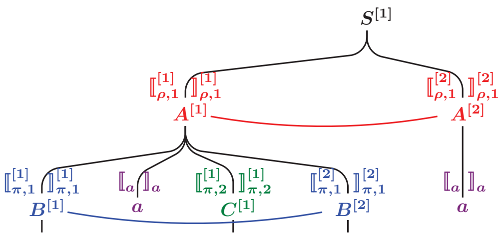
- Original mcfg (L):
  $\rho : S\langle z_1 z_2 \rangle :\!-\ A\langle z_1, z_2 \rangle,$
  $\pi : A\langle x_1 a y_1 x_2,\ a \rangle :\!-\ B\langle x_1, x_2 \rangle, C\langle y_1 \rangle$



MCFL's Case (naive application)

$L = h(D \cap R)$

- Original mcfg (L):
  $\rho : S\langle z_1 z_2\rangle \text{ :- } A\langle z_1, z_2\rangle,$
  $\pi : A\langle x_1 a y_1 x_2,\ a\rangle \text{ :- } B\langle x_1, x_2\rangle, C\langle y_1\rangle$



MCFL's Case (naive application)

$L = h(D \cap R)$

- Original mcfg ($L \in m$-MCFL(r)):
  $\pi : A\langle x_1 a y_1 x_2,\ a\rangle \text{ :- } B\langle x_1, x_2\rangle, C\langle y_1\rangle$

MCFL's Case (naive application)

$L = h(D \cap R)$

- Original mcfg ($L \in m$-MCFL(r)):
  $\pi : A\langle x_1 a y_1 x_2,\ a\rangle \text{ :- } B\langle x_1, x_2\rangle, C\langle y_1\rangle$
- Multiple Dyck decoration ($D \cap R \in m$-MCFL(r)):
  $\hat{\pi} : A\langle\ \llbracket^{[1]}_{\pi,1}\, x_1 \rrbracket^{[1]}_{\pi,1} \llbracket_a \rrbracket_a \llbracket^{[1]}_{\pi,2}\, y_1 \rrbracket^{[1]}_{\pi,2} \llbracket^{[2]}_{\pi,1}\, x_2 \rrbracket^{[2]}_{\pi,1}\ ,\ \llbracket_a \rrbracket_a\ \rangle$
  $\text{ :- } B\langle x_1, x_2\rangle, C\langle y_1\rangle$
  - Simulated by the multiple Dyck grammar in m-MCFL(r)

MCFL's Case (naive application)

$L = h(D \cap R)$

- Original mcfg ($L \in m$-MCFL(r)):
  $\pi : A\langle x_1 a y_1 x_2,\ a\rangle \text{ :- } B\langle x_1, x_2\rangle, C\langle y_1\rangle$
- Multiple Dyck decoration ($D \cap R \in m$-MCFL(r)):
  $\hat{\pi} : A\langle\ \llbracket^{[1]}_{\pi,1}\, x_1 \rrbracket^{[1]}_{\pi,1} \llbracket_a \rrbracket_a \llbracket^{[1]}_{\pi,2}\, y_1 \rrbracket^{[1]}_{\pi,2} \llbracket^{[2]}_{\pi,1}\, x_2 \rrbracket^{[2]}_{\pi,1}\ ,\ \llbracket_a \rrbracket_a\ \rangle$
  $\text{ :- } B\langle x_1, x_2\rangle, C\langle y_1\rangle$
- Control right linear grammar (R):
  - $T \to \varepsilon,$
  - $A^{[1]} \to \llbracket^{[1]}_{\pi,1} B^{[1]},\quad T \to \rrbracket^{[1]}_{\pi,1} \llbracket_a \rrbracket_a \llbracket^{[2]}_{\pi,2} C^{[1]},$
    $T \to \rrbracket^{[1]}_{\pi,2} \llbracket^{[2]}_{\pi,1} B^{[2]},\quad T \to \rrbracket^{[2]}_{\pi,1} T,$
  - $A^{[2]} \to \llbracket_a \rrbracket_a T,$

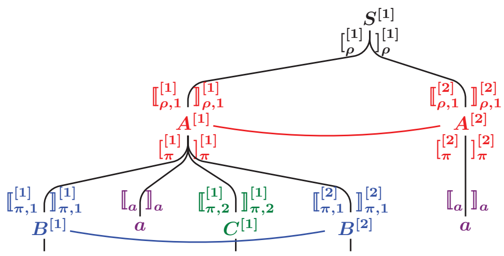
MCFL's Case (naive application)

$L = h(D \cap R)$

- Original mcfg ($L \in m$-MCFL(r)):
  $\pi : A\langle x_1 a y_1 x_2,\ a\rangle \text{ :- } B\langle x_1, x_2\rangle, C\langle y_1\rangle$
- Multiple Dyck decoration ($D \cap R \in m$-MCFL(r)):
  $\hat{\pi} : A\langle [^{[1]}_{\pi}\ \llbracket^{[1]}_{\pi,1}\, x_1 \rrbracket^{[1]}_{\pi,1} \llbracket_a \rrbracket_a \llbracket^{[1]}_{\pi,2}\, y_1 \rrbracket^{[1]}_{\pi,2} \llbracket^{[2]}_{\pi,1}\, x_2 \rrbracket^{[2]}_{\pi,1}\ ]^{[1]}_{\pi},\ [^{[2]}_{\pi} \llbracket_a \rrbracket_a]^{[2]}_{\pi} \rangle$
  $\text{ :- } B\langle x_1, x_2\rangle, C\langle y_1\rangle$
- Control right linear grammar (R):
  - $T \to \varepsilon,$
  - $A^{[1]} \to [^{[1]}_{\pi} \llbracket^{[1]}_{\pi,1} B^{[1]},\quad T \to \rrbracket^{[1]}_{\pi,1} \llbracket_a \rrbracket_a \llbracket^{[2]}_{\pi,2} C^{[1]},$
    $T \to \rrbracket^{[1]}_{\pi,2} \llbracket^{[2]}_{\pi,1} B^{[2]},\quad T \to \rrbracket^{[2]}_{\pi,1}]^{[1]}_{\pi} T,$
  - $A^{[2]} \to [^{[2]}_{\pi} \llbracket_a \rrbracket_a]^{[2]}_{\pi} T,$
- $[^{[1]}_{\pi},]^{[1]}_{\pi}, [^{[2]}_{\pi},]^{[2]}_{\pi}$ must co-occur in the multiple Dyck language D

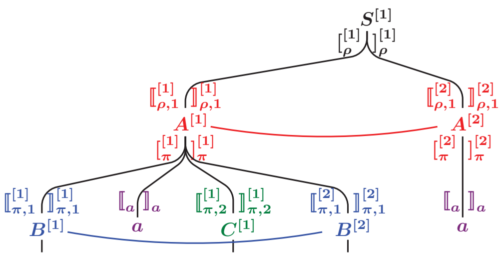
MCFL's Case (right decoration)

- $\rho : S\langle x_1, x_2\rangle \text{ :- } A\langle z_1, z_2\rangle,$
  $\pi : A\langle y_1 a z_1 y_2,\ a\rangle \text{ :- } B\langle x_1, x_2\rangle, C\langle y_1\rangle$



MCFL's Case (right decoration)

- $\rho : S\langle x_1, x_2\rangle :- A\langle z_1, z_2\rangle$,
  $\pi : A\langle y_1 a z_1 y_2,\ a\rangle :- B\langle x_1, x_2\rangle, C\langle y_1\rangle$,
  $\pi' : A\langle y_1 b z_1 y_2,\ b\rangle :- B\langle x_1, x_2\rangle, C\langle y_1\rangle$



Results

- Representation Theorem

$$L \in m\text{-MCFL}(r) \quad \text{if and only if} \quad L = h(D \cap R)$$

for some index alphabet Δ with $\dim(\Delta) = m$,
a regular language R over $\widehat{\Delta}$,
a homomorphism $h : \widehat{\Delta}^* \to \Sigma^*$,
where D is the multiple Dyck language on Δ of rank r.

Results

- Representation Theorem

$$L \in m\text{-MCFL}(r) \quad \text{if and only if} \quad L = h(D \cap R)$$

for some index alphabet Δ with $\dim(\Delta) = m$,
a regular language R over $\widehat{\Delta}$,
a homomorphism $h : \widehat{\Delta}^* \to \Sigma^*$,
where D is the multiple Dyck language on Δ of rank r.

- Generator Theorem
  For any $m, r \geq 1$ and Σ,
  there is an indexed alphabet Δ_0 and a homomorphism h_0 such that

$$L \in m\text{-MCFL}(r) \quad \text{if and only if} \quad L = h_0(D_0 \cap R)$$

for some regular language R,
where D_0 is the multiple Dyck language on Δ_0 of rank r.

Proof

- CFLs:
  - $\Delta_0 = \{\, [, \llbracket \,\} \cup \{\, \llbracket_a \mid a \in \Sigma \,\}$,
  - $\Delta = \{\, \llbracket_i \mid i = 1, \ldots, p \,\} \cup \{\, \llbracket_a \mid a \in \Sigma \,\}$,
  - $\llbracket_i \mapsto \llbracket \underbrace{[\ldots[}_{i\text{-times}} \llbracket.$

Proof

- CFLs:
  - $\Delta_0 = \{\, [, \llbracket \,\} \cup \{\, \llbracket_a \mid a \in \Sigma \,\}$,
  - $\Delta = \{\, \llbracket_i \mid i = 1, \ldots, p \,\} \cup \{\, \llbracket_a \mid a \in \Sigma \,\}$,
  - $\llbracket_i \mapsto \llbracket \underbrace{[\ldots[}_{i\text{-times}} \llbracket.$
- MCFLs:
  - $\Delta_0 = \{\, [_k, \llbracket_k \mid 1 \leq k \leq m \,\} \cup \{\, \llbracket_a \mid a \in \Sigma \,\}$
    with $\dim([_k) = \dim(\llbracket_k) = k$ and $\dim(\llbracket_a) = 1$,
  - $\Delta = \{\, \llbracket_{k,i} \mid i = 1, \ldots, p_k,\ 1 \leq k \leq m \,\} \cup \{\, \llbracket_a \mid a \in \Sigma \,\}$
    with $\dim(\llbracket_{k,i}) = k$,
  - $\llbracket^{[j]}_{k,i} \mapsto \llbracket^{[j]}_k \underbrace{[^{[j]}_k \cdots [^{[j]}_k}_{i\text{-times}} \llbracket^{[j]}_k.$

Thank you