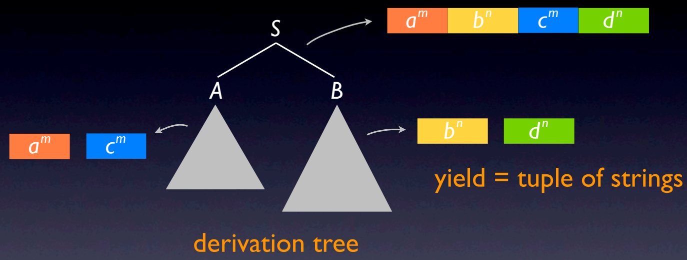
The Pumping Lemma for Well-Nested Multiple Context-Free Languages

Makoto Kanazawa
National Institute of Informatics
Tokyo, Japan

Multiple context-free grammars



$$S(x_1y_1x_2y_2) :- A(x_1,x_2), B(y_1,y_2).$$

$$A(\varepsilon,\varepsilon). \qquad B(\varepsilon,\varepsilon).$$

$$A(ax_1,cx_2) :- A(x_1,x_2). \qquad B(by_1,dy_2) :- B(y_1,y_2).$$

MCFGs have the same kind of derivation tree as CFGs, but the object produced by a derivation tree is a tuple of strings, rather than a string.
A nonterminal is like a predicate on strings.
A rule is a Horn clause.

m-multiple context-free grammars

Seki et al. 1991

$N = \bigcup_{r \le m} N^{(r)}$ ranked alphabet

$G = (N, \Sigma, P, S)$

$S \in N^{(1)}$

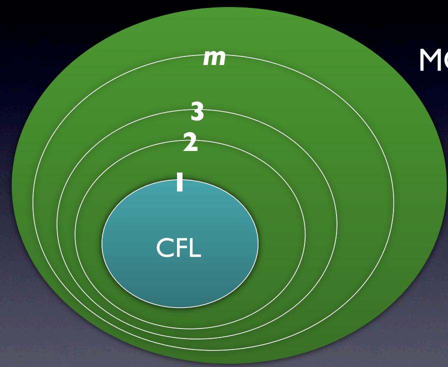
$$\underbrace{B(t_1,\dots,t_r)}_{\text{head}} :- \underbrace{B_1(x_{1,1},\dots,x_{1,r_1})}_{\text{subgoal}},\dots,\underbrace{B_n(x_{n,1},\dots,x_{n,r_n})}_{\text{subgoal}}.$$

- $B \in N^{(r)}, B_i \in N^{(r_i)}$
- $t_1 \dots t_r \in (\Sigma \cup X)^*$
- Each $x_{i,j}$ occurs at most once in $t_1 \dots t_r$

$$L(G) = \{ w \in \Sigma^* \mid P \vdash S(w) \}$$

X is the set of variables appearing in the right-hand side.
Use logic programming terminology.

An infinite hierarchy



$\mathrm{MCFL} = \bigcup_{m \ge 1} m\text{-MCFL}$

$$\{ a_1^n \dots a_{2m+1}^n \mid n \ge 0 \} \in (m+1)\text{-MCFL} - m\text{-MCFL}$$

MCFGs were introduced in the context of comp. ling., but natural.
Each level of the hierarchy is equivalently defined by various other formalisms, e.g., HR and yT_{fc}(REG).
Containment in LOGCFL, Parikh image semilinear.

The pumping lemma for MCFL

- A string $z \in L$ is k-pumpable in L if

  $z = u_0 v_1 u_1 v_2 u_2 \dots u_{k-1} v_k u_k$

  $v_1 v_2 \dots v_k \neq \epsilon$

  $u_0 v_1^i u_1 v_2^i u_2 \dots u_{k-1} v_k^i u_k \in L$ for every $i \geq 0$

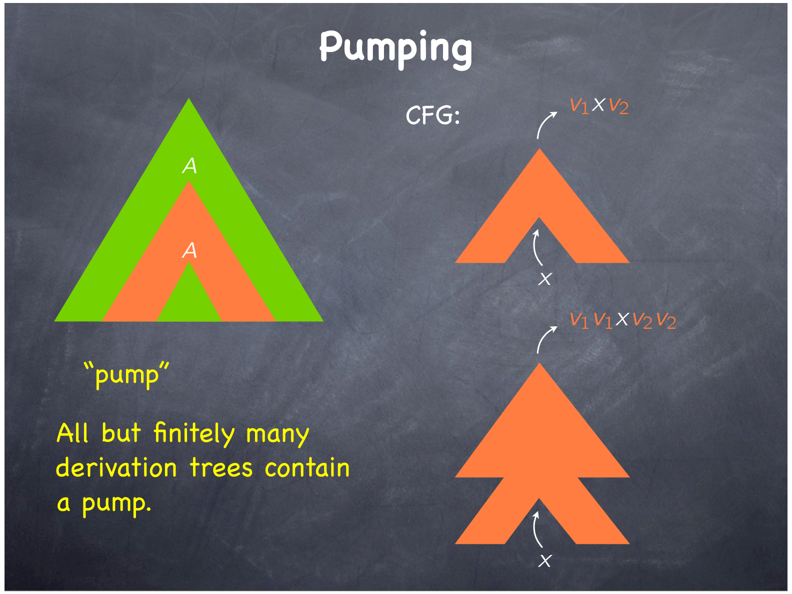
- **Theorem** (Seki et al. 1991). If L is an infinite m-MCFL, then there is a string $z \in L$ that is $2m$-pumpable.

- **Myth** (Radzinski 1991, Groenink 1997, Kracht 2003). If L is an m-MCFL, all but finitely many strings $z \in L$ are $2m$-pumpable.

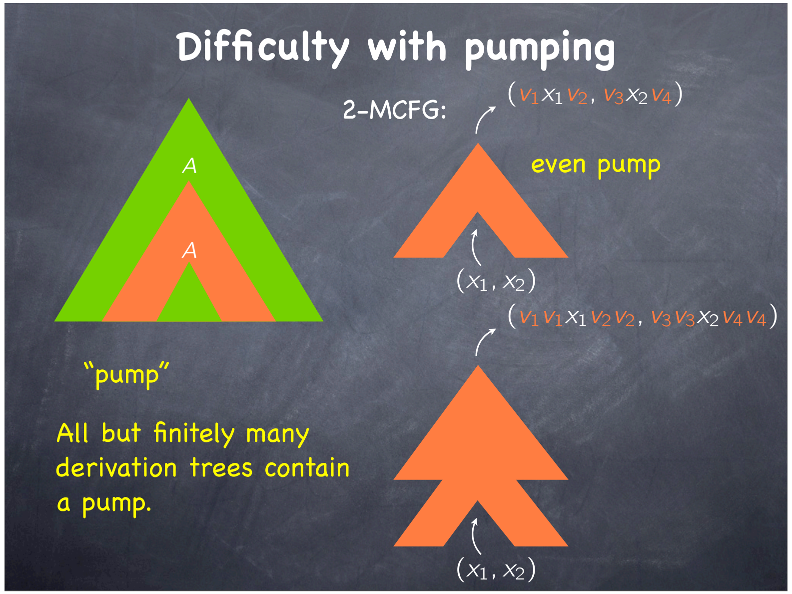
"universal pumping lemma"

Chinese number names (Radzinski)
crossed dependencies + coordination (Groenink)

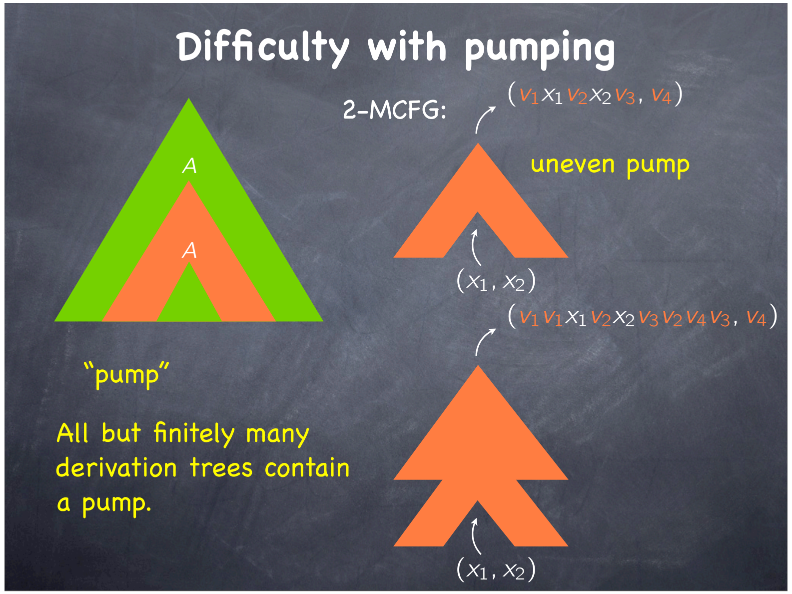
Seki et al.'s result is existential.
The Myth was appealed to in their attempts to show that these constructions go beyond the power of MCFGs.
Michaelis & Kracht (1997) showed the set of Chinese number names is not semilinear.



Let's see why it's not easy to prove the pumping lemma for MCFL.



Call such a pump an "even" pump because the components of an input tuple are evenly distributed among the components of the output tuple.
Not all pumps are even.

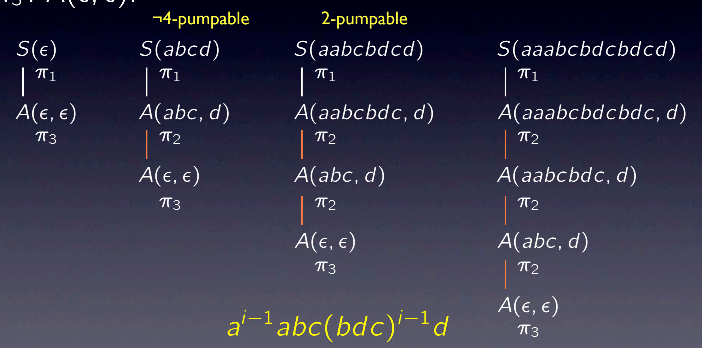


Example

$\pi_1: S(x_1 x_2) :- A(x_1, x_2).$

$\pi_2: A(a x_1 b x_2 c, d) :- A(x_1, x_2).$

$\pi_3: A(\epsilon, \epsilon).$



$a^{i-1}abc(bdc)^{i-1}d$

¬(All but finitely many derivation trees contain an even pump)

A concrete example of a grammar that gives rise to uneven pumps.
It almost seems as if aabcbdcd is 2-pumpable by accident.

The universal pumping lemma

- remains open for m-MCFGs
- holds for the subclass consisting of well-nested m-MCFGs

Well-nested MCFGs

✗ $S(x_1 y_1 x_2 y_2) :- A(x_1, x_2), B(y_1, y_2).$

✓ $S(x_1 y_1 y_2 x_2) :- A(x_1, x_2), B(y_1, y_2).$

✗ $C(x_1 y_1, y_2 z_1, z_2 x_2 z_3) :- A(x_1, x_2), B(y_1, y_2), C(z_1, z_2, z_3).$

✓ $C(z_1 x_1, x_2 z_2, y_1 y_2 z_3) :- A(x_1, x_2), B(y_1, y_2), C(z_1, z_2, z_3).$

Cf. Kuhlmann 2007

Well-nested MCFGs

- An MCFG rule

  $B(t_1, \ldots, t_r) :- B_1(x_{1,1}, \ldots, x_{1,r_1}), \ldots, B_n(x_{n,1}, \ldots, x_{n,r_n}).$

  is well-nested iff for every *i, i', j, j', k, k'* (*i*≠*i'*), it holds that

  $t_1 \ldots t_r \notin (\Sigma \cup X)^* x_{i,j} (\Sigma \cup X)^* x_{i',j'} (\Sigma \cup X)^* x_{i,k} (\Sigma \cup X^*) x_{i',k'} (\Sigma \cup X)^*$

- The well-nested (m-)MCFGs are the same as coupled-context-free grammars (Hotz & Pitsch 1995) (of rank m).

Coupled-context-free grammars take a top-down view of rules as rewriting instructions.

Well-nested vs. general MCFGs

- Universal recognition problem (Kaji et al. 1992, Satta 1992, Hotz & Pitsch 1995)

m-MCFG$_{wn}$	P-complete
m-MCFG	NP-complete ($m \geq 2$)

- **Theorem** (new). Well-nested (m-)MCFGs are equivalent to non-duplicating macro grammars (Fischer 1968) (of rank $m-1$).

m-MCFL vs. m-MCFL$_{wn}$

$\mathrm{RESP}_2 = \{ a_1^i a_2^i b_1^j b_2^j a_3^i a_4^i b_3^j b_4^j \mid i, j \geq 0 \}$ Weir 1989

$\mathrm{RESP}_2 \in$ 2-MCFL $-$ 2-MCFL$_{wn}$ Seki et al. 1991

$\mathrm{RESP}_m = \{ a_1^i a_2^i b_1^j b_2^j \ldots a_{2m-1}^i a_{2m}^i b_{2m-1}^j b_{2m}^j \mid i, j \geq 0 \}$

$\mathrm{RESP}_m \in m$-MCFL $- m$-MCFL$_{wn}$ for $m \geq 2$

Seki and Kato 2008

$\mathrm{RESP}_m \in 2m$-MCFL$_{wn}$

MCFL vs. MCFL$_{wn}$

- **Theorem** (Seki and Kato 2008). For all $m \geq 2$, m-MCFL$_{wn} \subsetneq m$-MCFL.
- **Theorem** (Staudacher 1993, Michaelis 2005). MCFL$_{wn} \subsetneq$ MCFL.

$$\{ w_1 \ldots w_n z_n w_n z_{n-1} \ldots z_1 w_1 z_0 w_1^R \ldots w_n^R \mid n \in \mathbb{N}, w_i \in \{a,b\}^+, z_n \ldots z_0 \in \underbrace{D_{[,]}^*}_{\text{Dyck language over [,]}} \} \in \mathrm{MCFL} - \mathrm{MCFL}_{wn}$$

There's just one example in the literature that purportedly shows the inclusion to be strict.

The pumping lemma for m-MCFG$_{wn}$

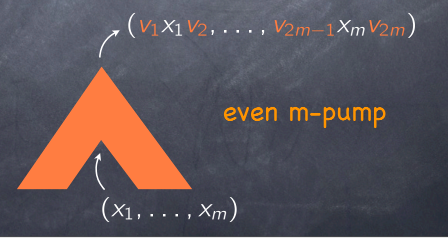
- If G is a w.n. m-MCFG ($m \geq 2$),

  { T | T is a derivation tree of G without even pumps }

  may not be finite.

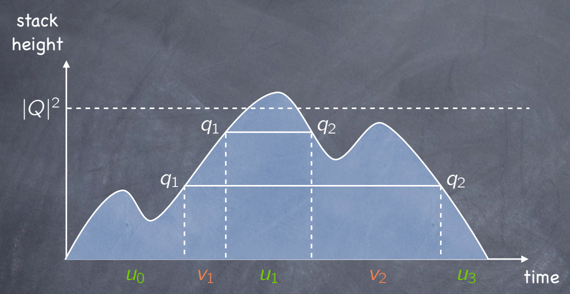
- But there is a w.n. (m-1)-MCFG generating

  { yield(T) | T is a derivation tree of G without even m-pumps }.



If the derivation tree contains an even m-pump, the string is 2m-pumpable.
Otherwise, the string is in the language of some w.n. (m-1)-MCFG, and therefore is 2(m-1)-pumpable.
Proof by induction on m.

The pumping lemma for PDA

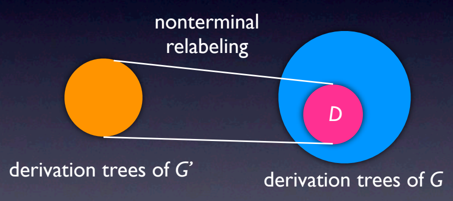


- ¬(All but finitely many accepting computations reach stack height $|Q|^2$)
- { w | w has an accepting computation that doesn't reach stack height $|Q|^2$ } is regular.

The proof generalizes to linear indexed grammars (CL_2) and to CL_k.

Proof of the pumping lemma

- **Lemma.** Let G be a w.n. m-MCFG, and let D be the set of derivation trees of G without even m-pumps. There is a w.n. m-MCFG G' without even m-pumps that generates $\{ \text{yield}(T) \mid T \in D \}$.



- **Lemma.** Any w.n. m-MCFG G without even m-pumps has an equivalent w.n. $(m-1)$-MCFG.

h is extended to h: T_P -> T_P'.
G' is well-nested if G is.
Implies yield(T) = yield(h(T))

Proof of the theorem

w.n. m-MCFG with no even m-pumps

↓

no m-proper rules

↓

total m-degree = 0

↓

w.n. $(m-1)$-MCFG

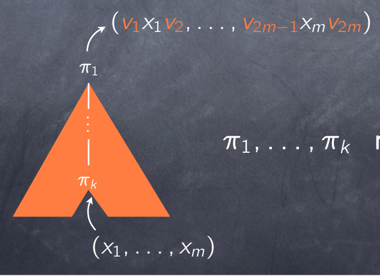
m-proper rules

$$\pi : B(t_1, \ldots, t_r) :- B_1(x_{1,1}, \ldots, x_{1,r_1}), \ldots, B_n(x_{n,1}, \ldots, x_{n,r_n}).$$

- π is m-proper on the i-th subgoal if

$$r_i = r = m,$$
$$t_j \in (\Sigma \cup X)^* x_{i,j} (\Sigma \cup X)^* \quad \text{for } j = 1, \ldots, m.$$

- Any even m-pump has the form



$\pi_1, \ldots, \pi_k$ m-proper

Elimination of m-proper rules

- **Lemma.** Any (w.n.) m-MCFG without even m-pumps has an equivalent (w.n.) m-MCFG that has no m-proper rules.

$$\pi_1: S(x_1x_2) :- B(x_1, x_2).$$
$$\pi_2: B(ax_1b, cx_2d) :- A(x_1, x_2). \quad \text{m-proper rule}$$
$$\pi_3: A(ax_1bx_2c, d) :- A(x_1, x_2).$$
$$\pi_4: A(\epsilon, \epsilon).$$

$\downarrow$ unfolding

$$\pi_1: S(x_1x_2) :- B(x_1, x_2).$$
$$\pi_2 \circ \pi_3: B(aax_1bx_2c_1b, cdd) :- A(x_1, x_2).$$
$$\pi_2 \circ \pi_4: B(ab, cd).$$
$$\pi_3: A(ax_1bx_2c, d) :- A(x_1, x_2).$$
$$\pi_4: A(\epsilon, \epsilon).$$

Unfolding

$$\pi: B(t_1, \ldots, t_r) :- C(y_1, \ldots, y_s), \Gamma$$
$$\pi': C(u_1, \ldots, u_s) :- \Delta$$

$\downarrow$ unfolding at the first subgoal

$$\pi \circ_1 \pi': B(t_1, \ldots, t_r)[y_i := u_i] :- \Delta, \Gamma$$

$$\pi: B(t_1, \ldots, t_r) :- B_1(x_{1,1}, \ldots, x_{1,r_1}), \ldots, B_n(x_{n,1}, \ldots, x_{n,r_n}).$$

Lemma.

$$P' = (P - \{\pi\}) \cup \{\, \pi \circ_i \pi' \mid \text{the head nonterminal of } \pi' \text{ is } B_i \,\}$$

is equivalent to P.

The operation of unfolding is familiar from logic programming.

w.n. m-MCFG with no even m-pumps

$\downarrow$ unfolding

no m-proper rules

$\downarrow$

total m-degree = 0

$\downarrow$

w.n. $(m-1)$-MCFG

Reduction of m-degrees

$$\pi: B(t_1, \ldots, t_r) :- B_1(x_{1,1}, \ldots, x_{1,r_1}), \ldots, B_n(x_{n,1}, \ldots, x_{n,r_n}).$$

- The m-degree of π =

$$\begin{cases} 0 & \text{if } r \neq m, \\ |\{\, i \mid r_i = m \,\}| & \text{if } r = m. \end{cases}$$

- **Lemma.** If $\pi: B(t_1, \ldots, t_m) :- C(y_1, \ldots, y_m), \Gamma$ is well-nested and not m-proper, then π can be replaced with

$$\pi_1: B(t'_1, \ldots, t'_m) :- D(z_1, \ldots, z_p), \Gamma_1$$
$$\pi_2: D(u_1, \ldots, u_p) :- C(y_1, \ldots, y_m), \Gamma_2$$

where D is a new nonterminal of arity p<m and

$$\Gamma = \Gamma_1, \Gamma_2,$$
$$\pi = \pi_1 \circ_1 \pi_2.$$

m-degree(π_1) < m-degree(π)
m-degree(π_2) = 0

Reduction of m-degrees

$$\pi : B(x_{1,1}y_1x_{2,1},\, x_{2,2}y_2ay_3b,\, cx_{1,2}d) :- C(y_1, y_2, y_3),\, A_1(x_{1,1}, x_{1,2}),\, A_2(x_{2,1}, x_{2,2}).$$

$$\downarrow$$

$$\pi_1 : B(x_{1,1}z_1,\, z_2b,\, cx_{1,2}d) :- D(z_1, z_2),\, A_1(x_{1,1}, x_{1,2}).$$
$$\pi_2 : D(y_1x_{2,1},\, x_{2,2}y_2ay_3) :- C(y_1, y_2, y_3),\, A_2(x_{2,1}, x_{2,2}).$$

$$\pi : B(x_{1,1}ax_{1,2}y_1x_{2,1},\, bx_{2,2}c,\, y_2y_3dx_{1,2}) :- C(y_1, y_2, y_3),\, A_1(x_{1,1}, x_{1,2}),\, A_2(x_{2,1}, x_{2,2}).$$

$$\downarrow$$

$$\pi_1 : B(z_1x_{2,1},\, bx_{2,2}c,\, z_2) :- D(z_1, z_2),\, A_2(x_{2,1}, x_{2,2}).$$
$$\pi_2 : D(x_{1,1}ax_{1,2}y_1,\, y_2y_3dx_{1,2}) :- C(y_1, y_2, y_3),\, A_1(x_{1,1}, x_{1,2}).$$

25

The 3-degree of π is 1.

D has arity < 3 because π is not 3-proper.
The converse of unfolding.
Well-nestedness is crucial here.

w.n. m-MCFG with no even m-pumps

↓ unfolding

no m-proper rules

↓ unfolding^{-1}

total m-degree = 0

↓

w.n. $(m-1)$-MCFG

26

Elimination of nonterminals of arity m

$$\pi : B(t_1, \ldots, t_r) :- C(y_1, \ldots, y_m), \Gamma.$$
$$\pi' : C(u_1, \ldots, u_m) :- \Delta.$$

$$\downarrow$$

$$\pi \circ_1 \pi' : B(t_1, \ldots, t_r)[y_i := u_i] :- \Delta, \Gamma.$$

27

Δ contains no arity-m nonterminal.

w.n. m-MCFG with no even m-pumps

↓ unfolding

no m-proper rules

↓ unfolding^{-1}

total m-degree = 0

↓ unfolding

w.n. $(m-1)$-MCFG

28

Conclusion

- **Theorem.** If L is a well-nested *m*-MCFL, all but finitely many $z \in L$ are 2*m*-pumpable.

$z = u_0 v_1 u_1 v_2 u_2 \dots u_{2m-1} v_{2m} u_{2m}$

$v_1 v_2 \dots v_k \neq \epsilon$

$u_0 v_1^i u_1 v_2^i u_2 \dots u_{2m-1} v_{2m}^i u_{2m} \in L \quad \text{for every } i \geq 0$

Conclusion

- **Theorem.** If L is a 2-MCFL, all but finitely many $z \in L$ are 4-pumpable.
- **Open question.** Does every *m*-MCFG without even *m*-pumps have an equivalent (*m*−1)-MCFG?