# Open Problems on Multiple Context-Free Grammars

# Double Copying Theorem for $\mathrm{MCFL}_{wn}$

$$\{\, w\#w \mid w \in L_0 \,\} \in \mathrm{MCFL}_{wn}$$

⇕

$$\{\, w\#w \mid w \in L_0 \,\} \in \mathrm{MCFL}(1)$$

⇕

$$L_0 \in \mathrm{MCFL}(1)$$

$\mathrm{EDT0L}_{FIN} = \mathrm{MCFL}(1)$

non-branching



The first open question concerns the copying theorem.
This theorem talks about the entire hierarchy, not each of its levels m-$\mathrm{MCFL}_{wn}$.

# Double Copying Theorem for CFL

$\{ w\#w \mid w \in L_0 \} \in \mathrm{CFL}$

$\Updownarrow$

$\{ w\#w \mid w \in L_0 \} \in \mathrm{CFL}(1)$

$\Updownarrow$

$L_0$ is a finite union of languages of the form *rRs*, where *R* is a regular subset of $t^*$



Members of CFL(1) are usually called “linear context-free languages”.

# Double Copying Theorem for *m*-MCFL$_{wn}$

$$\{\, w\#w \mid w \in L_0 \,\} \in m\text{-MCFL}_{wn}$$

$$\Updownarrow$$

$$\{\, w\#w \mid w \in L_0 \,\} \in m\text{-MCFL}(1)$$

$$\Downarrow$$

$$L_0 \in m\text{-MCFL}(1)$$

$$L_0 \in m\text{-MCFL}(1) \Rightarrow \{\, w\#w \mid w \in L_0 \,\} \in 2m\text{-MCFL}(1)$$

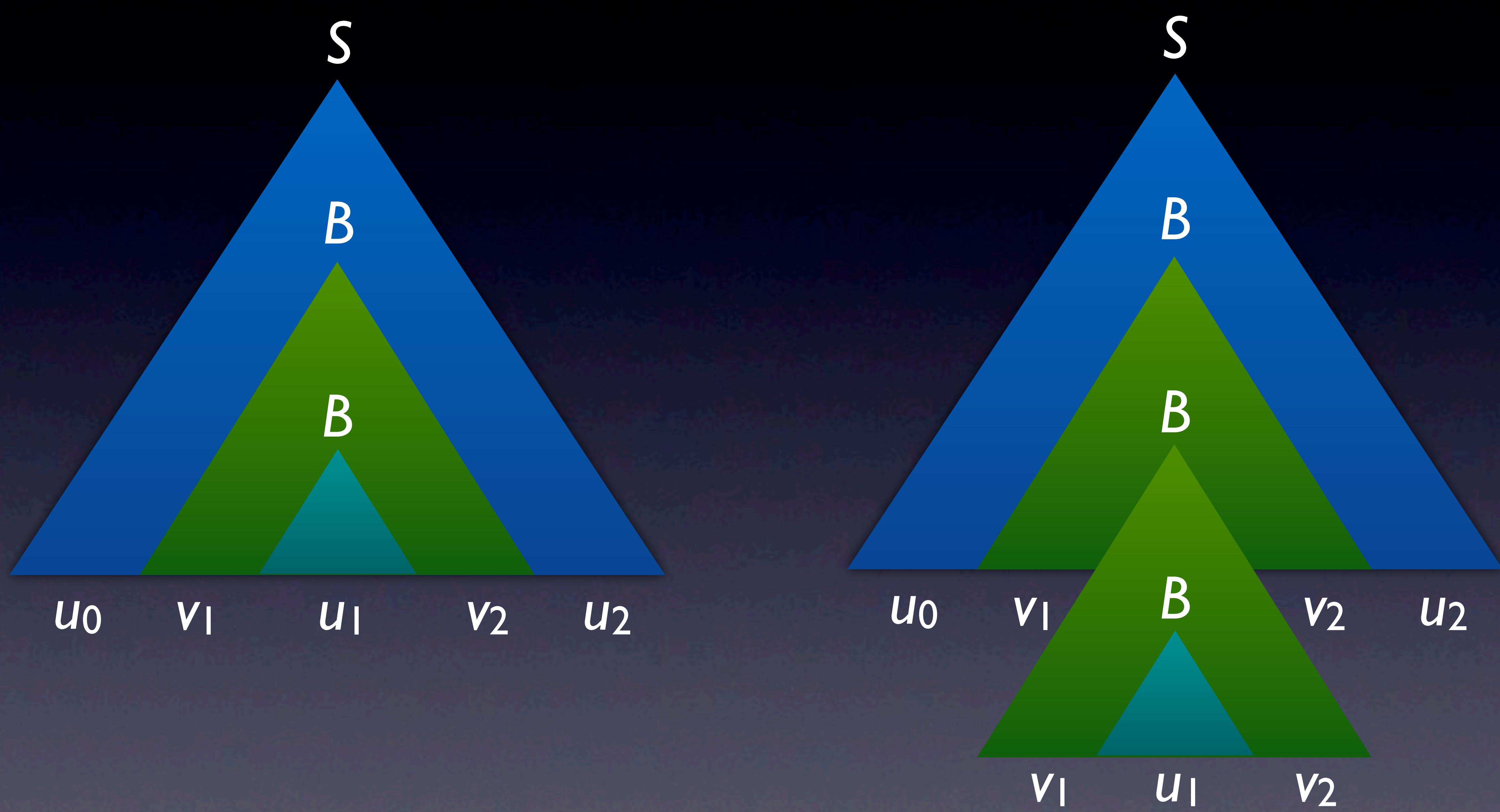
$\Leftarrow$ ?



This theorem talks about the entire hierarchy, not each of its levels m–MCFL$_{wn}$.

**C1.** $\{ w\#w \mid w \in L_0 \} \in 2m\text{-MCFL}(1) \Rightarrow L \in m\text{-MCFL}(1)$

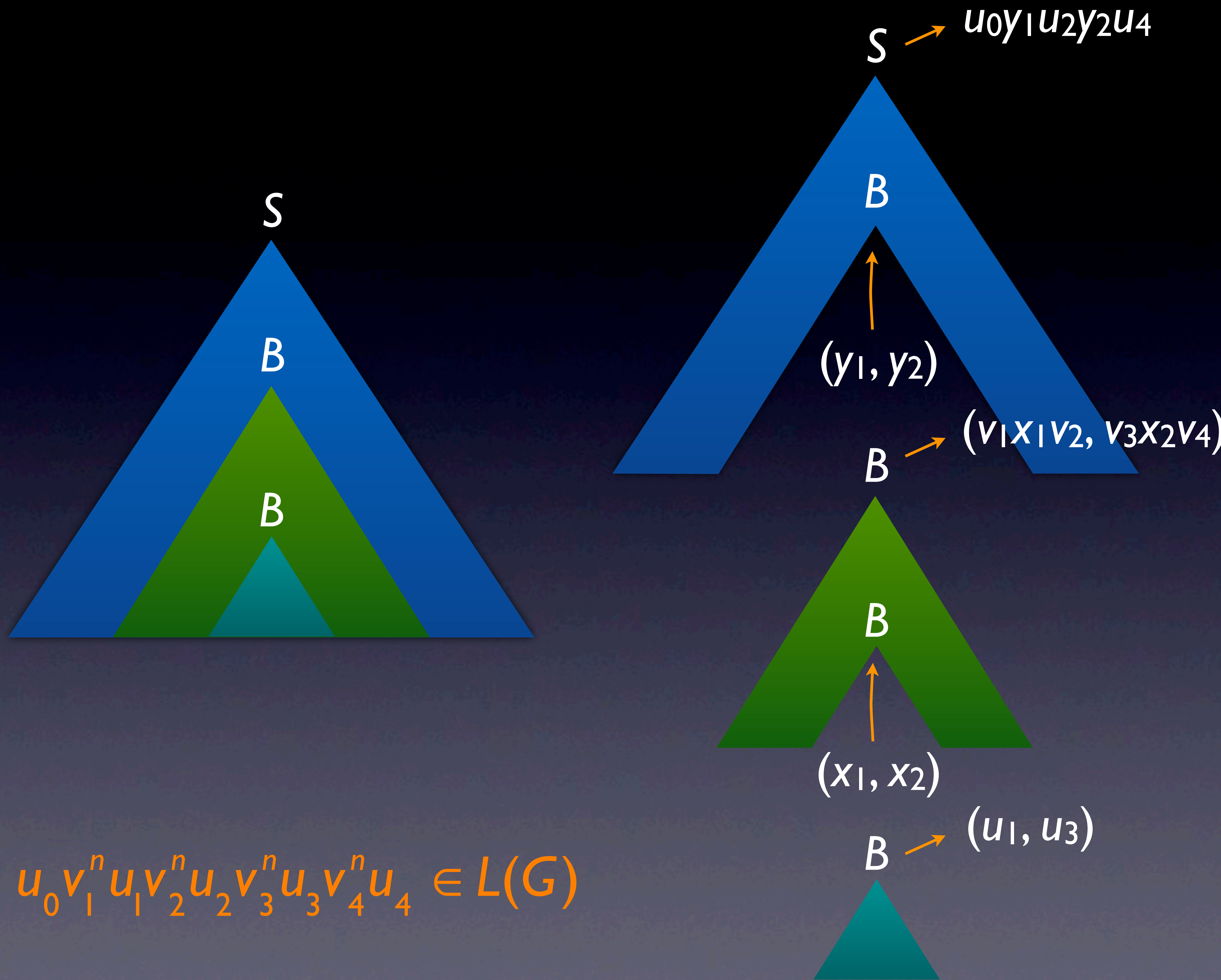
# Pumping




$u_0 v_1^n u_1 v_2^n u_2 \in L(G)$ for all $n \geq 0$



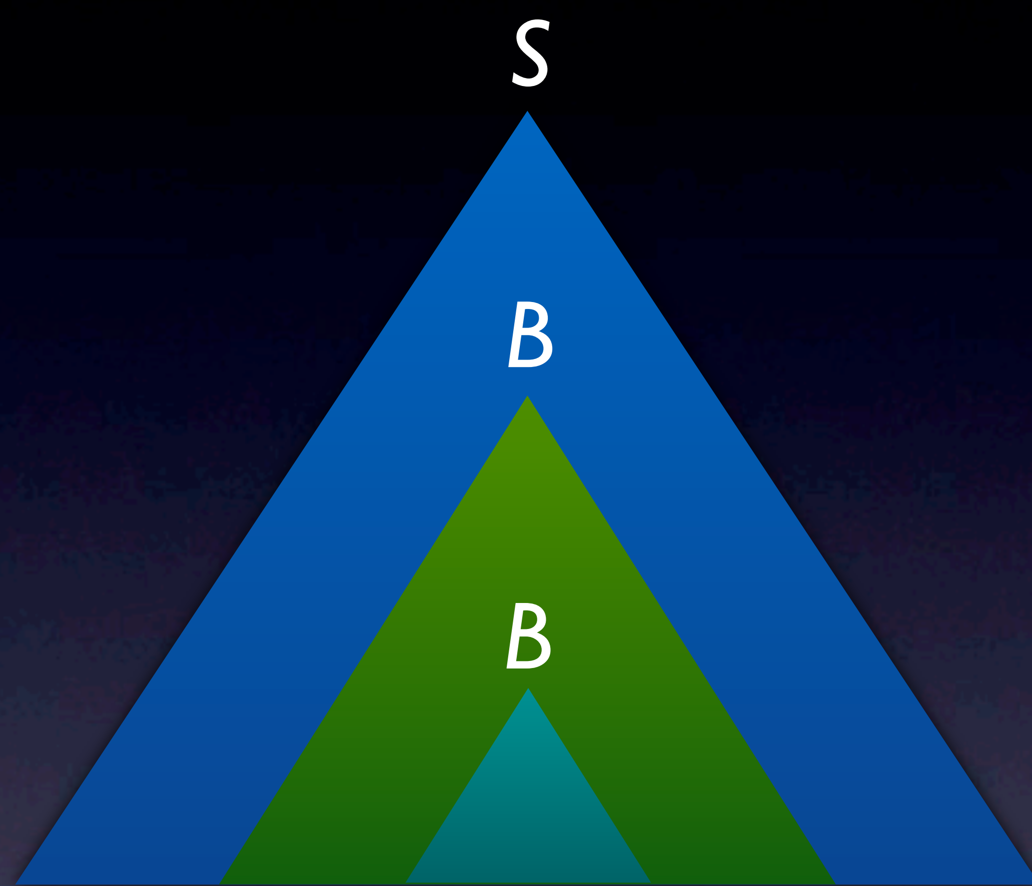
Let’s look at another question about MCFGs.
The case of CFG






The case of 2-MCFGs.
Is this the general picture?

# Difficulty with Pumping

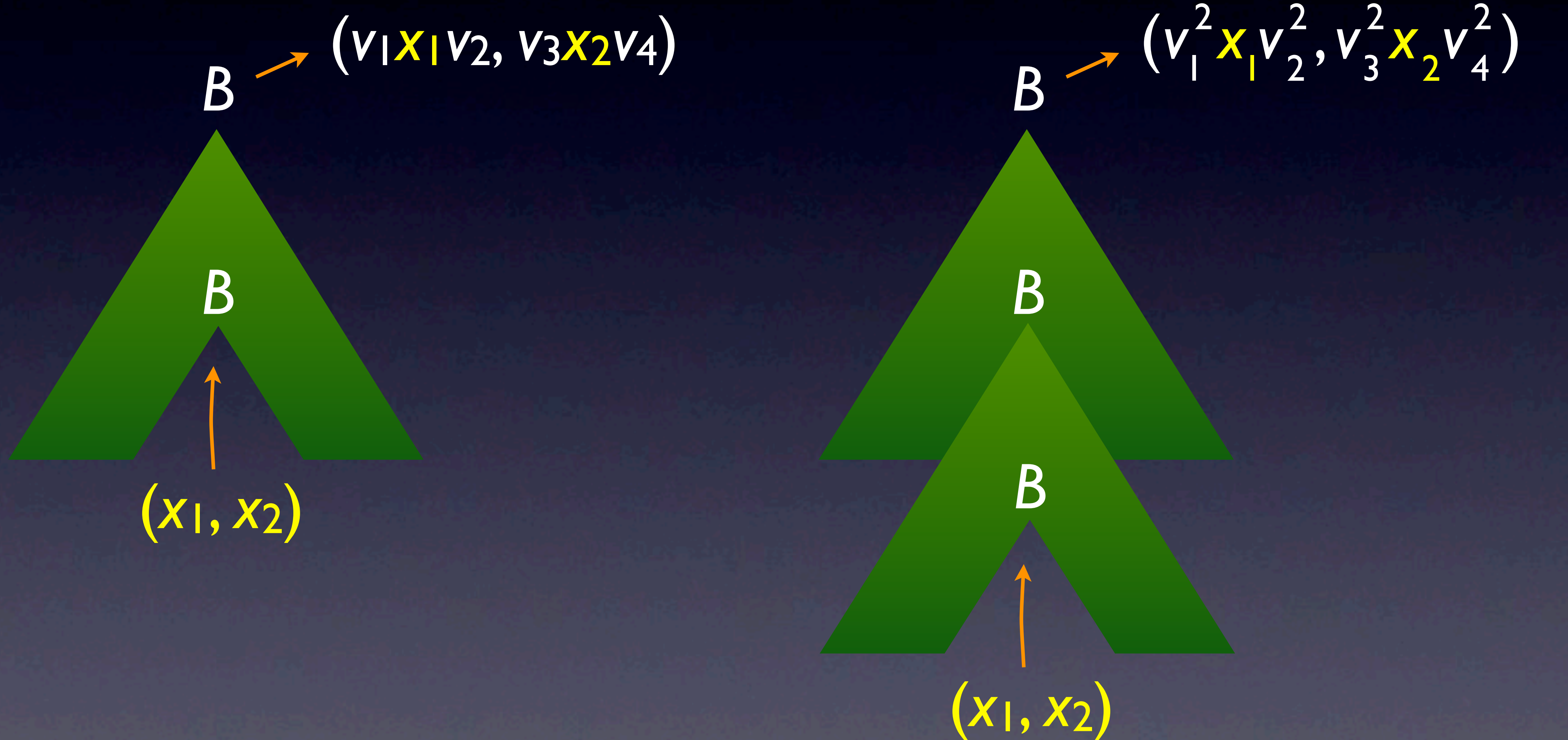



All but finitely many derivation trees contain a pump.



All sufficiently large derivation trees contain a part that can be repeated.

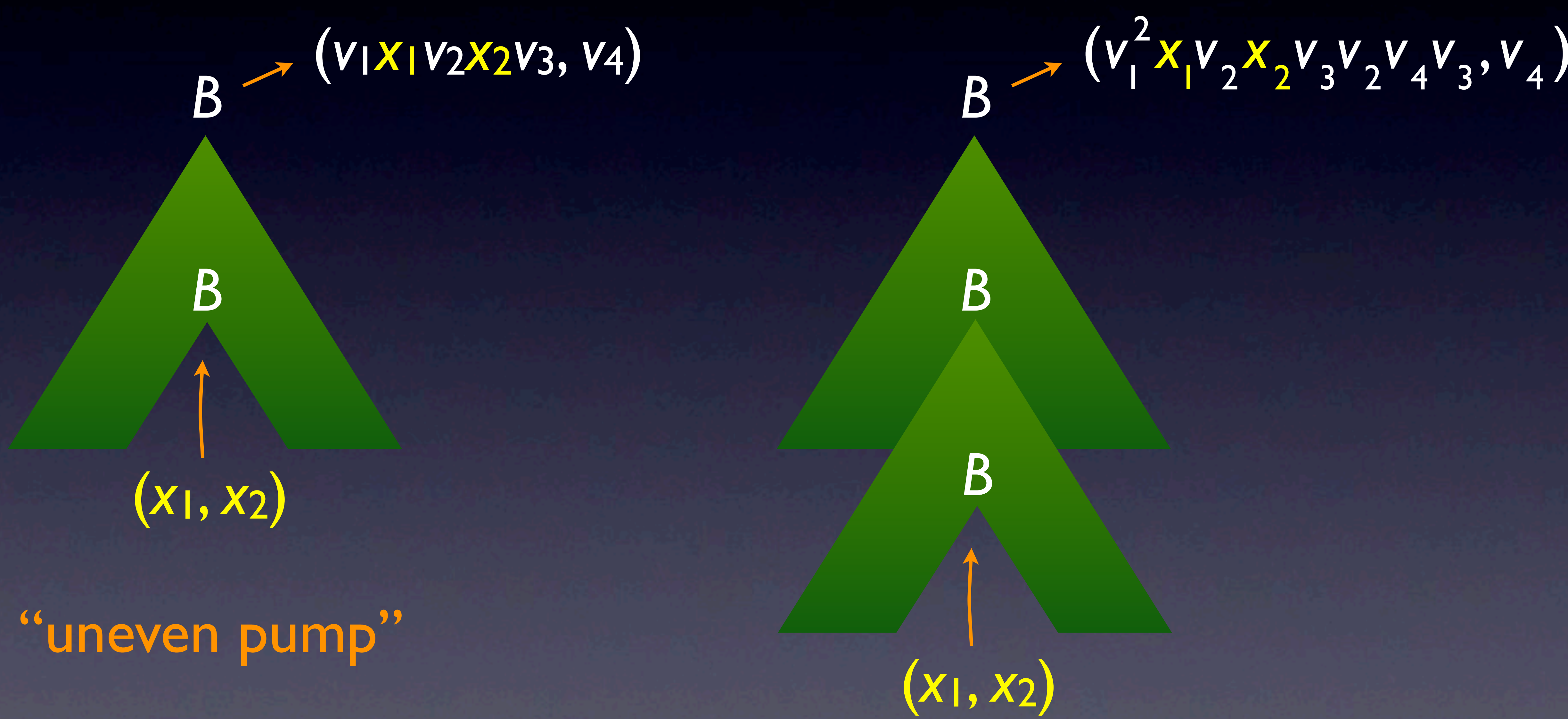
# Difficulty with Pumping






A derivation tree containing this pump yields a 4–pumpable string.

# Difficulty with Pumping






Rather complex pattern.

**C2.** $L \in m$-MCFL $\Rightarrow L$ is $2m$-iterative.

**Theorem** (Seki et al. 1991)**.**
$L \in m$-MCFL $\Rightarrow L$ is weakly $2m$-iterative.



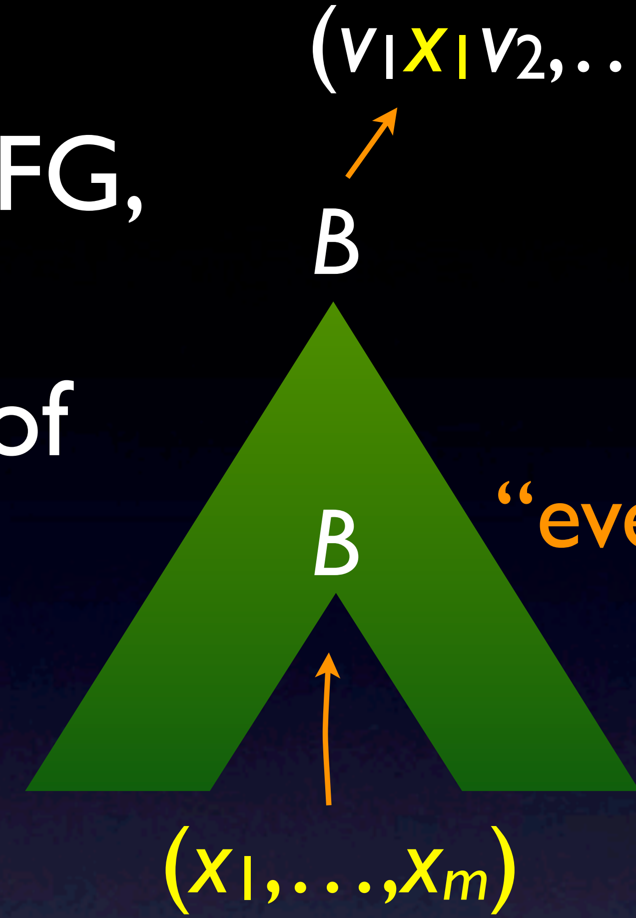
Many people erroneously believed that Seki et al. proved this conjecture.

**Theorem** (Kanazawa 2009)**.**
$L \in m\text{-MCFL}_{\mathrm{wn}} \Rightarrow L$ is $2m$-iterative.

- If $G$ is a well-nested $m$-MCFG,

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

  may not be finite.




- But there is a well-nested $(m-1)$-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 (disregarding finitely many exceptions).
Proof by induction on m.

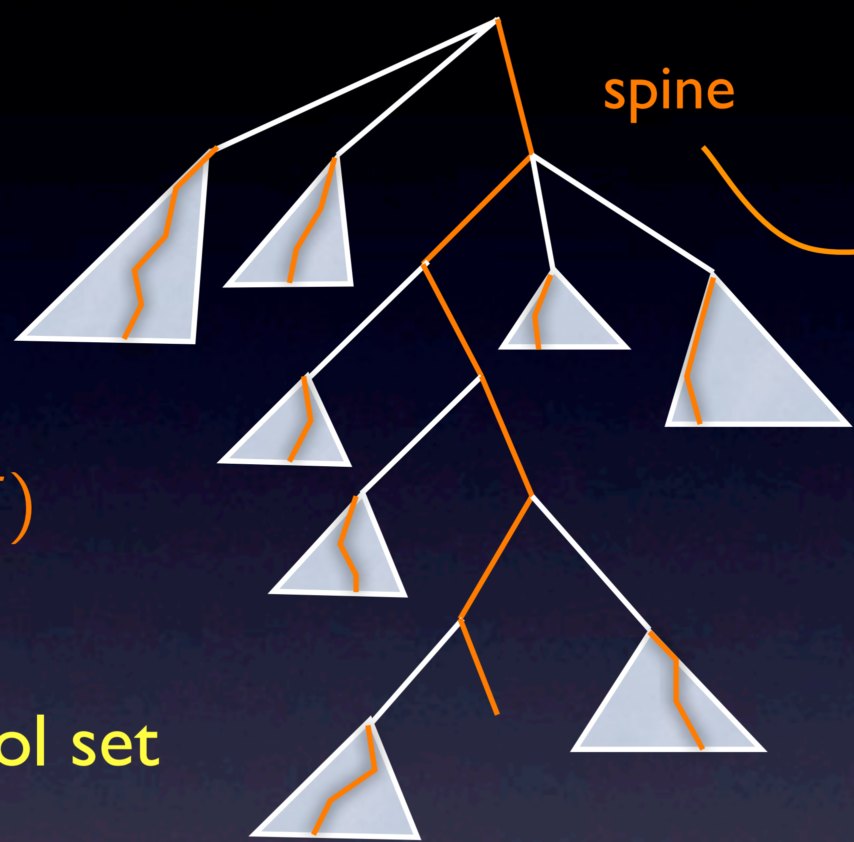
**Q2.1.** If G is an m-MCFG, there is an (m-1)-MCFG generating

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

?

# Control Languages

CFG derivation tree T



spine

head

$$\pi : A \to B_1 \ldots B_h \ldots B_n$$

control string = sequence of productions $\pi_1 \ldots \pi_n$

Spines($T$)

control set

non-regular tree language

$$DT(G, C) = \{ T \mid T \text{ is a complete derivation tree of } G \text{ and Spines}(T) \subseteq C \}$$

$$L(G, C) = \{ \text{yield}(T) \mid T \in DT(G, C) \}$$

$$\mathbf{C}_1 = \text{CFL} \quad \mathbf{C}_{k+1} = \{ L(G, C) \mid G \text{ is an LDG and } C \in \mathbf{C}_k \}$$



The Control Language Hierarchy of Weir (1988).

**Theorem** (Kanazawa and Salvati 2007).
$\mathbf{C}_k \subseteq 2^{k-1}$-MCFL

**C3.** $\bigcup_k \mathbf{C}_k \subsetneq$ MCFL



Both ∪k Ck and MCFL are closed under the "control" operation. Do the two fixed points coincide?

**Fact.**

$L_0 \in \mathbf{C}_k \Rightarrow \{ w\#w^R \mid w \in L_0 \} \in \mathbf{C}_{k+1}$

$L_0 \in \mathbf{C}_k \Rightarrow \{ w\#w \mid w \in L_0 \} \in \mathbf{C}_{k+2}$

$\mathbf{C}_3 - \mathrm{MCFL}_{\mathrm{wn}} \neq \varnothing$

**Q3.1.** $\mathrm{MCFL}_{\mathrm{wn}} \subseteq \bigcup_k \mathbf{C}_k$

?



If the answer to the question is no, is the closure of MCFLwn under control another fixed point, or does it coincide with MCFL?

**Theorem** (Salvati)**.** MIX $\in$ 2-MCFL

$$\mathrm{MIX}_k = \{\, w \in \{\, a_1,\ldots,a_k \,\}^* \mid \psi(w) = n \cdot (1,\ldots,1) \,\}$$

**Fact.** If $\mathcal{L}$ is a rational cone and contains $\mathrm{MIX}_k$ for all $k$, then $\mathcal{L}$ contains all permutaion-closed semilinear languages.



Open questions concerning MIX.

two-sided Dyck language

$$\mathrm{MIX} \equiv O_2 = \mathrm{Shuffle}(\hat{D}_1{}^*, \hat{D}_1{}^*)$$

**Fact.** $\mathrm{Shuffle}(D_1{}^*, D_1{}^*) \notin$ 2-MCFL



The one-sided analogue of O2 (the set of curves within the first quadrant) is not a 2-MCFL. This can be proved with the Pumping Lemma for 2-MCFL (Kanazawa 2009).

**Fact.**

$\mathrm{MIX} \notin 2\text{-}\mathrm{MCFL}(1)$

$\mathrm{MIX}_4 \notin 2\text{-}\mathrm{MCFL}_{\mathrm{wn}}$

$\mathrm{MIX}_{k+1} \notin k\text{-}\mathrm{MCFL}(1)$

$\mathrm{MIX}_{k+2} \notin k\text{-}\mathrm{MCFL}_{\mathrm{wn}}$



Appropriate refinements of the Pumping Lemma for MCFLwn give these facts.

**C4.** $\mathrm{MIX}_4 \notin \mathrm{MCFL}$

**C4.1.** $\mathrm{MIX} \notin \mathrm{MCFL}_{wn}$



Currently have no idea how to prove these.