XIII

Thirteenth Lecture
Automata & Formal Languages
3 November 2023

The context-free pumping lemma
("Bar-Hillel Lemma")

RPL: $|w| \geq n$, $w \in L$ $\Rightarrow$ $w = xyz$, $|xy| \leq n$, $|y| > 0$ s.t. $xy^kz \in L$

CFPL $|w| \geq n$, $w \in L$ $\Rightarrow$ $w = x\boxed{uyv}z$

$|uyv| \leq n$

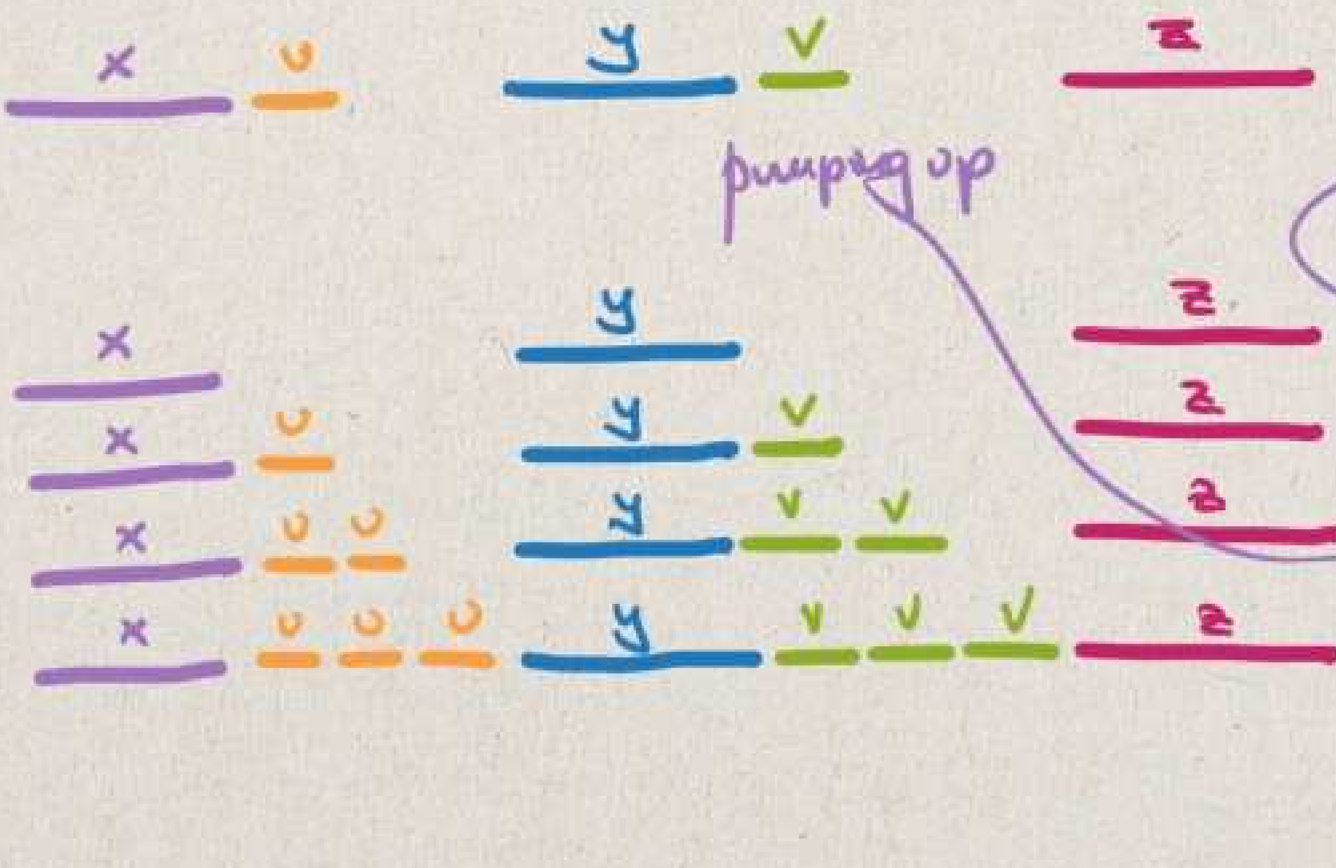
$|uv| > 0$

u, v are referred to as the pump.

Yehoshua Bar-Hillel

Born	September 8, 1915 Vienna, Austria-Hungary
Died	September 25, 1975 (aged 60) Jerusalem, Israel

Definition 3.10. Let $L \subseteq \mathbb{W}$ be a language. We say that *L satisfies the (context-free) pumping lemma with pumping number n* if for every word $w \in L$ such that $|w| \geq n$ there are words u, v, x, y, z such that $w = xuyvz$, $|uv| > 0$, $|uyv| \leq n$ and for all $k \in \mathbb{N}$, we have that $\boxed{xu^kyv^kz \in L.}$ We say that *L satisfies the (context-free) pumping lemma* if there is some n such that it satisfies the (context-free) pumping lemma with pumping number n.



$w = xuyvz$

"pumping down"

xyz

$xuyvz$

xu^2yv^2z

xu^3yv^3z

Definition 3.10. Let $L \subseteq \mathbb{W}$ be a language. We say that L *satisfies the (context-free) pumping lemma with pumping number* n if for every word $w \in L$ such that $|w| \geq n$ there are words u, v, x, y, z such that $w = xuyvz$, $|uv| > 0$, $|uyv| \leq n$ and for all $k \in \mathbb{N}$, we have that $xu^k yv^k z \in L$. We say that L *satisfies the (context-free) pumping lemma* if there is some n such that it satisfies the (context-free) pumping lemma with pumping number n.

Observations

(1) If L satisfies RPL, then it satisfies CFPL.

[$w = xuz$ as in RPL

$\Rightarrow$ $y := v := \varepsilon$

Then $xuyvz$ satisfies the condition of CFPL.]

(2) RPL identifies location of pump [since $|xy| \leq n$ and y is pump]; CFPL does not: no bound on $|x|$.

(3) Remember that RPL is NOT equivalent to "regular".

There are uncountably many languages satisfying RPL. [But only countably many regular languages.]

Therefore, CFPL does not characterise "context-free" either.

There are uncountably many languages that satisfy CFPL, but only countably many c-f languages.

Example (4c) on ES #1

$$L = \{ a^n b^n c^n ; n > 0 \}$$

On ES #1 showed that it is Type 1.
Show it's not context free:
we are going to show that it does not satisfy CFPL.

1. Suppose it does for pumping n.
2. Pick word $a^n b^n c^n = w$, $|w| = 3n \geq n$.
3. Apply CFPL and get
$w = xvyvz$ with $|vv| > 0$
$|vyv| \leq n$.

Case 1: vyv does not contain a's
Case 2: vyv does not contain c's.

In Case 1, if I pump down, I change the # of b's or c's or both, but not the # of a's.

$$a^n b^l c^m = xyz$$

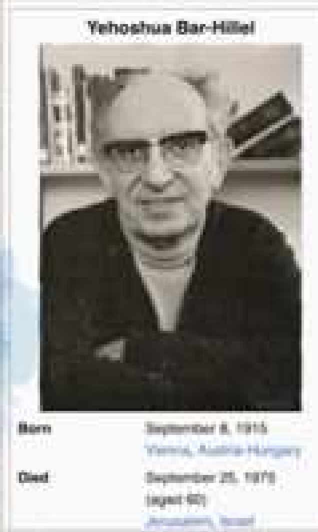
where $l, m \leq n$, $\min(l, m) < n$.

So $a^n b^l c^m \notin L$.

Case 2: entirely analogous.

THEOREM (Bar-Hillel)

Every context-free language satisfies the context-free pumping lemma for some pumping number n.



__Proof.__ Let G be context-free.

By Chomsky's Thm w.l.o.g., G is in CNF.

Let $G = (\Sigma, V, P, S)$ and $m := |V|$ and $n := 2^m$.

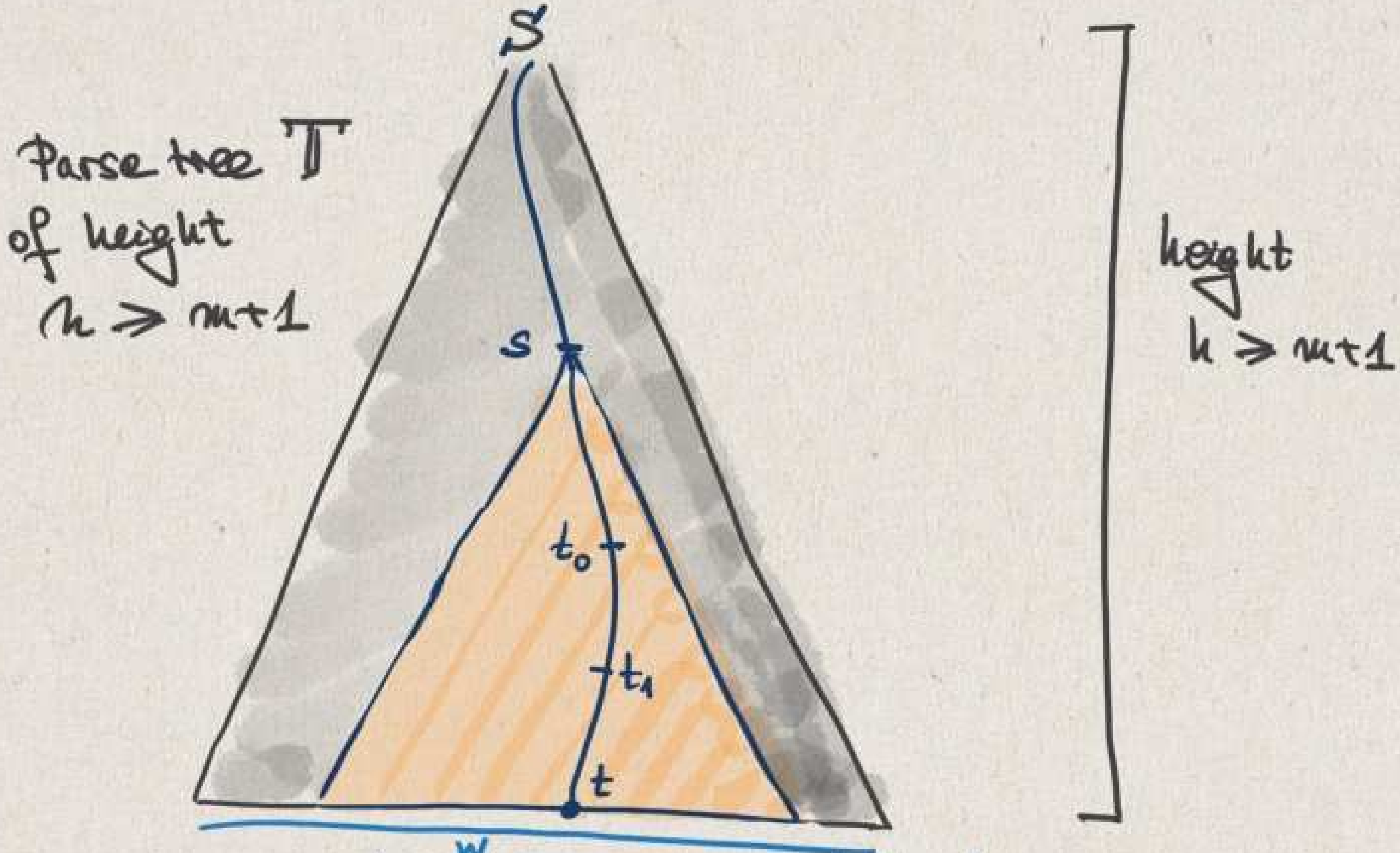
__Claim__ G satisfies CFPL with pumping # n.

Take $w \in \mathcal{L}(G)$ with $|w| \geq n = 2^m$.

By Lemma, we know that a parse tree of w must have height at least $m+1$.

Lecture XII, page 4

__Lemma__ If G CNF and T is a G-parse tree of w of height $h+1$. Then $|w| \leq 2^h$.



There is some leaf $t \in T$ s.t. $|t| = h$.
The branch leading from ε to t has length $h+1$
Find $s := t \upharpoonright h-(m+1)$, so that T_s has height $m+1$.

Consider $\{\ell(t \upharpoonright k);\ |s| \leq k \leq |t|\}$

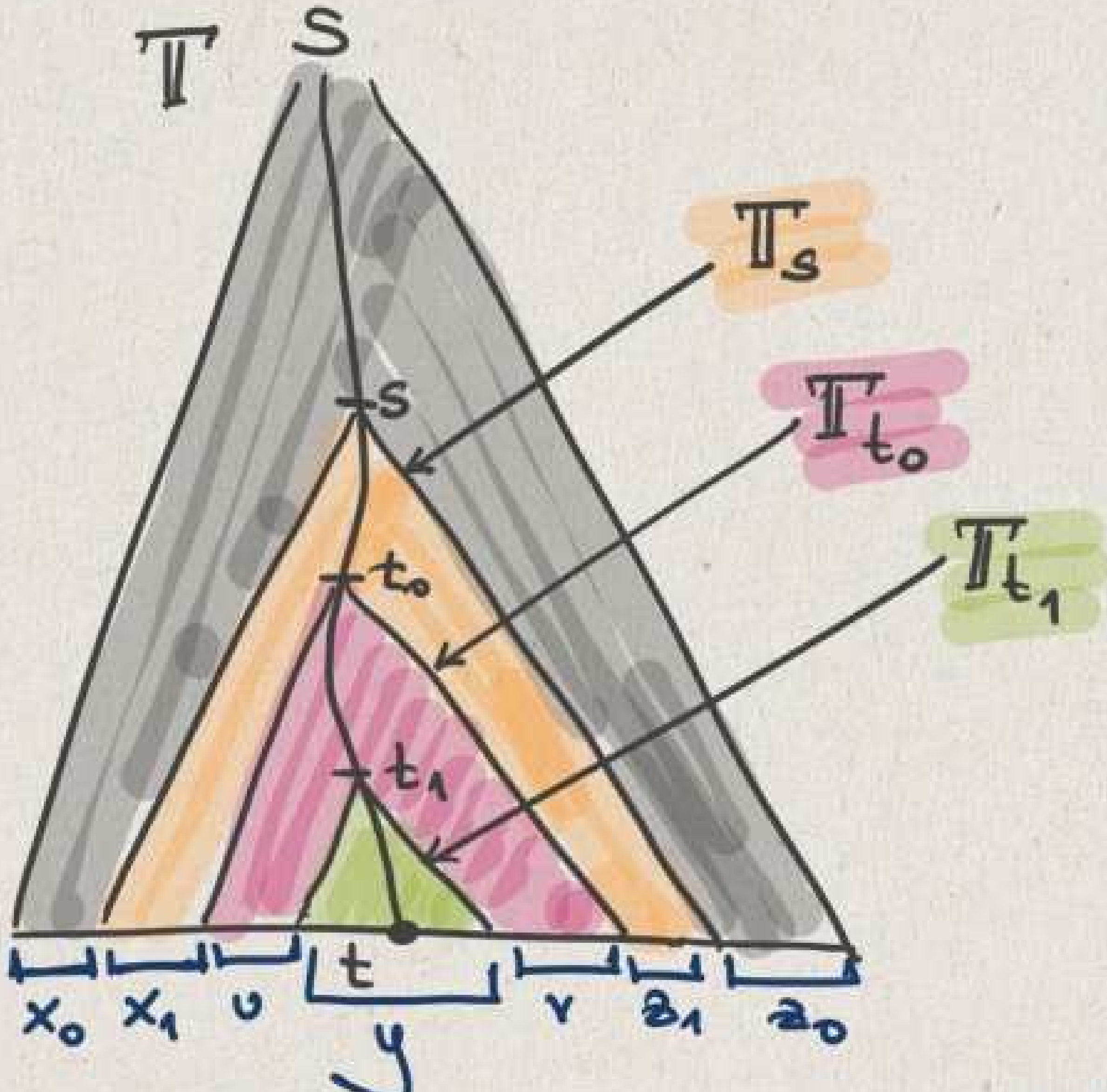
This has $m+1$ many variables and one letter.

Thus by PHP, there is a repeat: i.e., there are t_0, t_1 s.t.

$$s \subseteq t_0 \subsetneq t_1 \subsetneq t \quad \text{s.t.}$$

$$\ell(t_0) = \ell(t_1).$$

Note that $s = t_0$ is possible, $t_1 = t$ is not!



$\sigma_{\mathbb{T}} = x_0 \sigma_{\mathbb{T}_s} z_0$ $\quad$ $\sigma_{\mathbb{T}_{t_0}} = u \mathbb{T}_{t_1} v$

$\sigma_{\mathbb{T}_s} = x_1 \sigma_{\mathbb{T}_{t_0}} z_1$ $\quad$ $\sigma_{\mathbb{T}_{t_1}} = y$

Define $x := x_0 x_1$ $\quad$ $z := z_1 z_0$

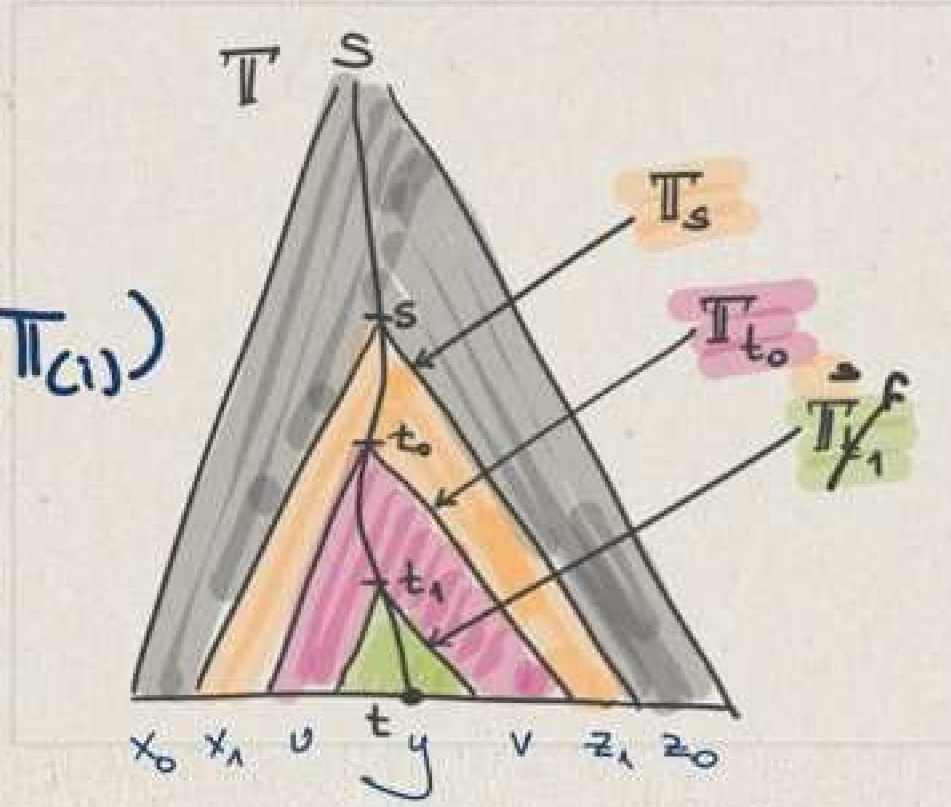
Clearly $|uv| > 0$ since $t_0 \neq t_1$, so something must be in $\sigma_{\mathbb{T}_{t_0}}$ that is not in $\sigma_{\mathbb{T}_{t_1}}$.

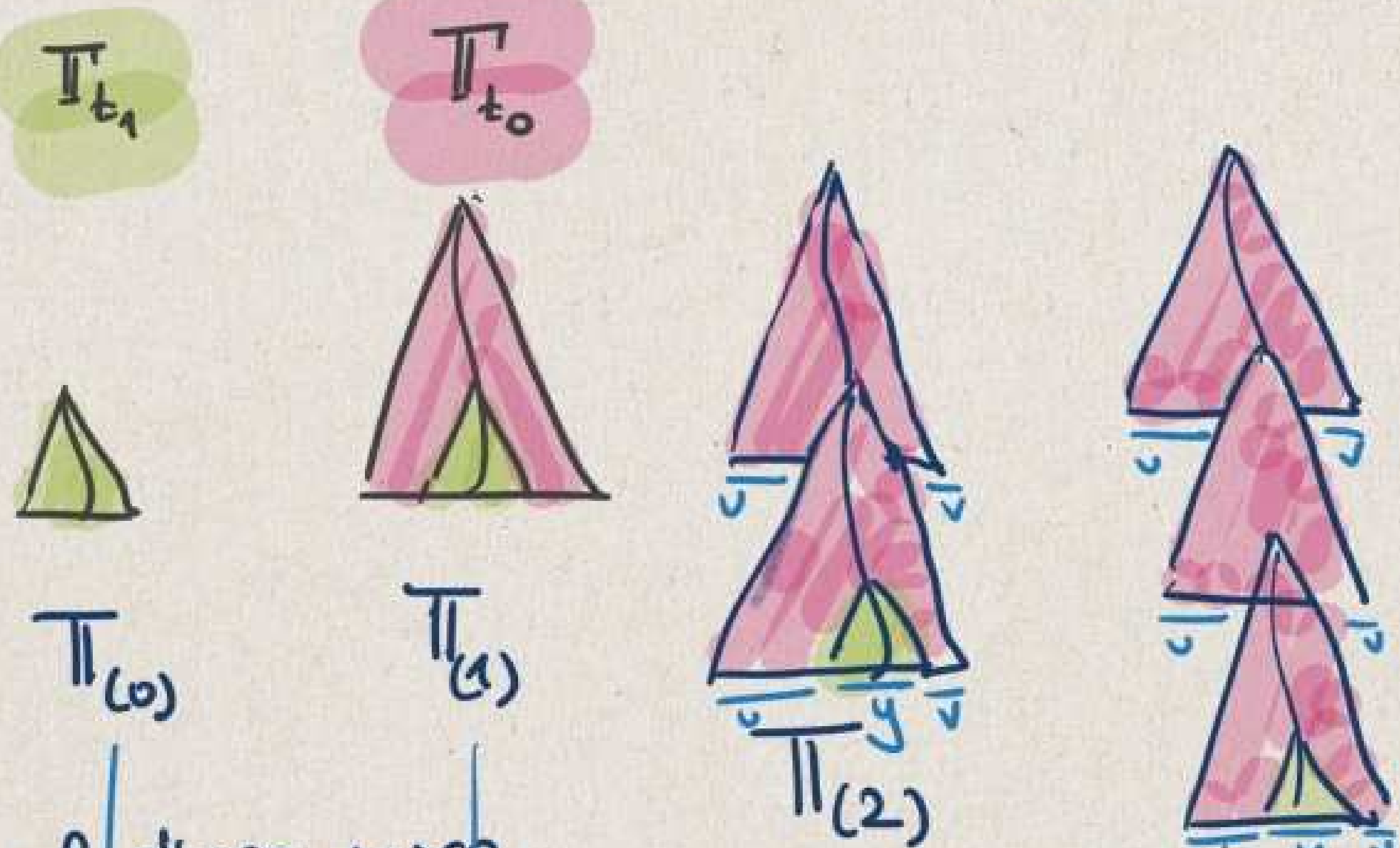
$|uyv| \leq n$: $uyv = \sigma_{\mathbb{T}_{t_0}}$, so it's a subword of $\sigma_{\mathbb{T}_s}$, but $\mathbb{T}_s$ had height $n+1$.

Define by recursion

$\mathbb{T}_{(0)} := \mathbb{T}_{t_1}$.

$\mathbb{T}_{(i+1)} := \mathrm{graft}(\mathbb{T}_{t_0}, t_1, \mathbb{T}_{(i)})$



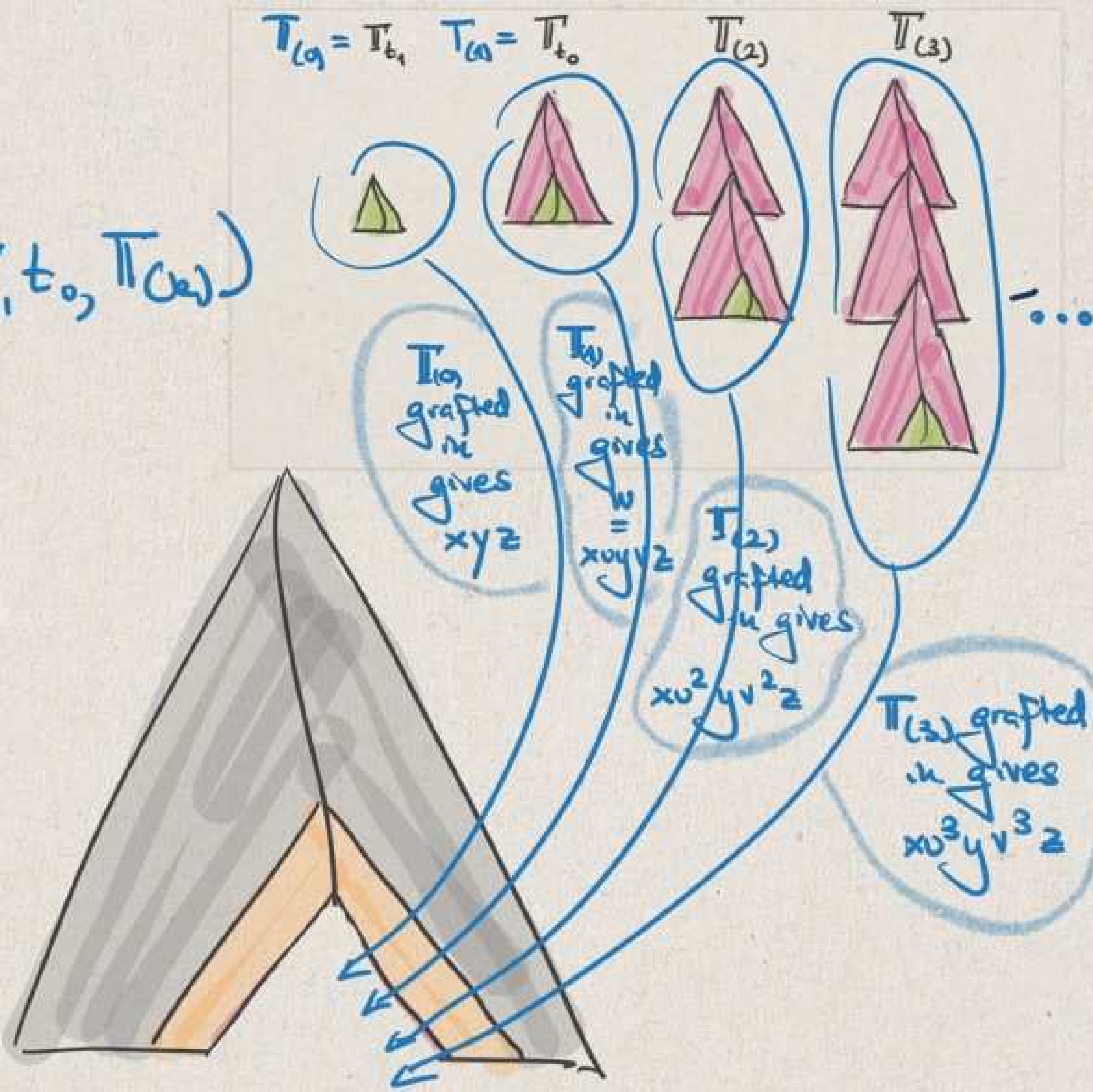


All of these parse trees start with $\ell(t_0)$.

y $\quad uyv \quad u^2yv^2 \quad u^3yv^3$

$\mathbb{T}_k :=$

$\triangleleft\text{graft}(\mathbb{T}, t_0, \mathbb{T}_{(k)})$



Consequently $\quad \sigma_{\mathbb{T}_k} = xu^kyv^kz \in \mathcal{L}(G)$

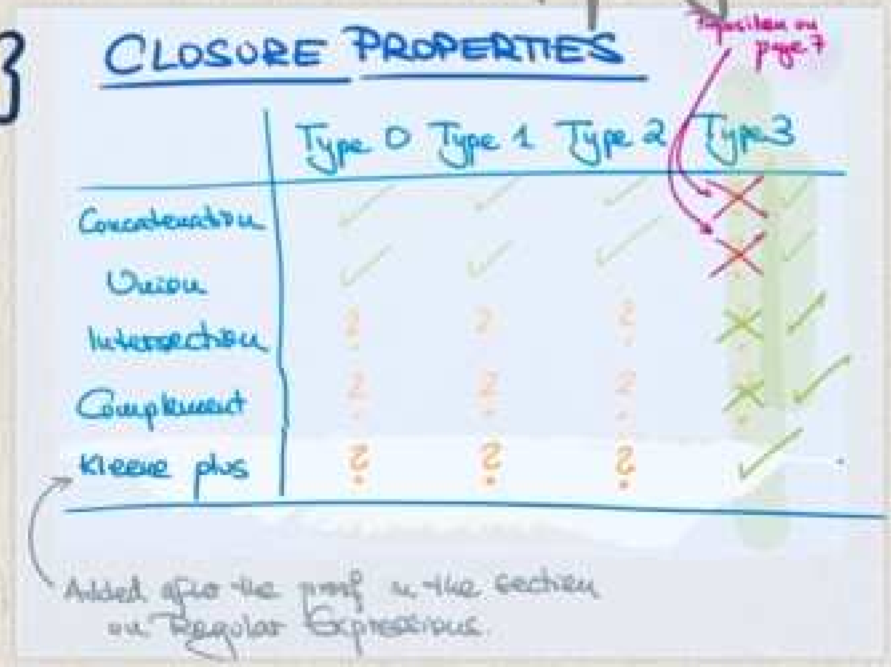
since there are

G-parse trees

q.e.d.

Closure properties

Lecture X, page 11:



CLOSURE PROPERTIES

	Type 0	Type 1	Type 2	Type 3
Concatenation	✓	✓	✓	✗ ✓
Union	✓	✓	✓	✗ ✓
Intersection	?	?	?	✗ ✓
Complement	?	?	?	✗ ✓
Kleene plus	?	?	?	✓

Added after the proof in the section on Regular Expressions.

Remember that $\{a^n b^n c^n ; n > 0\} = L$ is not context-free.

Corollary The c-f languages are not closed under intersections.

Proof.

$$L_0 := \{a^m b^m c^k ; m, k > 0\}$$
$$L_1 := \{a^k b^m c^m ; m, k > 0\}$$

Clearly, $L_0 \cap L_1 = L$.

However, $L_0 = L_{00} L_{01}$ with
$$L_{00} = \{a^m b^m ; m > 0\}$$
$$L_{01} = \{c^k ; k > 0\}$$

$L_1 = L_{10} L_{11}$ with
$$L_{10} = \{a^k ; k > 0\}$$
$$L_{11} = \{b^m c^m ; m > 0\}.$$

Now, $L_{00}, L_{01}, L_{10}, L_{11}$ are c-f & c-f languages closed under concat. Thus L_0, L_1 c-f.

q.e.d.

Corollary C-f languages are not closed under complement.

Pf. If so, then closure under union & complement implies closure under intersection. Contradiction!

q.e.d.

Remark ES#2 (28) shows c-f languages are closed under Kleene plus.

Decision problems

Know: Word problem solvable.

Remain: Emptiness & Equivalence Problem.

This (Emptiness) is precisely the same as in the regular case:

Lemma If L satisfies CFPL with pumping # n & $L \neq \emptyset$, then $\exists w \in L$ $|w| < n$.

Corollary The Emptiness Problem for c-f grammars is solvable.

Proof. STEP 1 Transform G to CNF.
Note that proof of Chomsky's theorem was an algorithm.

STEP 2 Take $|V|$ in CNF and calculate $2^{|V|} =: n$.

STEP 3 Check all words of length $\leq n$. [Word problem is solvable.]

q.e.d.

It turns out that the Equivalence Problem for c-f grammar is unsolvable. This result will not be proved in this course.

DECISION PROBLEMS

	Type 0	Type 1	Type 2	Type 3
WORD P.	?	✓	✓	✓
EMPTINESS P.	?	?	~~?~~ ✓	~~?~~ ✓
EQUIVALENCE P.	?	?	~~?~~ ✗	~~?~~ ✓

Proposition on page 7

CLOSURE PROPERTIES

	Type 0	Type 1	Type 2	Type 3
Concatenation	✓	✓	✓	~~?~~ ✓
Union	✓	✓	✓	~~?~~ ✓
Intersection	?	?	~~?~~ ✗	~~?~~ ✓
Complement	?	?	~~?~~ ✗	~~?~~ ✓
Kleene plus	?	?	~~?~~ ✓	✓