TDDD14 / TDDD85 – Lecture 11

Closure Properties and Pumping Lemma

August Ernstsson, 2026 (based on lecture notes by Jonas Wallgren)

LINKÖPING UNIVERSITY

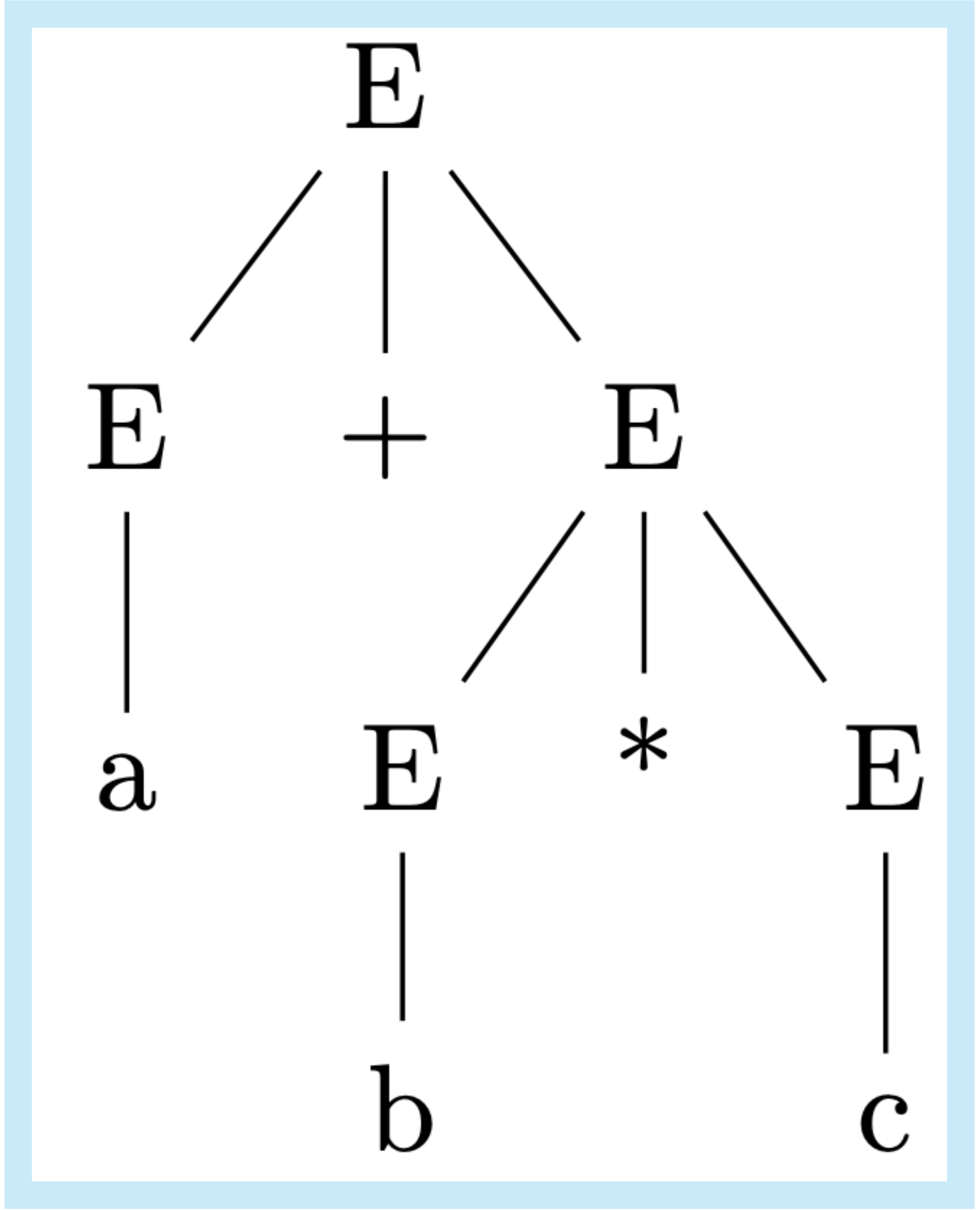
From previous lectures

Derivation trees

- A *derivation tree* is a tree such that:
  - The *root* of a derivation tree is S.
  - Each *leaf* of a derivation tree $\in \Sigma$.
  - Each *inner node* of a derivation tree $\in$ N.
  - **If** the node A has the children p, q, r, …
    - **then** there is a rule A $\rightarrow$ pqr… $\in$ P

Derivation trees

- Example grammar: $E \rightarrow E*E \mid E+E \mid a \mid b \mid c$
- Example derivation:
  - $E \Rightarrow E+E \Rightarrow E+E*E \Rightarrow E+E*c \Rightarrow E+b*c \Rightarrow a+b*c$
- Derivation (parse) tree:



Chomsky normal form

- A grammar is in Chomsky normal form
  - if all rules have the form A → a or A → BC.

- Example: **Grammar G_6**
  - S → aSb | ab | pTq | pq
  - T → pTq | pq

- **Grammar G_7**
  - S → ASB | AB | PTQ | PQ
  - T → PTQ | PQ
  - A → a
  - B → b
  - P → p
  - Q → q

- **Grammar G_8**
  - S → AE |AB | PF | PQ
  - T → PF | PQ
  - E → SB
  - F → TQ
  - A → a
  - B → b
  - P → p
  - Q → q

Let's start!

Today's topic

- In lecture 4, some *closure properties* for regular languages were presented.
- In lecture 6, the *pumping lemma* for regular languages was discussed.

- **In this lecture**: corresponding features for *context-free languages*.
  1. Some closure properties,
  2. The pumping lemma,
  3. Some more closure properties
     (that need the pumping lemma for their proofs).

Some closure properties

- The set of context-free languages are closed under
  - **Theorem 1**: Union
  - **Theorem 2**: Concatenation
  - **Theorem 3**: Star operation

- Proofs: Coming up.
  - Denote languages L_1 and L_2
  - With the grammars $G_1 = \langle N_1, \Sigma_1, P_1, S_1 \rangle$ and $G_2 = \langle N_2, \Sigma_2, P_2, S_2 \rangle$.
  - $L_1 = L(G_1)$ and $L_2 = L(G_2)$

Closure under union

- **Theorem 1**. CFLs are closed under union.
- **Proof**. Construct a new grammar:
  - $G_3 = \langle N_1 \cup N_2 \cup \{ S_3 \}, \quad \Sigma_1 \cup \Sigma_2, \quad P_1 \cup P_2 \cup \{ S_3 \rightarrow S_1 \mid S_2 \}, \quad S_3 \rangle$.
    - (If N_1 and N_2 are not disjoint sets, rename the nonterminals in N_2)
  - Starting from S_3, all derivations starting with **either** S_1 or S_2 can be done.
    - => All strings in $L(G_1)$ **and** $L(G_2)$ can be derived
    - => $L(G_3) = L(G_1) \cup L(G_2)$

Closure under concatenation

- **Theorem 2**. CFLs are closed under concatenation.
- **Proof**. Construct a new grammar:
  - $G_4 = \langle N_1 \cup N_2 \cup \{ S_4 \}, \quad \Sigma_1 \cup \Sigma_2, \quad P_1 \cup P_2 \cup \{ S_4 \rightarrow S_1 S_2 \}, \quad S_4 \rangle$.
    - (If N_1 and N_2 are not disjoint sets, rename the nonterminals in N_2)
  - Now, starting from S_4, all derivations consist of:
    - one part derived from S_1 in $L(G_1)$,
    - **followed by** one part derived from S_2 in $L(G_2)$
    - => $L(G_4) = L(G_1)L(G_2)$.

Closure under *

- **Theorem 3**. CFLs are closed under the star operation.
- **Proof**. Construct a new grammar:
  - $G_5 = \langle\, N_1 \cup \{ S_5 \},\quad \Sigma_1,\quad P_1 \cup \{ S_5 \rightarrow S_1 S_5 \mid \varepsilon \},\quad S_5 \,\rangle$.
  - Starting from S_5:
    - **either** ε can be derived,
    - **or** one string of $L(G_1)$,
      - **followed** once again by what can be derived from S_5.
    - => $L(G_5) = L(G_1)*$

Pumping lemma for CFLs – Motivation and proof sketch

- On the board.
- See lecture notes and course book for similar figures to what was shown during the lecture.

- **Reminders**:
- A grammar is in Chomsky normal form (CNF)
  - if all rules have the form A → a or A → BC.

Pumping lemma for CFLs – Direct form

- From the ideas in the figures, this form of the pumping lemma can be stated
- L is CF $\rightarrow \exists k > 0$:
  - $\forall z \in L : |z| \geq k \wedge$
    - $\exists u, v, w, x, y \in \Sigma^* : z = uvwxy \wedge vx \neq \varepsilon \wedge |vwx| \leq k \wedge$
      - $\forall i \geq 0 : uv^iwx^iy \in L$

1. If **L** is context-free then there is some constant **k** (pumping length) such that
2. if a string **z** in **L** is of enough length, then
3. it can be split into **uvwxy**—so that **vx** isn't empty and **vwx** isn't too long—so that
4. the **vx** parts can be repeated arbitrarily number of times and the string still is in L.

Pumping lemma for CFLs – Contrapositive form

- Like for regular pumping, instead of using the lemma on the form "**if A, then B**" we will use the contrapositive form "**if not B, then not A**".
- $(\forall k > 0 :$
  - $\exists z \in L : |z| \geq k \land$
    - $\forall u, v, w, x, y \in \Sigma_* : z = uvwxy \land vx \neq \varepsilon \land |vwx| \leq k \rightarrow$
      - $\exists i \geq 0 : uv^i wx^i y \notin L) \rightarrow L$ is not CF

1. For all constants **k** (pumping lengths)
2. there is a string **z** of enough length, and
3. for all ways to split **z** into **uvwxy**—such that **vx** isn't empty and **vwx** isn't too long—
4. you can pump the **vx** parts to get the string outside of the language.

Example 1

- *Prove that $L_6 = \{ a^n b^n c^n \mid n \geqslant 0 \}$ is not context-free.*
- For the pumping length **k**, let **z** = $a^k b^k c^k$.
- Now, since **vwx** cannot contain more than **k** symbols, **vx** cannot contain both a:s, b:s, and c:s.
- They have the same number of occurrences in **z**, but not all of them will change when pumping so the pumped string will be out of L_6.
- E.g. if **u** = a^p, **v** = a^q , **w** = a^r , **x** = a^s, **y** = $a^{k-p-q-r-s} b^k c^k$,
  - then pumping to $\mathbf{uv^2wx^2y}$ will give the string $a^{k+q+s} b^k c^k$.
- So L_6 is <u>not context-free</u>.

Example 2

- *Prove that the language $L_7 = \{ a^m b^n a^m b^n \mid m, n \geq 0 \}$ is not context-free.*
- For the pumping length **k**, let **z** = $a^k b^k a^k b^k$.
- We call each of a^k and b^k a *block*.
  - If **v** or **x** contains both ”a” and ”b”: pumping up **uv²wx²y** will in **v²** or **x²** result in a mixture of two symbols, and **z** will break the original $a{*}b{*}a{*}b{*}$ pattern in L_7.
  - If **v** and **x** are in the same block: pumping up **uv²wx²y** will make that block longer than the other block with the same symbol.
  - If **v** and **x** are in neighbouring blocks, then there will be different numbers of a:s and b:s in the blocks. E.g. if **u** = a^p, **v** = a^q , **w** = $a^{k-p-q}b^r$, **x** = b^s, **y** = $b^{k-r-s}a^k b^k$, then **uv²wx²y** = $a^{k+q}b^{k+s}a^k b^k$.
- So, for every possible split of **z**, it is pumpable out of L_7. <u>L_7 is not context-free.</u>

More closure properties

- The set of context-free languages are
  - **Theorem 1**: not closed under intersection
  - **Theorem 2**: are closed under intersection with regular languages
  - **Theorem 3**: are not closed under complement

- Proofs: Coming up.
  - In this section we will introduce a number of languages (L_8, L_9, …), their
  - grammars(G_8, G_9, …), and a number of automata (M_{10}, M_{11}, …)

Theorem 4: Not closed under intersection

- **Theorem 4**. CFLs are **not** closed under intersection.
- **Proof:** By counterexample.
  - Let $L_8 = \{ a^m b^m c^n \mid m, n \geq 0 \}$ and $L_9 = \{ a^m b^n c^n \mid m, n \geq 0 \}$.
  - Both languages are context-free.
    They can be defined with the following grammars:

    G_8: $S \to AC$ $\quad A \to aAb \mid \varepsilon$ $\quad C \to cC \mid \varepsilon$

    G_9: $S \to AC$ $\quad A \to aA \mid \varepsilon$ $\quad C \to bCc \mid \varepsilon$
  - $L_8 \cap L_9 = \{ a^n b^n c^n \mid n \geq 0 \}$
- Since L_8 and L_9 are context-free, and the result was proven not context-free in **Example 1**, context-free languages are not closed under intersection.

Theorem 5: Closure under intersection with regular languages

- **Theorem 5**. CFLs are closed under intersection **with regular languages**.
- **Proof**. Let CFL L_{10} be accepted by a PDA M_{10} and regular L_{11} by a DFA M_{11}.
- Let M_{12} be a PDA, where each state is a **pair** of a state in M_{10} and a state in M_{11}.
- The transitions in M_{12} follow simultaneously the transitions in M_{10} and M_{11}.
- A string is accepted in M_{12} if it is accepted in **both** M_{10} and M_{11}.
- The stack is handled as in M_{10}
  - Since M_{11} doesn't use the stack, the M_{10} stack handling won't be disturbed.

Theorem 6: Not closed under complement

- **Theorem 6**. CFLs are not closed under complement.
- **Proof**. $L_{13} \cap L_{14} = \sim(\sim L_{13} \cup \sim L_{14})$ (by De Morgan's Law)
- If CFLs were closed under complement:
  - then the union of two CFLs would have been constructed,
    - which would be context-free, according to **Theorem 1**;
  - and then the complement of that result should have been context-free.
- But the result is the intersection, that does not preserve context-freeness according to **Theorem 4.**

Example 3

- $L_{15} = \{ ww \mid w \in \{a, b\}^* \}$ *is not context-free.*
  - L_{15} contains strings consisting of two equal parts,
    - e.g. aabaab, ababaababa, and bbbbabbbba.
- **Proof**:
- Some of the strings in L_{15} are the same as some strings of the form $a^*b^*a^*b^*$.
- $L_{15} \cap a^*b^*a^*b^* = L_7$, with $L_7 = \{ a^m b^n a^m b^n \mid m, n \geq 0 \}$.
- Since CFLs are closed under intersection with regular languages, L_7 would be context-free if L_{15} were context-free.

Coming up soon …

- <u>Next week</u>:
  - **Monday**: LL(1) and LR(0) parsing
  - **Wednesday**: LR(1) parsing

- Then Victor is back for two lectures (14 and 15).
- I will return for the final lecture (16).

Thanks for today!