San José State University
Department of Computer Science

Ahmad Yazdankhah
ahmad.yazdankhah@sjsu.edu
www.cs.sjsu.edu/~yazdankhah

Grammars

(Part 2)

Lecture 21

Day 23/31

CS 154

Formal Languages and Computability

Fall 2019

Agenda of Day 23

- Summary of Lecture 20
- Quiz 8
- Lecture 21: Teaching ...
  - Grammars (Part 2)

Summary of Lecture 20: We learned ...

REGEXs

- Every REGEX represents a language.
- Can every language be represented by a REGEX?
- No, only regular languages ...
- A language is regular if a REGEX represents it.

- The limitation of REGEXs ...
  - They just represent regular languages, while more interesting languages are non-regular!
- Conclusion: we'd need a more powerful tool for representing formal languages.
- We don't have a standard REGEX.

Any Question?

Summary of Lecture 20: We learned ...

Grammars

- We were looking for a more powerful and practical tool to represent formal languages.
- We introduced grammars as the next choice.
- Roughly speaking, a set of production rules is called grammar.
- A sentence is well-formed over a grammar if ...
  - ... we can derive it from the grammar.

Any Question

Summary of Lecture 20: We learned ...

Grammar Terminologies

Starting Variables

Left-Hand Side (LHS)

Right-Hand Side (RHS)

<sentence> → <noun-phrase> <predicate>

<noun-phrase> → <article> <noun>

<article> → the

Production Rule

Variables

Terminal

"is defined as" symbol

Quiz 8

No Scantron

Formal Grammars

Formal Grammars

- We can generalize the natural languages grammars to formal grammars.

Example 4

- Consider the following set of "production rules":

$$\begin{cases} S \rightarrow aB \\ B \rightarrow baB \\ B \rightarrow \lambda \end{cases}$$

- This set of production rules is an example of a "formal grammar".
- Let's see its ingredients in detail?

Ingredients of the Production Rules

$$\begin{cases} S \rightarrow aB \\ B \rightarrow baB \\ B \rightarrow \lambda \end{cases}$$

- S and B are "variables" in this example.
  - Variables are represented by capital letters.
- By default, the "starting variable" is 'S' unless we mention something else.
- 'a' and 'b' in this example are called "terminal symbols".
  - Terminals are represented by lower-case letters.
  - λ is our familiar empty string.
  - Terminals can be any sequence of terminal symbols or λ.
- "aB" and "baB" contain both variable and terminals and are called "sentential form".

How a string can be derived from a grammar?

Example 5

- Let grammar G be:
  1. $S \rightarrow a\,S\,b$
  2. $S \rightarrow \lambda$
- Derive string "ab"

Solution

$$S \overset{1}{\Rightarrow} a\,S\,b \overset{2}{\Rightarrow} a\,\lambda\,b = ab$$

- Could we derive this string if we had started with rule #2?

Derivation of Strings

Example 6

- Let grammar G be:
  1. $S \rightarrow a\,S\,b$
  2. $S \rightarrow \lambda$
- Derive string "aabb".

Solution

$$S \overset{1}{\Rightarrow} a\,S\,b \overset{1}{\Rightarrow} aa\,S\,bb \overset{2}{\Rightarrow} aa\,\lambda\,bb = aabb$$

- We can summarize the above derivation like this:

$$S \overset{*}{\Rightarrow} aabb$$

- As we said before, this notation is used when we just want to show that S drives the string.

A Convention

- When the left-hand sides of two or more production rules are the same, we can combine the right-hand sides by separating them with a vertical bar "|".
- Here, "|" means "OR".

Example 7

- Let grammar G be:

  $S \rightarrow a\ S\ b$

  $S \rightarrow \lambda$

- We can represent it as:

  $S \rightarrow a\ S\ b \mid \lambda$

- It means: S is defined as "a S b" OR λ

Associated Language to Grammars

- We can apply the production rules "recursively" in any arbitrary orders.
- Therefore, a grammar can generate zero, one, or more strings.

Definition

- The set of all strings generated (aka produced) by the grammar G is called the "associated language to G" and is denoted by L(G).

Grammar $\rightarrow$ Language Examples

Grammar → Language Examples

Example 8

- Let grammar G be:

  $S \rightarrow a\,S \mid \lambda$

- L(G) = ?   // show it by a set-builder.

Solution

- How about this grammar?

  $S \rightarrow S\,a \mid \lambda$

- Is there any difference?

Grammar → Language Examples

Example 9

- Let grammar G be:

  $S \rightarrow a\,S\,b \mid \lambda$

- L(G) = ?   // show it by a set-builder.

Solution

Conclusion

- After this example, we know that grammars can represents more languages than just regular languages!
- So, they are more powerful tools!

Grammar → Language Examples

Example 10

- Let grammar G be:
  1. $S \rightarrow AB$
  2. $A \rightarrow aA \mid \lambda$
  3. $B \rightarrow bB \mid \lambda$
- $L(G) = ?$ // show it by a set-builder.

Solution

Language → Grammar Examples

Language → Grammar Examples

Example 11

- Find a grammar that generates the following language over $\Sigma = \{a, b\}$:

  $L = \{w : w \in \Sigma^*\}$

Solution

Example 12

- Find a grammar that generates the following language over $\Sigma = \{a, b\}$:

  $L = \{w : w \text{ contains exactly one } a\}$

Solution

Homework: Language → Grammar

- Find a grammar that generates the following languages over $\Sigma = \{a, b\}$:

1. $L = \{w : w \text{ contains at least one } a\}$
2. $L = \{w : w \text{ contains at least 2 } a\text{'s}\}$
3. $L = \{w : w \text{ contains no more than 3 } a\text{'s}\}$
4. $L = \{a^{2n} b^n : n \geq 0\}$
5. $L = \{a^{2n} b^m : n, m \geq 0\}$
6. $L = \{a^n b^m : n, m \geq 0, n \neq m\}$

Definitions

Formal Definition of Grammar

- A grammar G is defined by the quadruple:

$$G = (V, T, S, P)$$

- Where:
  - V is a nonempty finite set of variables.
  - T is a nonempty finite set of symbols (aka terminals) called terminal alphabet.
  - $S \in V$ is a special symbol called start variable.
  - P is a finite set of production rules (or simply rules) of the form

    $xAy \rightarrow z$

    where:

    $A \in V$ and $x, y, z \in (T \cup V)^*$

- Note that in this course, we'd always have only one variable in LHS.

Formal Definition of Grammar

$$G = (\overset{\text{Set of Variables}}{V}\ ,\ \underset{\text{Set of Terminals}}{T}\ ,\ \overset{\text{Start Variable}}{S}\ ,\ \underset{\text{Set of Production Rules}}{P})$$

Formal Definition of Grammar: Example

Example 13

- As we saw before, the following grammar

  $S \rightarrow aSb \mid \lambda$

  generates the language $L = \{a^n b^n : n \geq 0\}$.

- Write V, T, Starting variable, and P.

Solution

$V = \{S\}$

$T = \{a, b\}$

Start variable: $S \in V$

$P = \{S \rightarrow aSb , S \rightarrow \lambda\}$

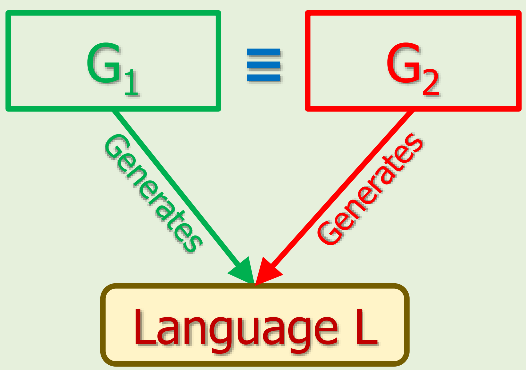
Equivalency of Grammars

- A given language can be generated by many grammars.

Definition

- Two grammars G_1 and G_2 are equivalent if both has the same associated language.

$$L(G_1) = L(G_2) \rightarrow G_1 \equiv G_2$$



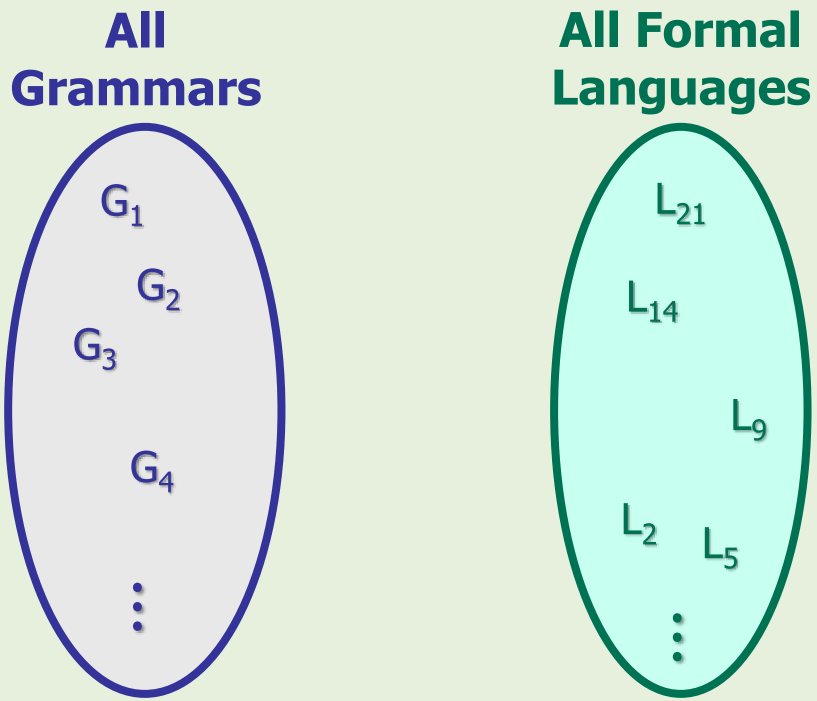
Grammars and Languages Association

Grammars and Languages Association

- What is the relationship between:

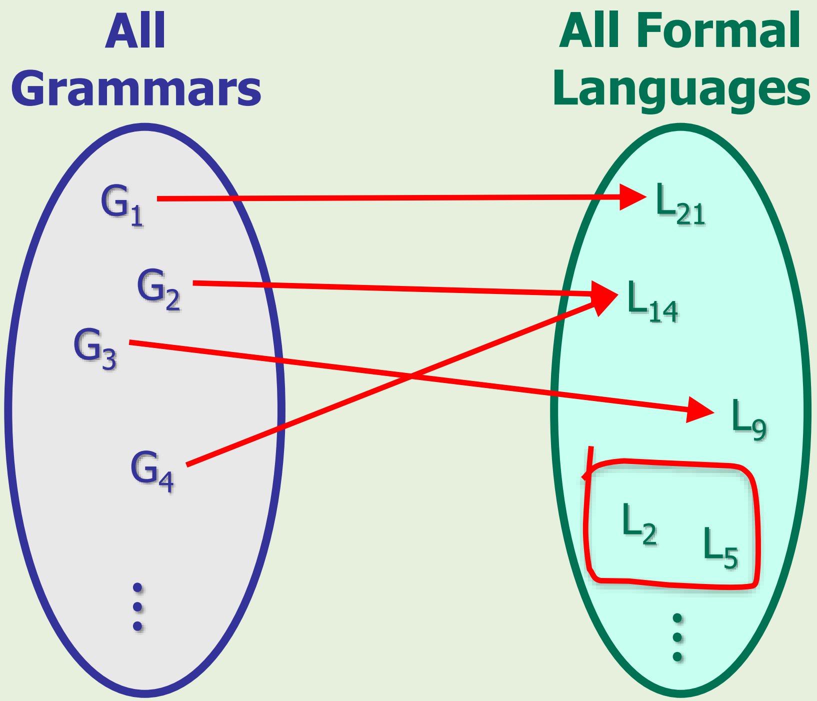
  the set of Grammars, and

  the set of all formal languages?



Grammars and Languages Association

- You agree that "every grammar represents a language".
- BUT we don't know yet whether we can represent every language, by a grammar or not!
  - Our knowledge is not enough yet.



Types of Grammars

Linear Grammars

Definition

- A grammar is linear if the right hand side of every production rule has at most one variable.
  - Again, in this course, we'd have one variable in the left hand side.

$A \rightarrow w \mid w\,B\,u$

Where $A, B \in V$ and $w, u \in T^*$

Example 14

- Is the following grammar linear?

  $S \rightarrow A$

  $A \rightarrow baBb \mid \lambda$

  $B \rightarrow Abb$

- Yes, because all production rules have at most one variable in the RHS.

Right-Linear Grammars

Definition

- A linear grammar is said to be right-linear if all production rules are of the form:

$A \rightarrow w \mid u\,B$

Where $A, B \in V$ and $w, u \in T^*$

- In the case of $A \rightarrow w$, we consider $A \rightarrow wB^0$.

Example 15

- Is the following grammar right-linear?

$S \rightarrow abS \mid a$

- Yes, it is.

Left-Linear Grammars

Definition

- A linear grammar is said to be left-linear if all production rules are of the form:

$A \rightarrow w \mid B\,u$

Where $A, B \in V$ and $w, u \in T^*$

- In the case of $A \rightarrow w$, we consider $A \rightarrow B^0 w$.

Example 16

- Is the following grammar left-linear?

$S \rightarrow Aab$

$A \rightarrow Bab \mid B$

$B \rightarrow a$

- Yes, it is.

Regular Grammars

Definition

- A grammar is said to be regular if it is either right-linear or left-linear.
  - In other words, all right-linear and left-linear grammars are regular.

Example 17

- Is the following grammar regular?

$S \rightarrow A$

$A \rightarrow aB \mid \lambda$

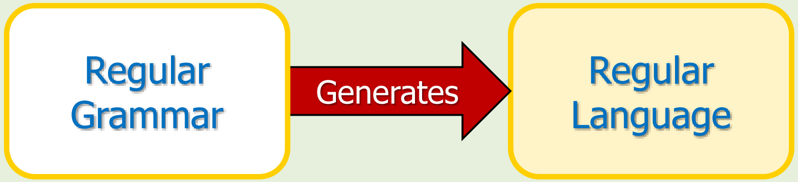
$B \rightarrow Ab$

- It is NOT regular because it is neither right-linear nor left-linear.

Regular Grammars and Regular Languages

Theorem

- Let G be a regular grammar, then L(G) is a regular language over T.

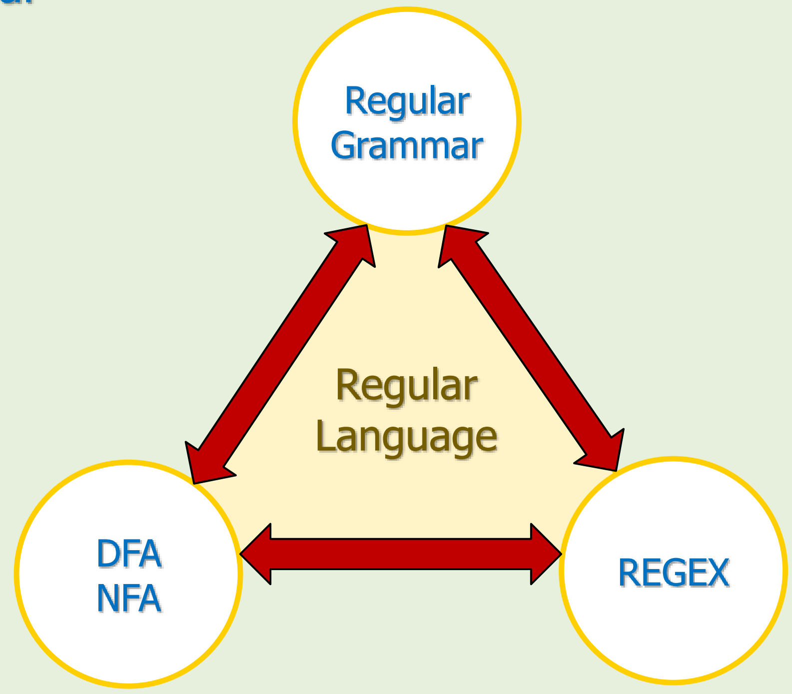


Theorem

- Let L be a regular language over Σ.
  Then there exists a regular grammar G that generates L.

Regular Languages Representations

- Now, we have three ways for representing Regular Languages:
  - DFA / NFA
  - REGEX
  - Regular Grammar



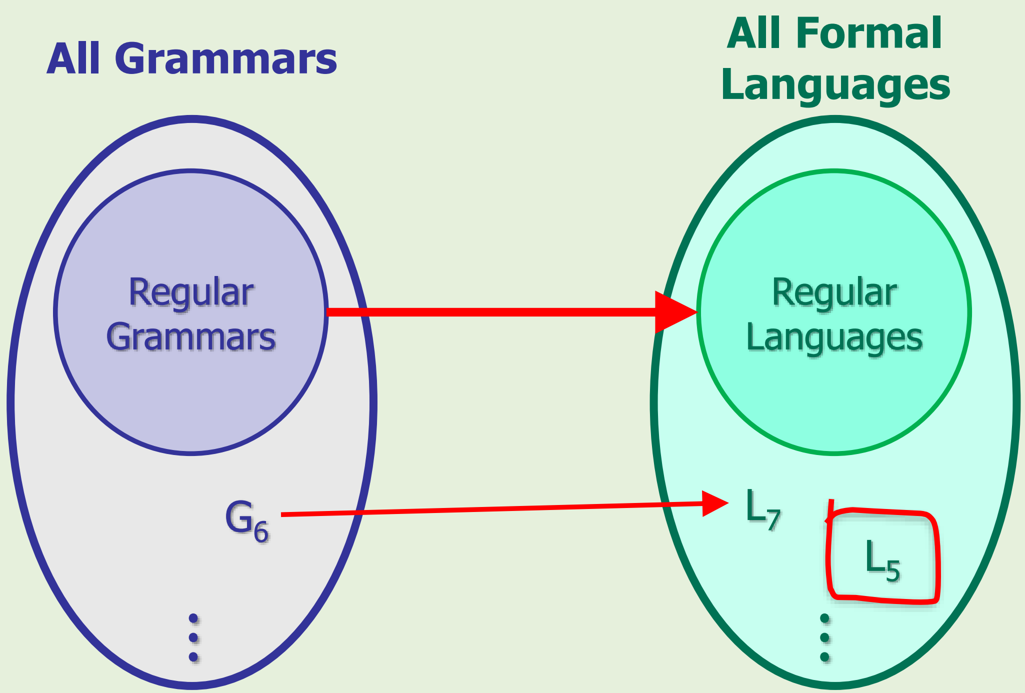
Grammars and Languages Association

Revisited

- We've already known that "every grammar represents a language".
- At this moment, we know that:

Regular grammars represent regular languages.

Every regular language can be represented by a regular grammar.



References

1. Linz, Peter, "An Introduction to Formal Languages and Automata, 5th ed.," Jones & Bartlett Learning, LLC, Canada, 2012
2. Michael Sipser, "Introduction to the Theory of Computation, 3rd ed.," CENGAGE Learning, United States, 2013
   ISBN-13: 978-1133187790
3. The ELLCC Embedded Compiler Collection, available at: http://ellcc.org/