

# CS441/CS241 Automata Theory and Formal Languages



ربيع 2024  
د. عدنان محمود عبدالله الشريف  
[adnan.sherif@uot.edu.ly](mailto:adnan.sherif@uot.edu.ly)



1

## القواعد Grammar

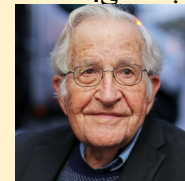
- تستخدم القواعد النحوية في اللغات للتعبير عن تركيبية الجمل على سبيل المثال في اللغة العربية الجملة الاسمية تتكون من مبتدأ وخبر.
- في علم الحاسب الآلي نستخدم قواعد مشابهة للتعبير عن مختلف أنواع اللغات.
- تختلف اللغات في نظام الحاسب الآلي من ناحية التركيب والتعقيد.
- درسنا في هذا المقرر نوع من اللغات هو اللغات المنتظمة وهي أبسط أنواع اللغات التي يمكن التعبير عنها ويمكن استخدام أوتومات منتهي للتعبير عنها أو استخدام التعبيرات المنتظمة.

2

## قواعد Grammar

نوع القاعد	القواعد المقبولة	اللغة المقبولة	الآلية المستخدمة
النوع 0 Type 0	القواعد غير المقيدة Unrestricted Grammar	معدودة بشكل متكرر Recursively enumerated languages	آلة تورينغ Turing Machine
النوع 1 Type 1	القواعد الحساسة للسياق Context Sensitive Grammar	اللغات الحساسة للسياق Context Sensitive Languages	الأتمتة ذات الحدود الخطية Linear Bounded Automata
النوع 2	القواعد خارج السياق Context-free Grammar	لغات خارج السياق Context-free Languages	الأتمتة بمكدس Pushdown Automata
النوع 3	القواعد المنتظمة Regular Grammar	اللغات المنتظمة Regular Language	الأوتومات المنتهية Finite Automata

• قام العالم نعوم تشومسكي (Noam Chomsky) بوضع تصنيف لأنواع القواعد المستخدمة في علم الحاسب الآلي وأنواع اللغات التي يمكن وصفها باستخدام القواعد المختلفة حسب التالي:



22/06/2024

CS441/CS241 Automata Theory and Formal Languages

3

3

## القواعد المنتظمة Regular Grammar

• يمكن وصف أي من أنواع القواعد باستخدام الرباعية  
 $G = (V, \Sigma, P, S)$

حيث

- $V$  مجموعة من المتغيرات وتعرف أيضا برموز لانتهائية.
- $\Sigma$  مجموعة من الرموز النهائية (الابجدية)
- $P$  علاقة معرفة على  $V \times (V \cup \Sigma)^*$  ويتم وصفها كمجموعة منتهية من القواعد على الشكل  $V \rightarrow (V \cup \Sigma)^*$
- $S \in V$  متغير البداية وهي نقطة بداية اشتقاق السلاسل بحيث

22/06/2024

CS441/CS241 Automata Theory and Formal Languages

4

4

## القواعد المنتظمة

## Regular Grammar

- القواعد المنتظمة تستخدم لتعريف اللغات المنتظمة التي تم دراستها في الدرس السابق.
- القواعد المنتظمة قسمين:

- قواعد منتظمة على اليمين وتكون جميع قواعدها على الشكل:

$$A \rightarrow aB$$

$$A \rightarrow a$$

- قواعد منتظمة على اليسار وتكون جميع قواعدها على الشكل:

$$A \rightarrow Ba$$

$$A \rightarrow a$$

حيث  $A \in V$  و  $B \in V$  و  $a \in \Sigma^*$

## القواعد المنتظمة

## Regular Grammar

- مثال 1 : قم بوصف قاعدة منتظمة  $G$  للغة كل السلاسل بها تنتهي بـ 01 على الابجدية

$$\Sigma = \{0,1\}$$

الحل: يمكن وصف القاعدة المنتظمة على اليسار باستخدام القواعد التالية:

$$A \rightarrow B01$$

$$B \rightarrow B0$$

$$B \rightarrow B1$$

$$B \rightarrow \varepsilon$$

يمكن الان تعريف

$$V = \{A, B\}, \Sigma = \{0,1\}, S = A$$

## القواعد المنتظمة

## Regular Grammar

- الاشتقاق (Derivation): هي عملية اشتقاق السلاسل من القاعدة ويرمز لها بالرمز  $\Rightarrow$ .
- يمكن تعريف عملية الاشتقاق كالتالي:
- دع السلسلة  $\alpha A \beta$  مكونة من متغير  $A \in V$  و  $\alpha \in (V \cup \Sigma)^*$  و  $\beta \in (V \cup \Sigma)^*$
- ودع أحد القواعد  $A \rightarrow \gamma$  حيث  $\gamma \in (V \cup \Sigma)^*$
- يمكن اشتقاق سلسلة جديدة باستخدام القاعدة ونرمز لها  $\alpha A \beta \Rightarrow \alpha \gamma \beta$

22/06/2024

CS441/CS241 Automata Theory and Formal Languages

7

7

## القواعد المنتظمة

## Regular Grammar

- مثال 2: باستخدام القواعد المنتظمة من المثال رقم 1 بين إمكانية اشتقاق السلاسل التالية:
  - 1.  $A \rightarrow B01$
  - 2.  $B \rightarrow B0$
  - 3.  $B \rightarrow B1$
  - 4.  $B \rightarrow \varepsilon$
- الحل: أولاً السلسلة 0010101 و 1100
- 0010101 السلسلة
- $A \xRightarrow{1} B01 \xRightarrow{3} B101 \xRightarrow{2} B0101 \xRightarrow{3} B10101 \xRightarrow{2} B010101 \xRightarrow{2} B0010101 \xRightarrow{4} 0010101$
- نلاحظ إمكانية اشتقاق السلسلة باستخدام القواعد إذن هذه السلسلة تنتمي إلى اللغة التي تصفها هذه القواعد

ثانياً: السلسلة 1100

 $A \xRightarrow{?}$ 

نلاحظ عدم وجود قاعدة يمكن استخدامها انطلاقاً من البداية  $A$  عليه هذه السلسلة لا تنتمي إلى اللغة ولا يمكن اشتقاقها باستخدام هذه القواعد.

22/06/2024

CS441/CS241 Automata Theory and Formal Languages

8

8

## القواعد المنتظمة

## Regular Grammar

• مثال 3 : قم بوصف قاعدة منتظمة  $G$  للغة كل السلاسل بها تحتوي على 01 على الأبجدية  $\Sigma = \{0,1\}$

الحل: يمكن وصف القاعدة المنتظمة على اليسار باستخدام القواعد التالية:

$$\begin{aligned} A &\rightarrow 0A \\ A &\rightarrow 1A \\ A &\rightarrow B \\ B &\rightarrow 01C \\ C &\rightarrow 0C \\ C &\rightarrow 1C \\ C &\rightarrow \varepsilon \end{aligned}$$

يمكن الآن تعريف

$$V = \{A, B, C\}, \Sigma = \{0,1\}, S = A$$

## القواعد المنتظمة

## Regular Grammar

1.  $A \rightarrow 0A$
2.  $A \rightarrow 1A$
3.  $A \rightarrow B$
4.  $B \rightarrow 01C$
5.  $C \rightarrow 0C$
6.  $C \rightarrow 1C$
7.  $C \rightarrow \varepsilon$

• مثال 4: باستخدام القواعد المنتظمة من المثال رقم 3 بين إمكانية اشتقاق السلاسل التالية:

1100 و 0010101

الحل: أولاً السلسلة 0010101

$$A \xrightarrow{1} 0A \xrightarrow{3} 0B \xrightarrow{4} 001C \xrightarrow{5} 0010C \xrightarrow{6} 00101C \xrightarrow{5} 001010C \xrightarrow{6} 0010101C \xrightarrow{7} 0010101$$

نلاحظ إمكانية اشتقاق السلسلة باستخدام القواعد إذن هذه السلسلة تنتمي إلى اللغة التي تصفها هذه القواعد

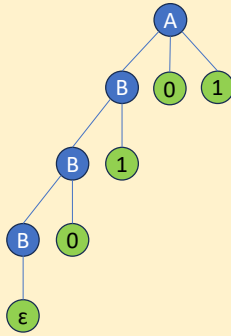
ثانياً: السلسلة 1100

$$A \xrightarrow{2} 1A \xrightarrow{2} 11A \xrightarrow{1} 110A \xrightarrow{1} 1100A$$

نلاحظ الناتج عن محاولة الاشتقاق سلسلة تحتوي على متغير رغم انتهاء السلسلة إذن هذه السلسلة لا تنتمي إلى هذه اللغة ولا يمكن اشتقاقها باستخدام هذه القواعد.

## القواعد المنتظمة Regular Grammar

• شجرة الاشتقاق (Derivation Tree): يمكن تمثيل اشتقاق سلسلة باستخدام قواعد لغة في شكل شجرة بالخصائص التالية:



• جذر الشجرة (Root) هو متغير البداية S.

• الرؤوس الداخلية في الشجرة تمثل المتغيرات في قواعد اللغة.

• أوراق الشجرة تمثل الرموز النهائية.

$$1. A \rightarrow B01$$

$$2. B \rightarrow B0$$

$$3. B \rightarrow B1$$

$$4. B \rightarrow \varepsilon$$

مثال 5: باستخدام قواعد اللغة في المثال رقم 2

الشجرة التالية تمثل اشتقاق السلسلة 0101

22/06/2024

CS441/CS241 Automata Theory and Formal Languages

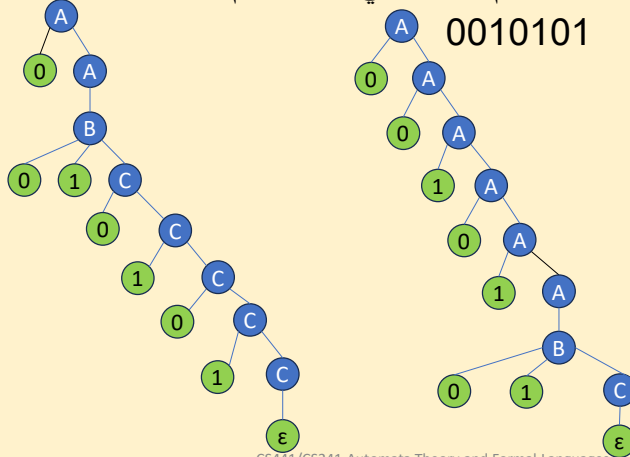
11

11

## القواعد المنتظمة Regular Grammar

1.  $A \rightarrow 0A$
2.  $A \rightarrow 1A$
3.  $A \rightarrow B$
4.  $B \rightarrow 01C$
5.  $C \rightarrow 0C$
6.  $C \rightarrow 1C$
7.  $C \rightarrow \varepsilon$

• مثال 6: ارسم شجرة الاشتقاق باستخدام القواعد في المثال رقم 3 للسلسلة التالية:



الحل:

نلاحظ ان بعض القواعد (كما في المثال) يمكن اشتقاق أكثر من شجرة لنفس السلسلة وفي هذه الحالة نقول ان القواعد مبهمه

22/06/2024

CS441/CS241 Automata Theory and Formal Languages

12

12

## قواعد واللغات خارج السياق

### Context Free Grammars and Languages

- بعد دراسة اللغات المنتظمة (Regular Languages) نوجه اهتمامنا الى دراسة مجموعة أكبر واشمل من اللغات تعرف باللغات خارج السياق (Context Free Languages - CFL).
- لهذه المجموعة من اللغات وصف بالتكرار يعرف بقواعد خارج السياق (Context Free Grammar - CFG).
- تستخدم قواعد خارج السياق لتعريف لغات البرمجة وتستخدم لبرمجة المحلل اللغوي للمترجمات (Parser).
- اللغات المنتظمة التي درسناها في السابق فئة جزئية من اللغات خارج السياق وعليه يمكن تعريف اللغات المنتظمة باستخدام قواعد خارج السياق.

## قواعد واللغات خارج السياق

### Context Free Grammars and Languages

- نستخدم نفس التعريف المستخدم للقواعد المنتظمة لتعريف قواعد خارج السياق:
- يمكن وصفها باستخدام الرباعية

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

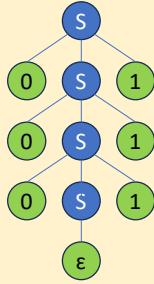
حيث

- $V$  مجموعة من المتغيرات (Variables or Nonterminal) وتعرف أيضا برموز لانهائية.
- $\Sigma$  مجموعة من الرموز النهائية (الابجدية) (Terminal).
- $P$  علاقة معرفة على  $V \times (V \cup \Sigma)^*$  ويتم وصفها كمجموعة منتهية من القواعد على الشكل  $V \rightarrow (V \cup \Sigma)^*$
- $S$  متغير البداية وهي نقطة بداية اشتقاق السلاسل بحيث  $S \in V$

## قواعد واللغات خارج السياق

### Context Free Grammars and Languages

- مثال: خلال دراستنا لتوطئة الضخ في الباب السابق تعرفنا على بعض اللغات غير المنتظمة منها  $L = \{0^n 1^n | n \geq 0\}$ . عرف اللغة باستخدام قواعد خارج السياق ثم بين اشتقاق السلسلة 000111.



- الحل: نستخدم القواعد التالية لتعريف اللغة

$$1. S \rightarrow 0S1$$

$$2. S \rightarrow \varepsilon$$

حيث  $V = \{S\}$  و  $\Sigma = \{0,1\}$  و  $S = S$

يمكن اشتقاق السلسلة 000111 باستخدام الخطوات التالية:

$$S \xrightarrow{1} 0S1 \xrightarrow{1} 00S11 \xrightarrow{1} 000S111 \xrightarrow{2} 000111$$

22/06/2024

CS441/CS241 Automata Theory and Formal Languages

اعداد د. عدنان محمود الشريف ، قسم الحاسب الآلي - كلية العلوم - جامعة طرابلس

15

15

## قواعد واللغات خارج السياق

### Context Free Grammars and Languages

- مثال 7: في جميع لغات البرمجة يمكن كتابة معادلات رياضية في شكل عمليات مثل  $(a + b) * a + b$

حيث  $a$  و  $b$  متغيرات البرنامج و  $+$  و  $*$  و  $()$  عمليات.  
اكتب قواعد خارج السياق للتعبير عن العملية الحسابية.

الحل:

بالنظر الى المثال نلاحظ ان  $\Sigma = \{a, b, +, *, (, )\}$

كما نعرف ان عمليات  $+$  و  $*$  ثنائية أي انها تعمل على معاملين اثنين كما ان الاقواس يجب ان تتقابل أي ان كل قوس مفتوح يجب ان يقابله قوس مغلق، من المثال

$$(a + b) * a + b$$

عليه سلاسل على الشكل  $a +$  و  $b + *$  و  $(a + (b * a))$  مرفوضة

22/06/2024

CS441/CS241 Automata Theory and Formal Languages

16

16



## قواعد واللغات خارج السياق Context Free Grammars and Languages

• تابع حل مثال 7:

على ما سبق يمكن تكوين القواعد التالية :

$$\begin{aligned} E &\rightarrow (E) \\ E &\rightarrow E + E \\ E &\rightarrow E * E \\ E &\rightarrow I \\ I &\rightarrow a \\ I &\rightarrow b \end{aligned}$$

حيث  $S = E$  و  $V = \{E, I\}$  و  $\Sigma = \{a, b, +, *, (, )\}$

## قواعد واللغات خارج السياق Context Free Grammars and Languages

• تابع حل مثال 7:

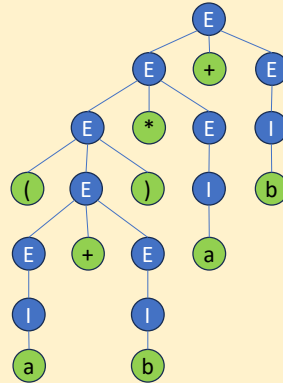
يمكن اشتقاق المثال  $(a + b) * a + b$  كالتالي:

$$\begin{aligned} E &\xRightarrow{2} E + E \xRightarrow{3} E * E + E \xRightarrow{1} (E) * E + E \xRightarrow{2} (E + E) * E + E \xRightarrow{4} \\ &\quad (I + E) * E + E \xRightarrow{5} (a + E) * E + E \xRightarrow{4} (a + I) * E + E \xRightarrow{6} \\ &\quad (a + b) * E + E \xRightarrow{4} (a + b) * I + E \xRightarrow{5} (a + b) * a + E \xRightarrow{4} \\ &\quad (a + b) * a + I \xRightarrow{6} (a + b) * a + b \end{aligned}$$

1.  $E \rightarrow (E)$
2.  $E \rightarrow E + E$
3.  $E \rightarrow E * E$
4.  $E \rightarrow I$
5.  $I \rightarrow a$
6.  $I \rightarrow b$

## قواعد واللغات خارج السياق Context Free Grammars and Languages

- تابع حل مثال 7: كما يمكن رسم شجرة الاشتقاق لنفس المثال  $(a + b) * a + b$  كالتالي:



22/06/2024

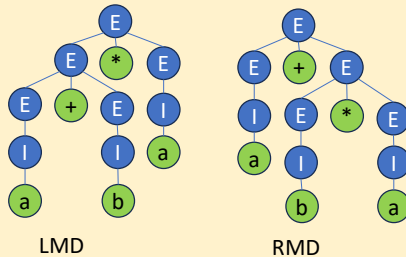
CS441/CS241 Automata Theory and Formal Languages

19

19

## قواعد واللغات خارج السياق Context Free Grammars and Languages

- في المثال السابق نلاحظ انه في حالة بعض السلاسل في اللغة يمكن اشتقاق أكثر من شجرة لنفس السلسلة (مثال السلسلة  $a + b * a$  يمكن ان نجد أكثر من اشتقاق لهذه السلسلة) ويمكن تقسيم أنواع الاشتقاق الى قسمين:



LMD

RMD

- اشتقاق من أقصى اليسار (Left Most Derivation):  
يتم فيها تكوين شجرة الاشتقاق بالتعويض عن المتغيرات في أقصى اليسار أولاً

- اشتقاق من أقصى اليمين (Right Most Derivation):  
يتم في هذه الحالة الاشتقاق بالتعويض عن المتغيرات أقصى اليمين أولاً

22/06/2024

CS441/CS241 Automata Theory and Formal Languages

20

20

## قواعد واللغات خارج السياق Context Free Grammars and Languages

- تحويل القواعد المبهمة الى قواعد غير مبهمة:  
في المثال السابق درسنا قواعد للتعبير عن معادلة رياضية في لغة برمجة ومن خلال دراستنا السابقة رأينا ان هذه القاعدة لها أكثر من شجرة اشتقاق لبعض السلاسل وعليه تعتبر مبهمة.  
لا توجد قاعدة او خوارزمية لتحويل القواعد المبهمة الى غير مبهمة ويجب دراسة كل حالة ومحاولة إعادة كتابة القواعد بحيث لا تكون مبهمة حسب أولوية او ترتيب معين.  
في المثال السابق نرى ان لعملية الضرب أولوية على الجمع وبالتالي شجرة الاشتقاق من أقصى اليمين (RMD) هي الصحيحة.

22/06/2024

CS441/CS241 Automata Theory and Formal Languages

21

21

## قواعد واللغات خارج السياق Context Free Grammars and Languages

- مثال 8: غير القواعد في حل المثال رقم 7 بحيث تصبح غير مبهمة.  
الحل: عند دراسة القواعد الناتجة عن حل المثال رقم 7 نلاحظ ان المشكلة  
تكمُن في  $E \rightarrow E + E$  و  $E \rightarrow E * E$  حيث ان لا تفرق القواعد بين  
عملية الجمع والضرب ويمكن تغيير هذه القواعد بإضافة متغيرين جديدين  
  1.  $E \rightarrow (E)$
  2.  $E \rightarrow E + E$
  3.  $E \rightarrow E * E$
  4.  $E \rightarrow I$
  5.  $I \rightarrow a$
  6.  $I \rightarrow b$
- متغير يعبر عن عملية الضرب  $T$
- متغير لا يمكن ان يحتوي على عملية ضرب او جمع ويحتوي فقط على قاعدة الاقواس والمتغير.

22/06/2024

CS441/CS241 Automata Theory and Formal Languages

22

22

## قواعد واللغات خارج السياق Context Free Grammars and Languages

تابع مثال 8: يمكن إعادة كتابة القواعد بالشكل التالي:

$$E \rightarrow E + T \mid T$$

$$T \rightarrow T * F \mid F$$

$$F \rightarrow I \mid (E)$$

$$I \rightarrow a \mid b$$

نلاحظ استخدام | للتعبير عن أكثر من قاعدة لنفس المتغير

$$E \rightarrow E + T \mid T \equiv \begin{matrix} E \rightarrow E + T \\ E \rightarrow T \end{matrix}$$

22/06/2024

CS441/CS241 Automata Theory and Formal Languages

23

23

## قواعد واللغات خارج السياق Context Free Grammars and Languages

1.  $E \rightarrow E + T \mid T$
2.  $T \rightarrow T * F \mid F$
3.  $F \rightarrow I \mid (E)$
4.  $I \rightarrow a \mid b$

• تابع مثال 8: نلاحظ عند اشتقاق السلسلة السابقة  $a + b * a$  لا يوجد أكثر من اشتقاق واحد مهما كان اتجاه الاشتقاق (من اليمين أو اليسار): الاشتقاق من اليسار:

$$\begin{aligned} E &\xRightarrow{1.1} E + T \xRightarrow{1.2} T + T \xRightarrow{2.2} F + T \xRightarrow{3.1} I + T \xRightarrow{4.1} \\ a + T &\xRightarrow{2.1} a + T * F \xRightarrow{2.2} a + F * F \xRightarrow{3.1} a + I * F \xRightarrow{4.2} \\ a + b * F &\xRightarrow{3.1} a + b * I \xRightarrow{4.1} a + b * a \end{aligned}$$

الاشتقاق من اليمين:

$$\begin{aligned} E &\xRightarrow{1.1} E + T \xRightarrow{2.1} E + T * F \xRightarrow{3.1} E + T * I \xRightarrow{4.1} E + T * a \xRightarrow{2.2} \\ E + F * a &\xRightarrow{3.1} E + I * a \xRightarrow{4.2} E + b * a \xRightarrow{1.2} T + b * a \xRightarrow{2.2} \\ F + b * a &\xRightarrow{3.1} I + b * a \xRightarrow{4.1} a + b * a \end{aligned}$$

22/06/2024

CS441/CS241 Automata Theory and Formal Languages

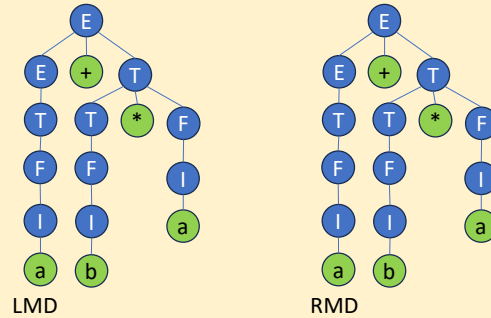
24

24

## قواعد واللغات خارج السياق

### Context Free Grammars and Languages

- تابع مثال 8: نلاحظ ان الاشتقاق من اليسار او اليمين يتبع نفس الخطوات، ولكن بترتيب مختلف. وعند تكوين شجرة الاشتقاق باستخدام الاشتقاق من اليسار او اليمين الناتج نفس الشجرة:



22/06/2024

CS441/CS241 Automata Theory and Formal Languages

25

25

## قواعد واللغات خارج السياق

### Context Free Grammars and Languages

- لغات خارج السياق: هي اللغات التي يمكن كتابة قواعد خارج السياق لتمثيلها ويمكن تعريفها كالتالي:

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

حيث  $L(G)$  هي اللغة الناتجة عن قواعد الاشتقاق خارج السياق  $G$  ويعبر عنها بمجموعة من السلاسل حيث كل سلسلة فيها  $w$  تنتمي الى  $\Sigma^*$  أي ان السلاسل تحتوي على رموز فقط ولا تحتوي على متغيرات وبشرط إمكانية اشتقاق  $w$  انطلاقاً من متغير البداية  $S$  بعدد من قواعد الاشتقاق  $\Rightarrow^*$ .

22/06/2024

CS441/CS241 Automata Theory and Formal Languages

26

26