cs401, 10 / 13 / 11

1. Project questions?
2. Regular vs. context-free languages
3. Grammar ambiguity
4. (fun) Grammars for NLP

Regular or Context-Free? #1

- Reminder: 0^n means n zeros
- L = $\{0^n 1^n \mid n >= 0\}$

$$E \rightarrow 0\ E\ 1$$
$$E \rightarrow \varepsilon$$

Regular or Context-Free? #2

- $L = \{0^n 1^m \mid n >= 0, m > n\}$

$E \rightarrow 0\ E_1\ 1$

$E \rightarrow E_1$

$E_1 \rightarrow E'\ 1$
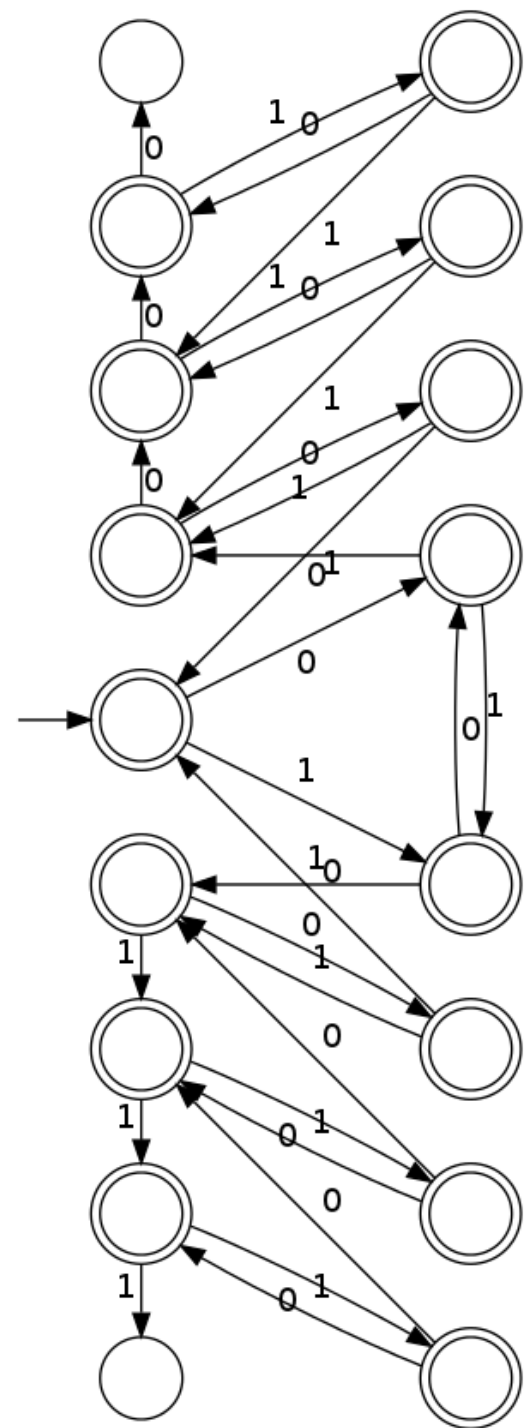
$E_1 \rightarrow E_1\ 1$

$E' \rightarrow E$

$E' \rightarrow \varepsilon$

Regular or Context-Free? #3

- #<x> counts occurrences of substring x
- L = **{**w s.t. |#00(w) - #11(w)| <= 3**}**

Regular or Context-Free? #3

- #<x> counts occurrences of substring x
- L = **{**w s.t. |#00(w) - #11(w)| <= 3**}**

Regular or Context-Free? #5

- #<x> counts occurrences of substring x
- L = **{**w | #0(w) == #1(w)**}**

$$E \rightarrow \varepsilon \mid 0\ E_1 \mid 1\ E_0$$
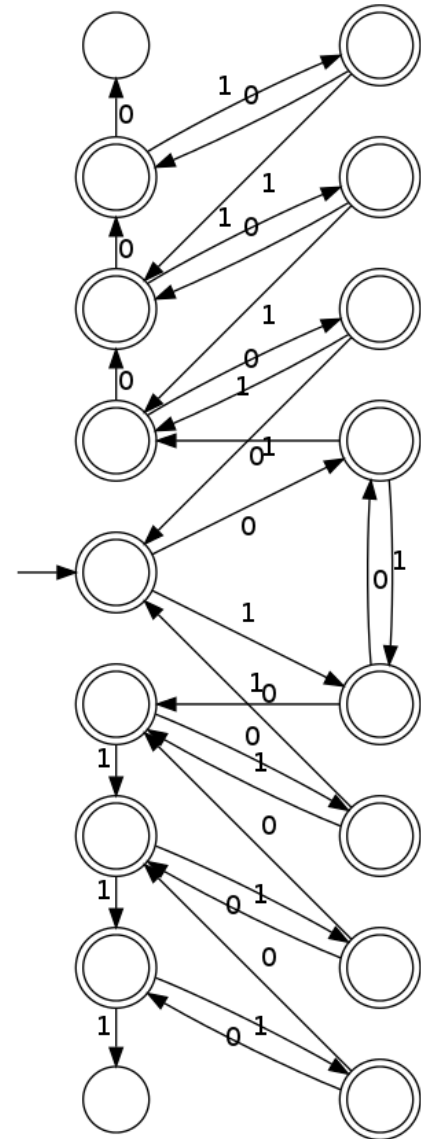$$E_1 \rightarrow 1\ E \mid E\ 1$$
$$E_0 \rightarrow 0\ E \mid E\ 0$$

Regular or Context-Free? #6

- Notation: #<x> counts occurrences of substring x
- L = {w | #01(w) == #10(w)}

$$(0+(1+0+)*|1+(0+1+)*)?$$

Some Things That Aren't Regular

- Counting to infinity
- Duplication of arbitrary-length strings

Ambiguous Grammars #1

- Example from class: calculator (eg 5+6*7)
  - Solve by changing the grammar

$$expr ::= expr + expr \mid expr - expr$$
$$\mid expr * expr \mid expr / expr \mid int$$
$$int ::= 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9$$

$$expr ::= expr + term \mid expr - term \mid term$$
$$term ::= term * factor \mid term / factor \mid factor$$
$$factor ::= int \mid (expr)$$
$$int ::= 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7$$

Ambiguous Grammars #3

- >> in C++

  list<list<int>> l;   *generics*

  int i = 318491 >> 5;   *right shift*

- Grammar for this?

Ambiguous Grammars #3

- >> in C++

  list<list<int>> l;   *generics*

  int i = 318491 >> 5;   *right shift*

  template <const int N> class factorialT {...}
  list<factorialT<5>> l;   *compile-time computation*

  template <const int N> class factorialT {...}
  list<factorialT<318491 >> 5>> l;   *???*

- How to avoid an ambiguous grammar?

- Do we *want* a nonambiguous grammar?

Natural Languages are Ambiguous

- First try at a cfg for English:

  S → NP VP
  NP → DET NP | NN | NP PP
  VP → V NP | VP PP
  PP → P NP
  N → boy | girl | house | UFO | radar | telescope | alien | government conspiracy
  V → sees | meets | leaves | knows | greets
  DET → a | an | the
  P → to | for | with | in

Natural Languages are Ambiguous

- First try at a cfg for English:

    S → NP VP
    NP → DET NP | NN | NP PP
    VP → V NP | VP PP
    PP → P NP
    N → boy | girl | house | UFO | radar | telescope | alien | government conspiracy
    V → sees | meets | leaves | knows | greets
    DET → a | an | the
    P → to | for | with | in

- How to parse?

    - the girl sees the boy in the telescope

Natural Languages are Not Context-Free

- Long-distance dependences
  - Who does the roll call?
  - For whom does the roll call?

- Context-sensitive verbs
  - I'd do what he would
  - I'd do what he would do
  - I'd run faster than he would
  - * I'd run faster than he would do