A Context-Free Grammar for a Repeated String

Zach Tomaszewski

ICS641

27 Apr 2012

Chomsky Hierarchy

- Classification of language classes
- Type / Grammar / Automata
  - Type 0 - Unrestricted - Turing Machine
  - Type 1 - Context-Sensitive - Linear Automata
  - Type 2 - Context-Free - ND Pushdown Automata
  - Type 3 - Regular - Finite Automata

- Useful to know what type of language you're dealing with

Recognizing a Language Class

- Not easy to do based on language definition alone
- Some examples...

- (all over the alphabet {**a**, **b**})

L_a = { *w* | *w* contains at least one **a**}

- Regular language
- Can recognize with a finite automaton:
  - start in non-accepting state
  - read left-to-right until an **a**
  - accepting state

  - Left-regular grammar would work similarly

$L_p = \{ww^R\}$

- Even-length palidromes
- Classic example of context-free (not regular)
  - Need a pushdown automata to recognize
  - has a stack, reads left-to-right
  - Note the need for nondeterminism

- Grammars often generate from center outwards

$L_d = \{ww\}$

- Repeated string
- Not context-free
  - pushdown automata doesn't work anymore (stack goes the wrong way)

- Seemingly simple language at first glance, but not.

Complement of a Language

- L = complement of L_d
  = {set of strings not of form *ww* for some *w*}
- Still over the alphabet {**a**, **b**}

- Complement of a language need not be in the same class
- Often higher, but here lower class

L= {set of strings not of form *ww*}

- Goal of this presentation:
  - Prove *L* is context-free by providing a grammar G_L
  - Prove that $L(G_L) = L$

Exploration: Odd Length Strings

- *ww* must be even-length...
- so any odd-length string is not-*ww*

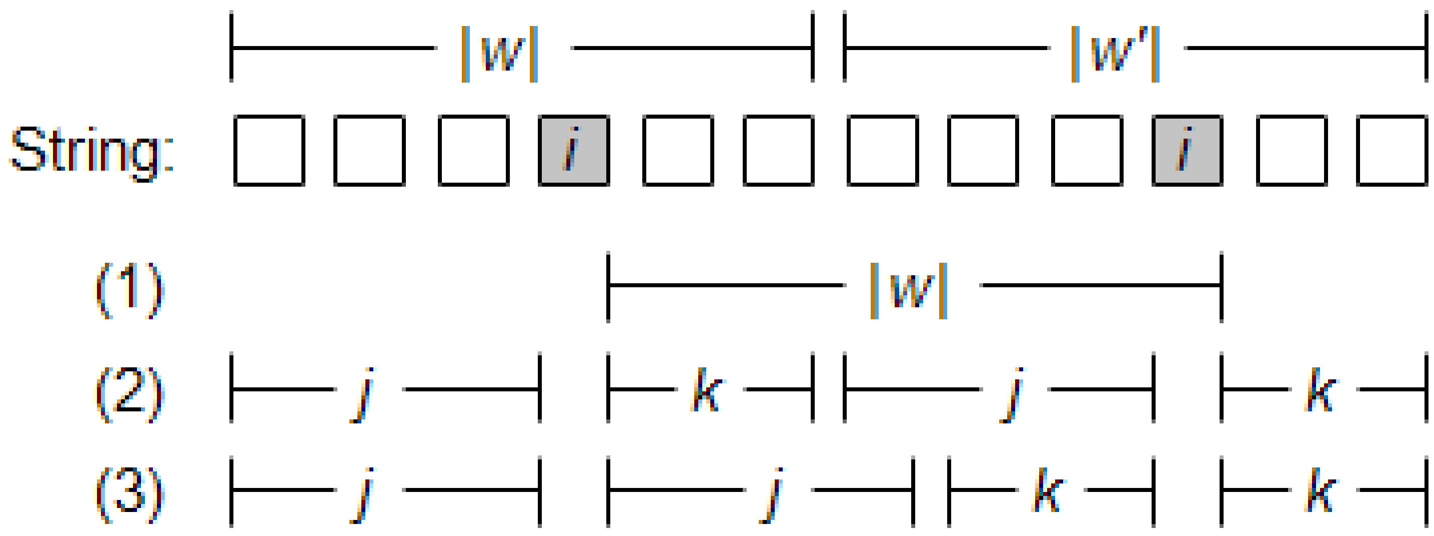
Grammar: Odd-Length (Unevens)

U --> ZUZ | Z

Z --> **a** | **b**

Exploration: Even-Length Strings

- Three key insights into minimal difference:



And $|w|= |x|/2$ = half the length of the string

Grammar: Even-Length (Evens)

E --> AB | BA

A --> ZAZ | **a**

B --> ZBZ | **b**

Z --> **a** | **b**

G_L: The Complete Grammar

S --> E | U | **ε**

E --> AB | BA

A --> ZAZ | **a**

B --> ZBZ | **b**

U --> ZUZ | Z

Z --> **a** | **b**

- non-*ww* includes ε.

But is G_L correct?

- Need to prove:
  - Every string produced by G_L is in L
  - Every string in L can be produced by G_L

Every string produced by G_L is in *L*

- S --> E | U | **ε**
  - **ε** is in *L* (ie, not-ww).
  - Now consider U and E

- U --> ZUZ | Z
- Z --> **a** | **b**
  - U => 1 symbol, 3 symbols, 5 symbols... always odd.
  - All in *L*.

Every string produced by G_L is in *L*

- E --> AB | BA

  A --> ZAZ | **a**

  B --> ZBZ | **b**

  Z --> **a** | **b**

- Consider AB =*=> (A expanded j + 1 times) and (B expanded k + 1 times)

  ==> ZZZ **a** ZZZ ZZ **b** ZZ = *x*

  - **a** with *j* pairs of Zs; **b** with *k* pairs of Zs
  - j and k >= 0; j + k + 1 = |x|/2
  - even-length; a and b distance of |x|/2 apart

Okay...

- So every string produced by G_L is in L
- But what about...

  Every string in L can be produced by G_L

Every string in *L* produced by G_L

- Zero case: ε
  - S => **ε**

Every string in *L* produced by G_L

- Odd-length case: proof by induction on |*x*|
  - *x* is in *L*

  - Basis: |x| = 1
  - S => U =*=> Z.   Z can then become any in {a, b}
  - So any *x* of length 1 in L can be generated.

  - Induction: |*x*| = n, n is odd, and S => U =*=> *x*.
  - next odd = n + 2 = |*x*| + 2 = ZxZ. (pre and post to *x*)
  - From induction hypothesis: ZUZ.
  - So we can generate any odd-length *x* in *L*

Every string in *L* produced by G_L

- Even-length case: reverse derivation.
  - Take any even-length *x* in *L*
  - As discussed, *x* must have two symbols that differ at a distance of |*x*|/2 for each other. Replace these a → A and b → B.
  - Replace all other symbols with Z.
  - This will produce a string of the form:
    $ZZZ_j\ A\ ZZZ_j\ ZZ_k\ B\ ZZ_k$ (or with A and B reversed)
  - Such a string is derivable from the grammar, so any even-length *L* is also in $L(G_L)$

Conclusion

- Chomsky hierarchy: different classes of languages out there, with different relationships (such as complement) between them
- While the language of ww is not context-free, not-ww is context-free.
- There is a CF grammar to prove it
- And there is a proof to prove the grammar

Thank You

Questions or comments?