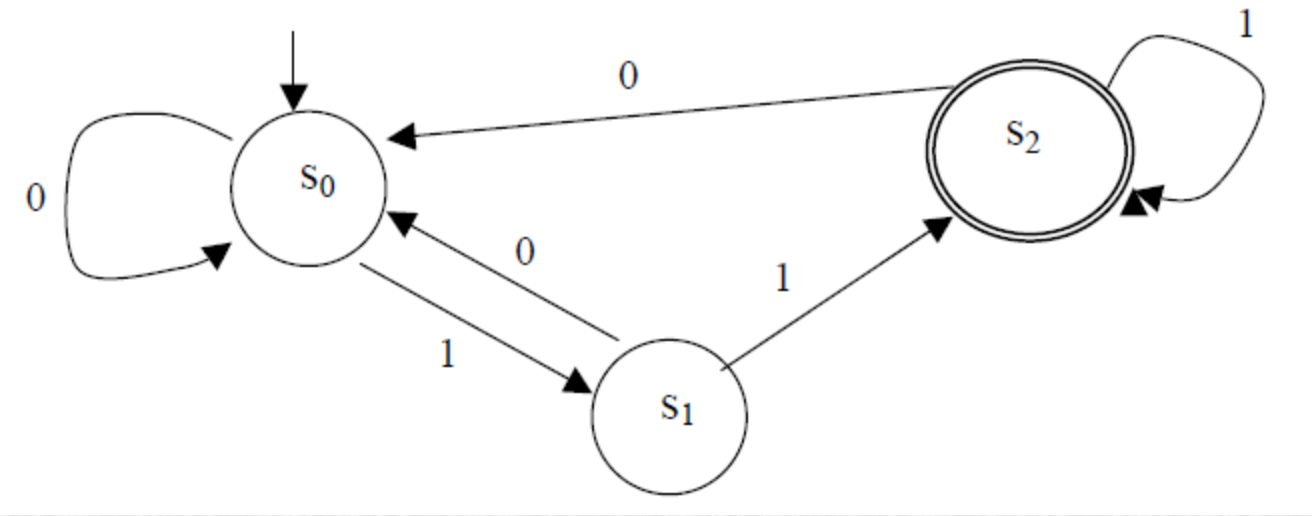
Computational Linguistics

Lecture 3

Dr. Dina Khattab

dina.khattab@cis.asu.edu.eg

FSA EXERCISES



- Consider the finite-state automaton *A*. Which of the strings are accepted by *A*?

i. 11

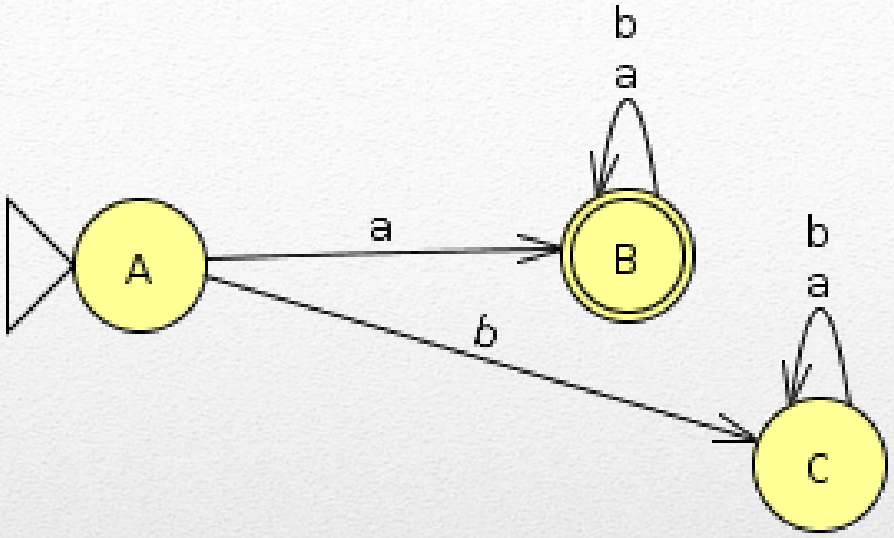
ii. 0101

iii. 011011

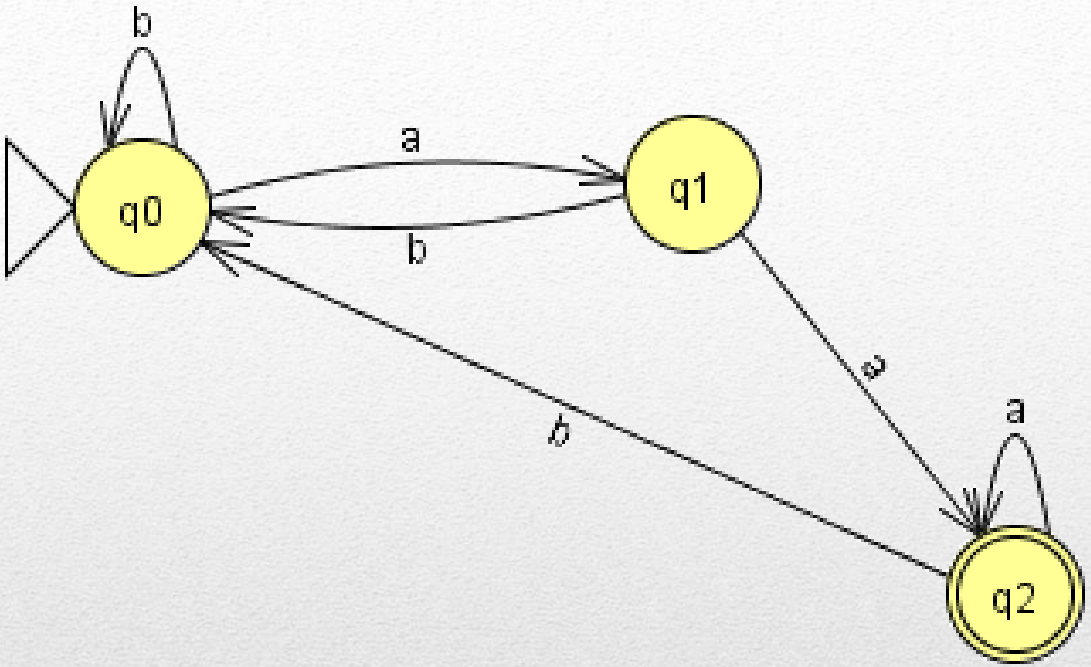
iv. 00110

- What is the language accepted by *A?*

Accepts any string that ends with at least two "1"s

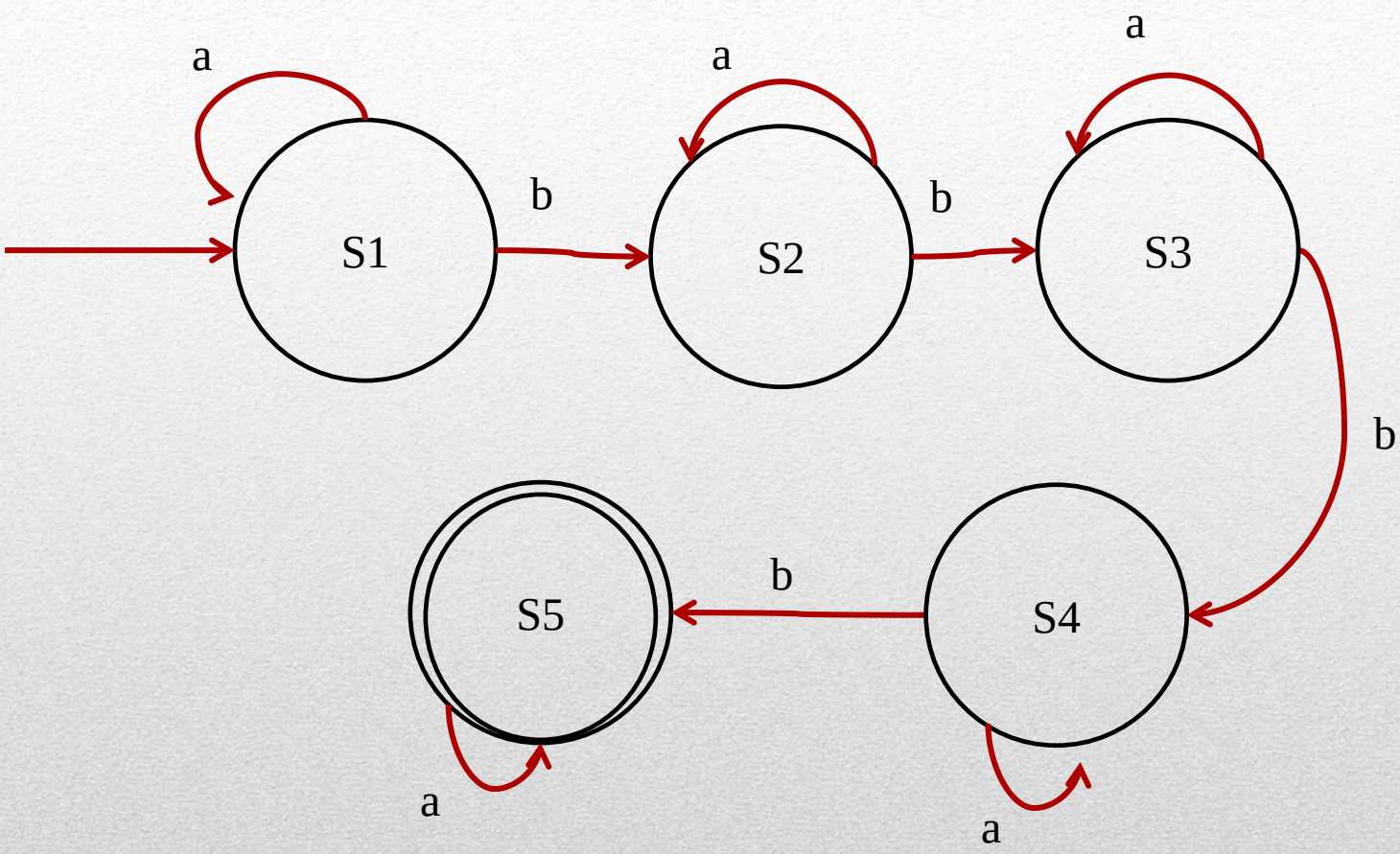


- *What language do you think this accepts?*
- This accepts all strings over a ,b that start with an ‘a’.



- *What language do you think this accepts?*
- This accepts all strings over a ,b that end with two ’a’s.

- Draw a graph for a FSA that accepts the language that consists of strings containing exactly 4 b's and $\Sigma = \{ a, b \}$.


a
a
a
S1
b
S2
b
S3
b
S4
b
S5
a
a

REGULAR EXPRESSION (REGEX)

Regular Expressions

- Easy way to generate a language that is accepted by FSA
- Rules:
  - ε is a regular expression
  - Any symbol in Σ is a regular expression

  If *r* and *s* are any regular expressions then so is:
  - *r* | *s* denotes union e.g. "*r* or *s*"
  - *rs* denotes *r* followed by *s (concatination)*
  - (*r*)* denotes concatination of r with itself zero or more times (Kleene closure).
  - (*r*)+ denotes concatination of r with itself one or more times (positive closure).
  - () used for controlling order of operations

Example Regular Expressions

Regular Expression	Corresponding Language
ε	ε
a	a
abc	abc
a|b|c	a, b, c
ab*	a, ab, abb, abbb, ….
ab+	ab, abb, abbb, …
(ab)*	ε, ab, abab, ababab, ….
(a|b|c)*	ε, a, b, c, aa, ab, ac, aaa, …
a|b*	ε, a, b, bb, bbb, ….
a|b|c|…|z|A|B|…|Z	Any letter
0|1|2|…|9	Any digit

Precedence in Regular Expressions

- * has highest precedence, left associative.
- Concatenation has second highest precedence, left associative.
- | has lowest associative, left associative.