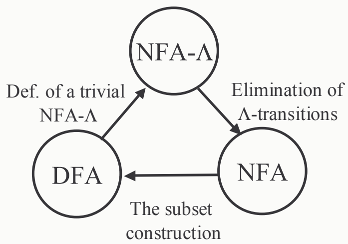
Session 11

Kleene's Theorem

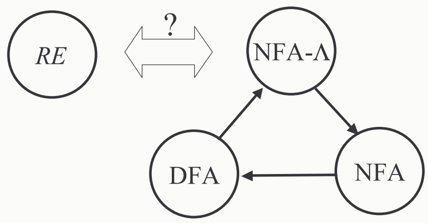
Kleene's Theorem

- A language L over the alphabet Σ is regular (can be expressed through a regular expression) if and only if there is an FA with input alphabet Σ that accepts L
  - Part 1 (only if): If there is a *RE* for L then there is a FA accepting L
  - Part 2 (if): If there is a FA accepting L then there is a *RE* for L

The Family of FA



RE and FA

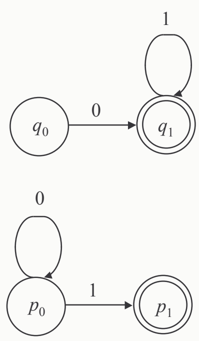


We can use any of these as the three are equivalent, but a particular choice may be better than the others!

Similarity of RE and NFA-Λ

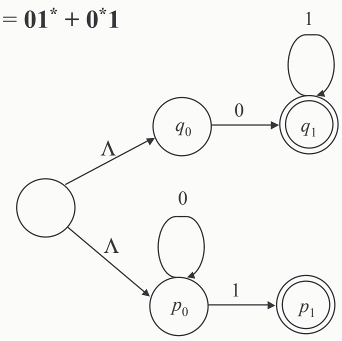
01^*

0^*1



Similarity of RE and NFA-Λ

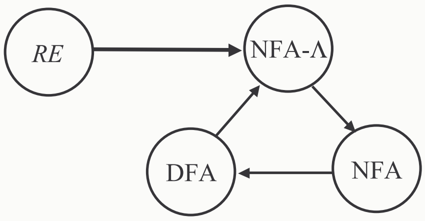
- $L_1 \cup L_2 = \mathbf{01^* + 0^*1}$



01^*

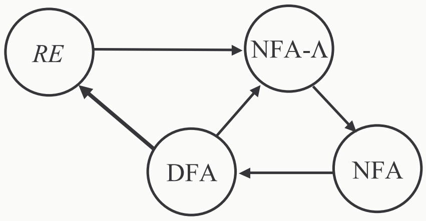
0^*1

The simplest strategy



If there are FA for the parts of a *RE*, we can compose a composite FA corresponding to the *RE* out of the simpler FA's (with the help of Λ-transitions)

The simplest strategy



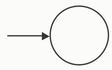
In this direction, we need to construct a *RE* that accounts for all strings of any length that are accepted by the corresponding FA

Proof of the theorem: Part 1

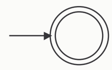
- All *RE* are built out of
  - Φ, Λ and All $\boldsymbol{a} \in \Sigma$
- And the composition operators
  - *E+F, EF* and E^*
- Automata Composition:
  - Define a FA for the basic parts
  - Define the form of FA for the three operators
  - Construct the FA in tandem with the corresponding *RE*

Proof of the theorem: Part 1

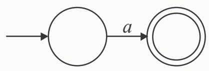
- Φ



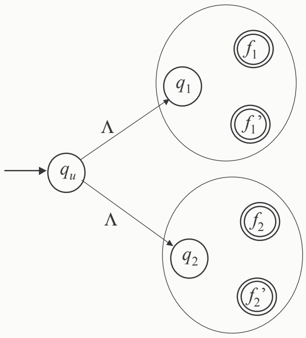
- Λ



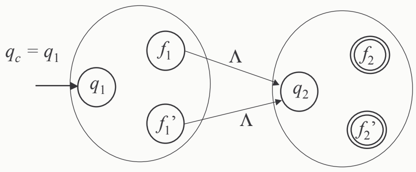
- All $\boldsymbol{a} \in \Sigma$



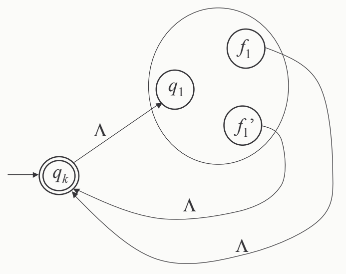
The union pattern: *E + F*



The concatenation pattern: *EF*



The closure pattern: E^*



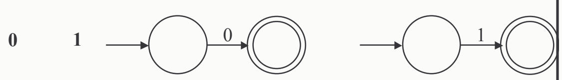
The construction

- With the FA for the basic *RE*
- With FA schemes for the composite *RE*
- Construct the composite FA in tandem with the structure of the *RE*!
- An example:

  Construct the FA for the RE: $(\mathbf{01})^* + (\mathbf{10})^*$

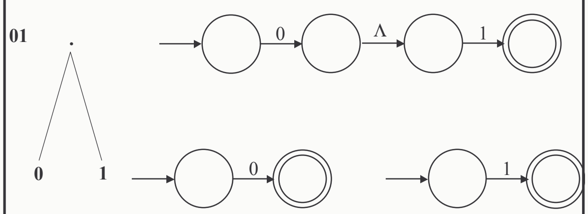
Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

Example: the FA for (**01**)*



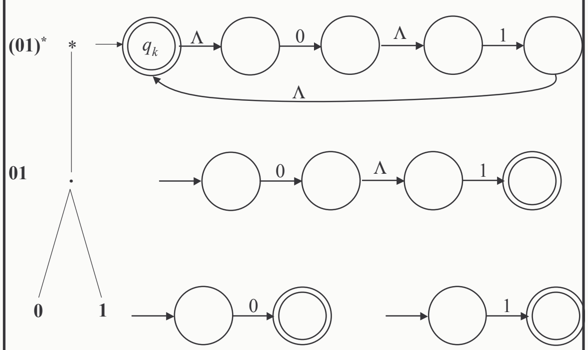
Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

Example: the FA for (**01**)*



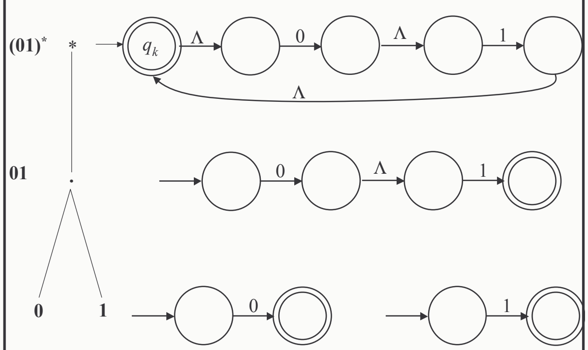
Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

Example: the FA for (**01**)*



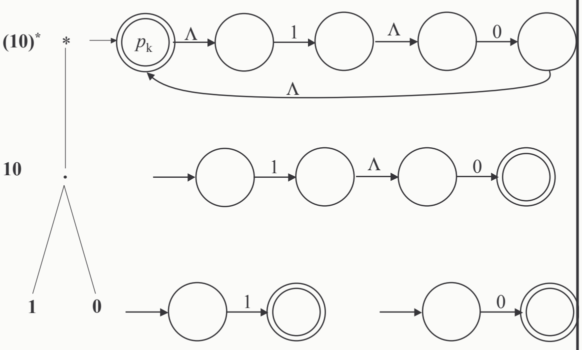
Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

Example: the FA for (**01**)*

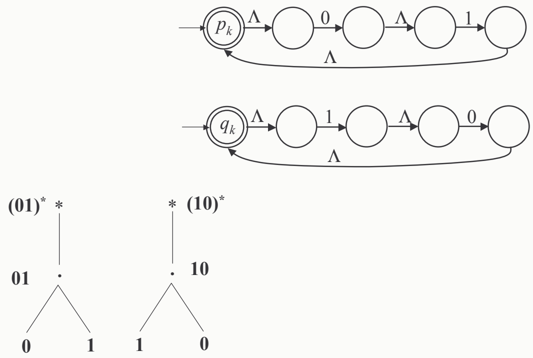


Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

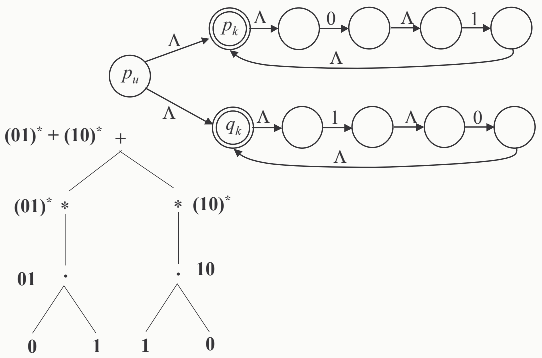
Example: the FA for **(10)***



Example: the FA for **(01)*** + **(10)***



Example: the FA for **(01)*** + **(10)***



Kleene's Theorem

✓Part 1 (only if): If there is a regular language *L* (expressed by *RE*) then there is a FA accepting *L*

Kleene's Theorem

✓Part 1 (only if): If there is a regular language *L* (expressed by *RE*) then there is a FA accepting *L*

● Part 2 (if): If there is a FA accepting *L* then there is a regular expression for *L*

The sets accepted by a FA

- Let $L \subseteq \Sigma^*$ be accepted by the FA M = $(Q, \Sigma, q_0, A, \delta)$.
  - this is:

    $L = \{x \in \Sigma^* \mid \delta^*(q_0, x) \in A\}$
  - If $A = \{q_i, \ldots, q_j\}$ then L is the union of a finite number of sets of the form (one for each accepting state):

    $L = \{x \in \Sigma^* \mid \delta^*(q_0, x) = q\}$
  - Consequently there must be a *RE* denoting each of these sets!
  - Also, if p and q are states of a FA, there must be a *RE* for the set:

    $L(p, q) = \{x \in \Sigma^* \mid \delta^*(p, x) = q\}$

The sets accepted by a FA

- The *RE* corresponding to a FA will be:
  - The union of the *RE*

    $L(q_0, q) = \{x \in \Sigma^* \mid \delta^*(q_0, x) = q\}$

    for every $q \in A = \{q_i, \ldots, q_j\}$
- In order to find the *RE* corresponding to this set, we construct the *RE* for the languages recognized by all possible paths of the form:

  $L(p, q) = \{x \in \Sigma^* \mid \delta^*(p, x) = q\}$

Path going through a state

- Let's assign a natural number n to every state of a FA
- For a string $x \in \Sigma^*$, x represents a path from state p to state q through state s if there are non null strings y and z so that
  - $x = yz$ $\quad \delta^*(p, y) = s \quad \delta^*(s, z) = q$
- A path can go from or to a state without going through it

  $p \xrightarrow{y} s \xrightarrow{z} q$
- Let the language of all paths that go through a state whose number is not higher than j be:

  $L(p, q, j)$ where $j \geq 0$
- There are n states, so $L(p, q, n) = L(p, q)$

Proof that a language is regular

- The induction:
  - We have to show that $L(p, q, n)$ is regular, and this will be the case if $L(p, q, j)$ is regular for every j such that $0 \leq j \leq n$
- The basis:
  - We need to show that $L(p, q, 0)$ is regular
  - This is, the languages of paths that go through no state in a FA are regular:

    $L(p, q, 0) \subseteq \Sigma \cup \{\Lambda\}$
  - $L(p, q, 0)$ is regular because the number of states is finite, so the number of paths!

Proof that a language is regular

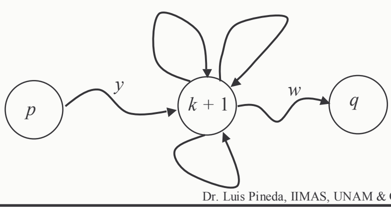
- The induction hypothesis:
  - $0 \leq k$
  - for every p and q satisfying $0 \leq p$ and $q \leq n$ the language $L(p, q, k)$ is regular
  - we want to show that $L(p, q, k + 1)$ is regular
- A string x is in $L(p, q, k + 1)$ if it represents a path from p to q that goes trough a state no higher than $k + 1$; there are two ways this can happen:
  - Case 1: The strings does not go through state $k + 1$, so it does not go through a state higher than k, then

    $x \in L(p, q, k)$

Proof that a language is regular

- Case 2: The strings goes through state $k + 1$, but not through a state whose number is higher than $k + 1$; for this:

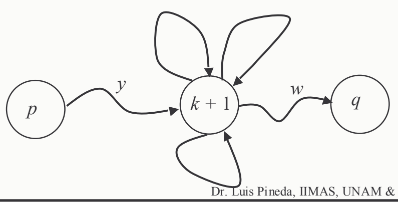
  $x = yzw$



Proof that a language is regular

- $x = yzw$
  - $y \in L(p, k + 1, k)$ — Reaches state $k + 1$
  - $z \in L(k + 1, k + 1, k)^*$ — Goes through state $k + 1$ any times
  - $w \in L(k + 1, q, k)$ — Starts in state $k + 1$
- x is constructed with concatenation and *-closure:

  $x \in L(p, k + 1, k)L(k + 1, k + 1, k)^*L(k + 1, q, k)$



The two cases together!

- Case 1:
  - $x \in L(p, q, k)$
- Case 2:
  - $x \in L(p, k+1, k)L(k+1, k+1, k)^*L(k+1, q, k)$
- Then

  $L(p, q, k+1) =$

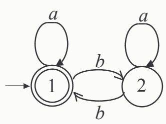
  $L(p, q, k) \cup L(p, k+1, k)L(k+1, k+1, k)^*L(k+1, q, k)$
- $L(p, q, k+1)$ is constructed with:
  - union
  - concatenation
  - *-closure!
- Then, $L(p, q, k+1)$ is regular!

So…?

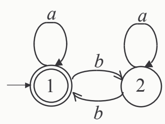
- The proof of the theorem provides an algorithm to find the *RE* corresponding to any FA:
- $L(p, q, 0) = \begin{cases} \{a \in \Sigma \mid \delta(p, a) = q\} & \text{if } p \neq q \\ \{a \in \Sigma \mid \delta(p, a) = q\} \cup \{\Lambda\} & \text{if } p = q \end{cases}$
- $L(p, q, k+1) =$

  $L(p, q, k) \cup L(p, k+1, k)L(k+1, k+1, k)^*L(k+1, q, k)$
- $L(p, q, n) = L(p, q)$ and $L = \bigcup_{q \in A} L(q_0, q)$

What is RE corresponding to a FA?



The RE corresponding to a FA



⟺

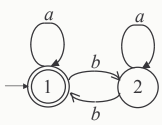
$\mathbf{(a^*ba^*ba^*)^*}$ **?**

$\mathbf{a^* + (ba^*ba^*)^*}$ **?**

$\mathbf{a^* + (a^*ba^*b)^*}$ **?**

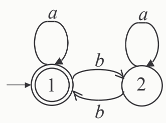
The basic case: $r(p,q,0)$

$L(p, q, 0) = \begin{cases} \{a \in \Sigma \mid \delta(p, a) = q\} & \text{if } p \neq q \\ \{a \in \Sigma \mid \delta(p, a) = q\} \cup \{\Lambda\} & \text{if } p = q \end{cases}$



p	$r(p,1,0)$	$r(p,2,0)$
1	$a + \Lambda$	b
2	b	$a + \Lambda$

The inductive part: $r(p,q,k+1)$

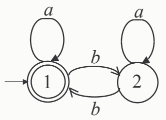


p	$r(p,1,k+1)$	$r(p,2,k+1)$
1		
2		

$L(p, q, k+1) =$

$L(p, q, k) \cup L(p, k+1, k)L(k+1, k+1, k)^*L(k+1, q, k)$

The inductive part: $r(p,q,1)$



p	$r(p,1,1)$	$r(p,2,1)$
1	a^*	
2		

$L(p, q, k + 1) =$

$L(p, q, k) \cup L(p, k + 1, k)L(k + 1, k + 1, k)^*L(k + 1, q, k)$

$k = 0, k + 1 = 1$

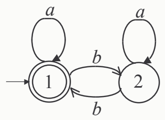
$L(1, 1, 1) = L(1, 1, 0) \cup L(1, 1, 0)L(1, 1, 0)^*L(1, 1, 0)$

$= (a + \Lambda) + (a + \Lambda)(a + \Lambda)^*(a + \Lambda)$

$= a^*$

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

The inductive part: $r(p,q,1)$



p	$r(p,1,1)$	$r(p,2,1)$
1	a^*	$b + a^*b$
2		

$L(p, q, k + 1) =$

$L(p, q, k) \cup L(p, k + 1, k)L(k + 1, k + 1, k)^*L(k + 1, q, k)$

$k = 0, k + 1 = 1$

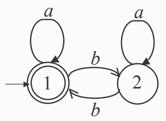
$L(1, 2, 1) = L(1, 2, 0) \cup L(1, 1, 0)L(1, 1, 0)^*L(1, 2, 0)$

$= b + (a + \Lambda)(a + \Lambda)^*b$

$= b + a^*b$

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

The inductive part: $r(p,q,1)$



p	$r(p,1,1)$	$r(p,2,1)$
1	a^*	$b + a^*b$
2		

$L(p, q, k + 1) =$

$L(p, q, k) \cup L(p, k + 1, k)L(k + 1, k + 1, k)^*L(k + 1, q, k)$

$k = 0, k + 1 = 1$

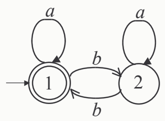
$L(1, 2, 1) = L(1, 2, 0) \cup L(1, 1, 0)L(1, 1, 0)^*L(1, 2, 0)$

$= b + (a + \Lambda)(a + \Lambda)^*b$

$= b + a^*b$

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

Simplifying…



p	$r(p,1,1)$	$r(p,2,1)$
1	a^*	a^*b
2		

$L(p, q, k + 1) =$

$L(p, q, k) \cup L(p, k + 1, k)L(k + 1, k + 1, k)^*L(k + 1, q, k)$

$k = 0, k + 1 = 1$

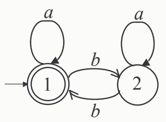
$L(1, 2, 1) = L(1, 2, 0) \cup L(1, 1, 0)L(1, 1, 0)^*L(1, 2, 0)$

$= b + (a + \Lambda)(a + \Lambda)^*b$

$= b + a^*b = a^*b$

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

The inductive part: $r(p,q,1)$



p	$r(p,1,1)$	$r(p,2,1)$
1	a^*	a^*b
2	ba^*	

$L(p, q, k + 1) =$

$L(p, q, k) \cup L(p, k + 1, k)L(k + 1, k + 1, k)^*L(k + 1, q, k)$

$k = 0, k + 1 = 1$

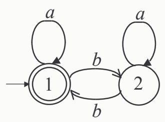
$L(2, 1, 1) = L(2, 1, 0) \cup L(2, 1, 0)L(1, 1, 0)^*L(1, 1, 0)$

$= b + b(a + \Lambda)^*(a + \Lambda)^*$

$= b + ba^* = ba^*$

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

The inductive part: $r(p,q,1)$



p	$r(p,1,1)$	$r(p,2,1)$
1	a^*	a^*b
2	ba^*	$\Lambda+a+ba^*b$

$L(p, q, k + 1) =$

$L(p, q, k) \cup L(p, k + 1, k)L(k + 1, k + 1, k)^*L(k + 1, q, k)$

$k = 0, k + 1 = 1$

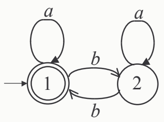
$L(2, 2, 1) = L(2, 2, 0) \cup L(2, 1, 0)L(1, 1, 0)^*L(1, 2, 0)$

$= (a + \Lambda) + b(a + \Lambda)^*b$

$= a + \Lambda + ba^*b$

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

The inductive part: $r(p,q,1)$



p	$r(p,1,1)$	$r(p,2,1)$
1	a^*	a^*b
2	ba^*	$\Lambda+a+ba^*b$

$L(p, q, k + 1) =$

$L(p, q, k) \cup L(p, k + 1, k)L(k + 1, k + 1, k)^*L(k + 1, q, k)$

$k = 0, k + 1 = 1$

$L(2, 2, 1) = L(2, 2, 0) \cup L(2, 1, 0)L(1, 1, 0)^*L(1, 2, 0)$

$= (a + \Lambda) + b(a + \Lambda)^*b$

$= a + \Lambda + ba^*b$

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

The inductive part: $r(p,q,2)$

p	$r(p,1,1)$	$r(p,2,1)$
1	a^*	a^*b
2	ba^*	$\Lambda+a+ba^*b$

p	$r(p,1,2)$	$r(p,2,2)$
1	$RE_{11\text{-}2}$	
2		

$L(p, q, k + 1) =$

$L(p, q, k) \cup L(p, k + 1, k)L(k + 1, k + 1, k)^*L(k + 1, q, k)$

$k = 1, k + 1 = 2$

$L(1, 1, 2) = L(1, 1, 1) \cup L(1, 2, 1)L(2, 2, 1)^*L(2, 1, 1)$

$= a^* + (a^*b)(\Lambda + a + ba^*b)^*(ba^*)$

$= RE_{11\text{-}2}$

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

The inductive part: $r(p,q,2)$

p	$r(p,1,1)$	$r(p,2,1)$
1	a^*	a^*b
2	ba^*	$\Lambda+a+ba^*b$

p	$r(p,1,2)$	$r(p,2,2)$
1	$RE_{11\text{-}2}$	$RE_{12\text{-}2}$
2		

$L(p, q, k + 1) =$

$L(p, q, k) \cup L(p, k + 1, k)L(k + 1, k + 1, k)^*L(k + 1, q, k)$

$k = 1, k + 1 = 2$

$L(1, 2, 2) = L(1, 2, 1) \cup L(1, 2, 1)L(2, 2, 1)^*L(2, 2, 1)$

$= a^*b + (a^*b)(\Lambda + a + ba^*b)^*(a + ba^*b)$

$= RE_{12\text{-}2}$

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

The inductive part: $r(p,q,2)$

p	$r(p,1,1)$	$r(p,2,1)$
1	a^*	a^*b
2	ba^*	$\Lambda+a+ba^*b$

p	$r(p,1,2)$	$r(p,2,2)$
1	$RE_{11\text{-}2}$	$RE_{12\text{-}2}$
2	$RE_{21\text{-}2}$	

$L(p, q, k + 1) =$

$L(p, q, k) \cup L(p, k + 1, k)L(k + 1, k + 1, k)^*L(k + 1, q, k)$

$k = 1, k + 1 = 2$

$L(2, 1, 2) = L(2, 1, 1) \cup L(2, 2, 1)L(2, 2, 1)^*L(2, 1, 1)$

$= ba^* + (a + ba^*b)(\Lambda + a + ba^*b)^*(ba^*)$

$= RE_{21\text{-}2}$

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

The inductive part: $r(p,q,2)$

p	$r(p,1,1)$	$r(p,2,1)$
1	a^*	a^*b
2	ba^*	$\Lambda+a+ba^*b$

p	$r(p,1,2)$	$r(p,2,2)$
1	$RE_{11\text{-}2}$	$RE_{12\text{-}2}$
2	$RE_{21\text{-}2}$	$RE_{22\text{-}2}$

$L(p, q, k + 1) =$

$L(p, q, k) \cup L(p, k + 1, k)L(k + 1, k + 1, k)^*L(k + 1, q, k)$

$k = 1, k + 1 = 2$

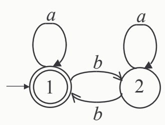
$L(2, 2, 2) = L(2, 2, 1) \cup L(2, 2, 1)L(2, 2, 1)^*L(2, 2, 1)$

$= (a + ba^*b) + (a + ba^*b)(\Lambda + a + ba^*b)^*(a + ba^*b)$

$= RE_{22\text{-}2}$

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

The inductive part: $r(p,q)$



p	$r(p,1,2)$	$r(p,2,2)$
1	$RE_{11\text{-}2}$	$RE_{12\text{-}2}$
2	$RE_{21\text{-}2}$	$RE_{22\text{-}2}$

The end of the induction: $(p, q, n) = L(p, q)$ and

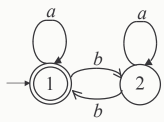
$$L = \bigcup_{q \in A} L(q_0, q)$$

$L(1, 1, 2) = L(1, 1, 1) \cup L(1, 2, 1)L(2, 2, 1)^*L(2, 1, 1)$

$= a^* + (a^*b)(\Lambda + a + ba^*b)^*(ba^*)$

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

The inductive part: $r(p,q)$



$(a^*ba^*ba^*)^*$?

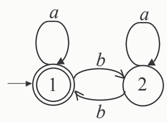
$a^* + (ba^*ba^*)^*$?

$a^* + (a^*ba^*b)^*$?

$L(1, 1, 2) = a^* + (a^*b)(\Lambda + a + ba^*b)^*(ba^*)$

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

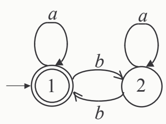
The inductive part: $r(p,q)$



$L(1, 1, 2) = a^* + (a^*b)(\Lambda + a + ba^*b)^*(ba^*)$

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

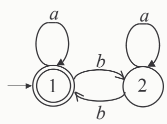
The inductive part: $r(p,q)$



$L(1, 1, 2) = a^* + (a^*b)(\Lambda + a + ba^*b)^*(ba^*)$

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

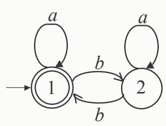
The inductive part: $r(p,q)$



$L(1, 1, 2) = a^* + (a^*b)(\Lambda + a + ba^*b)^*(ba^*)$

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

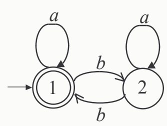
The inductive part: $r(p,q)$



$L(1, 1, 2) = a^* + (a^*b)(\Lambda + a + ba^*b)^*(ba^*)$

Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

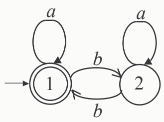
The inductive part: $r(p,q)$



$L(1, 1, 2) = a^* + (a^*b)(\Lambda + a + ba^*b)^*(ba^*)$

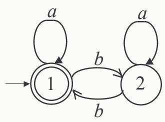
Dr. Luis Pineda, IIMAS, UNAM & OSU-CIS, 2003

The inductive part: $r(p,q)$



$$L(1, 1, 2) = a^* + (a^*b)(\Lambda + a + ba^*b)^*(ba^*)$$

The inductive part: $r(p,q)$



$$L(1, 1, 2) = a^* + (a^*b)(\Lambda + a + ba^*b)^*(ba^*)$$

The expression that reflects better the underlying structure of the FA

Kleene's Theorem

- ✓Part 1 (only if): If there is a regular language L (expressed by *RE*) then there is a FA accepting L
- ✓Part 2 (if): If there is a FA accepting L then there is a regular expression for L

Kleen's Theorem