Computational Learning Theory
Learning with EFSs

Akihiro Yamamoto 山本 章博
http://www.iip.ist.i.kyoto-u.ac.jp/member/akihiro/
akihiro@i.kyoto-u.ac.jp

Examples

Example 3

$C_3 = \{$aaabbb, ab, aaaabbbb, aaaaabbbbb, aabb$\}$

$D_3 = \{$a, b, bbbb, abb, baaaaba, babbb$\}$

- *Every string in C_3 consists of two strings: The first half consists only of* a*'s, and the second consists of the same number of* b*'s.*

No pattern represents the rule.

Example 3

C_3 ={aaabbb, ab, aaaabbbb, aaaaabbbbb, aabb}

- *Every string in C_3 consists of two strings: The first half consists only of* a*'s, and the second consists of the same number of* b*'s.*

- Assume a pattern π represents the rule.

If π has variables $x_1,\ldots, x_n$, a contradiction is derived by substituting an a to each of x_i, that is, applying $\theta = \{(x_1, \mathsf{a}), \ldots, (x_n, \mathsf{a})\}$ to π.

An EFS represents the rule.

Example 3

C_3 ={aaabbb, ab, aaaabbbb, aaaaabbbbb, aabb}

- *Every string in C_3 consists of two strings: The first half consists only of* a*'s, and the second consists of the same number of* b*'s.*

- Assume a pattern π represents the rule.

If π has variables $x_1, \ldots, x_n$, a contradiction is derived by substituting an a to each of x_i, that is, applying $\theta = \{(x_1, \mathsf{a}), \ldots, (x_n, \mathsf{a})\}$ to π.

Elementary Formal Systems

Mathematical Logic

- Mathematical logic is a subfield of mathematics exploring the applications of formal logic to mathematics.[Wikipedia]

All humans are mortal.

Socrates is human.

$\Downarrow$ Inference

Socrates is mortal.

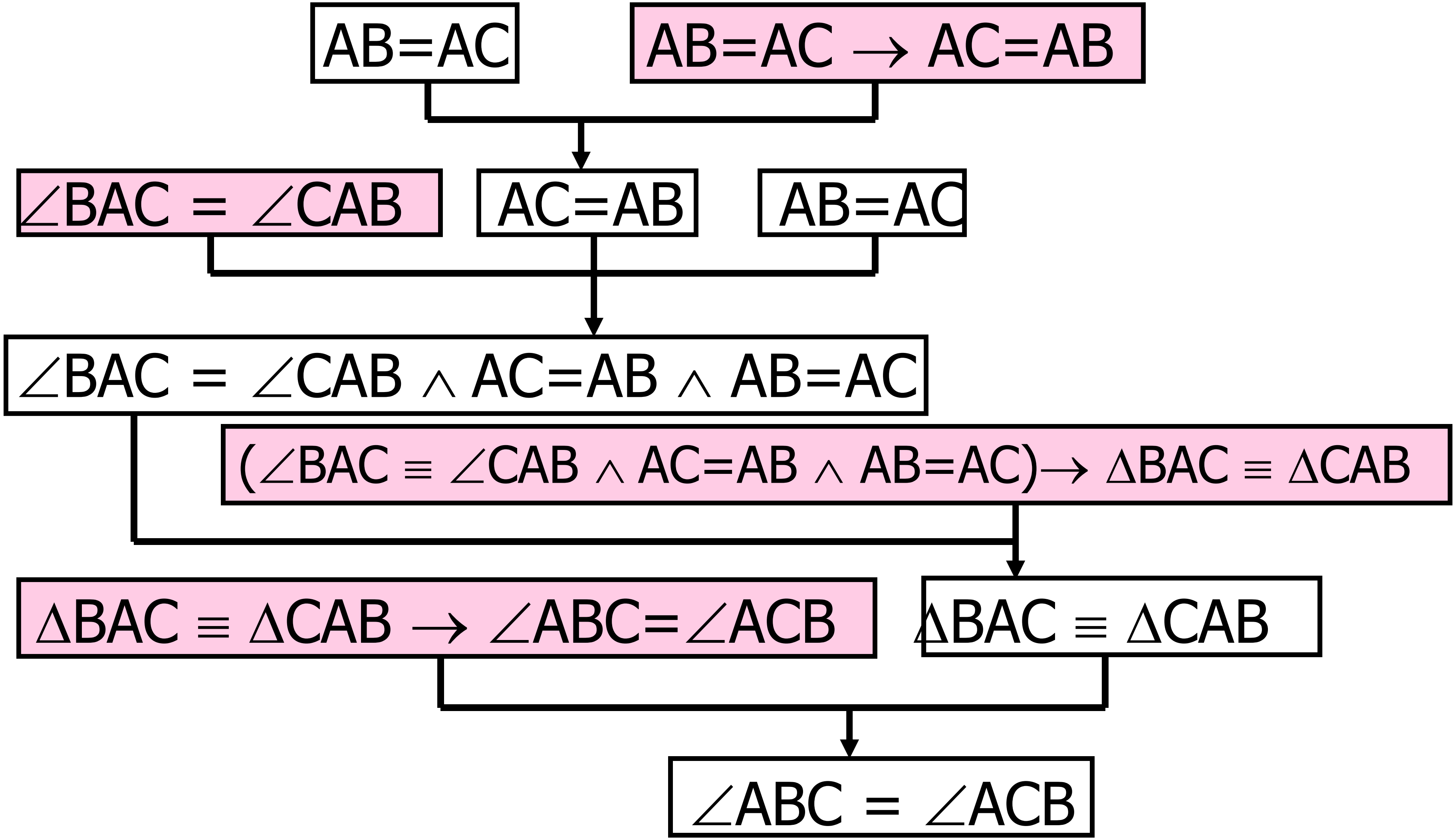
$\text{mortal}(x) \leftarrow \text{human}(x)$

$\text{human}(\text{Socrates})$

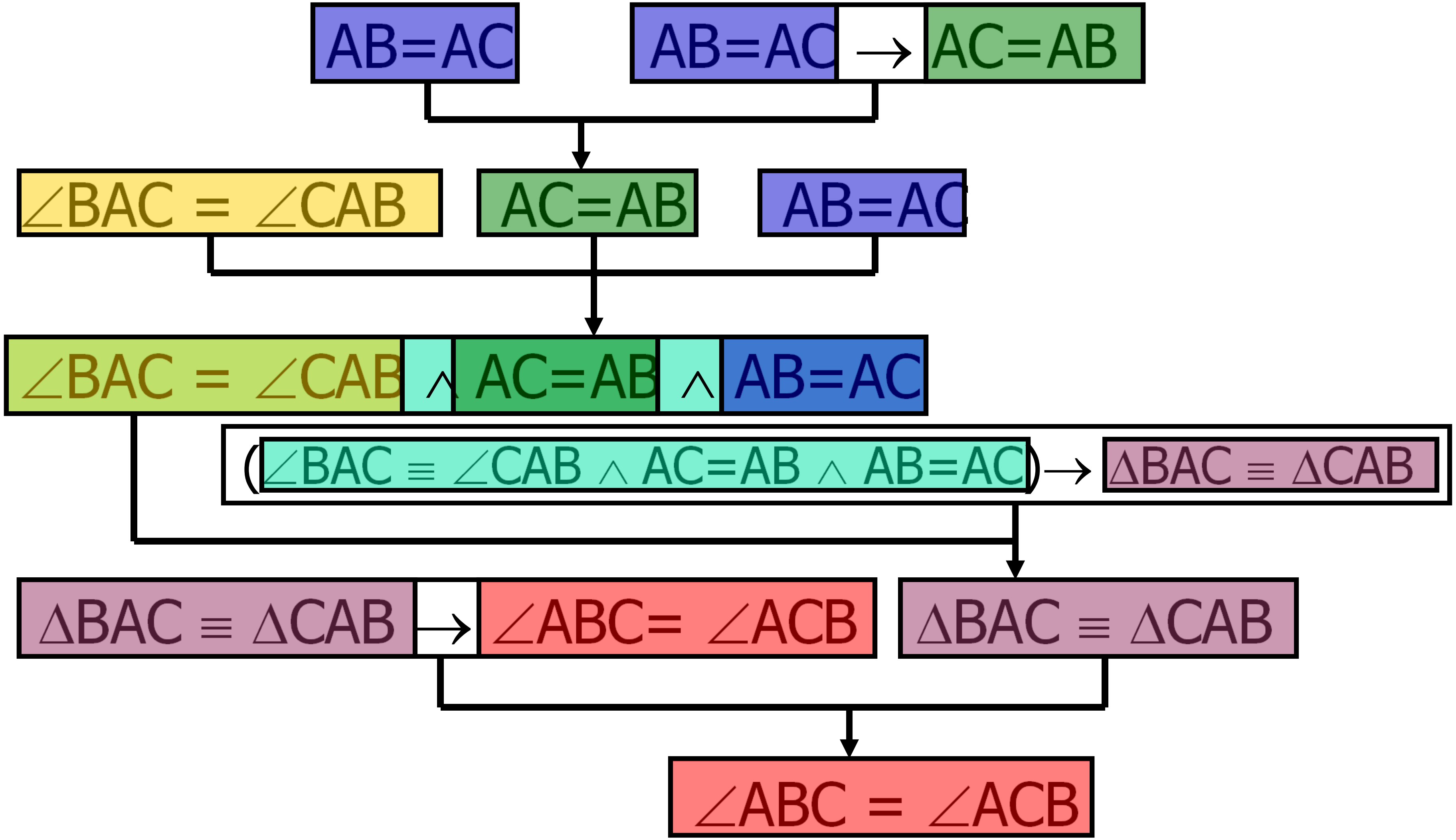
$\Downarrow$ Operation Algorithm

$\text{mortal}(\text{Socrates})$

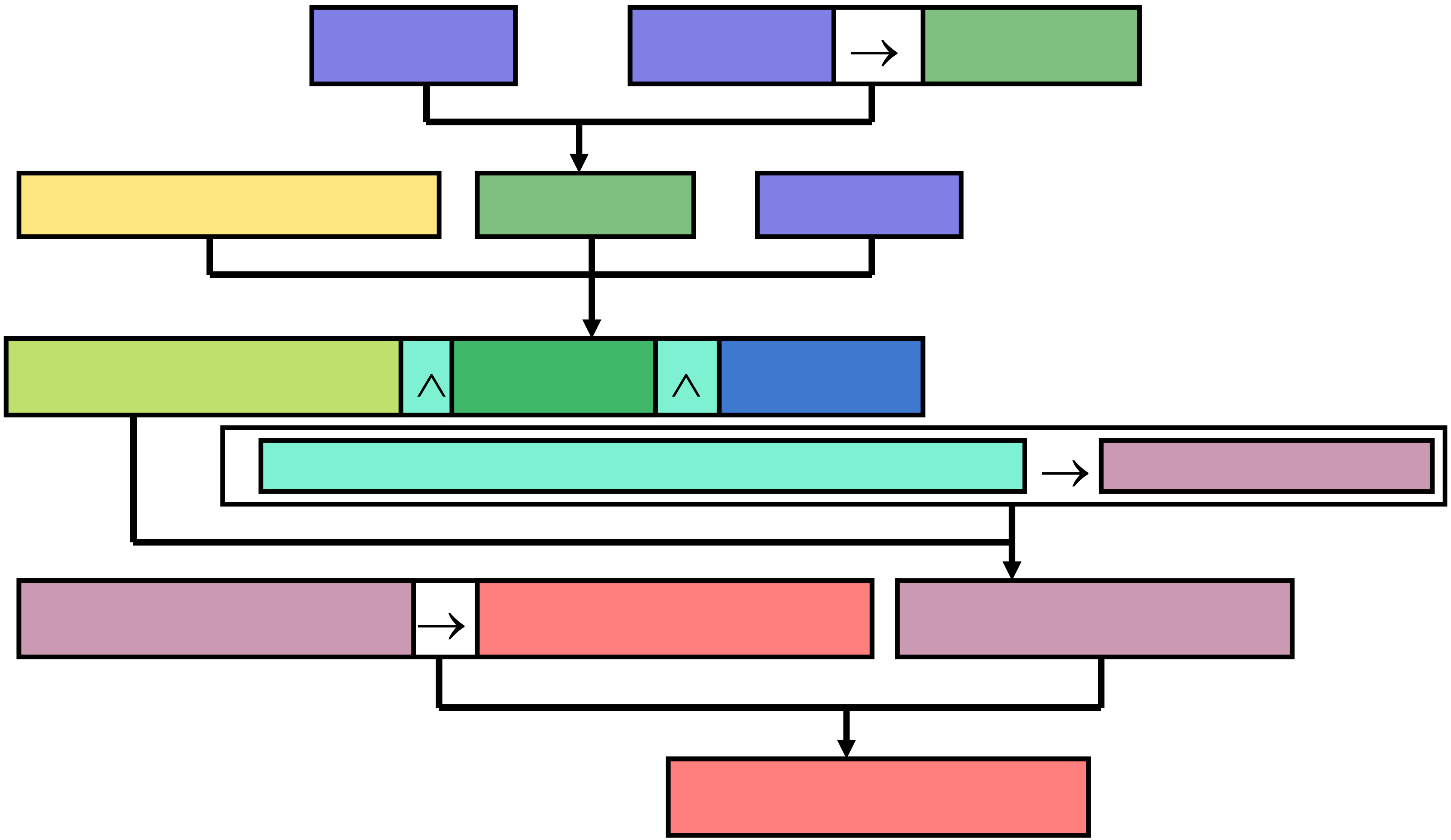
Relationship among Propositions



Relationship among Propositions



Relationship among Propositions



Predicate Symbols

- Introducing symbols of a new type.
  - Every predicate symbol is interpreted as a language (a set of strings) or a set of tuples $(s_1, s_2, \ldots, s_n)$ of strings.
  - In this course we use symbols p, q, r,… which are respectively interpreted as sets P, Q, R,….
- An atomic formula is a formula of the form

$$\mathsf{p}(\pi_1, \pi_2, \ldots, \pi_n)$$

where $\pi_1, \pi_2, \ldots, \pi_n$ are patterns. If n=1 and $\pi_1 = s$ is ground, $\mathsf{p}(s)$ is interpreted as $s \in \mathrm{P}$.

Example

Some examples of atomic formulae are $\mathsf{p}(\mathsf{a}x\mathsf{b})$, $\mathsf{q}(\mathsf{a}x, \mathsf{b}y)$, $\mathsf{q}(x, \mathsf{b}x\mathsf{b})$, $\mathsf{p}(\mathsf{aabb})$, $\mathsf{q}(\mathsf{aa}, \mathsf{bb})$. The last two formulae are ground.

Definite Clause (Rules) and EFS

- A definite clause is a formula of the form

$$p(\pi_1,\ldots,\pi_n) \leftarrow q_1(\tau_{11},\ldots), q_2(\tau_{21},\ldots),\ldots, q_k(\tau_{k1},\ldots)$$

where $\pi_1, \pi_2,\ldots, \tau_{11},\ldots, \tau_{k1},\ldots$ are patterns. The definite clause is interpreted as

"for any substitution θ, if $(\tau_{11}\theta,\ldots) \in Q_1, (\tau_{21}\theta,\ldots) \in Q_2,\ldots,$ $q_k(\tau_{k1}\theta,\ldots) \in Q_k$ then $(\pi_1\theta, \pi_2\theta,\ldots, \pi_n\theta) \in P$"

  - A clause $p(\pi_1,\ldots,\pi_n) \leftarrow$ which has no conditions is sometimes called a unit clause.

- A finite set of definite clause is called an elementary formal system (EFS). [Smullyan 61]

Examples

Some examples of definite clauses are

$$\mathrm{p}(\mathrm{a}x) \leftarrow \mathrm{r}(x)$$

$$\mathrm{r}(\mathrm{b}) \leftarrow$$

$$\mathrm{p}(\mathrm{a}x\mathrm{b}y) \leftarrow \mathrm{r}(x), \mathrm{r}(y)$$

$$\mathrm{q}(\mathrm{a}x, \mathrm{b}y) \leftarrow \mathrm{q}(x, y)$$

$$\ldots$$

$$\mathrm{animal}(x) \leftarrow \mathrm{human}(x)$$

$$\mathrm{objects}(x) \leftarrow \mathrm{animal}(x)$$

$$\mathrm{human}(xy) \leftarrow \mathrm{human}(x),\ \mathrm{human}(y)$$

$$\mathrm{human}(\mathrm{K}) \leftarrow$$

$$\mathrm{human}(\mathrm{N}) \leftarrow$$

Inference Rules for Definite Clauses

- We use the following two rules

  Instantiation

$$\frac{p(\pi_1,\ldots) \leftarrow q_1(\tau_{11},\ldots), q_2(\tau_{21},\ldots),\ldots, q_k(\tau_{k1},\ldots)}{(p(\pi_1,\ldots) \leftarrow q_1(\tau_{11},\ldots), q_2(\tau_{21},\ldots),\ldots, q_k(\tau_{k1},\ldots))\theta}$$

  Modus Pones

$$\frac{p(\pi_1,\ldots) \leftarrow p_1(\pi_{11},\ldots),\ldots, p_k(\pi_{k1},\ldots) \quad p_1(\pi_1,\ldots) \leftarrow q_1(\tau_{11},\ldots),\ldots}{p(\pi_1,\ldots) \leftarrow q_1(\tau_{11},\ldots),\ldots, p_2(\pi_{21},\ldots),\ldots, p_k(\pi_{k1},\ldots)}$$

- A proof is a continuous application of the inference rules.

Example of Proof (1)

$$
\begin{aligned}
S:\quad & \mathrm{a}(x) \leftarrow \mathrm{h}(x) \\
& \mathrm{o}(x) \leftarrow \mathrm{a}(x) \\
& \mathrm{h}(xy) \leftarrow \mathrm{h}(x),\ \mathrm{h}(y) \\
& \mathrm{h}(\mathrm{K}) \leftarrow \\
& \mathrm{h}(\mathrm{N}) \leftarrow \\
& \mathrm{h}(\mathrm{O}) \leftarrow
\end{aligned}
$$

$$
\cfrac{\cfrac{\cfrac{\mathrm{h}(xy) \leftarrow \mathrm{h}(x), \mathrm{h}(y)}{\mathrm{h}(\mathrm{NO}) \leftarrow \mathrm{h}(\mathrm{N}), \mathrm{h}(\mathrm{O})} \qquad \mathrm{h}(\mathrm{N}) \leftarrow}{\mathrm{h}(\mathrm{NO}) \leftarrow \mathrm{h}(\mathrm{N}) \qquad \mathrm{h}(\mathrm{O}) \leftarrow}}{\mathrm{h}(\mathrm{NO}) \leftarrow}
$$

Example of Proof (2)

$$
\begin{aligned}
S:\quad & \mathrm{a}(x) \leftarrow \mathrm{h}(x) \\
& \mathrm{o}(x) \leftarrow \mathrm{a}(x) \\
& \mathrm{h}(xy) \leftarrow \mathrm{h}(x),\ \mathrm{h}(y) \\
& \mathrm{h}(\mathrm{K}) \leftarrow \\
& \mathrm{h}(\mathrm{N}) \leftarrow \\
& \mathrm{h}(\mathrm{O}) \leftarrow
\end{aligned}
$$

$$
\dfrac{\begin{matrix}\mathrm{o}(x) \leftarrow \mathrm{a}(x) \\ \mathrm{o}(\mathrm{K}) \leftarrow \mathrm{a}(\mathrm{K})\end{matrix} \qquad \dfrac{\dfrac{\mathrm{a}(x) \leftarrow \mathrm{h}(x)}{\mathrm{a}(\mathrm{K}) \leftarrow \mathrm{h}(\mathrm{K})} \qquad \mathrm{h}(\mathrm{K}) \leftarrow}{\mathrm{a}(\mathrm{K}) \leftarrow}}{\mathrm{o}(\mathrm{K}) \leftarrow}
$$

Example of Proof (3)

S: $\mathrm{p}(\mathrm{a}x)\leftarrow\mathrm{q}(x)$ $\quad$ $\mathrm{q}(\mathrm{b}x)\leftarrow\mathrm{p}(x)$

$\mathrm{p}(\mathrm{a})\leftarrow$ $\quad$ $\mathrm{q}(\mathrm{b})\leftarrow$

$$\dfrac{\dfrac{\mathrm{p}(\mathrm{a}x)\leftarrow\mathrm{q}(x)}{\mathrm{p}(\mathrm{abab})\leftarrow\mathrm{q}(\mathrm{bab})} \quad \dfrac{\dfrac{\mathrm{q}(\mathrm{b}x)\leftarrow\mathrm{p}(x)}{\mathrm{q}(\mathrm{bab})\leftarrow\mathrm{p}(\mathrm{ab})} \quad \dfrac{\dfrac{\mathrm{p}(\mathrm{a}x)\leftarrow\mathrm{q}(x)}{\mathrm{p}(\mathrm{ab})\leftarrow\mathrm{q}(\mathrm{b})} \quad \mathrm{q}(\mathrm{b})\leftarrow}{\mathrm{p}(\mathrm{ab})\leftarrow}}{\mathrm{q}(\mathrm{bab})\leftarrow}}{\mathrm{p}(\mathrm{abab})\leftarrow}$$

Example of Proof (4)

S: $\mathrm{p(a}x\mathrm{b)} \leftarrow \mathrm{p(}x\mathrm{)}$

$\mathrm{p(ab)} \leftarrow$

$$\dfrac{\dfrac{\mathrm{p(a}x\mathrm{b)} \leftarrow \mathrm{p(}x\mathrm{)}}{\mathrm{p(aaabbb)} \leftarrow \mathrm{p(aabb)}} \qquad \dfrac{\dfrac{\mathrm{p(a}x\mathrm{b)} \leftarrow \mathrm{p(}x\mathrm{)}}{\mathrm{p(aabb)} \leftarrow \mathrm{p(ab)}} \qquad \mathrm{p(ab)} \leftarrow}{\mathrm{p(aabb)} \leftarrow}}{\mathrm{p(aaabbb)} \leftarrow}$$

Example of Proof (5)

S: p(axb)←p(x)

p(xy) ← p(x), p(y)

p(ab)←

$$\dfrac{\text{p}(\text{a}x\text{b})\leftarrow\text{p}(x)}{\text{p}(\text{aabb})\leftarrow\text{p}(\text{ab})}$$

$$\dfrac{\text{p}(\text{aabb})\leftarrow\text{p}(\text{ab}) \qquad \text{p}(\text{ab})\leftarrow}{\text{p}(\text{aabb})\leftarrow}$$

$$\dfrac{\text{p}(xy)\leftarrow \text{p}(x),\ \text{p}(y)}{\text{p}(\text{aabb}y)\leftarrow\text{p}(\text{aabb}),\text{p}(y)}$$

$$\dfrac{\text{p}(\text{aabb}y)\leftarrow\text{p}(\text{aabb}),\text{p}(y) \qquad \text{p}(\text{aabb})\leftarrow}{\text{p}(\text{aabb}y)\leftarrow \text{p}(y)}$$

$$\dfrac{\text{p}(\text{aabb}y)\leftarrow \text{p}(y)}{\text{p}(\text{aabbab})\leftarrow \text{p}(\text{ab})}$$

$$\dfrac{\text{p}(\text{aabbab})\leftarrow \text{p}(\text{ab}) \qquad \text{p}(\text{ab})\leftarrow}{\text{p}(\text{aabbab})\leftarrow}$$

Defining a language by proofs

- A ground atomic formula $p(s_1,\ldots,s_n)$ is provable from an EFS S if

  there is a proof which derives $p(s_1,\ldots,s_n)$ and S.

- We define a language with a proof from an EFS.

  $$L(p, S) = \{ s \mid p(s) \text{ is provable from } S \}$$

Example

$$S: \mathrm{p}(\mathrm{a}x\mathrm{b}) \leftarrow \mathrm{p}(x)$$

$$\mathrm{p}(\mathrm{ab}) \leftarrow$$

$$L(\mathrm{p}, S) = \{\mathrm{ab}, \mathrm{aabb}, \mathrm{aaabbb}, \mathrm{aaabbb}, \ldots\}$$

Refinement Operator for EFS's

Learning EFS

- Fix an effective enumeration of EFS on $\Sigma \cup X$:

  $S_1, S_2, \ldots,$

$k = 1, S = S_1$

for $n = 1$ forever

  receive $e_n = \langle s_n, b_n \rangle$

  while ($0 \leq \exists j \leq n$

    ($e_j = \langle s_j, + \rangle$ and $s_j \notin L(S)$) and

    ($e_j = \langle s_j, - \rangle$ and $s_j \in L(S)$)

      $S = S'$ for an appropriate S'; k ++

  output S

Enumerating EFS

- A simple method to enumerate EFS.
- We define the size of an EFS S as the total number of symbols in S but except "← ", " (", ")"and " , ".

Example size({p(axb)←p(x), p(ab)←})= 9

Enumeration of EFS

$\Sigma = \{\text{a}, \text{b}\}$

size(S)

2 $\{\ \text{p(a)} \leftarrow\ \}, \{\ \text{p(b)} \leftarrow\ \}, \{\ \text{p}(x) \leftarrow\ \}$

3 $\{\ \text{p(aa)} \leftarrow\ \}, \{\ \text{p(ab)} \leftarrow\ \}, \{\ \text{p(ba)} \leftarrow\ \}, \{\ \text{p(bb)} \leftarrow\ \},$

$\{\ \text{p}(xy) \leftarrow\ \}, \{\ \text{p}(xx) \leftarrow\ \}, \{\ \text{p(a}x) \leftarrow\ \}, \{\ \text{p(b}x) \leftarrow\ \},$

$\{\ \text{p}(x\text{a}) \leftarrow\ \}, \{\ \text{p}(x\text{b}) \leftarrow\ \},$

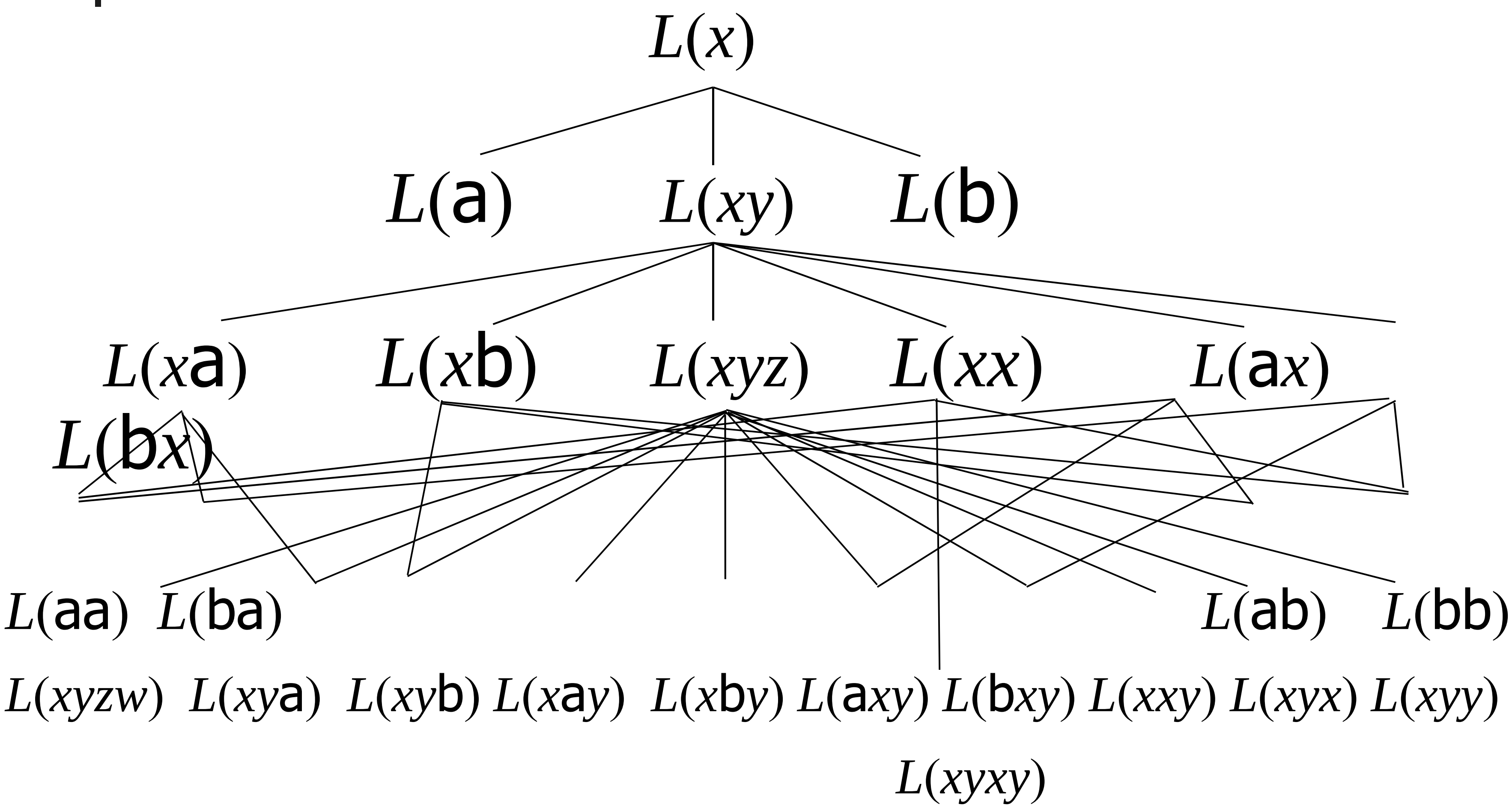
4 $\{\text{p(a)} \leftarrow \text{p(a)}\ \}, \ldots, \{\text{p}(x) \leftarrow \text{p}(x)\},$

$\{\text{p(aaa)} \leftarrow\ \}, \ldots,$

$\{\text{p(a)} \leftarrow, \text{p(b)} \leftarrow\}, \ldots, \{\text{p(b)} \leftarrow, \text{p}(x) \leftarrow\}$

…

Hasse Diagram (General Version)



Refinement Operator

- A refinement operator ρ defines, from a given rule set g, set of grammar satisfying:

  1. $\rho(g)$ is recursively enumerable,

  2. for all $h \in \rho(g)$ $L(h) \subseteq L(g)$, and

  3. ere is no sequence $g_1, g_2, \ldots, g_n$ of grammars such that $g_{i+1} = \rho(g_i)$ and $g_1 = g_n$.

  - The refinement operator works as the operator $\rho = f(x) \rightarrow f(x + \Delta x)$ $(= f'(x)\,\Delta x + f(x))$ for a usual mathematical function f .

Refinement of Patterns

- For patterns on Σ.

  $\sigma_x = \{ x := x\, y \}$ where y is a fresh variable

  $\theta_{xc} = \{ x := c \}$ where c is in Σ

  $\delta_{xy} = \{ x := z, y := x \}$ where z is a fresh variable

$$\rho(\pi) = \{ \pi\sigma_x \mid x \text{ is a variable occurring in } \pi \} \cup \{ \pi\theta_{xc} \mid x \text{ is a variable occurring in } \pi \text{ and } c \text{ is in } \Sigma \} \cup \{ \pi\delta_{xy} \mid x \text{ and } y \text{ are variables occurring in } \pi \}$$

Generating Patterns with Refinement

- Let C be a set of positive examples and D be a set of negative examples.
- Assume the set of variables $X = \{x_1, x_2, \ldots, x_n, \ldots\}$

Let $P := \{ x_1 \}$, $Q := \varnothing$
/* P is for keeping candidates, and Q is for minimal candidates.*/

while $P \neq \varnothing$ **do**
  choose π from P
  $P' := \varnothing$
  for each $\pi' \in \rho(\pi)$
    if $C \subseteq L(\pi')$ and $L(\pi') \cap D = \varnothing$
      $P' := P' \cup \{\pi'\}$
  if $P' = \varnothing$
    $Q := Q \cup \{\pi\}$
  else
    $P := P - \{\pi\} \cup P'$

Refinement for EFS's

- Because an EFS is a set of definite clauses, we define the refinement operator for EFSs by
  - defining the refinement of operator of definite clause
  - and then defining the refinement operator of the set of definite clauses.

Refinement for Definite Clauses

- For a definite clause $C = A \leftarrow B_1,\ldots,B_n$

  $\sigma_x = \{ x := x\,y \}$ where y is a fresh variable

  $\theta_{xc} = \{ x := c \}$ where c is in Σ

  $\delta_{xy} = \{ x := z, y := x \}$ where z is a fresh variable

$\rho(C) =$ $\{C\sigma_x \mid x$ is a variable occurring in $C \}$

$\cup$ $\{C\theta_{xc} \mid x$ is a variable occurring in C and c is in $\Sigma \}$

$\cup$ $\{C\delta_{xy} \mid x$ and y are variables occurring in C

}

$\cup$ $\{ A \leftarrow B_1,\ldots,B_n, p(x_1,\ldots,x_k) \mid$ where $x_1,\ldots,x_k$ are mutually distinct variables occuring in $A\}$

Refinement for Definite Clauses

- For a set S of definite clauses

$$\rho(S) = \{S \cup \{D\} \mid D \in \rho(C) \text{ for some } C \in S\} \cup \{S - \{C\} \mid C \in S\}$$

- The top element is a set of clauses of the form

$$\mathrm{T}: \left\{ \begin{array}{l} p_1(x_1,\ldots,x_{n1}) \leftarrow \\ p_2(x_1,\ldots,x_{n2}) \leftarrow \\ p_3(x_1,\ldots,x_{n3}) \leftarrow \\ \ldots \end{array} \right\}$$

- $\rho^n(P)$: The set of EFS which can be obtained by applying ρ repeatedly at most n times.

A Successful Case

- If we give some restrictions to EFS S, we can simply extend the learning algorithm for patterns.
- An example of such a restriction is:

  The number of definite clauses in S is bounded up to a given N and every clause is of the form

  $$p(\pi_1,\ldots,\pi_n) \leftarrow q_1(x_1), q_2(x_2),\ldots, q_k(x_k)$$

  where $x_1, x_2,\ldots, x_k$ appears in $\pi_1,\ldots,\pi_n$.

- The latter condition is just saying that S corresponds to a CFG.

A Key Property of Refinement

- For EFS S and T,

$$T \in \rho(S) \Rightarrow L(S) \supseteq L(T)$$

- The definition of $\rho(S)$ is rather mathematical, and a more practical method for finding hypotheses can be formalized with not using $\rho(S)$ but $\rho(C)$.
  - Starting with T, if a definite clause C generates any negative example, replace C with all of the clauses in $\rho(C)$.

Learning EFS

$S = \top$

for $n = 1$ forever

receive $e_n = \langle s_n , b_n \rangle$

while ($0 \leq \exists j \leq n \; e_j = \langle s_j , - \rangle$ and $s_j \in L(S)$)

delete a clause C in S and add all clauses

in $\rho(C)$

output S

Example

S : $\mathrm{p}(\mathsf{a}x\mathsf{b}) \leftarrow \mathrm{p}(y)$

$\mathrm{p}(\mathsf{ab}) \leftarrow$

$E_1 = \langle \mathrm{p}(\mathsf{aabb}), +\rangle$, $E_2 = \langle \mathrm{p}(\mathsf{ab}), +\rangle$

$E_3 = \langle \mathrm{p}(\mathsf{abb}), -\rangle$

Example

$E_1 = \langle \mathrm{p(aabb)}, + \rangle$, $E_2 = \langle \mathrm{p(ab)}, + \rangle$

$E_3 = \langle \mathrm{p(bba)}, - \rangle$

$S =$ $\cancel{\mathrm{p}(x) \leftarrow}$

$\mathrm{p}(xy) \leftarrow$, $\mathrm{p(a)} \leftarrow$, $\mathrm{p(b)} \leftarrow$, $\mathrm{p}(x) \leftarrow \mathrm{p}(x)$

Example

$E_1 = \langle p(aabb), +\rangle$, $E_2 = \langle p(ab), +\rangle$

$E_3 = \langle p(bba), -\rangle$

$S =$ $\cancel{p(x) \leftarrow}$

$\cancel{p(xy) \leftarrow}$, $p(a)\leftarrow$, $p(b)\leftarrow$, $p(x) \leftarrow p(x)$

$p(xyz)\leftarrow$, $p(ay)\leftarrow$, $p(by)\leftarrow$, $p(xa)\leftarrow$, $p(xb)\leftarrow$

$p(xy)\leftarrow p(x)$, $p(xy)\leftarrow p(y)$

Example

$E_1 = \langle \mathrm{p(aabb)}, + \rangle$, $E_2 = \langle \mathrm{p(ab)}, + \rangle$

$E_3 = \langle \mathrm{p(bba)}, - \rangle$

$S =$ $\cancel{\mathrm{p}(x) \leftarrow}$

$\cancel{\mathrm{p}(xy) \leftarrow}$, $\mathrm{p(a)} \leftarrow$, $\mathrm{p(b)} \leftarrow$, $\mathrm{p}(x) \leftarrow \mathrm{p}(x)$

$\cancel{\mathrm{p}(xyz) \leftarrow}$, $\mathrm{p(a}y) \leftarrow$, $\mathrm{p(b}y) \leftarrow$, $\mathrm{p}(x\mathrm{a}) \leftarrow$, $\mathrm{p}(x\mathrm{b}) \leftarrow$

$\mathrm{p}(xy) \leftarrow \mathrm{p}(x)$, $\mathrm{p}(xy) \leftarrow \mathrm{p}(y)$

$\mathrm{p}(xxz) \leftarrow$, $\mathrm{p}(xyx) \leftarrow$, $\mathrm{p}(xyy) \leftarrow$, $\mathrm{p}(xyz) \leftarrow \mathrm{p}(x)$,

$\mathrm{p}(xyz) \leftarrow \mathrm{p}(y)$, $\mathrm{p}(xyz) \leftarrow \mathrm{p}(z)$,

$\mathrm{p(a}yz) \leftarrow$, $\mathrm{p(b}yz) \leftarrow$, …, $\mathrm{p}(xy\mathrm{a}) \leftarrow$, $\mathrm{p}(xy\mathrm{b}) \leftarrow$,

Example

$E_1 = \langle \mathrm{p(aabb)}, + \rangle$, $E_2 = \langle \mathrm{p(ab)}, + \rangle$

$E_3 = \langle \mathrm{p(bba)}, - \rangle$

$S =$ $\cancel{\mathrm{p}(x) \leftarrow}$

$\cancel{\mathrm{p}(xy) \leftarrow}$, $\mathrm{p(a)} \leftarrow$, $\mathrm{p(b)} \leftarrow$, $\mathrm{p}(x) \leftarrow \mathrm{p}(x)$

$\cancel{\mathrm{p}(xyz) \leftarrow}$, $\mathrm{p(a}y) \leftarrow$, $\cancel{\mathrm{p(b}y) \leftarrow}$, $\mathrm{p}(x\mathrm{a}) \leftarrow$, $\mathrm{p}(x\mathrm{b}) \leftarrow$

$\mathrm{p}(xy) \leftarrow \mathrm{p}(x)$, $\mathrm{p}(xy) \leftarrow \mathrm{p}(y)$

$\mathrm{p}(xxz) \leftarrow$, $\mathrm{p}(xyx) \leftarrow$, $\mathrm{p}(xyy) \leftarrow$, $\mathrm{p}(xyz) \leftarrow \mathrm{p}(x)$,

$\mathrm{p}(xyz) \leftarrow \mathrm{p}(y)$, $\mathrm{p}(xyz) \leftarrow \mathrm{p}(z)$,

$\mathrm{p(a}yz) \leftarrow$, $\mathrm{p(b}yz) \leftarrow$, …, $\mathrm{p}(xy\mathrm{a}) \leftarrow$, $\mathrm{p}(xy\mathrm{b}) \leftarrow$,

$\mathrm{p(ba)} \leftarrow$, $\mathrm{p(bb)} \leftarrow$, $\cancel{\mathrm{p(b}xy) \leftarrow}$, $\mathrm{p(b}y) \leftarrow \mathrm{p}(y)$,

Example

$E_1 = \langle \mathrm{p(aabb)}, + \rangle$, $E_2 = \langle \mathrm{p(ab)}, + \rangle$

$E_3 = \langle \mathrm{p(bba)}, - \rangle$

$S =$ $\cancel{\mathrm{p}(x) \leftarrow}$

$\cancel{\mathrm{p}(xy)\leftarrow}$, $\mathrm{p(a)}\leftarrow$, $\mathrm{p(b)}\leftarrow$, $\mathrm{p}(x) \leftarrow \mathrm{p}(x)$

$\cancel{\mathrm{p}(xyz)\leftarrow}$, $\cancel{\mathrm{p(a}y)\leftarrow}$, $\mathrm{p(b}y)\leftarrow$, $\cancel{\mathrm{p}(x\mathrm{a})\leftarrow}$, $\mathrm{p}(x\mathrm{b})\leftarrow$

$\mathrm{p}(xy)\leftarrow\mathrm{p}(x)$, $\mathrm{p}(xy)\leftarrow\mathrm{p}(y)$

$\mathrm{p}(xxz)\leftarrow$, $\mathrm{p}(xyx)\leftarrow$, $\mathrm{p}(xyy)\leftarrow$, $\mathrm{p}(xyz)\leftarrow\mathrm{p}(x)$,

$\mathrm{p}(xyz)\leftarrow\mathrm{p}(y)$, $\mathrm{p}(xyz)\leftarrow\mathrm{p}(z)$,

$\mathrm{p(a}yz)\leftarrow$, $\mathrm{p(b}yz)\leftarrow$, …, $\mathrm{p}(xy\mathrm{a})\leftarrow$, $\mathrm{p}(xy\mathrm{b})\leftarrow$,

$\mathrm{p(aa)}\leftarrow$, $\mathrm{p(ab)}\leftarrow$, $\cancel{\mathrm{p(a}xy)\leftarrow}$, $\mathrm{p(a}y)\leftarrow\mathrm{p}(y)$,

$\mathrm{p(ba)}\leftarrow$, $\cancel{\mathrm{p(ab)}\leftarrow}$, $\cancel{\mathrm{p}(xy\mathrm{a})\leftarrow}$, $\mathrm{p}(x\mathrm{a})\leftarrow\mathrm{p}(x)$

Example

$E_1 = \langle \mathrm{p(aabb)}, + \rangle$, $E_2 = \langle \mathrm{p(ab)}, + \rangle$

$E_3 = \langle \mathrm{p(bba)}, - \rangle$

$S =$ $\mathrm{p(a)} \leftarrow$, $\mathrm{p(b)} \leftarrow$, $\mathrm{p}(x) \leftarrow \mathrm{p}(x)$

$\mathrm{p(b}y) \leftarrow$, $\mathrm{p}(x\mathrm{a}) \leftarrow$, $\mathrm{p}(xy) \leftarrow \mathrm{p}(x)$, $\mathrm{p}(xy) \leftarrow \mathrm{p}(y)$

$\mathrm{p}(xxz) \leftarrow$, $\mathrm{p}(xyx) \leftarrow$, $\mathrm{p}(xyy) \leftarrow$, $\mathrm{p}(xyz) \leftarrow \mathrm{p}(x)$,

$\mathrm{p}(xyz) \leftarrow \mathrm{p}(y)$, $\mathrm{p}(xyz) \leftarrow \mathrm{p}(z)$,

$\mathrm{p(a}yz) \leftarrow$, $\mathrm{p(b}yz) \leftarrow$, …, $\mathrm{p}(xy\mathrm{a}) \leftarrow$, $\mathrm{p}(xy\mathrm{b}) \leftarrow$,

$\mathrm{p(aa)} \leftarrow$, $\mathrm{p(ab)} \leftarrow$, $\mathrm{p(a}y) \leftarrow \mathrm{p}(y)$,

$\mathrm{p(bb)} \leftarrow$, $\mathrm{p}(xy\mathrm{b}) \leftarrow$, $\mathrm{p}(x\mathrm{b}) \leftarrow \mathrm{p}(x)$

Example

$E_1 = \langle p(\mathrm{aabb}), +\rangle$, $E_2 = \langle p(\mathrm{ab}), +\rangle$

$E_3 = \langle p(\mathrm{bba}), -\rangle$

$S =$ $p(\mathrm{a})\leftarrow$, $p(\mathrm{b})\leftarrow$, $p(x) \leftarrow p(x)$

$p(\mathrm{b}y)\leftarrow$, $p(x\mathrm{a})\leftarrow$, $p(xy)\leftarrow p(x)$, $p(xy)\leftarrow p(y)$

$p(xyx)\leftarrow$, $p(xyy)\leftarrow$, $p(xyz)\leftarrow p(x)$,

$p(xyz)\leftarrow p(y)$, $p(xyz)\leftarrow p(z)$,

$p(\mathrm{a}yz)\leftarrow$, $p(\mathrm{b}yz)\leftarrow$, …, $p(xy\mathrm{a})\leftarrow$, $p(xy\mathrm{b})\leftarrow$,

$p(\mathrm{aa})\leftarrow$, $p(\mathrm{ab})\leftarrow$, $p(\mathrm{a}y)\leftarrow p(y)$,

$p(\mathrm{bb})\leftarrow$, $p(xy\mathrm{b})\leftarrow$, $p(x\mathrm{b})\leftarrow p(x)$

$p(\mathrm{a}yy)\leftarrow$, $p(\mathrm{b}yy)\leftarrow$, $p(\mathrm{aa}z)\leftarrow$, $p(\mathrm{bb}z)\leftarrow$, $p(xyxyz)\leftarrow$,

$p(xxz)\leftarrow p(x)$, $p(xxz)\leftarrow p(z)$,

Example

$E_1 = \langle \mathrm{p(aabb)}, + \rangle$, $E_2 = \langle \mathrm{p(ab)}, + \rangle$

$E_3 = \langle \mathrm{p(bba)}, - \rangle$

$S =$ $\mathrm{p(a)} \leftarrow$, $\mathrm{p(b)} \leftarrow$, $\mathrm{p}(x) \leftarrow \mathrm{p}(x)$

$\mathrm{p(b}y) \leftarrow$, $\mathrm{p}(x\mathrm{a}) \leftarrow$, $\mathrm{p}(xy) \leftarrow \mathrm{p}(x)$, $\mathrm{p}(xy) \leftarrow \mathrm{p}(y)$

$\mathrm{p}(xyx) \leftarrow$, $\mathrm{p}(xyy) \leftarrow$, $\mathrm{p}(xyz) \leftarrow \mathrm{p}(x)$,

$\mathrm{p}(xyz) \leftarrow \mathrm{p}(y)$, $\mathrm{p}(xyz) \leftarrow \mathrm{p}(z)$,

$\mathrm{p(a}yz) \leftarrow$, $\mathrm{p(b}yz) \leftarrow$, …, $\mathrm{p}(xy\mathrm{a}) \leftarrow$, $\mathrm{p}(xy\mathrm{b}) \leftarrow$,

$\mathrm{p(aa)} \leftarrow$, $\mathrm{p(ab)} \leftarrow$, $\mathrm{p(a}y) \leftarrow \mathrm{p}(y)$,

$\mathrm{p(bb)} \leftarrow$, $\mathrm{p}(xy\mathrm{b}) \leftarrow$, $\mathrm{p}(x\mathrm{b}) \leftarrow \mathrm{p}(x)$

$\mathrm{p(a}yy) \leftarrow$, $\mathrm{p(b}yy) \leftarrow$, $\mathrm{p(aa}z) \leftarrow$, $\mathrm{p(bb}z) \leftarrow$, $\mathrm{p}(xyxyz) \leftarrow$,

$\mathrm{p}(xxz) \leftarrow \mathrm{p}(x)$, $\mathrm{p}(xxz) \leftarrow \mathrm{p}(z)$,

Example

$E_1 = \langle p(\text{aabb}), +\rangle$, $E_2 = \langle p(\text{ab}), +\rangle$

$E_3 = \langle p(\text{bba}), -\rangle$

$S =$ …

$p(\text{a}y\text{b}) \leftarrow p(y),$

$p(\text{ab}) \leftarrow ,$

…

Context Free Grammar

- A context free grammar is defined as $G = (N, \Sigma, P, S)$ where

  N is a finite set of non-terminal symbols (non-terminals, syntactic categories).

  Σ is a finite set of characters or terminals.

  P is a set of rules (productions).

  $S \in N$ is an initial state.

- A production is of the form $A \to \gamma$ where A is a non-terminal and γ is a sequence of terminals and non-terminals.
  - Note: In general definition γ can be an empty sequence ε.

    In this course we do not allow γ to be ε.

EFS and Formal Grammar

Context Free Grammar

- A context free grammar is defined as $G = (N, \Sigma, P, S)$ where

  N is a finite set of non-terminal symbols (non-terminals, syntactic categories).

  Σ is a finite set of characters or terminals.

  P is a set of rules (productions).

  $S \in N$ is an initial state.
- A production is of the form $A \rightarrow \gamma$ where A is a non-terminal and γ is a sequence of terminals and non-terminals.
  - Note: In general definition γ can be an empty sequence ε. In this course we do not allow γ to be ε.

Examples

$$G_1 = (N = \{S\}, \Sigma = \{\text{a, b}\}, P = \{S \to \text{ab},\ S \to \text{a}S\text{b}\}, S)$$

$$G_2 = (N = \{S, A, B\}, \Sigma = \{\text{a, b}\},$$
$$P = \{S \to \text{a}AB,\ A \to \text{b}BB, B \to \text{abb}\}, S)$$

$$G_3 = (N = \{S, A, B, C,\}, \Sigma = \{\text{a, b, c}\},$$
$$P = \{S \to \text{a}A,\ S \to \text{b}B, A \to \text{a}A, A \to \text{b}C,$$
$$A \to \text{b}$$
$$B \to \text{a}B, B \to \text{b}B, C \to \text{a}A, C \to \text{b}C\ ,$$
$$C \to \text{b}\}, S)$$

Examples(cont.)

$$G_5 = (N = \{S\}, \Sigma = \{\mathrm{a},\ \mathrm{b},\ \mathrm{c},\ +,\ *,\ (,\ )\},$$
$$P = \{S \rightarrow (S + S),\ S \rightarrow (S * S),$$
$$S \rightarrow \mathrm{a}, S \rightarrow \mathrm{b}, S \rightarrow \mathrm{c}\}, S)$$

Derivations

- An application of a production rule $A \to \gamma \in P$ is written as $\alpha\, A\, \beta \Rightarrow \alpha\, \gamma\, \beta$ where $\alpha, \beta \in (N \cup \Sigma)^*$
  - This means that only one occurrence of A in a string is replaced with γ.

$$S * (S + S) \Rightarrow V * (S + S)$$

$$S * (S + S) \Rightarrow S * (V + S)$$

$$S * (S + S) \not\Rightarrow V * (V + V)$$

- A derivation from α to β is a finite sequence of applications starting with α and ending with β:

$$\alpha = \gamma_1 \Rightarrow \gamma_2 \Rightarrow \ldots \Rightarrow \gamma_n = \beta$$

We write $\alpha \Rightarrow^* \beta$ if $\alpha = \beta$ or there is a derivation from α to β.

Context Free Languages

- The language defined a context free grammar $G = (N, \Sigma, P, S)$ is $L(G) = \{ w \in \Sigma^* \mid \text{there is a derivation } S \Rightarrow^* w \}$.

- A language L is context free if there is a context free grammar G such that $L = L(G)$.

Example For $G = (\{S\}, \{\mathsf{a}, \mathsf{b}\}, \{S \to \mathsf{ab},\ S \to \mathsf{a}S\mathsf{b}\}, S)$,
$L(G) = \{ \mathsf{a}^n\mathsf{b}^n \mid n \geq 1 \}$.

Example 1

- $\Sigma = \{\text{a, b}\}$
- $L = \{\ \text{a}^n\text{b}^n \mid n \geq 1\ \} = \{\ \underbrace{\text{a...a}}_{n \text{ times}}\underbrace{\text{b...b}}_{n \text{ times}} \mid n \geq 1\ \}$

  $= \{\text{ab, aabb, aaabbb, aaaabbbb,...}\ \}$
- The language is defined with a set of productions:

  $S \to \text{ab},\ S \to \text{a}S\text{b}$
- Some examples of derivations:

  $S \Rightarrow \text{ab}$

  $S \Rightarrow \text{a}S\text{b} \Rightarrow \text{aabb}$

  $S \Rightarrow \text{a}S\text{b} \Rightarrow \text{aa}S\text{bb} \Rightarrow \text{aaabbb}$

  $S \Rightarrow \text{a}S\text{b} \Rightarrow \text{aa}S\text{bb} \Rightarrow \text{aaa}S\text{bbb} \Rightarrow \text{aaaabbbb}$
- It is easy to show that there is no FA which accepts L.

Example 2

- Mathematical formulae consisting of x, y, +, *, (, and)

  correct: $x + y$, $y * (x + y)$, $((x * x) + x)$, x, …

  incorrect: $x +$, $(y * x + y$, $()$, …

  Note: We assume strict use of (and).

  correct: $(x + y)$, (x)

  incorrect: $x * x * x$, $y * x + x$

- The set are defined by a set of rules

$$S \to V \qquad V \to x$$

$$S \to S + S \qquad V \to y$$

$$S \to S * S$$

$$S \to (S)$$

Example 2 (cont.)

- An example of a derivation:

$$S \to V \qquad S \to S * S \qquad V \to x$$
$$S \to S + S \qquad S \to (S) \qquad V \to y$$

$$\begin{aligned} S &\Rightarrow S * S \Rightarrow S * (S) \Rightarrow S * (S + S) \\ &\Rightarrow V * (S + S) \Rightarrow V * (V + V) \\ &\Rightarrow y * (V + V) \Rightarrow y * (x + V) \Rightarrow y * (x + y) \end{aligned}$$

Transforming CFG to EFS

- For every production rule

  $$P \rightarrow w_1 Q_1 w_2 Q_2 \ldots Q_n w_{n+1} \qquad P, Q \in N,\ w_i \in \Sigma^*$$

  we define a definite clause

  $$\mathrm{p}(w_1 x_1 w_2 x_2 \ldots x_n w_{n+1}) \leftarrow \mathrm{q}_1(x_1), \mathrm{q}_2(x_2), \ldots, \mathrm{q}_n(x_n)$$

Example

$P \rightarrow \mathrm{a}P\mathrm{b}$ ⇨ $\mathrm{p}(\mathrm{a}x\mathrm{b}) \leftarrow \mathrm{p}(x)$

$P \rightarrow \mathrm{ab}$ ⇨ $\mathrm{p}(\mathrm{ab}) \leftarrow$

Derivation and Proof

$$P \rightarrow \mathrm{a}P\mathrm{b}$$

$$P \rightarrow \mathrm{ab}$$

$$P \Rightarrow \mathrm{a}P\mathrm{b} \Rightarrow \mathrm{aa}P\mathrm{bb} \Rightarrow \mathrm{aaabbb}$$

$$\mathrm{p}(\mathrm{a}x\mathrm{b}) \leftarrow \mathrm{p}(x)$$

$$\mathrm{p}(\mathrm{ab}) \leftarrow$$

$$\frac{\dfrac{\mathrm{p}(\mathrm{a}x\mathrm{b}) \leftarrow \mathrm{p}(x)}{\mathrm{p}(\mathrm{aaabbb}) \leftarrow \mathrm{p}(\mathrm{aabb})} \qquad \dfrac{\dfrac{\mathrm{p}(\mathrm{a}x\mathrm{b}) \leftarrow \mathrm{p}(x)}{\mathrm{p}(\mathrm{aabb}) \leftarrow \mathrm{p}(\mathrm{ab})} \qquad \mathrm{p}(\mathrm{ab}) \leftarrow}{\mathrm{p}(\mathrm{aabb}) \leftarrow}}{\mathrm{p}(\mathrm{aaabbb}) \leftarrow}$$

Notes

- There is a class of EFS which corresponds to the context sensitive grammar.
- There is a class of EFS which corresponds to TM.
- It is not easy to check which class $L(p, S)$ belongs to

$$\mathrm{p}(xx) \leftarrow \mathrm{p}(x)$$

$$\mathrm{p}(\mathrm{a}) \leftarrow$$