

# Automata on Finite Words

## Definition

A *nondeterministic finite automaton* (NFA) over  $\Sigma$  is a 4-tuple

$A = \langle S, I, T, F \rangle$ , where:

- $S$  is a finite set of *states*,
- $I \subseteq S$  is a set of *initial states*,
- $T \subseteq S \times \Sigma \times S$  is a *transition relation*,
- $F \subseteq S$  is a set of *final states*.

We denote  $T(s, \alpha) = \{s' \in S \mid (s, \alpha, s') \in T\}$ . When  $T$  is clear from the context we denote  $(s, \alpha, s') \in T$  by  $s \xrightarrow{\alpha} s'$ .

## Runs and Acceptance Conditions

Given a finite word  $w \in \Sigma^*$ ,  $w = \alpha_1\alpha_2 \dots \alpha_n$ , a *run* of  $A$  over  $w$  is a finite sequence of states  $s_1, s_2, \dots, s_n, s_{n+1}$  such that  $s_1 \in I$  and  $s_i \xrightarrow{\alpha_i} s_{i+1}$  for all  $1 \leq i \leq n$ .

A run over  $w$  between  $s_i$  and  $s_j$  is denoted as  $s_i \xrightarrow{w} s_j$ .

The run is said to be *accepting* if and only if  $s_{n+1} \in F$ . If  $A$  has an accepting run over  $w$ , then we say that  $A$  *accepts*  $w$ .

The language of  $A$ , denoted  $\mathcal{L}(A)$  is the set of all words accepted by  $A$ .

A set of words  $S \subseteq \Sigma^*$  is *recognizable* if there exists an automaton  $A$  such that  $S = \mathcal{L}(A)$ .

## Determinism and Completeness

**Definition 1** An automaton  $A = \langle S, I, T, F \rangle$  is **deterministic** (DFA) if and only if  $\|I\| \leq 1$  and, for each  $s \in S$  and for each  $\alpha \in \Sigma$ ,  $\|T(s, \alpha)\| \leq 1$ .

If  $A$  is deterministic we write  $T(s, \alpha) = s'$  instead of  $T(s, \alpha) = \{s'\}$ .

**Definition 2** An automaton  $A = \langle S, I, T, F \rangle$  is **complete** if and only if  $\|I\| \geq 1$  and, for each  $s \in S$  and for each  $\alpha \in \Sigma$ ,  $\|T(s, \alpha)\| \geq 1$ .

## Determinism and Completeness

**Proposition 1** *If  $A$  is deterministic, then it has **at most one run** for each input word.*

**Proposition 2** *If  $A$  is complete, then it has **at least one run** for each input word.*

## Determinization

**Theorem 1** *For every NFA  $A$  there exists a DFA  $A_d$  such that  $\mathcal{L}(A) = \mathcal{L}(A_d)$ .*

Let  $A_d = \langle 2^S, \{I\}, T_d, \{G \subseteq S \mid G \cap F \neq \emptyset\} \rangle$ , where

$$(S_1, \alpha, S_2) \in T_d \iff S_2 = \{s' \mid \exists s \in S_1 . (s, \alpha, s') \in T\}$$

This definition is known as **subset construction**.

**Exercise 1** *Let  $\Sigma = \{a, b\}$  and  $L_n = \{uav \mid u, v \in \Sigma^*, |v| = n - 1\}$ , for each integer  $n \geq 1$ . Build an NFA that recognizes  $L_n$  and apply subset construction to it.*

## Completion

**Lemma 1** *For every NFA  $A$  there exists a complete NFA  $A_c$  such that  $\mathcal{L}(A) = \mathcal{L}(A_c)$ .*

Let  $A_c = \langle S \cup \{\sigma\}, I, T_c, F \rangle$ , where  $\sigma \notin S$  is a new **sink state**. The transition relation  $T_c$  is defined as:

$$\forall s \in S \forall \alpha \in \Sigma . (s, \alpha, \sigma) \in T_c \iff \forall s' \in S . (s, \alpha, s') \notin T$$

and  $\forall \alpha \in \Sigma . (\sigma, \alpha, \sigma) \in T_c$ .

**Remark:** The subset construction yields a complete deterministic automaton, with sink state  $\emptyset$ .

## Closure Properties

**Theorem 2** Let  $A_1 = \langle S_1, I_1, T_1, F_1 \rangle$  and  $A_2 = \langle S_2, I_2, T_2, F_2 \rangle$  be two NFA, such that  $S_1 \cap S_2 = \emptyset$ . There exists automata  $\bar{A}_1$ ,  $A_{\cup}$  and  $A_{\cap}$  that recognize the languages  $\Sigma^* \setminus \mathcal{L}(A_1)$ ,  $\mathcal{L}(A_1) \cup \mathcal{L}(A_2)$ , and  $\mathcal{L}(A_1) \cap \mathcal{L}(A_2)$ , respectively.

Let  $A' = \langle S', I', T', F' \rangle$  be the **complete** and **deterministic** (why?) automaton such that  $\mathcal{L}(A_1) = \mathcal{L}(A')$ , and  $\bar{A}_1 = \langle S', I', T', S' \setminus F' \rangle$ .

Let  $A_{\cup} = \langle S_1 \cup S_2, I_1 \cup I_2, T_1 \cup T_2, F_1 \cup F_2 \rangle$ .

Let  $A_{\cap} = \langle S_1 \times S_2, I_1 \times I_2, T_{\cap}, F_1 \times F_2 \rangle$  where:

$$(\langle s_1, t_1 \rangle, \alpha, \langle s_2, t_2 \rangle) \in T_{\cap} \iff (s_1, \alpha, s_2) \in T_1 \text{ and } (t_1, \alpha, t_2) \in T_2$$

## On the Exponential Blowup of Complementation

**Theorem 3** *For every  $n \in \mathbb{N}$ ,  $n \geq 1$ , there exists an automaton  $A$ , with  $\text{size}(A) = n + 1$  such that no deterministic automaton with less than  $2^n$  states recognizes the complement of  $\mathcal{L}(A)$ .*

Let  $\Sigma = \{a, b\}$  and  $L_n = \{uav \mid u, v \in \Sigma^*, |v| = n - 1\}$ , for all  $n \geq 1$ .

There exists a NFA with exactly  $n + 1$  states which recognizes  $L_n$ .

Suppose that  $B = \langle S, \{s_0\}, T, F \rangle$ , is a (complete) DFA with  $\|S\| < 2^n$  that accepts  $\Sigma^* \setminus L_n$ .

## On the Exponential Blowup of Complementation

$\|\{w \in \Sigma^* \mid |w| = n\}\| = 2^n$  and  $\|S\| < 2^n$  (by the pigeonhole principle)

$\Rightarrow \exists uav_1, ubv_2 . |uav_1| = |ubv_2| = n$  and  $s \in S . s_0 \xrightarrow{uav_1} s$  and  $s_0 \xrightarrow{ubv_2} s$

Let  $s_1$  be the (unique) state of  $B$  such that  $s \xrightarrow{u} s_1$ .

Since  $|uav_1| = n$ , then  $uav_1u \in L_n \Rightarrow uav_1u \notin \mathcal{L}(B)$ , i.e.  $s$  is not accepting.

On the other hand,  $ubv_2u \notin L_n \Rightarrow ubv_2u \in \mathcal{L}(B)$ , i.e.  $s$  is accepting,  
**contradiction.**

## Projections

Let the input alphabet be  $\Sigma = \Sigma_1 \times \Sigma_2$ . Any word  $w \in \Sigma^*$  can be uniquely identified to a pair  $\langle w_1, w_2 \rangle \in \Sigma_1^* \times \Sigma_2^*$  such that  $|w_1| = |w_2| = |w|$ .

The *projection* operations are

$pr_1(L) = \{v \in \Sigma_2^* \mid \langle u, v \rangle \in L, \text{ for some } u \in \Sigma_1^*\}$  and

$pr_2(L) = \{u \in \Sigma_1^* \mid \langle u, v \rangle \in L, \text{ for some } v \in \Sigma_2^*\}$ .

**Theorem 4** *If the language  $L \subseteq (\Sigma_1 \times \Sigma_2)^*$  is recognizable, then so are the projections  $pr_i(L)$ , for  $i = 1, 2$ .*

## Remark

The operations of union, intersection and complement correspond to the boolean  $\vee$ ,  $\wedge$  and  $\neg$ .

The projection corresponds to the first-order existential quantifier  $\exists x$ .

## The Myhill-Nerode Theorem

Let  $A = \langle S, I, T, F \rangle$  be an automaton over the alphabet  $\Sigma^*$ .

Define the relation  $\sim_A \subseteq \Sigma^* \times \Sigma^*$  as:

$$u \sim_A v \iff [\forall s, s' \in S . s \xrightarrow{u} s' \iff s \xrightarrow{v} s']$$

$\sim_A$  is an **equivalence relation of finite index**

Let  $L \subseteq \Sigma^*$  be a language. Define the relation  $\sim_L \subseteq \Sigma^* \times \Sigma^*$  as:

$$u \sim_L v \iff [\forall w \in \Sigma^* . uw \in L \iff vw \in L]$$

$\sim_L$  is an **equivalence relation**

## The Myhill-Nerode Theorem

**Theorem 5** *A language  $L \subseteq \Sigma^*$  is recognizable iff  $\sim_L$  is of finite index.*

“ $\Rightarrow$ ” Suppose  $L = \mathcal{L}(A)$  for some automaton  $A$ .

$\sim_A$  is of finite index.

for all  $u, v \in \Sigma^*$  we have  $u \sim_A v \Rightarrow u \sim_L v$

index of  $\sim_L \leq$  index of  $\sim_A < \infty$

## The Myhill-Nerode Theorem

“ $\Leftarrow$ ”  $\sim_L$  is an equivalence relation of finite index, and let  $[u]$  denote the equivalence class of  $u \in \Sigma^*$ .

$A = \langle S, I, T, F \rangle$ , where:

- $S = \{[u] \mid u \in \Sigma^*\}$ ,
- $I = [\epsilon]$ ,
- $[u] \xrightarrow{\alpha} [v] \iff u\alpha \sim_L v$ ,
- $F = \{[u] \mid u \in L\}$ .

For DFA all minimal automata are isomorphic.

For NFA there may be more non-isomorphic minimal automata.

## Pumping Lemma

**Lemma 2 (Pumping)** *Let  $A = \langle S, I, T, F \rangle$  be a finite automaton with  $\text{size}(A) = n$ , and  $w \in \mathcal{L}(A)$  be a word of length  $|w| \geq n$ . Then there exists three words  $u, v, t \in \Sigma^*$  such that:*

1.  $|v| \geq 1$ ,
2.  $w = uvt$  and,
3. for all  $k \geq 0$ ,  $uv^k t \in \mathcal{L}(A)$ .

## Example

$L = \{a^n b^n \mid n \in \mathbb{N}\}$  is not recognizable:

Suppose that there exists an automaton  $A$  with  $size(A) = N$ , such that  $L = \mathcal{L}(A)$ .

Consider the word  $a^n b^n \in L = \mathcal{L}(A)$ , such that  $2n \geq N$ .

There exists words  $u, v, w$  such that  $|v| \geq 1$ ,  $uvw = a^n b^n$  and  $uv^k w \in L$  for all  $k \geq 1$ .

- $v = a^m$ , for some  $m \in \mathbb{N}$ .
- $v = a^m b^p$  for some  $m, p \in \mathbb{N}$ .
- $v = b^m$ , for some  $m \in \mathbb{N}$ .

## Decidability

Given nondeterministic finite automata  $A$  and  $B$ :

- **Emptiness**  $\mathcal{L}(A) = \emptyset$  ?
- **Inclusion**  $\mathcal{L}(A) \subseteq \mathcal{L}(B)$  ?
- **Equivalence**  $\mathcal{L}(A) = \mathcal{L}(B)$  ?
- **Infinity**  $\|\mathcal{L}(A)\| < \infty$  ?
- **Universality**  $\mathcal{L}(A) = \Sigma^*$  ?

## Emptiness

**Theorem 6** *Let  $A$  be an automaton with  $\text{size}(A) = n$ . If  $\mathcal{L}(A) \neq \emptyset$ , then there exists a word of length less than  $n$  that is accepted by  $A$ .*

Let  $u$  be the shortest word in  $\mathcal{L}(A)$ .

If  $|u| < n$  we are done.

If  $|u| \geq n$ , there exists  $u_1, v, u_2 \in \Sigma^*$  such that  $|v| > 1$  and  $u_1vu_2 = u$ .

Then  $u_1u_2 \in \mathcal{L}(A)$  and  $|u_1u_2| < |u_1vu_2|$ , contradiction.

## Everything is decidable

**Theorem 7** *The emptiness, equality, infinity and universality problems are decidable for automata on finite words.*

Although complexity varies from problem to problem:

- **Emptiness** ( $\mathcal{L}(A) = \emptyset$ ) belongs to NLOGSPACE
- **Inclusion** ( $\mathcal{L}(A) \subseteq \mathcal{L}(B)$ ) is PSPACE-complete
- **Equivalence** ( $\mathcal{L}(A) = \mathcal{L}(B)$ ) is PSPACE-complete
- **Infinity** ( $\|\mathcal{L}(A)\| < \infty$ ) belongs to NLOGSPACE
- **Universality** ( $\mathcal{L}(A) = \Sigma^*$ ) is PSPACE-complete

# Automata on Finite Words and WS1S

## WS1S

Let  $\Sigma = \{a, b, \dots\}$  be a finite alphabet.

Any finite word  $w \in \Sigma^*$  induces the *finite* sets  $p_a = \{p \mid w(p) = a\}$ .

- $x \leq y$  :  $x$  is less than  $y$ ,
- $s(x) = y$  :  $y$  is the successor of  $x$ ,
- $p_a(x)$  :  $a$  occurs at position  $x$  in  $w$

Remember that  $\leq$  and  $s(\cdot)$  can be defined one from another.

## Problem Statement

Given a sentence  $\varphi$  in WS1S, let  $\mathcal{L}(\varphi) = \{w \mid \mathfrak{m}_w \models \varphi\}$ , where  $\mathfrak{m}_w = \langle \text{dom}(w), \{\bar{p}_a\}_{a \in \Sigma}, \leq \rangle$ , such that:

- $\text{dom}(w) = \{0, 1, \dots, n - 1\}$ ,
- $\bar{p}_a = \{x \in \text{dom}(w) \mid w(x) = a\}$ ,

A language  $L \subseteq \Sigma^*$  is said to be *WS1S-definable* iff there exists a WS1S sentence  $\varphi$  such that  $L = \mathcal{L}(\varphi)$ .

1. Given  $A$  build  $\varphi_A$  such that  $\mathcal{L}(A) = \mathcal{L}(\varphi)$
2. Given  $\varphi$  build  $A_\varphi$  such that  $\mathcal{L}(A) = \mathcal{L}(\varphi)$

The recognizable and WS1S-definable languages coincide

## Coding of $\Sigma$

Let  $m \in \mathbb{N}$  be the smallest number such that  $\|\Sigma\| \leq 2^m$ .

W.l.o.g. assume that  $\Sigma = \{0, 1\}^m$ , and let  $X_1 \dots X_p, x_{p+1}, \dots, x_m$

A word  $w \in \Sigma^*$  induces an *interpretation* of  $X_1 \dots X_p, x_{p+1}, \dots, x_m$ :

- $i \in \iota_w(X_j)$  iff the  $j$ -th element of  $w_i$  is 1, and
- $\iota_w(x_j) = i$  iff  $w_i$  has 1 on the  $j$ -th position and, for all  $k \neq i$   $w_k$  has 0 on the  $j$ -th position.

## Example

*Example 1* Let  $\Sigma = \{a, b, c, d\}$ , encoded as  $a = (00)$ ,  $b = (01)$ ,  $c = (10)$  and  $d = (11)$ . Then the word  $abbaacdd$  induces the valuation  $X_1 = \{5, 6, 7\}$ ,  $X_2 = \{1, 2, 6, 7\}$ .  $\square$

## From Automata to Formulae

Let  $A = \langle S, I, T, F \rangle$  with  $S = \{s_1, \dots, s_p\}$ , and  $\Sigma = \{0, 1\}^m$ .

Build  $\Phi_A(X_1, \dots, X_m)$  such that  $\forall w \in \Sigma^* . w \in \mathcal{L}(A) \iff \llbracket \Phi_A \rrbracket_{\mathfrak{m}_w} = \text{true}$

Let  $a \in \{0, 1\}^m$ . Let  $\Phi_a(x, X_1, \dots, X_m)$  be the conjunction of:

- $X_i(x)$  if the  $a_i = 1$ , and
- $\neg X_i(x)$  otherwise.

For all  $w \in \Sigma^*$  we have  $\mathfrak{m}_w \models \forall x . \bigvee_{a \in \Sigma} \Phi_a(x, \vec{X})$

Notice that  $\Phi_a \wedge \Phi_b$  is unsatisfiable, for  $a \neq b$ .

## Coding of $S$

Let  $\{Y_1, \dots, Y_p\}$  be set variables.

$Y_i$  is the set of all positions labeled by  $A$  with state  $s_i$  during some run

$$\Phi_S(Y_1, \dots, Y_p) : \forall z . \bigvee_{1 \leq i \leq p} Y_i(z) \wedge \bigwedge_{1 \leq i < j \leq p} \neg \exists z . Y_i(z) \wedge Y_j(z)$$

## Coding of $I$

Every run starts from an initial state:

$$\Phi_I(Y_1, \dots, Y_p) : \exists x \forall y . x \leq y \wedge \bigvee_{s_i \in I} Y_i(x)$$

## Coding of $T$

Consider the transition  $s_i \xrightarrow{a} s_j$ :

$$\Phi_T(X_1, \dots, X_m, Y_1, \dots, Y_p) : \forall x . \neg last(x) \wedge Y_i(x) \wedge \Phi_a(x, \vec{X}) \rightarrow \bigvee_{(s_i, a, s_j) \in T} Y_j(s(x))$$

where  $last(x) = \forall y . y \leq x$

## Coding of $F$

The last state on the run is a final state:

$$\Phi_F(Y_1, \dots, Y_p) : \exists x . \textit{last}(x) \wedge \bigvee_{s_i \in F} Y_i(x)$$

$$\Phi_A = \exists Y_1 \dots \exists Y_p . \Phi_S \wedge \Phi_I \wedge \Phi_T \wedge \Phi_F$$

## From Formulae to Automata

Let  $\Phi(X_1, \dots, X_p, x_{p+1}, \dots, x_m)$  be a WS1S formula.

Build an automaton  $A_\Phi$  such that  $\forall w \in \Sigma^* . w \in \mathcal{L}(A) \iff \llbracket \Phi \rrbracket_{\iota_w}^{m_w} = \text{true}$

Let  $\Phi(X_1, X_2, x_3, x_4)$  be:

1.  $X_1(x_3)$
2.  $x_3 \leq x_4$
3.  $X_1 = X_2$

## From Formulae to Automata

$A_\Phi$  is built by induction on the structure of  $\Phi$ :

- for  $\Phi = \phi_1 \wedge \phi_2$  we have  $\mathcal{L}(A_\Phi) = \mathcal{L}(A_{\phi_1}) \cap \mathcal{L}(A_{\phi_2})$
- for  $\Phi = \phi_1 \vee \phi_2$  we have  $\mathcal{L}(A_\Phi) = \mathcal{L}(A_{\phi_1}) \cup \mathcal{L}(A_{\phi_2})$
- for  $\Phi = \neg\phi$  we have  $\mathcal{L}(A_\Phi) = \overline{\mathcal{L}(A_\phi)}$
- for  $\Phi = \exists X_i . \phi$ , we have  $\mathcal{L}(A_\Phi) = pr_i(\mathcal{L}(A_\phi))$ .

## Consequences

**Theorem 8** *A language  $L \subseteq \Sigma^*$  is definable in WS1S iff it is recognizable.*

**Corollary 1** *The SAT problem for WS1S is decidable.*

**Exercise 2** *Prove that there is no WS1S formula  $\varphi(x, y, z)$  that defines the relation  $\{(m, n, p) \in \mathbb{N}^3 \mid m + n = p\}$ .*

# Regular, Star Free and Aperiodic Languages

## Regular Languages

Let  $\Sigma$  be an alphabet, and  $X, Y \subseteq \Sigma^*$

$$XY = \{xy \mid x \in X \text{ and } y \in Y\}$$

$$X^* = \{x_1 \dots x_n \mid n \geq 0, x_1, \dots, x_n \in X\}$$

The class of *regular languages*  $\mathcal{R}(\Sigma)$  is the smallest class of languages  $L \subseteq \Sigma^*$  such that:

- $\emptyset, \{\epsilon\} \in \mathcal{R}(\Sigma)$
- $\{\alpha\} \in \mathcal{R}(\Sigma)$ , for all  $\alpha \in \Sigma$
- if  $X, Y \in \mathcal{R}(\Sigma)$  then  $X \cup Y, XY, X^* \in \mathcal{R}(\Sigma)$

## Regular, rational and recognizable languages

**Theorem 9 (Kleene)** *A set of finite words is recognizable if and only if it is regular.*

Proof in every textbook.

**Rational** = regular, in older books e.g.

Samuel Eilenberg. *Automata, Languages and Machines*. Academic Press, 1974

**NB:** if regular and recognizable languages are the same, then regular languages are closed under boolean operations

## Star Free Languages

The class of *star-free languages* is the smallest class  $SF(\Sigma)$  of languages  $L \in \Sigma^*$  such that:

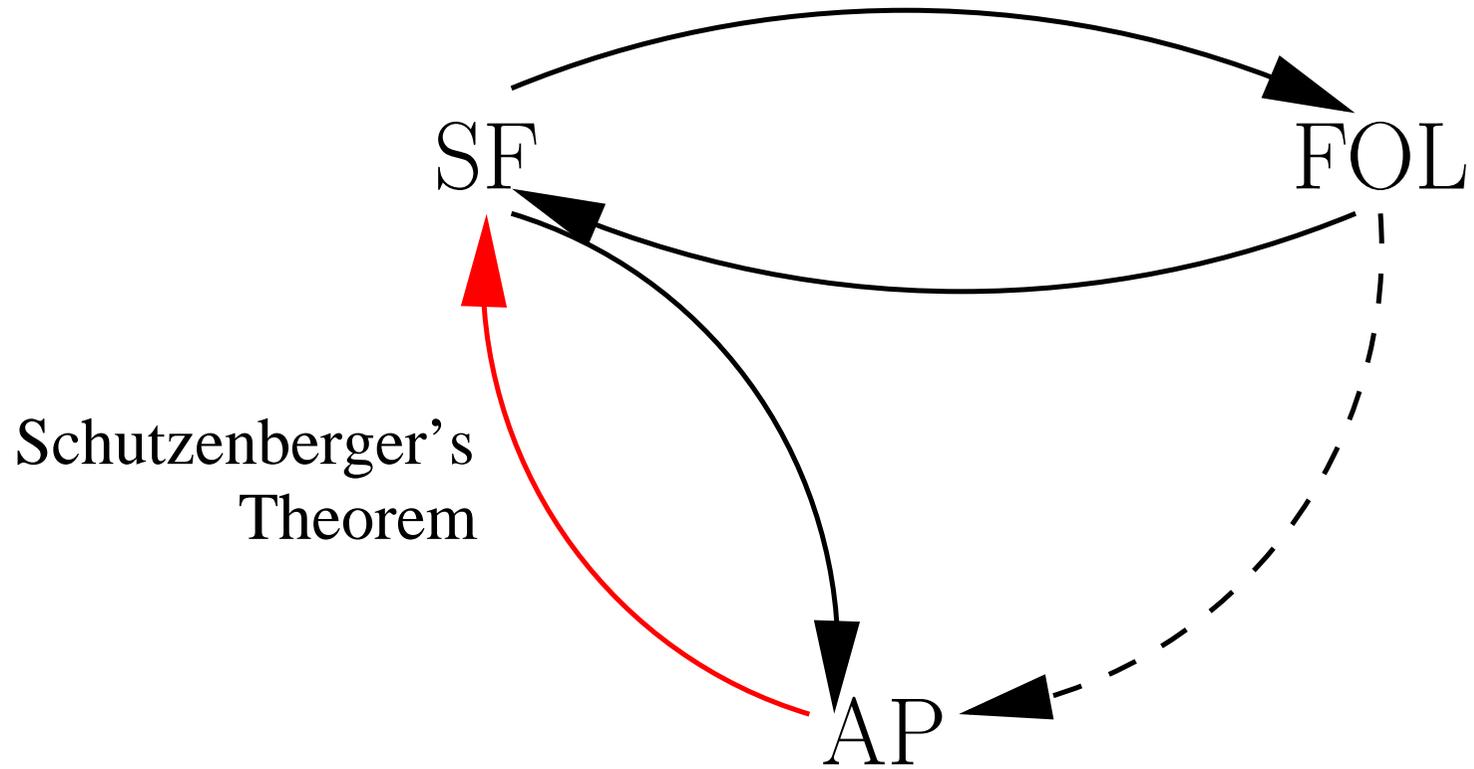
- $\emptyset, \{\epsilon\} \in SF(\Sigma)$  and  $\{a\} \in SF(\Sigma)$  for all  $a \in \Sigma$
- if  $X, Y \in SF(\Sigma)$  then  $X \cup Y, XY, \overline{X} \in SF(\Sigma)$  (hence  $X \cap Y \in SF(\Sigma)$ )

### *Example 2*

- $\Sigma^* = \overline{\emptyset}$  is star-free
- if  $B \subset \Sigma$ , then  $\Sigma^* B \Sigma^* = \bigcup_{b \in B} \Sigma^* b \Sigma^*$  is star-free
- if  $B \subset \Sigma$ , then  $B^* = \overline{\Sigma^* \overline{B} \Sigma^*}$  is star-free
- if  $\Sigma = \{a, b\}$ , then  $(ab)^* = \overline{b \Sigma^* \cup \Sigma^* a \cup \Sigma^* a a \Sigma^* \cup \Sigma^* b b \Sigma^*}$  is star-free

**Exercise 3** If  $\Sigma = \{a, b, c\}$ , write  $(ab)^*$  as a star-free language.

SF = FOL (= AP)



## The Splitting Lemma

**Lemma 3** *Let  $A, B \subseteq \Sigma$  be subalphabets such that  $A \cap B = \emptyset$ . Then, for each star-free language  $L \in SF(\Sigma)$ , we have:*

$$L \cap B^*AB^* = \bigcup_{1 \leq i \leq n} K_i a_i L_i$$

*where  $a_i \in A$  and  $K_i, L_i \in SF(B)$ , for all  $1 \leq i \leq n$ .*

W.l.o.g. we prove the case  $A = \{a\}$  (why?) by induction on  $L$ :

- If  $L = \{a\}$  then  $L \cap B^*AB^* = \{\epsilon\}a\{\epsilon\}$ .
- If  $L = \{a'\}$ ,  $a' \neq a$ , then  $L \cap B^*AB^* = \emptyset a \emptyset$ .
- If  $L = \Sigma^*$  then  $L \cap B^*AB^* = B^*AB^*$ .
- If  $L = L_1 \cup L_2$  then  $L \cap B^*AB^* = (L_1 \cap B^*AB^*) \cup (L_2 \cap B^*AB^*)$ .

## The Splitting Lemma

$$L \cap B^*AB^* = \bigcup_{1 \leq i \leq n} K_i a_i L_i$$

- If  $L = L_1 \cdot L_2$  then

$$L \cap B^*AB^* = (L_1 \cap B^*) \cdot (L_2 \cap B^*AB^*) \cup (L_1 \cap B^*AB^*) \cdot (L_2 \cap B^*).$$

- Else, if  $L = \Sigma^* \setminus L'$ , by the inductive hypothesis  $L' = \bigcup_{1 \leq i \leq n} K'_i a L'_i$ .

We assume w.l.o.g that  $\{K'_i\}_{i=1}^n$  form a partition of  $B^*$ :

- if  $K'_i \cap K'_j \neq \emptyset$ , rewrite

$$K'_i a L'_i \cup K'_j a L'_j = (K'_i \setminus K'_j) a L'_i \cup (K'_j \setminus K'_i) a L'_j \cup (K'_i \cap K'_j) a (L'_i \cup L'_j)$$

- if  $\bigcup_{i=1}^n K'_i \subsetneq B^*$ , add  $(B^* \setminus \bigcup_{i=1}^n K'_i) a \emptyset$  to  $\{K'_i a L'_i\}_{i=1}^n$

$$(\Sigma^* \setminus L') \cap B^* a B^* = \bigcup_{i=1}^n K'_i a (B^* \setminus L'_i)$$

## Subword Formulae

Let  $w = a_0a_1 \dots a_{n-1}$  be a finite word, and  $w(i, j) = a_i a_{i+1} \dots a_{j-1}$  be a subword of  $w$ ,  $0 \leq i < n$  and  $0 \leq j \leq n$ ,  $i < j$ .

**Proposition 3** *For each FOL sentence  $\varphi$  there exists a formula  $\varphi[x, y]$  such that, for each  $w \in \Sigma^*$  and each  $0 \leq i < j \leq |w|$ :*

$$\mathfrak{m}_{w(i,j)} \models \varphi \iff \llbracket \varphi[x, y] \rrbracket_{[x \leftarrow i][y \leftarrow j]}^{\mathfrak{m}_w} = \mathbf{true}$$

By induction on the structure of  $\varphi$ :

$$\begin{aligned}(\neg\varphi)[x, y] &= \neg(\varphi[x, y]) \\(\varphi \wedge \psi)[x, y] &= (\varphi[x, y]) \wedge (\psi[x, y]) \\(\exists z.\varphi)[x, y] &= \exists z . x \leq z \wedge z < y \wedge \varphi[x, y]\end{aligned}$$

## Star Free Languages are FOL-definable

For each  $L \in SF(\Sigma)$ , there exists an FOL sentence  $\varphi_L$  such that:

$$L = \{w \in \Sigma^* \mid \mathfrak{m}_w \models \varphi_L\}$$

By induction on the structure of  $L$ :

$$\emptyset = \{w \in \Sigma^* \mid \mathfrak{m}_w \models \perp\}$$

$$\{a\} = \{w \in \Sigma^* \mid \mathfrak{m}_u \models p_a(0) \wedge last(0)\}$$

$$X \cup Y = \{w \in \Sigma^* \mid \mathfrak{m}_u \models \varphi_X \vee \varphi_Y\}$$

$$\overline{X} = \{w \in \Sigma^* \mid \mathfrak{m}_u \models \neg\varphi_X\}$$

$$X \cdot Y = \{w \in \Sigma^* \mid \mathfrak{m}_u \models \exists y \exists z . 0 \leq y \leq z \wedge \varphi_X[0, y] \wedge \varphi_Y[y, z] \wedge last(z)\}$$

## FOL-definable Languages are Star Free

Let  $\varphi$  be an FOL formula with  $FV(\varphi) = V$  and let  $\Sigma_V = \Sigma \times \{0, 1\}^V$ .

Encode each pair  $(w, \iota)$ , with  $\iota : V \rightarrow [0, |w| - 1]$  as a word  $\overline{(w, \iota)} \in \Sigma_V^*$ :

$$\overline{(a_0 \dots a_{k-1}, \iota)} = (a_0, \tau_0) \dots (a_{k-1}, \tau_{k-1}), \quad \tau_i(x) = 1 \iff \iota(x) = i$$

and let  $\mathcal{N}_V = \{\overline{(w, \iota)} \mid w \in \Sigma^*, \iota : V \rightarrow [0, |w| - 1]\}$ .

Let  $\Sigma_V^{x=i} = \{(a, \tau) \mid a \in \Sigma, \tau(x) = i\}$ , for  $i = 0, 1$

$$\mathcal{N}_V = \bigcap_{x \in V} (\Sigma_V^{x=0})^* (\Sigma_V^{x=1}) (\Sigma_V^{x=0})^* \in SF(\Sigma_V)$$

## FOL-definable Languages are Star Free

$$\begin{aligned} \llbracket p_a(x) \rrbracket_V &= \{ \overline{(w, \iota)} \in \mathcal{N}_V \mid w = a_0 \dots a_{k-1}, a_{\iota(x)} = a \} \\ \llbracket x \leq y \rrbracket_V &= \{ \overline{(w, \iota)} \in \mathcal{N}_V \mid \iota(x) \leq \iota(y) \} \\ \llbracket \phi \vee \psi \rrbracket_V &= \llbracket \phi \rrbracket_V \cup \llbracket \psi \rrbracket_V \\ \llbracket \neg \phi \rrbracket_V &= \mathcal{N}_V \setminus \llbracket \phi \rrbracket_V \\ \llbracket \exists x . \phi \rrbracket_V &= \{ \overline{(w, \iota)} \in \mathcal{N}_V \mid \exists i \in [0, |w| - 1] . \overline{(w, \iota[x \leftarrow i])} \in \llbracket \phi \rrbracket_{V \cup \{x\}} \} \end{aligned}$$

**Proposition 4** *If  $\varphi \in FOL$  and  $FV(\varphi) \subseteq V$ , then  $\llbracket \varphi \rrbracket_V \in SF(\Sigma_V)$ .*

$$\begin{aligned} \llbracket p_a(x) \rrbracket_V &= \mathcal{N}_V \cap (\Sigma_V^* \cdot \{(a, \tau) \mid \tau(x) = 1\} \cdot \Sigma_V^*) \\ \llbracket x \leq y \rrbracket_V &= \mathcal{N}_V \cap (\Sigma_V^* \cdot \Sigma_V^{x=1} \cdot \Sigma_V^* \cdot \Sigma_V^{y=1} \cdot \Sigma_V^*) \end{aligned}$$

## FOL-definable Languages are Star Free

**Proposition 5** *If  $\varphi \in FOL$  and  $FV(\varphi) \subseteq V$ , then  $\llbracket \varphi \rrbracket_V \in SF(\Sigma_V)$ .*

If  $\varphi = \exists x . \phi$ , we assume w.l.o.g. that  $x \notin V$  ( $\alpha$ -conversion)

$$\begin{aligned}\llbracket \phi \rrbracket_{V \cup \{x\}} &= \llbracket \phi \rrbracket_{V \cup \{x\}} \cap (\Sigma_{V \cup \{x\}}^{x=0})^* (\Sigma_{V \cup \{x\}}^{x=1}) (\Sigma_{V \cup \{x\}}^{x=0})^* \\ &= \bigcup_{i=1}^n K'_i a'_i L'_i \text{ (Splitting Lemma)}\end{aligned}$$

where  $K'_i, L'_i \in SF(\Sigma_{V \cup \{x\}}^{x=0})$  and  $a'_i \in \Sigma_{V \cup \{x\}}^{x=1}$ , for all  $1 \leq i \leq n$

Let  $\pi : \Sigma_{V \cup \{x\}}^{x=0} \rightarrow \Sigma_V$  be the **bijective** (why?) renaming  $(a, \tau) \xrightarrow{\pi} (a, \tau \downarrow_V)$

Let  $K_i = \pi(K'_i)$ ,  $L_i = \pi(L'_i)$  and  $a_i = (a, \tau \downarrow_V) \iff a'_i = (a, \tau)$

$$\llbracket \exists x . \phi \rrbracket_V = \bigcup_{i=1}^n K_i a_i L_i$$

**NB:** SF languages are preserved by bijective renamings (why bijective ?)

## Aperiodic Languages

**Definition 3** A language  $L \subseteq \Sigma^*$  is said to be **aperiodic** iff:

$$\exists n_0 \forall n \geq n_0 \forall u, v, t \in \Sigma^* . uv^n t \in L \iff uv^{n+1} t \in L$$

$n_0$  is called the **index** of  $L$ .

**Example 3**  $0^*1^*$  is aperiodic. Let  $n_0 = 2$ . We have three cases:

1.  $u, v \in 0^*$  and  $t \in 0^*1^*$  :

$$\forall n \geq 2 . uv^n t \in L$$

2.  $u \in 0^*$ ,  $v \in 0^+1^+$  and  $t \in 1^*$  :

$$\forall n \geq 2 . uv^n t \notin L$$

3.  $u \in 0^*1^*$ ,  $v \in 1^*$  and  $t \in 1^*$  :

$$\forall n \geq 2 . uv^n t \in L$$

## Periodic Languages

Conversely, a language  $L \subseteq \Sigma^*$  is said to be *periodic* iff:

$$\forall n_0 \exists n \geq n_0 \exists u, v, t \in \Sigma^* . (uv^n t \notin L \wedge uv^{n+1} t \in L) \vee (uv^n t \in L \wedge uv^{n+1} t \notin L)$$

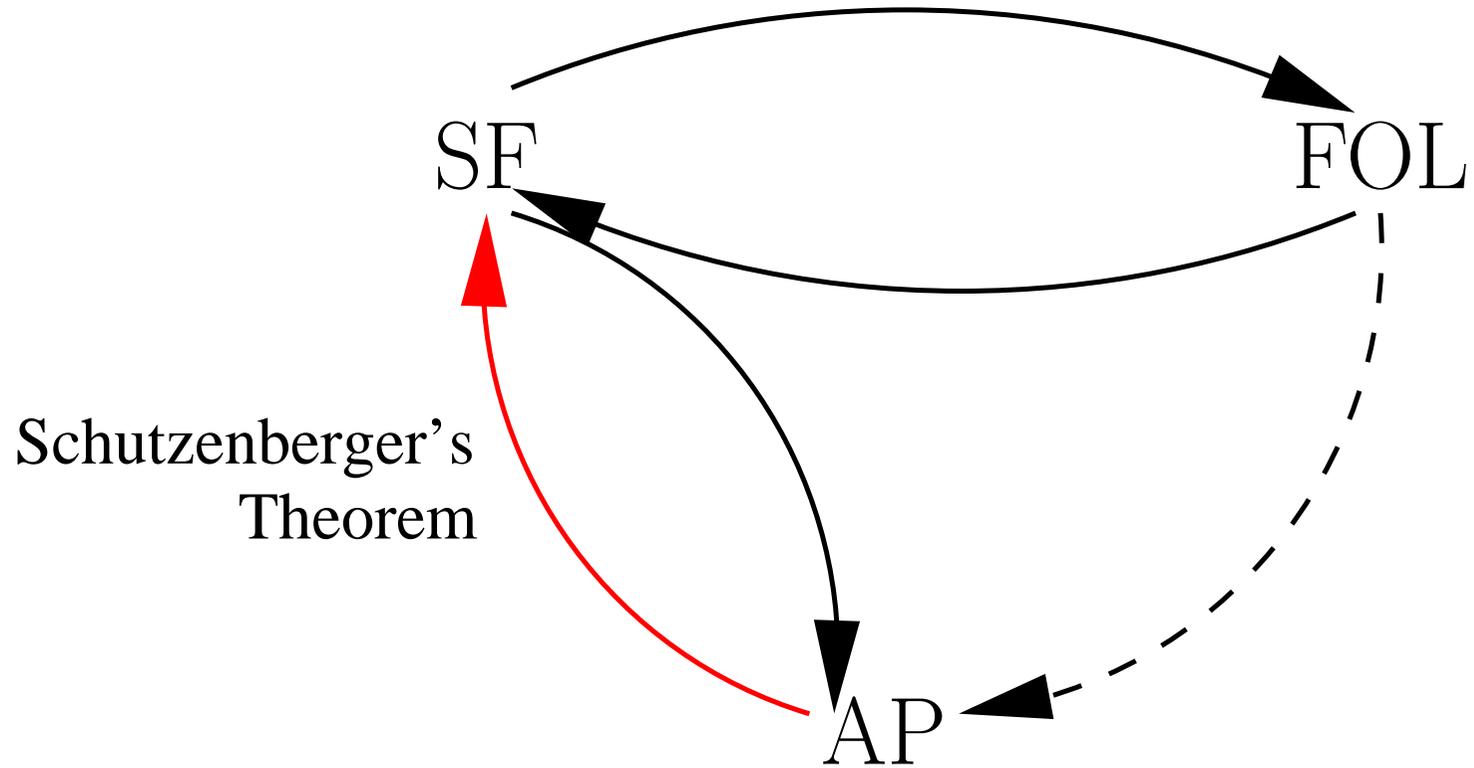
*Example 4*  $(00)^*1$  is periodic.

Given  $n_0$  take the next even number  $n \geq n_0$ ,  $u = \epsilon$ ,  $v = 0$  and  $t = 1$ . Then  $uv^n t \in (00)^*1$  and  $uv^{n+1} t \notin (00)^*1$ .  $\square$

*Exercise 4* Is  $(00)^*1$  WS1S-definable ?

*Exercise 5* Is the language  $(ab)^*$  periodic or aperiodic ?

# The Big Picture



## From Star-free to Aperiodic

**Proposition 6** *If  $L \in SF(\Sigma)$  then  $L$  is aperiodic.*

Prove the existence of an integer  $N(L)$  such that

$$\forall n \geq N(L) \forall u \forall v \forall t . uv^n t \in L \iff uv^{n+1} t \in L$$

. Suppose  $v \neq \epsilon$ . By induction on the structure of  $L$ :

- $\emptyset$  :  $N(\emptyset) = 0$ , since  $\forall n \geq 0 . uv^n t \notin L$
- $\{a\}, a \in \Sigma$  :  $N(\{a\}) = 2$ , since  $\forall n \geq 2 . uv^n t \notin L$
- $\overline{X}$  :  $N(\overline{X}) = N(X)$ , trivial
- $X \cup Y$  :  $N(X \cup Y) = \max\{N(X), N(Y)\}$ , trivial
- $XY$  :  $N(XY) = N(X) + N(Y) + 1$ , since for all  $n = n_1 + n_2 + 1 \geq N(X) + N(Y) + 1$ , we have either  $n_1 \geq N(X)$  or  $n_2 \geq N(Y)$ . Then  $uv^n t = (uv^{n_1} r)(sv^{n_2} t)$ , where  $rs = v$  and  $uv^{n_1} r \in X, sv^{n_2} t \in Y$ . If  $n_1 \geq N(X)$ ,  $uv^{n_1+1} r \in X \Rightarrow uv^{n+1} t \in XY \square$