Recapitulation:

Star-free languages

(1) $\emptyset \in SF_\Sigma$, $\{\varepsilon\} \in SF_\Sigma$, $\{a\} \in SF_\Sigma$ f.a. $a \in \Sigma$.

(2) If $L_1, L_2 \in SF_\Sigma$ then $L_1 \cup L_2$, $L_1 . L_2$, $\overline{L_1} \in SF_\Sigma$

=> Complement not an operation on regular sets (but we know that it can be derived)

=> Complement may yield $*$ in alternative representations of the language.

Theorem (McNaughton, Papert '71, part I, homework)
Let L be star-free. Then L is $FO[<]$-definable.

Goal: If L is $FO[<]$-definable then it is star-free.
This concludes our study:

$$\boxed{REG = WMSO\text{-definable} \quad \boxed{SF = FO[<]\text{-definable}} \quad (aa)^*}$$

Requires some insights:

↳ Relation $\equiv_{k,n}$ has finitely many equivalence classes

↳ Every class of $\equiv_{k,n}$ can be characterised by single formula
(The technique of characteristic formulas is often helpful
⇝ process algebra)

=> Afterwards give inductive construction of star-free representation for $FO[<]$-defined language.

Lemma 1:
For all $k, n \in \mathbb{N}$, relation $\equiv_{k,n}$ is an equivalence with finitely many classes.

Remember:

$(S_v, \vec{s}) \equiv_{k,n} (S_w, \vec{t})$ if $S_v, I[\vec{s}/\vec{x}] \models \varphi$ iff $S_w, I[\vec{t}/\vec{x}] \models \varphi$

f.a. formulas φ with qd $\leq k$ and n variables free.

Proof:

Equivalence: ✓

Finitely many classes:

Follows from:

Up to logical equivalence, there are only finitely many formulas of quantifier depth k with n free variables.

(φ, ψ logically equivalent if $S_v, I[\vec{s}/\vec{x}] \models \varphi$ iff $S_v, I[\vec{s}/\vec{x}] \models \psi$ f.a. $(S_v, \vec{s})$.)

To see this, note that every formula can be transformed into prenex normal form:

$$\underbrace{\exists/\forall x_1 : \ldots \exists/\forall x_k :}_{\text{Sequence of } k \text{ quantifiers}} \underbrace{\psi}$$

- Quantifier-free formula over $\underbrace{x_1, \ldots, x_k}_{\text{bound}}$ and $\underbrace{x_{k+1}, \ldots, x_{k+n}}_{\text{free variables}}$
- Boolean combination of atomic propositions $c = P_a(x)$ and $c = x < y$ (potentially negated)
- Wlog. in disjunctive normal form (DNF)

$$\bigvee_j \underbrace{\left(\bigwedge_i c_{ji}\right)}_{\text{disjunct}}$$

Estimate number of formulas in prenex normal form with qd $= k$ and n-free variables:

- $P_a(x)$ with $x \in \{x_1, \ldots, x_{k+n}\} \Rightarrow |\Sigma|(k+n)$ atomic propositions
- $x < y$ with $x, y \in \{x_1, \ldots, x_{k+n}\} \Rightarrow (k+n)^2$ atomic propositions
- Negation $\Rightarrow 2(|\Sigma|(k+n) + (k+n)^2)$ atomic propositions

· Thus, we have at most

$$2^{2(|\Sigma|(k+n)+(k+n)^2)} \quad \text{possible disjuncts for the DNF}$$

· Thus, we have at most

$$2^{2^{2(|\Sigma|(k+n)+(k+n)^2)}} \quad \text{possible DNFs.}$$

For quantified variables, we can choose between $\exists$ and $\forall$:

$$2^k \cdot 2^{2^{2(|\Sigma|(k+n)+(k+n)^2)}} \quad \text{formulas in prenex normal form.}$$

· Every equivalence class of $\equiv_{k,n}$ is determined by the formulas φ that hold with $qd(\varphi) \leq k$ and n free variables.

↳ We just realised that we (up to logical equivalence) finitely many such formulas, say $p \in \mathbb{N}$.

↳ Thus there are at most 2^p equivalence classes. □

<u>Lemma 2</u>:

For every $\equiv_{k,n}$ equivalence class $[S_v, \vec{s}]_{\equiv_{k,n}}$ there is a formula $\varphi_{[S_v,\vec{s}]_{\equiv_{k,n}}}$ of $qd \leq k$ so that

$$(S_w, \vec{t}) \in [S_v, \vec{s}]_{\equiv_{k,n}} \quad \text{iff} \quad S_w, I[\vec{t}/\vec{x}] \models \varphi_{[S_v,\vec{s}]_{\equiv_{k,n}}}$$

<u>Proof</u>:

Define

$$\varphi_{[S_v,\vec{s}]_{\equiv_{k,n}}} := \bigwedge_{\substack{\cdot\ \varphi \text{ of } qd \leq k \\ \cdot\ n\text{-free variables} \\ \cdot\ S_v, I[\vec{s}/\vec{x}] \models \varphi}} \varphi$$

// Conjunction of all formulas (up to logical equivalence) that hold in $[S_v, \vec{s}]_{\equiv_{k,n}}$.

By definition of $\equiv_{k,n}$, $\varphi_{[S_v,\vec{s}]_{\equiv_{k,n}}}$ independent of representative.

$\Rightarrow$ ✓

$\Leftarrow$ Left as an exercise.

Theorem (McNaughton, Papert '71, part II)

Let φ an $FO[<]$-sentence. Then $L(\varphi)$ is star-free.

Proof:

By induction on the quantifier-depth.

IA: The only sentences of quantifier depth 0 are true and false.
$k=0$ They yield

$$L(true) = \Sigma^* \quad \text{and} \quad L(false) = \emptyset$$

which are star-free.

IS: Assume formulas of qd $\leq k$ have star-free languages.
Consider

$$\varphi = \exists x : \psi \quad \text{where } \psi \text{ is of quantifier depth } k.$$

The remaining cases follow by the fact that star-free languages are closed under Boolean operations.

Claim:

$$L(\exists x : \psi) \overset{?}{=} \bigcup_{S_{u.a.v}, I[{}^{|u|}/_x] \models \psi} [u]_{\equiv_k} . a . [v]_{\equiv_k}$$

// Notation $u \equiv_k u'$ if $S_u \equiv_k S_{u'}$

By Lemma 1, this union is finite.

By Lemma 2, there are formulas $\varphi_{[u]_{\equiv_k}}$ and $\varphi_{[v]_{\equiv_k}}$ of qd $\leq k$ so that

$$u' \in [u]_{\equiv_k} \quad \text{iff} \quad S_{u'} \models \varphi_{[u]_{\equiv_k}}$$

By the hypothesis, $L(\varphi_{[u]_{\equiv_k}})$ and $L(\varphi_{[v]_{\equiv_k}})$ are star-free with expressions r_u and r_v.

Then

$$L(\exists x : \psi) = \bigcup_{S_{u.a.v}, I[{}^{|u|}/_x] \models \psi} r_u . a . r_v$$

All that remains to be shown is "$\overset{?}{=}$".

"$\subseteq$" Let $S_w \models \exists x: \psi$

Then there is a position with some letter $a \in \Sigma$ so that

$$w = u.a.v \quad \text{and} \quad S_{u.a.v}, I[{}^{|u|}/_x] \models \psi.$$

This means

$$w \in [u]_{\equiv_k} . a . [v]_{\equiv_k} \subseteq \bigcup_{\dots} \dots$$

"$\supseteq$" Let $w \in [u]_{\equiv_k} . a . [v]_{\equiv_k}$.

Then $w = u'.a.v'$ with $S_u \equiv_k S_{u'}$ and $S_v \equiv_k S_{v'}$.

With Ehrenfeucht-Fraïssé theorem (to be argued)

$$(S_{u'.a.v'}, |u'|) \equiv_{k,1} (S_{u.a.v}, |u|)$$

Since

$$S_{u.a.v}, I[{}^{|u|}/_x] \models \psi$$

we get

$$S_{u'.a.v'}, I[{}^{|u'|}/_x] \models \psi.$$

We conclude

$$S_{u'.a.v'} \models \exists x: \psi \qquad \text{which means } S_w \models \varphi$$

<u>On the Ehrenfeucht-Fraïssé argument:</u>

Since $S_u \equiv_k S_{u'}$ and $S_v \equiv_k S_{v'}$,

we have

$G_k(S_u, S_{u'})$ and $G_k(S_v, S_{v'})$ won by duplicator.

Consider

$$G_k((S_{u.a.v}, |u|), (S_{u'.a.v'}, |u'|))$$

Duplicator wins this game:

- The letters $a - a$ match
- For the left side, duplicator plays the winning strategy of $G_k(S_u, S_{u'})$.
- On the right side, duplicator wins $G_k(S_v, S_{v'})$.

$\square$

Recent results on FO-languages:

Gastin (Cachan), Diekert (Stuttgart) : FO for ω-languages '08

Darondeau (Rennes), Démri (Cachan), Mayer (KL), Morvan (Rennes) : FO for reachability graphs of Petri nets 'in preparation (predicates $x \to y$)