Formal Languages

[Linz § 1.2]

Start with an **alphabet**, Σ, a **finite**, **non-empty** set of **symbols**.

From symbols, construct **strings**.

Strings: Finite sequences of symbols from Σ.

Notation:

$a, b, c, \ldots$ for symbols

$u, v, w, \ldots$ for strings

Concatenation of Strings

If $u = a_1 \ldots a_n$, $v = b_1 \ldots b_m$

then $uv = a_1 \ldots a_n b_1 \ldots b_m$

Reverse of a String:

If $u = a_1 \ldots a_n$ then $u^{\mathsf{R}} = a_n \ldots a_1$

$|u|$ = length of u.

λ = **empty string**

Note:

$|\lambda| = 0$

$\lambda u = u\lambda = u$

Substring

If $w = u_1\, v\, u_2$ (possibly $u_1 = \lambda$, $u_2 = \lambda$),

then v is a **substring** of w.

If $w = uv$ (possibly $u = \lambda$, $v = \lambda$),

then u is a **prefix** or **initial substring** of w

and v is a **suffix** or **final substring** of w.

Example: [Linz, section 1.2]

Prove for any strings u, v

$$|uv| = |u| + |v|$$

(This much is obvious, but can you give a **formal proof**?)

First, we need a **formal definition** of the length $|u|$ of a string, u !

Notation

(1) $w^n = \underbrace{w\ w\ \ldots\ w}_{\text{n times}}$

Recursive Definition:

$w^0 = \underline{\lambda}$

$w^{n+1} = \underline{w^n\ w}$

(2) For any alphabet, Σ:

Σ^* is the set of all **strings** obtained by concatenating 0 or more symbols from Σ (including λ).

i.e. Σ^* is the set of all Σ-strings

$$\Sigma^+ = \Sigma^* \setminus \{\lambda\}$$

Or in other words, Σ^+ is all Σ^* except for λ.

Note:

Assume $\Sigma \neq \emptyset$

Then Σ^* and Σ^+ are always infinite!

Definition: A **language over** Σ or a **Σ-language** is a **set** of **strings** over Σ (or **Σ-strings**).

i.e. a **subset** of Σ^*.

Formal (recursive) definition of Σ^*

Base Clause:

$$\lambda \in \Sigma^*$$

Recursive Clause:

$$u \in \Sigma^*, a \in \Sigma \Longrightarrow u\, a \in \Sigma^*$$

Can give **modified BNF** (Backus-Naur Form):

Given Σ with symbols $a, b, \ldots$

define Σ^* with strings $u, v, \ldots$

$$u ::= \lambda \mid u\, a$$

Now give **formal definition** of length of string u by **recursion** on **construction** of $u \in \Sigma^*$, or **structural recursion** on $u \in \Sigma^*$.

$|\lambda| = 0$

$|u\, a| = |u| + 1$ (See Linz, Example 1.8)

Going back to our ***Example*** (p. 1-12):

Prove $\quad |u\, v| = |u| + |v| \qquad\qquad (*)$

Use **simple induction** on $|v|$
(or **structural induction** on $v \in \Sigma^*$).

Basis (base case) $\quad |v| = 0, v = \lambda$

$$|u\, v| = \underline{|u\, \lambda| = |u|}$$

$$|u| + |v| = \underline{|u| + 0 = |u|}$$

Induction step: **Assume** (*) for $|v| = n \qquad\qquad$ **(i.h.)**

Prove (*) for $|v| = n + 1$.

We will actually prove: $|u\, w| = |u| + |w|$ for $|w| = n + 1$.

So suppose $|w| = n + 1$, say $w = v\, a$.
Then $|v| = n$. So:

$$
\begin{aligned}
|\boldsymbol{u}\,\boldsymbol{w}| &= \underline{|\boldsymbol{u}\ (\boldsymbol{v}\ \boldsymbol{a})|} \\
&\underline{= |(\boldsymbol{u}\ \boldsymbol{v})\ \boldsymbol{a}|} \\
&\underline{= |\boldsymbol{u}\ \boldsymbol{v}| + \mathbf{1}} && \text{(def. of } |.|\text{)} \\
&\underline{= (|\boldsymbol{u}| + |\boldsymbol{v}|) + \mathbf{1}} && (\textbf{i.h.}) \\
&\underline{= |\boldsymbol{u}| + (|\boldsymbol{v}| + \mathbf{1})} \\
&\underline{= |\boldsymbol{u}| + |\boldsymbol{w}|} && \text{(def. of } |.|\text{)}
\end{aligned}
$$

So we have proved (*) for $|v| = n + 1$. $\square$

Definition: **Reverse u^{R} of string u.**

By **structural recursion on $u \in \Sigma^*$**

$\lambda^{\mathsf{R}} = \lambda$

$(u\, a)^{\mathsf{R}} = a\, u^{\mathsf{R}}$

Example: Prove $(uv)^{\mathsf{R}} = v^{\mathsf{R}} u^{\mathsf{R}}$ (*)

by **simple induction on $|v|$.**

Basis $|v| = 0, v = \lambda$.

So

$$(u\, \lambda)^{\mathsf{R}} = \underline{u^{\mathsf{R}}}$$

$$\lambda^{\mathsf{R}}\, u^{\mathsf{R}} = \underline{\lambda\, u^{\mathsf{R}} = u^{\mathsf{R}}}$$

Induction step: Assume (*) for $|v| = n$ **(i.h.)**

Prove (*) for $|v| = n + 1$.

We will actually prove: $(uw)^{\mathsf{R}} = w^{\mathsf{R}}u^{\mathsf{R}}$ for $|w| = n + 1$.

So let $|w| = n + 1,\ w = v\ a,\ |v| = n$.

We will show: $(u\ w)^{\mathsf{R}} = w^{\mathsf{R}}\ u^{\mathsf{R}}$ (**)

%beginalign*

$$
\begin{aligned}
\text{LHS of (**)} &= (u\ w)^{\mathsf{R}} \\
&\underline{= (u\ (v\ a))^{\mathsf{R}}} \\
&\underline{= ((u\ v)\ a)^{\mathsf{R}}} \\
&\underline{= a\ (u\ v)^{\mathsf{R}}} && \text{(def. of } u^{\mathsf{R}}\text{)} \\
&\underline{= a\ v^{\mathsf{R}}\ u^{\mathsf{R}}} && \textbf{(i.h.)} \\
\text{RHS of (**)} &\underline{= (v\ a)^{\mathsf{R}}\ u^{\mathsf{R}}} \\
&\underline{= a\ v^{\mathsf{R}}\ u^{\mathsf{R}}} && \text{(def. of } u^{\mathsf{R}}\text{)} \\
&\underline{= \textit{LHS of (**)}}
\end{aligned}
$$

Exercise: Prove $(u^{\mathsf{R}})^{\mathsf{R}} = u$.

Let L be a Σ-language, i.e. $L \subseteq \Sigma^*$.

A string in L is called a **sentence** of L.

Languages can be **finite** or **infinite**.

Examples: Let $\Sigma = \{a, b\}$

(1) $L = \{ab, a, aba\}$

(2) $L = \{a^n \, b^n \mid n \geq 0\}$

Definition: For any L, L_1, $L_2 \subseteq \Sigma^*$, we can define:

(1) $L_1 \cup L_2$

(2) $L_1 \cap L_2$

(3) **Complement** of $L = \overline{L} = \underline{\Sigma^* \backslash L}$

(4) **Reverse** of $L = L^{\mathsf{R}} = \underline{\{u^{\mathsf{R}} \mid u \in L\}}$

(5) **Concatenation** of L_1 and L_2:

$$L_1 \, L_2 = \underline{\{u \, v \mid u \in L_1, v \in L_2\}}$$

<u>***Note***</u>:

$\emptyset \, L = L \, \emptyset = \underline{\emptyset}$

$\{\lambda\} \, L = L \, \{\lambda\} = \underline{L}$

So $\{\lambda\}$ acts like an ***identity*** for ***concatenation*** of languages.
We write $I = \lambda$.

Powers of L:

$L^2 = L\,L = \{u\,v \mid u, v \in L\}$

$L^n = \{u_1\,u_2 \ldots u_n \mid u_i \in L \quad \text{for } i = 1, \ldots, n\}$

Recursive Definition of L^n

$L^0 = \underline{I = \{\lambda\}}$

$L^{n+1} = \underline{L^n \cdot L}$

Star-Closure

$$L^* = \bigcup_{n=0}^{\infty} L^n$$

$$= L^0 \cup L^1 \cup L^2 \cup \dots$$

$$= \underline{I \cup L \cup L^2 \cup \dots}$$

Positive Closure

$$L^+ = \bigcup_{n=1}^{\infty} L^n$$

$$= L \cup L^2 \cup \dots$$

$$= L^* \backslash \{\lambda\}$$

Example: [Linz, Example 1.10]

$\Sigma = \{a, b\}, \;\; L = \{a^n \, b^n \mid n \geq 0\}$

$L^2 = \underline{\{a^n \, b^n \, a^m \, b^m \mid n, m \geq 0\}}$

$L^{\mathsf{R}} = \underline{\{b^n \, a^n \mid n \geq 0\}}$