Theoretical importance of regular expressions

For the study of formal languages and grammars, the importance of regular expressions comes from the following theorem:

Theorem
A language is regular $\iff$ some regular expression describes it.

Sketch of proof: Recall the definition of a regular language as the language obtained in finitely many steps from finite subsets of words via union, concatenation or the Kleene star. We can construct a regular expression from the definition of the regular language in question, and vice versa starting w/ a regular expression we can define a finite sequence of L_i's such that each L_i is a finite set of words or is obtained from previous L_i's via union, concatenation or the Kleene star. (q.e.d.)

Finally, we can state the complete characterization of regular languages:

Theorem The following are equivalent:

(i) L is a regular language
(ii) L is recognized by a (deterministic or non-deterministic) finite state acceptor.
(iii) L is produced by a regular grammar.
(iv) L is given by a regular expression.

Remark It is possible to prove directly that (iv) $\iff$ (ii), but the construction is rather complicated. Instead, we sketched above the proof that (i) $\iff$ (iv), and we had previously stated that (i) $\iff$ (ii) $\iff$ (iii), so now we have that (i) $\iff$ (ii) $\iff$ (iii) $\iff$ (iv)

Example Let $L = \{0^m 1^n \mid m, n \in \mathbb{N}, m \geq 0, n \geq 0\}$ be the regular language we considered before. Give a regular expression for L. $L = 0^* \circ 1^*$. Recall we previously showed this language is regular from the definition of a regular language, so solving this problem is a concrete illustration of the proof that (i) $\Rightarrow$ (iv).

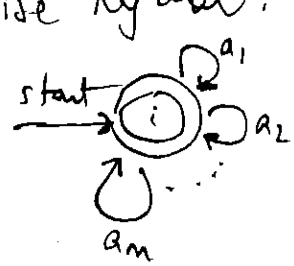
$\boxed{\text{The pumping lemma}}$

Task: Understand another criterion for figuring out when a language is regular.

Let a finite set A be the alphabet, and let L be a language over A. Then $L \subseteq A^*$. We make following two crucial observations

① If L is finite, then clearly there exists a finite state acceptor that recognizes $L \Rightarrow L$ is regular.

② If $L = A^*$, then L is likewise regular. Here is why: Let $A = \{a_1, \ldots, a_m\}$. The acceptor



with just one state i recognizes A^*.

Q: If L is infinite, but $L \subsetneq A^*$, how can we tell whether L is regular?

A: The Myhill-Nerode Theorem would have us look at equivalence classes of words, but that analysis can be complicated at times. The pumping lemma provides another way of checking whether L is regular.

The Pumping lemma: If L is a regular language, then there is a number p (the pumping length) where if w is any word in L of length at least p, then $w = xuy$ for words x, y, u satisfying

(1) $u \neq \varepsilon$ (i.e. $|u| > 0$, the length of u is positive)

(2) $|xu| \leq p$

(3) $xu^n y \in L \;\; \forall n \geq 0$.

Remark: p can be taken to equal the number of states of a deterministic finite state acceptor that recognizes L (we know such a finite state acceptor exists because L is regular). (3')

<u>Sketch of proof</u> The name of the lemma comes from the fact that if L is regular, then all of its words can be pumped through a finite state acceptor that recognizes L. We assume this acceptor is deterministic and has p states. We will show the pumping lemma is a consequence of the pigeonhole principle we studied in the unit on functions. If a word w has length l, then the finite state acceptor must process l pieces of information ($w = a_1 a_2 \cdots a_l$, where $a_k \in A$ $\forall k$, $1 \leq k \leq l$) $\Rightarrow$ it passes through $l+1$ states starting w/ the initial state. In the hypothesis of the lemma, we assume $|w| = l \geq p$,