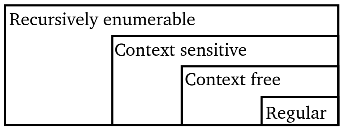
Formal languages

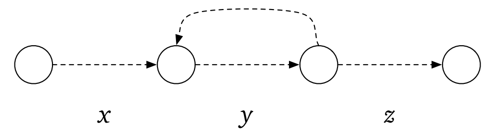
- Chomsky hierarchy



- Regular languages constitute a subset of possible formal languages
- A language is a regular language if and only if it can be described using a regular expression or an FSA
- It's easy to prove that a language is regular, harder to prove that it's not

Pumping Lemma

- The Pumping Lemma is a useful tool for showing that a language *isn't* regular
- The key intuition: any non-finite regular language must have a loop somewhere in its corresponding FSA



- The FSA accepts *xyz*, but must also accept *xyyz*, *xyyyz*, *xyyyyz*, etc., or xy^nz in general

Pumping Lemma

- **Pumping Lemma**. Let L be an infinite regular language. Then there are strings x, y, and z, such that $y \neq \epsilon$ and $xy^nz \in L$ for $n \geq 0$.
- Every regular language has some substring y that can be 'pumped'
- Non-regular languages may have strings that can be pumped: the pumping lemma is a necessary but not sufficient condition for showing a language is regular
- For example: a^nb^n
- Myhill–Nerode theorem

Pumping Lemma

- Is English syntax a regular language? (Partee et al. 1990; Chapter 16)
- Center embedding

  The cat likes tuna fish.
  The cat the dog chased likes tuna fish.
  The cat the dog the rat bit chased likes tuna fish.
  The cat the dog the rat the elephant admired bit chased likes tuna fish.

- These examples all have the form

  (*the* noun)^n (transitive verb)$^{n-1}$ *likes tuna fish.*

Pumping Lemma

- Take the intersection of the set of English sentences and the regular language */A* B* likes tuna fish/*
- This gives us the language:

  $L = x^n y^{n-1}$ *likes tuna fish,* $x \in A$, $y \in B$

  which is *not* regular (by the Pumping Lemma)
- Since regular languages are closed under intersection, English must not be a regular language