**Theory of Computation**

Natural languages consist of symbols, words constructed from these symbols and a set of rules as to how words are used.

In formal languages:

**Alphabet** - Formal nonempty set of symbols (generally lowercase letters)

**String** – Finite sequences of symbols from the alphabet set

**Concatenation** – Strings x and y are appending to each other (eg “ab” + “baab” = “abbaab”)

**Reverse** – Given a string x, the string of characters in reverse order of x

**Length** – Number of symbols in a string

**Empty string** – String of length 0, with no characters

If x is an empty string then

xw = wx = w

**Substring** – Given a string w, made up of 2 symbols u and v (w = uv), u and v are substrings of w. u is the prefix substring and v is the suffix substring.

If w is a string, w^n is the string obtained by concatenating w to itself n times. w^0 is the empty string, lambda.

If sigma is an alphabet,

**Sigma\*** is the set of all strings that can be constructed by concatenating 0 or more symbols from sigma. Sigma\* always contains lambda, the empty string.

**Sigma+** is the set of all strings that can be constructed by concatenating 1 or more symbols from sigma. It does not contain lambda.

Both of these are infinite sets

Sigma+ = sigma\* - {lambda}

**Language** – A subset of sigma\*. Any set of strings from sigma\* can be considered a language.

Ln = {|w| = n; w belongs to sigma\*} (all strings of length n from sigma\*)

**Complement of a language (L\_bar)** – Set of all strings in sigma\* that are not in L

L\_bar = sigma\* - L

**Reverse of a language (L^R)** – Set of all strings of a language L but reversed

L^R = {w^R: w belongs to L}

**Concatenation of languages** – Pairwise concatenation (all pairs) of strings of L1 and L2

L3 = L1L2 = {xy: x belongs to L1 and y belongs to L2}

L^n = {w^n: w belongs to L}

**Star closure of a language** – If L is a language, then the star closure of L, L\* is defined as

L\* = L^0 U L^1 U L^2 … (where L^i = L concatenated with itself i times)

**Positive Closure of a language** –

L+ = L^1 U L^2 U L^3 …

L+ = L\* - L^0

**Grammar**

Grammar is a language definition mechanism. It provides order to the words, so that the words together form sentences of a language.

In formal languages, grammar is given as a quadruple. If G is a grammar:

G = (V, T, S, P)

V = Set of objects called variables

T = Finite set of objects called terminal symbols

S = subset of V, called start symbol

P = final set of productions