African Journal of Physical Science, Volume 4, Number 2, 2011
Copyright© 2011 ISSN: 2141 – 0115 Devon Science Company

SYNTAX RECOGNITION IN SEMIGROUPS

D. Adeshola A. and T. Aliu
Department of Mathematics
Kwara State University, Malete, Kwara State, Nigeria

ABSTRACT

In this paper, we establish the fact that decimal integers are formed from the ten decimal or denary digits $0, \ldots, 9$; *English words are formed from the* 26 *letters of English alphabet (or* 52 *if we distinguish between lower and uppercase); English sentences are formed from the finite, but large number, of English words[2]. This preliminary study has attempted a description of a core aspect of both Syntax and Semigroup. It endeavors to succinctly explain what syntax entails and in chronological order, a review of major thoughts on syntax evaluation is highlighted. As a way of finding possible areas of relationship in the curious presence of Syntax and Semigroup , section one explains the relationship through logarithm algorithm, section two equally looks at the syntactic nature of prefix, suffix and subword, and in section three, results are analyzed and finally, conclusion are discussed in section four.*

Keywords: Syntax, Free Semigroups, Suffix and Prefix.

INTRODUCTION

Free Semigroups are first Mathematical objects every human being has to deal with. Even before attending school a child learns a "language". The child acquires "words" (i.e. generators of certain free semigroup) and "sentences" (i.e. sequences of generators) formed by Concatenation of "words". The early learning process relies very heavily on the fact that words and sentences carry meanings, the purpose of communication being the notation of objects, the expression of sensations and, later on, of thoughts and their recognition from or by other members of the social group. This aspect of language is called its semantic aspect. In systematic studies of languages, the semantic aspect of a language is usually distinguished from its syntax, that is, the set of rules leading to the formation or recognition of grammatically correct sentences (F. de Saussure, 1931). In other words, let S be a semigroup. A subset $X \subseteq S$ generates S freely, if $S = [X]_S$ and every mapping $\alpha_0 : X \to P$ (where P is any semigroup) can be extended to a homomorphism $\alpha : S \to P$ such that $\alpha \mid X = \alpha_0$. Here we say that α is a homomorphic extension of the mapping α_0. If S is freely generated by some subset, then S is a free semigroup [6].

Length and Logarithm

Given two positive reals x and y we have $\log(x \times y) = \log x + \log y$ and , similarly, given two words x and y we have $|xy| = |x| + |y|$. We have $\log(x \times 1) = \log x + \log 1 = \log x$, for a positive real x, and we have $|x\lambda| = |x| + |\lambda| = |x|$, for a word x. Note here that the empty word is the empty sequence and is denoted by λ, in other words, the empty word is a word over every alphabet [4].

Also, given a word x over an alphabet A, we define x^0 to be λ and for all $i \geq 1$, x^i to be xx^{i-1}. So $(12345)^0 = \lambda$, $(12345)^1$, $= 12345$, and $(12345)^2 = 1234512345$. Now $|x^i| = i.|x|$ in analogy with $\log(x^i) = i \,.\log x$ for reals. Hence, x^i is the ith power of x. For all words x, and all $i, j \geq 0$, $x^i x^j = x^{i+j} = x^j x^i$.

Theorem 1.1.1. : If a, x, and y are positive real numbers, and $a \neq 1$,
Then
$\log_a x + \log_a y = \log_a xy$
Proof

Let $\log_a x = m$ and $\log_a y = n$

then,

$x = a^m$ and $y = a^n$

but

$xy = (a^m)(a^n) = a^{m+n}$ (by indices)

Also by definition of logarithm, we have

$m + n = \log_a xy$

thus, $\log_a x + \log_a y = \log_a xy$.

Theorem 1.1.2.: If a and x are positive real numbers, and $a \neq 1$,

Then

$\log_a x^n = n \log_a x$

Proof

Let $\log_a x^n = m$

then,

$x = a^m$ so $x^n = (a^m)^n$ (by indices)

Also by definition of logarithm $mn = \log_a x^n$

Hence, $\log_a x^n = n \log_a x$.

Theorem 1.1.3.: Given logarithms base a of one equal zero

then

$\log_a 1 = 0$

Proof

Let $\log_a 1 = m$ then $a^m = 1$

$a^0 = 1$ (by indices)

We can thus write $a^m = a^0$ which gives $m = 0$ and so $\log_a 1 = 0$.

Prefix, Suffix and Subword

Given two words x and y over an alphabet A, we say that x equals y, written $x = y$, if they have same length and the same symbols at the same positions. We say x is a **prefix**[4] of y if there is a w z over A such that $xz = y$. For instance, let $x = world$ and $y = worldwide$, then x is a prefix y, since $xz = y$, where $z = wide$. In the same vein, we say x is a proper prefix of $x = \lambda, x \neq y$, and x is a prefix of y. Clearly, λ is always a prefix of any word x including itself a word x is always a prefix of itself. If $x = go$ and $y = goes$, then x is a **suffix**[4] of y and is, in fa proper suffix of y. In conclusion, we say that x is a **subword**[4] of y if there are words w and z $wxz = y$. We can visualize this given that $x = ma$ is a subword of $woman$ and it is a proper sub since $x \neq woman$ and $x \neq \lambda$.

Illustration

Consider the word $x = brother$. Then

$\lambda, b, br, bro, brot, broth, brothe, brother$

$\lambda, b, r, o, t, h, e, r, br, ro, ot, th, he, er, bro, rot, oth, the, her,$
$brot, roth, othe, ther, broth, rothe, other, brothe, rother, brother$
are all the subwords of x.

ANALYSIS AND DISCUSSION OF RESULTS

Words and its length

Let A be an alphabet. For all words, x, y, and z over A, $(xy)z = x(yz) = xyz$, then the catenation is associative. Also [5], let A be a non-empty set and let A^+ be the set of all finite, non-empty words $x_1x_2 \ldots x_m$ in the 'alphabet' A. A binary operation is defined on A^+ by juxtaposition

$(x_1x_2 \ldots x_m)(y_1y_2 \ldots y_n) = x_1x_2 \ldots y_mb_1y_2 \ldots y_n$

with respect to this operation. A^+ is a semigroup, called the free semigroup on the A. The set A is a generating set for A^+.

Given the alphabet $\{a, b, c, d, e\}$, the words

λ
dddab ce
dab
ceeab
aaaaa
dabbbcce
are words over $\{a, b, c, d, e\}$.

The length of a word x is denoted by $|x|$ and is simply the number of symbols in the given word, so $|13579| = 5$, $|hyskiema| = 8$, and $|\lambda| = 0$. Let x be a word over an alphabet A and let a be in A. Then the $a-length$ of x is denoted by $|x|_a$ and is the number of times a occurs in x. They have lengths

$|\lambda| = 0$
$|dddabce|_d = 3$
$|dab| = 3$
$|ceeab|_e = 2$
$|aaaaa| = 5$
$|dabbbcce|_b = 3$

It is important to note that the catenation of *dddabcedab* has length
$|dddabce| + |dab| = 3 + 3 = 6$ are expected.

Similarly, $(dab)^2(dab)^4 = (dab)^6 = dabdabdabdabdabdab$.

CONCLUSION

Elements of Free Semigroups are words in alphabet then they are strings of symbols in which we recognized that the strings that must be recognized in any application are constructed from set of symbols. These elements are strings of symbols and every input is an element of A^+, where A is the set of space and punctuation symbols together with both *upper* and *lower* case letters of the alphabet [3].

REFERENCES

Adeshola, A. D. and Aliu, T. (2009) Groups and Semigroups in Cryptography. *International Journal of Engineering and Technical Mathematics, Vol. 2, No 1 & 2, pp 152-158.*

Adeshola, A. D., Makanjuola, S. O. and Kehinde, R. (2008). Semigroups in Automata. *International Journal of Engineering and Technical Mathematics*, vol. I pp. 208-213.

Derick, W. (1987). *Theory of Computation*. John Wiley and Sons. Inc. Canada.

Howie, J. M. (1995). *Fundamentals of Semigroup Theory*. Clavendon Press, Oxford.

Tero, H. (1996). *Lecture Notes on Semigroups*. Department of Mathematics, University of Turku, Finland.