NEUROLOGICAL BASIS OF SENTENCE GENERATION*

K. H. KRISHNAMURTHY

Department of Biology,
Jawaharlal Institute of Postgraduate Medical Education and Research Pondicherry.

Summary: Even though a neurological basis of linguistic activity is conceded, the stages and steps of this activity atleast in the molar level are by no means clear. They are sought to be specified here with the help of the logistical and other models of sentence generation devised by the author. These are constructed on the depth hypothesis of a sentence viz. that the latter's visible structrue merely represents the final and the most superficial tier of a ramification process whose seat is in the depth of the brain below. To construct these models in a standardised manner, a notation for algebraic representation of the sentence is proposed. The logistical, matrix and other models designed, utilise the relevent informations known so far. Their main service is to suggest the functional and probably structural details of the underlying neuronic machinery. From this point of view, they seem to possess much promise, for instance, in suggesting atleast the limits of the cerebral tissues directly involved in an act of sentence generation. Insight into the generation of sentence is important, because, sentence is one of the significant and well recognised universals of languages and the processing of information in a sentence form is the crux of most thinking and speaking.

Key words: depth hypothesis formal analysis algebraic notation sentence generation matrix and logistical models

INTRODUCTION

The basic tenet of the modern discipline of biolinguistics is to hold that man's speech and language is but an offshoot of the same processes of organic evolution that prevails in the broader field of non-human biolagy. Nonetheless, a routine application of the scientific methodology of the latter to speech is beset with many difficulties, chief of which are: the immense complexity of speech even at the relatively simpler neuromuscular level and the unavoidable need to tackle with such ill understood, abstract phenomena as motivation, meaning, information and a host of others that become the central and integrated probelms in speech, unlike as in other areas of human physiology even, where they can be safely ignored. Apart from restating that the frontal lobe, recognised seat of man's intelligence, is involved here and that there are certain cortical areas specially designated as speech areas and of the two himispheres, one alone, mostly the left, is dominant for speech, neurologists can add very little to understanding the cellular matrix of speech. This is so despite some recent attempts at delineating even the molecular level of the processes involved (2,6,12). Trager's term neurolinguistics that aims at depicting the possible neruonic level of reactions i.e. the protoplasmic changes accompanying

* Presented in the XVII Annual Conference of the Association of Physiologists & Pharmacologists of India, March, 1972 at Trivandrum.

a linguistic act through it, is more an expression of our ambition rather than a sign of our achievement. The only anatomical differences (apart from the maturation schedule and the eventual size) between the human and non-human including the simian brain, are: the preponderent presence of the short rather than the long fibres in man connecting his primary cortical areas with their association areas and the excess development of his angular gyrus which is believed by some as towbe almost absent in the non-human brain (4,5,11).

It is mainly because of this, that even though it is conceded that the neurons of the cortex are the particular sites of cerebral speech activity, any further clarification in this regard, neurologically or neurophysiologically is difficult at present.

Another reason for this lacuna is the experimentational difficulty. For instance, the possibily applicable modern methods for this purpose such as the analysis of electroencephalographs, implanted electrodes, observation on the speech and language behaviour of the focal epilepts, study of dysphasic speech etc. appear to be too crude to look relevent even (See also 13). If for example, an elementary question is asked as to how does the brain process a sentence, the crux of speech behaviour, a neurologist can offer very little that is tangible. A deeper reason is that even a satisfactory theoretical analysis of what must be happening atleast in the molar level, at this juncture of the brain functioning has not been carried out. The logistical model (MM) of sentence generation developed by the author (9,10) is meant to clarify some aspects in this area precisely. We discuss below the construction and implications of a few other models also in this line.

MATERIALS AND METHODS

The modeling is developed on the analogy of the methods (for instance, the hypothetico-deductive method) common in the discipline of cybernetics (3). It rests on what is sometimes referred to as *depth hypothesis* (14) which presupposes that the visible i.e. surface structrue of a sentence is merely the last tier of a ramifying process whose details are worked out and elaborated below in the 'depths' of the brain. It attempts to picture the details of the development of the integration among the several components of a sentence; it is well known that any word in a sentence always indicates its role in the general information conveyance either by its place and/or morphenic shape; it never exists unconnected as in a vacum.

The advantages in selecting sentence generation for an intensive study like this are: much thinking and most talking is done in terms of sentences; sentence is an accepted universal of all languages; most importantly, the organisation of a sentence follows clearly ascertainable rules which are rigid and *repetitive* enough to justify postulation of a precise neuronic mechanism for the process. The models are conceived essentially as aids in investigation and clarification of the problem at present. They mainly utilise the tree from representation made popular from the works of Chomsky (1) and Yngve (14).

An algebraic notation: Empirically, a sentence is a succession of words which are semantically united in such a way that all of them together and in their entirety symbolise a single phenomenon *as if is actively being worked out* i.e. a sentence expresses a single phenomenon *operatively* (8). As such the minimum number of words in it are two, one signifying the *agent of work* and the other signifying the *act of work;* usually one more word occurs signifying the *product* or *goal of work*. These three (usually refered as the subject predicate and object) can be called the basic words x,y,z units of a sentence. There is no maximum for the possible number of words in a sentence though an optimal value usually prevails specially in spoken sentences. Despite this range, the three basic words x,y,z, are primary in the sentence context of the words and all the others are subsidiary. The latter can be called adjunct words. Each adjunct word is incomplete by itself i.e. it is clearly attached to any one the three primaries in whose coupling alone it will attain a completeness of meaning, We can thus say that its relation is always directed towards its own primary word. Their number may very for any primary word. They may precede or succeed and/or occur as seperated from their primaries by a foreign word i.e. a word whose relation is with an altogether different primary word. But their sign of relationship with the primary unit to which they are attached never varies i.e. this is always directed to the primary. Thus an x in xy sentence may have a. . . . n number of adjunets on either or both the sides or separated by a foreign(f) word but the sign (→) of them is always directed to itself. The notation for these cases would be $a \rightarrow n^{x}$, $x_{a} \leftarrow n$, $a — n^{x}\, a \leftarrow n$ and $a \rightarrow n^{fx}$.

Applying this simple notation we can represent any sentence algebraically. Such a representation is the first step to write our models in a standardised manner. We can now compute the basic types of sentences possible in *any* language and classify them on a criterion of increasing complexity. Three levels of such complexity can be recognised: (a) concerning the category of words composing a sentence i.e. whether the latter consists of only primary (xyz type) or primary and adjunct words ($xa..n^{yz}$ type)-*component level of complexity;* (b) when the order xyz of the basic units is changed (zyx type)-*order level* of *complexity* and (c) when the number of the phenomena the total sentence indicates worked out varies from one to more than one-*phenomenon level of complexity*. In each level we can also recognise subtypes brought about by minor variations. For instance in c, a sentence may indicate a signle phenomenon: $^{x}a..n^{yz}$ (the adjuncts present only for x) a simple sentence or one phenomenon completely and one or more other phenomena as worked out dependently with refernce to this: ($^{x}a..n^{yz}$) ($^{xy}a..n^{z}$) type-a complex sentence or more than one phenomena as worked out completely but joined together by a specific link word (a..n)+(a..n) type-*a compound sentence*. An advantage of such algebraic representation is that we can predict all the theoretical possiblities of senctence construction in advance. A preliminary attempt reveals that the total subtype possible in any language on any occasion is not more than 15. *A formal analysis and the models:* We shall carry out a detailed analysis and model making for only one sentence here. First, we shall attempt analysing its visible form, This is called *formal analysis* and consists of ascertaining the level of complexity as well as the total words in a sentence together with writing its algebraic representation.

The sentence selected is :

x y L x
A compelling assumption is that our image of the world, the seat of intellect and speech
(a n) (a.. .. ..
.. .. n) (a.) (a.. .. ..
is represented in a brain code preserved somehow in the circuitry of cerebral hemispheres
.. ..n)

This is a complex sentence and its total number of words is 30. Following the distribution of the super and subscripts in the sentence above, the algebraic representation would be $x_{a..n}{}^{yLz}(x_{a..n}{}^{y}\ a..n^{z}\ a..n)$. Here L is a word solely meant to link two major blocks (see the section on matrix model below). $z(=x_{a..n}{}^{y}\ a..n^{z}\ a..n)$ means that the z of the previous xyz pattern actually exists as ${}^{x}a..n^{y}\ a..n^{z}\ a..n$. Note that for the sake of simplicity even if the adjunct word is only one for a primary word (as in "is represented") we indicate it as a..n and we have also ignored variation in its place of occurnace with reference to the primary, always representing it as succeeding its primary (as in "a brain code preserved somehow" etc). *A Matrix Model* (*MM*)We now scan the words of the sentence complete and group them into *blocks*, each block consisting of one primary word along if it has no adjuncts (the first y of the Fig. 1) or one primary along with its adjuncts (the first x of the Fig. 1) if it has them. The adjuncts of any primary word may themselves be divisible into subblocks, each with its own (the second x of the Fig. 1) one main and other main words. In our scanning we recognise all these ultimate blocks first ("our image", "of the world" etc. of Fig. 1) and encircle

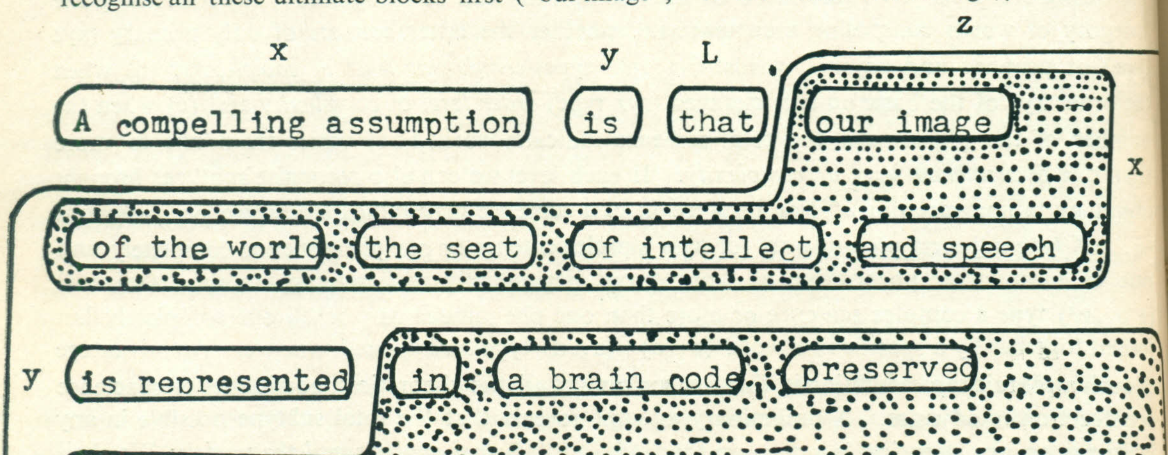

Fig. 1: Matirix model for a sentence viz. "A compelling assumption is that our image of the world, the seat of intellect and speech is represented in a brain code preserved somehow in the circuitry of cerebral hemispheres." This is the sentence for Fig. 2 and Fig 3 as well. The quantisable data here are: the presence of a single enclosure in the first x, y and L portion and a three tiered enclosure in the second z portion.

them in separate enclosures. All the subblocks or none primary unit are then enclosed in a single bigger envelope (the second x of Fig. 1). If the subblocks themselves display an xyz pattern i.e. they form a clause, the words are to be placed in three separate x,y,z blocks (as in the first x of Fig. 1). The resulting sketch of enclosures within enclosures (Fig. 1), the smallest representing the *ultimate working units* of words can be called a Matrix Model (MM) of a sentence since it reveals the grades of semantic relationship in a single plane. The hierarchy of this relationship is displayed extendedly in the logistical model (Fig. 2) constructed on the methods discussed elsewhere premilinarily (10).

Superficially many of the ultimate blocks of the MM are likely to be considered as equivalent to individual phrases. But the actual MM sketch would show that the cluster/in a brain code/for example which is often regarded as one phrase is split into two as/in/a brain code/. This is because in the internal contest of the words,/in/*can* stand as separate from/a brain code/, while the likelihood of/a/or/a brain/standing separate like this from /code/is very remote. In other words, it is simpler to imagine that the words of/a brain code/has a single immediate node of processing in terms of our LM while/in/can be resulting from a separate immediate node; our present concern is to recognise the *limits* of such possiblilities. It is in this light that each final block of the MM is considered as the ultimate working group of words. Grammatically,

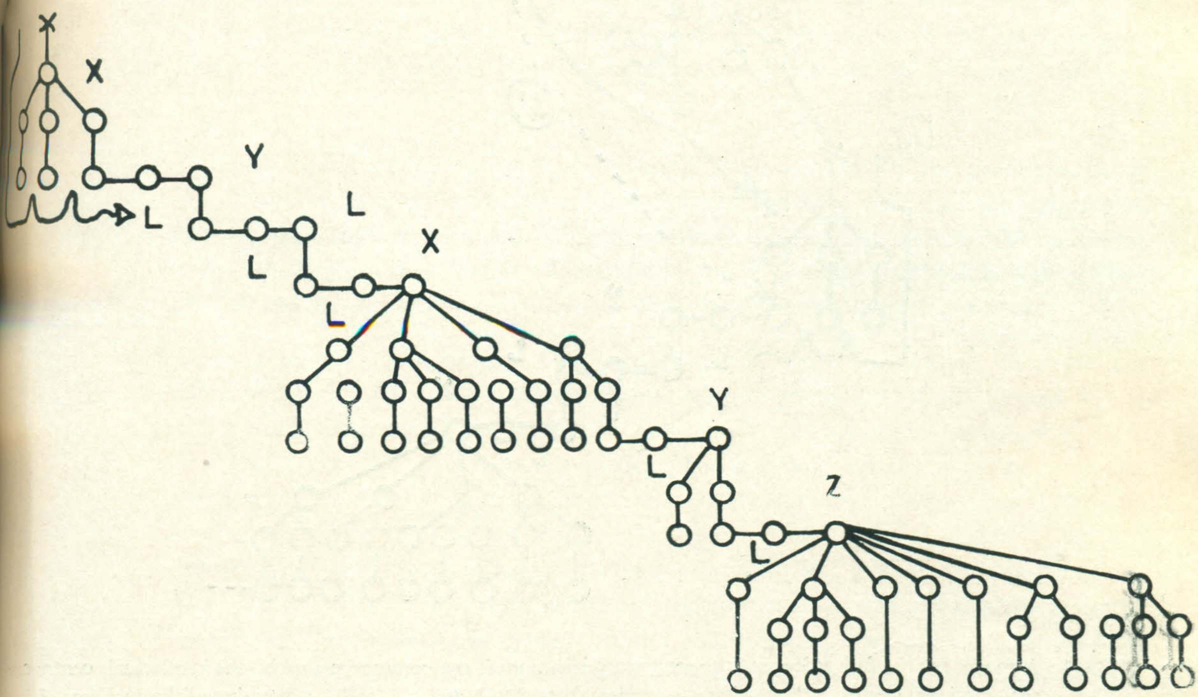

Fig. 2 : Blank logistical model of the sentence of Fig. 1. The quantisable data are : words 30, steps 78, words-steps ratio 1:2.6, links 5, levels 6. For further explanations, see the text.

this unit would be a single word that can stand independently (as in 'is', 'that') or, a group o words which is either a noun and its immediate ("our image") adjectivals or a verb and its im mediate grammatical parts ("is *represented*").

ANALYSIS

The implications of the matrix model are: 1) The words of a written sentence can b grouped into blocks within one another structurally. The degree of this enclosure reflects th intensity of their semantic relationship which is unity or closest and immediate in the smalles enclosure and weakest and more removed between the largest enclosures. Because of th latter reason, we may hypothesise that a special effort is necessary to link them and the sentenc chain is therefore weakest here and is likely to break down during the relevant disorder o speech, for instance, in laconic speech. A specific step for linking alone can thus be *postulate*

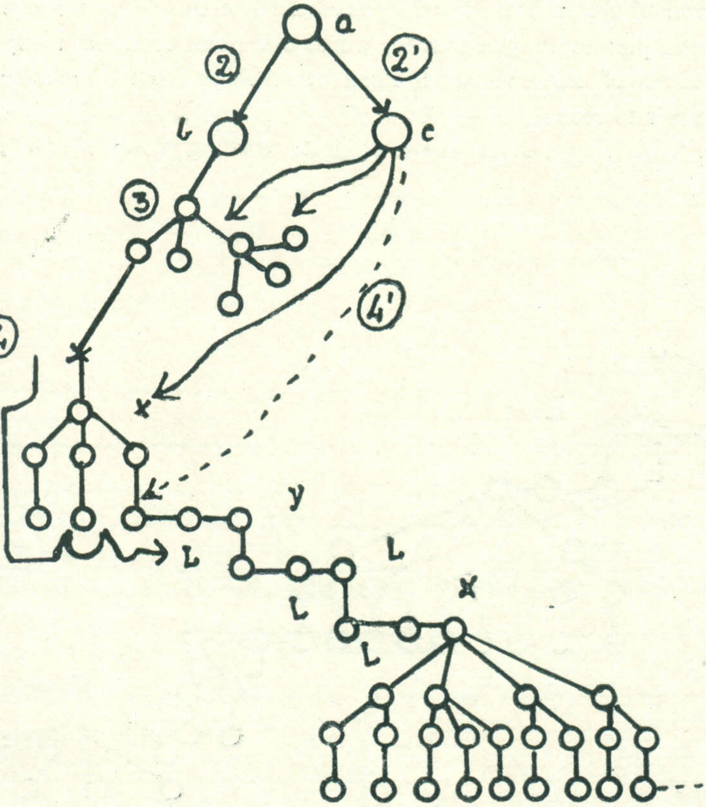

Fig. 3 : A model for the full process of the sentence generation a, the common centre b, the intellectual centre c, the emotional centre. 2,2', simultaneous orders from a to b and c. 3 and 4, generation of the sentence, 3 in the primary, 4 in the secondary plane. 3′ motor orders from to prepenultimate stages of the sentence. 4′ motor orders from c to influence penultimate stages of the sentence. For further explanations see the text.—indicates that the rest of the sketch is the same as in Fig. 2.

as *always* occuring between the basic blocks x and y and y and z; this may often be verbalised by a definite word (as in 'that of the MM Fig. 1). We may also note that each of these basic blocks is semantically stabler, more cohesive and earlier than the links. Moreover the need for such links would become very keen when the two adjoining blocks are large i.e. they contain may words. This postulated concept of link is incorporated appropriately in the LM. 2) The component words of any enclosure, smaller or bigger, have a common orign i.e. they are the elabroations (viz. the context or .elation indicators) of any one word among themselves. Apparently this particular word forms a *minor node* for attention focussing during the processing of the remaining date of that block. (3) Since x,y,z, are the *primary operatives,* all the words of a sentence are to be grouped under these largest enclosures finally, each representing a *major node* or a level of processing and hence requiring considerable time (plus an effort to link with next level). It is these ideas that are elaborated in the LM.

LM is not structured in a single plane; it consists of elements which we call as steps. These are distributed at different levels, each level corresponding to x,y, or z block or to a link step when that step is indicated by a specific word. If the numbering and direction signs of the complete LM are also taken into account, the model would show that all these steps are oriented in a single direction i.e. towards the phonation of the last word The model thus depicts a single continuous process which is how the actual verbalisation of the sentence takes place, the process normally not stepping till the last word is uttered. The course of the arrow however is tortuous (See Fig. 2) as it includes more than one descent and ascent whenever our model forks.

The heuristic value of the above analysis would become clear by the following deductions that we can make with the help and in the light of the LM, MM and the expressions made possible by the notation adapted. The deductions refer to two aspects: (a) a description of how does the brain works during the generation of a sentence i.e. the details of the processes concerned and (b) an exploration of the neurological site of these processes.

(a) *The Process of sentence generation:* This process is not single or sudden. It is gradual, stepwise, time-consuming and includes several stages: planes, levels links and steps (see Fig.2). It is hierarchial and cybernetic i.e. controlled by feed back throughout its progression. The process can be said to be taking place at two places, the *primary* and the *secondary* because: (1) In the information conveyance by the sentence, the elements of the primary plane of generating the basic x,y, and z units are more important than those of the secondary plane. In the mature sentence these units can stand by themselves and do not require the presence of the elements of the latter; they form the key words of the information conveyed, while the latter constitute the qualifications and/or relationals only of the former. They are more deep seated, probably more permanent and constitute the basic referents for the latter. (2) it is in view of this great difference in significance that we have postulated the existence of such a generation in two planes, the x,y,z, units in the first or the primary plane and the actual sentence chain in the later or the secondary plane. During any length of the sentence which may be often extensive, the elements x,y,z, of the primary plane are not forgotten. This can be easily accomplished if they are generated first and kept ready for reference when their time comes. In an adjunct containing sentence say $^{x}a..n^{yz}$ a..n the primary plane is not followed by phonation immediately (as is so in the sentence we have selected, Fig. 2) and represents an aspect of internal speech. The differentiation thus takes place in three stages: x (when the focal idea for the whole sentence is generated) xyz(when this focal idea is put is an operative way and the basic units are fixed) and the a..n (when the adjuncts for the zyx are elaborated).

Empiracally we are forced to presume that under normal circumstances the process of sentence generation is to be represented as a single chain of events, because from the time the concept of the first x arises till the last word is phonated, it does not stop. Whatever reckoning points (plancs, levels, links and steps) we can discern here, they are all to be regarded as different stages and aspects of a single process,. The LM infact is a picture of the temporal development of the sentence while MM is a dissected appearance of the matured structure. x may be conceived as to be arising as a flash i.e. at a single step. This is single, simple and without any parts. After generating this, attention is focussed on it continuously till the last word is phonated. During this time, numerous words arise which in the context of the sentence are all subordinate to x. The structural contents of a sentence are (specially as revealed by the MM and from superficial to deeper levels) : words, ultimate working groups of words, x,y,z, blocks and x. The developmental scheme of a sentence consists of (specially as reveald by the LM and from early to later stages): x, primary plane differentiation of x into xyz and adjunct differentiation of x,y, and z. Ontogenetically any sentence is thus an elaboration of x to final words through the xyz blocks as a single process and across the two planes and the requistie levels, links and steps. In terms of its contents, sentence generation is a recall of the words from the memory store, the words recalled thus are not unique, except in rare circumstances. But any natuarl sentence is always unique viz. it always involves a new ordering of the words. The speciallity of the sentence is that by utilising the known words and the known rules, it makes a new statement every time. To identify these known rules, a classification of the words and a consideration of the methods of their alteration in the sentence context become necessary.

(b) *The neurological site of the processes :* Any natural spoken sentence is always accompanied by emotion. This is expressed phenomically by the choice of particular individual words and/or word orders and non-phenomically by cries, laughter and prosodic features such as stress, accent pattern rate of phenome production pauses and the like. The motor orders for these are therefore to be superimposed on the sentence getting generated at the appropriate levels viz. those for the choice of the words and the word orders at the deeper levels and those for the nonphenomic phonatory features at the penultimate level (Fig. 2) It should also be noted that the order for firing the intellectual as well as the emotive behaviour which together is indicated by a spoken sentence must be coming from a common centre as both have to serve the same purpose viz. the psycholgical need for the organism. To incorporate this provision also in our model, we can extend our sketch as in Fig. 3. The actual neurological processes for a sentence generation would now be: (1) The activation of the common centre a to attempt sentence expression (as a means for fulfilling the organismic need) (2) Simultaneous orders from *a* to activate the intellectual and the emotive centres *b* and *c*. (3) and (4). The generation of the sentence, 3 in the primary plane and 4 in the secondary plane. 3) The motor orders from *c* to influence the deeper, mainly prepenultimate stages of the sentence. (4) The motor orders from *c* to influence the penultimate stages of the sentence. The quantitate data we get thus are: there are 3 centres a, b, and c whose activation and function we have to study: the total processes are: the activation and the relevant firing 2 and 2′ by a, the activation of b and c and the relevant firing by then of 3,4 and 3' and 4' respectively.

In exploring the neurological *site* of these processes we note that the centre *a* which triggers the psychological need should be the place where the general psyche is located viz. its seat is generalised in the brain. This is becuase, motivation, personality and similar 'gestalt' features are to be reckoned with here: an indication as to the role of the brain as a whole, specially its cortex of both the hemispheres, and not one, in the process of speech. The centre *b* is also to be located in the cortex of both the hemispheres, which by evolution is a seat of elaboration of input and out put; an important aspect of intellect. If this is so, it becomes a question as to what specific role does the dominant sphere alone play apart from the obvious one of sending motor orders from its Broca's area i.e. taking part in the penultimate stage of the sentence generation viz. the last tier of the LM. Beacuse of the specialisations in *a* and *b* the organism is capable of reacting to a stimulus by symbolisation. It is at this stage, the seat of memory is very much indispensable, because communicative symbolisation means encoding the symbol as a word or gesture taken from the permanent memory bank. Since evidence exists that such memory is stored in duplicate traces in both the hemisphers, particularly in the hippocampal region, the seat of long term memory, we should add these relevant sites to the list of our centres to be examined. Finally, *c* is presumed to be in the brain stem and the limbic girdle.

This picture represents the minimum extent of the centres whose activity, triggering and integration should form our subject matter. If we confine to *b* only where actually the sentence generation takes place, it is clear that apart from the encoding of the word symbols, the linking of these words in a single chain which is the crux of the secondary plane activity of the LM seems to be a function of *immediate memory span* and a question of feed back control at every step. It is also obvious that in this only, some cells of the cortex, not *all* are engaged concertedly viz. they *alone* are activated and also inhibited from noise and work in unision. The problem now is to locate these cells. This picture does not militate against Penfield's recognition of three cortical speech areas of the dominant hemisphere (as being of varying degree of significane to speech or rather aphasia); for, Penfield himself means "ideation" (i.e. the process of matching an idea with its appropriate word or, coupling the concept and the word) only here by his term speech. However, speech is not mere ideation for instance, syntactic ordering is not conveyed by this term of ideation. It is clear therefore that in deciding upon the neurologica of the full speech process, we should keep an extensive area of the brain tissue of both the h pheres in mind.

DISCUSSION

The value of the algebraic representation, the model construction and the prelim analysis presented above lies in directing our thinking on lines that may yield us a greate sight into the problems involved regarding the generation of a sentence in the brain. S features of two such aspects viz. the process of sentence generation and the extent of neurological sites involved, are pointed out. Other future possibilities of this line of w if carried out with the necessary rigour, are: algebraic representation of diverse sentence t in different languages to evaluate language differences on a common platform, an adequate onomy of the words, further elaborations of the models so as to secure more details of the sent generation process and an erection of a flow chart of the events involved in the brain durin sentence generation as well as fixing up of at least the limits of the machinery involved. Ne can we ignore the significance of the various quantisable data (e.g. word-step ratio) mentio above. Further work is under progress in these directions.

REFERENCES

1. Chomsky, N. *Syntactic Structures.* The Hague:Mouton, 1957.
2. Gaito, J. (Ed.) *Macromolecules and Behaviour.* Amsterdam: North Holland, 1966.
3. George, F.H. *The Brain as a Computer.* Pergamon Press. Chapter III 1961.
4. Geschwind, N. The development of the brain and the evolution of language. *George-town, Univer; Mo Ser. Lang. Ling.*, 17 : 155-169, 1964.
5. Geschwind, N. Disconnexion syndromes in animals and man. *Brain*, 88: 237-294 and 585-644, 1965.
6. Gurowitz, E.M. *Molecular Basis of Memory*, Englewood Cliffs. N.J. Prentice Hall, 1969.
7. Krishnamurthy, K.H. Psycholinguistic study of a Schizophrene's speech. *Language and Speech.* 256-257, 1969.
8. Krishnamurthy, K.H. Biolinguistics and the study of Language Disorders. *Proceedings of the Inter ciplinary Seminar on Linguistics and other Sciences*, Annamalai Nagar, India 1970.
9. Krishnamurthy, K.H. Some methods of analysing Speech and Language Samples. Presented at the W shop of Speech and Language Analysis in Indian Languages at the All India Institute of Speech and H ring, Mysore-5; V.R.A. February 1971 (Profect. Ind. 38-68).
10. Krishnamurthy, K.H. A New approch to the Science of Language. Special Lecture Series. Prasa ga, University of Mysore, 1971.
11. Ojemann, R.G. Correlations between specific brain lesions and memory changes. *Neurosciences Proog., Bull.*, 2 : 77-144, 1964.
12. Schmitt, F.O. Molecular and Ultrastructural Correlates of Function in Neurons, Neuronal Nets, And Brain. In *Neurosciences Research Symposium Summaries* Vol. 1 M.I.T. Press Massachusetts. 323-351, 1
13. Wooldridge D.E. *The Machinery of the Brain*, McGraw Hill Book Co., 1963.
14. Yngve, VH. A model and an hypothesis for language structure. *Proc. Am. Phil. Sec.*, 104:444-466, 19