

The P-structure C-structure Grammar (PCG) for the contrastive study of two or more languages

P. C. GANESHSUNDARAM

Foreign Languages Section, Indian Institute of Science, Bangalore 560 012

Received on September 29, 1977

Abstract

In this paper a basic philosophical view on language and a theory of language structure based on this view are outlined.

It is taken as a postulate that all languages have the same syntactic structure and so must be amenable to representation by a single set of sentence derivational rules.

Key words : Linguistic Theory, Language Structure, Contrastive Linguistics, Universal Grammar, Translation.

Introduction

1. *What is Contrastive Linguistics?*

Contrastive Linguistics is the discipline that attempts to study different language structures in relation to one another and, in the case of the P-structure C-structure Grammar (PCG), in relation to a common theoretical system of structures.

This kind of study leads us to the following, namely:

- (1) to better language teaching materials of one language intended for learners knowing another language, by systematically comparing and contrasting structures in them in a graded way;
- (2) to reference books that help recognise different structures for technical translation in terms of equivalent structures.
- (3) to mechanization of a greater or smaller part of language analysis and translation;
- (4) to mechanical compilation of specialized glossaries of various kinds: monolingual, bilingual or multilingual.

2. *What is the relevance of Contrastive Linguistics to science and engineering?*

- (a) Information processing is a branch of communication science or engineering, that deals with various types of information transfer, such as:
 - (i) optical,
 - (ii) acoustical, etc.,

To these could be added

- (iii) linguistic information processing, which among other things also deals with:
- (1) transfer of scientific information through a natural language, and
 - (2) transfer of scientific information across natural language boundaries (through translation).
- (b) All information (including scientific information) is communicated through the use of a natural language (that is, directly in any single language like English or any other, whether Indian or foreign).
- (c) Scientific information, as has just been noted, is frequently to be transferred (through translation) from a foreign language into English or from English into an Indian language.

All these could be processed systematically and mechanized step by step to different degrees.

3. *Mechanised, semi-mechanised or marginal'y mechanised translation*

Since

- (i) Information processing is a subject of scientific and engineering study, and
- (ii) translation (mechanised or otherwise) is information processing across language boundaries—and hence is also such a subject—, it may be seen that:
 - (a) when compared with the normal process of translation by a human translator, mechanised information processing and mechanised, semi-mechanised or even marginally mechanised translation are still more legitimately purely scientific and engineering subjects of investigation.
 - (b) information retrieval in a library also has to be done through language (or a sub-set language of linguistic symbolism). If this is to be partially or fully mechanised, we have again an engineering problem based on language processing.

4. *Theoretical framework available for language processing*

Theories of language structure now available for the description and analysis of linguistic structures are:

- (a) Chomsky's Transformational Generative Grammar (TGG),
- (b) The Neo-Firthian or Hallidayan Systemic Grammar (SG),
- (c) Fillmorean Case-Role Grammar (CRG) and
- (d) The P-structure C-structure Grammar (PCG).

5. Chomsky's Transformational Generative Grammar (TGG)

This grammar:

- (1) generates, according to rules, grammatical sentences based on certain deep-structure elements;
- (2) in addition to the generation of basic types of structures, many derived types are generated according to transformation rules for all possible grammatical sentences in a given language.

However :

- (1) It deals with only one language grammar at a time.
- (2) It divides the sentence (as per Western traditional grammar) *arbitrarily* into a subject and a predicate.
- (3) The predicate includes not only the verb but also all its objects and complements.
- (4) Its handling of complex sentence patterns is involved and cumbersome.
- (5) It cannot deal with two languages at the same time.
- (6) It appears to maintain that a grammar is rigid. Flexibility is ruled out.
- (7) It does not allow for equally tenable alternative structural analyses as part of *one* grammatical description.
- (8) It does not deal with verbless sentences, such cases being kept out of consideration.
- (9) The metalanguage used for the description of structures does not have unambiguous symbols for grouping of elements.

6. The Neo-Firthian or Hailidayan Systemic Grammar (SG)

- (1) It treats the predicate as representing only the verb, the sentence being made up of the elements S, P, C and A (subject, predicate, complement and adjunct).
- (2) It accounts for clauses serving as lower order elements (nominal group, etc.) by the concept of *rank* and the phenomenon of *rank-shift*.
- (3) It recognises *one word* and *verbless* sentences.
- (4) It accounts for the relation of subordinate clauses to the main clause in a simple way.
- (5) It recognises the possibility of giving several possible interpretations to any given structure.

However:

- (1) It does not have a symbolism to allow for alternative groupings of subordinate structures.

For example, ambiguous constructions like: 'He asked his children to do their homework when they went home' are not clearly demarcated in the symbolism used for the dependence relationship.

- (2) It also deals with only one language at a time, although translation questions are kept in view.
- (3) It cannot be handled in an automatic way, as the restricted metalinguistic symbols need human agency for interpretation.
- (4) It cannot adequately handle highly complex sentences.

7. *Fillmorean Case-Role Grammar (CRG)*

- (1) It deals with deep-structure logical relationships (irrespective of surface level voice, etc.).
- (2) It could, therefore, deal with the deep structures of several languages at a time.

However :

- (3) There is no simple formal symbolical system available in it for direct use in dealing with more than one language at a time.

8. *The need for a new theory*

In order to do away with the limitations listed above in §§ 5, 6 and 7, and to permit automatic handling of elements, a new theory is needed.

9. *The P-Structure C-Structure Grammar (PCG)*

- (1) This theory combines the important features of TGG, SG and CRG.
- (2) It could be handled with surface level elements themselves.
- (3) Transformations can be effected by algebraic rules.
- (4) It deals with several languages at the same time.
- (5) It can handle complex sentences of *any* degree or kind of complexity, as its inner structures are all of the same pattern as the outermost structure, namely the sentence itself, all of them being seen as telescopic structures, one within another.
- (6) It permits the definition of parts of speech in terms of the verb, which is taken to be a primitive self-evident entity.
- (7) It can handle all kinds of conjunct verbs and conjunct auxiliaries functioning as single units.
- (8) Combination of sentences can be dealt with in an algebraic way, taking common factors, etc.
- (9) It can give structural descriptions of the various possible interpretations of even apparently straight-forward sentences, like:
'He saw her peeping through the window.'

- (10) Without invoking rank-shift arguments as in SC, it can deal with clauses functioning as *nominals*, *adverbials*, etc., in terms of the telescopic structures and through the use of the concepts of *actual* and *virtual* parts of speech.
- (11) It can deal with any number of levels of ~~are~~ linguistic structures.
- (12) It is simple: the simplicity of the telescopic structures could lead to automatic parsing of sentences using the computer, with the help of a list of markers, auxiliaries, etc., in given languages.
- (13) It deals with the grammar of any one or more languages at a time, treating *morphology*, *syntax* and *lexicon* as interdependent areas of grammar that are flexible with respect to one another in an integrated manner.
- (14) It deals with the universal hierarchical structures and the language specific linearity of sentences in one representation, simultaneously handling deep and surface structures in one representation.

However :

- (1) It is still in its preliminary stages of development.
- (2) It has further scope for improvement towards greater precision.

10. Examples of ambiguous structures

Ambiguity is not a marginal phenomenon, but is part and parcel of any natural language. No language could be absolutely precise about any statement made in it about any given situation described by it. Even Sanskrit, with all its grandiose morphological machinery is ambiguous, for example, in a structure like the following:

‘Tasya grhasya samiipe.’

This could mean, ‘near *that* house’ or ‘near *his* house’.

We shall however give examples only from English.

(a) They are flying planes.

The ambiguity in this sentence is resolved by the hierarchical-linear ~~demarcations~~ according to PCG to get the two different meanings:

(i) ((They) are (flying planes))

and

(ii) ((They) are flying (planes))

(b) The children shrink from washing

is, according to PCG :

- (i) ((The children) shrink (+ from washing))
- (ii) ((The children) (shrink (from)) (washing))

(c) He saw her peeping through the window.

This has the PCG structures :

- (i) ((He) saw (her) ((peeping (+ through the window))))).
- (ii) ((He) saw ((her) ((peeping (+ through the window))))).

In addition to this structure, the morphological status of *her* tells us whether it is the qualifier or the qualified. Accordingly we have two interpretations.

- (iii) ((He) saw ((her) ((peeping))) (+ through the window)).

Here too we get two interpretations depending upon the morphological status of *her*.

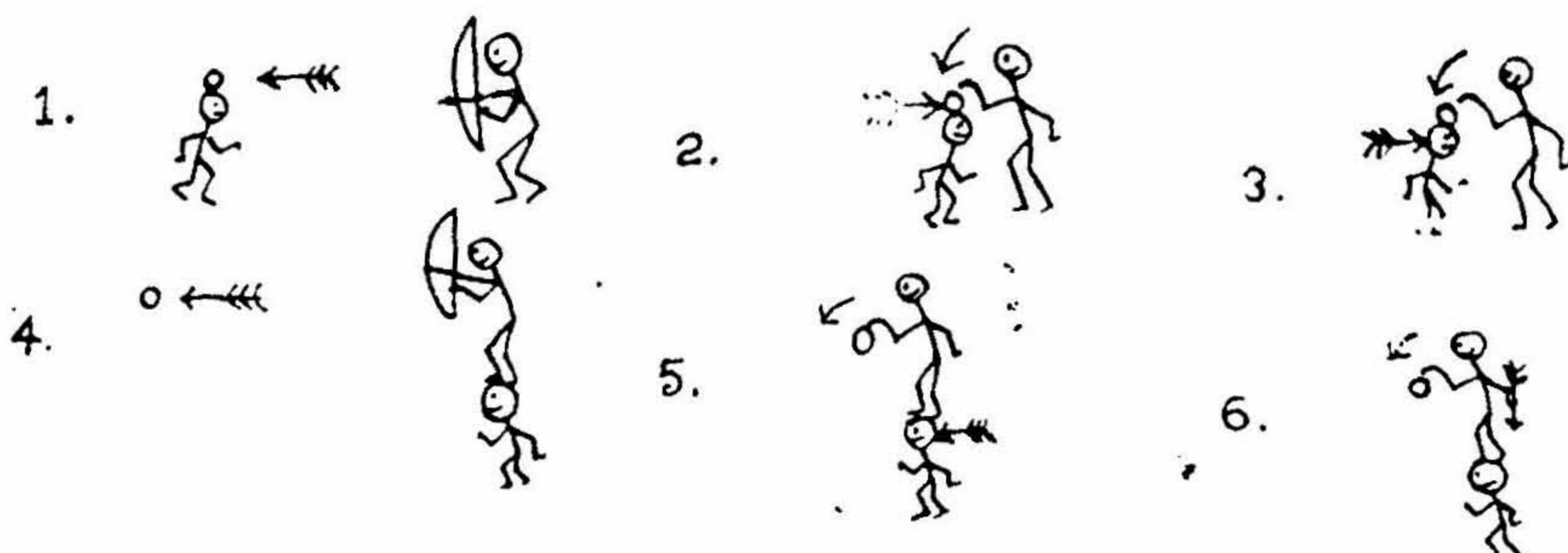
In the following example, it is seen that intonation in speech or punctuation in writing makes all the difference, although the linear succession of words is the same:

- (d) (i) What ! Do you think I will shave you for nothing and give you a drink ?
- (ii) What do you think ? I will shave you for nothing and give you a drink !

The following sentence has six interpretations, as shown by the PCG demarcations given below:

(e) William Tell hit an apple standing on his son's head with an arrow.

- (i) ((William Tell) hit (an apple ((standing (on his son's head)))) (with an arrow))
- (ii) ((William Tell) hit ((an apple ((standing (on his son's head)))) (with an arrow)))
- (iii) ((William Tell) hit (an apple ((standing (on his son's head (with an arrow))))))
- (iv) ((William Tell) hit (an apple) ((standing (on his son's head))) (with an arrow))
- (v) ((William Tell) hit (an apple) ((standing (on his son's head (with an arrow)))))
- (vi) ((William Tell) hit (an apple) ((standing (on his son's head) (with an arrow))))



We see from the above that every sentence in any language is actually or potentially ambiguous in any context or absence of it. That is why even the so-called 'carefully worded' legal documents could be interpreted differently by clever lawyers, who mint money not so much because there is a dispute but because there is ambiguity in language structure.

11. Automatic Parsing of English

If we take ambiguity as a general phenomenon and account for it in an explicit and formal way, it is quite possible to resolve the ambiguity, step by step, by mechanical means.

For this purpose we must have a checklist or lists of endings, auxiliaries, etc., to give us clues for alternative demarcations, often leading to a unique demarcation, when there is least ambiguity.

For example in the following three sentences, a correct demarcation depends on the interpretation of *is* :

- (1) He is here
- (2) He is coming
- (3) He is killed.

Pronouns are listed and so *He* is known. Auxiliaries are listed, but *is* is not unique. It could be a link verb, the auxiliary for the continuous tense or the auxiliary for the passive voice. Thus, if endings are listed and could be located, *-ing* and *-ed* could help to identify *is* as respectively the continuous or passive auxiliary. Otherwise *is* is a link verb. If in addition certain very common and frequently used adverbs and adjectives are also listed, *here* comes out to be an adverb. Since the sentence contains no more elements, the first sentence has a link verb. In this way we could identify the elements of these simple sentences without any ambiguity and arrive at the following PCG demarcations :

- (1') ((He) is (here))
- (2') ((He) is coming)
- (3') ((He) is killed).

where *is coming* and *is killed* are verb phrases.

Once the English sentence is grammatically analysed and demarcated, whether fully automatically or by a pre-editor, the stage is set for an automatic translation into another language, say Hindi.

12. *Automatic translation from English into Hindi*

If we have, in a mechanically retrievable form:

- (1) Check lists of various kinds
- (2) Grammatical tags of various kinds that could be mechanically added on to words on the basis of the check lists,

and

- (3) PCG demarcations, made either mechanically or by a pre-editor,

then by following the eight stages given below, we could mechanically translate from one language into another straight-forward sentences that are used to communicate scientific matter-of-fact information and logical arguments.

(It goes without saying that we exclude from our consideration all kinds of ornamental writing found in literary works of poetry or prose.)

The eight main stages for automatic translation from English into Hindi are given below (In each stage there may be many sub-routines):

Stage 1

Consultation of check lists and automatic or manual demarcation of the English text in terms of the PCG structures. (If this is done automatically many alternatives may come out as the outputs, of which one should be chosen by the pre-editor and given back as the input for the next stage.)

Stage 2

Machine processing of the demarcated English text, adding grammatical tags.

Stage 3

Substitutions from a machine readable English-Hindi dictionary.

Stage 4

Transposition of syntactic PCG brackets, obtaining an initial Hindi version with grammatical tags.

Stage 5

Grammatical processing of the initial Hindi version, obtaining an intermediate Hindi version.

Stage 6

Hindi substitutions for the grammatical tags by table look-up, obtaining a pre-final Hindi version with syntactic PCG brackets.

Stage 7

Dropping of brackets and getting the final version as a machine output.

Stage 8

Polishing the final version by a human post-editor.

We shall see only two simple examples below. We are not giving here the nature of the check lists or details about the grammatical tags.

We indicate in capitals all stages that pass through the machine from the input stage to the output stage.

*Example 1**Original Text*

He is my brother.

Stage 1:

((HE) IS (MY BROTHER)).

Stage 2:

((HE) BE *PRS3 (MY BROTHER)).

Stage 3:

((VAH + *PR3M) HO *VK1 *PRS3 (MER *TP1 BHAAII *NI)).

Stage 4:

((VAH + *PR3M) (MER *TP1 BHAAII *NI) HO *VK1 *PRS3).

Stage 5:

((VAH + *PR3M) (MER *TP1 *NI BHAAII *NI) HO *VK1 *PRS3 *PR3M).

Stage 6:

((VAH) (MER AA BHAAII) H AI).

Stage 7:

VAH MER AA BHAAII H AI.

Stage 8:

Vah meraa bhaaii hai.

Example 2:

Original Text :

She wants to go home.

Stage 1:

((SHE) WANTS (TO (GO (HOME))))).

Stage 2:

((SHE) WANT *PRS3 (TO (GO *INF (HOME))))).

Stage 3:

((VAH+ *PR3F) CAAH *VMI *PRS3 (*OB KO (JAA *VI4 N+ (GHAR *N1 *OB KO))))).

Stage 4:

((VAH+ *PR3F) (((GHAR *N1 *OB KO) JAA *VI4 N+) *OB KO) CAAH *VMI *PRS3).

(At this stage, if we do not want to take into account the effect of the element *OB KO, we could arrive at a final —shall we say—Muslim version of Hindi, such as:

Vah ghar ko jaane ko caahtii hai.

But we want a more polished secular Hindi and so we must take into account the effect of the element *OB KO, as shown below).

Stage 5:

((VAH *PR3F) (((GHAR *N1 *OB KO *VI4) JAA *VI4 N+) *OB KO *VMI) CAAH *VMI T+ *PR3F H+ *PR3F).

Stage 6:

((VAH) (((GHAR) JAA N) AA) CAAH T II H AI).

Stage 7:

VAH GHAR JAA N AA CAAH T II H AI.

Stage 8:

Vah ghar jaanaa caahtii hai.

The last stage could also be carried out mechanically to remove the gaps between stems and endings, leaving only stylistic polishing, if any, to the post-editor.

13. *Syntactical vs. morphological parts of speech*

A 'part of speech' is basically a syntactic unit, but we do find one-word members of such parts of speech, labelled as such at the morphological level. A noun in the nominative case, an adjective or an adverb could be both morphologically and syntactically represented by one and the same lexical element.

However, it is very rarely realised, conditioned as we are by traditional modes of thinking, that a very large structure consisting of several words could just be syntactically equivalent to a single word and they both belong, in a given sentence, to the same syntactic part of speech.

For example, the words underlined in the following two sentences belong to the same syntactic part of speech:

- (1) He went home.
- (2) He went to the one place on earth, where he could find comfort, joy and all that a man needs to forget the troubles of other people that he has to worry about at his office.

We are never taught to recognise these as equivalent, for, conditioned by traditional grammar, we have been trained to look at syntactic units through 'morphological thinking' and not from a syntactic angle.

In example (1) above, *home* is an *adverb* and so is the larger construction (indicating a *location* in the *direction* of which some movement takes place).

However, *home* is also considered to be a noun, which is true only in such cases as the words underlined in the following:

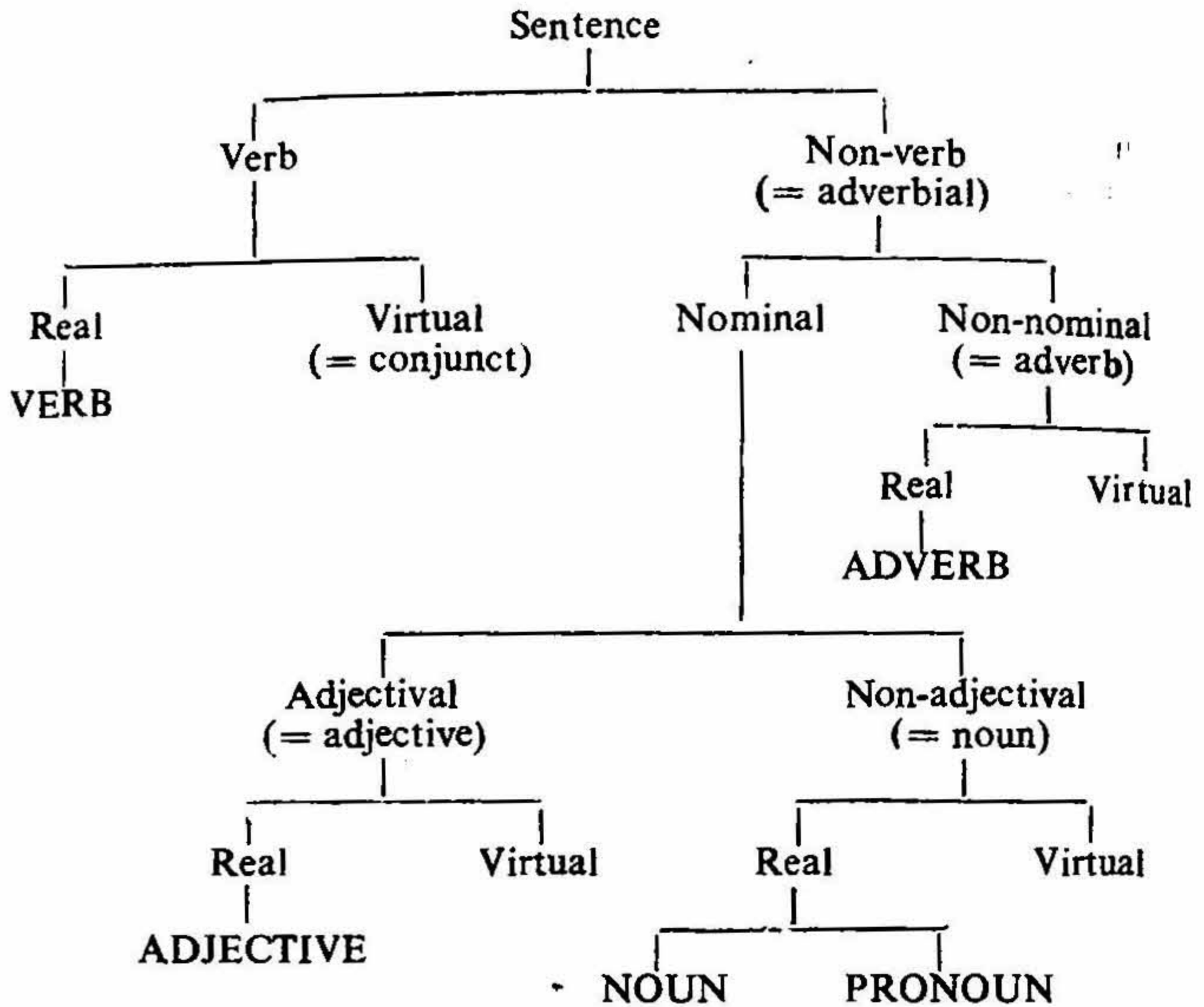
- (3) He likes home.
- (4) He likes the one place on earth ... at the office.

When *home* is an *adverb*, it is a *real* adverb in the sentence, for it consists of one word and is morphologically and syntactically the 'same' part of speech.

When *home* is *noun* too it is a *real* noun on the same grounds.

However, the equivalent larger constructions are respectively adverb and noun in the two sentences (2) and (4) only in a syntactic way. They are composed of a large number of different morphological parts of speech. Therefore, they are, respectively, *virtual adverb* and *virtual noun*.

A sentence consists of one verb and a number of non-verbs. Accordingly, the syntactic parts of speech in terms of the membership of any word or set of words in a sentence could be shown schematically as in the following diagram.



Apart from the fact that larger constructions could function as virtual parts of speech in a sentence, it must be noted that any morphological part of speech could serve as any other *virtual* syntactic part of speech.

For example, in the sentence:

(5) 'They *John* me *all the time*. But my name is Jones,' *John*, a morphological noun, is a *virtual verb*, and *all the time*, a noun phrase, is a *virtual adverb*.

In PCG demarcation, sentence (5) would be :

(5') ((They) ((John)) (me) (all the time)). ((But) (my name) is (Jones)),
where ((John)) is a conjunct structure with a non-verb functioning as a verb.

14. Why this PCG Theory?

In the history of mathematics, we have seen the limitations of the way in which numbers were represented in the Egyptian or the Roman systems.

It is only the Indian (or the so-called Arabic) numeral system that effected a breakthrough for further development in mathematics because of the decimal place value in this notation, facilitating a simple way of manipulating mathematical operations.

In a similar way, the existing grammars of languages, based on the various theories of language structure, have not hit upon a notation with a 'place value', very much akin to the number system. If that were done, going from one language to another, or from one structure in a language to another would be through operations that could be performed like the arithmetical operations on numbers.

Let us take, for instance, example 2 of §12, where we saw how we go step by step from the English sentence:

She wants to go home

A B C D E

to the Hindi sentence :

Vah ghar jaan aa caahtii hai.

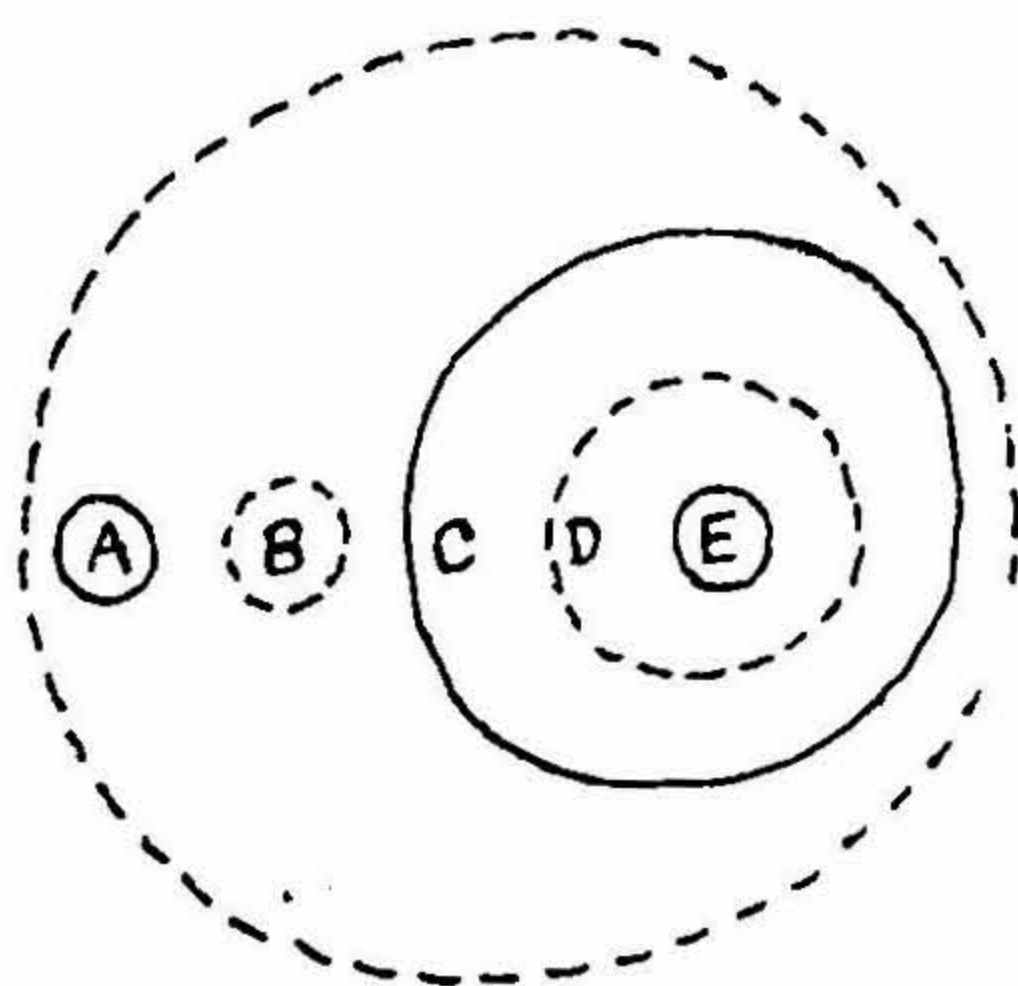
P Q R S T U

In PCG-demarcation, the English sentence would be:

((She) wants (to (go (home))))).

A B C D E

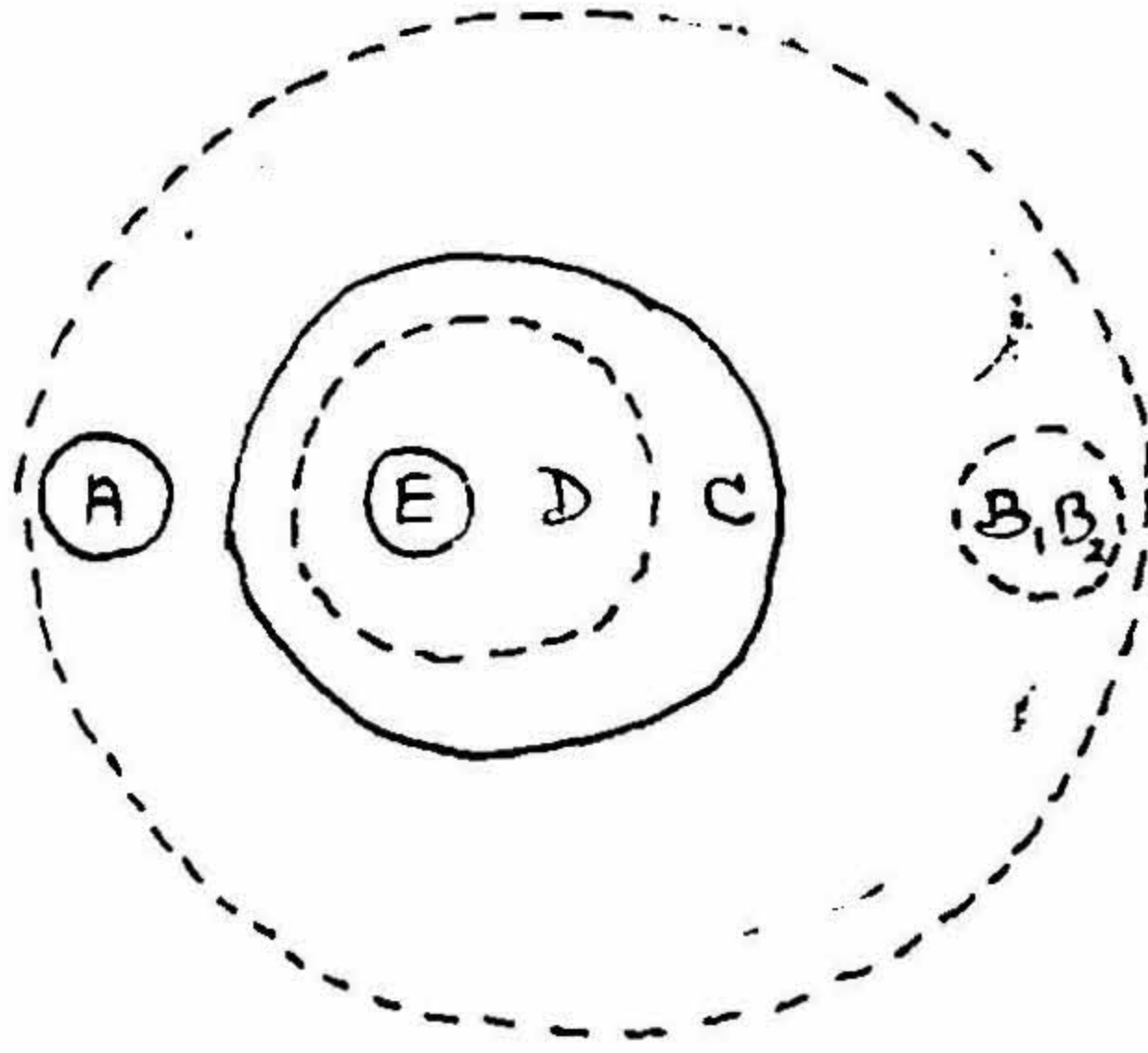
Replacing the words by the letters A-E, the brackets () by a dotted circle and the brackets < > by a solid circle we arrive at the spatial diagram:



In a similar way, corresponding to the Hindi version:

((Vah) (((ghar) jaan) aa) caahtii hai),
 P Q R S T U

we have the diagram:



If we compare the two diagrams, we see that they are spatially *identical* and that $P = A$, $B = T U$, $Q = E$, $R = D$ and $S = C$. These elements lie in their respective orbits, in spite of a change of position within their orbits (much like the positions of planets in the Solar system at different times of the year).

It is this syntactic spatial relation that has made it possible for us to go mechanically, step by step, from English to Hindi.

This spatial relation is brought into an explicit representation in the bracketing notation used in the PCG Theory.

Mechanical translation, in terms of such a theory, would be a meaningful engineering proposition, so long as we have to deal with only non-ornamental, matter-of-fact presentation of factual and logical information, and so long as we restrict the linguistic *style* within prescribed editorial limits.

After 1965 new information, about fresh attempts at *mechanical language processing* with more modest and limited objectives, began to be once again available.

It is now evident that several people (from several limited points of view) are interested in *mechanical language processing*: linguists (to gain greater insight into language structure), translators (to effect a practical and speedy translation, with pre- and post-editing, without laying undue emphasis on linguistic perfection) and information scientists (concerning themselves with the content or substance of a text, paying no great attention to the way it is expressed in the formal structure of natural language).

There is also a noticeable tendency¹ among some linguists and information scientists to move in each other's direction in their work. The technical translators interested

mechanical translation occupy a place midway between the (applied) linguists and the information scientists.

In the matter of mechanical processing of natural language material, therefore, there is already a convergence in the approaches of these different specialists, through their interest in semantics, although their respective goals are different.

Our own work has been concerned with the problem of mechanizing as much of the process of scientific translation as possible.

Translation is not a single process done at one go, although that is the impression we have of this process when it is gone through almost at one go by an experienced *human* translator. It is however a process made up of several component processes, gone through either successively or parallelly (but separately) and then finally combined to give the desired results.

These component processes take place in such linguistic areas as : syntactical analysis, morphological analysis, comparison and selection of equivalents from syntactic, morphological and lexical dictionaries, morphological and syntactic matching of separate dictionary equivalents in terms of the syntactic rules of the target language, syntactical reordering of words and phrases from their original order in the source language to a new order in the target language. Add to this the process of taking account of the semantic categories and the process of pre- and post-editing and we have a rough idea of the process of translation in a mechanized way.

However, all analyses of language, including Chomsky's Transformational-Generative Grammar, deal mainly with the syntactic structures of language, and other areas of language (that have anything to do with 'content' or 'substance') are relegated to a secondary position, if not altogether dropped out from consideration.

Language, when considered as a whole, has three main components : (1) The syntactico-semantic, (2) the morphologico-semantic and (3) the lexico-semantic—phonology being subsumed under morphology. (The key word here for any translation is 'semantic.')

Every component has to be dealt with not merely as *form*, but also in terms of its logical (and psychological) relations with respect to the *substance* (formally categorized in a logical way), which is what any language is required to communicate whatever be its own formal structure.

In other words, formal semantics, using logical (and psychological) dissection of the area of language activity referred to as *substance*,² would be necessary in addition to formal syntactical and morphological descriptions.

The lexicon (with these logical classifications incorporated in it) would then be an inalienable component of any linguistic description aimed at practical application to mechanized handling.

Even at the level of purely syntactic analysis one should aim at handling any type of complex sentence structure ever possible in a language. In order that two or more language structures could be compared, the syntactic structures considered should be 'universal', that is, common to *at least the two languages* that form the source and the target languages for translation.

Lastly, in most syntactic descriptions, analyses of sentences are given for an anticipated final hierarchical structure.

In ordinary writing, speaking or even reading a written sentence, what one does is to go from left to right, classifying and reclassifying the words and groups of words as one goes along.

Frequently the speaker or writer groups his words in one way at the beginning but, half way through, the same groups of words are regrouped as he formulates his sentence further. Quite as frequently, the speaker or writer groups his words in one way, but the reader, as he goes from left to right, could group the same words slightly differently, even when there is no pun involved.

When we consider a sentence not as an object already given in full but as a process going on from left to right, all that has been pointed out so far are the problems that are yet to be tackled either at the theoretical level or at the practical engineering level.

Here we are going to confine ourselves to the discussion of only a few limited problems:

- (1) Redefining the structure of a sentence with the *Verb* as its nucleus. [We do not split the sentence into an NP and a VP, as is done almost by all other linguists, with the possible exception of the Hallidayans. That approach, in our view is merely a formalised version of traditional grammatical treatment of the Western languages, wherein the sentence is split into a subject and a predicate and according to which the predicate includes (quite illogically in our view) a few noun phrases in addition to the verb, not to speak of adverbs considered also as part of the predicate. In our view all noun phrases including the subject are related to the verb in the sentence more or less in the same way.]
- (2) Considering idiomatic expressions and clichés as single indivisible lexical items in the two-language situation.
- (3) Establishing the structural constituents of a sentence as being 'universal' for more than one language, so that, structure-for-structure translations would lead to perfectly grammatical sentences in the target language.

The present paper is presented here serially, part by part.

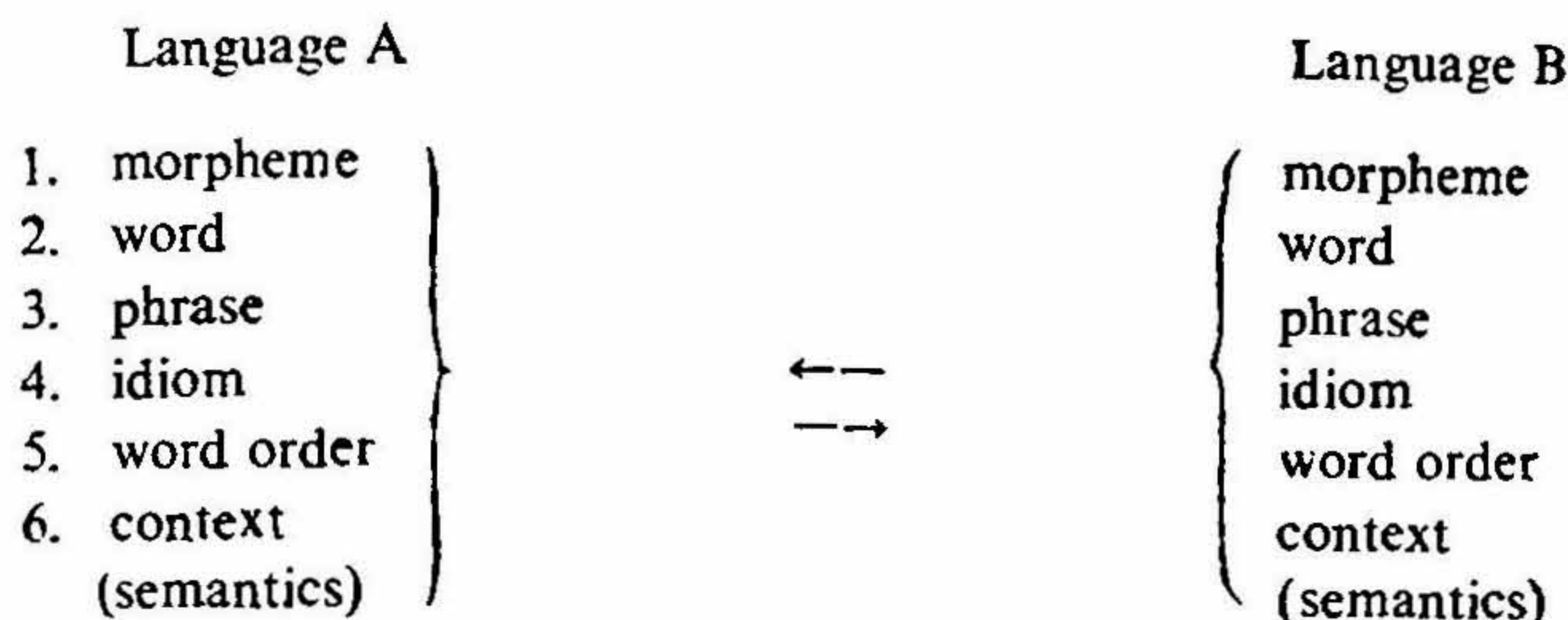
Part I on 'Problems of Translation' follows.

PART I

PROBLEMS OF TRANSLATION

1.1. *Linguistic problems of translation*

The linguistic problems involved in translation could be at several levels, as shown below:



Each level in language A may be equivalent to any of the levels in language B, leading to a one-many or many-one correspondence situation.

1.2. *Language as literary form vs. language as a vehicle for communication*

In poetry and other forms of imaginative writing meant to produce an aesthetic appeal the words and phrases in themselves do not necessarily mean what they are supposed to mean in matter-of-fact statements. Mood, rhythm, style, suggestion, allusion, etc., are some of the components poetic language is more likely to reflect. A word-to-word, phrase-to-phrase or even sentence-to-sentence translation would be ineffective here.

On the other hand, in matter of fact scientific communication, one could be a little more at home in sentence-to-sentence translations generally. Mostly phrase-to-phrase equivalents are available and in the case of purely technical terms even word-to-word translation is possible.

1.3. *Technical terms and fields of study*

The problem of finding proper equivalents is not completely absent even here.³

The technical terms in different languages may not exactly correspond. They too may reflect a one-many or many-one relationship. For example:

<i>English</i>	<i>Russian</i>
meaning } value } magnitude } quantity }	znachenie velichina

The choice of an equivalent, therefore, often depends on context, even in scientific communication.

Further, even in one given language, the same word could be used in different fields with different technical connotations. For example: *base, root, power, morphology, quantity, length, etc.*

1.4. *Style of presentation*

Under the expression 'style of presentation' we shall understand the differences in the vocabulary and structures of the language that are met with under different conditions. Some of these conditions could be:

1. A scientist formally communicating a report on his work to other scientists, who are specialists in his own field.
2. A scientist presenting a formal report on his work to administrative heads, who may not be specialists in his field.
3. A scientist informally chatting about his subject to his colleagues.
4. A scientist presenting his ideas formally or informally to a lay audience.

The intended audience, the formal or informal conditions under which the ideas are presented and the depth of specialization to which the audience is capable of being taken—all these determine to a large extent the form of the language: namely, types of sentences used, as well as phrases, clichés, and vocabulary (technical and non-technical).

1.5. *Minimization of the variation in these factors by restriction of field and style*

If translation is to be done without being overweighed by the interfering factors of style and fields of discourse (when all fields and all styles are treated at once), one could select one-style-situation and one field of study at a time for a detailed linguistic study, so as to work out rules for translation, that could be fairly uniform, leading to possible mechanization of a large number of processes involved in technical translation.

1.6. *Mechanization of translation*

Natural language structures and their semantic interpretations are so complex that an absolutely end to end mechanized translation would be almost impossible, even when the two languages and the subject field were taken to be finite and static, and the style of presentation to be rigidly uniform. However, such an 'ideal' limiting situation is not to be expected under real conditions.

A human translator, with all his cultural equipment and endowed by nature with a live learning mechanism, memory and reasoning power, capable of varied mental operations, as yet not imitated by machines, is still the best translator.

Having made a tentative translation, the human translator re-reads it with the question in his mind: 'Does it make sense?' Where logic, or factual rendering seem to be faulty, he re-examines the translation, consults other reference materials or a subject specialist and reformulates the doubtful portions in his translation. This is reasoned editing which perhaps a machine cannot take over from the human translator or editor.

However, the human translator too, goes through many mental operations that are of a repetitive, mechanical and time-consuming nature. Consulting a dictionary, analysing the sentence into meaningful parts, transposing the sentence elements into a different order, etc., are some of these.

For example, the German sentence:

Der eine Zigarette rauchende und auf dem Berg stehende Mann (ist X).

would sound as given below in English, if word-for-word substitution, without change in word order, were attempted:

The a cigarette smoking and on the hill standing man (is X).

With proper transpositions, we would have:

The man smoking a cigarette and standing on the hill.....

An experienced human translator does it almost automatically, although after a little reasoning in the case of complicated structures.

But this process of transposition could be mechanized, and the rules for doing so made explicit, if a pre-editor, unilingually, could prepare the text for mechanical operation much in the same way, as a man reading a proof or revising a draft goes on correcting and deleting and changing unilingual material.

Anticipating the structures of our Practical Theory of Language Structure (see part 3), let us put these into the structural frame-work of that theory. We then have:

$$(1) \langle_1 \text{Der} \langle_2 \langle_3 \langle_4 \langle_5 \text{eine Zigarette} \rangle_5 \text{rauchende} \rangle_4 \rangle_3 = \text{und} = \langle_6 \langle_7 \langle_8 + \text{auf dem Berg} \rangle_8 \text{stehende} \rangle_7 \rangle_6 \rangle_2 \text{Mann} \rangle_1$$

for German and:

$$(2) \langle_1 \text{The man} \langle_2 \langle_3 \langle_4 \text{smoking} \langle_5 \text{a cigarette} \rangle_5 \rangle_4 \rangle_3 = \text{and} = \langle_7 \text{standing} \langle_8 + \text{on the hill} \rangle_8 \rangle_7 \rangle_6 \rangle_2 \rangle_1$$

for English.

The expressions are equivalent, bracket for bracket, as numbered above, in the two languages.

Thus

$\langle_1 \text{Der Mann} \rangle_1 - \langle_1 \text{The man} \rangle_1$

The bracket $\langle_2 \rangle_2$ containing the qualifying expression for the noun in bracket 1, is placed after the noun in English and between the article and the noun in German.

As against examples (1) and (2) above, let us examine a structure semantically equivalent to (1) and (2) in the two languages:

We have:

(3) $\langle_1 \text{Der Mann} \langle_2 (\langle_3 \langle_4 \text{der} \rangle_4 \langle_5 (\langle_6 \langle_7 \text{eine Zigarette} \rangle_7 \text{raucht})_6 = \text{und} = (\langle_8 \langle_9 \text{+auf dem Berg} \rangle_9 \text{steht})_8)_5)_3 \rangle_2 \rangle_1$

and

(4) $\langle_1 \text{The man} \langle_2 (\langle_3 \langle_4 \text{who} \rangle_4 \langle_5 (\langle_6 \text{is smoking} \langle_7 \text{a cigarette} \rangle_7)_6 = \text{and} = (\langle_8 \text{is standing} \langle_9 \text{+on the hill} \rangle_9)_8)_5)_3 \rangle_2 \rangle_1$

or

(5) $\langle_1 \text{The man} \langle_2 (\langle_3 \langle_4 \text{who} \rangle_4 \text{is} \langle_5 (\langle_6 \text{smoking} \langle_7 \text{a cigarette} \rangle_7)_6 = \text{and} = (\langle_8 \text{standing} \langle_9 \text{+on the hill} \rangle_9)_8)_5)_3 \rangle_2 \rangle_1$

or again:

(6) $\langle_1 \text{Der Mann} \langle_2 (\langle_3 (\langle_4 \text{der} \rangle_4 \langle_5 (\langle_6 \langle_7 \text{eine Zigarette} \rangle_7 \text{raucht})_6 = \text{und} = (\langle_8 \langle_4 \text{der} \rangle_4 \langle_9 \text{+auf dem Berg} \rangle_9 \text{steht})_8)_3)_2 \rangle_1$

and

(7) $\langle_1 \text{The man} \langle_2 (\langle_3 (\langle_4 \text{who} \rangle_4 \text{is smoking} \langle_5 (\langle_6 \langle_7 \text{a cigarette} \rangle_7)_6 = \text{and} = (\langle_8 \langle_4 \text{who} \rangle_4 \text{is standing} \langle_9 \text{+on the hill} \rangle_9)_8)_3)_2 \rangle_1$

The structures (4)–(7) are different from (1) and (2) in that (3)–(7) have a relative clause construction within the qualifying bracket 2 that qualifies the noun which is within bracket 1.

In German the qualifying structure in bracket 2 of expression (1) contains a verb in the present participle. Since a present participle in this language behaves like an attribute even morphologically (that is, taking the gender and case endings), the bracket containing it is placed where an adjective would be placed, that is between the article and the noun.

In English too an *ing* form of the verb could be treated as an adjective. But, when it occurs with its own complement as in (2), the *ing* form is treated as a verb in this language and not as an adjective and consequently the bracket containing it is not placed between

the article and the noun [as a true (attributive) adjective should be in English too], but is put in a position after the noun, like a relative clause.

In both (1) and (2), irrespective of their use as attributes to the noun, the participle or *ing* forms, have the characteristics of a verb and so they determine a C-structure, with the complements of these verbal elements as their arguments having a P-structure. The C-structure as a whole qualifies the noun and therefore, with respect to that noun it is an inner element of a P-structure.

In our notation therefore we have placed all elements forming a C-structure within the bracket () and all elements forming a P-structure within the bracket ().

The relative clause or the participial construction is therefore a C-structure within a P-structure. [See Part 4, Section 1 for the *dual* character of structures, esp. 4/1.3 where a bracket like (()) has been explained.

In the relative clause constructions (3)–(5), both in German and English, the verb in bracket 3 (that is, in brackets 6 and 8), [bracket 5 being an algebraic bracket, as in $(_{3}({}_{6}abc)_{6} + ({}_{8}ade)_{8})_{3} = ({}_{3}a ({}_{5}({}_{6}bc)_{6} + ({}_{8}de)_{8})_{5})_{3}$, owing to a common element 'who' ('der') or 'who is' being taken out of the bracket], does not morphologically behave like an attributive adjective and so bracket 2 containing 3 is placed after the noun in bracket 1 in both the languages.

However, the relative clause in bracket 3 has an attributive function with respect to the noun in bracket 1, so it is placed in a P-structure bracket 2, since only a lexical adjective or a P-structure can be an attribute to a noun or to another P-structure.

In (6) and (7) no algebraic common factor is taken out of the brackets 6 and 8 and so each bracket retains its full relative clause structure. Therefore, no algebraic bracket like bracket 5 is necessary for them.

Nevertheless, since two C-structures connected by a logical connective ' = and = ' together form a C-structure, a bracket like 3 is necessary.

[The bracket numberings given here are arbitrary, but from (3) to (7) corresponding numbers are used to identify corresponding elements.]

We see from the above examples that irrespective of word order German and English have the same C- and P-structures, when we go from the outermost structure to the innermost, as shown by the bracketing system given here.

1.7. *Logic and the common man in his use of the natural language*

Even a linguist or a logician is a naive layman, when it comes to the practical use of a natural language, although they are both well-trained experts in their use of their own respective jargons.

The ordinary man, in his practical use of the natural language, is a 'naive naive-lay-man', when we compare him with the linguist or logician or linguist-logician.

The logician, linguist or any other scientist, as we said, is naive in the use of the natural language, because the full mechanism of the natural language has not yet been fully and unambiguously understood even by him.

The natural language, however, is also a common-man controlled mechanism, and is subject to the vagaries of his ways of using it, much in the same way as an untrained hand might often try to use a chisel as a screw-driver, if he doesn't fully understand the purpose for which the respective tool has been made or its nature and structure, its limits of operation, etc.

Add to this his mischievous bent of mind (reflected in the journalistic twists given to logical and scientific expressions, not to speak of other natural language expressions, in the ordinary use of a language) or his literary genius (resulting in expressions devised for giving multiple meanings, etc.) and we have the situation of a confusing tool worse confounded by the common man's handling of the natural language in imitation of the journalist and the literary genius.

(Who knows whether a great literary language, like Sanskrit, Old Greek or Latin, died a *natural* death or it was *killed* by pundits indulging in their literary tricks of producing expressions having multiple interpretations.)

One function of language is the work or task of *communication of information* in an undistorted form and another is the art of transmitting *aesthetic pleasure* through language used in literary productions, a third one is to exhibit one's emotional state through joyful or sorrowful or angry outpourings in linguistic form, and a fourth one, as even a child is able to know these days, is to use language in the *art of communicating misinformation* (that is communicating information with intentional *noise* and *distortion* through linguistic *jamming*), leaving the recipient guessing what is meant by what is said.

Work, aesthetic pleasure and the pleasure derived from artful deception are 'cultural' (and as a result linguistic) traits in man.

All this has made the common man controlled language (or language, controlled by the common interaction of men of all traits in society) a blunt and twisted tool or a fuzzy or hazy medium for communication.

Natural language, therefore, is such that, even in purely scientific writing for communication of facts without distortion, scientists have to *agree* that only some particular thing is meant by some particular expression. Quite often, even scientists (who unfortunately are laymen when it comes to their understanding of language) themselves do not realise that they use expressions without a common agreement of the particular desired interpretation of those expressions.

One example of this type is the use of a string of nouns in scientific writing. (This reminds me again of Sanskrit pundits who use compound nouns of variable interpretations for the mere pleasure of it !)

How does one interpret a string of nouns like:

$$N_1 N_2 N_3 N_4 N_5 ?$$

The associative law given in books of algebra namely:

$$a(b c) = (a b) c$$

does not hold good here. For, we have the situation:

$$N_1 (N_2 N_3) \neq (N_1 N_2) N_3.$$

For example:

- | | | | | | |
|-----|----------|-------|-----------|----------|--------|
| (1) | Electron | Tube | Vibration | Analyser | Design |
| | N_1 | N_2 | N_3 | N_4 | N_5 |

or

- | | | | | | |
|-----|----------|-------------|-------|-------------|------------|
| (2) | The City | Corporation | Road | Development | Department |
| | N_1 | N_2 | N_3 | N_4 | N_5 |

are two strings of nouns that are supposed to mean only one particular thing. What are the N's here that group themselves meaningfully ?

There are a number of theoretical alternatives of which the following, in which all the N's that precede together act as an attribute for the one that follows, could be one:

$$((((N_1) N_2) N_3) N_4) N_5).$$

This whole thing could further act as an attribute to N_6 .

This would then mean:

(1 a) (((Electron) tube) vibration) analyser) design, that is, 'design of an analyser for the vibration of an electron tube'.

Or,

(2 a) The (((City) Corporation) Road) Development) Department, that is, 'The Department for the Development of the City Corporation's road (The Department is not necessarily the City Corporation's Department).'

The proper grouping for (1) seems to be:

(1 b) (((Electron) tube) ((vibration) analyser)) design, that is, 'Design of a vibration analyser that uses an electron tube' (Here too, 'that uses' has to be added for clarification).

For (2), a proper grouping could be:

(2 *b*) ((City) Corporation) (((Road) Development) Department) that is, 'The City Corporation's Department for Road Development'.

We see that different groupings are required for a string of 5 nouns, if the nouns differ lexically. Other alternatives different from the above could be found for other strings.

Therefore, it may be necessary to prepare a grammar for such strings, classified according to combinations and described according to lexical or semantic criteria. It would also be useful to prepare special dictionaries in each field of science for different combinations of technical strings of nouns for any two languages for mechanized consultation or processing of a translation.

Corresponding to the English noun string $N_1N_2N_3N_4N_5$ one would have different alternatives, for example, in Russian:

<i>English</i>	<i>Russian</i>
$(((((N_1) N_2) N_3) N_4) N_5)$	$\rightarrow \begin{cases} N_5 (N_4 (N_3 (N_2 (N_1)))) \\ N_5 (N_4 (N_3 ((N_1) N_2))) \end{cases}$

In a bracket like $N_2 (N_1)$, N_1 would generally be a noun in the genitive case in Russian, but in a bracket like $(N_1) N_2$, N_1 would be morphologically an adjective.

Further in a combination like:

$$((N_1) N_2) N_3,$$

if N_2 is already an adjective morphologically, N_1 would also be an adjective compounded by a special morphological device (in Russian) to N_2 .

If we have:

	<i>English</i>	<i>Russian</i>
N_1	\rightarrow electron	elektron
N_2	\rightarrow tube	lampa
N_3	\rightarrow analyser	analizator

then, we could think of a combination like:

<i>English</i>		<i>Russian</i>
$((N_1) N_2) N_3$ Electron tube analyser	\rightarrow	$((N_1) N_2) N_3$ elektronno-lampovyj analizator

or

$((N_1) N_2) N_3$	\rightarrow	$N_3 ((N_1) N_2)$ analizator elektronnykh lamp
-------------------	---------------	---

that is, the bracket following N_3 is in the genitive case.

Such combinations are linguistic abbreviations for logical relations. They could be handled mechanically in many ways:

(1) Construct a dictionary of actual noun strings and give their equivalents in the other language. If there are alternative equivalents, provide them all for the post-editor to choose from.

(2) Carry out a semantic (logical) analysis of the combination for the source language and group the strings in different ways and give equivalence rules for the different groups.

For example, if we have in English the oft-repeated and discussed phrases :

‘The dream of a neurotic’

and

‘The dream of an apple’

we could ask the question : ‘Could we replace ‘of X’ by ‘by X’ or by ‘about X’?’

If only one of these replacements is possible, the pre-editor has to indicate a group tag for that particular possibility. If the context doesn’t tell anything and if both alternatives are possible, the group indication has to be by another tag.

Then, we have the replacement groups :

of X → by X	group 1
of X → about X	group 2
of X → {by X about X}	group 3

‘Of a neurotic’ in the absence of proper context would be group 3. ‘Of an apple’, except in a fairy tale, where ‘an apple’ could ‘dream of a neurotic’, would be only group 2.

If the context makes it clear, then ‘of a neurotic’ could be group 1. The unilingual source language pre-editor has to add the tag for translation into another language.

A syntactic sentence generally is an abbreviation for several different semantic sentences, much in the same way as for example, the abbreviation ‘M.E.’, which would mean something different in different contexts. For example we could say:

‘The war in the M.E. has nothing to do with the M.E. students of the M.E. department of the M.E. College,’

1. *Vide* Ref. 1, 2. 4.

2. *Vide* Ref. 3.

3. *Vide* Part 4, Section 4, (continuation of the present paper, to follow).

Where the different 'M.E.'s' in the order of their occurrences stand for : Middle East, Master of Engineering, Mechanical Engineering (all of which could more or less be guessed from the context and general knowledge) and Mohammed Ebrahim (which could not be guessed except through more particular knowledge).

1.8. *Meaning pre-linguistically determined :*

Meaning in general is a relationship that is determined at a pre-linguistic semantic (lexical, logical, psychological) 'deeper' structure level.

It is incidental that the imperfect, not-so-logical tool of natural language structures, (un) controlled by the common man, to some extent correspond in different languages :

Examples :

'My head'	(English)
'Mein Kopf'	(German)
'Moja Golova'	(Russian)

But the syntactic structures are different, the moment we try to express something done by or to 'my head'. For example, we say :

- 'I shook my head' in English,
- 'I shook myself the head' in German and
- 'I shook myself by the head' in Russian.

But these surface syntactic structures and even the syntactic deep-structures underlying them in the different languages mean the same thing at the pre-linguistic 'deeper' structure level.

This shows that the same things and actions of the external world are perceived, analysed and reported differently even in genetically related languages, however close.

Whorf's observations about the Hopi language, a so-called exotic language of the New World, and of the world view reflected in its grammatical and syntactic structures (or rather in the underlying pre-linguistic, logical and psychological structure) is worth examining a little closer.

References

1. *Current Research and Development in Scientific Documentation No. 15*, National Science Foundation, 1969, 91-162, 163-84 and 561-606.
2. GARDIN, J. C. *Documentation Analysis and Linguistic Theory. Journal of Documentation*, 1973, 29, 137-68.
3. MACKEY, W. F. *Language Teaching Analysis*, Longmans (1969), Introduction 3-33.
4. MONTGOMERY, C. *Linguistics and Information Science. J. Amer. Soc. for Inf. Science*, 1972, 23, 195-219.