The P-structure C-structure Grammar (PCG) for the contrastive study of two or more languages—Parts IV, V and VI*

P. C. GANESHSUNDARAM
Foreign Languages Section, Indian Institute of Science, Bangalore 560 012

Received on March 2, 1978; Revised on July 26, 1978

PART IV

LANGUAGE ANALYSIS AND TRANSLATION PROBLEMS

This part deals with a number of separate problems, related to the central theme of language structure and problems of translation (mechanical and otherwise). It is divided, therefore, into a number of separate sections, each of them dealing with one aspect of this central theme.

SECTION 1

Preparing language for a scheme of computer analysis towards partially mechanized translation with pre- and post-editing**

4.1.1. *Introduction*

For the purposes of this part, we understand by 'language' the written language in a suitably coded form.

The coded form may be in ordinary Roman spelling. For English and the Western European languages, this coded form will be in conventional spelling with very slight modifications to replace diacritical marks (like umlaut in German and the accents in French).

* Introduction & Part I appeared in JIISc, 1978, 60 (A), 167–192 and Parts II & III in JIISc, 1978, 60 (A), 239–266.

** This is a revised version of ref. 3.

Russian has to be completely Romanized. Czech, although in Roman script, ha to be re-Romanized to eliminate diacritical marks (by replacing letters with such mark by a combination of letters in a spelling system, more or less as in Polish).

Indian languages are to be Romanized also through a suitable spelling system.

4.1.2. *Romanization of Hindi for computer processing*

Hindi, for example, could be Romanized in different ways and different alternativ spelling systems devised.

It is possible to represent each Hindi (Devanagari) graphic unit by numbers or by one or two-letter combinations in the Roman script. This would be of advantage if ou aim were to devise a way for automatically converting the Romanized version into th Devanagari version.

However, if the script conversion is to be done more simply by human agency and the language itself is to be processed by the computer for different purposes other than script conversion, we need not bother to find equivalents for each graphic symbol.

We have used, for the preparation of a small English–Hindi and Hindi–English Glos sary, the following scheme of Romanization, which is easy to read, when once we ge used to the spelling system :

A (अ)	AA (आ)	I (इ)	II (ई)	U (उ)
UU (ऊ)	R (ऋ)	E (ए)	AI (ऐ)	O (ओ)
	AU (औ)	AW (अं)	AH (अ:)	

We note here that vowel length is represented by a repetition of the symbol for th short vowel.

R represents the special vocalic (syllabic) R.

We treat the letter W as an inverted M for indicating the nasalization of the preced ing vowel.

The consonants are given below :

K (क)	KH (ख)	G (ग)	GH (घ)	NG (ङ)
C (च)	CH (छ)	J (ज)	JH (झ)	NJ (ञ)
TX (ट)	TXH (ठ)	DX (ड)	DXH (ढ)	NX (ण)
T (त)	TH (थ)	D (द)	DH (ध)	N (न)

P (प)	PH (फ)	B (ब)	BH (भ)	M (म)
Y (य)	R (र)	L (ल)	V (व)	SH (श)
SX (ष)	S (स)	H (ह)	KSX (क्ष)	TR (त्र)
JNJ (ज्ञ)	F (फ़)	Z (ज़)	Q (क़)	KHH (ख)
	GHH (ग़)	RX (ड़)	RXH (ढ़)	

It may be noted that NG (ङ) and NJ (ञ) are often freely replaced by the *anuswàra* W (–). Thus the letter W would amply serve to represent them as an adequate equivalent:

For example:

अंक	or	अङ्क	AWK	or	ANGK
अंग	or	अङ्ग	AWG	or	ANGG
कुंज	or	कुञ्ज	KUWJ	or	KUNJJ

The letter X is used to represent retroflex sounds (or at least what were originally retroflex sounds in Sanskrit represented by the same Devanagari letters).

Thus we have:

विषय	VISXAY
लडका or लड़का	LADXKAA or LARXKAA
उठो	UTXHO
अक्षर	AKSXAR
क्षमा	KSXAMAA
ऋण	RNX
वक्ष	VRKSX

The Roman spelling closely follows the usual norm concerning the suppression of pronunciation of the short A (अ) at the end of a syllable or word.

Thus:

कमल	KAMAL (not KAMALA)
बिलकुल	BILKUL (not BILAKULA)

The Romanization presented here is purely a practical device for processing Hindi by a computer which accepts only the Roman letters.

4.1.3. *A pragmatic view of language structure*

For purposes of comparing and contrasting the structures of any two languages, we find it amply rewarding to turn away from absolutely rigid and exact theoretical formulations based on the theories established by different schools of linguistics.

Ours is a simple (if crude) rule of thumb theory of language structure.

We define, for our purposes, the following:

1. A *sentence* S (for practical purposes S′ is treated as equivalent to S—*vide* Part III and Part V) is that which lies between a *fullstop* and a *fullstop*, and is called an *S-structure.*

(If there is no fullstop, as it happens at the beginning of a paragraph, the pre-editor puts it there. This, we believe, is justified for purposes of ' language engineering ', since we note that even refined linguistic theoreticians often base their theories of language on that non-formalized thing called *intuition*, before they formalize their intuitive view of language.)

2. A sentence may be made up of *C-structures* and *P-structures* and one *Verb* (which may be a compound verb with several auxiliaries and one lexical verb and may then be called a *verb phrase*). It may thus consist of *at least one P-structure* (as in some one-word sentences of Tamil, Russian, etc.) or *one verb or verb phrase* (again as in Russian Tamil, etc.) or *at least one P-structure and one verb phrase* (as in English, Hindi, etc.).

We do not understand a verb phrase in the Chomskian sense. Our verb phrase does not contain any noun phrase in it. It is a verb phrase *per se*, consisting only of verb forms (*See* Section 2 of this Part for English verbs).

3. A *C-structure* is one that contains one verb phrase and one or more P-structures.

4. A *P-structure*, seen from without, does not contain a *verb phrase* as a direct component in it. It may (wholly or partly) consist of a C-structure or other P-structures. The C-structure, in its turn, has a verb as one of its components.

5. A C-structure, a P-structure or a verb may occur with a *marker* that helps to identify to some extent the C-structure type, P-structure type or verb phrase type. However, such a marker postulated on a semantic or logical basis, may be *zero* in form.

The outermost structure, *viz.*, the S-structure, contains *one* and only one verb-phrase as its immediate inner constituent, although its other constituents may have in their turn (if they are C-structures) an inner constituent that is a verb-phrase.

In other words a single verb phrase is the nucleus of every C-structure or every S-structure.

If a Sentence (S-structure) has two (or more) verb-phrases connected by *and* they are two (or more) sentences (S-structures) *logically* combined into one.

6. Certain structural units may be *dual* in character. Seen 'from without' (that is proceeding from the larger units that contain them and arriving at them) they may be looked upon as P-structures, under given conditions. But seen 'from within' (that is, in terms of the elements constituting these structural units), they may be looked upon as C-structures, if a verb phrase is one of these elements.

In such cases (*see* 4.1.5) the *inner bracket* will be a *C-structure bracket* and the *outer bracket* will be a *P-structure bracket* in our notation.

The telescopic structures given below are typical examples containing such dual structures. (Their analyses will be given in 4.1.5. below):

Examples of telescopic dual structures:

He knows that he knows that he knows that he knows that (1)

He knew she would say she knew he would say so (2)

Or, to take a less artificial and a more common (and meaningful) example:

This is the dogmatic linguist who said that he would cut the throat of any linguist who challenged his theory (3)

4.1.4. *Linear demarcation of structure*

Two-dimensional tree diagrams or cumbersome (and rigid !) rewrite formulas make it extremely difficult for an ordinary language user to make any practical analysis of language without having to undergo detailed training in linguistics in one of the established 'schools' of linguistics.

We would like to give a simple method (that could be easily learnt) of demarcating a sentence from left to right, as it is written or read, representing a more or less flexible rule-of-thumb method for practical purposes.

We do not also insist that a given structure should be interpreted only in one rigid way. It is common knowledge that, even when we exclude from our consideration deliberately ambiguous constructions or funny pairs of tricky sentences like 'she made him a good wife' and 'she made him a good husband', and the like, we still have ordinarily unambiguous constructions that do lend themselves to alternative interpretations.

Our view is that anyone is free enough to choose any one of the alternative interpretations. In fact that is what we do in perfectly normal situations. A little lee-way to allow

for flexibility is felt to be desirable in any method applied to the practical analysis of language structures.

Even if it were not logically justifiable to do so (from the point of view of any particular philosophy of language), we shun rigidity. For, our philosophy of language recognises flexibility as an important property of natural language.

In our linear demarcation system we make use of a few types of *brackets* to identify particular *types* of *structural units*. Other bracketing conventions could be adopted. But we found that the present choice, arbitrary as it is, proved to be quite convenient in punching cards.

4.1.5. *The system of brackets*

Table I tells us what we identify in terms of the bracketing signs specified:

Table I

Bracketing sign at the beginning	Bracketing sign at the end	Sign between units	Unit identified
(fullstop).	(fullstop).		Sentence or S-structure
(	)		C-structure
⟨	⟩		P-structure
–(	)–		Parenthetical C-structure
–⟨	⟩–		P-structure in apposition
=	=		Coordinating conjunction
+			Structural marker
		–	Units combined into a single item

The examples (1)–(3) given above would then be demarcated as:

. ⟨He⟩ knows ⟨+ that (⟨he⟩ knows ⟨+ that (⟨he⟩ knows ⟨+ that (⟨he⟩ knows ⟨that⟩)))))). (1)

. ⟨He⟩ knew (((she) would say ⟨((she⟩ knew ⟨((he⟩ would say ⟨so⟩))))))). (2)

. ⟨This⟩ is ⟨the dogmatic linguist ⟨((+ who⟩ said ⟨+ that (⟨he⟩ would cut ⟨the throat ⟨+ of any linguist ⟨(((+ who⟩ challenged ⟨his theory⟩)))))))))). (3)

There are words in many structural units that serve as *markers* to identify the *type* of these units. These markers are identified by putting a + sign at the beginning. Some of these have a double role: firstly, as markers of the structural units to which they are attached, indicating the relationship of these units to the larger unit, and secondly as smaller units forming part of those structural units which they identify.

For example:

(+ who) in the P-structure:

⟨The man ⟨(⟨ + who⟩ is ⟨here⟩)⟩⟩.

41.6. *The place of our method of structural analysis in relation to the different theories of linguistic structure*

We do not propose an alternative to what has been done by others, but try to take an engineering advantage of all of them.

Like all of them, we try to describe, transform, generate or interpret linguistic structures or analyse actual sentences in terms of the units of our description. Like all of them, therefore, we too operate on some basic units. (The basic units themselves may differ in detail).

We describe the 'inner structure' of a P-structure or of a verb phrase (as we have defined them here) in the Chomskian way.

We take account of the Hallidayan categories and the rank shift phenomenon among them in our own way, recognising the double roles of structural units as their inherent property.

We do not order the structural units in rank and then say a shift is possible. We recognise that units of different kinds can occur as parts of one another as a perfectly regular phenomenon. It is this interpenetrability of the units into one another that probably enables language to be used as a metalanguage to describe itself.

We take into account the role and case categories of Fillmore, Anderson *et al.*, and these are made explicit (where the situation demands it) by adding case and role markers within the P-structure or C-structure boundary immediately after the lexical or morphological marker. But this is done by us taking into account the structural relativity of the two languages taken up for contrastive study. We do not attempt any absolute or universal categorization of role and case.

At present, the initial step is taken by the demarcation of boundaries by a human pre-editor. Later, a smaller or greater part could be taken over by a more automatic system. (And this is found to be feasible as our further investigations, not reported here, show.)

We radically differ from everybody when we consider what is understood ordinarily as 'adverbs', 'adjectives', 'pronouns', 'nouns' and 'clauses' as alternative components of a P-structure. Since a P-structure is a syntactic category and 'noun', 'adverb', etc. are syntactic and morphological ones, we feel this is not a serious drawback from an engineering point of view. For, we could consider the second P-bracket in each pair of the following examples as semantically equivalent structures:

⟨I⟩ came ⟨+ into this room⟩ (A.1)

⟨I⟩ came ⟨here⟩ (A.2)

⟨He⟩ sings ⟨loudly⟩ (B.1)

⟨He⟩ sings ⟨+ in a loud voice⟩ (B.2)

⟨The machine⟩ is operated ⟨electrically⟩ (C.1)

⟨The machine⟩ is operated ⟨+ on electricity⟩ (C.2)

In the following pairs (where we have the link verb) an 'adjective' and a 'noun phrase' are equivalent:

⟨He⟩ is ⟨slow⟩ (D.1)

⟨He⟩ is ⟨a slow worker⟩ (D.2)

In the pair below a 'noun phrase' and a 'clause' are equivalent:

⟨He⟩ was garlanded ⟨+ on his arrival⟩ (E.1)

⟨He⟩ was garlanded ⟨+ when (⟨he⟩ arrived)⟩ (E.2)

Therefore, if ⟨+ into this room⟩ is a 'noun phrase', then ⟨here⟩ is also equivalent to this 'noun phrase' and both of these are one syntactic category, the P-structure.

4.1.7. ***Combining independent sentences into related or dependent structure in stages***

For the present discussion, let us consider the relative clause construction.

(1) *Stage zero* : Completely independent statements. The necessity for any context to relate them with each other does not arise.

. ⟨The man⟩ came ⟨here⟩ ⟨yesterday⟩. (4)

. ⟨The man⟩ is ⟨an engineer⟩. (5)

Let us assume that (4) is to be taken as a *frame sentence* and (5) as the potential *imbedded sentence*. Then:

(2) *Stage* 1 : Structurally independent, but contextually related for meaningful understanding:

. ⟨The man⟩ came ⟨here⟩ ⟨yesterday⟩. (6)

. ⟨He⟩ is ⟨an engineer⟩. (7)

(The noun is replaced by a pronoun in (7)).

(3) *Stage* 2 : Structurally loosely related. The pronominalized construction is placed parenthetically next to the noun to which the pronoun refers:

. ⟨The man⟩ -(⟨He⟩ is ⟨an engineer⟩)-came ⟨here⟩ ⟨yesterday⟩. (8)

(4) *Stage* 3 : Structurally united into one sentence by changing the pronoun in (8) into a relative pronoun, the whole construction beginning with the relative pronoun being treated as part of the noun phrase:

. ⟨The man ⟨((+ who⟩ is ⟨an engineer⟩)))⟩ came ⟨here⟩ ⟨yesterday⟩. (9)

(5) *Stage* 4 : Structurally united. Further, unnecessary repetitive elements elided, giving rise to an elliptical construction, by dropping the relative pronoun. When the relative pronoun is dropped, the link verb is also dropped. (Elliptical construction represented by a noun in apposition):

. ⟨The man -⟨an engineer⟩-⟩ came ⟨here⟩ ⟨yesterday⟩. (10)

(6) *Stage* 5: Converting the noun phrase into one with a single noun head, the noun in apposition of (10) being placed before it as an attribute:

. ⟨The engineer man⟩ came ⟨here⟩ ⟨yesterday⟩. (11)

4.1.8. *Flexibility in the demarcation of P-structure and C-structure boundaries*

In the initial stages a pre-editor (human) has to make the demarcation of C-structures and P-structures in a sentence. In doing so he may be required to make *ad hoc* semantic interpretations as he would normally do in reading a message in his own language. There would be alternative interpretations, even when the message is not intended to be a pun or any other tricky structure peculiar to a given language.

Structures are tricky, when we consider a number of languages together, to different degrees in the following way:

(1) *Tricky in only one language*

Example:

. ⟨She⟩ made ⟨him⟩ ⟨a good wife⟩.

. ⟨She⟩ made ⟨him -⟨a good husband⟩-⟩.

When a translation of this is made into another language, a semantic interpretation is essential, before a mechanized handling could be attempted. A semantic interpretation is partly provided by the brackets, But:

. ⟨She⟩ made ⟨him⟩ ⟨an apple pie⟩.

makes it still worse. An interpretation of $made_1$ and $made_2$ together with an interpretation of him_1 and him_2 is necessary.

(2) *Tricky in a group of languages as against another group of languages*

Example:

I look at the painting in this room (English)

Je r garde la peinture dans cette chambre (French).

These sentences are not tricky with respect to each other. But they are so in relation to a Dravidian language.

In both French and English, in the above two examples, one could understand the statement in one of two ways, namely:

(1) . ⟨I⟩ look ⟨+ at the painting⟩ ⟨+ in this room⟩.

. ⟨Je⟩ regarde ⟨la peinture⟩ ⟨+ dans cette chambre⟩.

or

(2) . ⟨I⟩ look ⟨+ at the painting ⟨+ in this room⟩⟩.

. ⟨Je⟩ regarde ⟨la peinture ⟨+ dans cette chambre⟩⟩.

The first analysis answers the question: where do I look at the painting?, and the second, the question: which painting do I look at?

In a Dravidian language, depending upon which of these questions is implicitly anticipated by the speaker, he has to choose a different structure (syntactically and morphologically). In Tamil, for example, we have to say:

(1) . ⟨Naan⟩ ⟨inta arhai + il⟩ ⟨cittirat + ai⟩ paarkkirheen.

or

(2) . ⟨Naan⟩ ⟨⟨⟨⟨inta arhai + il⟩ ullx) + a⟩ cittirat + ai⟩ paarkkirheen.

If we wish to analyse the sentence for translation between English and French, either analysis would lead to the same result. We would have the same set of possibilities of semantic interpretation as in the original, whichever analysis we followed.

However, if a translation from either of these languages into Tamil is attempted. the same semantic flexibility is not available. The pre-editor has to make a semantic deci-

sion as to what implicit question is answered and do the analysis accordingly. If he doesn't make a choice himself, but assumes that the original flexibility is available, his analysis imposes automatically a semantic interpretation.

Corresponding to the second analysis we have other structures in English and French which could be interpreted only in one way, such as:

. ⟨I⟩ look ⟨+ at the picture ⟨(⟨+ which⟩ is ⟨+ in this room⟩)⟩⟩.

By *eliding* the relative pronoun and consequently also the link verb, we have the following successive stages of *reduction* leading to the final *elliptical* structure:

. ⟨I⟩ look ⟨+ at the picture ⟨(ϕ⟨+ in this room⟩)⟩⟩.

Since within the C-structure bracket above there is only ϕ (that is, there is no verb) it is no longer a C-structure and the C-structure bracket is dropped. Since in that case, there would be a P-structure bracket *wholly* occupying another P-structure bracket, one of them becomes superfluous and is also dropped, giving us:

. ⟨I⟩ look ⟨+ at the picture ⟨+ in this room⟩⟩.

The P-structure obtained as the last stage of reduction is an elliptical structure, if we consider it as a reduction from an original C-structure. If, on the other hand, we look upon it as an unreduced structure, peculiar to English and unrelated to any other language, we could, if we choose, think of it as a full non-elliptical structure (as we often do). But, *in relation to Tamil* it is *elliptical in English*. This is a kind of syntactico-semantic *relativity* existing between languages.

(3) *Structures that are equally tricky in a large number of languages, if not in all of them*

Examples of this kind may be extremely rare, but if discovered it is presumed that they are likely to be tricky *logically* or *semantically* rather than *structurally* (that is, syntactically).

One type of this variety of trickiness may be proverbs, which do not mean generally what they literally say. For example:

. ⟨All ⟨(⟨that⟩ glitters)⟩⟩ is not ⟨gold⟩. ⟨English⟩

. ⟨(⟨Minnuv ⟨at⟩⟩) ellaam⟩ ⟨pon⟩ alla. ⟨Tamil⟩

Even if the structures in the two languages and the lexical and morphological elements could be equated and their literal meanings exactly equivalent, still in both the languages they mean something other than what they say.

In what follows, we try to present alternative analyses of English sentences. Each alternative analysis gives a slightly different semantic slant to the sentence. Depending upon this slant we would get a particular version of translation in another language.

(1) There is no place in civilization for the idler.

This could be analysed as:

(*a*) . (⟨There⟩ is) ⟨no place⟩ ⟨+ in civilization⟩ ⟨+ for the idler⟩.

(*b*) . (⟨There⟩ is) ⟨no place ⟨+ in civilization⟩⟩ ⟨+ for the idler⟩.

(*c*) . (⟨There⟩ is) ⟨no place -⟨+ in civilization⟩- ⟨+ for the idler⟩⟩.

(*d*) (⟨There⟩ is) ⟨no place⟩ ⟨+ in civilization ⟨+ for the idler⟩⟩.

(*e*) . (⟨There⟩ is) ⟨no place ⟨+ in civilization ⟨+ for the idler⟩⟩⟩.

A detailed analysis of the sentence-structure for each demarcated version abo would be as follows:

(C—C-structure, P—P-structure, V—Verb-phrase)

(*a*) S → V1 P1 P2 P3

V1 → there-is

P1 → no place

P2 → + in civilization

P3 → + for the idler

(*b*) S → V1 P1 P2

V1 → there-is

P1 → no place P3

P3 → + in civilization

P2 → + for the idler

(*c*) S → V1 P1

V1 → there-is

P1 → P2 P3 (P3 being a parenthetical P-structure)

P2 → no place P4

P4 → + for the idler

P3 → + in civilization

(*d*) S → V1 P1 P2

V1 → there-is

P1 → no place

P2 → + in civilization P3

P3 → + for the idler

(*e*) S → V1 P1
V1 → there-is
P1 → no place P2
P2 → + in civilization P3
P3 → + for the idler

(2) The faculties had a detailed discussion on the various measures to be adopted in improving the procedure of admission.

Only one possible alternative is given here in analysing this sentence:

. ⟨The faculties⟩ had ⟨a detailed discussion ⟨+ on the various measures ⟨+ to ⟨be adopted ⟨+ in ⟨improving ⟨the procedure ⟨+ of admission⟩⟩⟩⟩⟩⟩⟩.

Here:

S → P1 V1 P2
P1 → The faculties
V1 → had
P2 → a detailed discussion P3
P3 → + on the various measures P4
P4 → + to C1
C1 → be adopted P5
P5 → + in C2
C2 → improving P6
P6 → the procedure P7
P7 → + of admission

(3) I think that dress reform for women which seems to mean ugly clothes must always originate with plain women who want to make all other women look plain.

(*a*) . ⟨I⟩ think ⟨+ that ⟨⟨⟨dress reform ⟨+ for women⟩⟩ ⟨⟨⟨+ which⟩ seems ⟨+ to ⟨mean ⟨ugly clothes⟩⟩⟩⟩⟩⟩ must ⟨always⟩ originate ⟨+ with plain women ⟨⟨⟨+ who⟩ went ⟨+ to ⟨make ⟨⟨⟨all other women⟩ look ⟨plain⟩⟩⟩⟩⟩⟩⟩⟩⟩.

We then have:

S → P1 V1 P2
P1 → I
V1 → think
P2 → + that C1
C1 → P3 V2 P4 P5

P3 → P5′ P6′
P5 → dress reform P6
P6 → + for women
P6′ → C2
C2 → P7 V3 P8
P7 → + which
V3 → seems
P8 → + to C3
C3 → V4 P9
V4 → mean
P9 → ugly clothes
V2 → must originate
P4 → always
P5 → + with plain women P5″
P5″ → C4
C4 → P10 V5 P11
P10 → + who
V5 → want
P11 → + to C5
C5 → V6 P12
V6 → make
P12 → C6
C6 → P13 V7 P14
P13 → everyone else
V7 → look
P14 → plain

In the above analysis, V3 ' seems ' and V5 ' want ' have been treated in such a w that they are required to have a P-structure (when seen ' from without ') to go with the as a complement, similar to:

. ⟨He⟩ seems ⟨the ringleader⟩.

and

. ⟨I⟩ want ⟨bread⟩.

(*b*) In an alternative analysis ' (seems ⟨to⟩) mean ', ' (want ⟨to⟩) make ' may be con dered as verb-phrases, resulting in a different structural interpretation with or witho a different semantic slant.

In this case the Verb-phrase grammar must include such phrases as: (*have* ⟨*to*⟩) go, (*seem* ⟨*to*⟩) mean, (*want* ⟨*to*⟩) make, etc., in the same way as *must* go, *may* mean, *will* make. etc.. in order that the Sentence or C-structure should contain one and only one verb phrase as one of its immediate inner components. (But see, however, part IV, Sections 2 and 5).

From a practical point of view, nevertheless, it should be possible (within properly worked out limits) for a pre-editor to choose his alternatives in a flexible system of sentence interpretation.

(4) The farmer follows luck and his forefathers.

(*a*) . ⟨The farmer⟩ follows ⟨⟨luck⟩ =and= ⟨his forefathers⟩⟩.

This gives:

S → P1 V1 P2

P1 → the farmer

V1 → follows

P2 → P3 =and= P4

P3 → luck

P4 → his forefathers

It would appear that S → S1 =and=S2, where S1 → P1 V1 P3 and S2 → P1 V1 P4.

So that :

S → S1 =and= S2

→ ((P1 V1 P3)) =and= ((P1 V1 P4)).

Taking the common elements outside the ((algebraic)) brackets, we get:

S→ P1 V1 ((P3 =and P4))

→ P1 V1 P2.

(*b*) (Perhaps the above analysis is the only possible one for a simple structure like this.)

(5) Failure is only the opportunity more intelligently to begin again.

(*a*) ⟨Failure⟩ is ⟨only the opportunity ⟨–⟨more intelligently⟩– + to (begin ⟨again⟩)⟩⟩.

This gives:

S → P1 V1 P2

P1 → failure

V1 → is

P2 → only the opportunity P3′

P3′ → P3 + to C1

P3 → more intelligently

C1 → V2 P4

V2 → begin

P4 → again

(*b*) ⟨Failure⟩ is ⟨only⟩ ⟨the opportunity ⟨ – ⟨more intelligently⟩– + to ⟨begi ⟨again⟩⟩⟩⟩.

In this analysis:

S → P1 V1 P2 P3

(6) Whosoever does a thing best ought to be the one to do it.

(*a*) . ⟨⟨⟨+ Whosoever⟩ does ⟨a thing⟩ ⟨best⟩⟩⟩ (ought ⟨to⟩) be ⟨the one ⟨+ to ⟨d ⟨it⟩⟩⟩⟩.

Here we have:

S → P1 V1 P2

P1→ C1

C1 → P3 V2 P4 P5

P3 → + whosoever

V2 → does

P4 → a thing

P5 → best

V1 → ought to be

P2 → the one P2′

P2′ → + to C2

C2 → V3 P6

V3 → do

P6 → it

On the basis of Part IV, Sections 2 and 5, 'ought to be' could be analysed as: ought ⟨+ to (be)⟩.

Accordingly, we have:

(*b*) ⟨(⟨+ Whosoever⟩ does ⟨a thing⟩ ⟨best⟩)⟩ ought ⟨+ to (be ⟨the one ⟨+ to (do ⟨it⟩)⟩⟩)⟩.

This gives:

S → P1 V1 P2

P1 → C1

C1 → P3 V2 P4 P5

P3 → + Whosoever

V2 → does

P4 → a thing

P5 → best

V1 → ought

P2 → + to C2

C2 → be P6

P6 → the one P7

P7 → + to C3

C3 → do P8

P8 → it.

(7) Life, as I see it, is not a location but a journey.

(*a*) . ⟨Life - ⟨+ as (⟨I⟩ see ⟨it⟩)⟩ -⟩ is ⟨+ not ⟨a location⟩ =but= ⟨a journey⟩⟩.

S → P1 V1 P2

P1 → Life P3

P3 → + as C1

C1 → P4 V2 P5

P4 → I

V2 → see

P5 → it

V1 → is

P2 → + not P6 =but= P7

P6 → a location

P7 → a journey

Here perhaps the relation between ' + not ' and ' =but= ' is a logical one such as:

$\neg p \wedge q$.

Compare the following logical formulae with the corresponding ordinary language:

$p \wedge q$	sweet *and* tasty
$\neg p \wedge \neg q$	*not* sweet *and* tasty (that is *neither* sweet *nor* tasty).
$\neg p \wedge q$	*not* sweet *but* tasty
$p \wedge \neg q$	sweet *but not* tasty.

(*b*) . ⟨Life⟩ -⟨+ as (⟨I⟩ see ⟨it⟩)⟩ - is ⟨+ not ⟨a location⟩ =but= ⟨a journey⟩⟩.

In this analysis:

S → P1 P2 V1 P3

P1 → life

P2 → + as C1

C1 → P4 V2 P5

P4 → I

V2 → see

P5 → it

V1 → is

P3 → + not P6 =but= P7

P6 → a location

P7 → a journey.

In 7 (*a*) ' it ' refers to ' life ', whereas in 7 (*b*) it refers to the speaker's whole statement about ' life '. This seems to be a subtle semantic slant given in the alternative analysis. Either alternative is equally tenable. The pre-editor has this flexibility of choice.

4.1.9. *The complementarity of the dictionary and grammar in a flexible system of analysis*

The few examples given above have shown that semantic interpretation of a whole sentence cannot be a rigid one. (A whole situation could be viewed as a set of relations, as a

process or as a thing. The sentence ' He read the book ' and the noun phrase nominalized from it, ' His reading of the book ', refer to the same situation. In one case, it refers to an action completed. In the other, it is almost treated as a thing about which something more could be said, that is, in relation to which other relations or processes could be made explicit.) Even when only one type of syntactic structure is used, there could be alternative interpretations of its constituents.

The speaker (writer) may say something semantically. The listener (reader) may interpret or understand it semantically with a different slant. Further different readers may read with different semantic slants. Any syntactic sentence, therefore, is at once a number of different semantic sentences. A reader or writer chooses *one* semantic sentence when he reads or writes. When he re-reads perhaps he would find *another* semantic sentence as an alternative choice within the same syntactic sentence. This is probably one of the factors contributing to the fact that we are able to read new meaning into something we had already read earlier.

From the point of view of mechanical translation and in the light of this flexibility available to the pre-editor, the system that we develop should be capable of:

(1) including something in the dictionary that has been left out of the grammar in one analysis, and of

(2) including in the dictionary the same thing in a different way, if a feature of it has been accounted for in the grammar according to an alternative analysis.

We give examples below[1]:

1. (*a*) . ⟨I⟩ want ⟨+ to (look ⟨+ at it⟩ ⟨this way⟩)⟩.

   (*b*) . ⟨I⟩ want ⟨+ to ((look ⟨at⟩) ⟨it⟩ ⟨this way⟩)⟩.

2. (*a*) . ⟨I⟩ (want ⟨to⟩) look ⟨+ at it⟩ ⟨this way⟩.

   (*b*) . ⟨I⟩ (want ⟨to⟩) (look ⟨at⟩) ⟨it⟩ ⟨this way⟩.

The respective analyses would be:

1. (*a*) S → P1 V1 P2

   P1 → I

   V1 → want

   P2 → + to C1

   C1 → V2 P3 P4

1. *Vide* ref. 4, for a treatment of conjunct verbs, conjunct auxiliaries and of the sentence as a conjunct verb.

V2 → look

P3 → + at it

P4 → this way

1. (*b*) S → P1 V1 P2

P1 → I

V1 → want

P2 → + to C1

C1 → V2 P3 P4

V2 → (look ⟨at⟩)

P3 → it

P4 → this way

Here V2 is ' look at ', and consequently P3 represents the direct object of this verb syntactically and semantically (P3 being ' it '), whereas in 1. (*a*) above P3 was a locational or directional case element marked by the marker ' + at '. (One could compare this situation with the Russian idiom of using the verb with the accusative case indicating direction of the action:

' Ja khotel by smotretj *na èto* tak '

' I would like to look *at it* this way '.

2. (*a*) S → P1 V1 P2 P3

P1 → I

V1 → (want ⟨to⟩) look

P2 → + at it

P3 → this way

Here V1 is not a simple verb but a compound verb made up of two lexical verbs but treated as one verb phrase, the first lexical verb ' want ' being treated as a modal auxiliary. Thus such structures have to find a place in a verb phrase grammar. If the verb phrase grammar does not deal with such structures, then ' want to look ' should be a lexical item.

2. (*b*) S → P1 V1 P2 P3

P1 → I

V1 → (want ⟨to⟩) (look ⟨at⟩)

P2 → it

P3 → this way

The pre-editor is not a rigid system, but a human being. He is likely to analyse the same sentence differently at different times, thereby imposing a grouping of elements respectively into the grammar and the dictionary in different ways.

The grammar and the dictionary, therefore, must be flexible enough to accommodate the elements resulting from different alternative analyses in the respective places.

The pre-editor further is doing a quick editorial job, even when he is a trained linguist, taking *ad hoc* decisions (in accordance with some general guidelines) about the groupings of words in the source language sentences. [He is *not* a cranky linguist doing hairsplitting analysis towards writing an erudite dissertation for his Ph.D. degree (under the guidance of an equally cranky Professor), either already from an ivory tower or aiming ultimately to reach there one day to be able to continue his academic pursuits from above the clouds].

Therefore, his reference grammar and lexicon must provide him with alternatives for his *ad hoc* decisions according to general guidelines. He should be able to work with surface structures and should not have to worry about rigid and grandiose tree-structures with their left and right branchings.

4.1.10. *Problems and prospects*

As in the case of space travel, which was considered by many experts to be impossible only three decades ago (for it was said that an error of 1/2° in the launching angle would result in a space vehicle going off the mark by thousands of miles), mechanical translation has also been declared impossible.

With space launching taken as an engineering problem providing for techniques of correction in trajectory and restriction of the goal of space launching, a given space mission has been demonstrated to be quite within the realm of possibilities.

In similar way, suitable restrictions of goals, suitable mechanisms of corrections during the process and suitable initial processing by a human pre-editor, etc., could lead to partially mechanized translations.

Nobody has seriously considered the possibility or impossibility of mechanical essay writing, even as a unilingual venture. Writing in one language, revising, correcting proof, etc., are all done by human agency. Pre-editing of the unilingual text could be added to this list.

The translation process then would be a step by step, structure for structure, comparison operation looking for clichés, idioms, clauses, phrases and finally words. Substitutions have to be made as per alternative programmes at each stage. The final product could be stylistically either raw or polished. If raw, a post-editor (unilingual) could polish it unilingually in the target language.

The translation activity should be restricted to slices of subject areas, within which (when intended for a given type of audience under given conditions of delivery) terminology, style, usages and conventions are fairly uniform.

Our attempts are directed towards such restricted goals and objectives. Our methods, though based on some (if work-a-day) theory of language structure, are also flexible and rough-and-ready with plans for incorporating a self-corrective process at every step.

More of our work and results would be reported elsewhere.

PART IV

SECTION 2

English verbs

In accordance with the definition of an abstract *sentence* in a language as a *proposition* with one *Verb* and several *arguments*, presented elsewhere (*see* Part III and Part IV, Section 5), given by the formula:

$$S' \rightarrow {}^{*}S$$

where

$$S \rightarrow \langle P \rangle V$$

(S' being the *sentence*, * a set of *operations* or *markers*, P the set of *arguments* and V the *verb* or *predicate*), we are required to establish the identity of a verb in a sentence at the syntactic level (in any language). Such a requirement has led us to the following classification of English verbs:

4.2.1. *English verbs* (V_e)

The English verbs fall into two main categories:

(1) *Lexical verbs*: (V_l)

(*a*) Link verb (V_k)

(*b*) Intransitive verb (V_i)

(*c*) Transitive verb (V_t)

(2) *Auxiliary verbs* (V_x):

(i) Modal auxiliaries (V_m)

(ii) Aspect and voice auxiliaries (V_n)

where:

$$V_l \rightarrow \begin{Bmatrix} V_k \\ V_i \\ V_t \end{Bmatrix}$$

$$V_x \rightarrow \begin{Bmatrix} V_m \\ V_n \end{Bmatrix}$$

and

$V_k \rightarrow$ the link verb *be* (*i.e.*, BE_k)

$V_i \rightarrow$ {an intransitive verb like *go*, *run*, *sleep*, *rise*, etc.}

$V_t \rightarrow$ {a transitive verb like *read*, *give*, *take*, *do* (*i.e.*, DO_t), *have* (*i.e.*, $HAVE_t$), etc.}

$V_m \rightarrow$ {*shall*, *will*, *may*, *can*, *must*,..}[1a]

$$V_n \rightarrow \begin{Bmatrix} V'_n \\ be\ (i.e.,\ BE_n) \end{Bmatrix}$$

$$V'_n \rightarrow \begin{Bmatrix} have\ (i.e.,\ HAVE_n) \\ be\ (i.e.,\ BE_n) \end{Bmatrix}$$

4.2.2. *The five-slot verb-phrase of English*

The syntax of English verbs exhibits a high degree of regularity of patternment. The most complex structure of the English verb (a ' verb phrase ' in a non-Chomskian sense, representing a phrase that consists only of verb forms with *one* lexical verb and its auxiliaries) has five verb-slots: V_1 V_2 V_3 V_4 V_5.

Case 1 : *Active voice*

The modal auxiliaries (V_m) occupy the slot V_1 and no other slot.

In the absence of an 'aspect or voice auxiliary' (V_n), the lexical verb (V_t) occupies the slot V_2. The ' perfective aspect auxiliary ' $HAVE_n$, if present, takes the V_2 position and in that case the lexical verb V_t is pushed to the V_3 position. Instead, if there is the ' continuous aspect auxiliary ' BE_n present, it takes the V_2 position and in that case the lexical verb V_t is pushed to the V_4 position. If both $HAVE_n$ and BE_n are present, $HAVE_n$, BE_n, and V_t respectively occupy the V_2, V_3 and V_4 positions.

1a. V_m includes sub-classes of conjunct auxiliaries like (have ⟨to⟩), (went ⟨to⟩), (ought ⟨to⟩), etc.

If the slot V_1 is occupied (by a V_m), then the tense marker is attached to that. If V_1 is vacant, then the tense marker is attached to anything that occupies the V_2 position. The verb in the V_3 position always takes the *-en* morpheme and that in the V_4 position the *-ing* morpheme.

Case 2 : Passive voice

The lexical verb V_t goes to and stays throughout in the V_5 slot. The places it occupied in the active voice are now taken over by the ' passive auxiliary ' BE_p.

The lexical verb occupying the V_5 slot takes the *-en* morpheme.

4.2.3. *Tense, aspect and mood in English*

(1) English, in our present (unorthodox) view, has only *two tenses*: the *present* and the *past*.

(2) It has the *perfect* (vs. *imperfect*) and *continuous* (vs. *non-continuous*) *aspects.*

What is generally known (in English) as the ' future tense ' is, again in the present unorthodox view, no tense at all[1], but may be called the ' intentional mood '; (Volitional intention: *I will go*; non-volitional intention: *I shall go*). This ' intentional modal form ' could only be in one of the two tenses: present tense (' I will go ', ' I shall go ') or past tense (' I would go ', ' I should go '). In other words, the ' modal auxiliaries ' (like the lexical verbs used without an auxiliary) also have only the present and past forms.

This unconventional way of describing the English tenses seems to be inescapable. However, the ' intentional modal auxiliary ' used with the infinitive of the ' lexical verb ' corresponds to the *future tense* of other languages and hence it is traditionally taken to be a *tense* rather than a mood.

Other moods could be[2] :

(1) ' neutral ' or ' indicative ' mood: (no auxiliary)

(2) ' permissive ' mood: (*may*)

(3) ' capacitive ' mood: (*can*)

1. *See* ref. 7, wherein the author rejects the ' future tense ' as a separate tense in English.

2. The ' emphatic ' *do* and the ' dummy verb ' *do* (V_0' in Part III) occurring in the interrogative and negative sentences are not considered by us as auxiliaries in the sense described in this paper.

They behave slightly differently from the ' auxiliaries ' of this paper. Further, the ' emphatic ' form can occur in the ' imperative ' (' *Do listen to me, please*! ') while the ' modal auxiliaries ' (V_m) cannot.

We may also note that ' emphasis ' is not purely a phenomenon attached to the *verb*. It could, like the ' interrogative ' or ' negative ', be attached to any component of the ' proposition ', that is, to the ' predicate ' or any of the ' arguments '. We shall therefore deal with ' emphasis ' as a phenomenon of ' propositional transformation '.

and

(4) 'obligational' mood: (must).

4.2.4. *The five-slot verb-table*

The slots V_1, V_2, V_3, V_4 and V_5 stand for the following types of fillers:

V_1 — the modal auxiliaries (V_m)

V_2 — the lexical verb (V_t) or the aspect or voice aux. (V_s)

V_3 — the lexical verb (V_t) or BE_n (with morpheme *-en*)

V_4 — the lexical verb (V_t) or BE_g (with morpheme *-ing*)

V_5 — the lexical verb (V_t) (with morpheme *-en*).

In the table below are shown what different types of verbs and auxiliaries fill the slots and in what way. If T indicates the tense marker, we have the following verb phrases:

The Five Slot Verb-Table

No.	V_1	V_2	V_3	V_4	V_5
I. 1.		V_k–T			
2.		$HAVE_n$–T	V_k–en		
3.	V_m–T	V_k			
4.	V_m–T	$HAVE_n$	V_k–en		
II. 1.		V_t–T			
2.		$HAVE_n$–T	V_t–en		
3.		BE_n–T		V_t–ing	
4.		$HAVE_n$–T	BE_n–en	V_t–ing	
5.	V_m–T	V_t			
6.	V_m–T	$HAVE_n$	V_t–en		
7.	V_m–T	BE_n		V_t–ing	
8.	V_m–T	$HAVE_n$	BE_n–en	V_t–ing	
III. (*a*)					
1.		V_t–T			
2.		$HAVE_n$–T	V_t–en		
3.		BE_n–T		V_t–ing	

No.	V_1	V_2	V_3	V_4	V_5
4.		$HAVE_n$–T	BE_n–en	V_t–ing	
5.	V_m–T	V_t			
6.	V_m–T	$HAVE_n$	V_t–en		
7.	V_m–T	BE_n		V_t–ing	
8.	V_m–T	$HAVE_n$	BE_n–en	V_t–ing	
III. (*b*)					
1.		BE_p–T			V_t–en
2.		$HAVE_n$–T	BE_p–en		V_t–en
3.		BE_n–T		BE_p–ing	V_t–en
4.		$HAVE_n$–T	BE_n–en	BE_p–ing	V_t–en
5.	V_m–T	BE_p			V_t–en
6.	V_m–T	$HAVE_n$	BE_p–en		V_t–en
7.	V_m–T	BE_n		BE_p–ing	V_t–en
8.	V_m–T	$HAVE_n$	BE_n–en	BE_p–ing	V_t–en

4.2.5 *Examples*

I. $V_t \rightarrow$ be

1. $\left\{\begin{matrix}\text{am}\\\text{is}\\\text{are}\\\text{was}\\\text{were}\end{matrix}\right\}$

2. $\left\{\begin{matrix}\text{have}\\\text{has}\\\text{had}\end{matrix}\right\}$ been

3. $\left\{\begin{matrix}\text{Shall}\\\text{will}\\\text{can}\\\text{may}\\\text{must}\\\text{should}\\\text{would}\\\text{could}\\\text{might}\end{matrix}\right\}$ be

No.	V_1	V_2	V_3	V_4	V_5
4.	shall / will / can / may / must / should / would / could / might	have	been		

II. $V_i \rightarrow$ go

1.	go / goes / went				
2.	have / has / had		gone		
3.	am / is / are / was / were			going	
4.	have / has / had		been	going	
5.	shall / will / can / may / must / should / would / could / might	go			

No.	V_1	V_2	V_3	V_4	V_5
6.	shall / will / can / may / must / should / would / could / might	have	gone		
7.	shall / will / can / may / must / should / would / could / might	be		going	
8.	shall / will / can / may / must / should / would / could / might	have	been	going	

III. (*a*)

$V_t \rightarrow$ give

1.	give / gives / gave	
2.	have / has / had	given

No.	V_1	V_2	V_3	V_4	V_5
3.		am / is / are / was / were		giving	
4.		have / has / had	been	giving	
5.	shall / will / can / may / must / should / would / could / might		give		
6.	shall / will / can / may / must / should / would / could / might		have	given	
7.	shall / will / can / may / must / should / would / could / might		be	giving	

No.	V_1	V_2	V_3	V_4	V_5
8.	shall / will / can / may / must / should / would / could / might	have	been	giving	

III. (*b*)

$V_t \rightarrow$ give

1.	am / is / are / was / were				given
2.	have / has / had	been			given
3.	am / is / are / was / were			being	given
4.	have / has / had	been		being	given
5.	shall / will / can / may / must / should / would / could / might	be			given

No.	V_1	V_2	V_3	V_4	V_5
6.	shall / will / can / may / must / should / would / could / might	have	been		given
7.	shall / will / can / may / must / should / would / could / might	be		being	given
8.	shall / will / can / may / must / should / would / could / might	have	been	being	given[3]

4.2.6. *Generative-transformational rules for the English Verb*

It is seen from the above tables that the English finite verb, the VERB PHRASE in our terminology, may consist of one lexical verb and several auxiliary verbs, in different combinations. Their occurrences seem to be amenable to rules of generation and transformation. Accordingly, we give below a preliminary set of rules:

Among the units X, Y, Z, W described below, the distribution is as follows:

V_1 position: X T only.

V_2 position: W T, Y T, Z T, W ϕ, Y ϕ, Z ϕ.

3. *Vide* refs. 7 and 6, p. 150, for an attestation of these peculiar forms.

V_3 position: W–en, Z–en.

V_4 position: W–ing.

When W has two components, the second component always occupies the V_5 position.

Rules[4] :

(1) V → $\left\{\begin{matrix}(X) & W\\ (X)\ Y & W\\ (K)\ Z & W\end{matrix}\right\}$ E

(2) (X) W → $\left\{\begin{matrix}W\\ X\ W\end{matrix}\right\}$

(3) $\left[\begin{matrix}W\\ X\ W\end{matrix}\right]$ E → $\left[\begin{matrix}W\ T\\ X\ E\ W\text{–}\phi\end{matrix}\right]$

(4) $\left[\begin{matrix}(X)\ Y\ W)\\ (K)\ Z\ W\end{matrix}\right]$ E → $\left[\begin{matrix}(X)\ Y\ E\ W\text{–en}\\ (K)\ Z\ E\ W\text{–ing}\end{matrix}\right]$

(5) (K) Z → $\left\{\begin{matrix}Z\\ X\ Z\\ (X)\ Y\ Z\end{matrix}\right\}$

(6) $\left[\begin{matrix}Z\\ X\ Z\\ (X)\ Y\ Z\end{matrix}\right]$ E → $\left[\begin{matrix}Z\ T\\ X\ E\ Z\text{–}\phi\\ (X)\ Y\ E\ Z\text{–en}\end{matrix}\right]$

(7) (X) Y → $\left[\begin{matrix}Y\\ X\ Y\end{matrix}\right]$

(8) $\left[\begin{matrix}Y\\ X\ Y\end{matrix}\right]$ → $\left[\begin{matrix}Y\ T\\ X\ E\ Y\phi\end{matrix}\right]$

(9) X E → X T

(10) W $\left[\begin{matrix}T\\ \phi\\ en\\ ing\end{matrix}\right]$ → $\left\{\begin{matrix}A & \left[\begin{matrix}T\\ \phi\\ en\\ ing\end{matrix}\right]\\ B & \left[\begin{matrix}T\\ \phi\\ en\\ ing\end{matrix}\right] C\text{–en}\end{matrix}\right.$

4. The bracketing convention adopted here is the one described in ref. 1, pp. 17 ff.

(11) A → V_t

(12) B → BE_p

(13) C → V_t

(14) V_t → $\left\{ \begin{matrix} V_k \\ V_i \\ V_t \end{matrix} \right\}$

(15) V_k → BE_k

(16) V_i → ' intransitive verb '

(17) V_t → ' transitive verb '

(18) Z → BE_n

(19) Y → $HAVE_n$

(20) X → V_m (' modal auxiliary ')

(21) V_m → {*shall*, *will*, *may*, *can*, *must*}

(22) $HAVE_n$ → *have*

(23) $\left\{ \begin{matrix} BE_k \\ BE_n \\ BE_p \end{matrix} \right\}$ → *be*

(24) T → $\left\{ \begin{matrix} \text{' past '} \\ \text{' present '} \end{matrix} \right\}$

This, however, is only a tentative set of rules. Our theory of the ' proposition ' includes, in addition to the concept of the ' sentence ', the concept of a ' clause ', which takes not only the finite verb but also the ' infinitive ' and the ' participles ' (the ' clause ' in this sense being called the C-structure). The ' sentence ' too takes the ' imperative ' form of the verb. These would necessitate a slight modification in the above set of rules.

Discussion of these questions would be taken up in a subsequent paper that would deal with ' operations ' performed on ' propositions ' to get different types of ' sentences ' (See also Part III).

PART IV

SECTION 3

Some linguistic problems of machine translation*

4.3. In this Section of this Part a rapid survey of the newly evolving field of *Machine Translation* of languages is made with particular reference to the linguistic problems involved.

* This Section is the original English version of its published Hindi translation (ref. 2).

This new field of enquiry, like most other new fields of science which have grown rapidly during and after the last world war, is typical for its many-faceted character of having to tackle simultaneously problems in widely different fields such as those in linguistics on the one hand and those in electronic computer hardware and programming on the other in this particular case.

From an economic point of view this field has to deal with questions like time spent for human intellectual labour and time gained by employing mechanical analogues for the intellectual processes of man in making a translation from one language into another. The scope and limits of machine translation are also indicated.

4.3.1. *Mechanization and automation of physical labour*

In most fields of human endeavour one can observe a certain type of revolutionary movement taking place in a way that is characteristic of the present epoch.

The early industrial revolution brought in gadgets, machines and engines that effected a saving on man's *physical* labour and time. In all industrial and other fields of man's physical labour there began the revolutionary movement of *mechanization*. In a mechanized industry man has ever to be observing the working of the machine and controlling it as and when required to ensure an output according to required standards. His vigilance is also necessary to ensure an output which is uniform over any length of time. Thus, although man had largely been relieved of a good deal of physical exertion, still he had to go through a certain degree of *repetitive and routine operations* to keep the machine running in a specified manner.

The ultra-rapid hustle and bustle of industry and commerce of the present day has led to the attempts at the saving of energy and time involved in the repetitive and routine operations on the part of man. In achieving this, such developments have been brought in as to make the machine control itself, check its own output and make it correspond to predetermined specifications without the intervention of man. Thus, from mere *mechanization* of the previous decades, man has already started moving in the direction of *automation*.

4.3.2. *Mechanization and automation of mental processes*

Repetitive and routine *physical* operations of man have thus been largely taken over by machines in most modern undertakings. What about repetitive and routine *mental* operations? Even with regard to man's *mental* operations, particularly in the field of numerical computations, the processes of *mechanization* and *automation* are clearly evident to various degrees of perfection.

The latest field of man's mental activity to go through the preliminary stages of mechanization and automation is the field of *translation* from one language into another.

4.3.3. *Scope of mechanical translation*

Unlike ordinary numerical computers the mechanical translating machines must possess a large capacity to store linguistic data; they should be capable of performing *automatically* a great many different types of operations of analysis on the linguistic material fed into them. They should also store the results of such operations at each stage, select for the *target language* linguistic elements and processes corresponding to those of the *source language* from those stored linguistic materials and results of the analysing operations.

These requirements are still far from being satisfactorily fulfilled on the electronic engineering side. The solution of all the engineering problems connected with the process of automatic translation also depends to a large extent on the knowledge of linguistic structure that could be made in explicit form for the two languages involved in translation. One has to know in advance the linguistic methods of pairing elements and operations of one language with the elements and operations of the other language for any intelligible translation.

Now, three important questions arise:

(1) Is it possible to construct a machine which has as large a storage and operational capacity as we need ?

(2) Is it economically worthwhile (both with respect to expenditure of money and of time in doing all the preliminary building and operating work of the machine) when we consider the maximum speed and effectiveness of the translation output obtained from the machine ?

(3) How far can language be mechanically analysed and what kinds of material can be mechanically translated ? Between what kinds of languages translation could be mechanized ?

These are not questions that could be easily answered.

In this discussion we shall omit from consideration the first two questions, not because they are not important, but because we are concerned here only with the linguistic questions involved.

It is obvious that poetic, philosophical, religious and other types of material, which require for their translation a wide cultural background and not merely a few logical steps, cannot be mechanically translated. In fact they cannot be effectively translated even by an experienced human translator if he does not possess the neccssary cultural background and creative ability.

We have therefore to restrict our attention to matter-of-fact material involving largely logical steps that lead from one observation to another as perhaps in mathematics and the natural sciences.

Widely differing languages like say any one of the Indian languages and any one of the American Indian languages, which differ both in regard to their structure and in the manner of describing even the experiences of ordinary physical situations, cannot be handled mechanically for translations with respect to one another.

There remain thus only those languages that have more or less similar methods of describing experiences connected with ordinary physical situations.

Between any two such languages (as, for example, English and French, Hindi and Marathi or Tamil and Kannada), it now remains for us to see what method we have to and can adopt for working out schemes of mechanical translation.

4.3.4. *Problems of linguistic analysis*

There are innumerable difficulties even when we restrict the scope of our translations to similar languages and to specific fields of science.

In the very first instance we realise that no two languages are exactly alike. No word for word or even phrase for phrase substitution is likely to yield a translation. A word in one language may correspond to a phrase in another language. More fundamental than this, a grammatical inflexion in one language (as, for instance, in Sanskrit, Hindi, Marathi or Tamil) may correspond to word order in another language (like English or Chinese).

A further difficulty is that for any of the above aspects of grammar in one language there is invariably no corresponding substitute in another language. In cases where there are corresponding substitutes there is invariably no one-one correspondence. Correspondences, if present, are usually of the one-many or many-one type.

A grasp of these questions is sufficient to indicate that man and not machine is best suited for any translation work.

Our interest in machine translation is in the attempt at the discovery of the limit to which we can go to make the process of analysing the linguistic structure of a language *automatic*. Once again there are two ways of looking at the problem: the utilitarian way and the academic way. The utilitarian way does not bother as to whether any particular method employed leads to logically perfect results. 'Intelligible translations' alone are expected. That is, serious grammatical mistakes such as omissions and additions of grammatical elements, wrong or bad constructions, bad idioms, etc., do not

matter from this point of view. The translated version should just make the desired sense. No other question has any relevance.

If we look at the question of the automatic analysis of language from an academic point of view, our requirements are more stringent. We have to account for every result that we get. One exception, even one, could be a weak spot. Linguists point out that language is not a closed system. That is, it is not a system in which every new occurrence is an already known one. On the other hand language is known to be a growing dynamic system in which one has to be prepared to encounter elements and constructions that one has never met with before.

Mechanical translations could thus at the most deal with what elements and constructions have already been known or have already occurred. They cannot possibly deal with what elements and constructions are yet to come in the future, unless the computer is developed into a system that *learns* as it functions. Here is therefore another restriction on the scope of attempts at mechanical translation.

4.3.5. *Linguistic analysis by man* vs. *linguistic analysis by machine*

Bearing all these limitations in mind, we have, therefore, to see how far the existing methods of analysis of linguistic structure could be put into an automatic framework within the limits specified, and how far one could make a departure from established methods of linguistic analysis and in what way some of the serious difficulties could be solved or at least circumvented.

So far as the analysis of *one* language (or one dialect or idiolect) is concerned the traditional divisions of phonology, morphology, syntax and semantics may for convenience be considered as separate *fields* (or *levels* as they are normally referred to) and separately dealt with in a formal manner as is done in descriptive linguistics. Even in such a case no one ever makes the mistake of considering these fields as separate water-tight compartments. Indeed such formal analyses in all these different levels are conspicuous by the fact that no two descriptive statements, by different analysts, about even the same dialect or idiolect are identically the same. A closer examination of such descriptive statements reveals the fact that, ultimately, each analyst makes his own selections of the descriptive elements in a manner that is conditioned by an arbitrary personal preference, although all this takes place largely unconsciously. The rest of his description is formal to the extent permitted by his original selections.

So far as the formal part of his description goes, an analyst can justify his belief that he does not go by extraneous considerations of *meaning*, etc. In making some of his original choices, however, he is guided (particularly in the fields of morphology and syntax) entirely by implicit considerations of *meaning*.

A machine which has to do the classifications of morphemes, grammatical constructs, and so on, cannot by itself indulge in *implicit* considerations of meaning. Such considerations, for a machine, have to be *explicit operations*. In other words, if a particular process in our analysis is not purely formal, we have to make an *explicit* statement as to what implicit considerations are involved in arriving at a particular result in our analysis.

If this is done it would be possible to programme a machine in such a way that it could automatically analyse a given sentence in terms of its own built-in lexicon, morpheme inventory, morphological and syntactical rules and in terms of its own built-in inventory of other grammatical elements (that go by the names of parts of speech, markers, structural features, etc.).

This is as far as we can go with regard to the analysis by the machine of a sentence in any one given dialect by feeding the sentence into the machine in the form of a suitably coded input. Although all levels are involved in this, we may still manage to keep them apart so long as we confine ourselves to the analysis of one particular language or dialect.

4.3.6. *Analytical equivalents in a two-language situation*

The moment we try to compare or establish correspondences between the elements and structural features of a sentence in one language with those of an equivalent sentence in another language, all the differentiation of these *levels* crumbles. At any rate a level to level operation for establishing the necessary correspondences comes into being.

In this short and rapid survey of the field of machine translation the author does not wish to go into the details of the methods he has adopted in working out schemes of making *explicit* the formal steps to be taken in translating from one language into another. It has to be pointed out that no 'universal' scheme could ever be developed for translation between any language and any other. Schemes of formal translation could only be developed for any two specific languages at a time. Further, if a scheme is developed for formal translation from language A to language B, it cannot be applied for the reverse translation from language B to language A. A separate scheme would be necessary for it.

4.3.7. *Feasibility and economic worthiness of mechanical translation*

In conclusion, we may briefly touch upon the economic worthiness of such projects of machine translation. It has been estimated (taking into consideration the cost involved in building and operating a machine of such enormity as well as the time and labour involved in devising the specific schemes for particular languages and also the restriction in the scope of the translation) that, unless such a machine could work millions of times

as fast as a human being does in going through the same operations as the machine, a process of this type is likely to be a waste of time, labour and funds.

However, there are also hopes of a reasonable future for such projects, because we now have ultra-rapid electronic computers with much larger memory and operational capacity. These would be fruitful ventures, once the problems connected with the large-scale increase in the storage and operation capacity of the machines and those connected with the formalization of linguistic analysis are solved. The machine, for efficient functioning, must start analysing with the feeding in of the very first elements of a sentence to be analysed. An attempt at the development of a method (based on a corresponding linguistic approach) of such an analysis is now being made by the present author for languages like English, Tamil, Marathi, Hindi and Russian.

An attempt of this kind (even if it ultimately turned out to be unsuitable for the purposes of machine translation), it is hoped, would at least lead to a better understanding of some of the major problems connected with the analysis of linguistic structure.

PART IV

SECTION 4

TECHNICAL TERMS[1]

4.4. Words and technical terms

Ordinary words or lexical items do not, in any language, exactly denote a specific concept. The meaning of the words is mostly dependent on the context.

A technical term, on the other hand, is designed to denote a particular concept even when an immediate context is not available to provide further pointers towards the exact meaning of the term, as required for ordinary words.

4.4.1. *Mixing of common words and technical terms*

Quite often, however, we find that for clear understanding even technical terms do require a context, at least a larger context like the specification of the specialized field in which the terms are used. This necessity arises from the fact that common words like *force*, *energy*, *work*, *power* and *sputnik* (as used in Russian), etc., when used in ordinary parlance do not have a well-defined meaning and whatever meaning is to be attributed to them has to be culled out from the immediate context. At the most they have a qualitative meaning. But in physics, and other branches of science and technology, where physical concepts are used, these very words have a more exact and restricted meaning and (except the word *sputnik* in the above list) denote *quantities*—another common word

1. *See* also ref. 5.

used as a technical term!—that have a *magnitude* and can be *measured* and expressed by a *number*.

4.4.2. *Technical terms having different connotations in different fields*

Even when the common language is excluded from our consideration, we find that frequently there are difficulties in knowing a particular technical term in the absence of a larger context for the simple reason that the same word is used as a technical term in different fields of science denoting different concepts.

For example, we may list here just a few from a large number of terms, such as: *morphology* (biology, linguistics), *plasma* (biology, physics), *square* (geometry, algebra), *quantity* (mathematics, physics, chemistry, linguistics), *length* (geometry, linguistics), *stress* (physics, linguistics), and so on.

4.4.3. *Different concepts denoted by a technical term in the same field of science*

There are technical terms that denote different concepts even when used in the same field of science.

For example; *moment* (*i.e.*, instant, a point of time) and *moment* (as in *moment of inertia*, *bending moment*, *moment of momentum*, etc.).

To understand the meaning of the terms, the immediate context in addition to the specification of the larger context of the field is necessary here.

4.4.4. *Technical terms that are technical terms in their meaning, even when no context is given*

There are just a handful of terms that could be recognised as particular technical terms denoting particular concepts, even when they are merely given as a lexical item without additional specification of context.

The following terms are of this kind: *X-ray*, *radar*, *laser* and *sputnik* (as used in English).

4.4.5. *Degeneration of technical terms by being taken loosely into the common language*

Quite frequently, journalistic handling of technical terms, in contexts far removed from the scientific field in which they have the status of technical terms, leads to degeneration of this type.

For example, the expression ' He searched her very soul with his X-ray eyes' (Fiction), has nothing to do with the X-rays Röntgen discovered, although a component of the

properties of X-rays has been borrowed by the author of the fiction to express something special, probably the idea of ' penetration ' and ' revelation ' of the hidden truth.

4.4.6. *The purity of technical terms*

In the light of the facts noted above, it would be clear that the *purity* of a technical term depends on the principle: ' One term for one concept '. It is also clear that such purity cannot be achieved 100%.

4.4.7. *Language to language differences in technical terms*

Technical terms in one language differ from the corresponding technical terms in another language in the degree of their purity in the sense defined above. For example, the term *sputnik* is purer in English than in Russian, for, in the latter, it is also a common word meaning ' a fellow traveller ', ' a natural satellite of a celestial body ' as well as an ' artificial satellite ', whereas in English it has only the last meaning. Perhaps, in addition, it also has a further restriction in its meaning as ' an artificial satellite of the earth launched into space by the Russians '.

4.4.8. *Systems of technical terms*

In addition to all the other characteristics of technical terms, there is another feature among technical terms. This is the feature of their being members of a system of terms.

We could understand this system as being distributed in a two-dimensional space with the coordinates: linguistic (or grammatical) axis and conceptual axis.

Along the conceptual axis, for example, all the terms in physics, such as *length*, *mass*, *time*, etc., and the systematic combinations of these to form other concepts related to them, give rise to a conceptual system of terms.

Along the grammatical axis, any term standing for a concept varies its form according to the grammatical requirements in the language.

For example, in English, the term *integration* as used in Calculus varies its form grammatically to: *integral*, *integrand*, *to integrate*, *integrable*, *integrability*, *integrating*, etc.

The form *integrating*, however, has lost its mathematical purity in being extended to such expressions as *integrating circuit*. But insofar as it still refers to the concept of ' integration ' in mathematics, its purity is not seriously affected. If, on the other hand, a circuit resembling an ' integrating circuit ' in constructional details is used for performing an operation other than that of mathematical ' integration ', the term *integrating* loses its purity as a technical term.

In any two languages taken for comparison, systems of technical terms do not match 100% either in their grammatical forms or in their correspondence to the conceptual system or in their being used purely as technical terms without any common language undertones.

4.4.9. *The status of loan words as technical terms in another language*

As in the case of the technical term *sputnik* borrowed from Russian (in which the term is also a common word), technical terms could be borrowed as such into other languages, for example, from English into Russian: *laser*, *maser*, *radar*. The terms are purer as technical terms in the target language than in the source language, where some of the terms could be borrowings from the common language (e.g., *force*, *power*).

4.4.10. *Technical terms and expressions*

In addition to technical terms being single words (and therefore are terms properly so called), there could also be *technical expressions* (made of technical terms and associated grammatical devices represented by other words in the language).

For example, in English, we have the term *square* in algebra. The corresponding verb is *to square* (*squared*, *squaring*, etc.). It is still a single word and therefore it is the same technical term in different grammatical forms.

In Russian on the other hand, the noun is represented by *kvadrat*. No verb is derived from this, but in its place an expression made up of a verb and the noun *kvadrat* is used to denote the process of 'squaring' (*vozvyshatj v kvadrat*, *vozvedenie v kvadrat*, etc.).

4.4.11. *Inexactitude in the use of technical terms*

In many languages, especially those that have been recently adopted for the teaching and description of science, there is likely to be a lot of confusion of the following type in addition to all those noted above in previous sections :

1. One concept represented by many alternative terms;
2. Several concepts represented by one term;
3. Several concepts represented by several terms interchangeably.

These defects do exist to some extent even in languages that have been used in the description of science for a very long time (English, German, French and Russian).

Confusion matrices showing the concepts and terms could be prepared for different pairs of languages or for one language to match the terms with concepts.

4.4.12. *The birth and death of concepts in a developing science*

When science grows and newer phenomena are discovered the older concepts are no longer sufficient to explain the newer phenomena. Thus newer concepts are born. For example, the concepts of the *ether*, the *field*, the *plasma*, etc. Of these, the *field* and the *ether* are already being relegated to a secondary position, except in dealing with classical ideas in the subject field.

Thus technical terms are constantly created and discarded in unison with the progress of science. If not totally discarded, they are redefined.

This is a living process, and, in any language, systems of terms have to be critically examined continually or periodically in relation to the field of science. This is especially important for the developing languages of the world which are just being adopted for the teaching and description of science.

For such critical examination, no set of terms in existence could be considered as giving the final picture. The study of the confusion matrices between concepts and terms in one language or between corresponding terms in a pair of languages is likely to reveal large gaps and discrepancies in each language with respect to concepts and, in pairs of languages, with respect to the terms themselves.

PART IV

SECTION 5

Some English parts of speech

In English some parts of speech like 'adverbs', 'prepositions' and 'conjunctions' seem to reveal a sort of ordered relationship among themselves in terms of the composition of the P-structure in which they occur.

4.5.1. *Adverb-preposition relationship*

Let us examine a few pairs of sentences:

⟨He⟩ walked ⟨along⟩ (4.5.1)

⟨He⟩ walked ⟨+ along the road⟩ (4.5.2)

⟨He⟩ went ⟨along ⟨+ with him⟩⟩ (4.5.3)

⟨He⟩ went ⟨+ along-with him⟩ (4.5.4)

In (4.5.1) and (.3) *along* is an 'adverb', in the first case unmodified and the second modified by a 'prepositional phrase' that has an adverbial function ⟨+ with him⟩.

In (.2) and (.4) + *along* and + *along-with* are simple and compound prepositions, respectively[1].

Let us now consider the sentences:

⟨He⟩ apologised ⟨+ for negligence ⟨+ of his duty⟩⟩ (4.5.5)

⟨He⟩ apologised ⟨+ for (neglecting ⟨his duty⟩)⟩ (4.5.6)

⟨He⟩ apologised ⟨+ for (⟨he⟩ neglected ⟨his duty⟩)⟩ (4.5.7)

(⟨He⟩ apologised) = for = (⟨he⟩ neglected ⟨his duty⟩) (4.5.8)

. ⟨He⟩ apologised . ⟨For⟩ ⟨he⟩ neglected ⟨his duty⟩. (4.5.9)

In (.5) + *for* is a ' preposition '. In (.7) + *for* is a ' subordinating conjunction '. In (.6) + *for* is intermediate between a ' preposition ' and a ' subordinating conjunction '. In (.8) = *for* = is a ' coordinating conjunction'. In (.9) ⟨For⟩ is intermediate between a ' coordinating conjunction ' and a sentence modifying ' adverb '.

4.5.2. *The preposition-conjunction relationship*

If there is a marker within a P-structure within which there is no C-structure occurring as its immediate constituent, then that marker is a *preposition*.

If such a marker is accompanied by a C-structure, which is an immediate constituent of the P-structure to which the marker belongs, then:

(*a*) the marker is a *pre-junction*, if the C-structure contains a non-finite verb (*e.g.*, *-ing* form). For, the *-ing* form could be viewed either as a noun-like or a verb-like entity or as both at the same time. If it is noun-like, the marker is to be treated as a *preposition*. If it is verb-like, the marker is to be treated as a sub-class of a ' subordinating conjunction '. If it could be either, or both at the same time, the marker is mid-way between a pure ' preposition ' and a pure ' subordinating conjunction '. We could call it then a *pre-junction*. And

(*b*) the marker is a *sub-junction* (or ' subordinating conjunction '), if the C-structure contains a finite verb, the C-structure being a constituent of a P-structure.

If the whole sentence is made up of two C-structures each with a finite verb and connected by an element that has the function of a ' conjunction ', we could call it a ' *co-junction* '.

If the two C-structures are each considered as a separate sentence and an element like *for*, *since*, etc., is used to connect the two sentences across the sentence boundary, then that element has the status of a ' conjunction ' as well as a sentence modifying ' adverb '. But unlike the sentence modifying adverb, it cannot be shifted in position

1. ' He went along with him ' could also be analysed as:
(⟨He⟩ (went ⟨along⟩) ⟨+ with him⟩), where (went ⟨along⟩)
is a conjunct verb, within which ⟨along⟩ is an adverb.

but must occur as the first element in the second sentence. In such a case, we could call the element an *adverbial-conjunction*.

4.5.3. *The conjunction-adverb relation*

From 4.5.2, it follows that there are some elements which behave both as an adverbial modifier for the whole sentence, and at the same time as a conjunction connecting an idea across the sentence boundary: Such elements are *for*, *because*, *since*, etc., used at the beginning of the second of a pair of sentences. They cannot be shifted in position to any other place in the second sentence.

There are also adverbs, used as sentence modifiers, that could be shifted to other places in the second sentence, and connect an idea across the sentence boundary, like: *therefore*, *consequently*, *fortunately*, etc.

There are adverbs that do not relate to anything across the sentence boundary but are 'arguments' to the 'predicate' within a sentence or 'proposition':

For example:

'The train was moving *slowly*'

Adverbs modify adjectives or other adverbs, as in:

'The trend in the stock market was *cautiously* optimistic'.

Adverbs also modify adverbial phrases, or clauses:

'He could go back to his work *only* when the guests left'.

The above example, however, could be viewed as:

... ⟨only ⟨+ when (⟨the guests⟩ left)⟩⟩,

or as:

... ⟨+ only—when (⟨the guests⟩ left)⟩.

Using the words *adverb*-1 and *adverb*-2 to denote respectively the class of words that could be considered as being represented by *slowly*, *cautiously*, etc., and *therefore*, *consequently*, etc., we could now classify English prepositions, adverbs-1, conjunctions and adverbs-2, as follows.

4.5.4. *Classification of some English parts of speech lying between adverb-1 and adverb-2*

Without lingering to apologise for the unorthodox classification of some English parts of speech, we proceed to give below an ordered classification that is based on the practical system of sentence demarcation developed in Part III, so that we have, very much like

Mendeleev's periodic table, a definite place in a definite order for what we have called: adverb-1, preposition, pre-junction, sub-junction, co-junction, adverbial-conjunction and adverb-2. We don't try to give any theoretical justification for this classification. But we find it to be a comfortable system to take a practical decision about the nature of the elements talked about, and at the same time, we have a certain room for flexibility.

For example, if we should know what is *for* in a sentence, we should ask the question: what is the propositional structure into which it could go ? If there are alternative ways, *for* could be alternatively viewed as one or another part of speech based on its structural location.

Thus we have the following scheme of classification:

Adverb-1: DO_1 (*see* Part III)

It is an element that goes into a P-structure all by itself. At the same time it does not have a meaning connection across a sentence boundary. Examples:

(⟨The train⟩ moved ⟨*slowly*⟩)

(⟨The balloon⟩ flies ⟨*up*⟩)

(Come ⟨*along*⟩)

Preposition : Marker accompanying an NO :

⟨+ *along* the road⟩

⟨+ *in* a slow manner⟩

⟨+ *for* his delay⟩

Conjunction: This falls into three subgroups:

(1) *Pre-junction*: Marker used with a (C′), where (C′) is a C-structure having a *non-finite* verb. Examples:

⟨+ *for* (coming ⟨late⟩)⟩

⟨+ *in* (moving ⟨slowly⟩)⟩

⟨+ *to* (go ⟨there⟩)⟩

(2) *Sub-junction* : Marker used with a (C), where (C) is a C-structure having a finite verb. Examples:

... ⟨+ *for* ((he⟩ came ⟨late⟩)⟩

... ⟨+ *since* ((he⟩ could not do ⟨that⟩)⟩

(3) *Co-junction*: An element that connects two (C)'s within the same sentence. where (C) is a C-structure with a finite verb. Two independent sentences logically connected into a new sentence: $S \rightarrow S_1 = o = S_2$. Examples:

$(C_1) =$ *for* $= (C_2)$:

((He) apologised) = *for* = ((he) came (late))

((He) completed (his work)) = *and* = ((he) was (happy))

(4) *Adverbial-conjunction*: In the second of two successive sentences S_1. S_2, treated as separate sentences, if we have a logical connective that is analysed as belonging structurally only to the second sentence, then that logical connective is:

(1) An *adverbial-conjunction*, since structurally it is an adverb modifying the whole sentence S_2 and logically it is a conjunction across the sentence boundary. Examples:

((He) apologised). ((*For*) (he) was (late)).

((I) know (it)). ((*Because*) (he) told (me) (so)).

(The use of *because* as the first word in a sentence is common in India.)

The adverbial-conjunction is the first element in the second sentence.

A logical connective, which is also a sentence modifier adverb that has a freer choice of place in the second sentence, as opposed to an adverbial-conjunction, and has also a meaning connection across the sentence boundary, is:

(2) *Adverb*–2 : DO_2

Examples:

He, *therefore*, thought it over.

He was, *consequently*, more careful.

It is the peculiarity of the English language that different elements, that could be classified under different headings given above, could have the same *form*, as shown by the word *for*, in the examples discussed earlier.

However, we have a framework for the classification in terms of S, (C) and (P).

In our practical system of analysis and classification, we have to provide for all possible alternative interpretations for one and the same (non-tricky) structure. For, the speaker (or writer) may use an element as, say, a *sub-junction*, whereas the listener (or reader) may interpret it, according to his analysis framework, as (perhaps) an *adverbial-conjunction*.

We believe this flexibility of interpretation is available in any real-life communication situation.

PART V

5. A suggestion towards a syntactic algebra for language

Most formal language descriptions are axiomatic and therefore the theorems and conclusions reached in them, while being mathematically and logically well-founded, are often far removed from *natural language* phenomena.

On the other hand, most linguistic descriptions of natural language, while being linguistically sound, are not normally amenable to any exact mathematical or logical treatment.

This is so, perhaps, because:

(1) The mathematical relations that are used to describe formal language do not necessarily cover all natural language phenomena, and

(2) the natural language relations, that are *implicitly* taken into account by linguists, have not all been explicitly formalised.

In this chapter we would like to make a very modest attempt at formally stating certain linguistic relations and using these to deal with certain syntactic phenomena in the form of a simple 'school algebra'.

5.1. *The natural language sentence and the interrelationship of its e'ements*

Any natural language sentence is formulated by:

(1) an initial psychological dissection of the world of things, events and the relations among them into logical and linguistic categories ('Specification'),

(2) a subsequent logical 'association' of these things, events and their relations into a pre-linguistic 'proposition', and

(3) a final linguistic 'presentation' of this 'proposition' (resulting from the 'associations' of 'specifications') in one of several ways, using certain syntactic devices available in each language.

The elements of a natural language sentence are interrelated in several ways, showing themselves to be grammatically parallel to one another or grammatical abbreviations of larger structures or of a number of separate structures.

For example:

(1) In the linguistically 'presented' relations, we see that some lexical items in different languages take grammatically different forms to 'agree' with some other lexical items

in a sentence, if they have some particular grammatical relationship with them. Thus, (*a*) the form of the verb agrees with the 'person' and 'number' of the subject:

He comes (English)

Ja chitaju, On govorit } (Russian) { 'I read', 'He speaks'

or (*b*) the form of the adjective agrees with the 'gender', 'number' and 'case' of the noun. which it qualifies:

Les beaux arbres, La belle fille } (French) { 'The beautiful trees', 'The beautiful girl'

Ocharovateljnaja devushka, Lingvisticheskie voprosy } (Russian) { 'The charming girl', 'Linguistic problems'

(2) In addition to agreement, lexical items change their forms when some other features of 'presentation' are made explicit morphologically in different languages:

Thus,

(*a*) In almost any language the tense and aspect features of verbs, even when the subject and the lexical verb are unchanged, are represented by differences in form:

I go, I went, I look, I looked } (English)

(*b*) In a language like Japanese the predicative adjective changes its form with respect to tense:

Shiroi	'It is white'	} (Japanese)
Shirokatta	'It was white'	

(3) Some grammatico-lexical items could be grammatical substitutes for other grammatical, lexical or syntactic items or structures:

Thus,

(*a*) John had a book. He gave it to me.

(*b*) I said it was interesting. He didn't think so.

or

(*c*) Everybody writes a message and signs his name in this visitors' book. Why don't you too do so?

Thus we see that between any word in a sentence and some other word in it or in another sentence within a larger text, there are certain types of grammatical relationships that are to be made explicit in order to deal with them more effectively.

(4) There are linguistic ways of 'summing up', that could be done only with particular grammatical or lexical items in particular ways. The implicit relationships that permit this to be done are also to be made explicit.

For example (*a*) :

He came

She came and

It came

could be summed up as:

They came

or

Khoroshij otec

Khoroshaja matj

Khoroshie deti

could be summed up as:

Khoroshaja semjja.

Or (*b*):

I am here today

I shall be here tomorrow

I shall be here even next year

could be summed up as:

I shall always be here.

5.2. *Types of interrelationships among the elements of natural language sentences or among grammatical or lexical items*

If X and Y are elements of a sentence or grammatical or lexical items, we have one or more of the following relationships between them:

(i) $X \equiv Y$

that is X is *semantically* and *formally* the same as *Y*
For example, from

(*a*) She went *there* and I went *there*

we cannot conclude (or ' sum up ') that:

(*b*) She and I went *there*,

unless the word ' there ' means the same (that is, unless $X \equiv Y$, where $X \rightarrow there_1$ and $Y \rightarrow there_2$). If ' there ' in the two statements doesn't refer to the ' same place ', then we could sum it up as, perhaps:

(*c*) She and I went to different places

or

We went to different places.

In other words, ' there ' could be ambiguous in its referential meaning. However, the suggested alternative, ' We went to different places ', is also ambiguous in that it doesn't tell us whether ' We went *together* ' or ' *separately* '. Even when we are informed by context that ' separately ' is meant, we do not know whether ' to different places ' would mean ' to the *same* different places ' or ' to different places that are not the same '.

(ii) $X \not\equiv Y$, *i.e.*, X and Y may have the same form, but they are *semantically* different.

(iii) $X = Y$ that is, X is semantically equivalent to Y but is not formally equal to it.

For example:

If

$X_1 \equiv I$

$X_2 \equiv You$

and if $X \rightarrow X_1 \circ X_2$ (*see* below)

and if $Y \equiv We$, then

we have $X = Y$, that is, ' you and I ' semantically mean the same thing as ' we '. But formally the latter is different from X.

(iv) $X \neq Y$,

that is, X is neither semantically nor formally equivalent to Y, nor is it grammatically parallel to Y (*see* below).

(v) $X \parallel Y$,

that is, X ' is grammatically parallel ' to Y, but not lexically or semantically equivalent to it, that is X and Y have certain features of grammatical agreement (' gender ', ' number ' and ' case ' as between ' adjectives ' and ' nouns ', and ' gender ', ' number ' and ' person ' as between the ' subject ' and the verb in different ways in different languages.

(vi) $X \# Y$,

that is, X is not parallel to Y.

(vii) $X \rightarrow Y$,

that is X ' can be analysed as ' Y (*see* below).

(viii) $X \dashrightarrow Y$,

that is, X ' cannot be analysed as ' Y.

(ix) $X \times Y$,

that is, X and Y have an ' attributive ' relationship to each other, as between an ' adjective ' and a ' noun ', a ' relative clause ' and a ' noun ', etc.

(x) $X \circ Y$.

that is, X and Y have a ' coordinating relationship ', where ' o ' stands for ' and ', ' or ', ' but ', etc.

(xi) $X \cup Y$

that is, X and Y are ' grammatically typical ' (as for example two ' nouns ' with respect to each other).

This relationship is context oriented. For, two nouns N1 and N2 may be typical to each other as against a verb. But they may not be typical to each other in relation to another noun N3. If N1 and N3 are ' countable nouns ' and N2 is ' non-countable ', then $N1 \cup N3$ and $N1 \psi N2$, $N3 \psi N2$, within the contextual field of ' nouns '.

(xii) $X \psi Y$,

that is, X and Y are ' grammatically non-typical '. For example, an ' adjective ' and a ' noun ' are grammatically non-typical with respect to each other.

The relations given above may not be the only ones, they may not be interpretable in the way they have been given here, or they may be redundant with respect to one another. However, without waiting for a more thorough re-examination, we have given these here by way of a preliminary discussion.

5.3. *Natural language sentence and the underlying ' Proposition '*

We shall now list the elements of a ' propositional ' structure underlying a natural language sentence.

If S′ is an abstract sentence in a natural language (from which actual sentences are obtained), S the proposition, * the ' semantic determinant ', V the verb and J the set of arguments to the verb, then we have:

(1)	S′	→	* S
(2)	S	→	(' J ' ' V ')
(3)	J	→	Z ' Z′ '
(4)	Z′	→	J
(5)	Z	→	⟨P × ' Q '⟩
(6)	Q	→	P
(7)	P	→	$\begin{Bmatrix}\langle \Sigma\, P_{(i)}\rangle \\ \langle Z\rangle\end{Bmatrix}$
(8) (*a*) (*b*) (*c*) (*d*)	$P_{(i)}$	→	$\begin{Bmatrix}\langle C\rangle \\ \langle NO\rangle \\ \langle AO\rangle \\ \langle DO\rangle\end{Bmatrix}$
(9)	C	→	$(\Sigma C_{(i)})$
(10)	$C_{(i)}$	→	(' $J_{(i)}$ ' ' $V_{(i)}$ ')
(11)	NO	→	$\langle \Sigma NO_{(i)}\rangle$
(12)	$NO_{(i)}$	→	$\begin{Bmatrix}\langle \text{' T '}\ NA\rangle \\ \langle R_1\rangle \\ \langle R_2\rangle\end{Bmatrix}$
(13)	T	→	$\{\langle T_1\rangle, \langle T_2\rangle, \cdots\}$
(14)	NA	→	⟨ ' ⟨AO⟩ ' N⟩
(15)	N	→	$\{\langle N_1\rangle, \langle N_2\rangle, \cdots\}$
(16)	AO	→	$\langle \Sigma AO_{(i)}\rangle$
(17)	$AO_{(i)}$	→	⟨'⟨DO⟩' A⟩
(18)	DO	→	$\langle \Sigma DO_{(i)}\rangle$
(19)	$DO_{(i)}$	→	⟨D × ' Y '⟩
(20)	Y	→	$DO_{(i)}$

(21) A → {⟨A_1⟩, ⟨A_2⟩, ...}

(22) D → {⟨D_1⟩, ⟨D_2⟩, ...}

(23) R_1 → {⟨PR_1⟩, ⟨PR_2⟩, ...}

(24) R_2 → {⟨RR_1⟩, ⟨RR_2⟩, ...}

(25) {T_1, T_2, ...} → {Lexical Determiners like: *a*, *the*, *this*, *my*, etc.}

(26) {N_1, N_2, ...} → {Lexical Nouns}

(27) {A_1, A_2, ...} → {Lexical Adjectives}

(28) {D_1, D_2, ...} → {Lexical Adverbs}

(29) {PR_1, PR_2, ...} → {Lexical Pronouns}

(30) {RR_1, RR_2, ...} → {Relative Pronouns}

(31) *(a)*, *(b)* V → {VC, VO}

(32) VC → (X J)

(33) *(a)*, *(b)* X → {VC, VO}

(34) *(a)*, *(b)*, *(c)* VO → {VK, VI, VT}

(35) VK → {VK_1, VK_2, ...}

(36) VI → {VI_1, VI_2, ...}

(37) VT → {VT_1, VT_2, ...}

(38) * → +' ' + '

(39) +' → {., ?, !, !', ..., ∧, V, −}

(40) + → {The extended Fillmore ' case-role ' components}

and the brackets stand for what is written within them below:

(S-structures, C-structures or Conjunct Verbs),

⟨P-structures⟩,

{Alternatives}

and ' Optionals ',

5.4. *Some operations and relations among J, C, etc.*

We assume that:

(A1) C-structures cannot have a direct attributive relationship with one another or with a P-structure. They could only have conjunctive relationship with one another.

(A2) P-structures (that is, those that are in the form J or ⟨P⟩) can have either attributive or conjunctive relationship with other P-structures.

Let us now examine a number of isolated cases:

Case 1 :

(1) If	J	→	⟨P⟩
(2) and	P	→	$J_1 \times J_2$
(3) then	J	→	$\langle J_1 \times J_2 \rangle$
(4) If now	J_1	→	$\langle P_1 \rangle$
(5) and	J_2	→	$\langle P_2 \rangle$
(6) then	J	→	$\langle \langle P_1 \rangle \times \langle P_2 \rangle \rangle$

Further, if

(7)	P_2	→	$B_1 \circ B_2$

where

(8)	B_1	→	$\langle R_1 \rangle$
(9) and	B_2	→	$\langle R_2 \rangle$
(10) then	P	→	$\langle \langle P_1 \rangle \times \langle B_1 \circ B_2 \rangle \rangle$
		→	$\langle \langle P_1 \rangle \times \langle \langle R_1 \rangle \circ \langle R_2 \rangle \rangle \rangle$

If, now,

(11)	P_1	→	$E_1 \circ E_2$
(12) where	E_1	→	$\langle Q_1 \rangle$
(13) and	E_2	→	$\langle Q_2 \rangle$
(14) then	P	→	$\langle \langle \langle Q_1 \rangle \circ \langle Q_2 \rangle \rangle \times \langle \langle R_1 \rangle \circ \langle R_2 \rangle \rangle \rangle$

This is an underlying structure for such expressions as:

'Old and enthusiastic men and women'

'Old but enthusiastic men and women'

'Old and enthusiastic men or women', etc.,

if Q_1 and Q_2 are lexical adjectives and R_1 and R_2 are lexical nouns.

Case 2 :

If in (3) we put:

(15) $J_1 \rightarrow A'$

(16) and $J_2 \rightarrow N'$

then, (1) becomes:

(17) $J \rightarrow \langle A' \times N' \rangle$,

where A′ is an attribute to N′, under conditions to be specified, even when J_1 and J_2 are larger structures. A′ and N′ are not necessarily lexical adjectives and nouns, but they assume these roles and become *virtual* adjectives and nouns as elements of P-structures in attributive relationship. This would be the case even when the elements of these P-structures are C-structures.

Case 3 :

(*a*) When there are a number of J's and a V together within a structure (C-structure), all the J's are more directly and closely related to the V than to one another.

(18) $C \rightarrow (J\ V)$

$\rightarrow (J_1\ J_2\ V\ J_3\ J_4)$

Each such J then is characterised by a different case-role (answering, depending upon the V chosen and the completeness of the arguments of the V, varied questions like: Who or what ? whom or what ? to whom or to what ? at what place ? to which place ? in what manner ? under what causal, resultant or other conditions ? when ?, etc.).

(*b*) when there are a number of J's within a P-structure, however, they have:

(i) a coordinate relation:

J_1 o J_2 ('and', 'or', 'but', etc.)

or (ii) an attributive relation:

$J_1 \times J_2$ (where J_1 is an attribute of J_2 or J_2 is an attribute of J_1 under given conditions).

We then have:

(19) $\langle J_1 J_2 J_3 \text{ o } J_4 \rangle \equiv \langle J_1 \text{ o } J_2 \text{ o } J_3 \text{ o } J_4 \rangle$

except when o → ' but ', which connects only two J's at a time, and

(20) ⟨J_1 ... J_2 × J_3⟩, which can be interpreted only as:

⟨J_1 × ⟨J_2 × J_3⟩⟩ or ⟨⟨J_1 × J_2⟩ × J_3⟩.

There cannot be more than two elements in this relation as immediate inner members of a P-structure.

For example, if we are given a relational structure like (J_1 × J_2 × J_3 × J_4), then, for a correct interpretation, this must be reduced to:

(21) ⟨J_1 × J_2 × J_3 × J_4⟩ → ⟨J_1 × ⟨J_2 × ⟨J_3 × J_4⟩⟩⟩

or → ⟨J_1 × ⟨⟨J_2 × J_3⟩ × J_4⟩⟩

or → ⟨⟨J_1 × J_2⟩ × ⟨J_3 × J_4⟩⟩

or → ⟨⟨J_1 × ⟨J_2 × J_3⟩⟩ × J_4⟩

or → ⟨⟨⟨J_1 × J_2⟩ × J_3⟩ × J_4⟩,

where, within one P-structure bracket, we have no more than *two* elements in an attributive relationship.

If the inner structures of the different J's are known, the possibilities are correspondingly reduced for different languages.

Case 4 :

If there is a P-structure that could be expanded into the following form:

⟨⟨J_1 o J_2⟩ × ⟨C⟩⟩

then, the P-structure which is wholly composed of a C-structure with or without a marker could be a virtual attribute A′ to the P-structure which consists only of other P-structures. This P-structure of P-structures then assumes the role of a virtual noun N′.

For example, if

J_1 → ⟨P_1⟩

J_2 → ⟨P_2⟩

C → (J V)

→ (J_4 V J_5)

and if

J_4 → ⟨P_4⟩

and

J_5 → ⟨P_5⟩,

then

$$\langle\langle J_1 o J_2\rangle \times \langle C\rangle\rangle$$

$$\rightarrow \langle\langle\langle P_1\rangle o \langle P_2\rangle\rangle \times \langle\langle\langle P_4\rangle \text{ V } \langle P\ \rangle\rangle\rangle\rangle$$

$$\rightarrow \langle\langle\langle \text{The man}\rangle = \text{and} = \langle\text{the woman}\rangle\rangle \times \langle\langle\langle \text{who}\rangle \text{ were going } \langle + \text{along the road}\rangle\rangle\rangle\rangle.$$

Case 5:

If in a structure like:

$$\langle J_1 \times J_2\rangle$$

$$(a)\ J_2 - + \rightarrow \langle C\rangle$$

then, in English, generally

$$J_1 - + \rightarrow \langle C\rangle$$

and therefore J_1 is attribute to J_2, unless J_1 contains a ' real ' noun and J_2 has a noun phrase with a marker attached to it.

That is,

$$J_1 \rightarrow \langle A'\rangle$$

$$J_2 \rightarrow \langle N'\rangle.$$

And, if, in English:

$$(b)\ J_2 \rightarrow \langle C\rangle$$

where

$$C \rightarrow (K_j V_t),$$

then

$$J_1 \rightarrow \langle N'\rangle$$

$$J_2 \rightarrow \langle A'\rangle.$$

That is, J_2 is attribute to J_1.

Here K_j, when expanded, may contain an element in the role of a ' subject or agent ', in which case V_j would be a finite verb. If there is no element in the expansion of K_j that could have the role of a ' subject ', then V_j is not a finite verb, but is a form with the morpheme *-ing* or *-en* in English,

Case 6:

If in a structure like:

$$\langle J_1 \times J_2 \rangle$$

$$\left\{\begin{matrix} J_1 \\ J_2 \end{matrix}\right\} - + \rightarrow \langle C \rangle$$

then

(*a*) in English:

$$J_1 \rightarrow \langle A' \rangle$$

and

$$J_2 \rightarrow \langle N' \rangle$$

but

(*b*) in Russian:

(i) $J_1 \rightarrow \langle A' \rangle$ only if $J_1 \rightarrow \langle AO \rangle$

and $J_2 \langle N' \rangle$.

Or

(ii) $J_1 \rightarrow \langle N' \rangle$ if

$$J_1 - + \rightarrow \langle AO \rangle$$

and

$$J_2 \rightarrow \langle A' \rangle,$$

where

$$A' \rightarrow NO + G$$

where G indicates the Genitive case marker.

Case 7:

In a structure like $\langle J_1 \times J_2 \rangle$,
if, in Russian:

(*a*) $J_1 \rightarrow \langle C_1 \rangle$

and

$$J_2 - + \rightarrow \langle C_2 \rangle$$

then

$$J_1 \rightarrow \langle A' \rangle$$

$$J_2 \rightarrow \langle N' \rangle$$

and if

$$(b)\ J_1 \dashrightarrow \langle C_1 \rangle$$

and

$$J_2 \rightarrow \langle C_2 \rangle$$

then

$$J_2 \rightarrow \langle A' \rangle$$

and

$$J_1 \rightarrow (N')$$

5.5. *Combination of sentences*

Sentences could be combined in various ways, of which we shall examine a few cases here:

If we have S-structures of the type

$$S_1 \rightarrow (J'\ V_1)$$
$$\rightarrow (J_1 \dot{V}_1 J_3)$$

and

$$S_2 \rightarrow (J''\ V_2)$$
$$\rightarrow (J_2 V_2 J_4)$$

then, if

$$S_3 \rightarrow (S_1\ o\ S_2)$$

we have

$$S_3 \rightarrow ((J_1 V_1 J_3)\ o\ (J_2 V_2 J_4))$$

If now

$$V_1 \cup V_2$$
$$J_1 \cup J_2$$

and

$$J_3 \cup J_4$$

then

$$S_3 \rightarrow (\langle J_1\ o\ J_2 \rangle\ ((V_1 J_3)\ o\ (V_2 J_4)))$$
$$\rightarrow (\langle J_1\ o\ J_2 \rangle\ ((V_1)\ o\ (V_2))\ \langle J_3\ o\ J_4 \rangle)$$

If

$$((V_1)\ o\ (V_2)) \rightarrow V$$

then

$$S_3 \rightarrow (\langle J_1 \text{ o } J_2\rangle (V) \langle J_3 \text{ o } J_4\rangle)$$

But $(V) \rightarrow V$, since it is the only element within ( ), and therefore,

$$S_3 \rightarrow (\langle J_1 \text{ o } J_2\rangle \text{ V } \langle J_3 \text{ o } J_4\rangle)$$

Now if $\langle J_1 \text{ o } J_2\rangle \rightarrow J_5$ and $\langle J_3 \text{ o } J_4\rangle \rightarrow J_6$

then

$$S_3 \rightarrow (J_5 V\ J_6)$$

Examples:

(*a*) (i) $S_1 \rightarrow$ (⟨He⟩ is ⟨good⟩)

$S_2 \rightarrow$ (⟨She⟩ is ⟨good⟩)

and

$S_3 \rightarrow$ ((⟨He⟩ is ⟨good⟩) o (⟨She⟩ is ⟨good⟩))

→ ((⟨He⟩ o ⟨She⟩) ((is) o (is)) ⟨⟨good⟩ o ⟨good⟩⟩)

→ (⟨They⟩ (are) ⟨good⟩)

→ (⟨They⟩ are ⟨good⟩)

(ii) $S_1 \rightarrow$ (⟨On⟩ φ ⟨khoroshij⟩)

$S_2 \rightarrow$ (⟨Ona⟩ φ ⟨khoroshaja⟩)

$S_3 \rightarrow$ (⟨⟨On⟩ o ⟨ona⟩⟩ ((φ) o (φ)) ⟨⟨khoroshij⟩ o ⟨khoroshaja⟩⟩)

→ (⟨Oni⟩ (φ) ⟨khoroshie⟩)

→ (⟨Oni⟩ φ ⟨khoroshie⟩)

We have further:

$$(J_5 (V_5)\ J_6) \equiv ((J_5 V_5)\ J_6) \equiv (J_5 (V_5 J_6)) \equiv ((J_5 V_5 J_6)) \equiv (J_5 V_5 J_6)$$

(*b*) (i) $S_1 \rightarrow$ (⟨He⟩ is reading)

$S_2 \rightarrow$ (⟨He⟩ is writing)

$S_3 \rightarrow (S_1 \text{ o } S_2)$

→ $((J_1 V_1) \text{ o } (J_2 V_2))$

→ (⟨J_1 o J_2⟩ ((V_1) ∪ (V_2)))

Now (i *a*) if $J_1 \equiv J_2$, then $J_1 \equiv J_2 \equiv J$ and also $\langle J_1 \text{ o } J_2\rangle \rightarrow J$, then

$S_3 \rightarrow (J ((V_1) \text{ o } (V_2)))$

→ (⟨He⟩ is ((reading) = and = (writing)))

Further

(i *b*) If $J_1 \neq J_2$, then if we also have $J_1 \cup J_2$ and $V_1 = V_2$, then we may put

$$(J_1 \circ J_2) \rightarrow J'$$

and

$$((V_1) \circ (V_2)) \rightarrow V'$$

so that,

$$S_3 \rightarrow (\langle J_1 \circ J_2 \rangle \ ((V_1) \circ (V_2)))$$
$$\rightarrow (J' \ (V'))$$
$$\rightarrow (J' \ V')$$

since $(V') \rightarrow V'$

So that we have:

(1) $S_1 \rightarrow (\langle He \rangle \ reads)$

$S_2 \rightarrow (\langle You \rangle \ read)$

$S_3 \rightarrow ((\langle He \rangle \circ \langle you \rangle) \ ((reads) \circ (read)))$

$\rightarrow (\langle You \rangle \ (read))$

$\rightarrow (\langle You \rangle \ read)$

(2) When $V_1 \neq V_2$ and $J_1 \neq J_2$,

$S_1 \rightarrow \langle He_1 \rangle$ is reading

$S_2 \rightarrow \langle He_2 \rangle$ is writing

If S_1 and S_2 are combined into a single S_3 an additional explanatory element J_* is added to S_3 to indicate the distributive nature of the actions among members of the subject that have been summed up into a single lexical element.

Thus,

$$S_3 \rightarrow (S_1 \circ S_2)$$
$$\rightarrow ((J_1V_1) \circ (J_2V_2))$$
$$\rightarrow (\langle J_1 \circ J_2 \rangle \ ((V_1) \circ (V_2)) \ J_*)$$
$$\rightarrow (\langle \langle He_1 \rangle \circ \langle He_2 \rangle \rangle \ ((is \ reading) \circ (is \ writing)) \ (respectively))$$

That is,

$$S_3 \rightarrow (\langle They \rangle \ ((are \ reading) \circ (are \ writing)) \ \langle respectively \rangle)$$
$$\rightarrow (\langle They \rangle \ are \ ((reading) \circ (writing)) \ \langle respectively \rangle)$$

(*c*) In all cases when S_1 and S_2 are combined into S_3 they could be considered as clauses of S_3, that is, as C_1 and C_2.

Let

$S_1 \rightarrow$ (⟨The man⟩ works)

and

$S_2 \rightarrow$ (⟨The man⟩ lives)

If these are combined into S_3, then

$S_1 \rightarrow C_1$ and $S_2 \rightarrow C_2$. Let now

$C_1 \rightarrow (J_1V_1)$

$C_2 \rightarrow (J_2V_2)$

Case (1):

If

$J_1 \equiv J_2 \equiv J$,

then

$S_3 \rightarrow (C_1 \circ C_2)$

$\rightarrow ((J_1V_1) \circ (J_2V_2))$

$\rightarrow (J\ ((V_1) \circ (V_2)))$

$\rightarrow$ (⟨The man⟩ ((works) = and = (lives)))

Case (2):

If $J_1 = J_2 = J$, but we do not want to substitute J for J_1 and J_2 and at the same time want to avoid repetition, then we can put:

$J_1 \rightarrow \langle P_1 \rangle$

and

$J_2 \rightarrow \langle P_2 \rangle$,

where

$P_1 \| P_2$ and, if $P_1 \rightarrow NO_1$ and $P_2 \rightarrow R_1$, then:

$S_2 \rightarrow ((J_1V_1) \circ (J_2V_2))$

$\rightarrow ((\langle P_1 \rangle\ V_1) \circ (\langle P_2 \rangle\ V_2))$

$\rightarrow ((\langle NO_1 \rangle\ V_1) \circ (\langle R_1 \rangle\ V))$

$\rightarrow$ (⟨The man⟩ works) = and = (⟨he⟩ lives))

where

$NO_1 \rightarrow$ 'noun phrase'

$R \rightarrow$ 'pronoun'

$R_1 \rightarrow$ 'personal pronoun'

$R_2 \rightarrow$ 'relative pronoun'

Case (3):

If

$$C_1 \rightarrow (J_1V_1)$$

$$C_2 \rightarrow (J_2V_2)$$

and if

$$J_1 \rightarrow \langle P_3 \times J_3 \rangle$$

and

$$J_3 \rightarrow \langle C_2 \rangle$$

then:

$$C_1 \rightarrow S_1$$

and

$$S_1 \rightarrow (\langle P_3 \times J_3 \rangle\ V_1)$$

$$\rightarrow (\langle P_3 \times \langle C_2 \rangle \rangle\ V_1)$$

$$\rightarrow (\langle P_3 \times \langle (J_2V_2) \rangle \rangle\ V_1)$$

In such a case, if $P_3 \rightarrow NO_1$,

then

$$J_2 \rightarrow \langle P_4 \rangle \rightarrow \langle R_2 \rangle.$$

Also $NO_1 \parallel R_2$

So that:

$$S_1 \rightarrow (\langle \text{The man } \langle (\langle \text{who} \rangle \text{ lives}) \rangle \rangle \text{ works}).$$

Similarly, if we interchange C_1 and C_2 and have $C_2 \rightarrow S_2$, then we get:

$$S_2 \rightarrow (\langle \text{The man } \langle (\langle \text{who} \rangle \text{ works}) \rangle \rangle \text{ lives}).$$

5.6. *Relationship of P-structure to V*

Let us assume that V is of three main types:

That is,

$$V \rightarrow \begin{Bmatrix} V_k \\ V_i \\ V_t \end{Bmatrix}$$

where V_k is a 'link verb, V_i an 'intransitive' verb and V_t is a 'transitive' verb.

Case (i) :

If

$$V \rightarrow V_k,$$

then

$$C_1 \rightarrow (J_1 V_k J_2).$$

Here J_1 and J_2 have certain specific relations only.

(*a*) $J_1 \parallel J_2$

then

$$J_1 \rightarrow \langle NO \rangle$$

and $J_2 \rightarrow \langle AO \rangle$

where

$$NO \rightarrow \begin{Bmatrix} NO_1 \\ R_1 \end{Bmatrix},$$

and C_1 is of the type:

$$C_1 \rightarrow (\langle \text{The man} \rangle \text{ is } \langle \text{good} \rangle)$$

or

$$\rightarrow (\langle \text{He} \rangle \text{ is } \langle \text{good} \rangle)$$

(*b*) If $J_1 \cup J_2$, then

$$J_1 \rightarrow \langle NO_1 \rangle$$

and

$$J_2 \rightarrow \langle NO_2 \rangle$$

$$C_1 \rightarrow (\langle \text{The man} \rangle \text{ is } \langle \text{an engineer} \rangle).$$

(*c*) If $J_1 \psi J_2$, but $J_1 = J_2$, then

$$J_1 \rightarrow \langle NO \rangle \rightarrow \langle R_1 \rangle$$

and

$$J_2 \rightarrow \langle NO_2 \rangle$$

so that

$$C_1 \rightarrow (\langle \text{He} \rangle \text{ is an } \langle \text{engineer} \rangle).$$

(*d*) If we have: $J_b \parallel J_1$, and $J_a \rightarrow \langle J_b \times \langle C_1 \rangle \rangle$,

then

$$J_a \rightarrow \langle J_b \langle (J_1 V_k J_2) \rangle \rangle$$

where

$$J_3 \rightarrow \begin{Bmatrix} NO_1 \\ R_1 \end{Bmatrix}$$

and

$$J_1 \rightarrow R_2.$$

This leads to an expression like:

$J_s \rightarrow$ ⟨The man ⟨(⟨who⟩ is ⟨an engineer⟩)⟩⟩.

Case (ii) :

If

$$V \rightarrow V_1$$

and

$$C_1 \rightarrow (J_1 V_t J_2),$$

then

$$J_2 \neq J_2 \text{ and } J_1 \# J_2.$$

So that:

$$J_2 \rightarrow \begin{Bmatrix} \langle + \text{ M NO} \rangle \\ \langle \text{DO} \rangle \end{Bmatrix}$$

Case (iii):

If

$$V \rightarrow V_t$$

and

$$C_1 \rightarrow (J_1 V_t J_2)$$

then

$$J_2 \rightarrow J_3 J_4$$

and

$$J_3 \rightarrow \begin{Bmatrix} \langle \text{NO} \rangle \\ \langle + \text{ M NO} \rangle \end{Bmatrix}$$

If $J_3 \rightarrow \langle \text{NO} \rangle$, $J_4 \rightarrow \langle + \text{ M NO} \rangle$ and if $J_3 \rightarrow \langle + \text{ M NO} \rangle$ $J_4 \rightarrow \langle \text{NO} \rangle$ in English.

Then,

$C_1 \rightarrow (J_1 V_t J_2)$

$\rightarrow (J_1 V_t J_3 J_4)$

$\rightarrow$ (⟨He⟩ gave ⟨him⟩ ⟨the book⟩)

or

$\rightarrow$ (⟨He⟩ gave ⟨the book⟩ ⟨+ to him⟩)

That is, if $J_3 \rightarrow \langle + \text{ M NO} \rangle$, $+ \text{ M} \rightarrow \phi$ and if $J_4 \rightarrow \langle + \text{ M NO} \rangle$, $+ \text{ M} \rightarrow +$ to.

In other words + M is the marker for the indirect object (if we consider J_3 and J_4 as merely the positions the direct and indirect objects alternately occupy). If, however, we consider J_3 as the direct object and J_4 as the indirect object, then:

$$C_1 \rightarrow (J_1 V_t J_3 J_4)$$

where

$$J_3 \rightarrow \langle \text{NO} \rangle$$

and

$$J_4 \rightarrow \langle + \text{ to NO} \rangle$$

or

$$C_1 \rightarrow (J_1 V_t J_4 J_3)$$

where

$$J_3 \rightarrow \langle + \phi \text{ NO} \rangle$$

and

$$J_3 \rightarrow \langle \text{NO} \rangle.$$

Perhaps this alternative view is preferable.

Note : Any expansion like $J \rightarrow J_a J_b$ indicates that J_a and J_b are not the inner components of J, but are directly associated with the V with which J is associated.

Case (iv):

If

$$V \rightarrow \begin{Bmatrix} V_k \\ V_i \\ V_t \end{Bmatrix}$$

and

$$C_1 \rightarrow (J_1 \text{ V } J_2')$$

and if

$$J_2' \rightarrow J_2 J_6$$

then

$$C_1 \rightarrow (J_1 \text{ V } J_2 J_6)$$

J_2 could then be treated as in cases (i), (ii) and (iii).

Now, if J_6 could be analysed as:

$$J_6 \rightarrow J_7 J_8 J_9 J_{10},$$

then

$$J_7 \rightarrow \begin{Bmatrix} \langle \phi \rangle \\ \langle DO_{how} \rangle \\ \langle + M_{how}\ NO \rangle \end{Bmatrix}$$

$$J_8 \rightarrow \begin{Bmatrix} \langle \phi \rangle \\ \langle DO_{where} \rangle \\ \langle + M_{where}\ NO \rangle \\ \langle + where\ (C) \rangle \end{Bmatrix}$$

$$J_9 \rightarrow \begin{Bmatrix} \langle \phi \rangle \\ \langle DO_{when} \rangle \\ \langle + M_{when}\ NO \rangle \\ \langle + when\ (C) \rangle \end{Bmatrix}$$

$$J_{10} \rightarrow \begin{Bmatrix} \langle \phi \rangle \\ \langle + M_{cd}\ (C) \rangle \end{Bmatrix}$$

where $+ M_{cd}$ is some condition, cause, result, etc.

It has only been possible to list here examples of various kinds. A detailed discussion of the order in which J_7 J_8 J_9 J_{10}, etc., occur in any given language has to be postponed to a separate paper.

The various relationships (i)–(xii) given in this part have also to be critically evaluated. However, they are all given here in the present form, without a critical re-examination for 'whatever they are worth'.

PART VI

6. Ambiguities, abbreviations, paraphrase, etc.

In this part too we look at a few isolated phenomena at random. We have reserved more detailed exploration of the problems to a later study. But we give some indications here of our approach.

6.1. *Ambiguities*

Disambiguating ambiguities is not merely a question of finding the dominating node points in a deep structure. The logical connection underlying the structure has to be elicited through tests of (1) transformation, (2) substitution, (3) augmentation, (4) question frames and (5) grouping into propositional sets.

We shall not discuss these in detail here. We shall merely mention some cases at random.

For example, if we could put

(1) ((He⟩ is ⟨easy ⟨+ to (please))))

in the form

(2) ((It⟩ is ⟨easy ⟨+ to (please ⟨him)))))).

then (1) is different from:

(3) ((He⟩ is ⟨eager ⟨+ to (please))))

which cannot be put in that form, without change of meaning.

Further, in (1), we cannot ' augment ' the verb '(please)' by an expression '(please ⟨us⟩)' whereas (3) can be so ' augmented ' to give

(4) ((He⟩ is ⟨eager ⟨+ to (please ⟨us⟩)))))).

If we take the much discussed examples like:

(5) ((She⟩ made ⟨him ⟨a good husband⟩⟩)

(6) ((She⟩ made ⟨him⟩ ⟨a good wife⟩)

and

(7) ((She⟩ made ⟨him⟩ ⟨a pan cake⟩)

then, the ambiguities in them could be resolved only by asking questions like: What are the transformations possible ? What are the augmentations possible ? What are the elisions possible ? etc.

We find that (5) could go (without change in meaning) into the form:

(8) ((She⟩ made ⟨a good husband⟩ ⟨+ of him⟩)

(6) could go into the forms:

(9 *a*) ((She⟩ made (herself ⟨a good wife⟩⟩ ⟨+ for him⟩)

(9 *b*) ((She⟩ made ⟨a good wife⟩ ⟨+ of herself⟩ ⟨+ for him⟩)

and (7) could go into the form:

(10) ((She⟩ made ⟨a pan cake⟩ ⟨+ for him⟩)

Some surface structures, whatever they are in absolute terms, could from a practical point of view be disambiguated by finding out what questions they answer. For example:

(11) They are flying planes

could be an answer to the question:

(12) ((What⟩ are ⟨they⟩)?

giving the answer:

(13) (⟨They⟩ are ⟨flying planes⟩)

or to the question:

(14) (⟨What⟩ are ⟨they⟩ doing)?

giving the answer:

(15) (⟨They⟩ are flying ⟨planes⟩).

In the first case above 'are' is a link verb (V_k) and in the second case 'are' is an aspect auxiliary (V'_n) forming part of the verb phrase 'are flying' and 'they' is the agent and 'planes' is the patient of the lexical verb 'fly'.

6.2. *Complex structures*

There could be complex structures like :

(16) That she is not beautiful, which she is, is the view held by all other women.

Where does the relative clause fit in and what is its status in the sentence? It is rather difficult to answer this question except through a commentary. But let us follow our method of analysis:

Let

$S \rightarrow (J_1V_1J_2)$

$J_1 \rightarrow \langle +\ \text{that}\ C_1\rangle$

$V_1 \rightarrow$ is

$J_2 \rightarrow \langle P_2 \times J_3\rangle$

$P_2 \rightarrow NO_1 \rightarrow$ the view

$J_3 \rightarrow \langle C_3\rangle$

$C_3 \rightarrow (V_3J_4)$

$V_3 \rightarrow$ held

$J_4 \rightarrow \langle +\ \text{by}\ NO_2\rangle$

$NO_2 \rightarrow$ other women

$C_1 \rightarrow (J_5V_5J_6)$

$J_5 \rightarrow R_1 \rightarrow$ she

$V_5 \rightarrow$ is not

$J_6 \rightarrow \langle P_3 \times J_7\rangle$

$P_3 \rightarrow AO \rightarrow$ beautiful

$J_7 \rightarrow$ -⟨P_4⟩-

$P_4 \rightarrow C_4$

$C_4 \rightarrow (J_8 J_9 V_8)$

$J_8 \rightarrow \langle R_2 \rangle \rightarrow \langle$which$\rangle$

$J_9 \rightarrow \langle R_1 \rangle \rightarrow \langle$she$\rangle$

$V_8 \rightarrow$ is

This gives us the structure shown below:

(17) ((⟨+ That (⟨she⟩ is not ⟨beautiful -⟨((⟨which⟩ ⟨she⟩ is)⟩-⟩)⟩ is ⟨the view ((held ⟨+ by all other women⟩)⟩⟩)).

We have to decide that:

1. 'which she is' is a parenthetical comment made by the reporter on 'her being beautiful'. If this structure is acceptable (that is grammatical), then we have a case in which a relative pronoun has an adjective 'beautiful' as its antecedent.

2. If this relation between an adjective and a relative pronoun is held to be ungrammatical, then the analysis is wrong.

3. If the analysis for this structure is inescapable, then the construction is wrong, that is, nobody speaks like that. (This is a borrowed example and we have not had occasion to check it against a native speaker's innate feel of grammaticalness or otherwise of this structure.)

Another example from non-native speakers of English, not counterchecked with native English speakers, is the following:

(18) They say he is an expert, which is correct.

Assuming that this sentence is grammatical according to the native speaker, we have to decide what 'is correct': 'their saying that he is an expert' (irrespective of whether he is or not) or 'his being an expert' (irrespective of their saying so or not)?

We have the following alternative analyses:

(19) $S \rightarrow (JV)$

$V \rightarrow \phi$

$J \rightarrow \langle J_1 \times J_2 \rangle$

$J_1 \rightarrow \langle C_1 \rangle$

$C_1 \rightarrow$ ((They⟩ say ⟨((he⟩ is ⟨an expert⟩⟩⟩)

That is, we have a sentence S without a verb in it.
Or we have:

(20) S → (JV)
→ (J_1VJ_2)
V → say
J_1 → they
J_2 → ⟨$J_3 \times J_4$⟩
J_3 → ⟨C_1⟩ → ⟨(⟨he⟩ is ⟨an expert⟩)⟩
J_4 → —⟨C_2⟩—
C_2 → (⟨which⟩ is ⟨correct⟩).

In this analysis J_4 has to be a parenthetical comment by the reporter, since it is not a part of what 'they say'.

6.3. *Abbreviations*

Lexicalisation is one of the ways of abbreviation. But *what* is abbreviated could be classified according to its composition.

1. *Logical connective plus an adverb* :

We could state in general:

= and = .. ⟨DO⟩ → = but =

For, we could consider:

(21) (⟨I⟩ ((went ⟨home⟩) =but= (did not study))) is equivalent to:

(22) (⟨I⟩ ((went ⟨home⟩) =and= (⟨nevertheless⟩ did not study)))

(23) (((⟨We⟩ advised ⟨him ⟨+ to (be ⟨careful⟩)⟩⟩) =but= (⟨he⟩ ((continued ⟨+ to (be ⟨reckless⟩)⟩) =and= (got ⟨+ into trouble⟩)))))

is equivalent to:

(24) (((⟨We⟩ advised ⟨him ⟨+ to (be ⟨careful⟩)⟩⟩) =and= (⟨all—the—same⟩ (he) ((continued ⟨+ to (be ⟨reckless⟩)⟩) =and= (got ⟨+ into trouble⟩))))).

2. *Logical connective and sentence-modifying adverb combined into a clause-marker*

A sentence modifying adverb like 'therefore', 'however', etc., occurring in the second sentence could be combined with a logical connective and made into a clause marker for the first sentence, which is now a C-structure within a P-structure of the second sentence:

Thus. the two separate sentences:

(25) ((He⟩ came ⟨+ into the room⟩) ((Therefore⟩ ⟨she⟩ went ⟨out ⟨+ into the garden⟩⟩)

are combined into one sentence with two coordinate clauses:

(26) ((⟨He⟩ came ⟨+ into the room⟩) =and= ((therefore⟩ ⟨she⟩ went ⟨out ⟨+ into the garden⟩⟩).

and further combined as:

(27) ((⟨+ Since ((he⟩ came ⟨+ into the room⟩⟩) ⟨she⟩ went ⟨out ⟨+ into the garden⟩⟩)

having a traditional subordinate clause beginning with *since*, but in the present view, having a C-structure as part of a P-structure with a marker + *since* attached to it. This P-structure has the same status as a sentence-modifying adverb.

If morphologically and lexically a sentence-modifying adverb of this type belongs to the class of adverbs D_2, then a sentence-modifying P-structure with a marker and C-structure within it could be viewed as a *virtual* sentence-modifying adverb and referred to as D_2'.

3. *Adverb plus a clause marker*

An adverb and a clause marker could be combined and abbreviated into a new clause marker:

(28) ((⟨He⟩ sits ⟨+ in that chair⟩ ⟨every time⟩ ⟨+ when ((he⟩ comes ⟨here⟩⟩⟩⟩)

The time expression ⟨every time⟩, made up of a lexical adjective and noun is used in the sentence as a *virtual* adverb. This virtual adverb and the C-structure introduced by + *when* forming another *virtual* adverb are both expressions of time that could be considered as augmenting each other's meaning, so that we could consider one of them as an attribute or modifier to the other, giving:

(29) ((⟨He⟩ sits ⟨+ in that chair⟩ ⟨every time ⟨+ when ((he⟩ comes ⟨here⟩⟩⟩).

From this close-knit structure we could get an abbreviation such as + *whenever*, serving as a marker for the P-structure, giving:

(30) ((⟨He⟩ sits ⟨+ in that chair⟩ ⟨+ whenever ((he⟩ comes ⟨here⟩⟩⟩)

A further substitution and abbreviation for the time expression containing a C-structure that serves as a virtual time-adverb, would be by a regular (morphological or lexical) time-adverb. This would lead us to:

(31) ((⟨He⟩ ⟨always⟩ sits ⟨+ in that chair⟩).

4. *Any virtual part of speech replaced by an actual one*

In general, any *virtual* part of speech could be replaced by another of the same kind or by an actual morphological or lexical one.

For example:

(⟨He⟩ said ⟨+ that (⟨he⟩ would put ⟨every one⟩ ⟨+ in his place⟩)⟩)

could be replaced by:

(⟨He⟩ said ⟨that⟩)

or by:

(⟨He⟩ said ⟨it⟩)

A structure like:

(⟨The question⟩ is ⟨this⟩)

has an element ⟨this⟩ that could be a replacement of any possible complex structure, for example:

(⟨The question⟩ is ⟨⟨+ to (be)⟩ =or= ⟨+ not ⟨+ to (be)⟩⟩⟩)

or

(⟨The question⟩ is ⟨⟨this⟩ =or= ⟨that⟩⟩)

or, with a different emphasis:

(⟨⟨This⟩ =or= ⟨that⟩⟩ is ⟨the question⟩)

which is a replacement for:

(⟨⟨+ To (be)⟩ =or= ⟨+ not ⟨+ to (be)⟩⟩⟩ is ⟨the question⟩)

The whole 'subject' of this could be abbreviated as ⟨that⟩ giving us:

(⟨That⟩ is ⟨the question⟩)

For greater emphasis, the original statement and its replacement could be combined using a parenthetical device:

(⟨-⟨⟨+ to (be)⟩ =or= ⟨+ not ⟨+ to (be)⟩⟩⟩- that⟩ is ⟨the question⟩)

"To be, or not to be: that is the question".

6.4. *Syntactic abbreviation*

Another type of abbreviation is syntactic abbreviation (which is a sort of paraphrase).

If S is a proposition and it gives rise to the sentence S′, we could think of converting S into a virtual noun to serve as the object or subject of another proposition S_1.

That is, if

$$S \rightarrow (J_1 V_1 J_2)$$

and

$$S_1 \rightarrow (J_3 V_2 J_4)$$

where

$J_3 \rightarrow \langle *_1 S \rangle$

then $*_1$ is a transformation that converts a proposition into a virtual noun.

Thus.

$S_1 \rightarrow (\langle *_1 S \rangle \ V_2 J_4)$

$\rightarrow (\langle *_1 \ (J_1 V_1 J_2 \rangle) \ V_2 J_4)$

Since in this structure $*_1 S$ is to serve as the 'subject' of the verb V_2, a nominalization of whatever is within its bracket is involved here. So that, if

$S \rightarrow ((He\rangle \text{ came } \langle here))$

then

$*_1 S \rightarrow *_1 \ ((He\rangle \text{ came } \langle here))$

$\rightarrow \langle His\rangle \times \langle (coming \langle here))\rangle$

which leads to:

$S_1 \rightarrow (\langle *_1 S\rangle \ V_2 J_4)$

and if $V_2 \rightarrow$ is and $J_4 \rightarrow \langle good\rangle$

$S_1 \rightarrow (\langle *_1 \ (\langle He\rangle \text{ came } \langle here\rangle)\rangle \text{ is } \langle good\rangle)$

$\rightarrow (\langle \langle His\rangle \quad \langle (coming \langle here\rangle)\rangle\rangle \text{ is } \langle good\rangle)$

Now, if

$S_2 \rightarrow (\langle *_9 S\rangle \ V_2 J_4)$

$\rightarrow (\langle *_2 \ (\langle He\rangle \text{ came } \langle here\rangle)\rangle \text{ is } \langle a \text{ good thing}\rangle)$

$\rightarrow (\langle + \text{ that } (\langle he\rangle \text{ came } \langle here\rangle)\rangle \text{ is } \langle a \text{ good thing}\rangle)$

if S_3 is an abbreviation of S_2, we have:

$S_3 \rightarrow (\langle It\rangle \text{ is } \langle a \text{ good thing}\rangle)$

A parenthetical combination of S_2 and S_3 would give:

$(\langle It \text{ -}\langle + \text{ that } (\langle he\rangle \text{ came } \langle here\rangle)\rangle\text{-}\rangle \text{ is } \langle a \text{ good thing}\rangle)$

or

$(\langle It\rangle \text{ is } \langle a \text{ good thing -}\langle + \text{ that } (\langle he\rangle \text{ came } \langle here\rangle)\rangle\text{-}\rangle)$

Taking another example, we find that a structure consisting of a preposition plus a noun functions as a virtual adverb, like:

⟨ + on the road⟩

which could be abbreviated to or replaced by a lexical adverb like:

⟨there⟩.

Now, a sentence or clause could be nominalised and subsequently converted into a virtual adverb too. Thus,

(⟨He⟩ was going ⟨+ along the road⟩)

could be converted into:

⟨⟨His⟩ × ⟨(going ⟨+ along the road⟩)⟩⟩.

This virtual noun, when accompanied by a preposition, could form a virtual adverb, as in:

⟨+ in ⟨⟨his⟩ × ⟨(going ⟨+ along the road⟩)⟩⟩⟩.

A virtual adverb could also be obtained from the structure consisting of a marker and a C-structure together forming a P-structure. For example:

⟨+ While ⟨(⟨he⟩ was going ⟨+ along the road⟩)⟩⟩

which could be abbreviated to:

⟨+ at that time⟩

or finally to:

⟨then⟩

which is a morphologico-lexical ' adverb '.

6.5. *Abbreviations and nominalizations of sorts*

In the sentences

(32) (⟨I⟩ promised ⟨him⟩ ⟨+ that (⟨I⟩ would do ⟨the job⟩⟩))

and

(33) (⟨I⟩ ordered ⟨him⟩ ⟨+ that (⟨he⟩ should do ⟨the job⟩⟩))

the expression ' that I would ' or ' that he should ' could be abbreviated to '+ to '. In that case, if the agent of the two verbs in the sentence is the same, the abbreviation is a separate P-structure.

If the agent of the second verb is the patient of the first verb in the sentence, then the abbreviation is an 'apparent' attribute to that patient, and is a P-structure within that P-structure. Thus,

(34) ((I) promised ⟨him⟩ ⟨+ to (do ⟨the job⟩))

(35) ((I) ordered ⟨him ⟨+ to (do ⟨the job⟩))⟩)

Also:

(35 *a*) ((I) ordered ⟨a cab ⟨+ to (take ⟨him⟩ ⟨home⟩))⟩)

(35 *b*) ((I) promised ⟨a cab ⟨+ to (take ⟨him⟩ ⟨home⟩))⟩).

In (34) 'him' could be elided, in (35) it cannot be elided without eliding its 'apparent' attribute too. On the other hand the 'apparent' attribute could be elided.

Thus,

(36 *a*) ((I) promised ⟨+ to (do ⟨the job⟩))

(36 *b*) ((I) promised ⟨him⟩)

(37) ((I) ordered ⟨him⟩).

A. *Active voice nominalization*

From:

(38) ((I) gave ⟨the book⟩ ⟨+ to him⟩)

we get a nominalization:

(39) ⟨My⟩ × ⟨(giving ⟨the book⟩ ⟨+ to him⟩)⟩.

From (39) either of the P-structures accompanying the verb could be elided, leading to

(40) ⟨My⟩ × ⟨(giving ⟨the book⟩)⟩

or

(41) ⟨My⟩ × ⟨(giving ⟨+ to him⟩)⟩,

In (40) the *-ing* form and its arguments together are nominalized ('syntactic nominalization'), from (38). The *-ing* form in (40) could itself be further nominalized ('lexical nominalization'), in which case its argument becomes an attribute to it, as in:

(42) ⟨My⟩ × ⟨⟨book⟩ giving⟩

Now the attribute *book* and the 'noun' *giving* form a structure, which as a whole could be further 'nominalized' into a single lexical noun (a 'compound noun'), as in:

(43) ⟨My⟩ × ⟨book-giving⟩.

B. *Passive voice nominalizations*

In the structure:

(44) (⟨The book⟩ was given ⟨+ to him⟩ ⟨+ by me⟩)

the passive view of the 'action' (represented by the verb form *was given*) could be replaced by a view that represents a passive 'state', and we could get:

(45) (⟨The book⟩ was ⟨(given ⟨+ to him⟩ ⟨+ by me⟩)⟩)

The verb in this structure could be 'nominalized' leading to:

(46) ⟨The Book's⟩ × ⟨(being ⟨(given ⟨+ to him⟩ ⟨+ by me⟩)⟩)⟩

In this case, as in so many structures in the present view, an element like *being* or *given* is a 'virtual noun' or 'virtual adjective' with respect to the elements outside the P-structure containing them, but within the C-structure bracket containing them they are 'verbs'. Or, retaining *the book* as the head, we could consider the C-structure as an 'attributivised' form, giving:

(47) ⟨The book⟩ × ⟨(being ⟨(given ⟨+ to him⟩ ⟨+ by me⟩)⟩)⟩

where 'being' represents the 'attributivised process' and 'given' the 'attributivised state'. If we do not want to look at it as a 'process', but only as a 'state', we could elide 'being' and get

(48) ⟨The Book⟩ × ⟨(given ⟨+ to him⟩ ⟨+ by me⟩)⟩.

We could elide now the modifying P-structures of 'given' and convert 'given' into a 'lexicalized attribute', so that we get

(49) ⟨The book⟩ × ⟨(given)⟩.

The position of 'given' in English after the noun still retains in it the verbal colour.

This colour is removed by placing ⟨(given)⟩ → ⟨given⟩ before the noun, which gives:

(50) ⟨The ⟨given⟩ book⟩.

Lexical or 'lexicalized' adjectives, under certain conditions, could be 'nominalized'. For example:

(51) ⟨The ⟨poor⟩ people⟩

gives rise to the 'nominalization' of 'poor' by simply eliding 'people', leading to:

(52) ⟨The ⟨poor⟩⟩

which reduces to:

(53) ⟨The poor⟩.

If we take a sentence like:

(54) ⟨The people⟩ were trodden ⟨down⟩,

from this 'process', we could arrive at the 'state':

(55) ⟨The people⟩ were ⟨⟨trodden ⟨down⟩⟩⟩.

When this is 'nominalized' we get

(56) ⟨The people's⟩ × ⟨⟨being ⟨⟨trodden ⟨down⟩⟩⟩⟩⟩.

Or to an 'attributivization' as

(57) ⟨The people⟩ × ⟨⟨being ⟨⟨trodden ⟨down⟩⟩⟩⟩⟩.

Paying attention to the 'state' rather than the 'process', we could say, by eliding 'being'

(58) ⟨The people⟩ × ⟨⟨trodden ⟨down⟩⟩⟩.

The 'verbal colour' of the attribute could be reduced by placing it before the noun

(59) ⟨The ⟨⟨trodden ⟨down⟩⟩⟩ people⟩.

'Trodden' could be 'lexicalised' as an attribute, modified by the adverb 'down', in which case, the positions are rearranged

(60) ⟨The ⟨⟨down⟩ trodden⟩ people⟩.

'Down' and 'trodden' could be combined into a single adjective, and made into a lexicalized compound adjective

(61) ⟨The ⟨down-trodden⟩ people⟩.

It could be finally 'nominalized' by eliding the word 'people' which it qualifies

(62) ⟨The ⟨down-trodden⟩⟩

which reduces to

(63) ⟨The down-trodden⟩.

Like the rest of this paper, this last part too has been 'unorthodox' in its approach. We have moved away from any 'standard form' that others have used. Any rigid adherence to a 'standard form' would have confined us to a 'standard groove'. We

take this alternative course in the hope that we could see in languages something more than what a ' standard form ' would allow us to see.

If the reader has been led through these pages to see something more than what is within the rigid grooves through which he has been trained to view language structures, the purpose of this preliminary attempt should be served.

If no such result is seen and, on the other hand, the whole exercise has been an exercise in futility, we do not worry on that account. We are in the active process of retracing our steps, and revising and refining our system, until we find, or somebody else does, what we are seeking (*cf* for instance Part V with Part III).

Conclusion

Our consideration of the *verb* as the nucleus of a sentence (unlike the NP + VP division of CHOMSKY) and of all other elements of the sentence as arguments to the verb has made it possible for us to deal with several alternative *slants* of meaning contained in a sentence.

Our separation of parts of speech into *real* and *virtual* has made it possible to put any sentence, simple, compound or complex, into one and the same simple form:

$$S \rightarrow (VJ_1J_2J_3 \ldots).$$

Our idea of ' conjunct verb formation ' (not discussed here) and the bracketing notation adopted by us lend themselves to a sort of algebraic treatment of sentence elements as in the mathematical formula: $a(b + c) = ab + ac$. But this too has not been elaborated here.

CHOMSKY's VP, in our view, is only a particular case of conjunct verb formation.

Our chief limitations are in the, as yet, not reported development of the *semantic determinant* *.

We may use the semantic treatments suggested by Sydney M. LAMB (' Lexicology and Semantics ' in *Linguistics*, V.O.A. Forum Lectures, Washington D.C., 1973, pp. 45-56) or use a method of attack similar to what Y. A. WILKS (*Grammar, Meaning and the Machine Analysis of Language*, Routledge and Kegan Paul, London, 1972) has outlined, or we may develop our own ideas of *content-form matrices*.

How the * element in our formulas affects the transformational extension of our rewrite rules is also not touched upon here.

We ultimately aim at a synthesis of several viewpoints on language. Confrontation between rival theories, in our opinion, has outlived its time.

We have only expressed our belief that structures in different languages could be expressed by the same 'universal' abstract formula. Our belief is that the rewrite formulas are universal. Language specificity on the other hand comes through transformations, unlike most other treatments in which the very first few steps are already language specific

References

1. BACH, EMMON, *An Introduction to Transformation Grammars*, Holt, Rhinehart and Winston, Inc., New York, 1964.
2. GANESHSUNDARAM, P. C. Yaantrik Anuvaad kii Bhaasxaashaastriiya Samasyaacw, *Hindii Anushiilan*, Dr. Dhirendra Varma Jubilee Volume, 1959.
3. GANESHSUNDARAM, P. C. Preparing Language for a Scheme of Computer Analysis towards partially mechanised Translation with Pre- and Post-Editing, *JISTA* (New Delhi), 1972, 1, 19–35.
4. GANESHSUNDARAM, P. C. Structural Relativity in Languages, *Jour. Ind. Inst. Sci.*, 1975, 57, 329–357.
5. GANESHSUNDARAM, P. C. AND RANGACHARYA, S. K. On the Confusion Matrices of Technical Terms in English, Russian and Telugu in the Field of Algebra, *JISTA*, 1972, 1, 31–49.
6. HALLIDAY, M. A. K., MACINTOSH, ANGUS AND STREVENS, PETER *The Linguistic Sciences and Language Teaching*, ELBS and Longmans, London, 1970.
7. PALMER, F. R. *A Linguistic Study of the English Verb*, Longmans, London, 1965.