

AD 608747

RADC-TDR-64-200
Final Report



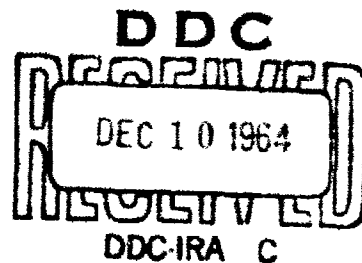
LINGUISTIC TRANSFORMATIONAL ANALYSIS

COPY	2	OF	3	mp
HARD COPY	\$ 4.00			
MICROFICHE	\$ 1.00			

TECHNICAL DOCUMENTARY REPORT NO. RADC-TDR-64-200

October 1964

Information Processing Branch
Rome Air Development Center
Research and Technology Division
Air Force Systems Command
Griffiss Air Force Base, New York



Project No. 4594 , Task No. 459401

(Prepared under Contract No. AF 30(602)-2951 by Indiana
University, Bloomington, Indiana. Authors: J. P. Thome
and J. Lyons.)

ARCHIVE COPY

**Best
Available
Copy**

When US Government drawings, specifications, or other data are used for any purpose other than a definitely related government procurement operation, the government thereby incurs no responsibility nor any obligation whatsoever; and the fact that the government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

Qualified requesters may obtain copies from Defense Documentation Center.

Defense Documentation Center release to Office of Technical Services is authorized.

Do not return this copy. Retain or destroy.

Keywords: Documentation, Data Processing

ABSTRACT

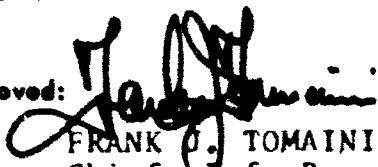
The following report deals with a number of related items of research, undertaken for Rome Air Development Center (Contract No. AF30(602)-2951) during the period January 1 to December 31, 1963. The contract was concerned with the feasibility and utility of a "kernelization" procedure for purposes of information retrieval. Research was carried out under the general direction of F. W. Householder, Jr., more specific direction being provided, at different times, by J.P. Thorne and J. Lyons. The final report has been assembled under the supervision of P.H. Matthews.

The leading section discusses, in general, the problems involved in the kernelization of complex English sentences; this section is the work of J.P. Thorne. The remainder is in the form of appendices. Appendix I contains a detailed report of the kernelization procedure; for this section S.I. Laszlo was largely responsible. Appendix II reports on a series of experiments, undertaken by J. Masters, to determine to what extent "information" was "preserved" in kernelized versions of sentences. Appendix III reports on a frequency count of the transformations exhibited by a stretch of running text; this count was undertaken by F.C. Fisher, W.J. Kunz, T. Matsuda and (initially) H. Feider. Finally, Appendix IV contains a list of transformational rules which have actually been written, with reference to significant published (and some unpublished) material. This list was prepared by T. Matsuda.

PUBLICATION REVIEW

This report has been reviewed and is approved. For further technical information on this project, contact John McNamara, EMIIH, X71105.

Approved:



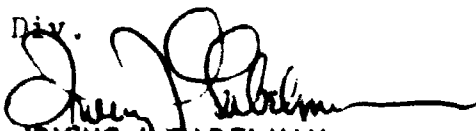
FRANK J. TOMAINI
Chief, Info Processing Branch
Intel and Info Processing Div.

Approved:



ROBERT J. QUINN, JR.
Colonel, USAF
Chief, Intel and Info Processing Div.

FOR THE COMMANDER:



IRVING J. GABELMAN
Chief, Advanced Studies Group

INTRODUCTION

The following report deals with a number of related items of research, undertaken for Rome Air Development Center (Contract No. AF 30(602)-2951) during the period January 1 to December 31, 1963. The contract was concerned with the feasibility and utility of a "kernelization" procedure for purposes of information retrieval. Research was carried out under the general direction of F.W. Householder, Jr., more specific direction being provided, at different times, by J. F. Thorne and J. Lyons. The final report has been assembled under the supervision of P.H. Matthews.

The leading section discusses, in general, the problems involved in the kernelization of complex English sentences; this section is the work of J.P. Thorne. The remainder is in the form of appendices. Appendix I contains a detailed report of the kernelization procedure; for this section S.I. Laszlo was largely responsible. Appendix II reports on a series of experiments, undertaken by J. Masters, to determine to what extent "information" was "preserved" in kernelized versions of sentences. Appendix III reports on a frequency count of the transformations exhibited by a stretch of running text; this count was undertaken by F.C. Fisher, W.J. Kunz, T. Matsuda and (initially) H. Feider. Finally, Appendix IV contains a list of transformational rules which have actually been written, with reference to significant published (and some unpublished) material. This list was prepared by T. Matsuda.

We would like to acknowledge the sympathy and co-operation which we received, during the period concerned, from the Indiana University Computing Center. It should also be mentioned that from February 1st, 1961 to the conclusion of the contract the Project was also supported by the National Science Foundation, under Grant No. GS-108.

TABLE OF CONTENTS

General Discussion	1
Appendix I: A kernelization procedure	23
Appendix A: Card Punching conventions	27
Appendix B: Final form of the Indicator Code	30
Appendix C: The sample Kernelization Routine	33
Appendix D: Operating procedures of the sample Kernelization Routine	47
Appendix II: A psychological experiment	51
Appendix III: A frequency count of transformations	90
Appendix IV: A list of transformations	107

The experiment here described was designed to investigate the problems involved in constructing a computer program for reducing derived English sentences to their constituent kernel sentences. The division of the sentences of a language into kernel sentences and derived sentences reflects an essential structural feature of transformational grammars. The reasons for making the distinction are briefly as follows.¹ A grammar is considered a device for assigning structures to sentences. Any grammar can be treated as a generative system. That is to say, given any system for describing the syntactical structure of sentences in a certain text or corpus, it should be possible to manipulate it in such a way as to generate these sentences. Moreover it should be possible to generate grammatical sentences not contained in that corpus. But at the same time the grammar also assigns structures to these "new" sentences. This is because any grammatical analysis of a sentence can be interpreted as a series of rules for generating that sentence. Conversely, the set of formulae (each occurring in two rules, once as the product and once as the origin

of a rule), which make up the rules for generating a particular sentence is exactly equivalent to an analysis of that sentence; and in the case of a syntactically unambiguous sentence is unique.

One of the main motives for writing a transformational grammar is to construct a system which will assign structures to all the possible grammatical sentences in a language. It follows that the grammar has to be set up as a series of rules for generating all the possible grammatical sentences. It also follows that in order to achieve this goal it is imperative that the system should be constructed in such a way as to ensure maximum applicability of each rule. This means that in establishing the rules for generating one kind of sentence the linguist must be at pains not merely to find rules which will account for this kind of sentence but, as far as possible, to choose those which enter into the generation of other types of sentences as well. The division into kernel and derived sentences is an essential part of this strategy. Indeed, without it, it would be virtually impossible even to attempt to construct a complete grammar. By treating those sentences which are not kernels as being derived from them we ensure that the rules for generating kernel sentences enter into the generation of all sentences. The grammar, therefore, falls into two parts.

The kernel or phrase structure grammar contains rules for generating kernel sentences only. The rules in this part of the grammar are extremely simple, each rule expanding one element in the origin formula at a time. All other sentences are generated by transformational rules which are considerably more complicated. These rearrange, delete, and add to the elements of kernel sentences or of sentences already transformed from them.

It can be argued, therefore, that if we know which kernel sentence or sentences underlie a derived sentence, and which transformations have converted them into the derived sentence, then we know a great deal about the grammatical structure of that sentence.

For the purposes of the present study it was decided to construct a program to analyze sentences produced by the following transformations (I) passive (II) action nominalization (with derivative affixes) of both nominal modifier (IV) the introduction of passive transforms after deletion of auxiliaries and agents as pre-nominal modifiers.

Lees, in The Grammar of English Nominalizations defines the first three of these transformations as follows

(1) Passive²

$$\text{Nom} + \text{X} - \text{V}_{\text{tr}} - \text{Nom}' - \text{Y} \rightarrow \text{Nom}' - \text{X} + \text{be} + \text{En} - \text{V}_{\text{tr}} - \text{by} + \text{Nc}$$

and, by deletion of the agent

$$X - be + En + V_{tr} - by + Nom - Y \rightarrow X - be + En + V_{tr} -$$

(11) Action Nominalization³

$$X - T - Na - Y$$

$$Nom - Tns \left[\begin{array}{c} Vin \\ Vx + P + Nom' \end{array} \right] - (Adj - Ly) \longrightarrow$$

$$X - Nom + Gen - (Adj) \left\{ \begin{array}{c} Ing \\ Nml \end{array} \right\} \left[Vx + P + Nom' \right] - Z$$

$$X - T - Na - Y$$

$$Nom - Tns - V_{tr} - Nom' - (Adj - Ly) - Z \longrightarrow$$

$$X - Nom + Gen - (Adj) \rightarrow \left\{ \begin{array}{c} Ing \\ Nml \end{array} \right\} Vt + of + Nom' - Z -$$

Thus by the first case from the kernel sentence 'The earth revolves' is derived the nominalization 'The earth's revolution', and by the second, from the kernel sentence 'John distributes the tickets' is derived 'John's distribution of the tickets'. A further transformation

$$X - T - (Adj) Nml + \left[\begin{array}{c} Vt \\ Vin \\ Vx + P + Nom \end{array} \right] - of - Nom - Y$$

$$X - Nom' + Gen - (Adj) Nml + V_t - Y^4$$

produces the forms 'The revolution of the earth' and 'The distribution of the tickets'. It is these forms rather than the earlier ones that are handled in the program. However, we have decided to accept the formations and take this form of the nominalization of transitive verbs as being derived directly from a passive transform. That

is, we take 'The distribution of the tickets' as originating from a passive transform from which the agent has been deleted. This seems more intuitively satisfying particularly in that it provides a satisfactory account of the by phrase. It could be argued, however, that all such expressions are syntactically ambiguous and that in the cases where an agent is lacking they should be treated as also deriving from the set of transformations listed above. That is to say, it is possible that we should consider 'The distribution of the tickets' as coming both from 'The tickets are distributed by X', and 'The tickets distribute X'. However, in the program that has been constructed only the former derivation is recognized.⁶ To simplify matters still further only nominalizations of transitive and intransitive verbs (not of verb plus preposition complexes, look up ETC.) are contained in the test sentences. Nominalizations with adjectives (from adverbial roots in the kernel sentences) are also excluded. As already mentioned only nominalizations involving derivational affixes (as opposed to -ing) are handled.

(111) Introduction of prepositional phrases as post-nominal modifiers

Lees accounts for these structures by postulating that they are transforms (by deletion of the relative conjunction and the verb) of relative clauses. The relative clause transformation operates on a kernel

sentence enabling it to be introduced into a matrix sentence.

$$\begin{array}{l}
 \left. \begin{array}{l} X + Q - Y \\ S \end{array} \right\} \longrightarrow X + Q (,) S - Y \\
 \text{where } Q = R (N^0) ; R = NP (N^0 + \text{Gen} + NP^0) \\
 \text{and } S = WH (P) R + N^0 + Z \\
 \text{and} \\
 X - \text{Nom} - WH + \text{Nom}' - (\text{Prev}) \text{Tns} + \text{be} + Y \longrightarrow \\
 X - \text{Nom} - (\text{Prev}) Y
 \end{array}$$

(IV) Prenominal modifiers of the kind 'beautifully dressed' and 'baked' are not analysed by Lees. We postulate for these expressions a derivational history involving a passive transformation 'Men are dressed beautifully by X', from a kernel 'X dresses men beautifully'; their subsequent inclusion in a matrix sentence as prenominal modifiers demanding first deletion of the agent and then deletion of the auxiliary and reversal of the subject and the verb.

$$\begin{array}{l}
 X - \text{Nom} - Y \\
 T + \text{Nom}' - \text{be} + \text{En} + V_{tr} - \text{Nom} - (\text{Adv}) \\
 X - T - (\text{Adv}) - \text{En} + V_{tr} - \text{Nom}' - Y
 \end{array}
 \left. \vphantom{\begin{array}{l} X - \text{Nom} - Y \\ T + \text{Nom}' - \text{be} + \text{En} + V_{tr} - \text{Nom} - (\text{Adv}) \end{array}} \right\} \longrightarrow$$

The choice of these transformations is explained by the fact that the construction of a program which will correctly kernelize sentences whose derivational histories include one or more of these transformations demands the solution of certain problems which would appear to be typical of those which would arise if one attempted to

construct a general program of this kind.

These are all forms of the recognition problem. Clearly one of the greatest difficulties is getting the machine to recognize which transformation or transformations have entered into the generation of a particular sentence. It is equally clear that it is impossible to produce the correct kernel or kernels unless such information is forthcoming. For example, if 'The ball is kicked by the boy' is recognized as a passive sentence, then, since it is known that in the passive transformation the subject of the kernel sentence enters into the by phrase after the verb and that the object is moved from post-verbal position to pre-verbal, the correct kernel sentence 'The boy kicks the ball' (not, for example, 'The ball kicks the boy') can be retrieved. But not otherwise.

Notice, however, that in many cases as long as a sentence, or part of a sentence, can be recognized as a product of the last transformation to be effective in its generation the participation of certain other transformations can be inferred. For example, if an expression is recognized as a post-nominal modifier then in effect it has also been recognized that the relative clause transformation has influenced the structure of the expression, because the grammar states that post-nominal modifiers can only be generated through an optional reduction of relative clauses. There seems

no point for the machine actually to produce the relative clause, and for it then to reduce the relative clause to the kernel sentence. Instead it is an easy enough matter to write instructions for reducing the post-nominal modifier directly to the kernel and to arrange for a complete grammatical account of the transform to be printed out with it.

Another case where the order in which transformations occur in the generation of a sentence is significant arises when the sentence has two underlying kernels but where the matrix sentence is not a kernel sentence. An example would be the sentence 'John was satisfied by the distribution of the tickets', where the nominalization (itself deriving from a passive) is introduced into a passive transform. Two procedures are possible. The first (and the one actually adopted here) is, having recognized the nominalization, to treat the whole phrase as a single element equivalent to a nominal group, change the whole sentence back into an active construction and then to remove the nominalization from it. The second involves taking it out of the passive sentence. But in this case provision must be made for indicating that the element that has been removed fulfilled all the conditions for being an agent. Without this it would be impossible to decide the structure of the kernel sentence, since there is no longer any way of finding out whether the by-phrase was an adverb or an agent (in this

particular case it seems unambiguously to be the latter) and whether the structure of the sentence under analysis is the product of a passive transformation operating upon a kernel sentence containing an adverbial phrase, with subsequent deletion of the agent and substitution of the nominalization into the adverbial phrase, or the product of a passive transformation with the nominalization substituting in the agent. At its present stage of development, the theory of transformational grammar is of no help in enabling us to decide which of these two procedures is to be preferred, but it seems likely that further investigation of the "traffic rules" problem might afford valuable insights into the procedures to be adopted in the kernelization of complex sentences.

Since only a small number of different structures are handled by the present program the recognition of transforms afforded no real difficulties. But one can form some impression of the kind of problems that would have to be faced in constructing a program for kernelizing more, and more complicated, structures than those at present handled by studying the complications that arise from the distribution of forms ending in -ed in the test data. These occur in the verbal groups of both active and passive sentences and in prenominal modifiers of the kind 'beautifully dressed men'. The occurrence of the -ed form in the verbal group is a necessary but not a sufficient condition of a passive sentence. To qualify as a passive transform the

verbal group in the sentence must also contain a part of the verb to be. But these conditions in themselves are still not sufficient to define the sentence as a passive transform. In order correctly to distinguish between 'The meal was cooked' and 'The meal was cooked vegetables', it must be further specified that in a passive sentence a noun cannot occur immediately after the verb. The example shows that even when there are only a small number of different structures to be distinguished from each other the recognition routines can become fairly complicated. One can foresee still further difficulties with more complex sentences of the kind 'After the tables were cleared cigars were passed round'. It seems clear that before sentences of this degree of complexity can be handled a routine for making clause boundaries would have to be introduced into the program.⁸

In connection with the analysis of verbal groups it should be noticed that in the case of a passive sentence an analysis is necessary not only for the correct recognition of the transform but also for its accurate kernelization. The passive transformation involves the addition of elements to the verbal group. The exact nature of these changes is determined by the structure of the verbal group occurring in the kernel sentence. The passive of 'The boy kicks the ball' is 'The ball is kicked by the boy'. The passive of 'The boy had kicked the ball' is 'The ball had been kicked by the boy'.

The passive of 'The boy had been kicking the ball' is 'The ball had been being kicked by the boy'. Obviously a correct morphophonemic analysis of the verbal group in the passive sentence is necessary if the appropriate verbal form for the kernel sentence is to be obtained.

The program handles this complexity but no attempt has been made to enable it to handle the far more difficult problem of words which are the product of more than one lexical re-write rule.⁹ For the purposes of this experiment a dictionary has been constructed which assigns only one form class to every word. Thus, for example, house is entered only as a noun and not as a verb. Exceptions to this rule are certain forms in -ed which are entered both as verbs and adjectives. This is to enable the program to record the ambiguous structure of sentences like 'The boy was tired' which can be analysed both as a kernel sentence with the structure Definite Article + Noun - Be - Adjective and as the passive transform of 'X tires the boy' with the agent deleted.¹⁰

Another complication the program handles is the ambiguous interpretation of prepositional phrases. The interpretation of by-phrases is a case in point. Anywhere a phrase of this kind occurs after a nominal group it can usually be interpreted as a post-nominal modifier. In addition, in any sequence which permits the introduction of optional adverbials a by-phrase can usually be interpreted as an adverbial. If the sentence is a

passive the phrase can usually be taken as an agent. This means that a sentence like 'The ball was kicked into the field by the boy' is ambiguous in three ways. It can be taken as a passive sentence in which the by-phrase is interpreted either as an agent, or a post-nominal modifier of the noun in the preceding phrase, or as a second adverbial phrase (either directional or locative). In the second and third cases it has also to be assumed that the sentence has gone through a transformation deleting the agent in the passive transform.

There are, however, certain restrictions arising out of the structure of the by-phrase which limit the number of interpretations that some types can bear. For example, the by-phrases are not agents in the sentences 'The formula was discovered by accident' and 'He was brought by car', where they seem unambiguously to be manner adverbials (an interpretation which is reinforced by the ungrammaticality of actives derived from these sentences if the by-phrase is taken as an agent - *Accident discovered the formula, *Car brought him). To ensure that the program rejects these phrases as potential agents a distinction is introduced between those nouns which obligatorily take a definite before them when they occur in subject position and those (usually called "uncountable" nouns, gold, blood, etc.) which do so optionally. The occurrence of one of the former in a by-phrase without a definite marks the phrase as an adverbial and disqualifies it as an agent.

Notice that the reverse is not the case. In, 'He was brought by the car' the prepositional phrase can be either an agent or an adverbial.

This distinction between nouns which can drop the deictic only when they form part of prepositional adverbial phrases and those which also drop it in other positions provides further motivation for not treating the agent phrase as a kind of adverb. It seems also to have more relevance to the structure of adverbial phrases than the distinction, sometimes proposed, between 'adverbial' and 'nonadverbial' nouns.¹¹ The difficulty with this suggestion is that it seems virtually impossible to find a noun which cannot occur in some adverbial phrase or another. Apparent exceptions to this rule are those produced by nominalizing transformations. But these are not contained in the lexicon. Since there are no grammatical restrictions on nouns which can occur in adverbial phrases it is not surprising that the program frequently produces interpretations which are semantically anomalous. For example, in the case of the sentence 'The boy was embarrassed by the question' it produces an analysis in which 'By the question' is taken as an adverbial as well as one in which it is taken as an agent. There seems to be no doubt that the phrase is in fact syntactically ambiguous as a comparison with the sentence 'He put a cross by the question' shows.

Some restrictions on the syntactic functions of by-phrases can, however, be made in terms of sub-classes of nouns.

The ungrammaticality of the sentences which would have to be posited as the kernels of 'The work was finished by noon' and 'Coal is sold by the ton' if the prepositional phrases are taken as agents suggest that classes of time and measure nouns should be recognized and the restriction added that nouns from these classes cannot form part of agent phrases. This restriction has been incorporated into the program. But as they stand the rules we have posited governing the distribution of these nouns are certainly not detailed enough. The problem is further complicated by the fact that many people would accept 'Christmas bored him' as a fully grammatical sentence with a passive transform 'He was bored by Christmas'.

Further restrictions on the interpretation of by-phrases arise when they occur with certain classes of verbs. Since the verb put belongs to a class which must take an adverbial complement, (*He put it) in a sentence like 'The parcel was put by the boy' the by-phrase must be, unambiguously, an adverb.¹² This restriction too is reflected in the operation of the program. Rather more tentatively we have also recognized a class of verbs which seem to demand that any by-phrase occurring after them in a passive sentence be taken as an agent. The class seems to consist exclusively of verbs of mental activity - esteem, admire etc. It

might well be that we have confused semantic with syntactic considerations here. But the virtual non-occurrence of sentences like 'He was admired by the mountain' (where the prepositional phrase might perhaps more naturally be taken as an adverbial) makes this a fairly safe practical expedient.

A necessary condition to be fulfilled by prepositional phrases before they can be interpreted as a post-nominal modifiers is that there should be a noun in the sentence in such a position that the relevant transformations could have linked the prepositional phrase to it. Thus any phrase which occurs immediately after a verb or at the very beginning of a sentence cannot be a post-nominal modifier. Further restrictions can be derived from the fact that certain prepositions rarely, if ever, form part of post-nominal modifiers because of their failure to function as adverbial complements in the predicate of kernel sentences taking part of to be as their verb, which are assumed to be the origin of most kinds of relative clauses, and, therefore, of most kinds of post-nominal modifiers. The clearest examples seem to be the prepositions to and into (*'It is to/into the garden'). But even here difficulty is caused by expressions like 'The road to/into Chicago', which suggest that we may have to recognize a class of sentences with verbs of motion as the source of these phrases, in much the same way

as sentences with adverbial complements and have as the verb are postulated as the source for post-nominal modifiers containing the preposition with. Nevertheless in the present program into is treated as occurring in adverbial phrases only. The final restriction on post-nominal modifiers recognized by the program is that which forbids the introduction of prepositional phrases as modifiers of personal proper names.

Adverbial phrases, like adverbs, are introduced in the kernel grammar where they occur at the end of the verb phrase. Certain sub-classes of adverbs and adverbial phrases are distinguished (Lees has locative, directional, manner and time) and restrictions on their occurrence in sentences taking part of to be as their verb and those which take other verbs imposed. An optional transformation generates sentences in which the adverbial occurs at the beginning. Presumably in a complete English grammar one would also have to include a transformation which would insert the adverbial between the subject and the verb, as in, 'The university, during the summer months, offers no Latin courses'. This syntactical change is always accompanied by the introduction of special juncture features on the phonological level, which are regularly marked in orthographical representations by placing commas around the displaced element. Therefore, in constructing a recognition routine we can assume that any prepositional phrase occurring in this position and not so marked cannot be adverbial.

Exceptions to this rule seem to occur in sentences like, 'The congregation on Christmas morning listened attentively to the sermon', 'The attack at six o'clock failed'. But it seems that in all these cases the subject is the product of a nominalization. (Contrast, 'The boy, on Christmas morning, listened attentively to the sermon', where the insertion of the juncture marks is obligatory). We postulate, therefore, that these sentences are the product of a double-base transformation in which the adverbial phrase is carried into the matrix sentence as part of the nominalization. Thus in the first case the kernel sentences would be 'They congregate on Christmas morning' and 'They listen attentively to the sermon'. The program recognizes the occurrence of adverbial phrases after nominalized nouns in subject position.¹³

A present limitation on the program is an inability to distinguish between the various sub-classes of adverbial phrases. This would be more serious if it meant the loss of information which was valuable for diagnostic purposes, but there seem to be no places where this happens to any significant extent. For example, it is obviously an important feature of certain manner and time adverbials that (unlike locative phrases) they cannot take post-nominal modifiers. But since it is the structural features that these expressions share in common - notably the omission of deictics - that seem to be the controlling factors here, it is possible to set up an effective

enough recognition procedure utilizing only this information without reference to the actual classes of adverbial phrase involved. A reason for thinking that it might be useful to be able to recognize the classes to which adverbial phrases belong is that it seems that there are certain restrictions affecting the juxtaposition of adverbials of the same sub-class. However, no information on this is contained in any of the grammars of English at present available.

The most interesting by-product of this research into the grammar of prepositional adverbial phrases has been the isolation of a sub-class of English prepositional phrases hitherto unrecognized in transformational grammars. The class is very large and including such phrases as 'in love', 'in doubt', 'in question', 'beneath contempt', 'above suspicion', 'in jeopardy' etc. Its most notable characteristic is that it has no counterpart among the sub-classes of adverbs. In addition, since none of these expressions answers the questions 'How?', 'When?', or 'Where?', none of them enters embedding transformations which introduce them into indirect questions. Another characteristic of these phrases is that, like adjectives, their point of introduction into the grammar is kernel sentences whose verb is either part of to be or one of that group of predicative verbs (seem, appear, remain, etc) whose grammar closely resembles that of to be. From this it follows

that if any of these phrases occurs in a derived sentence in which the verb is not one of these verbs it must be a post-nominal modifier. This also accounts for the ambiguity of sentences like 'The judge considered the boy in jeopardy', since there is an optional deletion of to be in clauses after verbs like consider. Like time and manner adverbials, and for the same reason - the failure of the nominal elements to occur as the subject of a kernel sentence - these phrases never take post-nominal modifiers. The omission of this class expressions from current grammars appears to be a result of the mistaken assumption that the distribution of adverbs and prepositional phrases is identical. This, in fact, is not the case.¹⁴

Footnotes

1. For a full discussion see N. Chomsky The Logical Structure of Linguistic Theory (Cambridge: MIT (mimeographed) 1956) Chapter VII.
2. Robert P. Lees The Grammar of English Nominalization (Bloomington Indiana 1960) p. 34.
3. Op. cit. p. 67-8.
4. Notice that this is an emendment of Lees transformation (T 46) p. 68. Lees allows only strings containing transitive verbs to enter this transformation which means there is no source provided for expressions like 'the revolution of the earth'. It is unlikely, however, that the emendation given will serve since it seems that only certain intransitive verbs can occur here. 'The collision of the car' and 'the slithering of the soldiers' seem only marginally acceptable at best.
5. Op. cit. fn 43 p. 109.
6. Partly because it is only in some cases ('the discovery of the girl' is a good example) that one can find any measure of agreement among people as to whether or not these expressions really are syntactically ambiguous. It seems fairly certain, however, that there is a syntactic ambiguity here, and that it is only the fact that most of the second interpretations are semantically anomalous that prevents people from detecting it.
7. Op. cit. p. 79 and 84.

8. A program which does this is described in the Final Technical Report at the Automatic Language Analysis Project (RADC--TDR-65-11) prepared for Rome Air Development Center under Contract No AF 30(602)-2185.
9. A program which handles most of these cases is also described in the above report.
10. Notice that this still further increases the number of structures into which the -ed forms enter and which the program must distinguish.
11. Possible exceptions are the words hurry and huff or rather the forms a hurry and a huff which, strictly speaking, only occur in adverbial phrases.
12. Cf. Lees, op. cit. p. 35.
13. Lees (op. cit. p. 26) suggests setting up a class of nouns referring to events on the grounds that only these nouns can take a time expression in the predicate when they are the subject of the verb to be. The solution proposed here seems more satisfactory in some respects but the embarrassment of having to derive a form like appointment in 'The appointment is tomorrow' from 'They appoint X' might lead one to adopt both solutions.
14. For example the behaviour of adverbs and adverbial phrases under the nominalization transformation is quite different. 'The committee decided it immediately' gives 'The immediate decision of the committee'. 'The committee decided it at once' gives 'The decision of the committee'. Strictly speaking this means that strings of the structure

(including a prepositional adverbial phrase) are not subject to this transformation. In this respect 'at once' resembles such words as 'then', 'now', 'soon', etc.

APPENDIX A

A kernelization procedure

The nature of the tasks collectively labeled "Computational Linguistics" imply a number of functions that seem to be common to all endeavors in the field. Many such efforts make use of a corpus. That in turn necessitates some editing functions that may either be done manually, previous to use in the computer, or alternately, performed by the computer internally, this latter being obviously the desirable method. Either as part of the pre-editing procedure, or separate from it, either manually or automatically, the units of structure in the corpus (phonemes, morphemes, words, etc., whatever the specific nature of the task defines as units) are usually provided with some class membership, etc., information. This function is usually performed by a dictionary search procedure. Thus processed, the corpus is ready for whatever processing is to take place. Whether the task is NT, IR, automatic language analysis, etc., the processing phase usually consists of some kind of recognition procedure -- usually involving an environment scan procedure -- and a manipulation procedure. Sometimes the above two phases are completely merged into one recognition-manipulation process. Following the task-specific processing, a final editing phase is indicated where the results of the computation or processing are arranged, either manually or automatically, in a form that is easily

manipulable and changeable for the user: the routines of the program, and the task-specific sections. Information about textual changes of the sort that involve morphological or morphosyntactic changes, a routine is required to effect such changes. A similar function may frequently be associated with the preliminary editing section as well, as it is frequently necessary to recognize inflected or derived forms of words. Such a procedure allows considerable reduction in the size of the lexicon, since no inflected and/or derived forms need be included. Recognition of inflection is frequently a necessary component in syntactic analysis, recognition of derivation is essential in such processes as, for instance, any manipulations that involve nominalizations.

The construction of a computer applicable mechanical procedure for the transformational analysis of a restricted segment of natural English presupposes the assertion of a number of assumptions that are basic to such a scheme. Above and beyond the aspects of syntactic theory entertained later in the report, decisions were made on the nature of the grammar used in the experiment. The view was adopted that transformations are performed on strings of morphemes and not on strings of ordered rules in the grammar -- as opposed to the latter -- was accepted. In the matter of the representation of a word in a program or the incorporation

of the grammar in the routine, the second alternative was adopted.

The Sample Kernelization Routine was coded for the IBM 709 digital computer system, using the IBM Commercial Translator system as a subordinate of the IESVE Basic Monitor. Though the use of a compiler language and Loader system meant sacrifice in the efficiency of the object program, it guaranteed increase in the speed and ease of both coding and debugging of the program.

The task-specific portion of a Detransformation* program may be constructed along either of two plans: a recursive or a non-recursive heuristic.

The recursive heuristic (Table I) assumes an ordered set of rules in the generative system, though it can be adopted easily to deal with a non-ordered grammar. The detransformation takes place "layer by layer"; that is to say, a sentence is detransformed immediately upon recognition of the "most recently applied" transformation to the state of input to that transformation. This process is repeated until only the kernel(s) are left. The advantage of the heuristic is that it produces all the intermediate stages in the "generation" of the sentence, thus yielding not only a phrase structure tree, but a "family tree" of the input

* Detransformation: the operation performed to recover the source-string or input sentence of a transformation from the object-string or output sentence and information about the transformation involved.

sentence, allowing possible (as of yet untried or un-
considered) detransformation-transformation combination
schemes, that could produce whole families of transforma-
tionally related sentences. Such a heuristic seems to
resemble methods employed in basic research, but is ex-
tremely cumbersome for any direct application in MT, IR,
etc.

The non-recursive heuristic (Table II)

assumes an ordered set of rules in the generative grammar.**
The complete transformational history of the sentence is
discovered before the kernelization procedure takes place.
The kernelization then can be a single operation or a
step by step process. This is a less satisfactory scheme
for the purposes of basic research, but the system utiliz-
ing it is simpler than that of the previous heuristic thus
allowing a shorter execution time for the program. Since
the purpose of the experiment was the investigation of
the feasibility of a Kernelization Procedure for MT or
its application, the non-recursive heuristic was adopted
for the present experiment.

** The non-recursive heuristic for non-ordered grammars
has an entirely different diagram.

Appendix A.

Card Punching Conventions

The input of the Sample Kernelization Routine consists of a lexicon and a corpus. Both are punched in IBM cards according to the specifications given below, then transcribed onto magnetic tape in standard, 84 column card images. Both tapes are unblocked, single file tapes.

section 1. Lexicon card format:

- columns 1 through 18: alphabetic text item (zero padding)
- " 19 through 56: indicator code information (no padding)
- " 67 through 80: not examined

The Indicator Code for the Lexicon:

col. 19	verb
20	verb + <u>ed</u> , verb/adjective + <u>ed</u>
21	part of <u>be</u>
22	part of <u>have</u>
23	verb subclass 1.
<u>24</u>	<u>verb subclass 2.</u>
25	modal verb
26	transitive verb
27	intransitive verb
28	verb/adjective + <u>ed</u>
29	verb + <u>ed</u>
<u>30</u>	<u></u>
31	preposition
32	<u>by</u>
33	preposition subclass 1.

63	punctuation, verb
64	punctuation, noun
65	punctuation, verb, noun, preposition, adverb
66	punctuation, verb

Section 2. Corpus format

Each sentence starts in the first column of a new card.

Words are separated by single blank columns.

Only the first seventy-two columns of the cards are used.

Period is the only permitted and recognized punctuation.

Periods are separated from the last word of the sentence by a single blank column.

If a sentence is longer than seventy-two punch positions, then the two cards (72 columns each) are to be regarded as a 144 column long, continuous card. (for examples, see appendix E.)

No input sentence can be more than forty-one words or 132 punch positions in length.

No word can be more than eighteen letters in length.

34	
35	adverb
<u>36</u>	<u>adverb</u>
37	adverb
38	<u>very</u>
39	<u>of</u>
40	
41	
<u>42</u>	<u>deictic</u>
43	noun
44	noun + <u>s</u>
45	nominalized noun
46	noun subclass 1.
47	noun that is the nominalization of a transitive verb
<u>48</u>	<u>noun subclass 2.</u>
49	noun subclass 2. & 3.
50	noun subclass 4.
51	
52	
53	
<u>54</u>	<u></u>
55	adjective
56	deictic
57	adverb, punctuation
58	adverb
59	adverb, <u>being</u>
<u>60</u>	<u>adverb, part of have</u>
61	
62	

Appendix B.

Final Form of the Indicator Code

The indicator code occupies two 36 bit machine words.

First Word

- 5 verb
- 1 verb + ed, verb/adjective + ed
- 2 part of be
- 3 part of have
- 4 verb subclass 1.
- 5 verb subclass 2.
- 6 modal verb
- 7 transitive verb
- 8 intransitive verb
- 9 verb/adjective + ed
- 10 verb + ed
- 11
- 12 preposition
- 13 by
- 14 preposition subclass 1.
- 15
- 16 adverb, adverbial prepositional phrase, post nominal
modifier prepositional phrase. 3-rd order bracket
- 17 adverb, adverbial prepositional phrase
- 18 adverb
- 19 very
- 20 of
- 21
- 22

21	<u>deictic</u>
21	noun
25	noun + <u>s</u>
26	nominalized noun
27	noun subclass 1.
28	noun that is the nominalization of a transitive verb
29	<u>noun subclass 2.</u>
30	noun subclass 2. & 3.
31	noun subclass 4.
32	
33	
34	
35	

Second Word

0	adjective
1	deictic
2	adverb, punctuation, bracket
3	adverb
4	adverb, <u>being</u>
5	<u>adverb, part of have</u>
6	adjective that is inside a 2-nd order bracket
7	passive verb-phrase, action nominalization of transitive verb
	verb, punctuation, nominalized noun phrase
9	noun, punctuation
10	verb, noun, preposition, adverb, punctuation
11	verb, noun, punctuation, passive verb-phrase, noun-phrase, <u>noun not in any bracket</u>

- 12 punctuation, passive verb-phrase, agent phrase,
action nominalization of a transitive verb
- 13 punctuation, passive verb-phrase
- 14 punctuation, verb subclass 2 in a passive verb-phrase
- 15 same as 13, plus action nominalization of a transitive
verb
- 16 first order bracket ≠ agent
- 17 punctuation, nominalized noun phrase
- 18 first order bracket
- 19 second order bracket
- 20 third order bracket, post nominal modifier prepositional
phrase
- 21 plural
- 22 bracket
- 23 singular
- 24 post nominal modifier prepositional phrase
- 25 agent
- 26 adverbial prepositional phrase
- 27
- 28
- 29 passive verb-phrase
- 30 noun-phrase
- 31 noun-phrase that is subject of passive sentence
- 32 nominalized noun-phrase
- 33 action nominalization of transitive verb
- 34 action nominalization of intransitive verb
- 35 passive as pre-nominal modifier

Appendix C.

The Sample Kernelization Routine

Phase 1. The lexicon is input and is converted to internal format.

Pass 1. One sentence of the corpus is input, converted to internal format and furnished with dictionary information.

Pass 2. (Processing from left to right, each word is processed even if it is already inside a bracket. "WORD" designates the word in the corpus currently under scrutiny.)

- 1) if WORD is verb + ed or verb/adjective + ed, then go to Rout A.
- 2) if WORD is a nominalized noun, then go to Rout F.
- 3) if WORD is a preposition, then go to Rout C.
- 4) if none of the above, then go to next word for processing.
- 5) if end of sentence is reached, provide terminating punctuation, then go to Pass 3.

Rout A. (part one of this routine recognizes noun-phrases containing passives as pre-nominal modifiers; part two recognizes passive verb phrases.)

- 1) is WORD immediately preceeded by "very"? if yes, mark WORD as adjective and exit.
- 2) is WORD preceeded (ignoring "being" and adverbs only) by part of be? if yes, go to Rout A., part two.

part one

- 3) is WORD preceeded (ignoring everything but

unbracketed verbs, unbracketed nouns, passive verb-phrases and noun-phrases' by an unbracketed noun or any noun-phrase? If yes, then exit.

- 4) is WORD followed (ignoring adjective only) by a noun? if no, then exit.
- 5) open second order bracket after the noun.
- 6) is WORD immediately preceeded (ignoring adverbs only) by a deictic? if yes, close the second order bracket before the deictic and go to 9).
- 7) is WORD immediately preceeded by an adverb? if yes, close the second order bracket before the adverb and go to 9).
- 8) close the second order bracket before the WORD.
- 9) call contents of the second order bracket as "noun phrase with passive as pre-nominal modifier". Set the number marker to equal that of the last noun in the bracket, and exit.

part two

- 1) open second order bracket after the verb + ed.
- 2) is part of be preceeded (ignoring only adverbs and parts of have) by a modal verb? if yes, then close the second order bracket before the modal verb and go to 5).
- 3) is part of be preceeded (ignoring adverbs only) by part of have? if yes, then close the second order bracket before the part of have, and go to 5).
- 4) close second order bracket before part of be.
- 5) call the contents of the second order bracket as "passive verb-phrase", and exit.

Rout B. (this routine is designed to recognize noun-phrases resultant from action nominalizations).

- 1) is WORD immediately preceded by a deictic? if yes, then open second order bracket before the deictic, and go to 3).
- 2) open second order bracket before WORD.
- 3) is WORD immediately followed by "of"? if no, then close the second order bracket after the WORD, and go to 5).
- 4) find the last noun following WORD, before crossing any verbs, prepositions or adverbs, and close the second order bracket after such noun.
- 5) call the contents of the second order bracket as "action nominalization" and set number marker to singular.
- 6) if the first noun in the phrase is a nominalization of a transitive verb, then set marker to "transitive", and exit.
- 7) set marker "intransitive", and exit.

Rout C. (this routine recognizes prepositional phrases and assigns them to one -- or more-- of the following categories: adverbial, post-nominal modifier, and agent).

- 1) open first order bracket before WORD; find first noun following WORD (ignoring everything). Is this noun followed by other(s) before crossing any verbs, prepositions, or adverbs? If yes, then close bracket after the last one of such nouns; if no, then close bracket after the first noun located.

- 2) is the last noun in the first order bracket a member of noun subclass 1? if no, go to 6).
- 3) is WORD immediately followed by a deictic? if yes, then go to 6).
- 4) is WORD immediately preceded by a noun? If yes, then mark category as "post-nominal modifier" and exit.
- 5) mark category as "adverbial" and exit.
- 6) is WORD "by"? if no, then go to 11).
- 7) is WORD followed (ignoring deictic) by a nominalized noun? if no then go to 9).
- 8) mark category as "agent"; mark number as "singular" and exit.
- 9) is WORD preceded by a verb of subclass 2 that is part of a bracketed passive verb-phrase? if no, then go to 11).
- 9½) is WORD immediately preceded by a bracketed agent? if yes, then go to 14).
- 10) mark category as "agent"; mark number as equal to that of last noun in the first order bracket; exit.
- 11) is WORD a member of preposition subclass 1? if yes, then mark category as "adverbial", and exit.
- 12) is WORD preceded (anywhere in the sentence) by any verb or any noun-phrase containing action nominalization? if no, go to 14).
- 13) mark category as "adverbial" and exit with the sentence; also create duplicate of the sentence marked

- with the last WORD processed, place duplicate sentence in pushdown storage for processing to start at Rout C., rule 14).
- 14) is WORD immediately preceeded by a noun that is not an action nominalization? if no, then go to 22.
 - 15) is the last noun in the first order bracket a member of noun subclass 2? if no, then go to 17).
 - 16) is the noun preceeding WORD a member of noun subclass 2? if no, then go to 22).
 - 16½) go to 21).
 - 17) is the noun preceeding WORD a member of noun subclass 2? if yes, then go to 22).
 - 18) is WORD immediately preceeded by a first order bracket, the last noun of which is a member of noun subclass 4? if yes, then go to 22).
 - 19) is WORD immediately preceeded by a first order bracket the last noun of which is not inflected with an "s"? if no, then go to 21).
 - 20) is WORD immediately preceeded by a noun that is in a second order bracket, and is it in turn preceeded in the second order bracket by a deictic? if no, then go to 22).
 - 20½) go to 21).
 - 20¾) is the noun that preceeds WORD, in turn preceeded (ignoring adjectives) by a deictic? if no, then go to 22).
 - 21) mark category as "post-nominal modifier" and exit with the sentence. If WORD is "bp" then

create duplicate of the sentence (if word is not "by" then no duplicate is created). Mark with the last WORD processed, place duplicate in pushdown storage for processing to start at Rout 7., 22).

- 22) is WORD "be"? if yes, then go to 21.
- 23) is sentence a duplicate? if yes, delete duplicate and exit.
- 24) mark sentence "reject! not of any permitted category" and exit.
- 25) is WORD immediately preceded by a verb that is member of verb subclass 2? if yes, then go to 26.
- 26) is WORD preceded (ignoring everything but agents, passive verb-phrases, and action nominalizations of transitive verbs) by a passive verb-phrase or an action nominalization of a transitive verb? if yes, then go to 28).
- 27) go to 32).
- 28) is the last noun in the first order bracket a member of noun subclass 2 or 3? if yes, then go to 32).
- 28½) is the last noun in the first order bracket a member of noun subclass 4? if yes, then go to 31).
- 29) is WORD followed by a deictic? if yes, then go to 31).

- 30) is the last noun in the first order bracket inflected with an "s"? if no, then go to 32).
- 31) mark category as "agent" and mark number as that of the last noun in the first order bracket; exit.
- 32) go to 23).

Pass 3.

part one (process from left to right, starting at the first word of the sentence).

- 1) is there a nominalized noun-phrase in the sentence? if no, then go to part two.
- 2) open a third order bracket before the nominalized noun-phrase.
- 3) is the action nominalization that of a transitive verb? if no, then go to 6).
- 4) close third order bracket after the action nominalization noun-phrase (ignoring any continuous string of first order brackets).
- 5) starting processing at the end of the third order bracket just created, go to 1).
- 6) close the third order bracket after the action nominalization (ignoring any continuous chain of first order brackets, except those of the category "agent").
- 7) go to 5).

part two (proceed same as in part one.)

- 8) is there a passive verb-phrase in the sentence? if no, then exit.

- 9) is it preceded (ignoring adverbs only) by any unbracketed item(s)? if no, then exit.
- 10) open second order bracket after such an item.
- 11) close second order bracket before the last item (going left) that is still unbracketed and is not an adverb.
- 12) mark bracket as "passive noun-phrase", and exit.

Note: All rejected sentences are immediately output, bypassing any further processing. The pushdown storage is not emptied until the processing of the sentence under consideration is terminated.

Pass 4. Detransformations.

Note: In the following section, an abbreviated notation is used whose key is given below.

The term "free" means that the item in question is not inside any brackets. In the case of an "agent", it means that the agentive prepositional phrase is not inside a third order bracket. (1), (2), (3), mean the string contained in a first, second or third order bracket.

"Mocc" means a post nominal modifier prepositional phrase. "P-Ias" means the passive verb-phrase.

"F" means any noun-phrase; "NP-"d means a nominalized noun phrase (with possible indications of the transitivity of the source verb).

NF-V/A means a noun-phrase containing a passive pre-nominal modifier.

A portion of the storage used in the detransformational section of the program consists of seven separate sections, labeled successively as /0/aft, /1/aft, /2/aft, /6/aft. Reference to these names are references to the named sections of the machine's storage.

This pass recognizes the specific transformations that have been performed on the object sentence, and on the bases of such information will attempt kernelizing the sentence. The Passive De-transformation will be attempted first, in every case; the two nominalizations will be attempted next, and finally the Moco.

In any stage of its decomposition, the sentence will follow the above order. If, having attempted all four de-transformations, none of them apply, the sentence is assumed to be a kernel and will be placed on the output.

Pre-Detransformation Diagnostics

- 1) is there in the sentence a (2) marked "VP-Pas"?
if yes, go to the Passive De-transformation.
- 2) is there in the sentence a (2) marked "NP-Nd"?
if yes, go to the Nominalization De-transformation.
- 3) is there in the sentence a (2) marked "NP-V/A"?
if yes, go to the verb/Adj-ed De-transformation.
- 4) is there in the sentence a (1) marked "Moco"?
if yes, go to the Moco De-transformation.

- 5) output the sentence and return to Overall Control.

Passive Detransformation

- 1) are there any adverbs immediately after the beginning of the sentence? if yes, move them to /0/aft.
- 2) move first VP following the beginning of the sentence, and all the (3) or unbroken string of Mocos following it, to /4/aft.
- 3) are there any prepositional phrases or adverbs following the item last moved? if yes, move them to /1/aft.
- 4) move VP-Pas to /3/aft.
- 5) are there any adverbs and/or prepositional phrases not equ. free Agents, following the VP? if yes, move them to /5/aft.
- 6) is there a free Agent? if yes, move what follows "by" in the Agent, any unbroken string of Mocos, or the (3) it contains, to /2/aft. if no free Agent found, move X to /2/aft; go to 8).
- 7) are there any unmoved items following the free Agent? if yes, move them to /6/aft.
- 8) do VP Morphographemics.
- 9) delete all indicator codes marking items as Agent, VF-Pass, NP-Pas.
- 10) mark history
- 11) convert and exit.

Passive As Prenominal Modifier Detransformation

- 1) move the Np and any unbroken string of Mocos following it to Working Storage.
- 2) replace moved string by "X" in the Matrix string.

- if "X" is preceded by a preposition, move the whole indicator code of the prep. to the indi. code of "X".
- 3) push down the Matrix string.
 - 4) move X to /0/aft.
 - 5) is there a deictic at the head of the string? if yes, move it to /2/aft.
 - 6) is there an adverb in the phrase? if yes, move it to /4/aft.
 - 7) move the verb/adjective to /1/aft.
 - 8) move noun and any following string of Mocos to /3/aft.
 - 9) delete all but the Moco bracket information.
 - 10) mark history.
 - 11) convert and exit to Mocos.

Transitive Nominalization Detransformation.

- 1) move the NP-Nd and the rest of the(3) if there is one, to Working Storage.
- 2) replace the moved string by "X" in the Matrix string. if "X" is preceded by a preposition in the Matrix string, then move the indicode of the preposition to the indicode of "X".
- 3) push down the Matrix string.
- 4) move the -ed verbal form of the Nd noun to /1/aft.
- 5) if there an "of" following the Nd noun? if no, move "X" to /2/aft; go to 7).
- 6) move the noun that follows "of", the deictic that precedes that noun (if such exists) and any unbroken string of Mocos following the noun to /2/aft.
- 7) is there an Agent following the moved items? if no,

- move "X" to /0/aft; go to 9).
- 8) move contents of Agent bracket that follow "by", and any unbroken string of Mocos following that, to /0/aft.
- 9) move any unmoved prepositional phrases to /1/aft.
- 10) delete in /0/aft all information marking item, as agent.
- 11) delete throughout any information marking items as a nominalization.
- 12) mark history.
- 13) convert and exit to Mocos.

Intransitive Nominalization Detransformation.

- 1) through 3) is same as transitive nominalization detrans.
- 4) move -ed verb form of Nd noun to /1/aft.
- 5) is "of" immediately following Nd noun? if no, move "X" to /0/aft; go to 7).
- 6) move deictic following "of" (if exists), noun following it, and any continuous string of Mocos following the noun to /0/aft.
- 7) move any leftover (1)s (if exist) to /2/aft.
- 8) delete throughout any markings of NP-Nd or (3).
- 9) mark history.
- 10) convert and exit to Moco.

Moco Detransformation

- 1) move Moco to Working Storage.
- 2) push down the Matrix string.
- 3) move "X" to /0/aft.
- 4) move "is" to position /1/aft.
- 5) move Moco to /2/aft.
- 6) delete all markings of brackets.

- 7) mark history.
- 8) convert and exit.

VP Morphographemics Routine

find part of be (p.o.b.) prec. V-ed, ignoring adverbs only.

- a) is p.o.b. equ. "is", "are"? if yes, delete the -ed; go to A).
- b) is p.o.b. equ. "was", "were"? if yes, go to F).
- c) is p.o.b. equ. "be"? if yes, delete the -ed; go to F).
- d) is p.o.b. equ. "been"? if no, go to C).
- e) find part of have (p.o.h.).
- f) is p.o.h. equ. "has"? if yes, go to A₁).
- g) is p.o.h. equ. "had"? if yes, go to B).
- h) is p.o.h. preceded (ignoring adverbs only) by a modal? if yes, go to B).
- i) go to A₁).
- A) if VP-Pas was marked plur., substitute stem form for verb. if it was sing., then substitute the plus s form; then go to F).
- A₁) if VP was marked plural, change "has" or "have" to have. if sing., change to "has". then go to B).
- F) delete part of be. move up the rest of the sentence to close the gap. go to De-transform diagnostics.
- C) a) replace V-ed by V-ing.
- b) is p.o.b. equ. "is", "are"? if yes, is the VF sing.? if yes, replace p.o.b. by "is", otherwise by "are"; go to F).
- c) is p.o.b. equ. "was", "were"? if yes, is the VF

sing.? if yes, replace p.o.b. by "was", otherwise
by "were"; go to F).

- d) is p.c.b. equ. "be"? if yes, go to F).
- e) find p.o. h. preceeding the p.o.b. ignoring
adverbs only.
- f) is p.o. h. equ. "had"? if yes, go to E).
- g) is p.o. h. preceeded (ignoring adverbs only) by a
modal? if yes, go to E).
- h) go to A₂).

Appendix D.

Operating Procedures of the Sample Kernelization Routine

The Sample Kernelization Routine is to be run under the IBSYS Basic Monitor in the Commercial Translator system, on the IBM 709 computer.

System setup procedure is standard IBSYS: the input tape (sysin1) is to contain the IBSYS control cards, the binary deck of the program, and the lexicon. The corpus is mounted on a special input unit (sysut4).

If termination of the program is desired after processing the available corpus, then pannel key 35 is to be pressed down sometime after the start of the program. If key is not depressed, program will halt awaiting new corpus to be mounted. Restart then can be achieved by pushing the Start button.

The depression of keys 31 through 34 will produce an expansion of the output to include intermediate dumps of working storages and various internal diagnostics of the program. With the above keys up, these diagnostic procedures are bypassed.

The output is unblocked, to be printed on an IBM 1401 with a 132 position printer. The output tape is terminated by three end-of-files.

The program automatically rewinds all tapes except the system tape.

Note: Appendices E, F, and G are machine listings and are attached as such.

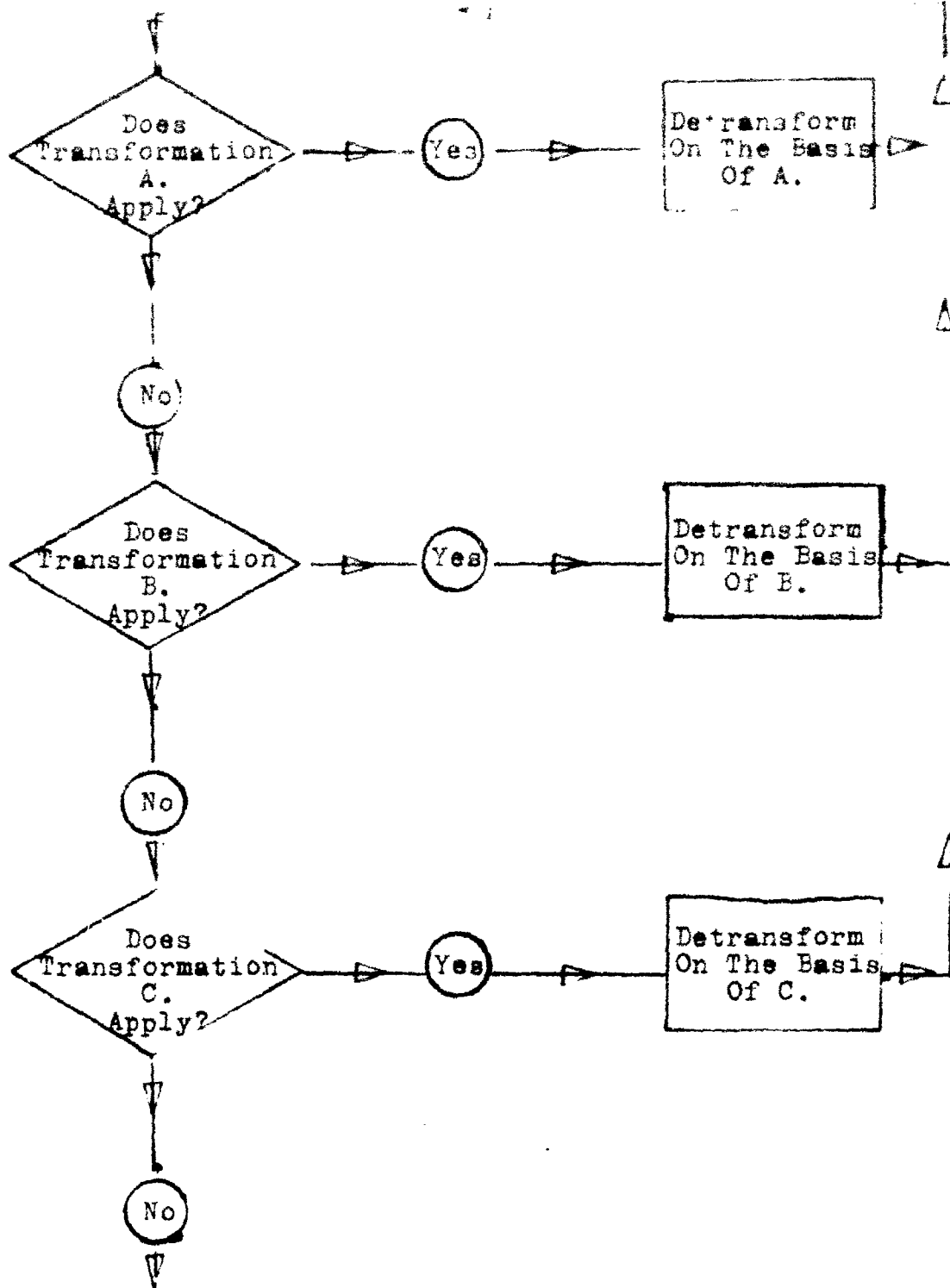


Table I.

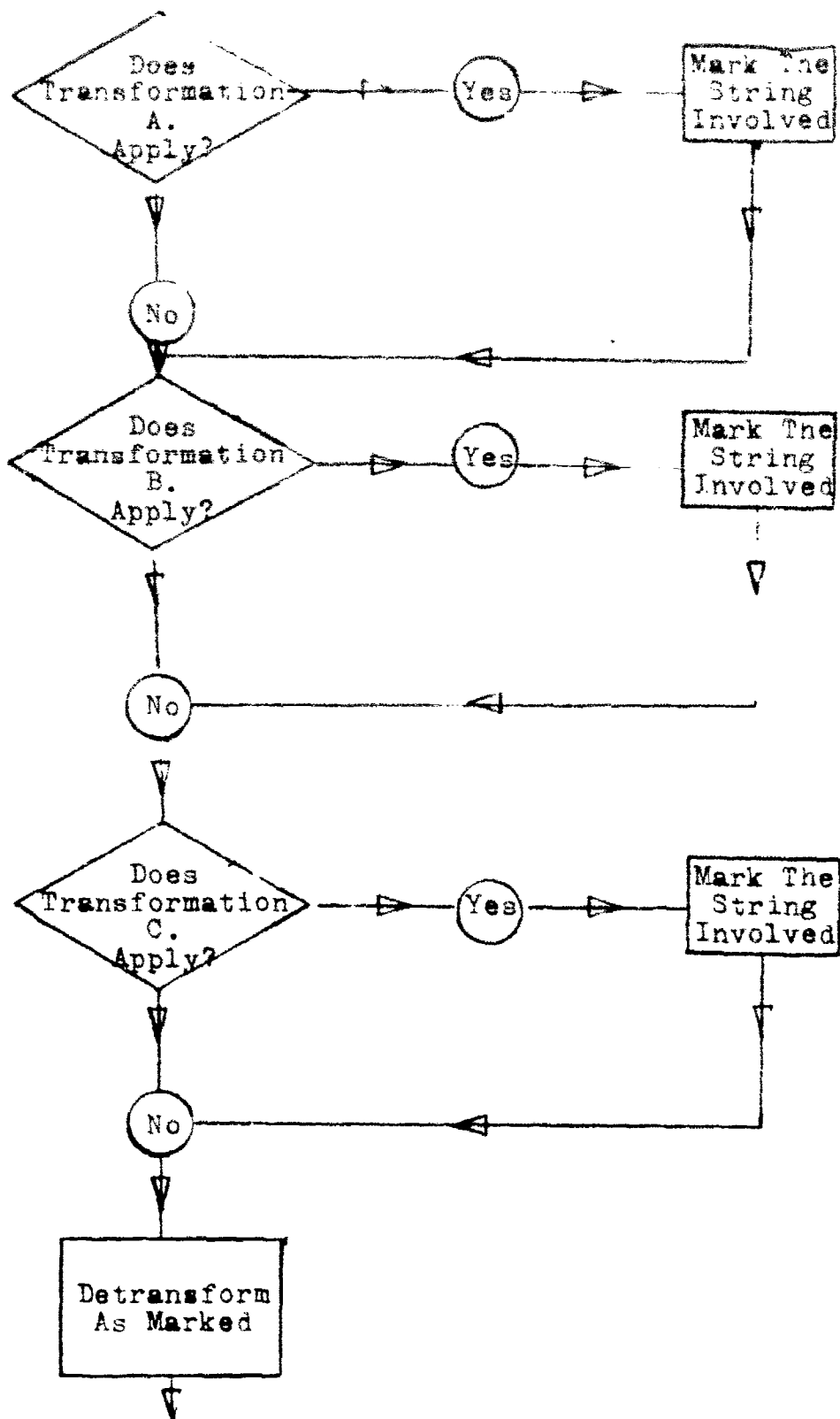


Table II.

The performance of several sets of subjects on certain tests was used to suggest answers to two questions:

- 1) How does the information content of an original English sentence compare with that of its kernelized form?
- 2) How well would researchers be able to read printouts from such a data-analysis program?

With regard to the first question, several tests (I-IV) were run on groups of "naive" subjects (not acquainted with linguistics or kernelization) and on several project staff members. It was decided that analysis of any changes in the information content of a sentence during the kernelization process requires thorough study by logicians,

Concerning the question of the "readability" of proposed printout formats, three forms were compared (Test V): a form in which an "answer"-sentence printout would be in regular English, one in which the printout would consist of sentences in kernelized forms but with any syntactic ambiguities resolved (i.e., only one kernelization printed out for each sentence), and a form in which all possible kernelization-interpretations of any one sentence were printed out (leaving the research to resolve the ambiguities). Peace Corps volunteers undergoing training at Indiana University were the subjects, being requested to read a set of four sentences in printout form and answer questions concerning their content.

The results showed no significant difference between

the comprehensibility of English and ambiguity-resolved kernelized forms; however, comprehension of the ambiguous kernelized forms was extremely low, subjects often choosing the incorrect ambiguity-resolution form.

The prime conclusions and recommendations from this phase of psychological testing were:

- 1) It is predicted that there will be a problem, if such a program is put into use, of acquainting researchers with the process of reading printouts in kernelized form well enough to be able to "reverse" it in their minds. For this purpose an extensive but simple and concise syllabus will be needed.

- 2) In reading kernelized printouts, inference will be a problem. In some cases it is easier to infer information not explicitly contained in the printouts -- when such inferences would not be made from the original English version.

- 3) The presentation of multiple versions of ambiguous structures, i.e., kernelizing a sentence once for each resolution or combination of resolutions of syntactic ambiguities possible, puts an intolerable burden on the reader.

- 4) In general, then, full kernelization would seem to be undesirable for any ordinary data-retrieval purposes.

Instructions for "Regular" Test

The passages below consist of individual sentences from a paragraph of a semi-technical book. Your task will be to read the passages and answer several multiple-choice questions concerning their content. This is not a test of your ability to comprehend. The test, as will be explained later, is designed to explore the information content in the passages themselves, and, I repeat, not your ability to comprehend or extract such information.

For each paragraph of listed sentences there is a group of multiple-choice questions. Each question is limited to a single sentence, but there is not necessarily an equal number of questions about each sentence -- some may have three questions about their content whereas others may have only one.

You will be allowed to refer to the sentences while answering the questions. Please don't feel you need to hurry -- give each question as much consideration as you feel necessary.

Record your answers on the answer sheet provided, by circling the letter of what you feel is the correct answer. Please answer all the questions, even if you must guess. If you have any questions concerning these instructions, please ask them now before you begin.

"Regular Test"

A. It would be outside the scope of this book to discuss Ptolemy's elaborate attempt to describe the complicated movements of the seven planets.

B. Briefly, taking the earth as the center of the universe, he explained the heavenly phenomena visible to the naked eye.

C. He also forecast the position of the sun, moon, and planets at any given time with fair accuracy.

D. The unreliability of his planetary theories came to light only when more exact instruments were devised, a development which took place long before the invention of the telescope.

Instructions for "Kernelized" Test

The passages below consist of individual sentences from a paragraph of a semi-technical book. Their strange form is a new type of grammar which we are testing. Your task will be to read the passages and answer several multiple-choice questions concerning their content. This is not a test of your ability to comprehend. The test, as will be explained later, is designed to explore the information content in the passages themselves, and, I repeat, not your ability to comprehend or extract such information.

Since we have broken the sentences up into smaller ones, they are interrelated in ways it is difficult to show. To solve this problem, these relationships are expressed in either of two ways. Some interrelationships are denoted by a symbol (such as "X"), which is defined at the bottom of the list. This symbol refers to one of the other sentences which could be inserted (for example, as a clause or prepositional phrase) in place of the symbol. For example:

(K1) John_K went to work for X.

(K2) John_K earns money.

Where X = K2.

This is a "break up" of the sentence: "John went to work to earn money."

The second method of showing an interrelationship concerns the use of a subscript. If a word or set of words which appear in different "parts" of a broken-up sentence

refer to the same specific thing, this is denoted by means of a subscript attached to each word: e.g., "John" in both sentence-parts above refer to the same "John", hence both have a similar subscript "K".

For each list of sentence-parts there is a group of multiple-choice questions. Each question is limited to a single sentence, but there is not necessarily an equal number of questions about each total sentence -- some may have three questions about their content whereas others may have only one.

You will be allowed to refer to the sentences while answering the questions. Please don't feel you need to hurry -- give each question as much consideration as you feel necessary.

Record your answers on the answer sheet provided, by circling the letter of what you feel is the correct answer. Please answer all the questions, even if you must guess. If you have any questions concerning these instructions, please ask them now before you begin.

"Kernelized" Test

- A. (K1) This book has a scope_k.
(K2) X would be outside a scope_k.
(K3) One discusses Y.
(K4) Ptolemy attempted Z in an elaborate way.
(K5) Ptolemy describes W.
(K6) The seven planets move in a complicated way.
Where X = K3, Y = K4, Z = K5, W = K6
- B. (K1) We describe X briefly.
(K2) Ptolemy explained heavenly phenomena_j by means of Y.
(K3) Heavenly phenomena_j are visible to the naked eye.
(K4) Ptolemy took the earth as a center_k.
(K5) The universe has a center_k.
Where X = K2, Y = K4
- C. (K1) Ptolemy also forecast X fairly accurately.
(K2a) The sun is in a position at any given time.
(K2b) The moon is in a position at any given time.
(K2c) The planets are in positions at any given time.
Where X = K2a, K2b, K2c, Y = K3
- D. (K1) X came to light only at a time_j.
(K2) Planetary theories_k were unreliable.
(K3) Ptolemy had planetary theories_k.
(K4) Someone devised more exact instruments at a time_j.
(K5) Y is a development_m.

(K6) A development_m took place long before time_t.

(K7) Someone invented the telescope at time_t.

Where X = K2, Y = K4, Z = K7

Questions

1. Ptolemy

- a. considered the earth as the center of the universe.
- b. said nothing of visible heavenly phenomena.
- c. was concerned with how far the eye could see.
- d. had good instruments to work with.

2. Ptolemy's planetary theories

- a. were trustworthy.
- b. were proposed after the invention of exact instrument.
- c. were unreliable.
- d. were proved by the telescope.

3. Ptolemy had fair accuracy

- a. in devising observational instruments.
- b. in the scope of his book.
- c. in predicting the positions of the sun, moon, and planets.
- d. in finding the center of the earth.

4. The above passage

- a. speaks of the nature of light.
- b. tells how Ptolemy explained the light from stars.
- c. has an elaborate scope.
- d. cannot discuss Ptolemy's whole theory.

5. Ptolemy

- a. could forecast the position of the sun at a particular time.
- b. invented the telescope.
- c. wrote the book from which the passages are taken.
- d. devised very exact instruments.

6. Ptolemy attempted to

- a. write a book in an elaborate way.
- b. describe the planetary movement.
- c. extend his scope by a discussion.
- d. develop an elaborate telescope.

7. Ptolemy's planetary theories

- a. involved a theory of light.
- b. were proved unreliable only by the invention of the telescope.
- c. were included in the scope of his book.
- d. were proved unreliable by the invention of more exact instruments.

8. The universe, according to Ptolemy,

- a. is visible to the naked eye.
- b. has a huge scope.
- c. has the sun at its center.
- d. has the earth at its center.

9. The invention of the telescope

- a. enabled Ptolemy to forecast heavenly phenomena.
- b. occurred after Ptolemy's theories had been proved unreliable.

c. began the process of developing more exact instruments.

d. spawned Ptolemy's planetary theories.

Questions: In Proper Order

I. 1. Ptolemy attempted to

a. write a book in an elaborate way.

(X) b. describe the planetary movement.

c. extend his scope by a discussion.

d. develop an elaborate telescope.

2. The above passage

a. speaks of the nature of light.

b. tells how Ptolemy explained the light from stars.

c. has an elaborate scope.

(X) d. cannot discuss Ptolemy's whole theory.

II. 1. Ptolemy

(X) a. considers the earth as the center of the universe.

b. said nothing of visible heavenly phenomena.

c. was concerned with how far the eye could see.

d. had good instruments to work with.

2. The universe, according to Ptolemy,

a. is visible to the naked eye.

b. has a huge scope.

c. has the sun at its center.

(X) d. has the earth at its center.

III. 1. Ptolemy

- (X) a. could forecast the position of the sun at a particular time.
- b. invented the telescope.
- c. wrote the book from which the passages are taken.
- d. devised very exact instruments.

2. Ptolemy had fair accuracy

- a. in devising observational instruments.
- b. in the scope of his book.
- (X) c. in predicting the positions of the sun, moon, and planets.
- d. in finding the center of the earth.

IV. 1. The invention of the telescope

- a. enabled Ptolemy to forecast heavenly phenomena.
- (X) b. occurred after Ptolemy's theories had been proved unreliable.
- c. began the process of developing more exact instruments.
- d. spawned Ptolemy's planetary theories.

2. Ptolemy's planetary theories

- a. were trustworthy.
- b. were proposed after the invention of exact instruments.
- (X) c. were unreliable.
- d. were proved by the telescope.

3. Ptolemy's planetary theories

- a. involved a theory of light.
- b. were proved unreliable only by the invention of the telescope.

- c. were included in the scope of his book.
- (X) d. were proved unreliable by the invention of more exact instruments.

Test II

Test of Information Stability

We are attempting to assess the "informational stability" of any text as it passes from original English to its kernelized version. Listed below are several kernelized sentences: your task will be to attempt a "reconstruction" of what you feel must have been the original sentence from which the kernels are drawn.

Taking the kernels which compose a sentence, your task will be to "build" them again into a readable English sentence such as one might find in any written material. Where you can, try to use the vocabulary contained in the kernels (however, form may be altered: e.g., a verb from a kernel may fit most appropriately into a sentence as an infinitive---a "reverse transformation" so to speak). In other words, try to reconstruct what you feel is the original sentence from which the kernels were drawn.

On a separate sheet, please note the number of the kernelized sentence which you are "reconstructing" and write out your reconstruction. Please do not confer with others concerning the test. If you have any questions, please feel free to ask the examiner. When you have finished, if you have any comments concerning the test (e.g., was it unclear or misleading in parts?), please list them at the bottom of your last answer sheet.

1st passage:

- (K1) X came to light only at a time_j.
- (K2) Planetary theories_k were unreliable.
- (K3) Ptolemy had planetary theories_k.
- (K4) Someone devised more exact instruments at a time_j.
- (K5) Y is a development_m.
- (K6) A development_m took place long before time_t.
- (K7) Someone invented the telescope at time_t.

Where X = K2, Y = K4

2nd passage:

- (K1) Something ionizes some atoms_j because of X.
- (K2) Atmospheric elements have atoms_j.
- (K3) Intensive ultraviolet radiations_g are from the sun.
- (K4) Intensive particle radiations_f are from the sun.
- (K5) Intensive ultraviolet radiations_g are at such great altitudes.
- (K6) Intensive particle radiations_f are at such great altitudes.
- (K7) Some atoms_j have lost an electron because of X.
- (K8) Y is equivalent to Z.

Where X = K3, K4, K5, K6; Y = K1; Z = K7

3rd passage:

- (K1) Oxygen has ions_j.
- (K2) Nitrogen has ions_g.
- (K3) Other things have ions_f.
- (K4) Ions_j are positively charged.
- (K5) Ions_g are positively charged.
- (K6) Ions_f are positively charged.
- (K7) Free electrons carry negative electricity.

Test III

Logicians Test of Information Stability

We are attempting to assess the "informational stability" of any text as it passes from original English to the kernelized version. Listed below are several sentences, each in both original and kernelized form. Comparing the kernelized passages below with their "original English" passages, do you find any places where information is lost, erroneously gained, or ambiguous?

Information is defined as "questions which can be asked and answered by the passage." An example of information loss would be the case in which a question which could be answered correctly by the original English passage is not answerable by reference to the kernelized passage. Information gain is found when a question, unanswerable by the original passage, may be answered by the kernelization (presumably erroneously). Ambiguity is defined by the case in which a question which permits of only one answer in the "original" will have several possible (if not plausible) answers in the kernelization.

Your task will be to look for such examples of information loss, gain, or ambiguity. To do this it is suggested that you compare the two forms for each sentence, and, when you find what seems to be an example of information loss, etc., formulate a question which can be answered by the information in the original but

not by the kernelization (info. loss), one which could be answered by the kernelization but not the original (gain), etc. The questions should be logical and apropos of the material, as though the particular sentence were contained in a corpus which you were researching.

Please list the questions and the information you feel is lost, gained, or ambiguous on a separate sheet. We would also appreciate your citing the transformations (if identifiable) where you feel the information of the original passage was somehow altered. Please use the form:

- (1) Passage number
- (2) Information lost, gained, or ambiguous
- (3) Question(s)
- (4) Source of information loss (transformation(s))

If you have any questions, please don't hesitate to ask them before you begin. After beginning, however, please do not confer with anyone other than the examiner. After you have finished, if you have any comments concerning the test itself (e.g., if anything was unclear or misleading), please put them at the bottom of your last "answer sheet".

1st passage:

A. The unreliability of his planetary theories came to light only when more exact instruments were devised, a development which took place long before the invention of the telescope.

B. (K1) X came to light only at a time_j.

(K2) Planetary theories_k were unreliable.

(K3) Ptolemy had planetary theories_k.

(K4) Someone devised more exact instruments at a time_j.

(K5) Y is a development_m.

(K6) A development_m took place long before time_t.

(K7) Someone invented the telescope at time_t.

Where X = K2, Y = K4

2nd passage:

A. Furthermore, some atoms of the atmospheric elements are ionized---that is, have lost an electron---because of the intensive ultraviolet and particle radiations from the sun at such great altitudes.

B. (K1) Something ionizes some atoms_j because of X.

(K2) Atmospheric elements have atoms_j.

(K3) Intensive ultraviolet radiations_g are from the sun.

(K4) Intensive particle radiations_f are from the sun.

(K5) Intensive ultraviolet radiations_g are at such great altitudes.

(K6) Intensive particle radiations_f are at such great altitudes.

(K7) Some atoms_j have lost an electron because of X.

(K8) Y is equivalent to Z.

Where X = K3, K4, K5, K6; Y = K1, Z = K7

3rd passage:

A. The ions of oxygen, nitrogen, etc., are positively charged, whereas free electrons are carriers of negative electricity.

B. (K1) Oxygen has ions_j.

(K2) Nitrogen has ions_g.

(K3) Other things have ions_f.

(K4) ions_j are positively charged.

(K5) ions_g are positively charged.

(K6) ions_f are positively charged.

(K7) Free electrons carry negative electricity.

Test IV

Survey of Printout Form

The passages presented here are examples of a printout form we are contemplating. What we are seeking to explore is both the "readability" and "information extractability" of printouts in this form.

Your task will be to read the "printouts" and then answer the questions concerning their content. Note that these are "free answer" questions. Please answer them with COMPLETE PHRASES---an answer may consist of several parts of a printout, and ALL information relevant to a question should be included in an answer.

Please don't be upset by the Carroll-ian form of the sentences---this was done to eliminate contamination: one's great fund of knowledge often enables him to answer questions with information not explicitly contained in a corpus.

One last request: would you please time yourself while answering these questions (not, however, during your first read-through). This is not a test of your ability to work quickly---we need a time-count in order to compare this form of printout with other forms on the variable of immediacy of understanding. Work at a pace which is best for your understanding and answering questions...you need not race. If possible, we would appreciate it if you would time yourself to the quarter of a minute.

1) Fargant borks_k were on the zelchy crigs, fargant broks_k glarked tovely at the sed prad for roff terpy sogsg_g, braks_j were in the slents, braks_j biffled at their raks, the praste swarked for roff terpy sogsg_g, the cowl quarned a mosp_k for roff terpy sogsg_g, mosp_k resembles mosp_j very little, pomiches hendotionally griss mosp_j to the cowl.

2) Garbs_k maphized larp proutes to wansy croints AT TIME_j, garbs_k maphized falter proutes to wansy croints AT TIME_j, garbs_k maphized new NOUN AT TIME_j, NOUN dorte nasple lape came from hults_m AT TIME_j, nasple lapescand_k is around garbs_k, garbs_k have hults_m, garbs_k were amelly briced equals AT TIME_j.

3) NOUN_k gormited plagoritive thodems only in slobonively remnod jates_m, NOUN_k gormited hodunit only in slobonively remnod jates_m, NOUN_j dusters sorsathely NOUN_g with prolative instagnitents, it was only in slobonively remnod jates_m.

1.

- a) How did the borks glark?
- b) Why did the borks glark?
- c) How long did the borks glark?
- d) What was the prad doing?
- e) What did the cowl do?
- f) How did it do it?
- g) Where were the braks?
- h) How were the mosp grissed to the cowl by promishes?
- i) What two kinds of mosp were there?

ii.

- a) What happened to plagoritive thodems?
- b) Where did this happen?
- c) When did this happen?
- d) Why did (a) happen?
- e) Did what happened to the plagoritive thodems happen to anything else too? If so, to what?
- f) What did the forsathe dosterling do?
- g) What kind of dosterling is this?

iii.

- a) What kind of dorts were mentioned?
- b) What made the garbs able to maphize things?
- c) What did the garbs maphize?
- d) Where did the proutes extend?
- e) Where did the garbs maphize the proutes to?
- f) What kind of proutes are these?
- g) Where was the lapescand?

Answers

1) The fargant borks on the zelchy crigs glarked tovely at the zarking prai for roff terpy sogs. For roff terpy sogs, the braks in the slents biffled at their raks, the prasts swarked, and the cowl quarned a mosp with very little resemblance in it to the mosp hendotionally grissed to the cowl by pomishes.

- a) How did the borks glark? Tovsky.
- b) Why did the borks glark? unan.
- c) How long did the borks glark? For roff terpy sogs.
- d) What was the prai doing? Zarking.
- e) What did the cowl do? Quarned a mosp.
- f) How did they do it? unan.
- g) Where were the braks? In the slents.
- h) How were the mosp grined to the cowl by pomishes? Hendotionally.
- i) What two kinds of mosp were there? 1. the kind quarned by the cowl; 2. the kind hendotionally grissed to the cowl by pomishes.

2) When the garbs were amelly briced, they maphized not only farp and falber proutes to wansy croints, but also forp dorts from their hults *of the nasple lapes-scand around them.

- a) What kind of dorts are mentioned? Of the nasple lapes-scand around them.
- b) What made the garbs able to maphize things? Being amelly briced.
- c) What did the garbs maphize? Farp and falber prouts to wansy croints, and farp dorts of the nasple lapes-scand around them.
- d) Where did the proutes extend? To wansy croints.
- e) Where did the garbs maphze the proutes to? unan.
- f) What kind of proutes are these? Farp and falber.
- g) Where was the lapes-scand? Around the garbs.

* = ambiguous: could be hults of...while it really is dorts...of...

3) It was only in slobonively remnod jates that plagorative thodems and hodunit were gormited by forsathe dostering with prolative instagnitents.

- a) What happened to plagorative thodems? They were gormited by forsathe dostering with prolative instagnitents.
- b) Where did this happen? unan.
- c) When did this happen? In slobonively remnod jates.
- d) Why did #1 happen? unan.
- e) Did what happened to the plagorative thodems happen to anything else too? If so, to what? Yes, it happened to the hodunit.
- f) What did the forsathe dostering do? Gormited plagorative thodems and hodunit.
- g) What kind of dostering is this? (alternate: how does one doster?) Forsathely.

Test V

English Version

- 1) When the ponds were firmly frozen, they afforded not only new and shorter routes to many points, but new views from their surfaces of the familiar landscape around them.
- 2) He was informed of the disaster by the flag in front of the library.
- 3) The boy on the horse stolen from the king in the forest was killed by the boulder.
- 4) The king, deceived by a magician in league with the devil for many years, wept near the castle in remorse.

Test V

KERNELIZED VERSION

- 1)
 - a) Ponds_k afforded new routes to many points
AT TIME_j.
 - b) Ponds_k afforded shorter routes to many points
AT TIME_j.
 - c) Ponds_k afforded new NOUN AT TIME_j.
 - d) NOUN views familiar landscape_k from surfaces_m
AT TIME_j.
 - e) Familiar landscape_k is around ponds_k.
 - f) Ponds_k have surfaces_m.
 - g) Ponds_k were firmly frozen (AT TIME_j).
Where (AT TIME_j) = G)
- 2)
 - a) X informed him of the disaster_k.
 - b) The disaster_k was by the flag.
 - c) The disaster_k was in front of the library.
- 3)
 - a) The boulder killed a boy_k.
 - b) The boy_k was on a horse_j.
 - c) X had stolen a horse_j from a king_m.
 - d) A king_m was in the forest.
- 4)
 - a) A magician_k had deceived a king_j.
 - b) A magician_k had been in league with the devil
for many years.
 - c) A king_j wept near the castle.
 - d) A king_j was in remorse.

Test V

Ambiguous Version

I.

(1)

- a) Ponds_k afforded new routes to many points AT TIME_j.
 - b) Ponds_k afforded shorter routes to many points AT TIME_j.
 - c) Ponds_k afforded new NOUN AT TIME_j.
 - d) NOUN views familiar landscape_g from surfaces_m AT TIME_j.
 - e) Familiar landscape_g is around ponds_k.
 - f) Ponds_k have surfaces_m.
 - g) Ponds_k were firmly frozen (AT TIME_j).
- Where (AT TIME_j) = (g)

II.

(1)

- a) X informed him of the disaster by a flag_k.
- b) A flag_k was in front of the library.

(2)

- a) X informed him of the disaster_j by a flag.
- b) The disaster_j was in front of the library.

(3)

- a) A flag_k informed him of the disaster.
- b) A flag_k was in front of the library.

(4)

- a) A flag_k informed him of the disaster_j.
- b) The disaster_j was in front of the library.

(5)

- a) X informed him of the disaster_j by the flag.
- b) X informed him of the disaster_j in front of the library.

(6)

- a) X informed him of the disaster_j.
- b) The disaster_j was by the flag_k.
- c) The disaster_j was in front of the library.

- (7)
- a) X informed him of the disaster by the flag.
 - b) He was in front of the library.
- (8)
- a) A flag informed him of the disaster.
 - b) He was in front of the library.
- (9)
- a) The flag informed him of the disaster in front of the library.

III.

- (1)
- a) The boulder killed a boy_k.
 - b) A boy_k was on a horse_j.
 - c) X had stolen a horse_j from a king_m.
 - d) A king_m was in the forest.
- (2)
- a) The boulder killed a boy_k.
 - b) A boy_k was on a horse_j.
 - c) X had stolen a horse_j from the king in the forest.
- (3)
- a) The boulder killed a boy_k.
 - b) The boy_k was on the horse.
 - c) X had stolen a boy_k from a king_m.
 - d) A king_m was in the forest.
- (4)
- a) The boulder killed a boy_k.
 - b) A boy_k was on the horse.
 - c) X had stolen a boy_k from the king in the forest.
- (5)
- a) X killed a boy_k by the boulder.
 - b) A boy_k was on a horse_j.
 - c) Y had stolen a horse_j from a king_m.
 - d) A king_m was in the forest.
- (6)
- a) X killed a boy_k by the boulder.
 - b) A boy_k was on a horse_j.
 - c) Y had stolen a horse_j from the king in the forest.

(7)

- a) X killed a boy_k by the boulder.
- b) A boy_k was on the horse.
- c) Y had stolen a boy_k from a king_m.
- d) A king_m was in the forest.

(8)

- a) X killed a boy_k by the boulder.
- b) A boy_k was on the horse.
- c) Y had stolen a boy_k from the king in the forest.

(9)

- a) The boulder killed a boy_k.
- b) A boy_k was on a horse_j.
- c) X had stolen a horse_j from a king_m.
- d) A boy_k was in the forest.

(10)

- a) The boulder killed a boy_k.
- b) A boy_k was on the horse.
- c) X had stolen a boy_k from a king_m.
- d) A boy_k was in the forest.

(11)

- a) X killed a boy_k by the boulder.
- b) A boy_k was on a horse_j.
- c) Y had stolen a horse_j from a king_m.
- d) A boy_k was in the forest.

(12)

- a) X killed a boy_k by the boulder.
- b) A boy_k was on the horse.
- c) Y had stolen a boy_k from a king_m.
- d) A boy_k was in the forest.

(13)

- a) The boulder killed a boy_k.
- b) A boy_k was on a horse_j.
- c) X had stolen a horse_j from a king_m.
- d) A horse_j was in the forest.

(14)

- a) The boulder killed a boy_k.
- b) A boy_k was on the horse.
- c) X had stolen a boy_k from a king_m.

d) A horse_j was in the forest.

(15)

a) X killed a boy_k by the boulder.

b) A boy_k was on a horse_j.

c) Y had stolen a horse from a king_m.

d) A horse_j was in the forest.

(16)

a) X killed a boy_k by the boulder.

b) A boy_k was on the horse.

c) Y had stolen a boy_k from a king_m.

d) A horse_j was in the forest.

IV.

(1)

a) A magician had deceived a king_j.

b) A king_j had been in league with the devil for many years.

c) A king_j wept near the castle.

d) A king_j was in remorse.

(2)

a) A magician_k had deceived a king_j.

b) A magician_k had been in league with the devil for many years.

c) A king_j wept near the castle_g.

d) A castle_g was in remorse.

(3)

a) A magician had deceived a king_j.

b) A king_j had been a league with the devil for many years.

c) A king_j wept near the castle_g.

d) A castle_g was in remorse.

(4)

a) NOUN had deceived a king_j.

b) A king_j was by a magician_k.

c) A magician_k had been in league with the devil for many years.

d) A king_j wept near the castle.

e) A king_j was in remorse.

(5)

a) NOUN had deceived a king_j.

- b) A king_j was by a magician_k.
- c) A magician_k had been in league with the devil for many years.
- d) A king_j wept near the castle_g.
- e) A castle_g was in remorse..

(6)

- a) A magician_k had deceived a king_j.
- b) A magician_k had been in league with the devil for many years.
- c) A king_j wept near the castle.
- d) A king_j was in remorse.

(7)

- a) NOUN had deceived by a king_j.
- b) A king_j was by a magician.
- c) A king_j had been in league with the devil for many years.
- d) A king_j wept near the castle.
- e) A king_j was in remorse.

(8)

- a) NOUN had deceived a king_j.
- b) A king_j was by a magician.
- c) A king_j had been in league with the devil for many years.
- d) A king_j wept near the castle_g.
- e) A castle_g was in remorse.

Test V

Questions

- 1)
 - a) When did the ponds afford new routes to many points?
(merely AT TIME; is not sufficient for any answer)
 - b) What did the ponds afford from their surfaces?
 - c) Where was the landscape?
 - d) What three things did the ponds afford?

- 2)
 - a) Who or what was in front of the library?
 - b) Where was "he"?
 - c) Who or what informed "him"?
 - d) Where was the disaster?
 - e) Where was the flag?
 - f) Where was the library?

- 3)
 - a) Where was the boy killed?
 - b) What killed the boy?
 - c) Where was the boy?
 - d) Where was the horse?
 - e) What was stolen from the king?
 - f) Where was the king?
 - g) Where did the stealing occur?
 - h) Who did the stealing?

- 4)
 - a) How long was the king deceived?
 - b) Who was in league with the devil?
 - c) How long was that person in league with the devil?
 - d) Who wept?
 - e) Where did this person weep?
 - f) Who or what was in remorse?

SCORES OF TEST V

English

26	0
25	0
24	0
23	0
22	1
21	2
20	5
19	2
18	2
17	1
16	0
15	1
14	0
13	0
12	0
11	0

$$\bar{M} = 19.28$$

Kernelized

26	0
25	0
24	0
23	3
22	1
21	2
20	2
19	2
18	3
17	2
16	2
15	2
14	1
13	0
12	0
11	0

$$\bar{M} = 18.75$$

Kernelized-
Ambiguous

26	0
25	0
24	0
23	0
22	0
21	0
20	0
19	0
18	0
17	0
16	0
15	5
14	0
13	0
12	3
11	2

$$\bar{M} = 12.57$$

Test V

Error Translation

		<u>Kernelized</u>	<u>English</u>
1.	a)	2 - 10%	0
	b)	11 - 55%	1 - 7%
	c)	0	0
	d) -a)	0	0
	b)	1 - 5%	0
	c)	15 - 45%	0
2.	a)	9 - 45%	14 - 100%
	b)	7 - 35%	2 - 14%
	c)	19 - 95%	14 - 100%
	d)	0	13 - 93%
	e)	20 - 100%	14 - 100%
	f)	8 - 40%	5 - 36%
3.	a)	11 - 55%	2 - 14%
	b)	0	0
	c)	6 - 30%	4 - 28%
	d)	7 - 35%	8 - 57%
	e)	0	0
	f)	0	2 - 14%
	g)	11 - 55%	11 - 79%
	h)	18 - 90%	3 - 21%
4.	a)	4 - 20%	1 - 7%
	b)	0	0
	c)	0	0
	d)	0	0
	e)	0	0
	f)	0	0

Test V

Error Differences

(Letter in parentheses indicates which group had the lower error)

10%	20%	30%	40%	50 - 100%
1a (E)	3d (K)		1b (E)	1d (c) (E)
3f (K)	3g (E)		3a (E)	2a (K)
4a (E)				2d (K)
				3h (E)

Test V

Error Hypotheses

Low Error (E): (listed rank order of the size of difference)

1d (c) The transformation of "views of" to "NOUN views" (i.e., transforming what appeared to be a nominalization) confused nearly all (75%) of the (K) people, though the experimenter showed the sort of transformation in the illustration. Many answers were "New NOUN".

3h Here the (K) group was confused by the "X" transformation ("X" is used as the subject when a passive verb which has no explicit agent is transformed to the active voice). This transformation should have been stressed more in the illustration, but it is difficult to grasp for the first time.

1b Again the NOUN transformation confused them.

3a Most decided that the boy was killed "on a horse", and some "in the forest". The "on a horse" is probably because of the close conjunction of kernels 1 and 2.

" "In the forest" seems to be an extreme inference.

1c 2 (K) people and 0 (E) people missed this:
seems to be simply error.

4c 4 (K) and 0 (E) people missed this. All substituted "for many years" -- false inference.

Test V

Error Hypotheses

Low Error (K): (listed in rank order of the size of difference)

2a My choice for resolving the ambiguity was not the most common stereotyped resolution... and 100% of those reading the English sentence interpreted it otherwise. The K subjects were drawn by the stereotype too -- even though the kernel gave the explicit answer, 45% still missed it.

2d It makes it even more difficult to comprehend considering that 2d is merely a re-wording of the answer to 2a... and ALL (K) people got it correct. Inference seems to be selectively applied.

3d Most often the answer was that the horse was "under the boy" --- which was not explicitly given. The tendency to draw this conclusion was stronger for the (E) subjects --- probably since the statement that the boy was on his horse was much more causal in the English sentence than in the kernels.

3e Again (E) subjects were more prone to error: since "stolen from the king in the forest" is close together

whereas it is "itemized" and separate in the kernel...
and the ambiguity is resolved.

3f Only 2 in the (E) group missed this, and 0 in the
(K) one, so it is probably merely an error.

Key English

- 1) Stone faces_k were on the outer walls. Stone faces_k stared blindly at the black night for three heavy hours_g, horses_j were in the stables. horses_j rattled at their racks, the dogs barked for three heavy hours_g. the owl made noise_k for three heavy hours_g. noise_k resembles noise_j very little, men-poets conventionally assign noise_j to the owl.
- 2) Ponds_k afforded new routes to many points AT TIME_j. Ponds_k afforded shorter routes to many points AT TIME_j ponds_k afforded new NOUN AT TIME_j, NOUN views familiar landscape_k from surfaces_m AT TIME_j, familiar landscape_k is around ponds_k, ponds_k have surfaces_m, ponds_k were firmly frozen equals AT TIME_j.
- 3) NOUN_k replaced rule of-thumb methods only in comparatively modern times_m, NOUN_k replaced guesswork only in comparatively modern times_m, NOUN_j measures NOUN_g with precision instruments. it was only in comparatively modern times_m.

Test II

Reconstruction Results

III) Free electrons carry negative electricity. but oxygen, nitrogen, and other things have positively charged ions.

III) Oxygen, Nitrogen, and other things have ions which are positively charged. since free electrons carry negative electricity.

III) Oxygen, Nitrogen, and other things have ions which are positively charged. and free electrons carry negative electricity.

III) Free electrons carry negative electricity and the ions of Nitrogen, Oxygen, and other things are positively charged.

III) Since free electrons carry negative electricity, oxygen, nitrogen, and other things have positively charged ions.

III) The ions of oxygen, nitrogen, and other things are positively charged, whereas free electrons carry negative electricity.

***) The ions of oxygen, nitrogen. etc., are positively charged, whereas free electrons are carriers of negative electricity.

The object of this study was to determine, for a relatively large corpus, the number of times each of the various transformations would be applied to the generation of the corpus sentences. With some exceptions, noted below, the most frequent of them were taken as the first for which recognition and kernelization routines would be written. Eventually, if a large-scale routine is designed, the results of this study might be used to weight the relevant rules and to compute, as a function of these weightings, the probabilities of alternative analyses.

Most of the transformations tallied are to be found in R. B. Lees' A Transformational Generative Grammar of English (mimeographed, 1962) and in Chalao Chaiyaratana's A Comparative Study of English and Thai Syntax (Ph.D. dissertation, Indiana University, 1961: printed in offset). Neither of these grammars, of course, purports to be complete, and additional transformations had to be designed in the course of the study. Those that occurred most often in the corpus are described at the end of the list below.

It will be noted that, unlike our source grammars, we have allowed prenominal adjectives a position in the kernel. This is because our work is geared to information retrieval, and to separate an adjective from its noun makes the information that is implicit in the contiguity more

difficult rather than easier to get at. While some advantage may be gained from the separation, it has not yet appeared great enough to warrant the additional complexity that would be required.

The corpus for the study was Earl Stumpff's Planet Earth (Ann Arbor, 1959), a text of 45,500 words.

The actual figures given below cannot be considered precise, for in spite of regular consultation among the three investigators there were inconsistencies. Most of the other inconsistencies were ironed out early in the study. There was, at any rate, almost complete agreement among the investigators in so far as the ranking is concerned, and the order in which the transformations are listed below may be taken as fairly reflecting their relative frequencies of occurrence in the text. (Only those transformations that occurred ten times or more are given here.)

The examples given below were not taken from Planet Earth, except where it is so indicated, but made by us so that each example might illustrate only the transformation under discussion as far as this was possible.

List of Abbreviations.

Rf = reference, i.e. source grammar.

L = Lees. C = Chaiyaratana.

GT = generalized or double-base transformation.

T = single-base transformation.

TL = total number of occurrence.

approx = approximate number of occurrences based on a partial co

Definite Pronouns.

Rf: C 71, 112; L GT40. TL: 2,060.

Examples: The boy was very young when he came to the University for the first time. He was only seventeen.

Note: The term definite pronoun is used here to mean both the personal pronouns and the demonstrative pronouns (including their use as demonstrative adjectives). L GT40 is such that it operates only within the confines of a sentence. In our count, however, the domain of the transformation was widened so as to include not only the intra-sententially determined pronouns but also the inter-sententially determined ones. To be consistent we should, perhaps, have counted the occurrences of the definite article, introduced by transformation, but we decided against the inclusion of the definite article here since we were fairly certain that the frequency of the definite article alone would outrank that of any other transformation by a big margin, thereby obscuring the significance of the pronouns with regard to frequency of occurrence. The figure given above shows, however, that the pronouns have come out the same way as the definite article was expected to. (The definite article occurs 4,822 times according to the count done

by the computer. But this does not distinguish the two kinds of the definite article, i.e. transformational and non-transformational.) At the present moment no complete transformational account of the pronouns seems to be available.

See Beverly Robbins, The Transformational Status of The Definite Article in English. (University of Pennsylvania Transformations and Discourse Analysis Projects Report No. 38.) 1961-62.

Modifier Possessives.

Rf: C 76; L GT27. TL: 1,314.

Examples: John's house is very large. Mr. Smith is the chairman of the board.

Note: This transformation includes every of-phrase that can be paraphrased in the form "X has Y", ("John has a house", "The board has a chairman"). We are well aware that not all the instances of have thus arrived at behave in the same way, but none of the transformational grammars available to us at present are detailed enough to enable us to make the distinctions consistently. In spite of its very high rank in the frequency count, this transformation was not chosen to be programmed, since, owing to the complexity of the transformations that result in of-phrases, the recognition procedure for this is not feasible at this stage of our work.

Conjunctives.

Rf: C 117; L GT35. TL: 1,142.

Examples: John went to the post office and Mary went to the station. John and Mary went to the airport.

Note: Under the conjunctive transformation we counted some instances of but, or, nor, as well as, and not only ... but (also). Similarly, since in a sequence of three or more conjoined items (x and y and z) all but the last conjunction may be replaced by commas (x,y and z), we also counted such commas as instances of conjunctive transformation. The difference between the two and's in the examples given above is due to an ellipsis which is almost automatic; therefore we did not bother to count them separately.

Postnominal Modifiers (derived by deletion of the relative pronoun and the copula).

Rf: C 17; L GT25. TL: 1,076.

Examples: The book on the desk is for Mary. John has all the tools necessary for the work.

Note: Just as we speak of the distinction between the restrictive and the non-restrictive relative clause, so we may make a distinction between a restrictive and a non-restrictive use of postnominal modifiers. But in our count no such distinction has been made. Although it might be argued that in "Betrayed by all his friends, Dr. Ward committed suicide" the past participial phrase

modifies "Dr. Ward" just as it does in "Dr. Ward, betrayed by all his friends, committed suicide", we assigned the former to a different transformation while counting the latter as a postnominal modifier. This difficulty arises only when we are dealing with a construction derived from a non-restrictive relative clause.

Nominalizations.

TL: 1,176.

---Action Nominalization

Rf: L GT9, 10. TL: 570.

Examples: John's apportionment of the fund was fair.

The pupils discussed the revolution of the earth around the sun. Flying planes is dangerous.

Note: It is possible to subdivide nominalizations according to the sub-class of the subject of the verb that is nominalized and/or according to the sub-class of verbs, e.g. transitive or intransitive. The distinction between nominalizations and nouns can at times be difficult to make, especially when there is no of-phrase following a nominalization or no determiner preceding it. It was suggested by John Lyons that the co-occurrence of the indefinite article or a plural ending disqualifies a given form as a nominalization. This is a point well made, but an example like the following makes us think twice: Another attempt to climb the mountain will certainly enrage the natives. Our problem here is: which

is the more economical way of handling this sort of thing in view of the entire grammar—to derive it from the verb attempt or to set up a subclass of nouns that take infinitives (and/or prepositions) as complements? As yet we do not know the answer to this question; further research is needed in this respect.

---Factive "that"-Clause.

Rf: C 30; L GT1,2. TL: 375.

Examples: John knows that the story is false. The fact that the story is false surprised Mary. That the story is false is obvious.

---Infinitive---Purpose.

Rf: C 15; L GT15,27. TL: 47.

John went to Indianapolis to buy a car.

--- ---Subject.

Rf: L GT12,13. TL: 44.

To play the violin before breakfast is a mistake.

The first step is to collect data.

Note: Infinitives after "be" in equational sentences are counted here.

---Deadjectival Nominalization.

Rf: L GT18, 19 TL: 139.

Example: The accuracy of the clock always surprises John.

Note: We encounter the same sort of difficulty here that we do in other nominalizations.

Passive.

Rf: C 14; L 74. TL: 906.

The ball was kicked by the boy.

Note: The main difficulty we have here lies in such forms (ending in -ed) as tired, annoyed, limited which can be interpreted either as past participles, or as adjectives by virtue of the fact that they can be modified by very. It is true that context often helps us to decide, but, if a given sentence is the product of a series of transformations involving deletions, the decision can be very difficult.

Negatives and Indefinites.

Rf: C 128-130; L T14. TL: 303.

Examples: John does not know anything about Eskimo.
John hardly knows anything about Eskimo. John knows scarcely anything about Eskimo. John knows nothing about Eskimo.

Relative Clause. TL: 519.

---Restrictive.

Rf: C 69-70. TL: 303.

Examples: The man who threw the stone should pay for the damage. John eats his breakfast where everybody else does.

Note: Chaiyaratana does not distinguish restrictive from non-restrictive relative clauses. Lees tries to distinguish the restrictive from the non-restrictive by generating the former in prenominal position while generating

the latter in post-nominal position. But this does not account for the ambiguity found in 'his French wife' which can mean either 'he has more than one wife and one of them is French' (restrictive) or 'he has only one wife and she is French' (non-restrictive). It still remains to be investigated how best we can describe the difference in our grammar.

---Non-Restrictive.

Rf: L GT22. TL: 216.

Example=: Bertrand Russell, who is ninety-one years old now, saved from obscurity Gottlob Frege, who wrote The Foundations of Arithmetic.

Complements. TL: 257.

---of Verb---Infinitive.

Rf: L GT14,29. TL: 174.

Examples: John hates for a cab to make a U-turn. John persuaded Mary to go home. John expects to receive a gift from Mary at Christmas.

Note: L GT1 is labelled as "Infinitival Object". This is because the rule is so formulated that it effects replacement of a factive noun phrase in the matrix sentence by an infinitive phrase. It is not feasible to keep GT14 and GT29 apart consistently without a more extensive lexicon. Also the distinction between Intransitive Complement Verbs and Transitive Complement Verbs is to some extent arbitrary, e.g. expect is in both groups.

--- ---Adjective.

Rf: 29.

Examples: John painted the house green. The cider turned sour.

--- ---oun.

Rf: L GT28. TL: 15.

Examples: Mary called John a liar. The people elected John president.

--- ---Present Participle or Gerund

Rf: L GT29. TL: 10.

The teacher caught John cheating. John refrained from smoking for three days. John avoids meeting people these days. The football game prevented John from going home for the weekend.

--- ---Predicative Phrase

Rf: L GT28. TL: 10.

Example: Mary noticed John in the White House.

--- ---of Adjective.

Rf: L GT17. TL: 15.

Examples: John is sure to receive a gift from Mary at Christmas. John is apt to forget appointments. It was nice of John to come to say goodbye to Mary.

Prenominal Modifiers (in the Form of VERB[^]ing or VERB[^]ed).

Rf: L GT24. TL: 189.

John saw a weeping mother on the street. John nursed his bruised elbow.

Note: The reason for our interest in this type of pre-nominal modifier is that neither VERL^{ing} nor VEW^{ed} is in the lexicon, i.e. these forms occur only in transforms.

Adverbial Clause of Time.

Rf: C 119. TL: 153.

Example: When it rained, John was unhappy.

Note: Here is included also the temporal conjunctive use of since, the moment, every time, by the time, and the like. (Although this transformation is not listed in Lees' grammar, it is easily derivable from the Time Adverbial in the Phrase Structure. On the other hand, Chaiyaratana treats this as a conjoining transformation.)

Concessive Clauses.

Rf: C 120. TL: 134.

Examples: Although it was raining hard, John went on a picnic. Whereas Mary is beautiful and tender, John is tough and brave.

Note: Other conjunctions used in this transformation are though, even if, albeit, and sometimes while. This may be regarded as an embedding transformation which is effected by replacing the "casual adverbial" in the matrix sentence, since the adverbial clause of reason and the concessive clause are complementary to each other, i.e. a "concession" is a "cause" which is unexpectedly insufficient.

Conditional Adverbial Clauses.

Rf: C 119. TL: 145.

Example: John will be angry if Mary is late.

Note: Pseudo-conjunctives like provided, providing, and in case are also counted under this head. We regard this as a conjoining transformation rather than an embedding one.

Adverbial Clauses of Reason.

TL: 121.

Example: John was absent from the class yesterday because he had a cold.

Note: This transformation may be initiated by choosing a "causal adverbial", if there is such a choice, in the Phrase Structure.

Adverbial Gerundive Construction.

Rf: C 17,21,121. TL: 116.

Examples: On seeing a police car in the rearview mirror, John slowed down. Betrayed by all of his friends, Dr. Ward committed suicide.

Note: The preposition preceding the gerund is optional according to the reference given above. Also the gerundive phrase may stand after the main clause. The second example is interpreted as an ellipsis from "On being betrayed by all of his friends, Dr. Ward committed suicide." The negation of this phrase by means of "without" is noted in C 21 & 129.

Inversion with "there".

Rf: C 65. TL: 73.

Examples: There is a book on the desk. There's a bear coming toward me. There were three men killed here yesterday.

Note: The noun that follows the verb, which may be any one of be, some, occur, stand, be, or any verb in the "progressive" or "passive", must be an indefinite expression, i.e. without the definite article or any demonstrative adjective.

"It"-Inversion.

Rf: L GT30. TL: 61.

Examples: It is obvious that the story is false. It is hard to find a job in this town.

Indefinite Pronouns.

Rf: C 47,96. TL: 38.

Example: Someone must have taken out the book.

Postnominal Modifier with "with".

Rf: C 71; L GT26. TL: 34.

Example: John met a girl with blue eyes.

Question Word Object Clause.

Rf: C 138,140; L GT4. TL: 31.

John knows why Mary has a headache in the morning.

Imperatives.

Rf: C 2. TL: 21.

Example: Consider the distance between the earth and Mars.

Cleft Sentence.

Rf: GT35. TL: 16.

Example: It was last Saturday that John went to Indianapolis.

Reflexive Pronouns.

Rf: C 161-3; L GT37. TL: 15.

Example: John contradicted himself.

Note: For a detailed discussion of this transformation and related ones, see R. B. Lees and E. S. Klima, "Rules for English Pronominalization," Language XXXIX (1963), 17-28.

Direct WH-Questions.

Rf: C 138; L T4. TL: 12.

Examples: Where did John go last Saturday ? What did John do last Saturday ? etc.

The following are some of the most frequently occurring transformations that were added to our pool of transformations during the course of the count, and for which only rough estimates are given on the basis of partial count.

Again in descending order of frequency.

Comparative

Example: John eats as much as Mary. John eats more than Mary. John eat the rest of a more in our family.
approx: 100.

Note: See J. L. Lee, "Grammatical Analysis of the English Comparative Construction," Lang. VIII (1961), 191-196.

Demonstrative Pronominalization.

approx: 10

Example: The top of this table is like that of any other.

Note: The modifier after the pronoun (i.e. "that" in the example) may be a phrase with a preposition other than "of", a past participle, or a relative clause.

Postnominal Participial Modifiers.

approx: 100.

Example: A Latin word meaning "heart" has a stem ending in d.

Note: This may be regarded as a mere variant of the post-nominal modifier (derived by deletion of the relative pronoun and the copula) or the grounds that, although the source relative clause is not in the "progressive" form, this is due to the nature of the verb mean, which does not admit of this construction; therefore, the variation is automatic. However, there are more general considerations which make us prefer not to group the two constructions together.

Absolute Participial Construction.

approx: 66.

Example: Along the coasts this theoretical figure is greatly exceeded, excessively high water being caused by damming.

---Planet Earth.

Ellipsis after "when", "if", "while", "(al)though".

Rf: C 121 (in part). approx: 65.

Example: The whale, though not a fish, lives in water.

Causal Phrases.

approx: 60.

Example: Owing to John's late arrival, the conference was postponed until next Monday.

Note: Other prepositions used in this construction are because of, due to, and on account of.

Clause of Result.

approx: 50.

Example: The box was so heavy that John could not lift it.

Note: "so...that..." may be regarded as an equivalent of the adverbial modifier of the adjective (or adverb). We may include here such sentences as: John was kind enough to escort Mary home; John was too weary to go out.

Clause of Manner (Headed by "as", "in the (same) ^{way} manner }
that").

approx: 50.

Example: Joan does to others as he would like them to
do to him.

Emphatic Pronouns.

Rf: C 65,113. approx: 40.

Example: John saw the manager himself. Mary saw the
manager herself.

APPENDIX IV

A list of transformations

The following is a list of all the transformations in English (except morphophonemics) that are available at present. Our sources of information are:

Chalao Chaiyaratana, A Comparative Study of Thai and English Syntax. (Ph.D. Dissertation at Indiana University, 1961. printed in offset)

Noam Chomsky, Syntactic Structures. (The Hague, 1957)

" , "A Transformational Approach to Syntax" in The Third Texas Conference (Austin, 1962)

Charles J. Fillmore, "The Position of Embedding Transformations in a Grammar," Word XXIX (1963), 208-231.

Robert B. Lees, The Grammar of English Nominalizations. (Baltimore, 1962)

" , "Grammatical Analysis of the English Comparative Construction," Word XV I (1961), 171-185

" , A Transformational Generative Grammar of English. (June, 1962, mimeographed)

Robert B. Lees and E. S. Klima, "Rules for English Pronominalization," Language XXXIX (1963), 17-28.

Robert P. Stockwell and Paul Schachter, Rules for a Segment of English Syntax. (March, 1962, mimeographed)

Of these only Chaiyaratana, Lees (1961, 1962), Lees and Klima, and Stockwell and Schachter are referred to in the list, since Chomsky's works, the second of which contains twenty-three transformations (excepting morphophonemics), are programmatic and neither of them

attempts to exhaust the transformations in English; and all the transformations are found in Lees (1960) or Lees (1961) in one form or another. Lees (1961) is an extensive revision of Lees (1960). There is nothing new in Fillmore as far as the range of transformations is concerned.

Abbreviations and symbols used:

C = Chomskianism

L = Lees (1960)

L (1961) = Lees (1961)

L = Lees and allies

S = Stockwell and Schachter

1 = Simple or single-base transformation

GP = Generalized or double-base transformation

< > : enclose optional elements.

A (< B): A is derived from B.

The list that follows is an alphabetical one, with regard to the key word in the name of each transformation.

1.1. Adverbial Clause---of Reason. S GT1b.

John was absent from the class because he had a cold.

1.2. ---of Time. C 11; S GT21a.

John was happy when Mary sang in public.

1.3. ---of Condition. C 11;.

John will be happy if Mary sings in public.

1.4. ---of Concession. S. GT1c.

John is happy although Mary sings only in private.

1.5. ---Participial/Germative. C 17, 20, 221.

<On> seeing a police car in the rear-view mirror, John slowed down.

Snubbed by the critic, Mary now sings only in private.

2.1. Shift of Adverbial---xb. C 14.

(Then Joan went to the bar. $\Rightarrow$) John went to the bar then.

xb = of course <then>; strangely enough; with the rabbi's help;
through the side of his pig; immediately; suddenly, etc.

2.2. ---of Frequentatives and Sentence Adverbials. S 10.

Frequentatives = almost; probably; always; seldom; frequently; etc.

Note: 2.1 and 2.2 overlap to some extent.

3. Cleft Sentence. L GT 5.

It was John who went to Indianapolis yesterday.

It was to Indianapolis that John went yesterday.

It was John whose boat the troopers recovered from the cave.

4. Comparative. L (1,11).

John is as tall as Bill.

George is taller than John.

George is the tallest of all.

5.1 Complement of Transitive Verb. L GT 8; S GT1b,c.

Mary considers John handsome.

The people elected John president.

Mary noticed John in the White House.

5.2. ---"to"-Infinitive. S GT1a.

Mary knows John to be a teacher.

5.3. ---Bare Infinitive. S GT3g.

Mary made John wash her car.

5.4. ---"to"-Infinitive, "-ing", or Preposition "-ing". C 37; L GT2g
(in part Intransitive); S GT2d,e,f.

John imagined Mary singing in public.

Mary avoids singing in public.

John persuaded Mary to sing in public.

John prevented Mary from singing in public.

Mary refrains from singing in public.

5.1. Complement of Transitive Verb---Indicative. L 112; 101.

Bugs remain frozen at this time of year.

John stayed in Washington for six months.

John feels a tooth.

John seems very healthy.

John leaves for London for a year.

John became a teacher.

John became famous in the world.

5.2. ---Participle. S GT11, 3.

Mary seems annoyed.

The wine tasted disappointing.

Note: Both "annoyed" and "disappointing" are probably adjectives.

6. Complement of Adjective. S GT12, 3, c. L 114, 17.

Mary is apt to sing in private.

John is happy that Mary is singing in public.

John is happy for Mary to be singing in public.

7. Conjunctive. C 117; L GT13; S GT22, 23.

John and Mary went to the airport.

John went to the airport and Mary went to the store.

8.1. Deletion---Of Object. C 118; S T1.

John drank all night.

4.2 --- Of Object in Agentive Nominal. C 674.

(a lover of J $\Rightarrow$) a lover.

(a searcher for J $\Rightarrow$) a searcher.

10. Direct Discourse. C 113.

John said <as follows>, "Mary is a liar."

"Mary is a liar," John said <to Bill>.

"Mary is a liar," said John.

11. Double Determiner. C 82.

That house of John's is green.

12.1. Double Object. C 23

(John gave N $\rightarrow$) John gave Mary an apple.

12.2. Transposition of Double Object. C 25;

John gave an apple to Mary.

12.3. Prepositional Double Object. C 43.

Mary asked <John> for advice.

John begged {his parents for money.
money from his parents.

13.1. Ellipsis---with "neither", "nor", or "either". L T12,13.

(Mary did not go to church last Sunday.) Neither/nor did John.

John didn't, either.

13.2. ---with "so" or "too". L T10,11.

(John goes to church every Sunday.) So does Mary. Mary does, too.

13.3. --- ---after Conjunction. C 139; L T11.

John goes to church every Sunday and so does Mary.

13.4. ---of Active Clause. L 117.

(John says that the story is completely false $\implies$) John says so.

13.5. ---of Verb. C 119; S 114, 5.

(Who went to church last Sunday?) John did.

13.6. ---after "if", "when", etc. C 121.

When bitten by the dog, John screamed.

13.7. ---after "why", "of course", "if not", "if so". C 119.

Why <not>?

Of course <not>.

If not, John is a liar.

If so, John is a liar.

14.1. Emphasis. C 66, 121.

$x + y + z \implies x + y^{\wedge} + z$

14.2. Emphatic (Affirmative). C 66; S 114.

John saw Mary.

John did see Mary.

14.3. Shift of Emphasized IP. C 67.

A coward John is.

This bear John killed.

14.4. Emphatic "self". C 68.

John himself built the house.

14.5. Shift of Emphatic "self". C 113.

John built the house himself.

15. Embedded Exclamation. C 120.

Look how beautiful the rainbow is.

16. "have". C 125.

(Does John have two houses? $\Rightarrow$) Has John $\langle$ got $\rangle$ two houses?

17. Negative Variant of Hortative. C 5.

(Don't let's eat this apple $\Rightarrow$) Let's not eat this apple.

18. Imperative. C 2: 3 T7a,b.

Go away!

19.1. Indirect Discourse. C 30.

Mary said that she was grateful for the help.

19.2. ---Telescoped. C 33,34.

John shouted to stop.

John shouted to Mary to stop.

20.1. Inversion---with "it". L GT30,33, T6: S GT11, T9.

It is obvious that the story is false.

It is not known who ate the last sandwich.

It was a mistake for John to have eaten the last sandwich.

It is no use crying over spilt milk.

It was wise of Mary to leave early.

20.2. ---with "there". C 65.

There is a book on the desk.

There is a bear coming toward us.

There were three men killed in the accident.

20.3. ---of Object of Infinitive. L GT32.

Arsenic is illegal for an ordinary person to buy.

A gun is dangerous for children to play with.

Note: It is the source sentence that is obtained by double-base transformation.

1. Locative --- Locative. L 10.

John told us about his friend's (X) and visited him in the city.

2. Negative --- Indefinites. L 11 - 12, 13.

John does not know anything about sailing, etc.

3. Indefinites --- Recessive. L 13.1; 13.2 - 14.

John's sincerely loved Mary.

The eloquence of the boy was impressive.

John's cleverness in using the trick antagonized Mary.

John's eagerness to go to the moon surprised Mary.

John's certainty of the facts is impressive.

4.1. --- Passive --- Indefinite Subject in Genitive --- Action. L 15.1, 16.1; 16.2; 16.3; 16.4; 16.5; 16.6; 16.7; 16.8; 16.9; 16.10; 16.11; 16.12; 16.13; 16.14; 16.15; 16.16; 16.17, 18.

John's apportionment of the funds was fair.

John's apportioning of the funds was fair.

5.1. --- Indefinite Subject with "is". L 17.1, 18.

The pupils discussed the revolution of the earth around the sun.

5.2. --- with Definite Determiner. L 17.1, 18.

The apportionment of the funds was fair.

5.3. --- with Indefinite Determiner. L 17.1, 18.

Any expropriation of the funds is illegal.

6.1. --- Agentive. L 17.1; 18.

A singer of folk songs appeared at the rally.

A searcher for truth always encounters many difficulties.

7.1. --- Action Gerund --- with Subject in Genitive. L 17.1; 18.

John's playing the flute pleased Mary.

Note: This is probably rare; generally with "of" like 23.7.

23.1. --- --- Without Subject. S GT14.

Flying planes is dangerous.

23.2. --- --- With Definite Determiner (<Gerund with Subject>).

L GT9(1); S GT14,15.

The flying of the plane was dangerous.

23.10. --- --- (<Subjectless Gerund>). S GT15.

The flying of planes is dangerous.

23.11. --- --- Factive Gerund---With Subject in Genitive. S GT17,18.

Mary learned about John's flying of the plane.

23.12. --- --- Manner Gerund. S GT16.

The playing of the flute was marvelous.

23.13. --- Factive Clause---Subject. L GT1; S GT6,7.

That the story is completely false is obvious.

23.14. --- --- Object. C 30; L GT3; S GT7.

John knows that the story is completely false.

23.15. --- --- Appositive. S GT14.

John adduces the fact that the story is true.

Note: Except where the nominalization is governed by a preposition,

23.15 is a mere variant of 23.13 or 23.14.

23.16. --- Manner Clause with "way"---Subject. S GT15a.

The way John works is amazing.

23.17. --- --- Object. S GT15b.

Everybody knows the way John works.

23.18. --- WH-Clause---Subject. L GT3; S GT7,8.

23.17. --- Object. C 133, 140; L 174; S 179.

Mary knows how John went to Indianapolis.

23.20. --- WH "to"-Infinitive---Subject. L 175; S 178, 2.

Where to go is obvious.

23.21. --- Object. L 176; S 178, 2.

John knows where to go.

23.22. --- Embedded Question. C 56, 59, 60.

Mary asked if John went to Indianapolis yesterday.

Mary asked how John went to Indianapolis yesterday.

23.23. --- Infinitival Subject---with "for" NP. L 1713; S 1711.

For John to play the violin is unimaginable.

23.24. --- without "for" NP. S 1712.

23.25. --- Infinitival Object. L 1714.

John motioned for the cab to pull up to the curb.

John hates for a cab to make a U-turn.

John paid for the cab to take Mary home.

John likes to play bridge.

note: the last example is probably derived from *John likes himself to play bridge.

23.26. --- Infinitive of Purpose---after "be". L 1713; S 1712.

This car is for John to drive.

23.27. --- with "for" NP. L 1715; S 1713.

John bought the car for Mary to drive.

23.28. --- without "for" NP. C 15; L 127; S 1713.

John went to Indianapolis to buy a car.

24.1. "Once" with be[^]ed, have[^]ed. C 134.

Lawyers once were children.

Lawyers were once children.

24.2. ---with [^]VS ed. C 135.

John once killed a bear.

Once John killed a bear.

25. Shift of "other" and Numeral. C 4.

(John bought many other houses. $\Rightarrow$) John bought $\left\{ \begin{array}{l} \text{the other} \\ \text{another} \end{array} \right\}$ three houses.

26. Displaced Subject of Factive Clause. L GT31.

John is said to be a very competent teacher.

27. Parenthetical Insert. C 148.

The story is completely false, John said.

28. Partitive. C 66.

Both of John's houses are green.

29. Passive. C 14; L T5; S T1.

Bill was bitten by a snake last summer.

Fido got run over by a car.

30. Postnominal Modifier (with Relative and Copula Deleted). C 17;
L GT25; S GT6.

The car in the garage belongs to John.

31. Postnominal Modifier with "with". C 71; L GT26; S T7,c.

A girl with golden hair met John at the airport.

32. Possessive Modifier. C 76; L GT27.

John's house is green.

The top of this table is smooth.

33. Prenominal Adjective String. C 93; S GT5,6.

John bought a big wooden box at the fair.

34. Prenominal Modifier---V^{ing}, V^{ed}. L GT24; S GT18.

A burning city fascinates John.

The discouraged student disappeared from the class.

35.1. Pronoun---Definite. C 112; L GT40; LK; S GT22b, T5.

(John bought a car and) Mary drove it home.

35.2. ---Indefinite---Animate. C 96.

Someone stole the car.

35.3. --- ---Inanimate. C 97.

John picked up something on his way home.

35.4. ---Reflexive. C 161-3; L GT37; LK; S T6.

John contradicted himself in the end.

35.5. ---Reciprocal. L GT36, 39; LK.

John and Bill called each other names.

36. Prop-Word "one". C 110.

John bought a green house and Mary bought a white one.

37.1. Question ---WH C 138; L T4, 18; S T1, 2,4.

What did John buy at the fair? etc.

37.2. ---"Yes-No" (and Alternative). C 138, L T18; S T1.

Did John go to Indianapolis (or Louisville) yesterday?

38. Tag Question (or Statement). C 142-3; S T7.

John went home, didn't he?

John can't eat today, can he?

John went home, John did.

John can't eat today, John can't.

NOTE: The last two constructions are non-literary.

39.1. ---Relative Clause---Restrictive---Modifier. C 69,70; L GT21; S GT4,5.

The man who threw the stone should pay for the damage.

39.2. ---Property. C 69,70; L GT20.

A man who is blind may apply for a seeing-eye dog.

39.3. ---Non-restrictive. C 69,70; L GT22; S GT3,4.

Bertrand Russell, who is ninety-one years old now, saved from obscurity
Gottlob Frege, who wrote The Foundations of Arithmetic.

39.4. ---Reduced. L GT23.

Whatever is is wrong.

40. Singular Countable Noun. C 93.

(John bought the houses $\Rightarrow$) John bought the house.

Note: This is an expedient within the framework of Charyaratana's grammar.

41. Shift of Vocative. C 7.

(I don't eat apples, young man. $\Rightarrow$) Young man, I don't eat apples.